Integrated Science, Life and Earth Science

**Compiled by Bernard Majdi
South Georgia State College**

taken by Pamela Majdi

Preface

In the state of Georgia all early childhood education majors are required to take four science classes. Two of these classes that fit in field of study area are Integrated science: life and earth science and integrated science: physical science. All students have to take GACE II exam for their teacher certification after completing student teaching. Student will encounter integrated science questions in this exam. Certified teachers are required to teach/cover standards of excellence for k-5 that can be found here: https://www.georgiastandards.org/Georgia-Standards/Pages/Science-K-5.aspx

This book is designed to expose the students to many of the topics that they are expected to teach. This book also utilizes links to YouTube videos to cover some topics.

This book has combined materials from five different OERs. These include;

Physical Geography and Natural Disasters by R. Adam Dastrup, MA, GISP is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0

OpenStax College, *Concepts of Biology*. OpenStax College. 25 April 2013. <http://openstaxcollege.org/textbooks/concepts-of-biology/get>

Physical Geology by Steven Earle used under a CC-BY 4.0 international license.

https://openstax.org/details/books/biology-2e

Principles of Earth Science Copyright © 2021 by Katharine Solada and K. Sean Daniels is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0

Some instructors consider astronomy part of earth science others consider it as a part of physics. For our purpose, we will let it be covered along with physics.

It is our hope that this OER book will reduce the cost of textbooks for early childhood education majors.

Table of contents
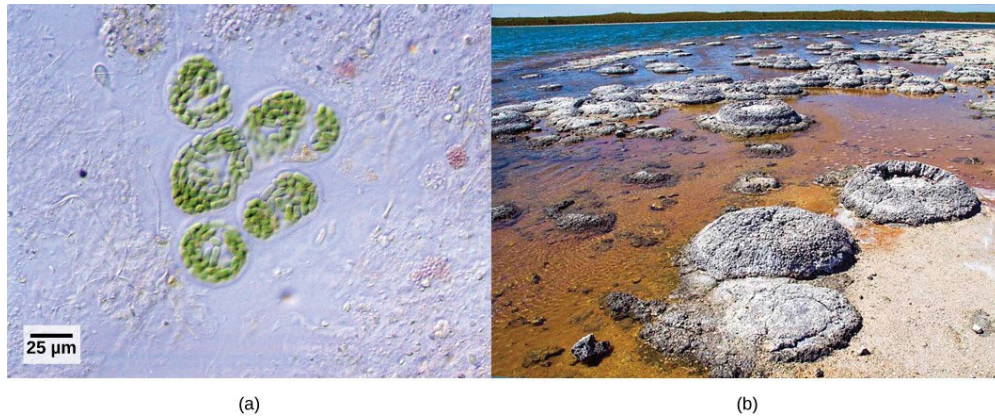
Chapter 1

The Scientific Method

The Process of Science

By the end of this section, you will be able to:

- Identify the shared characteristics of the natural sciences
- Understand the process of scientific inquiry
- Compare inductive reasoning with deductive reasoning
- Describe the goals of basic science and applied science

Figure 1.14 Formerly called blue-green algae, the (a) cyanobacteria seen through a light microscope are some of Earth's oldest life forms. These (b) stromatolites along the shores of Lake Thetis in Western Australia are ancient structures formed by the layering of cyanobacteria in shallow waters. (credit a: modification of work by NASA; scale-bar data from Matt Russell; credit b: modification of work by Ruth Ellison)

Like geology, physics, and chemistry, biology is a science that gathers knowledge about the natural world. Specifically, biology is the study of life. The discoveries of biology are made by a community of researchers who work individually and together using agreed-on methods. In this sense, biology, like all sciences is a social enterprise like politics or the arts. The methods of science include careful observation, record keeping, logical and mathematical reasoning, experimentation, and submitting conclusions to the scrutiny of others. Science also requires considerable imagination and creativity; a well-designed experiment is commonly described as elegant, or beautiful. Like politics, science has considerable practical implications and some science is dedicated to practical applications, such as the prevention of disease (see **Figure 1.15**). Other science proceeds largely motivated by curiosity. Whatever its goal, there is no doubt that science, including biology, has transformed human existence and will continue to do so.
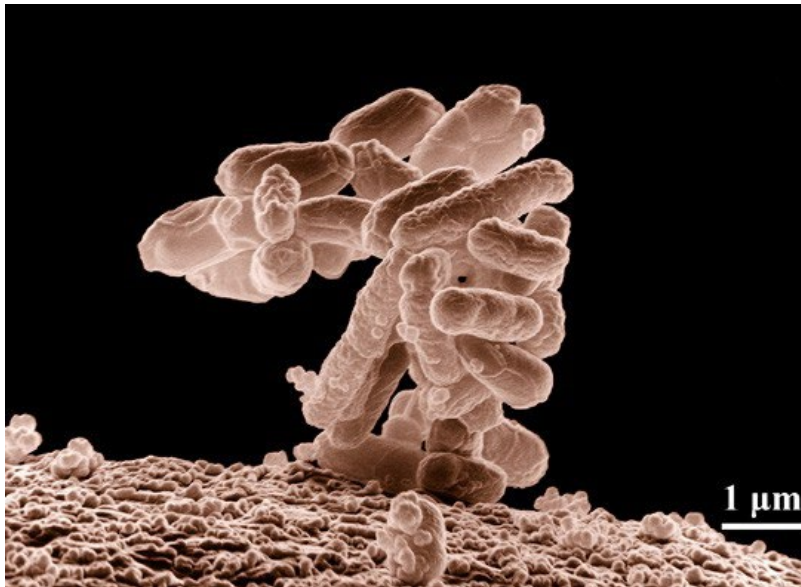
Figure 1.15 Biologists may choose to study *Escherichia coli* (*E. coli*), a bacterium that is a normal resident of our digestive tracts but which is also sometimes responsible for disease outbreaks. In this micrograph, the bacterium is visualized using a scanning electron microscope and digital colorization. (credit: Eric Erbe; digital colorization by Christopher Pooley, USDA-ARS)

The Nature of Science

Biology is a science, but what exactly is science? What does the study of biology share with other scientific disciplines? **Science** (from the Latin *scientia*, meaning "knowledge") can be defined as knowledge about the natural world.

Science is a very specific way of learning, or knowing, about the world. The history of the past 500 years demonstrates that science is a very powerful way of knowing about the world; it is largely responsible for the technological revolutions that have taken place during this time. There are however, areas of

knowledge and human experience that the methods of science cannot be applied to. These include such things as answering purely moral questions, aesthetic questions, or what can be generally categorized as spiritual questions. Science has cannot investigate these areas because they are outside the realm of material phenomena, the phenomena of matter and energy, and cannot be observed and measured.

The **scientific method** is a method of research with defined steps that include experiments and careful observation. The steps of the scientific method will be examined in detail later, but one of the most important aspects of this method is the testing of hypotheses. A **hypothesis** is a suggested explanation for an event, which can be tested. Hypotheses, or tentative explanations, are generally produced within the context of a **scientific theory**. A scientific theory is a generally accepted, thoroughly tested and confirmed explanation for a set of observations or phenomena. Scientific theory is the foundation of scientific knowledge. In addition, in many scientific disciplines (less so in biology) there are **scientific laws**, often expressed in mathematical formulas, which describe how elements of nature will behave under certain specific conditions. There is not an evolution of hypotheses through theories to laws as if they represented some increase in certainty about the world. Hypotheses are the day-to-day material that scientists work with and they are developed within the context of theories. Laws are concise descriptions of parts of the world that are amenable to formulaic or mathematical description.

Natural Sciences

What would you expect to see in a museum of natural sciences? Frogs? Plants? Dinosaur skeletons? Exhibits about how the brain functions? A planetarium? Gems and minerals? Or maybe all of the above? Science includes such diverse fields as astronomy, biology, computer sciences, geology, logic, physics, chemistry, and mathematics (**Figure 1.16**). However, those fields of science related to the physical world and its phenomena and processes are considered **natural sciences**. Thus, a museum of natural sciences might contain any of the items listed above.
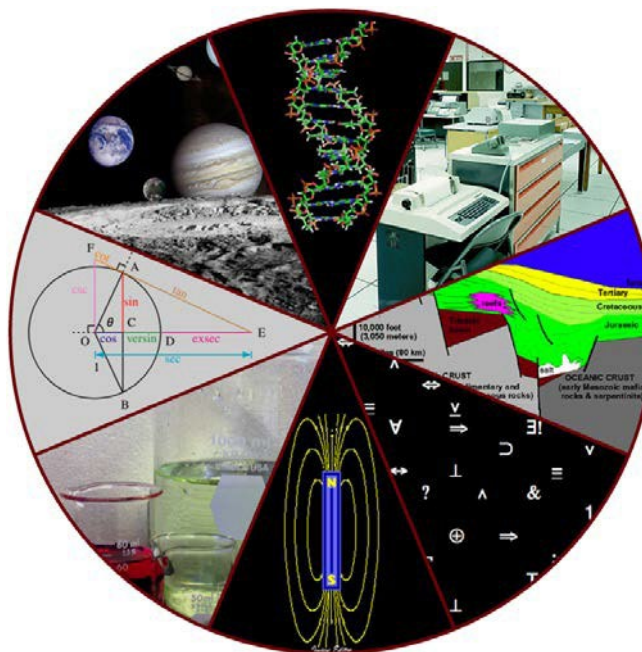
Figure 1.16 Some fields of science include astronomy, biology, computer science, geology, logic, physics, chemistry, and mathematics. (credit: "Image Editor"/Flickr)

There is no complete agreement when it comes to defining what the natural sciences include. For some experts, the natural sciences are astronomy, biology, chemistry, earth science, and physics. Other scholars choose to divide natural sciences into **life sciences**, which study living things and include biology, and **physical sciences**, which study nonliving matter and include astronomy, physics, and chemistry. Some disciplines such as biophysics and biochemistry build on two sciences and are interdisciplinary.

Scientific Inquiry

One thing is common to all forms of science: an ultimate goal "to know." Curiosity and inquiry are the driving forces for the development of science. Scientists seek to understand the world and the way it operates. Two methods of logical thinking are used: inductive reasoning and deductive reasoning.

Inductive reasoning is a form of logical thinking that uses related observations to arrive at a general conclusion. This type of reasoning is common in descriptive science. A life scientist such as a biologist

makes observations and records them. These data can be qualitative (descriptive) or quantitative

(consisting of numbers), and the raw data can be supplemented with drawings, pictures, photos, or videos. From many observations, the scientist can infer conclusions (inductions) based on evidence. Inductive reasoning involves formulating generalizations inferred from careful observation and the analysis of a large amount of data. Brain studies often work this way. Many brains are observed while people are doing a task. The part of the brain that lights up, indicating activity, is then demonstrated to be the part controlling the response to that task.

Deductive reasoning or deduction is the type of logic used in hypothesis-based science. In deductive reasoning, the pattern of thinking moves in the opposite direction as compared to inductive reasoning. **Deductive reasoning** is a form of logical thinking that uses a general principle or law to forecast specific results. From those general principles, a scientist can extrapolate and predict the specific results that would be valid as long as the general principles are valid. For example, a prediction would be that if the climate is becoming warmer in a region, the distribution of plants and animals should change. Comparisons have been made between distributions in the past and the present, and the many changes that have been found are consistent with a warming climate. Finding the change in distribution is evidence that the climate change conclusion is a valid one.

Both types of logical thinking are related to the two main pathways of scientific study: descriptive science and hypothesis-based science. **Descriptive** (or discovery) **science** aims to observe, explore, and discover, while **hypothesis-based science** begins with a specific question or problem and a potential answer or solution that can be tested. The boundary between these two forms of study is often blurred, because most scientific endeavors combine both approaches. Observations lead to questions, questions lead to forming a hypothesis as a possible answer to those questions, and then the hypothesis is tested. Thus, descriptive science and hypothesis-based science are in continuous dialogue.

Hypothesis Testing

Biologists study the living world by posing questions about it and seeking science-based responses. This approach is common to other sciences as well and is often referred to as the scientific method. The scientific method was used even in ancient times, but it was first documented by England's Sir Francis Bacon (1561–1626) (**Figure 1.17**), who set up inductive methods for scientific inquiry. The scientific method is not exclusively used by biologists but can be applied to almost anything as a logical problem-solving method.

Figure 1.17 Sir Francis Bacon is credited with being the first to document the scientific method.

The scientific process typically starts with an observation (often a problem to be solved) that leads to a question. Let's think about a simple problem that starts with an observation and apply the scientific method to solve the problem. One Monday morning, a student arrives at class and quickly discovers that the classroom is too warm. That is an observation that also describes a problem: the classroom is too warm. The student then asks a question: "Why is the classroom so warm?"

4

Recall that a hypothesis is a suggested explanation that can be tested. To solve a problem, several hypotheses may be proposed. For example, one hypothesis might be, "The classroom is warm because no one turned on the air conditioning." But there could be other responses to the question, and therefore other hypotheses may be proposed. A second hypothesis might be, "The classroom is warm because there is a power failure, and so the air conditioning doesn't work."

Once a hypothesis has been selected, a prediction may be made. A prediction is similar to a hypothesis but it typically has the format "If . . . then   " For example, the prediction for the first hypothesis might be, "*If* the student turns on the air conditioning, *then* the classroom will no longer be too warm."

A hypothesis must be testable to ensure that it is valid. For example, a hypothesis that depends on what a bear thinks is not testable, because it can never be known what a bear thinks. It should also be **falsifiable**, meaning that it can be disproven by experimental results. An example of an unfalsifiable hypothesis is "Botticelli's *Birth of Venus* is beautiful." There is no experiment that might show this statement to be false. To test a hypothesis, a researcher will conduct one or more experiments designed to eliminate one or more of the hypotheses. This is important. A hypothesis can be disproven, or eliminated, but it can never be proven. Science does not deal in proofs like mathematics. If an experiment fails to disprove a hypothesis, then we find support for that explanation, but this is not to say that down the road a better explanation will not be found, or a more carefully designed experiment will be found to falsify the hypothesis.

Each experiment will have one or more variables and one or more controls. A **variable** is any part of the experiment that can vary or change during the experiment. A **control** is a part of the experiment that does not change. Look for the variables and controls in the example that follows. As a simple example, an experiment might be conducted to test the hypothesis that phosphate limits the growth of algae in freshwater ponds. A series of artificial ponds are filled with water and half of them are treated by adding phosphate each week, while the other half are treated by adding a salt that is known not to be used by algae. The variable here is the phosphate (or lack of phosphate), the experimental or treatment cases are the ponds with added phosphate and the control ponds are those with something inert added, such as the salt. Just adding something is also a control against the possibility that adding extra matter to the pond has an effect. If the treated ponds show lesser growth of algae, then we have found support for our hypothesis. If they do not, then we reject our hypothesis. Be aware that rejecting one hypothesis does not determine whether or not the other hypotheses can be accepted; it simply eliminates one hypothesis that is not valid (**Figure 1.18**). Using the scientific method, the hypotheses that are inconsistent with experimental data are rejected.
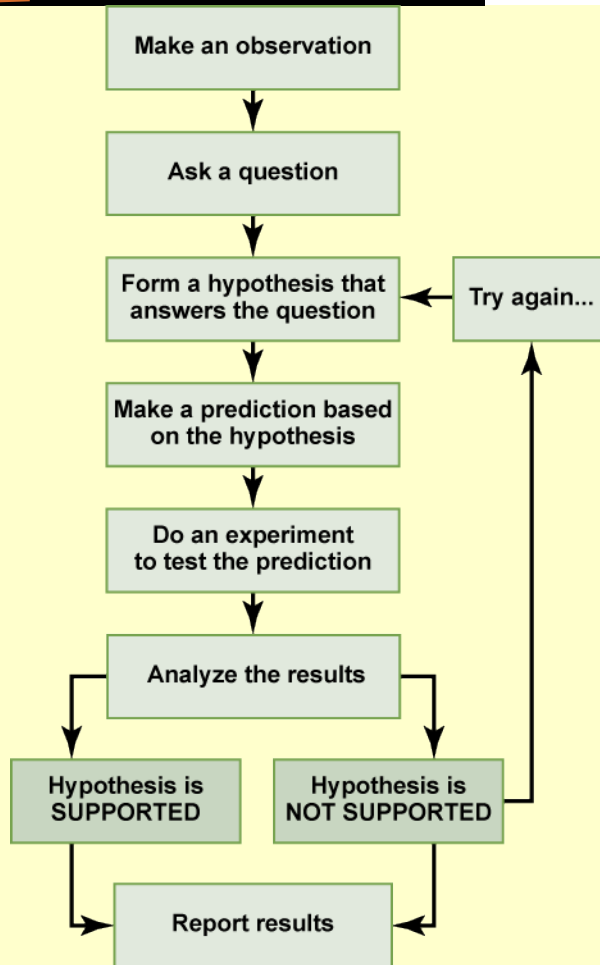
Figure 1.18 The scientific method is a series of defined steps that include experiments and careful observation. If a hypothesis is not supported by data, a new hypothesis can be proposed.

In the example below, the scientific method is used to solve an everyday problem. Which part in the example below is the hypothesis? Which is the prediction? Based on the results of the experiment, is the hypothesis supported? If it is not supported, propose some alternative hypotheses.

1. My toaster doesn't toast my bread.
2. Why doesn't my toaster work?
3. There is something wrong with the electrical outlet.
4. If something is wrong with the outlet, my coffeemaker also won't work when plugged into it.
5. I plug my coffeemaker into the outlet.
6. My coffeemaker works.

In practice, the scientific method is not as rigid and structured as it might at first appear. Sometimes an experiment leads to conclusions that favor a change in approach; often, an experiment brings entirely new scientific questions to the puzzle. Many times, science does not operate in a linear fashion; instead, scientists continually draw inferences and make generalizations, finding patterns as their research proceeds. Scientific reasoning is more complex than the scientific method alone suggests.

6

Basic and Applied Science

The scientific community has been debating for the last few decades about the value of different types of science. Is it valuable to pursue science for the sake of simply gaining knowledge, or does scientific knowledge only have worth if we can apply it to solving a specific problem or bettering our lives? This question focuses on the differences between two types of science: basic science and applied science.

Basic science or "pure" science seeks to expand knowledge regardless of the short-term application of that knowledge. It is not focused on developing a product or a service of immediate public or commercial value. The immediate goal of basic science is knowledge for knowledge's sake, though this does not mean that in the end it may not result in an application.

In contrast, **applied science** or "technology," aims to use science to solve real-world problems, making it possible, for example, to improve a crop yield, find a cure for a particular disease, or save animals threatened by a natural disaster. In applied science, the problem is usually defined for the researcher.

Some individuals may perceive applied science as "useful" and basic science as "useless." A question these people might pose to a scientist advocating knowledge acquisition would be, "What for?" A careful look at the history of science, however, reveals that basic knowledge has resulted in many remarkable applications of great value. Many scientists think that a basic understanding of science is necessary before an application is developed; therefore, applied science relies on the results generated through basic science. Other scientists think that it is time to move on from basic science and instead to find solutions to actual problems. Both approaches are valid. It is true that there are problems that demand immediate attention; however, few solutions would be found without the help of the knowledge generated through basic science.

One example of how basic and applied science can work together to solve practical problems occurred after the discovery of DNA structure led to an understanding of the molecular mechanisms governing DNA replication. Strands of DNA, unique in every human, are found in our cells, where they provide the instructions necessary for life. During DNA replication, new copies of DNA are made, shortly before a cell divides to form new cells. Understanding the mechanisms of DNA replication enabled scientists to develop laboratory techniques that are now used to identify genetic diseases, pinpoint individuals who were at a crime scene, and determine paternity. Without basic science, it is unlikely that applied science would exist.

Another example of the link between basic and applied research is the Human Genome Project, a study in which each human chromosome was analyzed and mapped to determine the precise sequence of DNA subunits and the exact location of each gene. (The gene is the basic unit of heredity; an individual's complete collection of genes is his or her genome.) Other organisms have also been studied as part of this project to gain a better understanding of human chromosomes. The Human Genome Project (**Figure 1.19**) relied on basic research carried out with non-human organisms and, later, with the human genome. An important end goal eventually became using the data for applied research seeking cures for genetically related diseases.
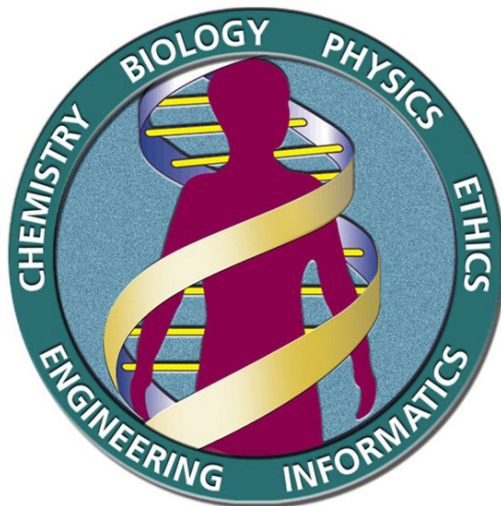
Figure 1.19 The Human Genome Project was a 13-year collaborative effort among researchers working in several different fields of science. The project was completed in 2003. (credit: the U.S. Department of Energy Genome Programs)

While research efforts in both basic science and applied science are usually carefully planned, it is important to note that some discoveries are made by serendipity, that is, by means of a fortunate accident or a lucky surprise. Penicillin was discovered when biologist Alexander Fleming accidentally left a petri dish of *Staphylococcus* bacteria open. An unwanted mold grew, killing the bacteria. The mold turned out to be *Penicillium*, and a new antibiotic was discovered. Even in the highly organized world of science, luck—when combined with an observant, curious mind—can lead to unexpected breakthroughs.

Reporting Scientific Work

Whether scientific research is basic science or applied science, scientists must share their findings for other researchers to expand and build upon their discoveries. Communication and collaboration within and between sub disciplines of science are key to the advancement of knowledge in science. For this reason, an important aspect of a scientist's work is disseminating results and communicating with peers. Scientists can share results by presenting them at a scientific meeting or conference, but this approach can reach only the limited few who are present. Instead, most scientists present their results in peer- reviewed articles that are published in scientific journals. **Peer-reviewed articles** are scientific papers that are reviewed, usually anonymously by a scientist's colleagues, or peers. These colleagues are qualified individuals, often experts in the same research area, who judge whether or not the scientist's work is suitable for publication. The process of peer review helps to ensure that the research described in a scientific paper or grant proposal is original, significant, logical, and thorough. Grant proposals, which are requests for research funding, are also subject to peer review. Scientists publish their work so other scientists can reproduce their experiments under similar or different conditions to expand on the findings. The experimental results must be consistent with the findings of other scientists.

There are many journals and the popular press that do not use a peer-review system. A large number of online open-access journals, journals with articles available without cost, are now available many of which use rigorous peer-review systems, but some of which do not. Results of any studies published in these forums without peer review are not reliable and should not form the basis for other scientific work. In one exception, journals may allow a researcher to cite a personal communication from another researcher about unpublished results with the cited author's permission.

Summary

A hypothesis is a tentative explanation for an observation. A scientific theory is a well-tested and consistently verified explanation for a set of observations or phenomena. A scientific law is a description, often in the form of a mathematical formula, of the behavior of an aspect of nature under certain circumstances. Two types of logical reasoning are used in science. Inductive reasoning uses results to produce general scientific principles. Deductive reasoning is a form of logical thinking that predicts results by applying general principles. The common thread throughout scientific research is the use of the scientific method. Scientists present their results in peer-reviewed scientific papers published in scientific journals.

Science can be basic or applied. The main goal of basic science is to expand knowledge without any expectation of short-term practical application of that knowledge. The primary goal of applied research, however, is to solve practical problems.

REVIEW QUESTIONS

1. Why doesn't my toaster work?
a. There is something wrong with the electrical outlet.
b. If something is wrong with the outlet, my coffeemaker also won't work when plugged into it.
c. I plug my coffeemaker into the outlet.
d. My coffeemaker works.

2. A suggested and testable explanation for an event is called a   .
a. hypothesis
b. variable
c. theory
d. control

3. The type of logical thinking that uses related observations to arrive at a general conclusion is called   .
a. deductive reasoning
b. the scientific method
c. hypothesis-based science
d. inductive reasoning

Chapter 2 Life, Cells and Metabolism

Properties of Life

All groups of living organisms share several key characteristics or functions: order, sensitivity or response to stimuli, reproduction, adaptation, growth and development, regulation, homeostasis, and energy processing. When viewed together, these eight characteristics serve to define life.

Order

Organisms are highly organized structures that consist of one or more cells. Even very simple, single-celled organisms are remarkably complex. Inside each cell, atoms make up molecules. These in turn make up cell components or organelles. Multicellular organisms, which may consist of millions of individual cells, have an advantage over single-celled organisms in that their cells can be specialized to perform specific functions, and even sacrificed in certain situations for the good of the organism as a whole. How these specialized cells come together to form organs such as the heart, lung, or skin in organisms like the toad shown in (Figure 2.1) will be discussed later.

Figure 2.1 A toad represents a highly organized structure consisting of cells, tissues, organs, and organ systems. (credit: "Ivengo(RUS)"/Wikimedia Commons)

Sensitivity or Response to Stimuli

Organisms respond to diverse stimuli. For example, plants can bend toward a source of light or respond to touch (Figure 2.2). Even tiny bacteria can move toward or away from chemicals (a process called chemotaxis) or light (phototaxis). Movement toward a stimulus is considered a positive response, while movement away from a stimulus is considered a negative response.

Figure 2.2 The leaves of this sensitive plant (*Mimosa pudica*) will instantly droop and fold when touched. After a few minutes, the plant returns to its normal state. (credit: Alex Lomas)

Watch this **video (http://openstaxcollege.org/l/thigmonasty)** to see how the sensitive plant responds to a touch stimulus.

Reproduction

Single-celled organisms reproduce by first duplicating their DNA, which is the genetic material, and then dividing it equally as the cell prepares to divide to form two new cells. Many multicellular organisms (those made up of more than one cell) produce specialized reproductive cells that will form new individuals. When reproduction occurs, DNA containing genes is passed along to an organism's offspring. These genes are the reason that the offspring will belong to the same species and will have characteristics similar to the parent, such as fur color and blood type.

Adaptation

All living organisms exhibit a "fit" to their environment. Biologists refer to this fit as adaptation and it is a consequence of evolution by natural selection, which operates in every lineage of reproducing organisms. Examples of adaptations are diverse and unique, from heat-resistant Archaea that live in boiling hot springs to the tongue length of a nectar-feeding moth that matches the size of the flower from which it feeds. All adaptations enhance the reproductive potential of the individual exhibiting them, including their ability to survive to reproduce. Adaptations are not constant. As an environment changes, natural selection causes the characteristics of the individuals in a population to track those changes.

Growth and Development

Organisms grow and develop according to specific instructions coded for by their genes. These genes provide instructions that will direct cellular growth and development, ensuring that a species' young (figure 2.3) will grow up to exhibit many of the same characteristics as its parents.

11

Figure 2.3 Although no two look alike, these kittens have inherited genes from both parents and share many of the same characteristics. (credit: Pieter & Renée Lanser)

Regulation

Even the smallest organisms are complex and require multiple regulatory mechanisms to coordinate internal functions, such as the transport of nutrients, response to stimuli, and coping with environmental stresses. For example, organ systems such as the digestive or circulatory systems perform specific functions like carrying oxygen throughout the body, removing wastes, delivering nutrients to every cell, and cooling the body.

Homeostasis

To function properly, cells require appropriate conditions such as proper temperature, pH, and concentrations of diverse chemicals. These conditions may, however, change from one moment to the next. Organisms are able to maintain internal conditions within a narrow range almost constantly, despite environmental changes, through a process called **homeostasis** or "steady state"—the ability of an organism to maintain constant internal conditions. For example, many organisms regulate their body temperature in a process known as thermoregulation. Organisms that live in cold climates, such as the polar bear (Figure 2.4), have body structures that help them withstand low temperatures and conserve body heat. In hot climates, organisms have methods (such as perspiration in humans or panting in dogs) that help them to shed excess body heat.

Figure 2.4 Polar bears and other mammals living in ice-covered regions maintain their body temperature by generating heat and reducing heat loss through thick fur and a dense layer of fat under their skin. (credit: "longhorndave"/Flickr)

Energy Processing

All organisms (such as the California condor shown in (Figure 2.5) use a source of energy for their metabolic activities. Some organisms capture energy from the Sun and convert it into chemical energy in food; others use chemical energy from molecules they take in.

12
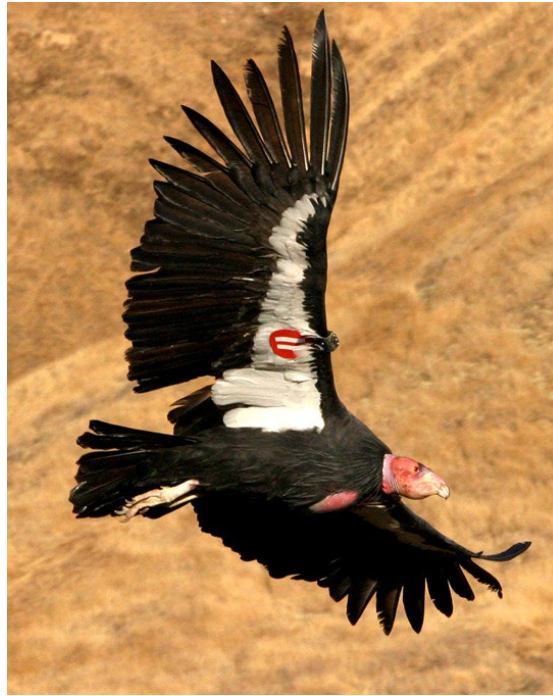
Figure 2.5 A lot of energy is required for a California condor to fly. Chemical energy derived from food is used to power flight. California condors are an endangered species; scientists have strived to place a wing tag on each bird to help them identify and locate each individual bird. (credit: Pacific Southwest Region U.S. Fish and Wildlife)

Cell Theory

The microscopes we use today are far more complex than those used in the 1600s by Antony van Leeuwenhoek, a Dutch shopkeeper who had great skill in crafting lenses. Despite the limitations of his now-ancient lenses, van Leeuwenhoek observed the movements of protists (a type of single-celled organism) and sperm, which he collectively termed "animalcules."

In a 1665 publication called *Micrographia*, experimental scientist Robert Hooke coined the term "cell" (from the Latin *cella*, meaning "small room") for the box-like structures he observed when viewing cork tissue through a lens. In the 1670s, van Leeuwenhoek discovered bacteria and protozoa. Later advances in lenses and microscope construction enabled other scientists to see different components inside cells.

By the late 1830s, botanist Matthias Schleiden and zoologist Theodor Schwann were studying tissues and proposed the **unified cell theory**, which states that all living things are composed of one or more cells, that the cell is the basic unit of life, and that all new cells arise from existing cells. These principles still stand today.

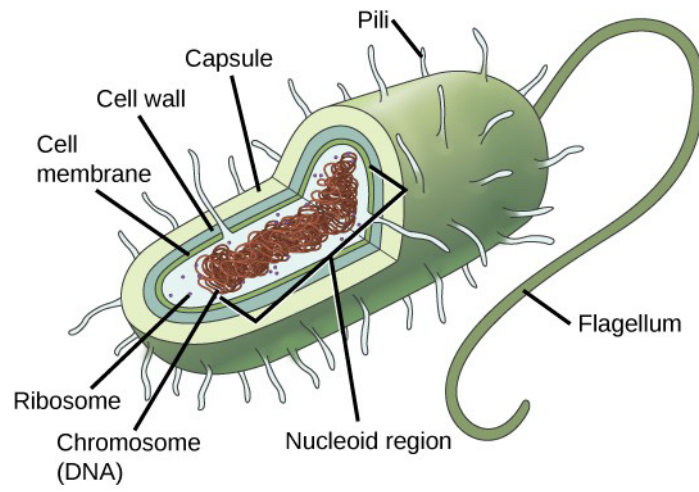## | Comparing Prokaryotic and Eukaryotic Cells

Cells fall into one of two broad categories: prokaryotic and eukaryotic. The predominantly single-celled organisms of the domains Bacteria and Archaea are classified as prokaryotes (*pro-* = before; *-karyon-* = nucleus). Animal cells, plant cells, fungi, and protists are eukaryotes (*eu-* = true).

Components of Prokaryotic Cells

All cells share four common components: 1) a plasma membrane, an outer covering that separates the cell's interior from its surrounding environment; 2) cytoplasm, consisting of a jelly-like region within the cell in which other cellular components are found; 3) DNA, the genetic material of the cell; and 4) ribosomes, particles that synthesize proteins. However, prokaryotes differ from eukaryotic cells in several ways.

A **prokaryotic cell** is a simple, single-celled (unicellular) organism that lacks a nucleus, or any other membrane-bound organelle. We will shortly come to see that this is significantly different in eukaryotes. Prokaryotic DNA is found in the central part of the cell: a darkened region called the nucleoid (**Figure 2.6**).

13

Figure 2.6 This figure shows the generalized structure of a prokaryotic cell.

Unlike Archaea and eukaryotes, bacteria have a cell wall made of peptidoglycan, comprised of sugars and amino acids, and many have a polysaccharide capsule (Figure 2.6). The cell wall acts as an extra layer of protection, helps the cell maintain its shape, and prevents dehydration. The capsule enables the cell to attach to surfaces in its environment. Some prokaryotes have flagella, pili, or fimbriae. Flagella are used for locomotion, while most pili are used to exchange genetic material during a type of reproduction called conjugation.
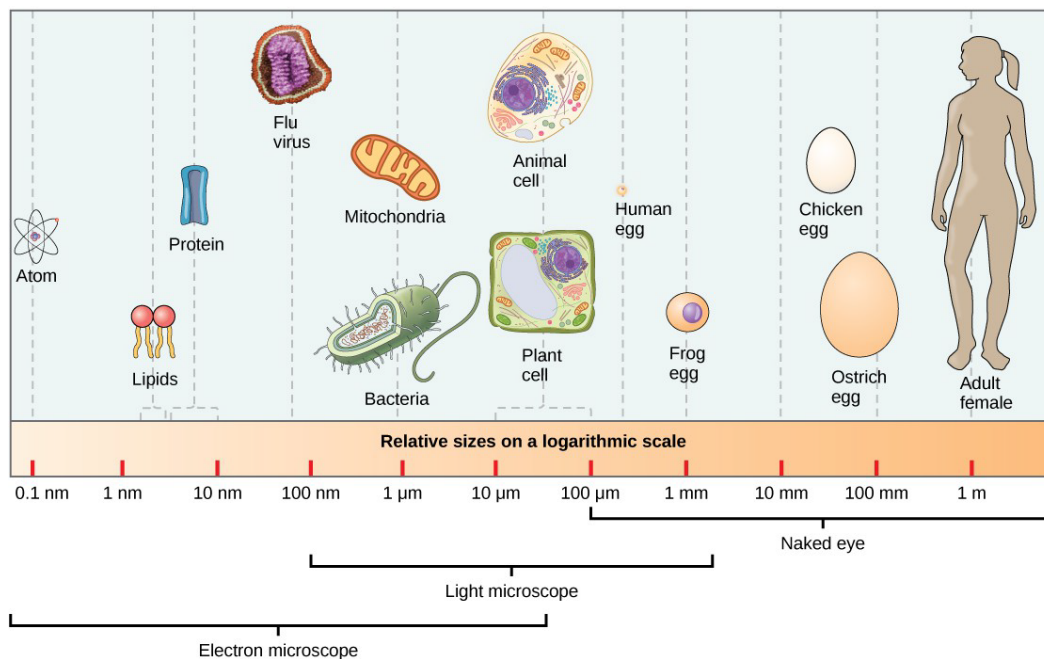
Eukaryotic Cells

In nature, the relationship between form and function is apparent at all levels, including the level of the cell, and this will become clear as we explore eukaryotic cells. The principle "form follows function" is found in many contexts. For example, birds and fish have streamlined bodies that allow them to move quickly through the medium in which they live, be it air or water. It means that, in general, one can deduce the function of a structure by looking at its form, because the two are matched.

A **eukaryotic cell** is a cell that has a membrane-bound nucleus and other membrane-bound compartments or sacs, called **organelles**, which have specialized functions. The word eukaryotic means "true kernel" or "true nucleus," alluding to the presence of the membrane-bound nucleus in these cells. The word "organelle" means "little organ," and, as already mentioned, organelles have specialized cellular functions, just as the organs of your body have specialized functions.
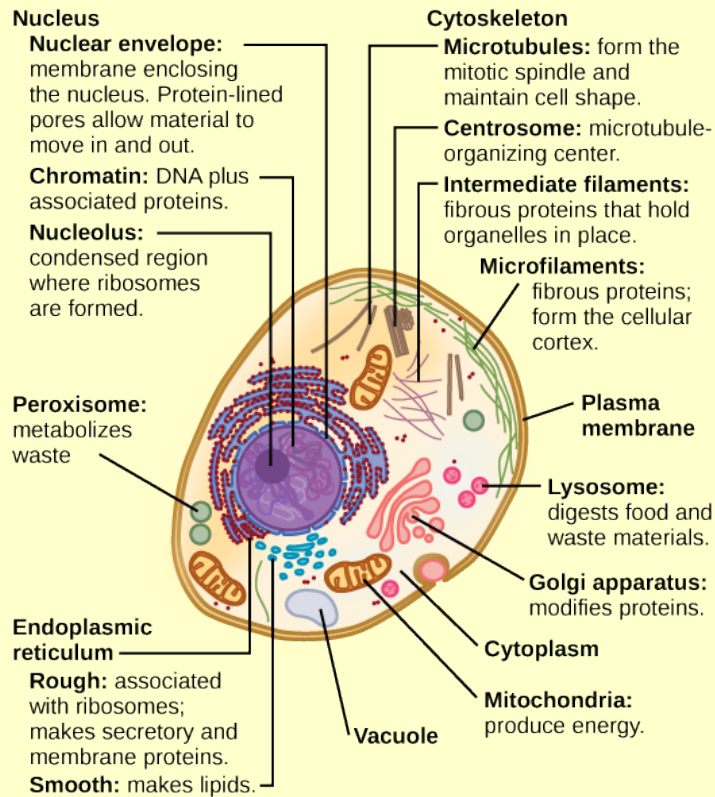
Cell Size

At 0.1–5.0 μm in diameter, prokaryotic cells are significantly smaller than eukaryotic cells, which have diameters ranging from 10–100 μm (Figure 2.7). The small size of prokaryotes allows ions and organic molecules that enter them to quickly spread to other parts of the cell. Similarly, any wastes produced within a prokaryotic cell can quickly move out. However, larger eukaryotic cells have evolved different structural adaptations to enhance cellular transport. Indeed, the large size of these cells would not be possible without these adaptations. In general, cell size is limited because volume increases much more quickly than does cell surface area. As a cell becomes larger, it becomes more and more difficult for the cell to acquire sufficient materials to support the processes inside the cell, because the relative size of the surface area across which materials must be transported declines.

Figure 2.7 This figure shows the relative sizes of different kinds of cells and cellular components. An adult human is shown for comparison.

At this point, it should be clear that eukaryotic cells have a more complex structure than do prokaryotic cells. Organelles allow for various functions to occur in the cell at the same time. Before discussing the functions of organelles within a eukaryotic cell, let us first examine two important components of the cell: the plasma membrane and the cytoplasm.

Nucleus
Nuclear envelope: membrane enclosing the nucleus. Protein-lined pores allow material to move in and out.
Chromatin: DNA plus associated proteins.
Nucleolus: condensed region where ribosomes are formed.

Peroxisome: metabolizes waste

Endoplasmic reticulum
Rough: associated with ribosomes; makes secretory and membrane proteins.
Smooth: makes lipids.

Cytoskeleton
Microtubules: form the mitotic spindle and maintain cell shape.
Centrosome: microtubule-organizing center.
Intermediate filaments: fibrous proteins that hold organelles in place.
Microfilaments: fibrous proteins; form the cellular cortex.

Plasma membrane

Lysosome: digests food and waste materials.

Golgi apparatus: modifies proteins.

Cytoplasm

Mitochondria: produce energy.

Vacuole

(a)

(a)

Plasmodesmata: channels connect two plant cells
Cell wall: maintains cell shape
Plasma membrane
Cytoplasm
Central vacuole: filled with cell sap that maintains pressure against cell wall
Cytoskeleton: microtubules intermediate filaments microfilaments

Endoplasmic reticulum smooth rough

Chloroplast: site of photosynthesis

Nucleus: contains chromatin, a nuclear envelope, and a nucleolus, as in an animal cell

Ribosomes
Golgi apparatus
Mitochondria
Peroxisome

**Plastid:** stores pigments

(b)

(b)

Figure 2.8 This figure shows (a) a typical animal cell and (b) a typical plant cell.

What structures does a plant cell have that an animal cell does not have? What structures does an animal cell have that a plant cell does not have?

The Plasma Membrane

Like prokaryotes, eukaryotic cells have a **plasma membrane** (**Figure 2.9**) made up of a phospholipid bilayer with embedded proteins that separates the internal contents of the cell from its surrounding environment. A phospholipid is a lipid molecule composed of two fatty acid chains, a glycerol backbone, and a phosphate group. The plasma membrane regulates the passage of some substances, such as organic molecules, ions, and water, preventing the passage of some to maintain internal conditions, while actively bringing in or removing others. Other compounds move passively across the membrane.
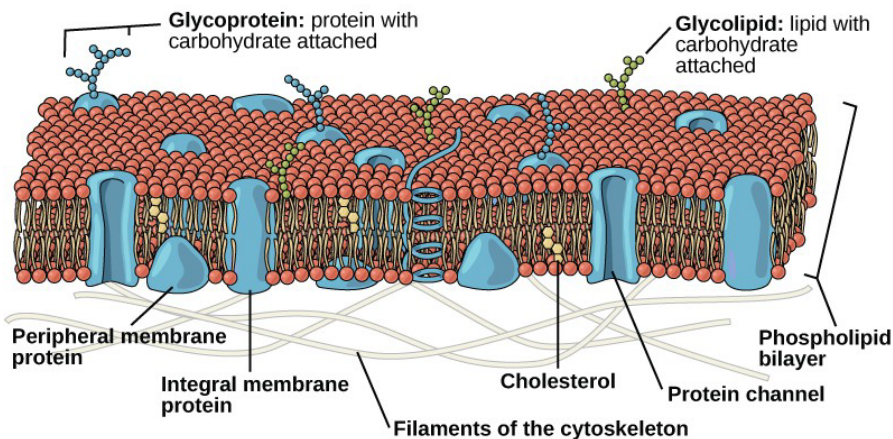
Figure 2.9 The plasma membrane is a phospholipid bilayer with embedded proteins. There are other components, such as cholesterol and carbohydrates, which can be found in the membrane in addition to phospholipids and protein.

The plasma membranes of cells that specialize in absorption are folded into fingerlike projections called microvilli (singular = microvillus). This folding increases the surface area of the plasma membrane. Such cells are typically found lining the small intestine, the organ that absorbs nutrients from digested food. This is an excellent example of form matching the function of a structure.

People with celiac disease have an immune response to gluten, which is a protein found in wheat, barley, and rye. The immune response damages microvilli, and thus, afflicted individuals cannot absorb nutrients. This leads to malnutrition, cramping, and diarrhea. Patients suffering from celiac disease must follow a gluten-free diet.

The Cytoplasm

The **cytoplasm** comprises the contents of a cell between the plasma membrane and the nuclear envelope (a structure to be discussed shortly). It is made up of organelles suspended in the gel-like **cytosol**, the cytoskeleton, and various chemicals (Figure 2.8). Even though the cytoplasm consists of 70 to 80 percent water, it has a semi-solid consistency, which comes from the proteins within it. However, proteins are not the only organic molecules found in the cytoplasm. Glucose and other simple sugars, polysaccharides, amino acids, nucleic acids, fatty acids, and derivatives of glycerol are found there too. Ions of sodium, potassium, calcium, and many other elements are also dissolved in the cytoplasm. Many metabolic reactions, including protein synthesis, take place in the cytoplasm.

The Cytoskeleton

If you were to remove all the organelles from a cell, would the plasma membrane and the cytoplasm be the only components left? No. Within the cytoplasm, there would still be ions and organic molecules, plus a network of protein fibers that helps to maintain the shape of the cell, secures certain organelles in specific positions, allows cytoplasm and vesicles to move within the cell, and enables unicellular organisms to move independently. Collectively, this network of protein fibers is known as the **cytoskeleton**. There are three types of fibers within the cytoskeleton: microfilaments, also known as actin filaments, intermediate filaments, and microtubules (Figure 2.10).
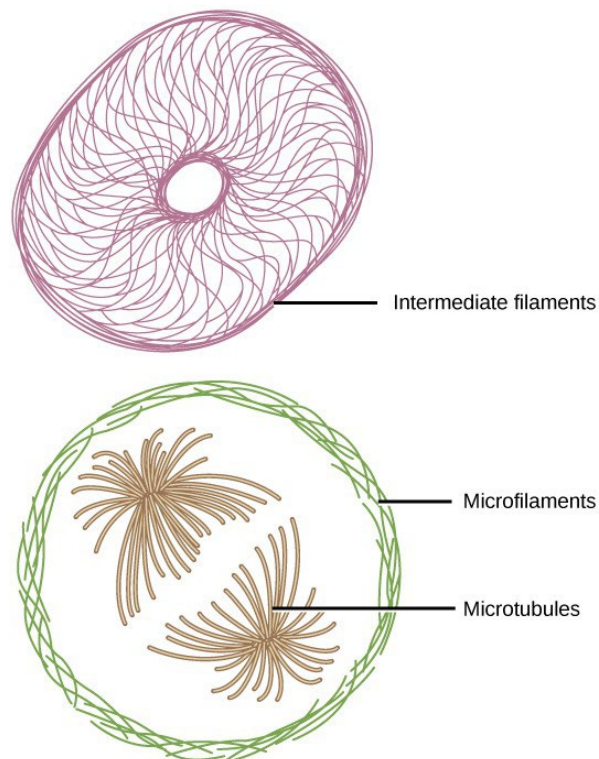
Figure 2.10 Microfilaments, intermediate filaments, and microtubules compose a cell's cytoskeleton.

Microfilaments are the thinnest of the cytoskeletal fibers and function in moving cellular components, for example, during cell division. They also maintain the structure of microvilli, the extensive folding of the plasma membrane found in cells dedicated to absorption. These components are also common in muscle cells and are responsible for muscle cell contraction. Intermediate filaments are of intermediate diameter and have structural functions, such as maintaining the shape of the cell and anchoring organelles. Keratin, the compound that strengthens hair and nails, forms one type of intermediate filament. Microtubules are the thickest of the cytoskeletal fibers. These are hollow tubes that can dissolve and reform quickly. Microtubules guide organelle movement and are the structures that pull chromosomes to their poles during cell division. They are also the structural components of flagella and cilia. In cilia and flagella, the microtubules are organized as a circle of nine double microtubules on the outside and two microtubules in the center.

The centrosome is a region near the nucleus of animal cells that functions as a microtubule-organizing center. It contains a pair of centrioles, two structures that lie perpendicular to each other. Each centriole is a cylinder of nine triplets of microtubules.

The centrosome replicates itself before a cell divides, and the centrioles play a role in pulling the duplicated chromosomes to opposite ends of the dividing cell. However, the exact function of the centrioles in cell division is not clear, since cells that have the centrioles removed can still divide, and plant cells, which lack centrioles, are capable of cell division.

Flagella and Cilia

Flagella (singular = flagellum) are long, hair-like structures that extend from the plasma membrane and are used to move an entire cell, (for example, sperm, *Euglena*). When present, the cell has just one flagellum or a few flagella. When **cilia** (singular = cilium) are present, however, they are many in number and extend along the entire surface of the plasma membrane. They are short, hair-like structures that are used to move entire cells (such as paramecium) or move substances along the outer surface of the cell (for example, the cilia of cells lining the fallopian tubes that move the ovum toward the uterus, or cilia lining the cells of the respiratory tract that move particulate matter toward the throat that mucus has trapped).

The Endomembrane System

The **endomembrane system** (*endo* = within) is a group of membranes and organelles (Figure 2.11) in eukaryotic cells that work together to modify, package, and transport lipids and proteins. It includes the nuclear envelope, lysosomes, and vesicles, the endoplasmic reticulum and Golgi apparatus, which

19

we will cover shortly. Although not technically *within* the cell, the plasma membrane is included in the endomembrane system because, as you will see, it interacts with the other endomembranous organelles.

The Nucleus

Typically, the nucleus is the most prominent organelle in a cell (Figure 2.8). The **nucleus** (plural = nuclei) houses the cell's DNA in the form of chromatin and directs the synthesis of ribosomes and proteins. Let us look at it in more detail (Figure 2.11).
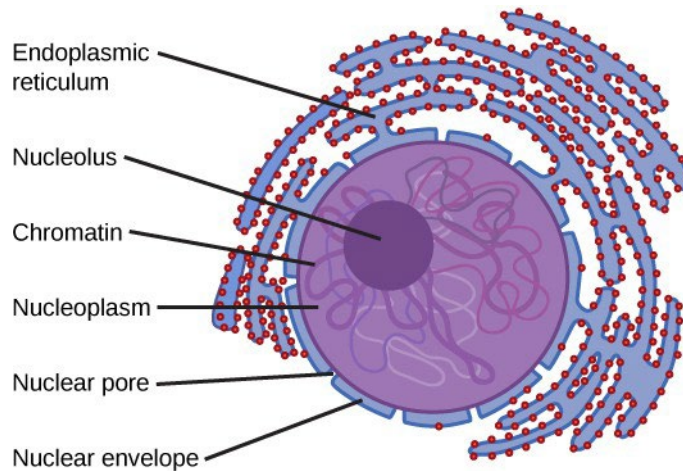
Figure 2.11 The outermost boundary of the nucleus is the nuclear envelope. Notice that the nuclear envelope consists of two phospholipid bilayers (membranes)—an outer membrane and an inner membrane—in contrast to the plasma membrane (Figure 2.9), which consists of only one phospholipid bilayer. (credit: modification of work by NIGMS, NIH)

The **nuclear envelope** is a double-membrane structure that constitutes the outermost portion of the nucleus (figure 2.11). Both the inner and outer membranes of the nuclear envelope are phospholipid bilayers.

The nuclear envelope is punctuated with pores that control the passage of ions, molecules, and RNA between the nucleoplasm and the cytoplasm.

To understand chromatin, it is helpful to first consider chromosomes. Chromosomes are structures within the nucleus that are made up of DNA, the hereditary material, and proteins. This combination of DNA and proteins is called chromatin. In eukaryotes, chromosomes are linear structures. Every species has a specific number of chromosomes in the nucleus of its body cells. For example, in humans, the chromosome number is 46, whereas in fruit flies, the chromosome number is eight.

Chromosomes are only visible and distinguishable from one another when the cell is getting ready to divide. When the cell is in the growth and maintenance phases of its life cycle, the chromosomes resemble an unwound, jumbled bunch of threads.

We already know that the nucleus directs the synthesis of ribosomes, but how does it do this? Some chromosomes have sections of DNA that encode ribosomal RNA. A darkly staining area within the nucleus, called the **nucleolus** (plural = nucleoli), aggregates the ribosomal RNA with associated proteins to assemble the ribosomal subunits that are then transported through the nuclear pores into the cytoplasm.

The Endoplasmic Reticulum

The **endoplasmic reticulum (ER)** (Figure 2.12) is a series of interconnected membranous tubules that collectively modify proteins and synthesize lipids. However, these two functions are performed in separate areas of the endoplasmic reticulum: the rough endoplasmic reticulum and the smooth endoplasmic reticulum, respectively.

The hollow portion of the ER tubules is called the lumen or cisternal space. The membrane of the ER, which is a phospholipid bilayer embedded with proteins, is continuous with the nuclear envelope.

The **rough endoplasmic reticulum (RER)** is so named because the ribosomes attached to its cytoplasmic surface give it a studded appearance when viewed through an electron microscope.

The ribosomes synthesize proteins while attached to the ER, resulting in transfer of their newly synthesized proteins into the lumen of the RER where they undergo modifications such as folding or addition of sugars. The RER also makes phospholipids for cell membranes.

If the phospholipids or modified proteins are not destined to stay in the RER, they will be packaged within vesicles and transported from the RER by budding from the membrane (Figure 2.13). Since the

- Download for free at http://openstaxcollege.org/textbooks/concepts-of- biology/get.

RER is engaged in modifying proteins that will be secreted from the cell, it is abundant in cells that secrete proteins, such as the liver.

The **smooth endoplasmic reticulum (SER)** is continuous with the RER but has few or no ribosomes on its cytoplasmic surface (see Figure 2.8). The SER's functions include synthesis of carbohydrates, lipids (including phospholipids), and steroid hormones; detoxification of medications and poisons; alcohol metabolism; and storage of calcium ions.

The Golgi Apparatus

We have already mentioned that vesicles can bud from the ER, but where do the vesicles go? Before reaching their final destination, the lipids or proteins within the transport vesicles need to be sorted, packaged, and tagged so that they wind up in the right place. The sorting, tagging, packaging, and distribution of lipids and proteins take place in the **Golgi apparatus** (also called the Golgi body), a series of flattened membranous sacs (Figure 2.12).
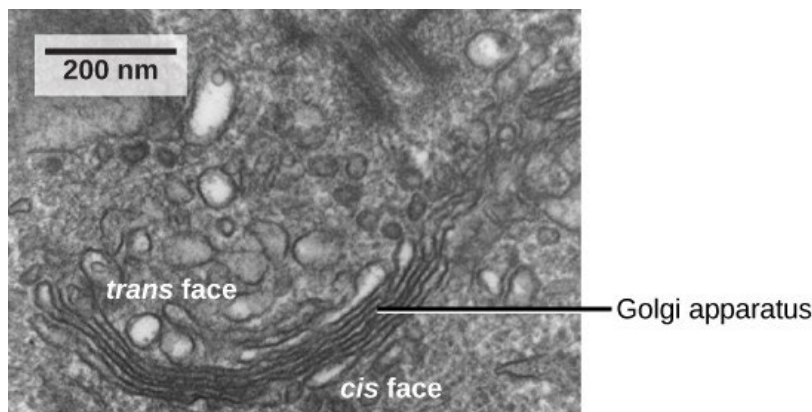
Figure 2.12 The Golgi apparatus in this transmission electron micrograph of a white blood cell is visible as a stack of semicircular flattened rings in the lower portion of this image. Several vesicles can be seen near the Golgi apparatus. (credit: modification of work by Louisa Howard; scale-bar data from Matt Russell)

The Golgi apparatus has a receiving face near the endoplasmic reticulum and a releasing face on the side away from the ER, toward the cell membrane. The transport vesicles that form from the ER travel to the receiving face, fuse with it, and empty their contents into the lumen of the Golgi apparatus. As the proteins and lipids travel through the Golgi, they undergo further modifications. The most frequent modification is the addition of short chains of sugar molecules. The newly modified proteins and lipids are then tagged with small molecular groups to enable them to be routed to their proper destinations.

Finally, the modified and tagged proteins are packaged into vesicles that bud from the opposite face of the Golgi. While some of these vesicles, transport vesicles, deposit their contents into other parts of the cell where they will be used, others, secretory vesicles, fuse with the plasma membrane and release their contents outside the cell.

The amount of Golgi in different cell types again illustrates that form follows function within cells. Cells that engage in a great deal of secretory activity (such as cells of the salivary glands that secrete digestive enzymes or cells of the immune system that secrete antibodies) have an abundant number of Golgi.

In plant cells, the Golgi has an additional role of synthesizing polysaccharides, some of which are incorporated into the cell wall and some of which are used in other parts of the cell.

Lysosomes

In animal cells, the **lysosomes** are the cell's "garbage disposal." Digestive enzymes within the lysosomes aid the breakdown of proteins, polysaccharides, lipids, nucleic acids, and even worn-out organelles. In single-celled eukaryotes, lysosomes are important for digestion of the food they ingest and the recycling of organelles. These enzymes are active at a much lower pH (more acidic) than those located in the cytoplasm. Many reactions that take place in the cytoplasm could not occur at a low pH, thus the advantage of compartmentalizing the eukaryotic cell into organelles is apparent.

Lysosomes also use their hydrolytic enzymes to destroy disease-causing organisms that might enter the cell. A good example of this occurs in a group of white blood cells called macrophages, which are part of your body's immune system. In a process known as phagocytosis, a section of the plasma membrane of the macrophage invaginates (folds in) and engulfs a pathogen. The invaginated section, with the pathogen inside, then pinches itself off from the plasma membrane and becomes a vesicle. The vesicle fuses with a lysosome. The lysosome's hydrolytic enzymes then destroy the pathogen (Figure 2.13).
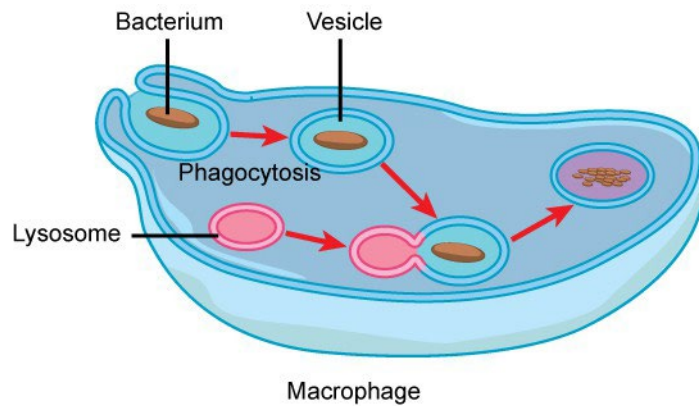
Figure 2.13 A macrophage has phagocytized a potentially pathogenic bacterium into a vesicle, which then fuses with a lysosome within the cell so that the pathogen can be destroyed. Other organelles are present in the cell, but for simplicity, are not shown.

Vesicles and Vacuoles

Vesicles and **vacuoles** are membrane-bound sacs that function in storage and transport. Vacuoles are somewhat larger than vesicles, and the membrane of a vacuole does not fuse with the membranes of other cellular components. Vesicles can fuse with other membranes within the cell system. Additionally, enzymes within plant vacuoles can break down macromolecules.

art CONNECTION
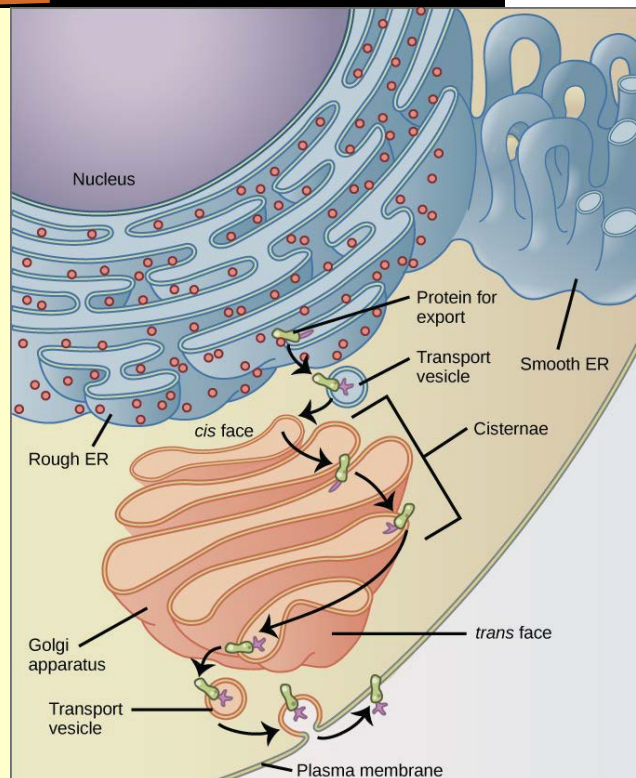
Figure 2.14 The endomembrane system works to modify, package, and transport lipids and proteins. (credit: modification of work by Magnus Manske)

21

Ribosomes

Ribosomes are the cellular structures responsible for protein synthesis. When viewed through an electron microscope, free ribosomes appear as either clusters or single tiny dots floating freely in the cytoplasm. Ribosomes may be attached to either the cytoplasmic side of the plasma membrane or the cytoplasmic side of the endoplasmic reticulum (). Electron microscopy has shown that ribosomes consist of large and small subunits. Ribosomes are enzyme complexes that are responsible for protein synthesis.

Because protein synthesis is essential for all cells, ribosomes are found in practically every cell, although they are smaller in prokaryotic cells. They are particularly abundant in immature red blood cells for the synthesis of hemoglobin, which functions in the transport of oxygen throughout the body.

Mitochondria

Mitochondria (singular = mitochondrion) are often called the "powerhouses" or "energy factories" of a cell because they are responsible for making adenosine triphosphate (ATP), the cell's main energy-carrying molecule. The formation of ATP from the breakdown of glucose is known as cellular respiration. Mitochondria are oval-shaped, double-membrane organelles (Figure 2.15) that have their own ribosomes and DNA. Each membrane is a phospholipid bilayer embedded with proteins. The   inner layer has folds called cristae, which increase the surface area of the inner membrane. The area surrounded by the folds is called the mitochondrial matrix. The cristae and the matrix have different roles in cellular respiration.

In keeping with our theme of form following function, it is important to point out that muscle cells have a very high concentration of mitochondria because muscle cells need a lot of energy to contract.
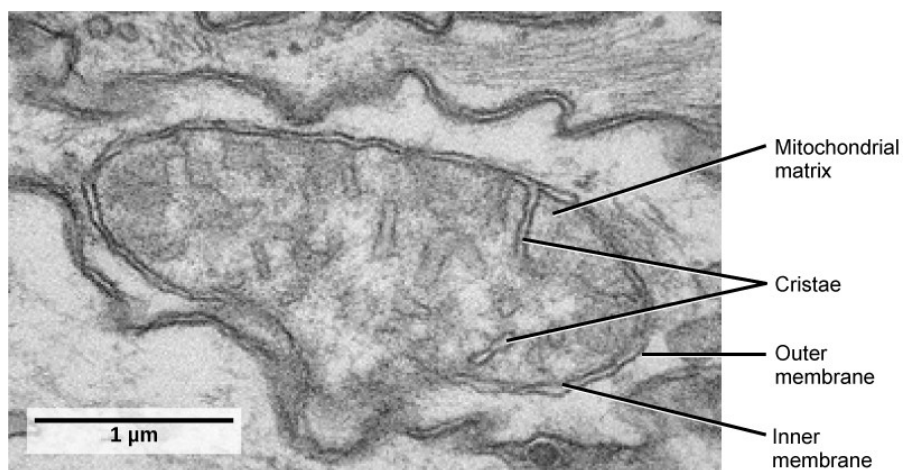
Figure 2.15 This transmission electron micrograph shows a mitochondrion as viewed with an electron microscope. Notice the inner and outer membranes, the cristae, and the mitochondrial matrix. (credit: modification of work by Matthew Britton; scale-bar data from Matt Russell)

Peroxisomes

Peroxisomes are small, round organelles enclosed by single membranes. They carry out oxidation reactions that break down fatty acids and amino acids. They also detoxify many poisons that may enter the body. Alcohol is detoxified by peroxisomes in liver cells. A byproduct of these oxidation reactions is hydrogen peroxide, H_2O_2, which is contained within the peroxisomes to prevent the chemical from causing damage to cellular components outside of the organelle. Hydrogen peroxide is safely broken down by peroxisomal enzymes into water and oxygen.

Animal Cells versus Plant Cells

Despite their fundamental similarities, there are some striking differences between animal and plant cells (see Table 2.1). Animal cells have centrioles, centrosomes (discussed under the cytoskeleton), and lysosomes, whereas plant cells do not. Plant cells have a cell wall, chloroplasts, plasmodesmata, and plastids used for storage, and a large central vacuole, whereas animal cells do not.

The Cell Wall

In Figure 2.8**b**, the diagram of a plant cell, you see a structure external to the plasma membrane called the cell wall. The **cell wall** is a rigid covering that protects the cell, provides structural support, and gives shape to the cell. Fungal and protist cells also have cell walls.

While the chief component of prokaryotic cell walls is peptidoglycan, the major organic molecule in the plant cell wall is cellulose, a polysaccharide made up of long, straight chains of glucose units. When nutritional information refers to dietary fiber, it is referring to the cellulose content of food.

Chloroplasts

Like mitochondria, chloroplasts also have their own DNA and ribosomes. **Chloroplasts** function in photosynthesis and can be found in eukaryotic cells such as plants and algae. In photosynthesis, carbon dioxide, water, and light energy are used to make glucose and oxygen. This is the major difference between plants and animals: Plants (autotrophs) are able to make their own food, like glucose, whereas animals (heterotrophs) must rely on other organisms for their organic compounds or food source.

Like mitochondria, chloroplasts have outer and inner membranes, but within the space enclosed by a chloroplast's inner membrane is a set of interconnected and stacked, fluid-filled membrane sacs called thylakoids (Figure 2.16). Each stack of thylakoids is called a granum (plural = grana). The fluid enclosed by the inner membrane and surrounding the grana is called the stroma.
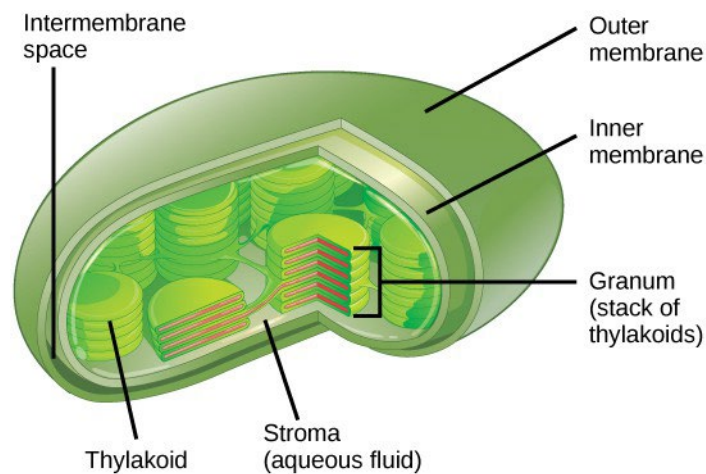
Figure 2.16 This simplified diagram of a chloroplast shows the outer membrane, inner membrane, thylakoids, grana, and stroma.

The chloroplasts contain a green pigment called chlorophyll, which captures the energy of sunlight for photosynthesis. Like plant cells, photosynthetic protists also have chloroplasts. Some bacteria also perform photosynthesis, but they do not have chloroplasts. Their photosynthetic pigments are located in the thylakoid membrane within the cell itself.

23

Endosymbiosis

We have mentioned that both mitochondria and chloroplasts contain DNA and ribosomes. Have you wondered why? Strong evidence points to endosymbiosis as the explanation.

Symbiosis is a relationship in which organisms from two separate species live in close association and typically exhibit specific adaptations to each other. Endosymbiosis (*endo-=* within) is a relationship in which one organism lives inside the other. Endosymbiotic relationships abound in nature. Microbes that produce vitamin K live inside the human gut. This relationship is beneficial for us because we are unable to synthesize vitamin K. It is also beneficial for the microbes because they are protected from other organisms and are provided a stable habitat and abundant food by living within the large intestine.

Scientists have long noticed that bacteria, mitochondria, and chloroplasts are similar in size. We also know that mitochondria and chloroplasts have DNA and ribosomes, just as bacteria do. Scientists believe that host cells and bacteria formed a mutually beneficial endosymbiotic relationship when the host cells ingested aerobic bacteria and cyanobacteria but did not destroy them. Through evolution, these ingested bacteria became more specialized in their functions, with the aerobic bacteria becoming mitochondria and the photosynthetic bacteria becoming chloroplasts.

The Central Vacuole

Previously, we mentioned vacuoles as essential components of plant cells. If you look at Figure 2.8, you will see that plant cells each have a large, central vacuole that occupies most of the cell. The **central vacuole** plays a key role in regulating the cell's concentration of water in changing environmental conditions. In plant cells, the liquid inside the central vacuole provides turgor pressure, which is the outward pressure caused by the fluid inside the cell. Have you ever noticed that if you forget to water a plant for a few days, it wilts? That is because as the water concentration in the soil becomes lower than the water concentration in the plant, water moves out of the central vacuoles and cytoplasm and into the soil. As the central vacuole shrinks, it leaves the cell wall unsupported. This loss of support to the cell walls of a plant results in the wilted appearance. Additionally, this fluid has a very bitter taste, which discourages consumption by insects and animals. The central vacuole also functions to store proteins in developing seed cells.

Extracellular Matrix of Animal Cells

Most animal cells release materials into the extracellular space. The primary components of these materials are glycoproteins and the protein collagen. Collectively, these materials are called the **extracellular matrix** (Figure 2.17). Not only does the extracellular matrix hold the cells together to form a tissue, but it also allows the cells within the tissue to communicate with each other.
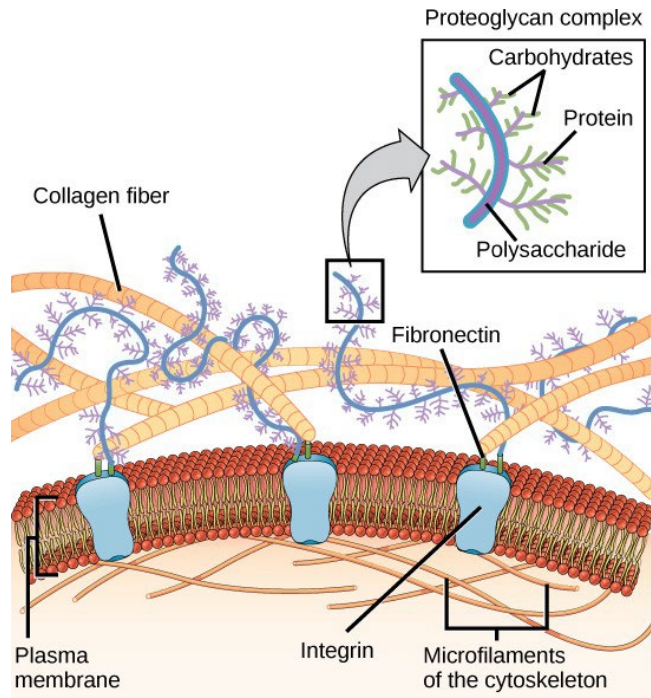
Figure 2.17 The extracellular matrix consists of a network of substances secreted by cells.

Blood clotting provides an example of the role of the extracellular matrix in cell communication. When the cells lining a blood vessel are damaged, they display a protein receptor called tissue factor. When tissue factor binds with another factor in the extracellular matrix, it causes platelets to adhere to the wall of the damaged blood vessel, stimulates adjacent smooth muscle cells in the blood vessel to contract (thus constricting the blood vessel), and initiates a series of steps that stimulate the platelets to produce clotting factors.

Intercellular Junctions

Cells can also communicate with each other by direct contact, referred to as intercellular junctions. There are some differences in the ways that plant and animal cells do this. **Plasmodesmata** (singular = plasmodesma) are junctions between plant cells, whereas animal cell contacts include tight and gap junctions, and desmosomes.

In general, long stretches of the plasma membranes of neighboring plant cells cannot touch one another because they are separated by the cell walls surrounding each cell. Plasmodesmata are numerous channels that pass between the cell walls of adjacent plant cells, connecting their cytoplasm and enabling signal molecules and nutrients to be transported from cell to cell (**Figure 2.18a**).

25
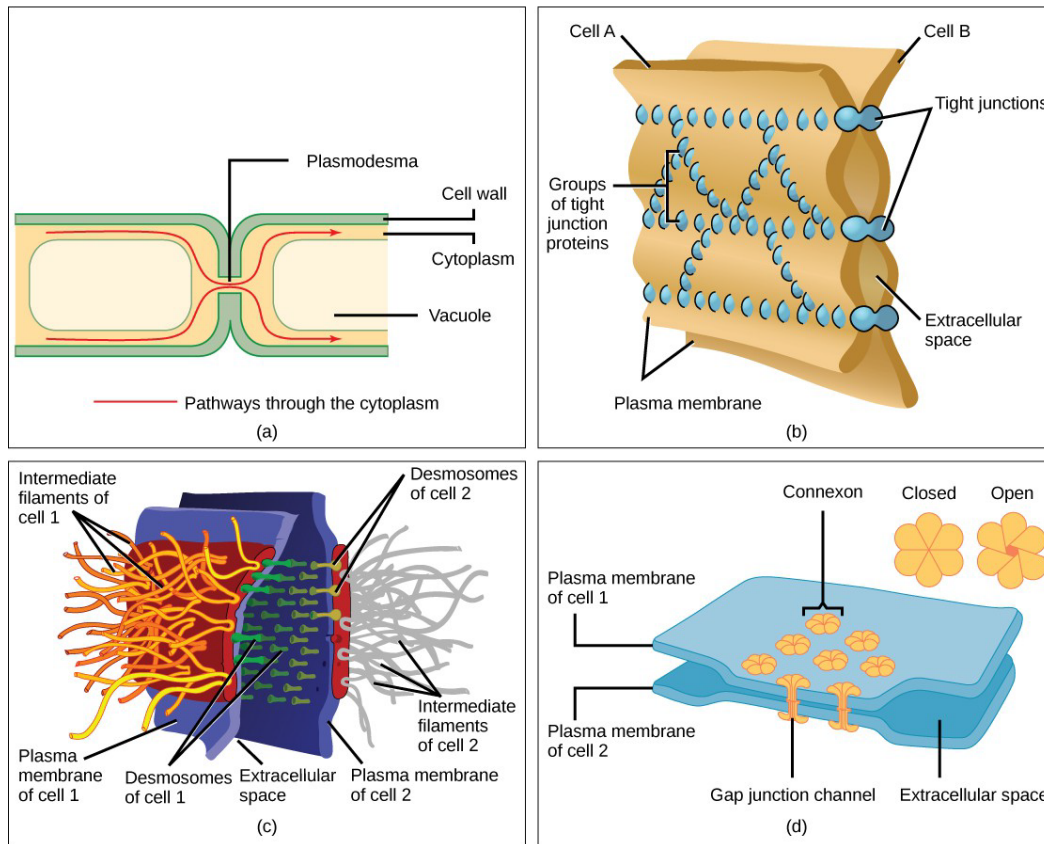
Figure 2.18 There are four kinds of connections between cells. (a) A plasmodesma is a channel between the cell walls of two adjacent plant cells. (b) Tight junctions join adjacent animal cells. (c) Desmosomes join two animal cells together. (d) Gap junctions act as channels between animal cells. (credit b, c, d: modification of work by Mariana Ruiz Villareal)

A **tight junction** is a watertight seal between two adjacent animal cells (Figure 2.18**b**). Proteins hold the cells tightly against each other. This tight adhesion prevents materials from leaking between the cells. Tight junctions are typically found in the epithelial tissue that lines internal organs and cavities, and composes most of the skin. For example, the tight junctions of the epithelial cells lining the urinary bladder prevent urine from leaking into the extracellular space.

Also found only in animal cells are **desmosomes**, which act like spot welds between adjacent epithelial cells (Figure 2.18**c**). They keep cells together in a sheet-like formation in organs and tissues that stretch, like the skin, heart, and muscles.

Gap junctions in animal cells are like plasmodesmata in plant cells in that they are channels between adjacent cells that allow for the transport of ions, nutrients, and other substances that enable cells to communicate (Figure 2.18**d**). Structurally, however, gap junctions and plasmodesmata differ.

Components of Prokaryotic and Eukaryotic Cells and Their Functions

Cell Component	Function	Present in Prokaryotes?	Present in Animal Cells?	Present in Plant Cells?
Plasma membrane	Separates cell from external environment; controls passage of organic molecules, ions, water, oxygen, and wastes into and out of the cell	Yes	Yes	Yes

Table 3.1

26

Components of Prokaryotic and Eukaryotic Cells and Their Functions

Cell Component	Function	Present in Prokaryotes?	Present in Animal Cells?	Present in Plant Cells?
Cytoplasm	Provides structure to cell; site of many metabolic reactions; medium in which organelles are found	Yes	Yes	Yes
Nucleoid	Location of DNA	Yes	No	No
Nucleus	Cell organelle that houses DNA and directs synthesis of ribosomes and proteins	No	Yes	Yes
Ribosomes	Protein synthesis	Yes	Yes	Yes
Mitochondria	ATP production/cellular respiration	No	Yes	Yes
Peroxisomes	Oxidizes and breaks down fatty acids and amino acids, and detoxifies poisons	No	Yes	Yes
Vesicles and vacuoles	Storage and transport; digestive function in plant cells	No	Yes	Yes
Centrosome	Unspecified role in cell division in animal cells; organizing center of microtubules in animal cells	No	Yes	No
Lysosomes	Digestion of macromolecules; recycling of worn-out organelles	No	Yes	No
Cell wall	Protection, structural support and maintenance of cell shape	Yes, primarily peptidoglycan in bacteria but not Archaea	No	Yes, primarily cellulose
Chloroplasts	Photosynthesis	No	No	Yes
Endoplasmic reticulum	Modifies proteins and synthesizes lipids	No	Yes	Yes
Golgi apparatus	Modifies, sorts, tags, packages, and distributes lipids and proteins	No	Yes	Yes
Cytoskeleton	Maintains cell's shape, secures organelles in specific positions, allows cytoplasm and vesicles to move within the cell, and enables unicellular organisms to move independently	Yes	Yes	Yes

Table 2.1

## Components of Prokaryotic and Eukaryotic Cells and Their Functions

Cell Component	Function	Present in Prokaryotes?	Present in Animal Cells?	Present in Plant Cells?
Flagella	Cellular locomotion	Some	Some	No, except for some plant sperm.
Cilia	Cellular locomotion, movement of particles along extracellular surface of plasma membrane, and filtration	No	Some	No

Table 2.1

This table provides the components of prokaryotic and eukaryotic cells and their respective functions.

| The Cell Cycle

The **cell cycle** is an ordered series of events involving cell growth and cell division that produces two new daughter cells. Cells on the path to cell division proceed through a series of precisely timed and carefully regulated stages of growth, DNA replication, and division that produce two genetically identical cells. The cell cycle has two major phases: interphase and the mitotic phase (Figure 2.19). During **interphase**, the cell grows and DNA is replicated. During the **mitotic phase**, the replicated DNA and cytoplasmic contents are separated and the cell divides. A link back to the SoftChalk module.
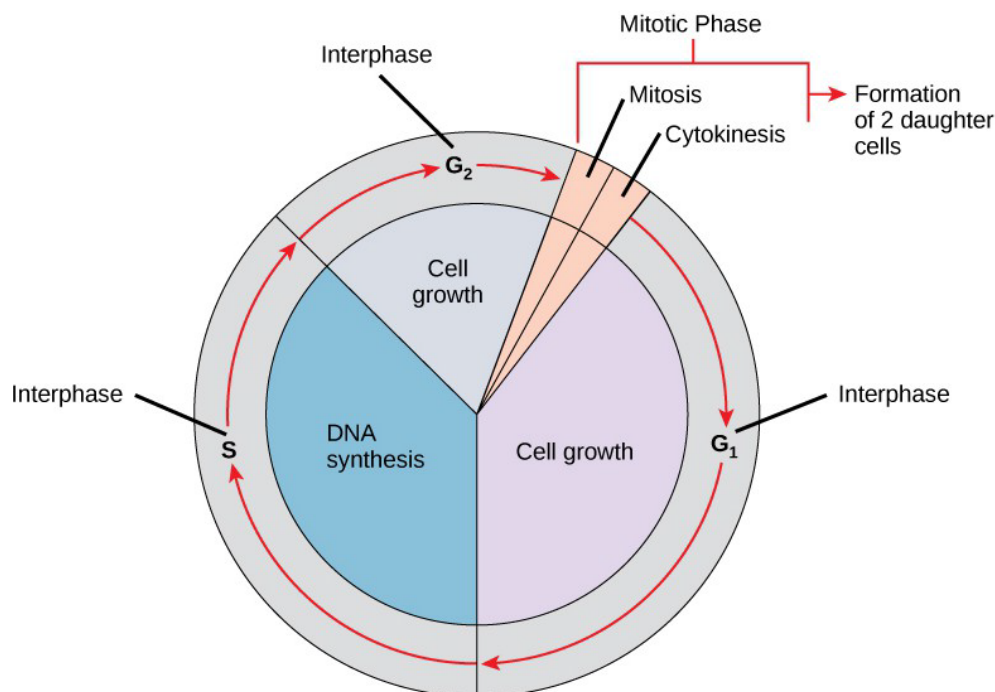
Figure 2.19 A cell moves through a series of phases in an orderly manner. During interphase, G_1 involves cell growth and protein synthesis, the S phase involves DNA replication and the replication of the centrosome, and G_2 involves further growth and protein synthesis. The mitotic phase follows interphase. Mitosis is nuclear division during which duplicated chromosomes are segregated and distributed into daughter nuclei. Usually the cell will divide after mitosis in a process called cytokinesis in which the cytoplasm is divided and two daughter cells are formed.

28

During interphase, the cell undergoes normal processes while also preparing for cell division. For a cell to move from interphase to the mitotic phase, many internal and external conditions must be met. The three stages of interphase are called G_1, S, and G_2.

G_1 Phase

The first stage of interphase is called the **G_1 phase**, or first gap, because little change is visible. However, during the G_1 stage, the cell is quite active at the biochemical level. The cell is accumulating the

29

building blocks of chromosomal DNA and the associated proteins, as well as accumulating enough energy reserves to complete the task of replicating each chromosome in the nucleus.

S Phase

Throughout interphase, nuclear DNA remains in a semi-condensed chromatin configuration. In the **S phase** (synthesis phase), DNA replication results in the formation of two identical copies of each chromosome—sister chromatids—that are firmly attached at the centromere region. At this stage, each chromosome is made of two sister chromatids and is a duplicated chromosome. The centrosome is duplicated during the S phase. The two centrosomes will give rise to the **mitotic spindle**, the apparatus that orchestrates the movement of chromosomes during mitosis. The centrosome consists of a pair of rod-like **centrioles** at right angles to each other. Centrioles help organize cell division. Centrioles are not present in the centrosomes of many eukaryotic species, such as plants and most fungi.
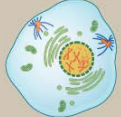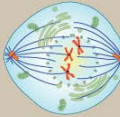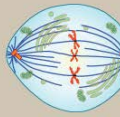
G_2 Phase

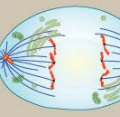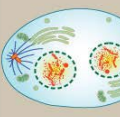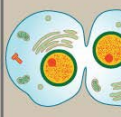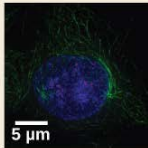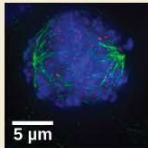In the **G_2 phase**, or second gap, the cell replenishes its energy stores and synthesizes the proteins necessary for chromosome manipulation. Some cell organelles are duplicated, and the cytoskeleton is dismantled to provide resources for the mitotic spindle. There may be additional cell growth during G_2. The final preparations for the mitotic phase must be completed before the cell is able to enter the first stage of mitosis.
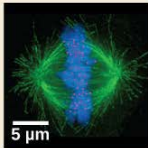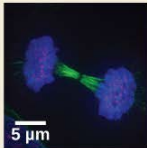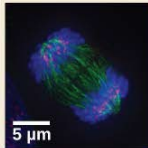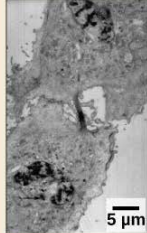
The Mitotic Phase

To make two daughter cells, the contents of the nucleus and the cytoplasm must be divided. The mitotic phase is a multistep process during which the duplicated chromosomes are aligned, separated, and moved to opposite poles of the cell, and then the cell is divided into two new identical daughter cells. The first portion of the mitotic phase, **mitosis**, is composed of five stages, which accomplish nuclear division. The second portion of the mitotic phase, called cytokinesis, is the physical separation of the cytoplasmic components into two daughter cells.

Mitosis

Mitosis is divided into a series of phases—prophase, prometaphase, metaphase, anaphase, and telophase—that result in the division of the cell nucleus (Figure 2.20).

Prophase	Prometaphase	Metaphase	Anaphase	Telophase	Cytokinesis
• Chromosomes condense and become visible • Spindle fibers emerge from the centrosomes • Nuclear envelope breaks down • Centrosomes move toward opposite poles	• Chromosomes continue to condense • Kinetochores appear at the centromeres • Mitotic spindle microtubules attach to kinetochores	• Chromosomes are lined up at the metaphase plate • Each sister chromatid is attached to a spindle fiber originating from opposite poles	• Centromeres split in two • Sister chromatids (now called chromosomes) are pulled toward opposite poles • Certain spindle fibers begin to elongate the cell	• Chromosomes arrive at opposite poles and begin to decondense • Nuclear envelope material surrounds each set of chromosomes • The mitotic spindle breaks down • Spindle fibers continue to push poles apart	• Animal cells: a cleavage furrow separates the daughter cells • Plant cells: a cell plate, the precursor to a new cell wall, separates the daughter cells

MITOSIS

Figure 2.20 Animal cell mitosis is divided into five stages—prophase, prometaphase, metaphase, anaphase, and telophase—visualized here by light microscopy with fluorescence. Mitosis is usually accompanied by cytokinesis, shown here by a transmission electron microscope. (credit "diagrams": modification of work by Mariana Ruiz Villareal; credit "mitosis micrographs": modification of work by Roy van Heesbeen; credit "cytokinesis micrograph": modification of work by the Wadsworth Center, NY State Department of Health; donated to the Wikimedia foundation; scale-bar data from Matt Russell)

Which of the following is the correct order of events in mitosis?

a. Sister chromatids line up at the metaphase plate. The kinetochore becomes attached to the mitotic spindle. The nucleus re-forms and the cell divides. The sister chromatids separate.

b. The kinetochore becomes attached to the mitotic spindle. The sister chromatids separate. Sister chromatids line up at the metaphase plate. The nucleus re-forms and the cell divides.

c. The kinetochore becomes attached to metaphase plate. Sister chromatids line up at the metaphase plate. The kinetochore breaks down and the sister chromatids separate. The nucleus re-forms and the cell divides.

d. The kinetochore becomes attached to the mitotic spindle. Sister chromatids line up at the metaphase plate. The kinetochore breaks apart and the sister chromatids separate. The nucleus re-forms and the cell divides.

During **prophase**, the "first phase," several events must occur to provide access to the chromosomes in the nucleus. The nuclear envelope starts to break into small vesicles, and the Golgi apparatus and endoplasmic reticulum fragment and disperse to the periphery of the cell. The nucleolus disappears. The centrosomes begin to move to opposite poles of the cell. The microtubules that form the basis of the mitotic spindle extend between the centrosomes, pushing them farther apart as the microtubule fibers lengthen. The sister chromatids begin to coil more tightly and become visible under a light microscope.

31

During **prometaphase**, many processes that were begun in prophase continue to advance and culminate in the formation of a connection between the chromosomes and cytoskeleton. The remnants of the nuclear envelope disappear. The mitotic spindle continues to develop as more microtubules assemble and stretch across the length of the former nuclear area. Chromosomes become more condensed and visually discrete. Each sister chromatid attaches to spindle microtubules at the centromere via a protein complex called the **kinetochore**.

During **metaphase**, all of the chromosomes are aligned in a plane called the **metaphase plate**, or the equatorial plane, midway between the two poles of the cell. The sister chromatids are still tightly attached to each other. At this time, the chromosomes are maximally condensed.

During **anaphase**, the sister chromatids at the equatorial plane are split apart at the centromere. Each chromatid, now called a chromosome, is pulled rapidly toward the centrosome to which its microtubule was attached. The cell becomes visibly elongated as the non-kinetochore microtubules slide against each other at the metaphase plate where they overlap.

During **telophase**, all of the events that set up the duplicated chromosomes for mitosis during the first three phases are reversed. The chromosomes reach the opposite poles and begin to decondense (unravel). The mitotic spindles are broken down into monomers that will be used to assemble cytoskeleton components for each daughter cell. Nuclear envelopes form around chromosomes.

This page of movies (http://openstaxcollege.org/l/divisn_newtcell) illustrates different aspects of mitosis. Watch the movie entitled "DIC microscopy of cell division in a newt lung cell" and identify the phases of mitosis.

Cytokinesis

Cytokinesis is the second part of the mitotic phase during which cell division is completed by the physical separation of the cytoplasmic components into two daughter cells. Although the stages of mitosis are similar for most eukaryotes, the process of cytokinesis is quite different for eukaryotes that have cell walls, such as plant cells.

In cells such as animal cells that lack cell walls, cytokinesis begins following the onset of anaphase. A contractile ring composed of actin filaments forms just inside the plasma membrane at the former metaphase plate. The actin filaments pull the equator of the cell inward, forming a fissure. This fissure, or "crack," is called the **cleavage furrow**. The furrow deepens as the actin ring contracts, and eventually the membrane and cell are cleaved in two (Figure 2.21).

In plant cells, a cleavage furrow is not possible because of the rigid cell walls surrounding the plasma membrane. A new cell wall must form between the daughter cells. During interphase, the Golgi apparatus accumulates enzymes, structural proteins, and glucose molecules prior to breaking up into vesicles and dispersing throughout the dividing cell. During telophase, these Golgi vesicles move on microtubules to collect at the metaphase plate. There, the vesicles fuse from the center toward the cell walls; this structure is called a **cell plate**. As more vesicles fuse, the cell plate enlarges until it merges with the cell wall at the periphery of the cell. Enzymes use the glucose that has accumulated between the membrane layers to build a new cell wall of cellulose. The Golgi membranes become the plasma membrane on either side of the new cell wall (Figure 2.21).

(a) Animal cell

Cleavage furrow

Contractile ring

(b) Plant cell

Cell plate

Golgi vesicles
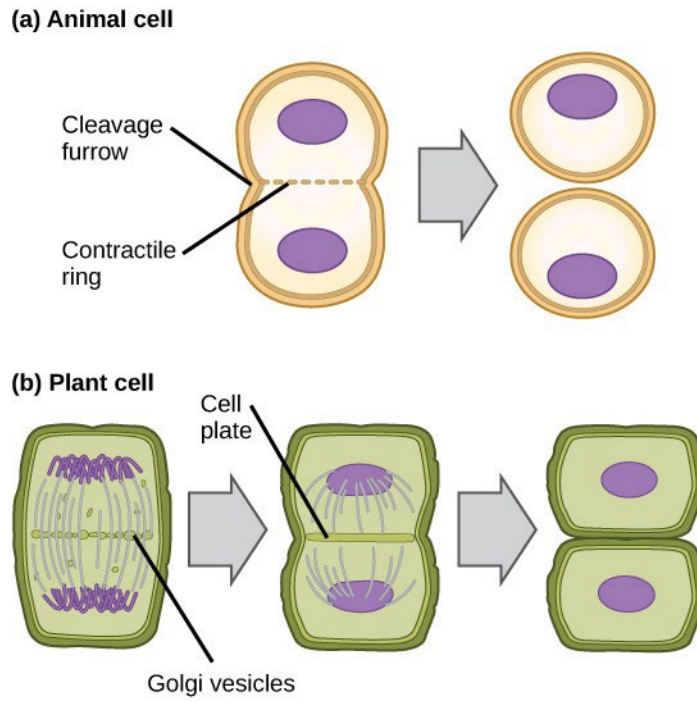
Figure 2.21 In part (a), a cleavage furrow forms at the former metaphase plate in the animal cell. The plasma membrane is drawn in by a ring of actin fibers contracting just inside the membrane. The cleavage furrow deepens until the cells are pinched in two. In part (b), Golgi vesicles coalesce at the former metaphase plate in a plant cell. The vesicles fuse and form the cell plate. The cell plate grows from the center toward the cell walls. New cell walls are made from the vesicle contents.

G_0 Phase

Not all cells adhere to the classic cell-cycle pattern in which a newly formed daughter cell immediately enters interphase, closely followed by the mitotic phase. Cells in the **G_0 phase** are not actively preparing to divide. The cell is in a quiescent (inactive) stage, having exited the cell cycle. Some cells enter G_0 temporarily until an external signal triggers the onset of G_1. Other cells that never or rarely divide, such as mature cardiac muscle and nerve cells, remain in G_0 permanently (Figure 2.22).
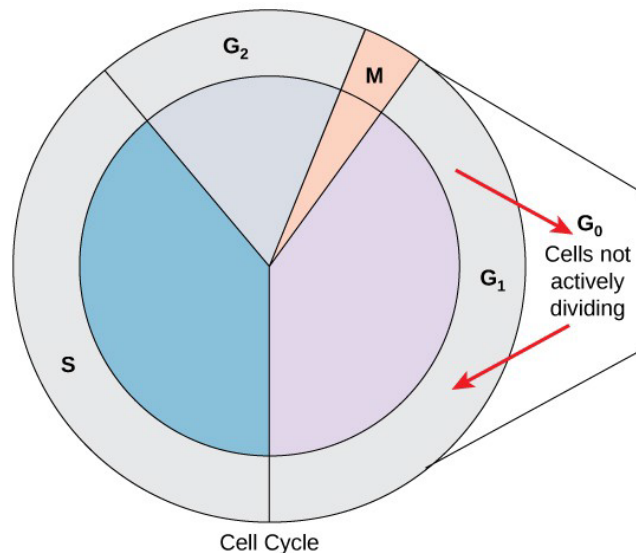

Figure 2.22 Cells that are not actively preparing to divide enter an alternate phase called G_0. In some cases, this is a temporary condition until triggered to enter G_1. In other cases, the cell will remain in G_0 permanently.

Control of the Cell Cycle

The length of the cell cycle is highly variable even within the cells of an individual organism. In humans, the frequency of cell turnover ranges from a few hours in early embryonic development to an average of two to five days for epithelial cells, or to an entire human lifetime spent in G_0 by specialized cells such as cortical neurons or cardiac muscle cells. There is also variation in the time that a cell spends in each phase of the cell cycle. When fast-dividing mammalian cells are grown in culture (outside the body under optimal growing conditions), the length of the cycle is approximately 24 hours. In rapidly dividing human cells with a 24-hour cell cycle, the G_1 phase lasts approximately 11 hours. The timing of events in the cell cycle is controlled by mechanisms that are both internal and external to the cell.

Regulation at Internal Checkpoints

It is essential that daughter cells be exact duplicates of the parent cell. Mistakes in the duplication or distribution of the chromosomes lead to mutations that may be passed forward to every new cell produced from the abnormal cell. To prevent a compromised cell from continuing to divide, there are internal control mechanisms that operate at three main **cell cycle checkpoints** at which the cell cycle can be stopped until conditions are favorable. These checkpoints occur near the end of G_1, at the G_2–M transition, and during metaphase (Figure 2..23).
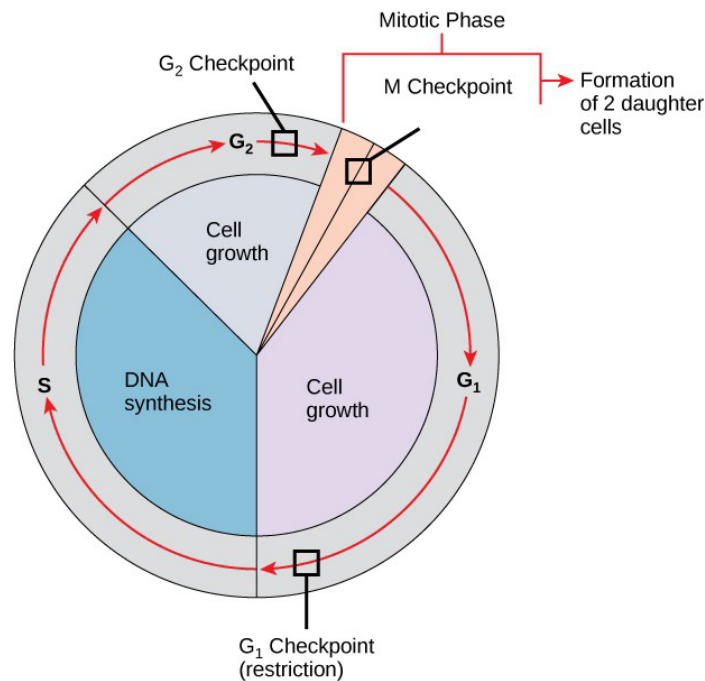
Figure 2.23 The cell cycle is controlled at three checkpoints. Integrity of the DNA is assessed at the G_1 checkpoint. Proper chromosome duplication is assessed at the G_2 checkpoint. Attachment of each kinetochore to a spindle fiber is assessed at the M checkpoint.

The G₁ Checkpoint

The G_1 checkpoint determines whether all conditions are favorable for cell division to proceed. The G_1 checkpoint, also called the restriction point, is the point at which the cell irreversibly commits to the cell-division process. In addition to adequate reserves and cell size, there is a check for damage to the genomic DNA at the G_1 checkpoint. A cell that does not meet all the requirements will not be released into the S phase.

The G₂ Checkpoint

The G_2 checkpoint bars the entry to the mitotic phase if certain conditions are not met. As in the G_1 checkpoint, cell size and protein reserves are assessed. However, the most important role of the G_2 checkpoint is to ensure that all of the chromosomes have been replicated and that the replicated DNA is not damaged.

The M Checkpoint

The M checkpoint occurs near the end of the metaphase stage of mitosis. The M checkpoint is also known as the spindle checkpoint because it determines if all the sister chromatids are correctly attached to the spindle microtubules. Because the separation of the sister chromatids during anaphase is an

34

irreversible step, the cycle will not proceed until the kinetochores of each pair of sister chromatids are firmly anchored to spindle fibers arising from opposite poles of the cell.

CONCEPT in ACTION

Watch what occurs at the G_1, G_2, and M checkpoints by visiting **this animation (http://openstaxcollege.org/l/cell_checkpnts2)** of the cell cycle.

| Sexual Reproduction

Sexual reproduction was an early evolutionary innovation after the appearance of eukaryotic cells. The fact that most eukaryotes reproduce sexually is evidence of its evolutionary success. In many animals, it is the only mode of reproduction. And yet, scientists recognize some real disadvantages to sexual reproduction. On the surface, offspring that are genetically identical to the parent may appear to be more advantageous. If the parent organism is successfully occupying a habitat, offspring with the same traits would be similarly successful. There is also the obvious benefit to an organism that can produce offspring by asexual budding, fragmentation, or asexual eggs. These methods of reproduction do not require another organism of the opposite sex. There is no need to expend energy finding or attracting a mate. That energy can be spent on producing more offspring. Indeed, some organisms that lead a solitary lifestyle have retained the ability to reproduce asexually. In addition, asexual populations only have female individuals, so every individual is capable of reproduction. In contrast, the males in sexual populations (half the population) are not producing offspring themselves. Because of this, an asexual population can grow twice as fast as a sexual population in theory. This means that in competition, the asexual population would have the advantage. All of these advantages to asexual reproduction, which are also disadvantages to sexual reproduction, should mean that the number of species with asexual reproduction should be more common.

However, multicellular organisms that exclusively depend on asexual reproduction are exceedingly rare. Why is sexual reproduction so common? This is one of the important questions in biology and has been the focus of much research from the latter half of the twentieth century until now. A likely explanation is that the variation that sexual reproduction creates among offspring is very important to the survival and reproduction of those offspring. The only source of variation in asexual organisms is mutation. This is the ultimate source of variation in sexual organisms. In addition, those different mutations are continually reshuffled from one generation to the next when different parents combine their unique genomes, and the genes are mixed into different combinations by the process of **meiosis**. Meiosis is the division of the contents of the nucleus that divides the chromosomes among gametes. Variation is introduced during meiosis, as well as when the gametes combine in fertilization.

35

The Red Queen Hypothesis

There is no question that sexual reproduction provides evolutionary advantages to organisms that employ this mechanism to produce offspring. The problematic question is why, even in the face of fairly stable conditions, sexual reproduction persists when it is more difficult and produces fewer offspring for individual organisms? Variation is the outcome of sexual reproduction, but why are ongoing variations necessary? Enter the Red

Queen hypothesis, first proposed by Leigh Van Valen in 1973. [1] The concept was named

in reference to the Red Queen's race in Lewis Carroll's book, *Through the Looking-Glass,*

in which the Red Queen says one must run at full speed just to stay where one is.

All species coevolve with other organisms. For example, predators coevolve with their prey, and parasites coevolve with their hosts. A remarkable example of coevolution between predators and their prey is the unique coadaptation of night flying bats and their moth prey. Bats find their prey by emitting high-pitched clicks, but moths have evolved simple ears to hear these clicks so they can avoid the bats. The moths have also adapted behaviors, such as flying away from the bat when they first hear it, or dropping suddenly to the ground when the bat is upon them. Bats have evolved "quiet" clicks in an attempt to evade the moth's hearing. Some moths have evolved the ability to respond to the bats' clicks with their own clicks as a strategy to confuse the bats echolocation abilities.

Each tiny advantage gained by favorable variation gives a species an edge over close competitors, predators, parasites, or even prey. The only method that will allow a coevolving species to keep its own share of the resources is also to continually improve its ability to survive and produce offspring. As one species gains an advantage, other species must also develop an advantage or they will be outcompeted. No single species progresses too far ahead because genetic variation among progeny of sexual reproduction provides all species with a mechanism to produce adapted individuals. Species whose individuals

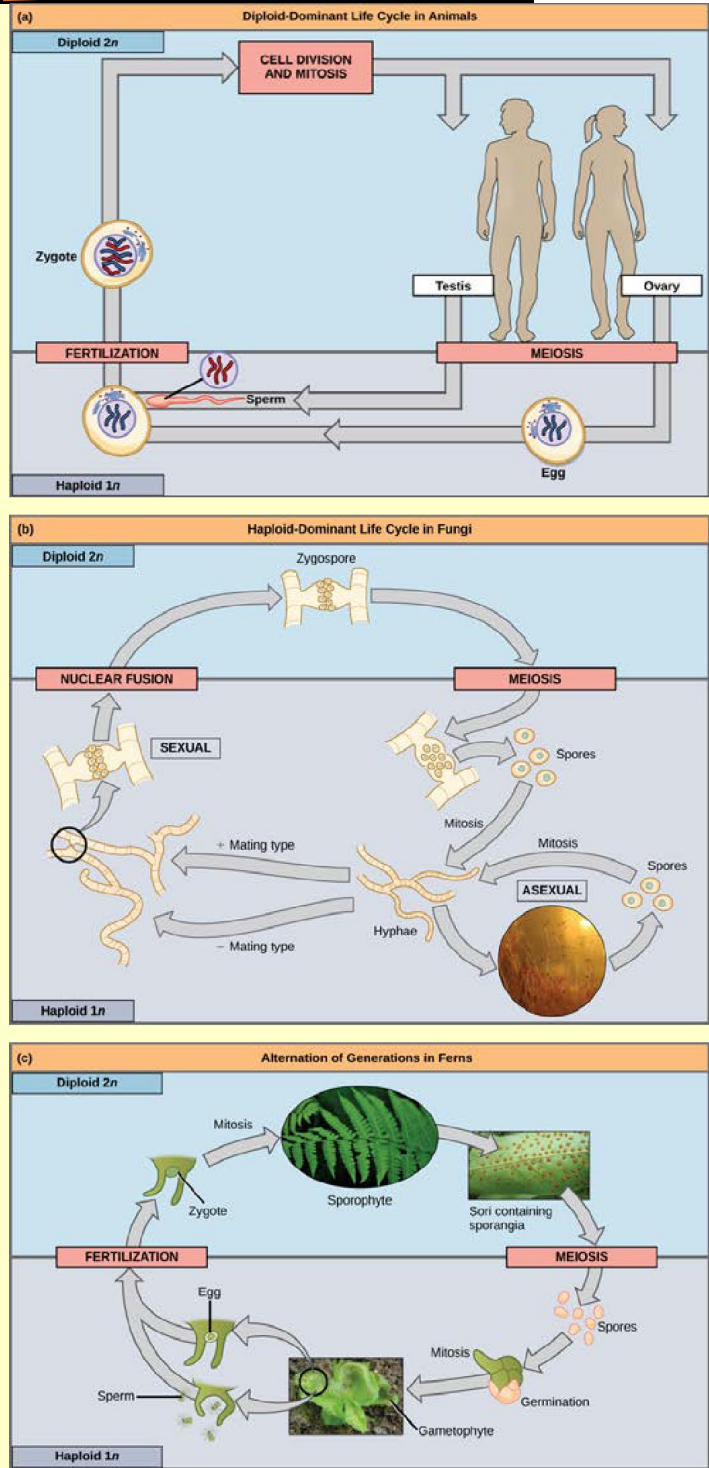### *Life Cycles of Sexually Reproducing Organisms*

Fertilization and meiosis alternate in sexual **life cycles**. What happens between these two events depends on the organism. The process of meiosis reduces the resulting gamete's chromosome number by half. Fertilization, the joining of two haploid gametes, restores the diploid condition. There are three main categories of life cycles in multicellular organisms: **diploid-dominant**, in which the multicellular diploid stage is the most obvious life stage (and there is no multicellular haploid stage), as with most animals including humans; **haploid-dominant**, in which the multicellular haploid stage is the most obvious life stage (and there is no multicellular diploid stage), as with all fungi and some algae; and **alternation of generations**, in which the two stages, haploid and diploid, are apparent to one degree or another depending on the group, as with plants and some algae.

Nearly all animals employ a diploid-dominant life-cycle strategy in which the only haploid cells produced by the organism are the gametes. The gametes are produced from diploid **germ cells**, a special cell line that only produces gametes. Once the haploid gametes are formed, they lose the ability to divide again. There is no multicellular haploid life stage. Fertilization occurs with the fusion of two gametes, usually from different individuals, restoring the diploid state (Figure 2.24**a**).

---

1. Leigh Van Valen, "A new evolutionary law," Evolutionary Theory 1 (1973): 1–30.

(a) Diploid-Dominant Life Cycle in Animals

Diploid 2n

CELL DIVISION AND MITOSIS

Zygote

Testis | Ovary

FERTILIZATION | MEIOSIS

Sperm

Egg

Haploid 1n

(b) Haploid-Dominant Life Cycle in Fungi

Diploid 2n

Zygospore

NUCLEAR FUSION | MEIOSIS

SEXUAL

Spores

Mitosis | Mitosis

+ Mating type

ASEXUAL

Spores

− Mating type | Hyphae

Haploid 1n

(c) Alternation of Generations in Ferns

Diploid 2n

Mitosis

Zygote

Sporophyte

Sori containing sporangia

FERTILIZATION | MEIOSIS

Egg

Spores

Mitosis

Sperm

Germination

Gametophyte

Haploid 1n

Figure 2.24 (a) In animals, sexually reproducing adults form haploid gametes from diploid germ cells. (b) Fungi, such as black bread mold (*Rhizopus nigricans*), have haploid-dominant life cycles. (c) Plants have a life cycle that alternates between a multicellular haploid organism and a multicellular diploid organism. (credit c "fern": modification of work by Cory Zanker; credit c "gametophyte": modification of work by "VImastra"/Wikimedia Commons)

If a mutation occurs so that a fungus is no longer able to produce a minus mating type, will it still be able to reproduce?

Most fungi and algae employ a life-cycle strategy in which the multicellular "body" of the organism is haploid. During sexual reproduction, specialized haploid cells from two individuals join to form a diploid zygote. The zygote immediately undergoes meiosis to form four haploid cells called spores (**Figure 2.24b**).

The third life-cycle type, employed by some algae and all plants, is called alternation of generations. These species have both haploid and diploid multicellular organisms as part of their life cycle. The haploid multicellular plants are called **gametophytes** because they produce gametes. Meiosis is not involved in the production of gametes in this case, as the organism that produces gametes is already haploid. Fertilization between the gametes forms a diploid zygote. The zygote will undergo many rounds of mitosis and give rise to a diploid multicellular plant called a **sporophyte**. Specialized cells of the sporophyte will undergo meiosis and produce haploid spores. The spores will develop into the gametophytes (Figure 2.24c).

| Meiosis

Sexual reproduction requires **fertilization**, a union of two cells from two individual organisms. If those two cells each contain one set of chromosomes, then the resulting cell contains two sets of chromosomes. The number of sets of chromosomes in a cell is called its ploidy level. Haploid cells contain one set of chromosomes. Cells containing two sets of chromosomes are called diploid. If the reproductive cycle is to continue, the diploid cell must somehow reduce its number of chromosome sets before fertilization can occur again, or there will be a continual doubling in the number of chromosome sets in every generation. So, in addition to fertilization, sexual reproduction includes a nuclear division, known as meiosis, that reduces the number of chromosome sets.

Most animals and plants are diploid, containing two sets of chromosomes; in each **somatic cell** (the nonreproductive cells of a multicellular organism), the nucleus contains two copies of each chromosome that are referred to as homologous chromosomes. Somatic cells are sometimes referred to as "body" cells. Homologous chromosomes are matched pairs containing genes for the same traits in identical locations along their length. Diploid organisms inherit one copy of each homologous chromosome from each parent; all together, they are considered a full set of chromosomes. In animals, haploid cells containing a single copy of each homologous chromosome are found only within gametes. Gametes fuse with another haploid gamete to produce a diploid cell.

The nuclear division that forms haploid cells, which is called meiosis, is related to mitosis. As you have learned, mitosis is part of a cell reproduction cycle that results in identical daughter nuclei that are also genetically identical to the original parent nucleus. In mitosis, both the parent and the daughter nuclei contain the same number of chromosome sets—diploid for most plants and animals. Meiosis employs many of the same mechanisms as mitosis. However, the starting nucleus is always diploid and the nuclei that result at the end of a meiotic cell division are haploid. To achieve the reduction in chromosome number, meiosis consists of one round of chromosome duplication and two rounds of nuclear division. Because the events that occur during each of the division stages are analogous to the events of mitosis, the same stage names are assigned. However, because there are two rounds of division, the stages are designated with a "I" or "II." Thus, **meiosis I** is the first round of meiotic division and consists of prophase I, prometaphase I, and so on. Meiosis I reduces the number of chromosome sets from two to one. The genetic information is also mixed during this division to create unique recombinant chromosomes. **Meiosis II**, in which the second round of meiotic division takes place in a way that is similar to mitosis, includes prophase II, prometaphase II, and so on.

Interphase

Meiosis is preceded by an interphase consisting of the G_1, S, and G_2 phases, which are nearly identical to the phases preceding mitosis. The G_1 phase is the first phase of interphase and is focused on cell growth.

38

In the S phase, the DNA of the chromosomes is replicated. Finally, in the G_2 phase, the cell undergoes the final preparations for meiosis.

During DNA duplication of the S phase, each chromosome becomes composed of two identical copies (called sister chromatids) that are held together at the centromere until they are pulled apart during meiosis II. In an animal cell, the centrosomes that organize the microtubules of the meiotic spindle also replicate. This prepares the cell for the first meiotic phase.

Meiosis I

Early in prophase I, the chromosomes can be seen clearly microscopically. As the nuclear envelope begins to break down, the proteins associated with homologous chromosomes bring the pair close to each other. The tight pairing of the homologous chromosomes is called **synapsis**. In synapsis, the genes on the chromatids of the homologous chromosomes are precisely aligned with each other. An exchange of chromosome segments between non-sister homologous chromatids occurs and is called **crossing over**. This process is revealed visually after the exchange as **chiasmata** (singular = *chiasma*) (Figure 2.25).

As prophase I progresses, the close association between homologous chromosomes begins to break down, and the chromosomes continue to condense, although the homologous chromosomes remain attached to each other at chiasmata. The number of chiasmata varies with the species and the length of the chromosome. At the end of prophase I, the pairs are held together only at chiasmata (Figure 2.25) and are called **tetrads** because the four sister chromatids of each pair of homologous chromosomes are now visible.

The crossover events are the first source of genetic variation produced by meiosis. A single crossover event between homologous non-sister chromatids leads to a reciprocal exchange of equivalent DNA between a maternal chromosome and a paternal chromosome. Now, when that sister chromatid is moved into a gamete, it will carry some DNA from one parent of the individual and some DNA from the other parent. The **recombinant** sister chromatid has a combination of maternal and paternal genes that did not exist before the crossover.
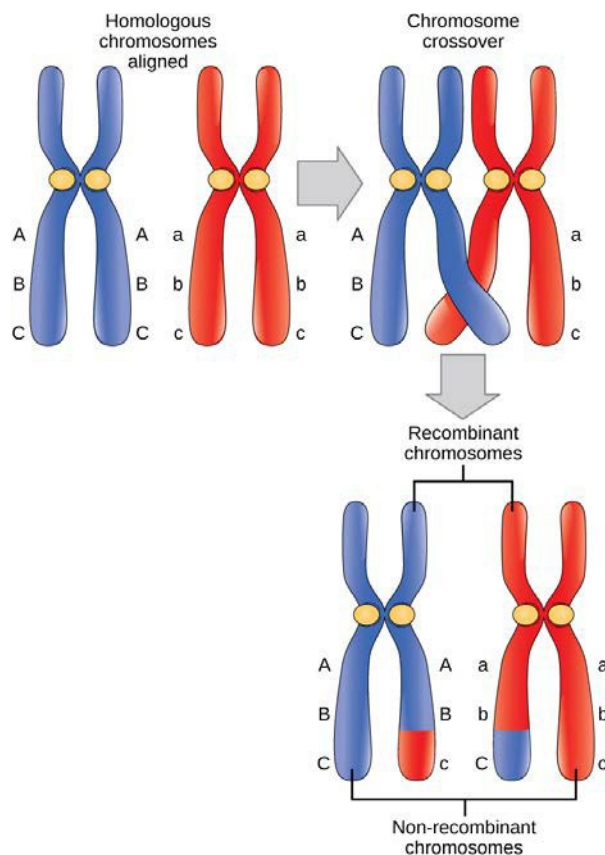
Figure 2.25 In this illustration of the effects of crossing over, the blue chromosome came from the individual's father and the red chromosome came from the individual's mother. Crossover occurs between non-sister chromatids of homologous chromosomes. The result is an exchange of genetic material between homologous chromosomes. The chromosomes that have a mixture of maternal and paternal sequence are called recombinant and the chromosomes that are completely paternal or maternal are called non-recombinant.

The key event in prometaphase I is the attachment of the spindle fiber microtubules to the kinetochore proteins at the centromeres. The microtubules assembled from centrosomes at opposite poles of the cell grow toward the middle of the cell. At the end of prometaphase I, each tetrad is attached to microtubules from both poles, with one homologous chromosome attached at one pole and the other homologous chromosome attached to the other pole. The homologous chromosomes are still held together at chiasmata. In addition, the nuclear membrane has broken down entirely.

During metaphase I, the homologous chromosomes are arranged in the center of the cell with the kinetochores facing opposite poles. The orientation of each pair of homologous chromosomes at the center of the cell is random.

This randomness, called independent assortment, is the physical basis for the generation of the second form of genetic variation in offspring. Consider that the homologous chromosomes of a sexually reproducing organism are originally inherited as two separate sets, one from each parent. Using humans as an example, one set of 23 chromosomes is present in the egg donated by the mother. The father provides the other set of 23 chromosomes in the sperm that fertilizes the egg. In metaphase I, these pairs line up at the midway point between the two poles of the cell. Because there is an equal chance that a microtubule fiber will encounter a maternally or paternally inherited chromosome, the arrangement of the tetrads at the metaphase plate is random. Any maternally inherited chromosome may face either pole. Any paternally inherited chromosome may also face either pole. The orientation of each tetrad is independent of the orientation of the other 22 tetrads.

In each cell that undergoes meiosis, the arrangement of the tetrads is different. The number of variations depends on the number of chromosomes making up a set. There are two possibilities for orientation (for each tetrad); thus, the possible number of alignments equals 2^n where n is the number of chromosomes per set. Humans have 23 chromosome pairs, which results in over eight million (2^{23}) possibilities. This number does not include the variability previously created in the sister chromatids by crossover. Given these two mechanisms, it is highly unlikely that any two haploid cells resulting from meiosis will have the same genetic composition (Figure 2.26).

To summarize the genetic consequences of meiosis I: the maternal and paternal genes are recombined by crossover events occurring on each homologous pair during prophase I; in addition, the random assortment of tetrads at metaphase produces a unique combination of maternal and paternal chromosomes that will make their way into the gametes.
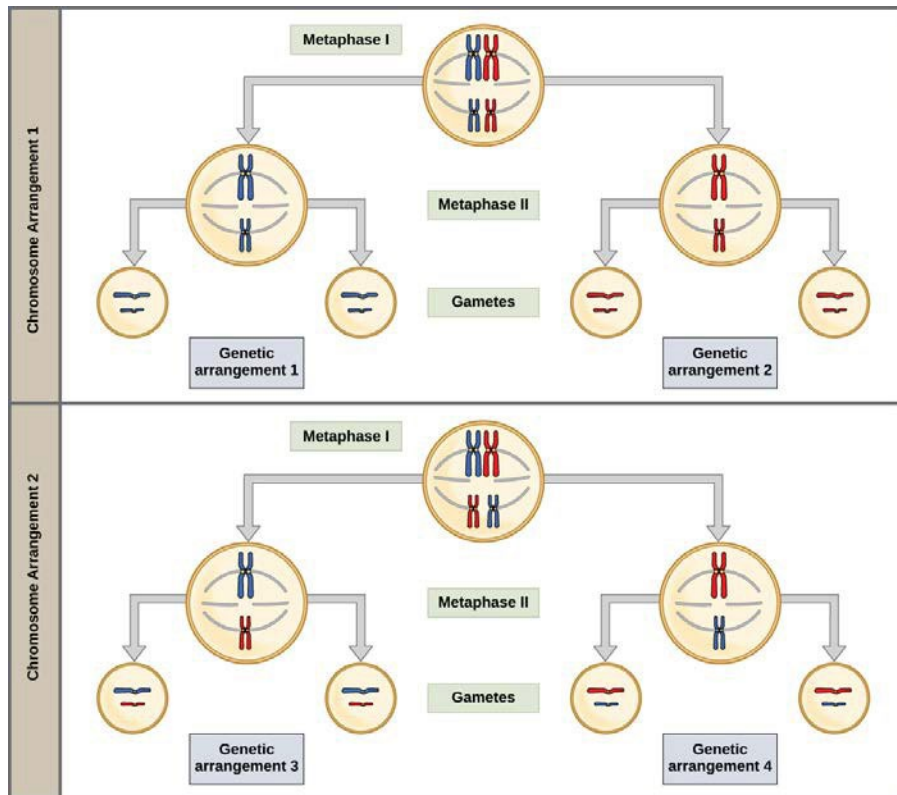
Figure 2.26 To demonstrate random, independent assortment at metaphase I, consider a cell with $n = 2$. In this case, there are two possible arrangements at the equatorial plane in metaphase I, as shown in the upper cell of each panel. These two possible orientations lead to the production of genetically different gametes. With more chromosomes, the number of possible arrangements increases dramatically.

In anaphase I, the spindle fibers pull the linked chromosomes apart. The sister chromatids remain tightly bound together at the centromere. It is the chiasma connections that are broken in anaphase I as the fibers attached to the fused kinetochores pull the homologous chromosomes apart (Figure 2.27).

In telophase I, the separated chromosomes arrive at opposite poles. The remainder of the typical telophase events may or may not occur depending on the species. In some organisms, the chromosomes decondense and nuclear envelopes form around the chromatids in telophase I.

Cytokinesis, the physical separation of the cytoplasmic components into two daughter cells, occurs without reformation of the nuclei in other organisms. In nearly all species, cytokinesis separates the cell contents by either a cleavage furrow (in animals and some fungi), or a cell plate that will ultimately lead to formation of cell walls that separate the two daughter cells (in plants). At each pole, there is just one member of each pair of the homologous chromosomes, so only one full set of the chromosomes is present. This is why the cells are considered haploid—there is only one chromosome set, even though there are duplicate copies of the set because each homolog still consists of two sister chromatids that are still attached to each other. However, although the sister chromatids were once duplicates of the same chromosome, they are no longer identical at this stage because of crossovers.

Review the process of meiosis, observing how chromosomes align and migrate, at **this site (http://openstaxcollege.org/l/animal_meiosis2)** .

Meiosis II

In meiosis II, the connected sister chromatids remaining in the haploid cells from meiosis I will be split to form four haploid cells. In some species, cells enter a brief interphase, or **interkinesis**, that lacks an S phase, before entering meiosis II. Chromosomes are not duplicated during interkinesis. The two cells produced in meiosis I go through the events of meiosis II in synchrony. Overall, meiosis II resembles the mitotic division of a haploid cell.

In prophase II, if the chromosomes decondensed in telophase I, they condense again. If nuclear envelopes were formed, they fragment into vesicles. The centrosomes duplicated during interkinesis move away from each other toward opposite poles, and new spindles are formed. In prometaphase II, the nuclear envelopes are completely broken down, and the spindle is fully formed. Each sister chromatid forms an individual kinetochore that attaches to microtubules from opposite poles. In metaphase II, the sister chromatids are maximally condensed and aligned at the center of the cell. In anaphase II, the sister chromatids are pulled apart by the spindle fibers and move toward opposite poles.
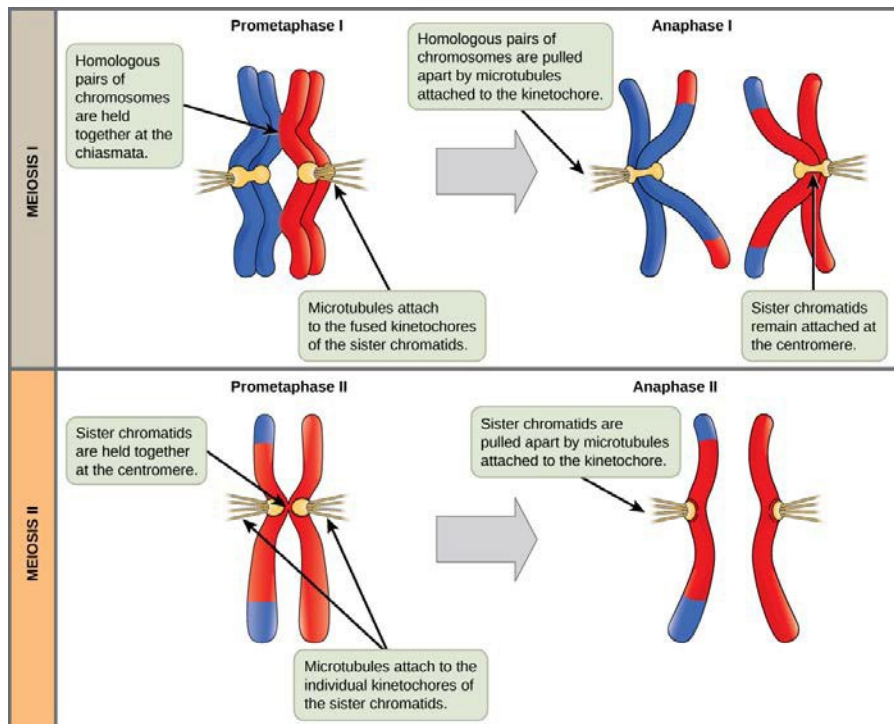
Figure 2.27 In prometaphase I, microtubules attach to the fused kinetochores of homologous chromosomes. In anaphase I, the homologous chromosomes are separated. In prometaphase II, microtubules attach to individual kinetochores of sister chromatids. In anaphase II, the sister chromatids are separated.

In telophase II, the chromosomes arrive at opposite poles and begin to decondense. Nuclear envelopes form around the chromosomes. Cytokinesis separates the two cells into four genetically unique haploid cells. At this point, the nuclei in the newly produced cells are both haploid and have only one copy of the single set of chromosomes. The cells produced are genetically unique because of the random assortment of paternal and maternal homologs and because of the recombination of maternal and paternal segments of chromosomes—with their sets of genes—that occurs during crossover.

Comparing Meiosis and Mitosis

Mitosis and meiosis, which are both forms of division of the nucleus in eukaryotic cells, share some similarities, but also exhibit distinct differences that lead to their very different outcomes. Mitosis is a single nuclear division that results in two nuclei, usually partitioned into two new cells. The nuclei resulting from a mitotic division are genetically identical to the original. They have the same number of sets of chromosomes: one in the case of haploid cells, and two in the case of diploid cells. On the other hand, meiosis is two nuclear divisions that result in four nuclei, usually partitioned into four new cells. The nuclei resulting from meiosis are never genetically identical, and they contain one chromosome set only—this is half the number of the original cell, which was diploid (Figure 2.28).
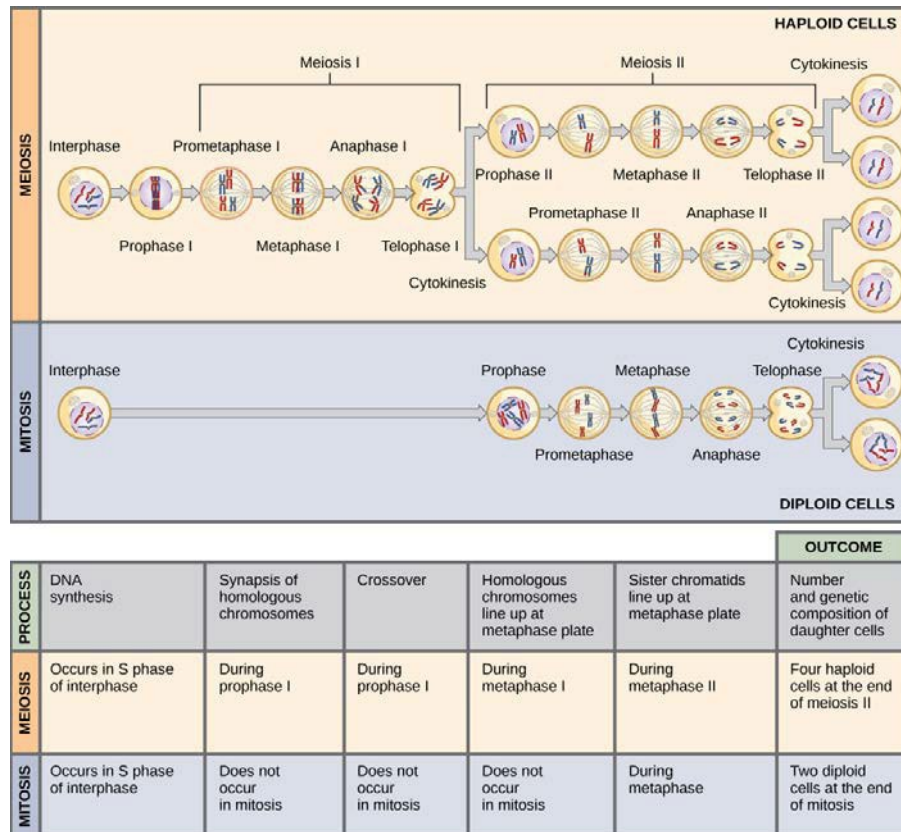
The differences in the outcomes of meiosis and mitosis occur because of differences in the behavior of the chromosomes during each process. Most of these differences in the processes occur in meiosis I,

which is a very different nuclear division than mitosis. In meiosis I, the homologous chromosome pairs become associated with each other, are bound together, experience chiasmata and crossover between sister chromatids, and line up along the metaphase plate in tetrads with spindle fibers from opposite spindle poles attached to each kinetochore of a homolog in a tetrad. All of these events occur only in meiosis I, never in mitosis.

Homologous chromosomes move to opposite poles during meiosis I so the number of sets of chromosomes in each nucleus-to-be is reduced from two to one. For this reason, meiosis I is referred to as a **reduction division**. There is no such reduction in ploidy level in mitosis.

Meiosis II is much more analogous to a mitotic division. In this case, duplicated chromosomes (only one set of them) line up at the center of the cell with divided kinetochores attached to spindle fibers from opposite poles. During anaphase II, as in mitotic anaphase, the kinetochores divide and one sister chromatid is pulled to one pole and the other sister chromatid is pulled to the other pole. If it were not for the fact that there had been crossovers, the two products of each meiosis II division would be identical as in mitosis; instead, they are different because there has always been at least one crossover per chromosome. Meiosis II is not a reduction division because, although there are fewer copies of the genome in the resulting cells, there is still one set of chromosomes, as there was at the end of meiosis I.

Cells produced by mitosis will function in different parts of the body as a part of growth or replacing dead or damaged cells. They may even be involved in asexual reproduction in some organisms. Cells produced by meiosis in a diploid-dominant organism such as an animal will only participate in sexual reproduction.

PROCESS	DNA synthesis	Synapsis of homologous chromosomes	Crossover	Homologous chromosomes line up at metaphase plate	Sister chromatids line up at metaphase plate	OUTCOME Number and genetic composition of daughter cells
MEIOSIS	Occurs in S phase of interphase	During prophase I	During prophase I	During metaphase I	During metaphase II	Four haploid cells at the end of meiosis II
MITOSIS	Occurs in S phase of interphase	Does not occur in mitosis	Does not occur in mitosis	Does not occur in mitosis	During metaphase	Two diploid cells at the end of mitosis

Figure 2.28 Meiosis and mitosis are both preceded by one round of DNA replication; however, meiosis includes two nuclear divisions. The four daughter cells resulting from meiosis are haploid and genetically distinct. The daughter cells resulting from mitosis are diploid and identical to the parent cell.

For an animation comparing mitosis and meiosis, go to **this website (http://openstaxcollege.org/l/ how_cells_dvid2)** .

Photosynthesis

No matter how complex or advanced a machine, such as the latest cellular phone, the device cannot function without energy. Living things, similar to machines, have many complex components; they too cannot do anything without energy, which is why humans and all other organisms must "eat" in some form or another. That may be common knowledge, but how many people realize that every bite of every meal ingested depends on the process of photosynthesis?

| Overview of Photosynthesis

All living organisms on earth consist of one or more cells. Each cell runs on the chemical energy found mainly in carbohydrate molecules (food), and the majority of these molecules are produced by one process: photosynthesis. Through photosynthesis, certain organisms convert solar energy (sunlight) into chemical energy, which is then used to build carbohydrate molecules. The energy used to hold these molecules together is released when an organism breaks down food. Cells then use this energy to perform work, such as cellular respiration.

The energy that is harnessed from photosynthesis enters the ecosystems of our planet continuously and is transferred from one organism to another. Therefore, directly or indirectly, the process of photosynthesis provides most of the energy required by living things on earth.

Photosynthesis also results in the release of oxygen into the atmosphere. In short, to eat and breathe, humans depend almost entirely on the organisms that carry out photosynthesis.

Click the following **link (http://openstaxcollege.org/l/photosynthesis2)** to learn more about photosynthesis.

Solar Dependence and Food Production

Some organisms can carry out photosynthesis, whereas others cannot. An **autotroph** is an organism that can produce its own food. The Greek roots of the word *autotroph* mean "self" (*auto*) "feeder" (*troph*). Plants are the best-known autotrophs, but others exist, including certain types of bacteria and algae (Figure 2.29). Oceanic algae contribute enormous quantities of food and oxygen to global food chains. Plants are also **photoautotrophs**, a type of autotroph that uses sunlight and carbon from carbon dioxide to synthesize chemical energy in the form of carbohydrates. All organisms carrying out photosynthesis require sunlight.

(a)   (b)   (c)

Figure 2.29 (a) Plants, (b) algae, and (c) certain bacteria, called cyanobacteria, are photoautotrophs that can carry out photosynthesis. Algae can grow over enormous areas in water, at times completely covering the surface. (credit a: Steve Hillebrand, U.S. Fish and Wildlife Service; credit b: "eutrophication&hypoxia"/Flickr; credit c: NASA; scale-bar data from Matt Russell)

Heterotrophs are organisms incapable of photosynthesis that must therefore obtain energy and carbon from food by consuming other organisms. The Greek roots of the word *heterotroph* mean "other" (*hetero*) "feeder" (*troph*), meaning that their food comes from other organisms. Even if the food organism is another animal, this food traces its origins back to autotrophs and the process of photosynthesis. Humans are heterotrophs, as are all animals. Heterotrophs depend on autotrophs, either directly or indirectly. Deer and wolves are heterotrophs. A deer obtains energy by eating plants. A wolf eating a deer obtains energy that originally came from the plants eaten by that deer. The energy in the plant came from photosynthesis, and therefore it is the only autotroph in this example (Figure 2.30). Using this reasoning, all food eaten by humans also links back to autotrophs that carry out photosynthesis.

45

Figure 2.30 The energy stored in carbohydrate molecules from photosynthesis passes through the food chain. The predator that eats these deer is getting energy that originated in the photosynthetic vegetation that the deer consumed. (credit: Steve VanRiper, U.S. Fish and Wildlife Service)

# biology IN ACTION

Photosynthesis at the Grocery Store

Figure 2.31 Photosynthesis is the origin of the products that comprise the main elements of the human diet. (credit: Associação Brasileira de Supermercados)

Major grocery stores in the United States are organized into departments, such as dairy, meats, produce, bread, cereals, and so forth. Each aisle contains hundreds, if not thousands, of different products for customers to buy and consume (Figure 2.31).

Although there is a large variety, each item links back to photosynthesis. Meats and dairy products link to photosynthesis because the animals were fed plant-based foods. The breads, cereals, and pastas come largely from grains, which are the seeds of photosynthetic plants. What about desserts and drinks? All of these products contain sugar—the basic carbohydrate molecule produced directly from photosynthesis. The photosynthesis connection applies to every meal and every food a person consumes.

Main Structures and Summary of Photosynthesis

Photosynthesis requires sunlight, carbon dioxide, and water as starting reactants (Figure 2.32). After the process is complete, photosynthesis releases oxygen and produces carbohydrate molecules, most commonly glucose. These sugar molecules contain the energy that living things need to survive.
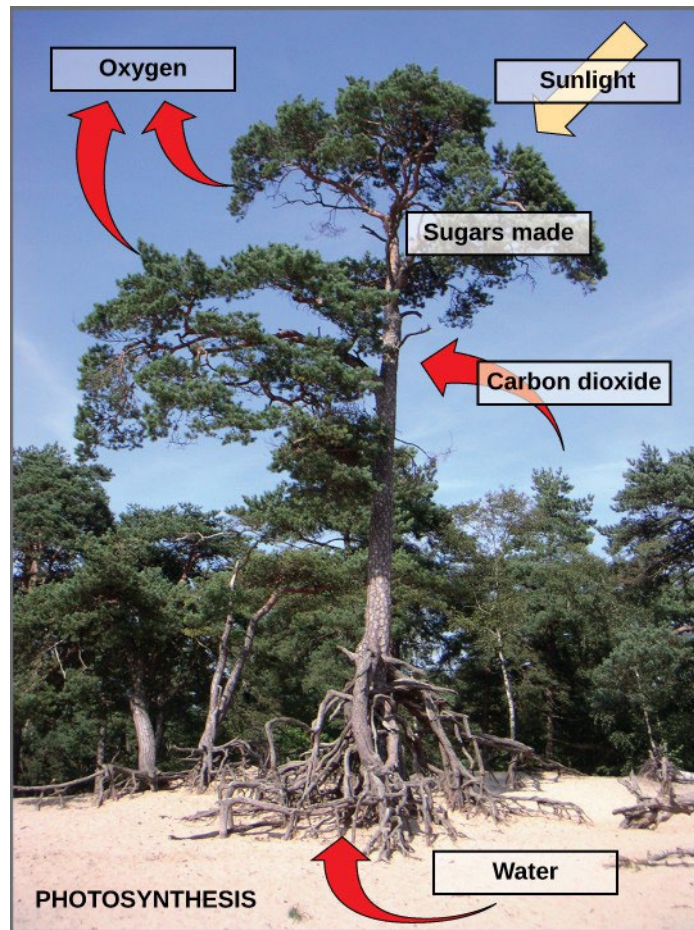
Figure 2.32 Photosynthesis uses solar energy, carbon dioxide, and water to release oxygen and to produce energy-storing sugar molecules.

The complex reactions of photosynthesis can be summarized by the chemical equation shown in **Figure 2.33**.

Photosynthesis Equation			
Carbon dioxide +	Water SUNLIGHT→	Sugar +	Oxygen
$6CO_2$	$6H_2O$	$C_6H_{12}O_6$	$6O_2$
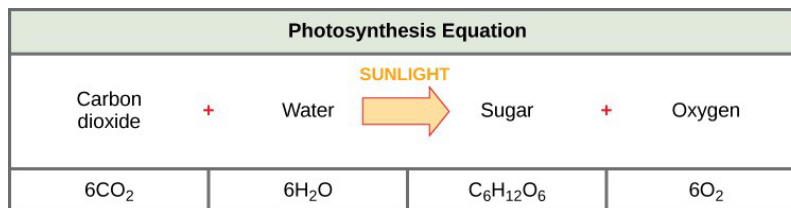
Figure 2.33 The process of photosynthesis can be represented by an equation, wherein carbon dioxide and water produce sugar and oxygen using energy from sunlight.

Although the equation looks simple, the many steps that take place during photosynthesis are actually quite complex, as in the way that the reaction summarizing cellular respiration represented many individual reactions. Before learning the details of how photoautotrophs turn sunlight into food, it is important to become familiar with the physical structures involved.

In plants, photosynthesis takes place primarily in leaves, which consist of many layers of cells and have differentiated top and bottom sides. The process of photosynthesis occurs not on the surface layers of the leaf, but rather in a middle layer called the **mesophyll** (Figure 2.34). The gas exchange of carbon dioxide and oxygen occurs through small, regulated openings called **stomata**.

In all autotrophic eukaryotes, photosynthesis takes place inside an organelle called a **chloroplast**. In plants, chloroplast-containing cells exist in the mesophyll. Chloroplasts have a double (inner and outer) membrane. Within the chloroplast is a third membrane that forms stacked, disc-shaped structures called **thylakoids**. Embedded in the thylakoid membrane are molecules of **chlorophyll**, a **pigment** (a molecule that absorbs light) through which the entire process of photosynthesis begins. Chlorophyll is responsible for the green color of plants. The thylakoid membrane encloses an internal space called the thylakoid space. Other types of pigments are also involved in photosynthesis, but chlorophyll is by far the most

47

important. As shown in **Figure 5.7**, a stack of thylakoids is called a **granum**, and the space surrounding the granum is called **stroma** (not to be confused with stomata, the openings on the leaves).
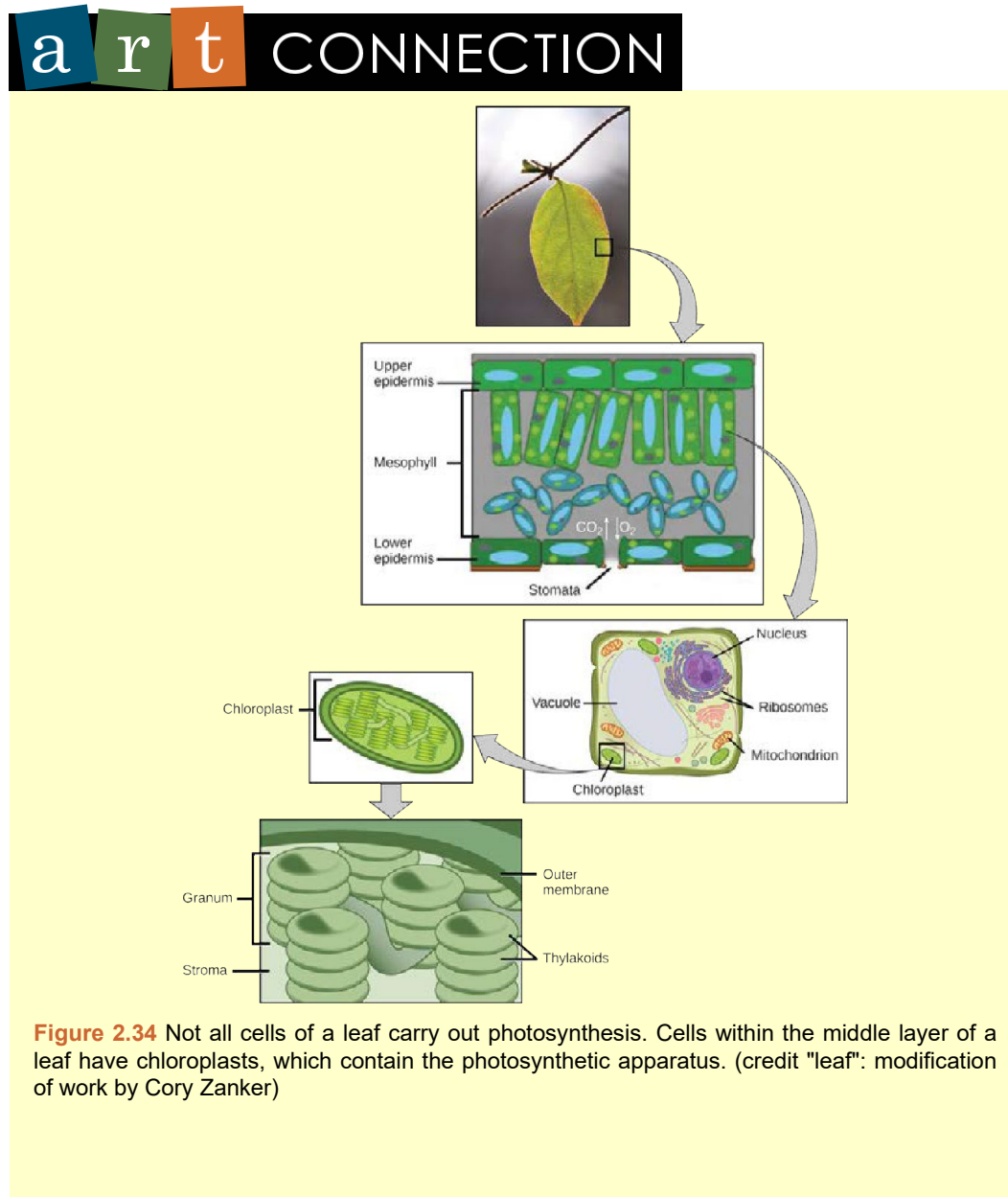
Figure 2.34 Not all cells of a leaf carry out photosynthesis. Cells within the middle layer of a leaf have chloroplasts, which contain the photosynthetic apparatus. (credit "leaf": modification of work by Cory Zanker)

The Two Parts of Photosynthesis

Photosynthesis takes place in two stages: the light-dependent reactions and the Calvin cycle. In the **light-dependent reactions**, which take place at the thylakoid membrane, chlorophyll absorbs energy from sunlight and then converts it into chemical energy with the use of water. The light-dependent reactions release oxygen from the hydrolysis of water as a byproduct. In the Calvin cycle, which takes place in the stroma, the chemical energy derived from the light-dependent reactions drives both the capture of carbon in carbon dioxide molecules and the subsequent assembly of sugar molecules. The two reactions use carrier molecules to transport the energy from one to the other. The carriers that move energy from the light-dependent reactions to the Calvin cycle reactions can be thought of as "full" because they bring energy. After the energy is released, the "empty" energy carriers return to the light-dependent reactions to obtain more energy.

In the light-dependent reactions (Figure 2.35), energy absorbed by sunlight is stored by two types of energy-carrier molecules: ATP and NADPH. The energy that these molecules carry is stored in a bond that holds a single atom to the molecule. For ATP, it is a phosphate atom, and for NADPH, it is a hydrogen atom.

After the energy from the sun is converted and packaged into ATP and NADPH, the cell has the fuel needed to build food in the form of carbohydrate molecules. The carbohydrate molecules made will have a backbone of carbon atoms. Where does the carbon come from? The carbon atoms used to build carbohydrate molecules comes from carbon dioxide, the gas that animals exhale with each breath. The **Calvin cycle** is the term used for the reactions of photosynthesis that use the energy stored by the light-dependent reactions to form glucose and other carbohydrate molecules.

In plants, carbon dioxide (CO_2) enters the chloroplast through the stomata and diffuses into the stroma of the chloroplast—the site of the Calvin cycle reactions where sugar is synthesized.
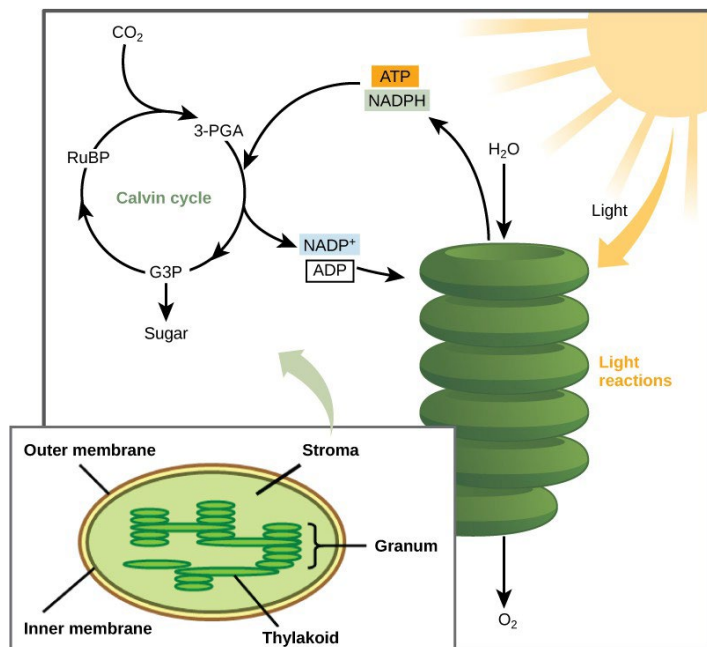

Figure 2.35 Light-dependent reactions harness energy from the sun to produce ATP and NADPH. These energy-carrying molecules travel into the stroma where the Calvin cycle reactions take place.

The Energy Cycle

Living things access energy by breaking down carbohydrate molecules. However, if plants make carbohydrate molecules, why would they need to break them down? Carbohydrates are storage molecules for energy in all living things. Although energy can be stored in molecules like ATP, carbohydrates are much more stable and efficient reservoirs for chemical energy. Photosynthetic organisms also carry out the reactions of respiration to harvest the energy that they have stored in carbohydrates, for example, plants have mitochondria in addition to chloroplasts.

You may have noticed that the overall reaction for photosynthesis:

$$6CO_2 + 6H_2O \longrightarrow C_6H_{12}O_6 + 6O_2$$
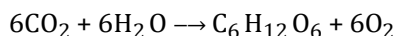
is the reverse of the overall reaction for cellular respiration:

$$6O_2 + C_6H_{12}O_6 \longrightarrow 6CO_2 + 6H_2O$$
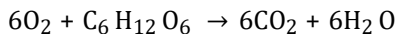
Photosynthesis produces oxygen as a byproduct, and respiration produces carbon dioxide as a byproduct.

In nature, there is no such thing as waste. Every single atom of matter is conserved, recycling indefinitely. Substances change form or move from one type of molecule to another, but never disappear (Figure 2.36).

CO_2 is no more a form of waste produced by respiration than oxygen is a waste product of photosynthesis. Both are byproducts of reactions that move on to other reactions. Photosynthesis absorbs energy to build carbohydrates in chloroplasts, and aerobic cellular respiration releases energy by using oxygen to break down carbohydrates. Both organelles use electron transport chains to generate the energy necessary to drive other reactions. Photosynthesis and cellular respiration function in a biological cycle, allowing organisms to access life-sustaining energy that originates millions of miles away in a star.
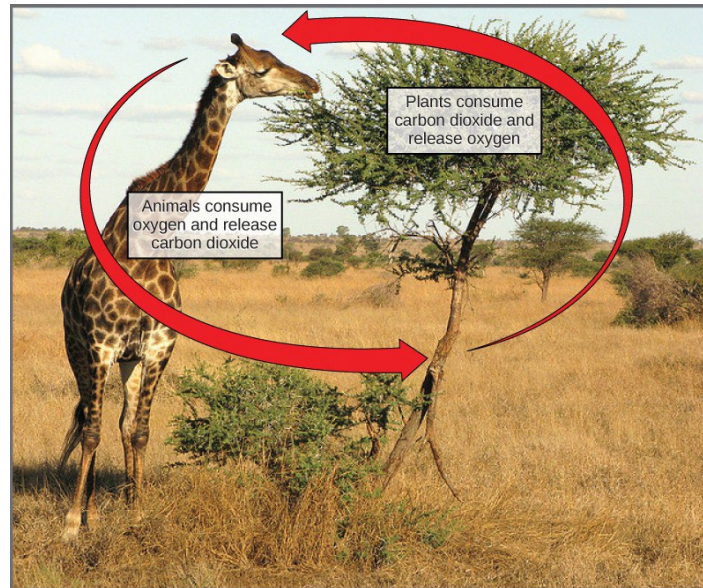
Figure 2.36 In the carbon cycle, the reactions of photosynthesis and cellular respiration share reciprocal reactants and products. (credit: modification of work by Stuart Bassil)

Glycolysis

You have read that nearly all of the energy used by living things comes to them in the bonds of the sugar, glucose. **Glycolysis** is the first step in the breakdown of glucose to extract energy for cell metabolism. Many living organisms carry out glycolysis as part of their metabolism. Glycolysis takes place in the cytoplasm of most prokaryotic and all eukaryotic cells.

Glycolysis begins with the six-carbon, ring-shaped structure of a single glucose molecule and ends with two molecules of a three-carbon sugar called pyruvate. Glycolysis consists of two distinct phases. In the first part of the glycolysis pathway, energy is used to make adjustments so that the six-carbon sugar molecule can be split evenly into two three-carbon pyruvate molecules. In the second part of glycolysis, ATP and nicotinamide-adenine dinucleotide (NADH) are produced (Figure 2.37).

If the cell cannot catabolize the pyruvate molecules further, it will harvest only two ATP molecules from one molecule of glucose. For example, mature mammalian red blood cells are only capable of glycolysis, which is their sole source of ATP. If glycolysis is interrupted, these cells would eventually die.
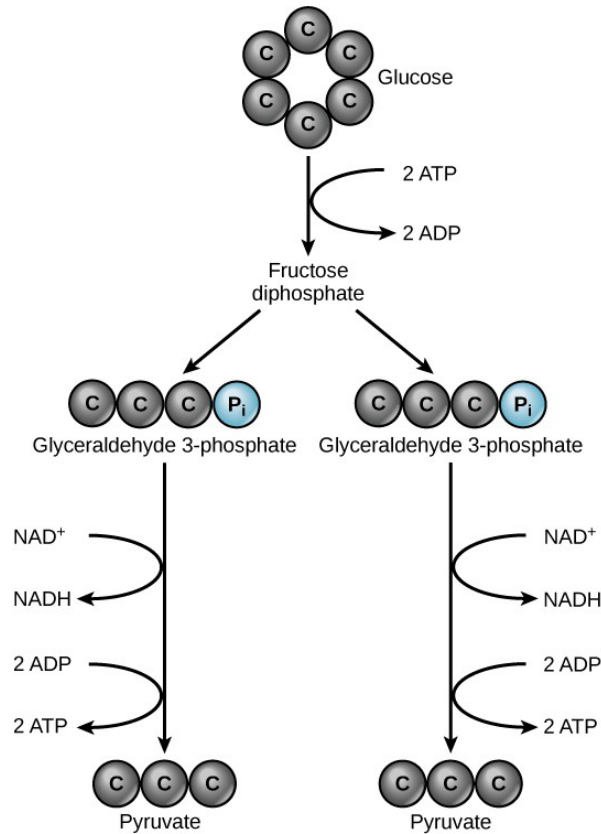
Figure 2.37 In glycolysis, a glucose molecule is converted into two pyruvate molecules.

The Citric Acid Cycle

In eukaryotic cells, the pyruvate molecules produced at the end of glycolysis are transported into mitochondria, which are sites of cellular respiration. If oxygen is available, aerobic respiration will go forward. In mitochondria, pyruvate will be transformed into a two-carbon acetyl group (by removing a molecule of carbon dioxide) that will be picked up by a carrier compound called coenzyme A (CoA), which is made from vitamin B_5. The resulting compound is called **acetyl CoA**. (Figure 2.38). Acetyl CoA can be used in a variety of ways by the cell, but its major function is to deliver the acetyl group derived from pyruvate to the next pathway in glucose catabolism.

51
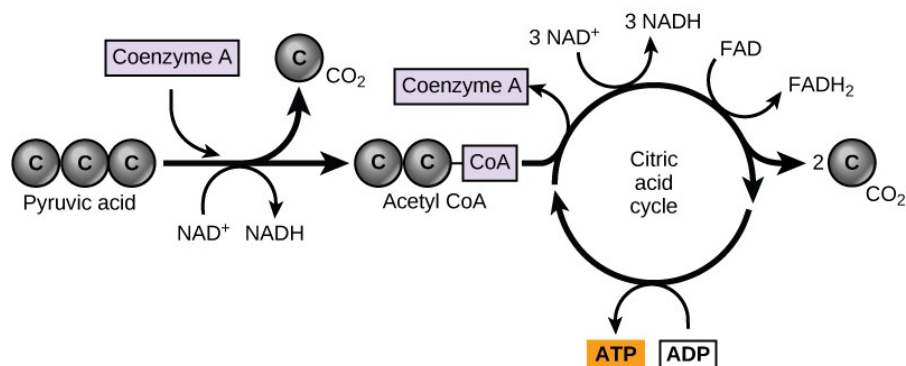
Figure 2.38 Pyruvate is converted into acetyl-CoA before entering the citric acid cycle.

Like the conversion of pyruvate to acetyl CoA, the **citric acid cycle** in eukaryotic cells takes place in the matrix of the mitochondria. Unlike glycolysis, the citric acid cycle is a closed loop: The last part of the pathway regenerates the compound used in the first step. The eight steps of the cycle are a series of chemical reactions that produces two carbon dioxide molecules, one ATP molecule (or an equivalent), and reduced forms (NADH and $FADH_2$) of NAD^+ and FAD^+, important coenzymes in the cell. Part of this is considered an aerobic pathway (oxygen-requiring) because the NADH and $FADH_2$ produced must transfer their electrons to the next pathway in the system, which will use oxygen. If oxygen is not present, this transfer does not occur.

Two carbon atoms come into the citric acid cycle from each acetyl group. Two carbon dioxide molecules are released on each turn of the cycle; however, these do not contain the same carbon atoms contributed by the acetyl group on that turn of the pathway. The two acetyl-carbon atoms will eventually be released on later turns of the cycle; in this way, all six carbon atoms from the original glucose molecule will be eventually released as carbon dioxide. It takes two turns of the cycle to process the equivalent of one glucose molecule. Each turn of the cycle forms three high-energy NADH molecules and one high-energy $FADH_2$ molecule. These high-energy carriers will connect with the last portion of aerobic respiration to produce ATP molecules. One ATP (or an equivalent) is also made in each cycle. Several of the intermediate compounds in the citric acid cycle can be used in synthesizing non-essential amino acids; therefore, the cycle is both anabolic and catabolic.

Oxidative Phosphorylation

You have just read about two pathways in glucose catabolism—glycolysis and the citric acid cycle—that generate ATP. Most of the ATP generated during the aerobic catabolism of glucose, however, is not generated directly from these pathways. Rather, it derives from a process that begins with passing electrons through a series of chemical reactions to a final electron acceptor, oxygen. These reactions take place in specialized protein complexes located in the inner membrane of the mitochondria of eukaryotic organisms and on the inner part of the cell membrane of prokaryotic organisms. The energy of the electrons is harvested and used to generate a electrochemical gradient across the inner mitochondrial membrane. The potential energy of this gradient is used to generate ATP. The entirety of this process is called **oxidative phosphorylation**.

The electron transport chain (Figure 2.39**a**) is the last component of aerobic respiration and is the only part of metabolism that uses atmospheric oxygen. Oxygen continuously diffuses into plants for this purpose. In animals, oxygen enters the body through the respiratory system. Electron transport is a series of chemical reactions that resembles a bucket brigade in that electrons are passed rapidly from one component to the next, to the endpoint of the chain where oxygen is the final electron acceptor and water is produced. There are four complexes composed of proteins, labeled I through IV in Figure 2.39**c**, and the aggregation of these four complexes, together with associated mobile, accessory electron carriers, is called the **electron transport chain**. The electron transport chain is present in multiple copies in the inner mitochondrial membrane of eukaryotes and in the plasma membrane of prokaryotes. In each transfer of an electron through the electron transport chain, the electron loses energy, but with some transfers, the energy is stored as potential energy by using it to pump hydrogen ions across the inner mitochondrial membrane into the intermembrane space, creating an electrochemical gradient.

52

art CONNECTION
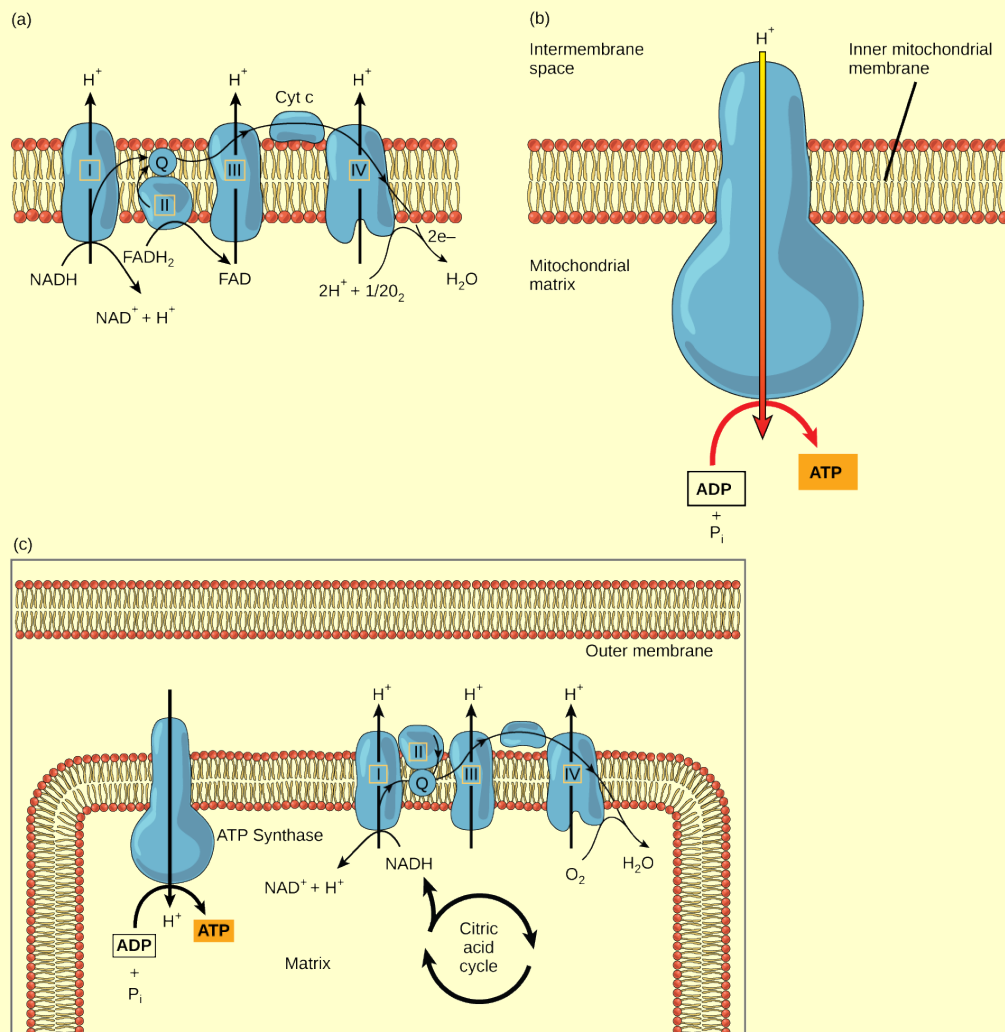
Figure 2.39 (a) The electron transport chain is a set of molecules that supports a series of oxidation-reduction reactions. (b) ATP synthase is a complex, molecular machine that uses an H^+ gradient to regenerate ATP from ADP. (c) Chemiosmosis relies on the potential energy provided by the H^+ gradient across the membrane.

Electrons from NADH and $FADH_2$ are passed to protein complexes in the electron transport chain. As they are passed from one complex to another (there are a total of four), the electrons lose energy, and some of that energy is used to pump hydrogen ions from the mitochondrial matrix into the intermembrane space. In the fourth protein complex, the electrons are accepted by oxygen, the terminal acceptor. The oxygen with its extra electrons then combines with two hydrogen ions, further enhancing the electrochemical gradient, to form water. If there were no oxygen present in the mitochondrion, the electrons could not be removed from the system, and the entire electron transport chain would back up and stop. The mitochondria would be unable to generate new ATP in this way, and the cell would ultimately die from lack of energy. This is the reason we must breathe to draw in new oxygen.

In the electron transport chain, the free energy from the series of reactions just described is used to pump hydrogen ions across the membrane. The uneven distribution of H^+ ions across the membrane establishes

an electrochemical gradient, owing to the H$^+$ ions' positive charge and their higher concentration on one side of the membrane.

Hydrogen ions diffuse through the inner membrane through an integral membrane protein called **ATP synthase** (Figure 2.39**b**). This complex protein acts as a tiny generator, turned by the force of the hydrogen ions diffusing through it, down their electrochemical gradient from the intermembrane space, where there are many mutually repelling hydrogen ions to the matrix, where there are few. The turning of the parts of this molecular machine regenerate ATP from ADP. This flow of hydrogen ions across the membrane through ATP synthase is called **chemiosmosis**.

Chemiosmosis (Figure 2.39**c**) is used to generate 90 percent of the ATP made during aerobic glucose catabolism. The result of the reactions is the production of ATP from the energy of the electrons removed from hydrogen atoms. These atoms were originally part of a glucose molecule. At the end of the electron transport system, the electrons are used to reduce an oxygen molecule to oxygen ions. The extra electrons on the oxygen ions attract hydrogen ions (protons) from the surrounding medium, and water is formed. The electron transport chain and the production of ATP through chemiosmosis are collectively called oxidative phosphorylation.

ATP Yield

The number of ATP molecules generated from the catabolism of glucose varies. For example, the number of hydrogen ions that the electron transport chain complexes can pump through the membrane varies between species. Another source of variance stems from the shuttle of electrons across the mitochondrial membrane. The NADH generated from glycolysis cannot easily enter mitochondria. Thus, electrons are picked up on the inside of the mitochondria by either NAD$^+$ or FAD$^+$. Fewer ATP molecules are generated when FAD$^+$ acts as a carrier. NAD$^+$ is used as the electron transporter in the liver and FAD$^+$ in the brain, so ATP yield depends on the tissue being considered.

Another factor that affects the yield of ATP molecules generated from glucose is that intermediate compounds in these pathways are used for other purposes. Glucose catabolism connects with the pathways that build or break down all other biochemical compounds in cells, and the result is somewhat messier than the ideal situations described thus far. For example, sugars other than glucose are fed into the glycolytic pathway for energy extraction. Other molecules that would otherwise be used to harvest energy in glycolysis or the citric acid cycle may be removed to form nucleic acids, amino acids, lipids, or other compounds. Overall, in living systems, these pathways of glucose catabolism extract about 34 percent of the energy contained in glucose.

careers IN ACTION

Mitochondrial Disease Physician

What happens when the critical reactions of cellular respiration do not proceed correctly? Mitochondrial diseases are genetic disorders of metabolism. Mitochondrial disorders can arise from mutations in nuclear or mitochondrial DNA, and they result in the production of less energy than is normal in body cells. Symptoms of mitochondrial diseases can include muscle weakness, lack of coordination, stroke-like episodes, and loss of vision and hearing. Most affected people are diagnosed in childhood, although there are some adult-onset diseases. Identifying and treating mitochondrial disorders is a specialized medical field. The educational preparation for this profession requires a college education, followed by medical school with a specialization in medical genetics. Medical geneticists can be board certified by the American Board of Medical Genetics and go on to become associated with professional organizations devoted to the study of mitochondrial disease, such as the Mitochondrial Medicine Society and the Society for Inherited Metabolic Disease.

| Fermentation

In aerobic respiration, the final electron acceptor is an oxygen molecule, O_2. If aerobic respiration occurs, then ATP will be produced using the energy of the high-energy electrons carried by NADH or $FADH_2$ to the electron transport chain. If aerobic respiration does not occur, NADH must be reoxidized to NAD^+ for reuse as an electron carrier for glycolysis to continue. How is this done? Some living systems use an organic molecule as the final electron acceptor. Processes that use an organic molecule to regenerate NAD^+ from NADH are collectively referred to as **fermentation**. In contrast, some living systems use an inorganic molecule as a final electron acceptor; both methods are a type of **anaerobic cellular respiration**. Anaerobic respiration enables organisms to convert energy for their use in the absence of oxygen.
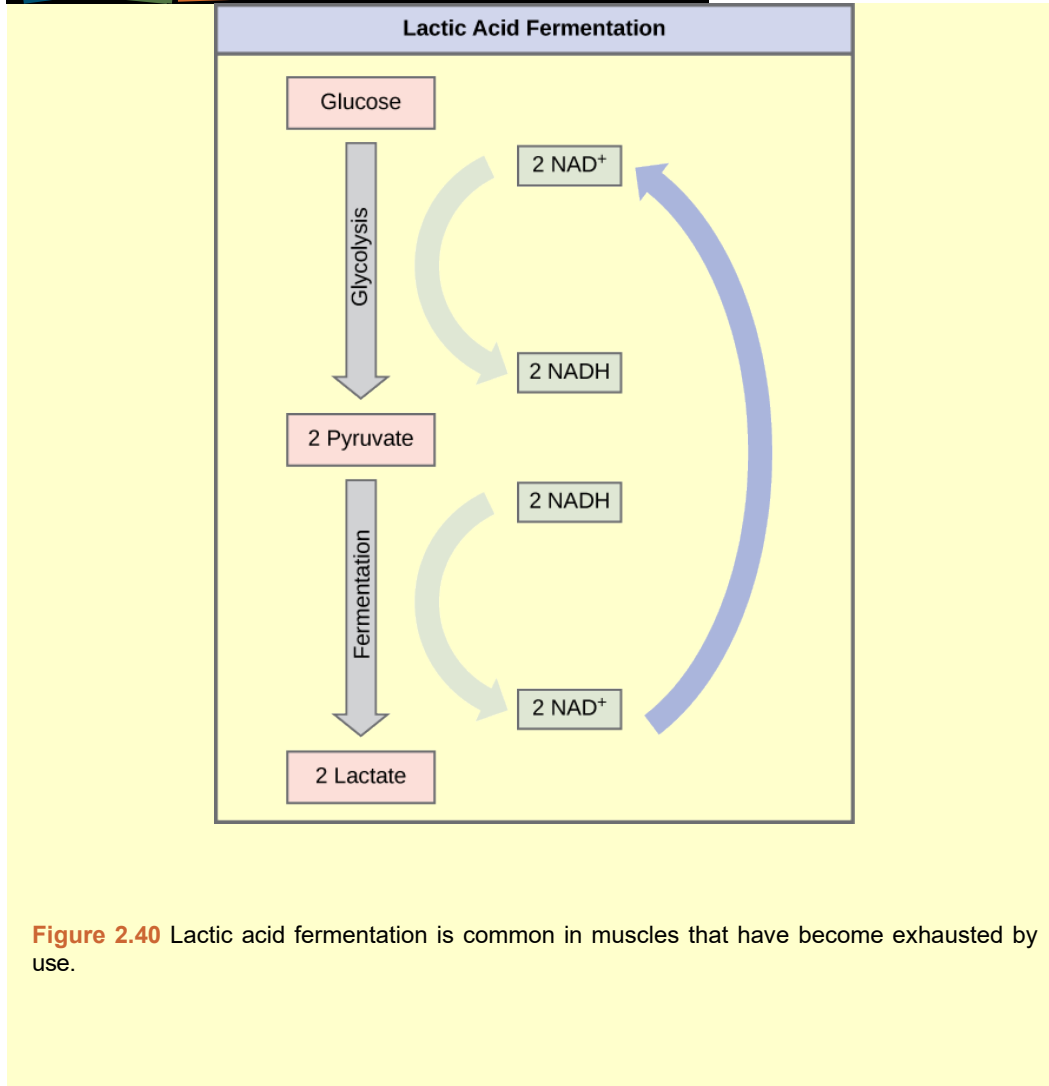
Lactic Acid Fermentation

The fermentation method used by animals and some bacteria like those in yogurt is lactic acid fermentation (Figure 2.40). This occurs routinely in mammalian red blood cells and in skeletal muscle that has insufficient oxygen supply to allow aerobic respiration to continue (that is, in muscles used to the point of fatigue). In muscles, lactic acid produced by fermentation must be removed by the blood circulation and brought to the liver for further metabolism. The chemical reaction of lactic acid fermentation is the following:

$$\text{Pyruvic acid } + \text{NADH} \leftrightarrow \text{lactic acid} + \text{NAD}^+$$
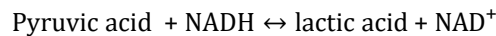
The enzyme that catalyzes this reaction is lactate dehydrogenase. The reaction can proceed in either direction, but the left-to-right reaction is inhibited by acidic conditions. This lactic acid build-up causes muscle stiffness and fatigue. Once the lactic acid has been removed from the muscle and is circulated to the liver, it can be converted back to pyruvic acid and further catabolized for energy.

55

art CONNECTION

Lactic Acid Fermentation

Glucose

Glycolysis

2 NAD$^+$

2 NADH

2 Pyruvate

2 NADH

Fermentation

2 NAD$^+$

2 Lactate

Figure 2.40 Lactic acid fermentation is common in muscles that have become exhausted by use.

Alcohol Fermentation

Another familiar fermentation process is alcohol fermentation (Figure 2.40), which produces ethanol, an alcohol. The alcohol fermentation reaction is the following:

Pyruvic acid ⟶ CO_2 + Acetaldehyde

Acetaldehyde ⟶ Ethanol

NADH   NAD$^+$
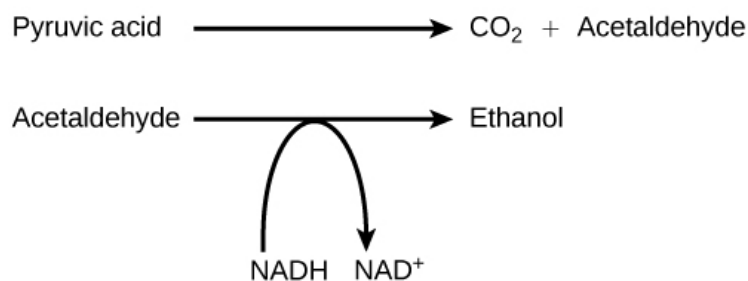
Figure 2.41 The reaction resulting in alcohol fermentation is shown.

In the first reaction, a carboxyl group is removed from pyruvic acid, releasing carbon dioxide as a gas. The loss of carbon dioxide reduces the molecule by one carbon atom, making acetaldehyde. The second reaction removes an electron from NADH, forming NAD$^+$ and producing ethanol from the acetaldehyde, which accepts the electron. The fermentation of pyruvic acid by yeast produces the ethanol found in alcoholic beverages (Figure 2.42). If the carbon dioxide produced by the reaction is not vented from the

56

fermentation chamber, for example in beer and sparkling wines, it remains dissolved in the medium until the pressure is released. Ethanol above 12 percent is toxic to yeast, so natural levels of alcohol in wine occur at a maximum of 12 percent.

Figure 2.42 Fermentation of grape juice to make wine produces CO_2 as a byproduct. Fermentation tanks have valves so that pressure inside the tanks can be released.

Summary

Comparing Prokaryotic and Eukaryotic Cells

Prokaryotes are predominantly single-celled organisms of the domains Bacteria and Archaea. All prokaryotes have plasma membranes, cytoplasm, ribosomes, a cell wall, DNA, and lack membrane-bound organelles. Many also have polysaccharide capsules. Prokaryotic cells range in diameter from 0.1–5.0 μm.

Like a prokaryotic cell, a eukaryotic cell has a plasma membrane, cytoplasm, and ribosomes, but a eukaryotic cell is typically larger than a prokaryotic cell, has a true nucleus (meaning its DNA is surrounded by a membrane), and has other membrane-bound organelles that allow for compartmentalization of functions. Eukaryotic cells tend to be 10 to 100 times the size of prokaryotic cells.

Eukaryotic Cells

Like a prokaryotic cell, a eukaryotic cell has a plasma membrane, cytoplasm, and ribosomes, but a eukaryotic cell is typically larger than a prokaryotic cell, has a true nucleus (meaning its DNA is surrounded by a membrane), and has other membrane-bound organelles that allow for compartmentalization of functions. The plasma membrane is a phospholipid bilayer embedded with proteins. The nucleolus within the nucleus is the site for ribosome assembly. Ribosomes are found in the cytoplasm or are attached to the cytoplasmic side of the plasma membrane or endoplasmic reticulum. They perform protein synthesis. Mitochondria perform cellular respiration and produce ATP. Peroxisomes break down fatty acids, amino acids, and some toxins. Vesicles and vacuoles are storage and transport compartments. In plant cells, vacuoles also help break down macromolecules.

Animal cells also have a centrosome and lysosomes. The centrosome has two bodies, the centrioles, with an unknown role in cell division. Lysosomes are the digestive organelles of animal cells.

Plant cells have a cell wall, chloroplasts, and a central vacuole. The plant cell wall, whose primary component is cellulose, protects the cell, provides structural support, and gives shape to the cell.

Photosynthesis takes place in chloroplasts. The central vacuole expands, enlarging the cell without the need to produce more cytoplasm.

The endomembrane system includes the nuclear envelope, the endoplasmic reticulum, Golgi apparatus, lysosomes, vesicles, as well as the plasma membrane. These cellular components work together to modify, package, tag, and transport membrane lipids and proteins.

The cytoskeleton has three different types of protein elements. Microfilaments provide rigidity and shape to the cell, and facilitate cellular movements. Intermediate filaments bear tension and anchor the nucleus and other organelles in place. Microtubules help the cell resist compression, serve as tracks for motor proteins that move vesicles through the cell, and pull replicated chromosomes to opposite ends of a dividing cell. They are also the structural elements of centrioles, flagella, and cilia.

Animal cells communicate through their extracellular matrices and are connected to each other by tight junctions, desmosomes, and gap junctions. Plant cells are connected and communicate with each other by plasmodesmata.

Glycolysis

ATP functions as the energy currency for cells. It allows cells to store energy briefly and transport it within itself to support endergonic chemical reactions. The structure of ATP is that of an RNA nucleotide with three phosphate groups attached. As ATP is used for energy, a phosphate group is detached, and ADP is produced. Energy derived from glucose catabolism is used to recharge ADP into ATP.

Glycolysis is the first pathway used in the breakdown of glucose to extract energy. Because it is used by nearly all organisms on earth, it must have evolved early in the history of life. Glycolysis consists of two parts: The first part prepares the six-carbon ring of glucose for separation into two three-carbon sugars. Energy from ATP is invested into the molecule during this step to energize the separation. The second half of glycolysis extracts ATP and high-energy electrons from hydrogen atoms and attaches them to NAD^+. Two ATP molecules are invested in the first half and four ATP molecules are formed during the second half. This produces a net gain of two ATP molecules per molecule of glucose for the cell.

Citric Acid Cycle and Oxidative Phosphorylation

The citric acid cycle is a series of chemical reactions that removes high-energy electrons and uses them in the electron transport chain to generate ATP. One molecule of ATP (or an equivalent) is produced per each turn of the cycle.

The electron transport chain is the portion of aerobic respiration that uses free oxygen as the final electron acceptor for electrons removed from the intermediate compounds in glucose catabolism. The electrons are passed through a series of chemical reactions, with a small amount of free energy used at three points to transport hydrogen ions across the membrane. This contributes to the gradient used in chemiosmosis. As the electrons are passed from NADH or $FADH_2$ down the electron transport chain, they lose energy. The products of the electron transport chain are water and ATP. A number of intermediate compounds can be diverted into the anabolism of other biochemical molecules, such as nucleic acids, non-essential amino acids, sugars, and lipids. These same molecules, except nucleic acids, can serve as energy sources for the glucose pathway.

Fermentation

If NADH cannot be metabolized through aerobic respiration, another electron acceptor is used. Most organisms will use some form of fermentation to accomplish the regeneration of NAD^+, ensuring the continuation of glycolysis. The regeneration of NAD^+ in fermentation is not accompanied by ATP production; therefore, the potential for NADH to produce ATP using an electron transport chain is not utilized.

Overview of Photosynthesis

The process of photosynthesis transformed life on earth. By harnessing energy from the sun, photosynthesis allowed living things to access enormous amounts of energy. Because of photosynthesis, living things gained access to sufficient energy, allowing them to evolve new structures and achieve the biodiversity that is evident today.

Only certain organisms, called autotrophs, can perform photosynthesis; they require the presence of chlorophyll, a specialized pigment that can absorb light and convert light energy into chemical energy.

 Photosynthesis uses carbon dioxide and water to assemble carbohydrate molecules (usually glucose) and releases oxygen into the air. Eukaryotic autotrophs, such as plants and algae, have organelles called chloroplasts in which photosynthesis takes place.

The Cell Cycle

The cell cycle is an orderly sequence of events. Cells on the path to cell division proceed through a series of precisely timed and carefully regulated stages. In eukaryotes, the cell cycle consists of a long preparatory period, called interphase. Interphase is divided into G1, S, and G2 phases. Mitosis consists of five stages: prophase, prometaphase, metaphase, anaphase, and telophase. Mitosis is usually accompanied by cytokinesis, during which the cytoplasmic components of the daughter cells are separated either by an actin ring (animal cells) or by cell plate formation (plant cells).

Each step of the cell cycle is monitored by internal controls called checkpoints.There are three major checkpoints in the cell cycle: one near the end of G1, a second at the G2–M transition, and the third during metaphase.

Meiosis

Sexual reproduction requires that diploid organisms produce haploid cells that can fuse during fertilization to form diploid offspring. The process that results in haploid cells is called meiosis. Meiosis is a series of events that arrange and separate chromosomes into daughter cells. During the interphase of meiosis, each chromosome is duplicated. In meiosis, there are two rounds of nuclear division resulting in four nuclei and usually four haploid daughter cells, each with half the number of chromosomes as the parent cell. During meiosis, variation in the daughter nuclei is introduced because of crossover in prophase I and random alignment at metaphase I. The cells that are produced by meiosis are genetically unique.

Meiosis and mitosis share similarities, but have distinct outcomes. Mitotic divisions are single nuclear divisions that produce daughter nuclei that are genetically identical and have the same number of chromosome sets as the original cell. Meiotic divisions are two nuclear divisions that produce four daughter nuclei that are genetically different and have one chromosome set rather than the two sets the parent cell had. The main differences between the processes occur in the first division of meiosis. The homologous chromosomes separate into different nuclei during meiosis I causing a reduction of ploidy level. The second division of meiosis is much more similar to a mitotic division.

Cancer and the Cell Cycle

Cancer is the result of unchecked cell division caused by a breakdown of the mechanisms regulating the cell cycle. The loss of control begins with a change in the DNA sequence of a gene that codes for one of the regulatory molecules. Faulty instructions lead to a protein that does not function as it should. Any disruption of the monitoring system can allow other mistakes to be passed on to the daughter cells. Each successive cell division will give rise to daughter cells with even more accumulated damage. Eventually, all checkpoints become nonfunctional, and rapidly reproducing cells crowd out normal cells, resulting in tumorous growth.

Review Questions

1. The _____ is the basic unit of life.

   a. organism
   b. cell
   c. tissue
   d. organ

2. Which of these do all prokaryotes and eukaryotes share?
   a. nuclear envelope
   b. cell walls
   c. organelles
   d. plasma membrane

3. A typical prokaryotic cell _____ compared to a eukaryotic cell.
   a. is smaller in size by a factor of 100
   b. is similar in size

    c. is smaller in size by a factor of one million

   d. is larger in size by a factor of 10

  4. Which of the following is found both in eukaryotic and prokaryotic cells?

   a. nucleus

   b. mitochondrion

   c. vacuole

   d. ribosome

  5. Which of the following is not a component of the endomembrane system?

   a. mitochondrion

   b. Golgi apparatus

   c. endoplasmic reticulum

    d. lysosome

6. The energy currency used by cells is .

 a. ATP

 b. ADP

 c. AMP

 d. adenosine

7. The glucose that enters the glycolysis pathway is split into two molecules of .

a. ATP

b. phosphate

c. NADH

d. pyruvate

8. Which of the following fermentation methods can occur in animal skeletal muscles?

a. lactic acid fermentation

b. alcohol fermentation

c. mixed acid fermentation

d. propionic fermentation

9. What two products result from photosynthesis?

a. water and carbon dioxide

b. water and oxygen

c. glucose and oxygen

d. glucose and carbon dioxide

10. From where does a heterotroph directly obtain its energy?

a. the sun

b. the sun and eating other organisms

c. eating other organisms

d.   simple chemicals in the environment

11.   Which molecule absorbs the energy of a photon in photosynthesis?

a.   ATP

b.   glucose

c.   chlorophyll

d.   water

12.   Plants produce oxygen when they photosynthesize. Where does the oxygen come from?

a.   splitting water molecules

b.   ATP synthesis

c.   the electron transport chain

d.   chlorophyll

13.   Chromosomes are duplicated during what portion of the cell cycle?

a.   G1 phase

b.   S phase

c.   prophase

d.   prometaphase

14.   Separation of the sister chromatids is a characteristic of which stage of mitosis?

a.   prometaphase

b.   metaphase

c.   anaphase

d.   telophase

15.   The individual chromosomes become visible with a light microscope during which stage of   mitosis?

a.   prophase

b.   prometaphase

c.   metaphase

d.   anaphase

16.   Meiosis produces _____ daughter cells.

a.   two haploid

b.   two diploid

c.   four haploid

d.   four diploid

17.   At which stage of meiosis are sister chromatids separated from each other?

a.   prophase I

b.   prophase II

c.   anaphase I

d.   anaphase II

63

Chapter 3

Nucleic Acids and Genetics

Nucleic Acids

Nucleic acids are key macromolecules in the continuity of life. They carry the genetic blueprint of a cell and carry instructions for the functioning of the cell.

The two main types of **nucleic acids** are **deoxyribonucleic acid (DNA)** and **ribonucleic acid (RNA)**. DNA is the genetic material found in all living organisms, ranging from single-celled bacteria to multicellular mammals.

The other type of nucleic acid, RNA, is mostly involved in protein synthesis. The DNA molecules never leave the nucleus, but instead use an RNA intermediary to communicate with the rest of the cell. Other types of RNA are also involved in protein synthesis and its regulation.

DNA and RNA are made up of monomers known as **nucleotides**. The nucleotides combine with each other to form a polynucleotide, DNA or RNA. Each nucleotide is made up of three components: a nitrogenous base, a pentose (five-carbon) sugar, and a phosphate group (Figure 3.1). Each nitrogenous base in a nucleotide is attached to a sugar molecule, which is attached to a phosphate group.
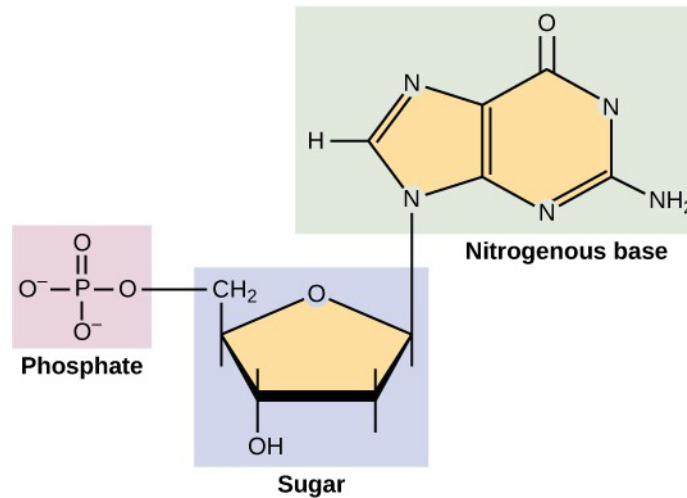
Figure 3.1 A nucleotide is made up of three components: a nitrogenous base, a pentose sugar, and a phosphate group.

DNA Double-Helical Structure

DNA has a double-helical structure (Figure 3.2). It is composed of two strands, or polymers, of nucleotides. The strands are formed with bonds between phosphate and sugar groups of adjacent nucleotides. The strands are bonded to each other at their bases with hydrogen bonds, and the strands coil about each other along their length, hence the "double helix" description, which means a double spiral.
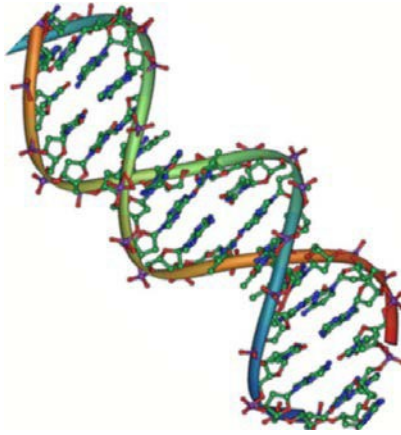
Figure 3.2 The double-helix model shows DNA as two parallel strands of intertwining molecules. (credit: Jerome Walker, Dennis Myts)

The alternating sugar and phosphate groups lie on the outside of each strand, forming the backbone of the DNA. The nitrogenous bases are stacked in the interior, like the steps of a staircase, and these bases

The Structure of DNA

By the end of this section, you will be able to:

- Describe the structure of DNA
- Describe how eukaryotic and prokaryotic DNA is arranged in the cell

In the 1950s, Francis Crick and James Watson worked together at the University of Cambridge, England, to determine the structure of DNA. Other scientists, such as Linus Pauling and Maurice Wilkins, were also actively exploring this field. Pauling had discovered the secondary structure of proteins using X-ray crystallography. X-ray crystallography is a method for investigating molecular structure by observing the patterns formed by X-rays shot through a crystal of the substance. The patterns give important information about the structure of the molecule of interest. In Wilkins' lab, researcher Rosalind Franklin was using X-ray crystallography to understand the structure of DNA. Watson and Crick were able to piece together the puzzle of the DNA molecule using Franklin's data (Figure 3.3). Watson and Crick also had key pieces of information available from other researchers such as Chargaff's rules. Chargaff had shown that of the four kinds of monomers (nucleotides) present in a DNA molecule, two types were always present in equal amounts and the remaining two types were also always present in equal amounts. This meant they were always paired in some way. In 1962, James Watson, Francis Crick, and Maurice Wilkins were awarded the Nobel Prize in Medicine for their work in determining the structure of DNA.


(a)   (b)

Figure 3.3 Pioneering scientists (a) James Watson and Francis Crick are pictured here with American geneticist Maclyn McCarty. Scientist Rosalind Franklin discovered (b) the X-ray diffraction pattern of DNA, which helped to elucidate its double helix structure. (credit a: modification of work by Marjorie McCarty; b: modification of work by NIH)

63

Now let's consider the structure of the two types of nucleic acids, deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). The building blocks of DNA are nucleotides, which are made up of three parts: a **deoxyribose** (5-carbon sugar), a **phosphate group**, and a **nitrogenous base** (Figure 3.4). There are four types of nitrogenous bases in DNA. Adenine (A) and guanine (G) are double-ringed purines, and cytosine (C) and thymine (T) are smaller, single-ringed pyrimidines. The nucleotide is named according to the nitrogenous base it contains.
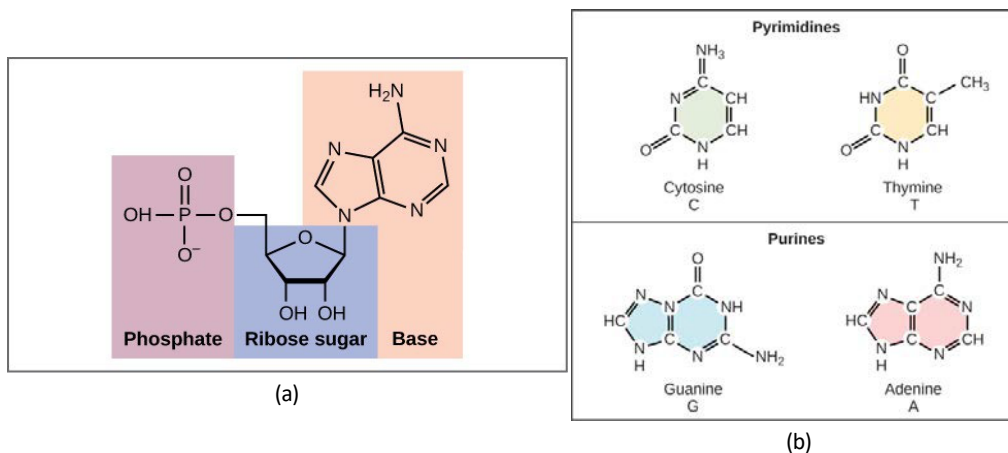
Figure 3.4 (a) Each DNA nucleotide is made up of a sugar, a phosphate group, and a base. (b) Cytosine and thymine are pyrimidines. Guanine and adenine are purines.

The phosphate group of one nucleotide bonds covalently with the sugar molecule of the next nucleotide, and so on, forming a long polymer of nucleotide monomers. The sugar–phosphate groups line up in a "backbone" for each single strand of DNA, and the nucleotide bases stick out from this backbone. The carbon atoms of the five-carbon sugar are numbered clockwise from the oxygen as 1', 2', 3', 4', and 5' (1' is read as "one prime"). The phosphate group is attached to the 5' carbon of one nucleotide and the 3' carbon of the next nucleotide. In its natural state, each DNA molecule is actually composed of two single strands held together along their length with hydrogen bonds between the bases.

Watson and Crick proposed that the DNA is made up of two strands that are twisted around each other to form a right-handed helix, called a **double helix**. Base-pairing takes place between a purine and pyrimidine: namely, A pairs with T, and G pairs with C. In other words, adenine and thymine are complementary base pairs, and cytosine and guanine are also complementary base pairs. This is the basis for Chargaff's rule; because of their complementarity, there is as much adenine as thymine in a DNA molecule and as much guanine as cytosine. Adenine and thymine are connected by two hydrogen bonds, and cytosine and guanine are connected by three hydrogen bonds. The two strands are anti-parallel in nature; that is, one strand will have the 3' carbon of the sugar in the "upward" position, whereas the other strand will have the 5' carbon in the upward position. The diameter of the DNA double helix is uniform throughout because a purine (two rings) always pairs with a pyrimidine (one ring) and their combined lengths are always equal. (Figure 3.5).
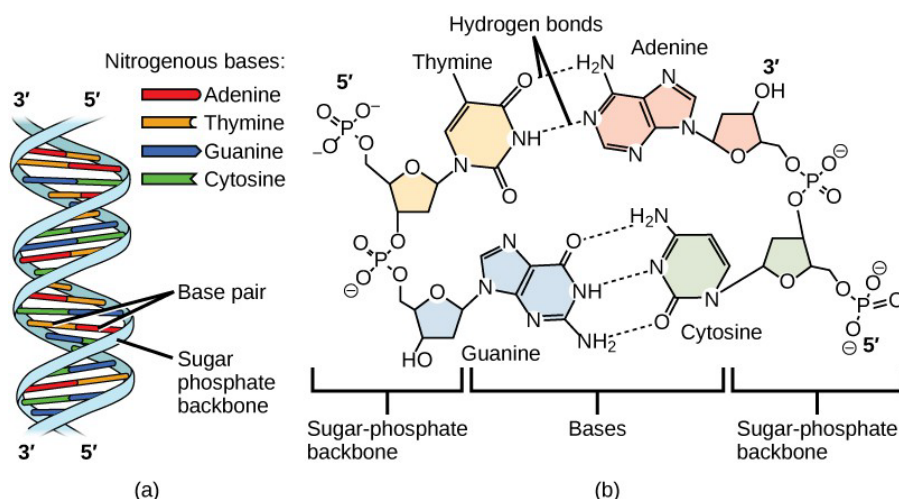

Figure 3.5 DNA (a) forms a double stranded helix, and (b) adenine pairs with thymine and cytosine pairs with guanine. (credit a: modification of work by Jerome Walker, Dennis Myts)

The Structure of RNA

There is a second nucleic acid in all cells called ribonucleic acid, or RNA. Like DNA, RNA is a polymer of nucleotides. Each of the nucleotides in RNA is made up of a nitrogenous base, a five-carbon sugar, and a phosphate group. In the case of RNA, the five-carbon sugar is ribose, not deoxyribose. Ribose has a hydroxyl group at the 2' carbon, unlike deoxyribose, which has only a hydrogen atom (Figure 3.6).

65
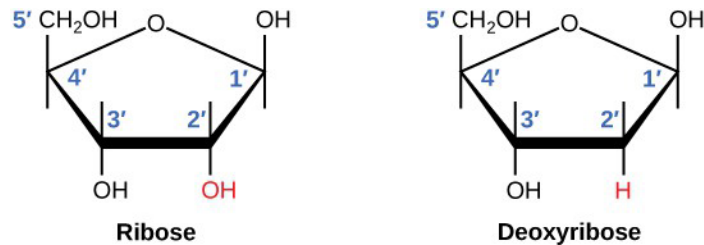
Figure 3.6 The difference between the ribose found in RNA and the deoxyribose found in DNA is that ribose has a hydroxyl group at the 2' carbon.

RNA nucleotides contain the nitrogenous bases adenine, cytosine, and guanine. However, they do not contain thymine, which is instead replaced by uracil, symbolized by a "U." RNA exists as a single-stranded molecule rather than a double-stranded helix. Molecular biologists have named several kinds of RNA on the basis of their function. These include messenger RNA (mRNA), transfer RNA (tRNA), and ribosomal RNA (rRNA)—molecules that are involved in the production of proteins from the DNA code.

How DNA Is Arranged in the Cell

DNA is a working molecule; it must be replicated when a cell is ready to divide, and it must be "read" to produce the molecules, such as proteins, to carry out the functions of the cell. For this reason, the DNA is protected and packaged in very specific ways. In addition, DNA molecules can be very long. Stretched end-to-end, the DNA molecules in a single human cell would come to a length of about 2 meters. Thus, the DNA for a cell must be packaged in a very ordered way to fit and function within a structure (the cell) that is not visible to the naked eye. The chromosomes of prokaryotes are much simpler than those of eukaryotes in many of their features (Figure 3.7). Most prokaryotes contain a single, circular chromosome that is found in an area in the cytoplasm called the nucleoid.
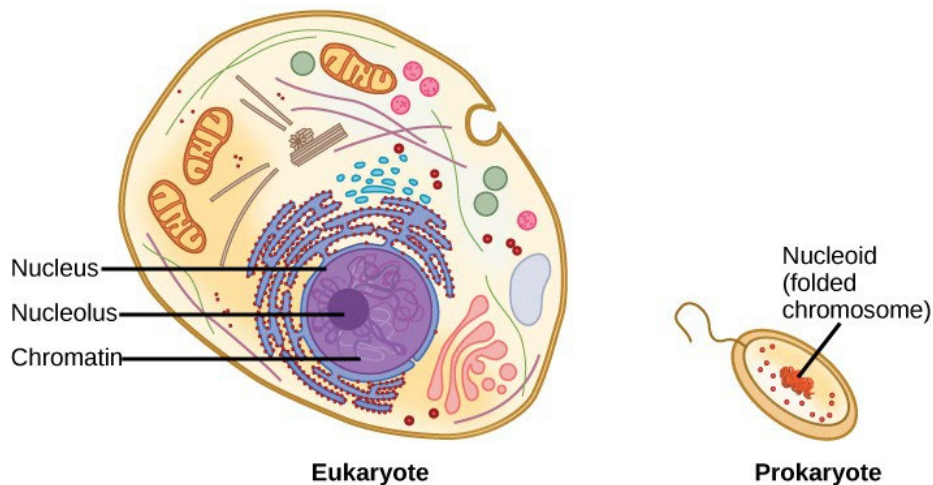
Figure 3.7 A eukaryote contains a well-defined nucleus, whereas in prokaryotes, the chromosome lies in the cytoplasm in an area called the nucleoid.

The size of the genome in one of the most well-studied prokaryotes, *Escherichia coli,* is 4.6 million base pairs, which would extend a distance of about 1.6 mm if stretched out. So how does this fit inside a small bacterial cell? The DNA is twisted beyond the double helix in what is known as supercoiling. Some proteins are known to be involved in the supercoiling; other proteins and enzymes help in maintaining the supercoiled structure.

Eukaryotes, whose chromosomes each consist of a linear DNA molecule, employ a different type of packing strategy to fit their DNA inside the nucleus (Figure 3.8). At the most basic level, DNA is wrapped around proteins known as histones to form structures called nucleosomes. The DNA is wrapped tightly around the histone core. This nucleosome is linked to the next one by a short strand of DNA that is free of histones. This is also known as the "beads on a string" structure; the nucleosomes are the "beads" and the short lengths of DNA between them are the "string." The nucleosomes, with their DNA coiled around them, stack compactly onto each other to form a 30-nm–wide fiber. This fiber is further coiled into a thicker and more compact structure. At the metaphase stage of mitosis, when the chromosomes are lined up in the center of the cell, the chromosomes are at their most compacted. They are approximately 700 nm in width, and are found in association with scaffold proteins.

66

In interphase, the phase of the cell cycle between mitoses at which the chromosomes are decondensed, eukaryotic chromosomes have two distinct regions that can be distinguished by staining. There is a tightly packaged region that stains darkly, and a less dense region. The darkly staining regions usually contain genes that are not active, and are found in the regions of the centromere and telomeres. The lightly staining regions usually contain genes that are active, with DNA packaged around nucleosomes but not further compacted.

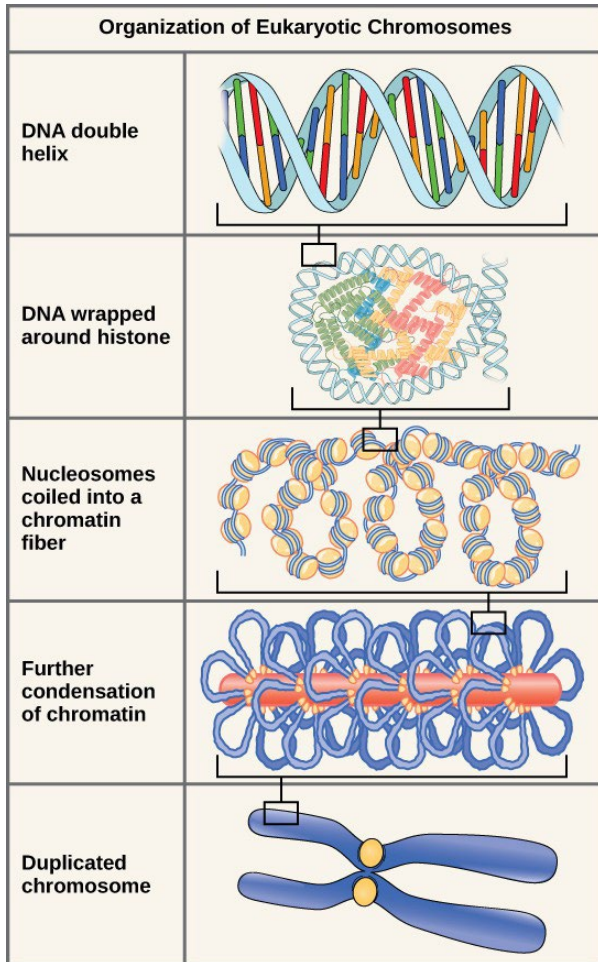| Organization of Eukaryotic Chromosomes	
DNA double helix	
DNA wrapped around histone	
Nucleosomes coiled into a chromatin fiber	
Further condensation of chromatin	
Duplicated chromosome	

Figure 3.8 Packing of DNA inside of cells

| DNA Replication

By the end of this section, you will be able to:

- Explain the process of DNA replication
- Explain the importance of telomerase to DNA replication
- Describe mechanisms of DNA repair

When a cell divides, it is important that each daughter cell receives an identical copy of the DNA. This is accomplished by the process of DNA replication. The replication of DNA occurs during the synthesis phase, or S phase, of the cell cycle, before the cell enters mitosis or meiosis.

The elucidation of the structure of the double helix provided a hint as to how DNA is copied. Recall that adenine nucleotides pair with thymine nucleotides, and cytosine with guanine. This means that the two strands are complementary to each other. For example, a strand of DNA with a nucleotide sequence of AGTCATGA will have a complementary strand with the sequence TCAGTACT (Figure 3.9).

67
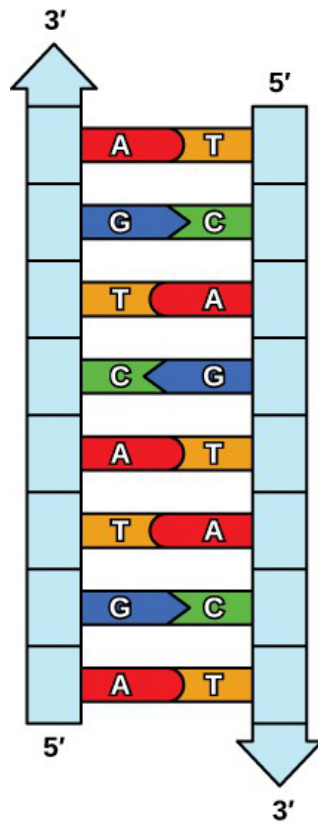
Figure 3.9 The two strands of DNA are complementary, meaning the sequence of bases in one strand can be used to create the correct sequence of bases in the other strand.

Because of the complementarity of the two strands, having one strand means that it is possible to recreate the other strand. This model for replication suggests that the two strands of the double helix separate during replication, and each strand serves as a template from which the new complementary strand is copied (Figure 3.10).
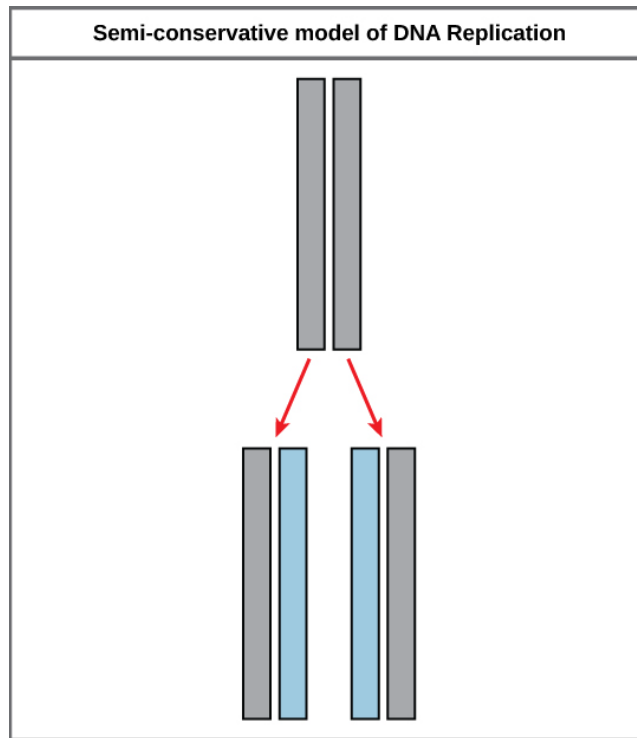
Figure 3.10 The semiconservative model of DNA replication is shown. Gray indicates the original DNA strands, and blue indicates newly synthesized DNA.

During DNA replication, each of the two strands that make up the double helix serves as a template from which new strands are copied. The new strand will be complementary to the parental or "old" strand. Each new double strand consists of one parental strand and one new daughter strand. This is known as **semiconservative replication**. When two DNA copies are formed, they have an identical sequence of nucleotide bases and are divided equally into two daughter cells.

DNA Replication in Eukaryotes

Because eukaryotic genomes are very complex, DNA replication is a very complicated process that involves several enzymes and other proteins. It occurs in three main stages: initiation, elongation, and termination.

Recall that eukaryotic DNA is bound to proteins known as histones to form structures called nucleosomes. During initiation, the DNA is made accessible to the proteins and enzymes involved in the replication process. How does the replication machinery know where on the DNA double helix to begin? It turns out that there are specific nucleotide sequences called origins of replication at which replication begins. Certain proteins bind to the origin of replication while an enzyme called **helicase** unwinds and opens up the DNA helix. As the DNA opens up, Y-shaped structures called **replication forks** are formed (). Two replication forks are formed at the origin of replication, and these get extended in both directions as replication proceeds. There are multiple origins of replication on the eukaryotic chromosome, such that replication can occur simultaneously from several places in the genome.

During elongation, an enzyme called **DNA polymerase** adds DNA nucleotides to the 3' end of the template. Because DNA polymerase can only add new nucleotides at the end of a backbone, a **primer** sequence, which provides this starting point, is added with complementary RNA nucleotides. This primer is removed later, and the nucleotides are replaced with DNA nucleotides. One strand, which is complementary to the parental DNA strand, is synthesized continuously toward the replication fork so the polymerase can add nucleotides in this direction. This continuously synthesized strand is known as the **leading strand**. Because DNA polymerase can only synthesize DNA in a 5' to 3' direction, the other new strand is put together in short pieces called **Okazaki fragments**. The Okazaki fragments each require a primer made of RNA to start the synthesis. The strand with the Okazaki fragments is known as the **lagging strand**. As synthesis proceeds, an enzyme removes the RNA primer, which is then replaced with DNA nucleotides, and the gaps between fragments are sealed by an enzyme called **DNA ligase**.

The process of DNA replication can be summarized as follows:

1. DNA unwinds at the origin of replication.

2. New bases are added to the complementary parental strands. One new strand is made continuously, while the other strand is made in pieces.

3. Primers are removed, new DNA nucleotides are put in place of the primers and the backbone is sealed by DNA ligase.

art CONNECTION
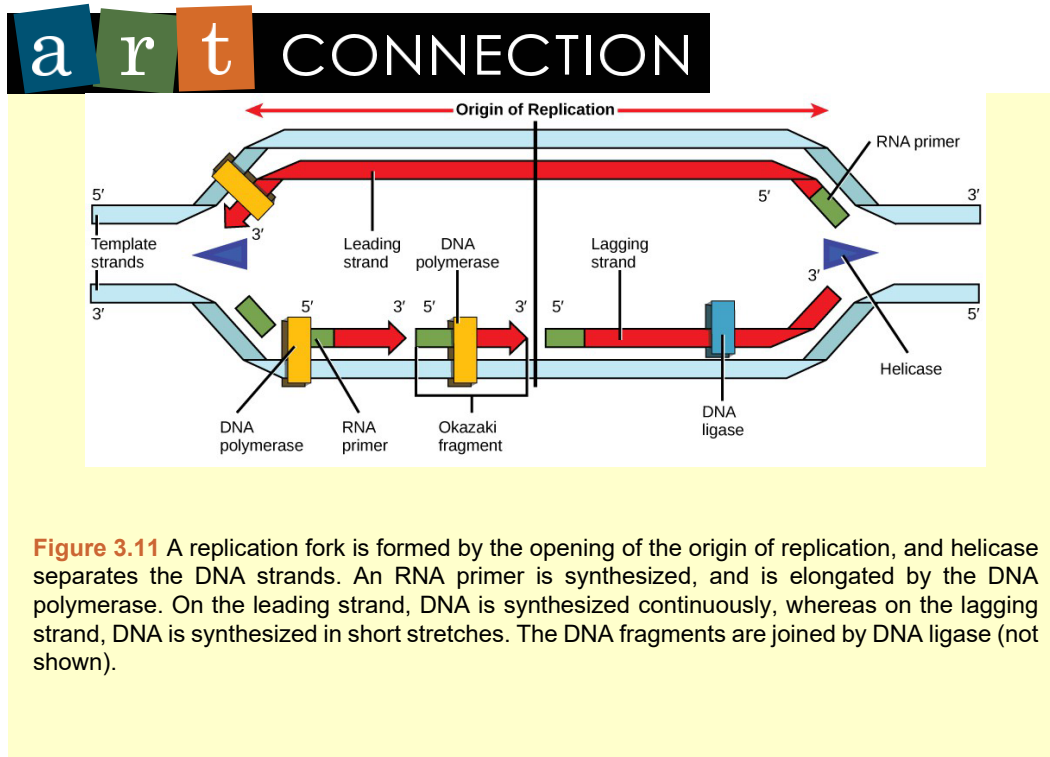
Figure 3.11 A replication fork is formed by the opening of the origin of replication, and helicase separates the DNA strands. An RNA primer is synthesized, and is elongated by the DNA polymerase. On the leading strand, DNA is synthesized continuously, whereas on the lagging strand, DNA is synthesized in short stretches. The DNA fragments are joined by DNA ligase (not shown).

Telomere Replication

Because eukaryotic chromosomes are linear, DNA replication comes to the end of a line in eukaryotic chromosomes. As you have learned, the DNA polymerase enzyme can add nucleotides in only one direction. In the leading strand, synthesis continues until the end of the chromosome is reached; however, on the lagging strand there is no place for a primer to be made for the DNA fragment to be copied at the end of the chromosome. This presents a problem for the cell because the ends remain unpaired, and over time these ends get progressively shorter as cells continue to divide. The ends of the linear chromosomes are known as **telomeres**, which have repetitive sequences that do not code for a particular gene. As a consequence, it is telomeres that are shortened with each round of DNA replication instead of genes. For example, in humans, a six base-pair sequence, TTAGGG, is repeated 100 to 1000 times. The discovery of the enzyme **telomerase** (**Figure 3.11**) helped in the understanding of how chromosome ends are maintained. The telomerase attaches to the end of the chromosome, and complementary bases to the RNA template are added on the end of the DNA strand. Once the lagging strand template is sufficiently elongated, DNA polymerase can now add nucleotides that are complementary to the ends of the chromosomes. Thus, the ends of the chromosomes are replicated.
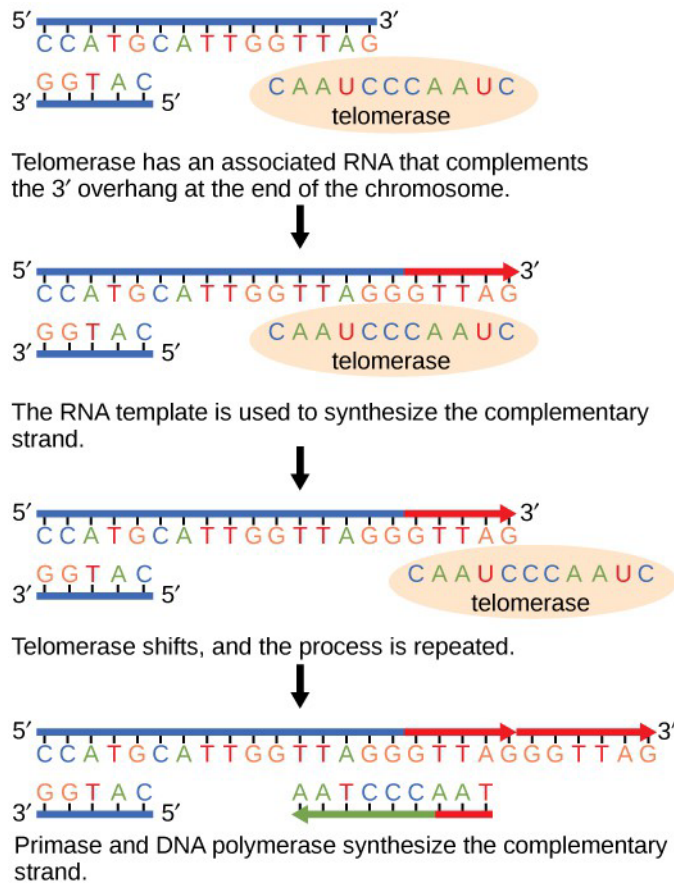
Figure 3.12 The ends of linear chromosomes are maintained by the action of the telomerase enzyme.

Telomerase is typically found to be active in germ cells, adult stem cells, and some cancer cells. For her discovery of telomerase and its action, Elizabeth Blackburn (Figure 3.13) received the Nobel Prize for Medicine and Physiology in 2009.

Figure 3.13 Elizabeth Blackburn, 2009 Nobel Laureate, was the scientist who discovered how telomerase works. (credit: U.S. Embassy, Stockholm, Sweden)

Telomerase is not active in adult somatic cells. Adult somatic cells that undergo cell division continue to have their telomeres shortened. This essentially means that telomere shortening is associated with aging. In 2010, scientists found that telomerase can reverse some age-related conditions in mice, and [1] this may have potential in regenerative medicine. Telomerase-deficient mice were used in these

1. Mariella Jaskelioff, et al., "Telomerase reactivation reverses tissue degeneration in aged telomerase-deficient mice," *Nature*, 469 (2011):102–7.

studies; these mice have tissue atrophy, stem-cell depletion, organ system failure, and impaired tissue injury responses. Telomerase reactivation in these mice caused extension of telomeres, reduced DNA damage, reversed neurodegeneration, and improved functioning of the testes, spleen, and intestines. Thus, telomere reactivation may have potential for treating age-related diseases in humans.

DNA Replication in Prokaryotes

Recall that the prokaryotic chromosome is a circular molecule with a less extensive coiling structure than eukaryotic chromosomes. The eukaryotic chromosome is linear and highly coiled around proteins. While there are many similarities in the DNA replication process, these structural differences necessitate some differences in the DNA replication process in these two life forms.

DNA replication has been extremely well-studied in prokaryotes, primarily because of the small size of the genome and large number of variants available. *Escherichia coli* has 4.6 million base pairs in a single circular chromosome, and all of it gets replicated in approximately 42 minutes, starting from a single origin of replication and proceeding around the chromosome in both directions. This means that approximately 1000 nucleotides are added per second. The process is much more rapid than in eukaryotes. Table 3.2 summarizes the differences between prokaryotic and eukaryotic replications.

Differences between Prokaryotic and Eukaryotic Replications

| Property	Prokaryotes	Eukaryotes
Origin of replication	Single	Multiple
Rate of replication	1000 nucleotides/s	50 to 100 nucleotides/s
Chromosome structure	circular	linear
Telomerase	Not present	Present

Table 3.2

Click through a **tutorial (http://openstaxcollege.org/l/DNA_replicatio2)** on DNA replication.

DNA Repair

DNA polymerase can make mistakes while adding nucleotides. It edits the DNA by proofreading every newly added base. Incorrect bases are removed and replaced by the correct base, and then polymerization continues (**Figure 3.14a**). Most mistakes are corrected during replication, although when this does not happen, the **mismatch repair** mechanism is employed. Mismatch repair enzymes recognize the wrongly incorporated base and excise it from the DNA, replacing it with the correct base (Figure 3.14**b**). In yet another type of repair, **nucleotide excision repair**, the DNA double strand is unwound and separated, the incorrect bases are removed along with a few bases on the 5' and 3' end, and these are replaced by copying the template with the help of DNA polymerase (Figure 3.14**c**). Nucleotide excision repair is particularly important in correcting thymine dimers, which are primarily caused by ultraviolet light. In a thymine dimer, two thymine nucleotides adjacent to each other on one strand are covalently bonded to each other rather than their complementary bases. If the dimer is not removed and repaired it will lead to a mutation. Individuals with flaws in their nucleotide excision repair genes show extreme sensitivity to sunlight and develop skin cancers early in life.

72
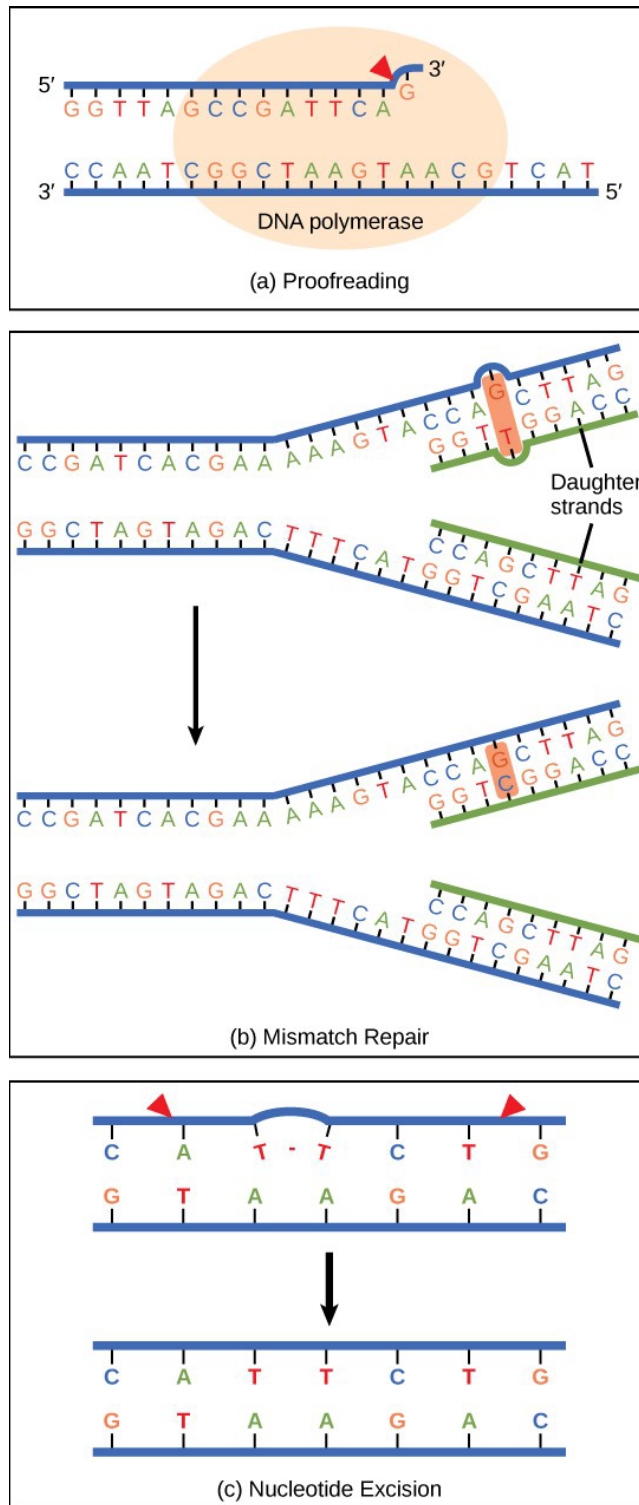
Figure 3.14 Proofreading by DNA polymerase (a) corrects errors during replication. In mismatch repair (b), the incorrectly added base is detected after replication. The mismatch repair proteins detect this base and remove it from the newly synthesized strand by nuclease action. The gap is now filled with the correctly paired base. Nucleotide excision (c) repairs thymine dimers. When exposed to UV, thymines lying adjacent to each other can form thymine dimers. In normal cells, they are excised and replaced.

Most mistakes are corrected; if they are not, they may result in a **mutation**—defined as a permanent change in the DNA sequence. Mutations in repair genes may lead to serious consequences like cancer.

By the end of this section, you will be able to:

- Explain the central dogma
- Explain the main steps of transcription
- Describe how eukaryotic mRNA is processed

In both prokaryotes and eukaryotes, the second function of DNA (the first was replication) is to provide the information needed to construct the proteins necessary so that the cell can perform all of its functions. To do this, the DNA is "read" or transcribed into an **mRNA** molecule. The mRNA then provides the code to form a protein by a process called translation. Through the processes of transcription and translation, a protein is built with a specific sequence of amino acids that was originally encoded in the DNA. This module discusses the details of transcription.

The Central Dogma: DNA Encodes RNA; RNA Encodes Protein

The flow of genetic information in cells from DNA to mRNA to protein is described by the central dogma (Figure 3.15), which states that genes specify the sequences of mRNAs, which in turn specify the sequences of proteins.
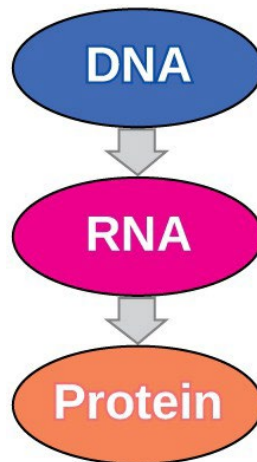
Figure 3.15 The central dogma states that DNA encodes RNA, which in turn encodes protein.

The copying of DNA to mRNA is relatively straightforward, with one nucleotide being added to the mRNA strand for every complementary nucleotide read in the DNA strand. The translation to protein is more complex because groups of three mRNA nucleotides correspond to one amino acid of the protein sequence. However, as we shall see in the next module, the translation to protein is still systematic, such that nucleotides 1 to 3 correspond to amino acid 1, nucleotides 4 to 6 correspond to amino acid 2, and so on.

Transcription: from DNA to mRNA

Both prokaryotes and eukaryotes perform fundamentally the same process of transcription, with the important difference of the membrane-bound nucleus in eukaryotes. With the genes bound in the nucleus, transcription occurs in the nucleus of the cell and the mRNA transcript must be transported to the cytoplasm. The prokaryotes, which include bacteria and archaea, lack membrane-bound nuclei and other organelles, and transcription occurs in the cytoplasm of the cell. In both prokaryotes and eukaryotes, transcription occurs in three main stages: initiation, elongation, and termination.

Initiation

Transcription requires the DNA double helix to partially unwind in the region of mRNA synthesis. The region of unwinding is called a **transcription bubble**. The DNA sequence onto which the proteins and enzymes involved in transcription bind to initiate the process is called a **promoter**. In most cases, promoters exist upstream of the genes they regulate. The specific sequence of a promoter is very

74

important because it determines whether the corresponding gene is transcribed all of the time, some of the time, or hardly at all (Figure 3.16).
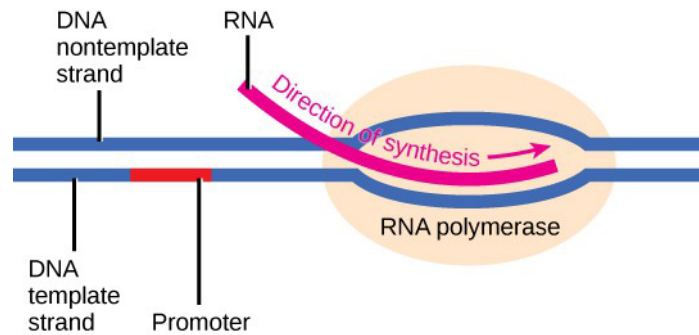
Figure 3.16 The initiation of transcription begins when DNA is unwound, forming a transcription bubble. Enzymes and other proteins involved in transcription bind at the promoter.

Elongation

Transcription always proceeds from one of the two DNA strands, which is called the **template strand**. The mRNA product is complementary to the template strand and is almost identical to the other DNA strand, called the **nontemplate strand**, with the exception that RNA contains a uracil (U) in place of the thymine (T) found in DNA. During elongation, an enzyme called **RNA polymerase** proceeds along the DNA template adding nucleotides by base pairing with the DNA template in a manner similar to DNA replication, with the difference that an RNA strand is being synthesized that does not remain bound to the DNA template. As elongation proceeds, the DNA is continuously unwound ahead of the core enzyme and rewound behind it (Figure 3.17).
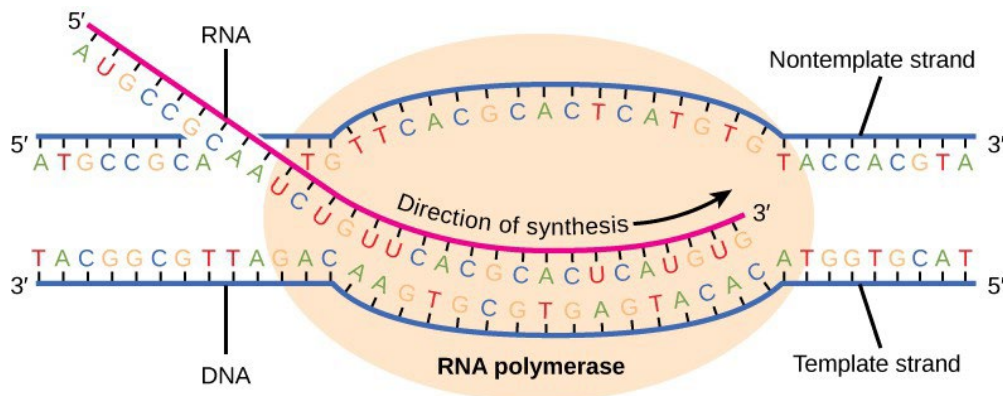

Figure 3.17 During elongation, RNA polymerase tracks along the DNA template, synthesizes mRNA in the 5' to 3' direction, and unwinds then rewinds the DNA as it is read.

Termination

Once a gene is transcribed, the prokaryotic polymerase needs to be instructed to dissociate from the DNA template and liberate the newly made mRNA. Depending on the gene being transcribed, there are two kinds of termination signals, but both involve repeated nucleotide sequences in the DNA template that result in RNA polymerase stalling, leaving the DNA template, and freeing the mRNA transcript.

On termination, the process of transcription is complete. In a prokaryotic cell, by the time termination occurs, the transcript would already have been used to partially synthesize numerous copies of the encoded protein because these processes can occur concurrently using multiple ribosomes (polyribosomes) (Figure 3.18). In contrast, the presence of a nucleus in eukaryotic cells precludes simultaneous transcription and translation.
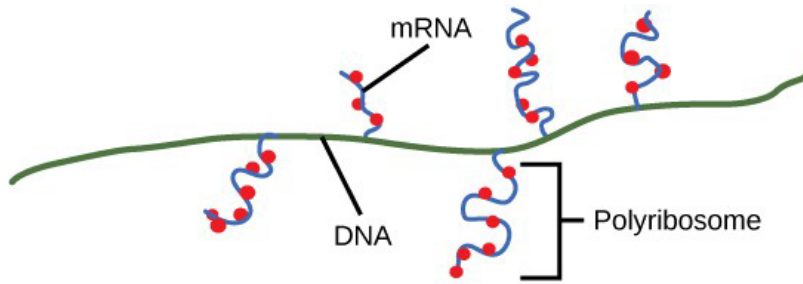
Figure 3.18 Multiple polymerases can transcribe a single bacterial gene while numerous ribosomes concurrently translate the mRNA transcripts into polypeptides. In this way, a specific protein can rapidly reach a high concentration in the bacterial cell.

Eukaryotic RNA Processing

The newly transcribed eukaryotic mRNAs must undergo several processing steps before they can be transferred from the nucleus to the cytoplasm and translated into a protein. The additional steps involved in eukaryotic mRNA maturation create a molecule that is much more stable than a prokaryotic mRNA. For example, eukaryotic mRNAs last for several hours, whereas the typical prokaryotic mRNA lasts no more than five seconds.

The mRNA transcript is first coated in RNA-stabilizing proteins to prevent it from degrading while it is processed and exported out of the nucleus. This occurs while the pre-mRNA still is being synthesized by adding a special nucleotide "cap" to the 5' end of the growing transcript. In addition to preventing degradation, factors involved in protein synthesis recognize the cap to help initiate translation by ribosomes.

Once elongation is complete, an enzyme then adds a string of approximately 200 adenine residues to the 3' end, called the poly-A tail. This modification further protects the pre-mRNA from degradation and signals to cellular factors that the transcript needs to be exported to the cytoplasm.

Eukaryotic genes are composed of protein-coding sequences called **exons** (*ex*-on signifies that they are *ex*pressed) and *int*ervening sequences called **introns** (*int*-ron denotes their *int*ervening role). Introns are removed from the pre-mRNA during processing. Intron sequences in mRNA do not encode functional proteins. It is essential that all of a pre-mRNA's introns be completely and precisely removed before protein synthesis so that the exons join together to code for the correct amino acids. If the process errs by even a single nucleotide, the sequence of the rejoined exons would be shifted, and the resulting protein would be nonfunctional. The process of removing introns and reconnecting exons is called **splicing** (Figure 3.19). Introns are removed and degraded while the pre-mRNA is still in the nucleus.
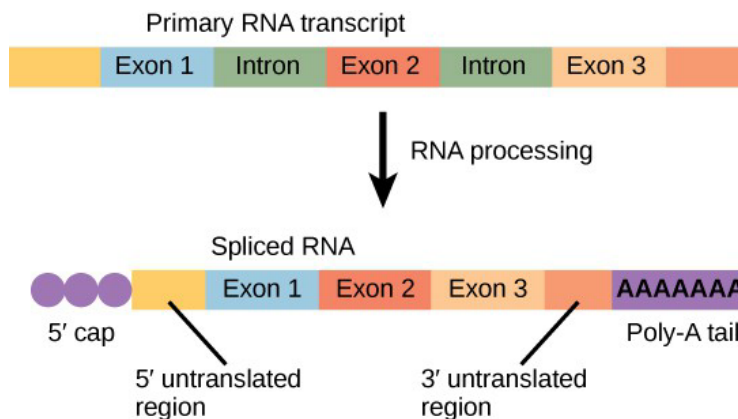
Figure 3.19 Eukaryotic mRNA contains introns that must be spliced out. A 5' cap and 3' tail are also added.

By the end of this section, you will be able to:

- Describe the different steps in protein synthesis
- Discuss the role of ribosomes in protein synthesis
- Describe the genetic code and how the nucleotide sequence determines the amino acid and the protein sequence

The synthesis of proteins is one of a cell's most energy-consuming metabolic processes. In turn, proteins account for more mass than any other component of living organisms (with the exception of water), and proteins perform a wide variety of the functions of a cell. The process of translation, or protein synthesis, involves decoding an mRNA message into a polypeptide product. Amino acids are covalently strung together in lengths ranging from approximately 50 amino acids to more than 1,000.

The Protein Synthesis Machinery

In addition to the mRNA template, many other molecules contribute to the process of translation. The composition of each component may vary across species; for instance, ribosomes may consist of different numbers of ribosomal RNAs (**rRNA**) and polypeptides depending on the organism. However, the general structures and functions of the protein synthesis machinery are comparable from bacteria to human cells. Translation requires the input of an mRNA template, ribosomes, tRNAs, and various enzymatic factors (Figure 3.20).
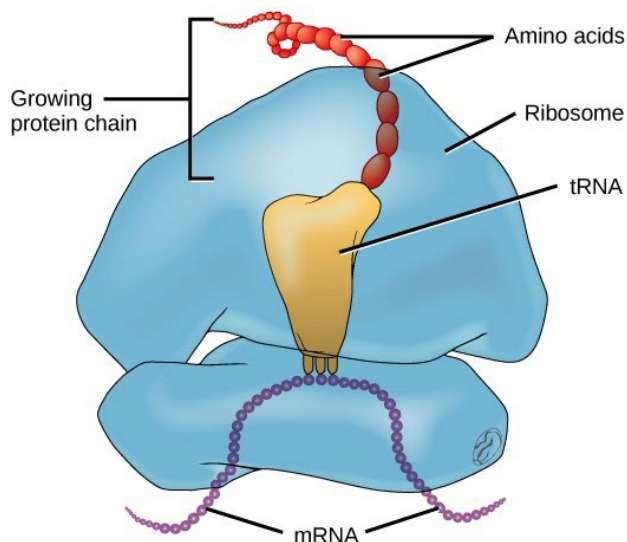
Figure 3.20 The protein synthesis machinery includes the large and small subunits of the ribosome, mRNA, and tRNA. (credit: modification of work by NIGMS, NIH)

In *E. coli*, there are 200,000 ribosomes present in every cell at any given time. A ribosome is a complex macromolecule composed of structural and catalytic rRNAs, and many distinct polypeptides. In eukaryotes, the nucleolus is completely specialized for the synthesis and assembly of rRNAs.

Ribosomes are located in the cytoplasm in prokaryotes and in the cytoplasm and endoplasmic reticulum of eukaryotes. Ribosomes are made up of a large and a small subunit that come together for translation. The small subunit is responsible for binding the mRNA template, whereas the large subunit sequentially binds **tRNAs**, a type of RNA molecule that brings amino acids to the growing chain of the polypeptide. Each mRNA molecule is simultaneously translated by many ribosomes, all synthesizing protein in the same direction.

Depending on the species, 40 to 60 types of tRNA exist in the cytoplasm. Serving as adaptors, specific tRNAs bind to sequences on the mRNA template and add the corresponding amino acid to the polypeptide chain. Therefore, tRNAs are the molecules that actually "translate" the language of RNA into the language of proteins. For each tRNA to function, it must have its specific amino acid bonded to it. In the process of tRNA "charging," each tRNA molecule is bonded to its correct amino acid.

The Genetic Code

To summarize what we know to this point, the cellular process of transcription generates messenger RNA (mRNA), a mobile molecular copy of one or more genes with an alphabet of A, C, G, and uracil (U). Translation of the mRNA template converts nucleotide-based genetic information into a protein product. Protein sequences consist of 20 commonly occurring amino acids; therefore, it can be said that the protein alphabet consists of 20 letters. Each amino acid is defined by a three-nucleotide sequence called the triplet **codon**. The relationship between a nucleotide codon and its corresponding amino acid is called the **genetic code**.

Given the different numbers of "letters" in the mRNA and protein "alphabets," combinations of nucleotides corresponded to single amino acids. Using a three-nucleotide code means that there are a total of 64 (4 × 4 × 4) possible combinations; therefore, a given amino acid is encoded by more than one nucleotide triplet (Figure 3.21).

Figure 3.21 This figure shows the genetic code for translating each nucleotide triplet, or codon, in mRNA into an amino acid or a termination signal in a nascent protein. (credit: modification of work by NIH)

Three of the 64 codons terminate protein synthesis and release the polypeptide from the translation machinery. These triplets are called **stop codons**. Another codon, AUG, also has a special function. In addition to specifying the amino acid methionine, it also serves as the **start codon** to initiate translation. The reading frame for translation is set by the AUG start codon near the 5' end of the mRNA. The genetic code is universal. With a few exceptions, virtually all species use the same genetic code for protein synthesis, which is powerful evidence that all life on Earth shares a common origin.

The Mechanism of Protein Synthesis

Just as with mRNA synthesis, protein synthesis can be divided into three phases: initiation, elongation, and termination. The process of translation is similar in prokaryotes and eukaryotes. Here we will explore how translation occurs in *E. coli*, a representative prokaryote, and specify any differences between prokaryotic and eukaryotic translation.

Protein synthesis begins with the formation of an initiation complex. In *E. coli*, this complex involves the small ribosome subunit, the mRNA template, three initiation factors, and a special initiator tRNA. The initiator tRNA interacts with the AUG start codon, and links to a special form of the amino acid methionine that is typically removed from the polypeptide after translation is complete.

In prokaryotes and eukaryotes, the basics of polypeptide elongation are the same, so we will review elongation from the perspective of *E. coli*. The large ribosomal subunit of *E. coli* consists of three compartments: the A site binds incoming charged tRNAs (tRNAs with their attached specific amino acids). The P site binds charged tRNAs carrying amino acids that have formed bonds with the growing polypeptide chain but have not yet dissociated from their corresponding tRNA. The E site releases dissociated tRNAs so they can be recharged with free amino acids. The ribosome shifts one codon at a time, catalyzing each process that occurs in the three sites. With each step, a charged tRNA enters the

78

complex, the polypeptide becomes one amino acid longer, and an uncharged tRNA departs. The energy for each bond between amino acids is derived from GTP, a molecule similar to ATP (Figure 3.22). Amazingly, the *E. coli* translation apparatus takes only 0.05 seconds to add each amino acid, meaning that a 200-amino acid polypeptide could be translated in just 10 seconds.
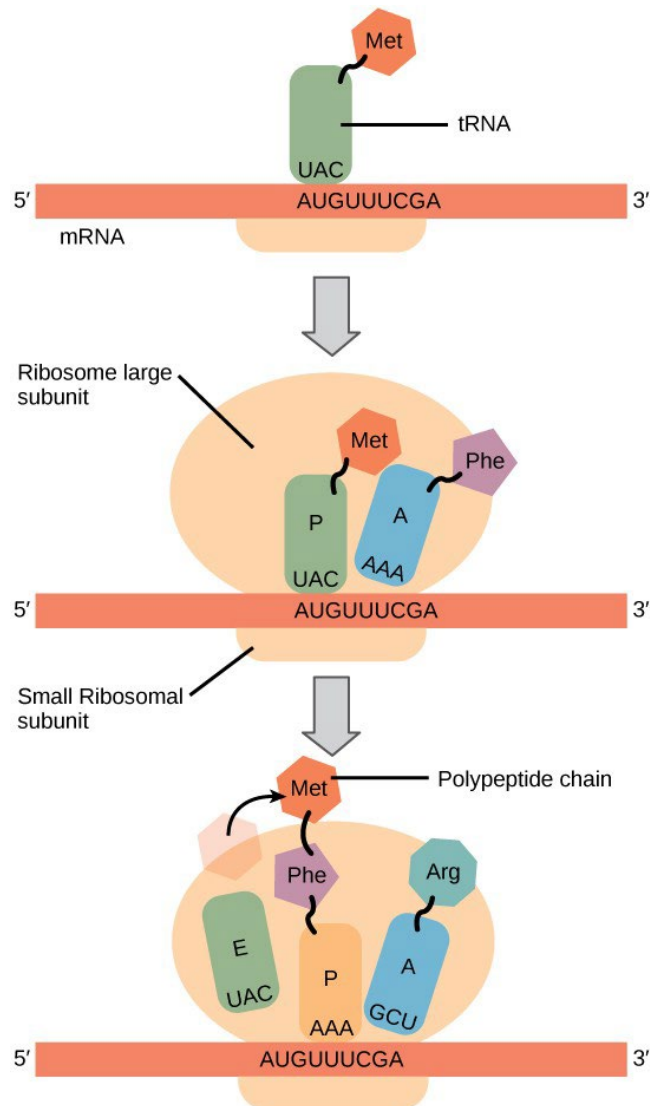
Figure 3.22 Translation begins when a tRNA anticodon recognizes a codon on the mRNA. The large ribosomal subunit joins the small subunit, and a second tRNA is recruited. As the mRNA moves relative to the ribosome, the polypeptide chain is formed. Entry of a release factor into the A site terminates translation and the components dissociate.

Termination of translation occurs when a stop codon (UAA, UAG, or UGA) is encountered. When the ribosome encounters the stop codon, the growing polypeptide is released and the ribosome subunits dissociate and leave the mRNA. After many ribosomes have completed translation, the mRNA is degraded so the nucleotides can be reused in another transcription reaction.

Transcribe a gene and translate it to protein using complementary pairing and the genetic code at **this site (http://openstaxcollege.org/l/create_protein2)** .

79

pair; the pairs are bound to each other by hydrogen bonds. The bases pair in such a way that the distance between the backbones of the two strands is the same all along the molecule.

Errors in Meiosis

By the end of this section, you will be able to:

- Explain how nondisjunction leads to disorders in chromosome number
- Describe how errors in chromosome structure occur through inversions and translocations

Inherited disorders can arise when chromosomes behave abnormally during meiosis. Chromosome disorders can be divided into two categories: abnormalities in chromosome number and chromosome structural rearrangements. Because even small segments of chromosomes can span many genes, chromosomal disorders are characteristically dramatic and often fatal.

Disorders in Chromosome Number

The isolation and microscopic observation of chromosomes forms the basis of cytogenetics and is the primary method by which clinicians detect chromosomal abnormalities in humans. A **karyotype** is the number and appearance of chromosomes, including their length, banding pattern, and centromere position. To obtain a view of an individual's karyotype, cytologists photograph the chromosomes and then cut and paste each chromosome into a chart, or **karyogram** (Figure 3.23).
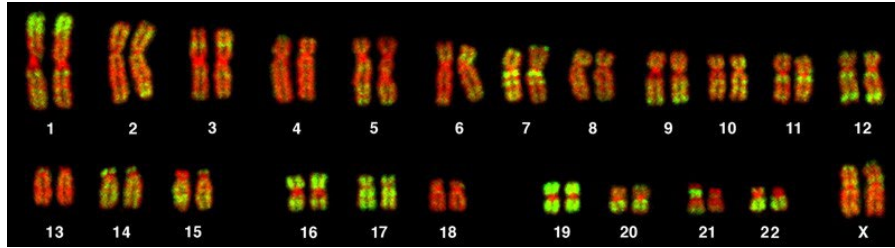
Figure 3.23 This karyogram shows the chromosomes of a female human immune cell during mitosis. (credit: Andreas Bolzer, et al)

Geneticists Use Karyograms to Identify Chromosomal Aberrations

The karyotype is a method by which traits characterized by chromosomal abnormalities can be identified from a single cell. To observe an individual's karyotype, a person's cells (like white blood cells) are first collected from a blood sample or other tissue. In the laboratory, the isolated cells are stimulated to begin actively dividing. A chemical is then applied to the cells to arrest mitosis during metaphase. The cells are then fixed to a slide.

The geneticist then stains chromosomes with one of several dyes to better visualize the distinct and reproducible banding patterns of each chromosome pair. Following staining, chromosomes are viewed using bright-field microscopy. An experienced cytogeneticist can identify each band. In addition to the banding patterns, chromosomes are further identified on the basis of size and centromere location. To obtain the classic depiction of the karyotype in which homologous pairs of chromosomes are aligned in numerical order from longest to shortest, the geneticist obtains a digital image, identifies each chromosome, and manually arranges the chromosomes into this pattern (**Figure 7.7**).

At its most basic, the karyogram may reveal genetic abnormalities in which an individual has too many or too few chromosomes per cell. Examples of this are Down syndrome, which is identified by a third copy of chromosome 21, and Turner syndrome, which is characterized by the presence of only one X chromosome in women instead of two. Geneticists can also identify large deletions or insertions of DNA. For instance, Jacobsen syndrome, which involves distinctive facial features as well as heart and bleeding defects, is identified by a deletion on chromosome 11. Finally, the karyotype can pinpoint **translocations**, which occur when a segment of genetic material breaks from one chromosome and reattaches to another chromosome or to a different part of the same chromosome. Translocations are implicated in certain cancers, including chronic myelogenous leukemia.

By observing a karyogram, geneticists can actually visualize the chromosomal composition of an individual to confirm or predict genetic abnormalities in offspring even before birth

Nondisjunctions, Duplications, and Deletions

Of all the chromosomal disorders, abnormalities in chromosome number are the most easily identifiable from a karyogram. Disorders of chromosome number include the duplication or loss of entire chromosomes, as well as changes in the number of complete sets of chromosomes. They are caused by **nondisjunction**, which occurs when pairs of homologous chromosomes or sister chromatids fail to separate during meiosis. The risk of nondisjunction increases with the age of the parents.

Nondisjunction can occur during either meiosis I or II, with different results (**Figure 3.24**). If homologous chromosomes fail to separate during meiosis I, the result is two gametes that lack that chromosome and two gametes with two copies of the chromosome. If sister chromatids fail to separate during meiosis II, the result is one gamete that lacks that chromosome, two normal gametes with one copy of the chromosome, and one gamete with two copies of the chromosome.
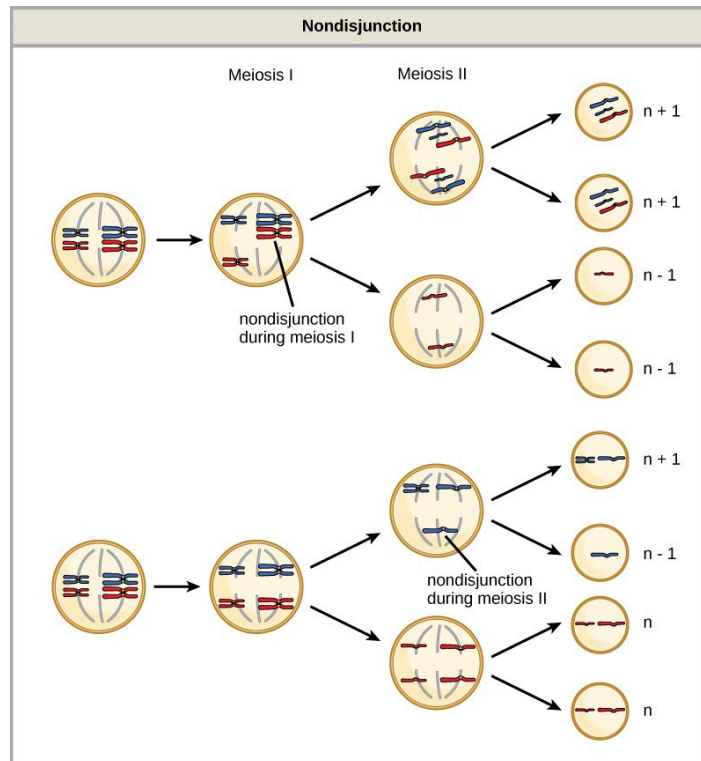
Figure 3.24 Following meiosis, each gamete has one copy of each chromosome. Nondisjunction occurs when homologous chromosomes (meiosis I) or sister chromatids (meiosis II) fail to separate during meiosis.

An individual with the appropriate number of chromosomes for their species is called **euploid**; in humans, euploidy corresponds to 22 pairs of **autosomes** and one pair of sex chromosomes. An individual with an error in chromosome number is described as **aneuploid**, a term that includes **monosomy** (loss of one chromosome) or **trisomy** (gain of an extraneous chromosome). Monosomic human zygotes missing any one copy of an autosome invariably fail to develop to birth because they have only one copy of essential genes. Most autosomal trisomies also fail to develop to birth; however, duplications of some of the smaller chromosomes (13, 15, 18, 21, or 22) can result in offspring that survive for several weeks to many years. Trisomic individuals suffer from a different type of genetic imbalance: an excess in gene dose. Cell functions are calibrated to the amount of gene product produced by two copies (doses) of each gene; adding a third copy (dose) disrupts this balance. The most common trisomy is that of chromosome 21, which leads to Down syndrome. Individuals with this inherited disorder have characteristic physical features and developmental delays in growth and cognition. The incidence of Down syndrome is correlated with maternal age, such that older women are more likely to give birth to children with Down syndrome (Figure 3.25).
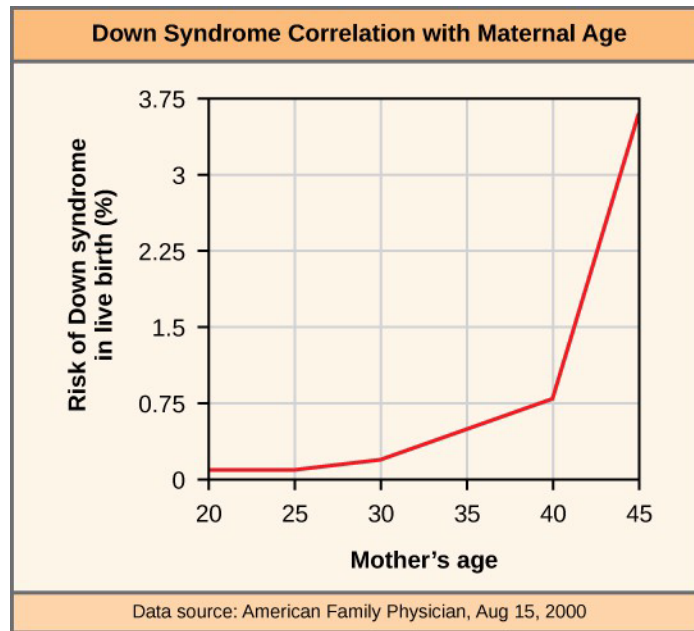
Figure 3.25 The incidence of having a fetus with trisomy 21 increases dramatically with maternal age.

Visualize the addition of a chromosome that leads to Down syndrome in this **video simulation (http://openstaxcollege.org/l/down_syndrome2)** .

Humans display dramatic deleterious effects with autosomal trisomies and monosomies. Therefore, it may seem counterintuitive that human females and males can function normally, despite carrying different numbers of the X chromosome. In part, this occurs because of a process called **X inactivation**. Early in development, when female mammalian embryos consist of just a few thousand cells, one X chromosome in each cell inactivates by condensing into a structure called a Barr body. The genes on the inactive X chromosome are not expressed. The particular X chromosome (maternally or paternally derived) that is inactivated in each cell is random, but once the inactivation occurs, all cells descended from that cell will have the same inactive X chromosome. By this process, females compensate for their double genetic dose of X chromosome.

In so-called "tortoiseshell" cats, X inactivation is observed as coat-color variegation (Figure 3.26). Females heterozygous for an X-linked coat color gene will express one of two different coat colors over different regions of their body, corresponding to whichever X chromosome is inactivated in the embryonic cell progenitor of that region. When you see a tortoiseshell cat, you will know that it has to be a female.

Figure 3.26 Embryonic inactivation of one of two different X chromosomes encoding different coat colors gives rise to the tortoiseshell phenotype in cats. (credit: Michael Bodega)

In an individual carrying an abnormal number of X chromosomes, cellular mechanisms will inactivate all but one X in each of her cells. As a result, X-chromosomal abnormalities are typically associated with mild mental and physical defects, as well as sterility. If the X chromosome is absent altogether, the individual will not develop.

Several errors in sex chromosome number have been characterized. Individuals with three X chromosomes, called triplo-X, appear female but express developmental delays and reduced fertility. The XXY chromosome complement, corresponding to one type of Klinefelter syndrome, corresponds to male individuals with small testes, enlarged breasts, and reduced body hair. The extra X chromosome undergoes inactivation to compensate for the excess genetic dosage. Turner syndrome, characterized as an X0 chromosome complement (i.e., only a single sex chromosome), corresponds to a female individual with short stature, webbed skin in the neck region, hearing and cardiac impairments, and sterility.

An individual with more than the correct number of chromosome sets (two for diploid species) is called **polyploid**. For instance, fertilization of an abnormal diploid egg with a normal haploid sperm would yield a triploid zygote. Polyploid animals are extremely rare, with only a few examples among the flatworms, crustaceans, amphibians, fish, and lizards. Triploid animals are sterile because meiosis cannot proceed normally with an odd number of chromosome sets. In contrast, polyploidy is very common in the plant kingdom, and polyploid plants tend to be larger and more robust than euploids of their species.

Chromosome Structural Rearrangements

Cytologists have characterized numerous structural rearrangements in chromosomes, including partial duplications, deletions, inversions, and translocations. Duplications and deletions often produce offspring that survive but exhibit physical and mental abnormalities. Cri-du-chat (from the French for "cry of the cat") is a syndrome associated with nervous system abnormalities and identifiable physical features that results from a deletion of most of the small arm of chromosome 5 (Figure 3.27). Infants with this genotype emit a characteristic high-pitched cry upon which the disorder's name is based.

84

Figure 3.27 This individual with cri-du-chat syndrome is shown at various ages: (A) age two, (B) age four, (C) age nine, and (D) age 12. (credit: Paola Cerruti Mainardi)

Chromosome inversions and translocations can be identified by observing cells during meiosis because homologous chromosomes with a rearrangement in one of the pair must contort to maintain appropriate gene alignment and pair effectively during prophase I.

A **chromosome inversion** is the detachment, 180° rotation, and reinsertion of part of a chromosome (Figure 3.28). Unless they disrupt a gene sequence, inversions only change the orientation of genes and are likely to have more mild effects than aneuploid errors.

- Download for free at http://openstaxcollege.org/textbooks/concepts-of- biology/get.
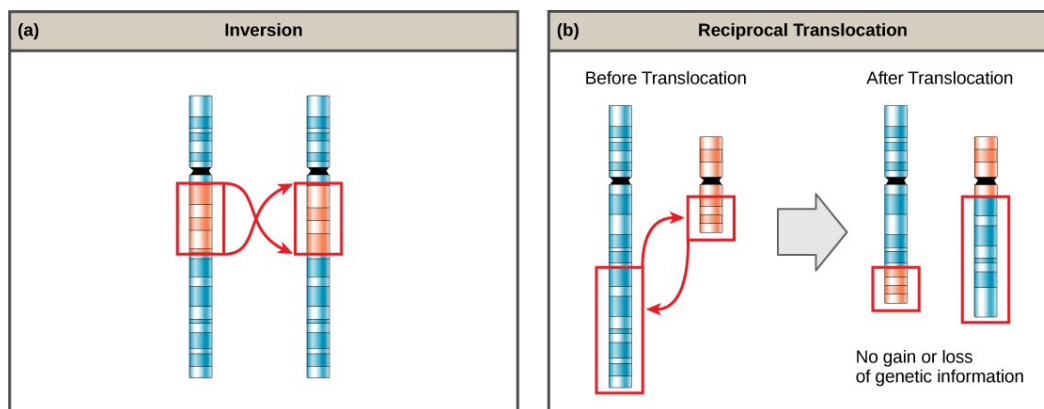
The Chromosome 18 Inversion

Not all structural rearrangements of chromosomes produce nonviable, impaired, or infertile individuals. In rare instances, such a change can result in the evolution of a new species. In fact, an inversion in chromosome 18 appears to have contributed to the evolution of humans. This inversion is not present in our closest genetic relatives, the chimpanzees.

The chromosome 18 inversion is believed to have occurred in early humans following their divergence from a common ancestor with chimpanzees approximately five million years ago. Researchers have suggested that a long stretch of DNA was duplicated on chromosome 18 of an ancestor to humans, but that during the duplication it was inverted (inserted into the chromosome in reverse orientation.

A comparison of human and chimpanzee genes in the region of this inversion indicates that two genes—*ROCK1* and *USP14*—are farther apart on human chromosome 18 than they are on the corresponding chimpanzee chromosome. This suggests that one of the inversion breakpoints occurred between these two genes. Interestingly, humans and chimpanzees express *USP14* at distinct levels in specific cell types, including cortical cells and fibroblasts. Perhaps the chromosome 18 inversion in an ancestral human repositioned specific genes and reset their expression levels in a useful way. Because both *ROCK1* and *USP14* code for enzymes, a change in their expression could alter cellular function.

It is not known how this inversion contributed to hominid evolution, but it appears to be a

A translocation occurs when a segment of a chromosome dissociates and reattaches to a different, nonhomologous chromosome. Translocations can be benign or have devastating effects, depending on how the positions of genes are altered with respect to regulatory sequences. Notably, specific translocations have been associated with several cancers and with schizophrenia. Reciprocal translocations result from the exchange of chromosome segments between two nonhomologous chromosomes such that there is no gain or loss of genetic information (Figure 3.27).

Figure 3.28 An (a) inversion occurs when a chromosome segment breaks from the chromosome, reverses its orientation, and then reattaches in the original position. A (b) reciprocal translocation occurs between two nonhomologous chromosomes and does not cause any genetic information to be lost or duplicated. (credit: modification of work by National Human Genome Research Institute (USA)

Genetics

Genetics is the study of heredity. Johann Gregor Mendel set the framework for genetics long before chromosomes or genes had been identified, at a time when meiosis was not well understood. Mendel selected a simple biological system and conducted methodical, quantitative analyses using large sample sizes. Because of Mendel's work, the fundamental principles of heredity were revealed. We now know that genes, carried on chromosomes, are the basic functional units of heredity with the ability to be replicated, expressed, or mutated. Today, the postulates put forth by Mendel form the basis of classical, or Mendelian, genetics. Not all genes are transmitted from parents to offspring according to Mendelian genetics, but Mendel's experiments serve as an excellent starting point for thinking about inheritance.

Figure 3.29 Johann Gregor Mendel set the framework for the study of genetics.

Johann Gregor Mendel (1822–1884) (Figure 3.29) was a lifelong learner, teacher, scientist, and man of faith. As a young adult, he joined the Augustinian Abbey of St. Thomas in Brno in what is now the Czech Republic. Supported by the monastery, he taught physics, botany, and natural science courses at the secondary and university levels. In 1856, he began a decade-long research pursuit involving inheritance

patterns in honeybees and plants, ultimately settling on pea plants as his primary **model system** (a system with convenient characteristics that is used to study a specific biological phenomenon to gain understanding to be applied to other systems). In 1865, Mendel presented the results of his experiments with nearly 30,000 pea plants to the local natural history society. He demonstrated that traits are transmitted faithfully from parents to offspring in specific patterns. In 1866, he published his work, [1]

Experiments in Plant Hybridization, in the proceedings of the Natural History Society of Brünn.

Mendel's work went virtually unnoticed by the scientific community, which incorrectly believed that the process of inheritance involved a blending of parental traits that produced an intermediate physical appearance in offspring. This hypothetical process appeared to be correct because of what we know now as continuous variation. **Continuous variation** is the range of small differences we see among individuals in a characteristic like human height. It does appear that offspring are a "blend" of their parents' traits when we look at characteristics that exhibit continuous variation. Mendel worked instead with traits that show **discontinuous variation**. Discontinuous variation is the variation seen among individuals when each individual shows one of two—or a very few—easily distinguishable traits, such as violet or white flowers. Mendel's choice of these kinds of traits allowed him to see experimentally that the traits were not blended in the offspring as would have been expected at the time, but that they were inherited as distinct traits. In 1868, Mendel became abbot of the monastery and exchanged his scientific pursuits for his pastoral duties. He was not recognized for his extraordinary scientific contributions during his lifetime; in fact, it was not until 1900 that his work was rediscovered, reproduced, and revitalized by scientists on the brink of discovering the chromosomal basis of heredity.

Mendel's Crosses

Mendel's seminal work was accomplished using the garden pea, *Pisum sativum*, to study inheritance. This species naturally self-fertilizes, meaning that pollen encounters ova within the same flower. The

---

1. Johann Gregor Mendel, "Versuche über Pflanzenhybriden." *Verhandlungen des naturforschenden Vereines in Brünn*, Bd. IV für das Jahr, 1865 Abhandlungen (1866):3–47. [for English translation, see http://www.mendelweb.org/Mendel.plain.html]

flower petals remain sealed tightly until pollination is completed to prevent the pollination of other plants. The result is highly inbred, or "true-breeding," pea plants. These are plants that always produce offspring that look like the parent. By experimenting with true-breeding pea plants, Mendel avoided the appearance of unexpected traits in offspring that might occur if the plants were not true breeding. The garden pea also grows to maturity within one season, meaning that several generations could be evaluated over a relatively short time. Finally, large quantities of garden peas could be cultivated simultaneously, allowing Mendel to conclude that his results did not come about simply by chance.

Mendel performed **hybridizations**, which involve mating two true-breeding individuals that have different traits. In the pea, which is naturally self-pollinating, this is done by manually transferring pollen from the anther of a mature pea plant of one variety to the stigma of a separate mature pea plant of the second variety.

Plants used in first-generation crosses were called **P**, or parental generation, plants (Figure 3.30). Mendel collected the seeds produced by the P plants that resulted from each cross and grew them the following season. These offspring were called the F_1, or the first filial (filial = daughter or son), generation. Once Mendel examined the characteristics in the F_1 generation of plants, he allowed them to self-fertilize naturally. He then collected and grew the seeds from the F_1 plants to produce the **F_2**, or second filial, generation. Mendel's experiments extended beyond the F_2 generation to the F_3 generation, F_4 generation, and so on, but it was the ratio of characteristics in the P, F_1, and F_2 generations that were the most intriguing and became the basis of Mendel's postulates.
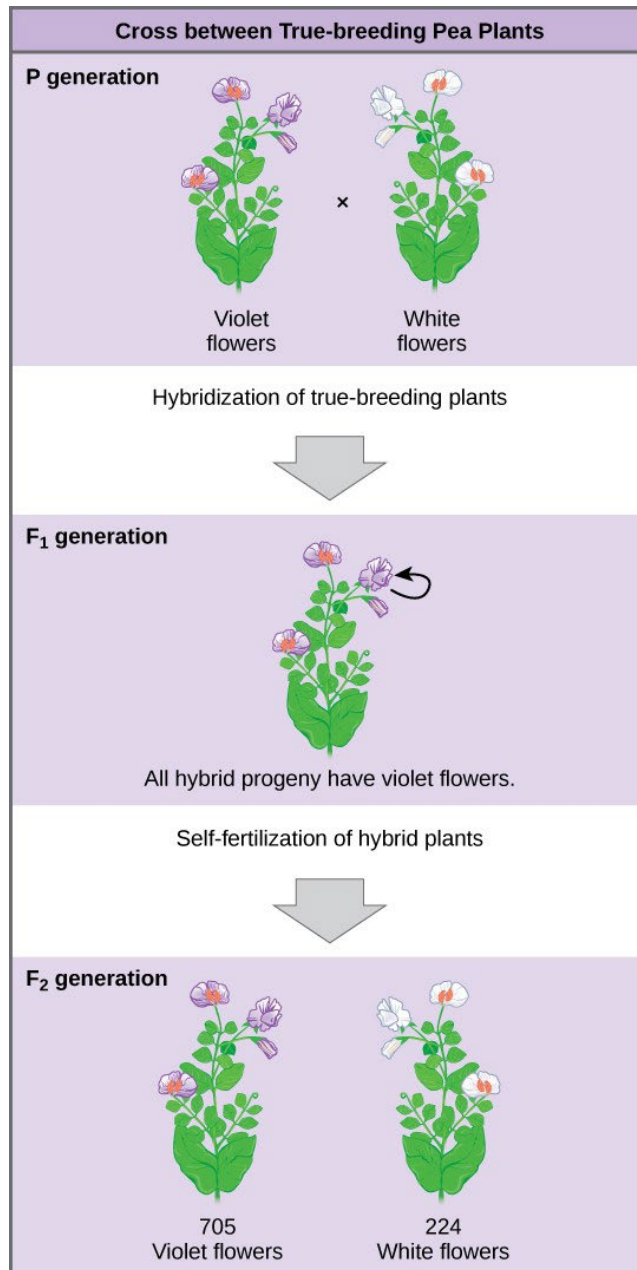
Figure 3.30 Mendel's process for performing crosses included examining flower color.

## Garden Pea Characteristics Revealed the Basics of Heredity

In his 1865 publication, Mendel reported the results of his crosses involving seven different characteristics, each with two contrasting traits. A **trait** is defined as a variation in the physical appearance of a heritable characteristic. The characteristics included plant height, seed texture, seed color, flower color, pea-pod size, pea-pod color, and flower position. For the characteristic of flower color, for example, the two contrasting traits were white versus violet. To fully examine each characteristic, Mendel generated large numbers of F_1 and F_2 plants and reported results from thousands of F_2 plants.

What results did Mendel find in his crosses for flower color? First, Mendel confirmed that he was using plants that bred true for white or violet flower color. Irrespective of the number of generations that Mendel examined, all self-crossed offspring of parents with white flowers had white flowers, and all self-crossed offspring of parents with violet flowers had violet flowers. In addition, Mendel confirmed that, other than flower color, the pea plants were physically identical. This was an important check to make sure that the two varieties of pea plants only differed with respect to one trait, flower color.

Once these validations were complete, Mendel applied the pollen from a plant with violet flowers to the stigma of a plant with white flowers. After gathering and sowing the seeds that resulted from this cross, Mendel found that 100 percent of the F_1 hybrid generation had violet flowers. Conventional wisdom at that time would have predicted the hybrid flowers to be pale violet or for hybrid plants to have equal numbers of white and violet flowers. In other words, the contrasting parental traits were expected to blend in the offspring. Instead, Mendel's results demonstrated that the white flower trait had completely disappeared in the F_1 generation.

Importantly, Mendel did not stop his experimentation there. He allowed the F_1 plants to self-fertilize and found that 705 plants in the F_2 generation had violet flowers and 224 had white flowers. This was a ratio of 3.15 violet flowers to one white flower, or approximately 3:1. When Mendel transferred pollen from a plant with violet flowers to the stigma of a plant with white flowers and vice versa, he obtained approximately the same ratio irrespective of which parent—male or female—contributed which trait. This is called a **reciprocal cross**—a paired cross in which the respective traits of the male and female in one cross become the respective traits of the female and male in the other cross. For the other six characteristics that Mendel examined, the F_1 and F_2 generations behaved in the same way that they behaved for flower color. One of the two traits would disappear completely from the F_1 generation, only to reappear in the F_2 generation at a ratio of roughly 3:1 (Figure 3.31).
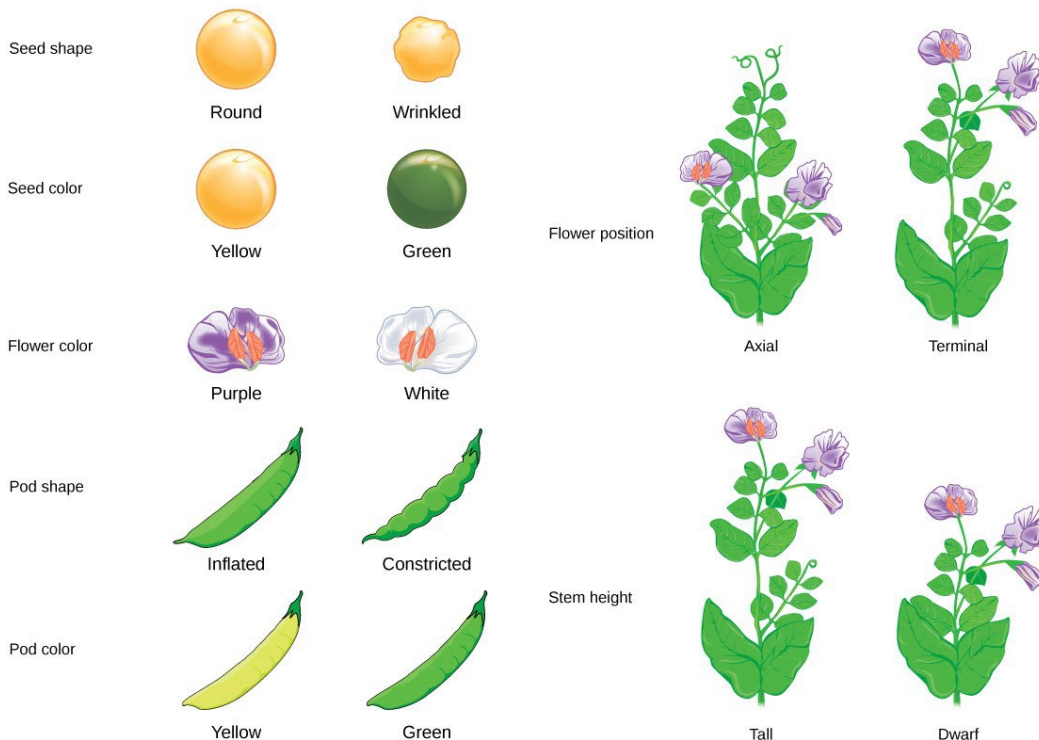
Figure 3.31 Mendel identified seven pea plant characteristics.

Upon compiling his results for many thousands of plants, Mendel concluded that the characteristics could be divided into expressed and latent traits. He called these dominant and recessive traits, respectively. **Dominant** traits are those that are inherited unchanged in a hybridization. **Recessive** traits become latent, or disappear in the offspring of a hybridization. The recessive trait does, however, reappear in the progeny of the hybrid offspring. An example of a dominant trait is the violet-colored flower trait. For this same characteristic (flower color), white-colored flowers are a recessive trait. The fact that the recessive trait reappeared in the F_2 generation meant that the traits remained separate (and were not blended) in the plants of the F_1 generation. Mendel proposed that this was because the plants possessed two copies of the trait for the flower-color characteristic, and that each parent transmitted one of their two copies to their offspring, where they came together. Moreover, the physical observation of a dominant trait could mean that the genetic composition of the organism included two dominant versions of the characteristic, or that it included one dominant and one recessive version. Conversely, the observation of a recessive trait meant that the organism lacked any dominant versions of this characteristic.

For an excellent review of Mendel's experiments and to perform your own crosses and identify patterns of inheritance, visit the **Mendel's Peas (http://openstaxcollege.org/l/mendels_peas)** web lab.

Laws of Inheritance

The seven characteristics that Mendel evaluated in his pea plants were each expressed as one of two versions, or traits. Mendel deduced from his results that each individual had two discrete copies of the characteristic that are passed individually to offspring. We now call those two copies genes, which are carried on chromosomes. The reason we have two copies of each gene is that we inherit one from each parent. In fact, it is the chromosomes we inherit and the two copies of each gene are located on paired chromosomes. Recall that in meiosis these chromosomes are separated out into haploid gametes. This separation, or segregation, of the homologous chromosomes means also that only one of the copies of the gene gets moved into a gamete. The offspring are formed when that gamete unites with one from another parent and the two copies of each gene (and chromosome) are restored.

For cases in which a single gene controls a single characteristic, a diploid organism has two genetic copies that may or may not encode the same version of that characteristic. For example, one individual may carry a gene that determines white flower color and a gene that determines violet flower color. Gene variants that arise by mutation and exist at the same relative locations on homologous chromosomes are called **alleles**. Mendel examined the inheritance of genes with just two allele forms, but it is common to encounter more than two alleles for any given gene in a natural population.

Phenotypes and Genotypes

Two alleles for a given gene in a diploid organism are expressed and interact to produce physical characteristics. The observable traits expressed by an organism are referred to as its **phenotype**. An organism's underlying genetic makeup, consisting of both the physically visible and the non-expressed alleles, is called its **genotype**. Mendel's hybridization experiments demonstrate the difference between phenotype and genotype. For example, the phenotypes that Mendel observed in his crosses between pea plants with differing traits are connected to the diploid genotypes of the plants in the P, F_1, and F_2 generations. We will use a second trait that Mendel investigated, seed color, as an example. Seed color is governed by a single gene with two alleles. The yellow-seed allele is dominant and the green-seed allele is recessive. When true-breeding plants were cross-fertilized, in which one parent had yellow seeds and one had green seeds, all of the F_1 hybrid offspring had yellow seeds. That is, the hybrid offspring were phenotypically identical to the true-breeding parent with yellow seeds. However, we know that the allele donated by the parent with green seeds was not simply lost because it reappeared in some of the F_2 offspring (Figure 3.32). Therefore, the F_1 plants must have been genotypically different from the parent with yellow seeds.

The P plants that Mendel used in his experiments were each homozygous for the trait he was studying. Diploid organisms that are **homozygous** for a gene have two identical alleles, one on each of their homologous chromosomes. The genotype is often written as *YY* or *yy*, for which each letter represents one of the two alleles in the genotype. The dominant allele is capitalized and the recessive allele is lower case. The letter used for the gene (seed color in this case) is usually related to the dominant trait (yellow allele, in this case, or "*Y*"). Mendel's parental pea plants always bred true because both produced gametes carried the same allele. When P plants with contrasting traits were cross-fertilized, all of the offspring were **heterozygous** for the contrasting trait, meaning their genotype had different alleles for the gene being examined. For example, the F₁ yellow plants that received a *Y* allele from their yellow parent and a *y* allele from their green parent had the genotype *Yy*.
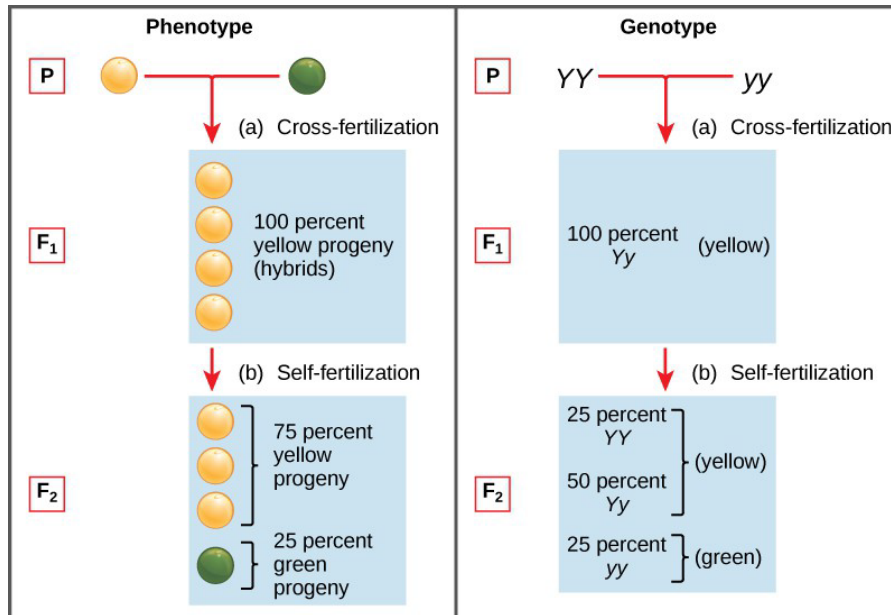
Figure 3.32 Phenotypes are physical expressions of traits that are transmitted by alleles. Capital letters represent dominant alleles and lowercase letters represent recessive alleles. The phenotypic ratios are the ratios of visible characteristics. The genotypic ratios are the ratios of gene combinations in the offspring, and these are not always distinguishable in the phenotypes.

Law of Dominance

Our discussion of homozygous and heterozygous organisms brings us to why the F₁ heterozygous offspring were identical to one of the parents, rather than expressing both alleles. In all seven pea-plant characteristics, one of the two contrasting alleles was dominant, and the other was recessive. Mendel called the dominant allele the expressed unit factor; the recessive allele was referred to as the latent unit factor. We now know that these so-called unit factors are actually genes on homologous chromosomes. For a gene that is expressed in a dominant and recessive pattern, homozygous dominant and heterozygous organisms will look identical (that is, they will have different genotypes but the same phenotype), and the recessive allele will only be observed in homozygous recessive individuals (**Table** 3.3).

Correspondence between Genotype and Phenotype for a Dominant- Recessive Characteristic.

	Homozygous	**Heterozygous**	**Homozygous**
Genotype	*YY*	*Yy*	*yy*
Phenotype	yellow	yellow	green

Table 3.3

Mendel's **law of dominance** states that in a heterozygote, one trait will conceal the presence of another trait for the same characteristic. For example, when crossing true-breeding violet-flowered plants with true-breeding white-flowered plants, all of the offspring were violet-flowered, even though they all had one allele for violet and one allele for white. Rather than both alleles contributing to a phenotype,

93

the dominant allele will be expressed exclusively. The recessive allele will remain latent, but will be transmitted to offspring in the same manner as that by which the dominant allele is transmitted. The recessive trait will only be expressed by offspring that have two copies of this allele (Figure 3.33), and these offspring will breed true when self-crossed.

Figure 3.33 The allele for albinism, expressed here in humans, is recessive. Both of this child's parents carried the recessive allele.

Extensions of the Laws of Inheritance

Mendel studied traits with only one mode of inheritance in pea plants. The inheritance of the traits he studied all followed the relatively simple pattern of dominant and recessive alleles for a single characteristic. There are several important modes of inheritance, discovered after Mendel's work, that do not follow the dominant and recessive, single-gene model.

Alternatives to Dominance and Recessiveness

Mendel's experiments with pea plants suggested that: 1) two types of "units" or alleles exist for every gene; 2) alleles maintain their integrity in each generation (no blending); and 3) in the presence of the dominant allele, the recessive allele is hidden, with no contribution to the phenotype. Therefore, recessive alleles can be "carried" and not expressed by individuals. Such heterozygous individuals are sometimes referred to as "carriers." Since then, genetic studies in other organisms have shown that much more complexity exists, but that the fundamental principles of Mendelian genetics still hold true. In the sections to follow, we consider some of the extensions of Mendelism.

Incomplete Dominance

Mendel's results, demonstrating that traits are inherited as dominant and recessive pairs, contradicted the view at that time that offspring exhibited a blend of their parents' traits. However, the heterozygote

94

phenotype occasionally does appear to be intermediate between the two parents. For example, in the snapdragon, *Antirrhinum majus* (Figure 3,34), a cross between a homozygous parent with white flowers ($C^W C^W$) and a homozygous parent with red flowers ($C^R C^R$) will produce offspring with pink flowers ($C^R C^W$). (Note that different genotypic abbreviations are used for Mendelian extensions to distinguish these patterns from simple dominance and recessiveness.) This pattern of inheritance is described as **incomplete dominance**, meaning that one of the alleles appears in the phenotype in the heterozygote, but not to the exclusion of the other, which can also be seen. The allele for red flowers is incompletely dominant over the allele for white flowers. However, the results of a heterozygote self-cross can still be predicted, just as with Mendelian dominant and recessive crosses. In this case, the genotypic ratio would be 1 $C^R C^R$:2 $C^R C^W$:1 $C^W C^W$, and the phenotypic ratio would be 1:2:1 for red:pink:white. The basis for the intermediate color in the heterozygote is simply that the pigment produced by the red allele (anthocyanin) is diluted in the heterozygote and therefore appears pink because of the white background of the flower petals.

Figure 3.34 These pink flowers of a heterozygote snapdragon result from incomplete dominance. (credit: "storebukkebruse"/Flickr)

Codominance

A variation on incomplete dominance is **codominance**, in which both alleles for the same characteristic are simultaneously expressed in the heterozygote. An example of codominance occurs in the ABO blood groups of humans. The A and B alleles are expressed in the form of A or B molecules present on the surface of red blood cells. Homozygotes ($I^A I^A$ and $I^B I^B$) express either the A or the B phenotype, and heterozygotes ($I^A I^B$) express both phenotypes equally. The $I^A I^B$ individual has blood type AB. In a self-cross between heterozygotes expressing a codominant trait, the three possible offspring genotypes are phenotypically distinct. However, the 1:2:1 genotypic ratio characteristic of a Mendelian monohybrid cross still applies (Figure 3.35).

- Download for free at http://openstaxcollege.org/textbooks/concepts-of- biology/get.
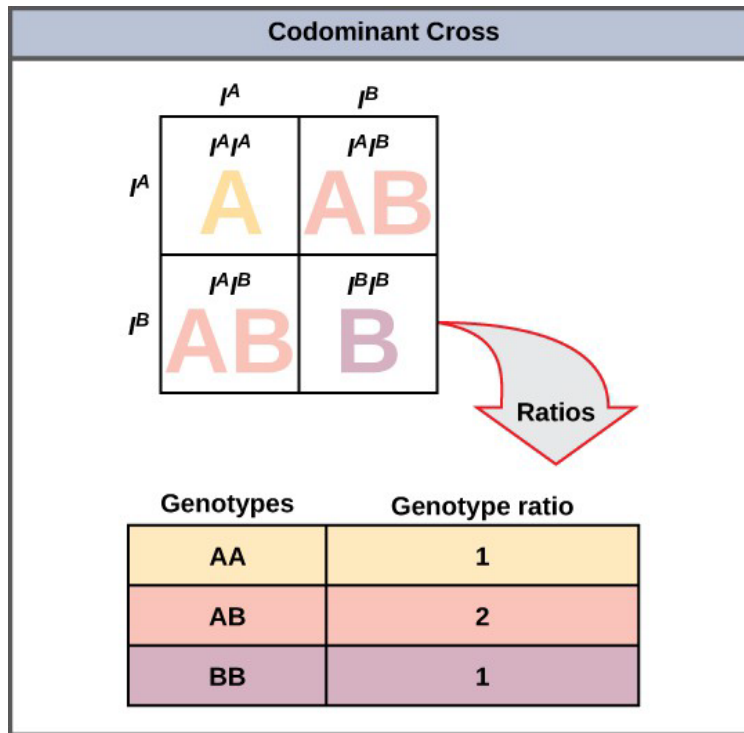
Figure 3.35 This Punnet square shows an AB/AB blood type cross

Multiple Alleles

Mendel implied that only two alleles, one dominant and one recessive, could exist for a given gene. We now know that this is an oversimplification. Although individual humans (and all diploid organisms) can only have two alleles for a given gene, multiple alleles may exist at the population level, such that many combinations of two alleles are observed. Note that when many alleles exist for the same gene, the convention is to denote the most common phenotype or genotype in the natural population as the **wild type** (often abbreviated "+"). All other phenotypes or genotypes are considered variants (mutants) of this typical form, meaning they deviate from the wild type. The variant may be recessive or dominant to the wild-type allele.

An example of multiple alleles is the ABO blood-type system in humans. In this case, there are three alleles circulating in the population. The I^A allele codes for A molecules on the red blood cells, the I^B allele codes for B molecules on the surface of red blood cells, and the i allele codes for no molecules on the red blood cells. In this case, the I^A and I^B alleles are codominant with each other and are both dominant over the i allele. Although there are three alleles present in a population, each individual only gets two of the alleles from their parents. This produces the genotypes and phenotypes shown in **Figure 3.36**. Notice that instead of three genotypes, there are six different genotypes when there are three alleles. The number of possible phenotypes depends on the dominance relationships between the three alleles.

96

Figure 3.36 Inheritance of the ABO blood system in humans is shown.

evolution IN ACTION

Multiple Alleles Confer Drug Resistance in the Malaria Parasite

Malaria is a parasitic disease in humans that is transmitted by infected female mosquitoes, including *Anopheles gambiae*, and is characterized by cyclic high fevers, chills, flu-like symptoms, and severe anemia. *Plasmodium falciparum* and *P. vivax* are the most common causative agents of malaria, and *P. falciparum* is the most deadly. When promptly and correctly treated, *P. falciparum* malaria has a mortality rate of 0.1 percent. However, in some parts of the world, the parasite has evolved resistance to commonly used malaria treatments, so the most effective malarial treatments can vary by geographic region.

In Southeast Asia, Africa, and South America, *P. falciparum* has developed resistance to the anti-malarial drugs chloroquine, mefloquine, and sulfadoxine-pyrimethamine. *P. falciparum*, which is haploid during the life stage in which it is infective to humans, has evolved multiple drug-resistant mutant alleles of the *dhps* gene. Varying degrees of sulfadoxine resistance are associated with each of these alleles. Being haploid, *P. falciparum* needs only one drug-resistant allele to express this trait.

In Southeast Asia, different sulfadoxine-resistant alleles of the *dhps* gene are localized to different geographic regions. This is a common evolutionary phenomenon that comes about because drug-resistant mutants arise in a population and interbreed with other *P. falciparum* isolates in close proximity. Sulfadoxine-resistant parasites cause considerable human hardship in regions in which this drug is widely used as an over-the-counter malaria remedy. As is common with pathogens that multiply to large numbers within an infection cycle, *P. falciparum* evolves relatively rapidly (over a decade or so) in response to the selective pressure of commonly used anti-malarial drugs. For this reason, scientists must constantly work to develop new drugs or drug combinations to combat the worldwide

[2]

Sex-Linked Traits

In humans, as well as in many other animals and some plants, the sex of the individual is determined by sex chromosomes—one pair of non-homologous chromosomes. Until now, we have only considered inheritance patterns among non-sex chromosomes, or autosomes. In addition to 22 homologous pairs of autosomes, human

females have a homologous pair of X chromosomes, whereas human males have

2. Sumiti Vinayak et al., "Origin and Evolution of Sulfadoxine Resistant *Plasmodium falciparum*," *PLoS Pathogens* 6 (2010): e1000830.

an XY chromosome pair. Although the Y chromosome contains a small region of similarity to the X chromosome so that they can pair during meiosis, the Y chromosome is much shorter and contains fewer genes. When a gene being examined is present on the X, but not the Y, chromosome, it is **X-linked**.

Eye color in *Drosophila*, the common fruit fly, was the first X-linked trait to be identified. Thomas Hunt Morgan mapped this trait to the X chromosome in 1910. Like humans, *Drosophila* males have an XY chromosome pair, and females are XX. In flies the wild-type eye color is red (X^W) and is dominant to white eye color (X^w) (Figure 3.37). Because of the location of the eye-color gene, reciprocal crosses do not produce the same offspring ratios. Males are said to be **hemizygous**, in that they have only one allele for any X-linked characteristic. Hemizygosity makes descriptions of dominance and recessiveness irrelevant for XY males. *Drosophila* males lack the white gene on the Y chromosome; that is, their genotype can only be $X^W Y$ or $X^w Y$. In contrast, females have two allele copies of this gene and can be $X^W X^W$, $X^W X^w$, or $X^w X^w$.

Figure 3.37 In *Drosophila*, the gene for eye color is located on the X chromosome. Red eye color is wild-type and is dominant to white eye color.

In an X-linked cross, the genotypes of F_1 and F_2 offspring depend on whether the recessive trait was expressed by the male or the female in the P generation. With respect to *Drosophila* eye color, when the P male expresses the white-eye phenotype and the female is homozygously red-eyed, all members of the F_1 generation exhibit red eyes (Figure 3.38). The F_1 females are heterozygous ($X^W X^w$), and the males are all $X^W Y$, having received their X chromosome from the homozygous dominant P female and their Y chromosome from the P male. A subsequent cross between the $X^W X^w$ female and the $X^W Y$ male would produce only red-eyed females (with $X^W X^W$ or $X^W X^w$ genotypes) and both red- and white-eyed males (with $X^W Y$ or $X^w Y$ genotypes). Now, consider a cross between a homozygous white-eyed female and a male with red eyes. The F_1 generation would exhibit only heterozygous red-eyed females ($X^W X^w$) and only white-eyed males ($X^w Y$). Half of the F_2 females would be red-eyed ($X^W X^w$) and half would be white-eyed ($X^w X^w$). Similarly, half of the F_2 males would be red-eyed ($X^W Y$) and half would be white-eyed ($X^w Y$).
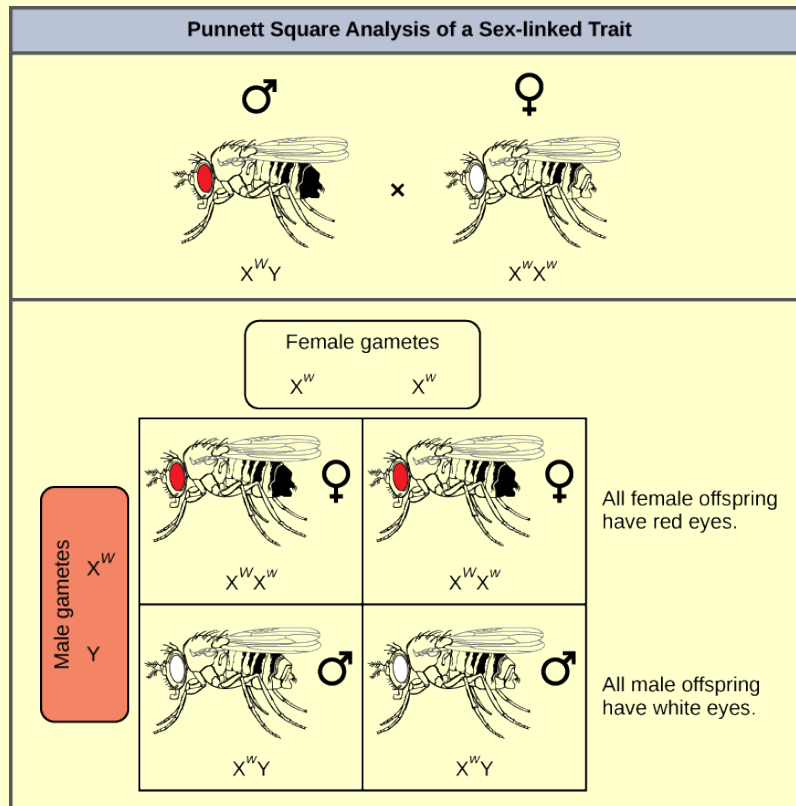
Figure 3.38 Crosses involving sex-linked traits often give rise to different phenotypes for the different sexes of offspring, as is the case for this cross involving red and white eye color in *Drosophila*. In the diagram, *w* is the white-eye mutant allele and *W* is the wild-type, red-eye allele.

Discoveries in fruit fly genetics can be applied to human genetics. When a female parent is homozygous for a recessive X-linked trait, she will pass the trait on to 100 percent of her male offspring, because the males will receive the Y chromosome from the male parent. In humans, the alleles for certain conditions (some color-blindness, hemophilia, and muscular dystrophy) are X-linked. Females who are heterozygous for these diseases are said to be carriers and may not exhibit any phenotypic effects. These females will pass the disease to half of their sons and will pass carrier status to half of their daughters; therefore, X-linked traits appear more frequently in males than females.

In some groups of organisms with sex chromosomes, the sex with the non-homologous sex chromosomes is the female rather than the male. This is the case for all birds. In this case, sex-linked traits will be more likely to appear in the female, in whom they are hemizygous.

99

Watch **this video (http://openstaxcollege.org/l/sex-linked_trts)** to learn more about sex-linked traits.

Summary

The Genome

Prokaryotes have a single loop chromosome, whereas eukaryotes have multiple, linear chromosomes surrounded by a nuclear membrane. Human somatic cells have 46 chromosomes consisting of two sets of 22 homologous chromosomes and a pair of nonhomologous sex chromosomes. This is the 2n, or diploid, state. Human gametes have 23 chromosomes or one complete set of chromosomes. This is the n, or haploid, state. Genes are segments of DNA that code for a specific protein or RNA molecule. An organism's traits are determined in large part by the genes inherited from each parent, but also by the environment that they experience. Genes are expressed as characteristics of the organism and each characteristic may have different variants called traits that are caused by differences in the DNA sequence for a gene.

Errors in Meiosis

The number, size, shape, and banding pattern of chromosomes make them easily identifiable in a karyogram and allow for the assessment of many chromosomal abnormalities. Disorders in chromosome number, or aneuploidies, are typically lethal to the embryo, although a few trisomic genotypes are viable. Because of X inactivation, aberrations in sex chromosomes typically have milder effects on an individual. Aneuploidies also include instances in which segments of a chromosome are duplicated or deleted. Chromosome structures also may be rearranged, for example by inversion or translocation. Both of these aberrations can result in negative effects on development, or death. Because they force chromosomes to assume contorted pairings during meiosis I, inversions and translocations are often associated with reduced fertility because of the likelihood of nondisjunction.

Mendel's Experiments

Working with garden pea plants, Mendel found that crosses between parents that differed for one trait produced F1 offspring that all expressed one parent's traits. The traits that were visible in the F1 generation are referred to as dominant, and traits that disappear in the F1 generation are described as recessive. When the F1 plants in Mendel's experiment were self-crossed, the F2 offspring exhibited the dominant trait or the recessive trait in a 3:1 ratio, confirming that the recessive trait had been transmitted faithfully from the original P parent. Reciprocal crosses generated identical F1 and F2 offspring ratios. By examining sample sizes, Mendel showed that traits were inherited as independent events.

When true-breeding, or homozygous, individuals that differ for a certain trait are crossed, all of the offspring will be heterozygous for that trait. If the traits are inherited as dominant and recessive, the F1 offspring will all exhibit the same phenotype as the parent homozygous for the dominant trait. If these heterozygous offspring are self-crossed, the resulting F2 offspring will be equally likely to inherit gametes carrying the dominant or recessive trait, giving rise to offspring of which one quarter are homozygous dominant, half are heterozygous, and one quarter are homozygous recessive.

Because homozygous dominant and heterozygous individuals are phenotypically identical, the observed traits in the F2 offspring will exhibit a ratio of three dominant to one recessive

Mendel postulated that genes (characteristics) are inherited as pairs of alleles (traits) that behave in a dominant and recessive pattern. Alleles segregate into gametes such that each gamete is equally likely to receive either one of the two alleles present in a diploid individual. In addition, genes are assorted into gametes independently of one another. That is, in general, alleles are not more likely to segregate into a gamete with a particular allele of another gene.

Extensions of the Laws of Inheritance

Alleles do not always behave in dominant and recessive patterns. Incomplete dominance describes situations in which the heterozygote exhibits a phenotype that is intermediate between the homozygous phenotypes. Codominance describes the simultaneous expression of both of the alleles in the heterozygote. Although diploid organisms can only have two alleles for any given gene, it is common for more than two alleles for a gene to exist in a population. In humans, as in many animals and some plants, females have two X chromosomes and males have one X and one Y chromosome. Genes that are present on the X but not the Y chromosome are said to be X-linked, such that males only inherit one allele for the gene, and females inherit two.

The Structure of DNA

The model of the double-helix structure of DNA was proposed by Watson and Crick. The DNA molecule is a polymer of nucleotides. Each nucleotide is composed of a nitrogenous base, a five-carbon sugar (deoxyribose), and a phosphate group. There are four nitrogenous bases in DNA, two purines (adenine and guanine) and two pyrimidines (cytosine and thymine). A DNA molecule is composed of two strands. Each strand is composed of nucleotides bonded together covalently between the phosphate group of one and the deoxyribose sugar of the next. From this backbone extend the bases. The bases of one strand bond to the bases of the second strand with hydrogen bonds. Adenine always bonds with thymine, and cytosine always bonds with guanine. The bonding causes the two strands to spiral around each other in a shape called a double helix. Ribonucleic acid (RNA) is a second nucleic acid found in cells. RNA is a single-stranded polymer of nucleotides. It also differs from DNA in that it contains the sugar ribose, rather than deoxyribose, and the nucleotide uracil rather than thymine. Various RNA molecules function in the process of forming proteins from the genetic code in DNA.

Prokaryotes contain a single, double-stranded circular chromosome. Eukaryotes contain double-stranded linear DNA molecules packaged into chromosomes. The DNA helix is wrapped around proteins to form nucleosomes. The protein coils are further coiled, and during mitosis and meiosis, the chromosomes become even more greatly coiled to facilitate their movement. Chromosomes have two distinct regions which can be distinguished by staining, reflecting different degrees of packaging and determined by whether the DNA in a region is being expressed (euchromatin) or not (heterochromatin).

Transcription

In prokaryotes, mRNA synthesis is initiated at a promoter sequence on the DNA template. Elongation synthesizes new mRNA. Termination liberates the mRNA and occurs by mechanisms that stall the RNA polymerase and cause it to fall off the DNA template. Newly transcribed eukaryotic mRNAs are modified with a cap and a poly-A tail. These structures protect the mature mRNA from degradation and help export it from the nucleus. Eukaryotic mRNAs also undergo splicing, in which introns are removed and exons are reconnected with single-nucleotide accuracy. Only finished mRNAs are exported from the nucleus to the cytoplasm.

Translation

The central dogma describes the flow of genetic information in the cell from genes to mRNA to proteins. Genes are used to make mRNA by the process of transcription; mRNA is used to synthesize proteins by the process of translation. The genetic code is the correspondence between the three- nucleotide mRNA codon and an amino acid. The genetic code is "translated" by the tRNA molecules, which associate a specific codon with a specific amino acid. The genetic code is degenerate because 64 triplet codons in mRNA specify only 20 amino acids and three stop codons.

This means that more than one codon corresponds to an amino acid. Almost every species on the planet uses the same genetic code.

Review Questions

1.   Which of the following does cytosine pair with?

a.   guanine

b.   thymine

c.   adenine

d.   a pyrimidine

2.   Prokaryotes contain a _____chromosome, and eukaryotes contain _____chromosomes.

a.   single-stranded circular; single-stranded linear

b.   single-stranded linear; single-stranded circular

c.   double-stranded circular; double- stranded linear

d.   double-stranded linear; double-stranded circular

3.   How long would the peptide be that is translated from this MRNA sequence: 5'-AUGGGCUACCGA-3'?

a.   0

b.   2

c.   3

d.   4

4.   The observable traits expressed by an organism are described as its   .

a.   phenotype

b.   genotype

c.   alleles

d.   zygote

5.   If black and white true-breeding mice are mated and the result is all gray offspring, what inheritance pattern would this be indicative of?

a.   dominance

b.   codominance

c.   multiple alleles

d.   incomplete dominance

6.   If a muscle cell of a typical organism has 32 chromosomes, how many chromosomes will be in a gamete of that same organism?

a.   8

b.   16

c.   32

d.   64

7.   The genotype XXY corresponds to:

a.   Klinefelter syndrome

b.   Turner syndrome

c.   Triplo-X

d.   Jacob syndrome

OpenStax College, *Concepts of Biology*. OpenStax College. 25 April 2013. <http://openstaxcollege.org/textbooks/concepts-of-biology/get>.

Chapter 4

Prokaryotes, Protists, Fungi and Plants
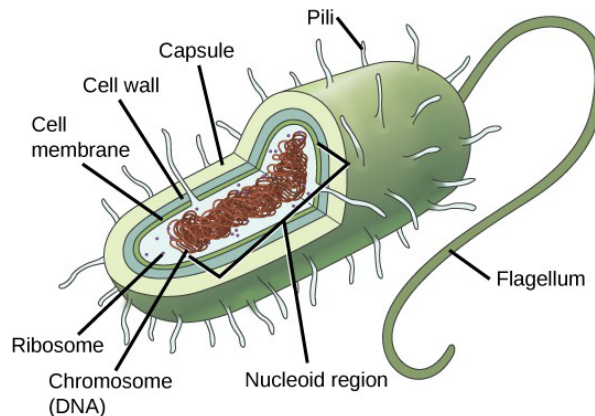
Characteristics of Prokaryotes

There are many differences between prokaryotic and eukaryotic cells. However, all cells have four common structures: a plasma membrane that functions as a barrier for the cell and separates the cell from its environment; cytoplasm, a jelly-like substance inside the cell; genetic material (DNA and RNA); and ribosomes, where protein synthesis takes place. Prokaryotes come in various shapes, but many fall into three categories: cocci (spherical), bacilli (rod-shaped), and spirilla (spiral-shaped) (Figure 4.1).
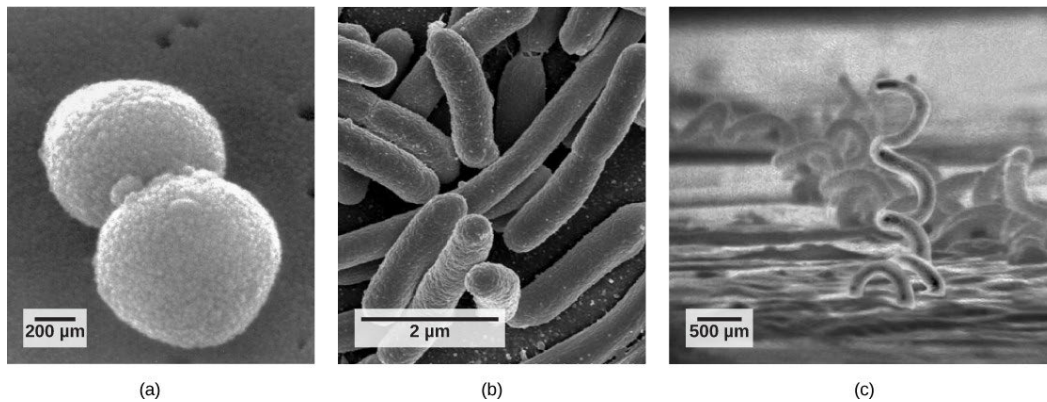
Figure 4.1 Many prokaryotes fall into three basic categories based on their shape: (a) cocci, or spherical; (b) bacilli, or rod-shaped; and (c) spirilla, or spiral-shaped. (credit a: modification of work by Janice Haney Carr, Dr. Richard Facklam, CDC; credit c: modification of work by Dr. David Cox, CDC; scale-bar data from Matt Russell)

The Prokaryotic Cell

Recall that prokaryotes (Figure 4.2) are unicellular organisms that lack organelles surrounded by membranes. Therefore, they do not have a nucleus but instead have a single chromosome—a piece of circular DNA located in an area of the cell called the nucleoid. Most prokaryotes have a cell wall lying outside the plasma membrane. The composition of the cell wall differs significantly between the domains Bacteria and Archaea (and their cell walls also differ from the eukaryotic cell walls found in plants and fungi.) The cell wall functions as a protective layer and is responsible for the organism's shape. Some other structures are present in some prokaryotic species, but not in others. For example, the **capsule** found in some species enables the organism to attach to surfaces and protects it from dehydration. Some species may also have flagella (singular, flagellum) used for locomotion, and pili (singular, pilus) used for attachment to surfaces and to other bacteria for conjugation. Plasmids, which consist of small, circular pieces of DNA outside of the main chromosome, are also present in many species of bacteria.

**Figure 4.2** The features of a typical bacterium cell are shown.

Both Bacteria and Archaea are types of prokaryotic cells. They differ in the lipid composition of their cell membranes and in the characteristics of their cell walls. Both types of prokaryotes have the same basic structures, but these are built from different chemical components that are evidence of an ancient separation of their lineages. The archaeal plasma membrane is chemically different from the bacterial membrane; some archaeal membranes are lipid monolayers instead of phospholipid bilayers.

The Cell Wall

The cell wall is a protective layer that surrounds some prokaryotic cells and gives them shape and rigidity. It is located outside the cell membrane and prevents osmotic lysis (bursting caused by increasing volume). The chemical compositions of the cell walls vary between Archaea and Bacteria, as well as between bacterial species. Bacterial cell walls contain **peptidoglycan**, composed of polysaccharide chains cross-linked to peptides. Bacteria are divided into two major groups: **Gram-positive** and **Gram- negative**, based on their reaction to a procedure called Gram staining. The different bacterial responses to the staining procedure are caused by cell wall structure. Gram-positive organisms have a thick wall consisting of many layers of peptidoglycan. Gram-negative bacteria have a thinner cell wall composed

- Download for free at http://openstaxcollege.org/textbooks/concepts-of- biology/get.

of a few layers of peptidoglycan and additional structures, surrounded by an outer membrane (**Figure 4.3**).
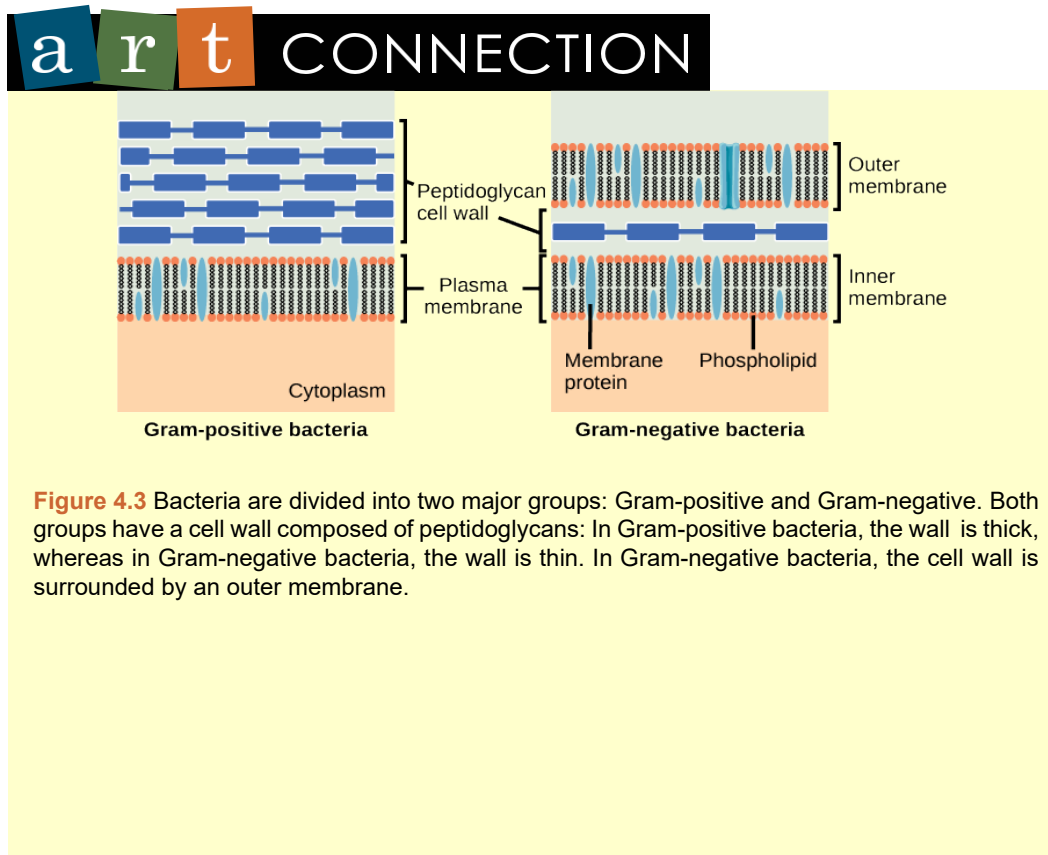
Figure 4.3 Bacteria are divided into two major groups: Gram-positive and Gram-negative. Both groups have a cell wall composed of peptidoglycans: In Gram-positive bacteria, the wall is thick, whereas in Gram-negative bacteria, the wall is thin. In Gram-negative bacteria, the cell wall is surrounded by an outer membrane.

Archaeal cell walls do not contain peptidoglycan. There are four different types of archaeal cell walls. One type is composed of **pseudopeptidoglycan**. The other three types of cell walls contain polysaccharides, glycoproteins, and surface-layer proteins known as S-layers.

Reproduction

Reproduction in prokaryotes is primarily asexual and takes place by binary fission. Recall that the DNA of a prokaryote exists usually as a single, circular chromosome. Prokaryotes do not undergo mitosis. Rather, the chromosome loop is replicated, and the two resulting copies attached to the plasma membrane move apart as the cell grows in a process called binary fission. The prokaryote, now enlarged, is pinched inward at its equator, and the two resulting cells, which are clones, separate. Binary fission does not provide an opportunity for genetic recombination, but prokaryotes can alter their genetic makeup in three ways.

In a process called **transformation**, the cell takes in DNA found in its environment that is shed by other prokaryotes, alive or dead. A **pathogen** is an organism that causes a disease. If a nonpathogenic bacterium takes up DNA from a pathogen and incorporates the new DNA in its own chromosome, it too may become pathogenic. In **transduction**, bacteriophages, the viruses that infect bacteria, move DNA from one bacterium to another. Archaea have a different set of viruses that infect them and translocate genetic material from one individual to another. During **conjugation**, DNA is transferred from one prokaryote to another by means of a pilus that brings the organisms into contact with one another. The DNA transferred is usually a plasmid, but parts of the chromosome can also be moved.

Cycles of binary fission can be very rapid, on the order of minutes for some species. This short generation time coupled with mechanisms of genetic recombination result in the rapid evolution of prokaryotes, allowing them to respond to environmental changes (such as the introduction of an antibiotic) very quickly.

Prokaryotes are metabolically diverse organisms. Prokaryotes fill many niches on Earth, including being involved in nutrient cycles such as the nitrogen and carbon cycles, decomposing dead organisms, and growing and multiplying inside living organisms, including humans. Different prokaryotes can use different sources of energy to assemble macromolecules from smaller molecules. Phototrophs obtain their energy from sunlight. Chemotrophs obtain their energy from chemical compounds.

Bacterial Diseases in Humans

Devastating pathogen-borne diseases and plagues, both viral and bacterial in nature, have affected and continue to affect humans. It is worth noting that all pathogenic prokaryotes are Bacteria; there are no known pathogenic Archaea in humans or any other organism. Pathogenic organisms evolved alongside humans. In the past, the true cause of these diseases was not understood, and some cultures thought that diseases were a spiritual punishment or were mistaken about material causes. Over time, people came to realize that staying apart from afflicted persons, improving sanitation, and properly disposing of the corpses and personal belongings of victims of illness reduced their own chances of getting sick.

Historical Perspective

There are records of infectious diseases as far back as 3,000 B.C. A number of significant **pandemics** caused by Bacteria have been documented over several hundred years. Some of the largest pandemics led to the decline of cities and cultures. Many were zoonoses that appeared with the domestication of animals, as in the case of tuberculosis. A zoonosis is a disease that infects animals but can be transmitted from animals to humans.

Infectious diseases remain among the leading causes of death worldwide. Their impact is less significant in many developed countries, but they are important determiners of mortality in developing countries. The development of antibiotics did much to lessen the mortality rates from bacterial infections, but access to antibiotics is not universal, and the overuse of antibiotics has led to the development of resistant strains of bacteria. Public sanitation efforts that dispose of sewage and provide clean drinking water have done as much or more than medical advances to prevent deaths caused by bacterial infections.

In 430 B.C., the plague of Athens killed one-quarter of the Athenian troops that were fighting in the Great Peloponnesian War. The disease killed a quarter of the population of Athens in over 4 years and weakened Athens' dominance and power. The source of the plague may have been identified recently when researchers from the University of Athens were able to analyze DNA from teeth recovered from a mass grave. The scientists identified nucleotide sequences from a pathogenic bacterium that causes [1]

typhoid fever.

From 541 to 750 A.D., an outbreak called the plague of Justinian (likely a bubonic plague) eliminated, by some estimates, one-quarter to one-half of the human population. The population in Europe declined by 50 percent during this outbreak. Bubonic plague would decimate Europe more than once.

One of the most devastating pandemics was the **Black Death** (1346 to 1361), which is believed to have been another outbreak of bubonic plague caused by the bacterium *Yersinia pestis*. This bacterium is carried by fleas living on black rats. The Black Death reduced the world's population from an estimated 450 million to about 350 to 375 million. Bubonic plague struck London hard again in the mid-1600s. There are still approximately 1,000 to 3,000 cases of plague globally each year. Although contracting bubonic plague before antibiotics meant almost certain death, the bacterium responds to several types of modern antibiotics, and mortality rates from plague are now very low.

CONCEPT in ACTION

Watch a **video (http://openstaxcollege.org/l/black_death2)** on the modern understanding of the Black Death (bubonic plague) in Europe during the fourteenth century.

---

1. Papagrigorakis M. J., Synodinos P. N., Yapijakis C, "Ancient typhoid epidemic reveals possible ancestral strain of *Salmonella enterica* serovar Typhi, *Infect Genet Evol* 7 (2007): 126-7.

Over the centuries, Europeans developed resistance to many infectious diseases. However, European conquerors brought disease-causing bacteria and viruses with them when they reached the Western hemisphere, triggering **epidemics** that completely devastated populations of Native Americans (who had no natural resistance to many European diseases).

Foodborne Diseases

Prokaryotes are everywhere: They readily colonize the surface of any type of material, and food is not an exception. Outbreaks of bacterial infection related to food consumption are common. A **foodborne disease** (colloquially called "food poisoning") is an illness resulting from the consumption of food contaminated with pathogenic bacteria, viruses, or other parasites. Although the United States has one of the safest food supplies in the world, the Center for Disease Control and Prevention (CDC) has reported that "76 million people get sick, more than 300,000 are hospitalized, and 5,000 Americans die each year [3] from foodborne illness."

The characteristics of foodborne illnesses have changed over time. In the past, it was relatively common to hear about sporadic cases of **botulism**, the potentially fatal disease produced by a toxin from the anaerobic bacterium *Clostridium botulinum*. A can, jar, or package created a suitable anaerobic environment where *Clostridium* could grow. Proper sterilization and canning procedures have reduced the incidence of this disease.

Most cases of foodborne illnesses are now linked to produce contaminated by animal waste. For example, there have been serious, produce-related outbreaks associated with raw spinach in the United States and with vegetable sprouts in Germany (Figure 4.4). The raw spinach outbreak in 2006 was produced by the bacterium *E. coli* strain O157:H7. Most *E. coli* strains are not particularly dangerous to humans, (indeed, they live in our large intestine), but O157:H7 is potentially fatal.

1. http://www.cdc.gov/ecoli/2006/september, Centers for Disease Control and Prevention, "Multi-state outbreak of *E. coli* O157:H7 infections from spinach," September-October (2006).

(a)   (b)
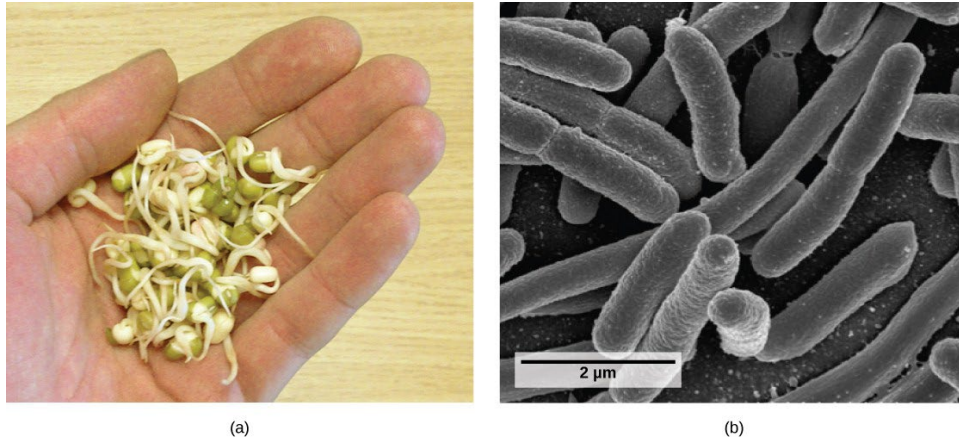
Figure 4.4 (a) Locally grown vegetable sprouts were the cause of a European *E. coli* outbreak that killed 31 people and sickened about 3,000 in 2010. (b) *Escherichia coli* are shown here in a scanning electron micrograph. The strain of *E. coli* that caused a deadly outbreak in Germany is a new one not involved in any previous *E. coli* outbreaks. It has acquired several antibiotic resistance genes and specific genetic sequences involved in aggregation ability and virulence. It has recently been sequenced. (credit b: Rocky Mountain Laboratories, NIAID, NIH; scale-bar data from Matt Russell)

All types of food can potentially be contaminated with harmful bacteria of different species. Recent outbreaks of *Salmonella* reported by the CDC occurred in foods as diverse as peanut butter, alfalfa sprouts, and eggs.

CAREERS IN ACTION

Epidemiologist

Epidemiology is the study of the occurrence, distribution, and determinants of health and disease in a population. It is, therefore, related to public health. An epidemiologist studies the frequency and distribution of diseases within human populations and environments.

Epidemiologists collect data about a particular disease and track its spread to identify the original mode of transmission. They sometimes work in close collaboration with historians to try to understand the way a disease evolved geographically and over time, tracking the natural history of pathogens. They gather information from clinical records, patient interviews, and any other available means. That information is used to develop strategies and design public health policies to reduce the incidence of a disease or to prevent its spread. Epidemiologists also conduct rapid investigations in case of an outbreak to recommend immediate measures to control it.

Epidemiologists typically have a graduate-level education. An epidemiologist often has a bachelor's degree in some field and a master's degree in public health (MPH). Many epidemiologists are also physicians (and have an MD) or they have a PhD in an associated field, such as biology or epidemiology.

Beneficial Prokaryotes

Not all prokaryotes are pathogenic. On the contrary, pathogens represent only a very small percentage of the diversity of the microbial world. In fact, our life and all life on this planet would not be possible without prokaryotes.

Prokaryotes, and Food and Beverages

According to the United Nations Convention on Biological Diversity, biotechnology is "any technological application that uses biological systems, living organisms, or derivatives thereof, to make [4]

or modify products or processes for specific use."   The concept of "specific use" involves some sort of

2. http://www.cbd.int/convention/articles/?a=cbd-02http://www.cbd.int/convention/articles/?a=cbd-02, United Nations Convention on Biological Diversity, "Article 2: Use of Terms."

commercial application. Genetic engineering, artificial selection, antibiotic production, and cell culture are current topics of study in biotechnology. However, humans have used prokaryotes to create products before the term biotechnology was even coined. And some of the goods and services are as simple as cheese, yogurt, sour cream, vinegar, cured sausage, sauerkraut, and fermented seafood that contains both bacteria and archaea (Figure 4.5).

(a)　　　　　　　　　　　　　　　(b)

(c)　　　　　　　　　　　　　　　(d)
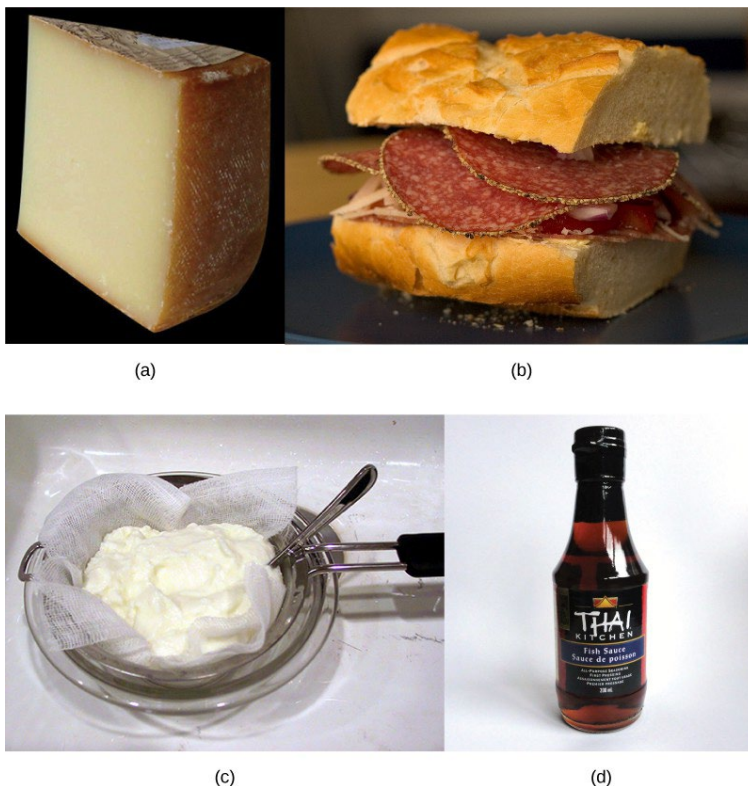
Figure 4.5 Some of the products derived from the use of prokaryotes in early biotechnology include

(a) cheese, (b) salami, (c) yogurt, and (d) fish sauce. (credit b: modification of work by Alisdair McDiarmid; credit c: modification of work by Kris Miller; credit d: modification of work by Jane Whitney)

Cheese production began around 4,000 years ago when humans started to breed animals and process their milk. Evidence suggests that cultured milk products, like yogurt, have existed for at least 4,000 years.

Using Prokaryotes to Clean up Our Planet: Bioremediation

Microbial **bioremediation** is the use of prokaryotes (or microbial metabolism) to remove pollutants. Bioremediation has been used to remove agricultural chemicals (pesticides and fertilizers) that leach from soil into groundwater. Certain toxic metals, such as selenium and arsenic compounds, can also be removed from water by bioremediation. The reduction of SeO_4^{2-} to SeO_3^{2-} and to Se^0 (metallic selenium) is a method used to remove selenium ions from water. Mercury is an example of a toxic metal that can be removed from an environment by bioremediation. Mercury is an active ingredient of some pesticides; it is used in industry and is also a byproduct of certain industries, such as battery production. Mercury is usually present in very low concentrations in natural environments but it is highly toxic because it accumulates in living tissues. Several species of bacteria can carry out the biotransformation of toxic mercury into nontoxic forms. These bacteria, such as *Pseudomonas aeruginosa*, can convert Hg^{2+} to Hg^0, which is nontoxic to humans.

Probably one of the most useful and interesting examples of the use of prokaryotes for bioremediation purposes is the cleanup of oil spills. The importance of prokaryotes to petroleum bioremediation has been demonstrated in several oil spills in recent years, such as the Exxon Valdez spill in Alaska (1989) (Figure 4.6), the Prestige oil spill in Spain (2002), the spill into the Mediterranean from a Lebanon power plant (2006,) and more recently, the BP oil spill in the Gulf of Mexico (2010). To clean up these spills, bioremediation is promoted by adding inorganic nutrients that help bacteria already present in the environment to grow. Hydrocarbon-degrading bacteria feed on the hydrocarbons in the oil droplet, breaking them into inorganic compounds. Some species, such as *Alcanivorax borkumensis*, produce surfactants that solubilize the oil, while other bacteria degrade the oil into carbon dioxide. In the case

110

of oil spills in the ocean, ongoing, natural bioremediation tends to occur, inasmuch as there are oil-consuming bacteria in the ocean prior to the spill. Under ideal conditions, it has been reported that up to 80 percent of the nonvolatile components in oil can be degraded within 1 year of the spill. Other oil fractions containing aromatic and highly branched hydrocarbon chains are more difficult to remove and remain in the environment for longer periods of time. Researchers have genetically engineered other bacteria to consume petroleum products; indeed, the first patent application for a bioremediation application in the U.S. was for a genetically modified oil-eating bacterium.

(a) (b)

Figure 4.6 (a) Cleaning up oil after the Valdez spill in Alaska, the workers hosed oil from beaches and then used a floating boom to corral the oil, which was finally skimmed from the water surface. Some species of bacteria are able to solubilize and degrade the oil. (b) One of the most catastrophic consequences of oil spills is the damage to fauna. (credit a: modification of work by NOAA; credit b: modification of work by GOLUBENKOV, NGO: Saving Taman)

Prokaryotes in and on the Body

Humans are no exception when it comes to forming symbiotic relationships with prokaryotes. We are accustomed to thinking of ourselves as single organisms, but in reality, we are walking ecosystems. There are 10 to 100 times as many bacterial and archaeal cells inhabiting our bodies as we have cells in our bodies. Some of these are in mutually beneficial relationships with us, in which both the human host and the bacterium benefit, while some of the relationships are classified as **commensalism**, a type of relationship in which the bacterium benefits and the human host is neither benefited nor harmed.

Human gut flora lives in the large intestine and consists of hundreds of species of bacteria and archaea, with different individuals containing different species mixes. The term "flora," which is usually associated with plants, is traditionally used in this context because bacteria were once classified as plants. The primary functions of these prokaryotes for humans appear to be metabolism of food molecules that we cannot break down, assistance with the absorption of ions by the colon, synthesis of vitamin K, training of the infant immune system, maintenance of the adult immune system, maintenance of the epithelium of the large intestine, and formation of a protective barrier against pathogens.

The surface of the skin is also coated with prokaryotes. The different surfaces of the skin, such as the underarms, the head, and the hands, provide different habitats for different communities of prokaryotes. Unlike with gut flora, the possible beneficial roles of skin flora have not been well studied. However, the few studies conducted so far have identified bacteria that produce antimicrobial compounds as probably responsible for preventing infections by pathogenic bacteria.

Researchers are actively studying the relationships between various diseases and alterations to the composition of human microbial flora. Some of this work is being carried out by the Human Microbiome Project, funded in the United States by the National Institutes of Healt

Protists

By the end of this section, you will be able to:

- Describe the main characteristics of protists
- Describe important pathogenic species of protists
- Describe the roles of protists as food sources and as decomposers
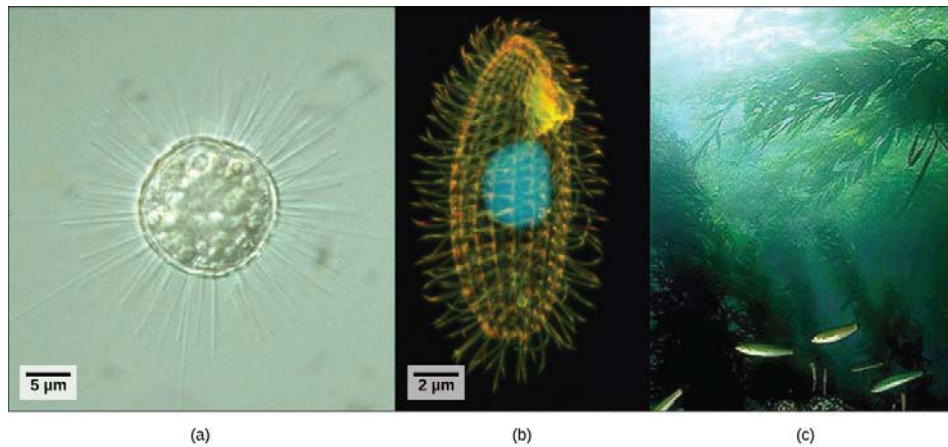
Figure 4.7 Protists range from the microscopic, single-celled (a) *Acanthocystis turfacea* and the

(b) ciliate *Tetrahymena thermophila* to the enormous, multicellular (c) kelps (Chromalveolata) that extend for hundreds of feet in underwater "forests." (credit a: modification of work by Yuiuji Tsukii; credit b: modification of work by Richard Robinson, Public Library of Science; credit c: modification of work by Kip Evans, NOAA; scale-bar data from Matt Russell)

Eukaryotic organisms that did not fit the criteria for the kingdoms Animalia, Fungi, or Plantae historically were called protists and were classified into the kingdom Protista. Protists include the single- celled eukaryotes living in pond water (Figure 4.7), although protist species live in a variety of other aquatic and terrestrial environments, and occupy many different niches. Not all protists are microscopic and single-celled; there exist some very large multicellular species, such as the kelps. During the past two decades, the field of molecular genetics has demonstrated that some protists are more related to animals, plants, or fungi than they are to other protists. For this reason, protist lineages originally classified into the kingdom Protista have been reassigned into new kingdoms or other existing kingdoms. The evolutionary lineages of the protists continue to be examined and debated. In the meantime, the term "protist" still is used informally to describe this tremendously diverse group of eukaryotes. As a collective group, protists display an astounding diversity of morphologies, physiologies, and ecologies.

Characteristics of Protists

There are over 100,000 described living species of protists, and it is unclear how many undescribed species may exist. Since many protists live in symbiotic relationships with other organisms and these relationships are often species specific, there is a huge potential for undescribed protist diversity that matches the diversity of the hosts. As the catchall term for eukaryotic organisms that are not animals, plants, fungi, or any single phylogenetically related group, it is not surprising that few characteristics are common to all protists.

Nearly all protists exist in some type of aquatic environment, including freshwater and marine environments, damp soil, and even snow. Several protist species are **parasites** that infect animals or plants. A parasite is an organism that lives on or in another organism and feeds on it, often without killing it. A few protist species live on dead organisms or their wastes, and contribute to their decay.

Protist Structure

The cells of protists are among the most elaborate of all cells. Most protists are microscopic and unicellular, but some true multicellular forms exist. A few protists live as colonies that behave in some ways as a group of free-living cells and in other ways as a multicellular organism. Still other protists are composed of enormous, multinucleate, single cells that look like amorphous blobs of slime or, in other cases, like ferns. In fact, many protist cells are multinucleated; in some species, the nuclei are different sizes and have distinct roles in protist cell function.

Single protist cells range in size from less than a micrometer to the 3-meter lengths of the multinucleate cells of the seaweed *Caulerpa*. Protist cells may be enveloped by animal-like cell membranes or plant-like cell walls. Others are encased in glassy silica-based shells or wound with **pellicles** of interlocking protein strips. The pellicle functions like a flexible coat of armor, preventing the protist from being torn or pierced without compromising its range of motion.

The majority of protists are motile, but different types of protists have evolved varied modes of movement. Some protists have one or more flagella, which they rotate or whip. Others are covered in rows or tufts of tiny cilia that they beat in coordination to swim. Still others send out lobe-like pseudopodia from anywhere on the cell, anchor the pseudopodium to a substrate, and pull the rest of the

112

cell toward the anchor point. Some protists can move toward light by coupling their locomotion strategy with a light-sensing organ.

How Protists Obtain Energy

Protists exhibit many forms of nutrition and may be aerobic or anaerobic. Photosynthetic protists (photoautotrophs) are characterized by the presence of chloroplasts. Other protists are heterotrophs and consume organic materials (such as other organisms) to obtain nutrition. Amoebas and some other heterotrophic protist species ingest particles by a process called phagocytosis, in which the cell membrane engulfs a food particle and brings it inward, pinching off an intracellular membranous sac, or vesicle, called a food vacuole (Figure 4.8). This vesicle then fuses with a lysosome, and the food particle is broken down into small molecules that can diffuse into the cytoplasm and be used in cellular metabolism. Undigested remains ultimately are expelled from the cell through exocytosis.
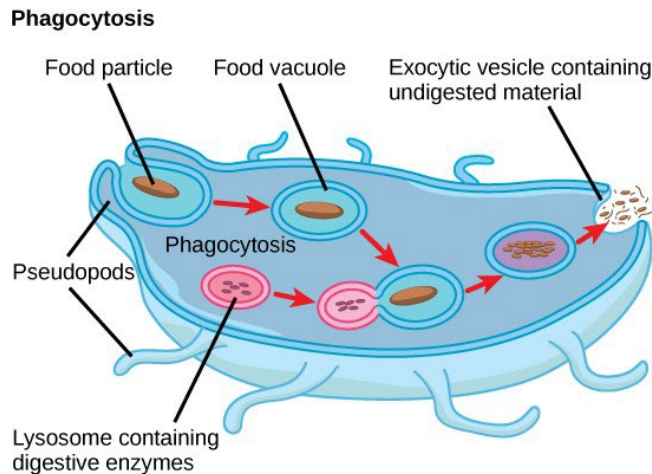
Figure 4.8 The stages of phagocytosis include the engulfment of a food particle, the digestion of the particle using hydrolytic enzymes contained within a lysosome, and the expulsion of undigested material from the cell.

Some heterotrophs absorb nutrients from dead organisms or their organic wastes, and others are able to use photosynthesis or feed on organic matter, depending on conditions.

Reproduction

Protists reproduce by a variety of mechanisms. Most are capable some form of asexual reproduction, such as binary fission to produce two daughter cells, or multiple fission to divide simultaneously into many daughter cells. Others produce tiny buds that go on to divide and grow to the size of the parental protist. Sexual reproduction, involving meiosis and fertilization, is common among protists, and many protist species can switch from asexual to sexual reproduction when necessary. Sexual reproduction is often associated with periods when nutrients are depleted or environmental changes occur. Sexual reproduction may allow the protist to recombine genes and produce new variations of progeny that may be better suited to surviving in the new environment. However, sexual reproduction is also often associated with cysts that are a protective, resting stage. Depending on their habitat, the cysts may be particularly resistant to temperature extremes, desiccation, or low pH. This strategy also allows certain protists to "wait out" stressors until their environment becomes more favorable for survival or until they are carried (such as by wind, water, or transport on a larger organism) to a different environment because cysts exhibit virtually no cellular metabolism.

Protist Diversity

With the advent of DNA sequencing, the relationships among protist groups and between protist groups and other eukaryotes are beginning to become clearer. Many relationships that were based on morphological similarities are being replaced by new relationships based on genetic similarities. Protists that exhibit similar morphological features may have evolved analogous structures because of similar selective pressures—rather than because of recent common ancestry. This phenomenon is called convergent evolution. It is one reason why protist classification is so challenging. The emerging classification scheme groups the entire domain Eukaryota into six "supergroups" that contain all of the protists as well as animals, plants, and fungi (Figure 4.9); these include the **Excavata**, **Chromalveolata**, **Rhizaria**, **Archaeplastida**, **Amoebozoa**, and **Opisthokonta**. The supergroups are

113

believed to be monophyletic; all organisms within each supergroup are believed to have evolved from a single common ancestor, and thus all members are most closely related to each other than to organisms outside that group. There is still evidence lacking for the monophyly of some groups.
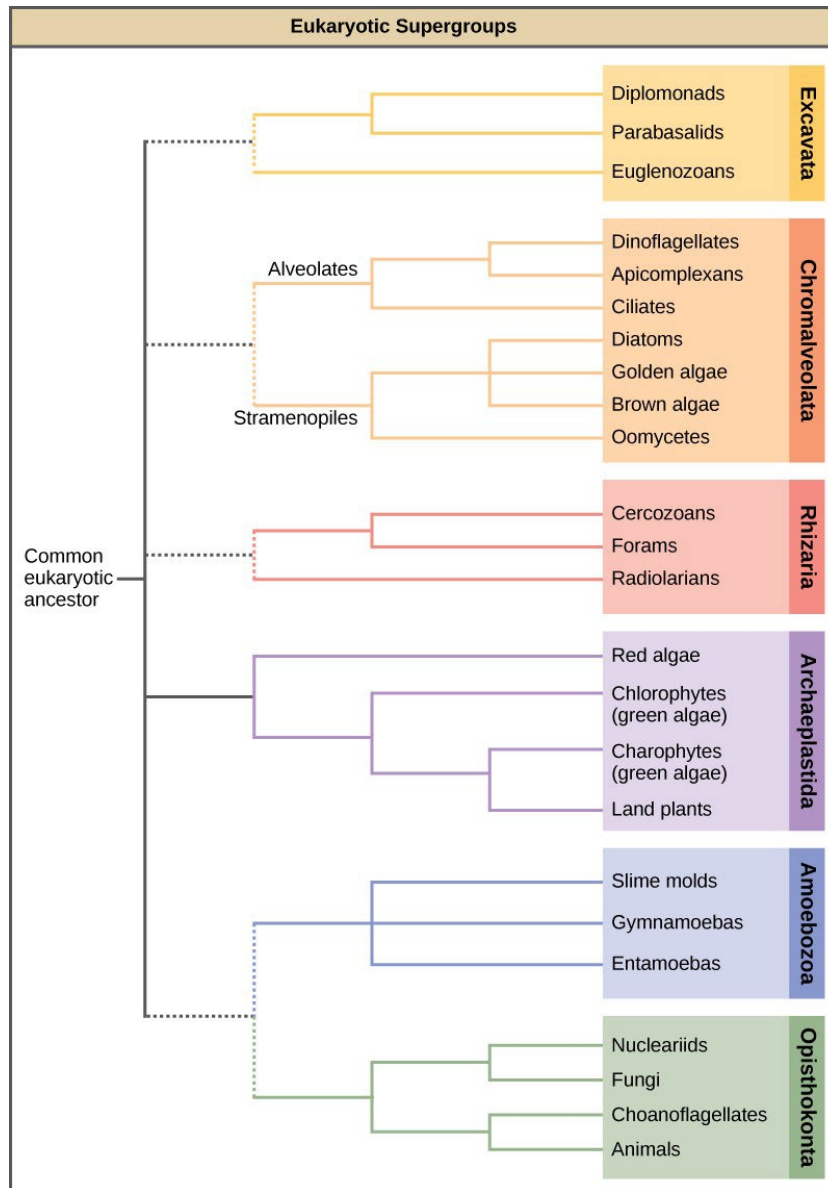
**Figure 4.9** Protists appear in all six eukaryotic supergroups.

Human Pathogens

Many protists are pathogenic parasites that must infect other organisms to survive and propagate. Protist parasites include the causative agents of malaria, African sleeping sickness, and waterborne gastroenteritis in humans. Other protist pathogens prey on plants, effecting massive destruction of food crops.

Plasmodium Species

Members of the genus *Plasmodium* must infect a mosquito and a vertebrate to complete their life cycle. In vertebrates, the parasite develops in liver cells and goes on to infect red blood cells, bursting from and destroying the blood cells with each asexual replication cycle (Figure 4.10). Of the four *Plasmodium* species known to infect humans, *P. falciparum* accounts for 50 percent of all malaria cases and is the primary cause of disease-related fatalities in tropical regions of the world. In 2010, it was estimated that malaria caused between 0.5 and 1 million deaths, mostly in African children. During the course of malaria, *P. falciparum* can infect and destroy more than one-half of a human's circulating blood cells, leading to severe anemia. In response to waste products released as the parasites burst from infected blood cells, the host immune system mounts a massive inflammatory response with delirium-inducing

114

fever episodes, as parasites destroy red blood cells, spilling parasite waste into the blood stream. *P. falciparum* is transmitted to humans by the African malaria mosquito, *Anopheles gambiae*. Techniques to kill, sterilize, or avoid exposure to this highly aggressive mosquito species are crucial to malaria control.
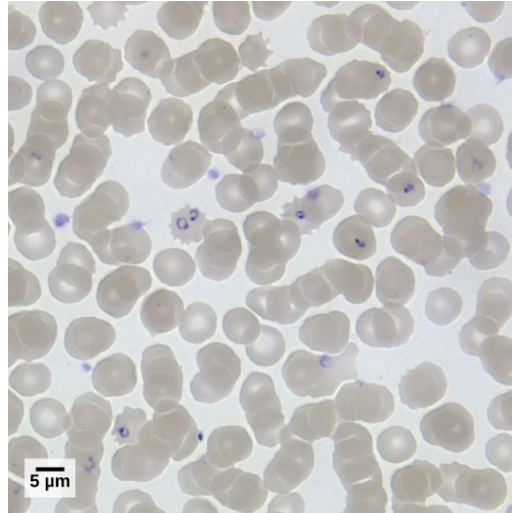
Figure 4.10 This light micrograph shows a 100× magnification of red blood cells infected with *P. falciparum* (seen as purple). (credit: modification of work by Michael Zahniser; scale-bar data from Matt Russell)
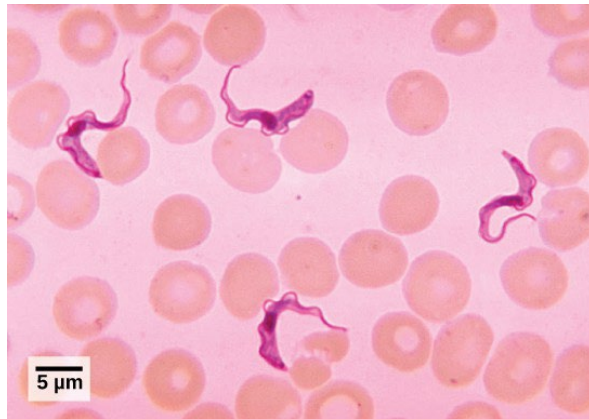

This **movie (http://openstaxcollege.org/l/malaria2)** depicts the pathogenesis of *Plasmodium falciparum*, the causative agent of malaria.

Trypanosomes

T. brucei, the parasite that is responsible for African sleeping sickness, confounds the human immune system by changing its thick layer of surface glycoproteins with each infectious cycle (Figure 4.11). The glycoproteins are identified by the immune system as foreign matter, and a specific antibody defense is mounted against the parasite. However, *T. brucei* has thousands of possible antigens, and with each subsequent generation, the protist switches to a glycoprotein coating with a different molecular structure. In this way, *T. brucei* is capable of replicating continuously without the immune system ever succeeding in clearing the parasite. Without treatment, African sleeping sickness leads invariably to death because of damage it does to the nervous system. During epidemic periods, mortality from the disease can be high. Greater surveillance and control measures have led to a reduction in reported cases; some of the lowest numbers reported in 50 years (fewer than 10,000 cases in all of sub-Saharan Africa) have happened since 2009.

In Latin America, another species in the genus, *T. cruzi*, is responsible for Chagas disease. *T. cruzi* infections are mainly caused by a blood-sucking bug. The parasite inhabits heart and digestive system tissues in the chronic phase of infection, leading to malnutrition and heart failure caused by abnormal heart rhythms. An estimated 10 million people are infected with Chagas disease, which caused 10,000 deaths in 2008.

Figure 4.11 Trypanosomes are shown in this light micrograph among red blood cells. (credit: modification of work by Myron G. Schultz, CDC; scale-bar data from Matt Russell)

This **movie (http://openstaxcollege.org/l/African_sleep2)** discusses the pathogenesis of

Trypanosoma brucei, the causative agent of African sleeping sickness.

Plant Parasites

Protist parasites of terrestrial plants include agents that destroy food crops. The oomycete *Plasmopara viticola* parasitizes grape plants, causing a disease called downy mildew (Figure 4.11**a**). Grape plants infected with *P. viticola* appear stunted and have discolored withered leaves. The spread of downy mildew caused the near collapse of the French wine industry in the nineteenth century.

Figure 4.12 (a) The downy and powdery mildews on this grape leaf are caused by an infection of

P. viticola. (b) This potato exhibits the results of an infection with *P. infestans*, the potato late blight. (credit a: modification of work by David B. Langston, University of Georgia, USDA ARS; credit b: USDA ARS)

Phytophthora infestans is an oomycete responsible for potato late blight, which causes potato stalks and stems to decay into black slime (Figure 4.12**b**). Widespread potato blight caused by *P. infestans* precipitated the well-known Irish potato famine in the nineteenth century that claimed the lives of approximately 1 million people and led to the emigration from Ireland of at least 1 million more. Late blight continues to plague potato crops in certain parts of the United States and Russia, wiping out as much as 70 percent of crops when no pesticides are applied.

116

Beneficial Protists

Protists play critically important ecological roles as producers particularly in the world's oceans. They are equally important on the other end of food webs as decomposers.

Protists as Food Sources

Protists are essential sources of nutrition for many other organisms. In some cases, as in plankton, protists are consumed directly. Alternatively, photosynthetic protists serve as producers of nutrition for other organisms by carbon fixation. For instance, photosynthetic dinoflagellates called zooxanthellae pass on most of their energy to the coral polyps that house them (Figure 4.13). In this mutually beneficial relationship, the polyps provide a protective environment and nutrients for the zooxanthellae. The polyps secrete the calcium carbonate that builds coral reefs. Without dinoflagellate symbionts, corals lose algal pigments in a process called coral bleaching, and they eventually die. This explains why reef- building corals do not reside in waters deeper than 20 meters: Not enough light reaches those depths for dinoflagellates to photosynthesize.

Figure 4.13 Coral polyps obtain nutrition through a symbiotic relationship with dinoflagellates.

Protists themselves and their products of photosynthesis are essential—directly or indirectly—to the survival of organisms ranging from bacteria to mammals. As primary producers, protists feed a large proportion of the world's aquatic species. (On land, terrestrial plants serve as primary producers.) In fact, approximately one-quarter of the world's photosynthesis is conducted by protists, particularly dinoflagellates, diatoms, and multicellular algae.

Protists do not create food sources only for sea-dwelling organisms. For instance, certain anaerobic species exist in the digestive tracts of termites and wood-eating cockroaches, where they contribute to digesting cellulose ingested by these insects as they bore through wood. The actual enzyme used to digest the cellulose is actually produced by bacteria living within the protist cells. The termite provides the food source to the protist and its bacteria, and the protist and bacteria provide nutrients to the termite by breaking down the cellulose.

Agents of Decomposition

Many fungus-like protists are **saprobes**, organisms that feed on dead organisms or the waste matter produced by organisms (saprophyte is an equivalent term), and are specialized to absorb nutrients from nonliving organic matter. For instance, many types of oomycetes grow on dead animals or algae. Saprobic protists have the essential function of returning inorganic nutrients to the soil and water. This process allows for new plant growth, which in turn generates sustenance for other organisms along the food chain. Indeed, without saprobic species, such as protists, fungi, and bacteria, life would cease to exist as all organic carbon became "tied up" in dead organisms

117

Fungi

By the end of this section, you will be able to:

- List the characteristics of fungi
- Describe fungal parasites and pathogens of plants and infections in humans
- Describe the importance of fungi to the environment
- Summarize the beneficial role of fungi in food and beverage preparation and in the chemical and pharmaceutical industry

(a)   (b)   (c)
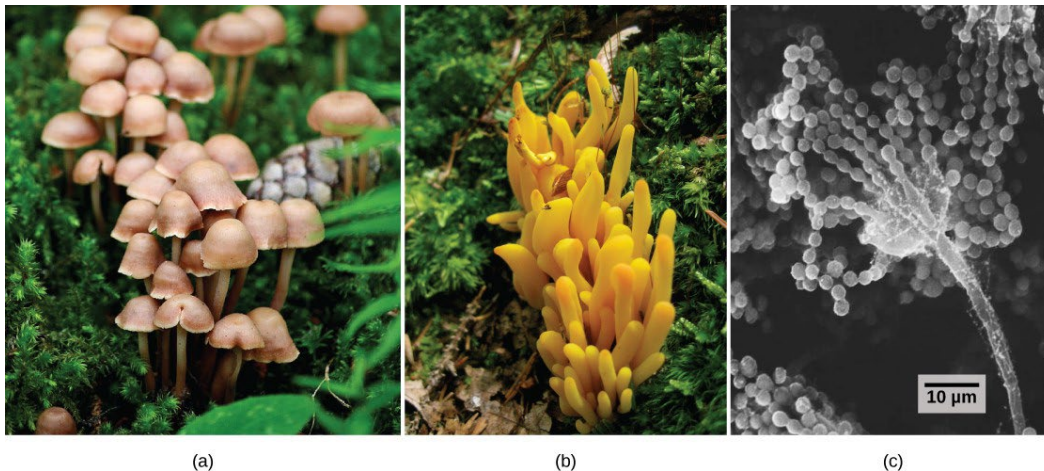
Figure 4.14 The (a) familiar mushroom is only one type of fungus. The brightly colored fruiting bodies of this (b) coral fungus are displayed. This (c) electron micrograph shows the spore-bearing structures of *Aspergillus*, a type of toxic fungi found mostly in soil and plants. (credit a: modification of work by Chris Wee; credit b: modification of work by Cory Zanker; credit c: modification of work by Janice Haney Carr, Robert Simmons, CDC; scale-bar data from Matt Russell)

The word *fungus* comes from the Latin word for mushroom. Indeed, the familiar mushrooms are fungi, but there are many other types of fungi as well (Figure 4.14). The kingdom Fungi includes an enormous variety of living organisms collectively referred to as Eumycota, or true fungi. While scientists have identified about 100,000 species of fungi, this is only a fraction of the over 1 million species likely present on Earth. Edible mushrooms, yeasts, black mold, and *Penicillium notatum* (the producer of the antibiotic penicillin) are all members of the kingdom Fungi, which belongs to the domain Eukarya. As eukaryotes, a typical fungal cell contains a true nucleus and many membrane-bound organelles.

Fungi were once considered plant-like organisms; however, DNA comparisons have shown that fungi are more closely related to animals than plants. Fungi are not capable of photosynthesis: They use complex organic compounds as sources of energy and carbon. Some fungal organisms multiply only asexually, whereas others undergo both asexual reproduction and sexual reproduction. Most fungi produce a large number of spores that are disseminated by the wind. Like bacteria, fungi play an essential role in ecosystems, because they are decomposers and participate in the cycling of nutrients by breaking down organic materials into simple molecules.

Fungi often interact with other organisms, forming mutually beneficial or mutualistic associations. Fungi also cause serious infections in plants and animals. For example, Dutch elm disease is a particularly devastating fungal infection that destroys many native species of elm (*Ulmus* spp.). The fungus infects the vascular system of the tree. It was accidentally introduced to North America in the 1900s and decimated elm trees across the continent. Dutch elm disease is caused by the fungus *Ophiostoma ulmi*. The elm bark beetle acts as a vector and transmits the disease from tree to tree. Many European and Asiatic elms are less susceptible than American elms.

In humans, fungal infections are generally considered challenging to treat because, unlike bacteria, they do not respond to traditional antibiotic therapy since they are also eukaryotes. These infections may prove deadly for individuals with a compromised immune system.

118

Fungi have many commercial applications. The food industry uses yeasts in baking, brewing, and wine making. Many industrial compounds are byproducts of fungal fermentation. Fungi are the source of many commercial enzymes and antibiotics.

Cell Structure and Function

Fungi are eukaryotes and as such have a complex cellular organization. As eukaryotes, fungal cells contain a membrane-bound nucleus. A few types of fungi have structures comparable to the plasmids (loops of DNA) seen in bacteria. Fungal cells also contain mitochondria and a complex system of internal membranes, including the endoplasmic reticulum and Golgi apparatus.

Fungal cells do not have chloroplasts. Although the photosynthetic pigment chlorophyll is absent, many fungi display bright colors, ranging from red to green to black. The poisonous *Amanita muscaria* (fly agaric) is recognizable by its bright red cap with white patches (Figure 4.15). Pigments in fungi are associated with the cell wall and play a protective role against ultraviolet radiation. Some pigments are toxic.


Figure 4.15 The poisonous *Amanita muscaria* is native to the temperate and boreal regions of North America. (credit: Christine Majul)

Like plant cells, fungal cells are surrounded by a thick cell wall; however, the rigid layers contain the complex polysaccharides chitin and glucan. Cellulose, the main component of plant cell walls, is found rarely in fungi. Chitin, also found in the exoskeleton of insects, gives structural strength to the cell walls of fungi. The cell wall protects the cell from desiccation and predators. Fungi have plasma membranes similar to other eukaryotes, except that the structure is stabilized by ergosterol, a steroid molecule that functions like the cholesterol found in animal cell membranes. Most members of the kingdom Fungi are nonmotile. Flagella are produced only by the gametes in the primitive division Chytridiomycota.

Growth and Reproduction

The vegetative body of a fungus is called a **thallus** and can be unicellular or multicellular. Some fungi are dimorphic because they can go from being unicellular to multicellular depending on environmental conditions. Unicellular fungi are generally referred to as **yeasts**.*Saccharomyces cerevisiae* (baker's yeast) and *Candida* species (the agents of thrush, a common fungal infection) are examples of unicellular fungi.

Most fungi are multicellular organisms. They display two distinct morphological stages: vegetative and reproductive. The vegetative stage is characterized by a tangle of slender thread-like structures called hyphae (singular, **hypha**), whereas the reproductive stage can be more conspicuous. A mass of hyphae is called a **mycelium** (Figure 4.16). It can grow on a surface, in soil or decaying material, in a liquid, or even in or on living tissue. Although individual hypha must be observed under a microscope, the mycelium of a fungus can be very large with some species truly being "the fungus humongous." The giant *Armillaria ostoyae* (honey mushroom) is considered the largest organism on Earth, spreading across over 2,000 acres of underground soil in eastern Oregon; it is estimated to be at least 2,400 years old.

119
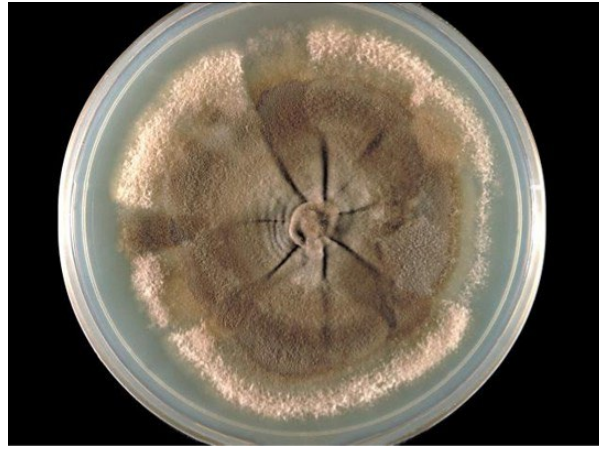
Figure 4.16 The mycelium of the fungus *Neotestudina rosati* can be pathogenic to humans. The fungus enters through a cut or scrape and develops into a mycetoma, a chronic subcutaneous infection. (credit: CDC)

Most fungal hyphae are divided into separate cells by end walls called septa (singular, **septum**). In most divisions (like plants, fungal phyla are called *divisions* by tradition) of fungi, tiny holes in the septa allow for the rapid flow of nutrients and small molecules from cell to cell along the hyphae. They are described as perforated septa. The hyphae in bread **molds** (which belong to the division Zygomycota) are not separated by septa. They are formed of large cells containing many nuclei, an arrangement described as coenocytic hyphae.

Fungi thrive in environments that are moist and slightly acidic, and can grow with or without light. They vary in their oxygen requirements. Most fungi are obligate aerobes, requiring oxygen to survive. Other species, such as the Chytridiomycota that reside in the rumen of cattle, are obligate anaerobes, meaning that they cannot grow and reproduce in an environment with oxygen. Yeasts are intermediate: They grow best in the presence of oxygen but can use fermentation in the absence of oxygen. The alcohol produced from yeast fermentation is used in wine and beer production, and the carbon dioxide they produce carbonates beer and sparkling wine, and makes bread rise.

Fungi can reproduce sexually or asexually. Fungi that propagate sexually and asexually are called perfect Fungi. Fungi that reproduce only asexually (by mitosis) are called imperfect Fungi.

In both sexual and asexual reproduction, fungi produce spores that disperse from the parent organism by either floating in the wind or hitching a ride on an animal. Fungal spores are smaller and lighter than plant seeds, but they are not usually released as high in the air. The giant puffball mushroom bursts open and releases trillions of spores: The huge number of spores released increases the likelihood of spores landing in an environment that will support growth (Figure 4.17).

(a)   (b)
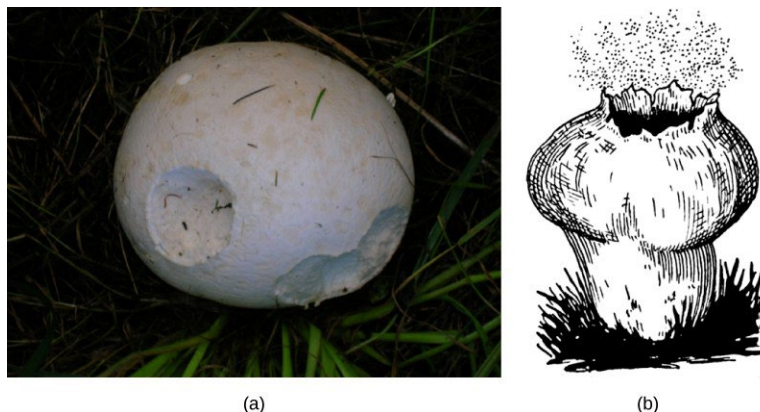
Figure 4.17 The (a) giant puffball mushroom releases (b) a cloud of spores when it reaches maturity. (credit a: modification of work by Roger Griffith; credit b: modification of work by Pearson Scott Foresman, donated to the Wikimedia Foundation)

How Fungi Obtain Nutrition

Like animals, fungi are heterotrophs: They use complex organic compounds as a source of carbon rather than fixing carbon dioxide from the atmosphere, as some bacteria and most plants do. In addition, fungi do not fix nitrogen from the atmosphere. Like animals, they must obtain it from their diet.

120

However, unlike most animals that ingest food and then digest it internally in specialized organs, fungi perform these steps in the reverse order. Digestion precedes ingestion. First, exoenzymes, enzymes that catalyze reactions on compounds outside of the cell, are transported out of the hyphae where they break down nutrients in the environment. Then, the smaller molecules produced by the external digestion are absorbed through the large surface areas of the mycelium. As with animal cells, the fungal storage polysaccharide is glycogen rather than starch, as found in plants.

Fungi are mostly saprobes, organisms that derive nutrients from decaying organic matter. They obtain their nutrients from dead or decomposing organic matter, mainly plant material. Fungal exoenzymes are able to break down insoluble polysaccharides, such as the cellulose and lignin of dead wood, into readily absorbable glucose molecules. Decomposers are important components of ecosystems, because they return nutrients locked in dead bodies to a form that is usable for other organisms. This role is discussed in more detail later. Because of their varied metabolic pathways, fungi fulfill an important ecological role and are being investigated as potential tools in bioremediation. For example, some species of fungi can be used to break down diesel oil and polycyclic aromatic hydrocarbons. Other species take up heavy metals such as cadmium and lead.

Fungal Diversity

The kingdom Fungi contains four major divisions that were established according to their mode of sexual reproduction. Polyphyletic, unrelated fungi that reproduce without a sexual cycle, are placed for convenience in a fifth division, and a sixth major fungal group that does not fit well with any of the previous five has recently been described. Not all mycologists agree with this scheme. Rapid advances in molecular biology and the sequencing of 18S rRNA (a component of ribosomes) continue to reveal new and different relationships between the various categories of fungi.

The traditional divisions of Fungi are the **Chytridiomycota** (chytrids), the **Zygomycota** (conjugated fungi), the **Ascomycota** (sac fungi), the **Basidiomycota** (club fungi), and the **Deuteromycota** (imperfect fungi). The **Glomeromycota** belong to a newly described group (**Figure 4.18**).

(a)   (b)   (c)

(d)   (e)
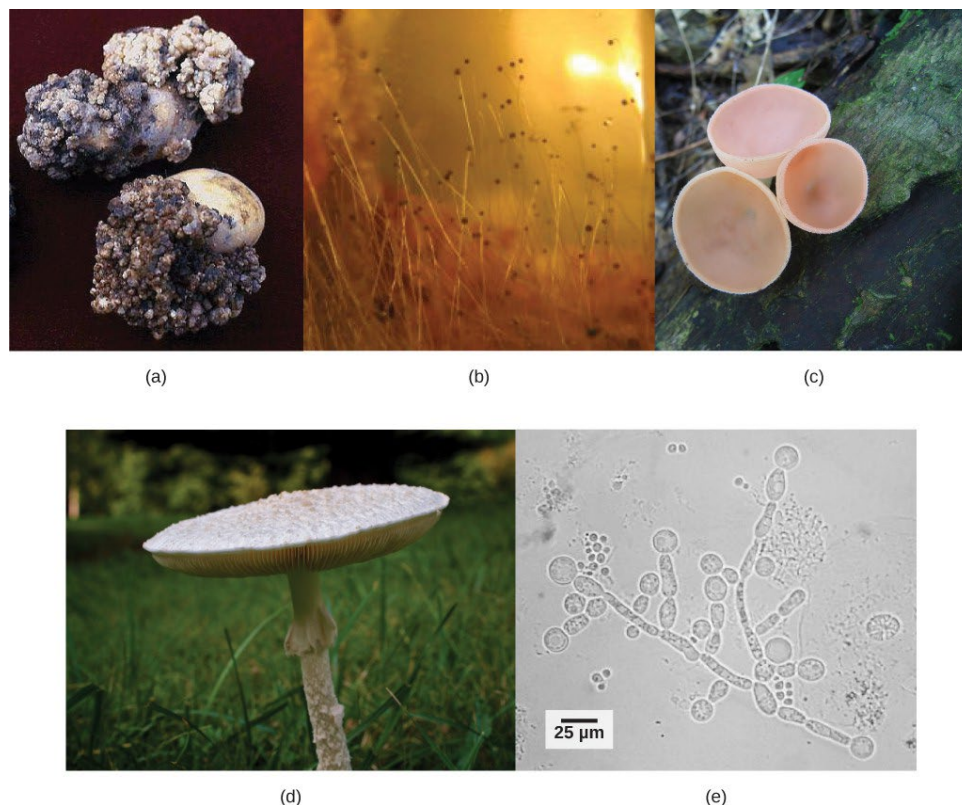
Figure 4.18 Divisions of fungi include (a) chytrids, (b) conjugated fungi, (c) sac fungi, (d) club fungi, and (e) imperfect fungi. (credit a: modification of work by USDA APHIS PPQ; credit c: modification of work by "icelight"/Flickr; credit d: modification of work by Cory Zanker; credit e: modification of work by CDC/ Brinkman; scale-bar data from Matt Russell)

Pathogenic Fungi

Many fungi have negative impacts on other species, including humans and the organisms they depend on for food. Fungi may be parasites, pathogens, and, in a very few cases, predators.

Plant Parasites and Pathogens

The production of enough good-quality crops is essential to our existence. Plant diseases have ruined crops, bringing widespread famine. Most plant pathogens are fungi that cause tissue decay and eventual death of the host (Figure 4.19). In addition to destroying plant tissue directly, some plant pathogens spoil crops by producing potent toxins. Fungi are also responsible for food spoilage and the rotting of stored crops. For example, the fungus *Claviceps purpurea* causes ergot, a disease of cereal crops (especially of rye). Although the fungus reduces the yield of cereals, the effects of the ergot's alkaloid toxins on humans and animals are of much greater significance: In animals, the disease is referred to as ergotism. The most common signs and symptoms are convulsions, hallucination, gangrene, and loss of milk in cattle. The active ingredient of ergot is lysergic acid, which is a precursor of the drug LSD. Smuts, rusts, and powdery or downy mildew are other examples of common fungal pathogens that affect crops.
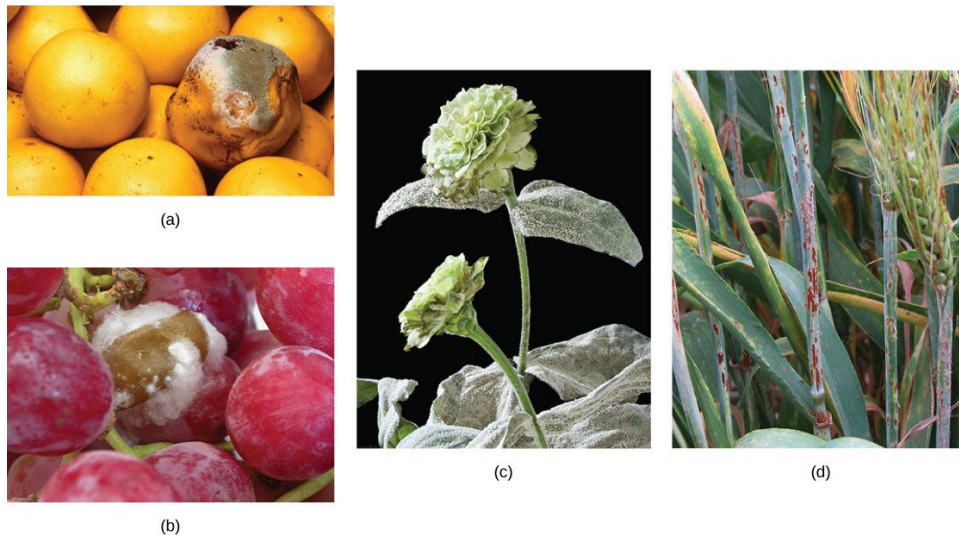
Figure 4.19 Some fungal pathogens include (a) green mold on grapefruit, (b) fungus on grapes, (c) powdery mildew on a zinnia, and (d) stem rust on a sheaf of barley. Notice the brownish color of the fungus in (b) *Botrytis cinerea*, also referred to as the "noble rot," which grows on grapes and other fruit. Controlled infection of grapes by *Botrytis* is used to produce strong and much-prized dessert wines. (credit a: modification of work by Scott Bauer, USDA ARS; credit b: modification of work by Stephen Ausmus, USDA ARS; credit c: modification of work by David Marshall, USDA ARS; credit d: modification of work by Joseph Smilanick, USDA ARS)

Aflatoxins are toxic and carcinogenic compounds released by fungi of the genus *Aspergillus*. Periodically, harvests of nuts and grains are tainted by aflatoxins, leading to massive recall of produce, sometimes ruining producers, and causing food shortages in developing countries.

Animal and Human Parasites and Pathogens

Fungi can affect animals, including humans, in several ways. Fungi attack animals directly by colonizing and destroying tissues. Humans and other animals can be poisoned by eating toxic mushrooms or foods contaminated by fungi. In addition, individuals who display hypersensitivity to molds and spores develop strong and dangerous allergic reactions. Fungal infections are generally very difficult to treat because, unlike bacteria, fungi are eukaryotes. Antibiotics only target prokaryotic cells, whereas compounds that kill fungi also adversely affect the eukaryotic animal host.

Many fungal infections (**mycoses**) are superficial and termed cutaneous (meaning "skin") mycoses. They are usually visible on the skin of the animal. Fungi that cause the superficial mycoses of the epidermis, hair, and nails rarely spread to the underlying tissue (Figure 4.20). These fungi are often misnamed "dermatophytes" from the Greek *dermis* skin and *phyte* plant, but they are not plants. Dermatophytes are also called "ringworms" because of the red ring that they cause on skin (although the ring is caused by fungi, not a worm). These fungi secrete extracellular enzymes that break down keratin (a protein found in hair, skin, and nails), causing a number of conditions such as athlete's foot, jock

122

itch, and other cutaneous fungal infections. These conditions are usually treated with over-the-counter topical creams and powders, and are easily cleared. More persistent, superficial mycoses may require prescription oral medications.
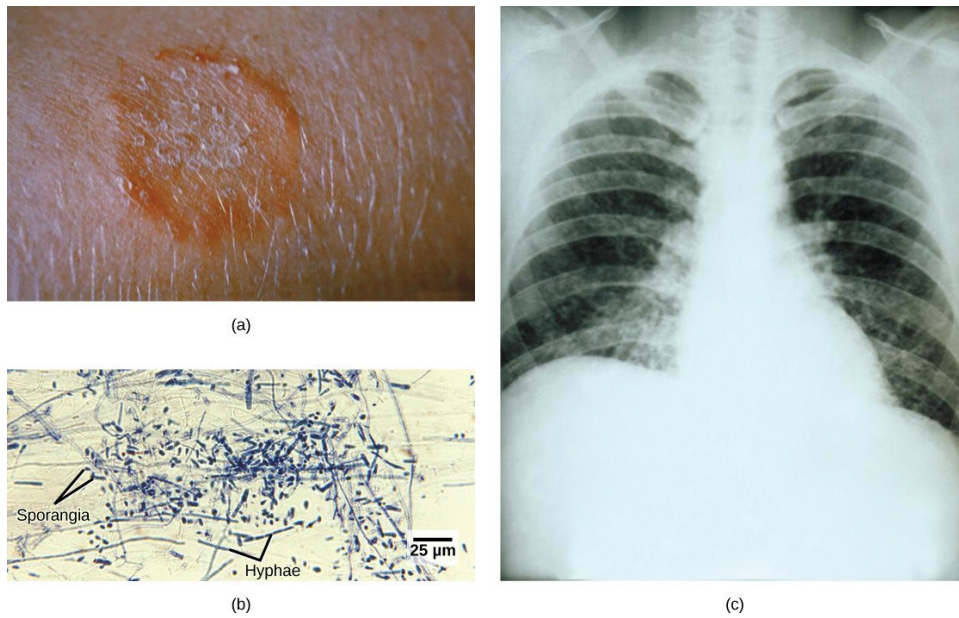
Figure 4.20 (a) Ringworm presents as a red ring on the skin. (b) *Trichophyton violaceum* is a fungus that causes superficial mycoses on the scalp. (c) *Histoplasma capsulatum*, seen in this X-ray as speckling of light areas in the lung, is a species of Ascomycota that infects airways and causes symptoms similar to the flu. (credit a, b: modification of work by Dr. Lucille K. Georg, CDC; credit c: modification of work by M Renz, CDC; scale-bar data from Matt Russell)

Systemic mycoses spread to internal organs, most commonly entering the body through the respiratory system. For example, coccidioidomycosis (valley fever) is commonly found in the southwestern United States, where the fungus resides in the dust. Once inhaled, the spores develop in the lungs and cause signs and symptoms similar to those of tuberculosis. Histoplasmosis (**Figure 13.26c**) is caused by the dimorphic fungus *Histoplasma capsulatum*; it causes pulmonary infections and, in rare cases, swelling of the membranes of the brain and spinal cord. Treatment of many fungal diseases requires the use of antifungal medications that have serious side effects.

Opportunistic mycoses are fungal infections that are either common in all environments or part of the normal biota. They affect mainly individuals who have a compromised immune system. Patients in the late stages of AIDS suffer from opportunistic mycoses, such as *Pneumocystis*, which can be life threatening. The yeast *Candida* spp., which is a common member of the natural biota, can grow unchecked if the pH, the immune defenses, or the normal population of bacteria is altered, causing yeast infections of the vagina or mouth (oral thrush).

Fungi may even take on a predatory lifestyle. In soil environments that are poor in nitrogen, some fungi resort to predation of nematodes (small roundworms). Species of *Arthrobotrys* fungi have a number of mechanisms to trap nematodes. For example, they have constricting rings within their network of hyphae. The rings swell when the nematode touches it and closes around the body of the nematode, thus trapping it. The fungus extends specialized hyphae that can penetrate the body of the worm and slowly digest the hapless prey.

Beneficial Fungi

Fungi play a crucial role in the balance of ecosystems. They colonize most habitats on Earth, preferring dark, moist conditions. They can thrive in seemingly hostile environments, such as the tundra, thanks to a most successful symbiosis with photosynthetic organisms, like lichens. Fungi are not obvious in the way that large animals or tall trees are. Yet, like bacteria, they are major decomposers of nature. With their versatile metabolism, fungi break down organic matter that is insoluble and would not be recycled otherwise.

Importance to Ecosystems

Food webs would be incomplete without organisms that decompose organic matter and fungi are key participants in this process. Decomposition allows for cycling of nutrients such as carbon, nitrogen, and phosphorus back into the environment so they are available to living things, rather than being trapped

123

in dead organisms. Fungi are particularly important because they have evolved enzymes to break down cellulose and lignin, components of plant cell walls that few other organisms are able to digest, releasing their carbon content.

Fungi are also involved in ecologically important coevolved symbioses, both mutually beneficial and pathogenic with organisms from the other kingdoms. **Mycorrhiza**, a term combining the Greek roots *myco* meaning fungus and *rhizo* meaning root, refers to the association between vascular plant roots and their symbiotic fungi. Somewhere between 80–90 percent of all plant species have mycorrhizal partners. In a mycorrhizal association, the fungal mycelia use their extensive network of hyphae and large surface area in contact with the soil to channel water and minerals from the soil into the plant. In exchange, the plant supplies the products of photosynthesis to fuel the metabolism of the fungus. Ectomycorrhizae ("outside" mycorrhiza) depend on fungi enveloping the roots in a sheath (called a mantle) and a net of hyphae that extends into the roots between cells. In a second type, the Glomeromycota fungi form arbuscular mycorrhiza. In these mycorrhiza, the fungi form arbuscles, a specialized highly branched hypha, which penetrate root cells and are the sites of the metabolic exchanges between the fungus and the host plant. Orchids rely on a third type of mycorrhiza. Orchids form small seeds without much storage to sustain germination and growth. Their seeds will not germinate without a mycorrhizal partner (usually Basidiomycota). After nutrients in the seed are depleted, fungal symbionts support the growth of the orchid by providing necessary carbohydrates and minerals. Some orchids continue to be mycorrhizal throughout their lifecycle.

Lichens blanket many rocks and tree bark, displaying a range of colors and textures. Lichens are important pioneer organisms that colonize rock surfaces in otherwise lifeless environments such as are created by glacial recession. The lichen is able to leach nutrients from the rocks and break them down in the first step to creating soil. Lichens are also present in mature habitats on rock surfaces or the trunks of trees. They are an important food source for caribou. Lichens are not a single organism, but rather a fungus (usually an Ascomycota or Basidiomycota species) living in close contact with a photosynthetic organism (an alga or cyanobacterium). The body of a lichen, referred to as a thallus, is formed of hyphae wrapped around the green partner. The photosynthetic organism provides carbon and energy in the form of carbohydrates and receives protection from the elements by the thallus of the fungal partner. Some cyanobacteria fix nitrogen from the atmosphere, contributing nitrogenous compounds to the association. In return, the fungus supplies minerals and protection from dryness and excessive light by encasing the algae in its mycelium. The fungus also attaches the symbiotic organism to the substrate.

Fungi have evolved mutualistic associations with numerous arthropods. The association between species of Basidiomycota and scale insects is one example. The fungal mycelium covers and protects the insect colonies. The scale insects foster a flow of nutrients from the parasitized plant to the fungus. In a second example, leaf-cutting ants of Central and South America literally farm fungi. They cut disks of leaves from plants and pile them up in gardens. Fungi are cultivated in these gardens, digesting the cellulose that the ants cannot break down. Once smaller sugar molecules are produced and consumed by the fungi, they in turn become a meal for the ants. The insects also patrol their garden, preying on competing fungi. Both ants and fungi benefit from the association. The fungus receives a steady supply of leaves and freedom from competition, while the ants feed on the fungi they cultivate.

Importance to Humans

Although we often think of fungi as organisms that cause diseases and rot food, fungi are important to human life on many levels. As we have seen, they influence the well-being of human populations on a large scale because they help nutrients cycle in ecosystems. They have other ecosystem roles as well. For example, as animal pathogens, fungi help to control the population of damaging pests. These fungi are very specific to the insects they attack and do not infect other animals or plants. The potential to use fungi as microbial insecticides is being investigated, with several species already on the market. For example, the fungus *Beauveria bassiana* is a pesticide that is currently being tested as a possible biological control for the recent spread of emerald ash borer. It has been released in Michigan, Illinois, Indiana, Ohio, West Virginia, and Maryland.

The mycorrhizal relationship between fungi and plant roots is essential for the productivity of farmland. Without the fungal partner in the root systems, 80–90% of trees and grasses would not survive. Mycorrhizal fungal inoculants are available as soil amendments from gardening supply stores and promoted by supporters of organic agriculture.

We also eat some types of fungi. Mushrooms figure prominently in the human diet. Morels, shiitake mushrooms, chanterelles, and truffles are considered delicacies (Figure 4.21). The humble meadow mushroom, *Agaricus campestris*, appears in many dishes. Molds of the genus *Penicillium* ripen many cheeses. They originate in the natural environment such as the caves of Roquefort, France, where wheels of sheep milk cheese are stacked to capture the molds responsible for the blue veins and pungent taste of the cheese.

124

Figure 4.21 The morel mushroom is an ascomycete that is much appreciated for its delicate taste. (credit: Jason Hollinger)

Fermentation—of grains to produce beer, and of fruits to produce wine—is an ancient art that humans in most cultures have practiced for millennia. Wild yeasts are acquired from the environment and used to ferment sugars into CO_2 and ethyl alcohol under anaerobic conditions. It is now possible to purchase isolated strains of wild yeasts from different wine-making regions. Pasteur was instrumental in developing a reliable strain of brewer's yeast, *Saccharomyces cerevisiae*, for the French brewing industry in the late 1850s. It was one of the first examples of biotechnology patenting. Yeast is also used to make breads that rise. The carbon dioxide they produce is responsible for the bubbles produced in the dough that become the air pockets of the baked bread.

Many secondary metabolites of fungi are of great commercial importance. Antibiotics are naturally produced by fungi to kill or inhibit the growth of bacteria, and limit competition in the natural environment. Valuable drugs isolated from fungi include the immunosuppressant drug cyclosporine (which reduces the risk of rejection after organ transplant), the precursors of steroid hormones, and ergot alkaloids used to stop bleeding. In addition, as easily cultured eukaryotic organisms, some fungi are important model research organisms including the red bread mold *Neurospora crassa* and the yeast, *S. cerevisiae.*

The Plant Kingdom

Plants are a large and varied group of organisms. There are close to 300,000 species of catalogued plants.   Of these, about 260,000 are plants that produce seeds. Mosses, ferns, conifers, and flowering plants are all members of the plant kingdom. The plant kingdom contains mostly photosynthetic organisms; a few parasitic forms have lost the ability to photosynthesize. The process of photosynthesis uses chlorophyll, which is located in organelles called chloroplasts. Plants possess cell walls containing cellulose. Most plants reproduce sexually, but they also have diverse methods of asexual reproduction. Plants exhibit indeterminate growth, meaning they do not have a final body form, but continue to grow body mass until they die.

The Major Divisions of Land Plants

Land plants are classified into two major groups according to the absence or presence of vascular tissue, as detailed in (Figure 4.22). Plants that lack vascular tissue formed of specialized cells for the transport of water and nutrients are referred to as **nonvascular plants**. The bryophytes, liverworts, mosses, and hornworts are seedless and nonvascular, and likely appeared early in land plant evolution. **Vascular plants** developed a network of cells that conduct water and solutes through the plant body. The first vascular plants appeared in the late Ordovician (461–444 million years ago) and were probably similar to lycophytes, which include club mosses (not to be confused with the mosses) and the pterophytes (ferns,

125

horsetails, and whisk ferns). Lycophytes and pterophytes are referred to as seedless vascular plants. They do not produce seeds, which are embryos with their stored food reserves protected by a hard casing. The seed plants form the largest group of all existing plants and, hence, dominate the landscape. Seed plants include gymnosperms, most notably conifers, which produce "naked seeds," and the most successful plants, the flowering plants, or angiosperms, which protect their seeds inside chambers at the center of a flower. The walls of these chambers later develop into fruits.

Embryophytes: The Land Plants						
Nonvascular Plants "Bryophytes"			Vascular Plants			
			Seedless Plants		Seed Plants	
			Lycophytes	Pterophytes	Gymno-sperms	Angio-sperms
Liverworts	Hornworts	Mosses	Club Mosses	Whisk Ferns		
			Quillworts	Horsetails		
			Spike Mosses	Ferns		

Figure 4.22 This table shows the major divisions of plants.

To learn more about the evolution of plants and their impact on the development of our planet, watch the BBC show "How to Grow a Planet: Life from Light" found at this **website (http://openstaxcollege.org/l/growing_planet2)** .

| Seedless Plants

By the end of this section, you will be able to:

- Describe the distinguishing traits of the three types of bryophytes
- Identify the new traits that first appear in seedless vascular plants
- Describe the major classes of seedless vascular plants

An incredible variety of seedless plants populates the terrestrial landscape. Mosses grow on tree trunks, and horsetails (Figure 4.23) display their jointed stems and spindly leaves on the forest floor. Yet, seedless plants represent only a small fraction of the plants in our environment. Three hundred million years ago, seedless plants dominated the landscape and grew in the enormous swampy forests of the Carboniferous period. Their decomposing bodies created large deposits of coal that we mine today.

Figure 4.23 Seedless plants like these horsetails (*Equisetum* sp.) thrive in damp, shaded environments under the tree canopy where dryness is a rare occurrence. (credit: Jerry Kirkhart)

Bryophytes

Bryophytes, an informal grouping of the nonvascular plants, are the closest extant relative of early terrestrial plants. The first bryophytes most probably appeared in the Ordovician period, about 490 million years ago. Because of the lack of lignin—the tough polymer in cell walls in the stems of vascular plants—and other resistant structures, the likelihood of bryophytes forming fossils is rather small, though some spores made up of sporopollenin have been discovered that have been attributed to early bryophytes. By the Silurian period (440 million years ago), however, vascular plants had spread throughout the continents. This fact is used as evidence that nonvascular plants must have preceded the Silurian period.

There are about 18,000 species of bryophytes, which thrive mostly in damp habitats, although some grow in deserts. They constitute the major flora of inhospitable environments like the tundra, where their small size and tolerance to desiccation offer distinct advantages. They do not have the specialized cells that conduct fluids found in the vascular plants, and generally lack lignin. In bryophytes, water and nutrients circulate inside specialized conducting cells. Although the name nontracheophyte is more accurate, bryophytes are commonly referred to as nonvascular plants.

In a bryophyte, all the conspicuous vegetative organs belong to the haploid organism, or gametophyte. The diploid sporophyte is barely noticeable. The gametes formed by bryophytes swim using flagella. The sporangium, the multicellular sexual reproductive structure, is present in bryophytes. The embryo also remains attached to the parent plant, which nourishes it. This is a characteristic of land plants.

The bryophytes are divided into three divisions (in plants, the taxonomic level "division" is used instead of phylum): the liverworts, or Marchantiophyta; the hornworts, or Anthocerotophyta; and the mosses, or true Bryophyta.

Liverworts

Liverworts (Marchantiophyta) may be viewed as the plants most closely related to the ancestor that moved to land. Liverworts have colonized many habitats on Earth and diversified to more than 6,000 existing species (Figure 4.24**a**). Some gametophytes form lobate green structures, as seen in **Figure 4.24b**. The shape is similar to the lobes of the liver and, hence, provides the origin of the common name given to the division.

127

Figure 4.24 (a) A 1904 drawing of liverworts shows the variety of their forms. (b) A liverwort, *Lunularia cruciata*, displays its lobate, flat thallus. The organism in the photograph is in the gametophyte stage.

Hornworts

The **hornworts** (Anthocerotophyta) have colonized a variety of habitats on land, although they are never far from a source of moisture. There are about 100 described species of hornworts. The dominant phase of the life cycle of hornworts is the short, blue-green gametophyte. The sporophyte is the defining characteristic of the group. It is a long and narrow pipe-like structure that emerges from the parent gametophyte and maintains growth throughout the life of the plant (Figure 4.25).

Figure 4.25 Hornworts grow a tall and slender sporophyte. (credit: modification of work by Jason Hollinger)

Mosses

More than 12,000 species of **mosses** have been catalogued. Their habitats vary from the tundra, where they are the main vegetation, to the understory of tropical forests. In the tundra, their shallow rhizoids allow them to fasten to a substrate without digging into the frozen soil. They slow down erosion, store moisture and soil nutrients, and provide shelter for small animals and food for larger herbivores, such as the musk ox. Mosses are very sensitive to air pollution and are used to monitor the quality of air. The sensitivity of mosses to copper salts makes these salts a common ingredient of compounds marketed to eliminate mosses in lawns (Figure 4.26).

Figure 4.26 This green feathery moss has reddish-brown sporophytes growing upward. (credit: "Lordgrunt"/Wikimedia Commons)

Vascular Plants

The vascular plants are the dominant and most conspicuous group of land plants. There are about 275,000 species of vascular plants, which represent more than 90 percent of Earth's vegetation. Several evolutionary innovations explain their success and their spread to so many habitats.

Vascular Tissue: Xylem and Phloem

The first fossils that show the presence of vascular tissue are dated to the Silurian period, about 430 million years ago. The simplest arrangement of conductive cells shows a pattern of xylem at the center surrounded by phloem. **Xylem** is the tissue responsible for long-distance transport of water and minerals, the transfer of water-soluble growth factors from the organs of synthesis to the target organs, and storage of water and nutrients.

A second type of vascular tissue is **phloem**, which transports sugars, proteins, and other solutes through the plant. Phloem cells are divided into sieve elements, or conducting cells, and supportive tissue. Together, xylem and phloem tissues form the vascular system of plants.

Roots: Support for the Plant

Roots are not well preserved in the fossil record; nevertheless, it seems that they did appear later in evolution than vascular tissue. The development of an extensive network of roots represented a significant new feature of vascular plants. Thin rhizoids attached the bryophytes to the substrate. Their rather flimsy filaments did not provide a strong anchor for the plant; neither did they absorb water and nutrients. In contrast, roots, with their prominent vascular tissue system, transfer water and minerals from the soil to the rest of the plant. The extensive network of roots that penetrates deep in the ground to reach sources of water also stabilizes trees by acting as ballast and an anchor. The majority of roots establish a symbiotic relationship with fungi, forming mycorrhizae. In the mycorrhizae, fungal hyphae grow around the root and within the root around the cells, and in some instances within the cells. This benefits the plant by greatly increasing the surface area for absorption.

Leaves, Sporophylls, and Strobili

A third adaptation marks seedless vascular plants. Accompanying the prominence of the sporophyte and the development of vascular tissue, the appearance of true leaves improved photosynthetic efficiency. Leaves capture more sunlight with their increased surface area.

In addition to photosynthesis, leaves play another role in the life of the plants. Pinecones, mature fronds of ferns, and flowers are all **sporophylls**—leaves that were modified structurally to bear sporangia. **Strobili** are structures that contain the sporangia. They are prominent in conifers and are known commonly as cones: for example, the pine cones of pine trees.

Seedless Vascular Plants

By the Late Devonian period (385 million years ago), plants had evolved vascular tissue, well-defined leaves, and root systems. With these advantages, plants increased in height and size. During the Carboniferous period (359–299 million years ago), swamp forests of club mosses and horsetails, with some specimens reaching more than 30 meters tall, covered most of the land. These forests gave rise

129

to the extensive coal deposits that gave the Carboniferous its name. In seedless vascular plants, the sporophyte became the dominant phase of the lifecycle.

Water is still required for fertilization of seedless vascular plants, and most favor a moist environment. Modern-day seedless vascular plants include club mosses, horsetails, ferns, and whisk ferns.

Club Mosses

The **club mosses**, or Lycophyta, are the earliest group of seedless vascular plants. They dominated the landscape of the Carboniferous period, growing into tall trees and forming large swamp forests. Today's club mosses are diminutive, evergreen plants consisting of a stem (which may be branched) and small leaves called microphylls (**Figure 14.13**). The division Lycophyta consists of close to 1,000 species, including quillworts (*Isoetales*), club mosses (Lycopodiales), and spike mosses (Selaginellales): none of which is a true moss.

Figure 4.27 *Lycopodium clavatum* is a club moss. (credit: Cory Zanker)

Horsetails

Ferns and whisk ferns belong to the division Pterophyta. A third group of plants in the Pterophyta, the horsetails, is sometimes classified separately from ferns. **Horsetails** have a single genus, *Equisetum*. They are the survivors of a large group of plants, known as Arthrophyta, which produced large trees and entire swamp forests in the Carboniferous. The plants are usually found in damp environments and marshes (**Figure 14.14**).

Figure 4.28 Horsetails thrive in a marsh. (credit: Myriam Feldman)

The stem of a horsetail is characterized by the presence of joints, or nodes: hence the name Arthrophyta, which means "jointed plant". Leaves and branches come out as whorls from the evenly spaced rings. The needle-shaped leaves do not contribute greatly to photosynthesis, the majority of which takes place in the green stem (Figure 4.29).

Figure 4.29 Thin leaves originating at the joints are noticeable on the horsetail plant. (credit: Myriam Feldman)

Ferns and Whisk Ferns

Ferns are considered the most advanced seedless vascular plants and display characteristics commonly observed in seed plants. Ferns form large leaves and branching roots. In contrast, **whisk ferns**, the psilophytes, lack both roots and leaves, which were probably lost by evolutionary reduction. Evolutionary reduction is a process by which natural selection reduces the size of a structure that is no longer favorable in a particular environment. Photosynthesis takes place in the green stem of a whisk fern. Small yellow knobs form at the tip of the branch stem and contain the sporangia. Whisk ferns have been classified outside the true ferns; however, recent comparative analysis of DNA suggests that this group may have lost both vascular tissue and roots through evolution, and is actually closely related to ferns.

With their large fronds, **ferns** are the most readily recognizable seedless vascular plants (Figure 4.30). About 12,000 species of ferns live in environments ranging from tropics to temperate forests. Although some species survive in dry environments, most ferns are restricted to moist and shaded places. They made their appearance in the fossil record during the Devonian period (416–359 million years ago) and expanded during the Carboniferous period, 359–299 million years ago (Figure 4.31).

Figure 4.30 Some specimens of this short tree-fern species can grow very tall. (credit: Adrian Pingstone)

EON	ERA	PERIOD	MILLIONS OF YEARS AGO
Phanerozoic	Cenozoic	Quaternary	---- 1.6 ----
		Tertiary	---- 66 ----
	Mesozoic	Cretaceous	---- 138 ----
		Jurassic	---- 205 ----
		Triassic	---- 240 ----
	Paleozoic	Permian	---- 290 ----
		Carboniferous	---- 360 ----
		Devonian	---- 410 ----
		Silurian	---- 435 ----
		Ordovician	---- 500 ----
		Cambrian	---- 570 ----
Proterozoic			----2500----
Archean			---- 3800? ----
Pre-Archean			

Figure 4.31 This chart shows the geological time scale, beginning with the Pre-Archean eon 3800 million years ago and ending with the Quaternary period in present time. (credit: modification of work by USGS)

CONCEPT in ACTION

openstax COLLEGE

Go to this **website (http://openstaxcollege.org/l/fern_life_cycl2)** to see an animation of the lifecycle of a fern and to test your knowledge.

Landscape Designer

Looking at the well-laid gardens of flowers and fountains seen in royal castles and historic houses of Europe, it is clear that the creators of those gardens knew more than art and design. They were also familiar with the biology of the plants they chose. Landscape design also has strong roots in the United States' tradition. A prime example of early American classical design is Monticello, Thomas Jefferson's private estate; among his many other interests, Jefferson maintained a passion for botany. Landscape layout can encompass a small private space, like a backyard garden; public gathering places, like Central Park in New York City; or an entire city plan, like Pierre L'Enfant's design for Washington, DC.

A landscape designer will plan traditional public spaces—such as botanical gardens, parks, college campuses, gardens, and larger developments—as well as natural areas and private gardens (Figure 4.32). The restoration of natural places encroached upon by human intervention, such as wetlands, also requires the expertise of a landscape designer.

With such an array of required skills, a landscape designer's education includes a solid background in botany, soil science, plant pathology, entomology, and horticulture. Coursework in architecture and design software is also required for the completion of the degree. The successful design of a landscape rests on an extensive knowledge of plant growth requirements, such as light and shade, moisture levels, compatibility of different species, and susceptibility to pathogens and pests. For example, mosses and ferns will thrive in a shaded area where fountains provide moisture; cacti, on the other hand, would not fare well in that environment. The future growth of the individual plants must be taken into account to avoid crowding and competition for light and nutrients. The appearance of the space over time is also of concern. Shapes, colors, and biology must be balanced for a well-maintained and sustainable green space. Art, architecture, and biology blend in a beautifully designed and implemented landscape.

Figure 4.32 This campus garden was designed by students in the horticulture and landscaping department of the college. (credit: Myriam Feldman)

| Seed Plants: Gymnosperms

By the end of this section, you will be able to:

- Discuss the type of seeds produced by gymnosperms, as well as other characteristics of gymnosperms
- List the four groups of modern-day gymnosperms and provide examples of each

The first plants to colonize land were most likely closely related to modern-day mosses (bryophytes) and are thought to have appeared about 500 million years ago. They were followed by liverworts (also bryophytes) and primitive vascular plants, the pterophytes, from which modern ferns are derived. The life cycle of bryophytes and pterophytes is characterized by the alternation of generations. The completion of the life cycle requires water, as the male gametes must swim to the female gametes. The male gametophyte releases sperm, which must swim—propelled by their flagella—to reach and fertilize the female gamete or egg. After fertilization, the zygote matures and grows into a sporophyte, which in turn will form sporangia, or "spore vessels," in which mother cells undergo meiosis and produce haploid spores. The release of spores in a suitable environment will lead to germination and a new generation of gametophytes.

The Evolution of Seed Plants

In seed plants, the evolutionary trend led to a dominant sporophyte generation, in which the larger and more ecologically significant generation for a species is the diploid plant. At the same time, the trend led to a reduction in the size of the gametophyte, from a conspicuous structure to a microscopic cluster of cells enclosed in the tissues of the sporophyte. Lower vascular plants, such as club mosses and ferns, are mostly homosporous (produce only one type of spore). In contrast, all seed plants, or spermatophytes, are heterosporous, forming two types of spores: megaspores (female) and microspores (male). Megaspores develop into female gametophytes that produce eggs, and microspores mature into male gametophytes that generate sperm. Because the gametophytes mature within the spores, they are not free-living, as are the gametophytes of other seedless vascular plants. Heterosporous seedless plants are seen as the evolutionary forerunners of seed plants.

Seeds and pollen—two adaptations to drought—distinguish seed plants from other (seedless) vascular plants. Both adaptations were critical to the colonization of land. Fossils place the earliest distinct seed plants at about 350 million years ago. The earliest reliable record of gymnosperms dates their appearance to the Carboniferous period (359–299 million years ago). Gymnosperms were preceded by the progymnosperms ("first naked seed plants"). This was a transitional group of plants that superficially resembled conifers ("cone bearers") because they produced wood from the secondary growth of the vascular tissues; however, they still reproduced like ferns, releasing spores to the environment. In the Mesozoic era (251–65.5 million years ago), gymnosperms dominated the landscape. Angiosperms took over by the middle of the Cretaceous period (145.5–65.5 million years ago) in the late Mesozoic era, and have since become the most abundant plant group in most terrestrial biomes.

The two innovative structures of pollen and seed allowed seed plants to break their dependence on water for reproduction and development of the embryo, and to conquer dry land. The pollen grains carry the male gametes of the plant. The small haploid ($1n$) cells are encased in a protective coat that prevents desiccation (drying out) and mechanical damage. Pollen can travel far from the sporophyte that bore it, spreading the plant's genes and avoiding competition with other plants. The seed offers the embryo protection, nourishment and a mechanism to maintain dormancy for tens or even thousands of years, allowing it to survive in a harsh environment and ensuring germination when growth conditions are optimal. Seeds allow plants to disperse the next generation through both space and time. With such evolutionary advantages, seed plants have become the most successful and familiar group of plants.
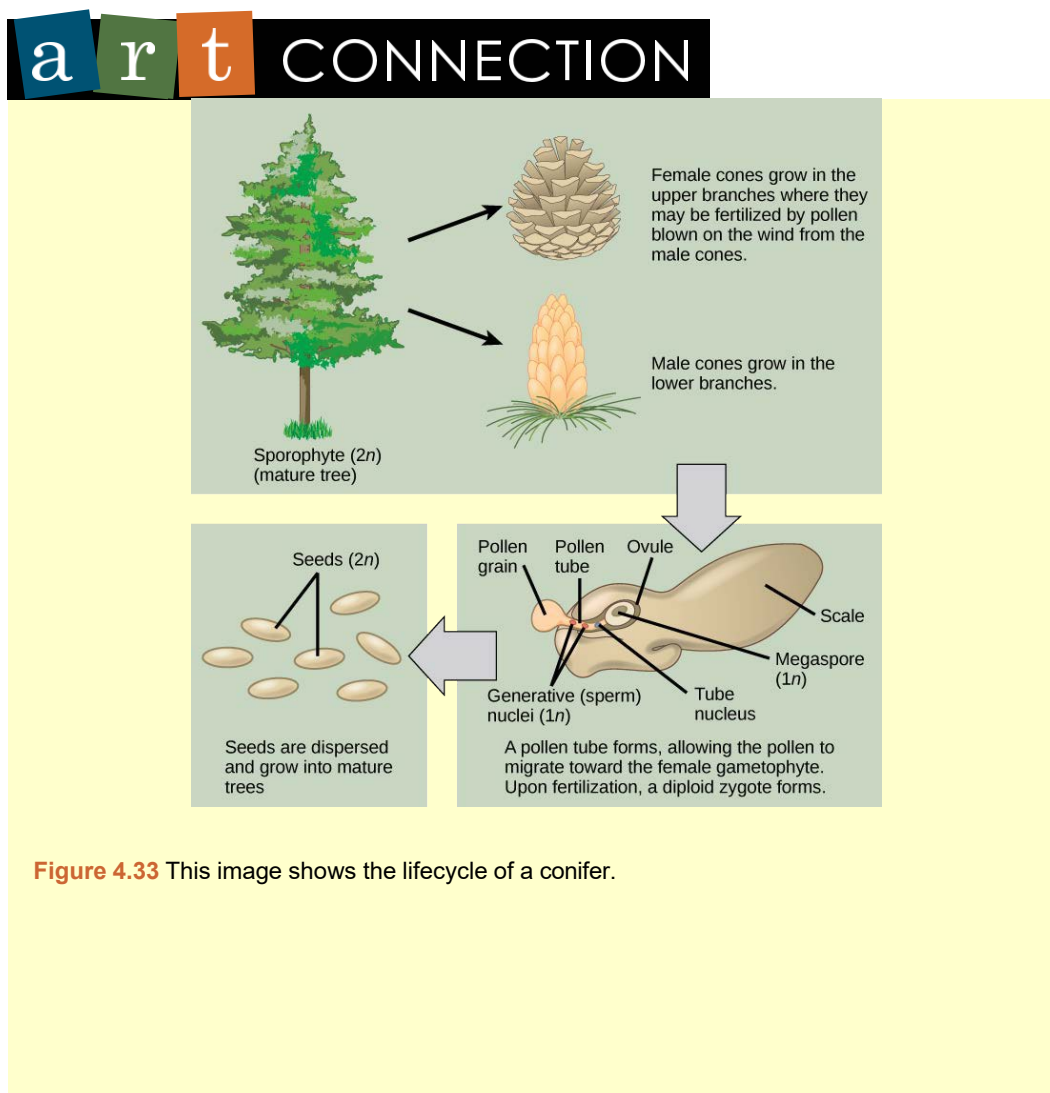
Gymnosperms

Gymnosperms ("naked seed") are a diverse group of seed plants and are paraphyletic. Paraphyletic groups do not include descendants of a single common ancestor. Gymnosperm characteristics include naked seeds, separate female and male gametes, pollination by wind, and tracheids, which transport water and solutes in the vascular system.

Life Cycle of a Conifer

Pine trees are conifers and carry both male and female sporophylls on the same plant. Like all gymnosperms, pines are heterosporous and produce male microspores and female megaspores. In the male cones, or staminate cones, the **microsporocytes** give rise to microspores by meiosis. The microspores then develop into pollen grains. Each pollen grain contains two cells: one generative cell that will divide into two sperm, and a second cell that will become the pollen tube cell. In the spring, pine trees release large amounts of yellow pollen, which is carried by the wind. Some gametophytes will land on a female cone. The pollen tube grows from the pollen grain slowly, and the generative cell in the pollen grain divides into two sperm cells by mitosis. One of the sperm cells will finally unite its haploid nucleus with the haploid nucleus of an egg cell in the process of fertilization.

Female **cones**, or ovulate cones, contain two ovules per scale. One **megasporocyte** undergoes meiosis in each ovule. Only a single surviving haploid cell will develop into a female multicellular gametophyte

134

that encloses an egg. On fertilization, the zygote will give rise to the embryo, which is enclosed in a seed coat of tissue from the parent plant. Fertilization and seed development is a long process in pine trees—it may take up to two years after pollination. The seed that is formed contains three generations of tissues: the seed coat that originates from the parent plant tissue, the female gametophyte that will provide nutrients, and the embryo itself. Figure 4.33 illustrates the life cycle of a conifer.

ART CONNECTION

Figure 4.33 This image shows the lifecycle of a conifer.

CONCEPT in ACTION

Watch this **video (http://openstaxcollege.org/l/gymnosperm)** to see the process of seed production in gymnosperms.

Diversity of Gymnosperms

Modern gymnosperms are classified into four major divisions and comprise about 1,000 described species. Coniferophyta, Cycadophyta, and Ginkgophyta are similar in their production of secondary cambium (cells that generate the vascular system of the trunk or stem) and their pattern of seed

development, but are not closely related phylogenetically to each other. Gnetophyta are considered the closest group to angiosperms because they produce true xylem tissue that contains both tracheids and vessel elements.

Conifers

Conifers are the dominant phylum of gymnosperms, with the most variety of species. Most are tall trees that usually bear scale-like or needle-like leaves. The thin shape of the needles and their waxy cuticle limits water loss through transpiration. Snow slides easily off needle-shaped leaves, keeping the load light and decreasing breaking of branches. These adaptations to cold and dry weather explain the predominance of conifers at high altitudes and in cold climates. Conifers include familiar evergreen trees, such as pines, spruces, firs, cedars, sequoias, and yews (Figure 4.34). A few species are deciduous and lose their leaves all at once in fall. The European larch and the tamarack are examples of deciduous conifers. Many coniferous trees are harvested for paper pulp and timber. The wood of conifers is more primitive than the wood of angiosperms; it contains tracheids, but no vessel elements, and is referred to as "soft wood."

Figure 4.34 Conifers are the dominant form of vegetation in cold or arid environments and at high altitudes. Shown here are the (a) evergreen spruce, (b) sequoia, (c) juniper, and (d) a deciduous gymnosperm: the tamarack *Larix larcinia*. Notice the yellow leaves of the tamarack. (credit b: modification of work by Alan Levine; credit c: modification of work by Wendy McCormac; credit d: modification of work by Micky Zlimen)

Cycads

Cycads thrive in mild climates and are often mistaken for palms because of the shape of their large, compound leaves. They bear large cones, and unusually for gymnosperms, may be pollinated by beetles, rather than wind. They dominated the landscape during the age of dinosaurs in the Mesozoic era (251–65.5 million years ago). Only a hundred or so cycad species persisted to modern times. They face possible extinction, and several species are protected through international conventions. Because of their attractive shape, they are often used as ornamental plants in gardens (Figure 4.35).

136

Figure 4.35 This *Encephalartos ferox* cycad exhibits large cones. (credit: Wendy Cutler)

Gingkophytes

The single surviving species of **ginkgophyte** is the *Ginkgo biloba* (Figure 4.36). Its fan-shaped leaves, unique among seed plants because they feature a dichotomous venation pattern, turn yellow in autumn and fall from the plant. For centuries, Buddhist monks cultivated *Ginkgo biloba,* ensuring its preservation. It is planted in public spaces because it is unusually resistant to pollution. Male and female organs are found on separate plants. Usually, only male trees are planted by gardeners because the seeds produced by the female plant have an off-putting smell of rancid butter.
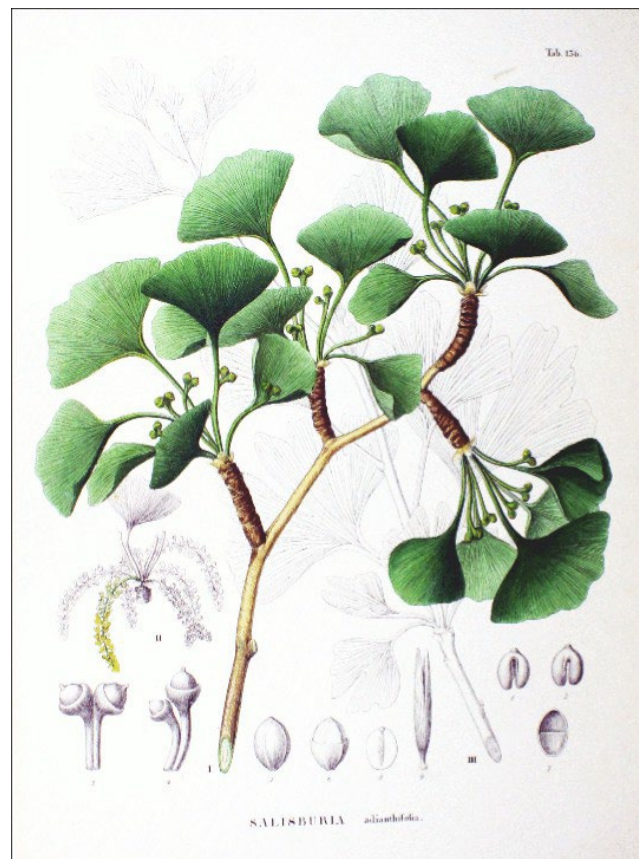
Figure 4.36 This plate from the 1870 book *Flora Japonica, Sectio Prima (Tafelband)* depicts the leaves and fruit of *Gingko biloba*, as drawn by Philipp Franz von Siebold and Joseph Gerhard Zuccarini.

Gnetophytes

Gnetophytes are the closest relatives to modern angiosperms, and include three dissimilar genera of plants. Like angiosperms, they have broad leaves. *Gnetum* species are mostly vines in tropical and subtropical zones. The single species of *Welwitschia* is an unusual, low-growing plant found in the deserts of Namibia and Angola. It may live for up to 2000 years. The genus *Ephedra* is represented in

138

North America in dry areas of the southwestern United States and Mexico (Figure 4.37). *Ephedra's* small, scale-like leaves are the source of the compound ephedrine, which is used in medicine as a potent decongestant. Because ephedrine is similar to amphetamines, both in chemical structure and neurological effects, its use is restricted to prescription drugs. Like angiosperms, but unlike other gymnosperms, all gnetophytes possess vessel elements in their xylem.

Figure 4.37 *Ephedra viridis*, known by the common name Mormon tea, grows in the western United States. (credit: US National Park Service, USDA-NRCS PLANTS Database)

Watch this **BBC video (http://openstaxcollege.org/l/welwitschia)** describing the amazing strangeness of Welwitschia.

Seed Plants: Angiosperms

By the end of this section, you will be able to:

- Describe the main parts of a flower and their purpose
- Detail the life cycle of an angiosperm
- Discuss the two main groups into which flower plants are divided, as well as explain how basal angiosperms differ from others

From their humble and still obscure beginning during the early Jurassic period (202–145.5 MYA), the angiosperms, or flowering plants, have successfully evolved to dominate most terrestrial ecosystems. Angiosperms include a staggering number of genera and species; with more than 260,000 species, the division is second only to insects in terms of diversification (Figure 4.38).

Figure 4.38 These flowers grow in a botanical garden border in Bellevue, WA. Flowering plants dominate terrestrial landscapes. The vivid colors of flowers are an adaptation to pollination by insects and birds. (credit: Myriam Feldman)

Angiosperm success is a result of two novel structures that ensure reproductive success: flowers and fruit. Flowers allowed plants to form cooperative evolutionary relationships with animals, in particular insects, to disperse their pollen to female gametophytes in a highly targeted way. Fruit protect the developing embryo and serve as an agent of dispersal. Different structures on fruit reflect the dispersal strategies that help with the spreading of seeds.

Flowers

Flowers are modified leaves or sporophylls organized around a central stalk. Although they vary greatly in appearance, all flowers contain the same structures: sepals, petals, pistils, and stamens. A whorl of **sepals** (the **calyx**) is located at the base of the peduncle, or stem, and encloses the floral bud before it opens. Sepals are usually photosynthetic organs, although there are some exceptions. For example, the corolla in lilies and tulips consists of three sepals and three petals that look virtually identical—this led botanists to coin the word tepal. **Petals** (collectively the **corolla**) are located inside the whorl of sepals and usually display vivid colors to attract pollinators. Flowers pollinated by wind are usually small and dull. The sexual organs are located at the center of the flower.

As illustrated in Figure 4.39, the stigma, style, and ovary constitute the female organ, the **carpel** or **pistil**, which is also referred to as the **gynoecium**. A gynoecium may contain one or more carpels within a single flower. The megaspores and the female gametophytes are produced and protected by the thick tissues of the carpel. A long, thin structure called a **style** leads from the sticky **stigma**, where pollen is deposited, to the **ovary** enclosed in the carpel. The ovary houses one or more ovules that will each develop into a seed upon fertilization. The male reproductive organs, the androecium or **stamens**, surround the central carpel. Stamens are composed of a thin stalk called a **filament** and a sac-like structure, the **anther**, in which microspores are produced by meiosis and develop into pollen grains. The filament supports the anther.

140
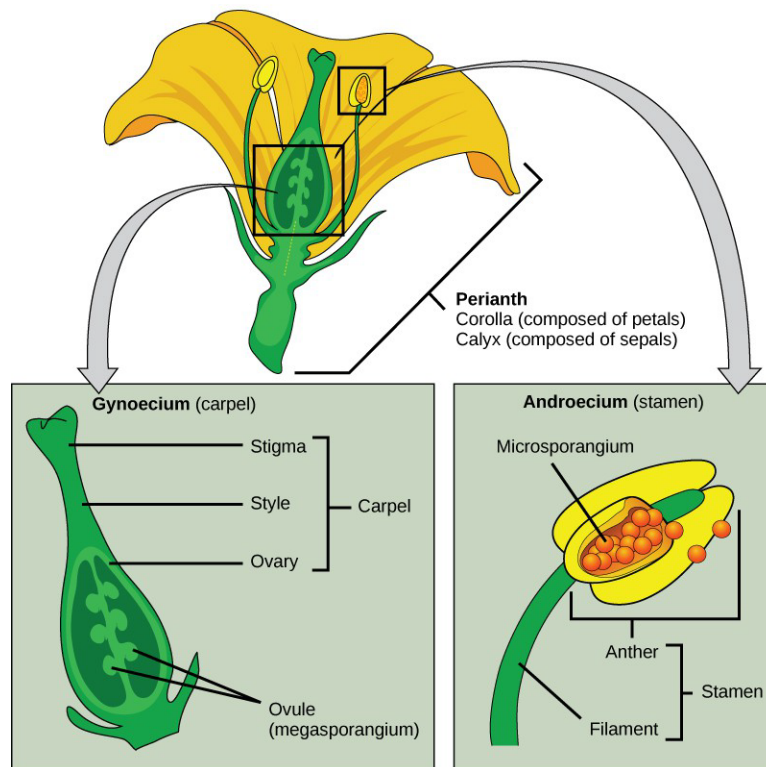
Figure 4.39 This image depicts the structure of a perfect and complete flower. Perfect flowers carry both male and female floral organs. (credit: modification of work by Mariana Ruiz Villareal)

Fruit

The seed forms in an ovary, which enlarges as the seeds grow. As the seed develops, the walls of the ovary also thicken and form the fruit. In botany, a fruit is a fertilized and fully grown, ripened ovary. Many foods commonly called vegetables are actually fruit. Eggplants, zucchini, string beans, and bell peppers are all technically fruit because they contain seeds and are derived from the thick ovary tissue. Acorns and winged maple keys, whose scientific name is a samara, are also fruit.

Mature fruit can be described as fleshy or dry. Fleshy fruit include the familiar berries, peaches, apples, grapes, and tomatoes. Rice, wheat, and nuts are examples of dry fruit. Another distinction is that not all fruits are derived from the ovary. Some fruits are derived from separate ovaries in a single flower, such as the raspberry. Other fruits, such as the pineapple, form from clusters of flowers. Additionally, some fruits, like watermelon and orange, have rinds. Regardless of how they are formed, fruits are an agent of dispersal. The variety of shapes and characteristics reflect the mode of dispersal. The light, dry fruits of trees and dandelions are carried by the wind. Floating coconuts are transported by water. Some fruits are colored, perfumed, sweet, and nutritious to attract herbivores, which eat the fruit and disperse the tough undigested seeds in their feces. Other fruits have burs and hooks that cling to fur and hitch rides on animals.

The Life Cycle of an Angiosperm

The adult, or sporophyte, phase is the main phase in an angiosperm's life cycle. Like gymnosperms, angiosperms are heterosporous. They produce microspores, which develop into pollen grains (the male gametophytes), and megaspores, which form an ovule containing the female gametophytes. Inside the anthers' microsporangia (Figure 4.40), male microsporocytes divide by meiosis, generating haploid microspores that undergo mitosis and give rise to pollen grains. Each pollen grain contains two cells: one generative cell that will divide into two sperm, and a second cell that will become the pollen tube cell.
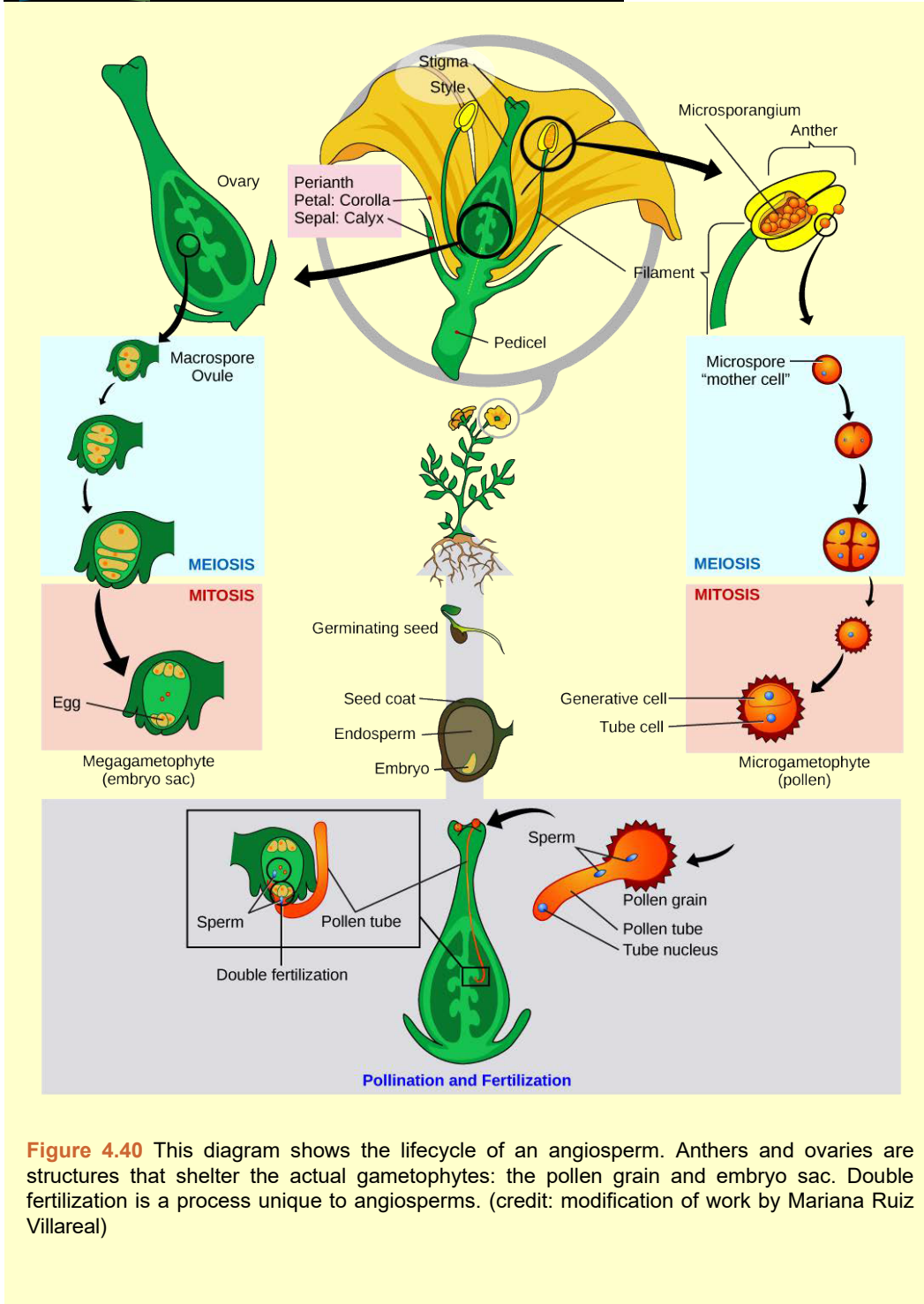
Figure 4.40 This diagram shows the lifecycle of an angiosperm. Anthers and ovaries are structures that shelter the actual gametophytes: the pollen grain and embryo sac. Double fertilization is a process unique to angiosperms. (credit: modification of work by Mariana Ruiz Villareal)

In the ovules, the female gametophyte is produced when a megasporocyte undergoes meiosis to produce four haploid megaspores. One of these is larger than the others and undergoes mitosis to form the female gametophyte or embryo sac. Three mitotic divisions produce eight nuclei in seven cells. The egg and two cells move to one end of the embryo sac (gametophyte) and three cells move to the other end. Two of the nuclei remain in a single cell and fuse to form a 2*n* nucleus; this cell moves to the center of the embryo sac.

142

When a pollen grain reaches the stigma, a pollen tube extends from the grain, grows down the style, and enters through an opening in the integuments of the ovule. The two sperm cells are deposited in the embryo sac.

What occurs next is called a double fertilization event (Figure 4.41) and is unique to angiosperms. One sperm and the egg combine, forming a diploid zygote—the future embryo. The other sperm fuses with the diploid nucleus in the center of the embryo sac, forming a triploid cell that will develop into the endosperm: a tissue that serves as a food reserve. The zygote develops into an embryo with a radicle, or small root, and one or two leaf-like organs called **cotyledons**. Seed food reserves are stored outside the embryo, and the cotyledons serve as conduits to transmit the broken-down food reserves to the developing embryo. The seed consists of a toughened layer of integuments forming the coat, the endosperm with food reserves and, at the center, the well-protected embryo.
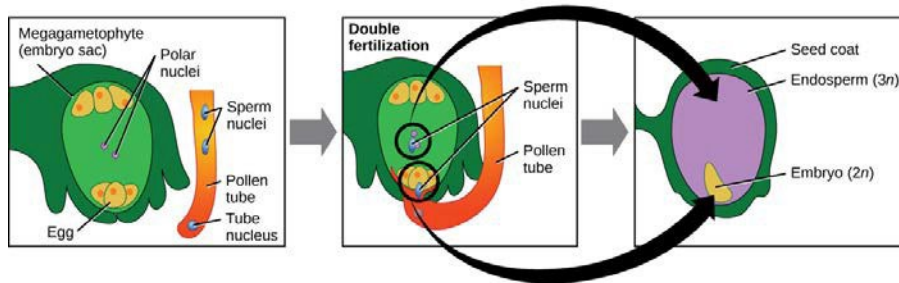
Figure 4.41 Double fertilization occurs only in angiosperms. (credit: modification of work by Mariana Ruiz Villareal)

Most flowers carry both stamens and carpels; however, a few species self-pollinate. These are known as "perfect" flowers because they contain both types of sex organs (Figure 4.42). Biochemical and anatomical barriers to self-pollination promote cross-pollination. Self-pollination is a severe form of inbreeding, and can increase the number of genetic defects in offspring.

A plant may have perfect flowers, and thus have both genders in each flower; or, it may have imperfect flowers of both kinds on one plant (Figure 4.42). In each case, such species are called monoecious plants, meaning "one house." Some botanists refer to plants with perfect flowers simply as hermaphroditic. Some plants are dioecious, meaning "two houses," and have male and female flowers ("imperfect flowers") on different plants. In these species, cross-pollination occurs all the time.

Figure 4.42 Monoecious plants have both male and female reproductive structures on the same flower or plant. In dioecious plants, males and females reproductive structures are on separate plants. (credit a: modification of work by Liz West; credit c: modification of work by Scott Zona)

- Download for free at http://openstaxcollege.org/textbooks/concepts-of- biology/get.

Diversity of Angiosperms

Angiosperms are classified in a single division, the **Anthophyta**. Modern angiosperms appear to be a monophyletic group, which means that they originate from a single ancestor. Flowering plants are divided into two major groups, according to the structure of the cotyledons, the pollen grains, and other features: **monocots**, which include grasses and lilies, and **eudicots** or **dicots**, a polyphyletic group. **Basal angiosperms** are a group of plants that are believed to have branched off before the separation into monocots and eudicots because they exhibit traits from both groups. They are categorized separately in many classification schemes, and correspond to a grouping known as the Magnoliidae. The Magnoliidae group is comprised of magnolia trees, laurels, water lilies, and the pepper family.

Basal Angiosperms

The Magnoliidae are represented by the magnolias: tall trees that bear large, fragrant flowers with many parts, and are considered archaic (Figure 4.43**d**). Laurel trees produce fragrant leaves and small inconspicuous flowers. The Laurales are small trees and shrubs that grow mostly in warmer climates. Familiar plants in this group include the bay laurel, cinnamon, spice bush (Figure 4.43**a**), and the avocado tree. The Nymphaeales are comprised of the water lilies, lotus (Figure 4.43**c**), and similar plants. All species of the Nymphaeales thrive in freshwater biomes, and have leaves that float on the water surface or grow underwater. Water lilies are particularly prized by gardeners, and have graced ponds and pools since antiquity. The Piperales are a group of herbs, shrubs, and small trees that grow in tropical climates. They have small flowers without petals that are tightly arranged in long spikes. Many species are the source of prized fragrances or spices; for example, the berries of *Piper nigrum* (Figure 4.43**b**) are the familiar black pepper that is used to flavor many dishes.

(a)   (b)
(c)   (d)
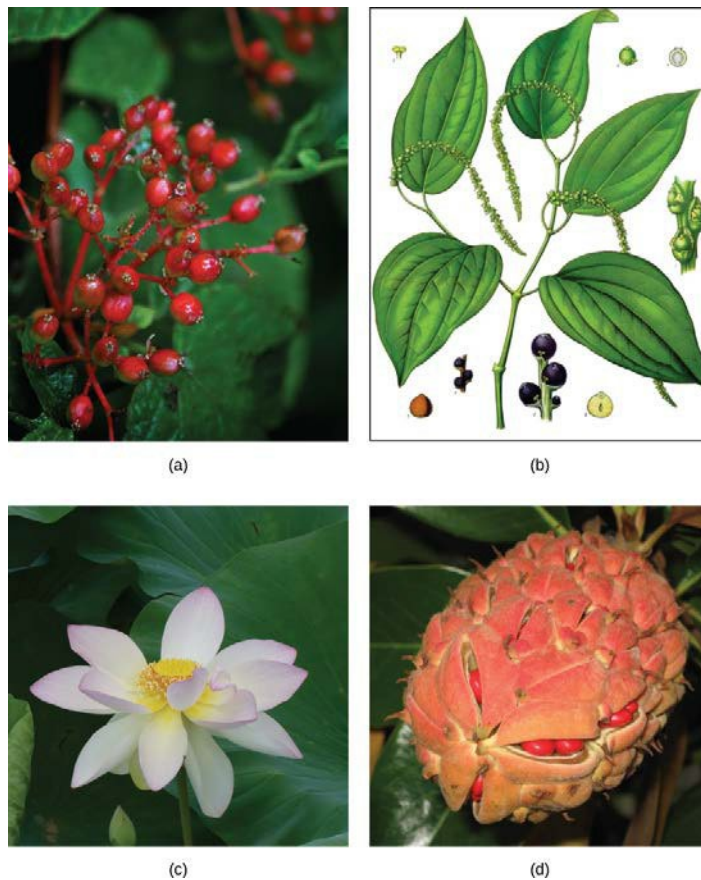
Figure 4.43 The (a) southern spicebush belongs to the *Laurales*, the same family as cinnamon and bay laurel. The fruit of (b) the *Piper nigrum* plant is black pepper, the main product that was traded along spice routes. Notice the small, unobtrusive clustered flowers. (c) Lotus flowers, *Nelumbo nucifera*, have been cultivated since antiquity for their ornamental value; the root of the lotus flower is eaten as a vegetable. The (d) red berries of a magnolia tree, characteristic of the final stage, are just starting to appear. (credit a: modification of work by Cory Zanker; credit b: modification of work by Franz Eugen Köhler; credit c: modification of work by "berduchwal"/Flickr; credit d: modification of work by "Coastside2"/Wikimedia Commons)

Monocots

Plants in the monocot group have a single cotyledon in the seedling, and also share other anatomical features. Veins run parallel to the length of the leaves, and flower parts are arranged in a three- or six-fold symmetry. The pollen from the first angiosperms was monosulcate (containing a single furrow or pore through the outer layer). This feature is still seen in the modern monocots. True woody tissue is rarely found in monocots, and the vascular tissue of the stem is not arranged in any particular pattern. The root system is mostly adventitious (unusually positioned) with no major taproot. The monocots include familiar plants such as the true lilies (not to be confused with the water lilies), orchids, grasses, and palms. Many important crops, such as rice and other cereals (Figure 4.44**a**), corn, sugar cane, and tropical fruit, including bananas and pineapple, belong to the monocots.

Figure 4.44 The major crops in the world are flowering plants. One staple food, (a) rice, is a monocot, as are other cereals, while (b) beans are eudicots. Some popular flowers, such as this (c) lily are monocots; while others, such as this (d) daisy are eudicots. (credit a: modification of work by David Nance; credit b: modification of work by USDA, ARS; credit c: modification of work by "longhorndave"/Flickr; credit d: modification of work by "Cellulaer"/NinjaPhoto)

Eudicots

Eudicots, or true dicots, are characterized by the presence of two cotyledons. Veins form a network in leaves. Flower parts come in four, five, or many whorls. Vascular tissue forms a ring in the stem. (In monocots, vascular tissue is scattered in the stem.) Eudicots can be **herbaceous** (like dandelions or violets), or produce woody tissues. Most eudicots produce pollen that is trisulcate or triporate, with three furrows or pores. The root system is usually anchored by one main root developed from the embryonic radicle. Eudicots comprise two-thirds of all flowering plants. Many species seem to exhibit characteristics that belong to either group; therefore, the classification of a plant as a monocot or a eudicot is not always clearly evident (**Table 4.1**).

Comparison of Structural Characteristics of Monocots and Eudicots

Characteristic	Monocot	Eudicot
Cotyledon	One	Two
Veins in leaves	Parallel	Network (branched)
Vascular tissue	Scattered	Arranged in ring pattern
Roots	Network of adventitious roots	Tap root with many lateral roots
Pollen	Monosulcate	Trisulcate

Table 4.1

145

Comparison of Structural Characteristics of Monocots and Eudicots

Characteristic	Monocot	Eudicot
Flower parts	Three or multiple of three	Four, five, multiple of four or five and whorls

Table 4.1

Explore this **website (http://openstaxcollege.org/l/pollinators)** for more information on poillinators.

Plant Tissues

Plants are multicellular eukaryotes with tissue systems made of various cell types that carry out specific functions. Plant tissue systems fall into one of two general types: meristematic tissue, and permanent (or non-meristematic) tissue. Cells of the meristematic tissue are found in **meristems**, which are plant regions of continuous cell division and growth. **Meristematic tissue** cells are either undifferentiated or incompletely differentiated, and they continue to divide and contribute to the growth of the plant. In contrast, **permanent tissue** consists of plant cells that are no longer actively dividing.

Meristematic tissues consist of three types, based on their location in the plant. **Apical meristems** contain meristematic tissue located at the tips of stems and roots, which enable a plant to extend in length. **Lateral**

146

meristems facilitate growth in thickness or girth in a maturing plant. **Intercalary meristems** occur only in monocots, at the bases of leaf blades and at nodes (the areas where leaves attach to a stem). This tissue enables the monocot leaf blade to increase in length from the leaf base; for example, it allows lawn grass leaves to elongate even after repeated mowing.

Meristems produce cells that quickly differentiate, or specialize, and become permanent tissue. Such cells take on specific roles and lose their ability to divide further. They differentiate into three main types: dermal, vascular, and ground tissue. **Dermal tissue** covers and protects the plant, and **vascular tissue** transports water, minerals, and sugars to different parts of the plant. **Ground tissue** serves as a site for photosynthesis, provides a supporting matrix for the vascular tissue, and helps to store water and sugars.

Secondary tissues are either simple (composed of similar cell types) or complex (composed of different cell types). Dermal tissue, for example, is a simple tissue that covers the outer surface of the plant and controls gas exchange. Vascular tissue is an example of a complex tissue, and is made of two specialized conducting tissues: xylem and phloem. Xylem tissue transports water and nutrients from the roots to different parts of the plant, and includes three different cell types: vessel elements and tracheids (both of which conduct water), and xylem parenchyma. Phloem tissue, which transports organic compounds from the site of photosynthesis to other parts of the plant, consists of four different cell types: sieve cells (which conduct photosynthates), companion cells, phloem parenchyma, and phloem fibers. Unlike xylem conducting cells, phloem conducting cells are alive at maturity. The xylem and phloem always lie adjacent to each other (Figure 4.45). In stems, the xylem and the phloem form a structure called a **vascular bundle**; in roots, this is termed the **vascular stele** or **vascular cylinder**.
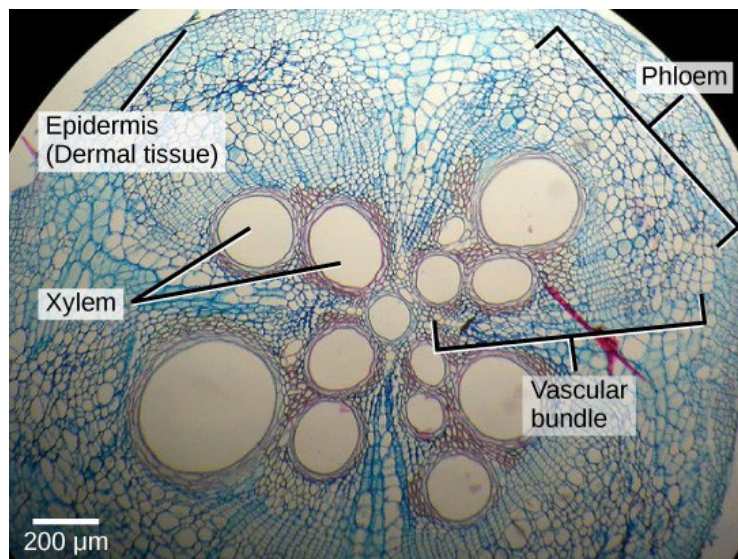
Figure 4.45 This light micrograph shows a cross section of a squash (*Curcurbita maxima*) stem. Each teardrop-shaped vascular bundle consists of large xylem vessels toward the inside and smaller phloem cells toward the outside. Xylem cells, which transport water and nutrients from the roots to the rest of the plant, are dead at functional maturity. Phloem cells, which transport sugars and other organic compounds from photosynthetic tissue to the rest of the plant, are living. The vascular bundles are encased in ground tissue and surrounded by dermal tissue. (credit: modification of work by "(biophotos)"/Flickr; scale-bar data from Matt Russell)

Stems
By the end of this section, you will be able to do the following:
- Describe the main function and basic structure of stems
- Compare and contrast the roles of dermal tissue, vascular tissue, and ground tissue
- Distinguish between primary growth and secondary growth in stems
- Summarize the origin of annual rings
- List and describe examples of modified stems

Stems are a part of the shoot system of a plant. They may range in length from a few millimeters to hundreds of meters, and also vary in diameter, depending on the plant type. Stems are usually above ground, although the stems of some plants, such as the potato, also grow underground. Stems may be herbaceous (soft) or woody in nature. Their main function is to provide support to

the plant, holding leaves, flowers and buds; in some cases, stems also store food for the plant. A stem may be unbranched, like that of a palm tree, or it may be highly branched, like that of a magnolia tree. The stem of the plant connects the roots to the leaves, helping to transport absorbed water and minerals to different parts of the plant. It also helps to transport the products of photosynthesis, namely sugars, from the leaves to the rest of the plant.

Plant stems, whether above or below ground, are characterized by the presence of nodes and internodes (Figure 4.46). **Nodes** are points of attachment for leaves, aerial roots, and flowers. The stem region between two nodes is called an **internode**. The stalk that extends from the stem to the base of the leaf is the petiole. An **axillary bud** is usually found in the axil—the area between the base of a leaf and the stem—where it can give rise to a branch or a flower. The apex (tip) of the shoot contains the apical meristem within the **apical bud**.
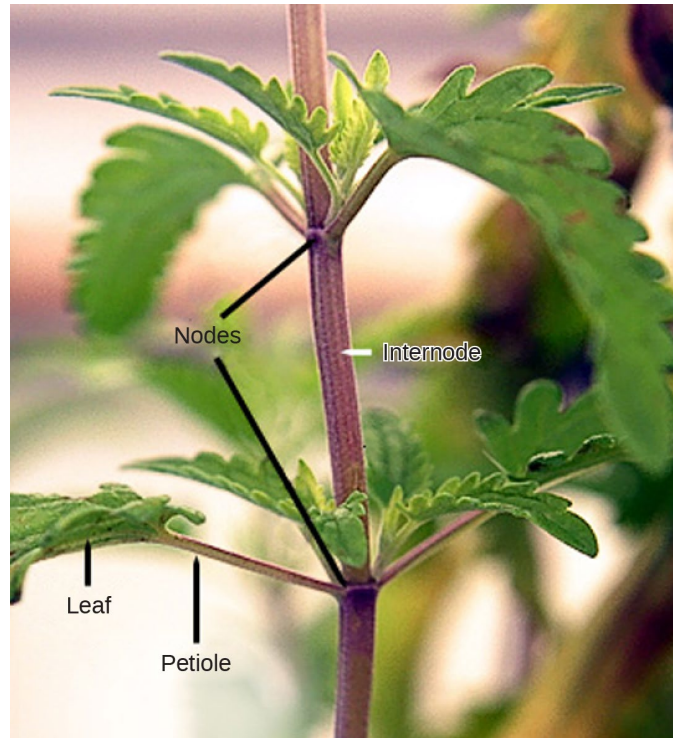
Figure 4.46 Leaves are attached to the plant stem at areas called nodes. An internode is the stem region between two nodes. The petiole is the stalk connecting the leaf to the stem. The leaves just above the nodes arose from axillary buds.

Stem Anatomy

The stem and other plant organs arise from the ground tissue, and are primarily made up of simple tissues formed from three types of cells: parenchyma, collenchyma, and sclerenchyma cells.

Parenchyma cells are the most common plant cells (Figure 4.47). They are found in the stem, the root, the inside of the leaf, and the pulp of the fruit. Parenchyma cells are responsible for metabolic functions, such as photosynthesis, and they help repair and heal wounds. Some parenchyma cells also store starch.

148
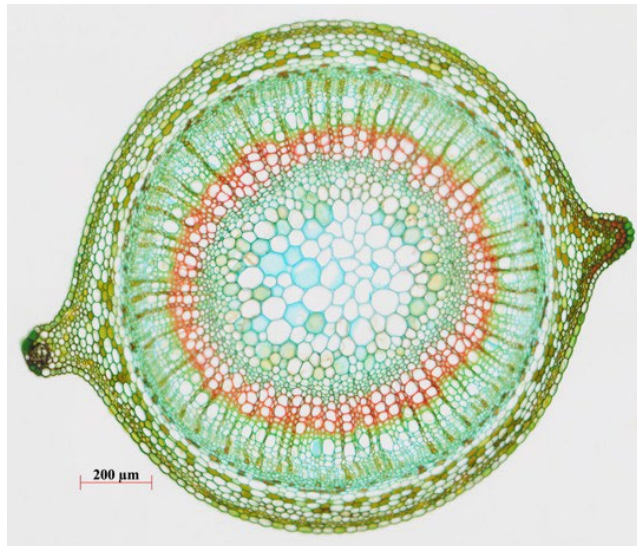
Figure 4.47 The stem of common St John's Wort (*Hypericum perforatum*) is shown in cross section in this light micrograph. The central pith (greenish-blue, in the center) and peripheral cortex (narrow zone 3–5 cells thick just inside the epidermis) are composed of parenchyma cells. Vascular tissue composed of xylem (red) and phloem tissue (green, between the xylem and cortex) surrounds the pith. (credit: Rolf-Dieter Mueller)

Collenchyma cells are elongated cells with unevenly thickened walls (Figure 4.48). They provide structural support, mainly to the stem and leaves. These cells are alive at maturity and are usually found below the epidermis. The "strings" of a celery stalk are an example of collenchyma cells.
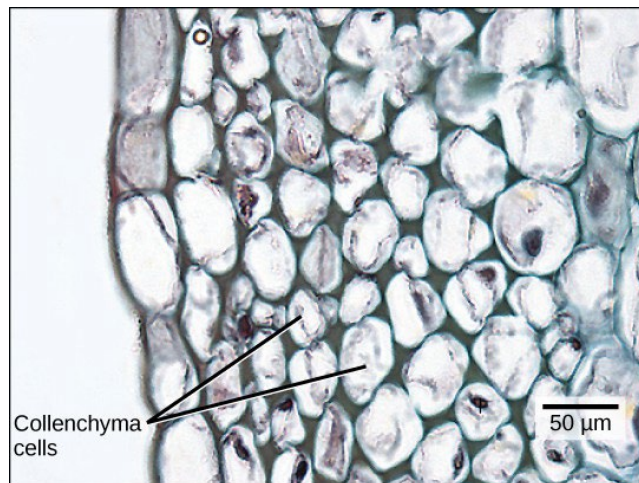
Figure 4.48 Collenchyma cell walls are uneven in thickness, as seen in this light micrograph. They provide support to plant structures. (credit: modification of work by Carl Szczerski; scale-bar data from Matt Russell)

Sclerenchyma cells also provide support to the plant, but unlike collenchyma cells, many of them are dead at maturity. There are two types of sclerenchyma cells: fibers and sclereids. Both types have secondary cell walls that are thickened with deposits of lignin, an organic compound that is a key component of wood. Fibers are long, slender cells; sclereids are smaller-sized. Sclereids give pears their gritty texture. Humans use sclerenchyma fibers to make linen and rope (Figure 4.49).

149

(a)

(b)

(c)
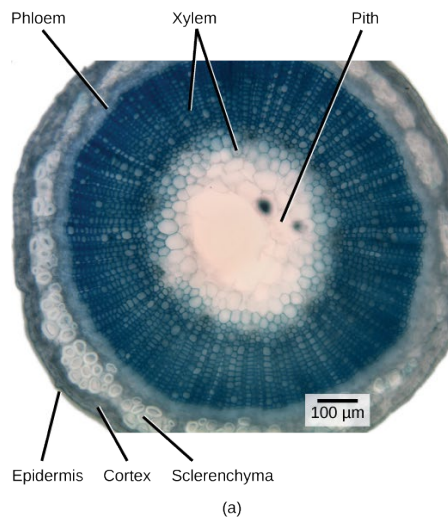
Figure 4.49 The central pith and outer cortex of the (a) flax stem are made up of parenchyma cells. Inside the cortex is a layer of sclerenchyma cells, which make up the fibers in flax rope and clothing. Humans have grown and harvested flax for thousands of years. In

(b) this drawing, fourteenth-century women prepare linen. The (c) flax plant is grown and harvested for its fibers, which are used to weave linen, and for its seeds, which are the source of linseed oil. (credit a: modification of work by Emmanuel Boutet based on original work by Ryan R. MacKenzie; credit c: modification of work by Brian Dearth; scale-bar data from Matt Russell)

Which layers of the stem are made of parenchyma cells?

a. cortex and pith
b. phloem
c. sclerenchyma
d. xylem

Like the rest of the plant, the stem has three tissue systems: dermal, vascular, and ground tissue. Each is distinguished by characteristic cell types that perform specific tasks necessary for the plant's growth and survival.

Dermal Tissue

The dermal tissue of the stem consists primarily of **epidermis**, a single layer of cells covering and protecting the underlying tissue. Woody plants have a tough, waterproof outer layer of cork cells commonly known as **bark**, which further protects the plant from damage. Epidermal cells are the most numerous and least differentiated of the cells in the epidermis. The epidermis of a leaf also contains openings known as stomata, through which the exchange of gases takes place (Figure 4.50). Two cells, known as **guard cells**, surround each leaf stoma, controlling its opening and closing and thus regulating the uptake of carbon dioxide and the release of oxygen and water vapor. **Trichomes** are hair-like structures on the epidermal surface. They help to reduce **transpiration** (the loss of water by aboveground plant parts), increase solar reflectance, and store compounds that defend the leaves against predation by herbivores.

150
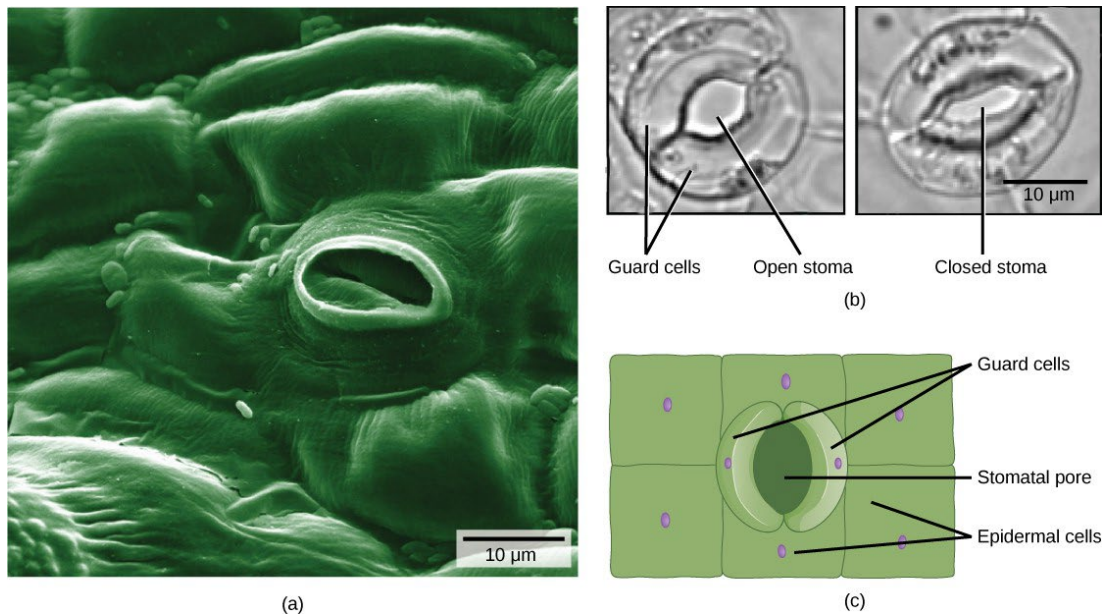
Figure 4.50 Openings called stomata (singular: stoma) allow a plant to take up carbon dioxide and release oxygen and water vapor. The (a) colorized scanning-electron micrograph shows a closed stoma of a dicot. Each stoma is flanked by two guard cells that regulate its (b) opening and closing. The (c) guard cells sit within the layer of epidermal cells. (credit a: modification of work by Louisa Howard, Rippel Electron Microscope Facility, Dartmouth College; credit b: modification of work by June Kwak, University of Maryland; scale-bar data from Matt Russell)

Vascular Tissue

The xylem and phloem that make up the vascular tissue of the stem are arranged in distinct strands called vascular bundles, which run up and down the length of the stem. When the stem is viewed in cross section, the vascular bundles of dicot stems are arranged in a ring. In plants with stems that live for more than one year, the individual bundles grow together and produce the characteristic growth rings. In monocot stems, the vascular bundles are randomly scattered throughout the ground tissue (Figure 4.51).
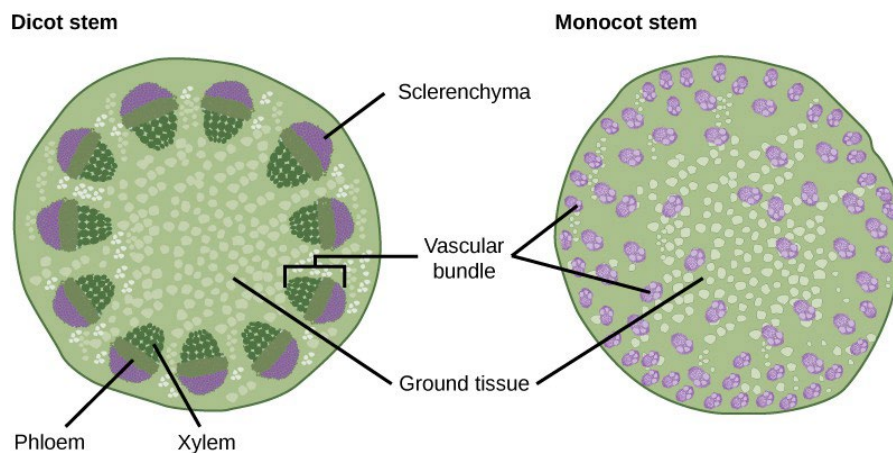
Figure 4.51 In (a) dicot stems, vascular bundles are arranged around the periphery of the ground tissue. The xylem tissue is located toward the interior of the vascular bundle, and phloem is located toward the exterior. Sclerenchyma fibers cap the vascular bundles. In (b) monocot stems, vascular bundles composed of xylem and phloem tissues are scattered throughout the ground tissue.

Xylem tissue has three types of cells: xylem parenchyma, tracheids, and vessel elements. The latter two types conduct water and are dead at maturity. **Tracheids** are xylem cells with thick secondary cell walls that are lignified. Water moves from one tracheid to another through regions on the side walls known as pits, where secondary walls are absent. **Vessel elements** are xylem cells with thinner walls; they are shorter than tracheids. Each vessel element is connected to the next by means of a perforation plate at the end walls of the element. Water moves through the perforation plates to travel up the plant.

151

Phloem tissue is composed of sieve-tube cells, companion cells, phloem parenchyma, and phloem fibers. A series of **sieve-tube cells** (also called sieve-tube elements) are arranged end to end to make up a long sieve tube, which transports organic substances such as sugars and amino acids. The sugars flow from one sieve-tube cell to the next through perforated sieve plates, which are found at the end junctions between two cells. Although still alive at maturity, the nucleus and other cell components of the sieve- tube cells have disintegrated. **Companion cells** are found alongside the sieve-tube cells, providing them with metabolic support. The companion cells contain more ribosomes and mitochondria than the sieve-tube cells, which lack some cellular organelles.

Ground Tissue

Ground tissue is mostly made up of parenchyma cells, but may also contain collenchyma and sclerenchyma cells that help support the stem. The ground tissue towards the interior of the vascular tissue in a stem or root is known as **pith**, while the layer of tissue between the vascular tissue and the epidermis is known as the **cortex**.

Growth in Stems

Growth in plants occurs as the stems and roots lengthen. Some plants, especially those that are woody, also increase in thickness during their life span. The increase in length of the shoot and the root is referred to as **primary growth**, and is the result of cell division in the shoot apical meristem. **Secondary growth** is characterized by an increase in thickness or girth of the plant, and is caused by cell division in the lateral meristem. Figure 4.52 shows the areas of primary and secondary growth in a plant.

Herbaceous plants mostly undergo primary growth, with hardly any secondary growth or increase in thickness. Secondary growth or "wood" is noticeable in woody plants; it occurs in some dicots, but occurs very rarely in monocots.

Primary growth   **Secondary growth**

Labels: Secondary phloem, Primary phloem, Cork, Bark, Pith, Cortex, Epidermis, Phloem, Sclerenchyma, Vascular cambium, Xylem, Primary xylem, Secondary xylem, Cork cambium
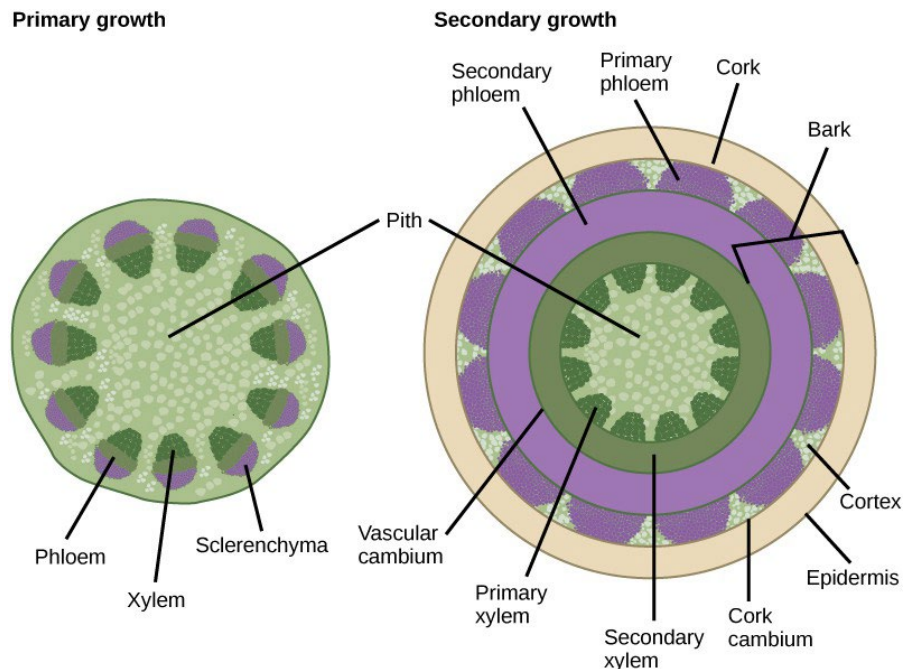
Figure 4.52 In woody plants, primary growth is followed by secondary growth, which allows the plant stem to increase in thickness or girth. Secondary vascular tissue is added as the plant grows, as well as a cork layer. The bark of a tree extends from the vascular cambium to the epidermis.

Some plant parts, such as stems and roots, continue to grow throughout a plant's life: a phenomenon called indeterminate growth. Other plant parts, such as leaves and flowers, exhibit determinate growth, which ceases when a plant part reaches a particular size.

Primary Growth

Most primary growth occurs at the apices, or tips, of stems and roots. Primary growth is a result of rapidly dividing cells in the apical meristems at the shoot tip and root tip. Subsequent cell elongation also contributes to primary growth. The growth of shoots and roots during primary growth enables plants to continuously seek water (roots) or sunlight (shoots).

The influence of the apical bud on overall plant growth is known as apical dominance, which diminishes the growth of axillary buds that form along the sides of branches and stems. Most coniferous trees exhibit strong apical dominance, thus producing

152

the typical conical Christmas tree shape. If the apical bud is removed, then the axillary buds will start forming lateral branches. Gardeners make use of this fact when they prune plants by cutting off the tops of branches, thus encouraging the axillary buds to grow out, giving the plant a bushy shape.

🔗 LINK TO LEARNING

Watch this BBC Nature video (http://openstax.org/l/motion_plants) showing how time-lapse photography captures plant growth at high speed.

Secondary Growth

The increase in stem thickness that results from secondary growth is due to the activity of the lateral meristems, which are lacking in herbaceous plants. Lateral meristems include the vascular cambium and, in woody plants, the cork cambium (see Figure 5.53). The vascular cambium is located just outside the primary xylem and to the interior of the primary phloem. The cells of the vascular cambium divide and form secondary xylem (tracheids and vessel elements) to the inside, and secondary phloem (sieve elements and companion cells) to the outside. The thickening of the stem that occurs in secondary growth is due to the formation of secondary phloem and secondary xylem by the vascular cambium, plus the action of cork cambium, which forms the tough outermost layer of the stem. The cells of the secondary xylem contain lignin, which provides hardiness and strength.

In woody plants, cork cambium is the outermost lateral meristem. It produces cork cells (bark) containing a waxy substance known as suberin that can repel water. The bark protects the plant against physical damage and helps reduce water loss. The cork cambium also produces a layer of cells known as phelloderm, which grows inward from the cambium. The cork cambium, cork cells, and phelloderm are collectively termed the **periderm**. The periderm substitutes for the epidermis in mature plants. In some plants, the periderm has many openings, known as **lenticels**, which allow the interior cells to exchange gases with the outside atmosphere (Figure 5.53). This supplies oxygen to the living and metabolically active cells of the cortex, xylem, and phloem.

Figure 4.53 Lenticels on the bark of this cherry tree enable the woody stem to exchange gases with the surrounding atmosphere. (credit: Roger Griffith)

Annual Rings

The activity of the vascular cambium gives rise to annual growth rings. During the spring growing season, cells of the secondary xylem have a large internal diameter and their primary cell walls are not extensively thickened. This is known as early wood, or spring wood. During the fall season, the secondary xylem develops thickened cell walls, forming late wood, or autumn wood, which is denser than early wood. This alternation of early and late wood is due largely to a seasonal decrease in the number of

153

vessel elements and a seasonal increase in the number of tracheids. It results in the formation of an annual ring, which can be seen as a circular ring in the cross section of the stem (Figure 5.54). An examination of the number of annual rings and their nature (such as their size and cell wall thickness) can reveal the age of the tree and the prevailing climatic conditions during each season.

Figure 4.54 The rate of wood growth increases in summer and decreases in winter, producing a characteristic ring for each year of growth. Seasonal changes in weather patterns can also affect the growth rate—note how the rings vary in thickness. (credit: Adrian Pingstone)

Stem Modifications

Some plant species have modified stems that are especially suited to a particular habitat and environment (Figure 5.55). A **rhizome** is a modified stem that grows horizontally underground and has nodes and internodes. Vertical shoots may arise from the buds on the rhizome of some plants, such as ginger and ferns. **Corms** are similar to rhizomes, except they are more rounded and fleshy (such as in gladiolus). Corms contain stored food that enables some plants to survive the winter. **Stolons** are stems that run almost parallel to the ground, or just below the surface, and can give rise to new plants at the nodes. **Runners** are a type of stolon that runs above the ground and produces new clone plants at nodes at varying intervals: strawberries are an example. **Tubers** are modified stems that may store starch, as seen in the potato (*Solanum* sp.). Tubers arise as swollen ends of stolons, and contain many adventitious or unusual buds (familiar to us as the "eyes" on potatoes). A **bulb**, which functions as an underground storage unit, is a modification of a stem that has the appearance of enlarged fleshy leaves emerging from the stem or surrounding the base of the stem, as seen in the iris.

Figure 4.55 Stem modifications enable plants to thrive in a variety of environments. Shown are (a) ginger (*Zingiber officinale*) rhizomes, (b) a carrion flower (*Amorphophallus titanum*) corm, (c) Rhodes grass (*Chloris gayana*) stolons, (d) strawberry (*Fragaria ananassa*) runners, (e) potato (*Solanum tuberosum*) tubers, and (f) red onion (*Allium*) bulbs. (credit a: modification of work by Maja Dumat; credit c: modification of work by Harry Rose; credit d: modification of work by Rebecca Siegel; credit e: modification of work by Scott Bauer, USDA ARS; credit f:

154

⊘ LINK TO LEARNING

Watch botanist Wendy Hodgson, of Desert Botanical Garden in Phoenix, Arizona, explain how agave plants were cultivated for food hundreds of years ago in the Arizona desert in this video: (http://openstax.org/l/ancient_crop) *Finding the Roots of an Ancient Crop.*

Some aerial modifications of stems are tendrils and thorns (Figure 4.56). **Tendrils** are slender, twining strands that enable a plant (like a vine or pumpkin) to seek support by climbing on other surfaces. **Thorns** are modified branches appearing as sharp outgrowths that protect the plant; common examples include roses, Osage orange, and devil's walking stick.

(a) (b)
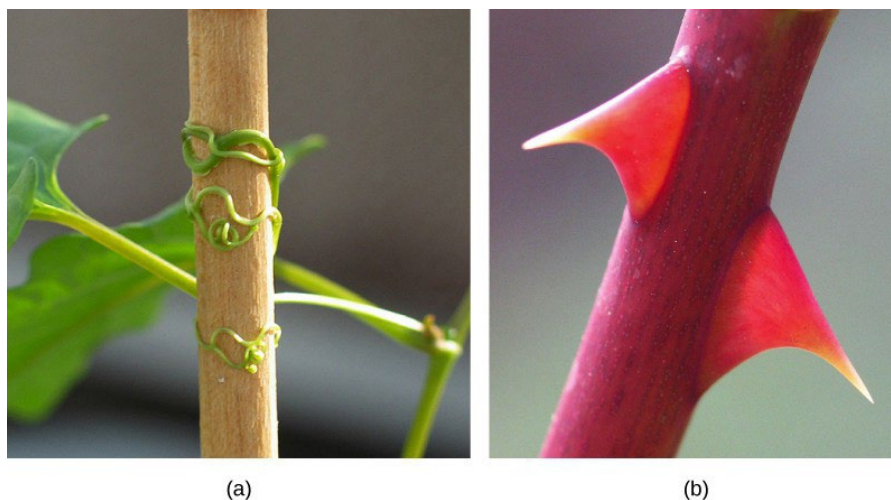
Figure 4.56 Found in southeastern United States, (a) buckwheat vine (*Brunnichia ovata*) is a weedy plant that climbs with the aid of tendrils. This one is shown climbing up a wooden stake. (b) Thorns are modified branches. (credit a: modification of work by Christopher Meloche, USDA ARS; credit b: modification of work by "macrophile"/Flickr)

Roots

By the end of this section, you will be able to do the following:

- Identify the two types of root systems
- Describe the three zones of the root tip and summarize the role of each zone in root growth
- Describe the structure of the root
- List and describe examples of modified roots

The roots of seed plants have three major functions: anchoring the plant to the soil, absorbing water and minerals and transporting them upwards, and storing the products of photosynthesis. Some roots are modified to absorb moisture and exchange gases. Most roots are underground. Some plants, however, also have **adventitious roots**, which emerge above the ground from the shoot.

Types of Root Systems

Root systems are mainly of two types (Figure 4.57). Dicots have a tap root system, while monocots have a fibrous root system. A **tap root system** has a main root that grows down vertically, and from which many smaller lateral roots arise. Dandelions are a good example; their tap roots usually break off when trying to pull these weeds, and they can regrow another shoot from the remaining root. A tap root system penetrates deep into the soil. In contrast, a **fibrous root system** is located closer to the soil surface, and forms a dense network of roots that also helps prevent soil erosion (lawn grasses are a good example, as are wheat, rice, and corn). Some plants have a combination of tap roots and fibrous roots. Plants that grow in dry areas often have deep root systems, whereas plants growing in areas with abundant water are likely to have shallower root systems.

155

(a) Taproot system   (b) Fibrous root system

Figure 4.57 (a) Tap root systems have a main root that grows down, while (b) fibrous root systems consist of many small roots. (credit b: modification of work by "Austen Squarepants"/Flickr)

Root Growth and Anatomy

Root growth begins with seed germination. When the plant embryo emerges from the seed, the radicle of the embryo forms the root system. The tip of the root is protected by the **root cap**, a structure exclusive to roots and unlike any other plant structure. The root cap is continuously replaced because it gets damaged easily as the root pushes through soil. The root tip can be divided into three zones: a zone of cell division, a zone of elongation, and a zone of maturation and differentiation (Figure 4.58). The zone of cell division is closest to the root tip; it is made up of the actively dividing cells of the root meristem. The zone of elongation is where the newly formed cells increase in length, thereby lengthening the root. Beginning at the first root hair is the zone of cell maturation where the root cells begin to differentiate into special cell types. All three zones are in the first centimeter or so of the root tip.
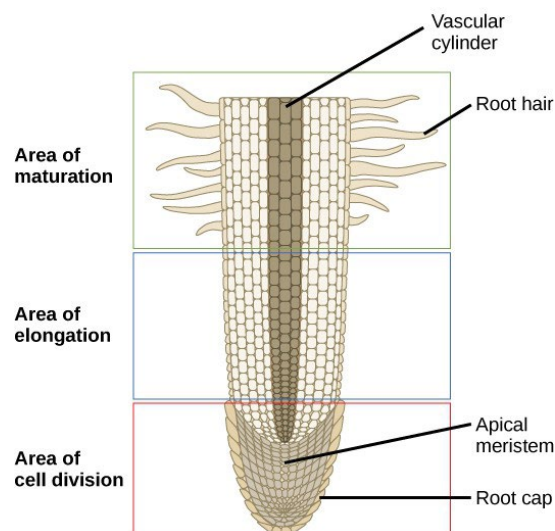
Figure 4.58 A longitudinal view of the root reveals the zones of cell division, elongation, and maturation. Cell division occurs in the apical meristem.

The root has an outer layer of cells called the epidermis, which surrounds areas of ground tissue and vascular tissue. The

156

epidermis provides protection and helps in absorption. **Root hairs**, which are extensions of root epidermal cells, increase the surface area of the root, greatly contributing to the absorption of water and minerals.

Inside the root, the ground tissue forms two regions: the cortex and the pith (Figure 4.59). Compared to stems, roots have lots of cortex and little pith. Both regions include cells that store photosynthetic products. The cortex is between the epidermis and the vascular tissue, whereas the pith lies between the vascular tissue and the center of the root.
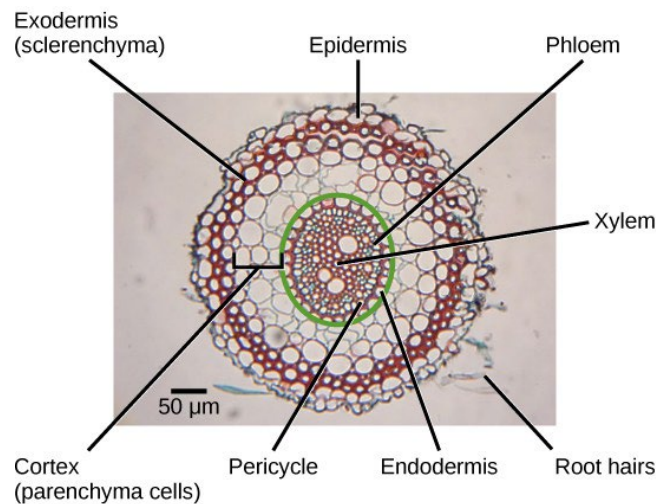
Figure 4.59 Staining reveals different cell types in this light micrograph of a wheat (*Triticum*) root cross section. Sclerenchyma cells of the exodermis and xylem cells stain red, and phloem cells stain blue. Other cell types stain black. The stele, or vascular tissue, is the area inside endodermis (indicated by a green ring). Root hairs are visible outside the epidermis. (credit: scale-bar data from Matt Russell)

The vascular tissue in the root is arranged in the inner portion of the root, which is called the **stele** (Figure 4.60). A layer of cells known as the **endodermis** separates the stele from the ground tissue in the outer portion of the root. The endodermis is exclusive to roots, and serves as a checkpoint for materials entering the root's vascular system. A waxy substance called suberin is present on the walls of the endodermal cells. This waxy region, known as the **Casparian strip**, forces water and solutes to cross the plasma membranes of endodermal cells instead of slipping between the cells. This ensures that only materials required by the root pass through the endodermis, while toxic substances and pathogens are generally excluded. The outermost cell layer of the root's vascular tissue is the **pericycle**, an area that can give rise to lateral roots. In dicot roots, the xylem and phloem of the stele are arranged alternately in an X shape, whereas in monocot roots, the vascular tissue is arranged in a ring around the pith.
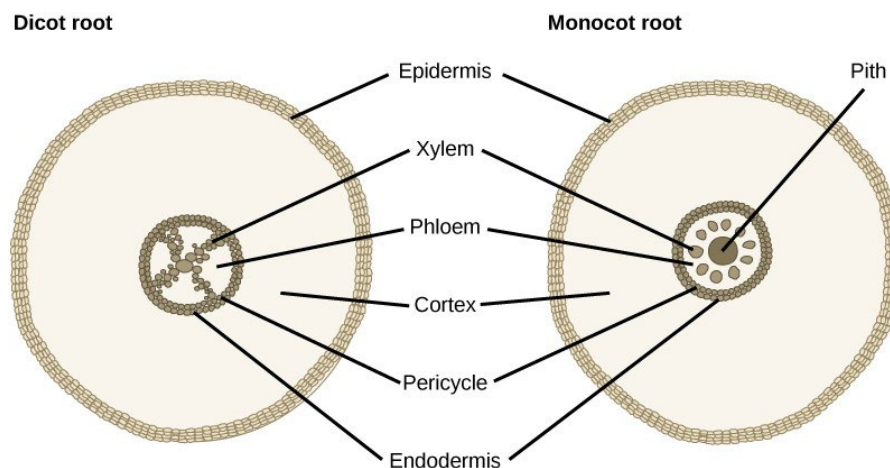
Figure 4.60 In (left) typical dicots, the vascular tissue forms an X shape in the center of the root. In (right) typical monocots, the phloem cells and the larger xylem cells form a characteristic ring around the central pith.

Root Modifications

Root structures may be modified for specific purposes. For example, some roots are bulbous and store starch. Aerial roots and

157

prop roots are two forms of aboveground roots that provide additional support to anchor the plant. Tap roots, such as carrots, turnips, and beets, are examples of roots that are modified for food storage (Figure 4.61).

Figure 4.61 Many vegetables are modified roots.

Epiphytic roots enable a plant to grow on another plant. For example, the epiphytic roots of orchids develop a spongy tissue to absorb moisture. The banyan tree (*Ficus* sp.) begins as an epiphyte, germinating in the branches of a host tree; aerial roots develop from the branches and eventually reach the ground, providing additional support (Figure 4.62). In screwpine (*Pandanus* sp.), a palm-like tree that grows in sandy tropical soils, aboveground prop roots develop from the nodes to provide additional support.

(a)   (b)

Figure 4.62 The (a) banyan tree, also known as the strangler fig, begins life as an epiphyte in a host tree. Aerial roots extend to the ground and support the growing plant, which eventually strangles the host tree. The (b) screwpine develops aboveground roots that help support the plant in sandy soils. (credit a: modification of work by "psyberartist"/Flickr; credit b: modification of work by David Eikhoff)

Leaves
By the end of this section, you will be able to do the following:
- Identify the parts of a typical leaf
- Describe the internal structure and function of a leaf
- Compare and contrast simple leaves and compound leaves
- List and describe examples of modified leaves

Leaves are the main sites for photosynthesis: the process by which plants synthesize food. Most leaves are usually green, due to the presence of chlorophyll in the leaf cells. However, some leaves may have different colors, caused by other plant pigments that mask the green chlorophyll.

The thickness, shape, and size of leaves are adapted to the environment. Each variation helps a plant species maximize its chances of survival in a particular habitat. Usually, the leaves of plants growing in tropical rainforests have larger surface areas than those of plants growing in deserts or very cold conditions, which are likely to have a smaller surface area to minimize water loss.

Structure of a Typical Leaf

Each leaf typically has a leaf blade called the **lamina**, which is also the widest part of the leaf. Some leaves are attached to the plant stem by a **petiole**. Leaves that do not have a petiole and are directly attached to the plant stem are called **sessile** leaves. Small green appendages usually found at the base of the petiole are known as **stipules**. Most leaves have a midrib, which travels the length of the leaf and branches to each side to produce veins of vascular tissue. The edge of the leaf is called the margin.
Figure 4.63 shows the structure of a typical eudicot leaf.
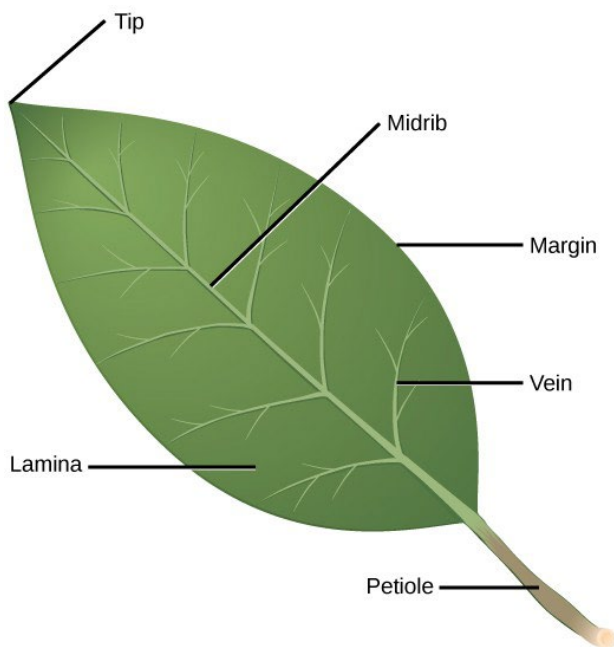
Figure 4.63 Deceptively simple in appearance, a leaf is a highly efficient structure.

Within each leaf, the vascular tissue forms veins. The arrangement of veins in a leaf is called the **venation** pattern. Monocots and dicots differ in their patterns of venation (Figure 4.64). Monocots have parallel venation; the veins run in straight lines across the length of the leaf without converging at a point. In dicots, however, the veins of the leaf have a net-like appearance, forming a pattern known as reticulate venation. One extant plant, the *Ginkgo biloba*, has dichotomous venation where the veins fork.

(a)   (b)   (c)
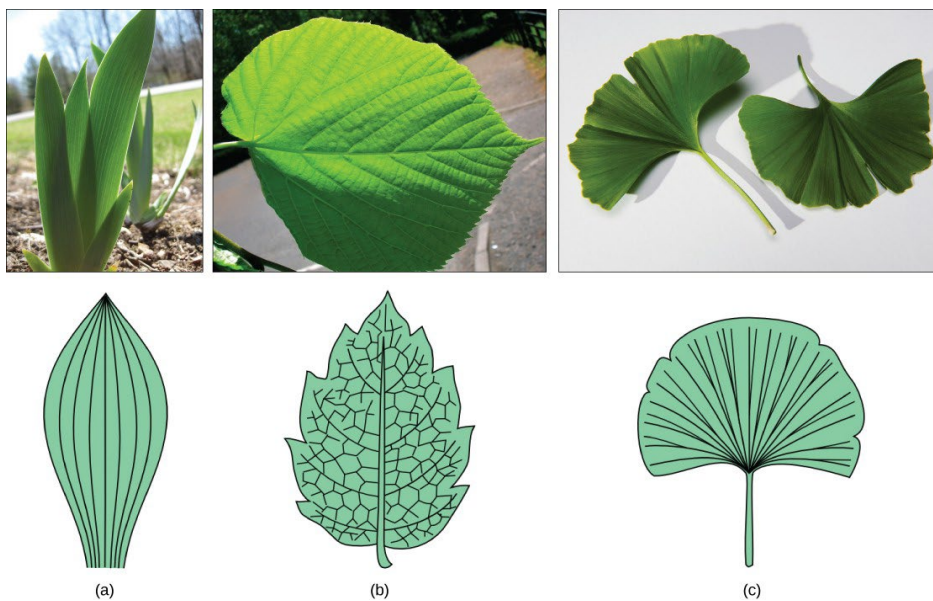
Figure 4.64 (a) Tulip (*Tulipa*), a monocot, has leaves with parallel venation. The netlike venation in this (b) linden (*Tilia cordata*) leaf distinguishes it as a dicot. The (c) *Ginkgo biloba* tree has dichotomous venation. (credit a photo: modification of work by

Leaf Arrangement

The arrangement of leaves on a stem is known as **phyllotaxy**. The number and placement of a plant's leaves will vary depending on the species, with each species exhibiting a characteristic leaf arrangement. Leaves are classified as either alternate, spiral, or opposite. Plants that have only one leaf per node have leaves that are said to be either alternate—meaning the leaves alternate on each side of the stem in a flat plane—or spiral, meaning the leaves are arrayed in a spiral along the stem. In an opposite leaf arrangement, two leaves arise at the same point, with the leaves connecting opposite each other along the branch. If there are three or more leaves connected at a node, the leaf arrangement is classified as **whorled**.

Leaf Form

Leaves may be simple or compound (Figure 4.65). In a **simple leaf**, the blade is either completely undivided—as in the banana leaf—or it has lobes, but the separation does not reach the midrib, as in the maple leaf. In a **compound leaf**, the leaf blade is completely divided, forming leaflets, as in the locust tree. Each leaflet may have its own stalk, but is attached to the rachis. A **palmately compound leaf** resembles the palm of a hand, with leaflets radiating outwards from one point. Examples include the leaves of poison ivy, the buckeye tree, or the familiar houseplant *Schefflera* sp. (common name "umbrella plant"). **Pinnately compound leaves** take their name from their feather-like appearance; the leaflets are arranged along the midrib, as in rose leaves (*Rosa* sp.), or the leaves of hickory, pecan, ash, or walnut trees.

(a) Simple   (b) Palmately compound

(c) Pinnately compound   (d) Doubly compound
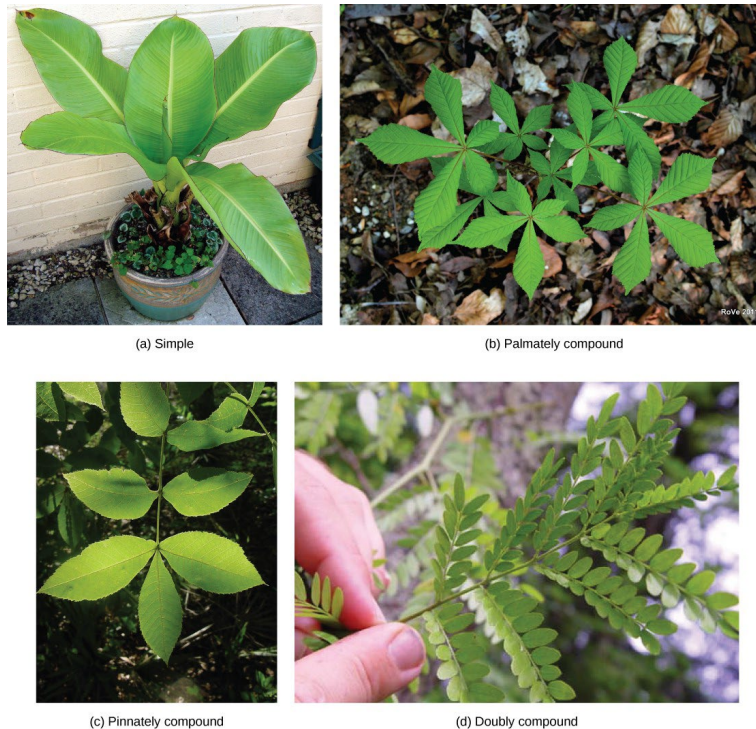
Figure 4.65 Leaves may be simple or compound. In simple leaves, the lamina is continuous. The (a) banana plant (*Musa* sp.) has simple leaves. In compound leaves, the lamina is separated into leaflets. Compound leaves may be palmate or pinnate. In (b) palmately compound leaves, such as those of the horse chestnut (*Aesculus hippocastanum*), the leaflets branch from the petiole. In (c) pinnately compound leaves, the leaflets branch from the midrib, as on a scrub hickory (*Carya floridana*). The (d) honey locust has double compound leaves, in which leaflets branch from the veins. (credit a: modification of work by "BazzaDaRambler"/Flickr; credit b: modification of work by Roberto Verzo; credit c: modification of work by Eric Dion; credit d: modification of work by Valerie Lykes)

Leaf Structure and Function

The outermost layer of the leaf is the epidermis; it is present on both sides of the leaf and is called the upper and lower epidermis, respectively. Botanists call the upper side the adaxial surface (or adaxis) and the lower side the abaxial surface (or abaxis). The epidermis helps in the regulation of gas exchange. It contains stomata (Figure 4.66): openings through which the

160

exchange of gases takes place. Two guard cells surround each stoma, regulating its opening and closing.

(a)   (b)   (c)
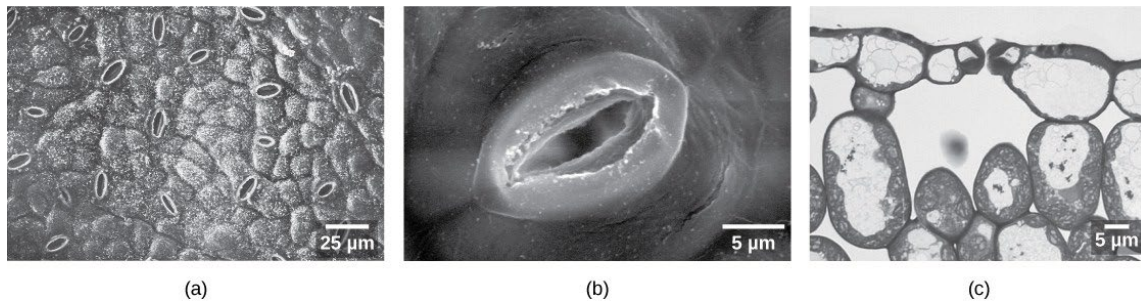
Figure 4.66 Visualized at 500x with a scanning electron microscope, several stomata are clearly visible on (a) the surface of this sumac (*Rhus glabra*) leaf. At 5,000x magnification, the guard cells of (b) a single stoma from lyre-leaved sand cress (*Arabidopsis lyrata)* have the appearance of lips that surround the opening. In this (c) light micrograph cross-section of an *A. lyrata* leaf, the guard cell pair is visible along with the large, sub-stomatal air space in the leaf. (credit: modification of work by Robert R. Wise; part c scale-bar data from Matt Russell)

The epidermis is usually one cell layer thick; however, in plants that grow in very hot or very cold conditions, the epidermis may be several layers thick to protect against excessive water loss from transpiration. A waxy layer known as the **cuticle** covers the leaves of all plant species. The cuticle reduces the rate of water loss from the leaf surface. Other leaves may have small hairs (trichomes) on the leaf surface. Trichomes help to deter herbivory by restricting insect movements, or by storing toxic or bad- tasting compounds; they can also reduce the rate of transpiration by blocking air flow across the leaf surface (Figure 4.67).

(a)   (b)   (c)
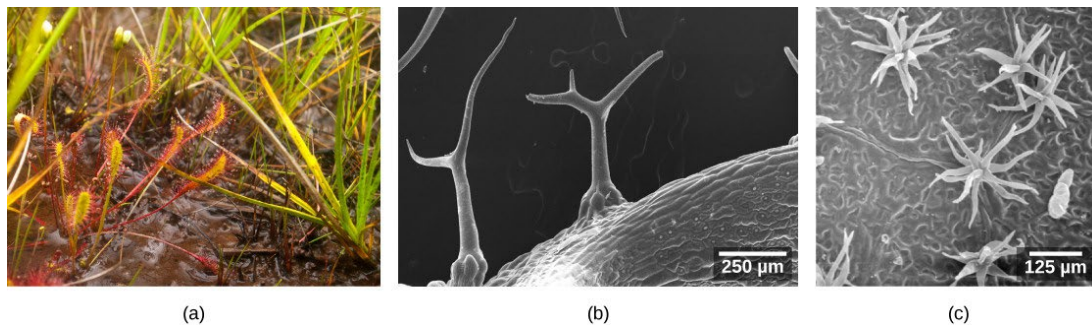
Figure 4.67 Trichomes give leaves a fuzzy appearance as in this (a) sundew (*Drosera* sp.). Leaf trichomes include (b) branched trichomes on the leaf of *Arabidopsis lyrata* and (c) multibranched trichomes on a mature *Quercus marilandica* leaf. (credit a: John Freeland; credit b, c: modification of work by Robert R. Wise; scale-bar data from Matt Russell)

Below the epidermis of dicot leaves are layers of cells known as the mesophyll, or "middle leaf." The mesophyll of most leaves typically contains two arrangements of parenchyma cells: the palisade parenchyma and spongy parenchyma (Figure 4.68). The palisade parenchyma (also called the palisade mesophyll) has column-shaped, tightly packed cells, and may be present in one, two, or three layers. Below the palisade parenchyma are loosely arranged cells of an irregular shape. These are the cells of the spongy parenchyma (or spongy mesophyll). The air space found between the spongy parenchyma cells allows gaseous exchange between the leaf and the outside atmosphere through the stomata. In aquatic plants, the intercellular spaces in the spongy parenchyma help the leaf float. Both layers of the mesophyll contain many chloroplasts. Guard cells are the only epidermal cells to contain chloroplasts.
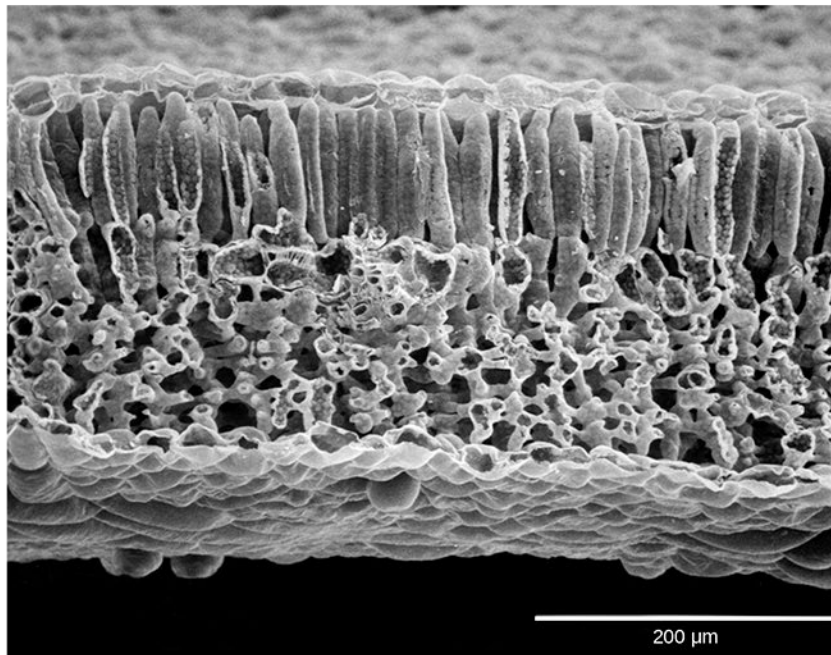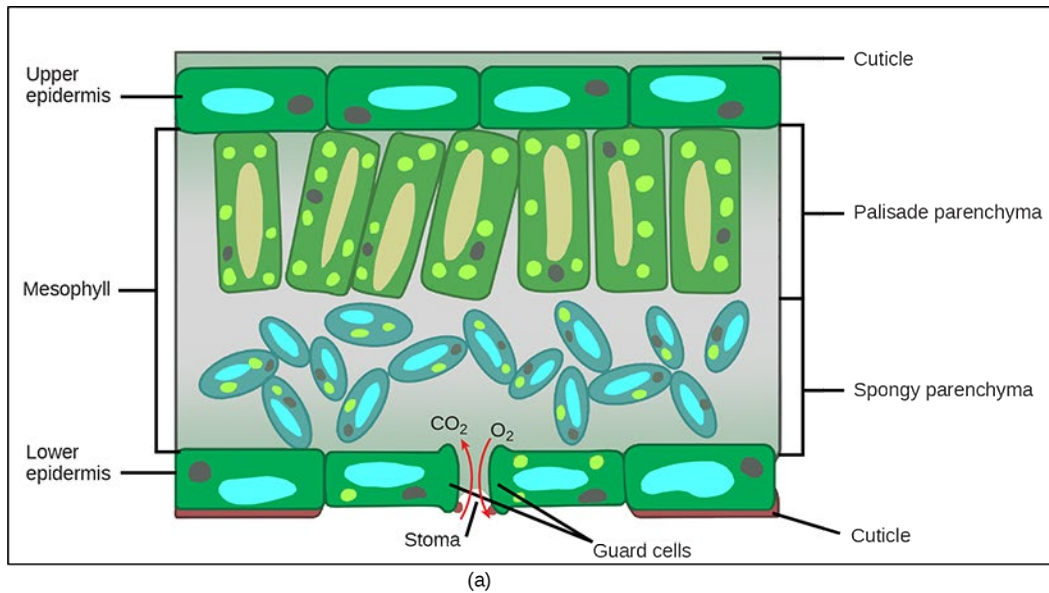
Figure 4.68 In the (a) leaf drawing, the central mesophyll is sandwiched between an upper and lower epidermis. The mesophyll has two layers: an upper palisade layer comprised of tightly packed, columnar cells, and a lower spongy layer, comprised of loosely packed, irregularly shaped cells. Stomata on the leaf underside allow gas exchange. A waxy cuticle covers all aerial surfaces of land plants to minimize water loss. These leaf layers are clearly visible in the (b) scanning electron micrograph. The numerous small bumps in the palisade parenchyma cells are chloroplasts. Chloroplasts are also present in the spongy parenchyma, but are not as obvious. The bumps protruding from the lower surface of the leave are glandular trichomes, which differ in structure from the stalked trichomes in Figure 30.25. (credit b: modification of work by Robert R. Wise)

Like the stem, the leaf contains vascular bundles composed of xylem and phloem (Figure 4.69). The xylem consists of tracheids and vessels, which transport water and minerals to the leaves. The phloem transports the photosynthetic products from the leaf to the other parts of the plant. A single vascular bundle, no matter how large or small, always contains both xylem and phloem tissues.

162
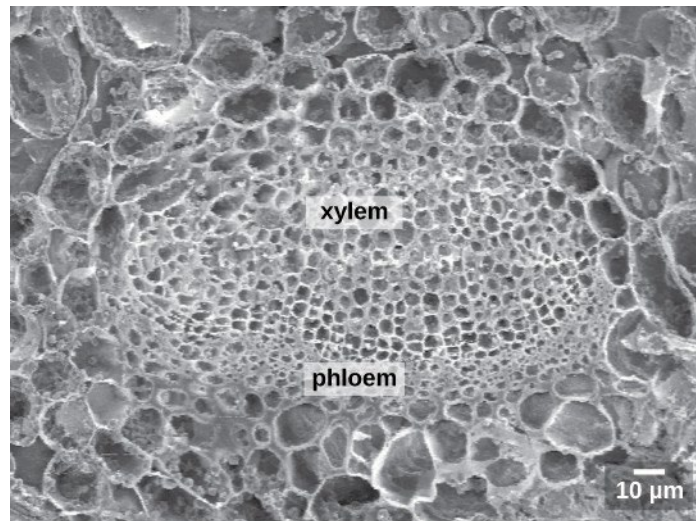
Figure 4.69 This scanning electron micrograph shows xylem and phloem in the leaf vascular bundle from the lyre-leaved sand cress (*Arabidopsis lyrata*). (credit: modification of work by Robert R. Wise; scale-bar data from Matt Russell)

Leaf Adaptations

Coniferous plant species that thrive in cold environments, like spruce, fir, and pine, have leaves that are reduced in size and needle-like in appearance. These needle-like leaves have sunken stomata and a smaller surface area: two attributes that aid in reducing water loss. In hot climates, plants such as cacti have leaves that are reduced to spines, which in combination with their succulent stems, help to conserve water. Many aquatic plants have leaves with wide lamina that can float on the surface of the water, and a thick waxy cuticle on the leaf surface that repels water.

🔗 LINK TO LEARNING

Watch "The Pale Pitcher Plant" episode of the video (http://openstax.org/l/plants_cool_too) series *Plants Are Cool, Too,* a Botanical Society of America video about a carnivorous plant species found in Louisiana.

🌱 EVOLUTION CONNECTION

Plant Adaptations in Resource-Deficient Environments

Roots, stems, and leaves are structured to ensure that a plant can obtain the required sunlight, water, soil nutrients, and oxygen resources. Some remarkable adaptations have evolved to enable plant species to thrive in less than ideal habitats, where one or more of these resources is in short supply.

In tropical rainforests, light is often scarce, since many trees and plants grow close together and block much of the sunlight from reaching the forest floor. Many tropical plant species have exceptionally broad leaves to maximize the capture of sunlight. Other species are epiphytes: plants that grow on other plants that serve as a physical support. Such plants are able to grow high up in the canopy atop the branches of other trees, where sunlight is more plentiful. Epiphytes live on rain and minerals collected in the branches and leaves of the supporting plant. Bromeliads (members of the pineapple family), ferns, and orchids are examples of tropical epiphytes (Figure 4.70). Many epiphytes have specialized tissues that enable them to efficiently capture and store water.

163

Figure 4.70 One of the most well known bromeliads is Spanish moss (*Tillandsia usneoides*), seen here in an oak tree. (credit: Kristine Paulus)

Some plants have special adaptations that help them to survive in nutrient-poor environments. Carnivorous plants, such as the Venus flytrap and the pitcher plant (Figure 4.71), grow in bogs where the soil is low in nitrogen. In these plants, leaves are modified to capture insects. The insect-capturing leaves may have evolved to provide these plants with a supplementary source of much-needed nitrogen.

(a)   (b)
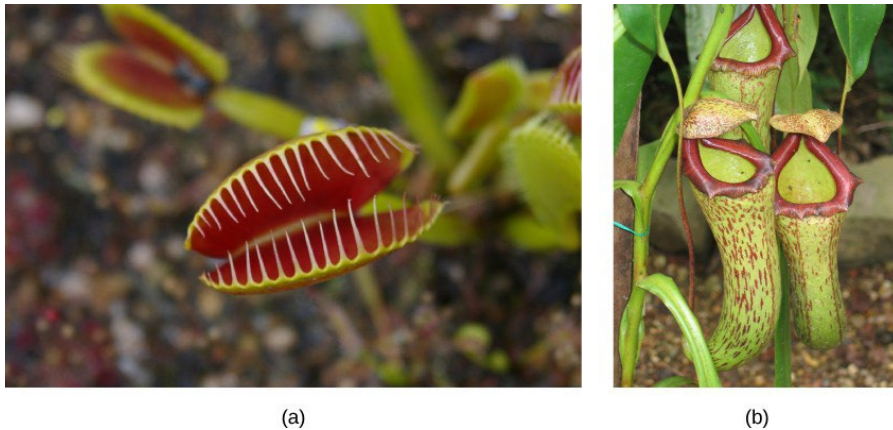
Figure 4.71 The (a) Venus flytrap has modified leaves that can capture insects. When an unlucky insect touches the trigger hairs inside the leaf, the trap suddenly closes. The opening of the (b) pitcher plant is lined with a slippery wax. Insects crawling on the lip slip and fall into a pool of water in the bottom of the pitcher, where they are digested by bacteria. The plant then absorbs the smaller molecules. (credit a: modification of work by Peter Shanks; credit b: modification of work by Tim Mansfield)

Many swamp plants have adaptations that enable them to thrive in wet areas, where their roots grow submerged underwater. In these aquatic areas, the soil is unstable and little oxygen is available to reach the roots. Trees such as mangroves (*Rhizophora* sp.) growing in coastal waters produce aboveground roots that help support the tree (Figure 4.72). Some species of mangroves, as well as cypress trees, have pneumatophores: upward-growing roots containing pores and pockets of tissue specialized for gas exchange. Wild rice is an aquatic plant with large air spaces in the root cortex. The air-filled tissue—called aerenchyma—provides a path for oxygen to diffuse down to the root tips, which are embedded in oxygen-poor bottom sediments.

164

Figure 4.72 The branches of (a) mangrove trees develop aerial roots, which descend to the ground and help to anchor the trees. (b) Cypress trees and some mangrove species have upward-growing roots called pneumatophores that are involved in gas exchange. Aquatic plants such as (c) wild rice have large spaces in the root cortex called aerenchyma, visualized here using scanning electron microscopy. (credit a: modification of work by Roberto Verzo; credit b: modification of work by Duane Burdick; credit c: modification of work by Robert R. Wise)

🔗 LINK TO LEARNING

atch *Venus Flytraps: Jaws of Death*, an extraordinary BBC close-up of the Venus flytrap in action.

165

Summary

Prokaryotic Diversity

Prokaryotes existed for billions of years before plants and animals appeared. Microbial mats are thought to represent the earliest forms of life on Earth, and there is fossil evidence, called stromatolites, of their presence about 3.5 billion years ago. During the first 2 billion years, the atmosphere was anoxic and only anaerobic organisms were able to live. Cyanobacteria began the oxygenation of the atmosphere.

The increase in oxygen concentration allowed the evolution of other life forms.

Prokaryotes (domains Archaea and Bacteria) are single-celled organisms lacking a nucleus. They have a single piece of circular DNA in the nucleoid area of the cell. Most prokaryotes have cell wall outside the plasma membrane. Bacteria and Archaea differ in the compositions of their cell membranes and the characteristics of their cell walls.

Bacterial cell walls contain peptidoglycan. Archaean cell walls do not have peptidoglycan. Bacteria can be divided into two major groups: Gram-positive and Gram-negative. Gram-positive organisms have a thick cell wall. Gram-negative organisms have a thin cell wall and an outer membrane. Prokaryotes use diverse sources of energy to assemble macromolecules from smaller molecules. Phototrophs obtain their energy from sunlight, whereas chemotrophs obtain it from chemical compounds.

Infectious diseases caused by bacteria remain among the leading causes of death worldwide. The excessive use of antibiotics to control bacterial infections has resulted in resistant forms of bacteria being selected. Foodborne diseases result from the consumption of contaminated food, pathogenic bacteria, viruses, or parasites that contaminate food. Prokaryotes are used in human food products. Microbial bioremediation is the use of microbial metabolism to remove pollutants. The human body contains a huge community of prokaryotes, many of which provide beneficial services such as the development and maintenance of the immune system, nutrition, and protection from pathogens.

Protists

Protists are extremely diverse in terms of biological and ecological characteristics due in large part to the fact that they are an artificial assemblage of phylogenetically unrelated groups. Protists display highly varied cell structures, several types of reproductive strategies, virtually every possible type of nutrition, and varied habitats. Most single-celled protists are motile, but these organisms use diverse structures for transportation.

The process of classifying protists into meaningful groups is ongoing, but genetic data in the past 20 years have clarified many relationships that were previously unclear or mistaken. The majority view at present is to order all eukaryotes into six supergroups. The goal of this classification scheme is to create clusters of species that all are derived from a common ancestor.

Fungi

Fungi are eukaryotic organisms that appeared on land over 450 million years ago. They are heterotrophs and contain neither photosynthetic pigments such as chlorophylls nor organelles such as chloroplasts.

Because they feed on decaying and dead matter, they are saprobes. Fungi are important decomposers and release essential elements into the environment. External enzymes digest nutrients that are absorbed by the body of the fungus called a thallus. A thick cell wall made of chitin surrounds the cell. Fungi can be unicellular as yeasts or develop a network of filaments called a mycelium, often described as mold. Most species multiply by asexual and sexual reproductive cycles, and display an alternation of generations.

The divisions of fungi are the Chytridiomycota, Zygomycota, Ascomycota, Basidiomycota, Glomeromycota, and the Deuteromycota, a polyphyletic group.

Fungi establish parasitic relationships with plants and animals. Fungal diseases can decimate crops and spoil food during storage. Compounds produced by fungi can be toxic to humans and other animals.

Mycoses are infections caused by fungi. Superficial mycoses affect the skin, whereas systemic mycoses spread through the body. Fungal infections are difficult to cure.

Fungi have colonized all environments on Earth but are most often found in cool, dark, moist places with a supply of decaying material. Fungi are important decomposers because they are saprobes. Many successful mutualistic relationships involve a fungus and another organism. They establish complex mycorrhizal associations with the roots of plants. Lichens are a symbiotic relationship between a fungus and a photosynthetic organism, usually an alga or cyanobacterium.

Fungi are important to everyday human life. Fungi are important decomposers in most ecosystems. Mycorrhizal fungi are essential for the growth of most plants. Fungi, as food, play a role in human nutrition in the form of mushrooms and as agents of fermentation in the production of bread, cheeses, alcoholic beverages, and numerous other food preparations. Secondary metabolites of fungi are used in medicine as antibiotics and anticoagulants. Fungi are used in research as model organisms for the study of eukaryotic genetics and metabolism.

The Plant Kingdom

Land plants evolved traits that made it possible to colonize land and survive out of water. Adaptations to life on land include vascular tissues, roots, leaves, waxy cuticles, and a tough outer layer that protects the spores. Land plants include nonvascular plants and vascular plants. Vascular plants, which include seedless plants and plants with seeds, have apical meristems, and embryos with nutritional stores. All land plants share the following characteristics: alternation of generations, with the haploid plant called a gametophyte and the diploid plant called a sporophyte; formation of haploid spores in a sporangium; and formation of gametes in a gametangium.

Seedless Plants

Seedless nonvascular plants are small. The dominant stage of the life cycle is the gametophyte. Without a vascular system and roots, they absorb water and nutrients through all of their exposed surfaces.

There are three main groups: the liverworts, the hornworts, and the mosses. They are collectively known as bryophytes.

Vascular systems consist of xylem tissue, which transports water and minerals, and phloem tissue, which transports sugars and proteins. With the vascular system, there appeared leaves—large photosynthetic organs—and roots to absorb water from the ground. The seedless vascular plants include club mosses, which are the most primitive; whisk ferns, which lost leaves and roots by reductive evolution; horsetails, and ferns.

Seed Plants: Gymnosperms

Gymnosperms are heterosporous seed plants that produce naked seeds. They appeared in the Carboniferous period (359–299 million years ago) and were the dominant plant life during the Mesozoic era (251–65.5 million years ago). Modern-day gymnosperms belong to four divisions. The division Coniferophyta—the conifers—are the predominant woody plants at high altitudes and latitudes. Cycads resemble palm trees and grow in tropical climates. Gingko biloba is the only species of the division Gingkophyta. The last division, the Gnetophytes, is a diverse group of species that produce vessel elements in their wood.

Seed Plants: Angiosperms

Angiosperms are the dominant form of plant life in most terrestrial ecosystems, comprising about 90 percent of all plant species. Most crop and ornamental plants are angiosperms. Their success results, in part, from two innovative structures: the flower and the fruit. Flowers are derived evolutionarily from modified leaves. The main parts of a flower are the sepals and petals, which protect the reproductive parts: the stamens and the carpels. The stamens produce the male gametes, which are pollen grains. The carpels contain the female gametes, which are the eggs inside ovaries. The walls of the ovary thicken after fertilization, ripening into fruit that can facilitate seed dispersal.

Angiosperms' life cycles are dominated by the sporophyte stage. Double fertilization is an event unique to angiosperms. The flowering plants are divided into two main groups—the monocots and eudicots—according to the number of cotyledons in the seedlings. Basal angiosperms belong to a lineage older than monocots and eudicots.

1.   Which group are vascular plants?

a.   liverworts

b.   mosses

c.   hornworts

d.   ferns

2.   Which of the following traits characterizes gymnosperms?

a.   The plants carry exposed seeds on modified leaves.

b.   Reproductive structures are located in a flower.

c.   After fertilization, the ovary thickens and forms a fruit.

3.   What adaptation do seed plants have in addition to the seed that is not found in seedless plants?

a.   gametophytes

b.   vascular tissue

c.   pollen

d.   chlorophyll

4.   Pollen grains develop in which structure?

a.   the anther

b.   the stigma

c.   the filament

d.   the carpel

5.   Corn develops from a seedling with a single cotyledon, displays parallel veins on its leaves, and produces monosulcate pollen. It is most likely:

a.   a gymnosperm

b.   a monocot

c.   a eudicot

d.   a basal angiosperm

6.   Which of the following consist of prokaryotic cells?

a.   bacteria and fungi

b.   archaea and fungi

c.   protists and animals

d.   bacteria and archaea

7.   What term describes the close association of a fungus with the root of a tree?

a.   a rhizoid

b.   a lichen

c.   a mycorrhiza

d.   an endophyte

OpenStax College, *Concepts of Biology*. OpenStax College. 25 April 2013.
<http://openstaxcollege.org/textbooks/concepts-of-biology/get>.

https://openstax.org/details/books/biology-2e

Chapter 5

Animals

Introduction

While we can easily identify dogs, lizards, fish, spiders, and worms as animals, other animals, such as corals and sponges, might be easily mistaken as plants or some other form of life. Yet scientists have recognized a set of common characteristics shared by all animals, including sponges, jellyfish, sea urchins, and humans.

The kingdom Animalia is a group of multicellular Eukarya. Animal evolution began in the ocean over 600 million years ago, with tiny creatures that probably do not resemble any living organism today. Since then, animals have evolved into a highly diverse kingdom. Although over one million currently living species of animals have been identified, scientists are continually discovering more species. The number of described living animal species is estimated to be about 1.4 million,[1] and there may be as many as 6.8 million.

Understanding and classifying the variety of living species helps us to better understand how to conserve and benefit from this diversity. The animal classification system characterizes animals based on their anatomy, features of embryological development, and genetic makeup. Scientists are faced with the task

---

1. "Number of Living Species in Australia and the World," A.D. Chapman, Australia Biodiversity Information Services, last modified August 26, 2010, http://www.environment.gov.au/biodiversity/abrs/publications/other/species-numbers/2009/03-exec-summary.html.

of classifying animals within a system of taxonomy that reflects their evolutionary history. Additionally, they must identify traits that are common to all animals as well as traits that can be used to distinguish among related groups of animals. However, animals vary in the complexity of their organization and exhibit a huge diversity of body forms, so the classification scheme is constantly changing as new information about species is learned.

| Features of the Animal Kingdom

By the end of this section, you will be able to:

- List the features that distinguish the animal kingdom from other kingdoms
- Explain the processes of animal reproduction and embryonic development
- Describe the hierarchy of basic animal classification
- Compare and contrast the embryonic development of protostomes and deuterostomes

Even though members of the animal kingdom are incredibly diverse, animals share common features that distinguish them from organisms in other kingdoms. All animals are eukaryotic, multicellular organisms, and almost all animals have specialized tissues. Most animals are motile, at least during certain life stages. Animals require a source of food to grow and develop. All animals are heterotrophic, ingesting living or dead organic matter. This form of obtaining energy distinguishes them from autotrophic organisms, such as most plants, which make their own nutrients through photosynthesis and from fungi that digest their food externally. Animals may be carnivores, herbivores, omnivores, or parasites (**Figure** 5.1). Most animals reproduce sexually: The offspring pass through a series of developmental stages that establish a determined body plan, unlike plants, for example, in which the exact shape of the body is indeterminate. The **body plan** refers to the shape of an animal.

(a)   (b)

Figure 5.1 All animals that derive energy from food are heterotrophs. The (a) black bear is an omnivore, eating both plants and animals. The (b) heartworm *Dirofilaria immitis* is a parasite that derives energy from its hosts. It spends its larval stage in mosquitos and its adult stage infesting the hearts of dogs and other mammals, as shown here. (credit a: modification of work by USDA Forest Service; credit b: modification of work by Clyde Robinson)

Complex Tissue Structure

A hallmark trait of animals is specialized structures that are differentiated to perform unique functions. As multicellular organisms, most animals develop specialized cells that group together into tissues with specialized functions. A tissue is a collection of similar cells that had a common embryonic origin. There are four main types of animal tissues: nervous, muscle, connective, and epithelial. Nervous tissue contains neurons, or nerve cells, which transmit nerve impulses. Muscle tissue contracts to cause all types of body movement from locomotion of the organism to movements within the body itself. Animals also have specialized connective tissues that provide many functions, including transport and structural support. Examples of connective tissues include blood and bone. Connective tissue is comprised of cells separated by extracellular material made of organic and inorganic materials, such as the protein and

172

mineral deposits of bone. Epithelial tissue covers the internal and external surfaces of organs inside the animal body and the external surface of the body of the organism.

CONCEPT in ACTION

View this **video (http://openstaxcollege.org/l/saving_life2)** to watch a presentation by biologist E.O. Wilson on the importance of animal diversity.

Animal Reproduction and Development

Most animals have diploid body (somatic) cells and a small number of haploid reproductive (gamete) cells produced through meiosis. Some exceptions exist: For example, in bees, wasps, and ants, the male is haploid because it develops from an unfertilized egg. Most animals undergo sexual reproduction, while many also have mechanisms of asexual reproduction.

Sexual Reproduction and Embryonic Development

Almost all animal species are capable of reproducing sexually; for many, this is the only mode of reproduction possible. This distinguishes animals from fungi, protists, and bacteria, where asexual reproduction is common or exclusive. During sexual reproduction, the male and female gametes of a species combine in a process called fertilization. Typically, the small, motile male sperm travels to the much larger, sessile female egg. Sperm form is diverse and includes cells with flagella or amoeboid cells to facilitate motility. Fertilization and fusion of the gamete nuclei produce a zygote. Fertilization may be internal, especially in land animals, or external, as is common in many aquatic species.

After fertilization, a developmental sequence ensues as cells divide and differentiate. Many of the events in development are shared in groups of related animal species, and these events are one of the main ways scientists classify high-level groups of animals. During development, animal cells specialize and form tissues, determining their future morphology and physiology. In many animals, such as mammals, the young resemble the adult. Other animals, such as some insects and amphibians, undergo complete metamorphosis in which individuals enter one or more larval stages. For these animals, the young and the adult have different diets and sometimes habitats. In other species, a process of incomplete metamorphosis occurs in which the young somewhat resemble the adults and go through a series of stages separated by molts (shedding of the skin) until they reach the final adult form.

Asexual Reproduction

Asexual reproduction, unlike sexual reproduction, produces offspring genetically identical to each other and to the parent. A number of animal species—especially those without backbones, but even some fish, amphibians, and reptiles—are capable of asexual reproduction. Asexual reproduction, except for occasional identical twinning, is absent in birds and mammals. The most common forms of asexual reproduction for stationary aquatic animals include budding and fragmentation, in which part of a parent individual can separate and grow into a new individual. In contrast, a form of asexual reproduction found in certain invertebrates and rare vertebrates is called parthenogenesis (or "virgin beginning"), in which unfertilized eggs develop into new offspring.

Sponges and Cnidarians

By the end of this section, you will be able to:

- Describe the organizational features of the simplest animals
- Describe the organizational features of cnidarians

The kingdom of animals is informally divided into invertebrate animals, those without a backbone, and vertebrate animals, those with a backbone. Although in general we are most familiar with vertebrate animals, the vast majority of animal species, about 95 percent, are invertebrates. Invertebrates include a huge diversity of animals, millions of species in about 32 phyla, which we can just begin to touch on here.

The sponges and the cnidarians represent the simplest of animals. Sponges appear to represent an early stage of multicellularity in the animal clade. Although they have specialized cells for particular functions, they lack true tissues in which specialized cells are organized into functional groups. Sponges are similar to what might have been the ancestor of animals: colonial, flagellated protists. The cnidarians, or the jellyfish and their kin, are the simplest animal group that displays true tissues, although they possess only two tissue layers.

Sponges

Animals in subkingdom Parazoa represent the simplest animals and include the sponges, or phylum **Porifera** (Figure 5.2). All sponges are aquatic and the majority of species are marine. Sponges live in intimate contact with water, which plays a role in their feeding, gas exchange, and excretion. Much of the body structure of the sponge is dedicated to moving water through the body so it can filter out food, absorb dissolved oxygen, and eliminate wastes.

Figure 5.2 Sponges are members of the phylum Porifera, which contains the simplest animals. (credit: Andrew Turner)

The body of the simplest sponges takes the shape of a cylinder with a large central cavity, the **spongocoel**. Water enters the spongocoel from numerous pores in the body wall. Water flows out through a large opening called the **osculum** (Figure 5.3). However, sponges exhibit a diversity of body forms, which vary in the size and branching of the spongocoel, the number of osculi, and where the cells that filter food from the water are located.

Sponges consist of an outer layer of flattened cells and an inner layer of cells called choanocytes separated by a jelly-like substance called **mesohyl**. The mesohyl contains embedded amoeboid cells that secrete tiny needles called **spicules** or protein fibers that help give the sponge its structural strength. The cell body of the **choanocyte** is embedded in mesohyl but protruding into the spongocoel is a mesh-like collar surrounding a single flagellum. The beating of flagella from all choanocytes moves water through the sponge. Food particles are trapped in mucus produced by the sieve-like collar of the choanocytes and are ingested by phagocytosis. This process is called **intracellular digestion**. **Amoebocytes** take up nutrients repackaged in food vacuoles of the choanocytes and deliver them to other cells within the sponge.

Osculum

Ostia

Spongocoel

Mesohyl

Basic sponge body plan
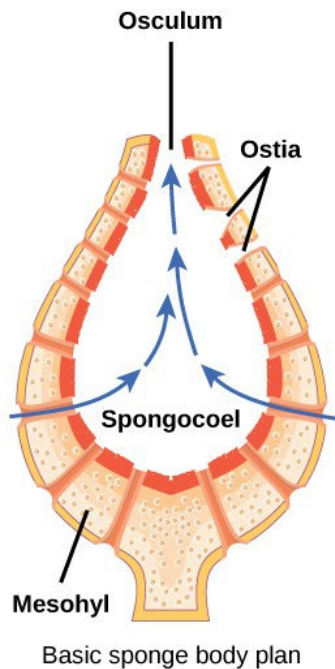
Figure 5.3 The sponge's basic body plan is shown.

Physiological Processes in Sponges

Despite their lack of complexity, sponges are clearly successful organisms, having persisted on Earth for more than half a billion years. Lacking a true digestive system, sponges depend on the intracellular digestive processes of their choanocytes for their energy intake. The limit of this type of digestion is that food particles must be smaller than individual cells. Gas exchange, circulation, and excretion occur by diffusion between cells and the water.

Sponges reproduce both sexually and asexually. Asexual reproduction is either by **fragmentation** (in which a piece of the sponge breaks off and develops into a new individual), or **budding** (an outgrowth from the parent that eventually detaches). A type of asexual reproduction found only in freshwater sponges occurs through the formation of **gemmules**, clusters of cells surrounded by a tough outer layer. Gemmules survive hostile environments and can attach to a substrate and grow into a new sponge.

Sponges are **monoecious** (or hermaphroditic), meaning one individual can produce both eggs and sperm. Sponges may be sequentially hermaphroditic, producing eggs first and sperm later. Eggs arise from amoebocytes and are retained within the spongocoel, whereas sperm arise from choanocytes and are ejected through the osculum. Sperm carried by water currents fertilize the eggs of other sponges. Early larval development occurs within the sponge, and free-swimming larvae are then released through the osculum. This is the only time that sponges exhibit mobility. Sponges are sessile as adults and spend their lives attached to a fixed substrate.

CONCEPT in ACTION

openstax COLLEGE™

Watch this **video (http://openstaxcollege.org/l/sponge_feed)** that demonstrates the feeding of sponges.

Cnidarians

The phylum **Cnidaria** includes animals that show radial or biradial symmetry and are diploblastic. Nearly all (about 99 percent) cnidarians are marine species. Cnidarians have specialized cells known as **cnidocytes** ("stinging cells") containing organelles called **nematocysts**. These cells are concentrated

176

around the mouth and tentacles of the animal and can immobilize prey with toxins. Nematocysts contain coiled threads that may bear barbs. The outer wall of the cell has a hairlike projection that is sensitive to touch. When touched, the cells fire the toxin-containing coiled threads that can penetrate and stun the predator or prey (see Figure 5.4).

(a) Nematocyst with stored thread and barb

(b) Nematocyst after firing
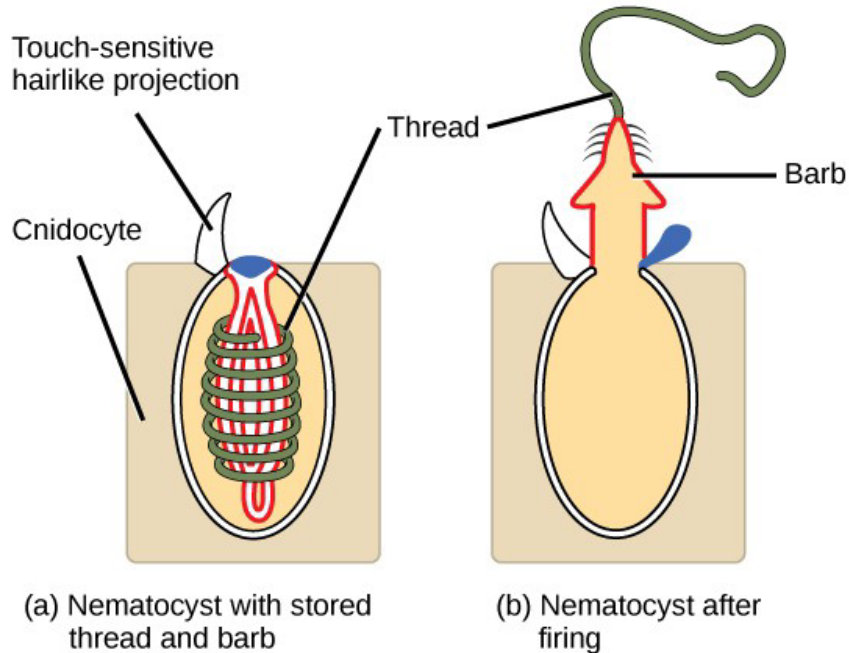
Figure 5.4 Animals from the phylum Cnidaria have stinging cells called cnidocytes. Cnidocytes contain large organelles called (a) nematocysts that store a coiled thread and barb. When hairlike projections on the cell surface are touched, (b) the thread, barb, and a toxin are fired from the organelle.

Cnidarians display two distinct body plans: **polyp** or "stalk" and **medusa** or "bell" (Figure 5.5). Examples of the polyp form are freshwater species of the genus *Hydra*; perhaps the best-known medusoid animals are the jellies (jellyfish). Polyps are sessile as adults, with a single opening to the digestive system (the mouth) facing up with tentacles surrounding it. Medusae are motile, with the mouth and tentacles hanging from the bell-shaped body. In other cnidarians, both a polyp and medusa form exist, and the life cycle alternates between these forms.
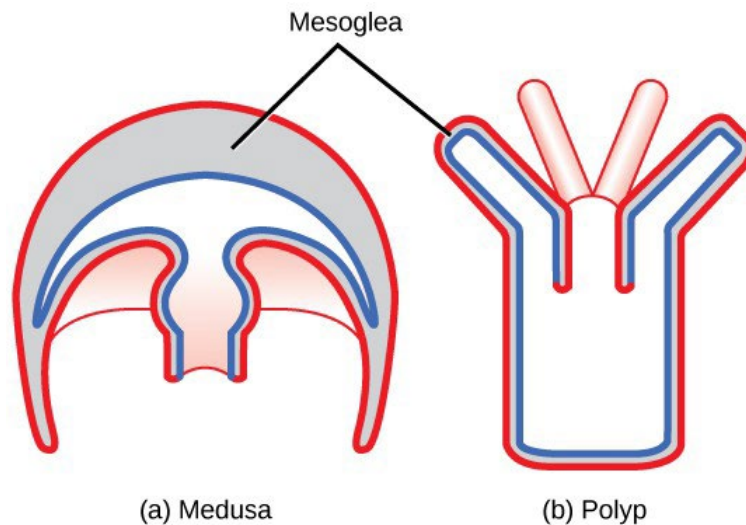
(a) Medusa

(b) Polyp

Figure 5.5 Cnidarians have two distinct body plans, the (a) medusa and the (b) polyp. All cnidarians have two tissue layers, with a jelly-like mesoglea between them.

Physiological Processes of Cnidarians

All cnidarians have two tissue layers. The outer layer is called the **epidermis**, whereas the inner layer is called the **gastrodermis** and lines the digestive cavity. Between these two layers is a non-living, jelly-like **mesoglea**. There are differentiated cell types in each tissue layer, such as nerve cells, enzyme-

177

secreting cells, and nutrient-absorbing cells, as well as intercellular connections between the cells. However, organs and organ systems are not present in this phylum.

The nervous system is primitive, with nerve cells scattered across the body in a network. The function of the nerve cells is to carry signals from sensory cells and to contractile cells. Groups of cells in the nerve net form nerve cords that may be essential for more rapid transmission. Cnidarians perform **extracellular digestion**, with digestion completed by intracellular digestive processes. Food is taken into the **gastrovascular cavity**, enzymes are secreted into the cavity, and the cells lining the cavity absorb the nutrient products of the extracellular digestive process. The gastrovascular cavity has only one opening that serves as both a mouth and an anus (an incomplete digestive system). Like the sponges, Cnidarian cells exchange oxygen, carbon dioxide, and nitrogenous wastes by diffusion between cells in the epidermis and gastrodermis with water.

Cnidarian Diversity

The phylum Cnidaria contains about 10,000 described species divided into four classes: Anthozoa, Scyphozoa, Cubozoa, and Hydrozoa.

The class Anthozoa includes all cnidarians that exhibit a sessile polyp body plan only; in other words, there is no medusa stage within their life cycle. Examples include sea anemones, sea pens, and corals, with an estimated number of 6,100 described species. Sea anemones are usually brightly colored and can attain a size of 1.8 to 10 cm in diameter. These animals are usually cylindrical in shape and are attached to a substrate. A mouth opening is surrounded by tentacles bearing cnidocytes (Figure 5.6).

Figure 5.6 Sea anemones are cnidarians of class Anthozoa. (credit: "Dancing With Ghosts"/Flickr)

Scyphozoans include all the jellies and are motile and exclusively marine with about 200 described species. The medusa is the dominant stage in the life cycle, although there is also a polyp stage. Species range from 2 cm in length to the largest scyphozoan species, *Cyanea capillata*, at 2 m across. Jellies display a characteristic bell-like body shape (Figure 5.7).
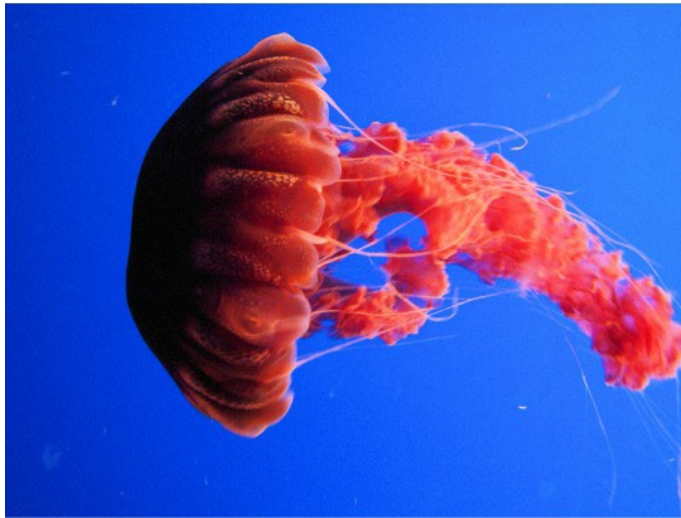
Figure 5.7 Scyphozoans include the jellies. (credit: "Jimg944"/Flickr)

Identify the life cycle stages of jellies using this **video animation game (http://openstaxcollege.org/l/amazing_jelly2)** from the New England Aquarium.

The class Cubozoa includes jellies that are square in cross-section and so are known as "box jellyfish." These species may achieve sizes of 15–25 cm. Cubozoans are anatomically similar to the jellyfish. A prominent difference between the two classes is the arrangement of tentacles. Cubozoans have muscular pads called pedalia at the corners of the square bell canopy, with one or more tentacles attached to each pedalium. In some cases, the digestive system may extend into the pedalia. Cubozoans typically exist in a polyp form that develops from a larva. The polyps may bud to form more polyps and then transform into the medusoid forms.

Watch this **video (http://openstaxcollege.org/l/box_jellyfish)** to learn more about the deadly toxins of the box jellyfish.

[2]

Hydrozoa includes nearly 3,500 species,   most of which are marine. Most species in this class have both polyp and medusa forms in their life cycle. Many hydrozoans form colonies composed of branches of specialized polyps that share a gastrovascular cavity. Colonies may also be free-floating and contain both medusa and polyp individuals in the colony, as in the Portuguese Man O'War (*Physalia*) or By-the-Wind Sailor (*Velella*). Other species are solitary polyps or solitary medusae. The characteristic shared by all of

2. "The Hydrozoa Directory," Peter Schuchert, Muséum Genève, last updated November 2012, http://www.ville-ge.ch/mhng/hydrozoa/hydrozoa-directory.htm.

these species is that their gonads are derived from epidermal tissue, whereas in all other cnidarians, they are derived from gastrodermal tissue (Figure 5.8**ab**).

(a) (b)
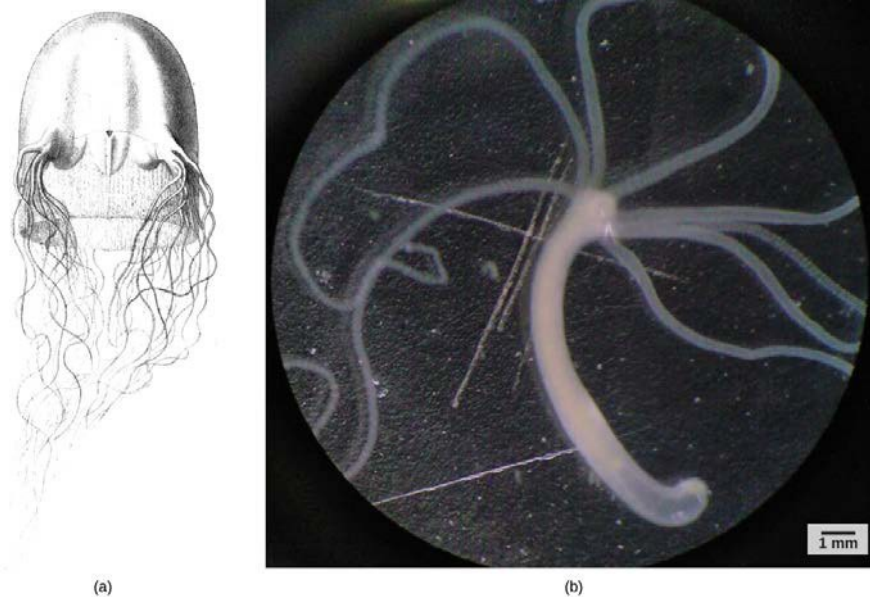
Figure 5.8 A (a) box jelly is an example from class Cubozoa. The (b) hydra is from class Hydrozoa. (credit b: scale-bar data from Matt Russell)

| Flatworms, Nematodes, and Arthropods

By the end of this section, you will be able to:

- Describe the structure and systems of flatworms
- Describe the structural organization of nematodes
- Compare the internal systems and the appendage specialization of arthropods

The animal phyla of this and subsequent modules are triploblastic and have an embryonic mesoderm sandwiched between the ectoderm and endoderm. These phyla are also bilaterally symmetrical, meaning that a longitudinal section will divide them into right and left sides that are mirror images of each other. Associated with bilateralism is the beginning of cephalization, the evolution of a concentration of nervous tissues and sensory organs in the head of the organism, which is where the organism first encounters its environment.

The flatworms are acoelomate organisms that include free-living and parasitic forms. The nematodes, or roundworms, possess a pseudocoelom and consist of both free-living and parasitic forms. Finally, the arthropods, one of the most successful taxonomic groups on the planet, are coelomate organisms with a hard exoskeleton and jointed appendages. The nematodes and the arthropods belong to a clade with a common ancestor, called Ecdysozoa. The name comes from the word *ecdysis*, which refers to the periodic shedding, or molting, of the exoskeleton. The ecdysozoan phyla have a hard cuticle covering their bodies that must be periodically shed and replaced for them to increase in size.

Flatworms

The relationships among flatworms, or phylum Platyhelminthes, is being revised and the description here will follow the traditional groupings. Most flatworms are parasitic, including important parasites of humans. Flatworms have three embryonic germ layers that give rise to surfaces covering tissues, internal tissues, and the lining of the digestive system. The epidermal tissue is a single layer of cells or a layer of fused cells covering a layer of circular muscle above a layer of longitudinal muscle. The mesodermal tissues include support cells and secretory cells that secrete mucus and other materials to the surface. The flatworms are acoelomate, so their bodies contain no cavities or spaces between the outer surface and the inner digestive tract.

180

Physiological Processes of Flatworms

Free-living species of flatworms are predators or scavengers, whereas parasitic forms feed from the tissues of their hosts. Most flatworms have an incomplete digestive system with an opening, the "mouth," that is also used to expel digestive system wastes. Some species also have an anal opening. The gut may be a simple sac or highly branched. Digestion is extracellular, with enzymes secreted into the space by cells lining the tract, and digested materials taken into the same cells by phagocytosis. One group, the cestodes, does not have a digestive system, because their parasitic lifestyle and the environment in which they live (suspended within the digestive cavity of their host) allows them to absorb nutrients directly across their body wall. Flatworms have an excretory system with a network of tubules throughout the body that open to the environment and nearby flame cells, whose cilia beat to direct waste fluids concentrated in the tubules out of the body. The system is responsible for regulation of dissolved salts and excretion of nitrogenous wastes. The nervous system consists of a pair of nerve cords running the length of the body with connections between them and a large ganglion or concentration of nerve cells at the anterior end of the worm; here, there may also be a concentration of photosensory and chemosensory cells (Figure 5.9).
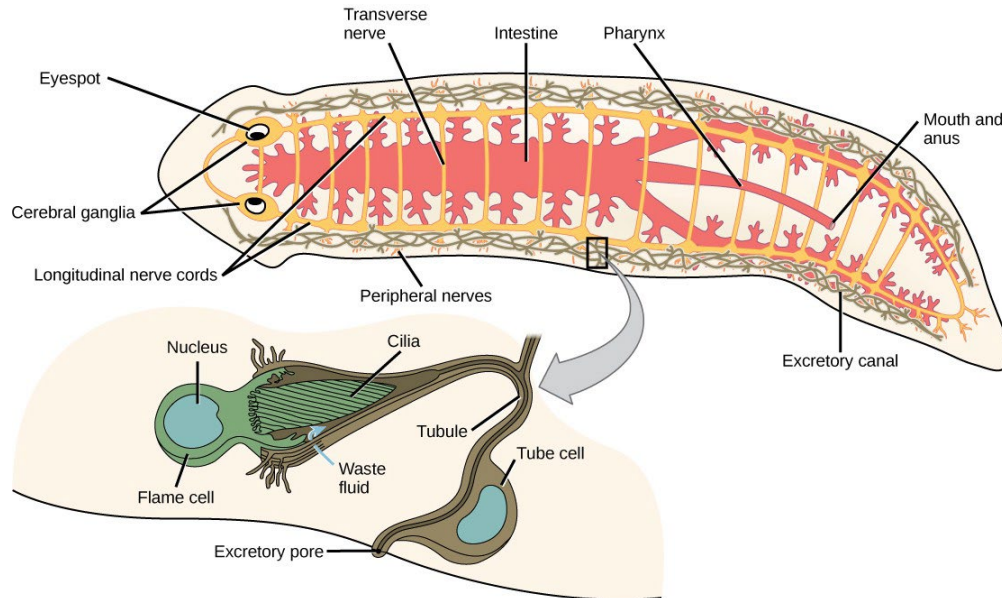
Figure 5.9 This planarian is a free-living flatworm that has an incomplete digestive system, an excretory system with a network of tubules throughout the body, and a nervous system made up of nerve cords running the length of the body with a concentration of nerves and photosensory and chemosensory cells at the anterior end.

Since there is no circulatory or respiratory system, gas and nutrient exchange is dependent on diffusion and intercellular junctions. This necessarily limits the thickness of the body in these organisms, constraining them to be "flat" worms. Most flatworm species are monoecious (hermaphroditic, possessing both sets of sex organs), and fertilization is typically internal. Asexual reproduction is common in some groups in which an entire organism can be regenerated from just a part of itself.

Diversity of Flatworms

Flatworms are traditionally divided into four classes: Turbellaria, Monogenea, Trematoda, and Cestoda (Figure 5.10). The turbellarians include mainly free-living marine species, although some species live in freshwater or moist terrestrial environments. The simple planarians found in freshwater ponds and aquaria are examples. The epidermal layer of the underside of turbellarians is ciliated, and this helps them move. Some turbellarians are capable of remarkable feats of regeneration in which they may regrow the body, even from a small fragment.
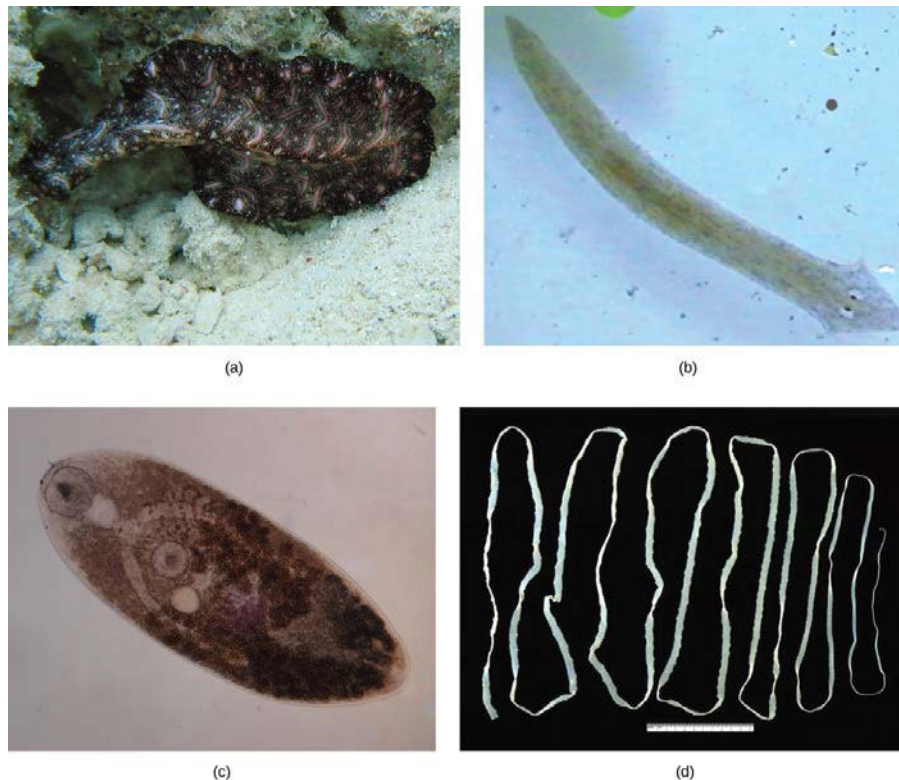
Figure 5.10 Phylum Platyhelminthes is divided into four classes: (a) Bedford's Flatworm (*Pseudobiceros bedfordi*) and the (b) planarian belong to class Turbellaria; (c) the Trematoda class includes about 20,000 species, most of which are parasitic; (d) class Cestoda includes tapeworms such as this *Taenia saginata*; and the parasitic class Monogenea (not shown). (credit a: modification of work by Jan Derk; credit c: modification of work by "Sahaquiel9102"/Wikimedia Commons; credit d: modification of work by CDC)

The monogeneans are external parasites mostly of fish with life cycles consisting of a free-swimming larva that attaches to a fish to begin transformation to the parasitic adult form. They have only one host during their life, typically of just one species. The worms may produce enzymes that digest the host tissues or graze on surface mucus and skin particles. Most monogeneans are hermaphroditic, but the sperm develop first, and it is typical for them to mate between individuals and not to self-fertilize.

The trematodes, or flukes, are internal parasites of mollusks and many other groups, including humans. Trematodes have complex life cycles that involve a primary host in which sexual reproduction occurs and one or more secondary hosts in which asexual reproduction occurs. The primary host is almost always a mollusk. Trematodes are responsible for serious human diseases including schistosomiasis, caused by a blood fluke (*Schistosoma*). The disease infects an estimated 200 million people in the tropics and leads to organ damage and chronic symptoms including fatigue. Infection occurs when a human enters the water, and a larva, released from the primary snail host, locates and penetrates the skin. The parasite infects various organs in the body and feeds on red blood cells before reproducing. Many of the eggs are released in feces and find their way into a waterway where they are able to reinfect the primary snail host.

The cestodes, or tapeworms, are also internal parasites, mainly of vertebrates. Tapeworms live in the intestinal tract of the primary host and remain fixed using a sucker on the anterior end, or scolex, of the tapeworm body. The remaining body of the tapeworm is made up of a long series of units called proglottids, each of which may contain an excretory system with flame cells, but will contain reproductive structures, both male and female. Tapeworms do not have a digestive system, they absorb nutrients from the food matter passing them in the host's intestine. Proglottids are produced at the scolex and are pushed to the end of the tapeworm as new proglottids form, at which point, they are "mature" and all structures except fertilized eggs have degenerated. Most reproduction occurs by cross-fertilization. The proglottid detaches and is released in the feces of the host. The fertilized eggs are eaten by an intermediate host. The juvenile worms emerge and infect the intermediate host, taking up residence, usually in muscle tissue. When the muscle tissue is eaten by the primary host, the cycle is completed. There are several tapeworm parasites of humans that are acquired by eating uncooked or poorly cooked pork, beef, and fish.

Nematodes

The phylum **Nematoda**, or roundworms, includes more than 28,000 species with an estimated 16,000 parasitic species. The name Nematoda is derived from the Greek word "nemos," which means "thread." Nematodes are present in all habitats and are extremely common, although they are usually not visible (Figure 5.11).

(a)
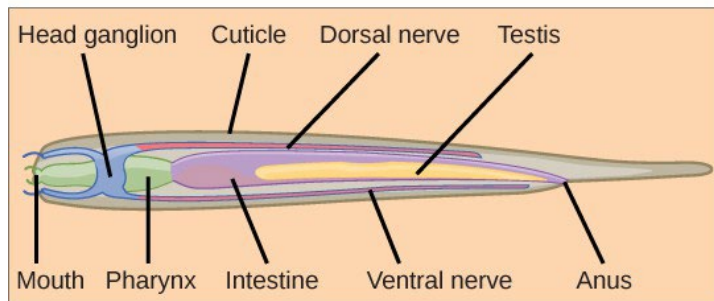
(b)

Figure 5.11 (a) An scanning electron micrograph of the nematode *Heterodera glycines* and (b) a schematic representation of the anatomy of a nematode are shown. (credit a: modification of work by USDA, ARS; scale-bar data from Matt Russell)

Most nematodes look similar to each other: slender tubes, tapered at each end (Figure 5.11). Nematodes are pseudocoelomates and have a **complete digestive system** with a distinct mouth and anus.

The nematode body is encased in a cuticle, a flexible but tough exoskeleton, or external skeleton, which offers protection and support. The cuticle contains a carbohydrate-protein polymer called **chitin**. The cuticle also lines the pharynx and rectum. Although the exoskeleton provides protection, it restricts growth, and therefore must be continually shed and replaced as the animal increases in size.

A nematode's mouth opens at the anterior end with three or six lips and, in some species, teeth in the form of cuticular extensions. There may also be a sharp stylet that can protrude from the mouth to stab prey or pierce plant or animal cells. The mouth leads to a muscular pharynx and intestine, leading to the rectum and anal opening at the posterior end.

Physiological Processes of Nematodes

In nematodes, the excretory system is not specialized. Nitrogenous wastes are removed by diffusion. In marine nematodes, regulation of water and salt is achieved by specialized glands that remove unwanted ions while maintaining internal body fluid concentrations.

183

Most nematodes have four nerve cords that run along the length of the body on the top, bottom, and sides. The nerve cords fuse in a ring around the pharynx, to form a head ganglion or "brain" of the worm, as well as at the posterior end to form the tail ganglion. Beneath the epidermis lies a layer of longitudinal muscles that permits only side-to-side, wave-like undulation of the body.

View this **video (http://openstaxcollege.org/l/nematode)** to see nematodes move about and feed on bacteria.

Nematodes employ a diversity of sexual reproductive strategies depending on the species; they may be monoecious, **dioecious** (separate sexes), or may reproduce asexually by parthenogenesis. *Caenorhabditis elegans* is nearly unique among animals in having both self-fertilizing hermaphrodites and a male sex that can mate with the hermaphrodite.

Arthropoda

The name "arthropoda" means "jointed legs," which aptly describes each of the enormous number of species belonging to this phylum. **Arthropoda** dominate the animal kingdom with an estimated 85 percent of known species, with many still undiscovered or undescribed. The principal characteristics of all the animals in this phylum are functional segmentation of the body and the presence of jointed appendages (Figure 5.12). As members of Ecdysozoa, arthropods also have an exoskeleton made principally of chitin. Arthropoda is the largest phylum in the animal world in terms of numbers of species, and insects form the single largest group within this phylum. Arthropods are true coelomate animals and exhibit prostostomic development.

Figure 5.12 Trilobites, like the one in this fossil, are an extinct group of arthropods. (credit: Kevin Walsh)

Physiological Processes of Arthropods

A unique feature of arthropods is the presence of a segmented body with fusion of certain sets of segments to give rise to functional segments. Fused segments may form a head, thorax, and abdomen, or a cephalothorax and abdomen, or a head and trunk. The coelom takes the form of a **hemocoel** (or blood cavity). The open circulatory system, in which blood bathes the internal organs rather than circulating in vessels, is regulated by a two-chambered heart. Respiratory systems vary, depending on the group of arthropod: Insects and myriapods use a series of tubes (**tracheae**) that branch throughout the body, open to the outside through openings called **spiracles**, and perform gas exchange directly between the cells and air in the tracheae. Aquatic crustaceans use gills, arachnids employ "book lungs," and aquatic chelicerates use "book gills." The book lungs of arachnids are internal stacks of alternating air pockets

184

and hemocoel tissue shaped like the pages of a book. The book gills of crustaceans are external structures similar to book lungs with stacks of leaf-like structures that exchange gases with the surrounding water (Figure 5.13).
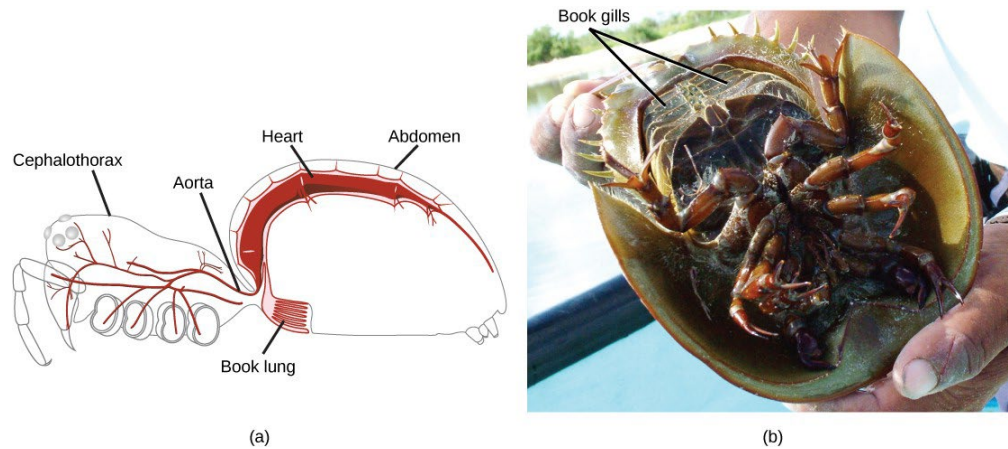

Figure 5.13 The book lungs of (a) arachnids are made up of alternating air pockets and hemocoel tissue shaped like a stack of books. The book gills of (b) crustaceans are similar to book lungs but are external so that gas exchange can occur with the surrounding water. (credit a: modification of work by Ryan Wilson based on original work by John Henry Comstock; credit b: modification of work by Angel Schatz)

Arthropod Diversity

Phylum Arthropoda includes animals that have been successful in colonizing terrestrial, aquatic, and aerial habitats. The phylum is further classified into five subphyla: Trilobitomorpha (trilobites), Hexapoda (insects and relatives), Myriapoda (millipedes, centipedes, and relatives), Crustacea (crabs, lobsters, crayfish, isopods, barnacles, and some zooplankton), and Chelicerata (horseshoe crabs, arachnids, scorpions, and daddy longlegs). Trilobites are an extinct group of arthropods found from the Cambrian period (540–490 million years ago) until they became extinct in the Permian (300–251 million years ago) that are probably most closely related to the Chelicerata. The 17,000 described species have been identified from fossils (Figure 5.15).

The Hexapoda have six legs (three pairs) as their name suggests. Hexapod segments are fused into a head, thorax, and abdomen (Figure 5.14). The thorax bears the wings and three pairs of legs. The insects we encounter on a daily basis—such as ants, cockroaches, butterflies, and bees—are examples of Hexapoda.
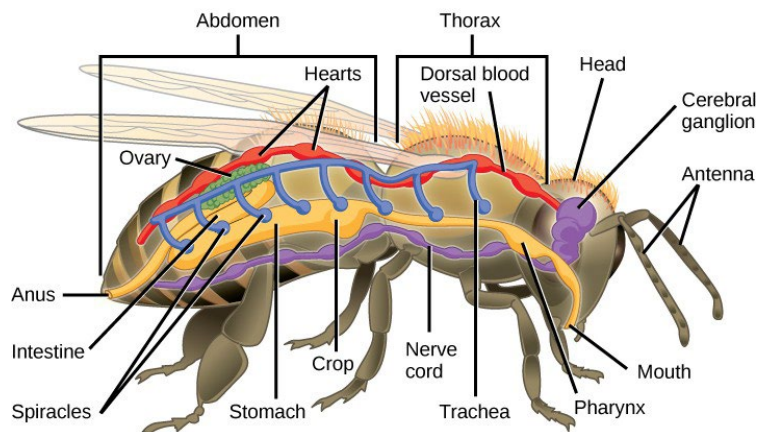
Figure 5.14 In this basic anatomy of a hexapod, note that insects have a developed digestive system (yellow), a respiratory system (blue), a circulatory system (red), and a nervous system (purple).

Subphylum Myriapoda includes arthropods with legs that may vary in number from 10 to 750. This subphylum includes 13,000 species; the most commonly found examples are millipedes and centipedes. All myriapods are terrestrial animals and prefer a humid environment (Figure 5.15).

185

Figure 5.15 (a) The centipede *Scutigera coleoptrata* has up to 15 pairs of legs. (b) This North American millipede (*Narceus americanus)* bears many legs, although not one thousand, as its name might suggest. (credit a: modification of work by Bruce Marlin; credit b: modification of work by Cory Zanker)

Crustaceans, such as shrimp, lobsters, crabs, and crayfish, are the dominant aquatic arthropods. A few crustaceans are terrestrial species like the pill bugs or sow bugs. The number of described crustacean [3] species stands at about 47,000.

Although the basic body plan in crustaceans is similar to the Hexapoda—head, thorax, and abdomen—the head and thorax may be fused in some species to form a **cephalothorax**, which is covered by a plate called the carapace (Figure 5.16). The exoskeleton of many species is also infused with calcium carbonate, which makes it even stronger than in other arthropods. Crustaceans have an open circulatory system in which blood is pumped into the hemocoel by the dorsal heart. Most crustaceans typically have separate sexes, but some, like barnacles, may be hermaphroditic. Serial hermaphroditism, in which the gonad can switch from producing sperm to ova, is also found in some crustacean species. Larval stages are seen in the early development of many crustaceans. Most crustaceans are carnivorous, but detritivores and filter feeders are also common.
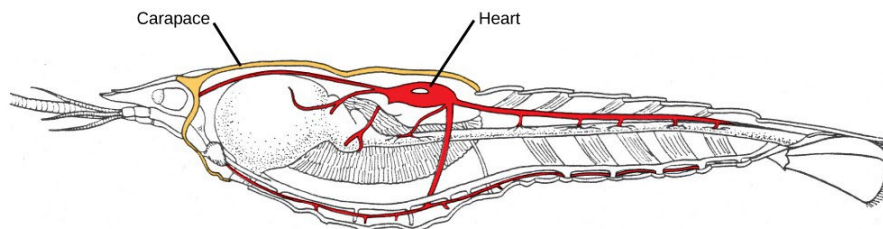
Figure 5.16 The crayfish is an example of a crustacean. It has a carapace around the cephalothorax and the heart in the dorsal thorax area. (credit: Jane Whitney)

Subphylum Chelicerata includes animals such as spiders, scorpions, horseshoe crabs, and sea spiders. This subphylum is predominantly terrestrial, although some marine species also exist. An estimated [4] 103,000 described species are included in subphylum Chelicerata.

The body of chelicerates may be divided into two parts and a distinct "head" is not always discernible. The phylum derives its name from the first pair of appendages: the **chelicerae** (Figure 5.17**a**), which are specialized mouthparts. The chelicerae are mostly used for feeding, but in spiders, they are typically modified to inject venom into their prey (Figure 5.17**b**). As in other members of Arthropoda, chelicerates also utilize an open circulatory system, with a tube-like heart that pumps blood into the large hemocoel that bathes the internal organs. Aquatic chelicerates utilize gill respiration, whereas terrestrial species use either tracheae or book lungs for gaseous exchange.

3. "Number of Living Species in Australia and the World," A.D. Chapman, Australia Biodiversity Information Services, last modified August 26, 2010, http://www.environment.gov.au/biodiversity/abrs/publications/other/species-numbers/2009/03-exec-summary.html.
4. "Number of Living Species in Australia and the World," A.D. Chapman, Australia Biodiversity Information Services, last modified August 26, 2010, http://www.environment.gov.au/biodiversity/abrs/publications/other/species-numbers/2009/03-exec-summary.html.

Figure 5.17 (a) The chelicerae (first set of appendages) are well developed in the Chelicerata, which includes scorpions (a) and spiders (b). (credit a: modification of work by Kevin Walsh; credit b: modification of work by Marshal Hedin)

Click through (http://openstaxcollege.org/l/arthropod2) this lesson on arthropods to explore interactive habitat maps and more.

| Mollusks and Annelids

By the end of this section, you will be able to:

- Describe the unique anatomical features of mollusks
- Describe the features of an animal classified in phylum Annelida

The mollusks are a diverse group (85,000 described species) of mostly marine species. They have a variety of forms, ranging from large predatory squid and octopus, some of which show a high degree of intelligence, to small grazing forms with elaborately sculpted and colored shells. The annelids traditionally include the oligochaetes, which include the earthworms and leeches, the polychaetes, which are a marine group, and two other smaller classes.

The phyla Mollusca and Annelida belong to a clade called the **Lophotrochozoa**, which also includes the phylum Nemertea, or ribbon worms. They are distinct from the Ecdysozoa (nematodes and arthropods) based on evidence from analysis of their DNA, which has changed our views of the relationships among invertebrates.

Phylum Mollusca

Mollusca is the predominant phylum in marine environments, where it is estimated that 23 percent of all known marine species belong to this phylum. It is the second most diverse phylum of animals with over 75,000 described species. The name "mollusca" signifies a soft body, as the earliest descriptions of mollusks came from observations of unshelled, soft-bodied cuttlefish (squid relatives). Although mollusk body forms vary, they share key characteristics, such as a ventral, muscular foot that is typically used for locomotion; the visceral mass, which contains most of the internal organs of the animal; and a dorsal mantle, which is a flap of tissue over the visceral mass that creates a space called the mantle cavity. The mantle may or may not secrete a shell of calcium carbonate. In addition, many mollusks have a scraping structure at the mouth, called a **radula** (Figure 5.18).

The muscular foot varies in shape and function, depending on the type of mollusk (described below in the section on mollusk diversity). It is a retractable as well as extendable organ, used for locomotion and anchorage. Mollusks are eucoelomates, but the coelomic cavity is restricted to a cavity around the heart in adult animals. The mantle cavity, formed inside the **mantle**, develops independently of the coelomic cavity. It is a multi-purpose space, housing the gills, the anus, organs for sensing food particles in the water, and an outlet for gametes. Most mollusks have an open circulatory system with a heart that circulates the hemolymph in open spaces around the organs. The octopuses and squid are an exception to this and have a closed circulatory system with two hearts that move blood through the gills and a third, systemic heart that pumps blood through the rest of the body.

art CONNECTION
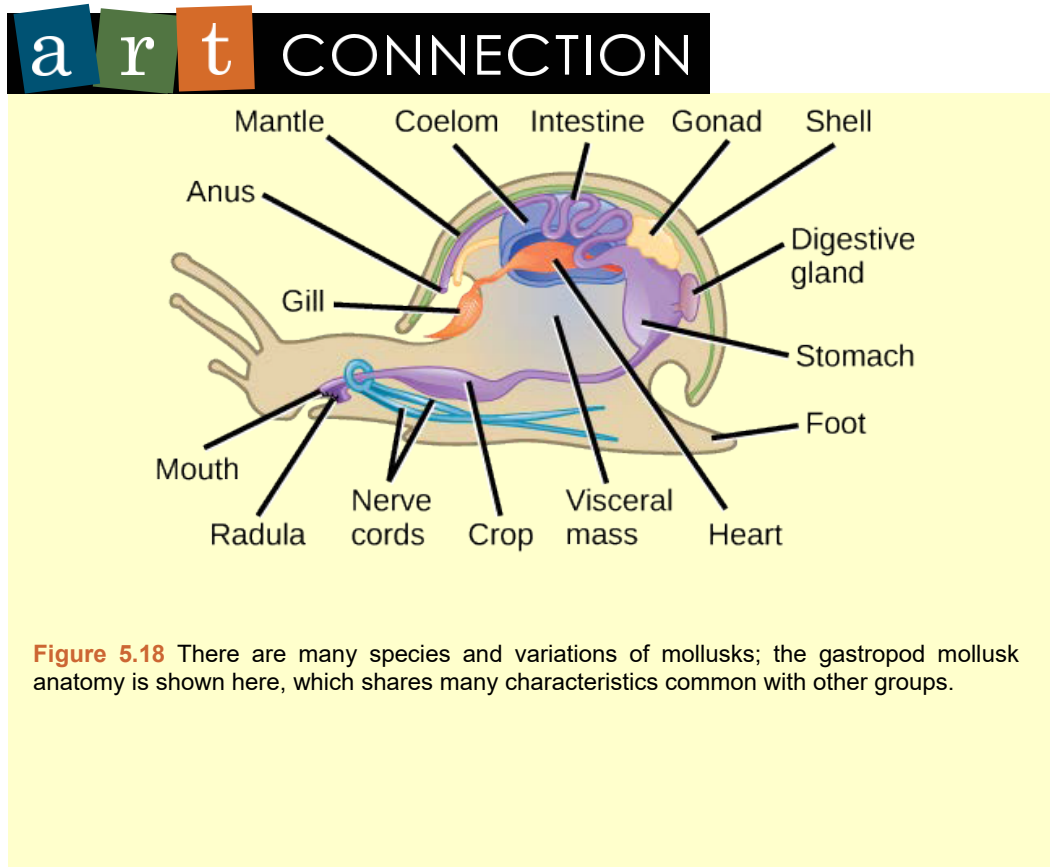
Figure 5.18 There are many species and variations of mollusks; the gastropod mollusk anatomy is shown here, which shares many characteristics common with other groups.

Mollusk Diversity

This phylum is comprised of seven classes: Aplacophora, Monoplacophora, Polyplacophora, Bivalvia, Gastropoda, Cephalopoda, and Scaphopoda.

Class Aplacophora ("bearing no plates") includes worm-like animals living mostly on deep ocean bottoms. These animals lack a shell but have aragonite spicules on their skin. Members of class Monoplacophora ("bearing one plate") have a single, cap-like shell enclosing the body. The monoplacophorans were believed extinct and only known as fossils until the discovery of *Neopilina galatheae* in 1952. Today, scientists have identified nearly two dozen living species.

Animals in the class Polyplacophora ("bearing many plates") are commonly known as "chitons" and bear an armor-like, eight-plated shell (Figure 5.19). These animals have a broad, ventral foot that is adapted for attachment to rocks and a mantle that extends beyond the shell in the form of a girdle. They breathe with **ctenidia** (gills) present ventrally. These animals have a radula modified for scraping. A single pair of nephridia for excretion is present.

- Download for free at http://openstaxcollege.org/textbooks/concepts-of-biology/get.

Figure 5.19 This chiton from the class Polyplacophora has the eight-plated shell indicative of its class. (credit: Jerry Kirkhart)

Class Bivalvia ("two shells") includes clams, oysters, mussels, scallops, and geoducks. They are found in marine and freshwater habitats. As the name suggests, bivalves are enclosed in a pair of shells (or valves) that are hinged at the dorsal side. The body is flattened on the sides. They feed by filtering particles from water and a radula is absent. They exchange gases using a pair of ctenidia, and excretion and osmoregulation are carried out by a pair of nephridia. In some species, the posterior edges of the mantle may fuse to form two siphons that inhale and exhale water. Some bivalves like oysters and mussels have the unique ability to secrete and deposit a calcareous **nacre** or "mother of pearl" around foreign particles that enter the mantle cavity. This property is commercially exploited to produce pearls.

Watch animations of **clams (http://openstaxcollege.org/l/clams2)** and **mussels (http://openstaxcollege.org/l/mussels2)** feeding to understand more about bivalves.

Gastropods ("stomach foot") include well-known mollusks like snails, slugs, conchs, sea hares, and sea butterflies. Gastropods include shell-bearing species as well as species with a reduced shell. These animals are asymmetrical and usually present a coiled shell (Figure 5.20).

(a)   (b)

Figure 5.20 (a) Like many gastropods, this snail has a stomach foot and a coiled shell. (b) This slug, which is also a gastropod, lacks a shell. (credit a: modification of work by Murray Stevenson; credit b: modification of work by Rosendahl)

The visceral mass in the shelled species is characteristically twisted and the foot is modified for crawling. Most gastropods bear a head with tentacles that support eyes. A complex radula is used to scrape food particles from the substrate. The mantle cavity encloses the ctenidia as well as a pair of nephridia.

The class Cephalopoda ("head foot" animals) includes octopuses, squids, cuttlefish, and nautilus. Cephalopods include shelled and reduced-shell groups. They display vivid coloration, typically seen in squids and octopuses, which is used for camouflage. The ability of some octopuses to rapidly adjust their colors to mimic a background pattern or to startle a predator is one of the more awe-inspiring feats of these animals. All animals in this class are predators and have beak-like jaws. All cephalopods have a well-developed nervous system, complex eyes, and a closed circulatory system. The foot is lobed and developed into tentacles and a funnel, which is used for locomotion. Suckers are present on the tentacles in octopuses and squid. Ctenidia are enclosed in a large mantle cavity and are serviced by large blood vessels, each with its own heart.

Cephalopods (Figure 5.27) are able to move quickly via jet propulsion by contracting the mantle cavity to forcefully eject a stream of water. Cephalopods have separate sexes, and the females of some species care for the eggs for an extended period of time. Although the shell is much reduced and internal in squid and cuttlefish, and absent altogether in octopus, nautilus live inside a spiral, multi-chambered shell that is filled with gas or water to regulate buoyancy.
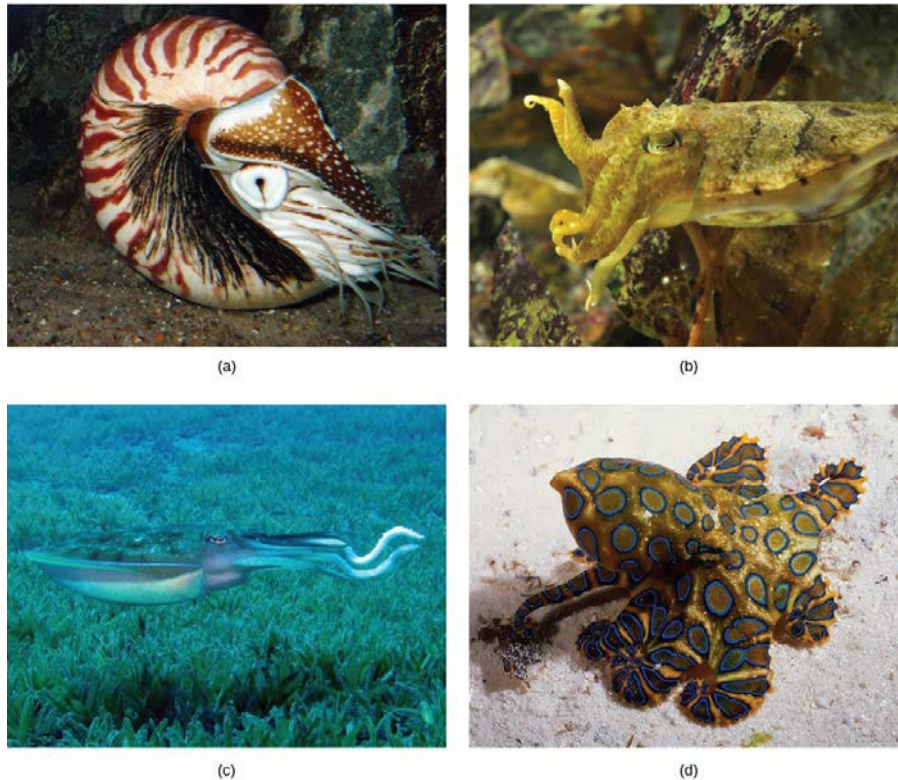
Figure 5.27 The (a) nautilus, (b) giant cuttlefish, (c) reef squid, and (d) blue-ring octopus are all members of the class Cephalopoda. (credit a: modification of work by J. Baecker; credit b: modification of work by Adrian Mohedano; credit c: modification of work by Silke Baron; credit d: modification of work by Angell Williams)

Members of the class Scaphopoda ("boat feet") are known colloquially as "tusk shells" or "tooth shells." Tooth shells are open at both ends and usually lie buried in sand with the front opening exposed to water and the reduced head end projecting from the back of the shell. Tooth shells have a radula and a foot modified into tentacles, each with a bulbous end that catches and manipulates prey (Figure 5.22).

Figure 5.22 *Antalis vulgaris* shows the classic Dentaliidae shape that gives these animals their common name of "tusk shell." (credit: Georges Jansoone)

Annelida

Phylum **Annelida** are segmented worms found in marine, terrestrial, and freshwater habitats, but the presence of water or humidity is a critical factor for their survival in terrestrial habitats. The name of the phylum is derived from the Latin word *annellus*, which means a small ring. Approximately 16,500 species have been described. The phylum includes earthworms, polychaete worms, and leeches. Like mollusks, annelids exhibit protostomic development.

191

Annelids are bilaterally symmetrical and have a worm-like appearance. Their particular segmented body plan results in repetition of internal and external features in each body segment. This type of body plan is called **metamerism**. The evolutionary benefit of such a body plan is thought to be the capacity it allows for the evolution of independent modifications in different segments that perform different functions. The overall body can then be divided into head, body, and tail.

Physiological Processes of Annelida

The skin of annelids is protected by a cuticle that is thinner than the cuticle of the ecdysozoans and does not need to be molted for growth. Chitinous hairlike extensions, anchored in the skin and projecting from the cuticle, called **chaetae**, are present in every segment in most groups. The chaetae are a defining character of annelids. Polychaete worms have paired, unjointed limbs called parapodia on each segment used for locomotion and breathing. Beneath the cuticle there are two layers of muscle, one running around its circumference (circular) and one running the length of the worm (longitudinal). Annelids have a true coelom in which organs are distributed and bathed in coelomic fluid. Annelids possess a well- developed complete digestive system with specialized organs: mouth, muscular pharynx, esophagus, and crop. A cross-sectional view of a body segment of an earthworm is shown in Figure 5.23; each segment is limited by a membrane that divides the body cavity into compartments.

Annelids have a closed circulatory system with muscular pumping "hearts" in the anterior segments, dorsal and ventral blood vessels that run the length of the body with connections in each segment, and capillaries that service individual tissues. Gas exchange occurs across the moist body surface. Excretion is carried out by pairs of primitive "kidneys" called metanephridia that consist of a convoluted tubule and an open, ciliated funnel present in every segment. Annelids have a well-developed nervous system with two ventral nerve cords and a nerve ring of fused ganglia present around the pharynx.
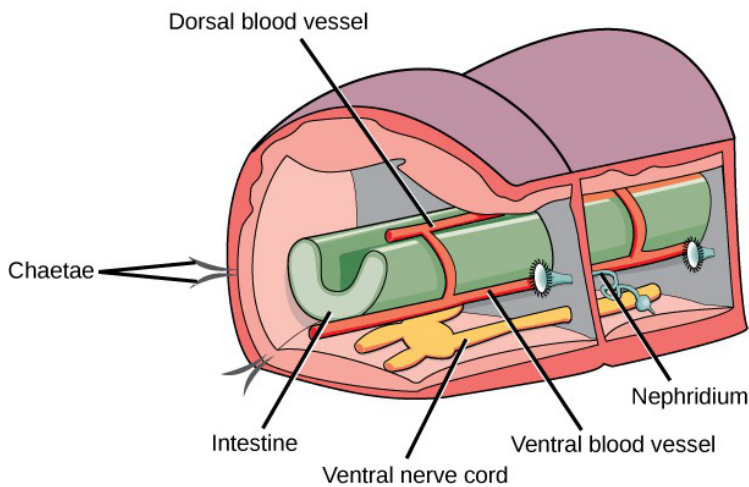
Figure 5.23 In this schematic showing the basic anatomy of annelids, the digestive system is indicated in green, the nervous system is indicated in yellow, and the circulatory system is indicated in red.

Annelids may be either monoecious with permanent gonads (as in earthworms and leeches) or dioecious with temporary or seasonal gonads (as in polychaetes).

This **video and animation (http://openstaxcollege.org/l/annelid2)** provides a close-up look at annelid anatomy.

192

Annelid Diversity

Phylum Annelida includes the classes Polychaeta and Clitellata (**Figure 15.30**); the latter contains subclasses Oligochaeta, Hirudinoidea, and Branchiobdellida.

Earthworms are the most abundant members of the subclass Oligochaeta, distinguished by the presence of the **clitellum**, a ring structure in the skin that secretes mucus to bind mating individuals and forms a protective cocoon for the eggs. They also have a few, reduced chaetae (oligo- = "few"; -chaetae = "hairs"). The number and size of chaetae is greatly diminished in oligochaetes as compared to the polychaetes (poly- = "many"; -chaetae = "hairs"). The chaetae of polychaetes are also arranged within fleshy, flat, paired appendages on each segment called parapodia.

The subclass Hirudinoidea includes leeches. Significant differences between leeches and other annelids include the development of suckers at the anterior and posterior ends, and the absence of chaetae. Additionally, the segmentation of the body wall may not correspond to internal segmentation of the coelomic cavity. This adaptation may allow leeches to swell when ingesting blood from host vertebrates. The subclass Branchiobdellida includes about 150 species that show similarity to leeches as well as oligochaetes. All species are obligate symbionts, meaning that they can only survive associated with their host, mainly with freshwater crayfish. They feed on the algae that grows on the carapace of the crayfish.

(a)   (b)

Figure 5.24 The (a) earthworm and (b) leech are both annelids. (credit a: modification of work by "schizoform"/Flickr; credit b: modification of work by "Sarah G..."/Flickr)

Echinoderms and Chordates

By the end of this section, you will be able to:

- Describe the distinguishing characteristics of echinoderms
- Describe the distinguishing characteristics of chordates

Deuterostomes include the phyla Echinodermata and Chordata (which includes the vertebrates) and two smaller phyla. Deuterostomes share similar patterns of early development.

Echinoderms

Echinodermata are named for their spiny skin (from the Greek "echinos" meaning "spiny" and "dermos" [5] meaning "skin"). The phylum includes about 7,000 described living species, such as sea stars, sea cucumbers, sea urchins, sand dollars, and brittle stars. **Echinodermata** are exclusively marine.

Adult echinoderms exhibit pentaradial symmetry and have a calcareous endoskeleton made of ossicles (Figure 5.25), although the early larval stages of all echinoderms have bilateral symmetry. The endoskeleton is developed by epidermal cells, which may also possess pigment cells, giving vivid colors to these animals, as well as cells laden with toxins. These animals have a true coelom, a portion of which is modified into a unique circulatory system called a **water vascular system**. An interesting feature of these animals is their power to regenerate, even when over 75 percent of their body mass is lost.

5. "Number of Living Species in Australia and the World," A.D. Chapman, Australia Biodiversity Information Services, last modified August 26, 2010, http://www.environment.gov.au/biodiversity/abrs/publications/other/species-numbers/2009/03-exec-summary.html.

Physiological Processes of Echinoderms

Echinoderms have a unique system for gas exchange, nutrient circulation, and locomotion called the water vascular system. The system consists of a central ring canal and radial canals extending along each arm. Water circulates through these structures allowing for gas, nutrient, and waste exchange. A structure on top of the body, called the **madreporite**, regulates the amount of water in the water vascular system. "Tube feet," which protrude through openings in the endoskeleton, may be expanded or contracted using the hydrostatic pressure in the system. The system allows for slow movement, but a great deal of power, as witnessed when the tube feet latch on to opposite halves of a bivalve mollusk, like a clam, and slowly, but surely pull the shells apart, exposing the flesh within.
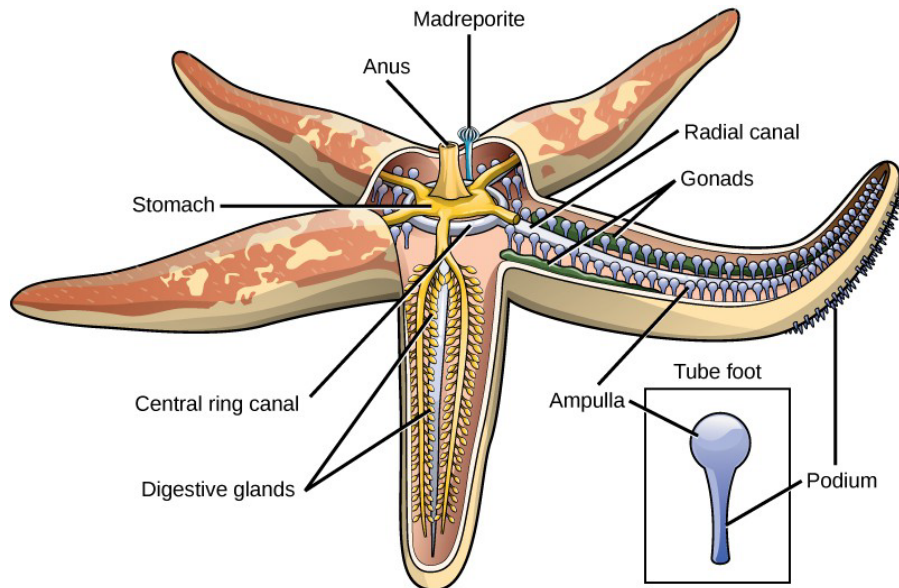
Figure 5.25 This diagram shows the anatomy of a sea star.

The echinoderm nervous system has a nerve ring at the center and five radial nerves extending outward along the arms. There is no centralized nervous control. Echinoderms have separate sexes and release their gametes into the water where fertilization takes place. Echinoderms may also reproduce asexually through regeneration from body parts.

Echinoderm Diversity

This phylum is divided into five classes: Asteroidea (sea stars), Ophiuroidea (brittle stars), Echinoidea (sea urchins and sand dollars), Crinoidea (sea lilies or feather stars), and Holothuroidea (sea cucumbers) (Figure 5.26).

Perhaps the best-known echinoderms are members of the class Asteroidea, or sea stars. They come in a large variety of shapes, colors, and sizes, with more than 1,800 species known. The characteristics of sea stars that set them apart from other echinoderm classes include thick arms that extend from a central disk where organs penetrate into the arms. Sea stars use their tube feet not only for gripping surfaces but also for grasping prey. Sea stars have two stomachs, one of which they can evert through their mouths to secrete digestive juices into or onto prey before ingestion. This process can essentially liquefy the prey and make digestion easier.

View this **video (http://openstaxcollege.org/l/echinoderm2)** to explore a sea star's body plan up close, watch one move across the sea floor, and see it devour a mussel.

Brittle stars have long, thin arms that do not contain any organs. Sea urchins and sand dollars do not have arms but are hemispherical or flattened with five rows of tube feet, which help them in slow movement. Sea lilies and feather stars are stalked suspension feeders. Sea cucumbers are soft-bodied and elongate with five rows of tube feet and a series of tube feet around the mouth that are modified into tentacles used in feeding.
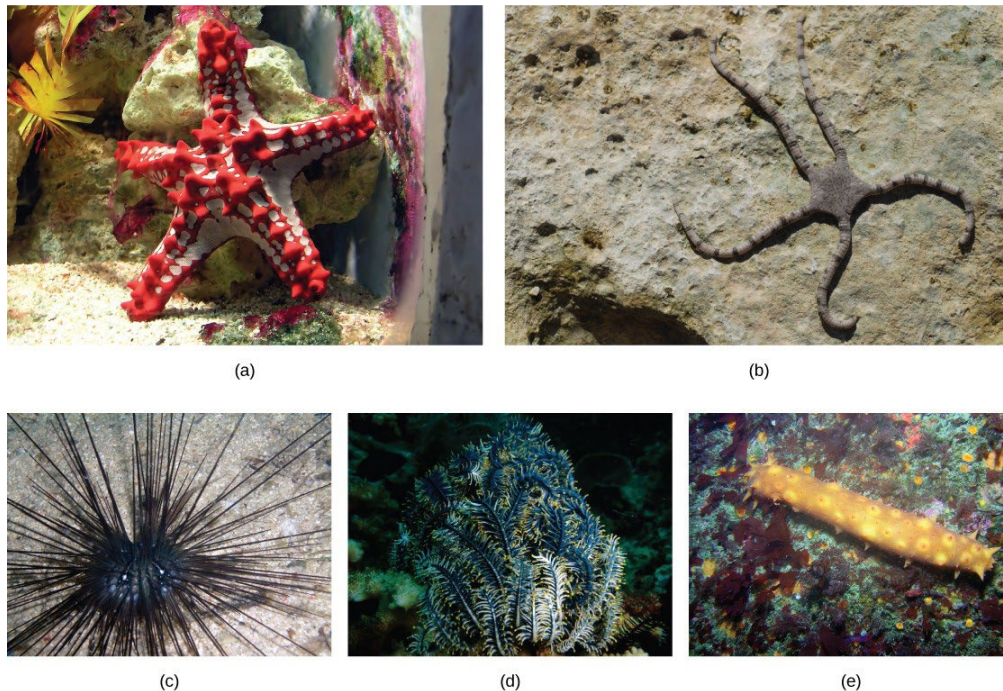
Figure 5.26 Different members of Echinodermata include the (a) sea star in class Asteroidea,

(b) the brittle star in class Ophiuroidea, (c) the sea urchins of class Echinoidea, (d) the sea lilies belonging to class Crinoidea, and (d) sea cucumbers representing class Holothuroidea. (credit a: modification of work by Adrian Pingstone; credit b: modification of work by Joshua Ganderson; credit c: modification of work by Samuel Chow; credit d: modification of work by Sarah Depper; credit e: modification of work by Ed Bierman)

Chordates

The majority of species in the phylum Chordata are found in the subphylum Vertebrata, which include many species with which we are familiar. The vertebrates contain more than 60,000 described species, divided into major groupings of the lampreys, fishes, amphibians, reptiles, birds, and mammals.

Animals in the phylum **Chordata** share four key features that appear at some stage of their development: a notochord, a dorsal hollow nerve cord, pharyngeal slits, and a post-anal tail (Figure 5.27). In certain groups, some of these traits are present only during embryonic development.

The chordates are named for the **notochord**, which is a flexible, rod-shaped structure that is found in the embryonic stage of all chordates and in the adult stage of some chordate species. It is located between the digestive tube and the nerve cord, and provides skeletal support through the length of the body. In some chordates, the notochord acts as the primary axial support of the body throughout the animal's lifetime. In vertebrates, the notochord is present during embryonic development, at which time it induces the development of the neural tube and serves as a support for the developing embryonic body. The notochord, however, is not found in the postnatal stage of vertebrates; at this point, it has been replaced by the **vertebral column** (the spine).

The **dorsal hollow nerve cord** is derived from ectoderm that sinks below the surface of the skin and rolls into a hollow tube during development. In chordates, it is located dorsally to the notochord. In contrast, other animal phyla possess solid nerve cords that are located either ventrally or laterally. The nerve cord found in most chordate embryos develops into the brain and spinal cord, which compose the central nervous system.

Pharyngeal slits are openings in the pharynx, the region just posterior to the mouth, that extend to the outside environment. In organisms that live in aquatic environments, pharyngeal slits allow for the exit of water that enters the mouth during feeding. Some invertebrate chordates use the pharyngeal slits to filter food from the water that enters the mouth. In fishes, the pharyngeal slits are modified into gill supports, and in jawed fishes, jaw supports. In tetrapods, the slits are further modified into components

195

of the ear and tonsils, since there is no longer any need for gill supports in these air-breathing animals. *Tetrapod* means "four-footed," and this group includes amphibians, reptiles, birds, and mammals. (Birds are considered tetrapods because they evolved from tetrapod ancestors.)

The **post-anal tail** is a posterior elongation of the body extending beyond the anus. The tail contains skeletal elements and muscles, which provide a source of locomotion in aquatic species, such as fishes. In some terrestrial vertebrates, the tail may also function in balance, locomotion, courting, and signaling when danger is near. In many species, the tail is absent or reduced; for example, in apes, including humans, it is present in the embryo, but reduced in size and nonfunctional in adults.
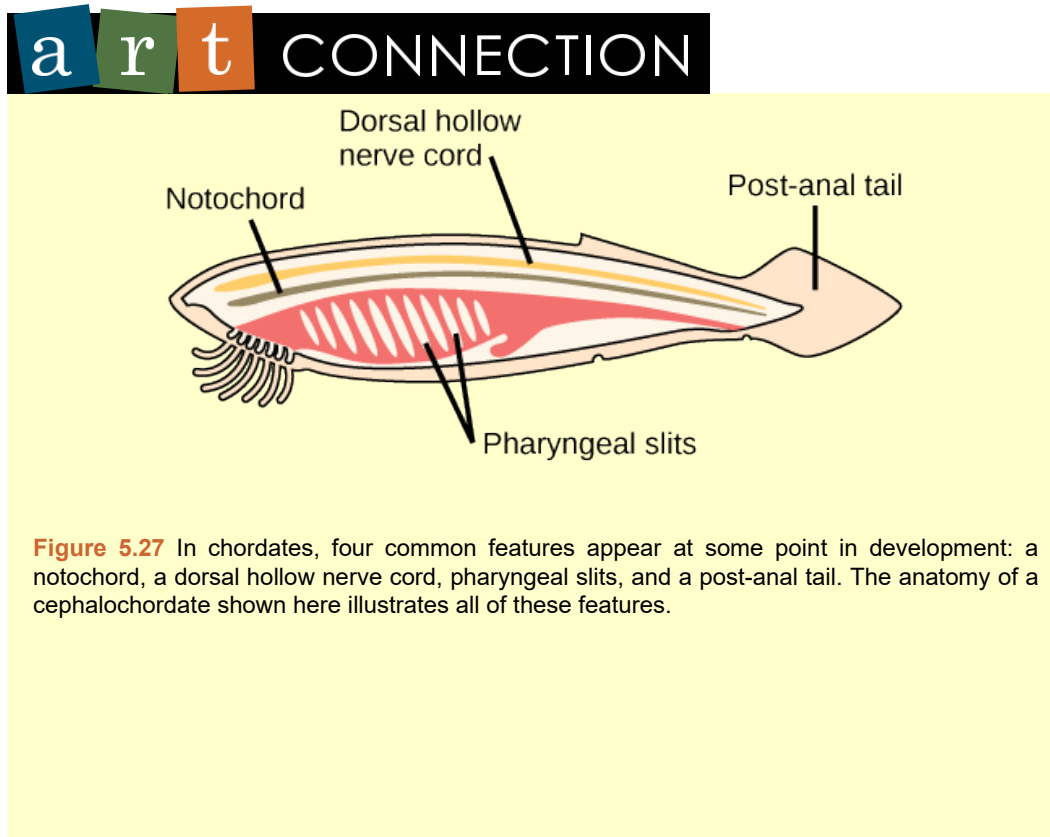
Figure 5.27 In chordates, four common features appear at some point in development: a notochord, a dorsal hollow nerve cord, pharyngeal slits, and a post-anal tail. The anatomy of a cephalochordate shown here illustrates all of these features.

Invertebrate Chordates

In addition to the vertebrates, the phylum Chordata contains two clades of invertebrates: **Urochordata** (tunicates) and **Cephalochordata** (lancelets). Members of these groups possess the four distinctive features of chordates at some point during their development.

The **tunicates** (Figure 5.28) are also called sea squirts. The name tunicate derives from the cellulose- like carbohydrate material, called the tunic, which covers the outer body. Although tunicates are classified as chordates, the adult forms are much modified in body plan and do not have a notochord, a dorsal hollow nerve cord, or a post-anal tail, although they do have pharyngeal slits. The larval form possesses all four structures. Most tunicates are hermaphrodites. Tunicate larvae hatch from eggs inside the adult tunicate's body. After hatching, a tunicate larva swims for a few days until it finds a suitable surface on which it can attach, usually in a dark or shaded location. It then attaches by the head to the substrate and undergoes metamorphosis into the adult form, at which point the notochord, nerve cord, and tail disappear.

196
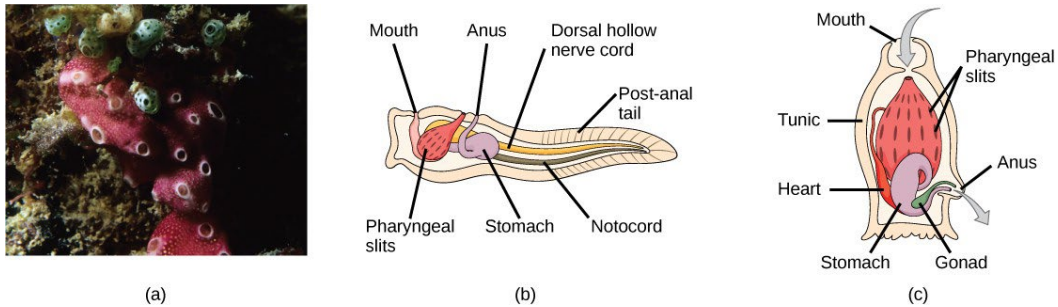
Figure 5.28 (a) This photograph shows a colony of the tunicate *Botrylloides violaceus*. In the (b) larval stage, the tunicate can swim freely until it attaches to a substrate to become (c) an adult. (credit a: modification of work by Dr. Dwayne Meadows, NOAA/NMFS/OPR)

Most tunicates live a sessile existence in shallow ocean waters and are suspension feeders. The primary foods of tunicates are plankton and detritus. Seawater enters the tunicate's body through its incurrent siphon. Suspended material is filtered out of this water by a mucus net (pharyngeal slits) and is passed into the intestine through the action of cilia. The anus empties into the excurrent siphon, which expels wastes and water.

Lancelets possess a notochord, dorsal hollow nerve cord, pharyngeal slits, and a post-anal tail in the adult stage (Figure 4.29). The notochord extends into the head, which gives the subphylum its name (Cephalochordata). Extinct fossils of this subphylum date to the middle of the Cambrian period (540–488 mya).The living forms, the lancelets, are named for their blade-like shape. Lancelets are only a few centimeters long and are usually found buried in sand at the bottom of warm temperate and tropical seas. Like tunicates, they are suspension feeders.
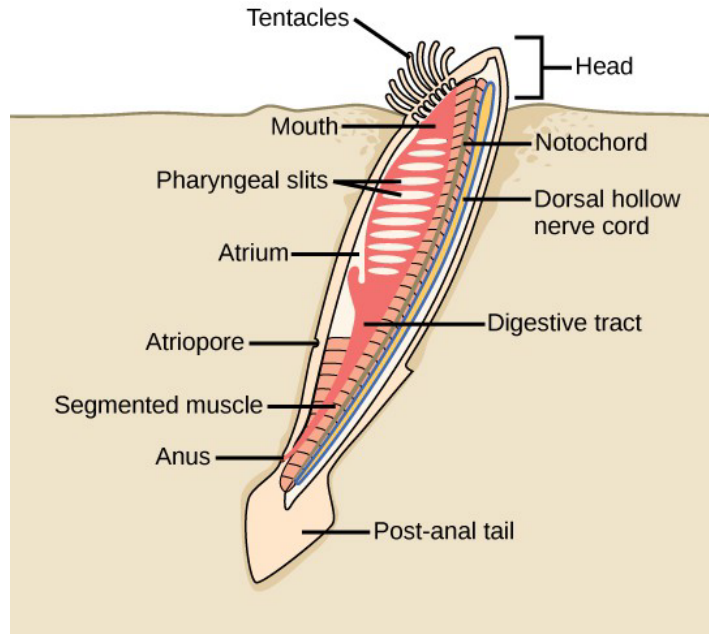
Figure 5.29 Adult lancelets retain the four key features of chordates: a notochord, a dorsal hollow nerve cord, pharyngeal slits, and a post-anal tail.

| Vertebrates

By the end of this section, you will be able to:

- Describe the difference between jawless and jawed fishes
- Explain the main characteristics of amphibians, reptiles, and birds
- Describe the derived characteristics in birds that facilitate flight
- Name and describe the distinguishing features of the three main groups of mammals
- Describe the derived features that distinguish primates from other animals

Vertebrates are among the most recognizable organisms of the animal kingdom (Figure 5.30). More than 62,000 vertebrate species have been identified. The vertebrate species now living represent only a small portion of the vertebrates that have existed. The best-known extinct vertebrates are the dinosaurs, a unique group of reptiles, reaching sizes not seen before or since in terrestrial animals. They were the dominant terrestrial animals for 150 million years, until they died out near the end of the Cretaceous period in a mass extinction. A great deal is known about the anatomy of the dinosaurs, given the preservation of their skeletal elements in the fossil record.

(a)   (b)   (c)
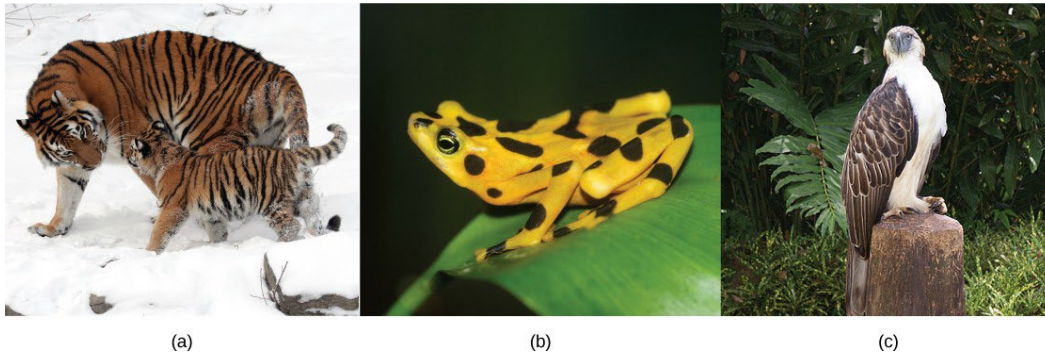
Figure 5.30 Examples of critically endangered vertebrate species include (a) the Siberian tiger *(Panthera tigris altaica)*, (b) the Panamanian golden frog *(Atelopus zeteki)*, and (c) the Philippine eagle *(Pithecophaga jefferyi)*. (credit a: modification of work by Dave Pape; credit b: modification of work by Brian Gratwicke; credit c: modification of work by "cuatrok77"/Flickr)

Fishes

Modern fishes include an estimated 31,000 species. Fishes were the earliest vertebrates, and jawless fishes were the earliest of these. Jawless fishes—the present day hagfishes and lampreys—have a distinct cranium and complex sense organs including eyes, distinguishing them from the invertebrate chordates. The jawed fishes evolved later and are extraordinarily diverse today. Fishes are active feeders, rather than sessile, suspension feeders.

Jawless Fishes

Jawless fishes are **craniates** (which includes all the chordate groups except the tunicates and lancelets) that represent an ancient vertebrate lineage that arose over one half-billion years ago. Some of the earliest jawless fishes were the **ostracoderms** (which translates as "shell-skin"). Ostracoderms, now extinct, were vertebrate fishes encased in bony armor, unlike present-day jawless fishes, which lack bone in their scales.

The clade **Myxini** includes 67 species of hagfishes. **Hagfishes** are eel-like scavengers that live on the ocean floor and feed on dead invertebrates, other fishes, and marine mammals (Figure 5.31**a**). Hagfishes are entirely marine and are found in oceans around the world except for the polar regions. A unique feature of these animals is the slime glands beneath the skin that are able to release an extraordinary amount of mucus through surface pores. This mucus may allow the hagfish to escape from the grip of predators. Hagfish are known to enter the bodies of dead or dying organisms to devour them from the inside.

-     Download for free at http://openstaxcollege.org/textbooks/concepts-of-biology/get.

Figure 5.31 (a) Pacific hagfishes are scavengers that live on the ocean floor. (b) These parasitic sea lampreys attach to their lake trout host by suction and use their rough tongues to rasp away flesh in order to feed on the trout's blood. (credit a: modification of work by Linda Snook, NOAA/ CBNMS; credit b: modification of work by USGS)

The skeleton of a hagfish is composed of cartilage, which includes a cartilaginous notochord, which runs the length of the body, and a skull. This notochord provides support to the fish's body. Although they are craniates, hagfishes are not vertebrates, since they do not replace the notochord with a vertebral column during development, as do the vertebrates.

The clade **Petromyzontidae** includes approximately 40 species of lampreys. **Lampreys** are similar to hagfishes in size and shape; however, lampreys have a brain case and incomplete vertebrae. Lampreys lack paired appendages and bone, as do the hagfishes. As adults, lampreys are characterized by a toothed, funnel-like sucking mouth. Some species are parasitic as adults, attaching to and feeding on the body fluids of fish (**Figure 15.37b**). Most species are free-living.

Lampreys live primarily in coastal and fresh waters and have a worldwide temperate region distribution. All species spawn in fresh waters. Eggs are fertilized externally, and the larvae are distinctly different from the adult form, spending 3 to 15 years as suspension feeders. Once they attain sexual maturity, the adults reproduce and die within days. Lampreys have a notochord as adults.

Jawed Fishes

Gnathostomes or "jaw-mouths" are vertebrates that have jaws and include both cartilaginous and bony fishes. One of the most significant developments in early vertebrate evolution was the origin of the jaw, which is a hinged structure attached to the cranium that allows an animal to grasp and tear its food. The evolution of jaws allowed early gnathostomes to exploit food resources that were unavailable to jawless fishes.

The clade **Chondrichthyes**, the cartilaginous fishes, is diverse, consisting of sharks (Figure 5.32**a**), rays, and skates, together with sawfishes and a few dozen species of fishes called *chimaeras*, or ghost sharks. Chondrichthyes have paired fins and a skeleton made of cartilage. This clade arose approximately 370 million years ago in the middle Devonian. They are thought to have descended from an extinct group that had a skeleton made of bone; thus, the cartilaginous skeleton of Chondrichthyes is a later development. Parts of the shark skeleton are strengthened by granules of calcium carbonate, but this is not the same as bone.

Most cartilaginous fishes live in marine habitats, with a few species living in fresh water for some or all of their lives. Most sharks are carnivores that feed on live prey, either swallowing it whole or using their jaws and teeth to tear it into smaller pieces. Shark teeth likely evolved from the jagged scales that cover their skin. Some species of sharks and rays are suspension feeders that feed on plankton.

199

(a)   (b)

Figure 5.32 (a) This hammerhead shark is an example of a predatory cartilaginous fish. (b) This stingray blends into the sandy bottom of the ocean floor when it is feeding or awaiting prey. (credit a: modification of work by Masashi Sugawara; credit b: modification of work by "Sailn1"/Flickr)

Sharks have well-developed sense organs that aid them in locating prey, including a keen sense of smell and electroreception, the latter being perhaps the most sensitive of any animal. Organs called **ampullae of Lorenzini** allow sharks to detect the electromagnetic fields that are produced by all living things, including their prey. Electroreception has only been observed in aquatic or amphibious animals. Sharks, together with most fishes, also have a sense organ called the **lateral line**, which is used to detect movement and vibration in the surrounding water, and a sense that is often considered homologous to "hearing" in terrestrial vertebrates. The lateral line is visible as a darker stripe that runs along the length of the fish's body.

Sharks reproduce sexually and eggs are fertilized internally. Most species are ovoviviparous, that is, the fertilized egg is retained in the oviduct of the mother's body, and the embryo is nourished by the egg yolk. The eggs hatch in the uterus and young are born alive and fully functional. Some species of sharks are oviparous: They lay eggs that hatch outside of the mother's body. Embryos are protected by a shark egg case or "mermaid's purse" that has the consistency of leather. The shark egg case has tentacles that snag in seaweed and give the newborn shark cover. A few species of sharks are viviparous, that is, the young develop within the mother's body, and she gives live birth.

Rays and skates include more than 500 species and are closely related to sharks. They can be distinguished from sharks by their flattened bodies, pectoral fins that are enlarged and fused to the head, and gill slits on their ventral surface (Figure 5.32**b**). Like sharks, rays and skates have a cartilaginous skeleton. Most species are marine and live on the sea floor, with nearly a worldwide distribution.

Bony Fishes

Members of the clade **Osteichthyes**, or bony fishes, are characterized by a bony skeleton. The vast majority of present-day fishes belong to this group, which consists of approximately 30,000 species, making it the largest class of vertebrates in existence today.

Nearly all bony fishes have an ossified skeleton with specialized bone cells (osteocytes) that produce and maintain a calcium phosphate matrix. This characteristic has only reverted in a few groups of Osteichthyes, such as sturgeons and paddlefish, which have primarily cartilaginous skeletons. The skin of bony fishes is often covered in overlapping scales, and glands in the skin secrete mucus that reduces drag when swimming and aids the fish in osmoregulation. Like sharks, bony fishes have a lateral line system that detects vibrations in water. Unlike sharks, some bony fish depend on their eyesight to locate prey. Bony fish are also unusual in possessing taste cells in the head and trunk region of the body that allow them to detect extremely small concentrations of molecules in the water.

All bony fishes, like the cartilaginous fishes, use gills to breathe. Water is drawn over gills that are located in chambers covered and ventilated by a protective, muscular flap called the operculum. Unlike sharks, bony fishes have a **swim bladder**, a gas-filled organ that helps to control the buoyancy of the fish. Bony fishes are further divided into two clades with living members: **Actinopterygii** (ray-finned fishes) and **Sarcopterygii** (lobe-finned fishes).

The ray-finned fishes include many familiar fishes—tuna, bass, trout, and salmon (Figure 5.33**a**), among others. Ray-finned fishes are named for the form of their fins—webs of skin supported by bony spines called rays. In contrast, the fins of lobe-finned fishes are fleshy and supported by bone (Figure 5.33**b**). Living members of lobe-finned fishes include the less familiar lungfishes and coelacanth.

200
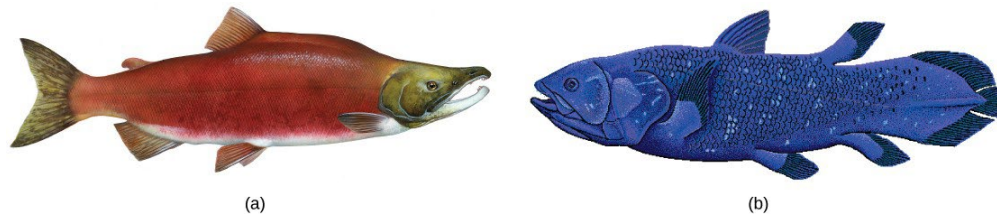
Figure 5.33 The (a) sockeye salmon and (b) coelacanth are both bony fishes of the Osteichthyes clade. The coelacanth, sometimes called a lobe-finned fish, was thought to have gone extinct in the Late Cretaceous period 100 million years ago until one was discovered in 1938 between Africa and Madagascar. (credit a: modification of work by Timothy Knepp, USFWS; credit b: modification of work by Robbie Cada)

Amphibians

Amphibians are vertebrate tetrapods. **Amphibia** includes frogs, salamanders, and caecilians. The term amphibian means "dual life," which is a reference to the metamorphosis that many frogs undergo from a tadpole to an adult and the mixture of aquatic and terrestrial environments in their life cycle. Amphibians evolved in the Devonian period and were the earliest terrestrial tetrapods.

As tetrapods, most amphibians are characterized by four well-developed limbs, although some species of salamanders and all caecilians possess only vestigial limbs. An important characteristic of extant amphibians is a moist, permeable skin, achieved by mucus glands. The moist skin allows oxygen and carbon dioxide exchange with the environment, a process called **cutaneous respiration**. All living adult amphibian species are carnivorous, and some terrestrial amphibians have a sticky tongue that is used to capture prey.

Amphibian Diversity

Amphibia comprise an estimated 6,500 extant species that inhabit tropical and temperate regions around the world. Amphibians can be divided into three clades: **Urodela** ("tailed-ones"), the salamanders and newts; **Anura** ("tail-less ones"), the frogs and toads; and **Apoda** ("legless ones"), the caecilians.

Living **salamanders** (Figure 5.34a) include approximately 500 species, some of which are aquatic, others terrestrial, and some that live on land only as adults. Adult salamanders usually have a generalized tetrapod body plan with four limbs and a tail. Some salamanders are lungless, and respiration occurs through the skin or external gills. Some terrestrial salamanders have primitive lungs; a few species have both gills and lungs.

Figure 5.34 (a) Most salamanders have legs and a tail, but respiration varies among species. (b) The Australian green tree frog is a nocturnal predator that lives in the canopies of trees near a water source. (credit a: modification of work by Valentina Storti; credit b: modification of work by Evan Pickett)

Watch this **video (http://openstaxcollege.org/l/river_monster2)** about an unusually large salamander species.

Frogs (Figure 5.34**b**) are the most diverse group of amphibians, with approximately 5,000 species that live on all continents except Antarctica. Frogs have a body plan that is more specialized than the salamander body plan for movement on land. Adult frogs use their hind limbs to jump many times their body length on land. Frogs have a number of modifications that allow them to avoid predators, including skin that acts as camouflage and defensive chemicals that are poisonous to predators secreted from glands in the skin.

Frog eggs are fertilized externally, as they are laid in moist environments. Frogs demonstrate a range of parental behaviors, with some species exhibiting little care, to species that carry eggs and tadpoles on their hind legs or backs. The life cycle consists of two stages: the larval stage followed by metamorphosis to an adult stage. The larval stage of a frog, the **tadpole**, is often a filter-feeding herbivore. Tadpoles usually have gills, a lateral line system, long-finned tails, but no limbs. At the end of the tadpole stage, frogs undergo a gradual metamorphosis into the adult form. During this stage, the gills and lateral line system disappear, and four limbs develop. The jaws become larger and are suited for carnivorous feeding, and the digestive system transforms into the typical short gut of a predator. An eardrum and air-breathing lungs also develop. These changes during metamorphosis allow the larvae to move onto land in the adult stage (Figure 5.35).

(a)   (b)   (c)

Figure 5.35 A frog begins as a (a) tadpole and undergoes metamorphosis to become (b) a juvenile and finally (c) an adult. (credit: modification of work by Brian Gratwicke)

Caecilians comprise an estimated 185 species. They lack external limbs and resemble giant earthworms. They inhabit soil and are found primarily in the tropics of South America, Africa, and southern Asia where they are adapted for a soil-burrowing lifestyle and are nearly blind. Unlike most of the other amphibians that breed in or near water, reproduction in a drier soil habitat means that caecilians must utilize internal fertilization, and most species give birth to live young (Figure 5.36).

Figure 5.36 Caecilians lack external limbs and are well adapted for a soil-burrowing lifestyle. (credit: modification of work by "cliff1066"/Flickr)

202

Reptiles and Birds

The **amniotes**—reptiles, birds, and mammals—are distinguished from amphibians by their terrestrially adapted (shelled) egg and an embryo protected by amniotic membranes. The evolution of amniotic membranes meant that the embryos of amniotes could develop within an aquatic environment inside the egg. This led to less dependence on a water environment for development and allowed the amniotes to invade drier areas. This was a significant evolutionary change that distinguished them from amphibians, which were restricted to moist environments due to their shell-less eggs. Although the shells of various amniotic species vary significantly, they all allow retention of water. The membranes of the amniotic egg also allowed gas exchange and sequestering of wastes within the enclosure of an eggshell. The shells of bird eggs are composed of calcium carbonate and are hard and brittle, but possess pores for gas and water exchange. The shells of reptile eggs are more leathery and pliable. Most mammals do not lay eggs; however, even with internal gestation, amniotic membranes are still present.

In the past, the most common division of amniotes has been into classes Mammalia, Reptilia, and Aves. Birds are descended, however, from dinosaurs, so this classical scheme results in groups that are not true clades. We will discuss birds as a group distinct from reptiles with the understanding that this does not reflect evolutionary history.

Reptiles

Reptiles are tetrapods. Limbless reptiles—snakes—may have vestigial limbs and, like caecilians, are classified as tetrapods because they are descended from four-limbed ancestors. Reptiles lay shelled eggs on land. Even aquatic reptiles, like sea turtles, return to the land to lay eggs. They usually reproduce sexually with internal fertilization. Some species display ovoviviparity, with the eggs remaining in the mother's body until they are ready to hatch. Other species are viviparous, with the offspring born alive.

One of the key adaptations that permitted reptiles to live on land was the development of their scaly skin, containing the protein keratin and waxy lipids, which prevented water loss from the skin. This occlusive skin means that reptiles cannot use their skin for respiration, like amphibians, and thus all must breathe with lungs. In addition, reptiles conserve valuable body water by excreting nitrogen in the form of uric acid paste. These characteristics, along with the shelled, amniotic egg, were the major reasons why reptiles became so successful in colonizing a variety of terrestrial habitats far from water.

Reptiles are ectotherms, that is, animals whose main source of body heat comes from the environment. Behavioral maneuvers, like basking to heat themselves, or seeking shade or burrows to cool off, help them regulate their body temperature,

Class Reptilia includes diverse species classified into four living clades. These are the Crocodilia, Sphenodontia, Squamata, and Testudines.

The **Crocodilia** ("small lizard") arose approximately 84 million years ago, and living species include alligators, crocodiles, and caimans. Crocodilians (Figure 5.37**a**) live throughout the tropics of Africa, South America, the southeastern United States, Asia, and Australia. They are found in freshwater habitats, such as rivers and lakes, and spend most of their time in water. Some species are able to move on land due to their semi-erect posture.

Figure 5.37 (a) Crocodilians, such as this Siamese crocodile, provide parental care for their offspring. (b) This Jackson's chameleon blends in with its surroundings. (c) The garter snake belongs to the genus *Thamnophis*, the most widely distributed reptile genus in North America. (d) The African spurred tortoise lives at the southern edge of the Sahara Desert. It is the third largest tortoise in the world. (credit a: modification of work by Keshav Mukund Kandhadai; credit c: modification of work by Steve Jurvetson; credit d: modification of work by Jim Bowen)

The **Sphenodontia** ("wedge tooth") arose in the Mesozoic Era and includes only one living genus, *Tuatara*, with two species that are found in New Zealand. There are many fossil species extending back to the Triassic period (250–200 million years ago). Although the tuataras resemble lizards, they are anatomically distinct and share characteristics that are found in birds and turtles.

Squamata ("scaly") arose in the late Permian; living species include lizards and snakes, which are the largest extant clade of reptiles (Figure 5.37**b**). Lizards differ from snakes by having four limbs, eyelids, and external ears, which are lacking in snakes. Lizard species range in size from chameleons and geckos that are a few centimeters in length to the Komodo dragon, which is about 3 meters in length.

Snakes are thought to have descended from either burrowing lizards or aquatic lizards over 100 million years ago (Figure 5.37**c**). Snakes comprise about 3,000 species and are found on every continent except Antarctica. They range in size from 10 centimeter-long thread snakes to 7.5 meter-long pythons and anacondas. All snakes are carnivorous and eat small animals, birds, eggs, fish, and insects.

Turtles are members of the clade **Testudines** ("having a shell") (Figure 5.37**d**). Turtles are characterized by a bony or cartilaginous shell, made up of the carapace on the back and the plastron on the ventral surface, which develops from the ribs. Turtles arose approximately 200 million years ago, predating crocodiles, lizards, and snakes. Turtles lay eggs on land, although many species live in or near water. Turtles range in size from the speckled padloper tortoise at 8 centimeters (3.1 inches) to the leatherback sea turtle at 200 centimeters (over 6 feet). The term "turtle" is sometimes used to describe only those species of Testudines that live in the sea, with the terms "tortoise" and "terrapin" used to refer to species that live on land and in fresh water, respectively.

Birds

Data now suggest that birds belong within the reptile clade, but they display a number of unique adaptations that set them apart. Unlike the reptiles, birds are endothermic, meaning they generate their own body heat through metabolic processes. The most distinctive characteristic of birds is their feathers, which are modified reptilian scales. Birds have several different types of feathers that are specialized for

- Download for free at http://openstaxcollege.org/textbooks/concepts-of-biology/get

specific functions, like contour feathers that streamline the bird's exterior and loosely structured **down feathers** that insulate (Figure 5.38**a**).

Feathers not only permitted the earliest birds to glide, and ultimately engage in flapping flight, but they insulated the bird's body, assisting the maintenance of endothermy, even in cooler temperatures. Powering a flying animal requires economizing on the amount of weight carried. As body weight increases, the muscle output and energetic cost required for flying increase. Birds have made several modifications to reduce body weight, including hollow or **pneumatic bones** (Figure 5.38**b**) with air spaces that may be connected to air sacs and cross-linked struts within their bones to provide structural reinforcement. Parts of the vertebral skeleton and braincase are fused to increase its strength while lightening its weight. Most species of bird only possess one ovary rather than two, and no living birds have teeth in their jaw, further reducing body mass.

Primary feathers

Secondary feathers

(a)   (b)
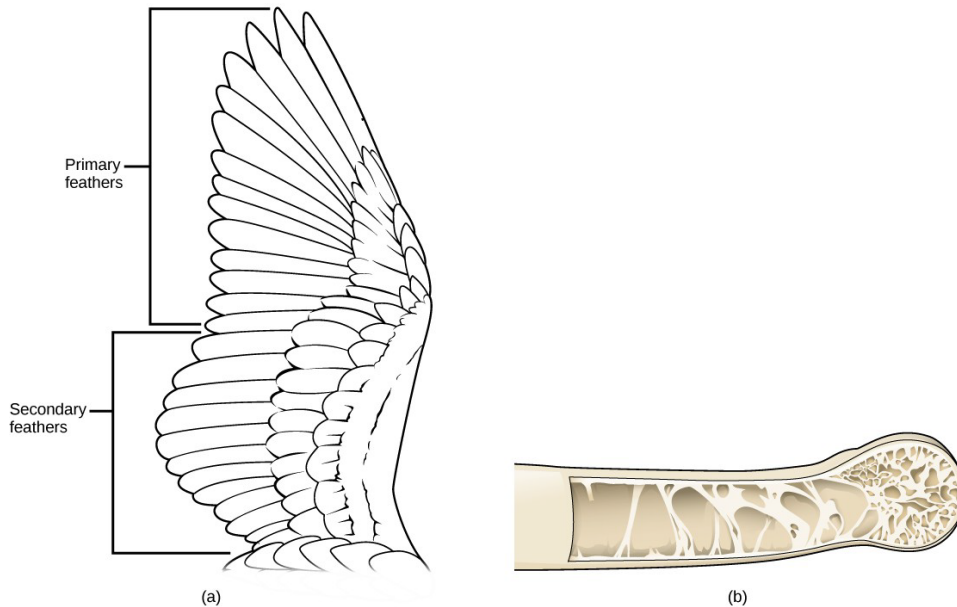
Figure 5.38 (a) Primary feathers are located at the wing tip and provide thrust; secondary feathers are located close to the body and provide lift. (b) Many birds have hollow pneumatic bones, which make flight easier.

Birds possess a system of air sacs branching from their primary airway that divert the path of air so that it passes unidirectionally through the lung, during both inspiration and expiration. Unlike mammalian lungs in which air flows in two directions as it is breathed in and out, air flows continuously through the bird's lung to provide a more efficient system of gas exchange.

Mammals

Mammals are vertebrates that have hair and mammary glands used to provide nutrition for their young. Certain features of the jaw, skeleton, skin, and internal anatomy are also unique to mammals. The presence of hair is one of the key characteristics of a mammal. Although it is not very extensive in some groups, such as whales, hair has many important functions for mammals. Mammals are endothermic, and hair provides insulation by trapping a layer of air close to the body to retain metabolic heat. Hair also serves as a sensory mechanism through specialized hairs called vibrissae, better known as whiskers. These attach to nerves that transmit touch information, which is particularly useful to nocturnal or burrowing mammals. Hair can also provide protective coloration.

Mammalian skin includes secretory glands with various functions. **Sebaceous glands** produce a lipid mixture called sebum that is secreted onto the hair and skin for water resistance and lubrication. Sebaceous glands are located over most of the body. **Sudoriferous glands** produce sweat and scent, which function in thermoregulation and communication, respectively. **Mammary glands** produce milk that is used to feed newborns. While male monotremes and eutherians possess mammary glands, male marsupials do not.

The skeletal system of mammals possesses unique features that differentiate them from other vertebrates. Most mammals have **heterodont teeth**, meaning they have different types and shapes of teeth that allow them to feed on different kinds of foods. These different types of teeth include the incisors, the canines, premolars, and molars. The first two types are for cutting and tearing, whereas the latter two types are for crushing and grinding. Different groups have different proportions of each type, depending on their diet.

Most mammals are also **diphyodonts**, meaning they have two sets of teeth in their lifetime: deciduous or "baby" teeth, and permanent teeth. In other vertebrates, the teeth can be replaced throughout life.

Modern mammals are divided into three broad groups: monotremes, marsupials, and eutherians (or placental mammals). The eutherians, or placental mammals, and the marsupials collectively are called therian mammals, whereas monotremes are called metatherians.

There are three living species of **monotremes**: the platypus and two species of echidnas, or spiny anteaters (Figure 3.39). The platypus and one species of echidna are found in Australia, whereas the other species of echidna is found in New Guinea. Monotremes are unique among mammals, as they lay leathery eggs, similar to those of reptiles, rather than giving birth to live young. However, the eggs are retained within the mother's reproductive tract until they are almost ready to hatch. Once the young hatch, the female begins to secrete milk from pores in a ridge of mammary tissue along the ventral side of her body. Like other mammals, monotremes are endothermic but regulate body temperatures somewhat lower (90 °F, 32 °C) than placental mammals do (98 °F, 37 °C). Like reptiles, monotremes have one posterior opening for urinary, fecal, and reproductive products, rather than three separate openings like placental mammals do. Adult monotremes lack teeth.

Figure 3.39 The platypus (left), a monotreme, possesses a leathery beak and lays eggs rather than giving birth to live young. An echidna, another monotreme, is shown in the right photo. (credit "echidna": modification of work by Barry Thomas)

Marsupials are found primarily in Australia and nearby islands, although about 100 species of opossums and a few species of two other families are found in the Americas. Australian marsupials number over 230 species and include the kangaroo, koala, bandicoot, and Tasmanian devil (Figure 5.40). Most species of marsupials possess a pouch in which the young reside after birth, receiving milk and continuing to develop. Before birth, marsupials have a less complex placental connection, and the young are born much less developed than in placental mammals.

Figure 5.40 The Tasmanian devil is one of several marsupials native to Australia. (credit: Wayne McLean)

Eutherians are the most widespread of the mammals, occurring throughout the world. There are several groups of eutherians, including Insectivora, the insect eaters; Edentata, the toothless anteaters; Rodentia, the rodents; Chiroptera, the bats; Cetacea, the aquatic mammals including whales; Carnivora, carnivorous mammals including dogs, cats, and bears; and Primates, which includes humans. **Eutherian**

mammals are sometimes called placental mammals, because all species have a complex placenta that connects a fetus to the mother, allowing for gas, fluid, waste, and nutrient exchange. While other mammals may possess a less complex placenta or briefly have a placenta, all eutherians have a complex placenta during gestation.

Primates

Order **Primates** of class Mammalia includes lemurs, tarsiers, monkeys, and the apes, which include humans. Non-human primates live primarily in tropical or subtropical regions of South America, Africa, and Asia. They range in size from the mouse lemur at 30 grams (1 ounce) to the mountain gorilla at 200 kilograms (441 pounds). The characteristics and evolution of primates are of particular interest to us as they allow us to understand the evolution of our own species.

All primate species have adaptations for climbing trees, as they all descended from tree-dwellers, although not all species are arboreal. This arboreal heritage of primates resulted in hands and feet that are adapted for **brachiation**, or climbing and swinging through trees. These adaptations include, but are not limited to 1) a rotating shoulder joint, 2) a big toe that is widely separated from the other toes and thumbs that are widely separated from fingers (except humans), which allow for gripping branches, and
3) **stereoscopic vision**, two overlapping visual fields, which allows for the depth perception necessary to gauge distance. Other characteristics of primates are brains that are larger than those of many other mammals, claws that have been modified into flattened nails, typically only one offspring per pregnancy, and a trend toward holding the body upright.

Order Primates is divided into two groups: prosimians and anthropoids. **Prosimians** include the bush babies of Africa, the lemurs of Madagascar, and the lorises, pottos, and tarsiers of Southeast Asia. **Anthropoids** include monkeys, lesser apes, and great apes (Figure 5.41). In general, prosimians tend to be nocturnal, smaller in size than anthropoids, and have relatively smaller brains compared to anthropoids.
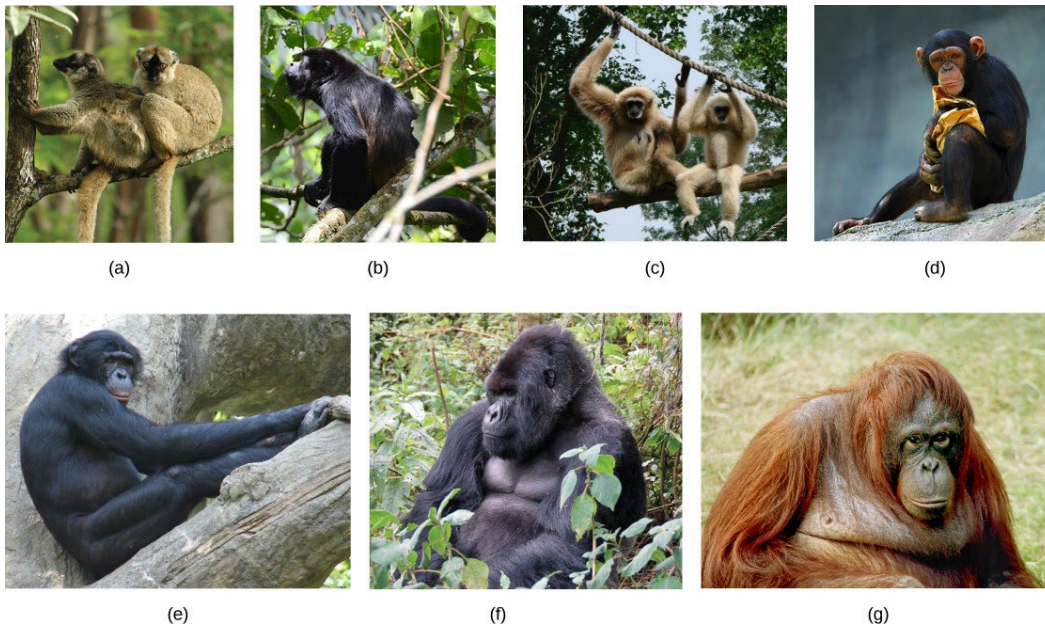
Figure 5.41 Primates can be divided into prosimians, such as the (a) lemur, and anthropoids. Anthropoids include monkeys, such as the (b) howler monkey; lesser apes, such as the (c) gibbon; and great apes, such as the (d) chimpanzee, (e) bonobo, (f) gorilla, and (g) orangutan. (credit a: modification of work by Frank Vassen; credit b: modification of work by Xavi Talleda; credit

d: modification of work by Aaron Logan; credit e: modification of work by Trisha Shears; credit f: modification of work by Dave Proffer; credit g: modification of work by Julie Langford)

Summary

Features of the Animal Kingdom

Animals constitute a diverse kingdom of organisms. Although animals range in complexity from simple sea sponges to human beings, most members share certain features. Animals are eukaryotic, multicellular, heterotrophic organisms that ingest their food and usually develop into motile creatures with a fixed body plan. Most members of the animal kingdom have differentiated tissues of four main classes—nervous, muscular, connective, and epithelial—that are specialized to perform different functions. Most animals reproduce sexually, leading to a developmental sequence that is relatively similar across the animal kingdom.

Organisms in the animal kingdom are classified based on their body morphology and development. True animals are divided into those with radial versus bilateral symmetry. Animals with three germ layers, called triploblasts, are further characterized by the presence or absence of an internal body cavity called a coelom. Animals with a body cavity may be either coelomates or pseudocoelomates, depending on which tissue gives rise to the coelom. Coelomates are further divided into two groups called protostomes and deuterostomes, based on a number of developmental characteristics.

Sponges and Cnidarians

Animals included in phylum Porifera are parazoans and do not possess true tissues. These organisms show a simple organization. Sponges have multiple cell types that are geared toward executing various metabolic functions.

Cnidarians have outer and inner tissue layers sandwiching a noncellular mesoglea. Cnidarians possess a well-formed digestive system and carry out extracellular digestion. The cnidocyte is a specialized cell for delivering toxins to prey and predators. Cnidarians have separate sexes. They have a life cycle that involves morphologically distinct forms—medusoid and polypoid—at various stages in their life cycle.

Flatworms, Nematodes, and Arthropods

Flatworms are acoelomate, triploblastic animals. They lack circulatory and respiratory systems, and have a rudimentary excretory system. The digestive system is incomplete in most species. There are four traditional classes of flatworms, the largely free-living turbellarians, the ectoparasitic monogeneans, and the endoparasitic trematodes and cestodes. Trematodes have complex life cycles involving a secondary mollusk host and a primary host in which sexual reproduction takes place.

Cestodes, or tapeworms, infect the digestive systems of primary vertebrate hosts.

Nematodes are pseudocoelomate members of the clade Ecdysozoa. They have a complete digestive system and a pseudocoelomic body cavity. This phylum includes free-living as well as parasitic organisms. They include dioecious and hermaphroditic species. Nematodes have a poorly developed

excretory system. Embryonic development is external and proceeds through larval stages separated by molts.

Arthropods represent the most successful phylum of animals on Earth, in terms of number of species as well as the number of individuals. They are characterized by a segmented body and jointed appendages. In the basic body plan, a pair of appendages is present per body segment. Within the phylum, classification is based on mouthparts, number of appendages, and modifications of appendages.

Arthropods bear a chitinous exoskeleton. Gills, tracheae, and book lungs facilitate respiration. Embryonic development may include multiple larval stages.

Mollusks and Annelids

The phylum Mollusca is a large, mainly marine group of invertebrates. Mollusks show a variety of morphologies. Many mollusks secrete a calcareous shell for protection, but in other species, the shell is reduced or absent. Mollusks are protostomes. The dorsal epidermis in mollusks is modified to form the mantle, which encloses the mantle cavity and visceral organs. This cavity is distinct from the coelomic cavity, which the adult animal retains, surrounding the heart. Respiration is facilitated by gills known as ctenidia. A chitinous scraper called the radula is present in most mollusks. Mollusks are mostly dioecious and are divided into seven classes.

The phylum Annelida includes worm-like, segmented animals. Segmentation is both external and internal, which is called metamerism. Annelids are protostomes. The presence of chitinous hairs called chaetae is characteristic of most members. These animals have well-developed nervous and digestive

systems. Polychaete annelids have parapodia that participate in locomotion and respiration. Suckers are seen in the order Hirudinea. Breeding systems include separate sexes and hermaphroditism.

Echinoderms and Chordates

Echinoderms are deuterostome marine organisms. This phylum of animals bear a calcareous endoskeleton composed of ossicles covered by a spiny skin. Echinoderms possess a water-based circulatory system. The madreporite is the point of entry and exit for water for the water vascular system.

The characteristic features of Chordata are a notochord, a dorsal hollow nerve cord, pharyngeal slits, and a post-anal tail. Chordata contains two clades of invertebrates: Urochordata (tunicates) and Cephalochordata (lancelets), together with the vertebrates. Most tunicates live on the ocean floor and are suspension feeders. Lancelets are suspension feeders that feed on phytoplankton and other microorganisms.

Vertebrates

The earliest vertebrates that diverged from the invertebrate chordates were the jawless fishes. Hagfishes are eel-like scavengers that feed on dead invertebrates and other fishes. Lampreys are characterized by a toothed, funnel-like sucking mouth, and some species are parasitic on other fishes. Gnathostomes include the jawed fishes (cartilaginous and bony fishes) as well as all other tetrapods. Cartilaginous fishes include sharks, rays, skates, and ghost sharks. Bony fishes can be further divided into ray-finned and lobe-finned fishes.

As tetrapods, most amphibians are characterized by four well-developed limbs, although some species of salamanders and all caecilians are limbless. Amphibians have a moist, permeable skin used for cutaneous respiration. Amphibia can be divided into three clades: salamanders (Urodela), frogs (Anura), and caecilians (Apoda). The life cycle of amphibians consists of two distinct stages: the larval stage and metamorphosis to an adult stage.

The amniotes are distinguished from amphibians by the presence of a terrestrially adapted egg protected by amniotic membranes. The amniotes include reptiles, birds, and mammals. A key adaptation that permitted reptiles to live on land was the development of scaly skin. Reptilia includes four living clades: Crocodilia (crocodiles and alligators), Sphenodontia (tuataras), Squamata (lizards and snakes), and Testudines (turtles).

Birds are endothermic amniotes. Feathers act as insulation and allow for flight. Birds have pneumatic bones that are hollow rather than tissue-filled. Airflow through bird lungs travels in one direction. Birds evolved from dinosaurs.

Mammals have hair and mammary glands. Mammalian skin includes various secretory glands. Mammals are endothermic, like birds. There are three groups of mammals living today: monotremes, marsupials, and eutherians. Monotremes are unique among mammals as they lay eggs, rather than giving birth to live young. Eutherian mammals have a complex placenta.

There are 16 extant (living) orders of eutherian mammals. Humans are most closely related to Primates, all of which have adaptations for climbing trees, although not all species are arboreal. Other characteristics of primates are brains that are larger than those of other mammals, claws that have been modified into flattened nails, and typically one young per pregnancy, stereoscopic vision, and a trend toward holding the body upright. Primates are divided into two groups: prosimians and anthropoids.

Review Questions

1.   Cnidocytes are found in   .

a.   phylum Porifera

b.   phylum Nemertea

c.   phylum Nematoda

d.   phylum Cnidaria

2.   A mantle and mantle cavity are present in   .

a.   class Oligochaeta

b.   class Bivalvia

c.   class Polychaeta

d.   class Hirudinea

3.   Which of the following is not a member of the phylum Chordata?

a.   Cephalochordata

b.   Echinodermata

c.   Urochordata

d.   Vertebrata

4.   Squamata includes   .

a.   crocodiles and alligators

b.   turtles

c.   tuataras

d.   lizards and snakes

5.   Sudoriferous glands produce   .

a.   sweat

b.   lipids

c.   sebum

d.   milk

6.   Which of the following is a Monotreme?

a.   kangaroo

b.   koala

c.   bandicoot

d.   platypus

Chapter 6

Ecology

Introduction

Ecosystem ecology is an extension of organismal, population, and community ecology. The ecosystem comprises all the biotic components (living things) and abiotic components (non-living things) in a particular geographic area. Some of the abiotic components include air, water, soil, and climate. Ecosystem biologists study how nutrients and energy are stored and moved among organisms and the surrounding atmosphere, soil, and water.

Wild lupine and Karner blue butterflies live in an oak-pine barren habitat in portions of Indiana, Michigan, Minnesota, Wisconsin, and New York. This habitat is characterized by natural disturbance in the form of fire and nutrient-poor soils that are low in nitrogen—important factors in the distribution of the plants that live in this habitat. Researchers interested in ecosystem ecology study the importance of limited resources in this ecosystem and the movement of resources (such as nutrients) through the biotic and abiotic portions of the ecosystem. Researchers also examine how organisms have adapted to their ecosystem.

| Energy Flow through Ecosystems

By the end of this section, you will be able to:
- Describe the basic types of ecosystems on Earth
- Differentiate between food chains and food webs and recognize the importance of each
- Describe how organisms acquire energy in a food web and in associated food chains
- Explain how the efficiency of energy transfers between trophic levels effects ecosystem

An **ecosystem** is a community of living organisms and their abiotic (non-living) environment. Ecosystems can be small, such as the tide pools found near the rocky shores of many oceans, or large, such as those found in the tropical rainforest of the Amazon in Brazil (Figure 6.1).

(a)   (b)
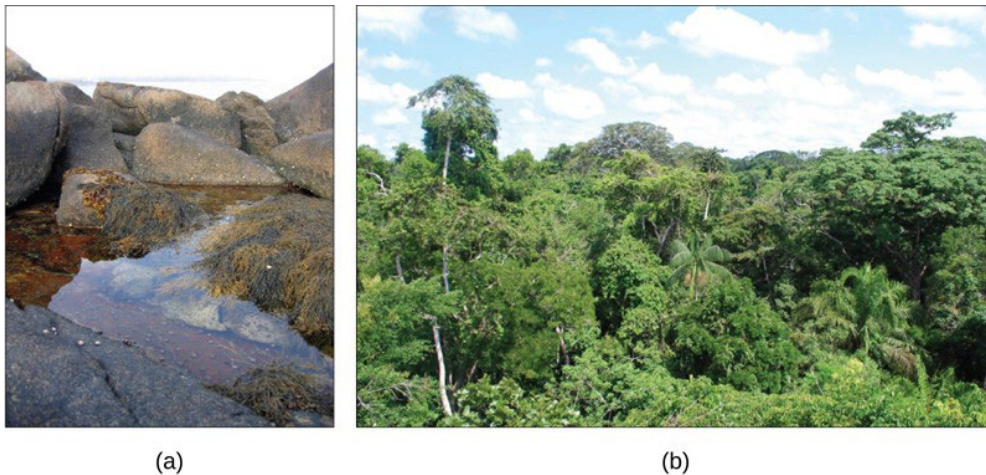
Figure 6.1 A (a) tidal pool ecosystem in Matinicus Island, Maine, is a small ecosystem, while the (b) Amazon rainforest in Brazil is a large ecosystem. (credit a: modification of work by Jim Kuhn; credit b: modification of work by Ivan Mlinaric)

There are three broad categories of ecosystems based on their general environment: freshwater, marine, and terrestrial. Within these three categories are individual ecosystem types based on the environmental habitat and organisms present.

Ecology of Ecosystems

Life in an ecosystem often involves competition for limited resources, which occurs both within a single species and between different species. Organisms compete for food, water, sunlight, space, and mineral nutrients. These resources provide the energy for metabolic processes and the matter to make up organisms' physical structures. Other critical factors influencing community dynamics are the components of its physical environment: a habitat's climate (seasons, sunlight, and rainfall), elevation, and geology. These can all be important environmental variables that determine which organisms can exist within a particular area.

Freshwater ecosystems are the least common, occurring on only 1.8 percent of Earth's surface. These systems comprise lakes, rivers, streams, and springs; they are quite diverse, and support a variety of animals, plants, fungi, protists and prokaryotes.

Marine ecosystems are the most common, comprising 75 percent of Earth's surface and consisting of three basic types: shallow ocean, deep ocean water, and deep ocean bottom. Shallow ocean ecosystems include extremely biodiverse coral reef ecosystems, yet the deep ocean water is known for large numbers of plankton and krill (small crustaceans) that support it. These two environments are especially important to aerobic respirators worldwide, as the phytoplankton perform 40 percent of all photosynthesis on Earth. Although not as diverse as the other two, deep ocean bottom ecosystems contain a wide variety of marine organisms. Such ecosystems exist even at depths where light is unable to penetrate through the water.

Terrestrial ecosystems, also known for their diversity, are grouped into large categories called biomes. A **biome** is a large-scale community of organisms, primarily defined on land by the dominant plant types that exist in geographic regions of the planet with similar climatic conditions. Examples of biomes

213

include tropical rainforests, savannas, deserts, grasslands, temperate forests, and tundras. Grouping these ecosystems into just a few biome categories obscures the great diversity of the individual ecosystems within them. For example, the saguaro cacti (*Carnegiea gigantean*) and other plant life in the Sonoran Desert, in the United States, are relatively diverse compared with the desolate rocky desert of Boa Vista, an island off the coast of Western Africa (Figure 6.2).

(a)   (b)

Figure 6.2 Desert ecosystems, like all ecosystems, can vary greatly. The desert in (a) Saguaro National Park, Arizona, has abundant plant life, while the rocky desert of (b) Boa Vista island, Cape Verde, Africa, is devoid of plant life. (credit a: modification of work by Jay Galvin; credit b: modification of work by Ingo Wölbern)

Ecosystems and Disturbance

Ecosystems are complex with many interacting parts. They are routinely exposed to various disturbances: changes in the environment that affect their compositions, such as yearly variations in rainfall and temperature. Many disturbances are a result of natural processes. For example, when lightning causes a forest fire and destroys part of a forest ecosystem, the ground is eventually populated with grasses, followed by bushes and shrubs, and later mature trees: thus, the forest is restored to its former state. This process is so universal that ecologists have given it a name—succession. The impact of environmental disturbances caused by human activities is now as significant as the changes wrought by natural processes. Human agricultural practices, air pollution, acid rain, global deforestation, overfishing, oil spills, and illegal dumping on land and into the ocean all have impacts on ecosystems.

Equilibrium is a dynamic state of an ecosystem in which, despite changes in species numbers and occurrence, biodiversity remains somewhat constant. In ecology, two parameters are used to measure changes in ecosystems: resistance and resilience. The ability of an ecosystem to remain at equilibrium in spite of disturbances is called **resistance**. The speed at which an ecosystem recovers equilibrium after being disturbed is called **resilience**. Ecosystem resistance and resilience are especially important when considering human impact. The nature of an ecosystem may change to such a degree that it can lose its resilience entirely. This process can lead to the complete destruction or irreversible altering of the ecosystem.

Food Chains and Food Webs

A **food chain** is a linear sequence of organisms through which nutrients and energy pass as one organism eats another; the levels in the food chain are producers, primary consumers, higher-level consumers, and finally decomposers. These levels are used to describe ecosystem structure and dynamics. There is a single path through a food chain. Each organism in a food chain occupies a specific **trophic level** (energy level), its position in the food chain or food web.

In many ecosystems, the base, or foundation, of the food chain consists of photosynthetic organisms (plants or phytoplankton), which are called **producers**. The organisms that consume the producers are herbivores: the **primary consumers**. **Secondary consumers** are usually carnivores that eat the primary consumers. **Tertiary consumers** are carnivores that eat other carnivores. Higher-level consumers feed on the next lower trophic levels, and so on, up to the organisms at the top of the food chain: the **apex consumers**. In the Lake Ontario food chain, shown in Figure 6.3, the Chinook salmon is the apex consumer at the top of this food chain.

214
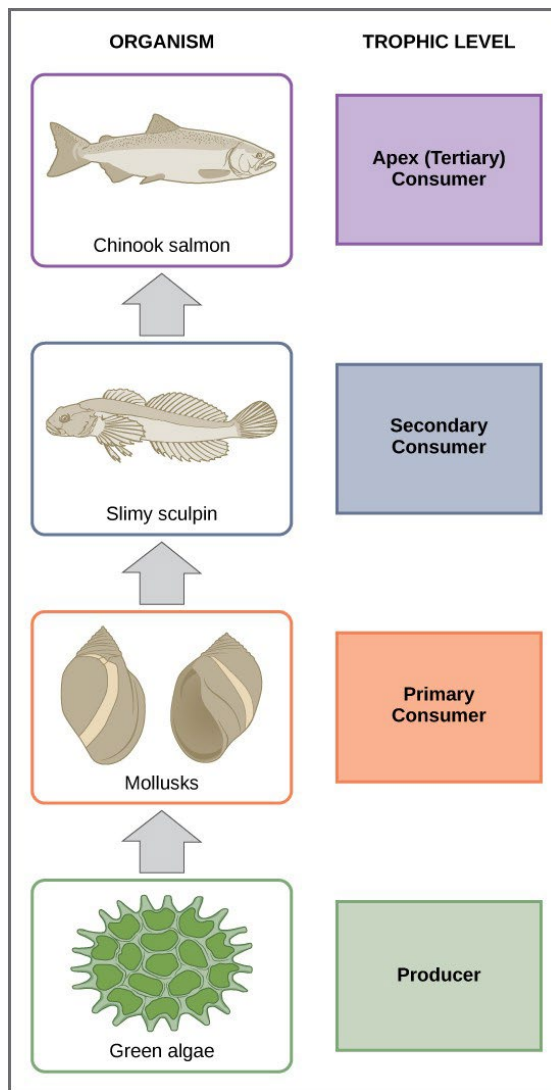
Figure 6.3 These are the trophic levels of a food chain in Lake Ontario at the United States–Canada border. Energy and nutrients flow from photosynthetic green algae at the base to the top of the food chain: the Chinook salmon. (credit: modification of work by National Oceanic and Atmospheric Administration/NOAA)

One major factor that limits the number of steps in a food chain is energy. Energy is lost at each trophic level and between trophic levels as heat and in the transfer to decomposers (Figure 6.4). Thus, after a limited number of trophic energy transfers, the amount of energy remaining in the food chain may not be great enough to support viable populations at yet a higher trophic level.

215
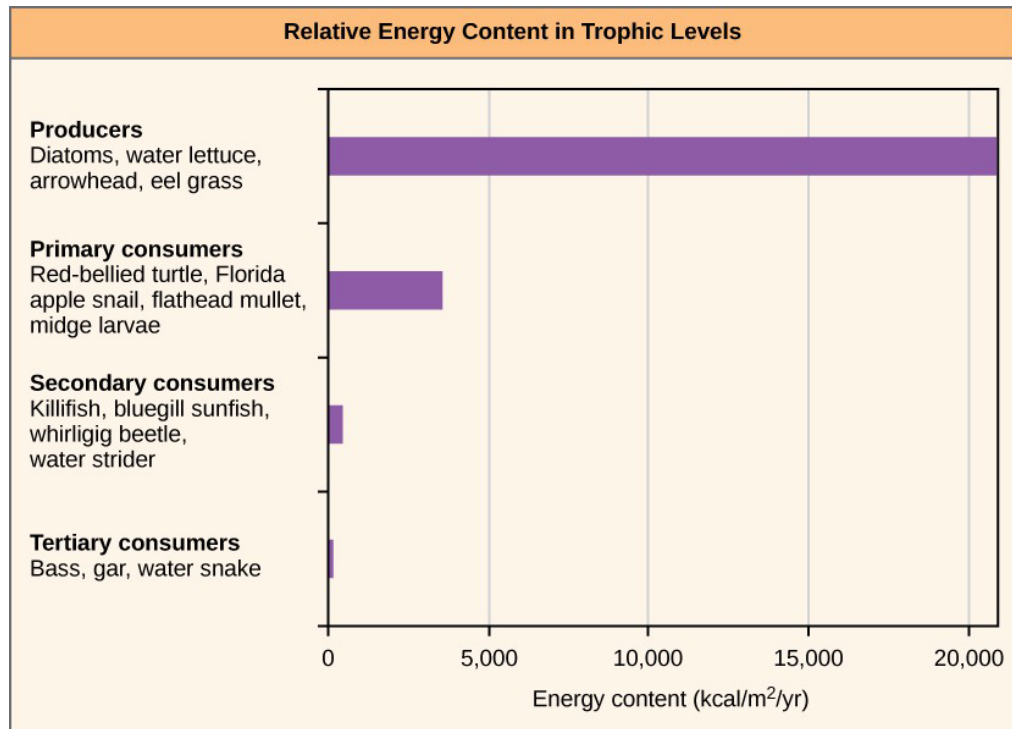
Figure 6.4 The relative energy in trophic levels in a Silver Springs, Florida, ecosystem is shown. Each trophic level has less energy available, and usually, but not always, supports a smaller mass of organisms at the next level.

There is a one problem when using food chains to describe most ecosystems. Even when all organisms are grouped into appropriate trophic levels, some of these organisms can feed on more than one trophic level; likewise, some of these organisms can also be fed on from multiple trophic levels. In addition, species feed on and are eaten by more than one species. In other words, the linear model of ecosystems, the food chain, is a hypothetical, overly simplistic representation of ecosystem structure. A holistic model—which includes all the interactions between different species and their complex interconnected relationships with each other and with the environment—is a more accurate and descriptive model for ecosystems. A **food web** is a concept that accounts for the multiple trophic (feeding) interactions between each species and the many species it may feed on, or that feed on it. In a food web, the several trophic connections between each species and the other species that interact with it may cross multiple trophic levels. The matter and energy movements of virtually all ecosystems are more accurately described by food webs (Figure 6.5).
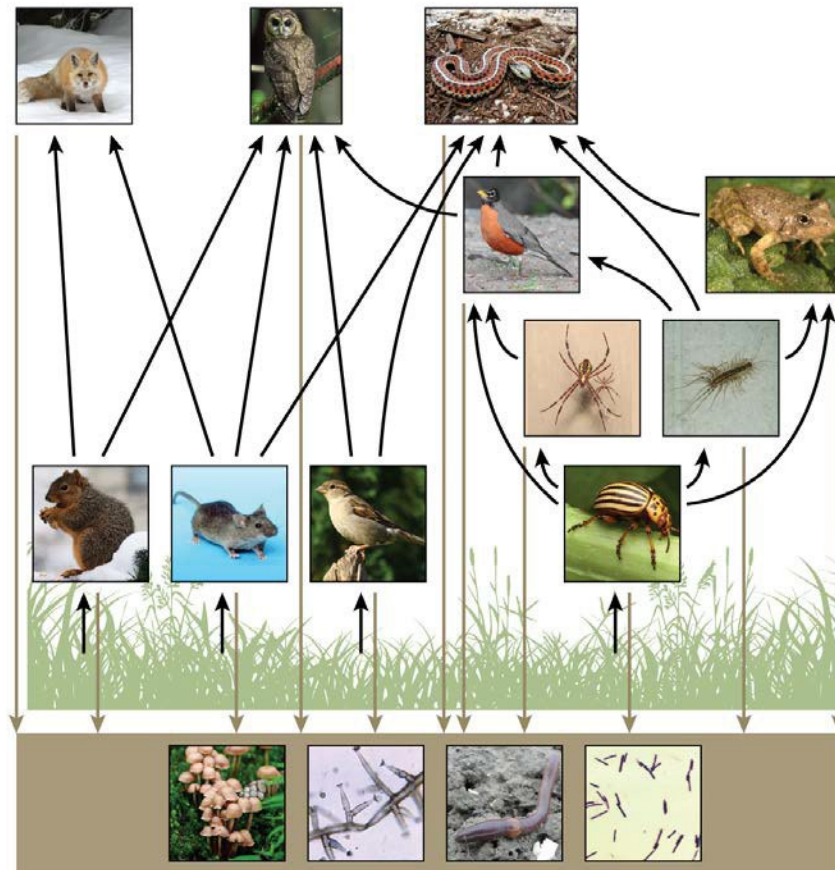
Figure 6.5 This food web shows the interactions between organisms across trophic levels. Arrows point from an organism that is consumed to the organism that consumes it. All the producers and consumers eventually become nourishment for the decomposers (fungi, mold, earthworms, and bacteria in the soil). (credit "fox": modification of work by Kevin Bacher, NPS; credit "owl": modification of work by John and Karen Hollingsworth, USFWS; credit "snake": modification of work by Steve Jurvetson; credit "robin": modification of work by Alan Vernon; credit "frog": modification of work by Alessandro Catenazzi; credit "spider": modification of work by "Sanba38"/Wikimedia Commons; credit "centipede": modification of work by "Bauerph"/Wikimedia Commons; credit "squirrel": modification of work by Dawn Huczek; credit "mouse": modification of work by NIGMS, NIH; credit "sparrow": modification of work by David Friel; credit "beetle": modification of work by Scott Bauer, USDA Agricultural Research Service; credit "mushrooms": modification of work by Chris Wee; credit "mold": modification of work by Dr. Lucille Georg, CDC; credit "earthworm": modification of work by Rob Hille; credit "bacteria": modification of work by Don Stalons, CDC)

Head to this **online interactive simulator (http://openstaxcollege.org/l/food_web)** to investigate food web function. In the *Interactive Labs* box, under Food Web, click **Step 1**. Read the instructions first, and then click **Step 2** for additional instructions. When you are ready to create a simulation, in the upper-right corner of the *Interactive Labs* box, click **OPEN SIMULATOR**.

Two general types of food webs are often shown interacting within a single ecosystem. A **grazing food web** has plants or other photosynthetic organisms at its base, followed by herbivores and various carnivores. A **detrital food web** consists of a base of organisms that feed on decaying organic matter (dead organisms), including decomposers (which break down dead and decaying organisms) and

217

detritivores (which consume organic detritus). These organisms are usually bacteria, fungi, and invertebrate animals that recycle organic material back into the biotic part of the ecosystem as they themselves are consumed by other organisms. As ecosystems require a method to recycle material from dead organisms, grazing food webs have an associated detrital food web. For example, in a meadow ecosystem, plants may support a grazing food web of different organisms, primary and other levels of consumers, while at the same time supporting a detrital food web of bacteria and fungi feeding off dead plants and animals. Simultaneously, a detrital food web can contribute energy to a grazing food web, as when a robin eats an earthworm.

How Organisms Acquire Energy in a Food Web

All living things require energy in one form or another. Energy is used by most complex metabolic pathways (usually in the form of ATP), especially those responsible for building large molecules from smaller compounds. Living organisms would not be able to assemble macromolecules (proteins, lipids, nucleic acids, and complex carbohydrates) from their monomers without a constant energy input.

Food-web diagrams illustrate how energy flows directionally through ecosystems. They can also indicate how efficiently organisms acquire energy, use it, and how much remains for use by other organisms of the food web. Energy is acquired by living things in two ways: autotrophs harness light or chemical energy and heterotrophs acquire energy through the consumption and digestion of other living or previously living organisms.

Photosynthetic and chemosynthetic organisms are **autotrophs**, which are organisms capable of synthesizing their own food (more specifically, capable of using inorganic carbon as a carbon source). Photosynthetic autotrophs (**photoautotrophs**) use sunlight as an energy source, and chemosynthetic autotrophs (**chemoautotrophs**) use inorganic molecules as an energy source. Autotrophs are critical for most ecosystems: they are the producer trophic level. Without these organisms, energy would not be available to other living organisms, and life itself would not be possible.

Photoautotrophs, such as plants, algae, and photosynthetic bacteria, are the energy source for a majority of the world's ecosystems. These ecosystems are often described by grazing and detrital food webs. Photoautotrophs harness the Sun's solar energy by converting it to chemical energy in the form of ATP (and NADP). The energy stored in ATP is used to synthesize complex organic molecules, such as glucose. The rate at which photosynthetic producers incorporate energy from the Sun is called **gross primary productivity**. However, not all of the energy incorporated by producers is available to the other organisms in the food web because producers must also grow and reproduce, which consumes energy. **Net primary productivity** is the energy that remains in the producers after accounting for these organisms' respiration and heat loss. The net productivity is then available to the primary consumers at the next trophic level.

Chemoautotrophs are primarily bacteria and archaea that are found in rare ecosystems where sunlight is not available, such as those associated with dark caves or hydrothermal vents at the bottom of the ocean (Figure 6.6). Many chemoautotrophs in hydrothermal vents use hydrogen sulfide (H_2S), which is released from the vents as a source of chemical energy; this allows them to synthesize complex organic molecules, such as glucose, for their own energy and, in turn, supplies energy to the rest of the ecosystem.

Figure 6.6 Swimming shrimp, a few squat lobsters, and hundreds of vent mussels are seen at a hydrothermal vent at the bottom of the ocean. As no sunlight penetrates to this depth, the ecosystem is supported by chemoautotrophic bacteria and organic material that sinks from the ocean's surface. This picture was taken in 2006 at the submerged NW Eifuku volcano off the coast of Japan by the National Oceanic and Atmospheric Administration (NOAA). The summit of this highly active volcano lies 1535 m below the surface.

Consequences of Food Webs: Biological Magnification

One of the most important consequences of ecosystem dynamics in terms of human impact is biomagnification. **Biomagnification** is the increasing concentration of persistent, toxic substances in organisms at each successive trophic level. These are substances that are fat soluble, not water soluble, and are stored in the fat reserves of each organism. Many substances have been shown to biomagnify, including classical studies with the pesticide dichlorodiphenyltrichloroethane (DDT), which were described in the 1960s bestseller, *Silent Spring* by Rachel Carson. DDT was a commonly used pesticide before its dangers to apex consumers, such as the bald eagle, became known. In aquatic ecosystems, organisms from each trophic level consumed many organisms in the lower level, which caused DDT to increase in birds (apex consumers) that ate fish. Thus, the birds accumulated sufficient amounts of DDT to cause fragility in their eggshells. This effect increased egg breakage during nesting and was shown to have devastating effects on these bird populations. The use of DDT was banned in the United States in the 1970s.

Other substances that biomagnify are polychlorinated biphenyls (PCB), which were used as coolant liquids in the United States until their use was banned in 1979, and heavy metals, such as mercury, lead, and cadmium. These substances are best studied in aquatic ecosystems, where predatory fish species accumulate very high concentrations of toxic substances that are at quite low concentrations in the environment and in producers. As illustrated in a study performed by the NOAA in the Saginaw Bay of Lake Huron of the North American Great Lakes (Figure 6.7), PCB concentrations increased from the producers of the ecosystem (phytoplankton) through the different trophic levels of fish species. The apex consumer, the walleye, has more than four times the amount of PCBs compared to phytoplankton. Also, based on results from other studies, birds that eat these fish may have PCB levels at least one order of magnitude higher than those found in the lake fish.

219
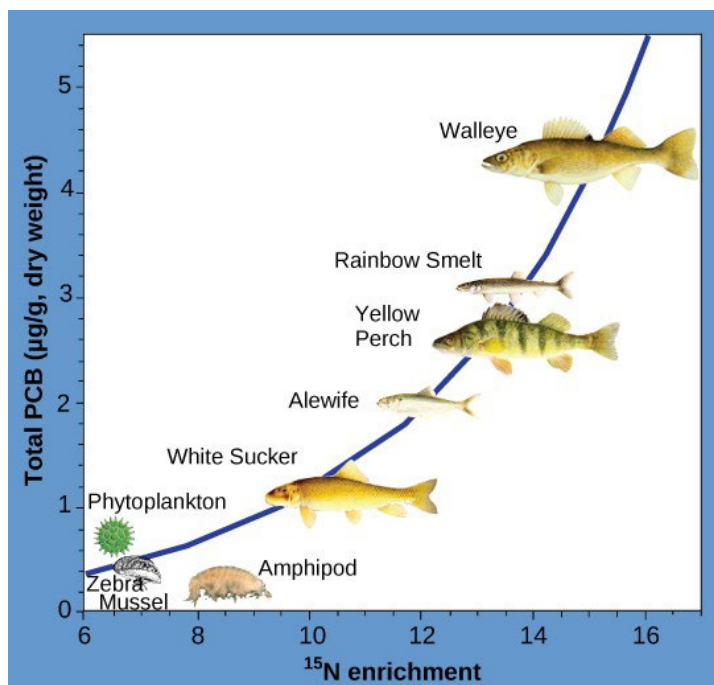
Figure 6.7 This chart shows the PCB concentrations found at the various trophic levels in the Saginaw Bay ecosystem of Lake Huron. Notice that the fish in the higher trophic levels accumulate more PCBs than those in lower trophic levels. (credit: Patricia Van Hoof, NOAA)

Other concerns have been raised by the biomagnification of heavy metals, such as mercury and cadmium, in certain types of seafood. The United States Environmental Protection Agency recommends that pregnant women and young children should not consume any swordfish, shark, king mackerel, or tilefish because of their high mercury content. These individuals are advised to eat fish low in mercury: salmon, shrimp, pollock, and catfish. Biomagnification is a good example of how ecosystem dynamics can affect our everyday lives, even influencing the food we eat.

| Biogeochemical Cycles

By the end of this section, you will be able to:

- Discuss the biogeochemical cycles of water, carbon, nitrogen, phosphorus, and sulfur

- Explain how human activities have impacted these cycles and the resulting potential consequences for Earth

Energy flows directionally through ecosystems, entering as sunlight (or inorganic molecules for chemoautotrophs) and leaving as heat during the transfers between trophic levels. Rather than flowing through an ecosystem, the matter that makes up living organisms is conserved and recycled. The six most common elements associated with organic molecules—carbon, nitrogen, hydrogen, oxygen, phosphorus, and sulfur—take a variety of chemical forms and may exist for long periods in the atmosphere, on land, in water, or beneath Earth's surface. Geologic processes, such as weathering, erosion, water drainage, and the subduction of the continental plates, all play a role in the cycling of elements on Earth. Because geology and chemistry have major roles in the study of this process, the recycling of inorganic matter between living organisms and their nonliving environment is called a **biogeochemical cycle**.

Water, which contains hydrogen and oxygen, is essential to all living processes. The **hydrosphere** is the area of Earth where water movement and storage occurs: as liquid water on the surface (rivers, lakes, oceans) and beneath the surface (groundwater) or ice, (polar ice caps and glaciers), and as water vapor in the atmosphere. Carbon is found in all organic macromolecules and is an important constituent of fossil fuels. Nitrogen is a major component of our nucleic acids and proteins and is critical to human agriculture. Phosphorus, a major component of nucleic acids, is one of the main ingredients (along with nitrogen) in artificial fertilizers used in agriculture, which has environmental impacts on our surface water. Sulfur, critical to the three-dimensional folding of proteins (as in disulfide binding), is released into the atmosphere by the burning of fossil fuels.

220

The cycling of these elements is interconnected. For example, the movement of water is critical for the leaching of nitrogen and phosphate into rivers, lakes, and oceans. The ocean is also a major reservoir for carbon. Thus, mineral nutrients are cycled, either rapidly or slowly, through the entire biosphere between the biotic and abiotic world and from one living organism to another.

CONCEPT in ACTION

Head to this **website (http://openstaxcollege.org/l/biogeochemical)** to learn more about biogeochemical cycles.

The Water Cycle

Water is essential for all living processes. The human body is more than one-half water and human cells are more than 70 percent water. Thus, most land animals need a supply of fresh water to survive. Of the stores of water on Earth, 97.5 percent is salt water (Figure 6.8). Of the remaining water, 99 percent is locked as underground water or ice. Thus, less than one percent of fresh water is present in lakes and rivers. Many living things are dependent on this small amount of surface fresh water supply, a lack of which can have important effects on ecosystem dynamics. Humans, of course, have developed technologies to increase water availability, such as digging wells to harvest groundwater, storing rainwater, and using desalination to obtain drinkable water from the ocean. Although this pursuit of drinkable water has been ongoing throughout human history, the supply of fresh water continues to be a major issue in modern times.
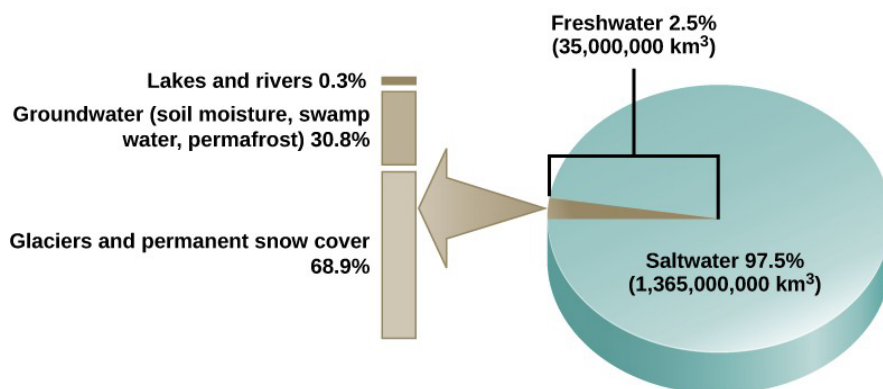
Figure 6.8 Only 2.5 percent of water on Earth is fresh water, and less than 1 percent of fresh water is easily accessible to living things.

The various processes that occur during the cycling of water are illustrated in Figure 6.9. The processes include the following:

- evaporation and sublimation
- condensation and precipitation
- subsurface water flow
- surface runoff and snowmelt
- streamflow

The water cycle is driven by the Sun's energy as it warms the oceans and other surface waters. This leads to evaporation (water to water vapor) of liquid surface water and sublimation (ice to water vapor) of frozen water, thus moving large amounts of water into the atmosphere as water vapor. Over time, this water vapor condenses into clouds as liquid or frozen droplets and eventually leads to precipitation (rain or snow), which returns water to Earth's surface. Rain reaching Earth's surface may evaporate again, flow over the surface, or percolate into the ground. Most easily observed is surface runoff: the flow of

221

fresh water either from rain or melting ice. Runoff can make its way through streams and lakes to the oceans or flow directly to the oceans themselves.

In most natural terrestrial environments rain encounters vegetation before it reaches the soil surface. A significant percentage of water evaporates immediately from the surfaces of plants. What is left reaches the soil and begins to move down. Surface runoff will occur only if the soil becomes saturated with water in a heavy rainfall. Most water in the soil will be taken up by plant roots. The plant will use some of this water for its own metabolism, and some of that will find its way into animals that eat the plants, but much of it will be lost back to the atmosphere through a process known as evapotranspiration. Water enters the vascular system of the plant through the roots and evaporates, or transpires, through the stomata of the leaves. Water in the soil that is not taken up by a plant and that does not evaporate is able to percolate into the subsoil and bedrock. Here it forms groundwater.

Groundwater is a significant reservoir of fresh water. It exists in the pores between particles in sand and gravel, or in the fissures in rocks. Shallow groundwater flows slowly through these pores and fissures and eventually finds its way to a stream or lake where it becomes a part of the surface water again. Streams do not flow because they are replenished from rainwater directly; they flow because there is a constant inflow from groundwater below. Some groundwater is found very deep in the bedrock and can persist there for millennia. Most groundwater reservoirs, or aquifers, are the source of drinking or irrigation water drawn up through wells. In many cases these aquifers are being depleted faster than they are being replenished by water percolating down from above.

Rain and surface runoff are major ways in which minerals, including carbon, nitrogen, phosphorus, and sulfur, are cycled from land to water. The environmental effects of runoff will be discussed later as these cycles are described.
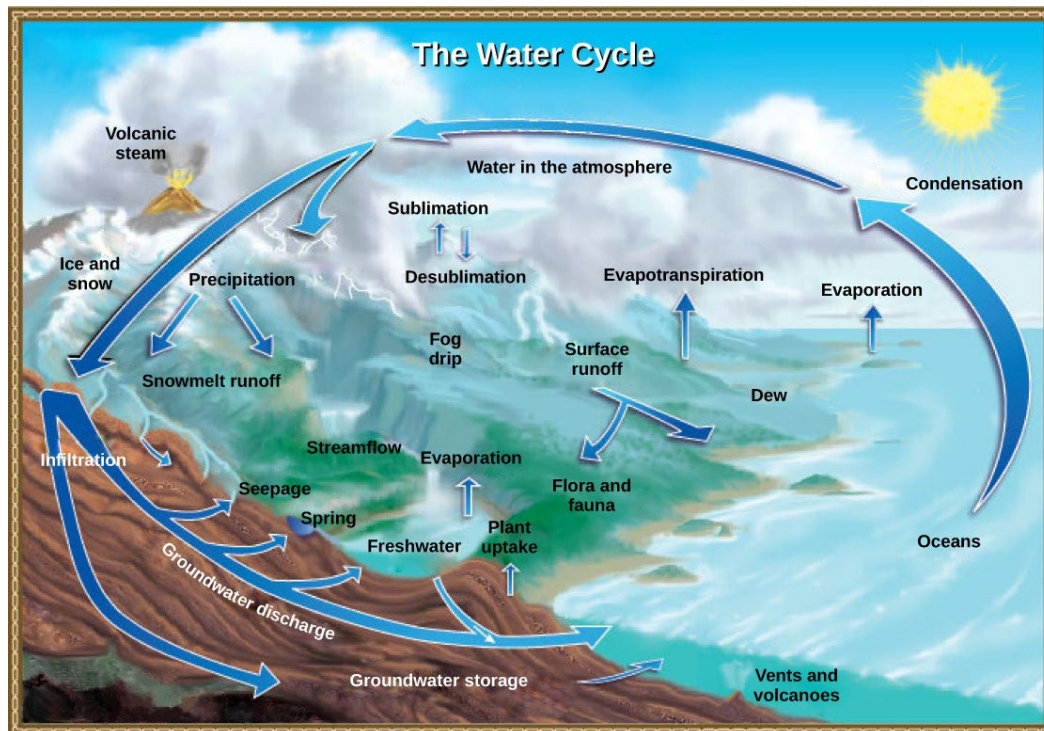
Figure 6.9 Water from the land and oceans enters the atmosphere by evaporation or sublimation, where it condenses into clouds and falls as rain or snow. Precipitated water may enter freshwater bodies or infiltrate the soil. The cycle is complete when surface or groundwater reenters the ocean. (credit: modification of work by John M. Evans and Howard Perlman, USGS)

The Carbon Cycle

Carbon is the fourth most abundant element in living organisms. Carbon is present in all organic molecules, and its role in the structure of macromolecules is of primary importance to living organisms. Carbon compounds contain energy, and many of these compounds from plants and algae have remained stored as fossilized carbon, which humans use as fuel. Since the 1800s, the use of fossil fuels has accelerated. As global demand for Earth's limited fossil fuel supplies has risen since the beginning of the Industrial Revolution, the amount of carbon dioxide in our atmosphere has increased as the fuels

222

are burned. This increase in carbon dioxide has been associated with climate change and is a major environmental concern worldwide.

The carbon cycle is most easily studied as two interconnected subcycles: one dealing with rapid carbon exchange among living organisms and the other dealing with the long-term cycling of carbon through geologic processes. The entire carbon cycle is shown in Figure 6.10.
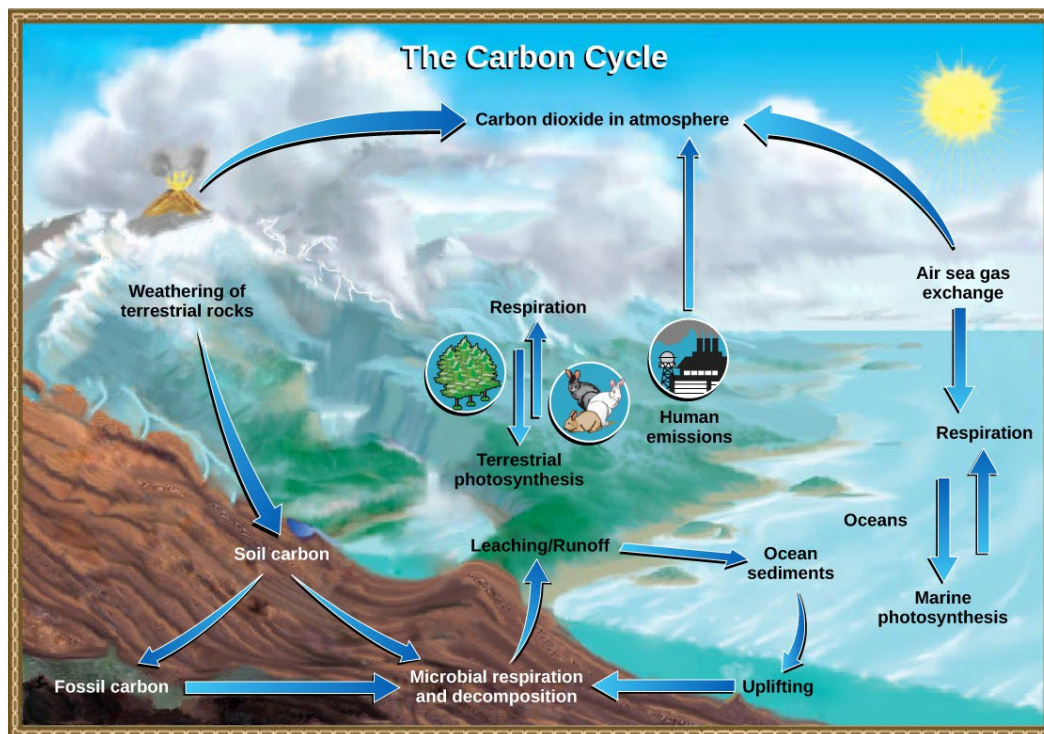
Figure 6.10 Carbon dioxide gas exists in the atmosphere and is dissolved in water. Photosynthesis converts carbon dioxide gas to organic carbon, and respiration cycles the organic carbon back into carbon dioxide gas. Long-term storage of organic carbon occurs when matter from living organisms is buried deep underground and becomes fossilized. Volcanic activity and, more recently, human emissions bring this stored carbon back into the carbon cycle. (credit: modification of work by John M. Evans and Howard Perlman, USGS)

The Biological Carbon Cycle

Living organisms are connected in many ways, even between ecosystems. A good example of this connection is the exchange of carbon between heterotrophs and autotrophs within and between ecosystems by way of atmospheric carbon dioxide. Carbon dioxide is the basic building block that autotrophs use to build multi-carbon, high-energy compounds, such as glucose. The energy harnessed from the Sun is used by these organisms to form the covalent bonds that link carbon atoms together. These chemical bonds store this energy for later use in the process of respiration. Most terrestrial autotrophs obtain their carbon dioxide directly from the atmosphere, while marine autotrophs acquire it in the dissolved form (carbonic acid, HCO_3^-). However the carbon dioxide is acquired, a byproduct of fixing carbon in organic compounds is oxygen. Photosynthetic organisms are responsible for maintaining approximately 21 percent of the oxygen content of the atmosphere that we observe today.

The partners in biological carbon exchange are the heterotrophs (especially the primary consumers, largely herbivores). Heterotrophs acquire the high-energy carbon compounds from the autotrophs by consuming them and breaking them down by respiration to obtain cellular energy, such as ATP. The most efficient type of respiration, aerobic respiration, requires oxygen obtained from the atmosphere or dissolved in water. Thus, there is a constant exchange of oxygen and carbon dioxide between the autotrophs (which need the carbon) and the heterotrophs (which need the oxygen). Autotrophs also respire and consume the organic molecules they form: using oxygen and releasing carbon dioxide. They release more oxygen gas as a waste product of photosynthesis than they use for their own respiration; therefore, there is excess available for the respiration of other aerobic organisms. Gas exchange through the atmosphere and water is one way that the carbon cycle connects all living organisms on Earth.

The Biogeochemical Carbon Cycle

The movement of carbon through land, water, and air is complex, and, in many cases, it occurs much more slowly geologically than the movement between living organisms. Carbon is stored for long

223

periods in what are known as carbon reservoirs, which include the atmosphere, bodies of liquid water (mostly oceans), ocean sediment, soil, rocks (including fossil fuels), and Earth's interior.

As stated, the atmosphere is a major reservoir of carbon in the form of carbon dioxide that is essential to the process of photosynthesis. The level of carbon dioxide in the atmosphere is greatly influenced by the reservoir of carbon in the oceans. The exchange of carbon between the atmosphere and water reservoirs influences how much carbon is found in each, and each one affects the other reciprocally. Carbon dioxide (CO_2) from the atmosphere dissolves in water and, unlike oxygen and nitrogen gas, reacts with water molecules to form ionic compounds. Some of these ions combine with calcium ions in the seawater to form calcium carbonate ($CaCO_3$), a major component of the shells of marine organisms. These organisms eventually form sediments on the ocean floor. Over geologic time, the calcium carbonate forms limestone, which comprises the largest carbon reservoir on Earth.

On land, carbon is stored in soil as organic carbon as a result of the decomposition of living organisms or from weathering of terrestrial rock and minerals. Deeper under the ground, at land and at sea, are fossil fuels, the anaerobically decomposed remains of plants that take millions of years to form. Fossil fuels are considered a non-renewable resource because their use far exceeds their rate of formation. A **non-renewable resource** is either regenerated very slowly or not at all. Another way for carbon to enter the atmosphere is from land (including land beneath the surface of the ocean) by the eruption of volcanoes and other geothermal systems. Carbon sediments from the ocean floor are taken deep within Earth by the process of **subduction**: the movement of one tectonic plate beneath another. Carbon is released as carbon dioxide when a volcano erupts or from volcanic hydrothermal vents.

Carbon dioxide is also added to the atmosphere by the animal husbandry practices of humans. The large number of land animals raised to feed Earth's growing human population results in increased carbon-dioxide levels in the atmosphere caused by their respiration. This is another example of how human activity indirectly affects biogeochemical cycles in a significant way. Although much of the debate about the future effects of increasing atmospheric carbon on climate change focuses on fossils fuels, scientists take natural processes, such as volcanoes, plant growth, soil carbon levels, and respiration, into account as they model and predict the future impact of this increase.

The Nitrogen Cycle

Getting nitrogen into living organisms is difficult. Plants and phytoplankton are not equipped to incorporate nitrogen from the atmosphere (where it exists as tightly bonded, triple covalent N_2) even though this molecule comprises approximately 78 percent of the atmosphere. Nitrogen enters the living world through free-living and symbiotic bacteria, which incorporate nitrogen into their macromolecules through specialized biochemical pathways leading to nitrogen fixation. Cyanobacteria live in most aquatic ecosystems where sunlight is present; they play a key role in nitrogen fixation. Cyanobacteria are able to "fix" nitrogen (from nitrogen gas) into ammonia (NH_3) that can be incorporated into the macromolecules of the organism. *Rhizobium* bacteria also fix nitrogen and live symbiotically in the root nodules of legumes (such as peas, beans, and peanuts) and provide them with the organic nitrogen they need. Free-living bacteria, such as *Azotobacter*, are also able to fix nitrogen.

Organic nitrogen is especially important to the study of ecosystem dynamics since many ecosystem processes, such as primary production and decomposition, are limited by the available supply of nitrogen. As shown in Figure 6.11, the nitrogen that enters living systems by nitrogen fixation is eventually converted from organic nitrogen back into nitrogen gas by bacteria. This process occurs in three steps in terrestrial systems: ammonification, nitrification, and denitrification. First, the ammonification process converts nitrogenous waste from living animals or from the remains of dead animals into ammonium (NH_4^+) by certain bacteria and fungi. Second, this ammonium is then converted to nitrites (NO_2^-) by nitrifying bacteria, such as *Nitrosomonas*, through nitrification. Subsequently, nitrites are converted to nitrates (NO_3^-) by similar organisms. Lastly, the process of denitrification occurs, whereby bacteria, such as *Pseudomonas* and *Clostridium*, convert the nitrates into nitrogen gas, thus allowing it to re-enter the atmosphere.
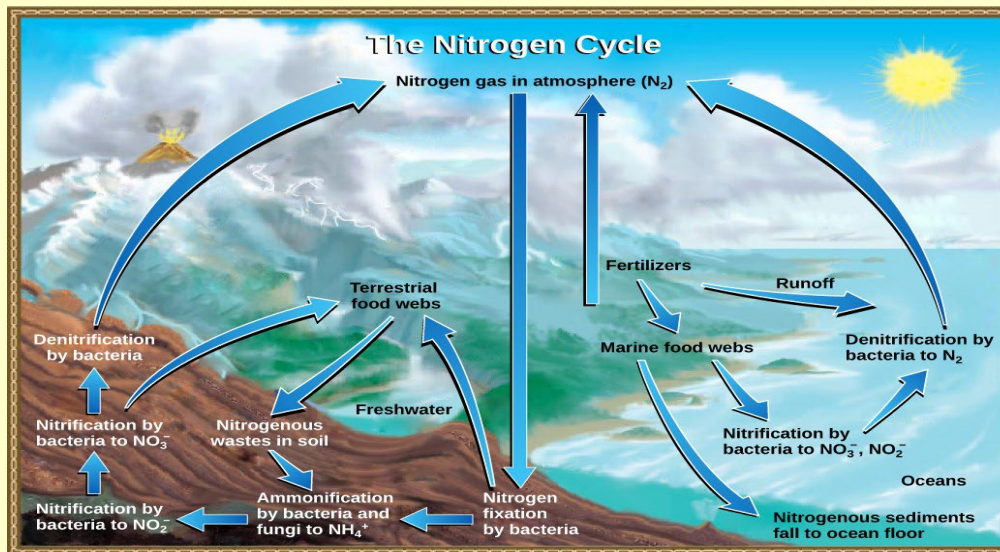
Figure 6.11 Nitrogen enters the living world from the atmosphere through nitrogen-fixing bacteria. This nitrogen and nitrogenous waste from animals is then processed back into gaseous nitrogen by soil bacteria, which also supply terrestrial food webs with the organic nitrogen they need. (credit: modification of work by John M. Evans and Howard Perlman, USGS)

Human activity can release nitrogen into the environment by two primary means: the combustion of fossil fuels, which releases different nitrogen oxides, and by the use of artificial fertilizers (which contain nitrogen and phosphorus compounds) in agriculture, which are then washed into lakes, streams, and rivers by surface runoff. Atmospheric nitrogen (other than N_2) is associated with several effects on Earth's ecosystems including the production of acid rain (as nitric acid, HNO_3) and greenhouse gas effects (as nitrous oxide, N_2O), potentially causing climate change. A major effect from fertilizer runoff is saltwater and freshwater **eutrophication**, a process whereby nutrient runoff causes the overgrowth of algae and a number of consequential problems.

A similar process occurs in the marine nitrogen cycle, where the ammonification, nitrification, and denitrification processes are performed by marine bacteria and archaea. Some of this nitrogen falls to the ocean floor as sediment, which can then be moved to land in geologic time by uplift of Earth's surface, and thereby incorporated into terrestrial rock. Although the movement of nitrogen from rock directly into living systems has been traditionally seen as insignificant compared with nitrogen fixed from the atmosphere, a recent study showed that this process may indeed be significant and should be included in any study of the global nitrogen cycle.[1]

1. Scott L. Morford, Benjamin Z. Houlton, and Randy A. Dahlgren, "Increased Forest Ecosystem Carbon and Nitrogen Storage from Nitrogen Rich Bedrock," *Nature* 477, no. 7362 (2011): 78–81.

225

The Phosphorus Cycle

Phosphorus is an essential nutrient for living processes; it is a major component of nucleic acids and phospholipids, and, as calcium phosphate, makes up the supportive components of our bones. Phosphorus is often the limiting nutrient (necessary for growth) in aquatic, particularly freshwater, ecosystems.

Phosphorus occurs in nature as the phosphate ion (PO_4^{3-}). In addition to phosphate runoff as a result of human activity, natural surface runoff occurs when it is leached from phosphate-containing rock by weathering, thus sending phosphates into rivers, lakes, and the ocean. This rock has its origins in the ocean. Phosphate-containing ocean sediments form primarily from the bodies of ocean organisms and from their excretions. However, volcanic ash, aerosols, and mineral dust may also be significant phosphate sources. This sediment then is moved to land over geologic time by the uplifting of Earth's surface. (Figure 6.12)

Phosphorus is also reciprocally exchanged between phosphate dissolved in the ocean and marine organisms. The movement of phosphate from the ocean to the land and through the soil is extremely slow, with the average phosphate ion having an oceanic residence time between 20,000 and 100,000 years.
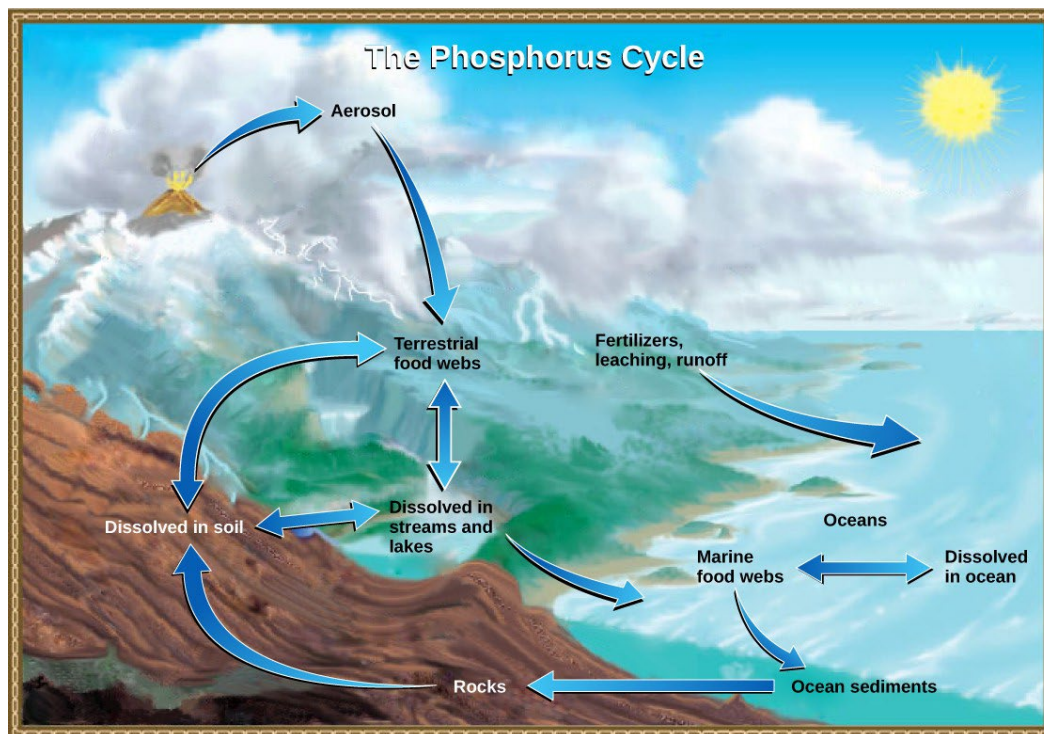
Figure 6.12 In nature, phosphorus exists as the phosphate ion (PO_4^{3-}). Weathering of rocks and volcanic activity releases phosphate into the soil, water, and air, where it becomes available to terrestrial food webs. Phosphate enters the oceans in surface runoff, groundwater flow, and river flow. Phosphate dissolved in ocean water cycles into marine food webs. Some phosphate from the marine food webs falls to the ocean floor, where it forms sediment. (credit: modification of work by John M. Evans and Howard Perlman, USGS)

Excess phosphorus and nitrogen that enter these ecosystems from fertilizer runoff and from sewage cause excessive growth of algae. The subsequent death and decay of these organisms depletes dissolved oxygen, which leads to the death of aquatic organisms, such as shellfish and finfish. This process is responsible for dead zones in lakes and at the mouths of many major rivers and for massive fish kills, which often occur during the summer months (see Figure 6.13).
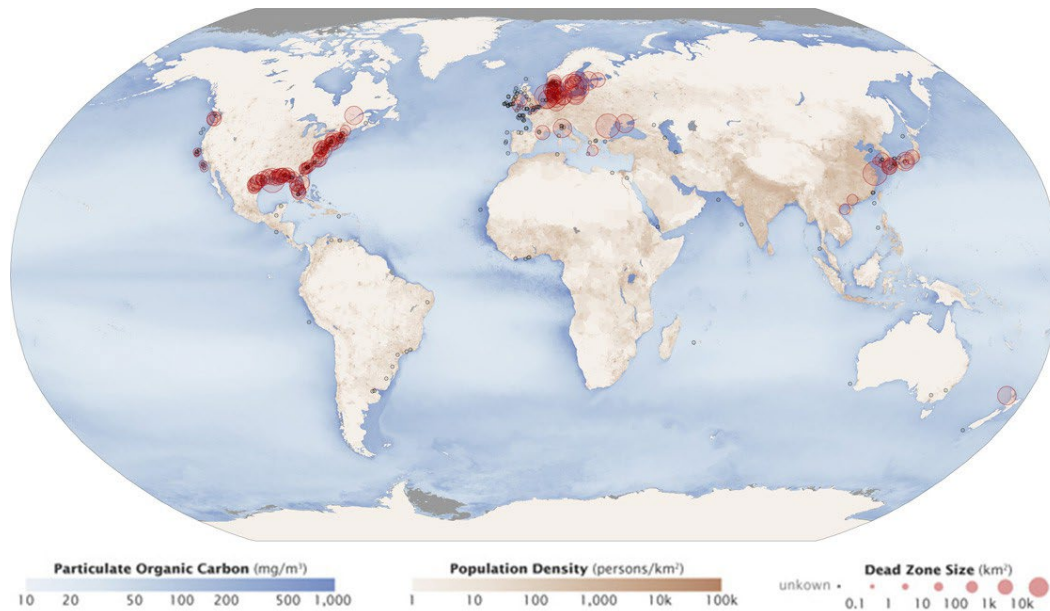
Figure 6.13 Dead zones occur when phosphorus and nitrogen from fertilizers cause excessive growth of microorganisms, which depletes oxygen and kills fauna. Worldwide, large dead zones are found in areas of high population density. (credit: Robert Simmon, Jesse Allen, NASA Earth Observatory)

A **dead zone** is an area in lakes and oceans near the mouths of rivers where large areas are periodically depleted of their normal flora and fauna; these zones can be caused by eutrophication, oil spills, dumping toxic chemicals, and other human activities. The number of dead zones has increased for several years, and more than 400 of these zones were present as of 2008. One of the worst dead zones is off the coast of the United States in the Gulf of Mexico: fertilizer runoff from the Mississippi River basin created a dead zone of over 8,463 square miles. Phosphate and nitrate runoff from fertilizers also negatively affect several lake and bay ecosystems including the Chesapeake Bay in the eastern United States.

- Download for free at http://openstaxcollege.org/textbooks/concepts-of-biology/get.

Chesapeake Bay

(a)   (b)
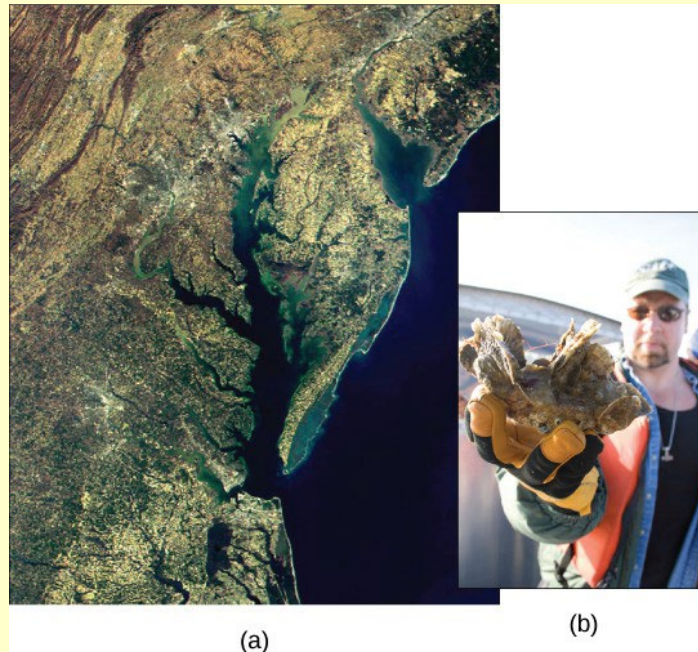
Figure 6.14 This (a) satellite image shows the Chesapeake Bay, an ecosystem affected by phosphate and nitrate runoff. A (b) member of the Army Corps of Engineers holds a clump of oysters being used as a part of the oyster restoration effort in the bay. (credit a: modification of work by NASA/MODIS; credit b: modification of work by U.S. Army)

The Chesapeake Bay (Figure 6.14**a**) is one of the most scenic areas on Earth; it is now in distress and is recognized as a case study of a declining ecosystem. In the 1970s, the Chesapeake Bay was one of the first aquatic ecosystems to have identified dead zones, which continue to kill many fish and bottom-dwelling species such as clams, oysters, and worms. Several species have declined in the Chesapeake Bay because surface water runoff contains excess nutrients from artificial fertilizer use on land. The source of the fertilizers (with high nitrogen and phosphate content) is not limited to agricultural practices. There are many nearby urban areas and more than 150 rivers and streams empty into the bay that are carrying fertilizer runoff from lawns and gardens. Thus, the decline of the Chesapeake Bay is a complex issue and requires the cooperation of industry, agriculture, and individual homeowners.

Of particular interest to conservationists is the oyster population (Figure 6.14**b**); it is estimated that more than 200,000 acres of oyster reefs existed in the bay in the 1700s, but that number has now declined to only 36,000 acres. Oyster harvesting was once a major industry for Chesapeake Bay, but it declined 88 percent between 1982 and 2007. This decline was caused not only by fertilizer runoff and dead zones, but also because of overharvesting. Oysters require a certain minimum population density because they must be in close proximity to reproduce. Human activity has altered the oyster population and locations, thus greatly disrupting the ecosystem.

The restoration of the oyster population in the Chesapeake Bay has been ongoing for several years with mixed success. Not only do many people find oysters good to eat, but the oysters also clean up the bay. They are filter feeders, and as they eat, they clean the water around them. Filter feeders eat by pumping a continuous stream of water over finely divided appendages (gills in the case of oysters) and capturing prokaryotes, plankton, and fine organic particles in their mucus. In the 1700s, it was estimated that it took only a few days for the oyster population to filter the entire volume of the bay. Today, with the changed water conditions, it is estimated that the present population would take nearly a year to do the same job.

228

Restoration efforts have been ongoing for several years by non-profit organizations such as the Chesapeake Bay Foundation. The restoration goal is to find a way to increase population density so the oysters can reproduce more efficiently. Many disease-resistant varieties (developed at the Virginia Institute of Marine Science for the College of William and Mary) are now available and have been used in the construction of experimental oyster reefs. Efforts by Virginia and Delaware to clean and restore the bay have been hampered because much of the pollution entering the bay comes from other states, which emphasizes the need for interstate cooperation to gain successful restoration.

The new, hearty oyster strains have also spawned a new and economically viable industry—oyster aquaculture—which not only supplies oysters for food and profit, but also has the added benefit of cleaning the bay.

The Sulfur Cycle

Sulfur is an essential element for the macromolecules of living things. As part of the amino acid cysteine, it is involved in the formation of proteins. As shown in Figure 6.15, sulfur cycles between the oceans, land, and atmosphere. Atmospheric sulfur is found in the form of sulfur dioxide (SO_2), which enters the atmosphere in three ways: first, from the decomposition of organic molecules; second, from volcanic activity and geothermal vents; and, third, from the burning of fossil fuels by humans.
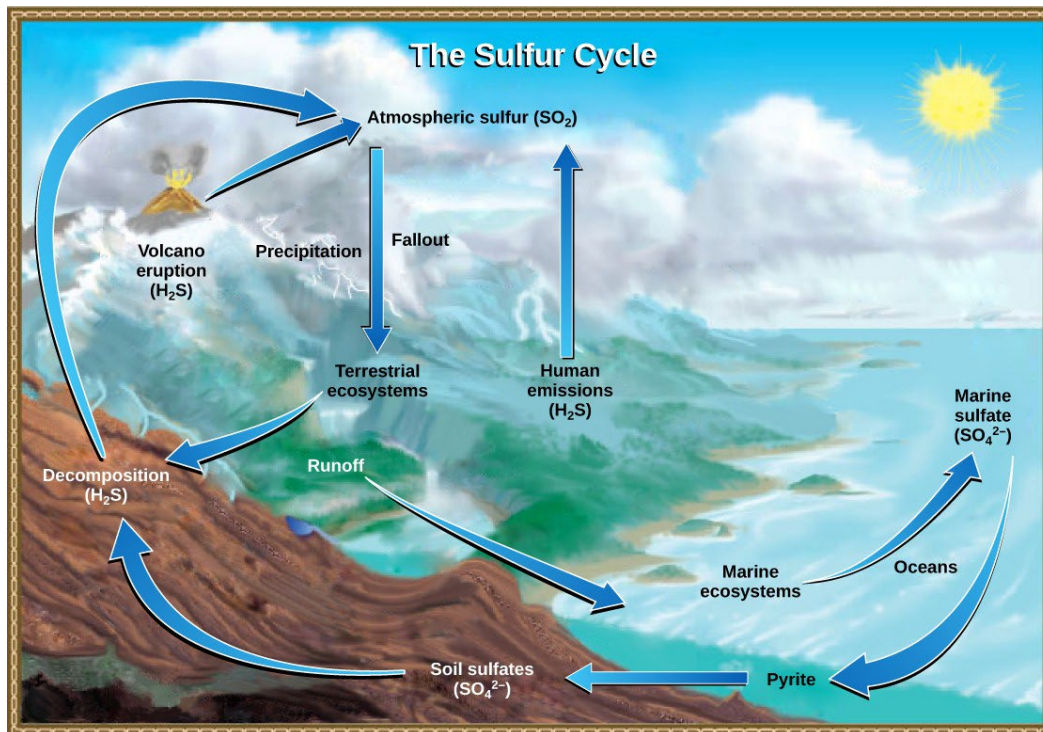
Figure 6.15 Sulfur dioxide from the atmosphere becomes available to terrestrial and marine ecosystems when it is dissolved in precipitation as weak sulfuric acid or when it falls directly to Earth as fallout. Weathering of rocks also makes sulfates available to terrestrial ecosystems. Decomposition of living organisms returns sulfates to the ocean, soil, and atmosphere. (credit: modification of work by John M. Evans and Howard Perlman, USGS)

On land, sulfur is deposited in four major ways: precipitation, direct fallout from the atmosphere, rock weathering, and geothermal vents (Figure 6.16). Atmospheric sulfur is found in the form of sulfur dioxide (SO_2), and as rain falls through the atmosphere, sulfur is dissolved in the form of weak sulfuric acid (H_2SO_4). Sulfur can also fall directly from the atmosphere in a process called **fallout**. Also, as sulfur-containing rocks weather, sulfur is released into the soil. These rocks originate from ocean sediments that are moved to land by the geologic uplifting of ocean sediments. Terrestrial ecosystems can then make use of these soil sulfates (SO_4^{2-}), which enter the food web by being taken up by plant roots. When these plants decompose and die, sulfur is released back into the atmosphere as hydrogen sulfide (H_2S) gas.

229

Figure 6.16 At this sulfur vent in Lassen Volcanic National Park in northeastern California, the yellowish sulfur deposits are visible near the mouth of the vent. (credit: "Calbear22"/Wikimedia Commons)

Sulfur enters the ocean in runoff from land, from atmospheric fallout, and from underwater geothermal vents. Some ecosystems rely on chemoautotrophs using sulfur as a biological energy source. This sulfur then supports marine ecosystems in the form of sulfates.

Human activities have played a major role in altering the balance of the global sulfur cycle. The burning of large quantities of fossil fuels, especially from coal, releases larger amounts of hydrogen sulfide gas into the atmosphere. As rain falls through this gas, it creates the phenomenon known as acid rain, which damages the natural environment by lowering the pH of lakes, thus killing many of the resident plants and animals. **Acid rain** is corrosive rain caused by rainwater falling to the ground through sulfur dioxide gas, turning it into weak sulfuric acid, which causes damage to aquatic ecosystems. Acid rain also affects the man-made environment through the chemical degradation of buildings. For example, many marble monuments, such as the Lincoln Memorial in Washington, DC, have suffered significant damage from acid rain over the years. These examples show the wide-ranging effects of human activities on our environment and the challenges that remain for our future.

| Terrestrial Biomes

By the end of this section, you will be able to:

- Identify the two major abiotic factors that determine the type of terrestrial biome in an area
- Recognize distinguishing characteristics of each of the eight major terrestrial biomes

Earth's biomes can be either terrestrial or aquatic. Terrestrial biomes are based on land, while aquatic biomes include both ocean and freshwater biomes. The eight major terrestrial biomes on Earth are each distinguished by characteristic temperatures and amount of precipitation. Annual totals and fluctuations of precipitation affect the kinds of vegetation and animal life that can exist in broad geographical regions. Temperature variation on a daily and seasonal basis is also important for predicting the geographic distribution of a biome. Since a biome is defined by climate, the same biome can occur in geographically distinct areas with similar climates (Figure 6.17). There are also large areas on Antarctica, Greenland, and in mountain ranges that are covered by permanent glaciers and support very little life. Strictly speaking, these are not considered biomes and in addition to extremes of cold, they are also often deserts with very low precipitation.

230
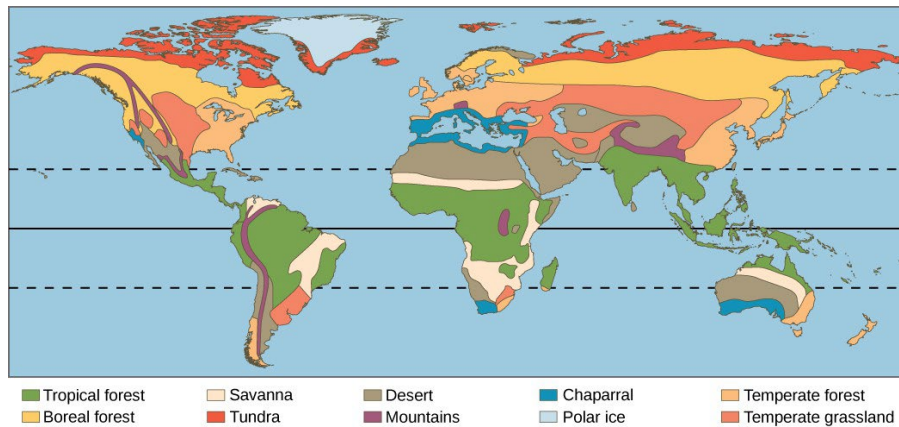
Figure 6.17 Each of the world's eight major biomes is distinguished by characteristic temperatures and amount of precipitation. Polar ice caps and mountains are also shown.

Tropical Forest

Tropical rainforests are also referred to as tropical wet forests. This biome is found in equatorial regions (Figure 6.17). Tropical rainforests are the most diverse terrestrial biome. This biodiversity is still largely unknown to science and is under extraordinary threat primarily through logging and deforestation for agriculture. Tropical rainforests have also been described as nature's pharmacy because of the potential for new drugs that is largely hidden in the chemicals produced by the huge diversity of plants, animals, and other organisms. The vegetation is characterized by plants with spreading roots and broad leaves that fall off throughout the year, unlike the trees of deciduous forests that lose their leaves in one season. These forests are "evergreen," year-round.

The temperature and sunlight profiles of tropical rainforests are stable in comparison to that of other terrestrial biomes, with average temperatures ranging from 20°C to 34°C (68°F to 93°F). Month-to-month temperatures are relatively constant in tropical rainforests, in contrast to forests further from the equator. This lack of temperature seasonality leads to year-round plant growth, rather than the seasonal growth seen in other biomes. In contrast to other ecosystems, a more constant daily amount of sunlight (11–12 hours per day) provides more solar radiation, thereby a longer period of time for plant growth.

The annual rainfall in tropical rainforests ranges from 250 cm to more than 450 cm (8.2–14.8 ft) with considerable seasonal variation. Tropical rainforests have wet months in which there can be more than 30 cm (11–12 in) of precipitation, as well as dry months in which there are fewer than 10 cm (3.5 in) of rainfall. However, the driest month of a tropical rainforest can still exceed the *annual* rainfall of some other biomes, such as deserts.

Tropical rainforests have high net primary productivity because the annual temperatures and precipitation values support rapid plant growth (Figure 6.18). However, the high rainfall quickly leaches nutrients from the soils of these forests, which are typically low in nutrients. Tropical rainforests are characterized by vertical layering of vegetation and the formation of distinct habitats for animals within each layer. On the forest floor is a sparse layer of plants and decaying plant matter. Above that is an understory of short, shrubby foliage. A layer of trees rises above this understory and is topped by a closed upper **canopy**—the uppermost overhead layer of branches and leaves. Some additional trees emerge through this closed upper canopy. These layers provide diverse and complex habitats for the variety of plants, animals, and other organisms within the tropical wet forests. Many species of animals use the variety of plants and the complex structure of the tropical wet forests for food and shelter. Some organisms live several meters above ground rarely ever descending to the forest floor.

Rainforests are not the only forest biome in the tropics; there are also tropical dry forests, which are characterized by a dry season of varying lengths. These forests commonly experience leaf loss during the dry season to one degree or another. The loss of leaves from taller trees during the dry season opens up the canopy and allows sunlight to the forest floor that allows the growth of thick ground-level brush, which is absent in tropical rainforests. Extensive tropical dry forests occur in Africa (including Madagascar), India, southern Mexico, and South America.

231

Figure 6.18 Species diversity is very high in tropical wet forests, such as these forests of Madre de Dios, Peru, near the Amazon River. (credit: Roosevelt Garcia)

Savannas

Savannas are grasslands with scattered trees, and they are found in Africa, South America, and northern Australia (Figure 6.17). Savannas are hot, tropical areas with temperatures averaging from 24°C –29°C (75°F –84°F) and an annual rainfall of 51–127 cm (20–50 in). Savannas have an extensive dry season and consequent fires. As a result, scattered in the grasses and forbs (herbaceous flowering plants) that dominate the savanna, there are relatively few trees (Figure 6.19). Since fire is an important source of disturbance in this biome, plants have evolved well-developed root systems that allow them to quickly re-sprout after a fire.

Figure 6.19 Although savannas are dominated by grasses, small woodlands, such as this one in Mount Archer National Park in Queensland, Australia, may dot the landscape. (credit: "Ethel Aardvark"/Wikimedia Commons)

Deserts

Subtropical deserts exist between 15° and 30° north and south latitude and are centered on the Tropic of Cancer and the Tropic of Capricorn (Figure 6.17). Deserts are frequently located on the downwind or lee side of mountain ranges, which create a rain shadow after prevailing winds drop their water

232

content on the mountains. This is typical of the North American deserts, such as the Mohave and Sonoran deserts. Deserts in other regions, such as the Sahara Desert in northern Africa or the Namib Desert in southwestern Africa are dry because of the high-pressure, dry air descending at those latitudes. Subtropical deserts are very dry; evaporation typically exceeds precipitation. Subtropical hot deserts can have daytime soil surface temperatures above 60°C (140°F) and nighttime temperatures approaching 0°C (32°F). The temperature drops so far because there is little water vapor in the air to prevent radiative cooling of the land surface. Subtropical deserts are characterized by low annual precipitation of fewer than 30 cm (12 in) with little monthly variation and lack of predictability in rainfall. Some years may receive tiny amounts of rainfall, while others receive more. In some cases, the annual rainfall can be as low as 2 cm (0.8 in) in subtropical deserts located in central Australia ("the Outback") and northern Africa.

The low species diversity of this biome is closely related to its low and unpredictable precipitation. Despite the relatively low diversity, desert species exhibit fascinating adaptations to the harshness of their environment. Very dry deserts lack perennial vegetation that lives from one year to the next; instead, many plants are annuals that grow quickly and reproduce when rainfall does occur, then they die. Perennial plants in deserts are characterized by adaptations that conserve water: deep roots, reduced foliage, and water-storing stems (Figure 6.20). Seed plants in the desert produce seeds that can lie dormant for extended periods between rains. Most animal life in subtropical deserts has adapted to a nocturnal life, spending the hot daytime hours beneath the ground. The Namib Desert is the oldest on the planet, and has probably been dry for more than 55 million years. It supports a number of endemic species (species found only there) because of this great age. For example, the unusual gymnosperm *Welwitschia mirabilis* is the only extant species of an entire order of plants. There are also five species of reptiles considered endemic to the Namib.

In addition to subtropical deserts there are cold deserts that experience freezing temperatures during the winter and any precipitation is in the form of snowfall. The largest of these deserts are the Gobi Desert in northern China and southern Mongolia, the Taklimakan Desert in western China, the Turkestan Desert, and the Great Basin Desert of the United States.

Figure 6.20 Many desert plants have tiny leaves or no leaves at all to reduce water loss. The leaves of ocotillo, shown here in the Chihuahuan Desert in Big Bend National Park, Texas, appear only after rainfall and then are shed. (credit "bare ocotillo": "Leaflet"/Wikimedia Commons)

Chaparral

The **chaparral** is also called scrub forest and is found in California, along the Mediterranean Sea, and along the southern coast of Australia (Figure 6.17). The annual rainfall in this biome ranges from 65 cm to 75 cm (25.6–29.5 in) and the majority of the rain falls in the winter. Summers are very dry and many chaparral plants are dormant during the summertime. The chaparral vegetation is dominated by shrubs and is adapted to periodic fires, with some plants producing seeds that germinate only after a hot fire. The ashes left behind after a fire are rich in nutrients like nitrogen that fertilize the soil and promote plant regrowth. Fire is a natural part of the maintenance of this biome and frequently threatens human habitation in this biome in the U.S. (Figure 6.21).

233

Figure 6.21 The chaparral is dominated by shrubs. (credit: Miguel Vieira)

Temperate Grasslands

Temperate grasslands are found throughout central North America, where they are also known as prairies, and in Eurasia, where they are known as steppes (Figure 6.17). Temperate grasslands have pronounced annual fluctuations in temperature with hot summers and cold winters. The annual temperature variation produces specific growing seasons for plants. Plant growth is possible when temperatures are warm enough to sustain plant growth, which occurs in the spring, summer, and fall.

Annual precipitation ranges from 25.4 cm to 88.9 cm (10–35 in). Temperate grasslands have few trees except for those found growing along rivers or streams. The dominant vegetation tends to consist of grasses. The treeless condition is maintained by low precipitation, frequent fires, and grazing (Figure 6.22). The vegetation is very dense and the soils are fertile because the subsurface of the soil is packed with the roots and rhizomes (underground stems) of these grasses. The roots and rhizomes act to anchor plants into the ground and replenish the organic material (humus) in the soil when they die and decay.

Figure 6.22 The American bison (*Bison bison*), more commonly called the buffalo, is a grazing mammal that once populated American prairies in huge numbers. (credit: Jack Dykinga, USDA ARS)

Fires, which are a natural disturbance in temperate grasslands, can be ignited by lightning strikes. It also appears that the lightning-caused fire regime in North American grasslands was enhanced by intentional burning by humans. When fire is suppressed in temperate grasslands, the vegetation eventually converts to scrub and dense forests. Often, the restoration or management of temperate grasslands requires the use of controlled burns to suppress the growth of trees and maintain the grasses.

234

Temperate Forests

Temperate forests are the most common biome in eastern North America, Western Europe, Eastern Asia, Chile, and New Zealand (Figure 6.17). This biome is found throughout mid-latitude regions. Temperatures range between –30°C and 30°C (–22°F to 86°F) and drop to below freezing on an annual basis. These temperatures mean that temperate forests have defined growing seasons during the spring, summer, and early fall. Precipitation is relatively constant throughout the year and ranges between 75 cm and 150 cm (29.5–59 in).

Deciduous trees are the dominant plant in this biome with fewer evergreen conifers. Deciduous trees lose their leaves each fall and remain leafless in the winter. Thus, little photosynthesis occurs during the dormant winter period. Each spring, new leaves appear as temperature increases. Because of the dormant period, the net primary productivity of temperate forests is less than that of tropical rainforests. In addition, temperate forests show far less diversity of tree species than tropical rainforest biomes.

The trees of the temperate forests leaf out and shade much of the ground; however, more sunlight reaches the ground in this biome than in tropical rainforests because trees in temperate forests do not grow as tall as the trees in tropical rainforests. The soils of the temperate forests are rich in inorganic and organic nutrients compared to tropical rainforests. This is because of the thick layer of leaf litter on forest floors and reduced leaching of nutrients by rainfall. As this leaf litter decays, nutrients are returned to the soil. The leaf litter also protects soil from erosion, insulates the ground, and provides habitats for invertebrates and their predators (Figure 6.23).

Figure 6.23 Deciduous trees are the dominant plant in the temperate forest. (credit: Oliver Herold)

Boreal Forests

The **boreal forest**, also known as taiga or coniferous forest, is found roughly between 50° and 60° north latitude across most of Canada, Alaska, Russia, and northern Europe (Figure 6.17). Boreal forests are also found above a certain elevation (and below high elevations where trees cannot grow) in mountain ranges throughout the Northern Hemisphere. This biome has cold, dry winters and short, cool, wet summers. The annual precipitation is from 40 cm to 100 cm (15.7–39 in) and usually takes the form of snow; little evaporation occurs because of the cold temperatures.

The long and cold winters in the boreal forest have led to the predominance of cold-tolerant cone-bearing plants. These are evergreen coniferous trees like pines, spruce, and fir, which retain their needle-shaped leaves year-round. Evergreen trees can photosynthesize earlier in the spring than deciduous trees because less energy from the Sun is required to warm a needle-like leaf than a broad leaf. Evergreen trees grow faster than deciduous trees in the boreal forest. In addition, soils in boreal forest regions tend to be acidic with little available nitrogen. Leaves are a nitrogen-rich structure and deciduous trees must produce a new set of these nitrogen-rich structures each year. Therefore, coniferous trees that retain nitrogen-rich needles in a nitrogen limiting environment may have had a competitive advantage over the broad-leafed deciduous trees.

- Download for free at http://openstaxcollege.org/textbooks/concepts-of-biology/get.

The net primary productivity of boreal forests is lower than that of temperate forests and tropical wet forests. The aboveground biomass of boreal forests is high because these slow-growing tree species are long-lived and accumulate standing biomass over time. Species diversity is less than that seen in temperate forests and tropical rainforests. Boreal forests lack the layered forest structure seen in tropical rainforests or, to a lesser degree, temperate forests. The structure of a boreal forest is often only a tree layer and a ground layer. When conifer needles are dropped, they decompose more slowly than broad leaves; therefore, fewer nutrients are returned to the soil to fuel plant growth (Figure 6.24).

Figure 6.24 The boreal forest (taiga) has low lying plants and conifer trees. (credit: L.B. Brubaker, NOAA)

Arctic Tundra

The **Arctic tundra** lies north of the subarctic boreal forests and is located throughout the Arctic regions of the Northern Hemisphere (Figure 6.17). Tundra also exists at elevations above the tree line on mountains. The average winter temperature is –34°C (–34°F) and the average summer temperature is 3°C–12°C (37°F –52°F). Plants in the Arctic tundra have a short growing season of approximately 50–60 days. However, during this time, there are almost 24 hours of daylight and plant growth is rapid. The annual precipitation of the Arctic tundra is low (15–25 cm or 6–10 in) with little annual variation in precipitation. And, as in the boreal forests, there is little evaporation because of the cold temperatures.

Plants in the Arctic tundra are generally low to the ground and include low shrubs, grasses, lichens, and small flowering plants (Figure 6.25). There is little species diversity, low net primary productivity, and low aboveground biomass. The soils of the Arctic tundra may remain in a perennially frozen state referred to as **permafrost**. The permafrost makes it impossible for roots to penetrate far into the soil and slows the decay of organic matter, which inhibits the release of nutrients from organic matter. The melting of the permafrost in the brief summer provides water for a burst of productivity while temperatures and long days permit it. During the growing season, the ground of the Arctic tundra can be completely covered with plants or lichens.

Figure 6.25 Low-growing plants such as shrub willow dominate the tundra landscape during the summer, shown here in the Arctic National Wildlife Refuge. (credit: Arctic National Wildlife Refuge, USFWS)

Watch this **Assignment Discovery: Biomes (http://openstaxcollege.org/l/biomes)** video for an overview of biomes. To explore further, select one of the biomes on the extended playlist: desert, savanna, temperate forest, temperate grassland, tropic, tundra.

| Aquatic and Marine Biomes

By the end of this section, you will be able to:

- Describe the effects of abiotic factors on the composition of plant and animal communities in aquatic biomes
- Compare the characteristics of the ocean zones
- Summarize the characteristics of standing water and flowing water in freshwater biomes

Like terrestrial biomes, aquatic biomes are influenced by abiotic factors. In the case of aquatic biomes the abiotic factors include light, temperature, flow regime, and dissolved solids. The aquatic medium—water— has different physical and chemical properties than air. Even if the water in a pond or other body of water is perfectly clear (there are no suspended particles), water, on its own, absorbs light. As one descends deep enough into a body of water, eventually there will be a depth at which the sunlight cannot reach. While there are some abiotic and biotic factors in a terrestrial ecosystem that shade light (like fog, dust, or insect swarms), these are not usually permanent features of the environment. The importance of light in aquatic biomes is central to the communities of organisms found in both freshwater and marine ecosystems because it controls productivity through photosynthesis.

In addition to light, solar radiation warms bodies of water and many exhibit distinct layers of water at differing temperatures. The water temperature affects the organisms' rates of growth and the amount of dissolved oxygen available for respiration.

237

The movement of water is also important in many aquatic biomes. In rivers, the organisms must obviously be adapted to the constant movement of the water around them, but even in larger bodies of water such as the oceans, regular currents and tides impact availability of nutrients, food resources, and the presence of the water itself.

Finally, all natural water contains dissolved solids, or salts. Fresh water contains low levels of such dissolved substances because the water is rapidly recycled through evaporation and precipitation. The oceans have a relatively constant high salt content. Aquatic habitats at the interface of marine and freshwater ecosystems have complex and variable salt environments that range between freshwater and marine levels. These are known as brackish water environments. Lakes located in closed drainage basins concentrate salt in their waters and can have extremely high salt content that only a few and highly specialized species are able to inhabit.

Marine Biomes

The ocean is a continuous body of salt water that is relatively uniform in chemical composition. It is a weak solution of mineral salts and decayed biological matter. Within the ocean, coral reefs are a second type of marine biome. Estuaries, coastal areas where salt water and fresh water mix, form a third unique marine biome.

The ocean is categorized by several zones (Figure 6.27). All of the ocean's open water is referred to as the **pelagic realm** (or zone). The **benthic realm** (or zone) extends along the ocean bottom from the shoreline to the deepest parts of the ocean floor. From the surface to the bottom or the limit to which photosynthesis occurs is the **photic zone** (approximately 200 m or 650 ft). At depths greater than 200 m, light cannot penetrate; thus, this is referred to as the **aphotic zone**. The majority of the ocean is aphotic and lacks sufficient light for photosynthesis. The deepest part of the ocean, the Challenger Deep (in the Mariana Trench, located in the western Pacific Ocean), is about 11,000 m (about 6.8 mi) deep. To give some perspective on the depth of this trench, the ocean is, on average, 4267 m or 14,000 ft deep.

Ocean

The physical diversity of the ocean has a significant influence on the diversity of organisms that live within it. The ocean is categorized into different zones based on how far light reaches into the water. Each zone has a distinct group of species adapted to the biotic and abiotic conditions particular to that zone.

The **intertidal zone** (Figure 6.27) is the oceanic region that is closest to land. With each tidal cycle, the intertidal zone alternates between being inundated with water and left high and dry. Generally, most people think of this portion of the ocean as a sandy beach. In some cases, the intertidal zone is indeed a sandy beach, but it can also be rocky, muddy, or dense with tangled roots in mangrove forests. The intertidal zone is an extremely variable environment because of tides. Organisms may be exposed to air at low tide and are underwater during high tide. Therefore, living things that thrive in the intertidal zone are often adapted to being dry for long periods of time. The shore of the intertidal zone is also repeatedly struck by waves and the organisms found there are adapted to withstand damage from the pounding action of the waves (Figure 6.26). The exoskeletons of shoreline crustaceans (such as the shore crab, *Carcinus maenas*) are tough and protect them from desiccation (drying out) and wave damage. Another consequence of the pounding waves is that few algae and plants establish themselves in constantly moving sand or mud.

Figure 6.26 Sea stars, sea urchins, and mussel shells are often found in the intertidal zone, shown here in Kachemak Bay, Alaska. (credit: NOAA)

The **neritic zone** (Figure 6.27) extends from the margin of the intertidal zone to depths of about 200 m (or 650 ft) at the edge of the continental shelf. When the water is relatively clear, photosynthesis can occur in the neritic zone. The water contains silt and is well-oxygenated, low in pressure, and stable in temperature. These factors all contribute to the neritic zone having the highest productivity and biodiversity of the ocean. Phytoplankton, including photosynthetic bacteria and larger species of algae, are responsible for the bulk of this primary productivity. Zooplankton, protists, small fishes, and shrimp feed on the producers and are the primary food source for most of the world's fisheries. The majority of these fisheries exist within the neritic zone.

Beyond the neritic zone is the open ocean area known as the **oceanic zone** (Figure 6.27). Within the oceanic zone there is thermal stratification. Abundant phytoplankton and zooplankton support populations of fish and whales. Nutrients are scarce and this is a relatively less productive part of the marine biome. When photosynthetic organisms and the organisms that feed on them die, their bodies fall to the bottom of the ocean where they remain; the open ocean lacks a process for bringing the organic nutrients back up to the surface.

Beneath the pelagic zone is the benthic realm, the deepwater region beyond the continental shelf (Figure 6.27). The bottom of the benthic realm is comprised of sand, silt, and dead organisms. Temperature decreases as water depth increases. This is a nutrient-rich portion of the ocean because of the dead organisms that fall from the upper layers of the ocean. Because of this high level of nutrients, a diversity of fungi, sponges, sea anemones, marine worms, sea stars, fishes, and bacteria exists.

The deepest part of the ocean is the **abyssal zone**, which is at depths of 4000 m or greater. The abyssal zone (Figure 6.27) is very cold and has very high pressure, high oxygen content, and low nutrient content. There are a variety of invertebrates and fishes found in this zone, but the abyssal zone does not have photosynthetic organisms. Chemosynthetic bacteria use the hydrogen sulfide and other minerals emitted from deep hydrothermal vents. These chemosynthetic bacteria use the hydrogen sulfide as an energy source and serve as the base of the food chain found around the vents.

239
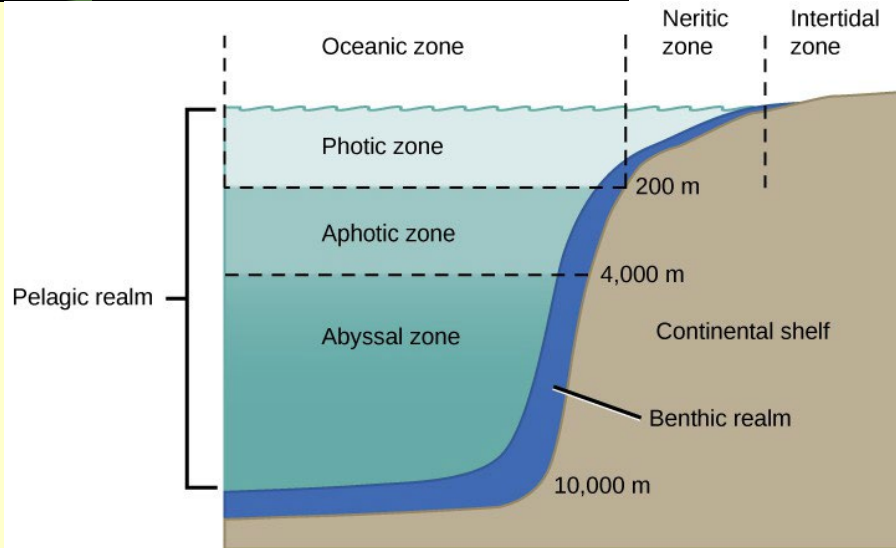
Figure 6.27 The ocean is divided into different zones based on water depth, distance from the shoreline, and light penetration.

Coral Reefs

Coral reefs are ocean ridges formed by marine invertebrates living in warm shallow waters within the photic zone of the ocean. They are found within 30° north and south of the equator. The Great Barrier Reef is a well-known reef system located several miles off the northeastern coast of Australia. Other coral reefs are fringing islands, which are directly adjacent to land, or atolls, which are circular reefs surrounding a former island that is now underwater. The coral-forming colonies of organisms (members of phylum Cnidaria) secrete a calcium carbonate skeleton. These calcium-rich skeletons slowly accumulate, thus forming the underwater reef (Figure 6.28). Corals found in shallower waters (at a depth of approximately 60 m or about 200 ft) have a mutualistic relationship with photosynthetic unicellular protists. The relationship provides corals with the majority of the nutrition and the energy they require. The waters in which these corals live are nutritionally poor and, without this mutualism, it would not be possible for large corals to grow because there are few planktonic organisms for them to feed on. Some corals living in deeper and colder water do not have a mutualistic relationship with protists; these corals must obtain their energy exclusively by feeding on plankton using stinging cells on their tentacles.

CONCEPT in ACTION

openstax COLLEGE

In this National Oceanic and Atmospheric Administration (NOAA) **video (http://openstaxcollege.org/l/coral_organisms)** , marine ecologist Dr. Peter Etnoyer discusses his research on coral organisms.

Coral reefs are one of the most diverse biomes. It is estimated that more than 4000 fish species inhabit coral reefs. These fishes can feed on coral, the **cryptofauna** (invertebrates found within the calcium carbonate structures of the coral reefs), or the seaweed and algae that are associated with the coral. These species include predators, herbivores, or planktivores. Predators are animal species that hunt and are carnivores or "flesh eaters." Herbivores eat plant material, and **planktivores** eat plankton.

Figure 6.28 Coral reefs are formed by the calcium carbonate skeletons of coral organisms, which are marine invertebrates in the phylum Cnidaria. (credit: Terry Hughes)

evolution IN ACTION

Global Decline of Coral Reefs

It takes a long time to build a coral reef. The animals that create coral reefs do so over thousands of years, continuing to slowly deposit the calcium carbonate that forms their characteristic ocean homes. Bathed in warm tropical waters, the coral animals and their symbiotic protist partners evolved to survive at the upper limit of ocean water temperature.

Together, climate change and human activity pose dual threats to the long-term survival of the world's coral reefs. The main cause of killing of coral reefs is warmer-than-usual surface water. As global warming raises ocean temperatures, coral reefs are suffering. The excessive warmth causes the coral organisms to expel their endosymbiotic, food-producing protists, resulting in a phenomenon known as bleaching. The colors of corals are a result of the particular protist endosymbiont, and when the protists leave, the corals lose their color and turn white, hence the term "bleaching."

Rising levels of atmospheric carbon dioxide further threaten the corals in other ways; as carbon dioxide dissolves in ocean waters, it lowers pH, thus increasing ocean acidity. As acidity increases, it interferes with the calcification that normally occurs as coral animals build their calcium carbonate homes.

When a coral reef begins to die, species diversity plummets as animals lose food and shelter. Coral reefs are also economically important tourist destinations, so the decline of coral reefs poses a serious threat to coastal economies.

Human population growth has damaged corals in other ways, too. As human coastal populations increase, the runoff of sediment and agricultural chemicals has increased, causing some of the once-clear tropical waters to become cloudy. At the same time, overfishing of popular fish species has allowed the predator species that eat corals to go unchecked.

Although a rise in global temperatures of 1°C–2°C (a conservative scientific projection) in the coming decades may not seem large, it is very significant to this biome. When change occurs rapidly, species can become extinct before evolution leads to newly adapted species. Many scientists believe that global warming, with its rapid (in terms of evolutionary time) and inexorable increases in temperature, is tipping the balance beyond the point at which many of the world's coral reefs can recover.

Estuaries: Where the Ocean Meets Fresh Water

Estuaries are biomes that occur where a river, a source of fresh water, meets the ocean. Therefore, both fresh water and salt water are found in the same vicinity; mixing results in a diluted (brackish) salt water. Estuaries form protected areas where many of the offspring of crustaceans, mollusks, and fish begin their lives. Salinity is an important factor that influences the organisms and the adaptations of the organisms found in estuaries. The salinity of estuaries varies and is based on the rate of flow of its freshwater sources. Once or twice a day, high tides bring salt water into the estuary. Low tides occurring at the same frequency reverse the current of salt water (Figure 6.29).

Figure 6.29 As estuary is where fresh water and salt water meet, such as the mouth of the Klamath River in California, shown here. (credit: U.S. Army Corps of Engineers)

The daily mixing of fresh water and salt water is a physiological challenge for the plants and animals that inhabit estuaries. Many estuarine plant species are halophytes, plants that can tolerate salty conditions. Halophytic plants are adapted to deal with salt water spray and salt water on their roots. In some halophytes, filters in the roots remove the salt from the water that the plant absorbs. Animals, such as mussels and clams (phylum Mollusca), have developed behavioral adaptations that expend a lot of energy to function in this rapidly changing environment. When these animals are exposed to low salinity, they stop feeding, close their shells, and switch from aerobic respiration (in which they use gills) to anaerobic respiration (a process that does not require oxygen). When high tide returns to the estuary, the salinity and oxygen content of the water increases, and these animals open their shells, begin feeding, and return to aerobic respiration.

Freshwater Biomes

Freshwater biomes include lakes, ponds, and wetlands (standing water) as well as rivers and streams (flowing water). Humans rely on freshwater biomes to provide aquatic resources for drinking water, crop irrigation, sanitation, recreation, and industry. These various roles and human benefits are referred to as **ecosystem services**. Lakes and ponds are found in terrestrial landscapes and are therefore connected with abiotic and biotic factors influencing these terrestrial biomes.

Lakes and Ponds

Lakes and ponds can range in area from a few square meters to thousands of square kilometers. Temperature is an important abiotic factor affecting living things found in lakes and ponds. During the summer in temperate regions, thermal stratification of deep lakes occurs when the upper layer of water is warmed by the Sun and does not mix with deeper, cooler water. The process produces a sharp transition between the warm water above and cold water beneath. The two layers do not mix until cooling temperatures and winds break down the stratification and the water in the lake mixes from top to bottom. During the period of stratification, most of the productivity occurs in the warm, well-illuminated, upper layer, while dead organisms slowly rain down into the cold, dark layer below where decomposing bacteria and cold-adapted species such as lake trout exist. Like the ocean, lakes and ponds have a photic layer in which photosynthesis can occur. Phytoplankton (algae and cyanobacteria) are found here and provide the base of the food web of lakes and ponds. Zooplankton, such as rotifers and small crustaceans, consume these phytoplankton. At the bottom of lakes and ponds, bacteria in the aphotic zone break down dead organisms that sink to the bottom.

Nitrogen and particularly phosphorus are important limiting nutrients in lakes and ponds. Therefore, they are determining factors in the amount of phytoplankton growth in lakes and ponds. When there is a large input of nitrogen and phosphorus (e.g., from sewage and runoff from fertilized lawns and farms), the growth of algae skyrockets, resulting in a large accumulation of algae called an **algal bloom**. Algal blooms (Figure 6.30) can become so extensive that they reduce light penetration in water. As a result, the lake or pond becomes aphotic and photosynthetic plants cannot survive. When the algae die and decompose, severe oxygen depletion of the water occurs. Fishes and other organisms that require oxygen are then more likely to die.

244

Figure 6.30 The uncontrolled growth of algae in this waterway has resulted in an algal bloom.

Rivers and Streams

Rivers and the narrower streams that feed into the rivers are continuously moving bodies of water that carry water from the source or headwater to the mouth at a lake or ocean. The largest rivers include the Nile River in Africa, the Amazon River in South America, and the Mississippi River in North America (Figure 6.31).

(a)   (b)

Figure 6.31 Rivers range from (a) narrow and shallow to (b) wide and slow moving. (credit a: modification of work by Cory Zanker; credit b: modification of work by David DeHetre)

Abiotic features of rivers and streams vary along the length of the river or stream. Streams begin at a point of origin referred to as **source water**. The source water is usually cold, low in nutrients, and clear. The **channel** (the width of the river or stream) is narrower here than at any other place along the length of the river or stream. Headwater streams are of necessity at a higher elevation than the mouth of the river and often originate in regions with steep grades leading to higher flow rates than lower elevation stretches of the river.

Faster-moving water and the short distance from its origin results in minimal silt levels in headwater streams; therefore, the water is clear. Photosynthesis here is mostly attributed to algae that are growing on rocks; the swift current inhibits the growth of phytoplankton. Photosynthesis may be further reduced by tree cover reaching over the narrow stream. This shading also keeps temperatures lower. An additional input of energy can come from leaves or other organic material that falls into a river or stream from the trees and other plants that border the water. When the leaves decompose, the organic material and nutrients in the leaves are returned to the water. The leaves also support a food chain of invertebrates that eat them and are in turn eaten by predatory invertebrates and fish. Plants and animals have adapted to this fast-moving water. For instance, leeches (phylum Annelida) have elongated bodies and suckers on both ends. These suckers attach to the substrate, keeping the leech anchored in place. In temperate

245

Figure 6.30 The uncontrolled growth of algae in this waterway has resulted in an algal bloom.

Rivers and Streams

Rivers and the narrower streams that feed into the rivers are continuously moving bodies of water that carry water from the source or headwater to the mouth at a lake or ocean. The largest rivers include the Nile River in Africa, the Amazon River in South America, and the Mississippi River in North America (Figure 6.31).

(a)   (b)

Figure 6.31 Rivers range from (a) narrow and shallow to (b) wide and slow moving. (credit a: modification of work by Cory Zanker; credit b: modification of work by David DeHetre)

Abiotic features of rivers and streams vary along the length of the river or stream. Streams begin at a point of origin referred to as **source water**. The source water is usually cold, low in nutrients, and clear. The **channel** (the width of the river or stream) is narrower here than at any other place along the length of the river or stream. Headwater streams are of necessity at a higher elevation than the mouth of the river and often originate in regions with steep grades leading to higher flow rates than lower elevation stretches of the river.

Faster-moving water and the short distance from its origin results in minimal silt levels in headwater streams; therefore, the water is clear. Photosynthesis here is mostly attributed to algae that are growing on rocks; the swift current inhibits the growth of phytoplankton. Photosynthesis may be further reduced by tree cover reaching over the narrow stream. This shading also keeps temperatures lower. An additional input of energy can come from leaves or other organic material that falls into a river or stream from the trees and other plants that border the water. When the leaves decompose, the organic material and nutrients in the leaves are returned to the water. The leaves also support a food chain of invertebrates that eat them and are in turn eaten by predatory invertebrates and fish. Plants and animals have adapted to this fast-moving water. For instance, leeches (phylum Annelida) have elongated bodies and suckers on both ends. These suckers attach to the substrate, keeping the leech anchored in place. In temperate

245

regions, freshwater trout species (phylum Chordata) may be an important predator in these fast-moving and colder river and streams.

As the river or stream flows away from the source, the width of the channel gradually widens, the current slows, and the temperature characteristically increases. The increasing width results from the increased volume of water from more and more tributaries. Gradients are typically lower farther along the river, which accounts for the slowing flow. With increasing volume can come increased silt, and as the flow rate slows, the silt may settle, thus increasing the deposition of sediment. Phytoplankton can also be suspended in slow-moving water. Therefore, the water will not be as clear as it is near the source. The water is also warmer as a result of longer exposure to sunlight and the absence of tree cover over wider expanses between banks. Worms (phylum Annelida) and insects (phylum Arthropoda) can be found burrowing into the mud. Predatory vertebrates (phylum Chordata) include waterfowl, frogs, and fishes. In heavily silt-laden rivers, these predators must find food in the murky waters, and, unlike the trout in the clear waters at the source, these vertebrates cannot use vision as their primary sense to find food. Instead, they are more likely to use taste or chemical cues to find prey.

When a river reaches the ocean or a large lake, the water typically slows dramatically and any silt in the river water will settle. Rivers with high silt content discharging into oceans with minimal currents and wave action will build deltas, low-elevation areas of sand and mud, as the silt settles onto the ocean bottom. Rivers with low silt content or in areas where ocean currents or wave action are high create estuarine areas where the fresh water and salt water mix.

Wetlands

Wetlands are environments in which the soil is either permanently or periodically saturated with water. Wetlands are different from lakes and ponds because wetlands exhibit a near continuous cover of emergent vegetation. **Emergent vegetation** consists of wetland plants that are rooted in the soil but have portions of leaves, stems, and flowers extending above the water's surface. There are several types of wetlands including marshes, swamps, bogs, mudflats, and salt marshes (Figure 6.32).

Figure 6.32 Located in southern Florida, Everglades National Park is vast array of wetland environments, including sawgrass marshes, cypress swamps, and estuarine mangrove forests. Here, a great egret walks among cypress trees. (credit: NPS)

Freshwater marshes and swamps are characterized by slow and steady water flow. Bogs develop in depressions where water flow is low or nonexistent. Bogs usually occur in areas where there is a clay bottom with poor percolation. Percolation is the movement of water through the pores in the soil or rocks. The water found in a bog is stagnant and oxygen depleted because the oxygen that is used during the decomposition of organic matter is not replaced. As the oxygen in the water is depleted, decomposition slows. This leads to organic acids and other acids building up and lowering the pH of the water. At a lower pH, nitrogen becomes unavailable to plants. This creates a challenge for plants because nitrogen is an important limiting resource. Some types of bog plants (such as sundews, pitcher plants, and Venus flytraps) capture insects and extract the nitrogen from their bodies. Bogs have low net primary productivity because the water found in bogs has low levels of nitrogen and oxygen.

Population Demographics and Dynamics

By the end of this section, you will be able to:

- Describe how ecologists measure population size and density
- Describe three different patterns of population distribution
- Use life tables to calculate mortality rates
- Describe the three types of survivorship curves and relate them to specific populations

Populations are dynamic entities. Their size and composition fluctuate in response to numerous factors, including seasonal and yearly changes in the environment, natural disasters such as forest fires and volcanic eruptions, and competition for resources between and within species. The statistical study of populations is called **demography**: a set of mathematical tools designed to describe populations and investigate how they change. Many of these tools were actually designed to study human populations. For example, **life tables**, which detail the life expectancy of individuals within a population, were initially developed by life insurance companies to set insurance rates. In fact, while the term "demographics" is sometimes assumed to mean a study of human populations, all living populations can be studied using this approach.

Population Size and Density

Populations are characterized by their **population size** (total number of individuals) and their **population density** (number of individuals per unit area). A population may have a large number of individuals that are distributed densely, or sparsely. There are also populations with small numbers of individuals that may be dense or very sparsely distributed in a local area. Population size can affect potential for adaptation because it affects the amount of genetic variation present in the population. Density can have effects on interactions within a population such as competition for food and the ability of individuals to find a mate. Smaller organisms tend to be more densely distributed than larger organisms (Figure 6.33).
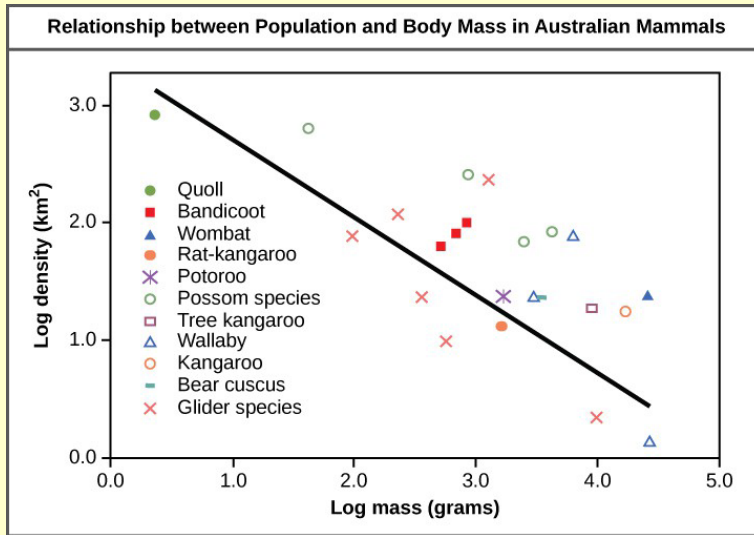
Figure 6.33 Australian mammals show a typical inverse relationship between population density and body size.

As this graph shows, population density typically decreases with increasing body size. Why do you think this is the case?

The most accurate way to determine population size is to count all of the individuals within the area. However, this method is usually not logistically or economically feasible, especially when studying large areas. Thus, scientists usually study populations by sampling a representative portion of each habitat and use this sample to make inferences about the population as a whole. The methods used to sample populations to determine their size and density are typically tailored to the characteristics of the organism being studied. For immobile organisms such as plants, or for very small and slow-moving organisms, a quadrat may be used. A **quadrat** is a wood, plastic, or metal square that is randomly located on the ground and used to count the number of individuals that lie within its boundaries. To obtain an accurate count using this method, the square must be placed at random locations within the habitat enough times to produce an accurate estimate. This counting method will provide an estimate of both population size and density. The number and size of quadrat samples depends on the type of organisms and the nature of their distribution.

For smaller mobile organisms, such as mammals, a technique called **mark and recapture** is often used. This method involves marking a sample of captured animals in some way and releasing them back into the environment to mix with the rest of the population; then, a new sample is captured and scientists determine how many of the marked animals are in the new sample. This method assumes that the larger the population, the lower the percentage of marked organisms that will be recaptured since they will have mixed with more unmarked individuals. For example, if 80 field mice are captured, marked, and released into the forest, then a second trapping 100 field mice are captured and 20 of them are marked, the population size (N) can be determined using the following equation:

$$\frac{\text{number marked first catch} \times \text{total number second catch}}{\text{number marked second catch}} =$$

Using our example, the population size would be 400.

$$\frac{80 \times 100}{20} = 400$$

These results give us an estimate of 400 total individuals in the original population. The true number usually will be a bit different from this because of chance errors and possible bias caused by the sampling methods.

Species Distribution

In addition to measuring density, further information about a population can be obtained by looking at the distribution of the individuals throughout their range. A **species distribution pattern** is the distribution of individuals within a habitat at a particular point in time—broad categories of patterns are used to describe them.

Individuals within a population can be distributed at random, in groups, or equally spaced apart (more or less). These are known as random, clumped, and uniform distribution patterns, respectively (**Figure** 6.34). Different distributions reflect important aspects of the biology of the species; they also affect the mathematical methods required to estimate population sizes. An example of random distribution occurs with dandelion and other plants that have wind-dispersed seeds that germinate wherever they happen to fall in favorable environments. A clumped distribution, may be seen in plants that drop their seeds straight to the ground, such as oak trees; it can also be seen in animals that live in social groups (schools of fish or herds of elephants). Uniform distribution is observed in plants that secrete substances inhibiting the growth of nearby individuals (such as the release of toxic chemicals by sage plants). It is also seen in territorial animal species, such as penguins that maintain a defined territory for nesting. The territorial defensive behaviors of each individual create a regular pattern of distribution of similar-sized territories and individuals within those territories. Thus, the distribution of the individuals within a population provides more information about how they interact with each other than does a simple density measurement. Just as lower density species might have more difficulty finding a mate, solitary species with a random distribution might have a similar difficulty when compared to social species clumped together in groups.
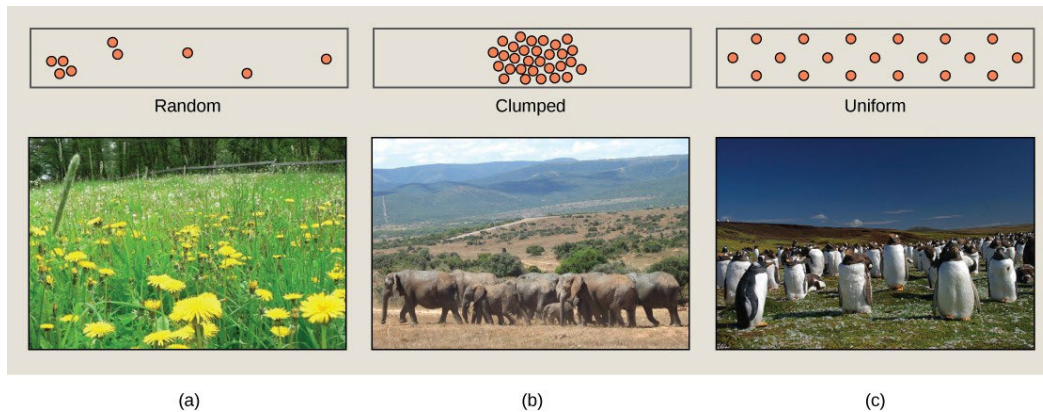
Figure 6.34 Species may have a random, clumped, or uniform distribution. Plants such as (a) dandelions with wind-dispersed seeds tend to be randomly distributed. Animals such as (b) elephants that travel in groups exhibit a clumped distribution. Territorial birds such as (c) penguins tend to have a uniform distribution. (credit a: modification of work by Rosendahl; credit b: modification of work by Rebecca Wood; credit c: modification of work by Ben Tubby)

Population Growth and Regulation

By the end of this section, you will be able to:

- Explain the characteristics of and differences between exponential and logistic growth patterns
- Give examples of exponential and logistic growth in natural populations
- Give examples of how the carrying capacity of a habitat may change
- Compare and contrast density-dependent growth regulation and density-independent growth regulation giving examples

Population ecologists make use of a variety of methods to model population dynamics. An accurate model should be able to describe the changes occurring in a population and predict future changes.

Population Growth

The two simplest models of population growth use deterministic equations (equations that do not account for random events) to describe the rate of change in the size of a population over time. The first of these models, exponential growth, describes theoretical populations that increase in numbers without any limits to their growth. The second model, logistic growth, introduces limits to reproductive growth that become more intense as the population size increases. Neither model adequately describes natural populations, but they provide points of comparison.

Exponential Growth

Charles Darwin, in developing his theory of natural selection, was influenced by the English clergyman Thomas Malthus. Malthus published his book in 1798 stating that populations with abundant natural resources grow very rapidly; however, they limit further growth by depleting their resources. The early pattern of accelerating population size is called **exponential growth**.

The best example of exponential growth in organisms is seen in bacteria. Bacteria are prokaryotes that reproduce largely by binary fission. This division takes about an hour for many bacterial species. If 1000 bacteria are placed in a large flask with an abundant supply of nutrients (so the nutrients will not become quickly depleted), the number of bacteria will have doubled from 1000 to 2000 after just an hour. In another hour, each of the 2000 bacteria will divide, producing 4000 bacteria. After the third hour, there should be 8000 bacteria in the flask. The important concept of exponential growth is that the growth rate—the number of organisms added in each reproductive generation—is itself increasing; that is, the

population size is increasing at a greater and greater rate. After 24 of these cycles, the population would have increased from 1000 to more than 16 billion bacteria. When the population size, N, is plotted over time, a **J-shaped growth curve** is produced (Figure 6.35a).

The bacteria-in-a-flask example is not truly representative of the real world where resources are usually limited. However, when a species is introduced into a new habitat that it finds suitable, it may show exponential growth for a while. In the case of the bacteria in the flask, some bacteria will die during the experiment and thus not reproduce; therefore, the growth rate is lowered from a maximal rate in which there is no mortality. The growth rate of a population is largely determined by subtracting the **death rate**, D, (number organisms that die during an interval) from the **birth rate**, B, (number organisms that are born during an interval). The growth rate can be expressed in a simple equation that combines the birth and death rates into a single factor: r. This is shown in the following formula:

$$\text{Population growth} = rN$$

The value of r can be positive, meaning the population is increasing in size (the rate of change is positive); or negative, meaning the population is decreasing in size; or zero, in which case the population size is unchanging, a condition known as **zero population growth**.

Logistic Growth

Extended exponential growth is possible only when infinite natural resources are available; this is not the case in the real world. Charles Darwin recognized this fact in his description of the "struggle for existence," which states that individuals will compete (with members of their own or other species) for limited resources. The successful ones are more likely to survive and pass on the traits that made them successful to the next generation at a greater rate (natural selection). To model the reality of limited resources, population ecologists developed the **logistic growth** model.

Carrying Capacity and the Logistic Model

In the real world, with its limited resources, exponential growth cannot continue indefinitely. Exponential growth may occur in environments where there are few individuals and plentiful resources, but when the number of individuals gets large enough, resources will be depleted and the growth rate will slow down. Eventually, the growth rate will plateau or level off (Figure 6.35b). This population size, which is determined by the maximum population size that a particular environment can sustain, is called the **carrying capacity**, or K. In real populations, a growing population often overshoots its carrying capacity, and the death rate increases beyond the birth rate causing the population size to decline back to the carrying capacity or below it. Most populations usually fluctuate around the carrying capacity in an undulating fashion rather than existing right at it.

The formula used to calculate logistic growth adds the carrying capacity as a moderating force in the growth rate. The expression "$K - N$" is equal to the number of individuals that may be added to a population at a given time, and "$K - N$" divided by "K" is the fraction of the carrying capacity available for further growth. Thus, the exponential growth model is restricted by this factor to generate the logistic growth equation:

$$\text{Population growth} = rN \left[\frac{K - N}{K} \right]$$

Notice that when N is almost zero the quantity in brackets is almost equal to 1 (or K/K) and growth is close to exponential. When the population size is equal to the carrying capacity, or $N = K$, the quantity in brackets is equal to zero and growth is equal to zero. A graph of this equation (logistic growth) yields the **S-shaped curve** (Figure 6.35b). It is a more realistic model of population growth than exponential growth. There are three different sections to an S-shaped curve. Initially, growth is exponential because there are few individuals and ample resources available. Then, as resources begin to become limited, the growth rate decreases. Finally, the growth rate levels off at the carrying capacity of the environment, with little change in population number over time.
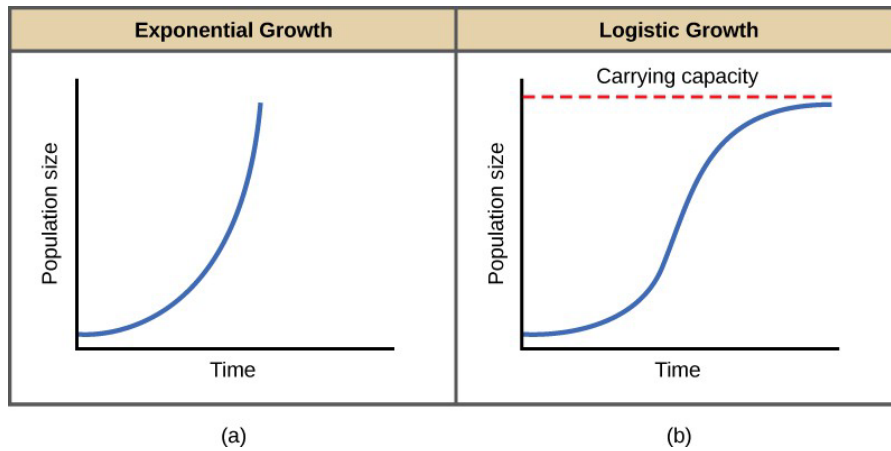
Figure 6.35 When resources are unlimited, populations exhibit (a) exponential growth, shown in a J-shaped curve. When resources are limited, populations exhibit (b) logistic growth. In logistic growth, population expansion decreases as resources become scarce, and it levels off when the carrying capacity of the environment is reached. The logistic growth curve is S-shaped.

Role of Intraspecific Competition

The logistic model assumes that every individual within a population will have equal access to resources and, thus, an equal chance for survival. For plants, the amount of water, sunlight, nutrients, and space to grow are the important resources, whereas in animals, important resources include food, water, shelter, nesting space, and mates.

In the real world, phenotypic variation among individuals within a population means that some individuals will be better adapted to their environment than others. The resulting competition for resources among population members of the same species is termed **intraspecific competition**. Intraspecific competition may not affect populations that are well below their carrying capacity, as resources are plentiful and all individuals can obtain what they need. However, as population size increases, this competition intensifies. In addition, the accumulation of waste products can reduce carrying capacity in an environment.

Examples of Logistic Growth

Yeast, a microscopic fungus used to make bread and alcoholic beverages, exhibits the classical S-shaped curve when grown in a test tube (Figure 6.36**a**). Its growth levels off as the population depletes the nutrients that are necessary for its growth. In the real world, however, there are variations to this idealized curve. Examples in wild populations include sheep and harbor seals (Figure 6.36**b**). In both examples, the population size exceeds the carrying capacity for short periods of time and then falls below the carrying capacity afterwards. This fluctuation in population size continues to occur as the population oscillates around its carrying capacity. Still, even with this oscillation, the logistic model is confirmed.
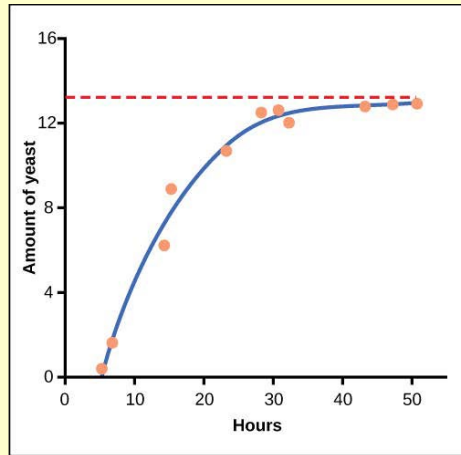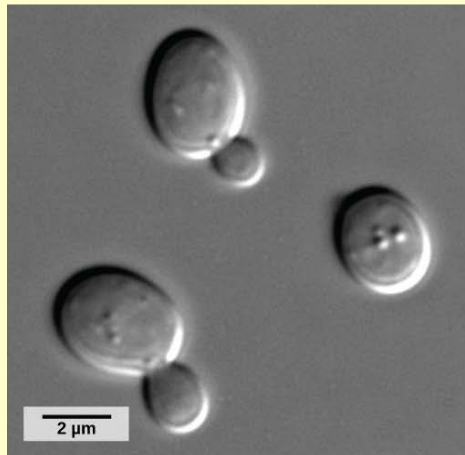
(a)
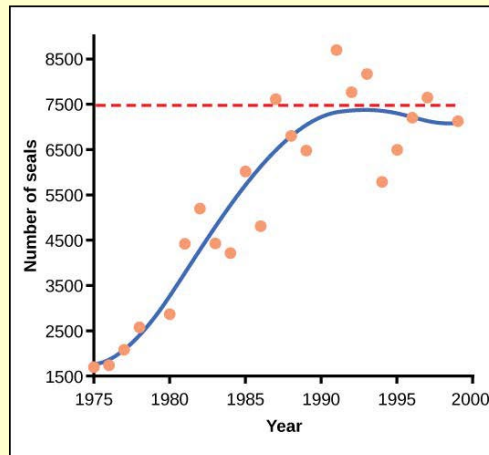
(b)

Figure 6.36 (a) Yeast grown in ideal conditions in a test tube shows a classical S-shaped logistic growth curve, whereas (b) a natural population of seals shows real-world fluctuation. The yeast is visualized using differential interference contrast light micrography. (credit a: scale- bar data from Matt Russell)

Population Dynamics and Regulation

The logistic model of population growth, while valid in many natural populations and a useful model, is a simplification of real-world population dynamics. Implicit in the model is that the carrying capacity of the environment does not change, which is not the case. The carrying capacity varies annually. For example, some summers are hot and dry whereas others are cold and wet; in many areas, the carrying capacity during the winter is much lower than it is during the summer. Also, natural events

253

such as earthquakes, volcanoes, and fires can alter an environment and hence its carrying capacity. Additionally, populations do not usually exist in isolation. They share the environment with other species, competing with them for the same resources (interspecific competition). These factors are also important to understanding how a specific population will grow.

Population growth is regulated in a variety of ways. These are grouped into **density-dependent** factors, in which the density of the population affects growth rate and mortality, and **density-independent** factors, which cause mortality in a population regardless of population density. Wildlife biologists, in particular, want to understand both types because this helps them manage populations and prevent extinction or overpopulation.

Density-dependent Regulation

Most density-dependent factors are biological in nature and include predation, inter- and intraspecific competition, and parasites. Usually, the denser a population is, the greater its mortality rate. For example, during intra- and interspecific competition, the reproductive rates of the species will usually be lower, reducing their populations' rate of growth. In addition, low prey density increases the mortality of its predator because it has more difficulty locating its food source. Also, when the population is denser, diseases spread more rapidly among the members of the population, which affect the mortality rate.

Density dependent regulation was studied in a natural experiment with wild donkey populations on two sites in Australia.[2] On one site the population was reduced by a population control program; the population on the other site received no interference. The high-density plot was twice as dense as the low-density plot. From 1986 to 1987 the high-density plot saw no change in donkey density, while the low-density plot saw an increase in donkey density. The difference in the growth rates of the two populations was caused by mortality, not by a difference in birth rates. The researchers found that numbers of offspring birthed by each mother was unaffected by density. Growth rates in the two populations were different mostly because of juvenile mortality caused by the mother's malnutrition due to scarce high-quality food in the dense population. Figure 6.37 shows the difference in age-specific mortalities in the two populations.
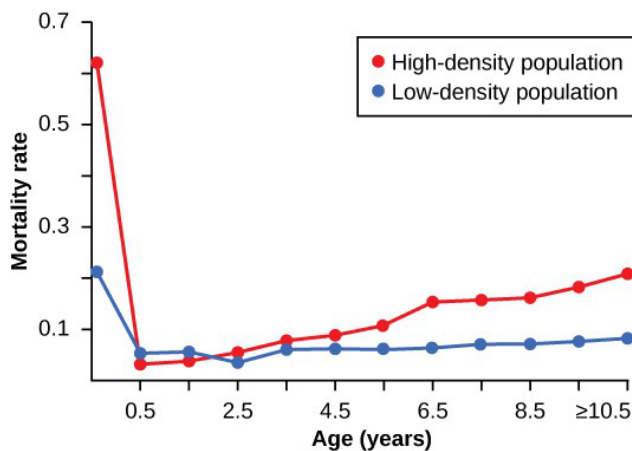
Figure 6.37 This graph shows the age-specific mortality rates for wild donkeys from high- and low-density populations. The juvenile mortality is much higher in the high-density population because of maternal malnutrition caused by a shortage of high-quality food.

Density-independent Regulation and Interaction with Density-dependent Factors

Many factors that are typically physical in nature cause mortality of a population regardless of its density. These factors include weather, natural disasters, and pollution. An individual deer will be killed in a forest fire regardless of how many deer happen to be in that area. Its chances of survival are the same whether the population density is high or low. The same holds true for cold winter weather.

In real-life situations, population regulation is very complicated and density-dependent and independent factors can interact. A dense population that suffers mortality from a density-independent cause will be able to recover differently than a sparse population. For example, a population of deer affected by a harsh winter will recover faster if there are more deer remaining to reproduce.

1. David Choquenot, "Density-Dependent Growth, Body Condition, and Demography in Feral Donkeys: Testing the Food Hypothesis," *Ecology* 72, no. 3 (June 1991):805–813.

evolution IN ACTION

Why Did the Woolly Mammoth Go Extinct?

(a)

(b)

(c)

Figure 6.38 The three images include: (a) 1916 mural of a mammoth herd from the American Museum of Natural History, (b) the only stuffed mammoth in the world is in the Museum of Zoology located in St. Petersburg, Russia, and (c) a one-month-old baby mammoth, named Lyuba, discovered in Siberia in 2007. (credit a: modification of work by Charles R. Knight; credit b: modification of work by "Tanapon"/Flickr; credit c: modification of work by Matt Howry)

Woolly mammoths began to go extinct about 10,000 years ago, soon after paleontologists believe humans able to hunt them began to colonize North America and northern Eurasia (Figure 6.38). A mammoth population survived on Wrangel Island, in the East Siberian Sea, and was isolated from human contact until as recently as 1700 BC. We know a lot about these animals from carcasses found frozen in the ice of Siberia and other northern regions.

It is commonly thought that climate change and human hunting led to their extinction. A 2008 study estimated that climate change reduced the mammoth's range from 3,000,000 square miles 42,000 years ago to 310,000 square miles 6,000 years ago.[3] Through archaeological evidence of kill sites, it is also well documented that humans hunted these animals. A 2012 study concluded that no single factor was exclusively responsible for the extinction of these magnificent creatures.[4] In addition to climate change and reduction of habitat, scientists demonstrated another important factor in the mammoth's extinction was the migration of human hunters across the Bering Strait to North America during the last ice age 20,000 years ago.

The maintenance of stable populations was and is very complex, with many interacting factors determining the outcome. It is important to remember that humans are also part of nature. Once we contributed to a species' decline using primitive hunting technology only.

---

2. David Nogués-Bravo et al., "Climate Change, Humans, and the Extinction of the Woolly Mammoth." *PLoS Biol* 6 (April 2008): e79, doi:10.1371/journal.pbio.0060079.
3. G.M. MacDonald et al., "Pattern of Extinction of the Woolly Mammoth in Beringia." *Nature Communications* 3, no. 893 (June 2012), doi:10.1038/ncomms1881.

Demographic-Based Population Models

Population ecologists have hypothesized that suites of characteristics may evolve in species that lead to particular adaptations to their environments. These adaptations impact the kind of population growth their species experience. Life history characteristics such as birth rates, age at first reproduction, the numbers of offspring, and even death rates evolve just like anatomy or behavior, leading to adaptations that affect population growth. Population ecologists have described a continuum of life-history "strategies" with *K*-selected species on one end and *r*-selected species on the other. **K-selected species** are adapted to stable, predictable environments. Populations of *K*-selected species tend to exist close to their carrying capacity. These species tend to have larger, but fewer, offspring and contribute large amounts of resources to each offspring. Elephants would be an example of a *K*-selected species. ***r*-selected species** are adapted to unstable and unpredictable environments. They have large numbers of small offspring. Animals that are *r*-selected do not provide a lot of resources or parental care to offspring, and the offspring are relatively self-sufficient at birth. Examples of *r*-selected species are marine invertebrates such as jellyfish and plants such as the dandelion. The two extreme strategies are at two ends of a continuum on which real species life histories will exist. In addition, life history strategies do not need to evolve as suites, but can evolve independently of each other, so each species may have some characteristics that trend toward one extreme or the other.

| The Human Population

By the end of this section, you will be able to:

- Discuss how human population growth can be exponential
- Explain how humans have expanded the carrying capacity of their habitat
- Relate population growth and age structure to the level of economic development in different countries
- Discuss the long-term implications of unchecked human population growth

Concepts of animal population dynamics can be applied to human population growth. Humans are not unique in their ability to alter their environment. For example, beaver dams alter the stream environment where they are built. Humans, however, have the ability to alter their environment to increase its carrying capacity, sometimes to the detriment of other species. Earth's human population and their use of resources are growing rapidly, to the extent that some worry about the ability of Earth's environment to sustain its human population. Long-term exponential growth carries with it the potential risks of famine, disease, and large-scale death, as well as social consequences of crowding such as increased crime.

Human technology and particularly our harnessing of the energy contained in fossil fuels have caused unprecedented changes to Earth's environment, altering ecosystems to the point where some may be in danger of collapse. Changes on a global scale including depletion of the ozone layer, desertification and topsoil loss, and global climate change are caused by human activities.

The world's human population is presently growing exponentially (Figure 6.39).

256
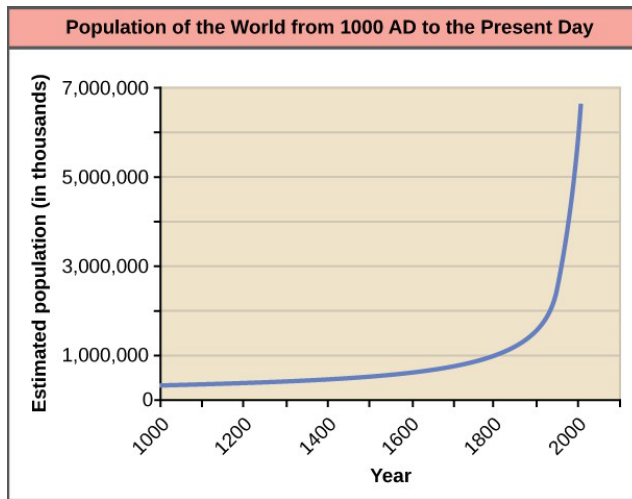
Figure 6.39 Human population growth since 1000 AD is exponential.

A consequence of exponential growth rate is that the time that it takes to add a particular number of humans to the population is becoming shorter. Figure 6.40 shows that 123 years were necessary to add 1 billion humans between 1804 and 1930, but it only took 24 years to add the two billion people between 1975 and 1999. This acceleration in growth rate will likely begin to decrease in the coming decades. Despite this, the population will continue to increase and the threat of overpopulation remains, particularly because the damage caused to ecosystems and biodiversity is lowering the human carrying capacity of the planet.
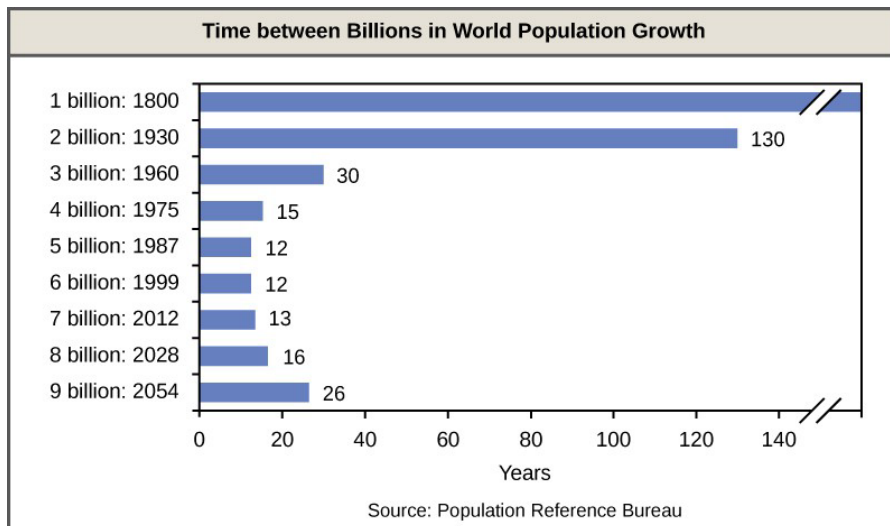
Figure 6.40 The time between the addition of each billion human beings to Earth decreases over time. (credit: modification of work by Ryan T. Cragun)

Click through this **interactive view (http://openstaxcollege.org/l/human_growth2)** of how human populations have changed over time.

257

Overcoming Density-Dependent Regulation

Humans are unique in their ability to alter their environment in myriad ways. This ability is responsible for human population growth because it resets the carrying capacity and overcomes density-dependent growth regulation. Much of this ability is related to human intelligence, society, and communication. Humans construct shelters to protect themselves from the elements and have developed agriculture and domesticated animals to increase their food supplies. In addition, humans use language to communicate this technology to new generations, allowing them to improve upon previous accomplishments.

Other factors in human population growth are migration and public health. Humans originated in Africa, but we have since migrated to nearly all inhabitable land on Earth, thus, increasing the area that we have colonized. Public health, sanitation, and the use of antibiotics and vaccines have decreased the ability of infectious disease to limit human population growth in developed countries. In the past, diseases such as the bubonic plaque of the fourteenth century killed between 30 and 60 percent of Europe's population and reduced the overall world population by as many as one hundred million people. Infectious disease continues to have an impact on human population growth. For example, life expectancy in sub-Saharan Africa, which was increasing from 1950 to 1990, began to decline after 1985 largely as a result of HIV/AIDS mortality. The reduction in life expectancy caused by HIV/AIDS was estimated to be 7 years for 2005.[5]

Declining life expectancy is an indicator of higher mortality rates and leads to lower birth rates.

The fundamental cause of the acceleration of growth rate for humans in the past 200 years has been the reduced death rate due to a development of the technological advances of the industrial age, urbanization that supported those technologies, and especially the exploitation of the energy in fossil fuels. Fossil fuels are responsible for dramatically increasing the resources available for human population growth through agriculture (mechanization, pesticides, and fertilizers) and harvesting wild populations.

Age Structure, Population Growth, and Economic Development

The age structure of a population is an important factor in population dynamics. **Age structure** is the proportion of a population in different age classes. Models that incorporate age structure allow better prediction of population growth, plus the ability to associate this growth with the level of economic development in a region. Countries with rapid growth have a pyramidal shape in their age structure diagrams, showing a preponderance of younger individuals, many of whom are of reproductive age (Figure 6.41). This pattern is most often observed in underdeveloped countries where individuals do not live to old age because of less-than-optimal living conditions, and there is a high birth rate. Age structures of areas with slow growth, including developed countries such as the United States, still have a pyramidal structure, but with many fewer young and reproductive-aged individuals and a greater proportion of older individuals. Other developed countries, such as Italy, have zero population growth. The age structure of these populations is more conical, with an even greater percentage of middle-aged and older individuals. The actual growth rates in different countries are shown in Figure 6.42, with the highest rates tending to be in the less economically developed countries of Africa and Asia.

---

4. Danny Dorling, Mary Shaw, and George Davey Smith, "Global Inequality of Life Expectancy due to AIDS," *BMJ* 332, no. 7542 (March 2006): 662-664, doi: 10.1136/bmj.332.7542.662.
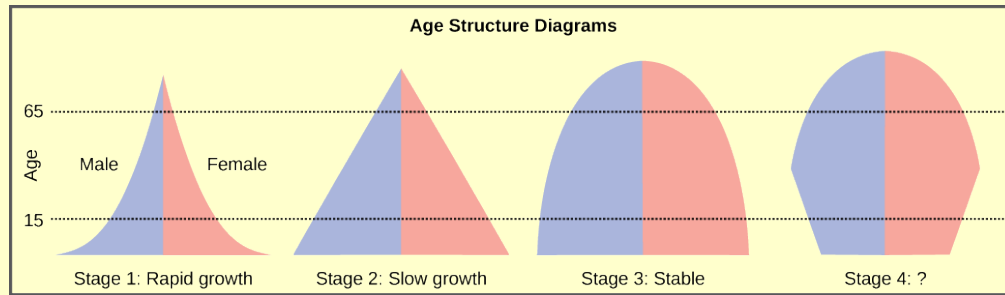
Figure 6.41 Typical age structure diagrams are shown. The rapid growth diagram narrows to a point, indicating that the number of individuals decreases rapidly with age. In the slow growth model, the number of individuals decreases steadily with age. Stable population diagrams are rounded on the top, showing that the number of individuals per age group decreases gradually, and then increases for the older part of the population.
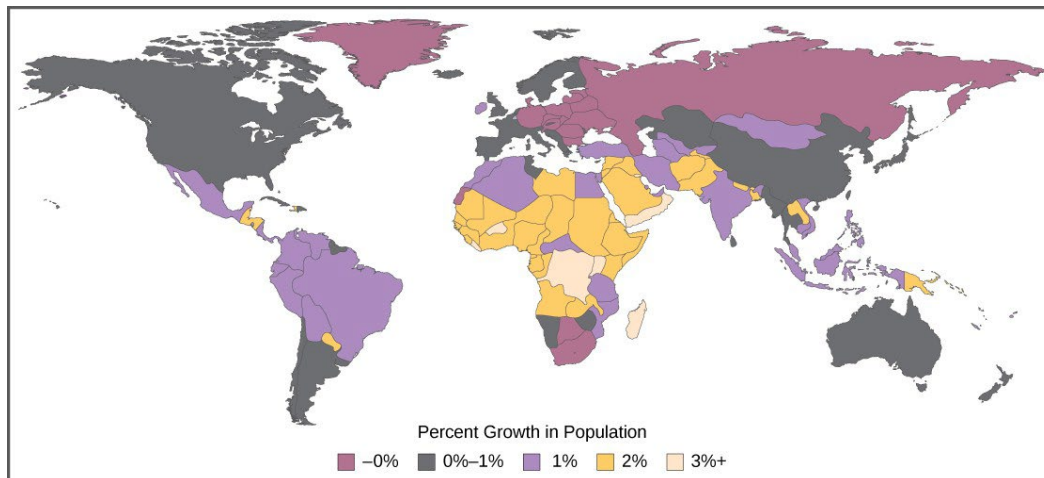
Figure 6.42 The percent growth rate of population in different countries is shown. Notice that the highest growth is occurring in less economically developed countries in Africa and Asia.

Long-Term Consequences of Exponential Human Population Growth

Many dire predictions have been made about the world's population leading to a major crisis called the "population explosion." In the 1968 book *The Population Bomb*, biologist Dr. Paul R. Ehrlich wrote, "The battle to feed all of humanity is over. In the 1970s hundreds of millions of people will starve to death in spite of any crash programs embarked upon now. At this late date nothing can prevent a substantial increase in the world death rate."[6] While many critics view this statement as an exaggeration, the laws of exponential population growth are still in effect, and unchecked human population growth cannot continue indefinitely.

Efforts to moderate population control led to the **one-child policy** in China, which imposes fines on urban couples who have more than one child. Due to the fact that some couples wish to have a male heir, many Chinese couples continue to have more than one child. The effectiveness of the policy in limiting overall population growth is controversial, as is the policy itself. Moreover, there are stories of female

5. Paul R. Erlich, prologue to *The Population Bomb*, (1968; repr., New York: Ballantine, 1970).

infanticide having occurred in some of the more rural areas of the country. Family planning education programs in other countries have had highly positive effects on limiting population growth rates and increasing standards of living. In spite of population control policies, the human population continues to grow. Because of the subsequent need to produce more and more food to feed our population, inequalities in access to food and other resources will continue to widen. The United Nations estimates the future world population size could vary from 6 billion (a decrease) to 16 billion people by the year 2100. There is no way to know whether human population growth will moderate to the point where the crisis described by Dr. Ehrlich will be averted.

Another consequence of population growth is the change and degradation of the natural environment. Many countries have attempted to reduce the human impact on climate change by limiting their emission of greenhouse gases. However, a global climate change treaty remains elusive, and many underdeveloped countries trying to improve their economic condition may be less likely to agree with such provisions without compensation if it means slowing their economic development. Furthermore, the role of human activity in causing climate change has become a hotly debated socio-political issue in some developed countries, including the United States. Thus, we enter the future with considerable uncertainty about our ability to curb human population growth and protect our environment to maintain the carrying capacity for the human species.

CONCEPT in ACTION

Visit this **website (http://openstaxcollege.org/l/populations2)** and select "Launch the movie" for an animation discussing the global impacts of human population growth.

Characteristics of Communities

Communities are complex systems that can be characterized by their structure (the number and size of populations and their interactions) and dynamics (how the members and their interactions change over time). Understanding community structure and dynamics allows us to minimize impacts on ecosystems and manage ecological communities we benefit from.

Biodiversity

Ecologists have extensively studied one of the fundamental characteristics of communities: biodiversity. One measure of biodiversity used by ecologists is the number of different species in a particular area and their relative abundance. The area in question could be a habitat, a biome, or the entire biosphere.

Species richness is the term used to describe the number of species living in a habitat or other unit. Species richness varies across the globe (Figure 6.43). Ecologists have struggled to understand the determinants of biodiversity. Species richness is related to latitude: the greatest species richness occurs near the equator and the lowest richness occurs near the poles. Other factors influence species richness as well. **Island biogeography** attempts to explain the great species richness found in isolated islands, and has found relationships between species richness, island size, and distance from the mainland.

Relative species abundance is the number individuals in a species relative to the total number of individuals in all species within a system. Foundation species, described below, often have the highest relative abundance of species.
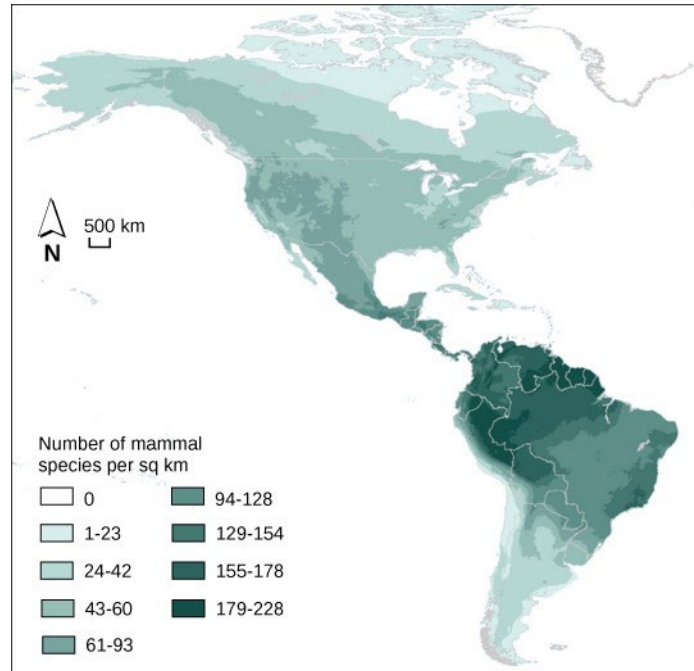
Figure 6.43 The greatest species richness for mammals in North America is associated in the equatorial latitudes. (credit: modification of work by NASA, CIESIN, Columbia University)

Foundation Species

Foundation species are considered the "base" or "bedrock" of a community, having the greatest influence on its overall structure. They are often primary producers, and they are typically an abundant organism. For example, kelp, a species of brown algae, is a foundation species that forms the basis of the kelp forests off the coast of California.

Foundation species may physically modify the environment to produce and maintain habitats that benefit the other organisms that use them. Examples include the kelp described above or tree species found in a forest. The photosynthetic corals of the coral reef also provide structure by physically modifying the environment (Figure 6.44). The exoskeletons of living and dead coral make up most of the reef structure, which protects many other species from waves and ocean currents.

261

Figure 6.44 Coral is the foundation species of coral reef ecosystems. (credit: Jim E. Maragos, USFWS)

Keystone Species

A **keystone species** is one whose presence has inordinate influence in maintaining the prevalence of various species in an ecosystem, the ecological community's structure, and sometimes its biodiversity. *Pisaster ochraceus*, the intertidal sea star, is a keystone species in the northwestern portion of the United States (Figure 6.45). Studies have shown that when this organism is removed from communities, mussel populations (their natural prey) increase, which completely alters the species composition and reduces biodiversity. Another keystone species is the banded tetra, a fish in tropical streams, which supplies nearly all of the phosphorus, a necessary inorganic nutrient, to the rest of the community. The banded tetra feeds largely on insects from the terrestrial ecosystem and then excretes phosphorus into the aquatic ecosystem. The relationships between populations in the community, and possibly the biodiversity, would change dramatically if these fish were to become extinct.

Figure 6.45 The *Pisaster ochraceus* sea star is a keystone species. (credit: Jerry Kirkhart)

262

Invasive Species

Invasive species are non-native organisms that, when introduced to an area out of its native range, alter the community they invade. In the United States, invasive species like the purple loosestrife (*Lythrum salicaria*) and the zebra mussel (*Dreissena polymorpha*) have altered aquatic ecosystems, and some forests are threatened by the spread of common buckthorn (*Rhamnus cathartica*) and garlic mustard (*Alliaria petiolata*). Some well-known invasive animals include the emerald ash borer (*Agrilus planipennis*) and the European starling (*Sturnus vulgaris*). Whether enjoying a forest hike, taking a summer boat trip, or simply walking down an urban street, you have likely encountered an invasive species.

One of the many recent proliferations of an invasive species concerns the Asian carp in the United States. Asian carp were introduced to the United States in the 1970s by fisheries (commercial catfish ponds) and by sewage treatment facilities that used the fish's excellent filter feeding abilities to clean their ponds of excess plankton. Some of the fish escaped, and by the 1980s they had colonized many waterways of the Mississippi River basin, including the Illinois and Missouri Rivers.

Voracious feeders and rapid reproducers, Asian carp may outcompete native species for food and could lead to their extinction. One species, the grass carp, feeds on phytoplankton and aquatic plants. It competes with native species for these resources and alters nursery habitats for other fish by removing aquatic plants. Another species, the silver carp, competes with native fish that feed on zooplankton. In some parts of the Illinois River, Asian carp constitute 95 percent of the community's biomass. Although edible, the fish is bony and not desired in the United States. Moreover, their presence now threatens the native fish and fisheries of the Great Lakes, which are important to local economies and recreational anglers. Asian carp have even injured humans. The fish, frightened by the sound of approaching motorboats, thrust themselves into the air, often landing in the boat or directly hitting boaters.

The Great Lakes and their prized salmon and lake trout fisheries are being threatened by Asian carp. The carp are not yet present in the Great Lakes, and attempts are being made to prevent its access to the lakes through the Chicago Ship and Sanitary Canal, which is the only connection between the Mississippi River and Great Lakes basins. To prevent the Asian carp from leaving the canal, a series of electric barriers have been used to discourage their migration; however, the threat is significant enough that several states and Canada have sued to have the Chicago channel permanently cut off from Lake Michigan. Local and national politicians have weighed in on how to solve the problem. In general, governments have been ineffective in preventing or slowing the introduction of invasive species.

The issues associated with Asian carp show how population and community ecology, fisheries management, and politics intersect on issues of vital importance to the human food supply and economy. Socio-political issues like the Asian carp make extensive use of the sciences of population ecology, the study of members of a particular species occupying a habitat; and community ecology, the study of the interaction of all species within a habitat.

Community Dynamics

Community dynamics are the changes in community structure and composition over time, often following **environmental disturbances** such as volcanoes, earthquakes, storms, fires, and climate change. Communities with a relatively constant number of species are said to be at equilibrium. The equilibrium is dynamic with species identities and relationships changing over time, but maintaining relatively constant numbers. Following a disturbance, the community may or may not return to the equilibrium state.

Succession describes the sequential appearance and disappearance of species in a community over time after a severe disturbance. In **primary succession**, newly exposed or newly formed rock is colonized

by living organisms; in **secondary succession**, a part of an ecosystem is disturbed and remnants of the previous community remain. In both cases, there is a sequential change in species until a more or less permanent community develops.

Primary Succession and Pioneer Species

Primary succession occurs when new land is formed, for example, following the eruption of volcanoes, such as those on the Big Island of Hawaii. As lava flows into the ocean, new land is continually being formed. On the Big Island, approximately 32 acres of land is added to it its size each year. Weathering and other natural forces break down the rock enough for the establishment of hearty species such as lichens and some plants, known as **pioneer species** (Figure 6.46). These species help to further break down the mineral-rich lava into soil where other, less hardy but more competitive species, such as grasses, shrubs, and trees, will grow and eventually replace the pioneer species. Over time the area will reach an equilibrium state, with a set of organisms quite different from the pioneer species.

Figure 6.46 During primary succession in lava on Maui, Hawaii, succulent plants are the pioneer species. (credit: Forest and Kim Starr)

Secondary succession

A classic example of secondary succession occurs in oak and hickory forests cleared by wildfire (**Figure 6.47**). Wildfires will burn most vegetation, and unless the animals can flee the area, they are killed. Their nutrients, however, are returned to the ground in the form of ash. Thus, although the community has been dramatically altered, there is a soil ecosystem present that provides a foundation for rapid recolonization.

Before the fire, the vegetation was dominated by tall trees with access to the major plant energy resource: sunlight. Their height gave them access to sunlight while also shading the ground and other low-lying species. After the fire, though, these trees are no longer dominant. Thus, the first plants to grow back are usually annual plants followed within a few years by quickly growing and spreading grasses and other pioneer species. Due, at least in part, to changes in the environment brought on by the growth of grasses and forbs, over many years, shrubs emerge along with small pine, oak, and hickory trees. These organisms are called intermediate species. Eventually, over 150 years, the forest will reach its equilibrium point and resemble the community before the fire. This equilibrium state is referred to as the **climax community**, which will remain until the next disturbance. The climax community is typically characteristic of a given climate and geology. Although the community in equilibrium looks the same once it is attained, the equilibrium is a dynamic one with constant changes in abundance and sometimes species identities. The return of a natural ecosystem after agricultural activities is also a well-documented secondary succession process.

Secondary Succession of an Oak and Hickory Forest

Pioneer species
Annual plants grow and are succeeded by grasses and perennials.

Intermediate species
Shrubs, then pines, and young oak and hickory begin to grow.

Climax community
The mature oak and hickory forest remains stable until the next disturbance.
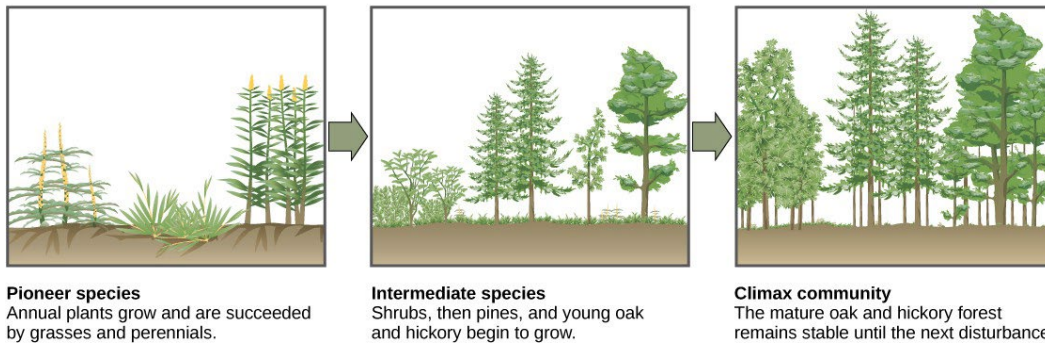
Figure 6.47 Secondary succession is seen in an oak and hickory forest after a forest fire. A sequence of the community present at three successive times at the same location is depicted.

Importance of Biodiversity

By the end of this section, you will be able to:

- Describe biodiversity as the equilibrium of naturally fluctuating rates of extinction and speciation
- Identify benefits of biodiversity to humans

Figure 6.48 This tropical lowland rainforest in Madagascar is an example of a high biodiversity habitat. This particular location is protected within a national forest, yet only 10 percent of the original coastal lowland forest remains, and research suggests half the original biodiversity has been lost. (credit: Frank Vassen)

Biodiversity is a broad term for biological variety, and it can be measured at a number of organizational levels. Traditionally, ecologists have measured **biodiversity** by taking into account both the number of species and the number of individuals in each of those species. However, biologists are using measures of biodiversity at several levels of biological organization (including genes, populations, and ecosystems) to help focus efforts to preserve the biologically and technologically important elements of biodiversity.

When biodiversity loss through extinction is thought of as the loss of the passenger pigeon, the dodo, or, even, the woolly mammoth there seems to be no reason to care about it because these events happened long ago. How is the loss practically important for the welfare of the human species? Would these species have made our lives any better? From the perspective of evolution and ecology, the loss of a particular individual species, with some exceptions, may seem unimportant, but the current accelerated extinction rate means the loss of tens of thousands of species within our lifetimes. Much of this loss is occurring in tropical rainforests like the one pictured in Figure 6.48, which are especially high-diversity ecosystems that are being cleared for timber and agriculture. This is likely to have dramatic effects on human welfare through the collapse of ecosystems and in added costs to maintain food production, clean air and water, and improve human health.

Biologists recognize that human populations are embedded in ecosystems and are dependent on them, just as is every other species on the planet. Agriculture began after early hunter-gatherer societies first settled in one place and heavily modified their immediate environment: the ecosystem in which they existed. This cultural transition has made it difficult for humans to recognize their dependence on living

things other than crops and domesticated animals on the planet. Today our technology smoothes out the extremes of existence and allows many of us to live longer, more comfortable lives, but ultimately the human species cannot exist without its surrounding ecosystems. Our ecosystems provide our food. This includes living plants that grow in soil ecosystems and the animals that eat these plants (or other animals) as well as photosynthetic organisms in the oceans and the other organisms that eat them. Our ecosystems have provided and will provide many of the medications that maintain our health, which are commonly made from compounds found in living organisms. Ecosystems provide our clean water, which is held in lake and river ecosystems or passes through terrestrial ecosystems on its way into groundwater.

Types of Biodiversity

A common meaning of biodiversity is simply the number of species in a location or on Earth; for example, the American Ornithologists' Union lists 2078 species of birds in North and Central America. This is one measure of the bird biodiversity on the continent. More sophisticated measures of diversity take into account the relative abundances of species. For example, a forest with 10 equally common species of trees is more diverse than a forest that has 10 species of trees wherein just one of those species makes up 95 percent of the trees rather than them being equally distributed. Biologists have also identified alternate measures of biodiversity, some of which are important in planning how to preserve biodiversity.

Genetic and Chemical Biodiversity

Genetic diversity is one alternate concept of biodiversity. **Genetic diversity** (or variation) is the raw material for adaptation in a species. A species' future potential for adaptation depends on the genetic diversity held in the genomes of the individuals in populations that make up the species. The same is true for higher taxonomic categories. A genus with very different types of species will have more genetic diversity than a genus with species that look alike and have similar ecologies. The genus with the greatest potential for subsequent evolution is the most genetically diverse one.

Most genes code for proteins, which in turn carry out the metabolic processes that keep organisms alive and reproducing. Genetic diversity can also be conceived of as **chemical diversity** in that species with different genetic makeups produce different assortments of chemicals in their cells (proteins as well as the products and byproducts of metabolism). This chemical diversity is important for humans because of the potential uses for these chemicals, such as medications. For example, the drug eptifibatide is derived from rattlesnake venom and is used to prevent heart attacks in individuals with certain heart conditions.

At present, it is far cheaper to discover compounds made by an organism than to imagine them and then synthesize them in a laboratory. Chemical diversity is one way to measure diversity that is important to human health and welfare. Through selective breeding, humans have domesticated animals, plants, and fungi, but even this diversity is suffering losses because of market forces and increasing globalism in human agriculture and migration. For example, international seed companies produce only a very few varieties of a given crop and provide incentives around the world for farmers to buy these few varieties while abandoning their traditional varieties, which are far more diverse. The human population depends on crop diversity directly as a stable food source and its decline is troubling to biologists and agricultural scientists.

Ecosystems Diversity

It is also useful to define **ecosystem diversity**: the number of different ecosystems on Earth or in a geographical area. Whole ecosystems can disappear even if some of the species might survive by adapting to other ecosystems. The loss of an ecosystem means the loss of the interactions between species, the loss of unique features of coadaptation, and the loss of biological productivity that an ecosystem is able to create. An example of a largely extinct ecosystem in North America is the prairie ecosystem (Figure 6.48). Prairies once spanned central North America from the boreal forest in northern Canada down into Mexico. They are now all but gone, replaced by crop fields, pasture lands, and suburban sprawl. Many of the species survive, but the hugely productive ecosystem that was responsible for creating our most productive agricultural soils is now gone. As a consequence, their soils are now being depleted unless they are maintained artificially at greater expense. The decline in soil productivity occurs because the interactions in the original ecosystem have been lost; this was a far more important loss than the relatively few species that were driven extinct when the prairie ecosystem was destroyed.

Figure 6.49 The variety of ecosystems on Earth—from coral reef to prairie—enables a great diversity of species to exist. (credit "coral reef": modification of work by Jim Maragos, USFWS; credit: "prairie": modification of work by Jim Minnerath, USFWS)

Current Species Diversity

Despite considerable effort, knowledge of the species that inhabit the planet is limited. A recent estimate suggests that the eukaryote species for which science has names, about 1.5 million species, account for less than 20 percent of the total number of eukaryote species present on the planet (8.7 million species, by one estimate). Estimates of numbers of prokaryotic species are largely guesses, but biologists agree that science has only just begun to catalog their diversity. Even with what is known, there is no centralized repository of names or samples of the described species; therefore, there is no way to be sure that the 1.5 million descriptions is an accurate number. It is a best guess based on the opinions of experts on different taxonomic groups. Given that Earth is losing species at an accelerating pace, science knows little about what is being lost. Table 6.1 presents recent estimates of biodiversity in different groups.

Estimated Numbers of Described and Predicted species

	Source: Mora et al 2011		Source: Chapman 2009		Source: Groombridge and Jenkins 2002	
	Described	Predicted	Described	Predicted	Described	Predicted
Animals	1,124,516	9,920,000	1,424,153	6,836,330	1,225,500	10,820,000
Photosynthetic protists	17,892	34,900	25,044	200,500	—	—
Fungi	44,368	616,320	98,998	1,500,000	72,000	1,500,000
Plants	224,244	314,600	310,129	390,800	270,000	320,000
Non-photosynthetic protists	16,236	72,800	28,871	1,000,000	80,000	600,000
Prokaryotes	—	—	10,307	1,000,000	10,175	—
Total	1,438,769	10,960,000	1,897,502	10,897,630	1,657,675	13,240,000

Table 6.1 This table shows the estimated number of species by taxonomic group—including both described (named and studied) and predicted (yet to be named) species.

There are various initiatives to catalog described species in accessible and more organized ways, and the internet is facilitating that effort. Nevertheless, at the current rate of species description, which according to the State of Observed Species[1] reports is 17,000–20,000 new species a year, it would take close to 500 years to describe all of the species currently in existence. The task, however, is becoming increasingly impossible over time as **extinction** removes species from Earth faster than they can be described.

---

1. International Institute for Species Exploration (IISE), *2011 State of Observed Species (SOS)*. Tempe, AZ: IISE, 2011. Accessed May, 20, 2012. http://species.asu.edu/SOS.

Naming and counting species may seem an unimportant pursuit given the other needs of humanity, but it is not simply an accounting. Describing species is a complex process by which biologists determine an organism's unique characteristics and whether or not that organism belongs to any other described species. It allows biologists to find and recognize the species after the initial discovery to follow up on questions about its biology. That subsequent research will produce the discoveries that make the species valuable to humans and to our ecosystems. Without a name and description, a species cannot be studied in depth and in a coordinated way by multiple scientists.

Patterns of Biodiversity

Biodiversity is not evenly distributed on the planet. Lake Victoria contained almost 500 species of cichlids (only one family of fishes present in the lake) before the introduction of an exotic species in the 1980s and 1990s caused a mass extinction. All of these species were found only in Lake Victoria, which is to say they were endemic. **Endemic species** are found in only one location. For example, the blue jay is endemic to North America, while the Barton Springs salamander is endemic to the mouth of one spring in Austin, Texas. Endemics with highly restricted distributions, like the Barton Springs salamander, are particularly vulnerable to extinction. Higher taxonomic levels, such as genera and families, can also be endemic.

Lake Huron contains about 79 species of fish, all of which are found in many other lakes in North America. What accounts for the difference in diversity between Lake Victoria and Lake Huron? Lake Victoria is a tropical lake, while Lake Huron is a temperate lake. Lake Huron in its present form is only about 7,000 years old, while Lake Victoria in its present form is about 15,000 years old. These two factors, latitude and age, are two of several hypotheses biogeographers have suggested to explain biodiversity patterns on Earth.

ca**reers** IN ACTION

Biogeography

Biogeography is the study of the distribution of the world's species both in the past and in the present. The work of biogeographers is critical to understanding our physical environment, how the environment affects species, and how changes in environment impact the distribution of a species.

There are three main fields of study under the heading of biogeography: ecological biogeography, historical biogeography (called paleobiogeography), and conservation biogeography. Ecological biogeography studies the current factors affecting the distribution of plants and animals. Historical biogeography, as the name implies, studies the past distribution of species. Conservation biogeography, on the other hand, is focused on the protection and restoration of species based upon the known historical and current ecological information. Each of these fields considers both zoogeography and phytogeography—the past and present distribution of animals and plants.

One of the oldest observed patterns in ecology is that biodiversity in almost every taxonomic group of organism increases as latitude declines. In other words, biodiversity increases closer to the equator (Figure 6.50).
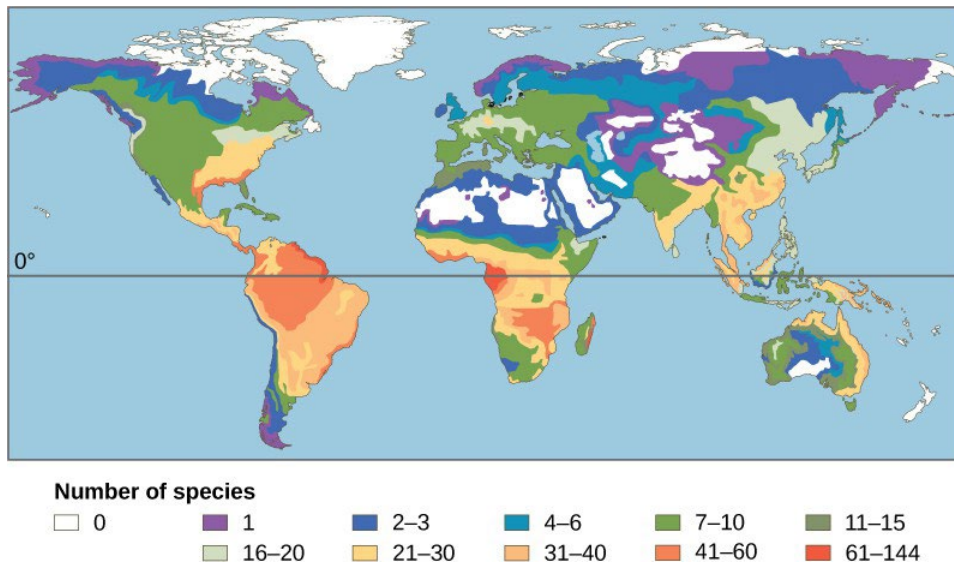
Figure 6.50 This map illustrates the number of amphibian species across the globe and shows the trend toward higher biodiversity at lower latitudes. A similar pattern is observed for most taxonomic groups.

It is not yet clear why biodiversity increases closer to the equator, but hypotheses include the greater age of the ecosystems in the tropics versus temperate regions, which were largely devoid of life or drastically impoverished during the last ice age. The greater age provides more time for speciation. Another possible explanation is the greater energy the tropics receive from the sun versus the lesser energy input in temperate and polar regions. But scientists have not been able to explain how greater energy input could translate into more species. The complexity of tropical ecosystems may promote speciation by increasing the **habitat heterogeneity**, or number of ecological niches, in the tropics relative to higher latitudes. The greater heterogeneity provides more opportunities for coevolution, specialization, and perhaps greater selection pressures leading to population differentiation. However, this hypothesis suffers from some circularity—ecosystems with more species encourage speciation, but how did they get more species to begin with? The tropics have been perceived as being more stable than temperate regions, which have a pronounced climate and day-length seasonality. The tropics have their own forms of seasonality, such as rainfall, but they are generally assumed to be more stable environments and this stability might promote speciation.

Regardless of the mechanisms, it is certainly true that biodiversity is greatest in the tropics. The number of endemic species is higher in the tropics. The tropics also contain more biodiversity hotspots. At the same time, our knowledge of the species living in the tropics is lowest and because of recent, heavy human activity the potential for biodiversity loss is greatest.

Importance of Biodiversity

Loss of biodiversity eventually threatens other species we do not impact directly because of their interconnectedness; as species disappear from an ecosystem other species are threatened by the changes in available resources. Biodiversity is important to the survival and welfare of human populations because it has impacts on our health and our ability to feed ourselves through agriculture and harvesting populations of wild animals.

Human Health

Many medications are derived from natural chemicals made by a diverse group of organisms. For example, many plants produce **secondary plant compounds**, which are toxins used to protect the plant from insects and other animals that eat them. Some of these secondary plant compounds also work as human medicines. Contemporary societies that live close to the land often have a broad knowledge of the medicinal uses of plants growing in their area. For centuries in Europe, older knowledge about the medical uses of plants was compiled in herbals—books that identified the plants and their uses. Humans are not the only animals to use plants for medicinal reasons. The other great apes, orangutans, chimpanzees, bonobos, and gorillas have all been observed self-medicating with plants.

Modern pharmaceutical science also recognizes the importance of these plant compounds. Examples of significant medicines derived from plant compounds include aspirin, codeine, digoxin, atropine, and vincristine (Figure 6.51). Many medications were once derived from plant extracts but are now synthesized. It is estimated that, at one time, 25 percent of modern drugs contained at least one plant

270

extract. That number has probably decreased to about 10 percent as natural plant ingredients are replaced by synthetic versions of the plant compounds. Antibiotics, which are responsible for extraordinary improvements in health and lifespans in developed countries, are compounds largely derived from fungi and bacteria.

Figure 6.51 *Catharanthus roseus*, the Madagascar periwinkle, has various medicinal properties. Among other uses, it is a source of vincristine, a drug used in the treatment of lymphomas. (credit: Forest and Kim Starr)

In recent years, animal venoms and poisons have excited intense research for their medicinal potential. By 2007, the FDA had approved five drugs based on animal toxins to treat diseases such as hypertension, chronic pain, and diabetes. Another five drugs are undergoing clinical trials and at least six drugs are being used in other countries. Other toxins under investigation come from mammals, snakes, lizards, various amphibians, fish, snails, octopuses, and scorpions.

Aside from representing billions of dollars in profits, these medications improve people's lives. Pharmaceutical companies are actively looking for new natural compounds that can function as medicines. It is estimated that one third of pharmaceutical research and development is spent on natural compounds and that about 35 percent of new drugs brought to market between 1981 and 2002 were from natural compounds.

Finally, it has been argued that humans benefit psychologically from living in a biodiverse world. The chief proponent of this idea is entomologist E. O. Wilson. He argues that human evolutionary history has adapted us to living in a natural environment and that built environments generate stresses that affect human health and well-being. There is considerable research into the psychologically regenerative benefits of natural landscapes that suggest the hypothesis may hold some truth.

Agricultural

Since the beginning of human agriculture more than 10,000 years ago, human groups have been breeding and selecting crop varieties. This crop diversity matched the cultural diversity of highly subdivided populations of humans. For example, potatoes were domesticated beginning around 7,000 years ago in the central Andes of Peru and Bolivia. The people in this region traditionally lived in relatively isolated settlements separated by mountains. The potatoes grown in that region belong to seven species and the number of varieties likely is in the thousands. Each variety has been bred to thrive at particular elevations and soil and climate conditions. The diversity is driven by the diverse demands of the dramatic elevation changes, the limited movement of people, and the demands created by crop rotation for different varieties that will do well in different fields.

Potatoes are only one example of agricultural diversity. Every plant, animal, and fungus that has been cultivated by humans has been bred from original wild ancestor species into diverse varieties arising from the demands for food value, adaptation to growing conditions, and resistance to pests. The potato demonstrates a well-known example of the risks of low crop diversity: during the tragic Irish potato famine (1845–1852 AD), the single potato variety grown in Ireland became susceptible to a potato blight—wiping out the crop. The loss of the crop led to famine, death, and mass emigration. Resistance to disease is a chief benefit to maintaining crop biodiversity and lack of diversity in contemporary crop species carries similar risks. Seed companies, which are the source of most crop varieties in developed countries, must continually breed new varieties to keep up with evolving pest organisms. These same seed companies, however, have participated in the decline of the number of varieties available as they focus on selling fewer varieties in more areas of the world replacing traditional local varieties.

The ability to create new crop varieties relies on the diversity of varieties available and the availability of wild forms related to the crop plant. These wild forms are often the source of new gene variants that

271

can be bred with existing varieties to create varieties with new attributes. Loss of wild species related to a crop will mean the loss of potential in crop improvement. Maintaining the genetic diversity of wild species related to domesticated species ensures our continued supply of food.

Since the 1920s, government agriculture departments have maintained seed banks of crop varieties as a way to maintain crop diversity. This system has flaws because over time seed varieties are lost through accidents and there is no way to replace them. In 2008, the Svalbard Global seed Vault, located on Spitsbergen island, Norway, (Figure 6.52) began storing seeds from around the world as a backup system to the regional seed banks. If a regional seed bank stores varieties in Svalbard, losses can be replaced from Svalbard should something happen to the regional seeds. The Svalbard seed vault is deep into the rock of the arctic island. Conditions within the vault are maintained at ideal temperature and humidity for seed survival, but the deep underground location of the vault in the arctic means that failure of the vault's systems will not compromise the climatic conditions inside the vault.

art CONNECTION

Figure 6.52 The Svalbard Global Seed Vault is a storage facility for seeds of Earth's diverse crops. (credit: Mari Tefre, Svalbard Global Seed Vault)

Although crops are largely under our control, our ability to grow them is dependent on the biodiversity of the ecosystems in which they are grown. That biodiversity creates the conditions under which crops are able to grow through what are known as ecosystem services—valuable conditions or processes that are carried out by an ecosystem. Crops are not grown, for the most part, in built environments. They are grown in soil. Although some agricultural soils are rendered sterile using controversial pesticide treatments, most contain a huge diversity of organisms that maintain nutrient cycles—breaking down organic matter into nutrient compounds that crops need for growth. These organisms also maintain soil texture that affects water and oxygen dynamics in the soil that are necessary for plant growth. Replacing the work of these organisms in forming arable soil is not practically possible. These kinds of processes are called ecosystem services. They occur within ecosystems, such as soil ecosystems, as a result of the diverse metabolic activities of the organisms living there, but they provide benefits to human food production, drinking water availability, and breathable air.

Other key ecosystem services related to food production are plant pollination and crop pest control. It is estimated that honeybee pollination within the United States brings in $1.6 billion per year; other pollinators contribute up to $6.7 billion. Over 150 crops in the United States require pollination to produce. Many honeybee populations are managed by beekeepers who rent out their hives' services to farmers. Honeybee populations in North America have been suffering large losses caused by a syndrome known as colony collapse disorder, a new phenomenon with an unclear cause. Other pollinators include a diverse array of other bee species and various insects and birds. Loss of these species would make growing crops requiring pollination impossible, increasing dependence on other crops.

Finally, humans compete for their food with crop pests, most of which are insects. Pesticides control these competitors, but these are costly and lose their effectiveness over time as pest populations adapt.

272

They also lead to collateral damage by killing non-pest species as well as beneficial insects like honeybees, and risking the health of agricultural workers and consumers. Moreover, these pesticides may migrate from the fields where they are applied and do damage to other ecosystems like streams, lakes, and even the ocean. Ecologists believe that the bulk of the work in removing pests is actually done by predators and parasites of those pests, but the impact has not been well studied. A review found that in 74 percent of studies that looked for an effect of landscape complexity (forests and fallow fields near to crop fields) on natural enemies of pests, the greater the complexity, the greater the effect of pest-suppressing organisms. Another experimental study found that introducing multiple enemies of pea aphids (an important alfalfa pest) increased the yield of alfalfa significantly. This study shows that a diversity of pests is more effective at control than one single pest. Loss of diversity in pest enemies will inevitably make it more difficult and costly to grow food. The world's growing human population faces significant challenges in the increasing costs and other difficulties associated with producing food.

Wild Food Sources

In addition to growing crops and raising food animals, humans obtain food resources from wild populations, primarily wild fish populations. For about one billion people, aquatic resources provide the main source of animal protein. But since 1990, production from global fisheries has declined. Despite considerable effort, few fisheries on Earth are managed sustainability.

Fishery extinctions rarely lead to complete extinction of the harvested species, but rather to a radical restructuring of the marine ecosystem in which a dominant species is so over-harvested that it becomes a minor player, ecologically. In addition to humans losing the food source, these alterations affect many other species in ways that are difficult or impossible to predict. The collapse of fisheries has dramatic and long-lasting effects on local human populations that work in the fishery. In addition, the loss of an inexpensive protein source to populations that cannot afford to replace it will increase the cost of living and limit societies in other ways. In general, the fish taken from fisheries have shifted to smaller species and the larger species are overfished. The ultimate outcome could clearly be the loss of aquatic systems as food sources.

CONCEPT in ACTION

openstax COLLEGE

Visit this **website (http://openstaxcollege.org/l/decliningfish2)** to view a brief video discussing a study of declining fisheries.

| Threats to Biodiversity

By the end of this section, you will be able to:

- Identify significant threats to biodiversity
- Explain the effects of habitat loss, exotic species, and hunting on biodiversity
- Identify the early and predicted effects of climate change on biodiversity

The core threat to biodiversity on the planet, and therefore a threat to human welfare, is the combination of human population growth and the resources used by that population. The human population requires resources to survive and grow, and those resources are being removed unsustainably from the environment. The three greatest proximate threats to biodiversity are habitat loss, overharvesting, and introduction of exotic species. The first two of these are a direct result of human population growth and resource use. The third results from increased mobility and trade. A fourth major cause of extinction, anthropogenic (human-caused) climate change, has not yet had a large impact, but it is predicted to become significant during this century. Global climate change is also a consequence of human population needs for energy and the use of fossil fuels to meet those needs (Figure 6.53). Environmental issues,

273

such as toxic pollution, have specific targeted effects on species, but are not generally seen as threats at the magnitude of the others.
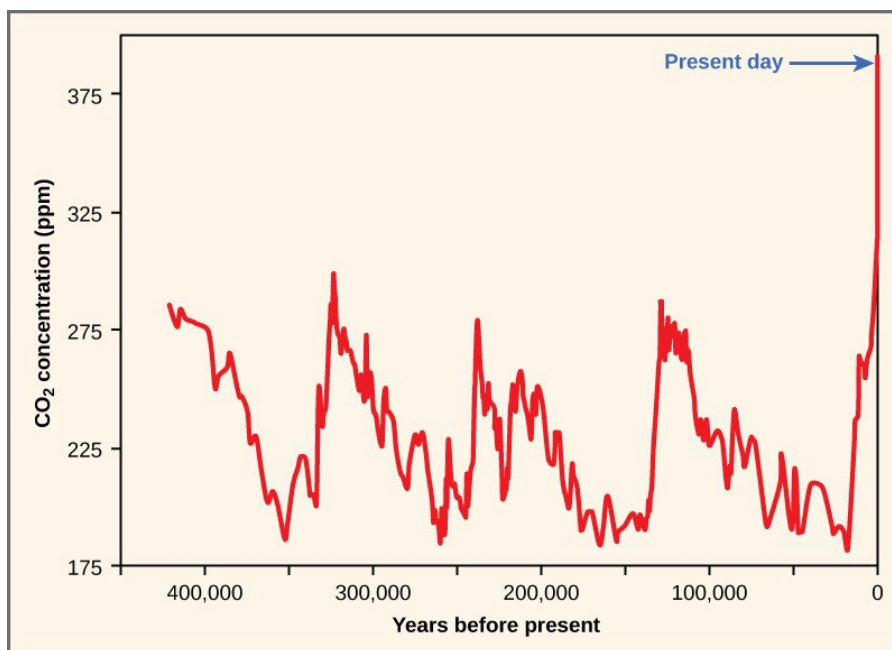
Figure 6.53 Atmospheric carbon dioxide levels fluctuate in a cyclical manner. However, the burning of fossil fuels in recent history has caused a dramatic increase in the levels of carbon dioxide in the Earth's atmosphere, which have now reached levels never before seen on Earth. Scientists predict that the addition of this "greenhouse gas" to the atmosphere is resulting in climate change that will significantly impact biodiversity in the coming century.

Habitat Loss

Humans rely on technology to modify their environment and replace certain functions that were once performed by the natural ecosystem. Other species cannot do this. Elimination of their habitat—whether it is a forest, coral reef, grassland, or flowing river—will kill the individuals in the species. Remove the entire habitat within the range of a species and, unless they are one of the few species that do well in human-built environments, the species will become extinct. Human destruction of habitats (habitats generally refer to the part of the ecosystem required by a particular species) accelerated in the latter half of the twentieth century. Consider the exceptional biodiversity of Sumatra: it is home to one species of orangutan, a species of critically endangered elephant, and the Sumatran tiger, but half of Sumatra's forest is now gone. The neighboring island of Borneo, home to the other species of orangutan, has lost a similar area of forest. Forest loss continues in protected areas of Borneo. The orangutan in Borneo is listed as endangered by the International Union for Conservation of Nature (IUCN), but it is simply the most visible of thousands of species that will not survive the disappearance of the forests of Borneo. The forests are removed for timber and to plant palm oil plantations (Figure 6.54). Palm oil is used in many products including food products, cosmetics, and biodiesel in Europe. A 5-year estimate of global forest cover loss for the years from 2000 to 2005 was 3.1 percent. Much loss (2.4 percent) occurred in the humid tropics where forest loss is primarily from timber extraction. These losses certainly also represent the extinction of species unique to those areas.

Figure 6.54 An oil palm plantation in Sabah province Borneo, Malaysia, replaces native forest habitat that a variety of species depended on to live. (credit: Lian Pin Koh)

biology IN ACTION

Preventing Habitat Destruction with Wise Wood Choices

Most consumers do not imagine that the home improvement products they buy might be contributing to habitat loss and species extinctions. Yet the market for illegally harvested tropical timber is huge, and the wood products often find themselves in building supply stores in the United States. One estimate is that 10 percent of the imported timber stream in the United States, which is the world's largest consumer of wood products, is potentially illegally logged. In 2006, this amounted to $3.6 billion in wood products. Most of the illegal products are imported from countries that act as intermediaries and are not the originators of the wood.

How is it possible to determine if a wood product, such as flooring, was harvested sustainably or even legally? The Forest Stewardship Council (FSC) certifies sustainably harvested forest products; therefore, looking for their certification on flooring and other hardwood products is one way to ensure that the wood has not been taken illegally from a tropical forest. Certification applies to specific products, not to a producer; some producers' products may not have certification while other products are certified. There are certifications other than the FSC, but these are run by timber companies creating a conflict of interest. Another approach is to buy domestic wood species. While it would be great if there was a list of legal versus illegal woods, it is not that simple. Logging and forest management laws vary from country to country; what is illegal in one country may be legal in another. Where and how a product is harvested and whether the forest from which it comes is being sustainably maintained all factor into whether a wood product will be certified by the FSC. It is always a good idea to ask questions about where a wood product came from and how the supplier knows that it was harvested legally.

Habitat destruction can affect ecosystems other than forests. Rivers and streams are important ecosystems and are frequently the target of habitat modification through building and from damming or water removal. Damming of rivers affects flows and access to all parts of a river. Altering a flow regime can reduce or eliminate populations that are adapted to seasonal changes in flow. For example, an estimated 91 percent of river lengths in the United States have been modified with damming or bank modifications. Many fish species in the United States, especially rare species or species with restricted distributions, have seen declines caused by river damming and habitat loss. Research has confirmed that species of amphibians that must carry out parts of their life cycles in both aquatic and terrestrial habitats are at greater risk of population declines and extinction because of the increased likelihood that one of their habitats or access between them will be lost. This is of particular concern because amphibians have been declining in numbers and going extinct more rapidly than many other groups for a variety of possible reasons.

275

Overharvesting

Overharvesting is a serious threat to many species, but particularly to aquatic species. There are many examples of regulated fisheries (including hunting of marine mammals and harvesting of crustaceans and other species) monitored by fisheries scientists that have nevertheless collapsed. The western Atlantic cod fishery is the most spectacular recent collapse. While it was a hugely productive fishery for 400 years, the introduction of modern factory trawlers in the 1980s and the pressure on the fishery led to it becoming unsustainable. The causes of fishery collapse are both economic and political in nature. Most fisheries are managed as a common resource, available to anyone willing to fish, even when the fishing territory lies within a country's territorial waters. Common resources are subject to an economic pressure known as the **tragedy of the commons**, in which fishers have little motivation to exercise restraint in harvesting a fishery when they do not own the fishery. The general outcome of harvests of resources held in common is their overexploitation. While large fisheries are regulated to attempt to avoid this pressure, it still exists in the background. This overexploitation is exacerbated when access to the fishery is open and unregulated and when technology gives fishers the ability to overfish. In a few fisheries, the biological growth of the resource is less than the potential growth of the profits made from fishing if that time and money were invested elsewhere. In these cases—whales are an example—economic forces will drive toward fishing the population to extinction.

Explore a U.S. Fish & Wildlife Service **interactive map (http://openstaxcollege.org/l/habitat_map2)** of critical habitat for endangered and threatened species in the United States. To begin, select "Visit the online mapper."

For the most part, fishery extinction is not equivalent to biological extinction—the last fish of a species is rarely fished out of the ocean. But there are some instances in which true extinction is a possibility. Whales have slow-growing populations and are at risk of complete extinction through hunting. Also, there are some species of sharks with restricted distributions that are at risk of extinction. The groupers are another population of generally slow-growing fishes that, in the Caribbean, includes a number of species that are at risk of extinction from overfishing.

Coral reefs are extremely diverse marine ecosystems that face peril from several processes. Reefs are home to 1/3 of the world's marine fish species—about 4000 species—despite making up only one percent of marine habitat. Most home marine aquaria house coral reef species that are wild-caught organisms—not cultured organisms. Although no marine species is known to have been driven extinct by the pet trade, there are studies showing that populations of some species have declined in response to harvesting, indicating that the harvest is not sustainable at those levels. There are also concerns about the effect of the pet trade on some terrestrial species such as turtles, amphibians, birds, plants, and even the orangutans.

View a **brief video (http://openstaxcollege.org/l/ocean_matters2)** discussing the role of marine ecosystems in supporting human welfare and the decline of ocean ecosystems.

Bush meat is the generic term used for wild animals killed for food. Hunting is practiced throughout the world, but hunting practices, particularly in equatorial Africa and parts of Asia, are believed to threaten several species with extinction. Traditionally, bush meat in Africa was hunted to feed families directly; however, recent commercialization of the practice now has bush meat available in grocery stores, which has increased harvest rates to the level of unsustainability. Additionally, human population growth has increased the need for protein foods that are not being met from agriculture. Species threatened by the bush meat trade are mostly mammals including many monkeys and the great apes living in the Congo basin.

Exotic Species

Exotic species are species that have been intentionally or unintentionally introduced by humans into an ecosystem in which they did not evolve. Human transportation of people and goods, including the intentional transport of organisms for trade, has dramatically increased the introduction of species into new ecosystems. These new introductions are sometimes at distances that are well beyond the capacity of the species to ever travel itself and outside the range of the species' natural predators.

Most exotic species introductions probably fail because of the low number of individuals introduced or poor adaptation to the ecosystem they enter. Some species, however, have characteristics that can make them especially successful in a new ecosystem. These exotic species often undergo dramatic population increases in their new habitat and reset the ecological conditions in the new environment, threatening the species that exist there. When this happens, the exotic species also becomes an invasive species. Invasive species can threaten other species through competition for resources, predation, or disease.

Explore this **interactive global database (http://openstaxcollege.org/l/exotic_invasiv2)** of exotic or invasive species.

Lakes and islands are particularly vulnerable to extinction threats from introduced species. In Lake Victoria, the intentional introduction of the Nile perch was largely responsible for the extinction of about 200 species of cichlids. The accidental introduction of the brown tree snake via aircraft (**Figure** 6.55) from the Solomon Islands to Guam in 1950 has led to the extinction of three species of birds and three to five species of reptiles endemic to the island. Several other species are still threatened. The brown tree snake is adept at exploiting human transportation as a means to migrate; one was even found on an aircraft arriving in Corpus Christi, Texas. Constant vigilance on the part of airport, military, and commercial aircraft personnel is required to prevent the snake from moving from Guam to other islands in the Pacific, especially Hawaii. Islands do not make up a large area of land on the globe, but they do contain a disproportionate number of endemic species because of their isolation from mainland ancestors.

Figure 6.55 The brown tree snake, *Boiga irregularis*, is an exotic species that has caused numerous extinctions on the island of Guam since its accidental introduction in 1950. (credit: NPS)

Many introductions of aquatic species, both marine and freshwater, have occurred when ships have dumped ballast water taken on at a port of origin into waters at a destination port. Water from the port of origin is pumped into tanks on a ship empty of cargo to increase stability. The water is drawn from the ocean or estuary of the port and typically contains living organisms such as plant parts, microorganisms, eggs, larvae, or aquatic animals. The water is then pumped out before the ship takes on cargo at the destination port, which may be on a different continent. The zebra mussel was introduced to the Great Lakes from Europe prior to 1988 in ship ballast. The zebra mussels in the Great Lakes have cost the industry millions of dollars in clean up costs to maintain water intakes and other facilities. The mussels have also altered the ecology of the lakes dramatically. They threaten native mollusk populations, but have also benefited some species, such as smallmouth bass. The mussels are filter feeders and have dramatically improved water clarity, which in turn has allowed aquatic plants to grow along shorelines, providing shelter for young fish where it did not exist before. The European green crab, *Carcinus maenas*, was introduced to San Francisco Bay in the late 1990s, likely in ship ballast water, and has spread north along the coast to Washington. The crabs have been found to dramatically reduce the abundance of native clams and crabs with resulting increases in the prey of native crabs.

Invading exotic species can also be disease organisms. It now appears that the global decline in amphibian species recognized in the 1990s is, in some part, caused by the fungus *Batrachochytrium dendrobatidis*, which causes the disease **chytridiomycosis** (Figure 6.56). There is evidence that the fungus is native to Africa and may have been spread throughout the world by transport of a commonly used laboratory and pet species: the African clawed frog, *Xenopus laevis*. It may well be that biologists themselves are responsible for spreading this disease worldwide. The North American bullfrog, *Rana catesbeiana*, which has also been widely introduced as a food animal but which easily escapes captivity, survives most infections of *B. dendrobatidis* and can act as a reservoir for the disease.

278

Figure 6.56 This Limosa harlequin frog (*Atelopus limosus*), an endangered species from Panama, died from a fungal disease called chytridiomycosis. The red lesions are symptomatic of the disease. (credit: Brian Gratwicke)

Early evidence suggests that another fungal pathogen, *Geomyces destructans*, introduced from Europe is responsible for **white-nose syndrome**, which infects cave-hibernating bats in eastern North America and has spread from a point of origin in western New York State (Figure 5.57). The disease has decimated bat populations and threatens extinction of species already listed as endangered: the Indiana bat, *Myotis sodalis*, and potentially the Virginia big-eared bat, *Corynorhinus townsendii virginianus*. How the fungus was introduced is unknown, but one logical presumption would be that recreational cavers unintentionally brought the fungus on clothes or equipment from Europe.

Figure 6.57 This little brown bat in Greeley Mine, Vermont, March 26, 2009, was found to have white-nose syndrome. (credit: modification of work by Marvin Moriarty, USFWS)

Climate Change

Climate change, and specifically the anthropogenic warming trend presently underway, is recognized as a major extinction threat, particularly when combined with other threats such as habitat loss. Anthropogenic warming of the planet has been observed and is hypothesized to continue due to past and continuing emission of greenhouse gases, primarily carbon dioxide and methane, into the atmosphere caused by the burning of fossil fuels and deforestation. These gases decrease the degree to which Earth is able to radiate heat energy created by the sunlight that enters the atmosphere. The changes in climate and energy balance caused by increasing greenhouse gases are complex and our understanding of them depends on predictions generated from detailed computer models. Scientists generally agree the present

279

warming trend is caused by humans and some of the likely effects include dramatic and dangerous climate changes in the coming decades. However, there is still debate and a lack of understanding about specific outcomes. Scientists disagree about the likely magnitude of the effects on extinction rates, with estimates ranging from 15 to 40 percent of species committed to extinction by 2050. Scientists do agree that climate change will alter regional climates, including rainfall and snowfall patterns, making habitats less hospitable to the species living in them. The warming trend will shift colder climates toward the north and south poles, forcing species to move with their adapted climate norms, but also to face habitat gaps along the way. The shifting ranges will impose new competitive regimes on species as they find themselves in contact with other species not present in their historic range. One such unexpected species contact is between polar bears and grizzly bears. Previously, these two species had separate ranges. Now, their ranges are overlapping and there are documented cases of these two species mating and producing viable offspring. Changing climates also throw off the delicate timing adaptations that species have to seasonal food resources and breeding times. Scientists have already documented many contemporary mismatches to shifts in resource availability and timing.

Range shifts are already being observed: for example, on average, European bird species ranges have moved 91 km (56.5 mi) northward. The same study suggested that the optimal shift based on warming trends was double that distance, suggesting that the populations are not moving quickly enough. Range shifts have also been observed in plants, butterflies, other insects, freshwater fishes, reptiles, amphibians, and mammals.

Climate gradients will also move up mountains, eventually crowding species higher in altitude and eliminating the habitat for those species adapted to the highest elevations. Some climates will completely disappear. The rate of warming appears to be accelerated in the arctic, which is recognized as a serious threat to polar bear populations that require sea ice to hunt seals during the winter months: seals are the only source of protein available to polar bears. A trend to decreasing sea ice coverage has occurred since observations began in the mid-twentieth century. The rate of decline observed in recent years is far greater than previously predicted by climate models (Figure 6.58).

Figure 6.58 The effect of global warming can be seen in the continuing retreat of Grinnell Glacier. The mean annual temperature in Glacier National Park has increased 1.33°C since 1900. The loss of a glacier results in the loss of summer meltwaters, sharply reducing seasonal water supplies and severely affecting local ecosystems. (credit: USGS, GNP Archives)

Finally, global warming will raise ocean levels due to meltwater from glaciers and the greater volume occupied by warmer water. Shorelines will be inundated, reducing island size, which will have an effect on some species, and a number of islands will disappear entirely. Additionally, the gradual melting and subsequent refreezing of the poles, glaciers, and higher elevation mountains—a cycle that has provided freshwater to environments for centuries—will be altered. This could result in an overabundance of salt water and a shortage of fresh water.

Conservation of Biodiversity

The threats to biodiversity at the genetic, species, and ecosystem levels have been recognized for some time. In the United States, the first national park with land set aside to remain in a wilderness state was Yellowstone Park in 1890. However, attempts to preserve nature for various reasons have occurred

280

for centuries. Today, the main efforts to preserve biodiversity involve legislative approaches to regulate human and corporate behavior, setting aside protected areas, and habitat restoration.

Changing Human Behavior

Legislation has been enacted to protect species throughout the world. The legislation includes international treaties as well as national and state laws. The Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) treaty came into force in 1975. The treaty, and the national legislation that supports it, provides a legal framework for preventing "listed" species from being transported across nations' borders, thus protecting them from being caught or killed in the first place when the purpose involves international trade. The listed species that are protected to one degree or another by the treaty number some 33,000. The treaty is limited in its reach because it only deals with international movement of organisms or their parts. It is also limited by various countries' ability or willingness to enforce the treaty and supporting legislation. The illegal trade in organisms and their parts is probably a market in the hundreds of millions of dollars.

Within many countries there are laws that protect endangered species and that regulate hunting and fishing. In the United States, the Endangered Species Act was enacted in 1973. When an at-risk species is listed by the Act, the U.S. Fish & Wildlife Service is required by law to develop a management plan to protect the species and bring it back to sustainable numbers. The Act, and others like it in other countries, is a useful tool, but it suffers because it is often difficult to get a species listed, or to get an effective management plan in place once a species is listed. Additionally, species may be controversially taken off the list without necessarily having had a change in their situation. More fundamentally, the approach to protecting individual species rather than entire ecosystems (although the management plans commonly involve protection of the individual species' habitat) is both inefficient and focuses efforts on a few highly visible and often charismatic species, perhaps at the expense of other species that go unprotected.

The Migratory Bird Treaty Act (MBTA) is an agreement between the United States and Canada that was signed into law in 1918 in response to declines in North American bird species caused by hunting. The Act now lists over 800 protected species. It makes it illegal to disturb or kill the protected species or distribute their parts (much of the hunting of birds in the past was for their feathers). Examples of protected species include northern cardinals, the red-tailed hawk, and the American black vulture.

Global warming is expected to be a major driver of biodiversity loss. Many governments are concerned about the effects of anthropogenic global warming, primarily on their economies and food resources. Since greenhouse gas emissions do not respect national boundaries, the effort to curb them is an international one. The international response to global warming has been mixed. The Kyoto Protocol, an international agreement that came out of the United Nations Framework Convention on Climate Change that committed countries to reducing greenhouse gas emissions by 2012, was ratified by some countries, but spurned by others. Two countries that were especially important in terms of their potential impact that did not ratify the Kyoto protocol were the United States and China. Some goals for reduction in greenhouse gasses were met and exceeded by individual countries, but, worldwide, the effort to limit greenhouse gas production is not succeeding. The intended replacement for the Kyoto Protocol has not materialized because governments cannot agree on timelines and benchmarks. Meanwhile, the resulting costs to human societies and biodiversity predicted by a majority of climate scientists will be high.

As already mentioned, the non-profit, non-governmental sector plays a large role in conservation effort both in North America and around the world. The approaches range from species-specific organizations to the broadly focused IUCN and Trade Records Analysis of Flora and Fauna in Commerce (TRAFFIC). The Nature Conservancy takes a novel approach. It purchases land and protects it in an attempt to set up preserves for ecosystems. Ultimately, human behavior will change when human values change. At present, the growing urbanization of the human population is a force that mitigates against valuing biodiversity, because many people no longer come in contact with natural environments and the species that inhabit them.

Conservation in Preserves

Establishment of wildlife and ecosystem preserves is one of the key tools in conservation efforts (**Figure 6.59**). A preserve is an area of land set aside with varying degrees of protection for the organisms that exist within the boundaries of the preserve. Preserves can be effective for protecting both species and ecosystems, but they have some serious drawbacks.

Figure 6.59 National parks, such as Grand Teton National Park in Wyoming, help conserve biodiversity. (credit: Don DeBold)

A simple measure of success in setting aside preserves for biodiversity protection is to set a target percentage of land or marine habitat to protect. However, a more detailed preserve design and choice of location is usually necessary because of the way protected lands are allocated and how biodiversity is distributed: protected lands tend to contain less economically valuable resources rather than being set aside specifically for the species or ecosystems at risk. In 2003, the IUCN World Parks Congress estimated that 11.5 percent of Earth's land surface was covered by preserves of various kinds. This area is greater than previous goals; however, it only represents 9 out of 14 recognized major biomes and research has shown that 12 percent of all species live outside preserves; these percentages are much higher when threatened species are considered and when only high quality preserves are considered. For example, high quality preserves include only about 50 percent of threatened amphibian species. The conclusion must be that either the percentage of area protected must be increased, the percentage of high quality preserves must be increased, or preserves must be targeted with greater attention to biodiversity protection. Researchers argue that more attention to the latter solution is required.

A **biodiversity hotspot** is a conservation concept developed by Norman Myers in 1988. Hotspots are geographical areas that contain high numbers of endemic species. The purpose of the concept was to identify important locations on the planet for conservation efforts, a kind of conservation triage. By protecting hotspots, governments are able to protect a larger number of species. The original criteria for a hotspot included the presence of 1500 or more species of endemic plants and 70 percent of the area disturbed by human activity. There are now 34 biodiversity hotspots (Figure 6.60) that contain large numbers of endemic species, which include half of Earth's endemic plants.
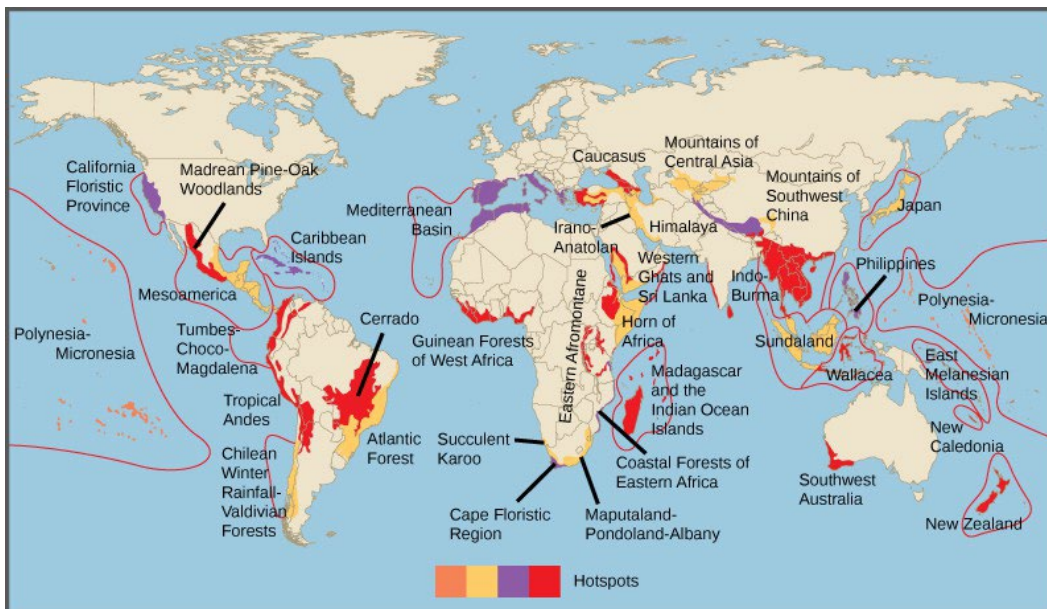
Figure 6.60 Conservation International has identified 34 biodiversity hotspots. Although these cover only 2.3 percent of the Earth's surface, 42 percent of the terrestrial vertebrate species and 50 percent of the world's plants are endemic to those hotspots.

There has been extensive research into optimal preserve designs for maintaining biodiversity. The fundamental principles behind much of the research have come from the seminal theoretical work of

282

Robert H. MacArthur and Edward O. Wilson published in 1967 on island biogeography.[2] This work sought to understand the factors affecting biodiversity on islands. Conservation preserves can be seen as "islands" of habitat within "an ocean" of non-habitat. In general, large preserves are better because they support more species, including species with large home ranges; they have more core area of optimal habitat for individual species; they have more niches to support more species; and they attract more species because they can be found and reached more easily.

Preserves perform better when there are partially protected buffer zones around them of suboptimal habitat. The buffer allows organisms to exit the boundaries of the preserve without immediate negative consequences from hunting or lack of resources. One large preserve is better than the same area of several smaller preserves because there is more core habitat unaffected by less hospitable ecosystems outside the preserve boundary. For this same reason, preserves in the shape of a square or circle will be better than a preserve with many thin "arms." If preserves must be smaller, then providing wildlife corridors between them so that species and their genes can move between the preserves; for example, preserves along rivers and streams will make the smaller preserves behave more like a large one. All of these factors are taken into consideration when planning the nature of a preserve before the land is set aside.

In addition to the physical specifications of a preserve, there are a variety of regulations related to the use of a preserve. These can include anything from timber extraction, mineral extraction, regulated hunting, human habitation, and nondestructive human recreation. Many of the decisions to include these other uses are made based on political pressures rather than conservation considerations. On the other hand, in some cases, wildlife protection policies have been so strict that subsistence-living indigenous populations have been forced from ancestral lands that fell within a preserve. In other cases, even if a preserve is designed to protect wildlife, if the protections are not or cannot be enforced, the preserve status will have little meaning in the face of illegal poaching and timber extraction. This is a widespread problem with preserves in the tropics.

Some of the limitations on preserves as conservation tools are evident from the discussion of preserve design. Political and economic pressures typically make preserves smaller, never larger, so setting aside areas that are large enough is difficult. Enforcement of protections is also a significant issue in countries without the resources or political will to prevent poaching and illegal resource extraction.

Climate change will create inevitable problems with the location of preserves as the species within them migrate to higher latitudes as the habitat of the preserve becomes less favorable. Planning for the effects of global warming on future preserves, or adding new preserves to accommodate the changes expected from global warming is in progress, but will only be as effective as the accuracy of the predictions of the effects of global warming on future habitats.

Finally, an argument can be made that conservation preserves reinforce the cultural perception that humans are separate from nature, can exist outside of it, and can only operate in ways that do damage to biodiversity. Creating preserves reduces the pressure on human activities outside the preserves to be sustainable and non-damaging to biodiversity. Ultimately, the political, economic, and human demographic pressures will degrade and reduce the size of conservation preserves if the activities outside them are not altered to be less damaging to biodiversity.

Check out this **interactive global data system (http://openstaxcollege.org/l/protected_area2)** of protected areas. Review data about specific protected areas by location or study statistics on protected areas by country or region.

Habitat Restoration

Habitat restoration holds considerable promise as a mechanism for maintaining or restoring biodiversity. Of course once a species has become extinct, its restoration is impossible. However, restoration can

2. Robert H. MacArthur and Edward O. Wilson, E. O., *The Theory of Island Biogeography* (Princeton, N.J.: Princeton University Press, 1967).

improve the biodiversity of degraded ecosystems. Reintroducing wolves, a top predator, to Yellowstone National Park in 1995 led to dramatic changes in the ecosystem that increased biodiversity. The wolves (Figure 6.61) function to suppress elk and coyote populations and provide more abundant resources to the guild of carrion eaters. Reducing elk populations has allowed revegetation of riparian (the areas along the banks of a stream or river) areas, which has increased the diversity of species in that habitat. Suppression of coyotes has increased the species previously suppressed by this predator. The number of species of carrion eaters has increased because of the predatory activities of the wolves. In this habitat, the wolf is a keystone species, meaning a species that is instrumental in maintaining diversity within an ecosystem. Removing a keystone species from an ecological community causes a collapse in diversity. The results from the Yellowstone experiment suggest that restoring a keystone species effectively can have the effect of restoring biodiversity in the community. Ecologists have argued for the identification of keystone species where possible and for focusing protection efforts on these species. It makes sense to return the keystone species to the ecosystems where they have been removed.

Figure 6.61 This photograph shows the Gibbon wolf pack in Yellowstone National Park, March 1, 2007. Wolves have been identified as a keystone species. (credit: Doug Smith, NPS)

Other large-scale restoration experiments underway involve dam removal. In the United States, since the mid-1980s, many aging dams are being considered for removal rather than replacement because of shifting beliefs about the ecological value of free-flowing rivers. The measured benefits of dam removal include restoration of naturally fluctuating water levels (often the purpose of dams is to reduce variation in river flows), which leads to increased fish diversity and improved water quality. In the Pacific Northwest, dam removal projects are expected to increase populations of salmon, which is considered a keystone species because it transports nutrients to inland ecosystems during its annual spawning migrations. In other regions, such as the Atlantic coast, dam removal has allowed the return of other spawning anadromous fish species (species that are born in fresh water, live most of their lives in salt water, and return to fresh water to spawn). Some of the largest dam removal projects have yet to occur or have happened too recently for the consequences to be measured. The large-scale ecological experiments that these removal projects constitute will provide valuable data for other dam projects slated either for removal or construction.

The Role of Zoos and Captive Breeding

Zoos have sought to play a role in conservation efforts both through captive breeding programs and education (Figure 6.62). The transformation of the missions of zoos from collection and exhibition facilities to organizations that are dedicated to conservation is ongoing. In general, it has been recognized that, except in some specific targeted cases, captive breeding programs for endangered species are inefficient and often prone to failure when the species are reintroduced to the wild. Zoo facilities are far too limited to contemplate captive breeding programs for the numbers of species that are now at risk. Education, on the other hand, is a potential positive impact of zoos on conservation efforts, particularly given the global trend to urbanization and the consequent reduction in contacts between people and wildlife. A number of studies have been performed to look at the effectiveness of zoos on people's attitudes and actions regarding conservation; at present, the results tend to be mixed.

Figure 6.62 Zoos and captive breeding programs help preserve many endangered species, such as this golden lion tamarin. (credit: Garrett Ziegler)

Summary

Population Demographics and Dynamics

Populations are individuals of a species that live in a particular habitat. Ecologists measure characteristics of populations: size, density, and distribution pattern. Life tables are useful to calculate life expectancies of individual population members. Survivorship curves show the number of individuals surviving at each age interval plotted versus time.

Population Growth and Regulation

Populations with unlimited resources grow exponentially—with an accelerating growth rate. When resources become limiting, populations follow a logistic growth curve in which population size will level off at the carrying capacity.

Populations are regulated by a variety of density-dependent and density-independent factors. Life- history characteristics, such as age at first reproduction or numbers of offspring, are characteristics that evolve in populations just as anatomy or behavior can evolve over time. The model of r- and K- selection suggests that characters, and possibly suites of characters, may evolve adaptations to population stability near the carrying capacity (K-selection) or rapid population growth and collapse (r- selection). Species will exhibit adaptations somewhere on a continuum between these two extremes.

The Human Population

Earth's human population is growing exponentially. Humans have increased their carrying capacity through technology, urbanization, and harnessing the energy of fossil fuels. The age structure of a population allows us to predict population growth. Unchecked human population growth could have dire long-term effects on human welfare and Earth's ecosystems.

Community Ecology

Communities include all the different species living in a given area. The variety of these species is referred to as biodiversity. Many organisms have developed defenses against predation and herbivory,

including mechanical defenses, warning coloration, and mimicry. Two species cannot exist indefinitely in the same habitat competing directly for the same resources. Species may form symbiotic relationships such as commensalism, mutualism, or parasitism. Community structure is described by its foundation and keystone species. Communities respond to environmental disturbances by succession: the predictable appearance of different types of plant species, until a stable community structure is established.

Energy Flow through Ecosystems
Ecosystems exist underground, on land, at sea, and in the air. Organisms in an ecosystem acquire energy in a variety of ways, which is transferred between trophic levels as the energy flows from the base to the top of the food web, with energy being lost at each transfer. There is energy lost at each trophic level, so the lengths of food chains are limited because there is a point where not enough energy remains to support a population of consumers. Fat soluble compounds biomagnify up a food chain causing damage to top consumers. even when environmental concentrations of a toxin are low.

Biogeochemical Cycles
Mineral nutrients are cycled through ecosystems and their environment. Of particular importance are water, carbon, nitrogen, phosphorus, and sulfur. All of these cycles have major impacts on ecosystem structure and function. As human activities have caused major disturbances to these cycles, their study and modeling is especially important. Ecosystems have been damaged by a variety of human activities that alter the natural biogeochemical cycles due to pollution, oil spills, and events causing global climate change. The health of the biosphere depends on understanding these cycles and how to protect the environment from irreversible damage.

Terrestrial Biomes
Earth has terrestrial and aquatic biomes. Aquatic biomes include both freshwater and marine environments. There are eight major terrestrial biomes: tropical rainforests, savannas, subtropical deserts, chaparral, temperate grasslands, temperate forests, boreal forests, and Arctic tundra. The same biome can occur in different geographic locations with similar climates. Temperature and precipitation, and variations in both, are key abiotic factors that shape the composition of animal and plant communities in terrestrial biomes. Some biomes, such as temperate grasslands and temperate forests, have distinct seasons with cold and hot weather alternating throughout the year. In warm, moist biomes, such as the tropical rainforest, net primary productivity is high as warm temperatures, abundant water, and a year-round growing season fuel plant growth. Other biomes, such as deserts and tundra, have low primary productivity due to extreme temperatures and a shortage of water.

Aquatic and Marine Biomes
Aquatic biomes include both saltwater and freshwater biomes. The abiotic factors important for the structuring of aquatic biomes can be different than those seen in terrestrial biomes. Sunlight is an important factor in bodies of water, especially those that are very deep, because of the role of photosynthesis in sustaining certain organisms. Other important factors include temperature, water movement, and salt content. Oceans may be thought of as consisting of different zones based on water depth, distance from the shoreline, and light penetrance. Different kinds of organisms are adapted to the

conditions found in each zone. Coral reefs are unique marine ecosystems that are home to a wide variety of species. Estuaries are found where rivers meet the ocean; their shallow waters provide nourishment and shelter for young crustaceans, mollusks, fishes, and many other species. Freshwater biomes include lakes, ponds, rivers, streams, and wetlands. Bogs are an interesting type of wetland characterized by standing water, a lower pH, and a lack of nitrogen.

Importance of Biodiversity

Biodiversity exists at multiple levels of organization, and is measured in different ways depending on the goals of those taking the measurements. These include numbers of species, genetic diversity, chemical diversity, and ecosystem diversity. The number of described species is estimated to be 1.5 million with about 17,000 new species being described each year. Estimates for the total number of eukaryotic species on Earth vary but are on the order of 10 million. Biodiversity is negatively correlated with latitude for most taxa, meaning that biodiversity is higher in the tropics. The mechanism for this pattern is not known with certainty, but several plausible hypotheses have been advanced.

Humans use many compounds that were first discovered or derived from living organisms as medicines: secondary plant compounds, animal toxins, and antibiotics produced by bacteria and fungi. More medicines are expected to be discovered in nature. Loss of biodiversity will impact the number of pharmaceuticals available to humans. Biodiversity may provide important psychological benefits to humans.

Crop diversity is a requirement for food security, and it is being lost. The loss of wild relatives to crops also threatens breeders' abilities to create new varieties. Ecosystems provide ecosystem services that support human agriculture: pollination, nutrient cycling, pest control, and soil development and maintenance. Loss of biodiversity threatens these ecosystem services and risks making food production more expensive or impossible. Wild food sources are mainly aquatic, but few are being managed for sustainability. Fisheries' ability to provide protein to human populations is threatened when extinction occurs.

Threats to Biodiversity

The core threats to biodiversity are human population growth and unsustainable resource use. To date, the most significant causes of extinction are habitat loss, introduction of exotic species, and overharvesting. Climate change is predicted to be a significant cause of extinction in the coming century. Habitat loss occurs through deforestation, damming of rivers, and other activities.

Overharvesting is a threat particularly to aquatic species, but the taking of bush meat in the humid tropics threatens many species in Asia, Africa, and the Americas. Exotic species have been the cause of a number of extinctions and are especially damaging to islands and lakes. Exotic species' introductions are increasing because of the increased mobility of human populations and growing global trade and transportation. Climate change is forcing range changes that may lead to extinction. It is also affecting adaptations to the timing of resource availability that negatively affects species in seasonal environments. The impacts of climate change are currently greatest in the arctic. Global warming will also raise sea levels, eliminating some islands and reducing the area of all others.

Preserving Biodiversity

Five mass extinctions with losses of more than 50 percent of extant species are observable in the fossil record. Recent extinctions are recorded in written history and are the basis for one method of estimating contemporary extinction rates. The other method uses measures of habitat loss and species-area relationships. Estimates of contemporary extinction rates vary but are as high as 500 times the background rate, as determined from the fossil record, and are predicted to rise.

There is a legislative framework for biodiversity protection. International treaties such as CITES regulate the transportation of endangered species across international borders. Legislation within individual countries protecting species and agreements on global warming have had limited success; there is at present no international agreement on targets for greenhouse gas emissions. In the United States, the Endangered Species Act protects listed species but is hampered by procedural difficulties and a focus on individual species. The Migratory Bird Act is an agreement between Canada and the United States to protect migratory birds. The non-profit sector is also very active in conservation efforts in a variety of ways.

Conservation preserves are a major tool in biodiversity protection. Presently, 11 percent of Earth's land surface is protected in some way. The science of island biogeography has informed the optimal design of preserves; however, preserves have limitations imposed by political and economic forces. In addition, climate change will limit the effectiveness of present preserves in the future. A downside of preserves is that they may lessen the pressure on human societies to function more sustainably outside the preserves. Habitat restoration has the potential to restore ecosystems to previous biodiversity levels before species become extinct. Examples of restoration include reintroduction of keystone species and removal of dams on rivers. Zoos have attempted to take a more active role in conservation and can have a limited role in captive breeding programs. Zoos also have a useful role in education.

1.   Converting a prairie to a farm field is an example of   .
a.   overharvesting
b.   habitat loss
c.   exotic species
d.   climate change

2.   Which two extinction risks may be a direct result of the pet trade?
a.   climate change and exotic species introduction
b.   habitat loss and overharvesting
c.   overharvesting and exotic species introduction
d.   habitat loss and climate change

3.   Certain species of parrot cannot be brought to the United States to be sold as pets. What is the name of the legislation that makes this illegal?

a.   Red List
b.   Migratory Bird Act
c.   CITES
d.   Endangered Species Act (ESA)

4.   Which of the following biomes is characterized by abundant water resources?
a.   deserts
b.   boreal forests
c.   savanna
d.   tropical wet forests

5.   Which of the following biomes is characterized by short growing seasons?
a.   deserts
b.   tropical wet forests
c.   Arctic tundra
d.   savanna

6.   Why is the tundra treeless?
a.   lack of sufficient water
b.   permanently frozen ground
c.   winters too harsh
d.   too many fires

7.   Where would you expect to find the most photosynthesis in an ocean biome?

a. aphotic zone
b. abyssal zone
c. benthic realm
d. intertidal zone
8. A key feature of estuaries is
a. low light conditions and high productivity
b. salt water and fresh water
c. frequent algal blooms
d. little or no vegetation

9. Species with limited resources usually exhibit a(n) _____ growth curve.
a. logistic
b. logical
c. experimental
d. exponential

10. The first species to live on new land, such as that formed from volcanic lava, are called .
a. climax community
b. keystone species
c. foundation species
d. pioneer species

Chapter 7
MINERALS

Chemistry Review: Atoms and Isotopes

A **chemical element** is a substance that cannot be made into a more straightforward form by ordinary chemical means the smallest unit of a chemical element is an atom. An **atom** has all the properties of that element. At the center of an atom is a nucleus made up of subatomic particles called protons and neutrons. **Protons** have a positive electrical charge. The number of protons in the nucleus determines what element the atom is. **Neutrons** are about the size of protons but have no charge. **Electrons**, each having a negative electrical charge, orbit the nucleus at varying energy levels in a region known as the **electron cloud**.
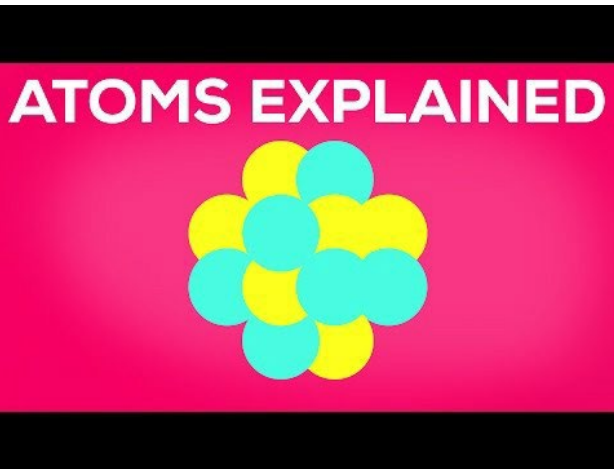
Video 7.1 Atoms explained
https://www.youtube.com/watch?v=_lNF3_30lUE

Because electrons are minuscule compared with protons and neutrons, the number of protons plus neutrons gives the atom its atomic mass. All atoms of a given element always have the same number of protons but may differ in the number of neutrons found in its nucleus. Atoms of an element with differing numbers of neutrons are called **isotopes**. (Matter Matters | Earth Science, n.d.)

Video 7.2 Isotopes
https://www.youtube.com/watch?v=92yZY7dd-Jc
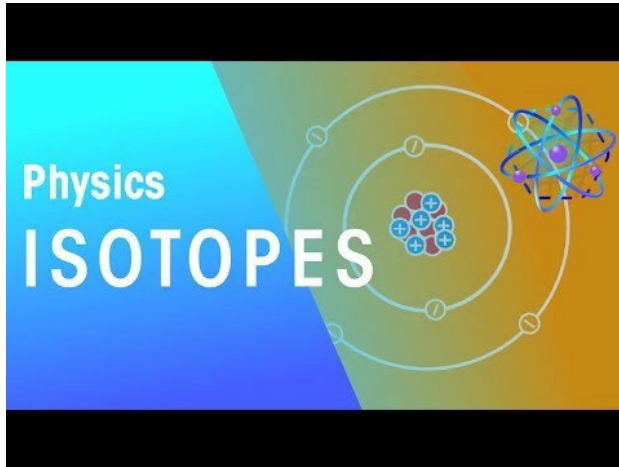
Ions and Molecules

Atoms are stable when they have a full outermost electron valence shell. For an atom to fill its outermost shell, it will give, take, or share electrons. When an atom either gains or loses electrons, this creates an ion. Ions have either a positive or a negative electrical charge. What is the charge of an ion if the atom loses an electron? An atom with the same number of protons and electrons has no overall charge, so if an atom loses the negatively charged electron, it has a positive charge. What is the charge of an ion if the atom gains an electron? If the atom gains an electron, it has a negative charge.

When atoms chemically bond, they form **compounds**. The smallest unit of a compound with all the properties of that compound is a **molecule**. When two or more atoms share electrons to form a chemical bond, they form a molecule. The molecular mass is the sum of the masses of all the atoms in the molecule.

Chemical Bonding

Ions come together to create a molecule so that electrical charges are balanced; the positive charges balance the negative charges, and the molecule has no electrical charge. An atom will balance its electrical charge by sharing an electron with another atom, giving it away, or receive an electron from another atom.

The joining of ions to make molecules is chemical bonding. There are three main types of chemical bonds. **Ionic bonds** occur when electrons are transferred between atoms. **Covalent bonds** occur when an atom shares one or more electrons with another atom. The sharing of electrons is not always evenly distributed within a molecule. If one atom has the electrons more often than another atom in the molecule, it has a positive and a negative side. It is a polar molecule because it acts a little bit as if it had poles, like a magnet. Finally, **hydrogen bonds** are weak intermolecular bonds that form when the positive side of one polar molecule is attracted to another polar molecule's negative side.

293
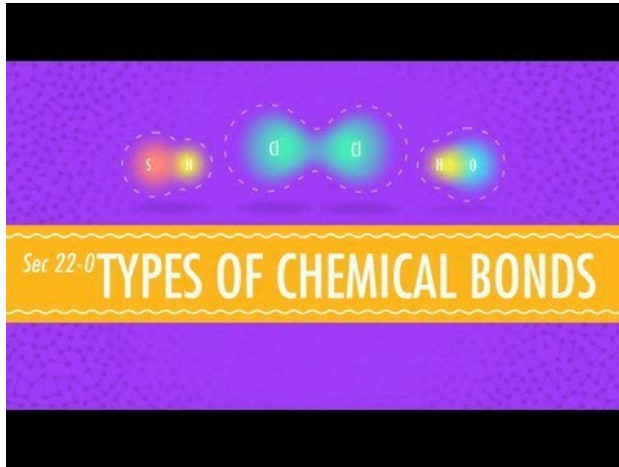
Video 7.3 Types of chemical bonds

https://www.youtube.com/watch?v=QXT4OVM4vXI

Minerals

Minerals are categorized based on their chemical composition. Owing to similarities in composition, minerals within the same group may have similar characteristics. Geologists have a precise definition of minerals. A material is characterized as a mineral if it meets all the following traits:

- Inorganic, crystalline solid.
- Formed through natural processes and has a definite chemical composition.

- Identified by their characteristic physical properties such as crystalline structure, hardness, density, flammability, and color.

Video 7.4 Identifying minerals

https://www.youtube.com/watch?v=KB-2pO7pSK8

Crystalline Solid

Minerals are crystalline solids. A **crystal** is a solid in which the atoms are arranged in a regular, repeating pattern. The pattern of atoms in different samples of the same mineral is the same. Is glass a mineral? Without a crystalline structure, even natural glass is not a mineral. (Minerals and Mineral Groups | Earth Science, n.d.)

295

Organic and Inorganic Substances

Organic substances are the carbon-based compounds made by living creatures and include proteins, carbohydrates, and oils. **Inorganic substances** have a structure that is not characteristic of living bodies. Coal is made of plant and animal remains. Is it a mineral? Coal is classified as a sedimentary rock but is not a mineral. (Minerals and Mineral Groups | Earth Science, n.d.)

Natural Processes

Minerals are made by natural processes, those that occur in or on Earth. A diamond created deep in Earth's crust is a mineral. Is a diamond created in a laboratory by placing carbon under high pressures a mineral? No. Do not buy a laboratory-made "diamond" for jewelry without realizing it is not technically a mineral.

Chemical Composition

Nearly all (98.5 percent) of Earth's crust is made up of only eight elements – oxygen, silicon, aluminum, iron, calcium, sodium, potassium, and magnesium – and these are the elements that make up most minerals.

All minerals have a specific chemical composition. The mineral silver is made up of only silver atoms and diamond is made only of carbon atoms, but most minerals are made up of chemical compounds. Each mineral has its chemical formula. Halite is $NaCl$ (sodium chloride). Quartz is always made of two oxygen atoms

bonded to a silicon atom, SiO_2. If a mineral contains any other elements in its crystal structure, it is not quartz.

A hard mineral with covalently bonded carbon is diamond, but a softer mineral containing calcium and oxygen along with carbon is calcite. Some minerals have a range of chemical composition. Olivine always has silicon and oxygen, as well as iron or magnesium or both. (Minerals and Mineral Groups | Earth Science, n.d.)

Mineral Identification

Their physical characteristics can identify minerals. The physical properties of minerals are related to their chemical composition and bonding. Some characteristics, such as a mineral's hardness, are more useful for mineral identification. Color is readily observable and undoubtedly visible, but it is usually less reliable than other physical properties. (Reading: Physical Characteristics of Minerals | Geology, n.d.)

Video 7.5 Identifying minerals
https://www.youtube.com/watch?v=NhkYFxeHghA

(This video provides examples of different minerals)

Color

Diamonds are popular gemstones because the way they reflect light makes them very sparkly. Turquoise is prized for its striking greenish-blue color. Notice that specific terms are being used to describe the appearance of minerals. Color is rarely useful for identifying a mineral.

Streak

Streak is the color of a mineral's powder. Streak is a more reliable

298

property than color because streak does not vary. Minerals that are the same color may have a different colored streak. Many minerals, such as the quartz, do not have a streak. To check streak, scrape the mineral across an unglazed porcelain plate. Yellow-gold pyrite has a blackish streak, another indicator that pyrite is not gold, with a golden yellow streak.

Luster

Luster describes the reflection of light off a mineral's surface. Mineralogists have specific terms to describe luster. One straightforward way to classify luster is based on whether the mineral is metallic or non-metallic. Minerals that are opaque and shiny, such as pyrite, have a metallic luster. Minerals such as quartz have a non-metallic luster.

Specific Gravity

Density describes how much matter is in a certain amount of space: density = mass/volume. Mass is a measure of the amount of matter in an object. Its volume describes the amount of space an object takes up. The density of an object depends on its mass and its volume. For example, the water in a drinking glass has the same density as the water in the same volume of a swimming pool. Gold has a density of about 19 g/cm3; pyrite has a density of about 5 g/cm3 – that is another way to tell pyrite from gold. Quartz is even less dense than pyrite and has a density of 2.7 g/cm3. The specific gravity of a substance compares its density to that of water. Denser substances have higher specific gravity.

Hardness

Hardness is a measure of whether a mineral will scratch or be scratched. **Mohs Hardness Scale** is a reference for mineral hardness. With a Mohs scale, anyone can evaluate an unknown mineral for its hardness. Imagine you have an unknown mineral. You find that it can scratch fluorite or even apatite, but feldspar scratches it. This examination of the mineral informs you that the mineral's hardness is between 5 and 6. Note that no other mineral can scratch diamond. Breaking a mineral breaks its chemical bonds. Since some bonds are weaker than other bonds, each type of mineral is likely to break where the bonds between the atoms are weakest. For that reason, minerals break apart in distinctive ways. (Mineral Identification | Earth Science, n.d.)

Cleavage

Cleavage is the tendency of a mineral to break along specific planes to make smooth surfaces. Halite breaks between layers of sodium and chlorine to form cubes with smooth surfaces. One reason gemstone is beautiful is that the cleavage planes make an attractive crystal shape with smooth faces. (Mineral Identification | Earth Science, n.d.)

Fracture

Fracture is a break in a mineral that is not along a cleavage plane. Fracture is not always the same in the same mineral because its structure does not determine fracture. Minerals may have natural

fractures. Metals usually fracture into jagged edges. If a mineral splinter like wood, it may be fibrous. Some minerals, such as quartz, form smooth curved surfaces when they fracture. (Reading: Physical Characteristics of Minerals | Geology, n.d.)

Video 7.6 Properties of common minerals

https://www.youtube.com/watch?v=KB-2pO7pSK8&t=202s

Other Identifying Characteristics

Mineral Formation

Minerals form under an enormous range of geologic conditions. There are more ways to form minerals than there are types of

minerals themselves. Minerals can form from volcanic gases, sediment formation, oxidation, crystallization from magma, or deposition from a saline fluid. (Reading: Physical Characteristics of Minerals | Geology, n.d.)

Formation from Hot Material

A rock is a collection of minerals. Imagine a rock that becomes so hot it melts. Many minerals start in liquids that are hot enough to melt rocks. Magma is melted rock inside Earth, a molten mixture of substances that can be hotter than $1,000^{o}$C. Magma cools slowly inside Earth, which gives mineral crystals time to grow large enough to be seen clearly. When magma erupts onto Earth's surface, it is called lava. Lava cools much more rapidly than magma when it is below the surface. In cooling lava, mineral crystals do not have time to form and are exceedingly small. The chemical composition will be the same as if the magma cooled slowly. Existing rocks may be heated enough to release the molecules from their structure and move around. The molecules may match up with different molecules to form new minerals as the rock cools, a process called metamorphism. (Matter Matters | Earth Science, n.d.)

Formation from Solutions

Water on Earth, such as the water in the oceans, contains chemical elements mixed into a solution. Various processes can cause these elements to combine to form solid mineral deposits.

Minerals from Salt Water

When water evaporates, it leaves behind a solid precipitate of minerals. Water can only hold a certain amount of dissolved minerals and salts. When the amount is too vast to stay dissolved in the water, the particles come together to form mineral solids, which sink. Halite readily precipitates out of the water, as does calcite. Some lakes, such as Mono Lake in California or The Great Salt Lake in Utah, contain many mineral precipitates.

Minerals from Hot Underground Water

Magma heats nearby underground water, which reacts with the rocks around it to pick up dissolved particles. As the water flows through open spaces in the rock and cools, it deposits solid minerals. The mineral deposits form when a mineral fills cracks in rocks are called veins when minerals are deposited in open spaces and large crystal forms.

Mining and Mineral Use

Some minerals are beneficial. An ore is a rock that contains minerals with useful elements. Aluminum in bauxite ore is extracted from the ground and refined to be used in aluminum foil and many other products. The cost of creating a product from a mineral depends on how abundant the mineral is and how much the extraction and refining processes cost. Environmental damage from these processes

is often not figured into a product's cost. It is crucial to use mineral resources wisely.

Finding and Mining Minerals

Geologic processes create and concentrate minerals that are valuable natural resources. Geologists study geological formations and test the physical and chemical properties of soil and rocks to locate possible ores and determine their size and concentration. A mineral deposit will only be mined if it is profitable. A concentration of minerals is only called an ore deposit if it is profitable to mine. There are many ways to mine ores. (Reading: Mining and Mineral Use | Geology, n.d.)

Surface mining allows the extraction of ores that are close to Earth's surface. Overlying rock is blasted, and the rock that contains the valuable minerals is placed in a truck and taken to a refinery. Surface mining includes open-pit mining and mountaintop removal. Other methods of surface mining include strip mining, placer mining, and dredging. Strip mining is like open-pit mining but with material removed along a strip.

Placers are valuable minerals found in stream gravels. California's nickname, the Golden State, can be traced to the discovery of placer deposits of gold in 1848. The gold weathered out of hard metamorphic rock in the western Sierra Nevada, which also contains deposits of copper, lead, zinc, silver, chromite, and other valuable minerals. The gold traveled down rivers and then settled in gravel deposits. Currently, California has active mines for gold and silver and for non-metal minerals such as sand and gravel, which are

Physical Geography and Natural Disasters by R. Adam Dastrup

used for construction. (Reading: Mining and Mineral Use | Geology, n.d.)

Figure 7.1 "Bingham Copper Mine" by David Guthrie is licensed under Creative Commons Attribution 2.0 Generic.

Underground mining is used to recover ores that are deeper into Earth's surface. Miners blast and tunnel into the rock to gain access to the ores. How underground mining is approached depends on the placement of the ore body, its depth, concentration of ore, and the surrounding rock's strength. Underground mining is costly and dangerous. Fresh air and lights must also be brought into the tunnels for the miners, and accidents are far too frequent.

The ore's journey to becoming a useable material is only just beginning when the ore leaves the mine. Rocks are crushed so that the valuable minerals can be separated from the waste rock. Then the minerals are separated from the ore. A few methods for extracting ore are:

Physical Geography and Natural Disasters by R. Adam Dastrup

- **Heap leaching**: the addition of chemicals, such as cyanide or acid, to remove ore.
- **Flotation**: the addition of a compound that attaches to the valuable mineral and floats.
- **Smelting**: roasting rock, causing it to segregate into layers so the mineral can be extracted.

To extract the metal from the ore, the rock is melted at a temperature higher than 900 degrees Celsius, which requires much energy. Extracting a metal from the rock is so energy-intensive that recycling just 40 aluminum cans will save the energy equivalent of one gallon of gasoline. (Mining and Mineral Use | Earth Science, n.d.)

Although mining provides people with many necessary resources, environmental costs can be high. Surface mining clears the landscape of trees and soil, and nearby streams and lakes are inundated with sediment. Pollutants from the mined rock, such as heavy metals, enter the sediment and water system. Acids flow from some mine sites, changing the composition of nearby waterways.

U.S. law has changed so that in recent decades a mine region must be restored to its natural state, a process called reclamation. This is not true of older mines. Pits may be refilled or reshaped, and vegetation planted. Pits may be allowed to fill with water and become lakes or maybe turned into landfills. Underground mines may be sealed off or left open as homes for bats.

Some minerals are valuable because they are beautiful. Jade has been used for thousands of years in China. Diamonds sparkle on many engagement rings. Minerals like jade, turquoise, diamonds,

306

and emeralds are gemstones. A gemstone, or gem, is a material that is cut and polished for jewelry. Gemstones are usually rare and do not break or scratch easily. Most are cut along cleavage faces and then polished so that light bounces back off the cleavage planes. Light does not pass through opaque gemstones, such as turquoise. Gemstones are not just used in jewelry. Diamonds are used to cut and polish other materials, such as glass and metals because they are so hard. The mineral corundum, of which ruby and sapphire are varieties, is used in sandpaper products.

Minerals are used in much less obvious places. The mineral gypsum is used for the sheetrock in homes. Window glass is made from sand, which is mostly quartz. Halite is mined for rock salt. Copper is used in electrical wiring, and bauxite is the source of the aluminum used in soda cans. (Reading: Mining and Mineral Use | Geology, n.d.)

Conflict Resouces

Blood Diamonds

Blood diamonds, also called **conflict diamonds**, are diamonds mined in conflict areas and sold to finance warlords or invading armies in a geographic region. The term "blood" diamonds are used to highlight the negative impacts and consequences of diamonds mined for these ends. Often, women and children are enslaved to mine blood diamonds, and other conflict resources, to fund these local and regional wars. The Kimberley Process was created to help combat and regulate the flow of diamonds mined around the world

to ensure that people were not purchasing blood diamonds. Blood diamonds have been mined from Angola, Ivory Coast, Sierra Leone, Liberia, Guinea, and Guinea Bissau.

Video 7.7 Blood diamonds

https://www.youtube.com/watch?v=FbT_b0obeg8

Cobalt

A new form of conflict mineral has developed in recent years. With the advancement of technology (i.e. laptops, smartphones, tablets), the need and demand for the mineral cobalt have grown rapidly. It is believed that Sub-Saharan Africa, specifically Central Africa, has roughly two-thirds of the world's cobalt. To meet the demand, women and children have been imprisoned to extract cobalt from

the mines. Currently, there is not a formalized process to track and monitor where cobalt comes from. Technology companies like Apple, Samsung, Dell, and others are looking at creating their own supply chain for cobalt so that where the cobalt comes from can be monitored.

Video 7.8 Cobalt mining

https://www.youtube.com/watch?v=MxKqVY2Sa6w

Creative Commons Attribution

An Introduction to Geology by Chris Johnson, Matthew D. Affolter, Paul Inkenbrandt, Cam Mosher is licensed under CC BY- NC-SA 4.0

Natural Disasters and Human Impacts by R. Adam Dastrup, MA, GISP is licensed under CC BY-NC-SA 4.0

Physical Geology – 2nd Edition by Steven Earle is licensed under CC BY-NC-SA 4.0

Geology by Lumen Learning is licensed under CC BY-NC-SA 4.0

Earth Science by Lumen Learning is licensed under CC BY-NC-SA 4.0

Chapter 8 Rocks

THE ROCK CYCLE

The most fundamental view of Earth materials is the **rock cycle**, which presents the primary materials that comprise the Earth and describes the processes by which they form and relate to each other. The rock cycle is usually said to begin with a hot molten liquid rock called magma or lava. **Magma** forms under the Earth's surface in the crust or mantle and erupts on Earth's surface as **lava**. When magma or lava cools, it solidifies by **crystallization** in which minerals grow within the magma or lava. The rock that results from this is an igneous rock from the Latin word ignis, meaning "fire."
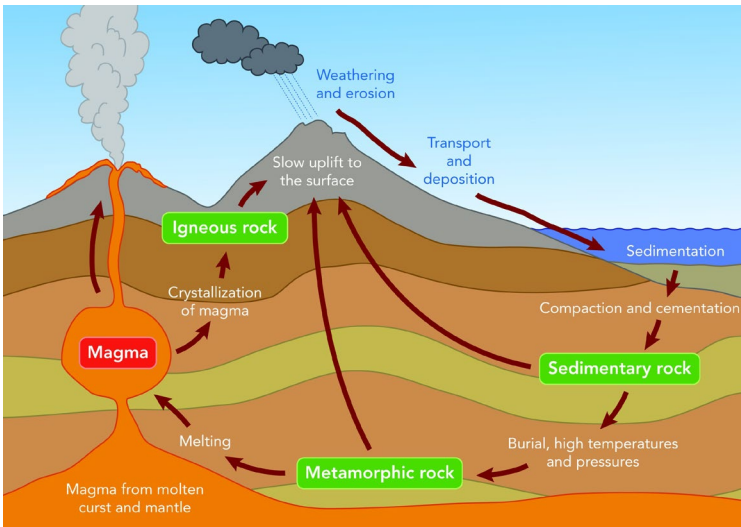
8.1 "Rock Cycle" by Siyavula Education is licensed under Creative Commons Attribution 2.0 Generic.

Physical Geography and Natural Disasters by R. Adam Dastrup,

Igneous rocks, as well as other types of rocks on Earth's surface, are exposed to weathering and erosion processes to produce sediments. **Weathering** is the physical and chemical breakdown of rocks into smaller fragments, and erosion is the removal of those fragments from their original location. Once igneous rocks are broken down and transported, these fragments or grains are considered sediments. Sediments such as gravel, sand, silt, and clay can be transported by water in the form of streams, ice in the form of glaciers, and air in the form of wind. Sediments ultimately come to rest in a process known as **deposition**. The deposited sediments accumulate in place, often underwater such as a shallow marine environment, get buried.

Within the burial process, the sediments go through compaction by the weight of overlying sediments and cementation as minerals in groundwater glue the sediments together. The process of compacting and cementing sediments together is lithification, and lithified sediments are considered a **sedimentary rock**, such as sandstone and shale. Other sedimentary rocks, known as **chemical sedimentary rocks**, are not made of weathered and eroded sedimentary fragments. The direct chemical precipitation instead makes them of minerals.

Pre-existing rocks may be metamorphosed into a **metamorphic rock**, meta- means "change", -morphos means "form" or "shape." When rocks are subjected to extreme increases in temperatures or pressures, the minerals alter into enlarged crystals, or entirely new minerals with a similar chemical make up. These elevated temperatures and pressures can occur when rocks are buried deep within the Earth's crust, or they encounter hot magma or lava. In

some cases, the temperature and pressure conditions can allow rocks to melt and create magma and lava, showing the cyclical nature of the rock cycle as new rocks are born. Click here to learn more about various igneous, sedimentary, and metamorphic rocks from the Utah Geologic Survey (UGS).

Video 1. The rock cycle

https://www.youtube.com/watch?v=kXV7D89S9sc
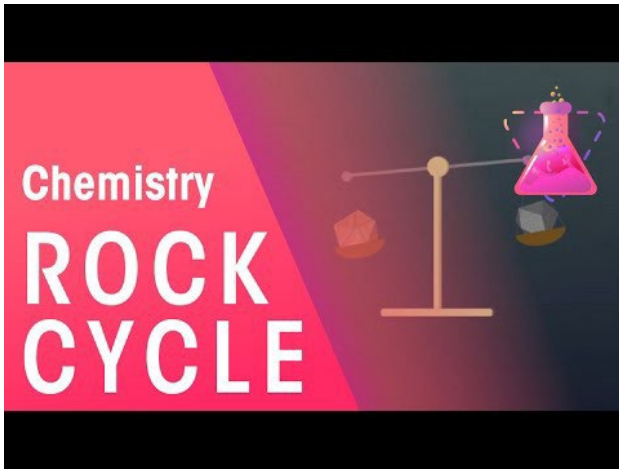
Igneous Rocks

Igneous rocks form from the cooling and hardening of molten magma in many different environments. Their composition and texture identify these rocks. More than 700 distinct types of igneous rocks are known.

Magma Composition

The rock beneath the Earth's surface is sometimes heated to high enough temperatures that it melts to create magma. Different magmas have a different composition and contain whatever elements were in the rock that melted. Magma also contains gases. (Reading: Types of Rocks | Geology, n.d.) The main elements are the same as the elements found in the crust. Whether rock melts to create magma depends on:

- **Temperature**: Temperature increases with depth, so melting is more likely to occur at greater depths.
- **Pressure**: Pressure increases with depth, but increased pressure raises the melting temperature, so melting is less likely to occur at higher pressures.

313

- **Water**: The addition of water changes the melting point of rock. As the number of water molecules increases, the melting point decreases.
- **Rock composition**: Minerals melt at different temperatures, so the temperature must be high enough to melt some minerals in the rock. The first mineral to melt from a rock will be quartz (if present), and the last will be olivine (if present).

As a rock heats up, the minerals that melt at the lowest temperatures will melt first. Partial melting occurs when the temperature on a rock is high enough to melt only some of the minerals in the rock. The minerals that will melt will be those that melt at lower temperatures. Fractional crystallization is the opposite of partial melting. This process describes the crystallization of different minerals as magma cools. (Igneous Rocks | Geology, n.d.)

Bowen's Reaction Series indicates the temperatures at which minerals melt or crystallize. An understanding of the way atoms joins to form minerals leads to an understanding of how different igneous rocks form. Bowen's Reaction Series also explains why some minerals are always found together, and some are never found together.

If the liquid separates from the solids at any time in partial melting or fractional crystallization, the liquid and solid chemical composition will be different. When that liquid crystallizes, the resulting igneous rock will have a different composition from the parent rock. (Reading: Types of Rocks | Geology, n.d.)

Intrusive Igneous Rock

Igneous rocks are called **intrusive** when they cool and solidify beneath the surface. Intrusive igneous rocks form **plutons**, and so are also called **plutonic**. A pluton is an igneous intrusive rock body that has cooled in the crust. When magma cools within the Earth, the cooling proceeds slowly. Intrusive igneous rocks cool slower than extrusive igneous rocks, which allows for more massive crystal structure to develop.

Igneous rocks make up most of the rocks on Earth. Most igneous rocks are buried below the surface and covered with sedimentary rocks or are buried beneath the ocean water. In some places, geological processes have brought igneous rocks to the surface. Yosemite is a classic example of intrusive igneous rock. The molten magma never reached Earth's surface, so the molten material had millions of years to cool down slowly to form granite. Later, geologic forces and erosion have caused those granite plutons to surface as they are today.

Extrusive Igneous Rock

Igneous rocks are called **extrusive** when they cool and solidify above the surface. These rocks usually form from a volcano, so they are also called volcanic rocks. Extrusive igneous rocks cool much more rapidly than intrusive rocks, reducing the time for crystal structure to form within the rocks.

Cooling rate and gas content create a variety of rock textures. Lavas that cool exceptionally rapidly may have a glassy texture. Those

with many holes from gas bubbles have a vesicular texture. (Igneous Rocks | Geology, n.d.)

Human Uses of Igneous Rock

Igneous rocks have a wide variety of uses, including to create buildings and statues, kitchen countertops, abrasive material for household products, or for smoothing skin. Ground-up pumice stone is sometimes added to toothpaste to function as an abrasive material to scrub teeth. Peridotite is sometimes mined for peridot, a type of olivine that is used in jewelry. Diorite was used extensively

by ancient civilizations for vases and other decorative artwork and is still used for art today.

Video 2. Igneous Rocks
https://www.youtube.com/watch?v=aCnAF1Opt8M

Sedimentary Rock

Sandstone is one of the common types of **sedimentary rocks** that form from **sediments**. There are many other types. Sediments may include:

- Fragments of other rocks that often have been worn down into small pieces, such as sand, silt, or clay.
- Organic materials, or the remains of once-living organisms.
- Chemical precipitates, which are materials that get left behind after the water evaporates from a solution.

Rocks at the surface undergo mechanical and chemical weathering. These physical and chemical processes break the rock into smaller pieces. **Physical weathering** breaks the rocks apart, while **chemical weathering** dissolves the less stable minerals. These original elements of the minerals end up in solution, and new

317

minerals may form. Sediments are removed and transported by water, wind, ice, or gravity in a process called **erosion**.

8.3 "Bryce Canyon at Sunset" by R. Adam Dastrup is copywritten.

Streams carry vast amounts of sediment. The more energy the water has, the larger the particle it can carry. A rushing river on a steep slope might be able to carry boulders. As this stream slows down, it no longer has the energy to carry large sediments and will drop them. A slower moving stream will only carry smaller particles.

Sediments are deposited on beaches and deserts, at the bottom of oceans, and in lakes, ponds, rivers, marshes, and swamps. Fast-moving avalanches and slow-moving glaciers can drop large piles of sediment. Wind can only transport sand and smaller particles. The type of sediment that is deposited will determine the type of sedimentary rock that can form. Distinct colors of sedimentary rock are determined by the environment where they are deposited. Red rocks form where oxygen is present, while darker sediments form when the environment is oxygen-poor. (Reading: Mining and Mineral Use | Geology, n.d.)

Video 3. What are sedimentary rocks

https://www.youtube.com/watch?v=Etu9BWbuDlY

Video 4. Sedimentary Rocks

https://www.youtube.com/watch?v=ptq3h6WZWYg

Sedimentary Rock Formation

Accumulated sediments harden into a rock by a process called **lithification**. Two essential steps are needed for sediments to lithify. Sediments are squeezed together by the weight of overlying sediments on top of them, called **compaction**. Cemented, non-organic sediments become **clastic rocks**. If organic material is included, they are **bioclastic rocks**.

Fluids fill in the spaces between the loose particles of sediment and crystallize to create a rock by **cementation**. When sediments settle out of calmer water, they form horizontal layers. One layer is

deposited first, and another layer is deposited on top of it. So, each layer is younger than the layer beneath it. When the sediments harden, the layers are preserved. Sedimentary rocks formed by the crystallization of chemical precipitates are called **chemical sedimentary rocks**.

Biochemical sedimentary rocks form in the ocean or a salt lake. Living creatures remove ions, such as calcium, magnesium, and potassium, from the water to make shells or soft tissue. When the organism dies, it sinks to the ocean floor to become biochemical sediment, which may become compacted and cemented into solid rock. (Reading: Mining and Mineral Use | Geology, n.d.)

Human Use of Sedimentary Rock

Sedimentary rocks are used as building stones, although they are not as hard as igneous or metamorphic rocks. Sedimentary rocks are used in construction. Sand and gravel are used to make concrete; they are also used in asphalt. Many economically valuable resources come from sedimentary rocks. Iron ore and aluminum are two examples. (Reading: Mining and Mineral Use | Geology, n.d.)

Metamorphic Rock

Metamorphism is the addition of heat and/or pressure to existing rocks, which causes them to change physically and/or chemically so that they become a new rock. **Metamorphic rocks** may change so much that they may not resemble the original rock.

Physical Geography and Natural Disasters by R. Adam Dastrup,

Any type of rock – igneous, sedimentary, or metamorphic – can become a metamorphic rock. All that is needed is enough heat and/ or pressure to alter the existing rock's physical or chemical makeup without melting the rock entirely. Rocks change during metamorphism be-cause the minerals need to be stable under the new temperature and pressure conditions. The need for stability may cause the structure of minerals to rearrange and form new minerals. Ions may move between minerals to create minerals of the different chemical compositions. Hornfels, with its alternating bands of dark and light crystals, is an excellent example of how minerals rearrange themselves during metamorphism.

Video 5. What are metamorphic rocks?
https://www.youtube.com/watch?v=1oQ1J0w3x0o

Extreme pressure may also lead to foliation, the flat layers that form

324

in rocks as pressure squeezes the rocks. **Foliation** forms typically when pressure is exerted in only one direction. Metamorphic rocks may also be **non-foliated**. Quartzite and limestone are non- foliated. The two main types of metamorphism are both related to heat within Earth:

- **Regional metamorphism**: Changes in enormous quantities of rock over a wide area caused by extreme pressure from overlying rock or compression caused by geologic processes. Deep burial exposes the rock to hot temperatures.
- **Contact metamorphism**: Changes in a rock that is in contact with magma because of the magma's extreme heat.

The Utah Geologic Survey has several resources related to landforms in Utah. They have also created a fun story map called GeoSights of popular geologic sights within the State of Utah.

Creative Commons Attribution

An Introduction to Geology by Chris Johnson, Matthew D. Affolter, Paul Inkenbrandt, Cam Mosher is licensed under CC BY- NC-SA 4.0

Natural Disasters and Human Impacts by R. Adam Dastrup, MA, GISP is licensed under CC BY-NC-SA 4.0

Physical Geology – 2nd Edition by Steven Earle is licensed under CC BY-NC-SA 4.0

Geology by Lumen Learning is licensed under CC BY-NC-SA 4.0

Earth Science by Lumen Learning is licensed under CC BY-NC-SA 4.0

Chapter 9
INTERNAL STRUCTURE OF EARTH

Layers of the Earth

To understand the details of plate tectonics, one must first understand the layers of the Earth. Humankind has insufficient first-hand information regarding what is below; most of what we know is pieced together from models, seismic waves, and assumptions based on meteorite material. In general, the Earth can be divided into layers based on chemical composition and physical characteristics. (2 Plate Tectonics – An Introduction to Geology, n.d.)
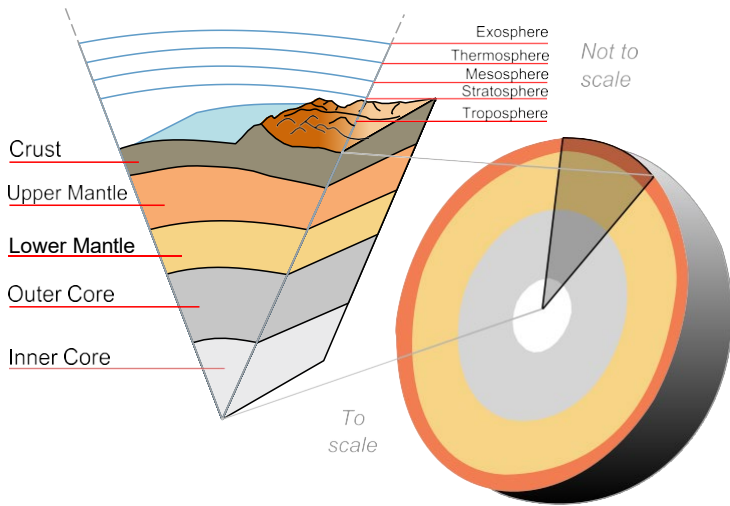
Chemical Layers

The Earth has three main divisions based on their chemical composition, which means chemical makeup. Indeed, there are countless variations in composition throughout the Earth, but only two significant changes take place, leading to three distinct chemical layers.
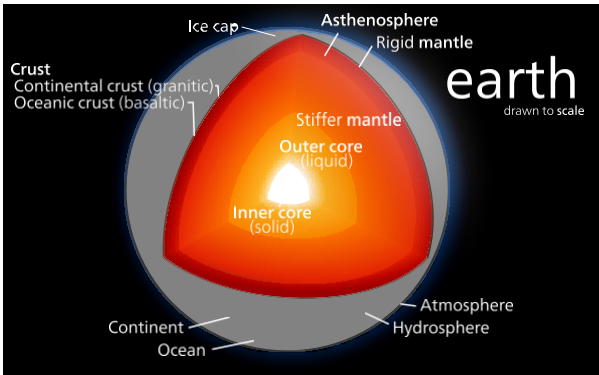
Physical Geography and Natural Disasters by R. Adam Dastrup,

Figure 9.2

Crust

The outermost chemical layer and the layer humans currently reside on is known as the **crust**. The crust has two types: **continental crust**, which is relatively low density and has a composition similar to granite, and **oceanic crust**, which is relatively high density (especially when it is cold and old) and has a composition similar to basalt. In the lower part of the crust, rocks start to be more ductile and less brittle, because of added heat. Earthquakes, therefore, generally occur in the upper crust.

At the base of the crust is a substantial change in seismic velocity called the **Mohorovičić Discontinuity**, or **Moho** for short, discovered by Andrija Mohorovičić (pronounced mo-ho-ro-vee-cheech) in 1909 by studying earthquake wave paths in his native Croatia. It is caused by the dramatic change in composition that occurs between the mantle and the crust. Underneath the oceans, the Moho is about 5 km down. Under continents, the average is about 30-40 km, except near a sizeable mountain-building event, known as an **orogeny**, where that thickness is about doubled.

Mantle

The **mantle** is the layer below the crust and above the core. It is the most substantial layer by volume, extending from the base of the crust to a depth of about 2900 km. Most of what we know about the mantle comes from seismic waves, though some direct information can be gathered from parts of the ocean floor that are brought to the surface, known as ophiolites. Also, carried within magma are xenoliths, which are small chunks of lower rock carried to the surface by eruptions. These xenoliths are made of the rock peridotite, which on the scale of igneous rocks is ultramafic. We assume the majority of the mantle is made of peridotite.

Core

The **core** of the Earth, which has both liquid and solid components, is made mostly of iron, nickel, and oxygen. First discovered in 1906 by looking into seismic data, it took the union of modeling, astronomical insight, and seismic data to arrive at the idea that the core is mostly metallic iron. Meteorites contain much more iron than typical surface rocks, and if meteoric material is what made the Earth, the core would have formed as dense material (including iron and nickel) sank to the center of the Earth via its weight as the planet formed, heating the Earth intensely.

Physical Layers

The Earth can also be broken down into five distinct physical layers based on how each layer responds to stress. While there is some

overlap in the chemical and physical designations of layers, precisely the core-mantle boundary, there are significant differences between the two systems. (2 Plate Tectonics – An Introduction to Geology, n.d.)
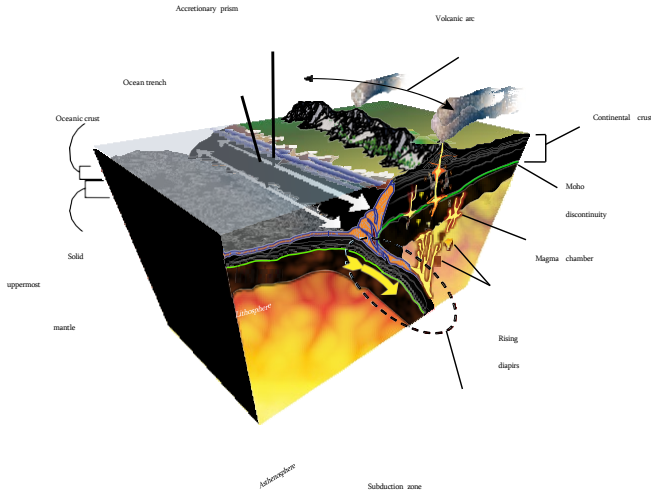
Lithosphere

The **lithosphere**, with 'litho' meaning rock, is the outermost physical layer of the Earth. Including the crust, it has both an oceanic component and a continental component. **Oceanic lithosphere**, ranging from a thickness of zero (at the forming of new plates on the mid-ocean ridge) to 140 km, is thin and rigid. **The continental lithosphere** is more plastic (especially with depth) and is overall thicker, from 40 to 280 km thick. Most importantly, the lithosphere is not continuous. It is broken into several segments that geologists call plates. A plate boundary is where two plates meet and move relative to each other. It is at and

330

near plate boundaries where plate tectonics' real action is seen, including mountain building, earthquakes, and volcanism.

Asthenosphere

The **asthenosphere**, with 'astheno' meaning weak, is the layer below the lithosphere. The most distinctive property of the asthenosphere is movement. While still solid, over geologic time scales, it will flow and move because it is mechanically weak. In this layer, partly driven by convection of intense interior heat, movement allows the lithospheric plates to move. Since certain types of seismic waves pass through the asthenosphere, we know that it is solid, at least at the short time scales of the passage of seismic waves. The depth and occurrence of the asthenosphere are dependent on heat and can be very shallow at mid-ocean ridges and very deep in plate interiors and beneath mountains.

Mesosphere

The **mesosphere**, or **lower mantle** as it is sometimes called, is more rigid and immobile than the asthenosphere, though still hot. This can be attributed to increased pressure with depth. Between approximately 410 and 660 km depth, the mantle is in a state of transition, as minerals with the same composition are changed to various forms, dictated by increasing pressure conditions. Changes in seismic velocity show this, and this zone also can be a physical barrier to movement. Below this zone, the mantle is uniform and homogeneous, as no significant changes occur until the core is reached.

The **outer core** is the only liquid layer found within Earth. It starts at 2,890 km (1,795 mi) depth and extends to 5,150 km (3,200 mi). Inge Lehmann, a Danish geophysicist, in 1936, was the first to prove that there was an inner core that was solid within the liquid outer core based on analyzing seismic data. The solid inner core is about 1,220 km (758 mi) thick, and the outer core is about 2,300 km (1,429 mi) thick.

It seems like a contradiction that the hottest part of the Earth is substantial, as hot temperatures usually lead to melting or boiling. The solid **inner core** can be explained by understanding that the immense pressure inhibits melting, though as the Earth cools by heat flowing outward, the inner core grows slightly larger over time. As the liquid iron and nickel in the outer core moves and convects, it becomes the most likely source for Earth's magnetic field. This is critically important to maintaining the atmosphere and conditions on Earth that make it favor-able to life. Loss of outer core convection and the Earth's magnetic field could strip the atmosphere of most of the gases essential to life and dry out the planet, much like what has happened to Mars.
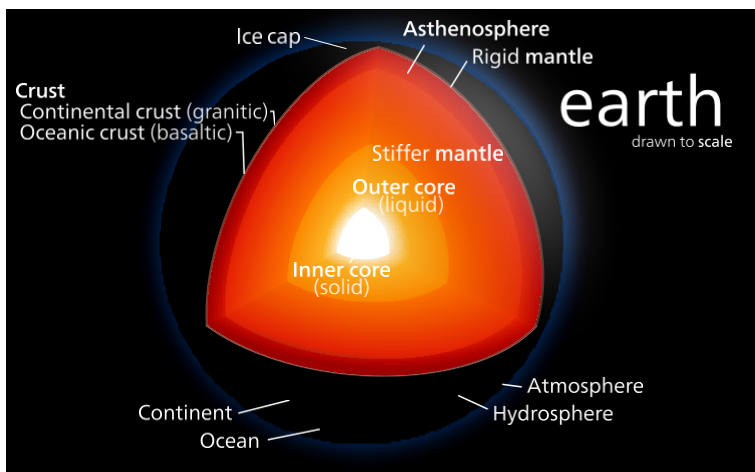
"Structure of Earth" is licensed under Creative Commons Attribution-ShareAlike 3.0 Unported.
Figure 9.4

Structure of Earth's Crust

The fundamental unifying principle of geology and the rock cycle is the theory of **Plate Tectonics**. Plate tectonics describes how the layers of the Earth move relative to each other. Specifically, the outer layer divided into tectonic or lithospheric plates. As the tectonic plates float on a mobile layer beneath called the asthenosphere, they collide, slide past each other, and split apart. Significant landforms are created at these plate boundaries, and rocks making up the tectonic plates move through the rock cycle.

The following is a summary of the Earth's layers based on chemical composition (or the chemical makeup of the layers). Earth has three main geological layers based on chemical composition – crust, mantle, and core. The outermost layer is the **crust** and is composed

333

of mostly silicon, oxygen, aluminum, iron, and magnesium. There are two types of crust, continental and oceanic crust. **Continental crust** is about 50 kilometers (30 miles) thick, represents most of the continents, and is composed of low-density igneous and sedimentary rocks. Oceanic crust is approximately 10 kilometers (6 miles) thick, makes up most of the ocean floor, and covers about 70 percent of the planet. **Oceanic crust** is high-density igneous basalt-type rocks. The moving tectonic plates are made of crust, and some of the next layers within the earth called the **mantle**. The crust and this portion of the upper mantle are rigid and called the **lithosphere** and make up the tectonic plates.

Continents

The oldest continental rocks are billions of years old, so the continents have had much time to happen to them. **Constructive forces** cause physical features on Earth's surface known as landforms to grow. **Crustal deformation** – when crust compresses, pulls apart, or slides past other crust – results in hills, valleys, and other landforms. Mountains rise when continents collide when one slab of ocean crust plunges beneath another or a slab of continental crust to create a chain of volcanoes. Sediments are deposited to form landforms, such as deltas. Volcanic eruptions can also be destructive forces that blow landforms apart. The **destructive forces** of weathering and erosion modify landforms. Water, wind, ice, and gravity are essential forces of erosion.

Oceanic Basins

The ocean basins are all younger than 180 million years. Although the ocean basins begin where the ocean meets the land, the continent extends downward to the seafloor, so the continental margin is made of continental crust.

The ocean floor itself is not flat. The most distinctive feature is the mountain range that runs through much of the ocean basin, known as the mid-ocean ridge. The ocean trenches are the deepest places of the ocean, many of which are found around the edge of the Pacific Ocean. Chains of volcanoes are also found in the center of the oceans, such as around Hawaii. Flat plains are found on the ocean floor with their features covered by mud.
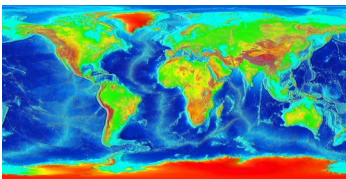
Figure 9.5
"Elevation" is NOAA is licensed as Public Domain.

Creative Commons Attribution

An Introduction to Geology by Chris Johnson, Matthew D. Affolter, Paul Inkenbrandt, Cam Mosher is licensed under CC BY- NC-SA 4.0

Natural Disasters and Human Impacts by R. Adam Dastrup, MA, GISP is licensed under CC BY-NC-SA 4.0

Physical Geology – 2nd Edition by Steven Earle is licensed under CC BY-NC-SA 4.0

Geology by Lumen Learning is licensed under CC BY-NC-SA 4.0

Earth Science by Lumen Learning is licensed under CC BY-NC-SA 4.0

Chapter 10
THEORY OF PLATE TECTONICS

Continental Drift Hypothesis

Alfred Wegener (1880-1930) was a German scientist who specialized in meteorology and climatology. He had a knack for questioning accepted ideas, and this started in 1910 when he disagreed with isostasy (vertical land movement due to the weight being removed or added) as the explanation for the Bering Land Bridge. After literary reviews, he published a hypothesis stating the continents had moved in the past. While he did not have the precise mechanism worked out, he had an extensive list of evidence that backed up his hypothesis of **continental drift**.

Video 1. The whole saga of supercontinents

https://www.youtube.com/watch?v=KfYn9KVya-Q

Early Evidence for Continental Drift

The first piece of evidence is that the shape of the coastlines of some continents fit together like pieces of a jigsaw puzzle. Since the first world map, people have noticed the similarities in the coastlines of South America and Africa, and the continents being ripped apart had even been mentioned as an explanation. Antonio Snider-Pellegrini even did preliminary work on continental separation and matching fossils in 1858.

What Wegener did differently than others was synthesized a significant amount of data in one place, as well as use the shape of

the continental shelf, the actual edge of the continent, instead of the current coastline, which fit even better than previous efforts. Wegener also compiled and added to evidence of similar rocks, fossils, and glacial formations across the oceans.

Video 2. What did Pangea look like?

https://www.youtube.com/watch?v=VKq0pr4rbRs

Fossil Evidence

For example, the primitive aquatic reptile Mesosaurus was found on the separate coastlines of the continents of Africa and South America, and the reptile Lystrosaurus was found on Africa, India, and Antarctica. These were land-dwelling creatures that could not have swam across an entire ocean; thus, this was explained away by

opponents of continental drift by land bridges. The land bridges had eroded, allowed animals and plants to move between the continents. However, some of the presumed land bridges would have had to have stretched across broad, deep oceans.
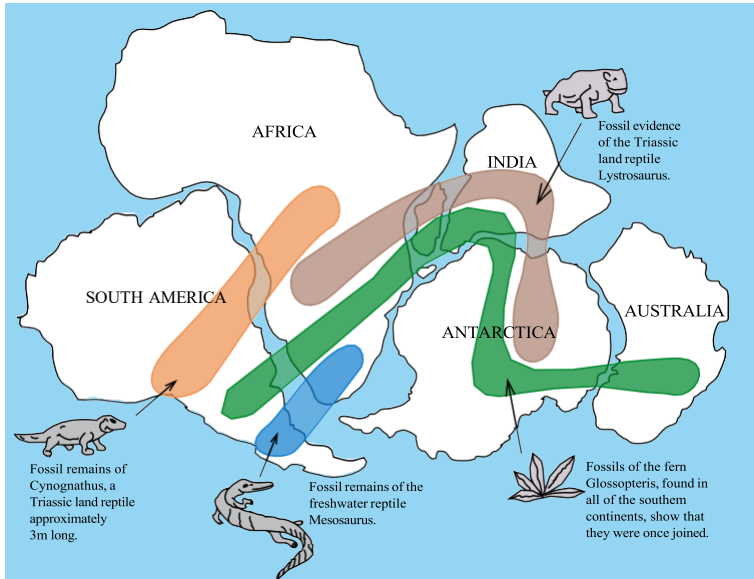
Figure 10.1 "Snider-Pellegrini Wegener Fossil Map" is licensed under Public Domain.

Geologic Evidence

Mountain ranges with the same rock types, structures, and ages are now on opposite sides of the Atlantic Ocean. The Appalachians of the eastern United States and Canada, for example, are just like mountain ranges in eastern Greenland, Ireland, Great Britain, and Norway. Wegener concluded that they formed a single mountain range that was separated as the continents drifted. (Dastrup, 2014)

Climatic Evidence

Another significant piece of evidence was climate anomalies. Late Paleozoic glacial evidence was found in widespread, warm areas like southern Africa, India, Australia, and the Arabian subcontinent. Wegener himself had found evidence of tropical plant fossils in areas north of the Arctic Circle. According to Wegener, the more straightforward explanation that fits all the climate, rock, and fossil observations, mainly as more data were collected, involved moving continents.

Grooves and rock deposits left by ancient glaciers are found today on different continents close to the equator. This would indicate that the glaciers either formed in the middle of the ocean or covered most of the Earth. Today glaciers only form on land and nearer the poles. Wegener thought that the glaciers were centered over the southern landmass close to the South Pole, and the continents later moved to their present positions.

Proposed Mechanism for Continental Drift

Wegener's work was considered a fringe theory for his entire life. One of the most significant flaws and easiest dismissals of Wegener's hypothesis was a mechanism for the movement of the continents. The continents did not appear to move, and extraordinary evidence would need to be provided to change the establishment's minds, including a mechanism for movement. Other pro-continental drift followers had used expansion, contraction, or even the origin of the

Moon as ideas to how the continents moved. Wegener used centrifugal forces and precession to explain the movement, but that was proven wrong. He had some speculation about **seafloor spreading**, with hints of convection within the earth, but these were unsubstantiated. As it turns out, convection within the mantle has been revealed as a significant force in driving plate movements, according to current knowledge.
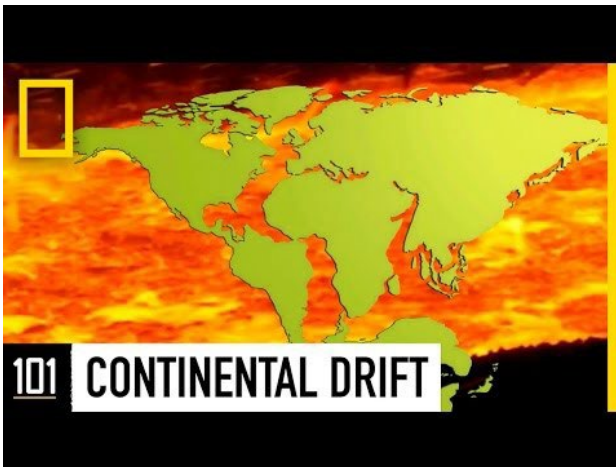
Video 3. Continental drift 101

https://www.youtube.com/watch?v=Wq9kLzm36h0

Theory of Plate Tectonics

Wegener died in 1930 on an expedition in Greenland. He was poorly respected in his lifetime, and his ideas of moving continents

seemed destined to be lost to history as a fringe idea. However, starting in the 1950s, evidence started to trickle in that made continental drift more viable. By the 1960s, there was enough evidence supporting Wegener's missing mechanism, seafloor spreading, allowing continental drift's hypothesis to develop into the **Theory of Plate Tectonics**. Widespread acceptance among scientists has transformed Wegener's hypothesis to a Theory. Today, GPS and earthquake data continue to back up the theory. Below are the pieces of evidence that allowed the transformation.
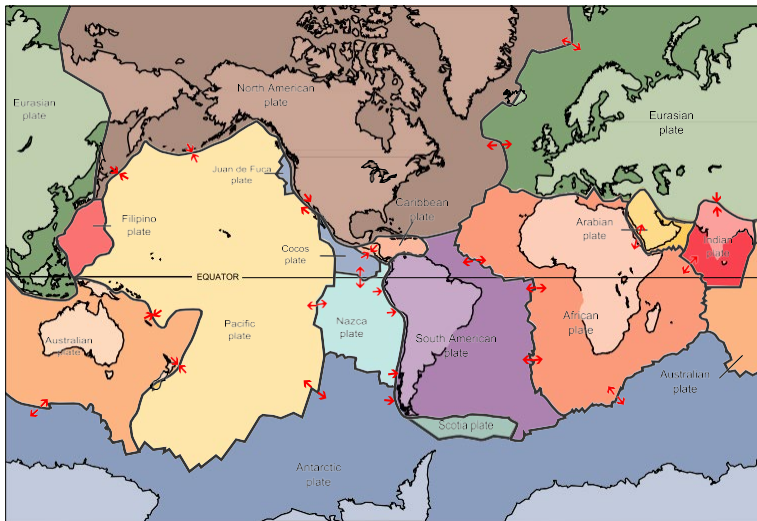
Figure 10.2 "Tectonic Plates" by the United States Geologic Survey (USGS) is licensed under Public Domain.

Mapping the Ocean Floors

Starting in 1947 and using an adaptation of sonar, researchers began to map a poorly understood topographic and thermal high in the mid-Atlantic. Bruce Heezen and Marie Tharp were the first to

Physical Geography and Natural Disasters by R. Adam Dastrup

make a detailed map of the ocean floor, and this map revealed the **mid-Atlantic Ridge**, a basaltic feature, unlike the continents. Initially, this was thought to be part of an expanding Earth or a mechanism for the ocean's growth. Transform faults were also added to explain movements more completely. When it was later realized that earthquake epicenters were also found within this feature, the idea that this was part of the continental movement took hold.

Video 4. Marie Tharp's maps

https://www.youtube.com/watch?v=vE2FK0B7gPo

Another way the seafloor was mapped was **magnetical**. Scientists had long known of strange magnetic anomalies (magnetic values that differ from expected values) associated with the ocean floor. This tool was adapted by geologists later for further study of the

343

ocean depths, including strange alternating symmetrical stripes on both sides of a feature (which would be discovered later as the mid-ocean ridge) showing reversing magnetic pole directions. By 1963, these magnetic stripes would be explained in concordance with Hess's spreading model and others.

Seafloor sediment was also an important feature that was measured in the oceans, both with dredging and with drilling. Sediment was believed to have been piling up on ocean floors for an exceptionally long time in a static model of accumulation. Initial studies showed less sediment than expected, and initial results were even used to argue against the continental movement. With more time, researchers discovered thinner sediment close to ridges, indicating a younger age.
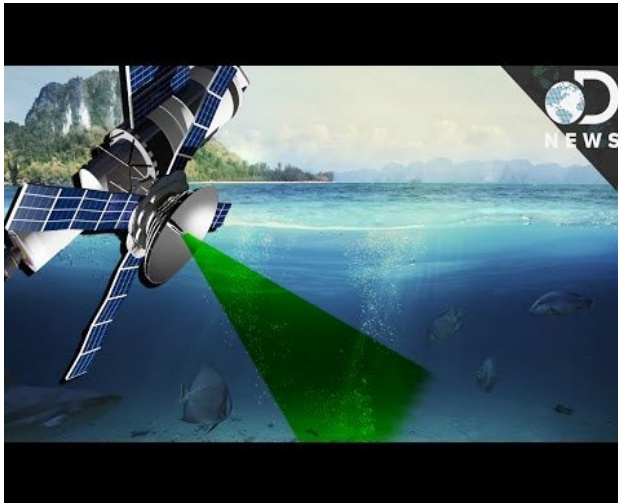
Video 5. How Satellites Mapped The Whole Ocean Floor

https://www.youtube.com/watch?v=HZVFsvjGx70&t=53s

Wadati-Benioff Zones

Around the same time that mid-ocean ridges were being investigated, ocean trenches and island arcs were also being linked to seismic action, explaining the opposite sides of the plates. A zone of deep earthquakes that lay along a plane, trending from the surface near the trenches to inside the Earth beneath the continents and island arcs, were recognized independently by several scientists. Today called the **Wadati-Benioff zone**; it was an essential piece of the puzzle.

Paleomagnetism

Magnetic field mapping, as mentioned above, was not the only way magnetism was used in the development of plate tectonics. The first new hard evidence that supported plate motion came from paleomagnetism. **Paleomagnetism** is the study of magnetic fields frozen within rocks, basically a fossil compass. This is typically most useful with igneous rocks where magnetic minerals like magnetite crystallizing in the magma align with the Earth's magnetic field and in the solid rock point to the paleo-magnetic north. The earth's magnetic field creates flux lines surrounding the magnetic north and south poles (like a bar magnet), which are both close to the

Physical Geography and Natural Disasters by R. Adam Dastrup

Earth's rotational north and south poles. In igneous rocks, magnetic minerals align parallel with these flux lines. Thus, both magnetic inclination, related to latitude, and declination related to magnetic north are preserved in the rocks.

Scientists had noticed that, for some time, the magnetic north, to which many rocks pointed, was nowhere close to the current magnetic north. This was explained by implying the magnetic north pole moved over time. Eventually, scientists started to realize that moving continents explained the data better than moving the pole around alone.
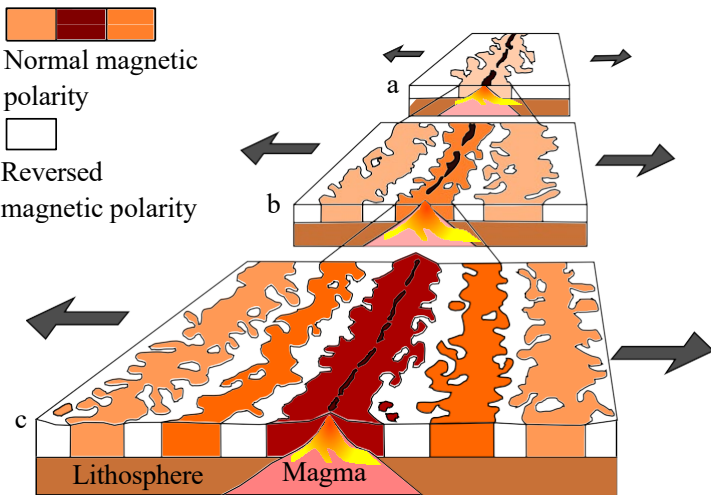
Figure 10.3 "Oceanic Stripe Magnetic Anomalies Scheme" by the United States Geologic Survey (USGS) is licensed under Public Domain.

Seafloor Spreading

World War II gave scientists the tools to find the mechanism for

continental drift that had eluded Wegener. Maps and other data gathered during the war allowed scientists to develop the **seafloor spreading hypothesis**. This hypothesis traces oceanic crust from its origin at a mid-ocean ridge to its destruction at a deep-sea trench and is the mechanism for continental drift. (Seafloor Spreading | Earth Science, n.d.)

During World War II, battleships and submarines carried echo sounders to locate enemy submarines. Echo sounders produce sound waves that travel outward in all directions, bounce off the nearest object, and then return to the ship. By knowing the sound speed in seawater, scientists calculate the distance to the object based on the time it takes for the wave to make a round trip. During the war, most of the sound waves ricocheted off the ocean bottom. This animation shows how sound waves are used to create pictures of the seafloor and ocean crust.
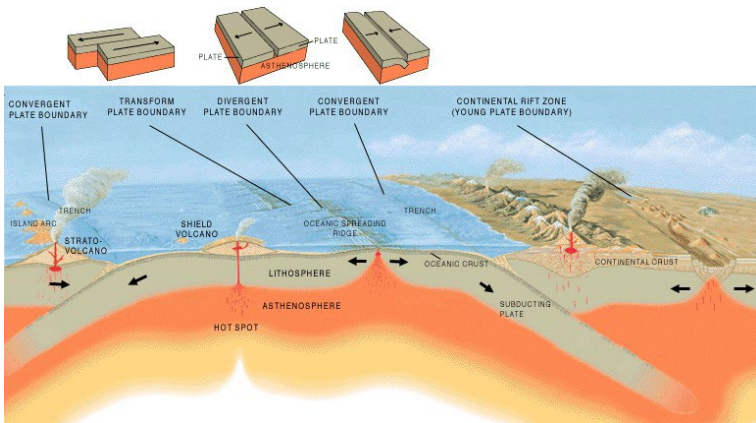
Figure 10.4 "Tectonic Plate Boundaries" by the United States Geologic Survey (USGS) is licensed under Public Domain.

After the war, scientists pieced together the ocean depths to

347

produce bathymetric maps, which revealed the ocean floor features as if the water were taken away. Even scientists were amazed that the seafloor was not completely flat. What was discovered was a large chain of mountains along the deep seafloor, called mid-ocean ridges. Scientists also discovered deep-sea trenches along the edges of continents or in the sea near active volcanic chains. Finally, large, flat areas called abyssal plains we found. When they first observed these bathymetric maps, scientists wondered what had formed these features.

Scientists brought these observations together in the early 1960s to create the **seafloor spreading hypothesis**. In this hypothesis, a hot, buoyant mantle creates a mid-ocean ridge, causing the ridge to rise upward. The hot magma at the ridge erupts as lava that forms new seafloor. When the lava cools, the magnetite crystals take on the current magnetic polarity, and as more lava erupts, it pushes the seafloor horizontally away from the ridge axis. (Seafloor Spreading | Earth Science, n.d.)

The magnetic stripes continue across the seafloor. As oceanic crust forms and spreads, moving away from the ridge crest, it pushes the continent away from the ridge axis. If the oceanic crust reaches a deep-sea trench, it sinks into the trench and is lost into the mantle. Scientists now know that the oldest crust is coldest and lies deepest in the ocean because it is less buoyant than the hot new crust.

Figure 10.5 "Age of Oceanic Lithosphere" by the National
Oceanic and Atmospheric Administration (NOAA) is licensed
under Public Domain.

Unifying Theory of Plate Tectonics

Using all the evidence mentioned, the theory of plate tectonics took
shape. In 1966, J. Tuzo Wilson was the first scientist to put the
entire picture together of an opening and closing ocean. Before
long, models were proposed showing the plates moving concerning
each other with clear boundaries between them, and scientists had
also started to piece together complicated tectonic histories. The
plate tectonic revolution had taken hold.

349

Video 6. Why it took scientists so long to accept this very obvious idea

https://www.youtube.com/watch?v=KB7HzF2O3Kg&t=58s

Seafloor and continents move around on Earth's surface, but what is moving? What part of the Earth makes up the "plates" in plate tectonics? This question was also answered because of technology developed during the Cold War. The **tectonic plates** are made up of the lithosphere. During the 1950s and early 1960s, scientists set up seismograph networks to see if enemy nations were testing atomic bombs. These seismographs also recorded all the earthquakes around the planet. The seismic records could be used to locate an earthquake's epicenter, the point on Earth's surface directly above the earthquake's place. Earthquake epicenters outline these tectonic plates. Mid-ocean ridges, trenches, and significant faults mark the edges of these plates along with where earthquakes

occur. (Reading: Theory of Plate Tectonics | Geology (Modification for Lehman College, CUNY), n.d.)

The lithosphere is divided into a dozen major and several minor tectonic plates. The plates' edges can be drawn by connecting the dots that mark earthquakes' epicenters. A single tectonic plate can be made of all oceanic lithosphere or all continental lithosphere, but nearly all plates are made of a combination of both. Movement of the plates over Earth's surface is termed **plate tectonics**. Plates move at a rate of a few centimeters a year, about the same rate fingernails grow. (Theory of Plate Tectonics | Earth Science, n.d.)

Creative Commons Attribution

An Introduction to Geology by Chris Johnson, Matthew D. Affolter, Paul Inkenbrandt, Cam Mosher is licensed under CC BY- NC-SA 4.0

Natural Disasters and Human Impacts by R. Adam Dastrup, MA, GISP is licensed under CC BY-NC-SA 4.0

Physical Geology – 2nd Edition by Steven Earle is licensed under CC BY-NC-SA 4.0

Geology by Lumen Learning is licensed under CC BY-NC-SA 4.0

Earth Science by Lumen Learning is licensed under CC BY-NC-SA 4.0

Chapter 11

Plates, Plate Motions, and Plate-Boundary Processes

Continental drift and sea-floor spreading became widely accepted around 1965 as more and more geologists started thinking in these terms. By the end of 1967, Earth's surface had been mapped into a series of plates (Figure 11.1). The major plates are Eurasia, Pacific, India, Australia, North America, South America, Africa, and Antarctic. There are also numerous small plates (e.g., Juan de Fuca, Nazca, Scotia, Philippine, Caribbean), and many very small plates or sub-plates. For example the Juan de Fuca Plate is actually three separate plates (Gorda, Juan de Fuca, and Explorer) that all move in the same general direction but at slightly different rates.
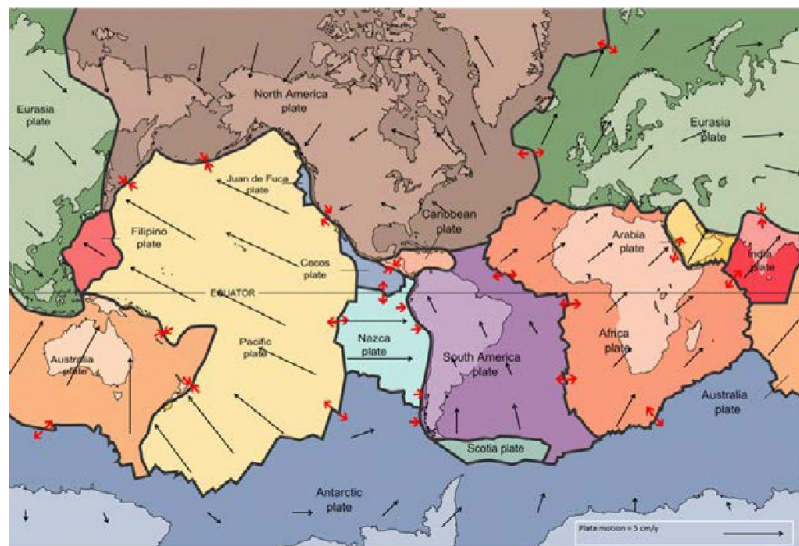
Figure 11.1 A map showing 15 of the Earth's tectonic plates and the approximate rates and directions of plate motions. [SE after USGS, http://en.wikipedia.org/wiki/Plate_tectonics#/media/ File:Plates_tect2_en.svg]

Rates of motions of the major plates range from less than 1 cm/y to over 10 cm/y. The Pacific Plate is the fastest at over 10 cm/y in some areas, followed by the Australian and Nazca Plates. The North American Plate is one of the slowest, averaging around 1 cm/y in the south up to almost 4 cm/y in the north.

Plates move as rigid bodies, so it may seem surprising that the North American Plate can be moving at different rates in different places. The explanation is that plates move in a rotational manner. The North American Plate, for example, rotates counter-clockwise; the Eurasian Plate rotates clockwise.

Boundaries between the plates are of three types: **divergent** (i.e., moving apart)*,* **convergent** (i.e., moving together), and **transform** (moving side by side). Before we talk about processes at plate boundaries, it's important to point out that there are never gaps between plates. The plates are made up of crust and the lithospheric part of the mantle (Figure 11.2), and even though they are moving all the time, and in different directions, there is never a significant amount of space between them. Plates are thought to move along the lithosphere-asthenosphere boundary, as the asthenosphere is the zone of partial melting. It is assumed that the relative lack of strength of the partial melting zone facilitates the sliding of the lithospheric plates.

352
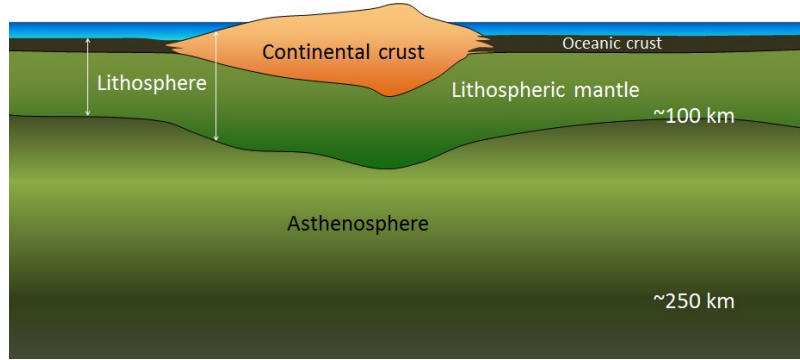
Figure 11.2 The crust and upper mantle. Tectonic plates consist of lithosphere, which includes the crust and the lithospheric (rigid) part of the mantle. [SE]
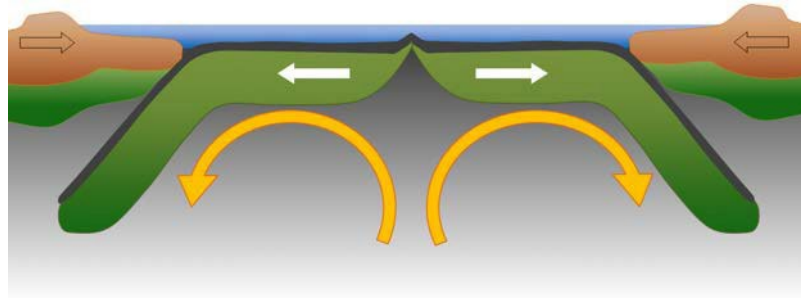
Figure 11.3 A representation of Harold Hess's model for sea-floor spreading and subduction [SE]

Phenomenon of reversals in Earth's magnetic field.

At spreading centers, the lithospheric mantle may be very thin because the upward convective motion of hot mantle material generates temperatures that are too high for the existence of a significant thickness of rigid lithosphere (Figure 11.2). The fact that the plates include both crustal material and lithospheric mantle material makes it possible for a single plate to be made up of both oceanic and continental crust. For example, the North American Plate includes most of North America, plus half of the northern Atlantic Ocean. Similarly the South American Plate extends across the western part of the southern Atlantic Ocean, while the European and African plates each include part of the eastern Atlantic Ocean. The Pacific Plate is almost entirely oceanic, but it does include the part of California west of the San Andreas Fault.

Divergent Boundaries

Divergent boundaries are spreading boundaries, where new oceanic crust is created from magma derived from partial melting of the mantle caused by decompression as hot mantle rock from depth is moved toward the surface (Figure 11.4). The triangular zone of partial melting near the ridge crest is approximately 60 km thick and the proportion of magma is about 10% of the rock volume, thus producing crust that is about 6 km thick. Most divergent boundaries are located at the oceanic ridges (although some are on land), and the crustal material created at a spreading boundary is always oceanic in character; in other words, it is mafic igneous rock (e.g., basalt or gabbro, rich in ferromagnesian minerals). Spreading rates vary considerably, from 1 cm/y to 3 cm/y in the Atlantic, to between 6 cm/y and 10 cm/y in the Pacific. Some of the processes taking place in this setting include:

- Magma from the mantle pushing up to fill the voids left by divergence of the two plates
- **Pillow lavas** forming where magma is pushed out into seawater (Figure 11.5)
- **Vertical** sheeted dykes intruding into cracks resulting from the spreading
- Magma cooling more slowly in the lower part of the new crust and forming gabbro bodies

Spreading is hypothesized to start within a continental area with up-warping or doming related to an underlying mantle plume or series of mantle plumes. The buoyancy of the mantle plume material creates a dome within the crust, causing it to fracture in a radial pattern, with three arms spaced at approximately 120° (Figure 11.6). When a series of mantle plumes exists beneath a large continent, the resulting rifts may align and lead to the formation of a rift valley (such as the present-day Great Rift Valley in eastern Africa). It is suggested that this type of valley eventually develops into a linear sea (such as the present-day Red Sea), and finally into an ocean (such as the Atlantic). It is likely that as many as 20 mantle plumes, many of which still exist, were responsible for the initiation of the rifting of Pangea along what is now the mid-Atlantic ridge (see Figure 11.6).

354
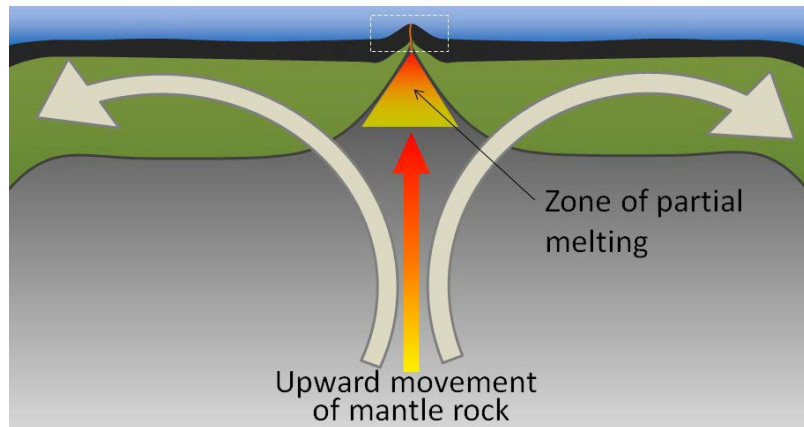
Figure 11.4 The general processes that take place at a divergent boundary. The area within the dashed white rectangle is shown in Figure 10.19. [SE]
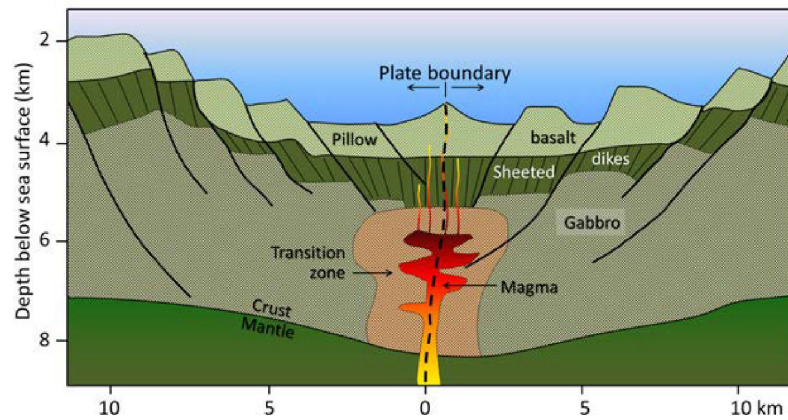
Figure 11.5 Depiction of the processes and materials formed at a divergent boundary [SE after Keary and Vine, 1996, Global Tectonics (2ed), Blackwell Science Ltd., Oxford]
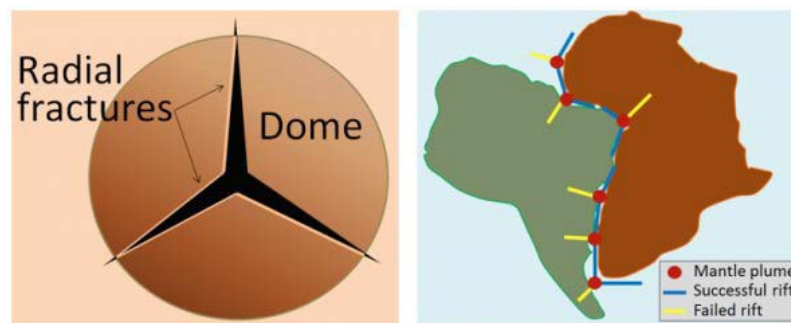
Figure 11.6 Depiction of the process of dome and three-part rift formation (left) and of continental rifting between the African and South American parts of Pangea at around 200 Ma (right) [SE]

355

Convergent Boundaries

Convergent boundaries, where two plates are moving toward each other, are of three types, depending on the type of crust present on either side of the boundary — oceanic or continental. The types are ocean-ocean, ocean-continent, and continent-continent.

At an ocean-ocean convergent boundary, one of the plates (oceanic crust and lithospheric mantle) is pushed, or subducted, under the other. Often it is the older and colder plate that is denser and subducts beneath the younger and hotter plate. There is commonly an ocean trench along the boundary. The subducted lithosphere descends into the hot mantle at a relatively shallow angle close to the subduction zone, but at steeper angles farther down (up to about 45°). Significant volume of water within the subducting material is released as the subducting crust is heated. This water is mostly derived from alteration of pyroxene and olivine to serpentine near the spreading ridge shortly after the rock's formation. It mixes with the overlying mantle, and the addition of water to the hot mantle lowers the crust's melting point and leads to the formation of magma (flux melting). The magma, which is lighter than the surrounding mantle material, rises through the mantle and the overlying oceanic crust to the ocean floor where it creates a chain of volcanic islands known as an island arc. A mature island arc develops into a chain of relatively large islands (such as Japan or Indonesia) as more and more volcanic material is extruded and sedimentary rocks accumulate around the islands.

Earthquakes take place close to the boundary between the subducting crust and the overriding crust. The largest earthquakes occur near the surface where the subducting plate is still cold and strong.
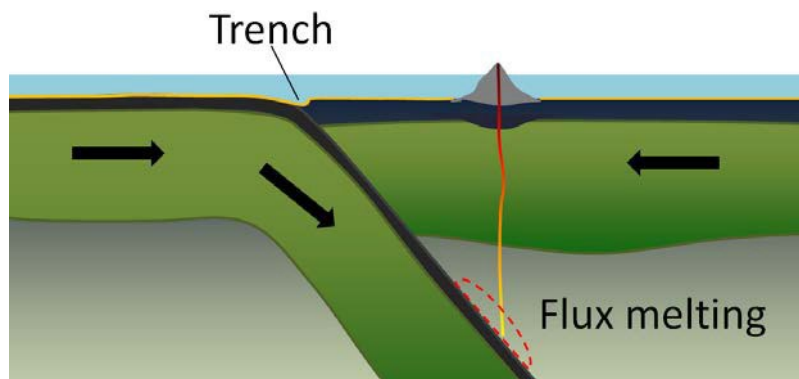
Figure 11.7 Configuration and processes of an ocean-ocean convergent boundary [SE]

Examples of ocean-ocean convergent zones are subduction of the Pacific Plate south of Alaska (Aleutian Islands) and west of the Philippines, subduction of the India Plate south of Indonesia, and subduction of the Atlantic Plate beneath the Caribbean Plate (Figure 11.7).

At an ocean-continent convergent boundary, the oceanic plate is pushed under the continental plate in the same manner as at an ocean-ocean boundary. Sediment that has accumulated on the **continental slope** is thrust up into an accretionary wedge, and compression leads to thrusting within the continental plate (Figure 11.8). The mafic magma produced adjacent to the subduction zone rises to the base of the continental crust and leads to partial melting of the crustal rock. The resulting magma ascends through the crust, producing a mountain chain with many volcanoes.

Examples of ocean-continent convergent boundaries are subduction of the Nazca Plate under South America (which has created the Andes Range) and subduction of the Juan de Fuca Plate under North America (creating the mountains Garibaldi, Baker, St. Helens, Rainier, Hood, and Shasta, collectively known as the Cascade Range).

A continent-continent collision occurs when a continent or large island that has been moved along with

356
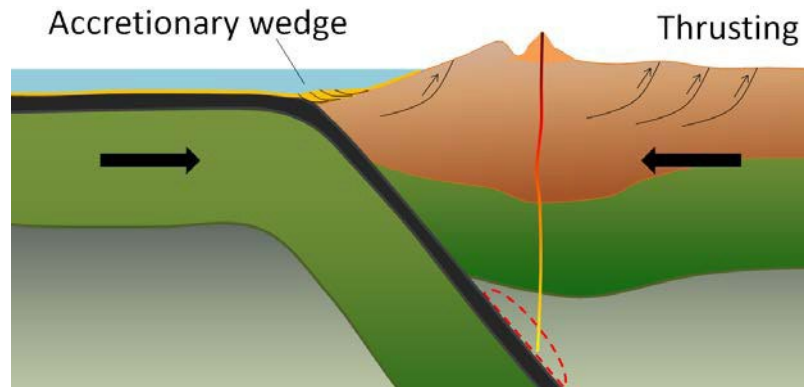
Figure 11.8 Configuration and processes of an ocean-continent convergent boundary [SE]

subducting oceanic crust collides with another continent (Figure 11.9). The colliding continental material will not be subducted because it is too light (i.e., because it is composed largely of light continental rocks [SIAL]), but the root of the oceanic plate will eventually break off and sink into the mantle. There is tremendous deformation of the pre-existing continental rocks, and creation of mountains from that rock, from any sediments that had accumulated along the shores (i.e., within geosynclines) of both continental masses, and commonly also from some ocean crust and upper mantle material.
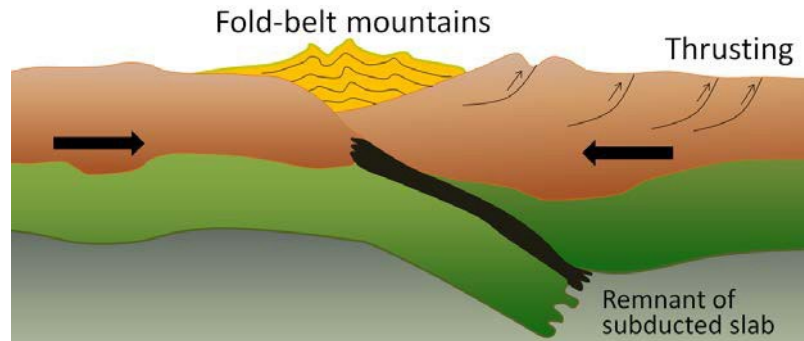
Figure 11.9 Configuration and processes of a continent-continent convergent boundary [SE]

Examples of continent-continent convergent boundaries are the collision of the India Plate with the Eurasian Plate, creating the **Himalaya Mountains**, and the collision of the African Plate with the Eurasian Plate, creating the series of ranges extending from the **Alps** in Europe to the **Zagros Mountains** in Iran. The **Rocky Mountains** in B.C. and Alberta are also a result of continent-continent collisions.

Transform Boundaries

Transform boundaries exist where one plate slides past another without production or destruction of crustal material. As explained above, most transform faults connect segments of mid-ocean ridges and are thus ocean-ocean plate boundaries (Figure 11.10). Some transform faults connect continental parts of plates. An example is the San Andreas Fault, which connects the southern end of the Juan de Fuca Ridge with the northern end of the East Pacific Rise (ridge) in the Gulf of California (Figures 11.10 and 11.11). The part of California west of the San Andreas Fault and all of Baja California are on the Pacific Plate. Transform faults do not just connect divergent boundaries. For example, the Queen Charlotte Fault connects the north end of the Juan de Fuca Ridge, starting at the north end of Vancouver Island, to the Aleutian subduction zone.

357
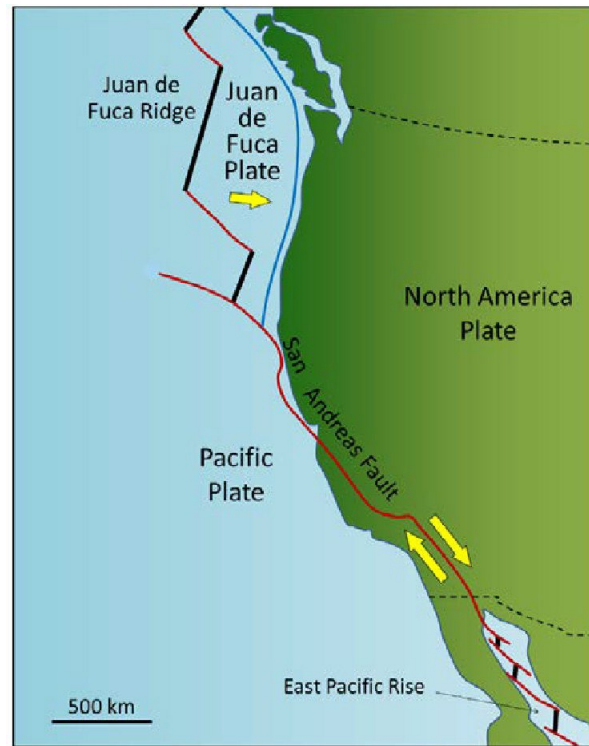
Figure 11.10 The San Andreas Fault extends from the north end of the

East Pacific Rise in the Gulf of California to the southern end of the Juan de Fuca Ridge. All of the red lines on this map are transform faults. [SE]

Figure 11.11 The San Andreas Fault at Parkfield in central California. The person with the orange shirt is standing on the Pacific Plate and the person at the far side of the bridge is on the North American Plate. The bridge is designed to slide on its foundation. [SE]

As originally described by Wegener in 1915, the present continents were once all part of a supercontinent, which he termed **Pangea** (*all land*). More recent studies of continental matchups and the magnetic ages of ocean-floor rocks have enabled us to reconstruct the history of the break-up of Pangea.

Pangea began to rift apart along a line between Africa and Asia and between North America and South America at around 200 Ma. During the same period, the Atlantic Ocean began to open up between northern Africa and North America, and India broke away from Antarctica. Between 200 and 150 Ma, rifting started between South America and Africa and between North America and Europe, and India moved north toward Asia. By 80 Ma, Africa had separated from South America, most of Europe had separated from North America, and India had separated from

Antarctica. By 50 Ma, Australia had separated from Antarctic, and shortly after that, India collided with Asia. To see the timing of these processes for yourself go to: http://barabus.tru.ca/geol1031/plates.html.

Within the past few million years, rifting has taken place in the Gulf of Aden and the Red Sea, and also within the Gulf of California. Incipient rifting has begun along the Great Rift Valley of eastern Africa, extending from Ethiopia and Djibouti on the Gulf of Aden (Red Sea) all the way south to Malawi.

Over the next 50 million years, it is likely that there will be full development of the east African rift and creation of new ocean floor. Eventually Africa will split apart. There will also be continued northerly movement of Australia and Indonesia. The western part of California (including Los Angeles and part of San Francisco) will split away from the rest of North America, and eventually sail right by the west coast of Vancouver Island, en route to Alaska. Because the oceanic crust formed by spreading on the mid-Atlantic ridge is not currently being subduct

(except in the Caribbean), the Atlantic Ocean is slowly getting bigger, and the Pacific Ocean is getting smaller. If this continues without changing for another couple hundred million years, we will be back to where we started, with one supercontinent.

Pangea, which existed from about 350 to 200 Ma, was not the first supercontinent. It was preceded by Pannotia (600 to 540 Ma), by Rodinia (1,100 to 750 Ma), and by others before that.

In 1966, Tuzo Wilson proposed that there has been a continuous series of cycles of continental rifting and collision; that is, break-up of supercontinents, drifting, collision, and formation of other supercontinents. At present, North and South America, Europe, and Africa are moving with their respective portions of the Atlantic Ocean. The eastern margins of North and South America and the western margins of Europe and Africa are called **passive margins** because there is no subduction taking place along them.

This situation may not continue for too much longer, however. As the Atlantic Ocean floor gets weighed down around its margins by great thickness of continental sediments (i.e., geosynclines), it will be pushed farther and farther into the mantle, and eventually the oceanic lithosphere may break away from the continental lithosphere (Figure 11.12). A subduction zone will develop, and the oceanic plate will begin to descend under the continent. Once this happens, the continents will no longer continue to move apart because the spreading at the mid-Atlantic ridge will be taken up by subduction. If spreading along the mid-Atlantic ridge continues to be slower than spreading within the Pacific Ocean, the Atlantic Ocean will start to close up, and eventually (in a 100 million years or more) North and South America will collide with Europe and Africa.
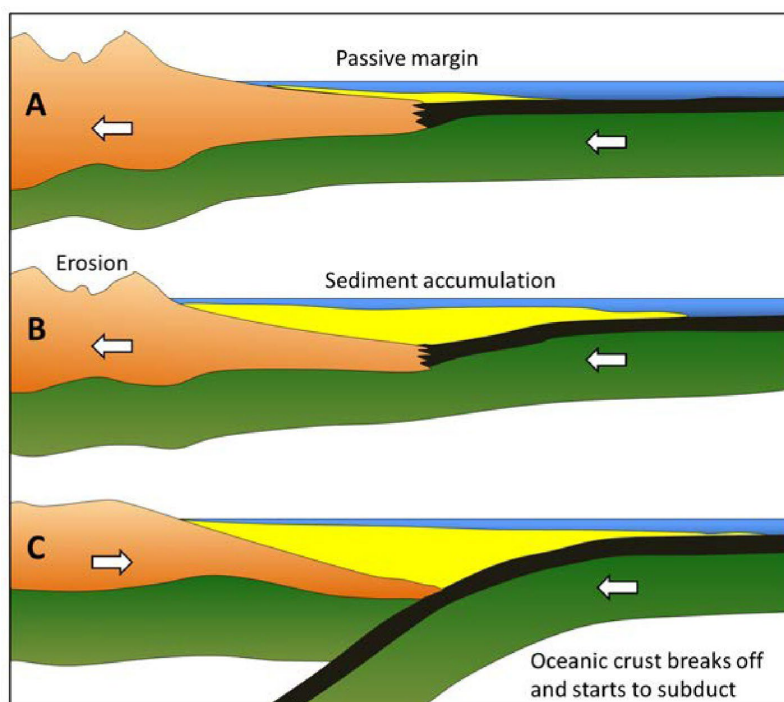
Figure 11.12 Development of a subduction zone at a passive margin. Times A, B, and C are separated by tens of millions of years. Once the oceanic crust breaks off and starts to subduct the continental crust (North America in this case) will no longer be pushed to the west and will likely start to move east because the rate of spreading in the Pacific basin is faster than that in the Atlantic basin. [SE]

There is strong evidence around the margins of the Atlantic Ocean that this process has taken place before. The roots of ancient mountain belts, which are present along the eastern margin of North America, the western margin of Europe, and the northwestern margin of Africa, show that these land masses once collided with each other to form a mountain chain, possibly as big as the Himalayas. The apparent line of collision runs between Norway

360

and Sweden, between Scotland and England, through Ireland, through Newfoundland, and the Maritimes, through the northeastern and eastern states, and across the northern end of Florida. When rifting of Pangea started at approximately 200 Ma, the fissuring was along a different line from the line of the earlier collision. This is why some of the mountain chains formed during the earlier collision can be traced from Europe to North America and from Europe to Africa.

That the Atlantic Ocean rift may have occurred in approximately the same place during two separate events several hundred million years apart is probably no coincidence. The series of hot spots that has been identified in the Atlantic Ocean may also have existed for several hundred million years, and thus may have contributed to rifting in roughly the same place on at least two separate occasions (Figure 11.13).
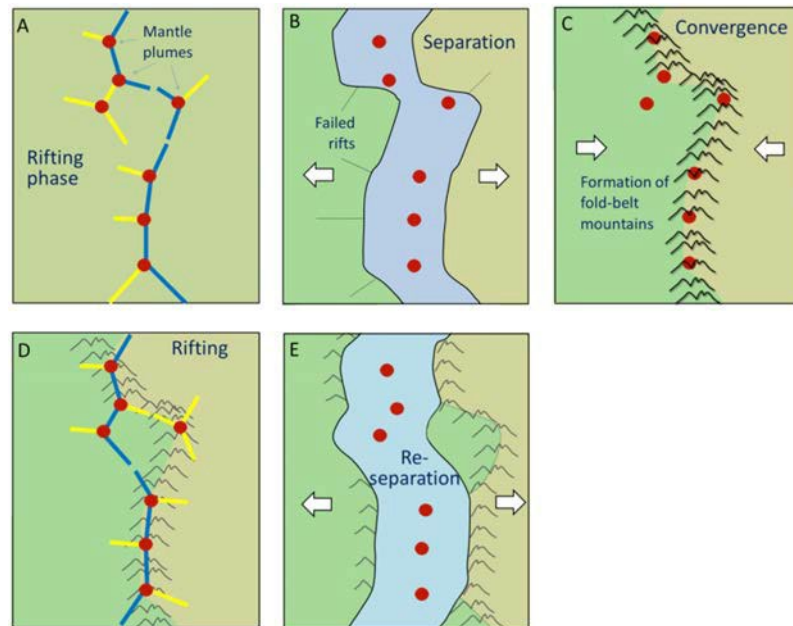
Figure 11.13 A scenario for the Wilson cycle. The cycle starts with continental rifting above a series of mantle plumes (A). The continents separate (B), and then re-converge some time later, forming a fold-belt mountain chain. Eventually rifting is repeated, possibly because of the same set of mantle plumes (D), but this time the rift is in a different place. [SE]

Isostasy

Theory holds that the mantle is able to convect because of its plasticity, and this property also allows for another very important Earth process known as **isostasy**. The literal meaning of the word isostasy is "equal standstill," but the importance behind it is the principle that Earth's crust is *floating* on the mantle, like a raft floating in the water, rather than *resting* on the mantle like a raft sitting on the ground.

The relationship between the crust and the mantle is illustrated in Figure 11.14. On the right is an example of a non-isostatic relationship between a raft and solid concrete. It's possible to load the raft up with lots of people, and it still won't sink into the concrete. On the left, the relationship is an isostatic one between two different rafts and a swimming pool full of peanut butter. With only one person on board, the raft floats high in the peanut butter, but with three people, it sinks dangerously low. We're using peanut butter here, rather than water, because its viscosity more closely represents the relationship between the crust and the mantle. Although it has about the same density as water, peanut butter is much more viscous (stiff), and so although the three-person raft will sink into the peanut butter, it will do so quite slowly.

361
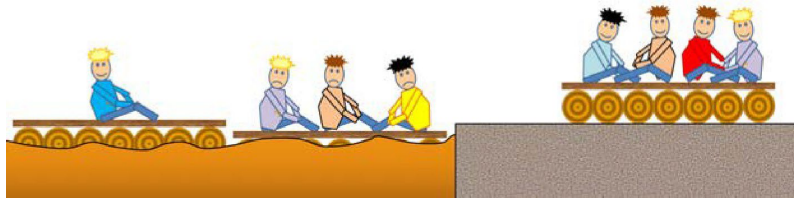
Figure 11.14 Illustration of a non-isostatic relationship between a raft and solid ground (right) and of isostatic relationships between rafts and peanut butter (left). [SE]

The relationship of Earth's crust to the mantle is similar to the relationship of the rafts to the peanut butter. The raft with one person on it floats comfortably high. Even with three people on it the raft is less dense than the peanut butter, so it floats, but it floats uncomfortably low for those three people. The crust, with an average density of around 2.6 grams per cubic centimetre (g/cm3), is less dense than the mantle (average density of approximately 3.4 g/cm3 near the surface, but more than that at depth), and so it is floating on the "plastic" mantle. When more weight is added to the crust, through the process of mountain building, it slowly sinks deeper into the mantle and the mantle material that was there is pushed aside (Figure 11.15, left). When that weight is removed by erosion over tens of millions of years, the crust rebounds and the mantle rock flows back (Figure 11.15, right).

Figure 11.15 Illustration of the isostatic relationship between the crust and the mantle. Following a period of mountain building, mass has been added to a part of the crust, and the thickened crust has pushed down into the mantle (left). Over the following tens of millions of years, the mountain chain is eroded and the crust rebounds (right). The green arrows represent slow mantle flow. [SE]

362

The crust and mantle respond in a similar way to glaciation. Thick accumulations of glacial ice add weight to the crust, and as the mantle beneath is squeezed to the sides, the crust subsides. This process is illustrated for the current ice sheet on Greenland in Figure 11.16. The Greenland Ice Sheet at this location is over 2,500 m thick, and the crust beneath the thickest part has been depressed to the point where it is below sea level over a wide area. When the ice eventually melts, the crust and mantle will slowly rebound, but full rebound will likely take more than 10,000 years.
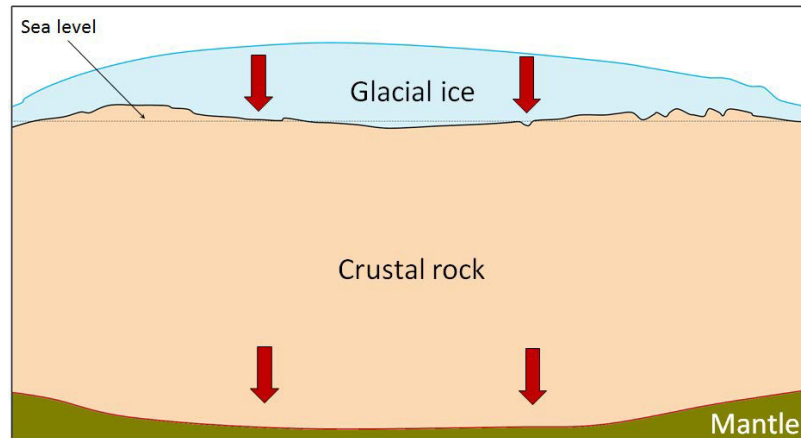
Figure 11.16a A cross-section through the crust in the northern part of Greenland (The ice thickness is based on data from NASA and the Center for Remote Sensing of Ice Sheets, but the crust thickness is less than it should be for the sake of illustration.) The maximum ice thickness is over 2,500 m. The red arrows represent downward pressure on the mantle because of the mass of the ice.
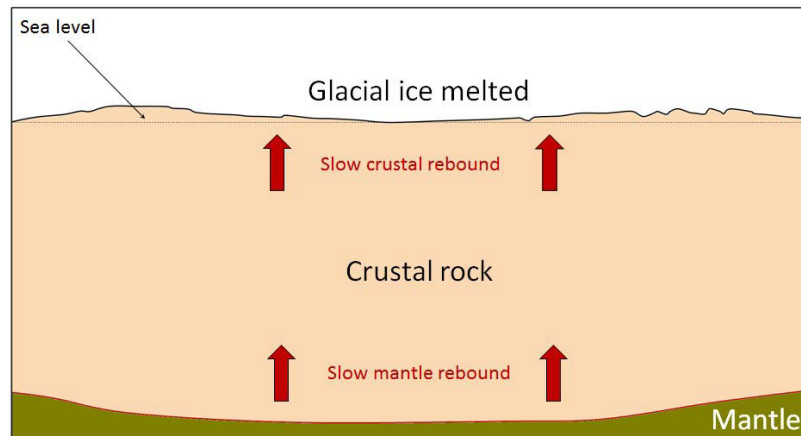
Figure 11.16b Depiction of the situation after complete melting of the ice sheet, a process that could happen within 2,000 years if people and their governments continue to ignore climate change. The isostatic rebound of the mantle would not be able to keep up with this rate of melting, so for several thousand years the central part of Greenland would remain close to sea level, in some areas even below sea level.
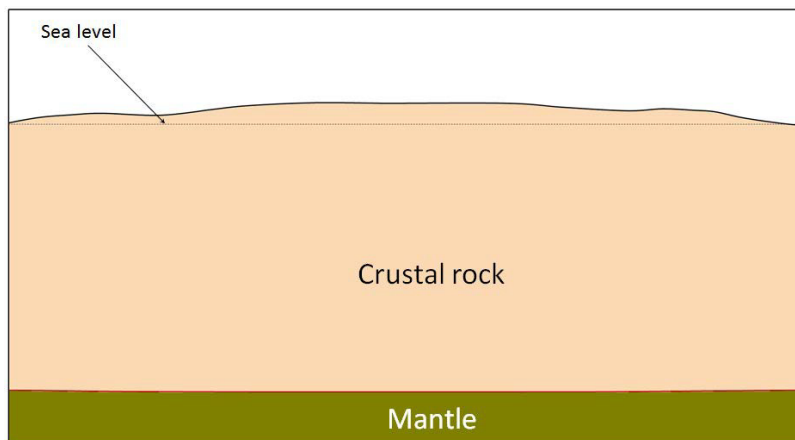
Figure 11.16c It is likely that complete rebound of the mantle beneath Greenland would take more than 10,000 years.

How can the mantle be both solid and plastic?

Figure 11.17. Picture of silly putty's non-Newtonian flow
[https://upload.wikimedia.org/wikipedia/commons/f/f3/Silly_putty_dripping.jpg]

You might be wondering how it is possible that Earth's mantle is rigid enough to break during an earthquake, and yet it convects and flows like a very viscous liquid. The explanation is that the mantle behaves as a non-Newtonian fluid, meaning that it responds differently to stresses depending on how quickly the stress is applied. A good example of this is the behavior of the material known as Silly Putty (Figure 11.17), which can bounce and will break if you pull on it sharply, but will deform in a liquid manner if stress is applied slowly. In this photo, Silly Putty was placed over a hole in a glass tabletop, and in response to gravity, it slowly flowed into the hole. The mantle will flow when placed under the slow but steady stress of a growing (or melting) ice sheet.

Large parts of Canada are still rebounding as a result of the loss of glacial ice over the past 12 ka, and as shown in Figure 11.18, other parts of the world are also experiencing isostatic rebound. The highest rate of uplift is in within a large area to the west of Hudson Bay, which is where the Laurentide Ice Sheet was the thickest (over 3,000 m). Ice finally left this region around 8,000 years ago, and the crust is currently rebounding at a rate of nearly 2 cm/year. Strong isostatic rebound is also occurring in northern Europe where the Fenno-Scandian Ice Sheet was thickest, and in the eastern part of Antarctica, which also experienced significant ice loss during the Holocene.

There are also extensive areas of subsidence surrounding the former Laurentide and Fenno-Scandian Ice

364
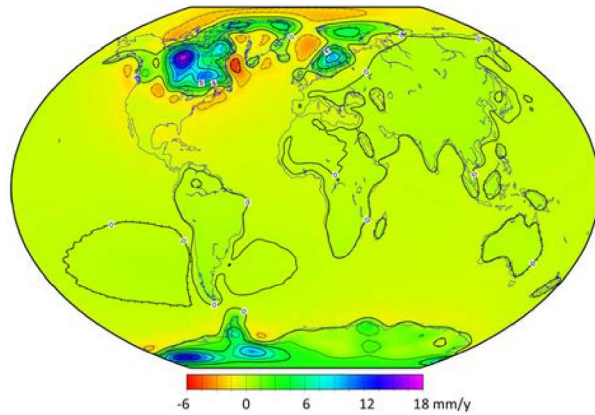
Figure 11.18 The current rates of post-glacial isostatic uplift (green, blue, and purple shades) and subsidence (yellow and orange). Subsidence is taking place where the mantle is slowly flowing back toward areas that are experiencing post-glacial uplift. [From: Paulson, A., S. Zhong, and

J. Wahr. Inference of mantle viscosity from GRACE and relative sea level data, Geophys. J. Int. (2007) 171, 497–508. Accessed at: http://en.wikipedia.org/wiki/Hudson_Bay#/media/ File:PGR_Paulson2007_Rate_of_Lithospheric_Uplift_due_to_PGR.png]

Sheets. During glaciation, mantle rock flowed away from the areas beneath the main ice sheets, and this material is now slowly flowing back, as illustrated in Figure 9.18b.

Physical Geology by Steven Earle used under a CC-BY 4.0 international license.

Chapter 12.

Earthquakes

What is an Earthquake?

Earthquake Shaking Comes from Elastic Deformation

Earthquakes occur when rock **ruptures** (breaks), causing rocks on one side of a fault to move relative to the rocks on the other side. Although motion along a fault is part of what happens when an earthquake occurs, rocks grinding past each other is not what creates the shaking. In fact, it could be said that the earthquake happens *after* rocks have undergone most of the displacement. Consider this: if rocks slide a few centimeters or even meters along a fault, would that motion *alone* explain the incredible damage caused by some earthquakes? If you were in a car that suddenly accelerated then stopped, you would feel a jolt. But earthquakes are not a single jolt. Buildings can swing back and forth until they shake themselves to pieces, train tracks can buckle and twist into s-shapes, and roads can roll up and down like waves on the ocean. During an earthquake, rock is not only slipping. It is also vibrating like a plucked guitar string.

Rocks might seem rigid, but when stress is applied they may stretch. If there hasn't been too much stretching, a rock will snap back to its original shape once the stress is removed. Deformation that is reversible is called **elastic deformation**. Rocks that are stressed beyond their ability to stretch can rupture, allowing the rest of the rock to snap back to its original shape. The snapping back of the rock returning to its original shape causes the rock to vibrate, and this is what causes the shaking during an earthquake. The snapping back is called **elastic rebound**.

Figure 12.1 (top) shows this sequence of events. Stress is applied to a rock and deforms it. The deformed rock ruptures, forming a fault

After rupturing, the rock above and below the fault snaps back to the shape it had before deformation.

Ruptures can also occur along pre-existing faults (Figure 12.1, bottom). The rocks on either side of the fault are locked together because bumps along the fault, called **asperities**, prevent the rocks from moving relative to each other. When the stress is great enough to break the asperities, the rocks on either side of the fault can slide again. While the rocks are locked together, stress can cause elastic deformation. When asperities break and release the stress, the rocks undergo elastic rebound and return to their original shape.

Rupture Surfaces Are Where the Action Happens

Images like 12.2 are useful for illustrating elastic deformation and rupture, but they can be misleading. The rupture that happens doesn't occur as in 12.2, with the block being ruptured through and through. The rupture and displacement only happen along a subsection of a fault, called the **rupture surface**. In Figure 12.3, the rupture surface is the dark pink patch. It takes up only a part of the **fault plane** (lighter pink). The fault plane represents the surface where the fault exists, and where ruptures have happened in the past. Although the fault plane is drawn as being flat in Figure 12.4, faults are not actually perfectly flat.

The location on the fault plane where the rupture happens is called the **hypocenter** or **focus** of the earthquake (Figure 12.4, right). The location on Earth's surface immediately above the hypocenter is the **epicenter** of the earthquake.

Shifting Stress Causes Foreshocks and Aftershocks

Earthquakes don't usually occur in isolation. There is often a sequence in which smaller earthquakes occur prior to a larger one, and then progressively smaller earthquakes occur after. The largest earthquake in the series is the **mainshock**. The smaller ones that come before are **foreshocks**, and the smaller ones that come after are **aftershocks**. These descriptions are relative, so it can be necessary to reclassify an earthquake. For example, the strongest earthquake in a series is classified as the mainshock, but if another even bigger one comes after it, the bigger one is called the mainshock, and the earlier one is reclassified as an aftershock.

A rupture surface does not fail all at once. A rupture in one place leads to another, which leads to another. Aftershocks and foreshocks represent the same thing, except on a much larger scale.

The rupture illustrated in Figure 9.4 reduced stress in one area, but in doing so, transferred stress to others (Figure 9.5). Imagine a frayed rope breaking strand by strand. When a strand breaks, the tension on that strand is released, but the remaining strands must still hold up the same amount of weight. If another strand breaks under the increased burden, the remaining strands have an even greater burden than before. In the same way that the stress causes one strand after another to fail, a rupture can trigger subsequent ruptures nearby.
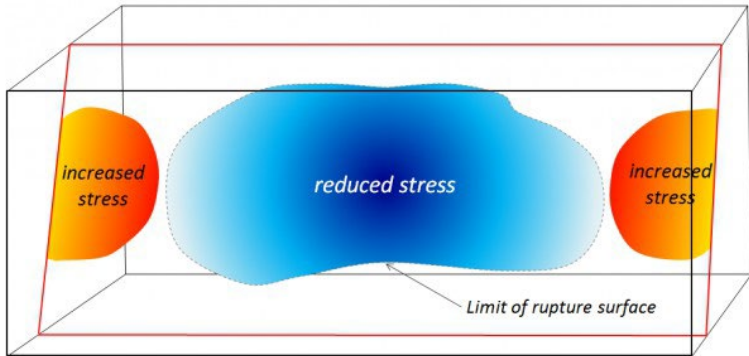
Figure 12.2 *Stress changes related to an earthquake. Stress decreases in the area of the rupture surface, but increases on adjacent parts of the fault.*

Source: Steven Earle (2015) CC BY 4.0 view source.

Numerous aftershocks were associated with the magnitude 7.8 earthquake that struck Haida Gwaii, B.C., in October of 2012 (Figure 12.3; mainshock in red, aftershocks in white). Some of the stress released by the mainshock was transferred to other nearby parts of the fault, and contributed to a cascade of smaller ruptures. But stress transfer need not be restricted to the fault along which an earthquake happened. It will affect the rocks in general around the site of the earthquake and may lead to increased stress on other faults in the region. The aftershocks from the Haida Gwaii earthquake are scattered rather than located only on the main faults.
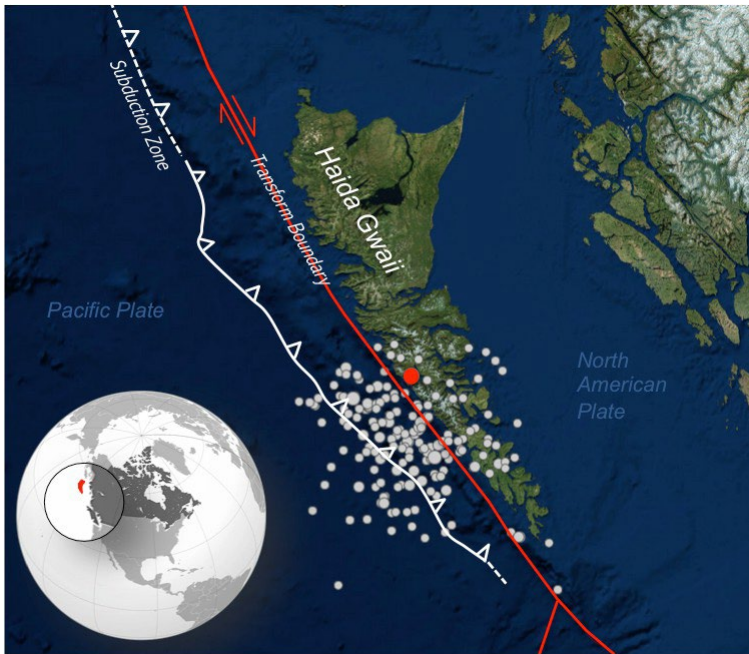
Figure 12.3 Magnitude 7.8 Haida Gwaii earthquake and aftershocks. Mainshock (red circle marks the epicenter) occurred on October 28th, 2012. Aftershocks are for the period from October 28th to November 10th of 2012. Although the epicenter is near a transform boundary, the rupture was influenced more by compression related to the subduction zone. Source: Karla Panchuk (2017) CC BY 4.0. Base map with epicenters from the U. S. Geological Survey Latest Earthquakes tool view interactive map. Subduction zone after Wang et al. (2015). Click the image for more attributions.

The effects of stress transfer may not show immediately. Aftershocks can be delayed for hours, days, weeks, or even years. Because stress transfer affects a region, not just a single fault, and because there can be delays between the event that transferred stress and the one that was triggered by the transfer, it can sometimes be hard to be know whether one earthquake is actually associated with another, and whether a **foreshock** or aftershock should be assigned to a particular mainshock.

Earthquakes at Divergent and Transform Plate Boundaries

Earthquakes along divergent and transform plate margins are shallow (usually less than 30 km deep) because below those depths, rock is too hot and weak to avoid being permanently deformed by the stresses in those settings. If deformation is permanent, then removing the stress does not result in the rocks snapping back to their original shape. No snapping back means no shaking.

Mid-ocean ridge divergent plate margins are offset by numerous transform faults (Figure 12.4). The locations of earthquakes along mid-ocean ridges, and the mechanisms for causing them, depend on how rapidly the mid-ocean ridges are spreading. The Pacific- Antarctic Ridge (left) is spreading relatively rapidly at 42 to 94 mm/ year, depending on the location along the ridge. Rapid spreading causes rocks near the axis of the spreading center to be hot and weak. As a result, most of the earthquakes (white dots) are located along transform faults, where rocks are cooler and stronger. Along rapidly spreading ridges, new ocean crust is bent upward into wide, high ridges. As spreading proceeds and crust moves away from the ridge, the bend is relaxed, and the crust stretches and breaks. This triggers earthquakes many kilometers away from the ridge.
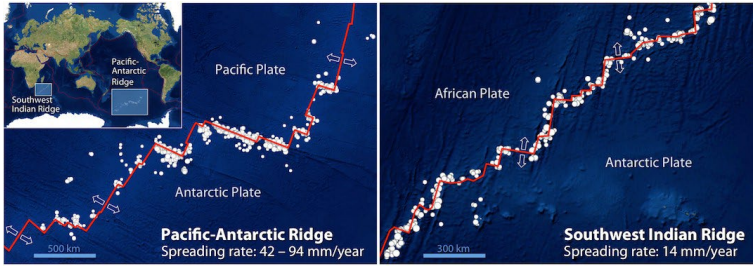
Figure 12.4 *Locations of earthquakes of magnitude 4 and greater from 1990 to 2010 along two mid-ocean ridges. Plate boundaries are marked in red. Arrows show the direction of plate motion. Left: Rapidly spreading Pacific-Antarctic ridge with earthquakes concentrated along transform faults. Right: Slowly spreading Southwest Indian Ridge, with earthquakes along both spreading segments and transform faults. Source: Karla Panchuk (2017) CC BY 4.0. Base maps with epicentres generated using the U. S. Geological Survey Latest Earthquakes website. Visit Latest Earthquakes*

The Southwest Indian Ridge (right) spreads very slowly, at approximately 14 mm/year. Rocks are cooler and stronger along the slowly spreading ridge than along the rapidly spreading one. In the slow-spreading environment, earthquakes are generated when rocks along the ridge axis stretch and break. Earthquakes are more evenly distributed between divergent and transform segments of the boundary than they are along fast-spreading ridges.

Earthquakes in continental rift zones are also shallow, but scattered more broadly than those along mid-ocean ridges. Lake Baikal (Figure 12.5), the world's oldest, deepest, and largest freshwater lake, formed 25 million years ago because of continental rifting. Note the scale in Figure 12.18, and compare how widely the earthquakes (blue dots) are spread in the Lake Baikal region, versus along the mid-ocean ridges in Figure 12.5.

374
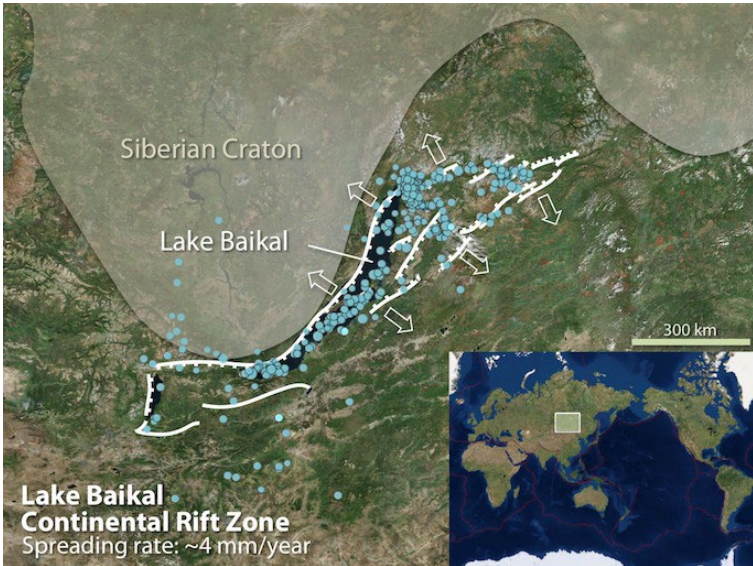
Figure 11.5 Small circles mark the locations of earthquakes of M4 and greater from 1990 to 2010 along the Lake Baikal rift zone. White lines show some of the faults in the region. White lines with tick marks are normal faults related to spreading. Arrows show the direction of spreading. White lines without tick marks are transform faults. The Siberian Craton (shaded region) is strong 2 billion year old crust. Source: Karla Panchuk (2017) CC BY 4.0. Faults after U.

S. Geological Survey (see references). Base maps (inset and rift views) with epicenters generated using the U. S. Geological Survey Latest Earthquakes website. Visit Latest Earthquakes

One reason for the difference in earthquake distribution in continental rift zones is that the rifts are only beginning to form. Faulting is "disorganized" within the continental crust. There is no well-established spreading center, unlike mid-ocean ridges. Another reason is that the locations of faults, and thus earthquakes, in continental rift zones are affected by pre-existing geological structures within continental crust. In the case of the Lake Baikal rift, the strong, ancient crust of the Siberian Craton influences the orientation of the faults forming the rift. Faults run parallel to the **craton** near Lake Baikal. As rifting extends to the east, the part

375

of the craton in the upper right of Figure 12.18 may deflect rifting southward.

Measuring Earthquakes

The shaking from an earthquake travels away from the rupture in the form of **seismic waves**. Seismic waves are measured to determine the location of the earthquake, and to estimate the amount of energy released by the earthquake (its **magnitude**).

Types of Seismic Waves

Seismic waves are classified according to where they travel, and how they move particles.

Body Waves

Seismic waves that travel through Earth's interior are called **body waves**. **P-waves** are body waves that move by alternately compressing and stretching materials in the direction the wave moves. For this reason, P-waves are also called compression waves. The "P" in P-wave stands for primary, because P-waves are the fastest of the seismic waves. They are the first to be detected when an earthquake happens.

   A P-wave can be simulated by fixing one end of a spring to a solid surface, then giving the other end a sharp push toward the surface (Figure 12.6, top). The compression will propagate (travel) along the length of the spring. Some parts of the spring will be stretched, and others compressed. Any one point on the spring will jiggle forward and backward as the compression travels along the spring.
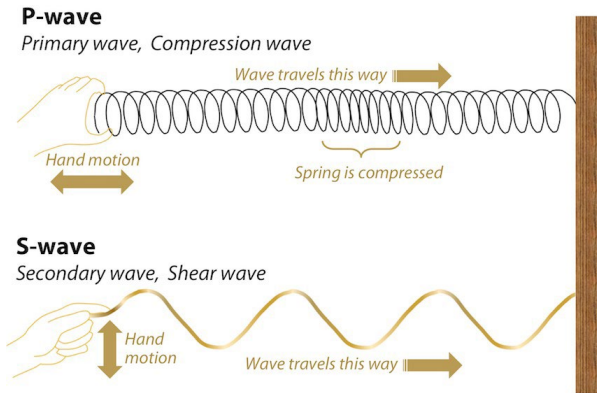
P-wave
Primary wave, Compression wave

Wave travels this way

Hand motion

Spring is compressed

S-wave
Secondary wave, Shear wave

Hand motion

Wave travels this way

S-waves are body waves that move with a shearing motion, shaking particles from side to side. S-waves can be simulated by fixing one end of a rope to a solid surface, then giving the other end a flick (Figure 12.6, bottom). Any one point on the rope will move from side to side at a right angle to the direction in which the snaking motion is traveling. The "S" in S-wave stands for secondary, because S-waves are slower than P-waves, and are detected after the P-waves are measured. S-waves cannot travel through liquids.

P-waves and S-waves can travel rapidly through geological materials, at speeds many times the speed of sound in air.

Surface Waves

When body waves reach Earth's surface, some of their energy is transformed into surface waves, which travel along Earth's surface. Two types of surface waves are **Rayleigh waves** and **Love waves** (Figure 12.7). Rayleigh waves (R-waves) are characterized by vertical motion of the ground surface, like waves rolling on water. Love waves (L-waves) are characterized by side-to-side motion. Notice that the effects of both kinds of surface waves diminish with depth in Figure 12.7.

Surface waves are slower than body waves, and are detected after the body waves. Surface waves typically cause more ground motion than body waves, and therefore do more damage than body waves.
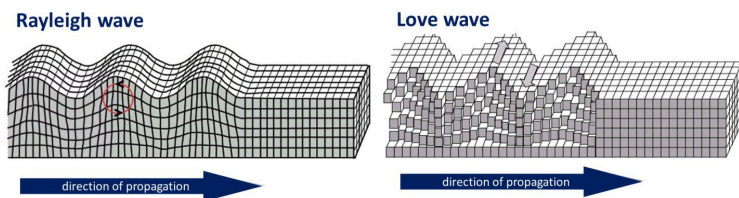
Figure 12.7 Surface waves travel along Earth's surface and have a diminished impact with depth. Rayleigh waves (left) cause a rolling motion, and Love waves (right) cause the ground to shift from side to side. Source: Steven Earle (2015) CC BY 4.0 view source. Click the image for more attributions.

Recording Seismic Waves Using a Seismograph

A **seismometer** is an instrument that detects seismic waves. An instrument that combines a seismometer with a device for recording the waves is called a s**eismograph**. The graphical output from a seismograph is called a **seismogram**. Figure 12.8 (right) shows how a seismograph works. The instrument consists of a frame or

housing that is firmly anchored to the ground. A mass is suspended from the housing, and can move freely on a spring. When the ground shakes, the housing shakes with it, but the mass remains fixed. A pen attached to the mass moves up and down on a rotating drum of paper, drawing a wavy line, the seismogram. The seismograph in Figure 12.8 (right) is oriented to measure vertical ground motion. The photo on the left shows a seismograph oriented to record horizontal ground motion.
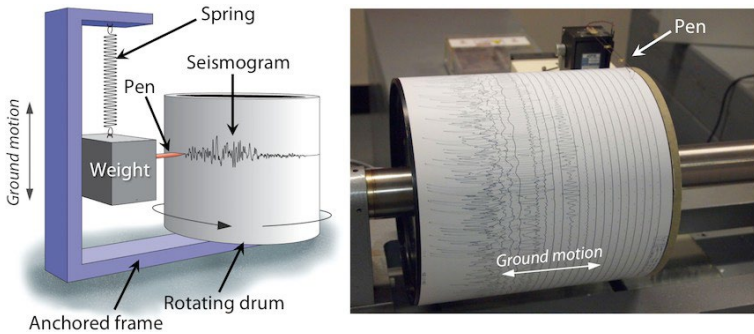
Figure 12.8 How a seismograph records earthquakes. Source: Left- Karla Panchuk (2018) CC BY-NC-SA 4.0 modified after IRIS (2012) "How Does a Seismometer Work?" view source; Right: Karla Panchuk (2018) CC BY-SA 4.0, photo by Z22 (2014) CC BY-SA 3.0 view source. Click the image for more attributions.

The pen and drum of a mechanical seismograph record the motion of the ground relative to the mass. However, unless an earthquake causes a large amount of ground motion directly beneath the seismograph, the height of the wave recorded on paper might be very small, making the seismogram difficult to analyze. The seismograph on the right has a device to amplify the ground motion, drawing larger waves that are easier to study.

Modern seismographs record shaking as electrical signals, and are able to transmit those signals. This means seismologists need not return to the instrument to collect recordings before the records can be examined.

Finding The Location of an Earthquake

P-waves travel faster than S-waves. As the waves travel away from the location of an earthquake, the P-wave gets farther and father ahead of the S-wave. Therefore, the farther a seismograph is from the location of an earthquake, the longer the delay between when the P-wave arrival is recorded, and the S-wave arrival is recorded. The delay between the P-wave and S-wave arrival appears as a widening gap in a diagram of P-wave and S-wave travel times (Figure 12.9, grey lines).

P-wave and S-wave arrival times can be identified on seismograms. In the three seismograms in Figure 12.9, the arrivals of the P-waves and S-waves are marked with arrows, and the interval in minutes between the P-wave and S-wave arrivals are noted. The seismograms were recorded at three different **seismic stations** (earthquake monitoring locations equipped with seismographs). The distance of each station from the earthquake is determined by finding the distance along the graph where the gap between the P-wave and S-wave travel-time curves matches the delay between P-wave and S-wave arrivals on the seismogram.
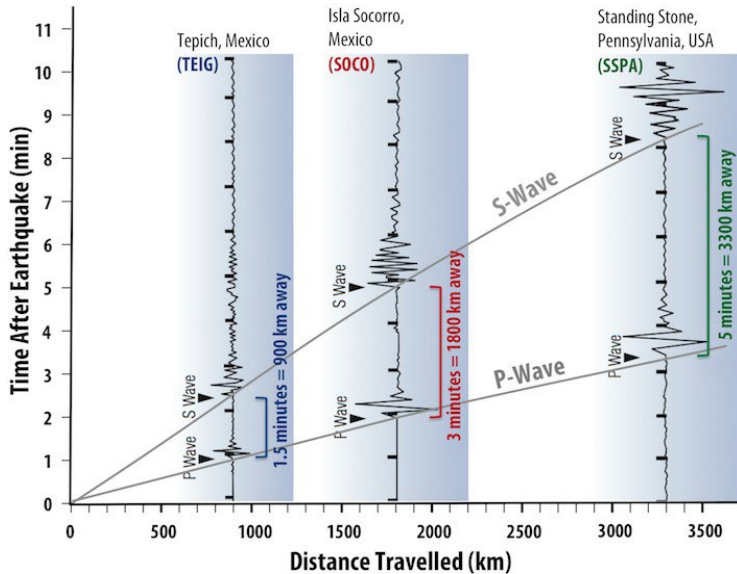
Figure 12.9 *Using P-wave and S-wave travel times to determine how far seismic waves have traveled. Grey curves show the distance traveled by*

P-waves and S-waves after an earthquake occurs. P-waves are faster than S-waves, and the gap between them increases with time and distance. The

delays between P-wave and S-wave arrivals on seismograms are matched to the curve to find the distances of seismic stations from the source of the seismic waves. Source: Karla Panchuk (2018) CC BY-NC-SA 4.0 modified after IRIS (n.d.) "How Are Earthquakes Located?" view source

The delay between the P-wave and S-wave arrival at a seismic station can indicate how far the station is from the source of the earthquake, but not the direction from which the seismic waves travelled. The possible locations of the earthquake can be represented on a map by drawing a circle around the seismic station, with the radius of the circle being the distance determined from the P-wave and S-wave travel times (Figure 12.10). If this is done for at least three seismic stations, the circles will intersect at the origin of the earthquake.
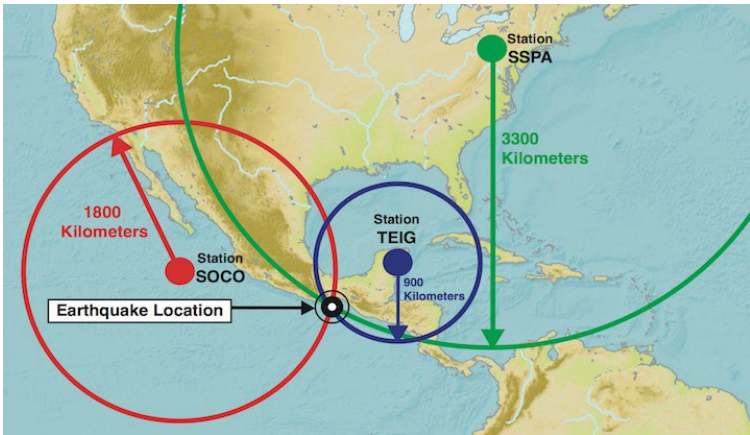
Figure 12.10 Locating earthquakes by drawing three circles with radii of lengths determined from P-wave and S-wave travel times. Where the radii intersect from all three circles is the epicenter of the earthquake. Station names (SOCO, TEIG, SSPA) correspond to seismograms in Figure 12.11. Source: IRIS (n.d.) "How Are Earthquakes Located?" view source Click the image for terms of use.

How Big Was It?

Earthquakes can be described in terms of their **magnitude**, which reflects the amount of energy released by the shaking. They can also be described in terms of **intensity**, which characterizes the impact of the shaking on people and their surroundings.

Earthquake Magnitude

Earthquake magnitudes are determined by measuring the amplitudes of seismic waves. The **amplitude** is the height of the wave relative to the baseline (Figure 12.11). Wave amplitude depends on the amount of energy carried by the wave. The amplitudes of

seismic waves reflect the amount of energy released by earthquakes.
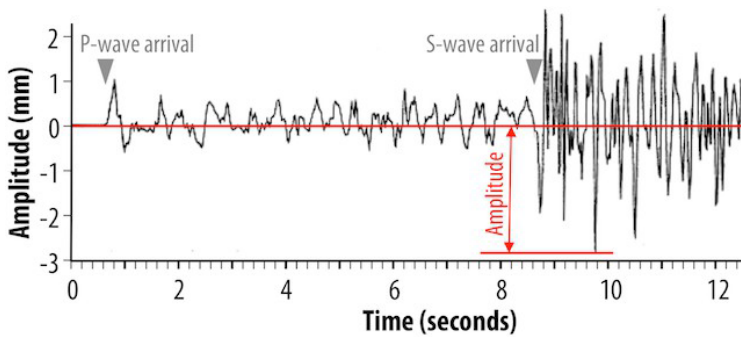
Figure 12.11 Seismogram for a small earthquake that occurred near Vancouver Island in 1997. The maximum amplitude of the S-wave is indicated. Source: Karla Panchuk (2018) CC BY 4.0 modified after Steven Earle (2015) CC BY 4.0 view source

The **Richter magnitude** scale uses the amplitudes of S-waves, and corrects for the decrease in amplitude that happens as the waves travel away from their source. The correction depends on how seismic waves interact with the specific rock types through which they travel, and therefore on local conditions, so the Richter magnitude is also referred to as the **local magnitude**.

While news reports about earthquakes might still refer to the "Richter scale" when describing magnitudes, the number they report is most likely the **moment magnitude**. The moment magnitude is calculated from the **seismic moment** of an earthquake. The seismic moment takes into account the surface area of the region that ruptured, how much displacement occurred, and the stiffness of the rocks. Moment magnitude can capture the difference between short earthquakes and longer ones resulting from larger ruptures, even of both types of earthquakes generate the same amplitude of waves. The moment magnitude scale is also better for earthquakes that are far from the seismic station. Seismic

wave measurements are still used to determine the moment magnitude, however different waves are used than for the local magnitude scale.

The magnitude scale is a logarithmic one rather than a linear one- an increase of one unit of magnitude corresponds to a 32 times increase in energy release (Figure 12.12). There are far more low-magnitude earthquakes than high-magnitude earthquakes. In 2017 there were 7 earthquakes of M7 (magnitude 7) or greater, but millions of tiny earthquakes.
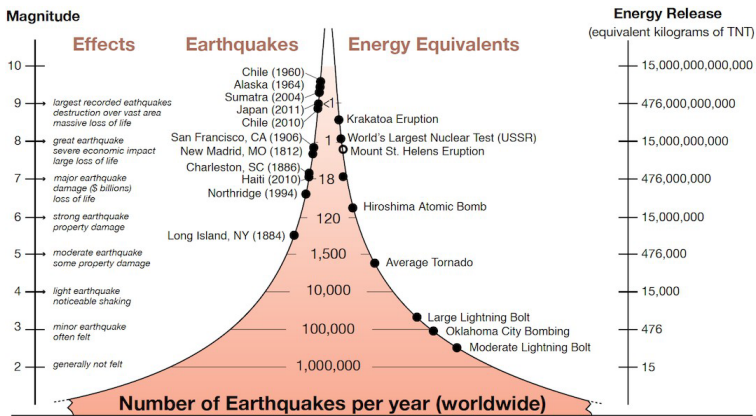
Figure **12.12** *Earthquake magnitude and corresponding energy release. Energy release increases by approximately 32 times for each unit change in magnitude. Source: IRIS (n.d.) "How Often Do Earthquakes Occur?" view source*

Earthquake Intensity

Intensity scales were first used in the late 19th century, and then adapted in the early 20th century by Giuseppe Mercalli and modified later by others to form what we now call the **Modified Mercalli Intensity Scale** (Table 12.1). To determine the intensity of an earthquake, reports are collected about what people felt and

how much damage was done. The reports are then used to assign intensity ratings to regions where the earthquake was felt.

Table 12.1 Modified Mercalli Intensity Scale	
I Not felt	Not felt except by a very few under especially favorable conditions
II Weak	Felt only by a few persons at rest, especially on upper floors of buildings
III Weak	Felt quite noticeably by persons indoors, especially on upper earthquake; standing motor cars may rock slightly; vibrations similar to the passing of a truck; duration estimated
IV Light	Felt indoors by many, outdoors by few during the day; at night, some awakened; dishes, windows, doors disturbed; walls make cracking sound; sensation like heavy truck striking building; standing motor cars rocked noticeably
V Moderate	Felt by nearly everyone; many awakened; some dishes, windows broken; unstable objects overturned; pendulum clocks may stop
VI Strong	Felt by all, many frightened; some heavy furniture moved; a few instances of fallen plaster; damage slight
VII Very Strong	Damage negligible in buildings of good design and construction; slight to moderate in well-built ordinary structures; considerable damage in poorly built or badly designed structures; some chimneys broken
VIII Severe	Damage slight in specially designed structures; considerable damage in ordinary substantial buildings with partial collapse; damage great in poorly built structures; fall of chimneys, factory stacks, columns, monuments, walls; heavy furniture overturned
IX Violent	Damage considerable in specially designed structures; well-designed frame structures thrown out of plumb; damage great in substantial buildings, with partial collapse; buildings shifted off foundations
X Extreme	Some well-built wooden structures destroyed; most masonry and frame structures destroyed with foundations; rails bent
XI Extreme	Few, if any (masonry), structures remain standing; bridges destroyed; broad fissures in ground; underground pipelines completely out of service; earth slumps and land slips in soft ground; rails bent greatly
XII Extreme	Damage total; waves seen on ground surfaces; lines of sight and level distorted; objects thrown upward into the air

Source: U. S. Geological Survey (1989). *The Severity of an Earthquake*. USGS General Interest Publication 1989-288-913 view source

Intensity values are assigned to locations, rather than to the earthquake itself. This means that intensity can vary for a given earthquake, depending on the proximity to the epicenter and local conditions. For the 1946 M7.3 Vancouver Island earthquake, intensity was greatest in the central island region (Figure 12.13). In some communities within this region, chimneys were damaged on more than 75% of buildings. Some roads were made impassable, and a major rock slide occurred. The earthquake was felt as far north as Prince Rupert, as far south as Portland Oregon, and as far east as the Rockies, but with less intensity.
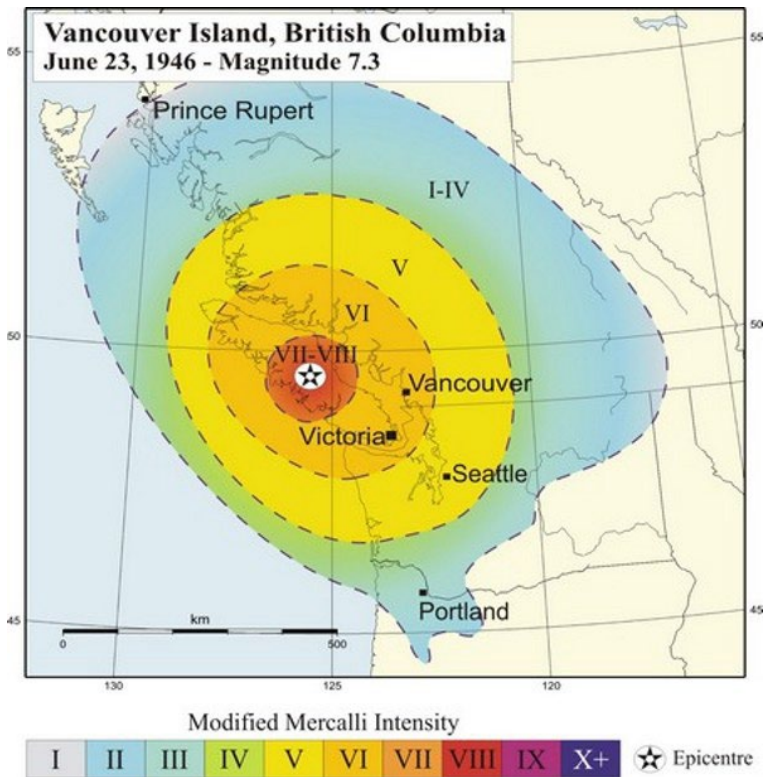
Figure 12.13 Intensity map for the M7.3 Vancouver Island earthquake on June 23, 1946. Source: Earthquakes Canada, Natural Resources Canada (2016) view source. Click the image for terms of use.

Intensity estimates are important as a way to identify regions that are especially prone to strong shaking. A key factor is the nature of the underlying geological materials. The weaker the underlying materials, the more likely it is that there will be strong shaking. Areas underlain by strong solid bedrock tend to experience far less shaking than those underlain by unconsolidated river or lake sediments.

An example of this effect is the 1985 M8 earthquake that struck the Michoacán region of western Mexico, southwest of Mexico City. There was relatively little damage near the epicenter, but 350 km away in heavily populated Mexico City there was tremendous damage and approximately 5,000 deaths. The reason is that Mexico City was built largely on the unconsolidated and water-saturated sediment of former Lake Texcoco. These sediments resonate at a frequency of about two seconds, which was similar to the frequency of the body waves that reached the city. Consequently, the shaking was amplified. Survivors of the disaster recounted that the ground in some areas moved up and down by approximately 20 cm every two seconds for over two minutes. Damage was greatest to buildings between 5 and 15 stories tall, because they also resonated at around two seconds, which amplified the shaking.

The Impacts of Earthquakes

Earthquakes can have direct impacts, such as structural damage to buildings from shaking, and secondary impacts, such as triggering landslides, fires, and tsunami. The types and extent of impacts will depend on local conditions where the earthquake strikes. The geological materials in the area matter, as does the type of terrain, and whether the region is near the coast or not. The extent of impact and type of damage will depend on whether the area is predominantly urban or rural, densely or sparsely populated, highly

developed or underdeveloped. It will depend on whether the infrastructure has been designed to withstand shaking.

Damage to Structures from Shaking

As with the example of the 1985 Mexico earthquake, the geological foundations on which structures are built will affect the amount of shaking that occurs. Earthquakes produce seismic waves that vibrate at different rates, or **frequencies**. Waves with rapid vibrations have a high frequency, and waves with slower vibrations have lower frequencies.

   The energy of the higher frequency waves tends to be absorbed by solid rock. Lower frequency waves pass through solid rock without being absorbed, but are absorbed and amplified by soft sediments. It is therefore very common to see much worse earthquake damage in areas underlain by soft sediments than in areas of solid rock. During the 1989 Loma Prieta earthquake, parts of a two-layer highway in the Oakland area near San Francisco collapsed where they were built on soft sediments (Figure 12.14).

Figure 12.14 *A collapsed section of the Cypress Freeway in Oakland California.*
Source: U. S. Geological Survey (1989) Public Domain view source

Building damage is also greatest in areas of soft sediments, and multi-story buildings tend to be more seriously damaged than smaller ones. Buildings can be designed to withstand most earthquakes, and this practice is increasingly applied in earthquake- prone regions. Turkey is one such region, but even though Turkey had a relatively strong building code in the 1990s, adherence to the code was poor. Builders did whatever they could to save costs, including using inappropriate materials in concrete, and reducing the amount of steel reinforcing. The result was more than 17,000 deaths in the 1999 M7.6 Izmit earthquake (Figure 12.15). After two devastating earthquakes that year, Turkish authorities strengthened the building code further, but the new code has been applied only in a few regions, and enforcement of the code is still weak, as revealed by the amount of damage from a M7.1 earthquake in eastern Turkey in 2011.

Figure 12.15 Damage from the 1999 M7.6 Izmit, Turkey earthquake. Source: Left; USGS (1999) Public Domain view source; Right: USGS (1999) Public Domain view source

Structures underlain by sediments may be at risk of another hazard, called **liquefaction**, in which sediment is transformed into a fluid. When water-saturated sediments are shaken, the grains may lose contact with each other, and no longer support one another. Water between the grains holds them apart, causing the sediment to turn to mud and flow. The loss of support can lead to the collapse of buildings or other structures that might otherwise have sustained little damage. During the 1964 M7.6 earthquake in Niigata, Japan, liquefaction caused buildings to sink into the sediments (Figure 12.16).

Figure 12.16 Collapsed apartment buildings in the Niigata area of Japan. The material beneath the buildings was liquefied by the 1964 earthquake. Source: DOC/NOAA/NESDIS/NCEI (1964) Public Domain view source

Parts of the Fraser River delta are also prone to liquefaction-related damage. The region is characterized by a 2 m to 3 m thick layer of fluvial silt and clay above a layer of water-saturated fluvial sand that is at least 10 m thick (Figure 12.17). Under these conditions, seismic shaking can be amplified, and the sandy sediments will liquefy. This could lead to subsidence and tilting of buildings. Liquefaction can also contribute to slope failures and to fountains of sandy mud (sand volcanoes) in areas where there is loose saturated sand beneath a layer of more cohesive clay. Current building-code regulations in the Fraser delta area require that measures be taken to strengthen the ground beneath multi-story buildings prior to construction.
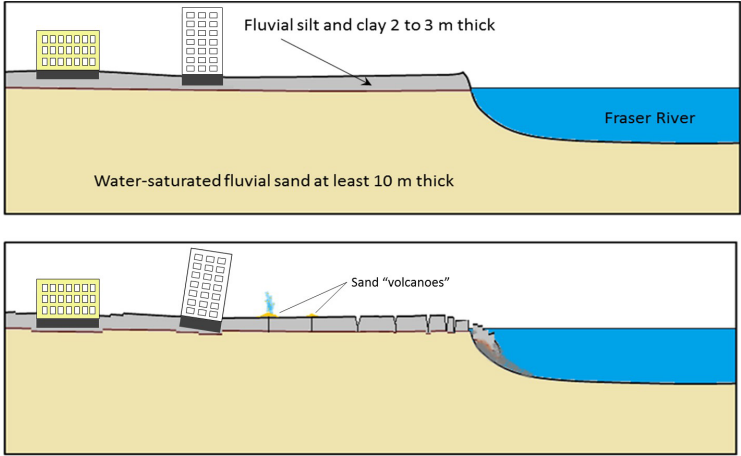
Figure 12.17 Recent unconsolidated sedimentary layers in the Fraser River delta area (top) and the potential consequences in the event of a damaging earthquake. Source: Steven Earle (2015) CC BY 4.0 view source

Slope Failure

Ground shaking during an earthquake can be enough to weaken rock and loose materials to the point of failure. Earthquakes can also trigger failures on slopes that are already weak. In January of 2011 a M7.6 earthquake offshore of El Salvador triggered slope failures that killed nearly 600 people (Figure 12.18).

Figure 12.18 *A slope gives way in a suburb of San Salvador after the January 2001 earthquake offshore of El Salvador. This is one of hundreds of slope failures resulting from the earthquake. Source: U. S. Geological Survey (2001) Public Domain view source*

Fires

Fires are commonly associated with earthquakes because gas lines rupture and electrical lines are damaged when the ground shakes. Most of the damage in the great 1906 San Francisco earthquake was caused by massive fires in the downtown area of the city (Figure 12.19). Some 25,000 buildings were destroyed by those fires, which were fueled by gas leaking from broken pipes. Fighting the fires was difficult because water mains had also ruptured. Today the risk of fires can be reduced through P-wave early warning systems if utility operators can decrease pipeline pressure and break electrical circuits.

Figure 12.19
Fires in San Francisco following the 1906
earthquake. Source: Pillsbury Picture Co. (1906) Public Domain, courtesy of the Library of Congress Prints and Photographs Division view source

Tsunami

Large earthquakes that take place beneath the ocean have the otential to displace large volumes of water. In a subduction zone,

for example, the overriding plate becomes distorted by elastic deformation. It is squeezed laterally and pushed up (Figure 12.20 top). When an earthquake happens, the plate rebounds over an area of thousands of square kilometers, generating waves- a **tsunami** (Figure 12.20, middle). The waves spread across the ocean at velocities of several hundred kilometers per hour. Tsunami can make it to the far side of an ocean in about the same time as a passenger jet.
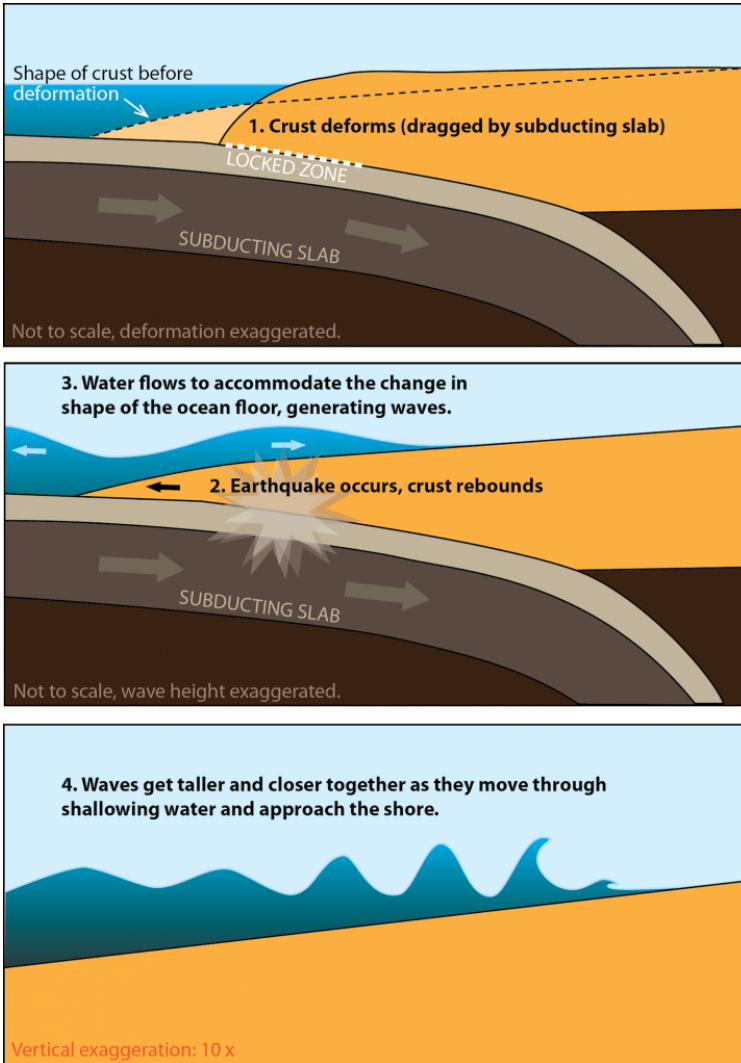
Figure **12.20** *Tsunami triggered along a subduction zone. Top: The overriding crust is deformed because it is locked to the subducting slab. Middle: When the locked zone ruptures, the crust rebounds, and waves are created. Bottom: Tsunami waves have small amplitudes in the deep ocean, but once in shallow water, they slow down, causing the waves to become taller and closer together. Source: Karla Panchuk (2018) CC BY-NC-SA 4.0. Top and middle modified after Steven Earle (2015) CC BY 4.0 view source. Bottom modified after COMET/UCAR (1997-2017) view source. Click the image for COMET/UCAR attribution and terms of use.*

Tsunami waves gain their height as they travel through shallower waters. In the deep ocean, the waves may be so small as to go undetected by ships, but when they are slowed down by interacting with the ocean floor, they can become much taller (Figure 12.20, bottom). In the tsunami following the 2004 Sumatra earthquake, the tallest waves were more than 30 m high.

Subduction earthquakes must be large to generate significant tsunami. Earthquakes with magnitude less than 7 do not typically generate significant tsunami because the amount of vertical displacement of the sea floor is minimal. Sea-floor transform earthquakes, even large ones (M7 to M8), don't typically generate tsunami either, because the motion is mostly side to side, not vertical.

Tsunami can have an impact across an entire ocean basin. They spread across the ocean at velocities of several hundred kilometers per hour, and can make it to the far side of an ocean in about the same time as a passenger jet. In 1700 a rupture along the Cascadia thrust fault running from Vancouver Island to northern California resulted in a M9 earthquake. It generated a tsunami that traveled across the Pacific Ocean, and was recorded in Japan nine hours later. A computer simulation of the tsunami (Figure 12.21) shows how long it took tsunami waves from the Cascadia earthquake to travel across the Pacific Ocean, and how high the waves were. Notice that over all, the waves decrease in height moving away from the rupture, but they increase in height again as they reach the opposite side of the Pacific Ocean.
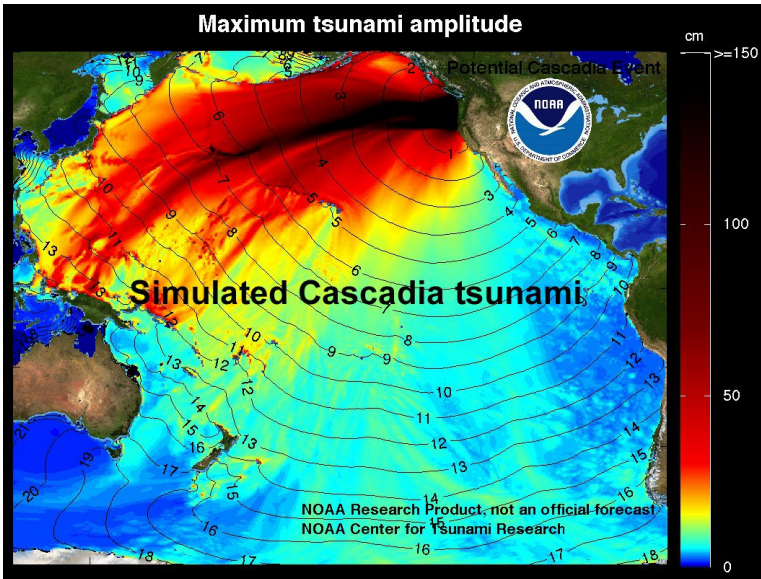
Figure 12.21 Computer simulation of the tsunami from the 1700 M9 Cascadia earthquake. Colours show open-ocean wave heights. Contours show travel time in hours. Tsunami wave heights increase as the tsunami reached the western margin of the Pacific ocean. Source: NOAA/PMEL/Center for Tsunami Research (2011) Public Domain view source / view context

References

Earthquakes Canada, Natural Resources Canada (2016). The M9 Cascadia Megathrust Earthquake of January 26, 1700. Visit website

University Corporation for Atmospheric Research (2010). Propagation. *Tsunamis.* Visit website

U. S. Geological Survey (2014). Indian Ocean Tsunami   Remembered — Scientists reflect on the 2004 Indian Ocean that killed thousands. Visit website

398

Summary

The topics covered in this chapter can be summarized as follows:

What Is an Earthquake?

An earthquake is the shaking that results when a deformed body of rock snaps back to its original shape. The rupture is initiated at a point but quickly spreads across the area of a fault, with aftershocks initiated by stress transfer. Episodic tremor and slip is a periodic slow movement, accompanied by harmonic tremors, along the middle part of a subduction zone boundary.

Measuring Earthquakes

Earthquakes produce seismic waves that can be measured by a seismograph. The amplitudes of seismic waves are used to determine the amount of energy released by an earthquake- its magnitude. For the moment magnitude scale used today, the amount of energy released by an earthquake is proportional to the size of the rupture surface, the amount of displacement, and the strength of the rock. Intensity is a measure of the amount of shaking that occurs, and damage done at locations that experience an earthquake. Intensity will vary depending on the distance to the epicenter, the depth of the earthquake, and the type of geological materials present.

Earthquakes and Plate Tectonics

Most earthquakes happen at or near plate boundaries. Along divergent and transform boundaries earthquakes are shallow (less than 30 km depth), but at convergent boundaries they can be hundreds of kilometers beneath the surface. The largest earthquakes happen when a broad segment of the locked zone of a subduction zone ruptures. Intraplate earthquakes happen away from plate boundaries. They can be caused by human activities, or renewed motion on ancient faults.

The Impacts of Earthquakes

Damage to buildings is the most serious consequence of most large earthquakes. The amount of damage is related to the type and size of buildings, how they are constructed, and the nature of the material on which they are built. Other important consequences are fires, damage to bridges and highways, slope failures, liquefaction, and tsunami.

Review Questions

1. What causes the shaking during an earthquake?
2. What is a rupture surface, and how does the area of a rupture surface relate to earthquake magnitude?
3. What is an aftershock and how are aftershocks related to stress transfer?
4. Episodic slip on the middle part of the Cascadia subduction zone is thought to increase the stress on the locked zone. Why?
5. What is the difference between the magnitude of an earthquake and its intensity?
6. How much more energy is released by a magnitude 7 earthquake compared to a magnitude 5 earthquake?
7. The images below show earthquake locations for three regions of ocean lithosphere. The colour scheme indicates the depths of earthquakes. a) The images show a subduction zone, a slowly spreading mid-ocean ridge, and a rapidly spreading mid-ocean ridge. Which is which? b) In the image with the subduction zone, which side is the subducting plate, and which is the overriding plate?
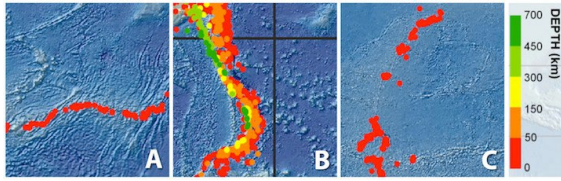
Figure 12.22 *Earthquake s along plate margins.*

Dots are color coded according to depth.

Source: Map details from Lisa Christiansen, Caltech Tectonics Observatory (2008) view source.

Click the image for terms of use.

8. Why is earthquake damage likely to be more severe for buildings built on unconsolidated sediments as opposed to on solid rock?
9. Why are fires common during earthquakes?
10. What type of earthquake is likely to lead to a tsunami?

Chapter 13
VOLCANOES

Magma Generation

Magma and **lava** contain three components – melt, solids, and volatiles (dissolved gases). The liquid part, called **melt**, is made of ions from minerals that have already melted. The solid part, called **solids**, are crystals of minerals that have not melted (higher melting temperature) and are floating in the melt. Volatiles are gaseous components dissolved in the magma, such as water vapor, carbon dioxide, sulfur, and chlorine. The presence and amount of these three components affect the physical behavior of the magma. (4 Igneous Processes and Volcanoes – An Introduction to Geology, n.d.).
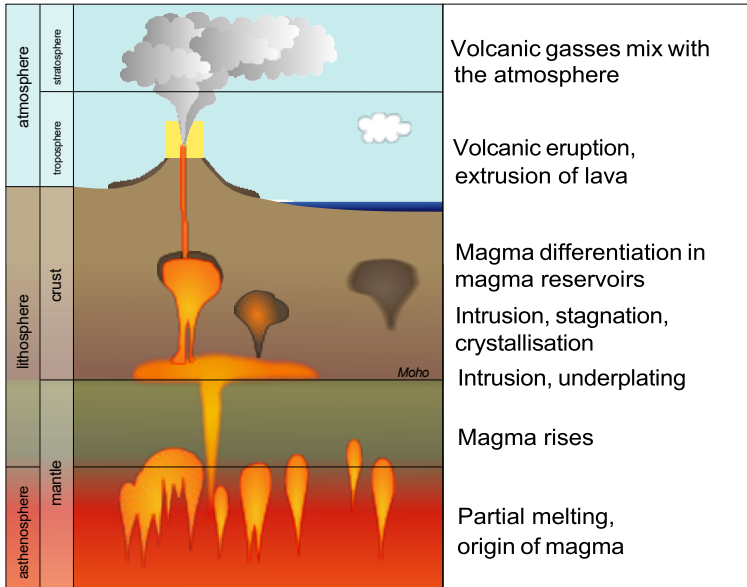
Figure 13.1 "Magmatism and Volcanism" is licensed as Creative Commons Attribution-ShareAlike 3.0 Unported.

Geothermal Gradient

Although it is scorching under the Earth's surface, the crust and mantle are mostly solid. This heat inside the Earth is caused by residual heat left over from the original formation of Earth and radioactive decay. The rate at which temperature increases with depth is called the **geothermal gradient**. The average geothermal gradient in the upper 100 kilometers of the crust is generally about 25 degrees Celsius per kilometer (km). So, for every kilometer of depth, the temperature increases by about 25 degrees Celsius. (4 Igneous Processes and Volcanoes – An Introduction to Geology, n.d.)
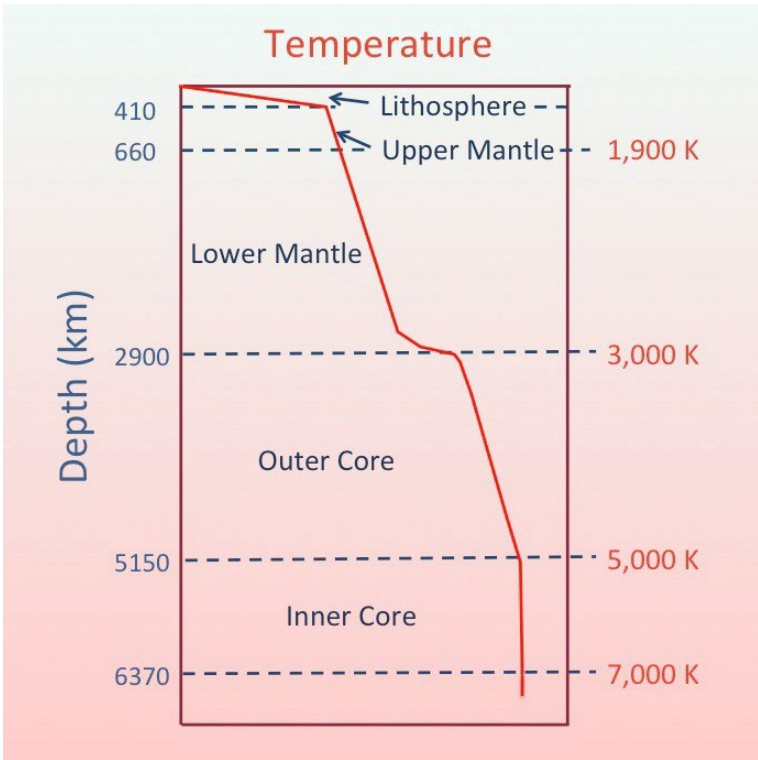
Physical Geography and Natural Disasters by R. Adam Dastrup

Decompression Melting

Magma is created at the mid-ocean ridge by decompression melting. The mantle is solid but is slowly flowing under enormous pressure and temperatures due to convection. Rock is not a good conductor of heat, so as mantle rock rises, the pressure is reduced along with the melting point (the green line), but the rock temperature remains about the same, and the rising rock begins to melt. Pressure changes instantaneously as the rock rises, but temperature changes

slowly because of the low heat conductivity of rock. In the figure above, setting B: mid-ocean ridge shows a mass of mantle rock at a pressure-temperature X on the P-T diagram and its geographical location on the cross-section under a mid-ocean ridge. At this location, the P-T diagram shows the red arrow increasing to the right. Thus, hotter rock is now shallower, at a lower pressure, and the new geothermal gradient (red line) shifts past the solidus (green line) and melting starts. As this magma continues to rise at divergent boundaries and encounters seawater, it cools and crystallizes to form new lithospheric crust. (4 Igneous Processes and Volcanoes – An Introduction to Geology, n.d.)

Flux Melting

Another way that rocks melt is when volatile gases (e.g., water vapor) are added to mantle rock from a descending subducting slab in a process called flux melting (or fluid-induced melting). The subducting slab contains oceanic lithosphere and hydrated minerals. As the slab descends and slowly increases in temperature, volatiles are expelled from these hydrated minerals, like squeezing water out of a sponge. The volatiles then rises into the overlying asthenospheric mantle, which lowers the melting point of the peridotite minerals (olivine and pyroxene). The pressure and temperature of the overlying mantle rock do not change, but the addition of volatiles lowers the melting temperature. This is analogous to adding salt to an icy roadway. The salt lowers the melting/crystallization temperature of the solid water (ice) so that it melts. Another example is welders adding flux to lower the melting point of their welding materials.

Flux melting is illustrated in setting D: island arc (subduction zone) of the P-T diagram above. Volatiles added to mantle rock at location "Z" act as a flux to lower the melting temperature. This is shown in the P-T diagram by the solidus (green line) shifting to the left. The solidus line moves past the geothermal gradient (red line), and melting begins. Magmas producing the volcanoes of the Ring of Fire, associated with the circum-Pacific subduction zones, are a result of flux melting. As introduced in the minerals chapter, water ions can bond with other ions in the crystal structures of amphibole (and other silicates), which is essential in considering how magmas form in subduction zones by "flux melting." Such hydrated minerals in subducting slabs contribute water to the flux melting process. (4 Igneous Processes and Volcanoes – An Introduction to Geology, n.d.)
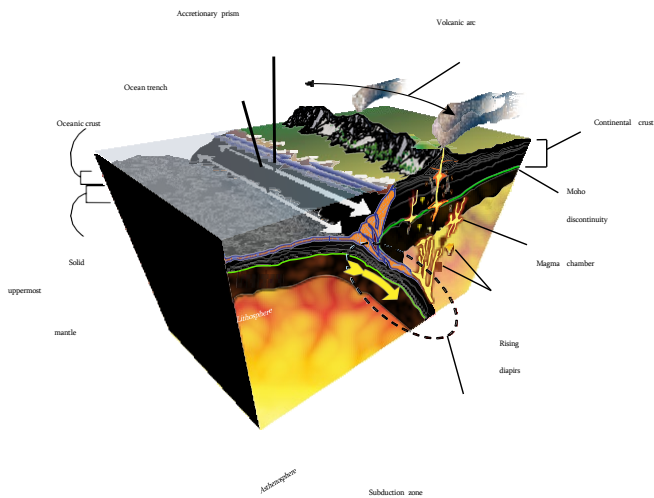
407

Magma Composition

In 1980, Mount St. Helens blew up in the costliest and deadliest volcanic eruption in United States history. The eruption killed 57 people, destroyed 250 homes, and swept away 47 bridges. Mount St. Helens today still has minor earthquakes and eruptions, and now has a horseshoe-shaped crater with a lava dome inside. The dome is formed of viscous lava that oozes into place.

It should first be noted that **magma** is molten material inside the earth, while **lava** is molten on the earth's surface. The reason for the distinction is because lava can cool quickly from the air and solidify into rock rapidly, while magma may never reach the earth's surface. Volcanoes do not always erupt in the same way. Each volcanic eruption is unique, differing in size, style, and composition of the erupted material. One key to what makes the eruption unique is the chemical composition of the magma that feeds a volcano, which determines (1) the eruption style, (2) the type of volcanic cone that forms, and (3) the composition of rocks that are found at the volcano.

Different minerals within rocks melt at different temperatures, and the amount of partial melting and the composition of the original rock determine the magma's composition. Magma collects in magma chambers in the crust at 160 kilometers (100 miles) beneath the surface of a volcano.

The words that describe the composition of igneous rocks also describe magma composition. Mafic magmas are low in silica and have darker magnesium, and iron-rich mafic minerals, such as

olivine and pyroxene. Felsic magmas are higher in silica and have lighter colored minerals such as quartz and orthoclase feldspar. The higher the amount of silica in the magma, the higher is its viscosity. **Viscosity** is a liquid's resistance to flow.

Viscosity determines what the magma will do. **Mafic magma** is not viscous and will flow smoothly to the surface. **Felsic magma** is viscous and does not flow smoothly. Most felsic magma will stay deeper in the crust and will cool to form intrusive igneous rocks such as granite and granodiorite. If felsic magma rises into a magma chamber, it may be too viscous to move, and so it gets stuck. Dissolved gases become trapped by thick magma, and the magma chamber begins to build pressure.

Explosive Eruptions

The type of magma in the chamber determines the type of volcanic eruption. A massive explosive eruption creates even more devastation than the force of the atom bomb dropped on Nagasaki at the end of World War II, in which more than 40,000 people died. A massive explosive volcanic eruption is 10,000 times as powerful. Felsic magmas erupt explosively because of hot, gas-rich magma churning within its chamber. The pressure becomes so great that the magma eventually breaks the seal and explodes, just like when a cork is released from a champagne bottle. Magma, rock, and ash burst upward in an enormous explosion creating volcanic ash called **tephra**. It should be noted that when looked under a microscope, the volcanic "ash" is actual microscopic shards of glass. That is why it is so dangerous to inhale the air following an eruption.

Scorching hot tephra, ash, and gas may speed down the volcano's slopes at 700 km/h (450 mph) as a pyroclastic flow. Pyroclastic flows knock down everything in their path. The temperature inside a pyroclastic flow may be as high as 1,000 degrees Celsius (1,800 degrees Fahrenheit). Before the Mount St. Helens eruption in 1980, the Lassen Peak eruption on May 22, 1915, was the most recent Cascade eruption. A column of ash and gas shot 30,000 feet into the air. This triggered a high-speed pyroclastic flow, which melted snow and created a volcanic mudflow known as a lahar. Lassen Peak currently has geothermal activity and could erupt explosively again. Mt. Shasta, the other active volcano in California, erupts every 600 to 800 years. An eruption would create a large pyroclastic flow, and probably a lahar. Of course, Mt. Shasta could explode and collapse like Mt. Mazama in Oregon.

Video 13.1 Size of volcanic eruptions

https://www.youtube.com/watch?v=YluS9jOjQdM&t=17s

Volcanic gases can form toxic and invisible clouds in the atmosphere that could contribute to environmental problems such as acid rain and ozone destruction. Particles of dust and ash may stay in the atmosphere for years, disrupting weather patterns and blocking sunlight.

411
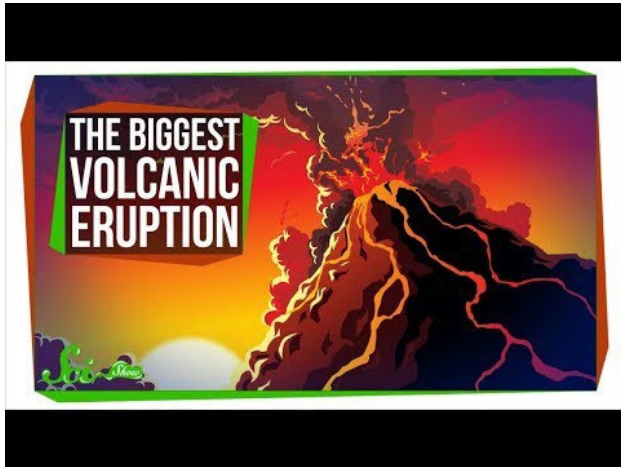
Video 13.2 The biggest volcano eruption
https://www.youtube.com/watch?v=tYzEwLtqKNo

Effusive Eruptions

Mafic magma creates gentler effusive eruptions. Although the pressure builds enough for the magma to erupt, it does not erupt with the same explosive force as felsic magma. People can usually be evacuated before an effusive eruption, so they are much less deadly. Magma pushes toward the surface through fissures and reaches the surface through volcanic vents. Click here to view a lava stream within the vent of a Hawaiian volcano using a thermal camera.

Low-viscosity lava flows down mountainsides. Differences in composition and where the lavas erupt result in lava types like a

ropy form **pahoehoe** and a chunky form called **aa**. Although effusive eruptions rarely kill anyone, they can be destructive. Even when people know that a lava flow is approaching, there is not much anyone can do to stop it from destroying a building, road, or infrastructure.

Figure 13.5 "Pahoehoe Lava" is licensed under Creative Commons Attribution-ShareAlike 4.0 International.

Volcano Features and Types
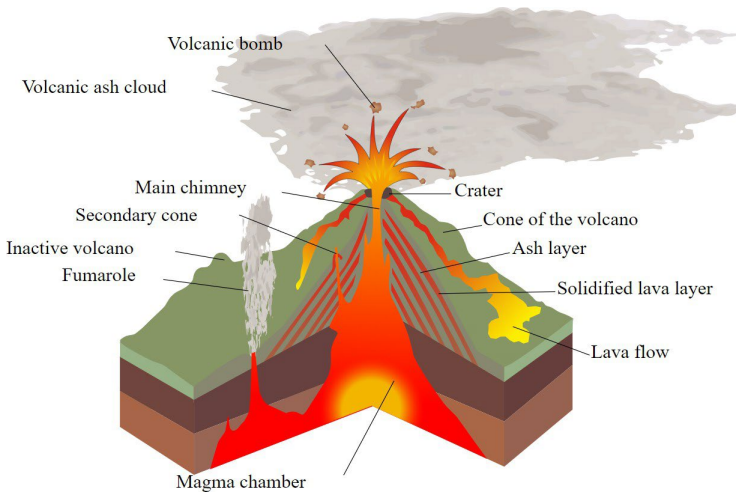
A **volcano** occurs where lava erupts at the surface and solidifies into rock. There are several types of volcanoes based on their shape, eruption style, magmatic composition, and other aspects. The most massive craters are called a caldera, such as the Crater Lake Caldera in Oregon. Many volcanic features are produced by viscosity, a fundamental property of lava. Viscosity is the resistance to flowing by a fluid. Low viscosity magma flows easily more like syrup, the basaltic volcanism that occurs in Hawaii on shield volcanoes. High viscosity means a sticky magma, typically felsic or intermediate, that flows slowly, like toothpaste. (4 Igneous Processes and Volcanoes – An Introduction to Geology, n.d.)

Video 13.3 Volcanoes 101

https://www.youtube.com/watch?v=VNGUdObDoLk

Shield Volcanoes

The largest volcano is a **shield volcano** and is characterized by broad, low-angle flanks, a small vent, or groups of vents at the top, and basaltic magma. The name "shield" comes from the side view resembling a medieval warrior's shield. They are typically associated with hotspots, midocean ridges, or continental rifts where upper mantle material rises, and build up slowly from many low-viscosity basaltic lava flows that can travel long distances, hence making the low-angle flanks. Because the magma is basaltic and low viscosity, the eruption style is not explosive but rather effusive, meaning that

415

volcanic eruptions are small, localized, and predictable. Therefore, this eruption style is not typically much of a hazard.

Video 13.4 The World' tallest mountain

https://www.youtube.com/watch?v=xmiyV7G4K2k

Mauna Loa (info) and the more active Kilauea (info) in Hawaii are good examples of vents on a shield volcano. The eruption of Kilauea from fissures in Hawaii in 2018, while not explosive, produced viscous lavas that did considerable damage to roads and structures. Shield volcanoes are also found in Iceland, the Galapagos Islands, Northern California, Oregon, and the East African Rift (USGS, 2011).
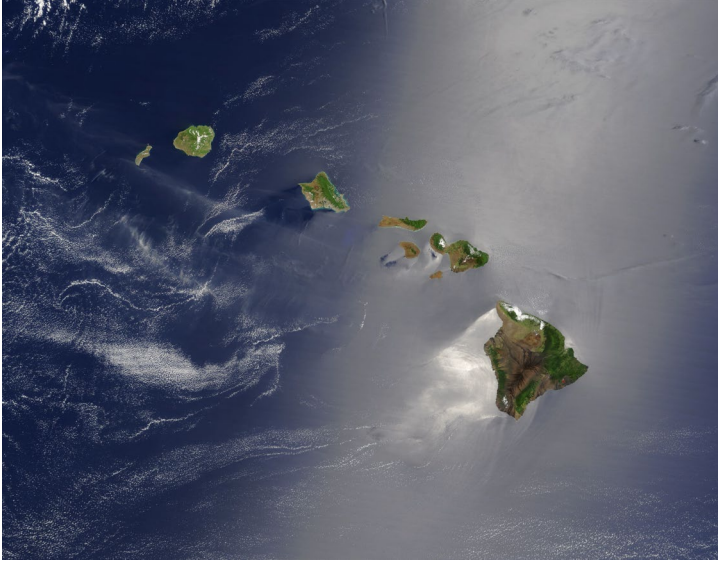
Basaltic magma can form several rock types and unique landforms. Based on magma temperature, composition, and content of dissolved gases and water vapor, there are two main types of basaltic volcanic rocks with Hawaiian names – pahoehoe and aa. Pahoehoe is a basaltic magma that flows smoothly into a "ropey" appearance. In contrast, aa (sometimes spelled a'a or 'a'ā and pronounced "ah-ah") has a crumbly blocky appearance. (Peterson and Tilling 1980). Felsic silica-rich lavas also form aa flows.

In **basaltic lava flows**, the low viscosity lava can smoothly flow, and it tends to harden on the outside but continue to flow internally within a tube. Once the interior flowing lava subsides, the tube may be left as an empty lava tube. Lava tubes famously make

Physical Geography and Natural Disasters by R. Adam Dastrup

caves (with or without collapsed roofs) in Hawaii, Northern California, the Columbia River Basalt Plateau of Washington and Oregon, El Malpais National Monument in New Mexico, and Craters of the Moon National Monument in Idaho. Fissures, cracks that originate from shield-style eruptions, are also common. Magmas from fissures are typically very fluid and mafic. The volcanic activity itself causes some fissures, and some can be influenced by tectonics, such as the common fissures parallel to the divergent boundary in Iceland. See above for fissure flows from Kilauea in 2018.

Since basalt flows are thick accumulations of lava with a homogeneous composition that flows quickly when the lava begins to cool, it can contract into columns with a hexagonal cross-section called columnar jointing. This feature is common in basaltic lava flows but can be found in more felsic lavas and tuffs. (4 Igneous Processes and Volcanoes – An Introduction to Geology, n.d.)

Cinder Cones

Cinder cones are small volcanoes with steep sides, made of cinders and volcanic bombs ejected from a clear central vent. Typically, they come from mafic lavas that have high volatile content. Cinders form when hot lava is ejected into the air, cooling and solidifying before they reach the flank of the volcano. The largest cinders are called volcanic bombs. Cinder cones form in short-lived eruption events that are common in the western United States.

A recent and striking example of a short-lived cinder cone is the 1943 eruption near Parícutin, Mexico. The cinder cone started with

an explosive eruption shooting cinders out of a vent in the middle of a farmer's field. Quickly, volcanism continued building the cone to a height of over 300 feet in a week and 1,200 feet in the first eight months. After the initial explosive gases and cinders were released, growing the cone, basaltic lava poured out around the base of the cone. This order of events is typical for cinder cones: first violent eruption, then the formation of cone and crater, followed by a low-viscosity lava flow from the base (the cone of cinders is not strong enough to support a column of lava rising to the top of the crater). The Parícutin cinder cone was built over nine years and covered about 100-square miles with ashes and destroyed the town of San Juan.

Figure 13.8 "Groundhog Cinder Cone" by the United States Department of Agriculture is licensed under Public Domain.

Lava Domes

Lava domes are a relatively small accumulation of silica-rich volcanic rocks, such as rhyolite and obsidian, that is too viscous to flow, and therefore, pile high close to the vent. The domes often form within the collapsed crater of a stratovolcano near the vent and grow by expansion from within. As it grows, its outer surface cools and hardens, then shatters, spilling loose fragments down its sides. An excellent example of a lava dome is inside of a collapsed stratovolcano crater is Mount Saint Helens. Examples of a stand-alone lava dome are Chaiten in Chile and the Mammoth Mountain in California. (Volcanoes: Principal Types of Volcanoes, n.d.)
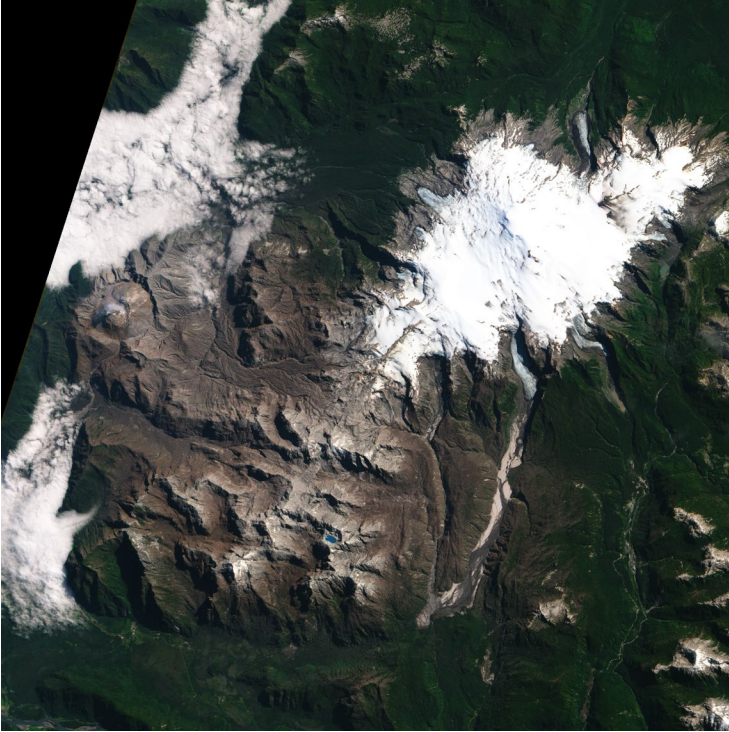
Figure 13.9 "Chaiten Volcano Lava Dome, Chile" by NASA's Earth Observatory is licensed under Public Domain.

Composite Volcanoes

Composite volcanoes, also called **stratovolcanos**, has steep flanks, a symmetrical cone shape, a distinct crater, and rises prominently above the surrounding landscape. The figure at the beginning of this section shows a stratovolcano. Examples include Mount Rainier in the Cascade Range in Washington and Mount Fuji in Japan. Stratovolcanoes can have magma with felsic to mafic composition. However, felsic to intermediate magmas are most

common. The term "composite" refers to the alternating layers of pyroclastic materials (like ash) and lava flows. The viscous nature of the intermediate and felsic magmas in subduction zones results in steep flanks and explosive eruption styles. Stratovolcanoes are made of alternating lava flows and ash. (4 Igneous Processes and Volcanoes – An Introduction to Geology, n.d.)

Figure 13.10 "Mount Rainier and the Emmons Glacier" by the National Park Service is licensed under Public Domain.

Video 13.5 Volcano: The Power of Sleeping Giants

https://www.youtube.com/watch?v=D7fR2Z880z8

Calderas

Calderas are generally large (some up to 15 miles in diameter), steep-walled, basin-shaped depressions formed by the collapse of a volcanic edifice into an emptying magma chamber and although the word caldera only refers to the vent, many use calderas as a volcano type, typically formed by high-viscosity felsic volcanism with high volatile content. Crater Lake, Yellowstone, and Long Valley Caldera are good examples. At Crater Lake National Park in Oregon, about 6,800 years ago, Mount Mazama was a composite volcano that erupted in a sizeable explosive blast, ejecting massive volcanic ash. The eruption rapidly drained the underlying magma

chamber causing the top to collapse into it, forming a significant depression that later filled with water. Today a resurgent dome is found rising through the lake as a cinder cone, called Wizard Island.
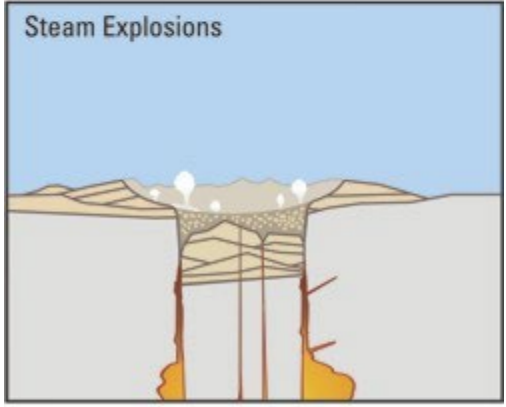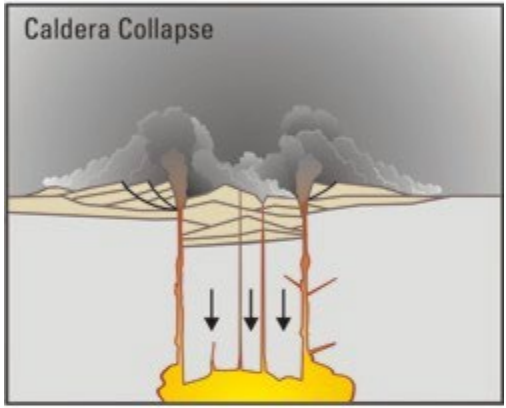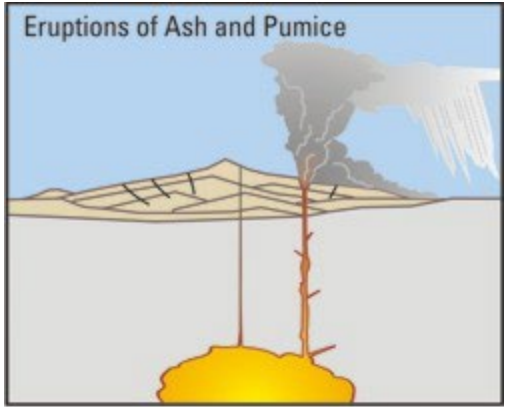
Eruptions of Ash and Pumice

Caldera Collapse

Steam Explosions

Today

Garfield Peak   Wizard Island   Llao Rock

Super-volcanoes

Video 13.6 Super-volcanoes
https://www.youtube.com/watch?v=kAlawvE8lVw

The Yellowstone caldera erupted three times in the recent past, at 2.1, 1.3, and 0.64 million years ago. Each eruption created large rhyolite flows and pyroclastic clouds of ash that solidified into tuff. These extra-large eruptions rapidly emptied the magma chamber causing the roof to collapse and form a caldera. Three calderas are still preserved from these eruptions, and most of the roads and

hotels of Yellowstone National Park are located within the caldera. Two resurgent domes are located within the last caldera.
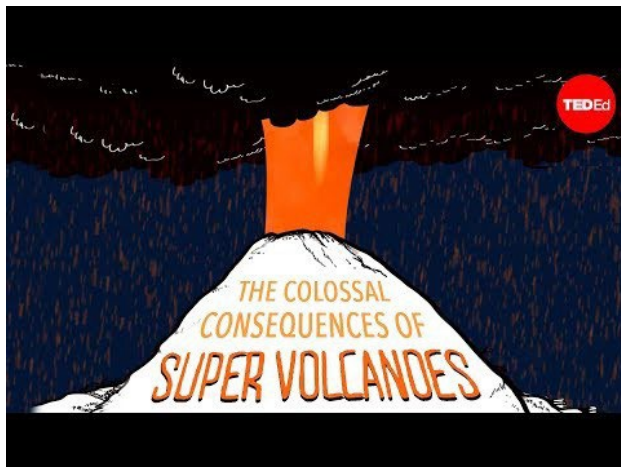
Video 13.7 The colossal consequence of super volcanoes

https://www.youtube.com/watch?v=hDNlu7Qf6_E

Yellowstone volcanism started as a hot spot under the North American lithosphere about 17-million years ago near the Oregon/ Nevada border. As the North American plate slid southwestward over the stationary hotspot, surface volcanism followed and helped form Idaho's Snake River Plain, eventually arriving at its current location in northwestern Wyoming. As the plate moved to the southwest over the stationary hotspot, it left a track of past volcanic activities.
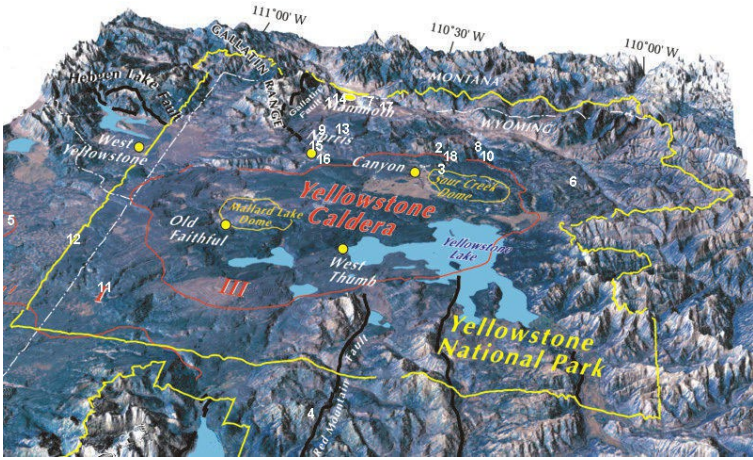
The Long Valley Caldera near Mammoth California is a massive explosive volcano that erupted 760,000 years ago and dumped a significant amount of ash throughout the United States, similar to the Yellowstone eruptions. This ash formed the large Bishop Tuff deposit. Like the Yellowstone caldera, the Long Valley Caldera contains the town of Mammoth Lakes, a major ski resort, an airport, and a major highway. Further, there is a resurgent dome in the middle and active hot springs. (4 Igneous Processes and Volcanoes – An Introduction to Geology, n.d.)

Video 13.8 Don't worry about the Yellowstone volcano

https://www.youtube.com/watch?v=v0c7wcpJslg

Distribution of Volcanic Activity

Most volcanoes are located at active plate boundaries called **interplate volcanism**. The prefix "inter-" means "between." In contrast, some volcanoes are not associated with plate boundaries but are found within the plate far from plate boundaries. These are called intraplate volcanoes, and hotspots and fissure eruptions form many. The prefix "intra-" means "within." The following discusses volcanism's location in more detail with mid-ocean ridges, subduction zones, continental rifts representing interplate volcanism, and hot spots representing intraplate volcanism.

Physical Geography and Natural Disasters by R. Adam Dastrup

Volcanoes along Mid-Oceanic Ridges

Although most volcanism occurs on the ocean floor along the mid-ocean ridge (a type of divergent plate boundary), they are the least observed since most are under 10,000 to 15,000 feet of ocean, an exception being Iceland. As the oceanic plates diverge and thin, hot mantle rock is allowed to rise, pressure from depth is released, which causes the ultramafic mantle rock (peridotite) to melt partially. The resulting magma is basaltic in composition based on the concept of partial melting discussed earlier. Because most volcanoes on the ocean floor are basaltic, most of the oceanic lithosphere is also basaltic near the surface with phaneritic gabbro and ultramafic peridotite forming underneath. Icelandic volcanism is an example of this, but lying above sea level.

An underwater volcanic eruption occurs when basaltic magma erupts underwater, forming **pillow basalts** or in small explosive eruptions. Lava erupting into seawater forms pillow-shaped structures, hence the name. In association with these seafloor eruptions, an entire underwater ecosystem thrives in parts of the mid-ocean ridge. This ecosystem exists around tall vents emitting black, hot mineral-rich water called **deep-sea hydrothermal vents** (also known as **black smokers**).

This hot water, up to 380 degrees Celsius (716 degrees Fahrenheit), is heated by the magma and dissolves many elements that support the ecosystem. Deep underwater where the sun cannot reach, this ecosystem of organisms depends on the heat of the vent for energy and vent chemicals as its foundation of life called **chemosynthesis**. The foundation of the ecosystem is hydrogen sulfide-oxidizing

bacteria that live symbiotically with the larger organisms. Hydrogen sulfide (H_2S, the gas that smells like rotten eggs) needed by these bacteria is contained in the volcanic gases emitted from the hydrothermal vents. The source of most of this sulfur and other elements is the Earth's interior. (4 Igneous Processes and Volcanoes – An Introduction to Geology, n.d.)
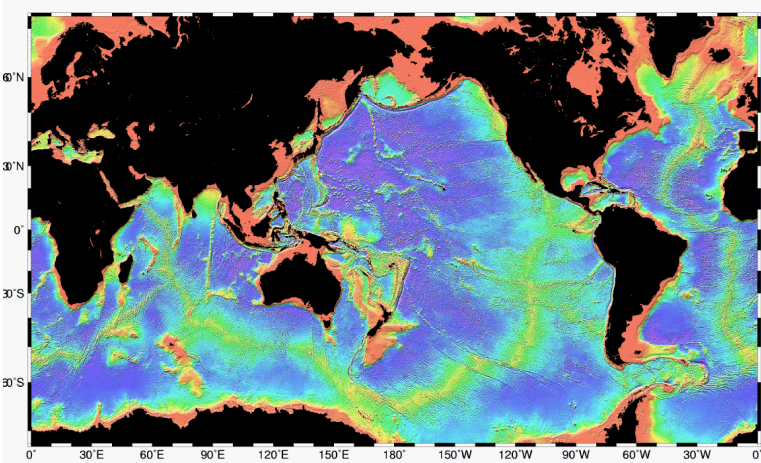
Figure 13.14 "Mid-Ocean Ridge System" by the National Oceanic and Atmospheric Administration is licensed under Public Domain.

Volcanoes along Convergent Boundaries

Volcanoes are a vibrant manifestation of plate tectonics processes. Volcanoes are common along convergent and divergent plate boundaries but are also found within lithospheric plates away from plate boundaries. Wherever mantle can melt, volcanoes may be the result. Volcanoes erupt because mantle rock melts. The first stage in creating a volcano is when mantle rock begins to melt because of

scorching temperatures, lithospheric pressure lowers, or water is added.

Subduction Zones

During the process of **subduction**, water is expelled from the hydrated minerals causing partial melting by flux melting in the overlying mantle rock. This creates a mafic magma that rises through the lithosphere and can change the composition by interacting with surrounding continental crust as well as by magma differentiation. These changes then evolve basaltic magma into a more silica-rich rock in volcanoes and plutons. These silica-rich rocks are felsic to intermediate rocks such as andesite, rhyolite, pumice, and tuff. The **Ring of Fire** surrounding the Pacific Ocean is dominated by subduction and contains volcanoes with silica-rich magma. These volcanoes are discussed in more detail in the stratovolcano section.

Large earthquakes are prevalent along convergent plate boundaries. Since the Pacific Ocean is rimmed by convergent and transform boundaries, 80 percent of all earthquakes occur around the Pacific Ocean basin, called the Ring of Fire. A description of the Pacific Ring of Fire along western North America is below:

- Subduction at the Middle American Trench creates volcanoes in Central America.
- The San Andreas Fault is a transform boundary.
- Subduction of the Juan de Fuca plate beneath the North American plate creates the Cascade volcanoes like Mount St. Helens, Mount Rainer, Mount Hood, and more.

- Subduction of the Pacific plate beneath the North American plate in the north creates the long chain of the Aleutian Islands volcanoes near Alaska.
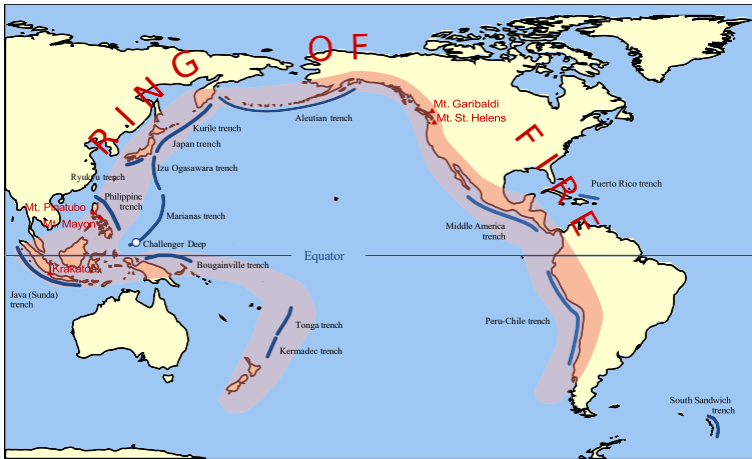
Creative Commons Attribution

An Introduction to Geology by Chris Johnson, Matthew D. Affolter, Paul Inkenbrandt, Cam Mosher is licensed under CC BY- NC-SA 4.0

Natural Disasters and Human Impacts by R. Adam Dastrup, MA, GISP is licensed under CC BY-NC-SA 4.0

Physical Geology – 2nd Edition by Steven Earle is licensed under CC BY-NC-SA 4.0

Geology by Lumen Learning is licensed under CC BY-NC-SA 4.0

Earth Science by Lumen Learning is licensed under CC BY-NC-SA 4.0

Chapter 14 WEATHERING

Weathering is what takes place when a body of rock is exposed to the "weather" — in other words, to the forces and conditions that exist at Earth's surface. Most rocks are formed at some depth within the crust except for volcanic rocks and some sedimentary rocks. They experience relatively constant temperature, high pressure, no contact with the atmosphere, and little or no moving water. Once a rock is exposed at the surface, which is what happens when the overlying rock is eroded, conditions change dramatically. Temperatures vary widely, there is much less pressure, oxygen and other gases are plentiful, and in most climates, water is abundant.

Weathering includes two main processes that are entirely different. One is the mechanical breakdown of rock into smaller fragments, and the other is the chemical change of the minerals within the rock to forms that are stable in the surface environment. Mechanical weathering provides fresh surfaces for attack by chemical processes, and chemical weathering weakens the rock so that it is more susceptible to mechanical weathering. Together, these processes create two significant products, one being the sedimentary clasts and ions in solution that can eventually become sedimentary rock, and the other is the soil necessary for our existence on Earth.

Mechanical Weathering

Intrusive igneous rocks form at depths of several hundreds of meters to several tens of kilometers. Sediments are turned into sedimentary rocks only when other sediments bury them to depths more than several hundreds of meters. Most metamorphic rocks are formed at depths of kilometers to tens of kilometers. Weathering cannot even begin until these rocks are uplifted through various mountain-building processes — most of which are related to plate tectonics — and the overlying material has been eroded, and the rock is exposed as an **outcrop**. The critical agents of mechanical weathering are:

- The decrease in pressure that results from the removal of

436

overlying rock

- Freezing and thawing of water in cracks in the rock
- Formation of salt crystals within the rock
- Cracking from plant roots and exposure by burrowing animals

When a mass of rock is exposed by weathering and removal of the overlying rock, there is a decrease in the confining pressure on the rock, and the rock expands. This unloading promotes cracking of the rock, known as **exfoliation**.

Granitic rock tends to exfoliate parallel to the exposed surface because it is typically homogenous, and it does not have predetermined planes along which it must fracture. Sedimentary and metamorphic rocks, on the other hand, tend to exfoliate along predetermined planes.

Frost wedging, also called **ice wedging**, is the process by which water seeps into cracks in a rock, expands on freezing, and thus enlarges the cracks. The effectiveness of frost wedging is related to the frequency of freezing and thawing. Frost wedging is most effective in mountainous climates. In warm areas where freezing is infrequent, in very cold areas where thawing is infrequent, or in arid areas, where there is little water to seep into cracks, the role of frost wedging is limited.

In many mountainous regions, the transition between freezing nighttime temperatures and thawing daytime temperatures is frequent — tens to hundreds of times a year. Even in warm coastal areas, freezing and thawing transitions are frequent at higher elevations. A common feature in areas of active frost wedging is a

437

talus slope — a fan-shaped deposit of fragments removed by frost wedging from the steep rocky slopes above.
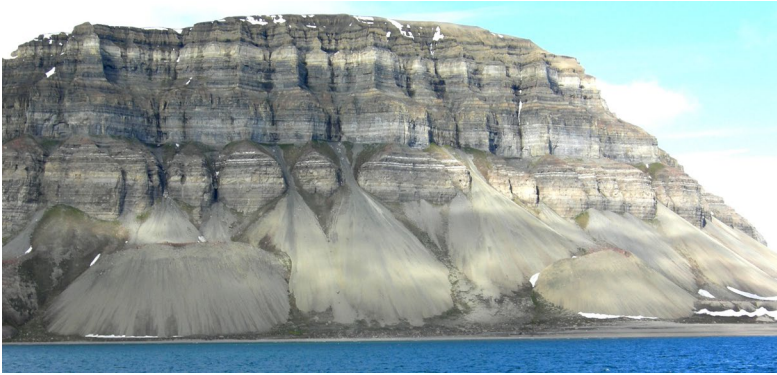
A related process, **frost heaving**, takes place within unconsolidated materials on gentle slopes. In this case, water in the soil freezes and expands, pushing the overlying material up. Frost heaving is responsible for winter damage to roads all over North America.

When saltwater seeps into rocks and then evaporates on a hot sunny day, salt crystals grow within cracks and pores. The growth of these crystals exerts pressure on the rock and can push grains apart, causing the rock to weaken and break. Salt weathering can also occur away from the coast because most environments have some salt in them.

The effects of plants and animals are significant in mechanical weathering. Roots can force their way into even the tiniest cracks, and they exert tremendous pressure on the rocks as they grow, widening the cracks and breaking the rock. Although animals do generally not burrow through solid rock, they can excavate and

remove vast volumes of soil, and thus expose the rock to weathering by other mechanisms.

Mechanical weathering is greatly facilitated by erosion, which is the removal of weathering products, allowing for the exposure of more rock for weathering. On the steep rock faces at the top of the cliff, rock fragments have been broken off by ice wedging, and then removed by gravity. This is a form of **mass wasting**, which will be looked at later in this chapter. Other essential agents of erosion that have the effect of removing the products of weathering include water in streams, ice in glaciers, and waves on the coasts.

Chemical Weathering

Chemical weathering results from chemical changes to minerals that become unstable when they are exposed to surface conditions. The kinds of changes that take place are highly specific to the mineral and the environmental conditions. Some minerals, like quartz, are virtually unaffected by chemical weathering, while others, like feldspar, are easily altered. In general, the degree of chemical weathering is most significant in warm and wet climates and least in cold and dry climates. The essential characteristics of surface conditions that lead to chemical weathering are the presence of water (in the air and on the ground surface), the abundance of oxygen, and the presence of carbon dioxide, which produces weak carbonic acid when combined with water.

The Products of Weathering and

Erosion

The products of weathering and erosion are the unconsolidated materials that we find around us on slopes, beneath glaciers, in stream valleys, on beaches, and in deserts. The nature of these materials — their composition, size, degree of sorting, and degree of rounding — is determined by the type of rock that is being weathered, the nature of the weathering, the erosion, and transportation processes, and the climate.

A summary of the weathering products of some of the common minerals present in rocks is provided below.

Common Mineral	Typical Weathering Products
Quartz	Quartz as sand grains
Feldspar	Clay minerals plus potassium, sodium, and calcium in solution
Biotite and Amphibole	Chlorite plus iron and magnesium in solution
Pyroxene and Olivine	Serpentine plus iron and magnesium in solution
Calcite	Calcium and carbonate in solution
Pyrite	Iron oxide minerals plus iron in solution and sulphuric acid

The produces created from weathering range widely in size and shape, depending on the processes involved. If and when deposits like these are turned into sedimentary rocks, the textures of those

440

rocks will vary significantly. Importantly, when we describe sedimentary rocks that formed millions of years in the past, we can use those properties to make inferences about the conditions that existed during their formation.

WEATHERING AND SOIL FORMATION

Weathering is a vital part of the process of soil formation, and soil is critical to our existence on Earth. Many people refer to any loose material on Earth's surface as soil, but to geographers, soil is the material that includes organic matter, lies within the top few tens of centimeters of the surface, and is vital in sustaining plant growth.

Soil is a complex mixture of minerals (approximately 45 percent), organic matter (approximately 5 percent), and empty space (approximately 50 percent, filled to varying degrees with air and water). The mineral content of soils is variable but is dominated by clay minerals and quartz, along with minor amounts of feldspar and small fragments of rock. The types of weathering that take place within a region have a significant influence on soil composition and texture. For example, in a warm climate, where chemical weathering dominates, soils tend to be more abundant in clay. Soil scientists describe soil texture in terms of the relative proportions of sand, silt, and clay. The sand and silt components in this diagram are dominated by quartz, with lesser amounts of feldspar and rock fragments, while the clay minerals dominate the clay component.

Soil forms through accumulation and decay of organic matter and the mechanical and chemical weathering processes described above. The factors that affect the nature of soil and the rate of its formation include climate (especially average temperature and

441

precipitation amounts, and the following types of vegetation), the type of parent material, the slope of the surface, and the amount of time available.

Formation of Soil

Climate

Soils develop because of the weathering of materials on Earth's surface, including the mechanical breakup of rocks, and the chemical weathering of minerals. The downward percolation of water facilitates soil development. Soil forms most readily under temperate to tropical conditions (not cold), and precipitation amounts are moderate (not dry, but not too wet). Chemical weathering reactions (especially the formation of clay minerals) and biochemical reactions proceed fastest under warm conditions, and plant growth is enhanced in warm climates. Too much water (e.g., in rainforests) can lead to the leaching of critical chemical nutrients and, hence, acidic soils. In humid and poorly drained regions, swampy conditions may prevail, producing soil that is dominated by organic matter. Too little water (e.g., in deserts and semi-deserts), results in minimal downward chemical transportation and the accumulation of salts and carbonate minerals (e.g., calcite) from upward-moving water. Soils in dry regions also suffer from a lack of organic material.

Parent Material

Soil **parent materials** can include all different bedrock

types and any unconsolidated sediments, such as glacial deposits and stream deposits. Soils are described as residual soils if they develop on bedrock, and transported soils if they develop on transported material such as glacial sediments. However, the term "transported soil" is misleading because it implies that the soil itself has been transported, which is not the case. When referring to such soil, it is better to be specific and say "soil developed on unconsolidated material," because that distinguishes it from soil developed on bedrock.

Parent materials provide essential nutrients to residual soils. For example, a minor constituent of granitic rocks is the calcium-phosphate mineral apatite, which is a source of the critical soil nutrient phosphorus. Basaltic parent material tends to generate very fertile soils because it also provides phosphorus, along with significant amounts of iron, magnesium, and calcium.

Some unconsolidated materials, such as river-flood deposits, make for especially good soils because they tend to be rich in clay minerals. Clay minerals have large surface areas with negative charges that are attractive to positively charged elements like calcium, magnesium, iron, and potassium, which are essential nutrients for plant growth.
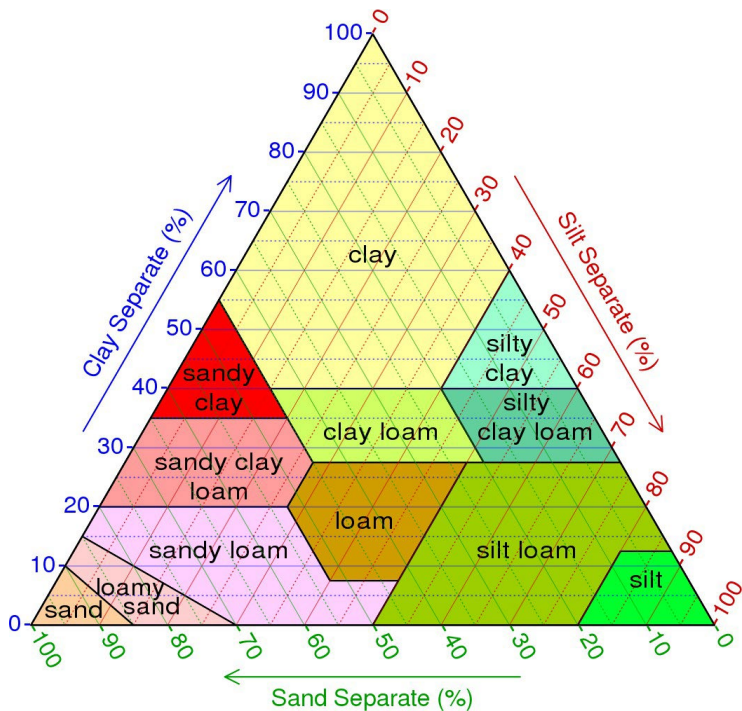
Figure 14.3 "Soil Texture" by the United States Department of Agriculture is licensed under Creative Commons Attribution-Share Alike 4.0 International.

Slope

Soil can only develop where surface materials remain in place and are not frequently moved away by mass wasting. Soils cannot develop where the rate of soil formation is less than the rate of erosion, so steep slopes tend to have little or no soil.

Time

Even under ideal conditions, soil takes thousands of years to develop. Virtually all of northern North America was still glaciated up until 14 ka. Glaciers still dominated the central and northern parts of Canada until around 10 ka, and so, at that time, conditions were still not ideal for soil development even in the southern regions. Therefore, soils in Canada, and especially in central and northern Canada, are relatively young and not well developed.

The same applies to soils forming on newly created surfaces, such as recent deltas or sandbars, or in areas of mass wasting.

Soil Horizons

The process of soil formation generally involves the downward movement of clay, water, and dissolved ions, and a typical result of that is the development of chemically and texturally different layers known as **soil horizons**. The typically developed soil horizons are:

- O — the layer of organic matter
- A — the layer of partially decayed organic matter mixed with mineral material
- E— the eluviated (leached) layer from which some of the clay and iron have been removed to create a pale layer that may be sandier than the other layers
- B — the layer of accumulation of clay, iron, and other elements from the overlying soil
- C — the layer of incomplete weathering

445
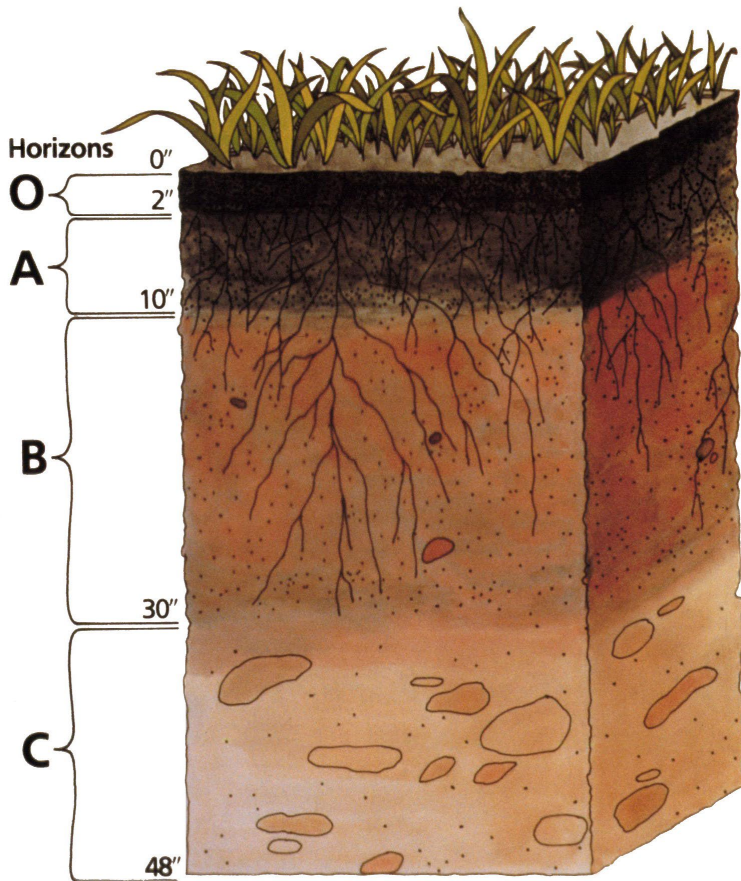
Figure 14.4 "Soil Profile" by the United State Department of Agriculture is licensed under Public Domain.

Another type of layer develops in hot, arid regions known as **caliche** (pronounced ca-lee-chee). It forms from the downward (or in some cases upward) movement of calcium ions, and the precipitation of calcite within the soil. When well developed,

caliche cements the surrounding material together to form a layer that has the consistency of concrete.

Like all geological materials, soil is subject to erosion, although, under natural conditions on gentle slopes, the rate of soil formation either balances or exceeds the rate of erosion. Human practices related to forestry and agriculture have significantly upset this balance.

Figure 14.5 "Soil Profile" is licensed under Creative Commons Attribution-ShareAlike 4.0 International.

Soils are held in place by vegetation. When vegetation is removed, either through cutting trees or routinely harvesting crops and tilling the soil, that protection is either temporarily or permanently lost. The primary agents of the erosion of unprotected soil are water and wind.

447

Soil is an important reservoir for organic components necessary for plants, animals, and microorganisms to live. The organic component of soil, called **humus**, is a rich source of bioavailable nitrogen. Nitrogen is the most common element in the atmosphere, but it exists in a form most life forms are unable to use. Special bacteria found only in soil provide most nitrogen compounds that are usable, bioavailable, by life forms.
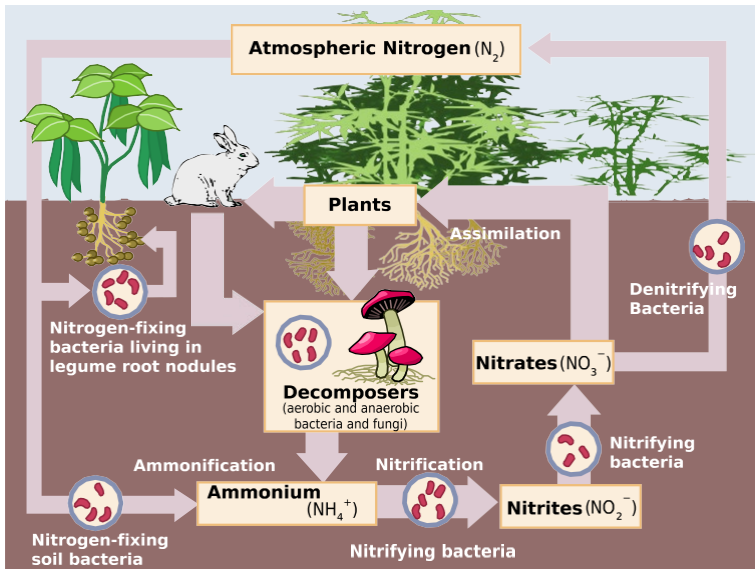
Figure 14.6 "Nitrogen Cycle" by the Environmental Protection Agency is licensed under Public Domain.

These nitrogen-fixing bacteria absorb nitrogen from the atmosphere and convert it into nitrogen compounds. These compounds are absorbed by plants and used to make DNA, amino acids, and enzymes. Animals obtain bioavailable nitrogen by eating plants, and this is the source of most of the nitrogen used by life. That nitrogen is an essential component of proteins and DNA.

Soils range from poor to rich, depending on the amount of humus they contain. Soil productivity is determined by water and nutrient content. Freshly created volcanic soils, called **andisols**, and clay-rich soils that hold nutrients and water are examples of productive soils.
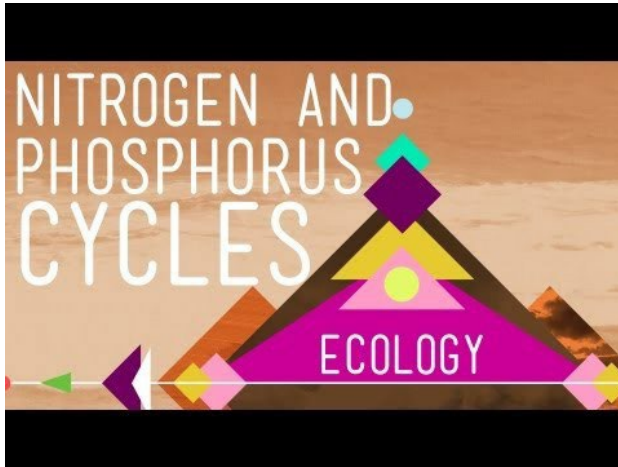
Video 14.1 Nitrogen and phosphorous cycles

https://www.youtube.com/watch?v=leHy-Y_8nRs

Water erosion is accentuated on sloped surfaces because fast-flowing water has higher eroding power than still water. Raindrops can disaggregate exposed soil particles, putting the finer material (e.g., clays) into suspension in the water. **Sheetwash**, unchannelled flow across a surface carries suspended material away, and channels erode right through the soil layer, removing both fine and coarse material.

Wind erosion is exacerbated by the removal of trees that act as

449

windbreaks and by agricultural practices that leave bare soil exposed.

Tillage is also a factor in soil erosion, especially on slopes, because each time a cultivator lifts the soil, it is moved a few centimeters down the slope.

MASS WASTING

Mass wasting, which is synonymous with *slope failure*" is the failure and downslope movement of rock or unconsolidated materials in response to gravity. The term *landslide* is almost synonymous with mass wasting, but not entirely because some people reserve "landslide" for relatively rapid slope failures, while others do not. Other than the video below, this textbook will avoid using the term "landslide."

450

Video 14.2 What causes rockslides

https://www.youtube.com/watch?v=RCxvbosa4fU

Factors That Control Slope Stability

Mass wasting happens because tectonic processes have created uplift. **Erosion**, driven by gravity, is the inevitable response to that uplift, and various types of erosion, including mass wasting, have created slopes in the uplifted regions. Slope stability is ultimately determined by two factors: the angle of the slope and the strength of the materials on it.

A block of rock is typically situated on a rock slope that is being pulled toward Earth's center (vertically down) by gravity. The vertical gravitational force can be split into two components relative to the slope: one pushing the block down the slope, called the **shear force**, and the other pushing into the slope, called the **normal force**. The shear force, which wants to push the block down the slope, has to overcome the strength of the connection between the block and the slope, which may be quite weak if the block has split away from the main body of rock, or may be very strong if the block is still a part of the rock. If the **shear strength** is greater than the shear force, the block should not move. However, if the shear force becomes stronger than the shear strength, the block of rock will slide down the slope.

As already noted, slopes are created by uplift, followed by erosion. In areas with relatively recent uplift, slopes tend to be quite steep. This is especially true where glaciation has taken place because glaciers in mountainous terrain create steep-sided valleys. In areas without recent uplift, slopes are less steep because hundreds of millions of years of erosion (including mass wasting) have made

Physical Geography and Natural Disasters by R. Adam Dastrup

them that way. However, as we will see, some mass wasting can happen even on relatively gentle slopes.

The strength of the materials on slopes can vary widely. Solid rocks tend to be strong, but there is an extensive range of rock strength. If we consider just the strength of the rocks and ignore issues like fracturing and layering, then most crystalline rocks, like granite, basalt, or gneiss, are very strong, while some metamorphic rocks, like schist, are moderately strong. Sedimentary rocks have variable strength. Dolostone and some limestone are strong, most sandstone and conglomerate are moderately strong, and some sandstone and all mudstones are quite weak.

Fractures, metamorphic foliation, or bedding can significantly reduce the strength of a body of rock, and in the context of mass wasting, this is most critical if the planes of weakness are parallel to the slope and least critical if they are perpendicular to the slope.

Internal variations in the composition and structure of rocks can significantly affect their strength. Schist, for example, may have layers rich in sheet silicates (mica or chlorite), and these will tend to be weaker than other layers. Some minerals tend to be more susceptible to weathering than others, and the weathered products are commonly quite weak (e.g., the clay formed from feldspar).

Unconsolidated sediments are generally weaker than sedimentary rocks because they are not cemented and, in most cases, have not been significantly compressed by overlying materials. This binding property of sediment is sometimes referred to as cohesion. Sand and silt tend to be particularly weak, clay is generally a little stronger, and sand mixed with clay can be stronger still. Finer deposits are

relatively strong (they maintain a steep slope), while the overlying sand is relatively weak, and has a shallower slope that has recently failed. Glacial till, typically a mixture of clay, silt, sand, gravel, and larger clasts, forms and is compressed beneath tens to thousands of meters of glacial ice so it can be as strong as some sedimentary rock.

Apart from the type of material on a slope, the amount of water that the material contains is the most critical factor controlling its strength. This is especially true for unconsolidated materials, but it also applies to bodies of rock. Granular sediments, like the sand at Point Grey, have lots of spaces between the grains. Those spaces may be arid (filled only with air); or moist (often meaning that some spaces are water-filled, some grains have a film of water around them, and small amounts of water are present where grains are touching each other); or completely saturated. Unconsolidated sediments tend to be strongest when they are moist because the small amounts of water at the grain boundaries hold the grains together with surface tension. Dry sediments are held together only by the friction between grains, and if they are well sorted or well rounded, or both, that cohesion is weak. Saturated sediments tend to be the weakest of all because the large amount of water pushes the grains apart, reducing the mount friction between grains. This is especially true if the water is under pressure.

Water will also reduce the strength of solid rock, especially if it has fractures, bedding planes, or clay-bearing zones. This effect is even more significant when the water is under pressure, which is why holes are drilled into rocks on road cuts to relieve this pressure.

Water also has a particular effect on clay-bearing materials. All clay minerals will absorb a little bit of water, and this reduces their

454

strength. The **smectite clays**, such as the **bentonite** used in cat litter, can absorb much water, and that water pushes the sheets apart at a molecular level and makes the mineral swell. Smectite that has expanded in this way has almost no strength; it is incredibly slippery.

Moreover, finally, water can significantly increase the mass of the material on a slope, which increases the gravitational force pushing it down. A body of sediment that has 25% porosity and is saturated with water weighs approximately 13% more than it does when it is completely dry, so the gravitational shear force is also 13% higher.

Mass-Wasting Triggers

The shear force and the shear strength of materials on slopes, and about factors that can reduce the shear strength. Shear force is primarily related to the slope angle, and this does not change quickly. However, shear strength can change quickly for various reasons, and events leading to a rapid reduction in shear strength are considered to be *triggers* for mass waste.

An increase in water content is the most common mass-wasting trigger. This can result from the rapid melting of snow or ice, heavy rain, or some event that changes the pattern of water flow on the surface. Rapid melting can be caused by a dramatic increase in temperature (e.g., in spring or early summer) or by a volcanic eruption. Heavy rains are typically related to storms. Earthquakes can cause changes in water flow patterns, previous slope failures that dam up streams, or human structures that interfere with runoff (e.g., buildings, roads, or parking lots).

In some cases, a decrease in water content can lead to failure. This is most common with clean sand deposits, which lose strength when there is no more water around the grains. Freezing and thawing can also trigger some forms of mass wasting. More specifically, the thawing can release a block of rock attached to a slope by a film of ice. One other process that can weaken a body of rock or sediment is shaking. The most apparent source of shaking is an earthquake, but shaking from highway traffic, construction, or mining will also do the job. Several deadly mass-wasting events (including snow avalanches) were triggered by the M7.8 earthquake in Nepal in April 2015.

Classification of Mass Wasting

It is crucial to classify slope failures so that we can understand what causes them and how to mitigate them. The three criteria used to describe slope failures are:

- The type of material that failed (typically either bedrock or unconsolidated sediment).
- The mechanism of the failure (how the material moved).
- The rate at which it moved.

The type of motion is the essential characteristic of slope failure, and there are three different types of motion:

- If the material drops through the air, vertically or nearly vertically, it is known as a **fall**.
- If the material moves as a mass along a sloping surface

(without internal motion within the mass), it is a **slide**.

- If the material has internal motion, like a fluid, it is a **flow**.
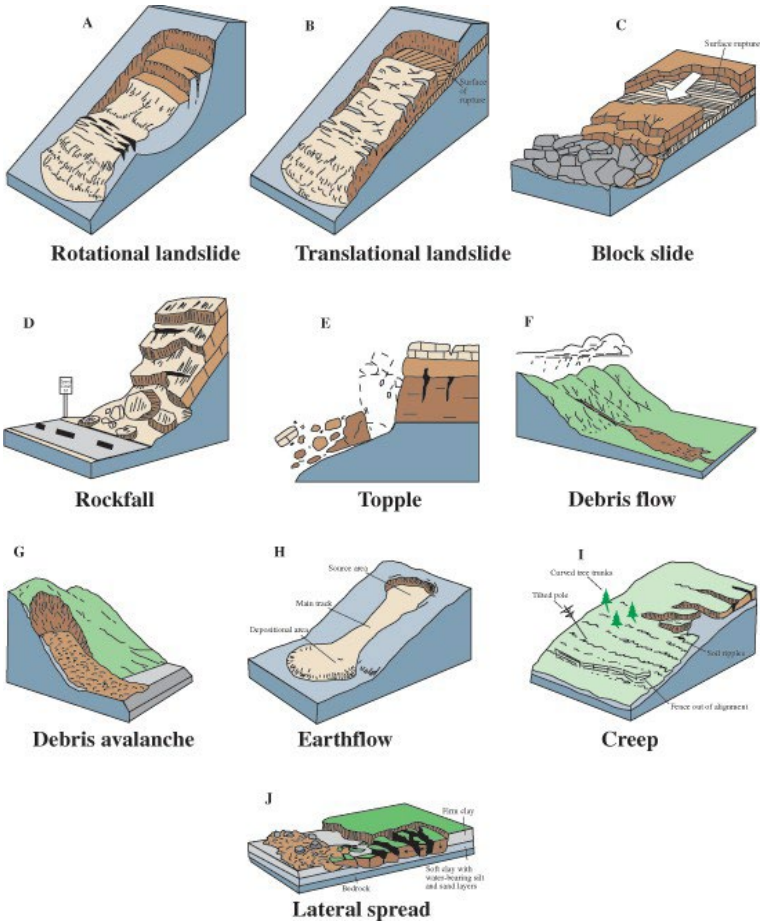
Figure 14.7 "Mass Wasting Types" by the United States Geologic Survey is licensed under Public Domain.

Unfortunately, it is not typically that simple. Many slope failures involve two of these types of motion, some involve all three, and in

many cases, it is not easy to tell how the material moved. The types of slope failure are summarized below.

Failure Type	Type of Material	Type of Motion	Rate of Motion
Rock fall	Rock fragments	Vertical or near-vertical fall (plus bouncing in many cases)	Very fast (>10s m/s)
Rock slide	A large rock body	Motion as a unit along a planar surface (translational sliding)	Typically very slow (mm/y to cm/y), some can be faster
Rock avalanche	A large rock body that slides and then breaks into smaller fragments	Flow (at high speeds, the mass of rock fragments is suspended on a cushion of air)	Very fast (>10s m/s)
Creep or solification	Soil or other overburden in some small cases, mixed with ice	Flow (although sliding motion may also occur)	Very slow (mm/y to cm/y)
Slump	Thick deposits of unconsolidated sediment	Motion as a unit along a curved surface (rotational sliding)	Slow (cm/y to m/y)
Mudflow	Loose sediment with a significant component of silt and clay	Flow (a mixture of sediment and water moves down a channel)	Moderate to fast (cm/s to m/s)

Debris flow	Sand, gravel, and larger fragments	Flow (similar to a mudflow, but typically faster)	Fast (m/s)

Rock Fall

A **rock fall** is when fragments or rock break off relatively easily from steep bedrock slopes, most commonly due to frost-wedging in areas where there are many freeze-thaw cycles per year. When hiking along a steep mountain trail on a cool morning, one might have heard the occasional fall of rock fragments onto a **talus slope**. This happens because the water between cracks freezes and expands overnight. When that same water thaws in the morning sun, the fragments that had been pushed beyond their limit by the ice fall to the slope below.

Physical Geography and Natural Disasters by R. Adam Dastrup

Rock Slide

A **rock slide** is the sliding motion of rock along a sloping surface. In most cases, the movement is parallel to a fracture, bedding, or metamorphic foliation plane, and it can range from very slow to moderately fast. The word **sackung** describes a slope's very slow motion of a block of rock (mm/y to cm/y).

Rock Avalanche

If a rock slides and then starts moving quickly (m/s), the rock is likely to break into many small pieces, and at that point, it turns into a **rock avalanche**, in which the large and small fragments of

461

rock move in a fluid manner supported by a cushion of air within and beneath the moving mass.

Figure 14.9 "Goodell Creek Debris Avalanche" is licensed under Public Domain.

Creep or Solifluction

The very slow, millimeters per year to centimeters per year, movement of soil or other unconsolidated material on a slope is known as creep. **Creep**, which generally only affects the upper several centimeters of loose material, is typically a very slow flow, but in some cases, sliding may occur. Creep can be facilitated by freezing and thawing because particles are lifted perpendicular to the surface by the growth of ice crystals within the soil, and then let down vertically by gravity when the ice melts. The same effect can be produced by frequent wetting and drying of the soil. In cold environments, **solifluction** is a more intense form of freeze-thaw-triggered creep.

462

Creep is most noticeable on moderate-to-steep slopes where trees, fence posts, or grave markers are consistently leaning in a downhill direction. In the case of trees, they try to correct their lean by growing upright, and this leads to a curved lower trunk known as a "pistol butt."

Figure 14.10 "Soil Creep" by Derek Harper is licensed under Creative Commons Share-Alike 2.0 Unported.

Slump is a type of slide (movement as a mass) that takes place within thick unconsolidated deposits (typically thicker than 10 m). Slumps involve movement along one or more curved failure surfaces, with downward motion near the top and outward motion toward the bottom. They are typically caused by an excess of water within these materials on a steep slope.

463

Mudflows and Debris Flows

When a mass of sediment becomes wholly saturated with water, the mass loses strength, to the extent that the grains are pushed apart, and it will flow, even on a gentle slope. This can happen during rapid spring snowmelt or heavy rains and is also relatively common during volcanic eruptions because of the rapid melting of snow and ice. (A mudflow or debris flow on a volcano or during a volcanic eruption is a lahar.) If the material involved is primarily sand-sized or smaller, it is known as a **mudflow**.

If the material involved is gravel-sized or larger, it is known as a **debris flow**. Because it takes more gravitational energy to move more massive particles, a debris flow typically forms in an area with steeper slopes and more water than a mudflow. In many cases, a debris flow takes place within a steep stream channel and is triggered by the collapse of bank material into the stream. This creates a temporary dam and a significant flow of water and debris when the dam breaks.

Video 14.3 Debris flow in San Bernardino mountains amid massive mudslide

https://www.youtube.com/watch?v=TfyBALcKKr4

Erosion

Erosion is a mechanical process, usually driven by water, gravity, wind, or ice that removes sediment from weathering. Liquid water is the principal agent of erosion. Erosion resistance is essential in the creation of distinctive geological features. This is well demonstrated in the cliffs of the Grand Canyon. The cliffs are made of rock left standing after less resistant materials have weathered and eroded. Rocks with different levels of erosion resistance also create unique-looking features called hoodoos in Bryce Canyon National Park and Goblin Valley State Park in Utah.

465

Figure 14.11 "Bryce Canyon" by R. Adam Dastrup, is licensed under Creative Commons Attribution-NoncCommercial-ShareAlike 4.0 International.

Erosion by Streams

Streams, any running water from a rivulet to a raging river, complete the hydrologic cycle by returning precipitation that falls on land to the oceans. Some of this water moves over the surface, and some moves through the ground as groundwater. Flowing water does the work of both erosion and deposition.

Flowing streams pick up and transport weathered materials by eroding sediments from their banks. Streams also carry ions and ionic compounds that dissolve easily in the water. Sediments are carried as the following loads: dissolved, suspended, and bed. A dissolved load is composed of ions in solution. These ions are usually carried in the water to the ocean.

Sediments carried as solids as the stream flows are called a **suspended load**. The stream's velocity determines the size of particles that can be carried within a load. Faster streams can carry larger particles. Streams that carry larger particles have greater competence. Streams with a steep gradient (slope) have a faster velocity and greater competence.

TRANSPORT BY WIND

The power of the wind to erode depends on particle size, wind strength, and whether the particles can be picked up. Wind is a crucial erosional force in arid than humid regions. Wind transports small particles, such as silt and clay, over great distances, even halfway across a continent or an entire ocean basin. Particles may be suspended for days. Wind more easily picks up particles on the ground that has been disturbed, such as a construction site or a

dune. Just like flowing water, wind transports particles as both bed load and suspended load. For wind, **bed load** is made of sand-sized particles, many of which move by **saltation**. The suspended load is tiny particles of silt and clay.

Wind Erosion

Wind is a more potent erosional force in arid regions than it is in humid regions. In humid areas, water and vegetation bind the soil, so it is harder to pick up. In arid regions, small particles are selectively picked up and transported. As they are removed, the ground surface gets lower and rockier, causing deflation. What is left is desert pavement, a surface covered by gravel-sized particles that are not easily moved by wind.

Figure 14.12 "Delicate Arch" by John Buie is licensed under Creative Commons Attribution 2.0 Generic.

Particles moved by the wind do the work of **abrasion**. As a grain strikes another grain or surface, it erodes that surface. Abrasion by wind may polish natural or human-made surfaces, such as buildings. Stones that have become polished and faceted due to abrasion by sand particles are called **ventifacts**.

Exposed rocks in desert areas often develop a dark brown to black coating called **desert varnish**. Wind transports clay-sized particles that chemically react with other substances at high temperatures. The coating is formed of iron and manganese oxides. Often petroglyphs are carved into the desert varnish by earlier civilizations in arid regions.

469

Wind Deposition

Deserts and seashores sometimes have **dunes**. **Beach dunes** have different compositions depending on their location, and are usually quartz because, in humid areas, other minerals weather into clays. In the tropics, dunes may be composed of calcium carbonate, which is typical. In deserts, **sand dunes** may be composed of a variety of minerals. There is little weathering, and so less stable minerals are left behind.

Sand dunes are usually very uniform in size and shape. Particles are sand-sized because more massive particles are too heavy for the

wind to transport by suspension. They are rounded since rounded grains roll more easily than angular grains.

Figure 14.14 "Sand Dunes in Namibia, Africa" is Free for Commercial Use by Pikrepo.

For dunes to form, there must be an abundant supply of sand and steady winds. A strong wind slows down, often over some obstacles, such as a rock or some vegetation, and drops its sand. As the wind moves up and over the obstacle, it increases in speed. It carries the sand grains up the gently sloping, upwind side of the dune by saltation. As the wind passes over the dune, its speed decreases — sand cascades down the crest, forming the slip face of the dune. The slip face is steep because it is at the angle of repose for dry sand, about 34 degrees.

Wind deposits dunes layer by layer. If the wind changes directions, cross-beds form. Cross beds are named for the way each layer is formed at an angle to the ground. The types of a dune that forms depend on the amount of sand available, the character and

direction of the wind, and the type of ground the sand is moving over.

Formation and Movement of Glaciers

Glaciers cover about 10 percent of the land surface near Earth's poles, and they are also found in high mountains. During the Ice Ages, glaciers covered as much as 30 percent of Earth. Around 600 to 800 million years ago, geologists think that almost all of the Earth was covered in snow and ice, called the **Snowball Theory**. Scientists use the evidence of erosion and deposition left by glaciers to do a kind of detective work to figure out where the ice once was and where it came from.

Video 14.4 How volcanoes froze the earth

https://www.youtube.com/watch?v=4ONwQV26L-k

Glaciers are solid ice that moves exceptionally slowly along the land surface. They erode and shape the underlying rocks. Glaciers also deposit sediments in characteristic landforms. The two types of glaciers are: continental and alpine. **Continental glaciers** are large ice sheets that cover relatively flat ground. These glaciers flow outward from where the most considerable amount of snow and ice accumulate. **Alpine** or **valley glaciers** flow downhill through mountains along existing valleys.

Figure 14.15 "Thwaites Glacier" is a continental glacier taken by NASA and is licensed under Public Domain.

Figure 14.16 "Perito Moreno Glacier" by Luca Galuzzi is an alpine glacier and is licensed under the Creative Commons Attribution-ShareAlike 2.5 Generic license.

Glacial Erosion

Glaciers erode the underlying rock by abrasion and plucking. Glacial meltwater seeps into cracks of the underlying rock, the water freezes and pushes pieces of rock outward. The rock is then plucked out and carried away by the flowing ice of the moving glacier. With the weight of the ice over them, these rocks can scratch deeply into the underlying bedrock making long, parallel grooves in the bedrock, called **glacial striations**.

Mountain glaciers leave behind unique erosional features. When a glacier cuts through a 'V' shaped river valley, the glacier pucks rocks from the sides and bottom. This widens the valley and steepens the walls, making a 'U' shaped valley.

Smaller **tributary glaciers**, like tributary streams, flow into the main glacier in their own shallower 'U' shaped valleys. A **hanging valley** forms where the main glacier cuts off a tributary glacier and creates a cliff. Streams plunge over the cliff to create waterfalls. Up high on a mountain, where a glacier originates, rocks are pulled away from valley walls.

Figure 14.17 "U-Shaped Valley" by Dan Hobley is licensed under the Creative Commons Attribution-ShareAlike 3.0 Unported license.

Depositional Features of Glaciers
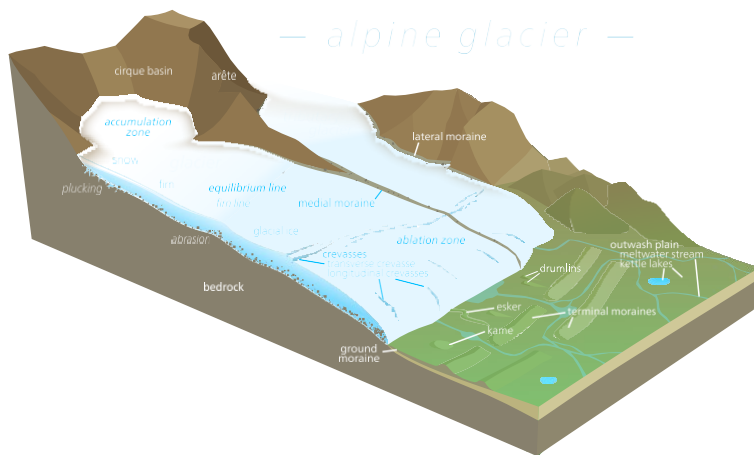
As glaciers flow, mechanical weathering loosens rock on the valley walls, which falls as debris on the glacier. Glaciers can carry rocks of any size, from giant boulders to silt. These rocks can be carried for many miles over many years and decades. These rocks that are different in type or origin from the surrounding bedrock are glacial erratics. Melting glaciers deposit all the big and small bits of solid material they are carrying in a pile. These unsorted deposits of rock are called **glacial till**.

Avalanches

Video 14.5 Avalanche types and their trigger mechanisms
https://www.youtube.com/watch?v=ZFWII5bAlQI

Creative Commons Attribution

An Introduction to Geology by Chris Johnson, Matthew D. Affolter, Paul Inkenbrandt, Cam Mosher is licensed under CC BY- NC-SA 4.0

Natural Disasters and Human Impacts by R. Adam Dastrup, MA, GISP is licensed under CC BY-NC-SA 4.0

Physical Geology – 2nd Edition by Steven Earle is licensed under CC BY-NC-SA 4.0

Geology by Lumen Learning is licensed under CC BY-NC-SA 4.0

Earth Science by Lumen Learning is licensed under CC BY-NC-SA 4.0

Chapter 15 Landforms

Stress and Strain

Stress is the force exerted per unit area, and strain is a material's response to that force. **Strain** is deformation caused by stress. Strain in rocks can be represented as a change in rock volume and rock shape, as well as fracturing the rock. There are three types of stress: tensional, compressional, and shear. **Tensional stress** involves pulling something apart in opposite directions, stretching and thinning the material. **Compressional stress** involves things coming together and pushing on each other, thickening the material. **Shear stress** involves transverse movement of the material moving past each other, like a scissor. (9 Crustal Deformation and Earthquakes – An Introduction to Geology, n.d.)

Deformation

When rocks are stressed, the resulting strain can be elastic, ductile, or brittle, called **deformation**. **Elastic deformation** is strain that is reversible after the stress is released. For example, when compressing a spring, it elastically returns to its original shape after releasing it.

Ductile deformation occurs when enough stress is applied to a material that the changes are permanent, and the material is no longer able to revert to its original shape. For example, if a spring is stretched too far, it can be permanently bent out of shape. Note that concepts related to ductile deformation apply at the visible (macro) scale, and deformation is more complicated at a microscopic scale. Research of plastic deformation, which touches on the atomic scale, is beyond the scope of introductory texts. The yield point is the amount of strain at which elastic deformation is surpassed, and permanent deformation is measurable. Brittle deformation is when the material undergoes another critical point of no return. When sufficient stress to pass that point occurs, it fails and fractures. (9 Crustal Deformation and Earthquakes – An Introduction to Geology, n.d.)
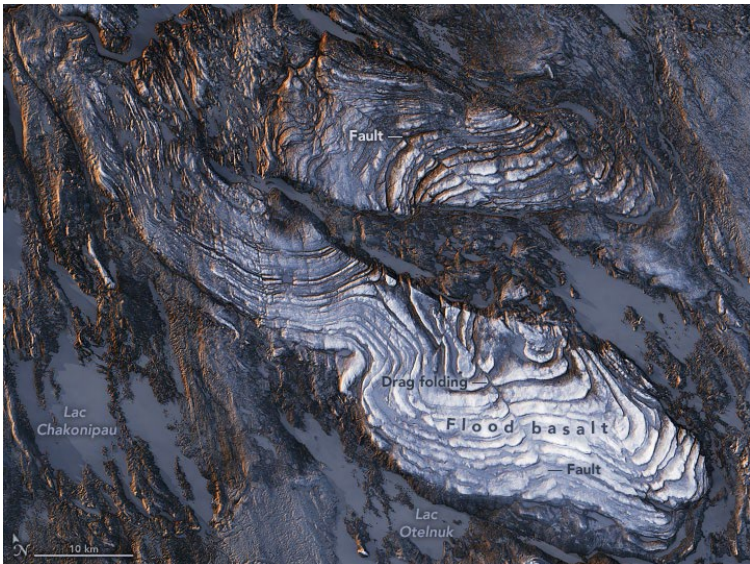
Figure 15.1 "Deformation" by NASA's Earth Observatory is licensed under Public Domain.

Important factors that influence how a rock will undergo elastic, ductile, or brittle deformation is the intensity of the applied stress, time, temperature, confining pressure, pore pressure, strain rate, and rock strength. **Pore pressure** is the pressure exerted by fluids inside of the open spaces (pores) inside of a rock or sediment. **Strain rate** is how quickly material is deformed. **Rock strength** is a measure of how readily a rock will respond to stress. Shale has low strength, and granite has high strength.

Removing heat, such as decreasing temperature, makes the material more rigid. Likewise, heating materials make them more ductile. Heating glass makes it capable of bending and stretching. Regarding strain response, it is easier to bend a piece of wood slowly without breaking it.

Sedimentary rocks are essential for deciphering the geologic history of a region because they follow specific rules. First, sedimentary rocks are formed with the oldest layers on the bottom and the youngest on top. Second, sediments are deposited horizontally, so sedimentary rock layers are originally horizontal, as are some volcanic rocks, such as ash falls. Finally, sedimentary rock layers that are not horizontal are deformed in some manner – often looking like they are tiling into the earth. Scientists can trace the deformation a rock has experienced by seeing how it differs from its original horizontal, oldest-on-bottom position. This deformation produces geologic structures such as folds, joints, and faults that are caused by stresses.

Stress and Mountain Building

The sheer power and strength of two or more converging continental plates smash upwards that create mountain ranges. Stresses from geologic uplift cause folds, reverse faults, and thrust faults, which allow the crust to rise upwards. The subduction of oceanic lithosphere at convergent plate boundaries also builds mountain ranges. When tensional stresses pull crust apart, it breaks into blocks that slide up and drop down along normal faults. The result is alternating mountains and valleys, known as a basin-and-range.

Figure 15.2 "Folding" is licensed under Creative Commons Attribution-ShareAlike 4.0 International.

Folds

Geologic folds are layers of rock that are curved or bent by **ductile deformation**. Terms involved with folds include axis, which is the line along which the bending occurred, and limbs, which are the dipping beds that make up the sides of the folds. Compressional forces most commonly form folds at depth, where hotter

482

temperatures and higher confining pressures allow ductile deformation to occur.

Folds are described by the orientation of their axes, axial planes, and limbs. They are made up of two or more dipping beds, dipping in opposite directions, which come together along a line, called the axis. Each set of dipping beds is known as a **fold limb**. The plane that splits the fold into two halves is known as the **axial plane**. (9 Crustal Deformation and Earthquakes – An Introduction to Geology, n.d.)

Symmetrical folds have mirrored limbs across their axial planes. The limbs of a symmetrical fold are inclined at the same, but opposite, angle indicating equal compression on both sides of the fold. **Asymmetrical folds** have dipping, non-vertical axial planes, where limbs dip into the ground at different angles. **Recumbent folds** are very tight folds with limbs compressed near the axial planes and are generally horizontal, and overturned folds are where the angles on both limbs dip in the same direction. The fold axis is where the axial plane intersects the strata involved in the fold. A horizontal fold has a horizontal fold axis. When the axis of the fold plunges into the ground, the fold is called a **plunging fold**.

483

Anticline

Anticlines are arch-like ("A"-shaped) folds, with downward
curving limbs that have beds that dip away from the central axis of
the fold. They are convex-upward in shape. In anticlines, the oldest
rock strata are in the center of the fold along the axis, and the
younger beds are on the outside. An antiform has the same shape as
an anticline, but in antiforms, the relative ages of the beds in the
fold cannot be determined. Oil geologists have an interest in
anticlines because they can form oil traps, where oil migrates up
along the limbs of the fold and accumulates in the high point along
the axis of the fold. (9 Crustal Deformation and Earthquakes – An
Introduction to Geology, n.d.)
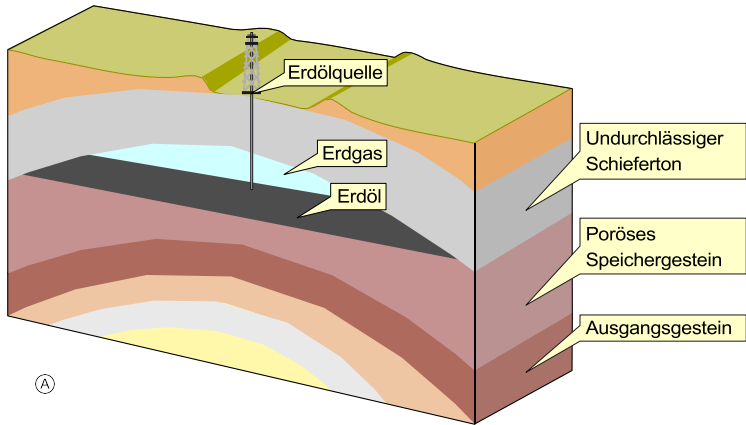
Figure 15.4 "Anticline Trap" is licensed under the
Creative Commons Attribution-Share Alike 3.0
Unported.

Syncline

Synclines are trough-like ("U" shaped), upward-curving folds that
have beds that dip in towards the fold's central axis. They are
concave-upward in shape. In synclines, the older rock is on the
outside of the fold, and the youngest rock is on the inside of the
fold along the axis. A synform has the shape of a syncline but, like
an antiform, does not distinguish between the ages of the units.

ANTICLINE

CENTER OR AXIS

SYNCLINE
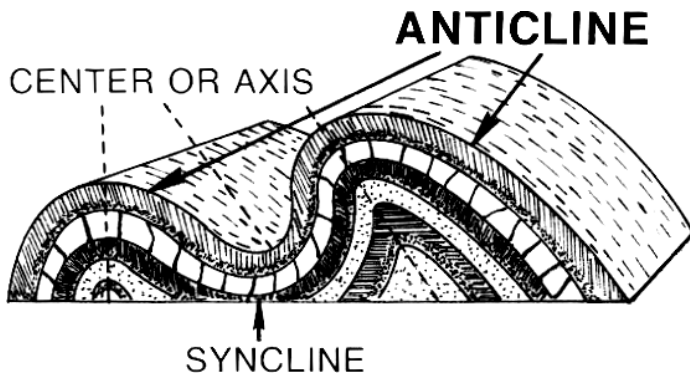
Figure 15.5 "Anticline" by Pearson Scott Foresman is licensed under Public Domain.

Monocline

Monoclines are step-like folds, in which flat rocks are upwarped or downwarped, then continue flat. They are relatively common on the Colorado Plateau, where they form "reefs," ridges that act as topographic barriers and should not be confused with ocean reefs. Capitol Reef National Park is an example of a monocline in Utah. Monoclines can be caused by bending of shallower sedimentary strata as faults grow below them. These faults are commonly called "blind faults" because they end before reaching the surface and can be either normal or reverse faults. (9 Crustal Deformation and Earthquakes – An Introduction to Geology, n.d.)

Dome

A **dome** is a symmetrical to semi-symmetrical upwarping of rock beds, like in Utah's San Rafael Swell. Domes have a shape like an

486

inverted bowl, similar to domes on buildings, like the Capitol Building. Some domes are formed from compressional forces, while other domes are formed from underlying igneous intrusions, by salt diapirs, or even impacts, like upheaval dome in Canyonlands National Park. (9 Crustal Deformation and Earthquakes – An Introduction to Geology, n.d.)

Basin

A **basin** is the inverse of a dome. The basin is when rock forms a bowl-shaped depression. The Uinta Basin is an example of a basin in Utah. Technically, geologists refer to rocks folded into a bowl- shape as structural basins. Sometimes structural basins can also be sedimentary basins in which large quantities of sediment accumulate over time. Sedimentary basins can form as a result of folding but are much more commonly produced in mountain building, between mountain blocks or via faulting. Regardless of the cause, as the basin sinks, called **subsidence**, it can accumulate even more sediment as the sediment's weight causes more subsidence in a positive-feedback loop. There are active sedimentary basins all over the world.

For mountain types refer to video 1

Video 15.1 Types of mountains and how they are formed
https://www.youtube.com/watch?v=UaaWoqlv9no

Creative Commons Attribution

Earth Science by Lumen Learning is licensed under CC BY-NC-SA 4.0

488

Chapter 16
Surface WATER

Water is composed of two atoms of hydrogen and one atom of oxygen bonded together. Despite its simplicity, water has remarkable properties. Water expands when it freezes and has high surface tension because of the molecules' polar nature that they tend to stick together. Without water, life might not exist on Earth, and it certainly would not have the tremendous complexity and diversity that we see.

Where is Earth's Water?
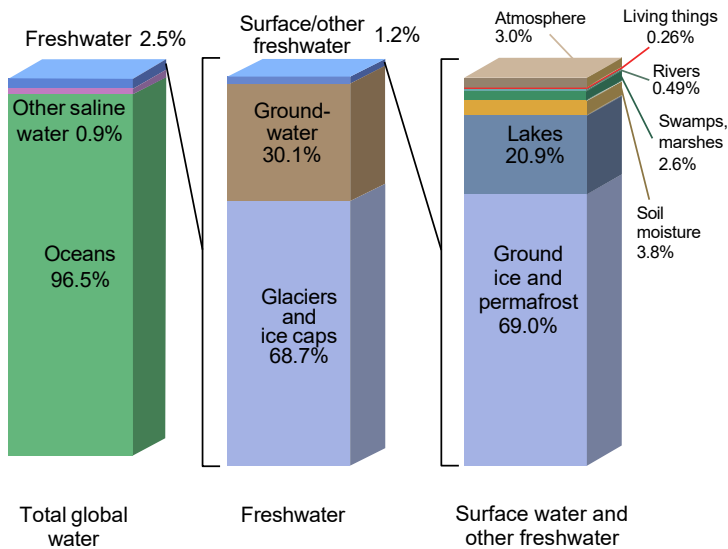
Earth's oceans contain 97 percent of the planet's water, and just 3 percent is freshwater with relatively low concentrations of salts. Most freshwater is trapped as ice in the vast glaciers and ice sheets of Greenland and Antarctica. A storage location for water such as an ocean, glacier, pond, or even the atmosphere is known as a **reservoir**. A water molecule may pass through a reservoir very quickly or may remain for much longer. The amount of time a molecule stays in a reservoir is known as its **residence time**.

Hydrologic Cycle

Because of the unique properties of water, water molecules can cycle through almost any-where on Earth. The water molecule found in a glass of water today could have erupted from a volcano early in Earth history. In the intervening billions of years, the molecule probably spent time in a glacier or far below the ground. The molecule surely was high up in the atmosphere and maybe deep in the belly of a dinosaur.

Water is the only substance on Earth that is present in all three states of matter – as a solid, liquid, or gas. Along with that, Earth is the only planet where water is present in all three states. Because of the ranges in temperature in specific locations around the planet, all three phases may be present in a specific location or region. The three phases are solid (ice or snow), liquid (water), and gas (water vapor).
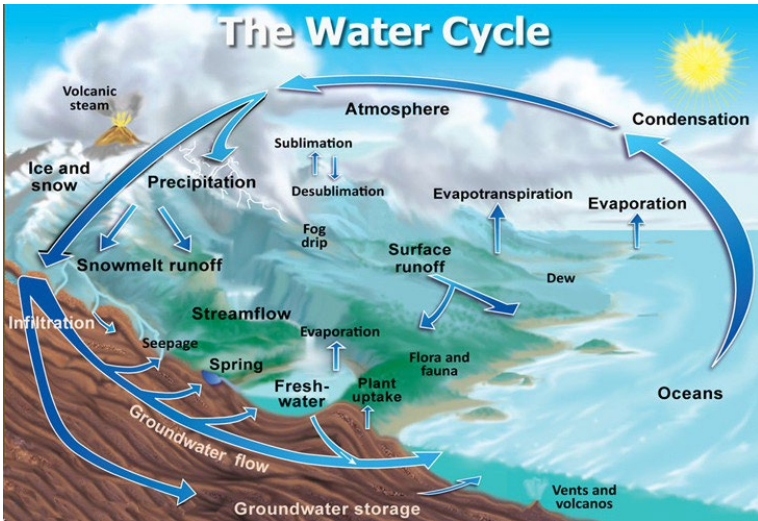
Figure 16.2 "The Water Cycle" by the United States Geologic Survey is licensed under Public Domain.

Water is continuously on the move. It is **evaporated** from the oceans, lakes, streams, the surface of the land, and plants (**transpiration**) by solar energy. It is moved through the atmosphere by winds and condenses to form clouds of water droplets or ice crystals. It comes back down as rain or snow and flows through streams, into lakes, and eventually back to the ocean. Water on the surface and in streams and lakes infiltrates the ground to become groundwater. Groundwater slowly moves through the rock and surficial materials. Some groundwater returns to other streams and lakes, and some go directly back to the oceans. (Earle, 2019)

Because Earth's water is present in all three states, it can get into various environments around the planet. The movement of water around the Earth's surface is the hydrologic (water) cycle. Water

changes from a liquid to a gas by evaporation to become water vapor. The Sun's energy can evaporate water from the ocean surface or lakes, streams, or puddles on land. Only the water molecules evaporate; the salts remain in the ocean or a freshwater reservoir. The water vapor stays in the atmosphere until it undergoes condensation to become tiny droplets of liquid. The droplets gather in clouds, which are blown about the globe by the wind. As the water droplets in the clouds collide and grow, they fall from the sky as precipitation. **Precipitation** can be rain, sleet, hail, or snow. Sometimes precipitation falls back into the ocean, and sometimes it falls onto the land surface.

When water falls from the sky as rain, it may enter streams and rivers that flow down to oceans and lakes. Water that falls as snow may sit on a mountain for several months. Snow may become part of the ice in a glacier, where it may remain for hundreds or thousands of years. Snow and ice may go directly back into the air by **sublimation**, the process in which a solid changes directly into a gas without first becoming a liquid. Although it is hard to see water vapor sublimate from a glacier, it is possible to see dry ice sublimate in the air.

Video 16.1 NASA | Earth's Water Cycle video

https://www.youtube.com/watch?v=oaDkph9yQBs

Snow and ice slowly melt over time to become liquid water, which provides a steady flow of freshwater to streams, rivers, and lakes below. A water droplet falling as rain could also become part of a stream or a lake. At the surface, the water may eventually evaporate and reenter the atmosphere.
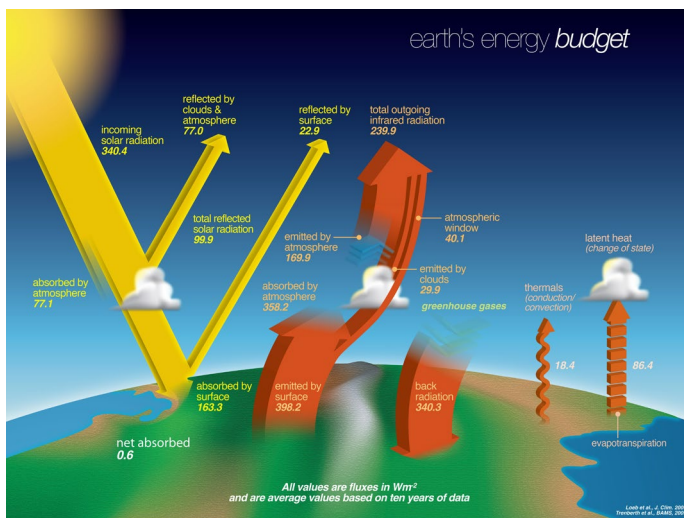
Figure 16.3 "Earth's Energy Budget" from NASA is licensed under Public Domain.

A significant amount of water infiltrates into the ground, and soil moisture is an important reservoir for that water. Water trapped in the soil is essential for plants to grow. Water may seep through dirt and rock below the soil through pores infiltrating the ground to go into Earth's groundwater system. Groundwater may enter aquifers that may store freshwater for centuries. Alternatively, the water may come to the surface through springs or find its way back to the oceans. Plants and animals depend on water to live, and they also play a role in the water cycle. Plants take water from the soil and release substantial amounts of water vapor into the air through their leaves, a process known as transpiration. NASA has an excellent online animation of the hydrologic cycle.

495

Video 16.2 Water Cycle | How the Hydrologic Cycle Works video

https://www.youtube.com/watch?v=al-do-HGuIk

People also depend on water as a natural resource. Not content to get water directly from streams or ponds, humans create canals, aqueducts, dams, and wells to collect water and direct it to where they want it.

Physical Geography and Natural Disasters by R. Adam Dastrup

Use	United States	Global
Agriculture	34 percent	70 percent
Domestic (drinking, cooking, bath, etc.)	12 percent	10 percent
Industry	5 percent	20 percent
Power Plant Cooling	49 percent	small

The table above displays water use in the United States and globally (Estimated Use of Water in the United States in 2005, USGS). It is important to note that water molecules cycle around. If climate cools and glaciers and ice caps grow, there is less water for the oceans, and sea level will fall. The reverse can also happen.

Rivers

Rivers are the largest types of stream, moving substantial amounts of water from higher to lower elevations. The Amazon River, the planet's most enormous river, has a flow rate of nearly 220,000 cubic meters per second! People have used rivers since the beginning

of civilization as a source of water, food, transportation, defense, power, recreation, and waste disposal.

497

WETLANDS

Wetlands are lands that are wet for significant periods. They are common where water and land meet. Wetlands can be large flat areas or small and steep areas. Wetlands are vibrant and unique ecosystems with many species that rely on both the land and the water for survival. Only specialized plants can grow in these conditions.

Wetlands tend to have a great deal of biological diversity. Wetland ecosystems can also be fragile systems sensitive to the amounts and quality of water present within them. Learn more about wetlands further from the U.S. Environmental Protection Agency.

Figure 16.5 "Sunrise at Viru Bog, Estonia" is licensed under the Creative Commons Attribution-ShareAlike 3.0 Unported license.

Types of Wetlands

There are a variety of distinct types of wetlands. **Marshes** are

shallow wetlands around lakes, streams, or the ocean where grasses and reeds are common, but trees are not. Frogs, turtles, muskrats, and many varieties of birds are at home in marshes.

Figure 16.6 "Bride Book Salt Marsh" is licensed under the Creative Commons Attribution-ShareAlike 3.0 Unported license.

A **swamp** is a wetland with lush trees and vines found in a low-lying area beside slow-moving rivers. Like marshes, they are often or always inundated with water. Since the water in a swamp moves slowly, oxygen in the water is often scarce. Swamp plants and animals must be adapted for these low-oxygen conditions. Like marshes, swamps can be freshwater, saltwater, or a mixture of both.

Figure 16.7 "National Wildlife Refuge, Georgia" by the United States Fish and Wildlife Service is licensed as Public Domain.

In an **estuary**, salt water from the sea mixes with fresh water from a stream or river. These semi-enclosed areas are home to plants and animals that can tolerate the sharp changes in salt content that the constant motion and mixing of waters create. Estuaries contain brackish water, water that has more salt than freshwater but less than seawater. Because of the rapid changes in salt content, estuaries have many different habitats for plants, animals, and extremely high biodiversity.

501

Ecological Role of Wetlands

As mentioned above, wetlands are home to many varied species of organisms. Although they make up only 5 percent of the United States area, wetlands contain more than 30 percent of the plant types. Many endangered species live in wetlands, so wetlands are protected from human use. Wetlands also offer protection from storm surges from tropical storm systems like hurricanes and supply safe, hidden protection for hatchlings.

Wetlands also play a critical biological role by removing pollutants from water. For example, they can trap and use fertilizer that has washed off a farmer's field, and therefore they prevent that fertilizer from contaminating another body of water. Since wetlands

Physical Geography and Natural Disasters by R. Adam Dastrup

naturally purify water, preserving wetlands also help to maintain clean supplies of water.

Ponds and Lakes

Ponds and lakes are bordered by hills or low rises so that the water is blocked from flowing directly downhill. **Ponds** are small bodies of freshwater that usually have no outlet; ponds are often are fed by underground springs. **Lakes** are more substantial bodies of water. Lakes are usually freshwater, although the Great Salt Lake in Utah is just one exception. Water usually drains out of a lake through a river or a stream, and all lakes lose water to evaporation.

Large lakes have tidal systems and currents, and can even affect weather patterns. The Great Lakes in the United States contain 22 percent of the world's fresh surface water. The largest them, Lake Superior, has a tide that rises and falls several centimeters each day. The Great Lakes are large enough to alter the Northeastern United States' weather system by the "lake effect," which is an increase in snow downwind of the relatively warm lakes. The Great Lakes are home to countless species of fish and wildlife.

Lakes form in various ways: in depressions carved by glaciers, in calderas, and along tectonic faults, to name a few. Subglacial lakes are even found below an ice cap. As a result of geologic history and landmasses' arrangement, most lakes are in the Northern Hemisphere.

Limnology is the study of bodies of freshwater and the organisms

that live there. The ecosystem of a lake is divided into three distinct sections:

- **Surface (littoral) zone** is the sloped area closest to the edge of the water.
- **Open-water zone** (also the photic or limnetic zone) has abundant sunlight.
- **Deep-water zone** (also the aphotic or profundal zone) has little or no sunlight.

There are several life zones found within a lake. In the littoral zone, sunlight promotes plant growth, which provides food and shelter to animals such as snails, insects, and fish. In the open-water zone, other plants and fish, such as bass and trout, live. The deep-water zone does not have photosynthesis since there is no sunlight. Most deep-water organisms are scavengers, such as crabs and catfish that feed on dead organisms that fall to the bottom of the lake. Fungi and bacteria aid in the decomposition in the deep zone. Though different creatures live in the oceans, ocean waters also have these same divisions based on sunlight with similar types of creatures that live in each of the zones.

Lakes are not permanent features of a landscape. Some come and go with the seasons, as water levels rise and fall. Over a longer time, lakes disappear when they fill with sediments, if the springs or streams that fill them diminish, or if their outlets grow because of erosion. When the climate of an area changes, lakes can either expand or shrink. Lakes may disappear if precipitation significantly diminishes.

Chapter 16
FLOODS

Floods, an overflow of water in one place, are a natural part of the water cycle, but they can be terrifying forces of destruction. Floods can occur for a variety of reasons, and their effects can be minimized in several ways. Perhaps unsurprisingly, floods tend to affect low-lying areas most severely. Floods usually occur when precipitation falls more quickly than that water can be absorbed into the ground or carried away by rivers or streams. Waters may build up gradually throughout weeks when an extended period of rainfall or snowmelt fills the ground with water and raises stream levels.

Flash floods are sudden and unexpected, taking place when very heavy rains fall over a very brief period. A flash flood may do its damage miles from where the rain falls if the water travels far down a dry streambed so that the flash flood occurs far from the location of the original storm.

Densely vegetated lands are less likely to experience flooding. Plants slow down water as it runs over the land, giving it time to enter the ground. Even if the ground is too wet to absorb more water, plants still slow the water's passage and increase the time between rainfall and the water's arrival in a stream; this could keep all the water falling over a region to hit the stream at once. Wetlands function as a buffer between land and high-water levels and play a key role in minimizing the impacts of floods. Flooding is often more severe in areas that have been recently logged.

When a dam breaks along a reservoir, flooding can be catastrophic. High water levels have also caused small dams to break, wreaking havoc downstream. People try to protect areas that might flood with dams, and dams are remarkably effective. People may also line a riverbank with **levees**, high walls that keep the stream within its banks during floods. A levee in one location may force the high water up or downstream and cause flooding there. The New Madrid Overflow was created with the recognition that the Mississippi River sometimes cannot be contained by levees and must be allowed to flood.

Video 16.4 How "levee wars" are making floods worse
https://www.youtube.com/watch?v=LTv6RkFnelM

Not all the consequences of flooding are negative. Rivers deposit new nutrient-rich sediments when they flood, and so floodplains have traditionally been suitable for farming. Flooding as a source of nutrients was essential to Egyptians along the Nile River until the Aswan Dam was built in the 1960s. Although the dam protects crops and settlements from the annual floods, farmers must now use fertilizers to feed their crops.

Floods are also responsible for moving substantial amounts of sediments about within streams. These sediments provide habitats for animals, and the periodic movement of sediment is crucial to the lives of several types of organisms. Plants and fish along the

508

Colorado River, for example, depend on seasonal flooding to rearrange sand bars.

GROUNDWATER

Groundwater is stored in the open spaces within rocks and within unconsolidated sediments. Rocks and sediments near the surface are under less pressure than those at significant depth and therefore tend to have more open space. For this reason, and because it is expensive to drill deep wells, most of the groundwater that is accessed by individual users is within the first 100 m of the surface. Some municipal, agricultural, and industrial groundwater users get their water from greater depth, but deeper groundwater tends to be of lower quality than shallow groundwater, so there is a limit to how deep we can go.
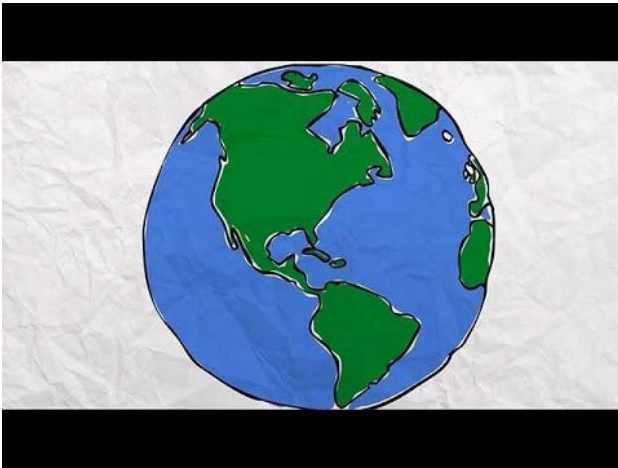
Porosity is the percentage of open space within unconsolidated sediment or rock. **Primary porosity** is represented by the spaces between grains in a sediment or sedimentary rock. **Secondary porosity** is porosity that has developed after the rock has formed. It can include **fracture porosity**, space within fractures in any rock. Some volcanic rock has a particular type of porosity related to vesicles, and some limestone has increased porosity related to cavities within fossils.
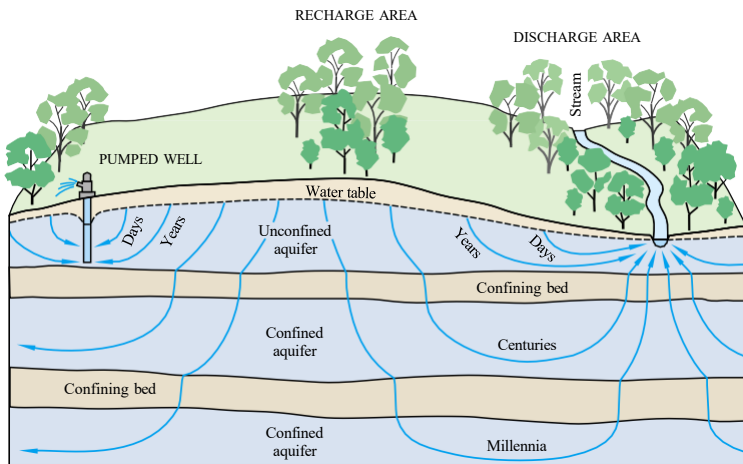
Figure 16.9 "Groundwater Flow" by the United States Geologic Survey is licensed under Public Domain.

Porosity is expressed as a percentage calculated from the volume of open space in a rock compared with the total volume of rock. Unconsolidated sediments tend to have higher porosity than consolidated ones because they have no cement, and most have not been strongly compressed. Finer-grained materials (e.g., silt and clay) tend to have greater porosity, some as high as 70 percent, than coarser materials (e.g., gravel). Primary porosity tends to be higher in well-sorted sediments compared to poorly sorted sediments, where there is a range of smaller particles to fill the spaces made by the larger particles. **Glacial till**, which has a wide range of grain sizes and is typically formed under compression beneath glacial ice, has relatively low porosity.

Figure 16.10 "Terminus of the Fox Glacier, New Zealand" by Paul Morris is licensed under the Creative Commons Attribution-ShareAlike 2.0 Generic license.

Consolidation and cementation during the process of lithification of unconsolidated sediments into sedimentary rocks reduce primary porosity. Sedimentary rocks generally have porosities in the range of 10 percent to 30 percent, some of which may be secondary (fracture) porosity. The grain size, sorting, compaction, and degree of cementation of the rocks all primary influence porosity. For example, poorly sorted and well-cemented sandstone and well-compressed mudstone can have incredibly low porosity. Igneous or metamorphic rocks have the lowest primary porosity because they commonly form at depth and have interlocking crystals. Most of their porosity comes in the form of secondary porosity in fractures. Of the consolidated rocks, well-fractured volcanic rocks and limestone with cavernous openings produced by dissolution have the highest potential porosity, while intrusive igneous and metamorphic rocks, which formed under enormous pressure, have the lowest. Porosity is a measure of how much water can be stored in geological materials. Most rocks have some porosity and therefore contain groundwater. Groundwater is found under the ground and everywhere on the planet. Considering that sedimentary rocks and unconsolidated sediments cover about 75 percent of the continental crust with an average thickness of a few hundred meters and that they are likely to have around 20 percent porosity on average, it is easy to see that a considerable volume of water can be stored in the ground.

Porosity is a description of how much space there could be to hold water under the ground, and permeability describes how those pores are shaped and interconnected. This determines how easy it is for water to flow from one pore to the next. Larger pores mean there is less friction between flowing water and the sides of the

pores. Smaller pores mean more friction along pore walls, but also more twists and turns for the water to have to flow-through. A permeable material has a sizable number of well-connected pores spaces, while an impermeable material has fewer, smaller pores that are poorly connected. Permeability is the most crucial variable in groundwater. Permeability describes how easily water can flow through the rock or unconsolidated sediment and how easy it will be to extract it for our purposes. The characteristic of permeability of a geological material is quantified by geoscientists and engineers using some different units, but the most common is the hydraulic conductivity. The symbol used for hydraulic conductivity is K. Although hydraulic conductivity can be expressed in a range of different units, in this book, we will always use m/s. (Earle, 2015)

Unconsolidated materials are generally more permeable than the corresponding rocks (compare sand with sandstone, for example), and the coarser materials are much more permeable than the finer ones. The least permeable rocks are unfractured intrusive igneous and metamorphic rocks, followed by unfractured mudstone, sandstone, and limestone. The permeability of sandstone can vary widely depending on the degree of sorting and the amount of cement present. Fractured igneous and metamorphic rocks, especially fractured volcanic rocks, can be highly permeable, as can limestone dissolve along fractures and bedding planes to create solutional openings. Both sand and clay deposits (and sandstone and mudstone) are quite porous (30 percent to 50 percent for sand and 40 percent to 70 percent for silt and clay), but while sand can be quite permeable, clay and mudstone are not.

We have now seen a wide range of porosity in geological materials

and an even more extensive range of permeability. Groundwater exists everywhere there is porosity. However, whether groundwater can flow in significant quantities depends on the permeability. An **aquifer** is a body of rock or unconsolidated sediment with sufficient permeability to allow water to flow through it. Unconsolidated materials like gravel, sand, and even silt make relatively good aquifers, as do rocks like sandstone. Other rocks can be suitable aquifers if they are well fractured. An **aquitard** is a body that does not transmit a significant amount of water, such as clay, a till, or a poorly fractured igneous or metamorphic rock. These are relative terms, not absolute, and are usually defined based on someone's desire to pump groundwater; an aquifer to someone who does not need much water may be an aquitard to someone else who does. An aquifer that is exposed at the ground surface is called an **unconfined aquifer**. An aquifer where there is a lower permeability material between the aquifer and the ground surface is known as a **confined aquifer**, and the aquitard separating the ground surface, and the aquifer is known as the **confining layer**.

If a person were to go out into their garden or a forest or a park and start digging, they would find that the soil is moist (unless you are in a desert), but it is not saturated with water. This means that some of the pore space in the soil is occupied by water, and some of the pore space is occupied by air (unless you are in a swamp). This is known as the **unsaturated zone**. If a person were to dig down far enough, they would get to the point where all of the pore spaces are 100 percent filled with water (saturated), and the bottom of your hole would fill up with water. The level of water in the hole represents the water table, which is the surface of the **saturated zone**.

Water falling on the ground surface as precipitation (rain, snow, hail, fog, etc.) may flow off a hill slope directly to a stream in the

form of runoff, or it may infiltrate the ground, where it is stored in the unsaturated zone. The water in the unsaturated zone may be used by plants (transpiration), evaporate from the soil (evaporation), or continue past the root zone and flow down to the water table, where it recharges the groundwater.

In areas with topographic relief, the water table generally follows the land surface, but tends to come closer to surface in valleys, and intersects the surface where there are streams or lakes. The water table can be determined from the depth of water in a well that is not being pumped. Although, that only applies if the well is within an unconfined aquifer. In this case, most of the hillside forms the recharge area, where water from precipitation flows downward through the unsaturated zone to reach the water table. The area at the stream or lake to which the groundwater is flowing is a discharge area.

STREAMS AND RIVER SYTEMS

Freshwater in streams, ponds, and lakes is an essential part of the hydrologic cycle if only because of its importance to living creatures. Along with wetlands, these freshwater regions contain a tremendous variety of organisms. **Streams** are bodies of water that consist of constant motion, called a **current**. Geologists recognize many categories of streams depending on their size, depth, speed, and location. Creeks, brooks, tributaries, bayous, and rivers might all be lumped together as streams. In streams, water always flows downhill, but the form that downhill movement takes varies with rock type, topography, and many other factors.

Streams are the most important agents of erosion and transportation of sediments on Earth's surface. They are responsible for the creation of much of the topography that we see around us. They are also places of great beauty and tranquility, and of course, they supply much of the water that is essential to our existence. However, streams are not always peaceful and soothing. During large storms and rapid snowmelts, they can become raging torrents capable of moving cars and houses and destroying roads and bridges. When they spill over their banks, they can flood vast areas, devastating populations, and infrastructure. (11 Water – An Introduction to Geology, n.d.)

Parts of a Stream

There are a variety of distinct types of streams. A stream originates at its **source**, such as high mountains where snows collect in winter and melt in summer, or a source might be a **spring**. A stream may have more than one source, and when two streams come together, it is called a **confluence**. The smaller of the two streams is a **tributary** of the larger stream. A stream may create a pool where the water slows and becomes more profound.

Figure 16.12 "River Source" by Tony Ferrie is licensed under the Creative Commons Attribution-ShareAlike 2.0 Generic.
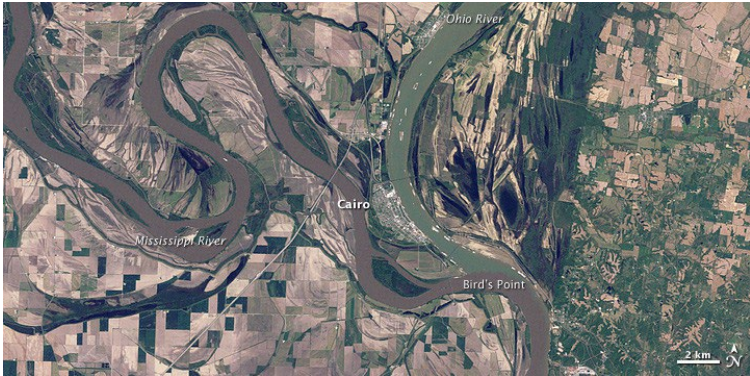
519

The point at which a stream comes into a large body of water, like an ocean or a lake, is called the **mouth**. Where the stream meets the ocean or lake, it is called an **estuary**. The mix of fresh and saltwater, where a river runs into the ocean, creates a diversity of environments where many different organisms create unique ecosystems.
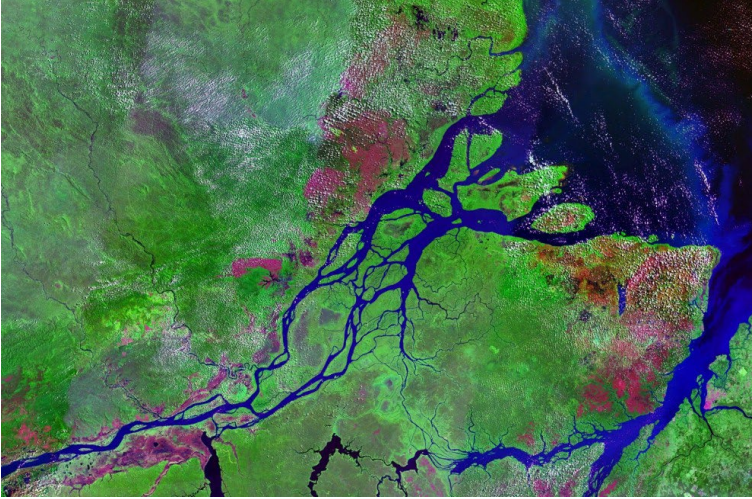
Physical Geography and Natural Disasters by R. Adam Dastrup

The Water Table

For a groundwater aquifer to hold the same amount of water, the amount of recharge must equal the amount of discharge. In wet regions, streams are fed by groundwater; the stream's surface is the top of the water table. In dry regions, water seeps down from the stream into the aquifer, often dry much of the year. Water leaves a groundwater reservoir in streams or springs, where people take water from aquifers.

Although groundwater levels do not rise and fall as rapidly as at the surface, over time, the water table will rise during wet periods and

fall during droughts. One of the most interesting but extremely atypical types of aquifers are found in Florida. Although aquifers are very rarely underground rivers, in Florida, water has dissolved the limestone so that streams travel underground and aboveground. (Components of Groundwater | Geology, n.d.)

Groundwater Use

Groundwater is a significant water source for people. Groundwater can be a renewable resource, as long as when the water pumped from the aquifer is replenished. It is essential for anyone who intends to dig a well to know how deep the water table is beneath the surface. Because groundwater involves interaction between the Earth and the water, the study of groundwater is called **hydrogeology**. Some aquifers are overused; people pump out more water than is replaced. As the water is pumped out, the water table slowly falls, requiring wells to be dug deeper, which takes more money and energy. Wells may go completely dry if they are not deep enough to reach into the lowered water table.

The Ogallala Aquifer supplies about one-third of the irrigation water in the United States. The aquifer is found from 30 to 100 meters deep over about 440,000 square kilometers! The water in the aquifer is mostly from the last ice age. People widely use the Ogallala Aquifer for municipal and agricultural needs. About eight times more water is taken from the Ogallala Aquifer each year than is replenished. Much of the water is used for irrigation of crops in the Breadbasket of the central plains. Currently, there is great concern about the long-term health of this vast aquifer because it is being tapped into and used at a higher rate than being replenished by natural processes. This could have huge implications regarding food production in the country if this critical water source is

depleted. At current rates of use, 70 percent of the aquifer could be gone by 2050.
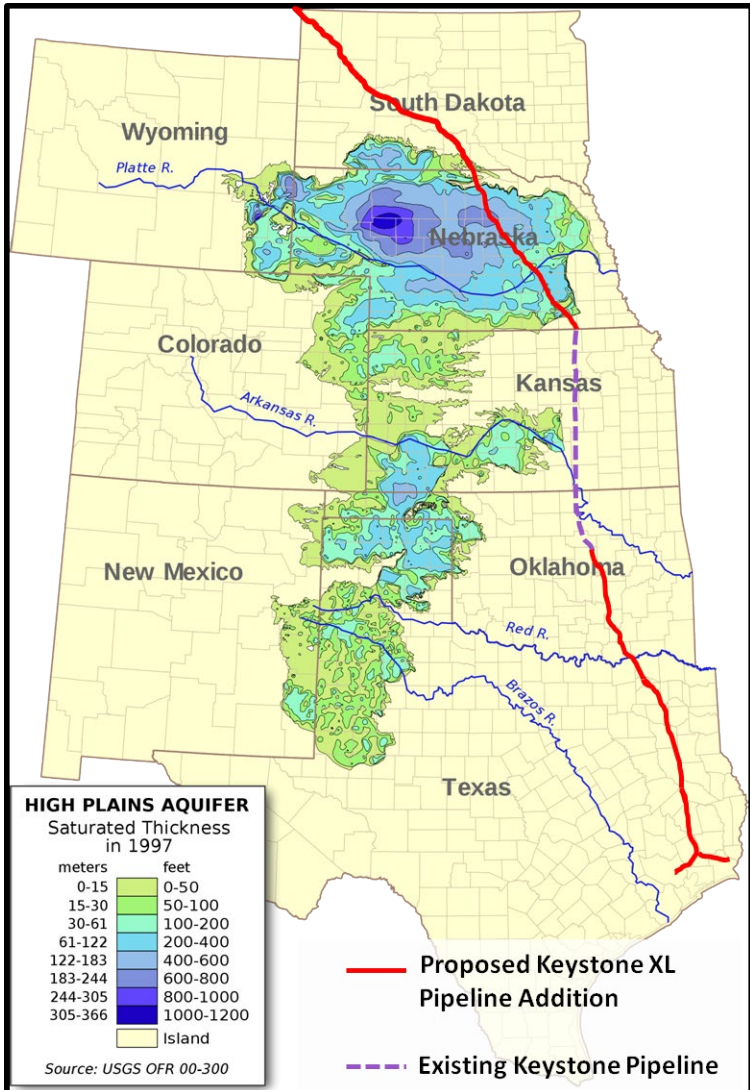
524

Overuse and lowering of the water tables of aquifers could have other impacts as well. Lowering the water table may cause the ground surface to sink. **Subsidence** may occur beneath houses and other structures. When coastal aquifers are overused, salt water from the ocean may enter the aquifer, contaminating the aquifer and making it less useful for drinking and irrigation. Saltwater incursion is a problem in developed coastal regions, such as in Hawaii.

Video 16.7 Why parched California land is collapsing at rapid rate

https://www.youtube.com/watch?v=TxqWUFyq-ZA&list=PLbN8RIyCRl35vq_yh5HhF4eAKYiYuaBF

Springs and Wells

Groundwater meets the surface in a stream or a spring. A spring

525

may be constant, or may only flow at certain times of the year. Towns in many locations depend on water from springs. Springs can also be a vital source of water in locations where surface water is scarce.

A well is created by digging or drilling to reach groundwater. When the water table is close to the surface, wells are a convenient method for extracting water. When the water table is far below the surface, specialized equipment must be used to dig a well. Most wells use motorized pumps to bring water to the surface, but some still require people to use a bucket to draw water up.

Finally, it is also essential to understand how water is cleaned, filtered, and delivered to our homes and work. Many of us do not know where our water comes from, and we take it for granted. This often leads to the wasteful water use of our lawns, showers, and other appliances.

Anthropogenic Contamination of Groundwater

Groundwater can become contaminated by pollution at the surface (or at depth), and there are many different anthropogenic (human-caused) sources of contamination. The vulnerability of aquifers to pollution depends on several factors, including the depth to the water table, the permeability of the material between the surface and the aquifer, the aquifer's permeability, the slope of the surface, and the amount of precipitation. Confined aquifers tend to be much less vulnerable than unconfined ones, and deeper aquifers are less vulnerable than shallow ones. Steeper slopes mean that surface water tends to run off rather than infiltrate (and this can reduce the

possibility of contamination). Contamination risk is also less in dry areas than in areas with heavy rainfall. (Earle, 2019)

The principal sources of anthropogenic groundwater contamination include the following:

- Chemicals and animal waste related to agriculture, and chemicals applied to golf courses and domestic gardens
- Agriculture
- Landfills
- Industrial operations

- Mines, quarries, and other rock excavations
- Leaking fuel storage tanks (especially those at gas stations)
- Septic systems
- Runoff from roads (e.g., winter salting) or chemical spills of materials being transported

Landfills

In the past, domestic and commercial refuse was commonly trucked to a "dump" (typically a hole in the ground), and when the hole was filled, it was covered with soil and forgotten. In situations like this, rain and melting snow can easily pass through the soil used to cover the refuse. This water passes into the waste itself, and the resulting landfill leachate that flows from the bottom of the landfill can seriously contaminate the surrounding groundwater and surface water. In the past few decades, regulations around refuse disposal have been significantly strengthened, and significant steps have been taken to reduce the amount of landfill waste by diverting recyclable and compostable materials to other locations.

A modern engineered landfill has an impermeable liner (typically heavy plastic, although engineered clay liners or natural clay may be adequate in some cases), a plumbing system for draining leachate (the rainwater that flows through the refuse and becomes contaminated), and a network of monitoring wells both within and around the landfill. Once a part or all of a landfill site is full, it is sealed over with a plastic cover, and a system is put in place to extract landfill gas (typically a mixture of carbon dioxide and methane). That gas can be sent to a nearby location where it is burned to create heat or used to generate electricity. The leachate

must be treated, and that can be done in a typical sewage treatment plant.

The monitoring wells are used to assess the water table level around the landfill and collect groundwater samples so that any leakage can be detected. Because some leakage is almost inevitable, the ideal placement for landfills is in areas where the depth to the water table is significant (tens of meters if possible) and where the aquifer material is relatively impermeable. Landfills should also be situated far from streams, lakes, or wetlands to avoid contamination of aquatic habitats.

Today, hundreds of abandoned dumps are scattered across the country; most have been left to contaminate groundwater that we might wish to use sometime in the future. In many cases, it is unlikely that we will be able to do so. (Earle, 2019)

Mines, Quarries, and Rock Excavations

Mines and other operations that involve the excavation of substantial amounts of rock (e.g., highway construction) have the potential to create severe environmental damage. The exposure of rock that has previously not been exposed to air and water can lead to the oxidation of sulfide-bearing minerals, such as pyrite, within the rock. The combination of pyrite, water, oxygen, and a particular type of bacteria (Acidithiobacillus ferrooxidans) that thrives in acidic conditions leads to the generation of acidity, in some cases, to pH less than 2. Water that acidic is hazardous by itself, but the low pH also increases the solubility of certain heavy metals. The water that is generated by this process is known as acid rock drainage (ARD). ARD can occur naturally where sulfide-

529

bearing rocks are near the surface. The issue of ARD is a primary environmental concern at both operating mines and abandoned mines. Groundwater adjacent to the contaminated streams in the area is very likely contaminated as well.

Leaking Fuel Tanks

Underground storage tanks (USTs) are used to store fuel at gas stations, industrial sites, airports, and anywhere that large volumes of fuel are used. They do not last forever, and eventually, they start to leak their contents into the ground. This is a problem at older gas stations, although it may also become a future problem at newer gas stations. Sometimes a gas station can be seen that is closed and surrounded by a chain-link fence. In virtually all such cases, the discovery of leaking USTs and the requirement has triggered the closure to cease operations and remediate the site.

Petroleum fuels are complex mixtures of hydrocarbon compounds and the properties of their components – such as density, viscosity, solubility in water, and volatility – tend to vary widely. As a result, a petroleum spill is like several spills for the price of one. The petroleum liquid slowly settles through the unsaturated zone and then tends to float on the surface of the groundwater. The more readily soluble components of the spill dissolve in the groundwater and are dispersed along with the normal groundwater flow, and the more volatile components of the spill rise toward the surface, potentially contaminating buildings. (Earle, 2019)

Creative Commons Attributions

An Introduction to Geology by Chris Johnson, Matthew D. Affolter, Paul Inkenbrandt, Cam Mosher is licensed under CC BY- NC-SA 4.0

Natural Disasters and Human Impacts by R. Adam Dastrup, MA, GISP is licensed under CC BY-NC-SA 4.0

Physical Geology – 2nd Edition by Steven Earle is licensed under CC BY-NC-SA 4.0

Geology by Lumen Learning is licensed under CC BY-NC-SA 4.0

Earth Science by Lumen Learning is licensed under CC BY-NC-SA 4.0

Chapter 17 Oceans SIGNIFICANCE OF OCEAN

The oceans makeup 70 percent of the planet and contain 97 percent of all the water on Earth. It also makes up most of the water stores the majority of the planet's moisture, terrestrial energy, and heat from the Sun. This energy is transferred between the equator and the two poles by larger surface currents by winds and deep ocean currents driven by ocean density differences. It also supplies the moisture and energy for storm systems and, ultimately, global climates.

Phytoplankton, microscopic plants, and animals in the oceans provide the foundation of the **global food web** of species. The earth's oceans are so vital for life that over 40 percent of the world's population lives near coastal areas. (Ch 4 Earth's Hydrosphere, n.d.)
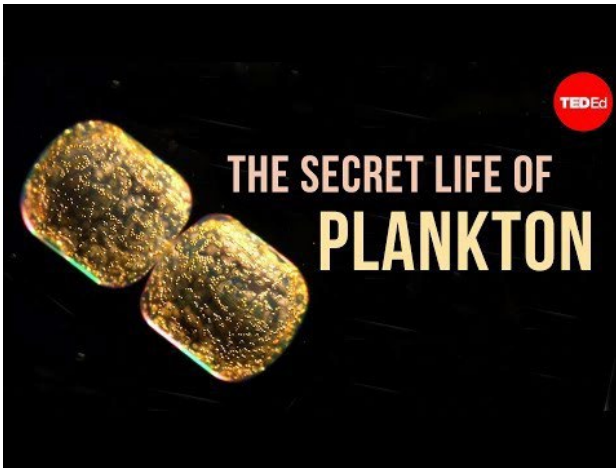
Video17. 2 The secret life of plankton video
https://www.youtube.com/watch?v=xFQ_fO2D7f0

https://www.youtube.com/watch?v=fS422O4SLc4

Moderate Climates

As terrestrial creatures, humans think of the importance of the planet's land surfaces, yet Earth is a planet consisting of 70 percent water. From space, the dominance of water is clear because most of it is stored in Earth's oceans. (Oceans and Coastal Environments | Earth Science, n.d.)

Earth would not be the same planet without its oceans. The oceans, along with the atmosphere, keep Earth's surface temperatures relatively constant worldwide. Some places on Earth reach as cold as -47 degrees Celsius, while other places reach as hot as 55 degrees

Celsius. On other planets like Mercury, temperatures range from -180 degrees Celsius to 430 degrees Celsius. (Significance of the Oceans | Physical Geography, n.d.)
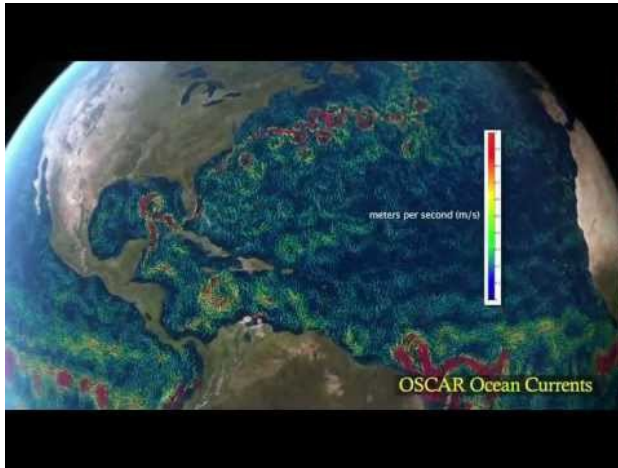
Video 17.4 The Ocean: A Driving Force for Weather and Climate

https://www.youtube.com/watch?v=6vgvTeuoDWY

The oceans, along with the atmosphere, distribute heat around the planet. The oceans absorb heat near the equator and then transport that solar energy to polar regions. The oceans also moderate the climate within a region. At the same latitude, the temperature range is smaller and coastal areas compared to areas farther inland. Along coastal areas, summer temperatures are not as hot, and winter temperatures are not as cold, because water takes a long time to heat up or cool down.

534

Biologically Rich

The oceans are an essential part of Earth's water cycle. Since they cover so much of the planet, most evaporation comes from the ocean, and most precipitation falls on the oceans.

Figure 17.1 "Coral Reef at the Andaman Islands" is licensed under the Creative Commons Attribution-ShareAlike 3.0 Unported license.

The oceans are also home to an enormous amount of life. That is, they have tremendous biodiversity. Tiny ocean plants create the base of a food web that supports all sorts of life forms. Marine life makes up most of all biomass on Earth. **Biomass** is the total mass of living organisms in a given area. These organisms supply us with food and even the oxygen created by marine plants.

535

Continental Margins

Recall from the chapter on Plate Tectonics that the ocean floor is not flat. Mid-ocean ridges, deep-sea trenches, and other features all rise sharply above or plunge deeply below the abyssal plains. Earth's tallest mountain is Mauna Kea volcano, which rises 10,203 m (33,476 ft.) from the Pacific Ocean floor to become one of the volcanic mountains of Hawaii. The deepest canyon is also on the ocean floor, the Challenger Deep in the Marianas Trench, 10,916 m (35,814 ft). The mapping of the ocean floor and coastal margins is called **bathymetry**. (Introduction to the Oceans | Earth Science, n.d.)
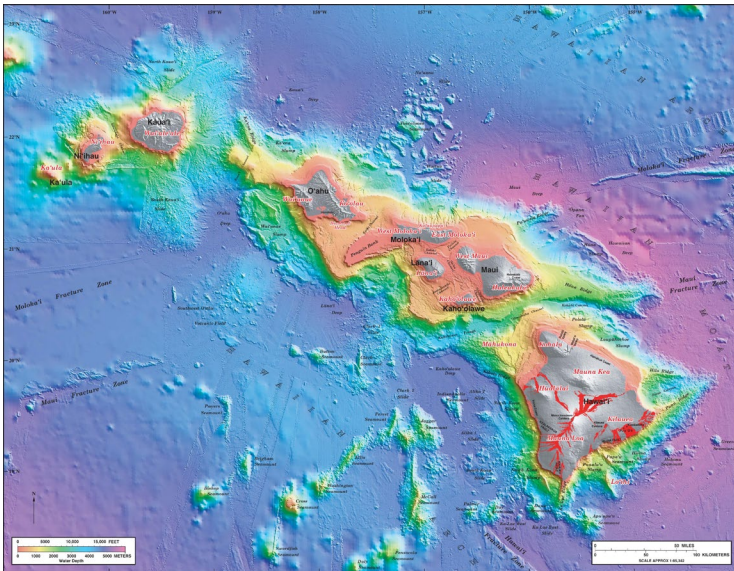
Figure 17.2 "Bathymetry of the Hawaiian Archipelago" by the USGS is licensed as Public Domain.

The **continental margin** is the transition from the land to the deep sea or, geologically speaking, from continental crust to oceanic crust. More than one-quarter of the ocean basin is the continental margin.

Composition of Ocean Water

The ocean's water is a complex system of organic and inorganic substances. Water is a polar molecule that can dissolve many substances such as salts, sugars, acids, bases, and organic molecules. Saltwater comes come water that moves through rock and soil on the land and picks up ions. This is the flip side of weathering. Salts comprise about 3.5 percent of the mass of ocean water, but the salt content or **salinity** is different in various locations.

Legend:
- Chlorine — 55 %
- Sodium — 31 %
- Magnesium — 7.7 %
- Sulfur — 3.7%
- Calcium — 1.2%
- Other — 1.4%
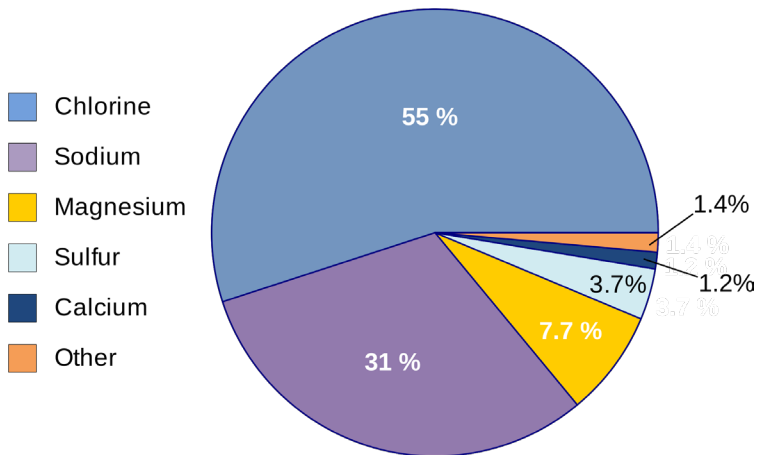
Figure 17.3 "Composition of Ocean Water" is licensed under the Creative Commons Attribution-ShareAlike 3.0 Unported license.

In places like estuaries, seawater mixes with freshwater, causing salinity to be much lower than average. Where there is lots of evaporation, but minimal circulation of water, salinity can be much higher. The Dead Sea has 30 percent salinity, nearly nine times the average salinity of ocean water. It is called the Dead Sea because nothing can survive within it because of its salinity. (Significance of the Oceans | Physical Geography, n.d.)

Layers of the Ocean

In 1960, two scientists in a specially designed submarine called the **Trieste** descended into a submarine trench called the Challenger Deep (10,910 meters). The average depth of the ocean is 3,790 m, a lot shallower than the deep trenches but still an incredible depth for sea creatures to live in. Three significant factors make the deep ocean hard to inhabit: the absence of light, low temperature, and extremely high pressure. The National Weather Service as information on the layers of the ocean. (Layers of the Ocean | Physical Geography, n.d.)

Vertical Divisions

To better understand regions of the ocean, scientists define the **water column** by depth. They divide the entire ocean into two zones vertically, based on the light level. Large lakes are divided into similar regions. Sunlight only penetrates the sea surface to a depth of about 200 m, creating the **photic zone** (consisting of the Sunlight Zone and Twilight Zone). Organisms that photosynthesize depend on sunlight for food and so are restricted

to the photic zone. Since tiny photosynthetic organisms, known as **phytoplankton**, supply nearly all of the energy and nutrients to the rest of the marine food web, most other marine organisms live in or at least visit the photic zone. In the aphotic zone (consisting of the Midnight Zone and the Abyss), there is not enough light for photosynthesis. The **aphotic zone** makes up most of the ocean, but has a small amount of its life, both in the diversity of type and numbers. (Introduction to the Oceans | Earth Science, n.d.)
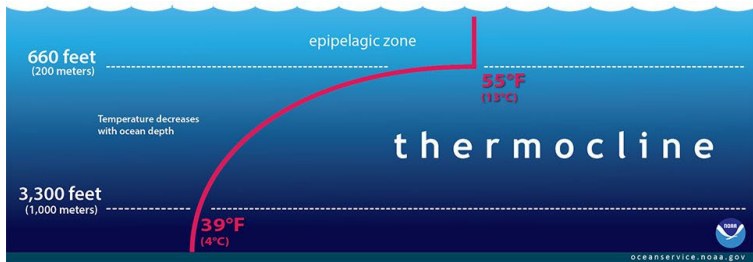
Figure 17.4 "Thermocline" by NOAA is licensed as Public Domain.

Horizontal Divisions

The seabed is also divided into the zones described above, but the ocean itself is separated horizontally by distance from the shore. Nearest to the shore lies the **intertidal zone**, the region between the high and low tidal marks. This hallmark of the intertidal is changed, where water is in constant motions from ocean waves, tides, and currents. The land is sometimes underwater and sometimes is exposed. The **neritic zone** is from low tide mark and slopes gradually down to the edge of the seaward side of the continental shelf. Some sunlight penetrates the seabed here. The

Physical Geography and Natural Disasters by R. Adam Dastrup

oceanic zone is the entire rest of the ocean from the bottom edge of the neritic zone, where sunlight does not reach the bottom.
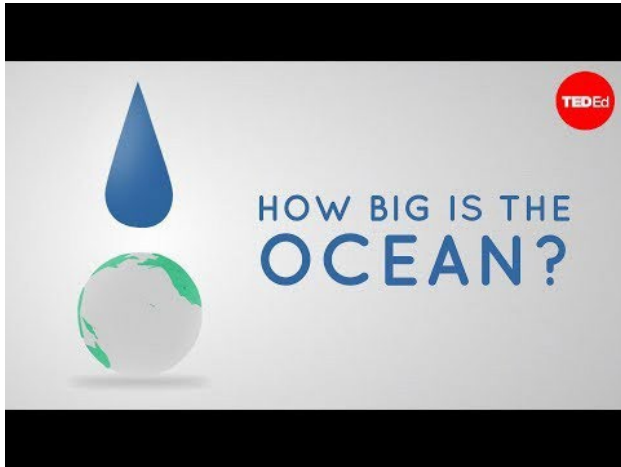
Video 17.5 How big is the ocean

https://www.youtube.com/watch?v=QUW_Zv_jJb8

Tsunamis

A particular type of wave is generated by an energetic event affecting the seafloor, such as earthquakes, submarine landslides, and volcanic eruptions. Called **tsunamis**, these waves are created when a portion of the seafloor is suddenly elevated by movement in the crustal rocks below that are involved in an earthquake. The water is suddenly lifted, and a wave train spreads out in all directions from the mound carrying enormous energy and traveling very fast (hundreds of miles per hour).

Tsunamis may pass unnoticed in the open ocean because the wavelength is very long, and the wave height is shallow. However, as the wave train approaches the shore, each wave makes contact with the shallow seafloor, friction increases, and the wave slows down. Wave height builds up, and the wave strikes the shore as a wall of water a hundred or more feet high. The massive wave may sweep inland well beyond the beach. This is called the tsunami **run-up**, which destroys structures far inland. Tsunamis deliver a catastrophic blow to observers at the beach as the water in the trough in front of it is drawn back toward the tsunami wave, exposing the seafloor. Curious and unsuspecting people on the beach may run out to see exposed offshore sea life only to be

540

overwhelmed when the breaking crest hits. (12 Coastlines – An Introduction to Geology, n.d.)

Tidal Waves

Tides are the daily rise and fall of sea level at any given place. The pull of the Moon's gravity on Earth is the primary cause of tides, and the pull of the Sun's gravity on Earth is the secondary cause. The Moon has a more significant effect because, although it is much smaller than the Sun, it is much closer. The Moon's pull is about twice that of the Sun's.

Video 17.7 How tides work.
https://www.youtube.com/watch?v=pwChk4S99i4
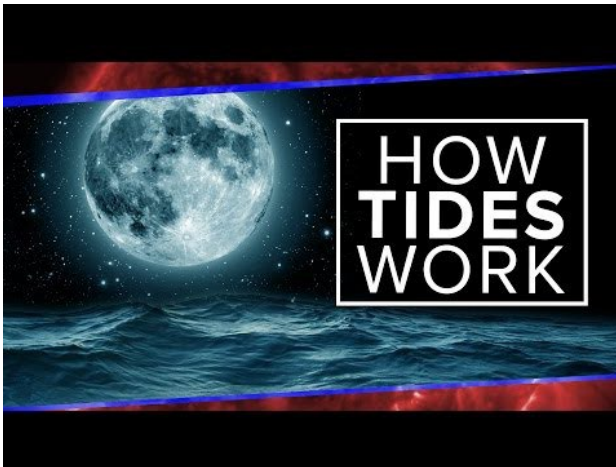
Daily Tide Patterns

To understand the tides, it is easiest to start with the effect of the Moon on Earth. As the Moon revolves around our planet, its gravity pulls Earth toward it. The lithosphere is unable to move much, but the gravity pulls the water above it, and a bulge is created. This bulge is the high tide beneath the Moon. The Moon's gravity then pulls the Earth toward it, leaving the water on the opposite side of the planet behind. This creates a second-high tide bulge on the opposite side of Earth from the Moon. These two water bulges on opposite sides of the Earth aligned with the Moon are the high tides. (Tides | Earth Science, n.d.)

Since so much water is pulled into the two high tides, **low tides** form between the two high tides. As the Earth rotates beneath the Moon, a single spot will experience two high tides and two low tides every day.

The **tidal range** is the difference between the ocean level at high tide and the ocean at low tide. The tidal range in a location depends on some factors, including the slope of the seafloor. Water appears to move a greater distance on a gentle slope than on a steep slope. (Ocean Movements | Earth Science, n.d.)

Monthly Tidal Patterns

Waves are additive, so when the gravitational pull of the Sun and Moon are in the same direction, the high tides add, and the low tides add. High tides are higher and low tides are lower than at other times throughout the month. These more extreme tides, with

542

a greater tidal range, are called **spring tides**. Spring tides do not just occur in the spring; they occur whenever the Moon is in a new-moon or full-moon phase, about every 14 days. (Tidal Waves | Physical Geography, n.d.)

Neap tides are tides that have the smallest tidal range, and they occur when the Earth, the Moon, and the Sun form a 90-degree angle. They occur precisely halfway between the spring tides when the Moon is at first or last quarter. How do the tides add up to create neap tides? The Moon's high tide occurs in the same place as the Sun's low tide and the Moon's low tide in the same place as the Sun's high tide. At neap tides, the tidal range relatively small.

High tides occur about twice a day, about every 12 hours and 25 minutes. The reason is that the Moon takes 24 hours and 50 minutes to rotate once around the Earth, so the Moon is over the same location 24 hours and 50 minutes later. Since high tides occur twice a day, one arrives every 12 hours and 25 minutes. (Tides | Earth Science, n.d.)

Some coastal areas do not follow this pattern at all. These coastal areas may have one high and one low tide per day or a different amount of time between two high tides. These differences are often because of local conditions, such as the shape of the coastline that the tide is entering. The National Ocean Service has a wealth of information on tides and water levels.

543

OCEAN CURRENTS

Surface Currents

Ocean water moves in predictable ways along the ocean surface. **Surface currents** can flow for thousands of kilometers and can reach depths of hundreds of meters. These surface currents do not depend on the weather; they remain unchanged even in large storms because they depend on factors that do not change. (Surface Currents | Physical Geography, n.d.)

Physical Geography and Natural Disasters by R. Adam Dastrup
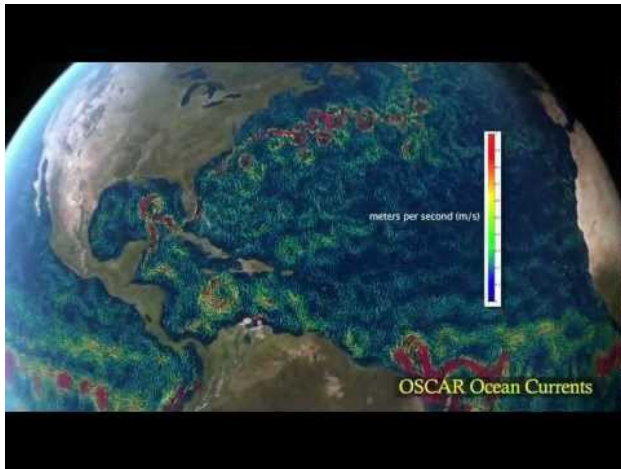
Surface currents are created by three things: global wind patterns, the rotation of the earth, and ocean basins' shape. Surface currents are significant because they distribute heat around the planet and are a significant factor influencing climate around the globe.

Global Wind Currents

Winds on Earth are either global or local. **Global winds** blow in the same directions all the time and are related to the unequal heating of Earth by the Sun, that is that more solar radiation strikes the equator than the polar regions, and the rotation of the Earth called the Coriolis effect. The causes of the global wind patterns will be described later when we look at the atmosphere. (Surface Currents | Physical Geography, n.d.) Water in the surface currents is pushed in the direction of the significant wind belts:

- **Trade winds** are consistent winds that flow east to west between the equator and 30 degrees North and 30 degrees South.

Physical Geography and Natural Disasters by R. Adam Dastrup

- **Westerlies** are winds that flow west to east in the middle latitudes.
- **Polar easterlies** are winds that flow east to west between 50 degrees and 60 degrees north and south of the equator and the north and south pole.

Topology of the Sea Floor

Oceans cover 71 percent of Earth's surface and hold 97 percent of Earth's water. The water the oceans hold is critical to plate tectonics, volcanism, and, of course, life on Earth. We know more about the surface of the Moon than the floor of the oceans. Whether this is true or not, the critical point is that the ocean floor is covered with an average of nearly 4,000 m of water, and it is pitch black below a few hundred meters, so it is not easy to discover what is down there. We know a lot more about the oceans than we used to, but there is still a great deal more to discover. (Earle, 2015)

Earth has had oceans for a very long time, dating back to the point where the surface had cooled enough to allow liquid water, only a few hundred million years after Earth's formation. At that time, there were no continental rocks, so the water that was here was likely spread out over the surface in one giant (but relatively shallow) ocean. (Earle, 2019)

We examined the seafloor's topography from the perspective of plate tectonics, but here we are going to take another look at the essential features from an oceanographic perspective. The essential features are the extensive **continental shelves** less than 250 m deep

(pink); the vast deep **ocean plains** between 4,000 and 6,000 m deep (light and dark blue); the mid-Atlantic ridge, in many areas shallower than 3,000 m; and the deep ocean trench north of Puerto Rico. (18.1 The Topography of the Sea Floor – Physical Geology, n.d.)
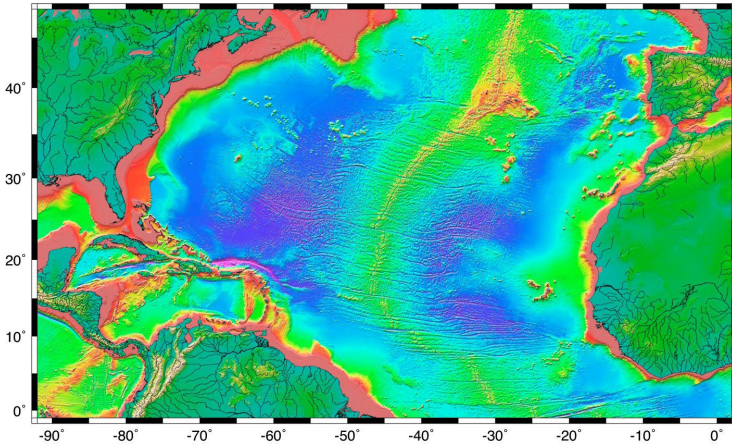
Figure 17.5 "Topography of the Atlantic Ocean Sea Floor" by NASA is licensed as Public Domain.

The main features of the Pacific Ocean floor are the continental slopes, which drop from about 200 m to several thousand meters over a distance of a few hundred kilometers; the abyssal plains – exceedingly flat and from 4,000 m to 6,000 m deep; volcanic seamounts and islands; and trenches at subduction zones that are up to 11,000 m deep.
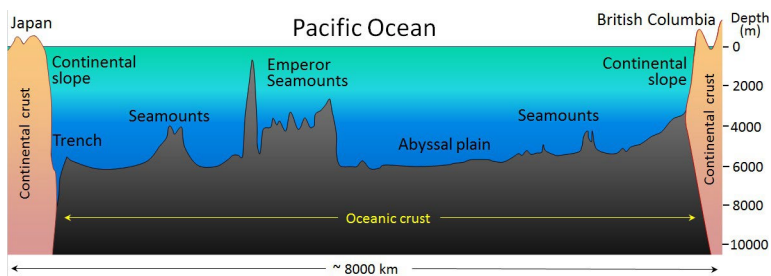
The ocean floor is entirely underlain by mafic oceanic crust, while the continental slopes are underlain by felsic continental crust (mostly granitic and sedimentary rocks). Moreover, the denser oceanic crust floats lower on the mantle than continental crust does, and that is why oceans are oceans. Although the temperature of the ocean surface varies widely, from a few degrees either side of freezing in polar regions to over 25°C in the tropics, in most parts of the ocean, the water temperature is around 10°C at 1,000 m depth and about 4°C from 2,000 m depth to the bottom. (18.1 The Topography of the Sea Floor – Physical Geology, n.d.)

The deepest parts of the ocean are within the subduction trenches, and the deepest of these is the Marianas Trench in the southwestern Pacific (near Guam) at 11,000 m. Other trenches in the southwestern Pacific are over 10,000 m deep; the Japan Trench is over 9,000 m deep, and the Puerto Rico and Chile-Peru Trenches are over 8,000 m deep. Shallow trenches tend to be that way because they have significant sediment infill. There is no recognizable trench along the subduction zone of the Juan de Fuca

Plate because it has been filled with sediments from the Fraser and Columbia Rivers. (Earle, 2015).

Barrier islands are elongated islands composed of sand that form a few kilometers away from the mainland. They are common along the US Gulf Coast from Texas to Florida and along the US Atlantic Coast from Florida to Massachusetts. Georgia barrier islands include Jekyll Island, St. Simons Island, and Sea Island.

Creative Commons Attributions

Chapter 18 Weather

COMPOSITION AND STRUCTURE OF THE ATMOSPHERE

Composition of the Atmosphere

Vertical Structure of the Atmosphere
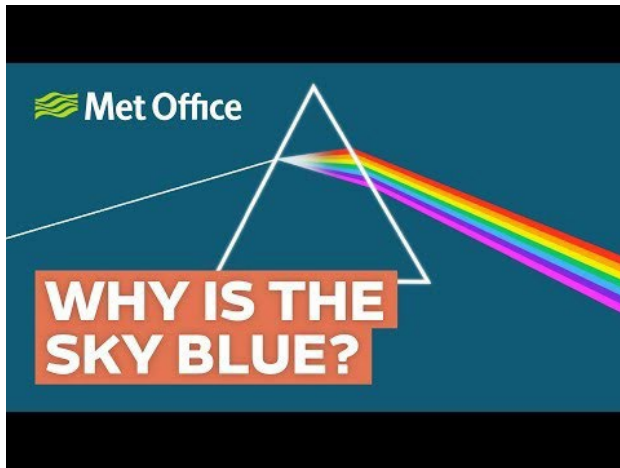
Video 18.1: Why is the sky blue
https://slcc.pressbooks.pub/physicalgeography/?p=614

Physical Geography and Natural Disasters by R. Adam Dastrup
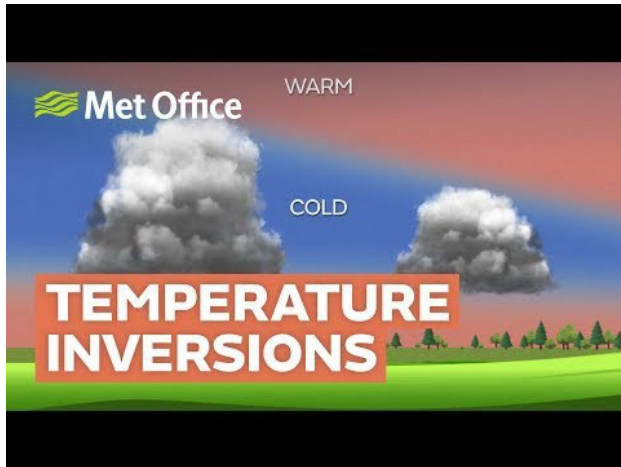
the text. You can view it online here:

Video 18.2: Temperature Inversions

https://slcc.pressbooks.pub/physicalgeography/?p=614

Physical Geography and Natural Disasters by R. Adam Dastrup

WARMING THE ATMOSPHERE

Incoming Solar Radiation

The sun is the driving force of our weather and climate. Its diameter is about 865,000 miles or 109 Earths lined up side-by-side and is so large that it makes up 99.99 percent of the mass of our Solar System. According to astronomers, our sun is the most common type of star in the known universe and has existed for nearly 4.6 billion years.

The sun's composition is about 92 percent hydrogen (the lightest element in the universe) and almost 8 percent helium (the second lightest element). Using these elements, the sun generates energy through a process called **thermonuclear reaction**. This energy is created when the sun fuses hydrogen atoms to make a heavier element called helium. This process not only creates new elements but also releases an enormous amount of energy.

552

Video 18.3 Sun 101

https://www.youtube.com/watch?v=2HoTK_Gqi2Q

The energy emitted by the sun is called **electromagnetic energy**, a range of wavelengths from the shortwave to the longwave spectrum of light, which travels at the speed of light. Recall from the module on tsunamis that a wavelength is the distance between two wave crests, and frequency is the number of wavelengths that pass a given point in a given amount of time.
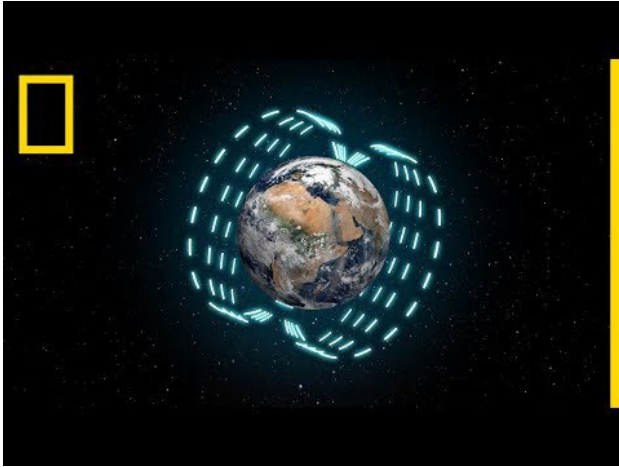
Video 18.4 Electromagnetism 101

https://www.youtube.com/watch?v=Elv3WpL32UE

The sun's life expectancy is already half over. In another 4 billion years, it will consume all the hydrogen within it, leaving behind helium and a few other newly created elements. When the hydrogen runs out, in about 5 billion years, the sun will grow into a red giant, expanding its diameter well past Mercury and possibly Venus and Earth. The sun will continue to fuse helium into carbon until the star begins to release its outer layers to form a **planetary nebula**. Left behind will be the white-hot core of our star, called a **white dwarf** that will slowly fade out over billions of years.

Why would you want to wear a black shirt in the winter and a white shirt in the summer? Would you rather walk barefoot on black pavement or grass? The answers to these questions relate to the idea that different objects absorb and reflect energy differently on the planet. Darker objects tend to absorb more energy, while lighter objects reflect less energy. The percent of energy an object reflects is an object's **albedo**. Because the planet has different albedos from cities, forests, deserts, oceans, rocks, glaciers, and more, the planet has uneven heating. The result is that of all the energy the Earth receives from the sun, only 51 percent is absorbed by the planet, while 49 percent is reflected into space.
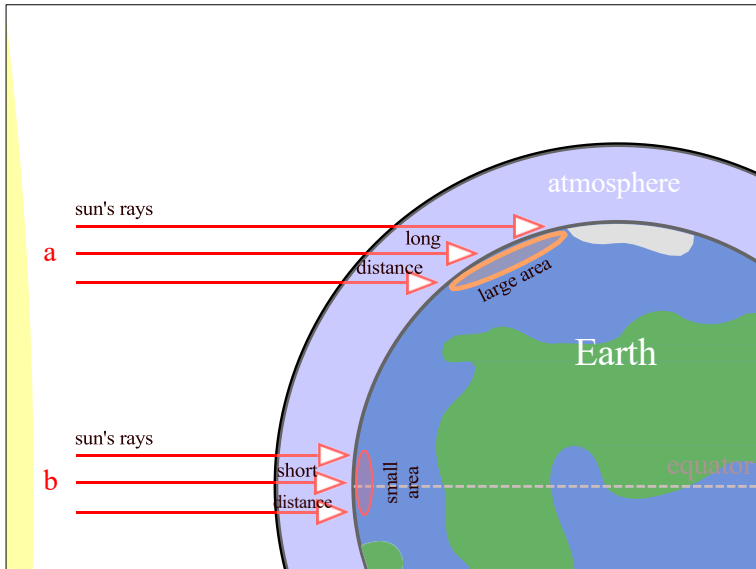
Uneven heating of the planet also occurs because of the curvature of the Earth. The planet receives most of its energy near the equator, where the sun's rays are directly hitting the planet. Towards the poles, the sun's energy becomes more diffused and spread out over a greater area. The equator receives more energy from the sun than it can radiate back into space, producing a surplus of energy at 0 degrees latitude. At the poles, the planet radiates more energy out into space than it receives from the sun. There is a deficit of energy in these regions. The ultimate purpose of weather is to transfer the surplus of energy and heat from the equator to the poles, bring colder air toward the equator and find equilibrium.

Global Transfer of Heat

The source of energy for the weather on Earth is the sun. The earth-atmosphere energy balance is the balance of energy the earth absorbs, and the energy reflected into space. Roughly 51 percent of the energy the earth receives from the sun is absorbed, while approximately 49 percent of the sun's energy is reflected into space. To balance the energy deficit at the earth's poles and the energy surplus at the equator, the planet will transfer energy by conduction, convection, latent heat, and radiation.

Video 18.5 Heat transfer
https://www.youtube.com/watch?v=YK7G6l_K6sA

Conduction

Conduction is the transfer of heat from a warmer object to a colder object through molecule interaction. As the sun heats the ground, energy is transferred to the atmosphere by conduction. However, the atmosphere is an inferior conductor of heat. In calm weather, the heated ground only warms the first few centimeters of air. The atmosphere's temperature can be up to 50 percent colder five feet above the ground than at the surface. Since the atmosphere is such a poor conductor of heat, there must be other ways to transfer the energy.

Convection

Convection is the transfer of heat by the mass movement of a fluid (such as water and air). It occurs mostly in liquids and gases because they are free to move around. Heat is transferred upward and outward away from its heat source, and cooler air is brought in to replace the rising air. On a local scale, convection in the summer can produce afternoon thunderstorms. On a global scale, convection transfers energy between the equator and poles.

Physical Geography and Natural Disasters by R. Adam Dastrup
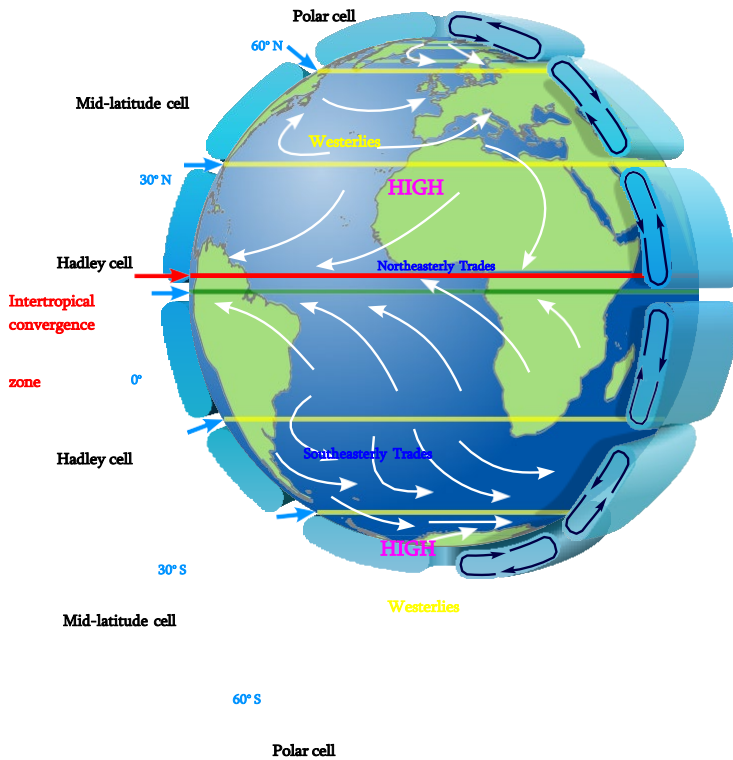
Figure 18.2 "Earth Global Circulation" by NASA is licensed as Public Domain.

Radiation

The final process of transferring heat around the planet is called **radiation**. The energy coming from the sun comes in various wavelengths. A **wavelength** is the distance measured along a wave of energy from one crest to another. The shorter the wavelength, the higher the energy. The energy received from the sun passes through the atmosphere without warming it. All objects that absorb radiation from the sun radiate some of that energy back into space, but in a weaker form of energy. This weaker energy becomes longwave radiation and is typically observed as heat. All living

objects (including humans) radiate longwave radiation. The Earth

itself radiates the energy absorbed from the sun in the form of longwave radiation, which is sometimes called **Earthlight**.

earth's energy *budget*

incoming solar radiation 340.4

reflected by clouds & atmosphere 77.0

reflected by surface 22.9

total outgoing infrared radiation 239.9

total reflected solar radiation 99.9

atmospheric window 40.1

emitted by atmosphere 169.9

latent heat (change of state)

absorbed by atmosphere 77.1

absorbed by atmosphere 358.2

emitted by clouds 29.9

thermals (conduction/ convection)

greenhouse gases

absorbed by surface 163.3

emitted by surface 398.2

back radiation 340.3

18.4   86.4

net absorbed 0.6

evapotranspiration

All values are fluxes in Wm⁻² and are average values based on ten years of data

Loeb et al., J. Clim. 2009
Trenberth et al., BAMS, 2009

Figure 18.3 "Earth's Energy Budget" by NASA is licensed as Public Domain.

Latent Heat

Latent heat is a powerful force in the weather. When water transforms from gas to liquid or solid, or vice versa, it is called a **phase change**. The heat required to change phases is called latent heat. For water to change from a liquid to a gas, energy/heat must be taken from external sources such as the surrounding atmosphere. Therefore, evaporation is a cooling process because the water is taking heat from the surrounding air to evaporate.
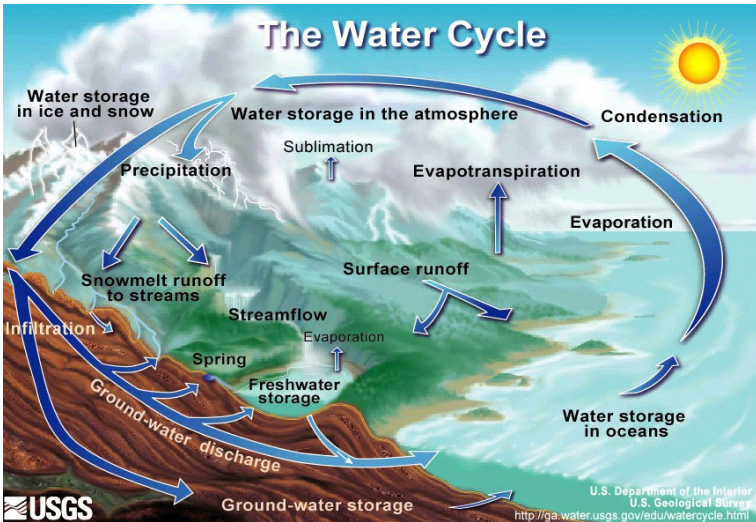
Figure 18.4 "The Water Cycle" by the United States Geologic Survey is licensed as Public Domain.

The process of evaporation and condensation transfers substantial amounts of heat around the planet. The heat released in a typical thunderstorm is equal to one of the original atomic bombs. The heat and energy within a hurricane are equal to thousands of nuclear weapons or could power the United States for over a year.
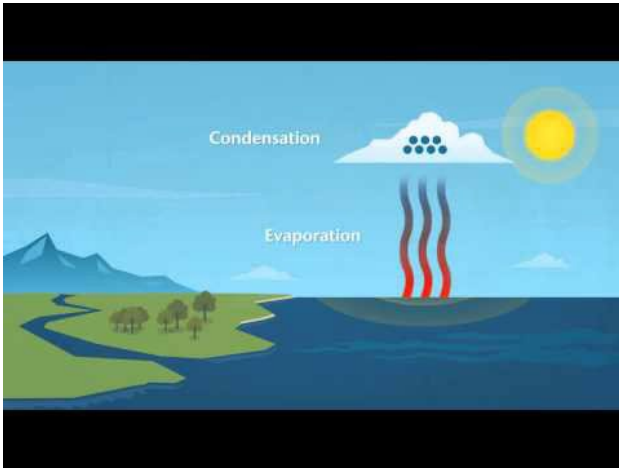
Video 18.6 How does rain form

https://www.youtube.com/watch?v=zBnKgwnn7i4

CONTROLS OF WEATHER AND CLIMATE

Latitude

Several controlling factors determine global temperatures. The first and most significant is **latitude**. Because of the Earth's shape and the sun's angle hitting the planet, temperatures are highest near the equator and decrease toward the poles. In fact, at the equator, more energy is absorbed from the sun than is radiated back into space. At the poles, more energy is radiated back into space than is absorbed by the sun. The purpose of weather and ocean currents is to balance out these two extremes.
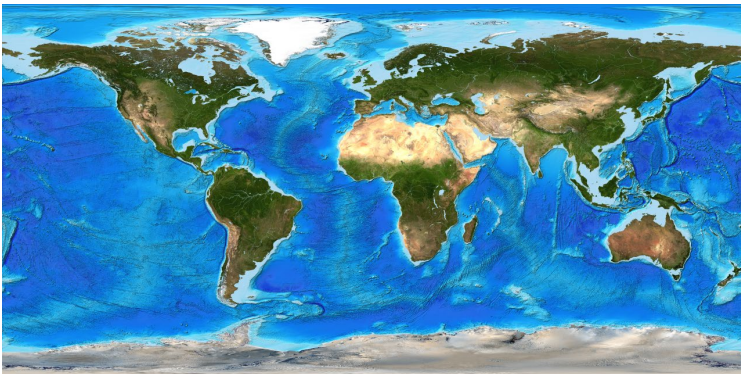
Figure 18.5 "Large World Physical Map" is licensed as Creative Commons Attribution-ShareAlike 4.0 International.

Physical Geography and Natural Disasters by R. Adam Dastrup

Land-Water Distribution

The next impact on temperature is the **land-water distribution** on the planet. Places near the ocean tend to have milder climates year-round versus regions surrounded by land. This is because the earth can heat up and cool down faster and with more significant fluctuation than the ocean. The reason is that sunlight must heat a larger volume of area in the ocean because light can pass through water. Water requires five times more energy to heat one degree Celsius than for landmasses, called **specific heat**. Thus, the region's temperatures found near large bodies of water temperatures change slowly compared to land. Ocean currents are also vital controls in transferring heat around the planet. In the Northern Hemisphere, ocean currents rotate clockwise, bringing cold water from the North Pole toward the equator and warm water from the equator toward the North Pole. The opposite occurs in the Southern Hemisphere, where ocean currents rotate counterclockwise.

Elevation

The last control of temperature is **elevation**. On average, atmospheric temperature decreases 3.6 degrees Fahrenheit per 1,000 feet rise in elevation. This is called the **normal lapse rate** or also called the **temperature lapse rate**.

565

Moisture and Humidity

For liquid water to evaporate, water molecules must absorb enough energy to break bonds between each other. To do this, the liquid water must absorb energy and heat from the surrounding environment. This release of energy is called **latent heat**. If the water vapor absorbs enough energy, they will begin to vibrate fast enough to break their molecular bonds and become individual water molecules or gas. **Evaporation** is a cooling process because it takes heat from the surrounding environment. The concept of latent heat is essential to understand and will be revisited later when cloud formation and severe weather are discussed.

566

The opposite must occur for water vapor to condense into liquid water. For fast vibrating water molecules to condense into liquid, it must release latent heat to the surrounding environment. Releasing energy allows the water molecules to slow down their vibration and attach to other water molecules to become liquid. However, one step is missing. For water vapor to become liquid, it needs something to condense onto condensation nuclei. **Condensation nuclei** consist of microscopic dust, smoke, salt particles, or even bacteria that float in the air. It is believed that bacteria make up nearly half of all condensation nuclei. To summarize, for water vapor to condense into small liquid or ice cloud droplets, condensation nuclei must be present.

Figure 18.9 "Dew" is by Piklist.

Humidity is defined as the amount of water vapor in the atmosphere. There are several ways to classify humidity, but we will focus on **relative humidity** for this course. Relative humidity is the ratio of the atmosphere's actual water vapor content divided by the amount of water vapor required for atmospheric saturation at that temperature; it is usually expressed as a percentage. If the relative humidity is 25 percent, the atmosphere is only holding a quarter of what it could hold. If the relative humidity is at 100 percent, the atmosphere is saturated.

There are two ways to change relative humidity: **moisture content** and **temperature**. If the air temperature stays the same, but the amount of water vapor increases or decreases, relative humidity will change. Next, it should first be noted that warm air can "hold" more moisture than cooler air. If the water content stays the same, but temperature increases, relative humidity will decrease. If the

water content stays the same, but the temperature decreases, relative humidity will increase.
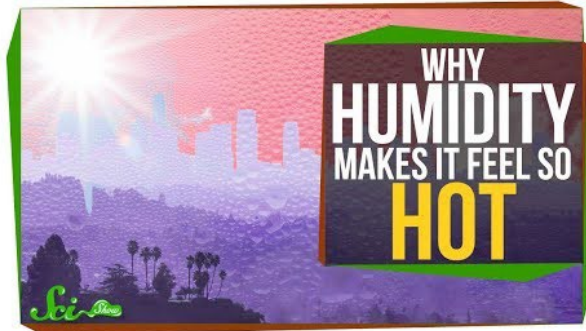
Video 18.7 Why humidity make it feel so hot?

https://www.youtube.com/watch?v=SGHRz8wpj3E

Relative humidity is just as the name implies; it is a relative measurement. A more direct measurement and analysis of humidity is **dew point**; the atmospheric temperature air must cool down to in order for it to condense into liquid water or solid ice crystals. So if the dew point is 42 degrees Fahrenheit for a particular geographic location at a particular time, then the current temperature must fall to 42 degrees for the air to become saturated. The higher the dew point reading, the less air must cool to become saturated and condense. The lower the dew point reading, the more air must cool to become saturated; thus, the air is quite dry. Dew

point analysis is vital for weather forecasting in the summer to determine the likelihood of afternoon thunderstorms. If the humidity is high, providing a high dew point measurement, then afternoon convection does not require the unstable moisture to rise as high in order for condensation and thunderstorms genesis to occur. Recall that condensation from water vapor to liquid water or ice crystals releases latent heat is a crucial ingredient for the formation of thunderstorms.

Atmospheric Pressure and Wind

Atmospheric pressure is a force created by the weight of the atmosphere. Because of gravity, air pressure is highest at sea level and decreases with height. There is also high pressure and low pressure. **High pressure**, also called an **anticyclone**, occurs when descending air molecules "pile-up" at the surface and spread outward in a clockwise rotation in the Northern Hemisphere. In the Southern Hemisphere, the air within high pressure flows counterclockwise. In either case, the descending air will warm, which prevents water vapor from cooling and condensing into clouds to produce storms. Instead, regions under high pressure tend to experience clear skies. **Low pressure**, also called a **cyclone**, occurs when converging air is forced upward (in a counterclockwise manner in the Northern Hemisphere) where it cools and condenses into clouds and possible storms. Ultimately, air flows from high pressure to low pressure, and this is called **wind**.
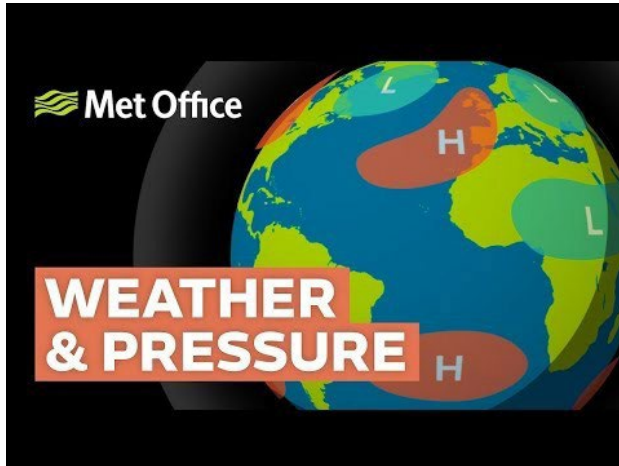
When atmospheric high pressure is near atmospheric pressure, there is an imbalance between the atmospheric pressure. The force to balance these two pressure imbalances is called the **pressure gradient force**, which creates wind. Wind is the horizontal movement of air from high pressure to low pressure to balance atmospheric pressure.

Atmospheric Stability

To have cloud formation, the air must be unstable. **Stable air** means air does not want to rise, cool, and condense. Thus, weather

conditions tend to be clear skies with stable air. **Unstable air** means the air wants to rise, cool, and condense into clouds and potential storms. The forces that cause air to rise are convection, orographic uplift, convergence, and weather fronts.

Convection occurs when air rises, much like a hot air balloon. Because of albedo, some areas on the ground can get heated more than other areas. Where the land heats up more, the air above also warms, becomes less dense, and rises. If the air rises high enough, it may cool and condense to create clouds and possibly thunderstorms.

Orographic uplift is when mountains help destabilize air and occur when air must rise over a mountain range. As the air rises over the mountain, the moisture within it may begin to cool and condense to form thunderstorms. Often with orographic uplift, one side of the mountain will be very moist from the storms, while the other side is arid. The dry side of the mountain is called the **rainshadow effect**. Later we will discuss how this process can generate what is called dry thunderstorms and wildfires.
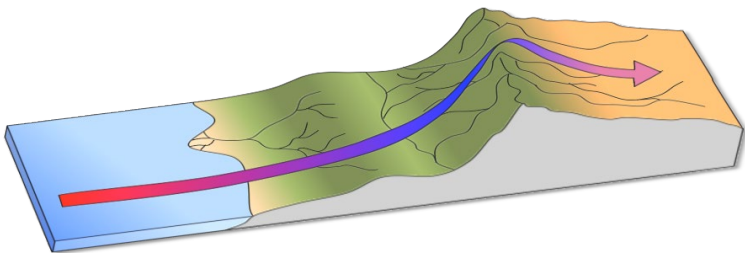
Figure 18.10 "Rainshadow Effect Caused by Orographic Uplift" is licensed as Public Domain.

Convergence occurs when air is forced to rise because of low pressure above, causing the rising air to cool and condense into clouds. One of the best examples of this is over Florida. Because Florida is a peninsula, surrounded by water on three sides, the land heats up more than the surrounding water. This causes the air above the land to rise. To replace this rising air, humid and cooler air from the Gulf of Mexico and the Atlantic Ocean converges inward over Florida. This moist air is heated by the land and is forced upward to create powerful thunderstorms. Florida has more thunderstorms and lightning than any other state in the nation. Another excellent example of convergence is the eye of a hurricane because winds and moisture are rotating around the eye until they converge within the eye.

573

Finally, **weather fronts**, such as cold fronts, warm fronts, stationary fronts, and occluded fronts, can force air to rise. For example, a cold front occurs when a cold, dense air mass replaces a warm, lighter air mass. The cold air mass plows through, forcing the warmer air mass upward to cool and condense into clouds.

Coriolis Effect

All free moving objects appear to be deflected to the right in the Northern Hemisphere, and to the left in the Southern Hemisphere, because of the rotation of the Earth. This apparent deflection is called the **Coriolis effect**.
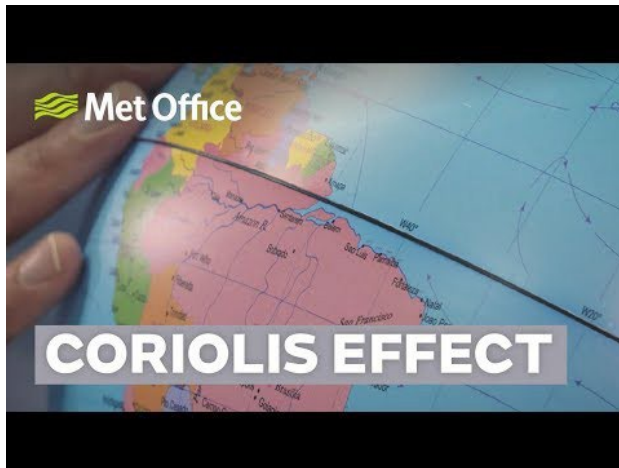
Video 18.9 Coriolis effect
https://www.youtube.com/watch?v=WB4dxpUS530

Aside from a football, what else is considered a "free moving" object? Free moving objects include airplanes, ocean currents, clouds, wind, hurricanes, or anything else that is not attached to the earth's surface. So how does this influence the atmosphere? As winds flow from high to low pressure, it deflects to the right (in the Northern Hemisphere) because of the earth's rotation. This deflection is nonexistent at the equator and progressively gets stronger towards the poles. Because of this, hurricanes can only form 5-20 degrees north or south of the equator. Near the equator, the water is warm enough, but the Coriolis effect is too weak to make the hurricane rotate. Beyond 20 degrees latitude, the Coriolis effect is strong enough, but the waters are too cold.

Types of Precipitation

For **precipitation** to occur, a variety of atmospheric conditions must occur. This includes a moisture source along with high and low pressure. Next, atmospheric instability needs to occur so that cloud formation may develop. Finally, moisture in the air needs to condense onto condensation nuclei so that the condensed moisture can become large enough to fall from the clouds in the form of precipitation. The atmosphere can produce a variety of **precipitation types**.

Rain is liquid water falling from nimbostratus or cumulonimbus clouds. Many times, in the midlatitudes, precipitation will fall from clouds in the form of ice or snow, which then melts on its way down toward the ground.
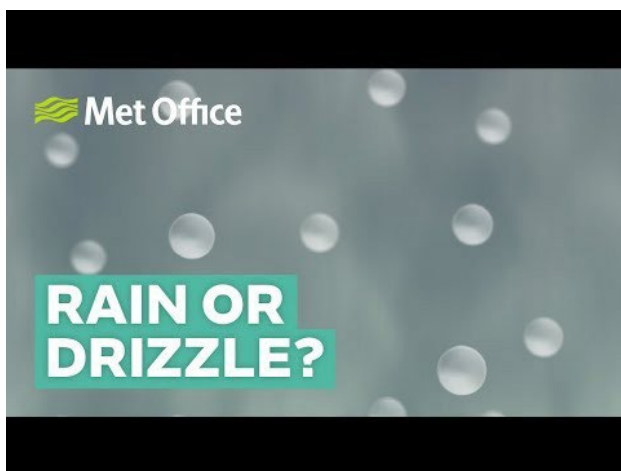
Video 18.10 Rain or drizzle

https://www.youtube.com/watch?v=OdesSenwQgM
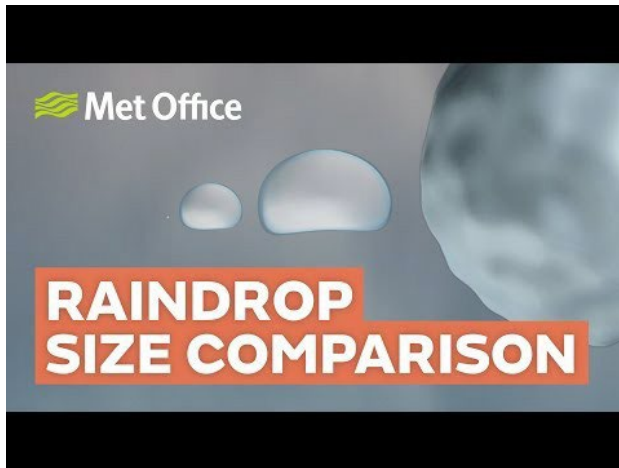
Video 18.11 Raindrop size comparison
https://www.youtube.com/watch?v=TjxEUjk9LLg

Snow is precipitation in the form of ice crystals. The temperature the snow forms will determine the size, shape, and concentration of snowflakes. When temperatures are freezing, snowflakes tend to be small, dry, to produce "powder" and powdery. When the temperatures are warmer, the snowflakes are larger.

Video 18.12 How does snow form?
https://www.youtube.com/watch?v=Cf6El0mI1fM

Sleet is precipitation that falls as ice pellets. It occurs when precipitation falls from the base of a cloud in the form of snow. As the snow falls, it enters a region of warm air and melts into rain. However, as the rain continues to fall, it enters a layer of cold air and refreezes in the form of ice pellets.

Video 18.13 What is sleet?

https://www.youtube.com/watch?v=XSS4xQNfmMs
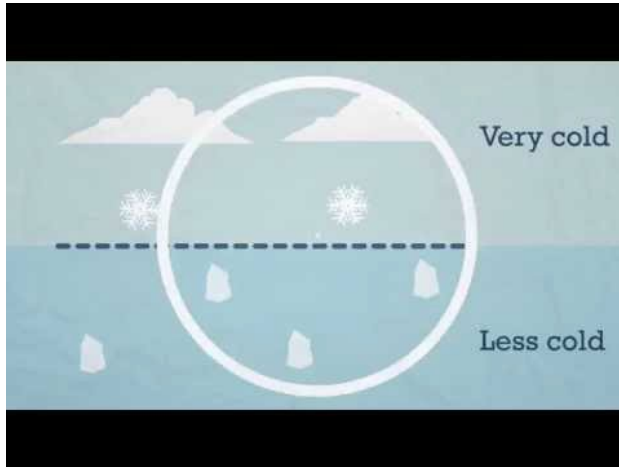
Freezing rain (also called **glaze**) is similar to sleet except for the last step. As the rain falls, it enters a layer of cold air. However, the rain is not in this cold region long enough to freeze. Instead is stays as supercooled raindrops. However, once the supercooled raindrops reach the ground, they freeze instantly on any object it touches.

Video 18.14 Freezing rain explained

https://www.youtube.com/watch?v=56QmJWpBwzk

Video 18.15 What is the difference between snow, sleet and freezing rain
https://www.youtube.com/watch?v=5Efr4oLEjLQ

Hail is precipitation in the form of hard pellets of ice and only forms in cumulonimbus clouds where the lower region of the cloud contains liquid water and is above freezing, while the upper region is below freezing. When an ice pellet falls within the cumulonimbus cloud, it enters the warm, liquid region and picks up moisture. Then the updrafts through the ice pellet back up above the freezing point hardening the newly gathered water. The ice pellet will fall again to collect liquid water and thrown back up to refreeze. This process will continue until the hailstone's weight becomes too heavy for the updrafts to hold it up. Once the hail becomes too heavy, the hail will precipitate from the cloud. (Dastrup, 2014)

Video 18.16 Hail and hailstones
https://www.youtube.com/watch?v=lxv_M0p_I2A

CLOUD TYPES AND FORMATION

Video 18.17 What are clouds and how do they form

https://www.youtube.com/watch?v=UZEETyzql0Q
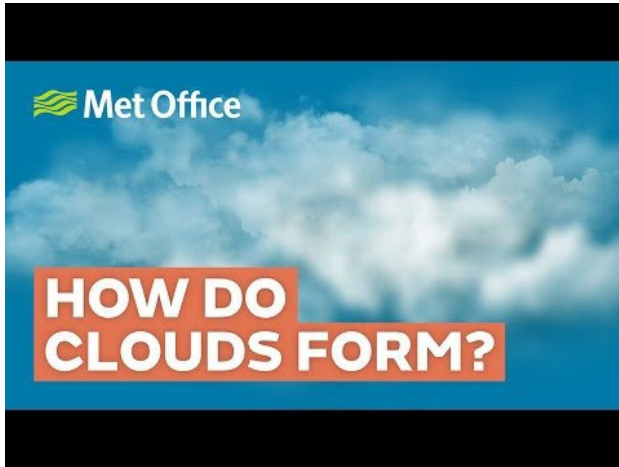
Video 18.18 How do clouds form?

https://www.youtube.com/watch?v=q87Ekar3emA

Video 18.19 How clouds were named.

Video 18.20 Cloud spotting guide

https://www.youtube.com/watch?v=JaAz3lp8Keg
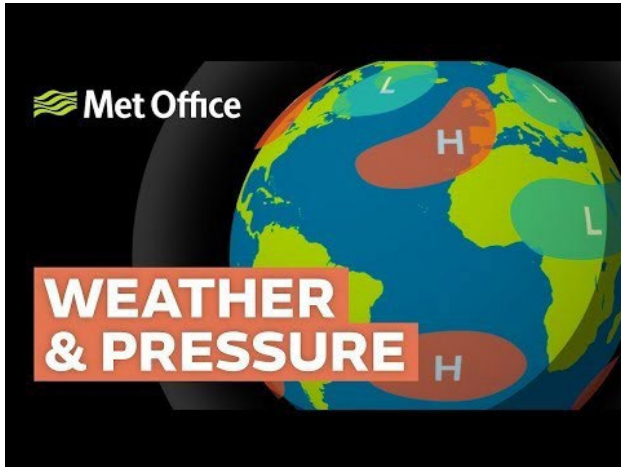
Video 18.21 Weather and pressure

https://www.youtube.com/watch?v=SWHj71qS_NA
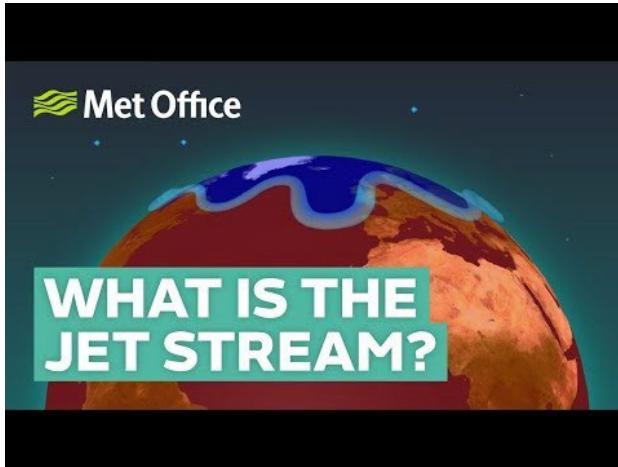
Video 18.22 What is jet stream?
https://www.youtube.com/watch?v=Lg91eowtfbw

AIR MASSES AND WEATHER FRONTS

Air Masses

An **air mass** is a large body of air covering a relatively wide area and exhibiting "horizontally uniform properties of moisture and temperature. An air mass originates from the source region and determines the moisture and temperature characteristics of an air mass. For an air mass to develop, the surface of the source region must be relatively flat and uniform in composition (i.e., oceans, deserts, glaciers.), but not a combination. If the air mass stays long enough within the source region, it will begin to develop the characteristics of that source region. An air mass is classified by its temperature and moisture and is identified by using a "letter code" system. (Dastrup, 2014)
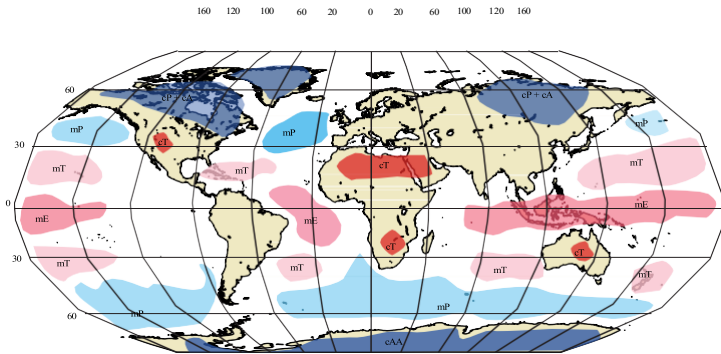
The first letter is always lower case and determines the moisture content within the air mass:

- m (maritime) and is moist
- c (continental) and is dry

The second letter is always capital and determines latitude.

- E (equatorial) and is hot
- T (tropical) and is warm
- P (polar) and is cold
- A (arctic) and is very cold

Video 18.23 Air masses

https://www.youtube.com/watch?v=QxvF1nnWRwg

Video 18.24 What are airmasses?
https://www.youtube.com/watch?v=kvk-hBFnBTI

Weather Fronts

A **weather front** is a transition zone between two air masses of relatively different densities, temperatures, and moisture. When two air masses come into contact with each other, they do not like to mix well because of their different densities (much like water and oil.) Along a weather front, the warmer, less dense air rises over the colder, denser air to form clouds. There are several types of weather fronts: stationary fronts, cold fronts, warm fronts, and occluded fronts. (Dastrup, 2014)

Video 18.25 Weather fronts
https://www.youtube.com/watch?v=o4lg8UfY5DM

A **stationary front** occurs when two contrasting air masses of moisture and temperature connect, but neither of them will ground the other. In many ways, the two air masses are stuck or at a stalemate until one of them begins to give ground to the other. Typically, but not always, stationary fronts produce mild but prolonged precipitation.

Cold fronts are zones separating two distinct air masses, of which the colder, denser mass is advancing and replacing the warmer. The colder, denser air pushes under the warm air, forcing the warm, lighter air upward. If the warm air rising is unstable enough, massive thunderstorms are likely to occur. Cumulous and cumulonimbus clouds are common.

596

Video 18.26 What are weather fronts?

https://www.youtube.com/watch?v=G7Ewqm0YHUI

Warm fronts mark the boundary between a warm air mass that is replacing a colder air mass. When a warm air mass advances over a cold air mass, the warm air rises over, but at a gentler rate than a cold front. Since the warm air does not rise as fast as a cold front, more stratus clouds form, and precipitation is not as heavy.
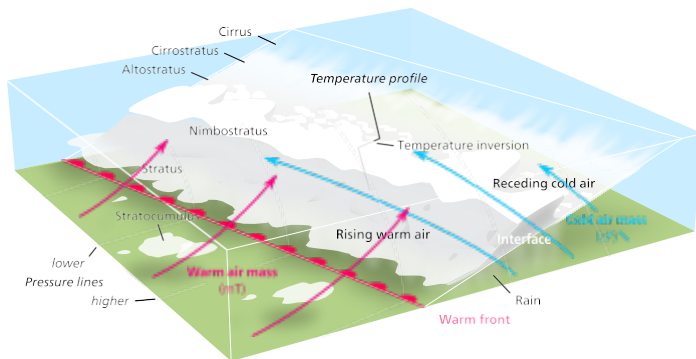
Finally, **occluded fronts** occur when a cold front overtakes a warm front, causing the warmer air to rise above, and meets up with another cold air mass.

Video 18. 27 What are weather fronts and how do they affect our weather?

https://www.youtube.com/watch?v=dwIQds-4I7I

THUNDERSTORMS

Thunderstorm Basics

A **thunderstorm** is defined as a storm that has lightning and thunder. Worldwide there are over 40,000 thunderstorms every day with the United States alone having 100,000 thunder-storms yearly. The essential ingredients for a thunderstorm are warm, moist, unstable air that is forced to rise either through convection, convergence, orographic uplift, or weather fronts. The rising parcel of air condenses into various types of cumulus clouds. (Dastrup, 2014)
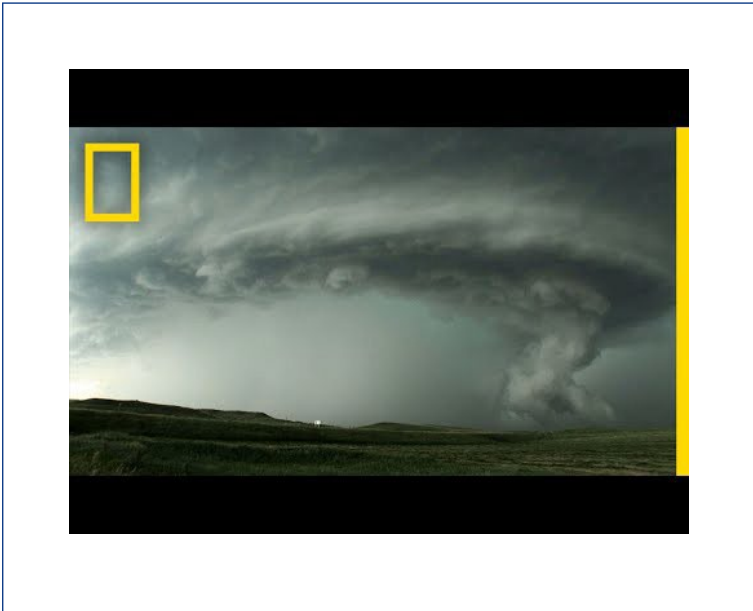
Thunderstorm Genesis

All thunderstorms go through a three-stage lifecycle. The first stage is called the **cumulus stage**, where an air parcel is forced to rise, cool, and condense, called the **lower condensation level**, to develop into a cumulus cloud. The process of water vapor condensing into liquid water releases large quantities of latent heat, which makes the air within the cloud warmer, and unstable causing the cloud continues to grow upward like a hot air balloon. These rising air parcels, called updrafts, prevent precipitation from falling from the cloud. However, once the precipitation becomes too heavy for the updrafts to hold up, the moisture begins to fall, creating downdrafts within the cloud. The downdrafts also begin to pull cold, dry air from outside the cloud toward the ground in a process called **entrainment**.
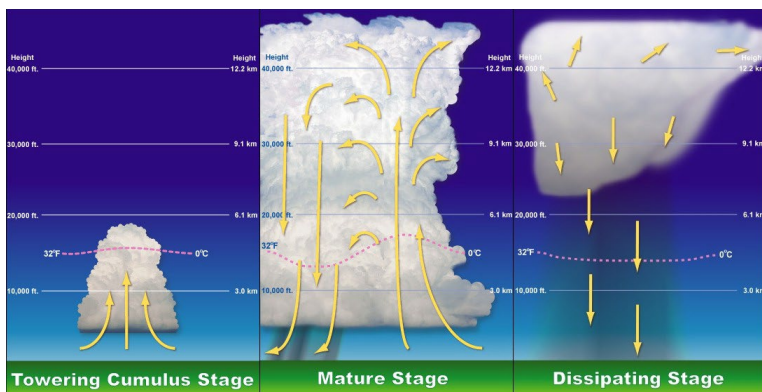
Figure 18.15 "Thunderstorm Formation" by NOAA Jetstream is licensed as Public Domain.

Once the precipitation begins to fall from the cloud, the storm has reached the mature stage. During this stage, updrafts and downdrafts exist side-by-side, and the cumulonimbus is called a **cell**. If the updrafts reach the top of the troposphere, the cumulus cloud will begin to spread outward, creating a defined **anvil**. At the same time, the downdrafts spread within the cloud and at first make the cloud more extensive, eventually overtaking the updrafts. Cold downdrafts form when precipitation and the cold air from entrainment are dragged down to the lower regions of a thunderstorm. It is also during the mature stage when the storm is most intense, producing strong, gusting winds, heavy precipitation, lightning, and possibly small hail.

Once the downdrafts overtake the updrafts, which prevents the release of latent heat energy, the thunderstorm will begin to weaken into the third and final stage, called the dissipating stage. During this stage, light precipitation and downdrafts become the dominate feature within the cloud as it weakens. Only twenty percent of the

602

moisture within the cloud fell as precipitation, whereas the other eighty percent evaporates back into the atmosphere. (Dastrup, 2014)

Video 18.29 What causes thunderstorms?
https://www.youtube.com/watch?v=Z7c4bZwvkrs

Types of Thunderstorms

Air Mass Thunderstorms

There is a variety of thunderstorms, all with defined characteristics. **Air mass thunderstorms**, sometimes called **ordinary thunderstorms**, go through the defined lifecycle sequence mentioned earlier. As the word implies, air mass thunderstorms

603

occur when a warm, moist (mT) mass of air from a source region such as the Gulf of Mexico migrates over the land like the Great Plains. However, rather than a large-scale storm system forcing the mT air mass upward, localized uplift such as convection, sea-breezes, or orographic uplift forces the moist air within the air mass upward.

Most thunderstorms can also create other thunderstorms when downdrafts within a thunderstorm slam the ground and spread outward in an arc-shape. These rippling waves are called, **outflow boundaries** or **gust fronts**, because the cold air from the downdrafts acts like a mini-cold front. These mini-cold fronts can force warm air upward to generate new thunderstorms with cumulus development. On average, ordinary thunderstorms last about an hour. Ultimately, because of these outflow boundaries, an area may have several thunderstorms in various stages of development. These complex systems of thunderstorms can last several hours and are usually called **multi-cell thunderstorms**.

Severe Thunderstorms

The **Storm Prediction Center** classifies a severe thunderstorm as having winds that exceed 58 miles per hour or produces hail with a diameter of 1.9 centimeters. Of the 100,000 thunderstorms that develop within the United States' experiences every year, about ten percent (10,000 storms) become severe thunderstorms. These can include air mass thunderstorms, squall lines, dryline thunderstorms, and supercells.

Figure 18.16 "Squall Line" by Piklist is licensed as Public Domain.

Severe thunderstorms go through the same stages as air-mass thunderstorms, but with one significant difference; severe thunderstorms last much longer in the mature stage. Ordinary thunderstorms do not last much longer than an hour because the downdrafts begin to cut off the updrafts. However, severe thunderstorms have vertical wind shear at different levels that keep the storm in the mature stage longer. This occurs when fast upper-level winds, such as the jet stream, causes the updrafts to pull away from the downdrafts. This prevents the downdrafts from cutting off the updrafts. Severe thunderstorms are capable of producing beautiful mammatus clouds, torrential downpours, flash flooding, large hail, lightning, and straight-line winds.

Squall lines are thunderstorms that develop linearly over hundreds of miles along the leading edge of outflow boundaries and are called gust fronts. Though these large storm systems are quite powerful,

they do not tend to produce tornadoes. Rather, straight-line winds and precipitation form because of the powerful, horizontal movement of the gust front.
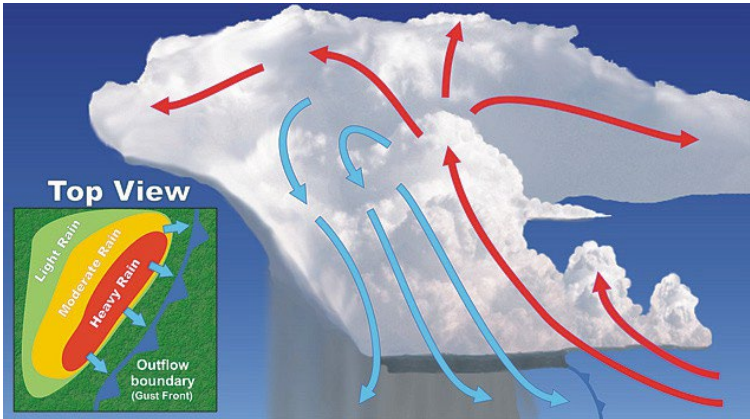
Figure 18.17 "Squall Line" by NOAA Jetstream is licensed as Public Domain.

A squall line can also form along a boundary called a **dryline**. Recall from the section on mid-latitude cyclones, that the warm air in front of the cold front is usually an mT air mass. Sometimes a dry, warm cT air mass can penetrate the mT air mass and cP air mass. Drier air is denser than moist air, so when it infiltrates a storm system, it can cause the mT air mass to rise much like a cold front without the cold. The hallmark of a dryline is that the moisture within the air dramatically drops following the boundary. Drylines can produce severe thunderstorms that are often more powerful than the cold front behind it.
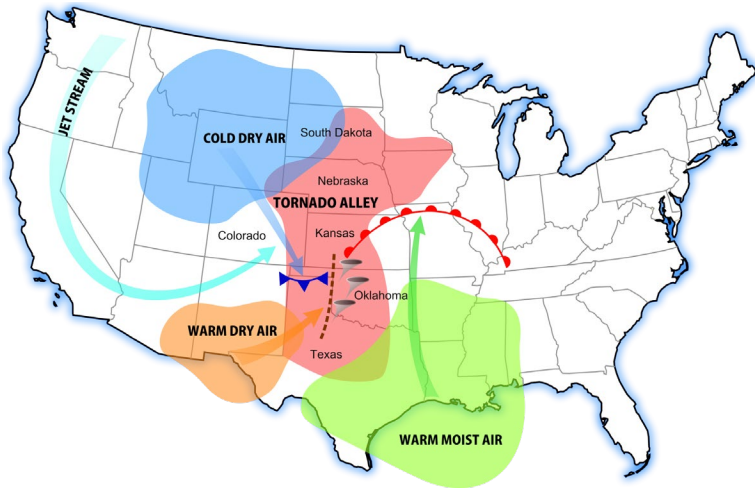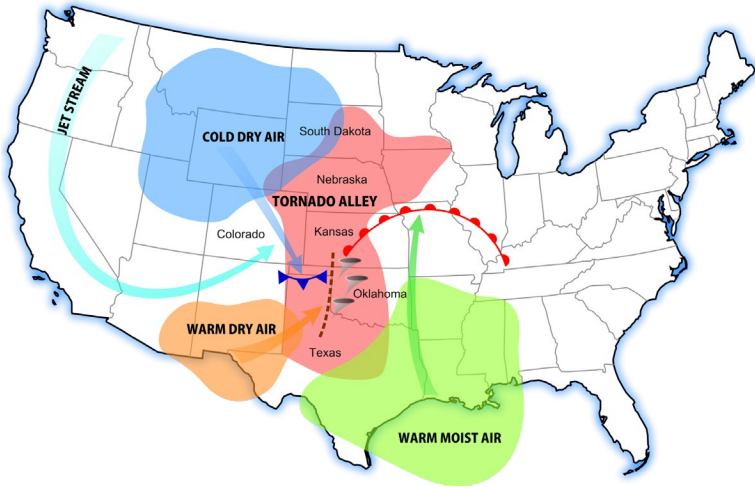
Figure 18.18 "Tornado Alley Diagram" by Dan Craggs is licensed under the Creative Commons Attribution-ShareAlike 3.0 Unported license.

Supercells

The most powerful thunderstorm is a **supercell**, which is oddly enough for a single-cell thunderstorm on steroids. The United States alone has anywhere from 2,000 to 3,000 super-cells a year. Supercells consist of a single powerful thunderstorm that can last several hours and grow to a height of 65,000 feet and last several hours in the mature stage with winds reaching over 100 mph and form the majority of all tornadoes.
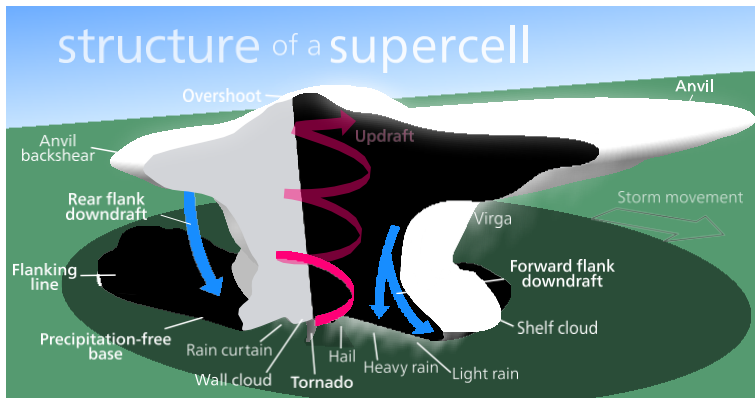
Like severe thunderstorms, supercells may have upper-level winds
that can pull the up-drafts away from the downdrafts. Wind shear
between the upper-level, mid-level, and surface winds can also cause
the air beneath a forming supercell to rotate. For example, if there
are upper-level and mid-level winds flowing from the northwest and
low-level, warm southern surface winds converging together, a
corkscrew flow of air could develop. This rotating column of air is
then picked up by the thunderstorm's updrafts, causing the entire
storm to rotate vertically, called a **mesocyclone**. Sometimes, the
rotating portion of the supercell, the mesocyclone, can extend
below the thunderstorm base, creating a cloud feature called a **wall
cloud**.

Physical Geography and Natural Disasters by R. Adam Dastrup
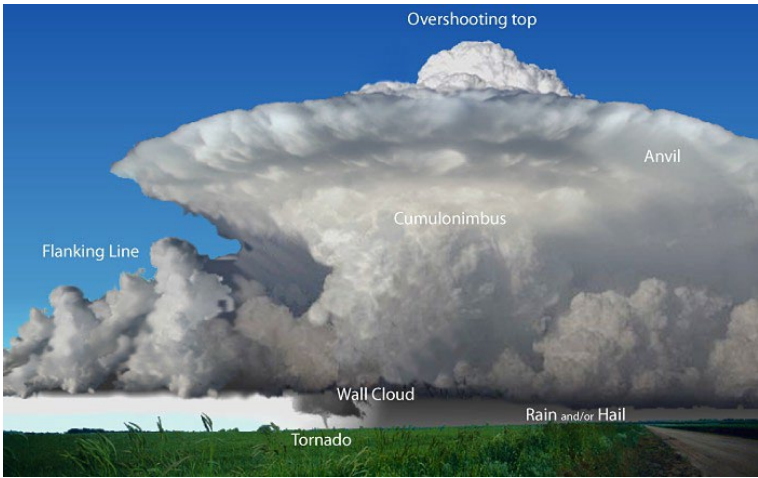
Figure 18.20 "Idealized Supercell" by NOAA Jetstream is licensed as Public Domain.

WEATHER HAZARDS

Tornadoes

One of the most violent and destructive forces of weather are **tornadoes**. The National Weather Service states that "a tornado is a violently rotating (usually counter-clockwise in the northern hemisphere) column of air descending from a thunderstorm and in contact with the ground."

Video 18.30 Tornadoes 101
https://www.youtube.com/watch?v=aacHWoB7cmY

It is pretty much an extension of a supercell's mesocyclone. They range in size from 300 feet to over two miles wide, last minutes to hours, travel a few miles to over 250 miles, at speeds of 30-65 mph. About 75 percent of all the tornadoes in the world occur in the United States; the United States has more tornadoes than the rest of the world combined.

What makes tornadoes so destructive are the wind speeds within them. Atmospheric pressure within a tornado can be 10 percent lower than the tornado's air, causing air to flow into the tornado from all directions. As the air flows into a tornado, the moisture begins to cool and condense into a cloud, allowing it to be seen. Debris picked up by the tornado will also cause it to darken.

Video 18.31 Tornado alley
https://www.youtube.com/watch?v=kE3Afc9uEf8

Tornado Formation

The anatomy and development of tornadoes are not fully understood, but they form from cold fronts, severe thunderstorms, squall lines, supercells, and hurricanes. Geography also plays a crucial role in determining where tornadoes can and can not form. The majority of thunderstorms in the United States form in the Midwest, called **Tornado Alley**. This region is where cP air masses from Canada collide with mT air from the Gulf of Mexico. This **wind shear** creates unstable atmospheric conditions and a rotating corkscrew column of air.
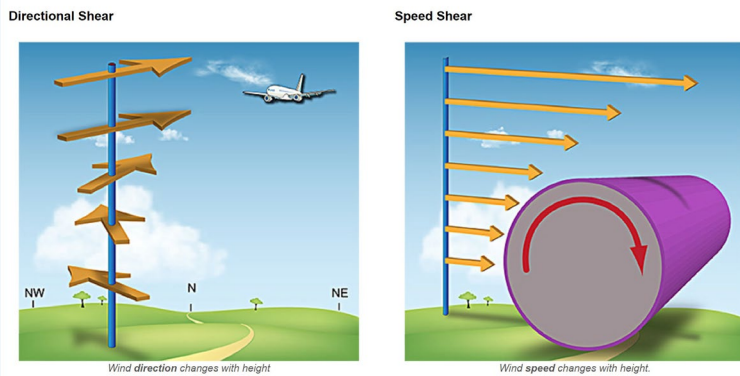
Figure 18.21 "Types of Atmospheric Wind Shear" by NOAA Jetstream is licensed as Public Domain.

Figure 18.22 "Tornado Alley Diagram" by Dan Craggs is licensed under the Creative Commons Attribution-ShareAlike 3.0 Unported license.

As the ground heats up through the day, updrafts pick up the rotating air into a portion of the thunderstorm to developing what is called a rotating **mesocyclone**. The updrafts stretch and tighten the now vertical column of air, causing it to rotate faster, much like an ice skater tightens to spin faster. As the rotating updrafts rise, a rotating wall cloud begins to form from the base of a mesocyclone. Sometimes a funnel cloud may begin to descend from the mesocyclone and may even be called a tornado if it reaches the ground.
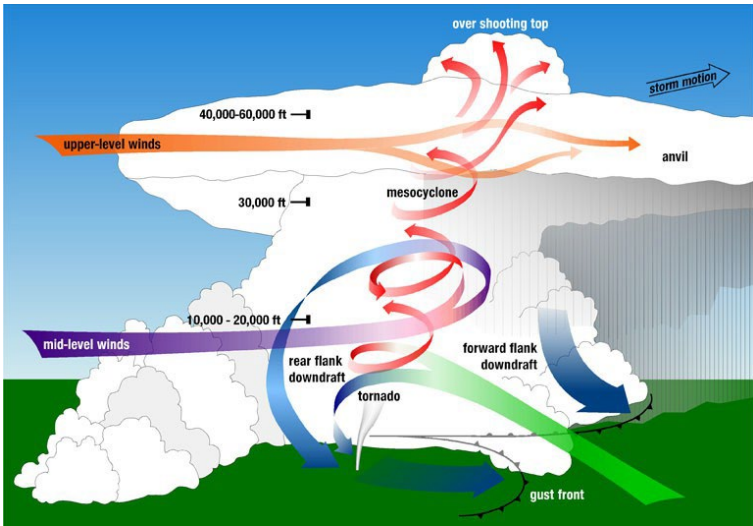
Figure 18.23 "Tornadic Supercell" by the NOAA National Severe Storm Laboratory is licensed as Public Domain.
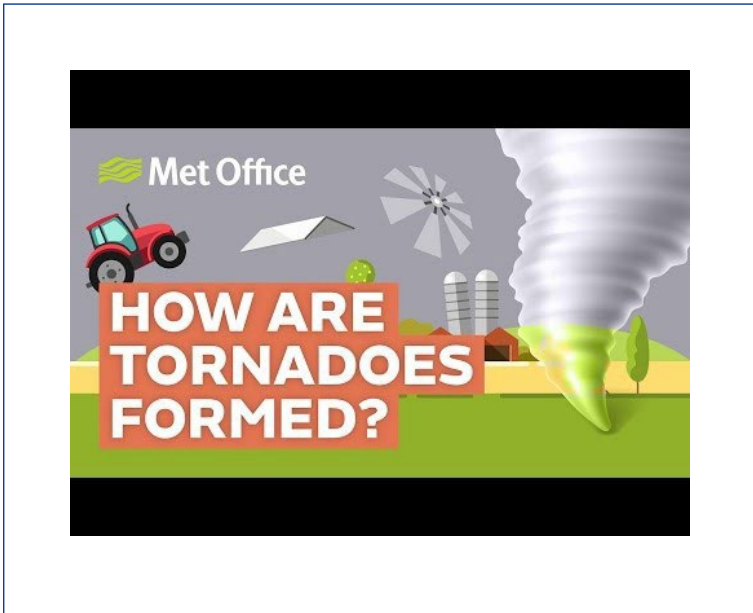
Since 1990, the United States has averaged 1,200 tornadoes reported every year. However, the exact number per year can vary. Because the contrast between cold polar air and warm subtropical air is most significant in the spring and fall, most tornadoes in the United States develop during those seasons with April through June being the greatest.

Video 18.33 Why do tornadoes hate America?

https://www.youtube.com/watch?v=0yiZveJAEp4

Tornado Destruction

The original way tornadoes were classified was based on a scale system developed by Dr. Fujita, which became called the **Fujita Intensity Scale** or simply the F-scale. The Storm Prediction Center states that the Fujita Intensity Scale's goal was to "categorize each tornado by its intensity and its area and estimate a wind speed associated with the damage caused by the tornado." The scale ranged from weak F0s to rare and obliterating F5s.

Enhanced Fujita Scale damage and windspeed estimates						
Damage	Little Damage	Minor Damage	Roof Gone	Walls Collapse	Blown Down	Blown Away
EF5	X	X	X	X	X	X
EF4	X	X	X	X	X	
EF3	X	X	X	X		
EF2	X	X	X			
EF1	X	X				
EF0	X					
Estimated windspeed (mph)	65-85	86-110	111-135	136-165	166-200	200+

Figure 18.24 "Enhanced Fujita Scale Damage and Windspeed Estimates" by the Air Force Medical Service is licensed as Public Domain.

In February 2007, a new classification system began called the **Enhanced Fujita Scale**, or the EF-scale. The scale still ranges from 0-5 and is still an estimate of the tornado's wind speeds. However, the new scale is based on the amount of damage to different structures instead of actual wind speed. The stronger focus on the

616

Physical Geography and Natural Disasters by R. Adam Dastrup

actual damage caused is because some structures and objects are built stronger than others. Examples of various strengths could include wood-framed homes, brick homes, malls, churches, and steel buildings.

The majority of deaths caused by a tornado are not caused by wind, but by the flying debris. Sometimes a tornado will pass through a neighborhood and destroy some houses while leaving others with little damage. Scientists have discovered that some massive, intense tornadoes might contain several smaller tornadoes within them called suction vortices that makeup what is called multiple vortex tornadoes. (Dastrup, 2014)

Tornado Forecasting

A lot of time, money, and research has gone into understanding tornadoes to provide better warnings to the public. When the atmospheric conditions are right, the Storm Prediction Center (SPC), which is part of the National Weather Service, analyzes severe weather and posts their forecasts to alert the public. If the conditions are right for a tornado to form, the SPC will send out a tornado watch alerting the public of the possibility of a tornado in a given area within a specific time. When a tornado has been cited, the National Weather Ser-vice will send a tornado warning out. A

tornado warning warns of a "high probability of imminent danger" to a specific location.
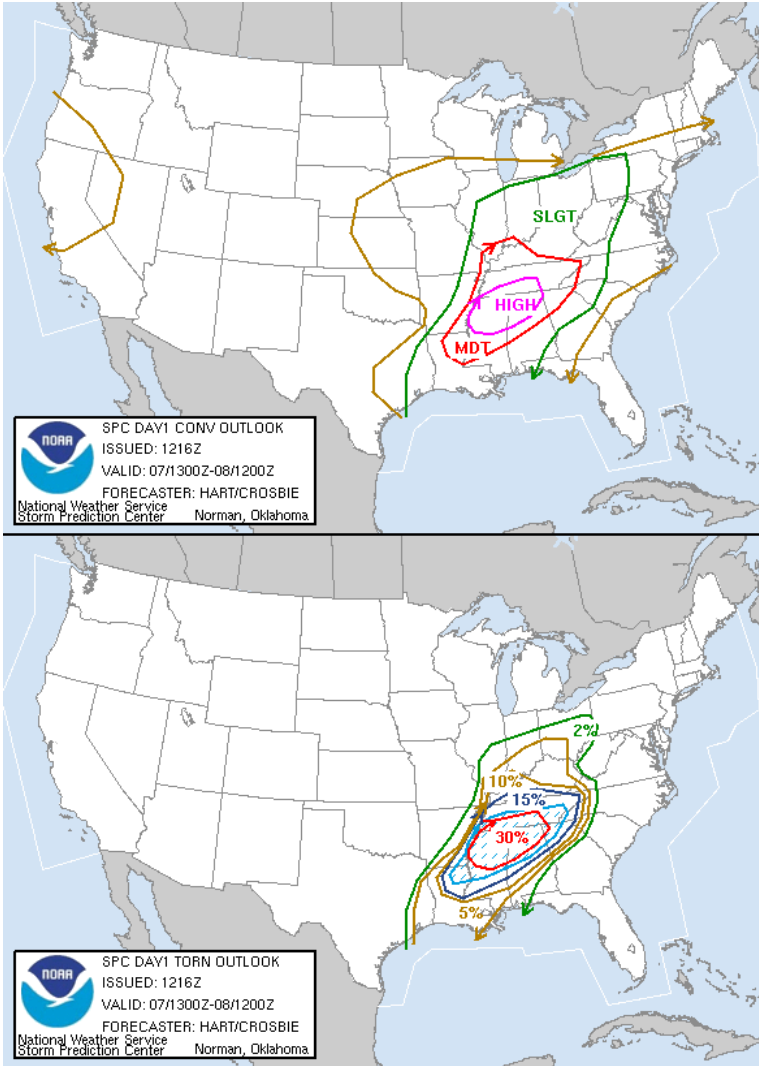
Figure 18.25 "SPC Severe Outlook" by the NOAA Storm Prediction Center is licensed as Public Domain.

Doppler radar has become vital in determining the location and intensity of precipitation and locating where mesocyclones and tornadoes may be forming. Doppler radar can do this by monitoring the wind's flow toward or away from the tower, called the **Doppler effect**. Click here to view an image of an EF5 tornado approaches the Doppler. Notice a hook echo form in the red and begin rotating counter-clockwise, as most low-pressure systems do in the Northern Hemisphere. Weather satellites have also become instrumental in weather forecasting and monitoring. (Dastrup, 2014)
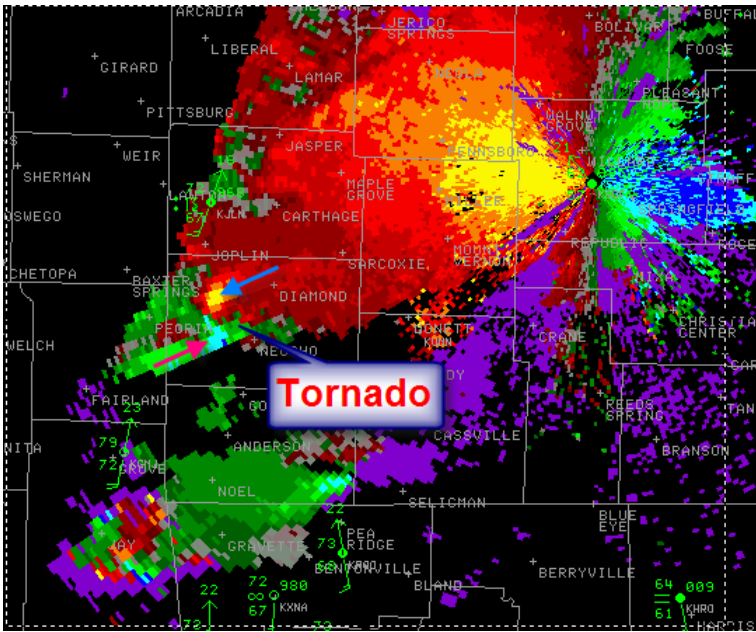
Figure 18.26 "Velocity Missouri Tornado" by the National Weather Service is licensed as Public Domain.

Lightning

Lighting is how the atmosphere balances out the buildup of opposite electrical charges within a cumulonimbus. The buildup occurs when charged atmospheric particles become segregated between a cloud and the ground, between clouds, or from a cloud and surrounding atmosphere. Scientists do not fully understand the reason for charged particle separation. However, some believe that ice crystals called **graupel** and supercooled raindrops interact with each other through collision, creating positive and negative charges that get separated within the cloud by updrafts and downdrafts. Others believe that ice pellet formation within the cloud charges the particles. As an ice pellet forms, the outer shell freezes first and becomes positively charged, and the liquid inner core is negatively charged. As the interior begins to freeze and expand, it breaks apart the positively charged outer layers. These positively charged fragments are taken into the cloud's upper regions, giving the top of the cloud a positive charge. The denser liquid water has a negative charge and is at the base of the cloud.

622
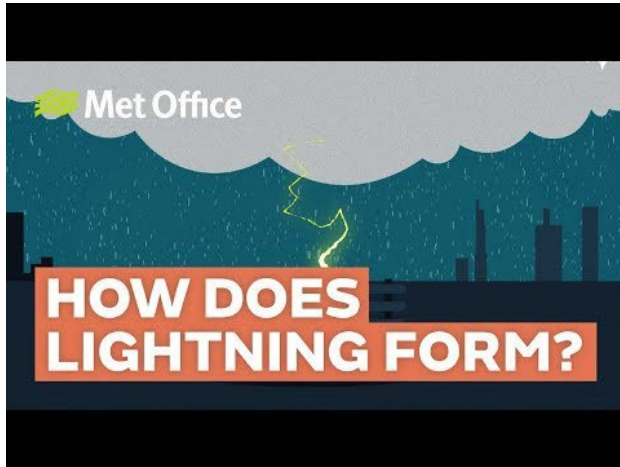
Video 18.34 How does lightening form?
https://www.youtube.com/watch?v=lQGFvKAujb8

Once charged particles are separated, an electrical attraction exists between the negative and positive charges. The air between the oppositely charged particles tries to keep them separated. However, when the air can not keep the two opposite charges separated any longer, the negative charges move towards the positive charge in a zigzag manner called a **stepped leader**. As the stepped leader approaches the ground, the surface's positive charges will rise toward the steeped leader. It may rise through trees, buildings, or humans. When the steeped leader and the rising positively charged particles connect, the negatively charged particles flow from the cloud to the ground. At the same time, an electrical discharge called a return stroke, flows from the ground toward the cloud along the

same path as the stepped leader. The return stroke is the part we see as lightning. When lightning flickers, it is because this process is repeating itself along the same path. Eighty percent of all lightning strikes occur cloud-to-cloud, but this process mentioned above can occur with cloud-to-ground, cloud-to-cloud, and cloud-to-air.

Thunderstorm gathers another pool of positively charged particles.

Negatively charged area in the storm will send out a charge.

Lightning channel develops
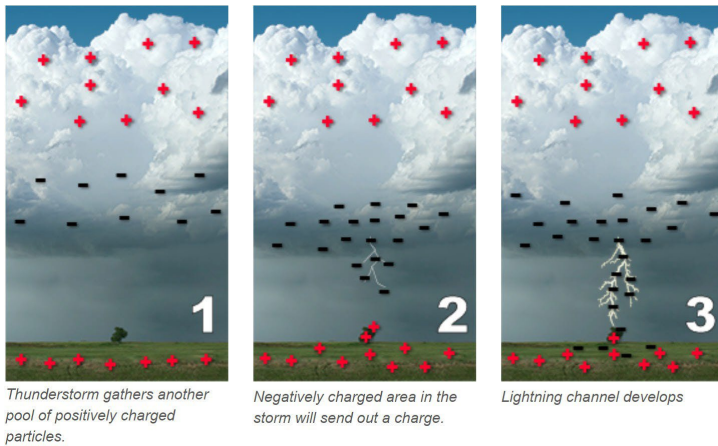
Figure 18.27 "Lightning Stages" by NOAA Jetstream is licensed as Public Domain.

Lightning is only a few centimeters thick but travels at 60,000 mph with a temperature of 54,000 degrees F (five times hotter than the surface of the sun). Lightning causes the surrounding air to become superheated, causing the air to expand violently, creating a shockwave called **thunder**. To estimate how far you are from lightning, count the number of seconds between the flash of lightning and when you hear the thunder reach you, then divide by five. However, if the lightning strike is more than 12 miles away, thunder will not be heard.

Physical Geography and Natural Disasters by R. Adam Dastrup

Video 18.35 How lightening works
https://www.youtube.com/watch?v=9K-v-RJ-z2A&t=1s

In the summertime, convection or orographic uplift can create dry thunderstorms. **Dry thunderstorms** typically mean storms that have lightning, but little to no precipitation falling. Often this occurs when thunderstorms develop in dry atmospheric conditions. Because the air is so dry, rising eddies of moisture must reach great heights before being able to condense. This creates thunderstorms with very high cloud-bases. Any precipitation that does fall in these dry atmospheric conditions is likely to evaporate before reaching the ground. Thus these thunderstorms can create lighting, which can cause ignite wildfires, and the evaporating precipitation can produce strong, dry winds that can "push" the fires around.

Physical Geography and Natural Disasters by R. Adam Dastrup

Video 18.36 The science of lightening

https://www.youtube.com/watch?v=fzNk4w2k2h0

Hail

Hail is precipitation in the form of ice pellets that only forms in cumulonimbus clouds where the lower region of the cloud contains liquid water and the upper region containing ice freezing. When an ice pellet falls within the cumulonimbus cloud, it enters the warm, liquid region and picks up moisture. Then the updrafts through the ice pellet back up above the freezing point hardening the newly gathered water. The ice pellet will fall again to collect liquid water and thrown back up to refreeze. This process will continue until the hailstone's weight becomes too heavy for the updrafts to hold it up.

Once the hail becomes too heavy, the hail will precipitate on the downdraft side of the cumulonimbus. (Weather Processes and Systems | Earth Science, n.d.)

Flash Floods

The number one weather-related cause of death in the United States are **flash floods**. The **National Weather Service** states that "flash floods are short-term events, occurring within 6 hours of the causative event (heavy rain, dam break, levee failure, rapid, snowmelt and ice jams) and often within 2 hours of the start of high-intensity rainfall. A flash flood is characterized by a rapid

stream rise with depths of water that can reach well above the creek banks. Flash flood damage and most fatalities tend to occur in areas immediately adjacent to a stream or arroyo. Additionally, heavy rain falling on steep terrain can weaken the soil and cause mudslides, damaging homes, roads, and property." Urbanized areas are susceptible to flash floods because soil and vegetation are removed and replaced by concrete, roads, and buildings. When intense precipitation occurs, the water has nowhere to go. Learn more about flash floods from the National Weather Service. (Ch 9 Living with Disasters, n.d.)

Video 18.37 The dangers of flash floods
https://www.youtube.com/watch?v=PvJuocemHS4

Ice Storms

Video 18.38 The science behind ice storms
https://slcc.pressbooks.pub/physicalgeography/?p=632

Physical Geography and Natural Disasters by R. Adam Dastrup

9.2 HURRICANES

Tropical cyclones are considered some of the most powerful weather systems on the planet because of their size, strength, and potential loss to life and property. Tropical cyclones go by different names depending on their geographic location. In North and Central America, they are called **hurricanes**. In the northwestern portion of the Pacific Ocean near China and Japan, they are called **typhoons**. Furthermore, in the Indian Ocean and Australia, they are named **cyclones**. They all have winds exceeding 74 mph, can be hundreds of miles wide, and tower over 40,000 feet above sea level. (Dastrup, 2014)

Physical Geography and Natural Disasters by R. Adam Dastrup

Anatomy of a Hurricane

The anatomy of a hurricane is relatively simple, though the processes involved are quite complex. As a low-pressure disturbance forms, the warm, moist air rushes towards the low pressure to rise upward to form towering thunderstorms. Around the low pressure, disturbance in a wall of clouds called an **eyewall**. Within the eyewall, the wind speeds are highest, the clouds are the tallest, atmospheric pressure is at its lowest, and precipitation is most intense.

Video 18.40 Hurricanes 101
https://www.youtube.com/watch?v=LlXVikDkyTg

At the center or heart of the hurricane is called the **eye**. Within the eye of a hurricane, winds are light, precipitation is minimal, and the skies above are occasionally clear. It is the calm region of the tropical storm, but that is what makes it so dangerous. Many people tend to go outside as the eye moves overhead because they believe the storm is over. However, what some do not realize is that "round two" is coming from behind. Moving away from the eyewall are organized, intense thunderstorms, called **spiral rain bands**, that rotate around and toward the storm's eyewall. (Dastrup, 2014)
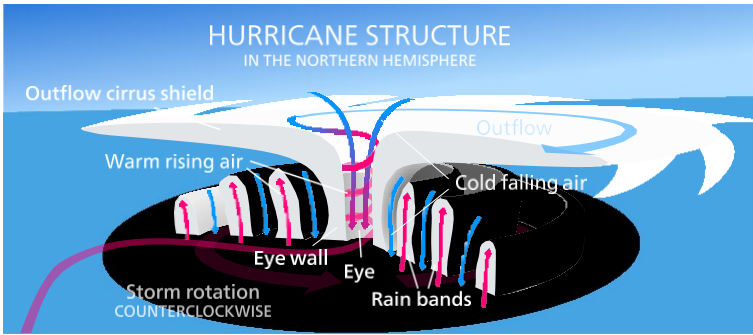
Development of a Hurricane

There are several requirements needed for a hurricane to form. First, hurricanes must form over warm, moist oceans with surface temperatures that are at least 80 degrees Fahrenheit for at least 150 feet deep. This allows for water to evaporate, causing the air to become warm and humid. The atmosphere must also be warm and humid so that latent heat energy could be released as the moisture condenses within a hurricane. The third ingredient is a low-pressure disturbance, called an **easterly wave** or **tropical wave**, which is needed for air to converge toward and ascend to produce powerful thunderstorms.

Once these storms form, most need to be 5-20 degrees north or south of the equator to develop into a hurricane. If the tropical storm forms between 0-5 degrees north or south, the Coriolis effect is not strong enough to cause the storm to organize into a system and rotate. If the storm forms too far north or south, the water is

too cold to develop into a hurricane. Finally, in order for a hurricane to form in the Atlantic Ocean, there must be little to no wind shear. If the jet stream is too low and blowing eastward while the trade winds are blowing west-ward, the wind shear will tear apart the forming cyclone before it can strengthen into a hurricane. Most hurricanes form under large regions of high pressure within the Atlantic Ocean. Recall that high pressure tends to have clear skies and mild winds.

Video 18.41 Engines of destruction
https://www.youtube.com/watch?v=Xtu2_ziBI_w

Now there are five primary stages a hurricane goes through. The first stage goes back to the low-pressure disturbance or easterly wave. Where easterly waves develop is currently being debated. However, it is believed that they form either off the west coast of

Africa near the Canary Islands or eastern Africa near Ethiopia. At this stage, warm, moist air begins to rise within the low-pressure disturbance releasing vast amounts of latent heat because of condensation, causing the storm to intensify. If the easterly waveforms under the right conditions, the Coriolis effect may allow the thunderstorms to organize and rotate, and it becomes a tropical depression. If the tropical storm continues to organize and strengthen with sustained winds between 39-73 mph, the storm is upgraded to a tropical storm and given a name by the National Hurricane Center. If intense rotation continues, the storm grows with winds reaching 74 mph or higher, it becomes a hurricane.

There are a variety of reasons why a hurricane will begin to die off. First, it could move in-to colder water, which does not supply the storm with as much latent heat energy. The idea is that colder water can not evaporate as easily, so no latent heat is released. Another way is if the storm moves over the land, causing it to lose its moisture source below. Finally, a third way a tropical storm could die-off is to move into an area where winds aloft begin to tear the storm apart.
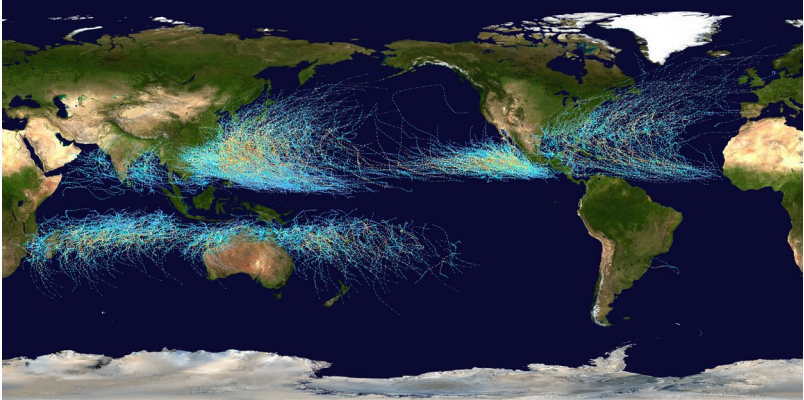
An atmospheric event that can limit hurricane development occurs in the Pacific Ocean called **El Nino**. During an El Nino, the jet stream over the eastern United States migrates into lower latitudes. This upper wind shear caused by the jet stream moving eastward and the trade winds moving westward tends to tear apart hurricanes before they can form.

Another phenomenon that can deduce or eliminate hurricane development is dust storms coming off the coast of the Sahara Desert and into the Atlantic Ocean. In the summer of 2006, following the record activity of 2005 with Hurricane Katrina, the Sahara Desert in Africa had several sandstorms coming off into the Atlantic Ocean. Now recall two things: hurricanes originate near Africa, and the Atlantic Ocean must be over 80 degrees Fahrenheit for them to form. It is believed that the sandstorms blocked the sun over the eastern Atlan-tic, which ultimately cooled the ocean enough to hinder hurricane genesis. (Dastrup, 2014)

Hurricane Classification and Naming

The **Saffir-Simpson Scale** was created to determine the strength and intensity of hurricanes. Just like the Fujita scale, it ranges from 1-5, with category five being the strongest. The intensity of a hurricane increases as the atmospheric pressure near the eye decreases and the pressure-gradient force becomes steeper (atmospheric pressure decreases rapidly towards the eye), causing the winds to intensify, resulting in the potential for more damage. (Dastrup, 2014)

Hurricane strength scale

Saffir-Simpson hurricane scale is used to estimate potential damage and flooding expected along a coast from a hurricane landfall.

CATEGORY 1   **Winds 74-95 mph (119-153 kph)**

Storm surge 4-5 ft. (1.2-1.5 m)
Damage Minimal; signs, tree branches and power lines blown down; damage to mobile homes

CATEGORY 2   **Winds 96-110 mph (154-177 kph)**

Storm surge 6-8 ft. (1.8-2.4 m)
Damage Moderate; some damage to roofs, windows; some downed trees

CATEGORY 3   **Winds 111-129 mph (178-208 kph)**

Storm surge 9-12 ft. (2.7-3.7 m)
Damage Extensive; minor damage to buildings, homes; large trees blown down

CATEGORY 4   **Winds 130-156 mph (209-251 kph)**

Storm surge 13-18 ft. (4.0-5.5 m)
Damage Extreme; almost total destruction of doors, windows; mobile homes destroyed
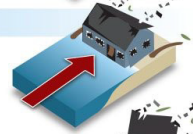
CATEGORY 5   **Winds more than 157 mph (252 kph)**

Storm surge Higher than 18 ft. (5.5 m)
Damage Catastrophic; buildings, roofs, structures destroyed; all trees, shrubs downed

Source: U.S. National Hurricane Center   Graphic: Staff, TNS

Figure 18.30 "Hurricane Strength Scale" by Thomas Cizauskas is licensed as Public Domain.

For many years, hurricanes were named based on where they struck land. For example, the Galveston Hurricane hit Galveston, Texas in 1900. In the early 1900s, scientists started naming hurricanes using female names; clearly a sexist action by mostly white, male scientists at the time. But by the 1950s, the National Hurricane Center started naming hurricanes alphabetically, starting with the letter A, and based on male and female names. If the first year started the

638

letter A with a female name, the following year, the first name would be male. The names are also determined six years in advance so that politics do not get involved with the naming of these deadly storms.

Hazards

Storm Surge

Storm surges, defined as rising ocean levels along a coastal area because of low atmospheric pressure within a tropical cyclone, are the most dangerous hazard created by a tropical cyclone. They are blamed for 90 percent of all deaths and property damage. The height of a storm surge is dependent on the strength of winds, which is dependent on the low pressure near the eye. The low pressure does not "pull up" ocean water to create a storm surge; instead, the wind blowing towards the low pressure causes the water underneath to pile up near the hurricane's eye.
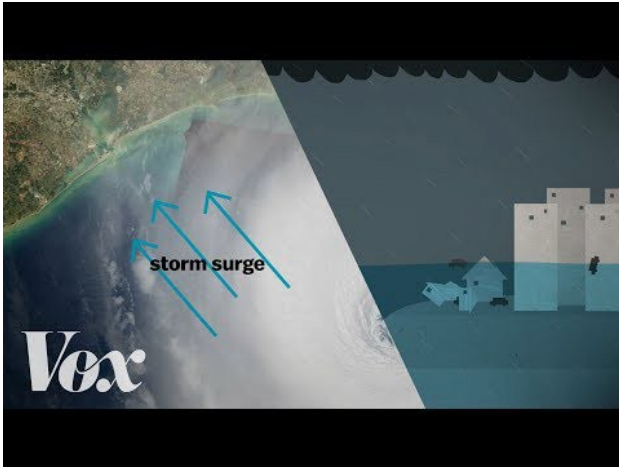
Video 18.42 Why a storm surge can be the deadliest part of a hurricane

https://www.youtube.com/watch?v=TJ5C7TsBsao

As it approaches the shore, the piling causes the sea level to rise and flood the area near the eye and eyewall. Storm surges are most damaging when the atmospheric pressure decrease, if the continental shelf rises slowly as it approaches the shore, or if the storm surge arrives during high tide. The storm surge is also highest on the right side of the eye as winds blow to-ward shore in a counter-clockwise rotation. New Orleans was struck on the left side of the storm surge, so they did not receive the highest portion of the surge. (Dastrup, 2014)

Wind Damage and Inland Flooding

Another hazard created by hurricanes is property damage caused by hurricane-force winds. Though property, such as mobile homes, is obvious, tall buildings are also susceptible to wind damage. People in tall buildings should get below the tenth floor because of powerful winds but should stay above the main floor in case of flooding. Hurricanes also produce tornadoes. It is believed that hurricanes Charley, Frances, Gaston, and Ivan that struck Florida in 2004 produced nearly 300 tornadoes!

Flooding, besides storm surges, is also a dangerous component of hurricanes. The 2004 hurricane season killed over 3,000. The majority of the deaths occurred in Haiti from flash floods and mudslides caused by heavy precipitation. Category 5 Hurricane Mitch in 1998 killed over 18,000 people, most in Central America, dropping nearly three feet of rain, causing massive flooding and mudslides.

Click on the story map below to learn more about tropical cyclones in the United States.

Figure 18.31 "Tropical Cyclone Wind Exposure in the United States" by NOAA is licensed as Public Domain.

Hurricane Forecasting

The governmental agency that analyzes and monitors hurricanes is NOAA's National Hurricane Center. Most hurricanes move westward as they approach North and Central America, then veer northeast. The reason is relatively simple. At around 30 degrees north, there is a sub-tropical high called the Bermuda High. As tropical storm systems move across the Atlantic Ocean, they are pushed westward from the underside of the subtropical high's clockwise rotation (a.k.a. the trade winds). As tropical storms move northward, the subtropical high's clockwise rotation (a.k.a. the westerly winds) pushes the storms northwestward into colder water. (Dastrup, 2014)

Click here to read about how the National Oceanic and Atmospheric Administration (NOAA) prepares for hurricanes season.

ATTRIBUTIONS AND REFERENCES

Creative Commons Attribution

An Introduction to Geology by Chris Johnson, Matthew D. Affolter, Paul Inkenbrandt, Cam Mosher is licensed under CC BY-NC-SA 4.0

Natural Disasters and Human Impacts by R. Adam Dastrup, MA, GISP is licensed under CC BY-NC-SA 4.0

Physical Geology – 2nd Edition by Steven Earle is licensed under CC BY-NC-SA 4.0

Geology by Lumen Learning is licensed under CC BY-NC-SA 4.0

Earth Science by Lumen Learning is licensed under CC BY-NC-SA 4.0

References

A Tornado Hits The Weather Channel. (2018, June 20). https://www.youtube.com/watch?v=0cODBQqaGTw

About NSSL. (n.d.). [Text]. NOAA National Severe Storms

643

Laboratory. Retrieved July 30, 2020, from
https://www.nssl.noaa.gov/about/

Air Masses. (2017, October 22). https://www.youtube.com/
watch?v=QxvF1nnWRwg&feature=emb_logo

Air Masses and Fronts. (2016, August 25).
https://www.youtube.com/
watch?v=o4lg8UfY5DM&feature=emb_logo

Causes of Long-term Climate Change | Physical Geography. (n.d.).
Retrieved May 5, 2020, from https://courses.lumenlearning.com/
geophysical/chapter/causes-of-long-term-climate-change/

Center, N. H. (n.d.). National Hurricane Center. Retrieved May 4,
2020, from https://www.nhc.noaa.gov/

Dastrup, R. A. (2014a). Ch 3 Earth's Atmosphere—Open
Geography Education. https://www.opengeography.org/
ch-3-earths-atmosphere.html

Dastrup, R. A. (2014b). Ch 7 Weather Hazards—Open Geography
Education. In Open Geography Education.
https://www.opengeography.org/ch-7-weather-hazards.html

Dastrup, R. A. (2014c). Ch 8 Severe Weather—Open Geography
Education. https://www.opengeography.org/ch-8-severe-
weather.html

Dastrup, R. A. (2014d). Ch 9 Hurricanes—Open Geography
Education. In Open Geography Education.
https://www.opengeography.org/ch-9-hurricanes.html

Dastrup, R. A. (2014e). Ch 9 Living with Disasters—Open
Geography Education. In Open Geography Education.
https://www.opengeography.org/ch-9-living-with-disasters.html

Dastrup, R. A. (2014f). Ch 10 Climatic Disasters—Open
Geography Education. In Open Geography Education.
https://www.opengeography.org/ch-10-climatic-disasters.html

Engines of Destruction: The Science of Hurricanes! (n.d.).
Retrieved May 4, 2020, from https://www.youtube.com/
watch?v=Xtu2_ziBI_w

Forces of Nature | National Geographic Society. (n.d.). Retrieved
May 4, 2020, from https://www.nationalgeographic.org/
interactive/forces-nature/

How are thunderstorms formed? | Weather Wise S2E1. (n.d.).
Retrieved May 4, 2020, from https://www.youtube.com/
watch?v=Z7c4bZwvkrs&feature=emb_logo

How are tornadoes formed? (n.d.). Retrieved May 4, 2020, from
https://www.youtube.com/watch?v=Ay69rQPMK3s

How does lightning form? (2018, April 6).
https://www.youtube.com/
watch?v=lQGFvKAujb8&feature=emb_logo

How does lightning work? Where does it come from? | Weather
Wise S2E2. (n.d.). Retrieved May 4, 2020, from
https://www.youtube.com/watch?v=9K-v-RJ-
z2A&feature=emb_logo

How Scientists Predict Tornado Damage | Mega Disaster. (n.d.). Retrieved August 2, 2020, from https://www.youtube.com/watch?v=HgZZnwcFytE

How to read a synoptic chart. (2013, July 8). https://www.youtube.com/watch?v=wl_FFK_HbjY&feature=emb_logo

Microbursts Explained. (n.d.). Retrieved May 4, 2020, from https://www.youtube.com/watch?v=Fdt6-si2UFo&feature=emb_logo

National Geographic. (n.d.). Hurricanes 101. Retrieved May 4, 2020, from https://www.youtube.com/watch?v=LlXVikDkyTg

National Oceanic and Atmospheric Administration. (n.d.). Retrieved May 4, 2020, from https://www.noaa.gov/

NHC Track and Intensity Models. (n.d.). Retrieved August 8, 2020, from https://www.nhc.noaa.gov/modelsummary.shtml

NOAA National Severe Storms Laboratory. (n.d.). Retrieved July 30, 2020, from https://www.nssl.noaa.gov/

NOAA/NWS Storm Prediction Center. (n.d.). Retrieved May 4, 2020, from https://www.spc.noaa.gov/

NWS JetStream—Thunderstorm Hazards: Damaging Wind. (n.d.). Retrieved May 4, 2020, from https://www.weather.gov/jetstream/wind_damage

NWS JetStream—Tropical Cyclone Structure. (n.d.). Retrieved

August 8, 2020, from https://www.weather.gov/jetstream/
tc_structure

Saffir-Simpson Hurricane Wind Scale. (n.d.). Retrieved August 8,
2020, from https://www.nhc.noaa.gov/aboutsshws.php

Service, N. N. W. (n.d.). Glossary—NOAA's National Weather
Service. Retrieved May 4, 2020, from https://w1.weather.gov/
glossary/index.php?word=outflow+boundary

Seven Incredible lightning types | Amazing Weather. (n.d.).
Retrieved May 4, 2020, from https://www.youtube.com/
watch?v=Tw_6U1iOUVQ&feature=emb_logo

Sprites, Jets, and Glowing Balls: The Science of Lightning. (2016,
June 23). https://www.youtube.com/
watch?v=fzNk4w2k2h0&feature=emb_logo

Storm Surge Overview. (n.d.). Retrieved August 8, 2020, from
https://www.nhc.noaa.gov/surge/

The Dangers of Flash Flooding. (n.d.). Retrieved May 4, 2020,
from https://www.youtube.com/
watch?v=PvJuocemHS4&feature=emb_logo

The Science Behind Ice Storms. (2019, January 17).
https://www.youtube.com/
watch?time_continue=178&v=wfzm1M_iSos&feature=emb_logo

Thunderstorms 101 | National Geographic. (n.d.). Retrieved May
4, 2020, from https://www.youtube.com/
watch?v=zUNEFefftt8&feature=emb_logo

Tornadoes 101 | National Geographic. (n.d.). Retrieved May 4, 2020, from https://www.youtube.com/ watch?v=aacHWoB7cmY&feature=emb_logo

Tropical Cyclone Names. (n.d.). Retrieved August 8, 2020, from https://www.nhc.noaa.gov/aboutnames.shtml

Tropical Cyclone Wind Exposure in the United States. (n.d.). Retrieved August 2, 2020, from https://www.arcgis.com/apps/ MapJournal/ index.html?appid=197837baeb0046aaa91cf3fc4d62d44d

US Department of Commerce, N. (n.d.-a). Hurricane Safety Tips and Resources. NOAA's National Weather Service. Retrieved August 8, 2020, from https://www.weather.gov/safety/hurricane

US Department of Commerce, N. (n.d.-b). NWS JetStream – Thunderstorm Hazards: Flash Floods. NOAA's National Weather Service. Retrieved May 4, 2020, from https://www.weather.gov/ jetstream/flood

US Department of Commerce, N. (n.d.-c). NWS JetStream – Thunderstorm Hazards: Hail. NOAA's National Weather Service. Retrieved May 4, 2020, from https://www.weather.gov/jetstream/ hail

US Department of Commerce, N. (n.d.-d). NWS JetStream—Air Masses. Retrieved May 4, 2020, from https://www.youtube.com/ watch?v=QxvF1nnWRwg&feature=emb_logo

US Department of Commerce, N. (n.d.-e). NWS JetStream—How to read "Surface" weather maps. NOAA's National Weather

Service. Retrieved August 8, 2020, from https://www.weather.gov/ jetstream/wxmaps

US Department of Commerce, N. (n.d.-f). NWS JetStream—How to read the NHC Tropical Cyclone Forecast/Advisory. NOAA's National Weather Service. Retrieved August 8, 2020, from https://www.weather.gov/jetstream/tc_tcm

US Department of Commerce, N. (n.d.-g). NWS JetStream—Hurricane Damage Potential. NOAA's National Weather Service. Retrieved August 8, 2020, from https://www.weather.gov/jetstream/tc_potential

US Department of Commerce, N. (n.d.-h). NWS JetStream—Introduction to Lightning. NOAA's National Weather Service. Retrieved May 4, 2020, from https://www.weather.gov/ jetstream/lightning_intro

US Department of Commerce, N. (n.d.-i). NWS JetStream—Introduction to Tropical Cyclones. NOAA's National Weather Service. Retrieved August 8, 2020, from https://www.weather.gov/jetstream/tc

US Department of Commerce, N. (n.d.-j). NWS JetStream—Life Cycle of a Thunderstorm. NOAA's National Weather Service. Retrieved May 4, 2020, from https://www.weather.gov/jetstream/ life

US Department of Commerce, N. (n.d.-k). NWS JetStream—Norwegian Cyclone Model. NOAA's National Weather Service. Retrieved August 8, 2020, from https://www.weather.gov/jetstream/cyclone

US Department of Commerce, N. (n.d.-l). NWS JetStream—Tropical Cyclone Classification. NOAA's National Weather Service. Retrieved August 8, 2020, from https://www.weather.gov/jetstream/tc_classification

US Department of Commerce, N. (n.d.-m). NWS JetStream—Tropical Cyclone Hazards. NOAA's National Weather Service. Retrieved August 8, 2020, from https://www.weather.gov/jetstream/tc_hazards

US Department of Commerce, N. (n.d.-n). NWS JetStream—Tropical Cyclone Names. NOAA's National Weather Service. Retrieved August 8, 2020, from https://www.weather.gov/jetstream/tc_names

US Department of Commerce, N. (n.d.-o). NWS JetStream—Tropical Cyclone Safety. NOAA's National Weather Service. Retrieved August 8, 2020, from https://www.weather.gov/jetstream/tc_safety

US Department of Commerce, N. (n.d.-p). NWS JetStream—Types of Thunderstorms. NOAA's National Weather Service. Retrieved May 4, 2020, from https://www.weather.gov/jetstream/tstrmtypes

Weather Processes and Systems | Earth Science. (n.d.). Retrieved May 5, 2020, from https://courses.lumenlearning.com/earthscience/chapter/weather-processes-and-systems/

What are air masses? (2012, May 8). https://www.youtube.com/watch?v=kvk-hBFnBTI

What are polar lows? (2018, November 9).
https://www.youtube.com/
watch?v=H5VNPfHIgpE&feature=emb_logo

What are weather fronts? (2012, May 8).
https://www.youtube.com/
watch?v=G7Ewqm0YHUI&feature=emb_logo

What are weather fronts and how do they affect our weather? (n.d.).
Retrieved May 4, 2020, from https://www.youtube.com/
watch?v=dwIQds-4I7I&feature=emb_logo

What is a blocking high? (2018, September 20).
https://www.youtube.com/watch?v=b-PvxQ1jobU

What is hail? How is hail formed and why does it happen? |
Weather Wise S2E3. (n.d.). Retrieved May 4, 2020, from
https://www.youtube.com/watch?v=6M-
ycZLSF1w&feature=emb_logo

What is the difference between hurricanes, typhoons and cyclones?
(n.d.). Retrieved May 4, 2020, from https://www.youtube.com/
watch?v=jImZLBmM_kA

Where is tornado alley and why does it exist? (n.d.). Retrieved May
4, 2020, from https://www.youtube.com/
watch?v=kE3Afc9uEf8&feature=emb_logo

Why a storm surge can be the deadliest part of a hurricane. (n.d.).
Retrieved May 4, 2020, from https://www.youtube.com/
watch?v=TJ5C7TsBsao

Why Do Tornadoes Hate America? (2013, July 4). https://www.youtube.com/ watch?v=0yiZveJAEp4&feature=emb_logo

Why the US has so many tornadoes. (n.d.). Retrieved May 4, 2020, from https://www.youtube.com/watch?v=lsEA9tGMFQQ

Chapter 19

Geological Time

In 1788, after many years of geological study, James Hutton, one of the great pioneers of geology, wrote the following about the age of Earth: *The result, therefore, of our present enquiry is, that we find no vestige of a beginning — no prospect of an end.*[1] Of course he wasn't exactly correct, there was a beginning and there will be an end to Earth, but what he was trying to express is that geological time is so vast that we humans, who typically live for less than a century, have no means of appreciating how much geological time there is. Hutton didn't even try to assign an age to Earth, but we now know that it is approximately 4,570 million years old. Using the scientific notation for geological time, that is 4,570 **Ma** (for *mega annum* or "millions of years") or 4.57 **Ga** (for *giga annum* or billions of years). More recent dates can be expressed in **ka** (*kilo annum*); for example, the last cycle of glaciation ended at approximately 11.7 ka or 11,700 years ago. This notation will be used for geological dates throughout this book.

Unfortunately, knowing how to express geological time doesn't really help us understand or appreciate its extent. A version of the geological time scale is included as Figure 19.1. Unlike time scales you'll see in other places, or even later in this book, this time scale is linear throughout its length, meaning that 50 Ma during the **Cenozoic** is the same thickness as 50 Ma during the **Hadean**—in each case about the height of the "M" in Ma. The Pleistocene glacial epoch began at about 2.6 Ma, which is equivalent to half the thickness of the thin grey line at the top of the yellow bar marked "Cenozoic." Most other time scales have earlier parts of Earth's history compressed so that more detail can be shown for the more recent parts. That makes it difficult to appreciate the extent of geological time.

To create some context, the **Phanerozoic** Eon (the last 542 million years) is named for the time during which visible (*phaneros*) life (*zoi*) is present in the geological record. In fact, large organisms — those that leave fossils visible to the naked eye — have existed for a little longer than that, first appearing around 600 Ma, or a span of just over 13% of geological time. Animals have been on land for 360 million years, or 8% of geological time. Mammals have

---

1. Hutton, J, 1788. Theory of the Earth; or an investigation of the laws observable in the composition, dissolution, and restoration of land upon the Globe. Transactions of the Royal Society of Edinburgh.
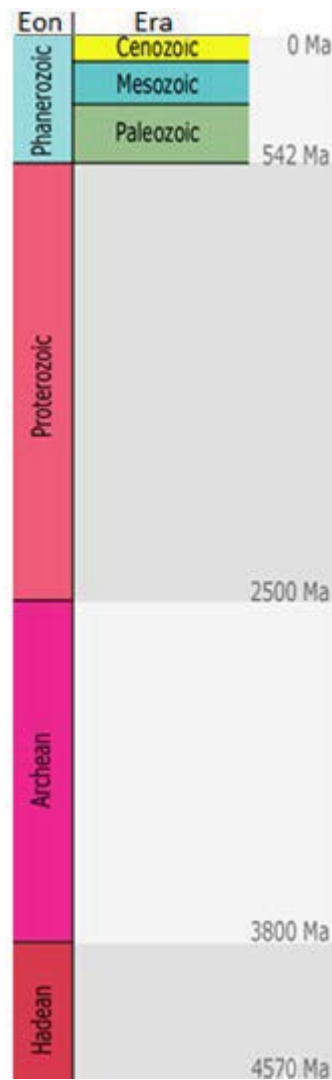
Figure 19.1 The geological time scale [SE]

dominated since the demise of the dinosaurs around 65 Ma, or 1.5% of geological time, and the genus *Homo* has existed since approximately 2.2 Ma, or 0.05% (1/2,000th) of geological time.

Geologists (and geology students) need to understand geological time. That doesn't mean simply memorizing the geological time scale; instead, it means getting your mind around the concept that although most geological processes are extremely slow, very large and important things can happen if such processes continue for enough time.

For example, the Atlantic Ocean between Nova Scotia and northwestern Africa has been getting wider at a rate of about 2.5 cm per year. Imagine yourself taking a journey at that rate — it would be impossibly and ridiculously slow. And yet, since it started to form around 200 Ma (just 4% of geological time), the Atlantic Ocean has grown to a width of over 5,000 km!

A useful mechanism for understanding geological time is to scale it all down into one year. The origin of the solar system and Earth at 4.57 Ga would be represented by January 1, and the present year would be represented by the last tiny fraction of a second on New Year's Eve. At this scale, each day of the year represents 12.5 million

654

years; each hour represents about 500,000 years; each minute represents 8,694 years; and each second represents 145 years. Some significant events in Earth's history, as expressed on this time scale, are summarized on Table 19.1.

Event	Approximate Date	Calendar Equivalent
Formation of oceans and continents	4.5 – 4.4 Ga	January
Evolution of the first primitive life forms	3.8 Ga	early March
Formation of British Columbia's oldest rocks	2.0 Ga	July
Evolution of the first multi-celled animals	0.6 Ga or 600 Ma	November 15
Animals first crawled onto land	360 Ma	December 1
Vancouver Island reached North America and the Rocky Mountains were formed	90 Ma	December 25
Extinction of the non-avian dinosaurs	65 Ma	December 26
Beginning of the Pleistocene ice age	2 Ma or 2000 ka	8 p.m., December, 31
Retreat of the most recent glacial ice from southern Canada	14 ka	11:58 p.m., December 31
Arrival of the first people in British Columbia	10 ka	11:59 p.m., December 31
Arrival of the first Europeans on the west coast of what is now Canada	250 years ago	2 seconds before midnight, December 31

Table 19.1 A summary of some important geological dates expressed as if all of geological time was condensed into one year [SE]

The geological time scale is currently maintained by the International Commission on Stratigraphy (ICS), which is part of the International Union of Geological Sciences. The time scale is continuously being updated

as we learn more about the timing and nature of past geological events. You can view the ICS time scale at https://stratigraphy.org/chart#latest-version It would be a good idea to print a copy (in colour) to put on your wall while you are studying geology.

Geological time has been divided into four eons: Hadean, Archean, Proterozoic, and Phanerozoic, and as shown in Figure 19.2, the first three of these represent almost 90% of Earth's history. The last one, the Phanerozoic (meaning "visible life"), is the time that we are most familiar with because Phanerozoic rocks are the most common on Earth, and they contain evidence of the life forms that we are all somewhat familiar with.
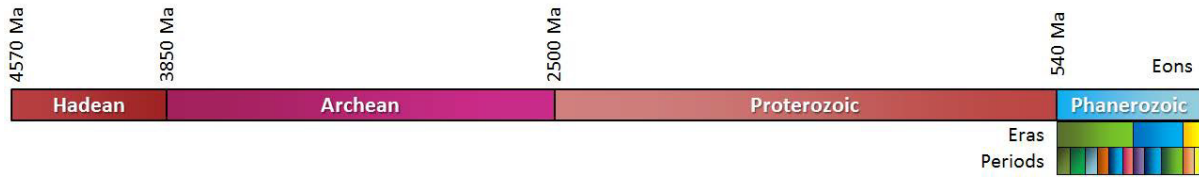

Figure 19.2 The eons of Earth's history [SE]

The Phanerozoic — the past 540 Ma of Earth's history — is divided into three eras: the Paleozoic ("early life"), the Mesozoic ("middle life"), and the Cenozoic ("new life"), and each of these is divided into a number of periods (Figure 19.3). Most of the organisms that we share Earth with evolved at various times during the Phanerozoic.
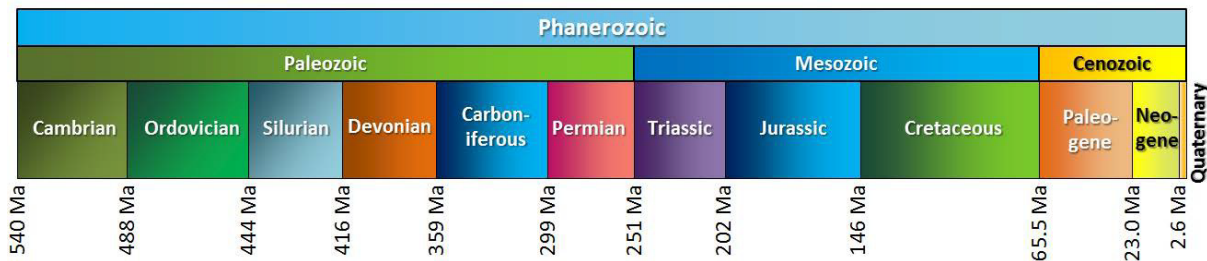

Figure 19.3 The eras (middle row) and periods (bottom row) of the Phanerozoic [SE]

The Cenozoic, which represents the past 65.5 Ma, is divided into three periods: Paleogene, Neogene, and Quaternary, and seven epochs (Figure 19.4). Dinosaurs became extinct at the start of the Cenozoic, after which birds and mammals radiated to fill the available habitats. Earth was very warm during the early Eocene and has steadily cooled ever since. Glaciers first appeared on Antarctica in the Oligocene and then on Greenland in the Miocene, and covered much of North America and Europe by the Pleistocene. The most recent of the Pleistocene glaciations ended around 11,700 years ago. The current epoch is known as the Holocene. Epochs are further divided into ages (a.k.a. stages), but we won't be going into that level of detail here.
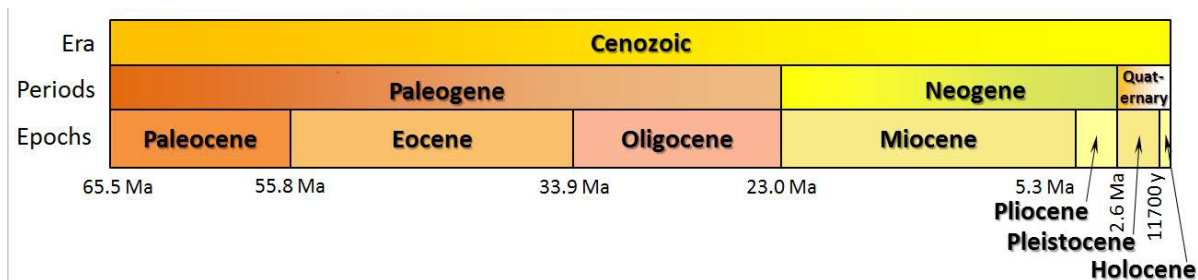

Figure 19.4 The periods (middle row) and epochs (bottom row) of the Cenozoic [SE]

Most of the boundaries between the periods and epochs of the geological time scale have been fixed on the basis

656

of significant changes in the fossil record. For example, as already noted, the boundary between the Cretaceous and the Paleogene coincides exactly with the extinction of the dinosaurs. That's not a coincidence. Many other types of organisms went extinct at this time, and the boundary between the two periods marks the division between sedimentary rocks with Cretaceous organisms below, and Paleogene organisms above.

Relative Dating Methods

The simplest and most intuitive way of dating geological features is to look at the relationships between them. There are a few simple rules for doing this. For example, the principle of superposition states that sedimentary layers are deposited in sequence, and, unless the entire sequence has been turned over by tectonic processes or disrupted by faulting, the layers at the bottom are older than those at the top. The **principle of inclusions** states that any rock fragments that are included in rock must be older than the rock in which they are included. For example, a **xenolith** in an igneous rock or a clast in sedimentary rock must be older than the rock that includes it (Figure 19.5a).

Figure 19.5a A xenolith of diorite incorporated into a basalt lava flow, Mauna Kea volcano, Hawaii. The lava flow took place some time after the diorite cooled, was uplifted, and then eroded. (Hammerhead for scale) [SE]

The **principle of cross-cutting relationships** states that any geological feature that cuts across, or disrupts another feature must be younger than the feature that is disrupted. An example of this is given in Figure 19.5b, which shows three different sedimentary layers. The lower sandstone layer is disrupted by two **faults**, so we can infer that the faults are younger than that layer. But the faults do not appear to continue into the coal seam, and they certainly do not continue into the upper sandstone. So we can infer that coal seam is younger than the faults (because it disrupts them), and of course the upper sandstone is youngest of all, because it lies on top of the coal seam.

Figure 19.5b Rip-up clasts of shale embedded in Gabriola Formation sandstone, Gabriola Island, B.C. The pieces of shale were eroded as the sandstone was deposited, so the shale is older than the sandstone. [SE]

Figure 19.6 Superposition and cross-cutting relationships in Cretaceous Nanaimo Group rocks in Nanaimo, B.C. The coal seam is about 50 cm thick. [SE]

An **unconformity** represents an interruption in the process of deposition of sedimentary rocks. Recognizing unconformities is important for understanding time relationships in sedimentary sequences. An example of an unconformity is shown in Figure 19.7. The Proterozoic rocks of the Grand Canyon Group have been tilted and then eroded to a flat surface prior to deposition of the younger Paleozoic rocks. The difference in time between the youngest of the Proterozoic rocks and the oldest of the Paleozoic rocks is close to 300 million years. Tilting and erosion of the older rocks took place during this time, and if there was any deposition going on in this area, the evidence of it is now gone.

There are four types of unconformities, as summarized in Table 8.1, and illustrated in Figure 19.8.

Figure 19.7 The great angular unconformity in the Grand Canyon, Arizona. The tilted rocks at the bottom are part of the Proterozoic Grand Canyon Group (aged 825 to 1,250 Ma). The flat-lying rocks at the top are Paleozoic (540 to 250 Ma). The boundary between the two represents a time gap of nearly 300 million years. [SE]

Unconformity Type	Description
Nonconformity	A boundary between non-sedimentary rocks (below) and sedimentary rocks (above)
Angular unconformity	A boundary between two sequences of sedimentary rocks where the underlying ones have been tilted (or folded) and eroded prior to the deposition of the younger ones (as in Figure 8.8)
Disconformity	A boundary between two sequences of sedimentary rocks where the underlying ones have been eroded (but not tilted) prior to the deposition of the younger ones (as in Figure 8.7)
Paraconformity	A time gap in a sequence of sedimentary rocks that does not show up as an angular unconformity or a disconformity

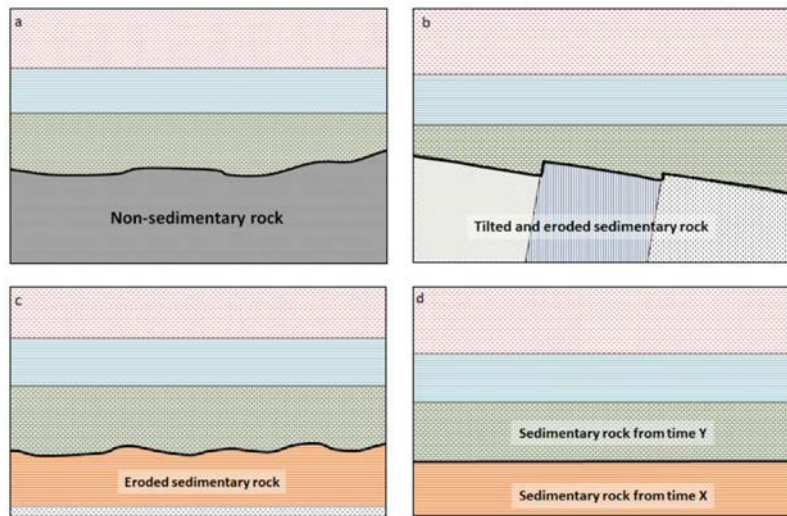Table 19.1 The characteristics of the four types of unconformities

Figure 19.8 The four types of unconformities: (a) a nonconformity between non-sedimentary rock and sedimentary rock, (b) an angular unconformity, (c) a disconformity between layers of sedimentary rock, where the older rock has been eroded but not tilted, and (d) a paraconformity where there is a long period (millions of years) of non-deposition between two parallel layers. [SE]

662

Dating Rocks Using Fossils

Geologists get a wide range of information from fossils. They help us to understand evolution and life in general; they provide critical information for understanding depositional environments and changes in Earth's climate; and, of course, they can be used to date rocks.

Although the recognition of fossils goes back hundreds of years, the systematic cataloguing and assignment of relative ages to different organisms from the distant past — paleontology — only dates back to the earliest part of the 19th century. The oldest undisputed fossils are from rocks dated around 3.5 Ga, and although fossils this old are typically poorly preserved and are not useful for dating rocks, they can still provide important information about conditions at the time. The oldest well-understood fossils are from rocks dating back to around 600 Ma, and the sedimentary record from that time forward is rich in fossil remains that provide a detailed record of the history of life. However, as anyone who has gone hunting for fossils knows, that does not mean that all sedimentary rocks have visible fossils or that they are easy to find. Fossils alone cannot provide us with numerical ages of rocks, but over the past century geologists have acquired enough isotopic dates from rocks associated with fossil-bearing rocks (such as igneous dykes cutting through sedimentary layers) to be able to put specific time limits on most fossils.

A very selective history of life on Earth over the past 600 million years is provided in Figure 19.9. The major groups of organisms that we are familiar with evolved between the late Proterozoic and the Cambrian (~600 Ma to ~520 Ma). Plants, which evolved in the oceans as green algae, came onto land during the Ordovician (~450 Ma). Insects, which evolved from marine arthropods, came onto land during the Devonian (400 Ma), and amphibians (i.e., vertebrates) came onto land about 50 million years later. By the late Carboniferous, trees had evolved from earlier plants, and reptiles had evolved from amphibians. By the mid-Triassic, dinosaurs and mammals had evolved from very different branches of the reptiles; birds evolved from dinosaurs during the Jurassic. Flowering plants evolved in the late Jurassic or early Cretaceous. The earliest primates evolved from other mammals in the early Paleogene, and the genus *Homo* evolved during the late Neogene (~2.8 Ma).
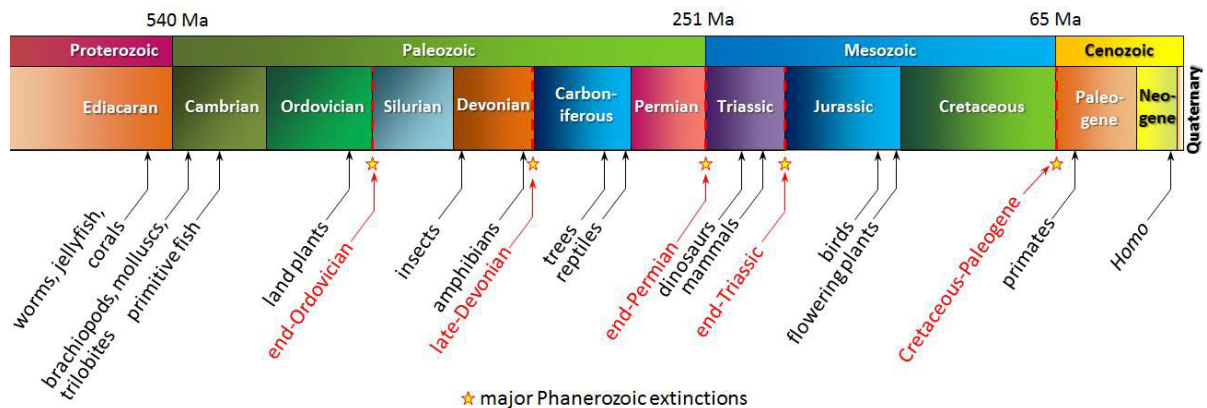

Figure 19.9 A summary of life on Earth during the late Proterozoic and the Phanerozoic. The top row shows geological eras, and the lower row shows the periods. [SE]

If we understand the sequence of evolution on Earth, we can apply knowledge to determining the relative ages of rocks. This is William Smith's principle of faunal succession, although of course it doesn't just apply to "fauna" (animals); it can also apply to fossils of plants and those of simple organisms.

The Phanerozoic has seen five major extinctions, as indicated in Figure 19.9. The most significant of these was at the end of the Permian, which saw the extinction of over 80% of all species and over 90% of all marine

species. Most well-known types of organisms were decimated by this event, but only a few became completely extinct, including trilobites. The second most significant extinction was at the Cretaceous-Paleogene boundary (K-Pg, a.k.a. the K-T extinction). At that time, about 75% of marine species disappeared. Again, a few well-known types of organisms disappeared altogether, including dinosaurs (but not birds) and the pterosaurs. Other types were badly decimated but survived, and then flourished in the Paleogene. The K-Pg extinction is thought to have been caused by the impact of a large extraterrestrial body (10 km to 15 km across), but it is generally agreed that the other four Phanerozoic extinctions had other causes, although their exact nature is not clearly understood.

As already stated, it is no coincidence that the major extinctions all coincide with boundaries of geological periods and even eras. Paleontologists have placed most of the divisions of the geological time scale at points in the fossil record where there are major changes in the type of organisms observed.

If we can identify a fossil to the species level, or at least to the genus level, and we know the time period when the organism lived, we can assign a range of time to the rock. That range might be several million years because some organisms survived for a very long time. If the rock we are studying has several types of fossils in it, and we can assign time ranges to those fossils, we might be able to narrow the time range for the age of the rock considerably. An example of this is given in Figure 19.10.
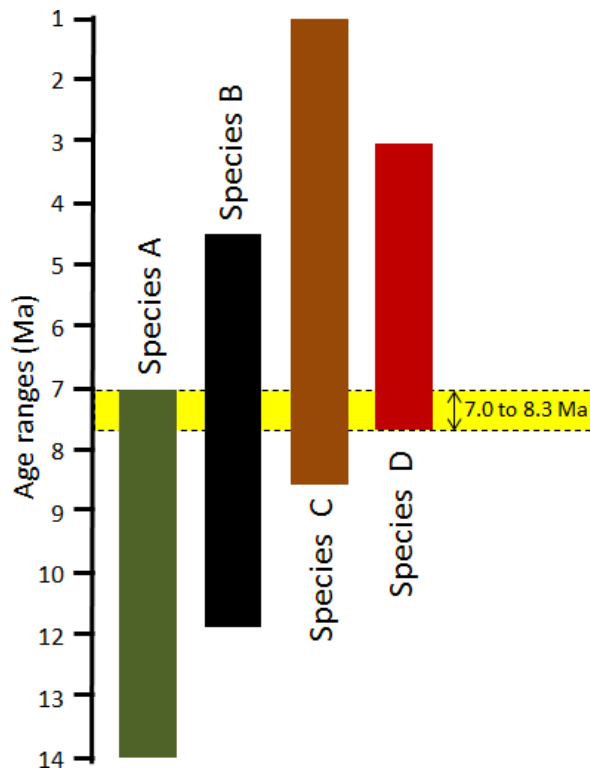

Figure 19.10 The application of bracketing to constrain the age of a rock based on several fossils. In this diagram, the coloured bar represents the time range during which each of the four species (A – D) existed on Earth. Although each species lived for several million years, we can narrow down the likely age of the rock to just 0.7 Ma during which all four species coexisted. [SE]

Some organisms survived for a very long time, and are not particularly useful for dating rocks. Sharks, for example, have been around for over 400 million years, and the great white shark has survived for 16 million years, so far. Organisms that lived for relatively short time periods are particularly useful for dating rocks, especially if they were distributed over a wide geographic area and so can be used to compare rocks from different regions. These are

664

known as **index fossils**. There is no specific limit on how short the time span has to be to qualify as an index fossil. Some lived for millions of years, and others for much less than a million years.

Some well-studied groups of organisms qualify as **biozone** fossils because, although the genera and families lived over a long time, each species lived for a relatively short time and can be easily distinguished from others on the basis of specific features. For example, ammonites have a distinctive feature known as the **suture** line — where the internal shell layers that separate the individual chambers (**septae**) meet the outer shell wall, as shown in Figure 19.11. These suture lines are sufficiently variable to identify species that can be used to estimate the relative or absolute ages of the rocks in which they are found.

Figure 19.11 The septum of an ammonite (white part, left), and the suture lines where the septae meet the outer shell (right). [SE]

Foraminifera (small, carbonate-shelled marine organisms that originated during the Triassic and are still around today) are also useful biozone fossils. As shown in Figure 19.12, numerous different foraminifera lived during the Cretaceous. Some lasted for over 10 million years, but others for less than 1 million years. If the foraminifera in a rock can be identified to the species level, we can get a good idea of its age.
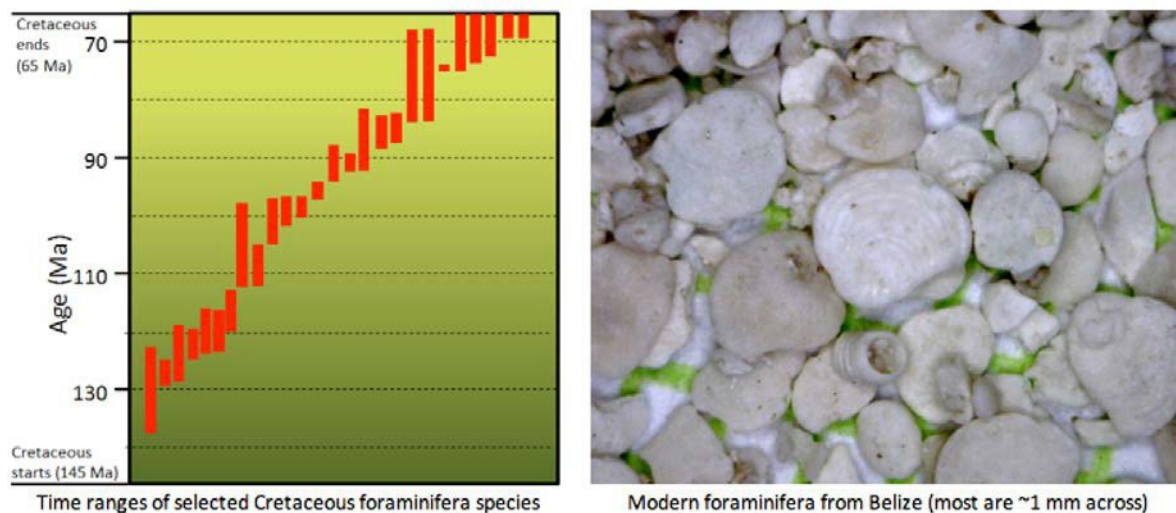

Figure 19.12 Time ranges for Cretaceous foraminifera (left) and modern foraminifera from the Ambergris area of Belize (right) [left: SE, from data in Scott, R, 2014, A Cretaceous chronostratigraphic database: construction and applications, Carnets de Géologie, Vol. 14., right : SE

Isotopic Dating Methods

Originally fossils only provided us with relative ages because, although early paleontologists understood biological succession, they did not know the absolute ages of the different organisms. It was only in the early part of the 20th century, when isotopic dating methods were first applied, that it became possible to discover the absolute ages of the rocks containing fossils. In most cases, we cannot use isotopic techniques to directly date fossils or the sedimentary rocks they are found in, but we can constrain their ages by dating igneous rocks that cut across sedimentary rocks, or volcanic ash layers that lie within sedimentary layers.

Isotopic dating of rocks, or the minerals in them, is based on the fact that we know the decay rates of certain unstable **isotopes** of elements and that these rates have been constant over geological time. It is also based on the premise that when the atoms of an element decay within a mineral or a rock, they stay there and don't escape to the surrounding rock, water, or air. One of the isotope pairs widely used in geology is the decay of ^{40}K to ^{40}Ar (potassium-40 to argon-40). ^{40}K is a radioactive isotope of potassium that is present in very small amounts in all minerals that have potassium in them. It has a half-life of 1.3 billion years, meaning that over a period of 1.3 Ga one-half of the ^{40}K atoms in a mineral or rock will decay to ^{40}Ar, and over the next 1.3 Ga one-half of the remaining atoms will decay, and so on (Figure 19.13).
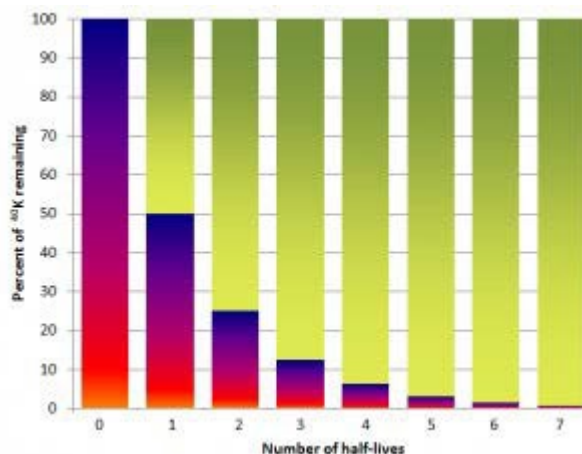

Figure 19.13 The decay of 40K over time. Each half-life is 1.3 billion years, so after 3.9 billion years (three half-lives) 12.5% of the original 40K will remain. The red-blue bars represent 40K and the green-yellow bars represent 40Ar. [SE]

In order to use the K-Ar dating technique, we need to have an igneous or metamorphic rock that includes a potassium-bearing mineral. One good example is granite, which normally has some potassium feldspar (Figure 19.14). Feldspar does not have any argon in it when it forms. Over time, the ^{40}K in the feldspar decays to ^{40}Ar. Argon is a gas and the atoms of ^{40}Ar remain embedded within the crystal, unless the rock is subjected to high temperatures after it forms. The sample must be analyzed using a very sensitive mass-spectrometer, which can detect the differences between the masses of atoms, and can therefore distinguish between ^{40}K and the much more abundant ^{39}K. Biotite and hornblende are also commonly used for K-Ar dating.
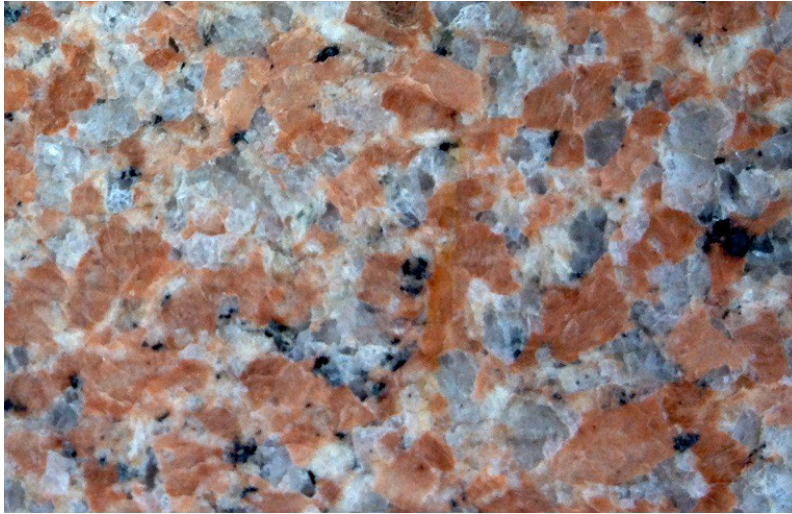
Figure 19.14 Crystals of potassium feldspar (pink) in a granitic rock are candidates for isotopic dating using the K-Ar method because they contained potassium and no argon when they formed. [SE]

Why can't we use isotopic dating techniques with sedimentary rocks?

An important assumption that we have to be able to make when using isotopic dating is that when the rock formed none of the daughter isotope was present (e.g., ^{40}Ar in the case of the K-Ar method). A clastic sedimentary rock is made up of older rock and mineral fragments, and when the rock forms it is almost certain that all of the fragments already have daughter isotopes in them. Furthermore, in almost all cases, the fragments have come from a range of source rocks that all formed at different times. If we dated a number of individual grains in the sedimentary rock, we would likely get a range of different dates, all older than the age of the rock. It might be possible to date some chemical sedimentary rocks isotopically, but there are no useful isotopes that can be used on old chemical sedimentary rocks. Radiocarbon dating can be used on sediments or sedimentary rocks that contain carbon, but it cannot be used on materials older than about 60 ka.

667

K-Ar is just one of many isotope-pairs that are useful for dating geological materials. Some of the other important pairs are listed in Table 19.2, along with the age ranges that they apply to and some comments on their applications. When radiometric techniques are applied to metamorphic rocks, the results normally tell us the date of metamorphism, not the date when the parent rock formed.

Isotope System	Half-Life	Useful Range	Comments
Potassium-argon	1.3 Ga	10 Ka – 4.57 Ga	Widely applicable because most rocks have some potassium
Uranium-lead	4.5 Ga	1 Ma – 4.57 Ga	The rock must have uranium-bearing minerals
Rubidium-strontium	47 Ga	10 Ma – 4.57 Ga	Less precision than other methods at old dates
Carbon-nitrogen (a.k.a. radiocarbon dating)	5,730 y	100 y to 60,000 y	Sample must contain wood, bone, or carbonate minerals; can be applied to young sediments

Table 19.2 A few of the isotope systems that are widely used for dating geological materials

Radiocarbon dating (using ^{14}C) can be applied to many geological materials, including sediments and sedimentary rocks, but the materials in question must be younger than 60 ka. Fragments of wood incorporated into young sediments are good candidates for carbon dating, and this technique has been used widely in studies involving late Pleistocene glaciers and glacial sediments. An example is shown in Figure 19.15; radiocarbon dates from wood

fragments in glacial sediments have been used to estimate the time of the last glacial advance along the Strait of Georgia.
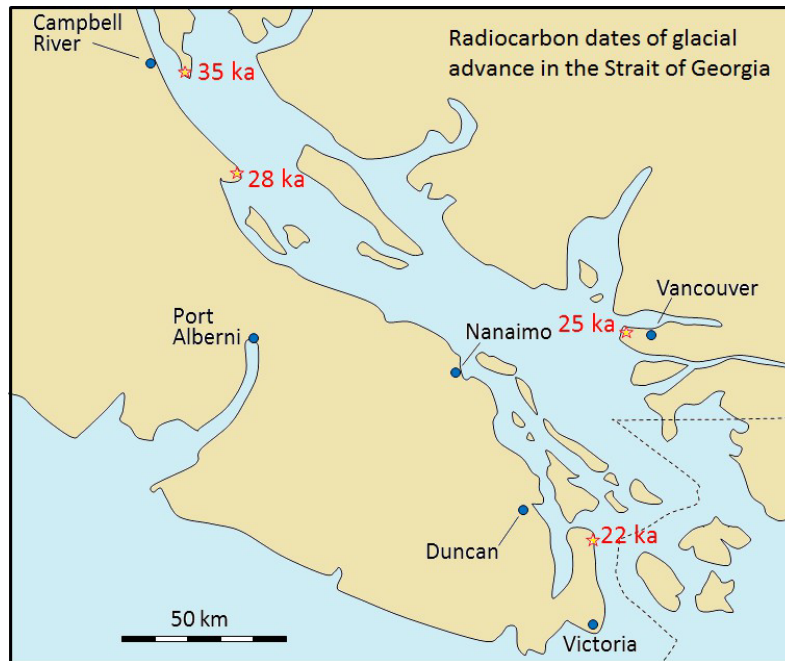

Figure 19.15 Radiocarbon dates on wood fragments in glacial sediments in the Strait of Georgia [SE after Clague, J, 1976, Quadra Sand and its relation to late Wisconsin glaciation of southeast British Columbia, Can. J. Earth Sciences

Understanding Geologic time

It's one thing to know the facts about geological time — how long it is, how we measure it, how we divide it up, and what we call the various periods and epochs — but it is quite another to really understand geological time. The problem is that our lives are short and our memories are even shorter. Our experiences span only a few decades, so we really don't have a way of knowing what 11,700 years means. What's more, it's hard for us to understand how 11,700 years differs from 65.5 Ma, or even from 1.8 Ga. It's not that we can't comprehend what the numbers mean — we can all get that figured out with a bit of practice — but even if we do know the numerical meaning of 65.5 Ma, we can't really appreciate how long ago it was.

You may be wondering why it's so important to really "understand" geological time. There are some very good reasons. One is so that we can fully understand how geological processes that seem impossibly slow can produce anything of consequence. For example, we are familiar with the concept of driving from one major city to another: a journey of several hours at around 100 km/h. Continents move toward each other at rates of a fraction of a millimetre per day, or something in the order of 0.00000001 km/h, and yet, at this impossibly slow rate (try walking at that speed!), they can move thousands of kilometres. Sediments typically accumulate at even slower rates — less than a millimetre per year — but still they are thick enough to be thrust up into monumental mountains and carved into breathtaking canyons.

Another reason is that for our survival on this planet, we need to understand issues like extinction of endangered species and **anthropogenic** (human-caused) climate change. Some people, who don't understand geological time, are quick to say that the climate has changed in the past, and that what is happening now is no different. And it certainly has changed in the past. For example, from the Eocene (50 Ma) to the present day, Earth's climate cooled by about 12°C. That's a huge change that ranks up there with many of the important climate changes of the distant past, and yet the rate of change over that time was only 0.000024°C/century. Anthropogenic climate change has been

1.1°C over the past 100 years,[1] and that is 45,800 times faster than the rate of natural climate change since the Eocene!

One way to wrap your mind around geological time is to put it into the perspective of single year, because we all know how long it is from one birthday to the next. At that rate, each hour of the year is equivalent to approximately 500,000 years, and each day is equivalent to 12.5 million years.

If all of geological time is compressed down to a single year, Earth formed on January 1, and the first life forms evolved in late March (~3,500 Ma). The first large life forms appeared on November 13 (~600 Ma), plants appeared on land around November 24, and amphibians on December 3. Reptiles evolved from amphibians during the first week of December and dinosaurs and early mammals evolved from reptiles by December 13, but the dinosaurs, which survived for 160 million years, were gone by Boxing Day (December 26). The Pleistocene Glaciation got started at around 6:30 p.m. on New Year's Eve, and the last glacial ice left southern Canada by 11:59 p.m.

It's worth repeating: on this time scale, the earliest ancestors of the animals and plants with which we are familiar did not appear on Earth until mid-November, the dinosaurs disappeared after Christmas, and most of Canada was periodically locked in ice from 6:30 to 11:59 p.m. on New Year's Eve. As for people, the first to inhabit B.C. got here about one minute before midnight, and the first Europeans arrived about two seconds before midnight.

It is common for the popular press to refer to distant past events as being "prehistoric." For example, dinosaurs are reported as being "prehistoric creatures," even by the esteemed National Geographic Society.[2] The written records of our history date back to about 6,000 years ago, so anything prior to that is considered "prehistoric." But to call the dinosaurs prehistoric is equivalent to — and about as useful as — saying that Singapore is beyond the

---

1. Climate change data from NASA Goddard Institute for Space Studies: http://data.giss.nasa.gov/gistemp/tabledata_v3/GLB.Ts.txt
2. http://science.nationalgeographic.com/science/prehistoric-world/

city limits of Kamloops! If we are going to become literate about geological time, we have to do better than calling dinosaurs, or early horses (54 Ma), or even early humans (2.8 Ma), "prehistoric."

Summary

The Geological Time Scale	The work of William Smith was critical to the establishment of the first geological time scale early in the 19th century, but it wasn't until the 20th century that geologists were able to assign reliable dates to the various time periods. The geological time scale is now maintained by the International Commission on Stratigraphy. Geological time is divided into eons, eras, periods, and epochs.
Relative Dating Methods	We can determine the relative ages of different rocks by observing and interpreting relationships among them, such as superposition, cross-cutting, and inclusions. Gaps in the geological record are represented by various types of unconformities.
Dating Rocks Using Fossils	Fossils are useful for dating rocks date back to about 600 Ma. If we know the age range of a fossil, we can date the rock, but some organisms lived for many millions of years. Index fossils represent shorter geological times, and if a rock has several different fossils with known age ranges, we can normally narrow the time during which the rock formed.
Isotopic Dating Methods	Radioactive isotopes decay at predictable and known rates, and can be used to date igneous and metamorphic rocks. Some of the more useful isotope systems are potassium-argon, rubidium-strontium, uranium-lead, and carbon-nitrogen. Radiocarbon dating can be applied to sediments and sedimentary rocks, but only if they are younger than 60 ka.
Understanding Geological Time	While knowing about geological time is relatively easy, actually comprehending the significance of the vast amounts of geological time is a great challenge. To be able to solve important geological problems and critical societal challenges, like climate change, we need to really understand geological time.

Review Question
    What are the features of a useful index fossil?