17 | BIOTECHNOLOGY AND GENOMICS

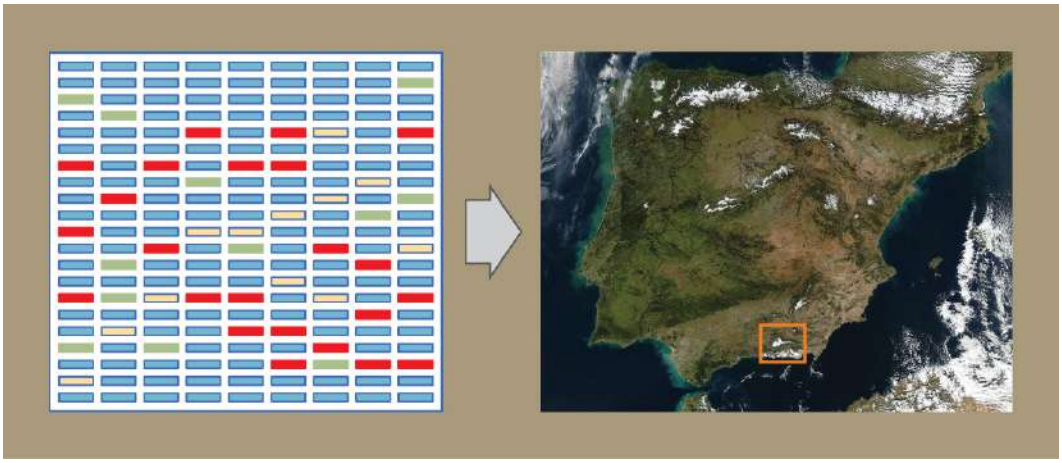


Figure 17.1 In genomics, the DNA of different organisms is compared, enabling scientists to create maps with which to navigate the DNA of different organisms. (credit "map": modification of photo by NASA)

Chapter Outline

Introduction

Some of the greatest accomplishments of biotechnology are in the fields of medicine and medical research. For example, intestinal failure due to missing or abnormal intestinal tissue is a frequent problem in premature babies. Intestinal problems are also common for people who have had parts of their small intestines removed for reasons , such as Crohn's Disease, cancer, and blockages. Complications from intestinal failure may include liver disease, bacterial overgrowth, dehydration, and malnutrition.

Scientists have recently developed a way to engineer human intestines from human cells using mice. Using a mixture of healthy mouse and human intestinal cells and placing it on scaffolding in the abdominal cavity of immunocompromised mice, functional human intestinal cells grow within four weeks. This could be the breakthrough needed to help patients suffering from intestinal failure. More details about this exciting research can be found **here (http://openstaxcollege.org/l/32grwinmouse)** .

17.1 | Biotechnology

In this section, you will explore the following questions:

- What are examples of basic techniques used to manipulate genetic material (DNA and RNA)?
- What is the difference between molecular and reproductive cloning?
- What are examples of uses of biotechnology in medicine and agriculture?

Connection for AP® Courses

Did you eat cereal for breakfast or tomatoes in your dinner salad? Do you know someone who has received gene therapy to treat a disease such as cancer? Should your school, health insurance provider, or employer have access to your genetic profile? Understanding how DNA works has allowed scientists to recombine DNA molecules, clone organisms, and produce mice that glow in the dark. We likely have eaten genetically modified foods and are familiar with how DNA analysis is used to solve crimes. Manipulation of DNA by humans has resulted in bacteria that can protect plants from insect pests and restore ecosystems. Biotechnologies also have been used to produce insulin, hormones, antibiotics, and medicine that dissolve blood clots. Comparative genomics yields new insights into relationships among species, and DNA sequences reveal our personal genetic make-up. However, manipulation of DNA comes with social and ethical responsibilities, raising questions about its appropriate uses.

Nucleic acids can be isolated from cells for analysis by lysing cell membranes and enzymatically destroying all other macromolecules. Fragmented or whole chromosomes can be separated on the basis of size (base pair length) by gel electrophoresis. Short sequences of DNA or RNA can be amplified using the polymerase chain reaction (PCR). Recombinant DNA technology can combine DNA from different sources using bacterial plasmids or viruses as vectors to carry foreign genes into host cells, resulting in genetically modified organisms (GMOs). Transgenic bacteria, agricultural plants such as corn and rice, and farm animals produce protein products such as hormones and vaccines that benefit humans. (It is important to remind ourselves that recombinant technology is possible because the genetic code is universal, and the processes of transcription and translation are fundamentally the same in all organisms.) Cloning produces genetically identical copies of DNA, cells, or even entire organisms (reproductive cloning). Genetic testing identifies disease-causing genes, and gene therapy can be used to treat or cure an inheritable disease. However, questions emerge from these technologies including the safety of GMOs and privacy issues.

Information presented and the examples highlighted in the section support concepts outlined in Big Idea 3 of the AP® Biology Curriculum Framework. The Learning Objectives listed in the Curriculum Framework provide a transparent foundation for the AP® Biology course, an inquiry-based laboratory experience, instructional activities, and AP® exam questions. A Learning Objective merges required content with one or more of the seven Science Practices.

Big Idea 3	Living systems store, retrieve, transmit and respond to information essential to life processes.
Enduring Understanding 3.A	Heritable information provides for continuity of life.
Essential Knowledge	**3.A.1** DNA, and in some cases RNA, is the primary source of heritable information.
Science Practice	**6.4** The student can make claims and predictions about natural phenomena based on scientific theories and models.
Learning Objective	**3.5** The student can justify the claim that humans can manipulate heritable information by identifying an example of a commonly used technology.
Big Idea 3	Living systems store, retrieve, transmit and respond to information essential to life processes.

Enduring Understanding 3.C	The processing of genetic information is imperfect and is a source of genetic variation.
Essential Knowledge	**3.C.1** Changes in genotype can result in changes in phenotype.
Science Practice	**7.2** The student can connect concepts in and across domain(s) to generalize or extrapolate in and/or across enduring understandings and/or big ideas.
Learning Objective	**3.24** The student is able to predict how a change in genotype, when expressed as a phenotype, provides a variation that can be subject to natural selection.

The Science Practice Challenge Questions contain additional test questions for this section that will help you prepare for the AP exam. These questions address the following standards:
[APLO 3.13][APLO 3.23][APLO 3.28][APLO 3.24][APLO 1.11][APLO 3.5][APLO 4.2][APLO 4.8]

Biotechnology is the use of biological agents for technological advancement. Biotechnology was used for breeding livestock and crops long before the scientific basis of these techniques was understood. Since the discovery of the structure of DNA in 1953, the field of biotechnology has grown rapidly through both academic research and private companies. The primary applications of this technology are in medicine (production of vaccines and antibiotics) and agriculture (genetic modification of crops, such as to increase yields). Biotechnology also has many industrial applications, such as fermentation, the treatment of oil spills, and the production of biofuels.

Basic Techniques to Manipulate Genetic Material (DNA and RNA)

To understand the basic techniques used to work with nucleic acids, remember that nucleic acids are macromolecules made of nucleotides (a sugar, a phosphate, and a nitrogenous base) linked by phosphodiester bonds. The phosphate groups on these molecules each have a net negative charge. An entire set of DNA molecules in the nucleus is called the genome. DNA has two complementary strands linked by hydrogen bonds between the paired bases. The two strands can be separated by exposure to high temperatures (DNA denaturation) and can be reannealed by cooling. The DNA can be replicated by the DNA polymerase enzyme. Unlike DNA, which is located in the nucleus of eukaryotic cells, RNA molecules leave the nucleus. The most common type of RNA that is analyzed is the messenger RNA (mRNA) because it represents the protein-coding genes that are actively expressed. However, RNA molecules present some other challenges to analysis, as they are often less stable than DNA.

DNA and RNA Extraction

To study or manipulate nucleic acids, the DNA or RNA must first be isolated or extracted from the cells. Various techniques are used to extract different types of DNA (**Figure 17.2**). Most nucleic acid extraction techniques involve steps to break open the cell and use enzymatic reactions to destroy all macromolecules that are not desired (such as degradation of unwanted molecules and separation from the DNA sample). Cells are broken using a **lysis buffer** (a solution which is mostly a detergent); lysis means “to split.” These enzymes break apart lipid molecules in the cell membranes and nuclear membranes. Macromolecules are inactivated using enzymes such as **proteases** that break down proteins, and **ribonucleases** (RNAses) that break down RNA. The DNA is then precipitated using alcohol. Human genomic DNA is usually visible as a gelatinous, white mass. The DNA samples can be stored frozen at –80°C for several years.

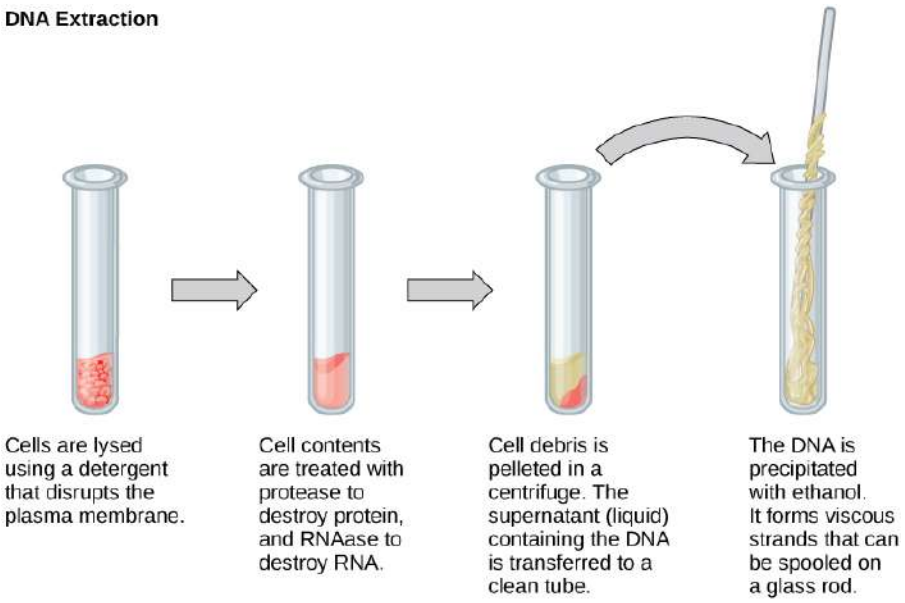


Figure 17.2 This diagram shows the basic method used for extraction of DNA.

RNA analysis is performed to study gene expression patterns in cells. RNA is naturally very unstable because RNAses are commonly present in nature and very difficult to inactivate. Similar to DNA, RNA extraction involves the use of various buffers and enzymes to inactivate macromolecules and preserve the RNA.

Gel Electrophoresis

Because nucleic acids are negatively charged ions at neutral or basic pH in an aqueous environment, they can be mobilized by an electric field. **Gel electrophoresis** is a technique used to separate molecules on the basis of size, using this charge. The nucleic acids can be separated as whole chromosomes or fragments. The nucleic acids are loaded into a slot near the negative electrode of a semisolid, porous gel matrix and pulled toward the positive electrode at the opposite end of the gel. Smaller molecules move through the pores in the gel faster than larger molecules; this difference in the rate of migration separates the fragments on the basis of size. There are molecular weight standard samples that can be run alongside the molecules to provide a size comparison. Nucleic acids in a gel matrix can be observed using various fluorescent or colored dyes. Distinct nucleic acid fragments appear as bands at specific distances from the top of the gel (the negative electrode end) on the basis of their size (**Figure 17.3**). A mixture of genomic DNA fragments of varying sizes appear as a long smear, whereas uncut genomic DNA is usually too large to run through the gel and forms a single large band at the top of the gel.

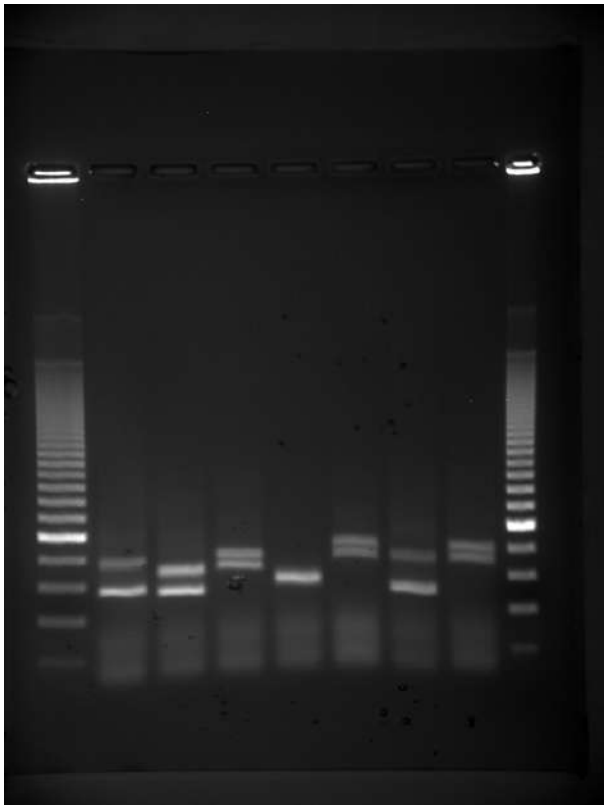


Figure 17.3 Shown are DNA fragments from seven samples run on a gel, stained with a fluorescent dye, and viewed under UV light. (credit: James Jacob, Tompkins Cortland Community College)

Amplification of Nucleic Acid Fragments by Polymerase Chain Reaction

Although genomic DNA is visible to the naked eye when it is extracted in bulk, DNA analysis often requires focusing on one or more specific regions of the genome. **Polymerase chain reaction** (**PCR**) is a technique used to amplify specific regions of DNA for further analysis (**Figure 17.4**). PCR is used for many purposes in laboratories, such as the cloning of gene fragments to analyze genetic diseases, identification of contaminant foreign DNA in a sample, and the amplification of DNA for sequencing. More practical applications include the detection of genetic diseases.

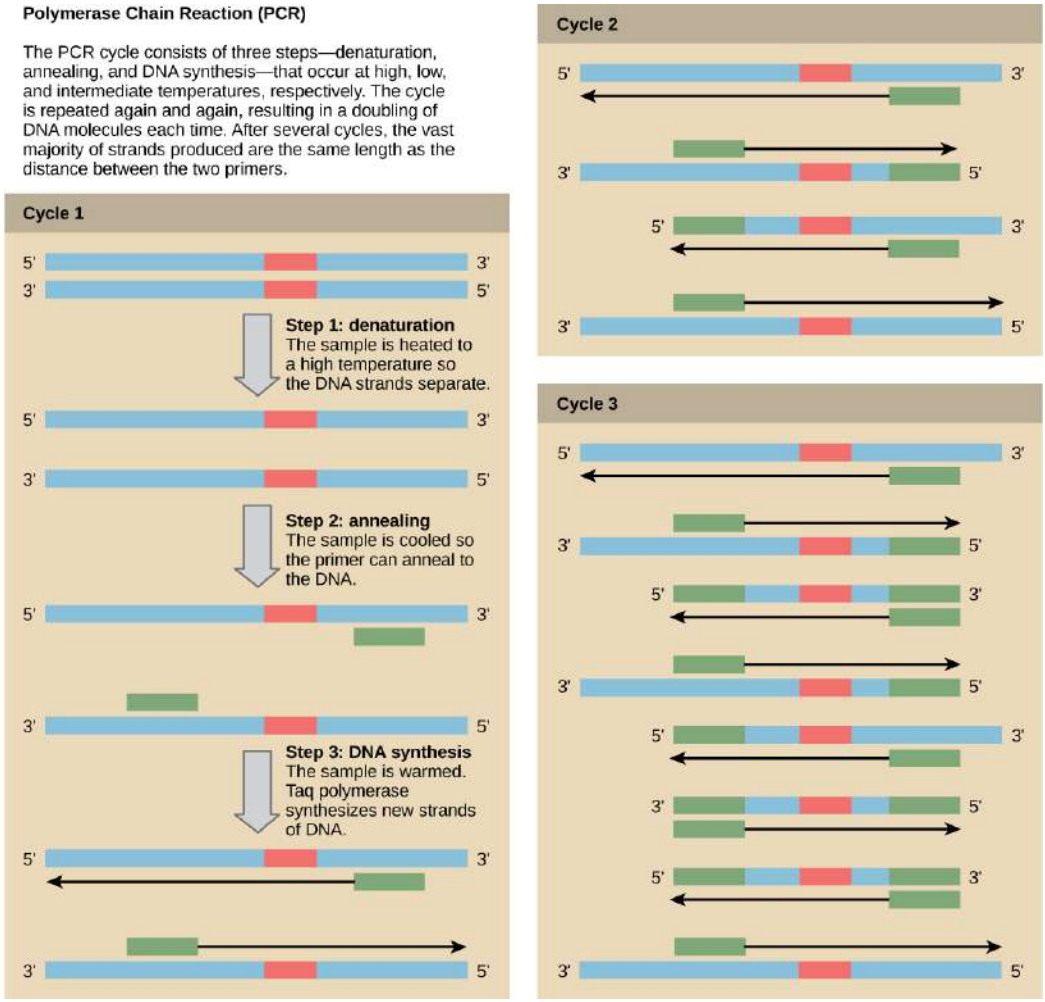


Figure 17.4 Polymerase chain reaction, or PCR, is used to amplify a specific sequence of DNA. Primers—short pieces of DNA complementary to each end of the target sequence—are combined with genomic DNA, Taq polymerase, and deoxynucleotides. Taq polymerase is a DNA polymerase isolated from the thermostable bacterium *Thermus aquaticus* that is able to withstand the high temperatures used in PCR. *Thermus aquaticus* grows in the Lower Geyser Basin of Yellowstone National Park. Reverse transcriptase PCR (RT-PCR) is similar to PCR, but cDNA is made from an RNA template before PCR begins.

DNA fragments can also be amplified from an RNA template in a process called **reverse transcriptase PCR (RT-PCR)**. The first step is to recreate the original DNA template strand (called cDNA) by applying DNA nucleotides to the mRNA. This process is called reverse transcription. This requires the presence of an enzyme called reverse transcriptase. After the cDNA is made, regular PCR can be used to amplify it.

Deepen your understanding of the polymerase chain reaction by clicking through **this interactive exercise (http://openstaxcollege.org/l/PCR)** .

Explain an advantage the polymerase chain reaction (PCR) has for DNA research.

a. The process of PCR can isolate a particular piece of DNA for copying, which allows scientists to copy millions of strands of DNA in a short amount of time.

b. The process of PCR can purify a particular piece of DNA, and very small amounts of DNA can be used for purification.

c. The process of PCR separates and analyzes DNA and its fragments, which requires very little DNA.

d. The process of PCR anneals DNA molecules to complementary DNA strands, which maintains the same amount of DNA.

Hybridization, Southern Blotting, and Northern Blotting

Nucleic acid samples, such as fragmented genomic DNA and RNA extracts, can be probed for the presence of certain sequences. Short DNA fragments called **probes** are designed and labeled with radioactive or fluorescent dyes to aid detection. Gel electrophoresis separates the nucleic acid fragments according to their size. The fragments in the gel are then transferred onto a nylon membrane in a procedure called **blotting** (**Figure 17.5**). The nucleic acid fragments that are bound to the surface of the membrane can then be probed with specific radioactively or fluorescently labeled probe sequences. When DNA is transferred to a nylon membrane, the technique is called **Southern blotting**, and when RNA is transferred to a nylon membrane, it is called **northern blotting**. Southern blots are used to detect the presence of certain DNA sequences in a given genome, and northern blots are used to detect gene expression.

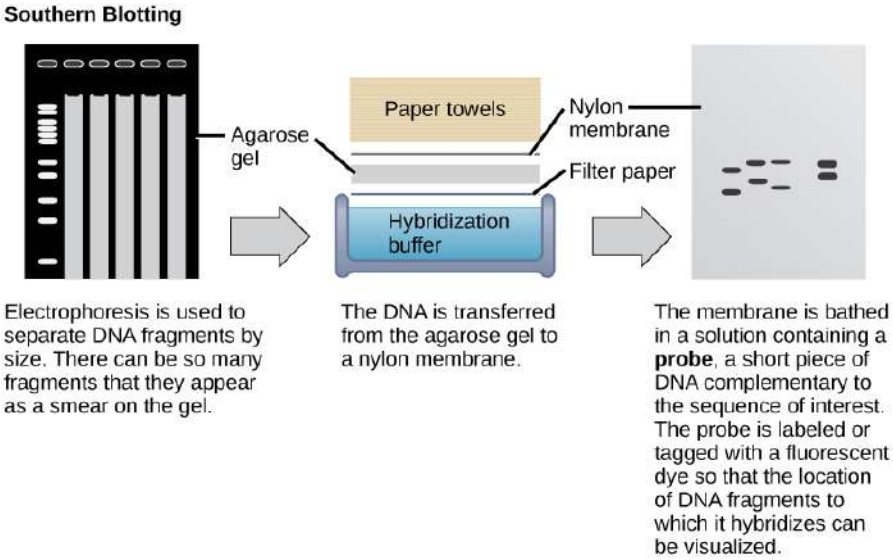


Figure 17.5 Southern blotting is used to find a particular sequence in a sample of DNA. DNA fragments are separated on a gel, transferred to a nylon membrane, and incubated with a DNA probe complementary to the sequence of interest. Northern blotting is similar to Southern blotting, but RNA is run on the gel instead of DNA. In western blotting, proteins are run on a gel and detected using antibodies.

Molecular Cloning

In general, the word “cloning” means the creation of a perfect replica; however, in biology, the re-creation of a whole organism is referred to as “reproductive cloning.” Long before attempts were made to clone an entire organism, researchers learned how to reproduce desired regions or fragments of the genome, a process that is referred to as molecular cloning.

Cloning small fragments of the genome allows for the manipulation and study of specific genes (and their protein products), or noncoding regions in isolation. A plasmid (also called a vector) is a small circular DNA molecule that replicates independently of the chromosomal DNA. In cloning, the plasmid molecules can be used to provide a "folder" in which to insert a desired DNA fragment. Plasmids are usually introduced into a bacterial host for proliferation. In the bacterial context, the fragment of DNA from the human genome (or the genome of another organism that is being studied) is referred to as **foreign DNA**, or a transgene, to differentiate it from the DNA of the bacterium, which is called the **host DNA**.

Plasmids occur naturally in bacterial populations (such as *Escherichia coli*) and have genes that can contribute favorable traits to the organism, such as **antibiotic resistance** (the ability to be unaffected by antibiotics). Plasmids have been repurposed and engineered as vectors for molecular cloning and the large-scale production of important reagents, such as insulin and human growth hormone. An important feature of plasmid vectors is the ease with which a foreign DNA fragment can be introduced via the **multiple cloning site (MCS)**. The MCS is a short DNA sequence containing multiple sites that can be cut with different commonly available restriction endonucleases. **Restriction endonucleases** recognize specific DNA sequences and cut them in a predictable manner; they are naturally produced by bacteria as a defense mechanism against foreign DNA. Many restriction endonucleases make staggered cuts in the two strands of DNA, such that the cut ends have a 2- or 4-base single-stranded overhang. Because these overhangs are capable of annealing with complementary overhangs, these are called “sticky ends.” Addition of an enzyme called DNA ligase permanently joins the DNA fragments via phosphodiester bonds. In this way, any DNA fragment generated by restriction endonuclease cleavage can be spliced between the two ends of a plasmid DNA that has been cut with the same restriction endonuclease (**Figure 17.6**).

Recombinant DNA Molecules

Plasmids with foreign DNA inserted into them are called **recombinant DNA** molecules because they are created artificially and do not occur in nature. They are also called chimeric molecules because the origin of different parts of the molecules can be traced back to different species of biological organisms or even to chemical synthesis. Proteins that are expressed from recombinant DNA molecules are called **recombinant proteins**. Not all recombinant plasmids are capable of expressing genes. The recombinant DNA may need to be moved into a different vector (or host) that is better designed for gene expression. Plasmids may also be engineered to express proteins only when stimulated by certain environmental factors, so that scientists can control the expression of the recombinant proteins.

visual CONNECTION

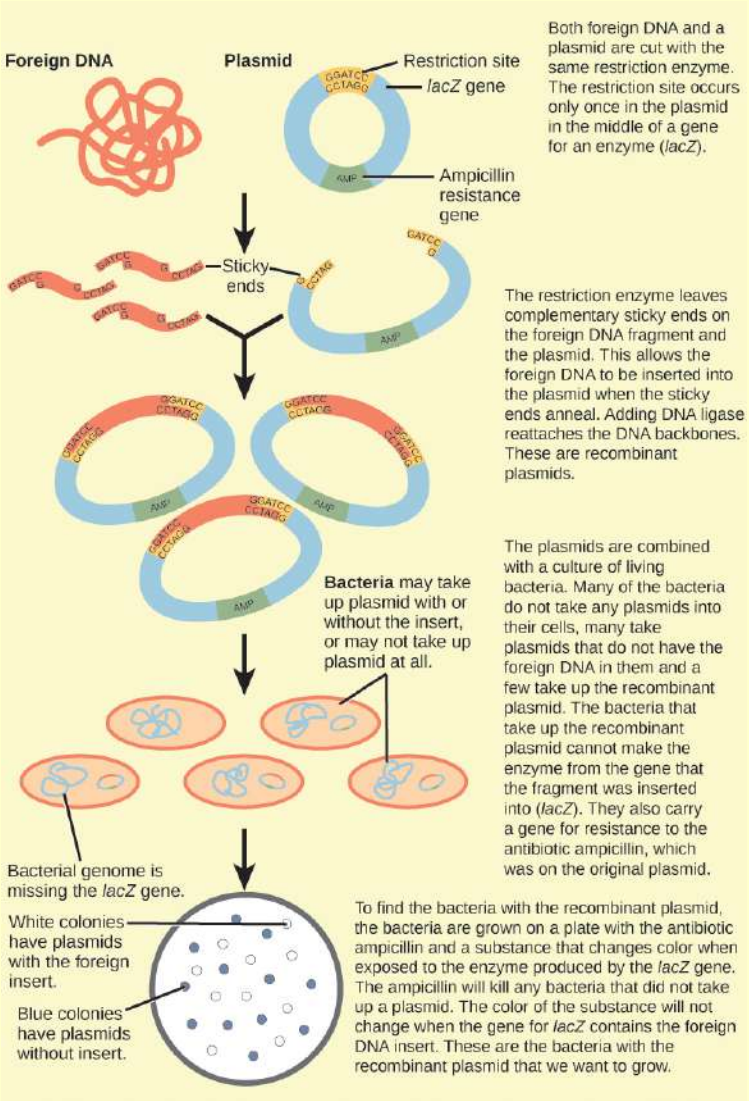
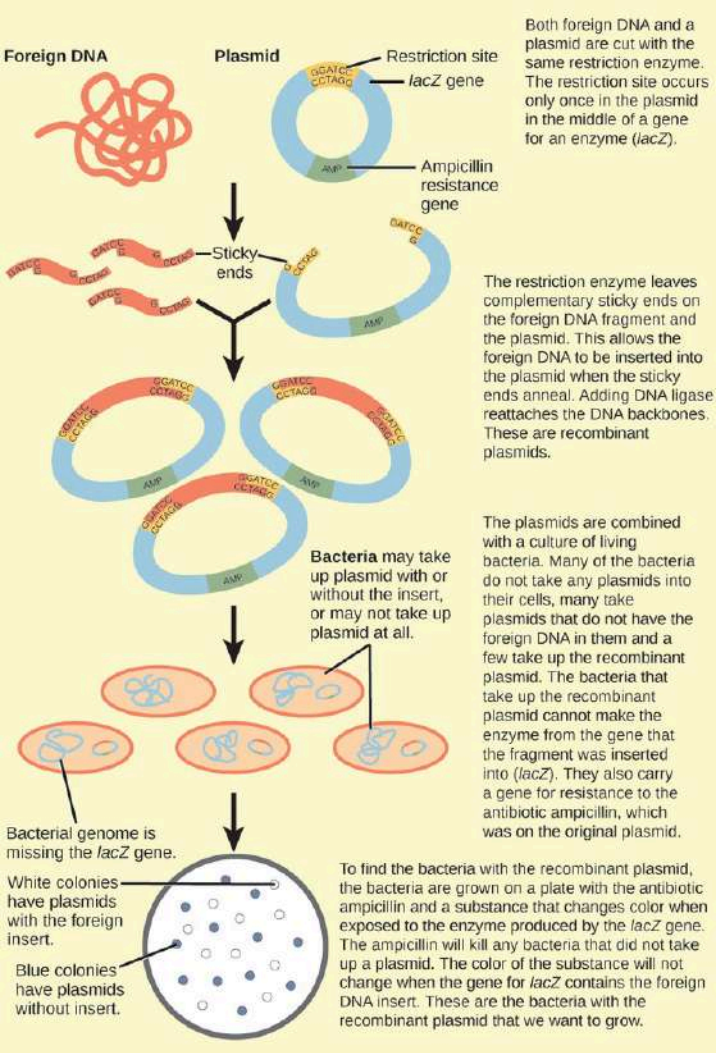


Figure 17.6 This diagram shows the steps involved in molecular cloning.



You are working in a molecular biology lab and, unbeknownst to you, your lab partner left the foreign genomic DNA you are planning to clone on the lab bench overnight instead of storing it in the freezer. As a result, it was degraded by nucleases, but still used in the experiment. The plasmid, on the other hand, is fine. What results would you expect from your molecular cloning experiment?

a. There will be no colonies on the bacterial plate.

b. There will be blue colonies only.

c. There will be blue and white colonies.

d. The will be white colonies only.

View an **animation of recombination in cloning (http://openstaxcollege.org/l/recombination)** from the DNA Learning Center.

What are the functions of the restriction enzymes and DNA ligase in recombination?

a. Restriction enzymes restrict the recombination process. DNA ligase ligates the products of recombination with each other.

b. DNA ligase splices DNA at a specific point, so the new piece of DNA can be inserted. Restriction enzymes stitch together the new gene to the existing piece of DNA.

c. Restriction enzymes splice DNA at a specific point, so the new piece of DNA can be inserted. DNA ligase stitches together the new foreign gene to the existing piece of DNA.

d. Restriction enzymes splice the existing piece of DNA only to accommodate the new piece of DNA. DNA ligase ligates the new DNA with the existing DNA.

science practices CONNECTION for AP® Courses

Activity

Cloning can be used to quickly replicate crop plants that have advantageous genes, such as greater disease resistance or greater fruit production. However, cloning also produces crop plants that have little genetic variation. In a group, discuss the advantages and disadvantages of using clones as human food sources in an era where the Earth is undergoing a period of climate change. How well will cloned populations of crop plants be able to adapt to climate change, compared to non-clone crop plants? Then, defend your group's position against those of other groups in a classroom debate.

Think About It

How would a scientist introduce a gene for herbicide resistance into a plant, such as corn?

Cellular Cloning

Unicellular organisms, such as bacteria and yeast, naturally produce clones of themselves when they replicate asexually by binary fission; this is known as **cellular cloning**. The nuclear DNA duplicates by the process of mitosis, which creates an exact replica of the genetic material.

Reproductive Cloning

Reproductive cloning is a method used to make a clone or an identical copy of an entire multicellular organism. Most multicellular organisms undergo reproduction by sexual means, which involves genetic hybridization of two individuals (parents), making it impossible for generation of an identical copy or a clone of either parent. Recent advances in biotechnology have made it possible to artificially induce asexual reproduction of mammals in the laboratory.

Parthenogenesis, or "virgin birth," occurs when an embryo grows and develops without the fertilization of the egg occurring; this is a form of asexual reproduction. An example of parthenogenesis occurs in species in which the female lays an egg and if the egg is fertilized, it is a diploid egg and the individual develops into a female; if the egg is not fertilized, it

remains a haploid egg and develops into a male. The unfertilized egg is called a parthenogenic, or virgin, egg. Some insects and reptiles lay parthenogenic eggs that can develop into adults.

Sexual reproduction requires two cells; when the haploid egg and sperm cells fuse, a diploid zygote results. The zygote nucleus contains the genetic information to produce a new individual. However, early embryonic development requires the cytoplasmic material contained in the egg cell. This idea forms the basis for reproductive cloning. Therefore, if the haploid nucleus of an egg cell is replaced with a diploid nucleus from the cell of any individual of the same species (called a donor), it will become a zygote that is genetically identical to the donor. Somatic cell nuclear transfer is the technique of transferring a diploid nucleus into an enucleated egg. It can be used for either therapeutic cloning or reproductive cloning.

The first cloned animal was Dolly, a sheep who was born in 1996. The success rate of reproductive cloning at the time was very low. Dolly lived for seven years and died of respiratory complications (**Figure 17.7**). There is speculation that because the cell DNA belongs to an older individual, the age of the DNA may affect the life expectancy of a cloned individual. Since Dolly, several animals such as horses, bulls, and goats have been successfully cloned, although these individuals often exhibit facial, limb, and cardiac abnormalities. There have been attempts at producing cloned human embryos as sources of embryonic stem cells, sometimes referred to as cloning for therapeutic purposes. Therapeutic cloning produces stem cells to attempt to remedy detrimental diseases or defects (unlike reproductive cloning, which aims to reproduce an organism). Still, therapeutic cloning efforts have met with resistance because of bioethical considerations.

visual CONNECTION

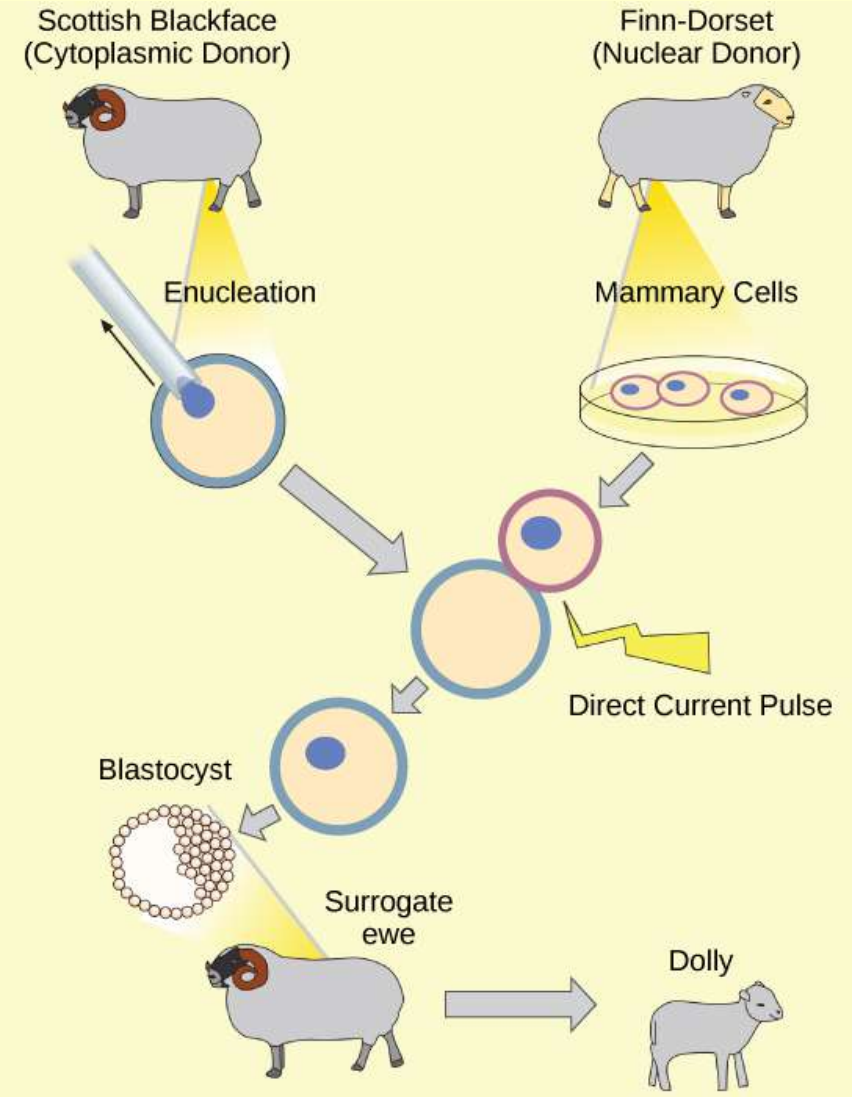
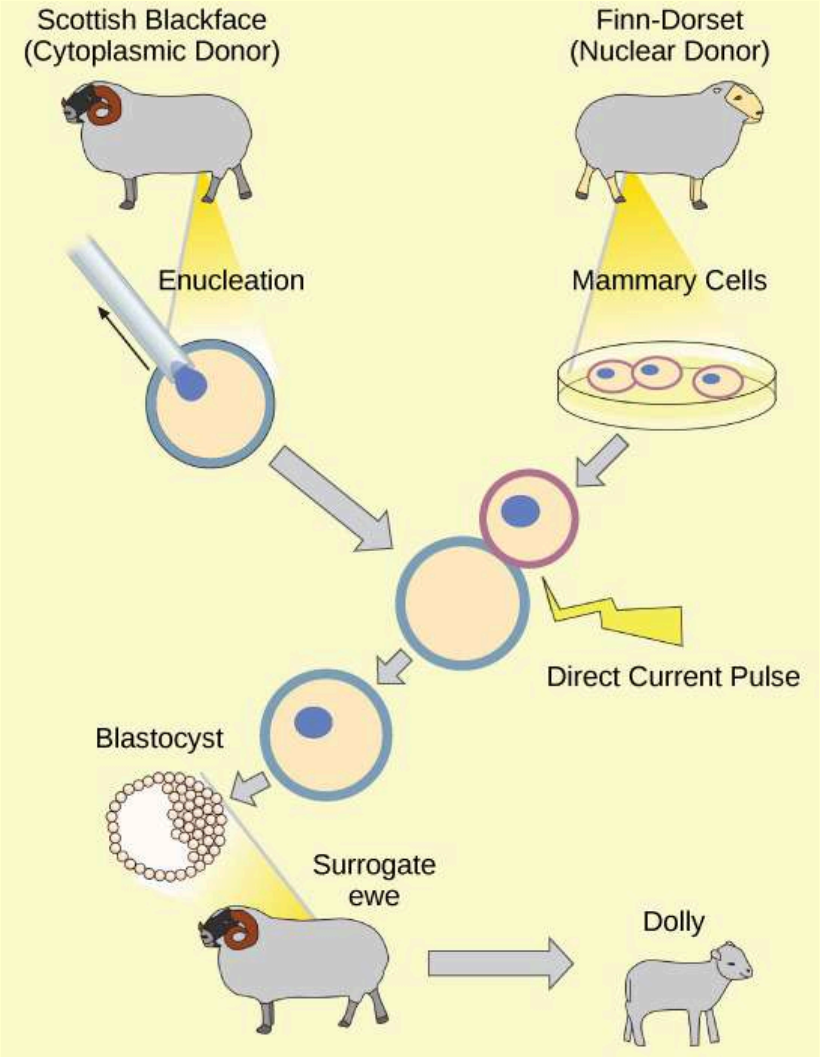


Figure 17.7 Dolly the sheep was the first mammal to be cloned. To create Dolly, the nucleus was removed from a donor egg cell. The nucleus from a second sheep was then introduced into the cell, which was allowed to divide to the blastocyst stage before being implanted in a surrogate mother. (credit: modification of work by "Squidonius"/Wikimedia Commons)



Do you think Dolly was a Finn-Dorset or a Scottish Blackface sheep?

a. Blackface because this follows cytoplasmic inheritance.
b. Dolly was a Finn-Dorset sheep; the nucleus in the process was taken from a Finn-Dorset mother.
c. Dorset sheep because this follows cytoplasmic inheritance.
d. Dolly the sheep was a Scottish blackface due to epigenetic inheritance.

Genetic Engineering

Genetic engineering is the alteration of an organism's genotype using recombinant DNA technology to modify an organism's DNA to achieve desirable traits. The addition of foreign DNA in the form of recombinant DNA vectors generated by molecular cloning is the most common method of genetic engineering. The organism that receives the recombinant DNA is called a **genetically modified organism** (GMO). If the foreign DNA that is introduced comes from a different species, the host organism is called **transgenic**. Bacteria, plants, and animals have been genetically modified since the early 1970s for academic, medical, agricultural, and industrial purposes. In the US, GMOs such as herbicide-resistant soybeans and borer-resistant corn are part of many common processed foods.

Gene Targeting

Although classical methods of studying the function of genes began with a given phenotype and determined the genetic basis of that phenotype, modern techniques allow researchers to start at the DNA sequence level and ask: "What does this gene or DNA element do?" This technique, called reverse genetics, has resulted in reversing the classic genetic

methodology. This method would be similar to damaging a body part to determine its function. An insect that loses a wing cannot fly, which means that the function of the wing is flight. The classical genetic method would compare insects that cannot fly with insects that can fly, and observe that the non-flying insects have lost wings. Similarly, mutating or deleting genes provides researchers with clues about gene function. The methods used to disable gene function are collectively called gene targeting. **Gene targeting** is the use of recombinant DNA vectors to alter the expression of a particular gene, either by introducing mutations in a gene, or by eliminating the expression of a certain gene by deleting a part or all of the gene sequence from the genome of an organism.

Biotechnology in Medicine and Agriculture

It is easy to see how biotechnology can be used for medicinal purposes. Knowledge of the genetic makeup of our species, the genetic basis of heritable diseases, and the invention of technology to manipulate and fix mutant genes provides methods to treat the disease. Biotechnology in agriculture can enhance resistance to disease, pest, and environmental stress, and improve both crop yield and quality.

Genetic Diagnosis and Gene Therapy

The process of testing for suspected genetic defects before administering treatment is called **genetic diagnosis** by **genetic testing**. Depending on the inheritance patterns of a disease-causing gene, family members are advised to undergo genetic testing. For example, women diagnosed with breast cancer are usually advised to have a biopsy so that the medical team can determine the genetic basis of cancer development. Treatment plans are based on the findings of genetic tests that determine the type of cancer. If the cancer is caused by inherited gene mutations, other female relatives are also advised to undergo genetic testing and periodic screening for breast cancer. Genetic testing is also offered for fetuses (or embryos with in vitro fertilization) to determine the presence or absence of disease-causing genes in families with specific debilitating diseases.

Gene therapy is a genetic engineering technique used to cure disease. In its simplest form, it involves the introduction of a good gene at a random location in the genome to aid the cure of a disease that is caused by a mutated gene. The good gene is usually introduced into diseased cells as part of a vector transmitted by a virus that can infect the host cell and deliver the foreign DNA (Figure 17.8). More advanced forms of gene therapy try to correct the mutation at the original site in the genome, such as is the case with treatment of severe combined immunodeficiency (SCID).

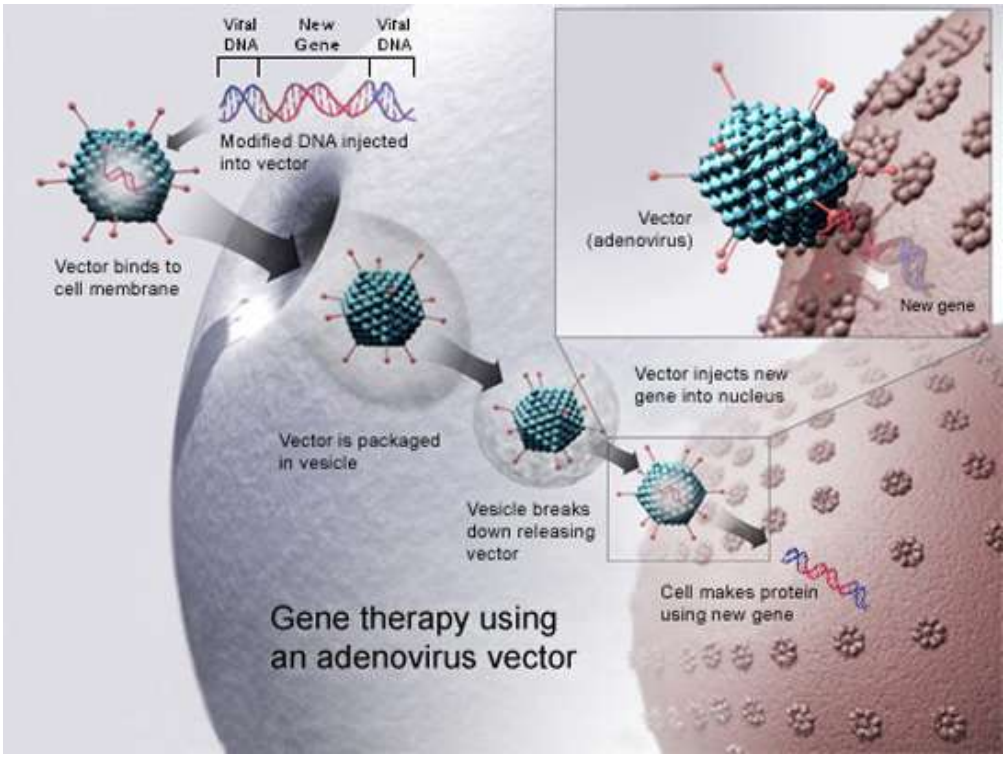


Figure 17.8 Gene therapy using an adenovirus vector can be used to cure certain genetic diseases in which a person has a defective gene. (credit: NIH)

Production of Vaccines, Antibiotics, and Hormones

Traditional vaccination strategies use weakened or inactive forms of microorganisms to mount the initial immune response.

Modern techniques use the genes of microorganisms cloned into vectors to mass produce the desired antigen. The antigen is then introduced into the body to stimulate the primary immune response and trigger immune memory. Genes cloned from the influenza virus have been used to combat the constantly changing strains of this virus.

Antibiotics are a biotechnological product. They are naturally produced by microorganisms, such as fungi, to attain an advantage over bacterial populations. Antibiotics are produced on a large scale by cultivating and manipulating fungal cells.

Recombinant DNA technology was used to produce large-scale quantities of human insulin in *E. coli* as early as 1978. Previously, it was only possible to treat diabetes with pig insulin, which caused allergic reactions in humans because of differences in the gene product. In addition, human growth hormone (HGH) is used to treat growth disorders in children. The HGH gene was cloned from a cDNA library and inserted into *E. coli* cells by cloning it into a bacterial vector.

Transgenic Animals

Although several recombinant proteins used in medicine are successfully produced in bacteria, some proteins require a eukaryotic animal host for proper processing. For this reason, the desired genes are cloned and expressed in animals, such as sheep, goats, chickens, and mice. Animals that have been modified to express recombinant DNA are called transgenic animals. Several human proteins are expressed in the milk of transgenic sheep and goats, and some are expressed in the eggs of chickens. Mice have been used extensively for expressing and studying the effects of recombinant genes and mutations.

Transgenic Plants

Manipulating the DNA of plants (i.e., creating GMOs) has helped to create desirable traits, such as disease resistance, herbicide and pesticide resistance, better nutritional value, and better shelf-life (**Figure 17.9**). Plants are the most important source of food for the human population. Farmers developed ways to select for plant varieties with desirable traits long before modern-day biotechnology practices were established.

Figure 17.9 Corn, a major agricultural crop used to create products for a variety of industries, is often modified through plant biotechnology. (credit: Keith Weller, USDA)

Plants that have received recombinant DNA from other species are called transgenic plants. Because they are not natural, transgenic plants and other GMOs are closely monitored by government agencies to ensure that they are fit for human consumption and do not endanger other plant and animal life. Because foreign genes can spread to other species in the environment, extensive testing is required to ensure ecological stability. Staples like corn, potatoes, and tomatoes were the first crop plants to be genetically engineered.

Transformation of Plants Using Agrobacterium tumefaciens

Gene transfer occurs naturally between species in microbial populations. Many viruses that cause human diseases act by incorporating their DNA into the human genome. In plants, tumors caused by the bacterium *Agrobacterium tumefaciens* occur by transfer of DNA from the bacterium to the plant. Although the tumors do not kill the plants, they make the plants stunted and more susceptible to harsh environmental conditions. Many plants, such as walnuts, grapes, nut trees, and beets, are affected by *A. tumefaciens*. The artificial introduction of DNA into plant cells is more challenging than in animal cells because of the thick plant cell wall.

Researchers used the natural transfer of DNA from *Agrobacterium* to a plant host to introduce DNA fragments of their choice into plant hosts. In nature, the disease-causing *A. tumefaciens* have a set of plasmids, called the **Ti plasmids** (tumor-inducing plasmids), that contain genes for the production of tumors in plants. DNA from the Ti plasmid integrates into the infected plant cell's genome. Researchers manipulate the Ti plasmids to remove the tumor-causing genes and insert the desired DNA fragment for transfer into the plant genome. The Ti plasmids carry antibiotic resistance genes to aid selection and can be propagated in *E. coli* cells as well.

The Organic Insecticide Bacillus thuringiensis

Bacillus thuringiensis (Bt) is a bacterium that produces protein crystals during sporulation that are toxic to many insect species that affect plants. Bt toxin has to be ingested by insects for the toxin to be activated. Insects that have eaten Bt toxin stop feeding on the plants within a few hours. After the toxin is activated in the intestines of the insects, death occurs within a couple of days. Modern biotechnology has allowed plants to encode their own crystal Bt toxin that acts against insects. The crystal toxin genes have been cloned from Bt and introduced into plants. Bt toxin has been found to be safe for the environment, non-toxic to humans and other mammals, and is approved for use by organic farmers as a natural insecticide.

Flavr Savr Tomato

The first GM crop was in 1994. It was a tomato that resisted rotting and maintained flavor for longer periods of time. Antisense RNA technology was used to slow down the process of softening and rotting caused by fungal infections, which led to increased shelf life of the GM tomatoes. Additional genetic modification improved the flavor of this tomato. This GM tomato did not successfully stay in the market because of problems maintaining and shipping the crop.

17.2 | Mapping Genomes

In this section, you will explore the following questions:

- What is genomics?
- What is a genetic map?
- What is an example of a genomic mapping method?

Connection for AP® Courses

Genome mapping is similar to solving a big, complicated puzzle with pieces of information collected from laboratories all over the world. Genetic maps provide an outline for the location of genes within a chromosome. Distances between genes and genetic markers are estimated on the basis of recombination (crossing over) frequencies during meiosis. The Human Genome Project helped researchers identify thousands of human genes and their protein products. Noncoding regions of DNA may be involved in regulating gene expression, and other sequences once considered "junk" may play an important role in genome evolution. Few differences exist between human DNA sequences and those of many other organisms.

Information presented and the examples highlighted in the section support concepts outlined in Big Idea 3 of the AP® Biology Curriculum Framework. The Learning Objectives listed in the Curriculum Framework provide a transparent foundation for the AP® Biology course, an inquiry-based laboratory experience, instructional activities, and AP® exam questions. A Learning Objective merges required content with one or more of the seven Science Practices.

Big Idea 3	Living systems store, retrieve, transmit and respond to information essential to life processes.

Enduring Understanding 3.A	Heritable information provides for continuity of life.
Essential Knowledge	**3.A.1** DNA, and in some cases RNA, is the primary source of heritable information.
Science Practice	**6.4** The student can make claims and predictions about natural phenomena based on scientific theories and models.
Learning Objective	**3.5** The student can justify the claim that humans can manipulate heritable information by identifying examples of commonly used technologies.
Essential Knowledge	**3.A.2** In eukaryotes, heritable information is passed to the next generation via processes that include the cell cycle and mitosis or meiosis.
Science Practice	**7.1** The student can connect phenomena and models across spatial and temporal scales.
Learning Objective	**3.10** The student is able to represent the connection between meiosis and increased genetic diversity necessary for evolution.
Essential Knowledge	**3.A.3** The chromosomal basis of inheritance provides an understanding of the pattern of passage (transmission) of genes from parent to offspring.
Science Practice	**1.1** The student can create representations and models of natural or man-made phenomena and systems in the domain.
Learning Objective	**7.2** The student can connect concepts in and across domain(s) to generalize or extrapolate in and/or across enduring understandings and/or big ideas.

Genomics is the study of entire genomes, including the complete set of genes, their nucleotide sequence and organization, and their interactions within a species and with other species. **Genome mapping** is the process of finding the locations of genes on each chromosome. The maps created by genome mapping are comparable to the maps that we use to navigate streets. A **genetic map** is an illustration that lists genes and their location on a chromosome. Genetic maps provide the big picture (similar to a map of interstate highways) and use genetic markers (similar to landmarks). A **genetic marker** is a gene or sequence on a chromosome that co-segregates (shows genetic linkage) with a specific trait. Early geneticists called this linkage analysis. Physical maps present the intimate details of smaller regions of the chromosomes (similar to a detailed road map). A **physical map** is a representation of the physical distance, in nucleotides, between genes or genetic markers. Both genetic linkage maps and physical maps are required to build a complete picture of the genome. Having a complete map of the genome makes it easier for researchers to study individual genes. Human genome maps help researchers in their efforts to identify human disease-causing genes related to illnesses like cancer, heart disease, and cystic fibrosis. Genome mapping can be used in a variety of other applications, such as using live microbes to clean up pollutants or even prevent pollution. Research involving plant genome mapping may lead to producing higher crop yields or developing plants that better adapt to climate change.

Genetic Maps

The study of genetic maps begins with **linkage analysis**, a procedure that analyzes the recombination frequency between genes to determine if they are linked or show independent assortment. The term *linkage* was used before the discovery of DNA. Early geneticists relied on the observation of phenotypic changes to understand the genotype of an organism. Shortly after Gregor Mendel (the father of modern genetics) proposed that traits were determined by what are now known as genes, other researchers observed that different traits were often inherited together, and thereby deduced that the genes were physically linked by being located on the same chromosome. The mapping of genes relative to each other based on linkage analysis led to the development of the first genetic maps.

Observations that certain traits were always linked and certain others were not linked came from studying the offspring of crosses between parents with different traits. For example, in experiments performed on the garden pea, it was discovered that the color of the flower and shape of the plant's pollen were linked traits, and therefore the genes encoding these traits were in close proximity on the same chromosome. The exchange of DNA between homologous pairs of chromosomes is called **genetic recombination**, which occurs by the crossing over of DNA between homologous strands of DNA, such as nonsister chromatids. Linkage analysis involves studying the recombination frequency between any two genes. The greater the distance between two genes, the higher the chance that a recombination event will occur between them, and the higher the recombination frequency between them. Two possibilities for recombination between two nonsister chromatids during

meiosis are shown in Figure 17.10. If the recombination frequency between two genes is less than 50 percent, they are said to be linked.

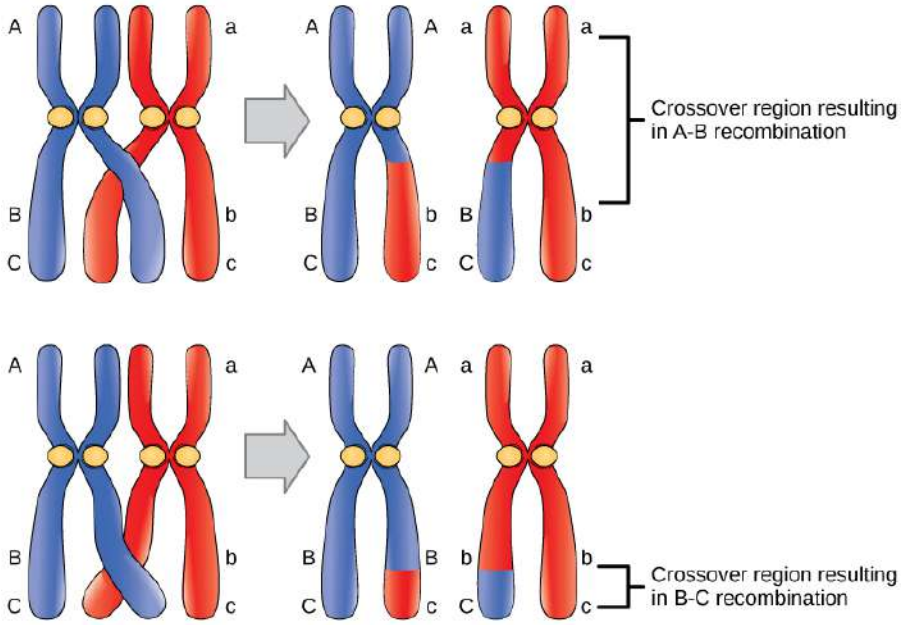


Figure 17.10 Crossover may occur at different locations on the chromosome. Recombination between genes *A* and *B* is more frequent than recombination between genes *B* and *C* because genes *A* and *B* are farther apart; a crossover is therefore more likely to occur between them.

The generation of genetic maps requires markers, just as a road map requires landmarks (such as rivers and mountains). Early genetic maps were based on the use of known genes as markers. More sophisticated markers, including those based on non-coding DNA, are now used to compare the genomes of individuals in a population. Although individuals of a given species are genetically similar, they are not identical; every individual has a unique set of traits. These minor differences in the genome between individuals in a population are useful for the purposes of genetic mapping. In general, a good genetic marker is a region on the chromosome that shows variability or polymorphism (multiple forms) in the population.

Some genetic markers used in generating genetic maps are **restriction fragment length polymorphisms** (RFLP), variable number of tandem repeats (VNTRs), **microsatellite polymorphisms**, and the **single nucleotide polymorphisms** (SNPs). RFLPs (sometimes pronounced “rif-lips”) are detected when the DNA of an individual is cut with a restriction endonuclease that recognizes specific sequences in the DNA to generate a series of DNA fragments, which are then analyzed by gel electrophoresis. The DNA of every individual will give rise to a unique pattern of bands when cut with a particular set of restriction endonucleases; this is sometimes referred to as an individual’s DNA “fingerprint.” Certain regions of the chromosome that are subject to polymorphism will lead to the generation of the unique banding pattern. VNTRs are repeated sets of nucleotides present in the non-coding regions of DNA. Non-coding, or “junk,” DNA has no known biological function; however, research shows that much of this DNA is actually transcribed. While its function is uncertain, it is certainly active, and it may be involved in the regulation of coding genes. The number of repeats may vary in individual organisms of a population. Microsatellite polymorphisms are similar to VNTRs, but the repeat unit is very small. SNPs are variations in a single nucleotide.

Because genetic maps rely completely on the natural process of recombination, mapping is affected by natural increases or decreases in the level of recombination in any given area of the genome. Some parts of the genome are recombination hotspots, whereas others do not show a propensity for recombination. For this reason, it is important to look at mapping information developed by multiple methods.

Physical Maps

A physical map provides detail of the actual physical distance between genetic markers, as well as the number of nucleotides. There are three methods used to create a physical map: cytogenetic mapping, radiation hybrid mapping, and sequence mapping. **Cytogenetic mapping** uses information obtained by microscopic analysis of stained sections of the chromosome (Figure 17.11). It is possible to determine the approximate distance between genetic markers using cytogenetic mapping, but not the exact distance (number of base pairs). **Radiation hybrid mapping** uses radiation, such as x-rays, to break the DNA into fragments. The amount of radiation can be adjusted to create smaller or larger fragments. This technique

overcomes the limitation of genetic mapping and is not affected by increased or decreased recombination frequency. **Sequence mapping** resulted from DNA sequencing technology that allowed for the creation of detailed physical maps with distances measured in terms of the number of base pairs. The creation of **genomic libraries** and **complementary DNA (cDNA) libraries** (collections of cloned sequences or all DNA from a genome) has sped up the process of physical mapping. A genetic site used to generate a physical map with sequencing technology (a sequence-tagged site, or STS) is a unique sequence in the genome with a known exact chromosomal location. An **expressed sequence tag (EST)** and a single sequence length polymorphism (SSLP) are common STSs. An EST is a short STS that is identified with cDNA libraries, while SSLPs are obtained from known genetic markers and provide a link between genetic maps and physical maps.

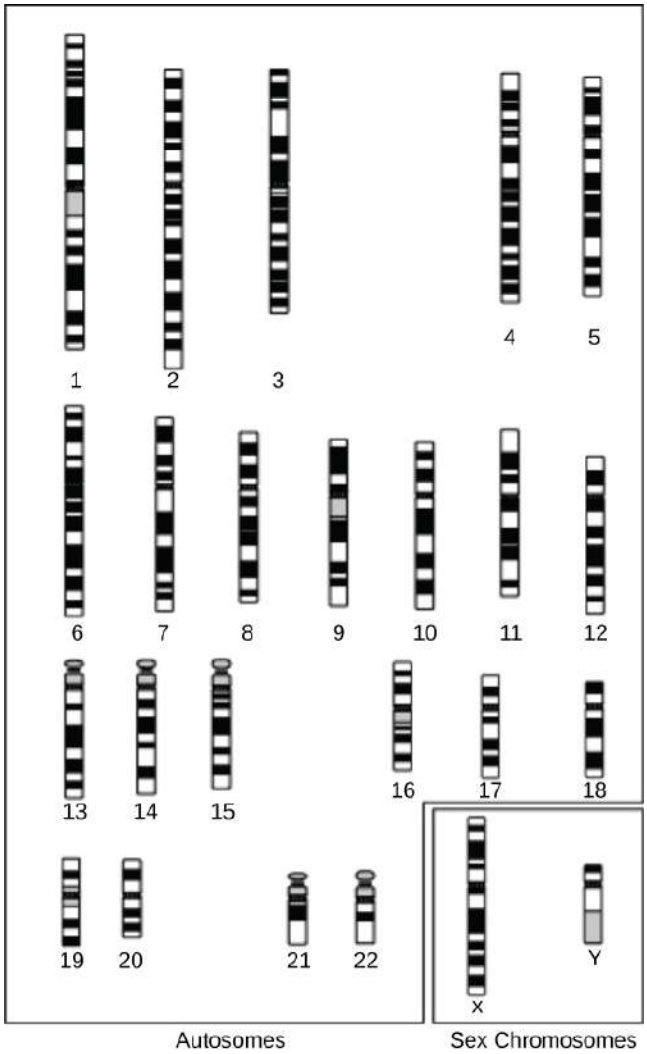


Figure 17.11 A cytogenetic map shows the appearance of a chromosome after it is stained and examined under a microscope. (credit: National Human Genome Research Institute)

Integration of Genetic and Physical Maps

Genetic maps provide the outline and physical maps provide the details. It is easy to understand why both types of genome mapping techniques are important to show the big picture. Information obtained from each technique is used in combination to study the genome. Genomic mapping is being used with different model organisms that are used for research. Genome mapping is still an ongoing process, and as more advanced techniques are developed, more advances are expected. Genome mapping is similar to completing a complicated puzzle using every piece of available data. Mapping information generated in laboratories all over the world is entered into central databases, such as GenBank at the National Center for Biotechnology Information (NCBI). Efforts are being made to make the information more easily accessible to researchers and the general public. Just as we use global positioning systems instead of paper maps to navigate through roadways, NCBI has created a genome viewer tool to simplify the data-mining process.

How to Use a Genome Map Viewer

Problem statement: Do the human, macaque, and mouse genomes contain common DNA sequences?

Develop a hypothesis.

To test the hypothesis, click this **link (http://openstaxcollege.org/l/32mapview)** .

In Search box on the left panel, type any gene name or phenotypic characteristic, such as iris pigmentation (eye color). Select the species you want to study, and then press Enter. The genome map viewer will indicate which chromosome encodes the gene in your search. Click each hit in the genome viewer for more detailed information. This type of search is the most basic use of the genome viewer; it can also be used to compare sequences between species, as well as many other complicated tasks.

Is the hypothesis correct? Why or why not?

Online Mendelian Inheritance in Man (OMIM) is a searchable online catalog of human genes and genetic disorders. This website shows genome mapping information, and also details the history and research of each trait and disorder. Click this **link (http://openstaxcollege.org/l/OMIM)** to search for traits (such as handedness) and genetic disorders (such as diabetes).

How can this database help to support and guide research for rare genetic conditions, like progeria?

a. The database provides information related to the prevention of a genetic disease.

b. The database provides all the information about genes for genetic diseases, their inheritance and their expression. It also provides suggestions for some treatments.

c. The database provides information about the symptoms of the disease.

d. The database provides information only about the early reported cases.

Think About It

Why is so much effort being poured into genome mapping applications? How could a genetic map of the human genome help find a treatment for genetically based cancers?

17.3 | Whole-Genome Sequencing

In this section, you will explore the following questions:

- What are three types of gene sequencing?
- What is whole-genome sequencing?

Connection for AP® Courses

Information presented in section is not in scope for AP®. However, you can study information in the section as optional or illustrative material.

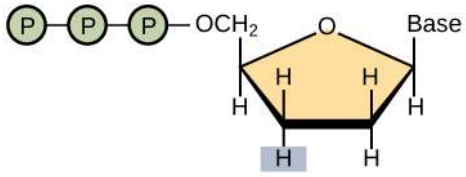
Although there have been significant advances in the medical sciences in recent years, doctors are still confounded by some diseases, and they are using whole-genome sequencing to get to the bottom of the problem. **Whole-genome sequencing** is a process that determines the DNA sequence of an entire genome. Whole-genome sequencing is a brute-force approach to problem solving when there is a genetic basis at the core of a disease. Several laboratories now provide services to sequence, analyze, and interpret entire genomes.

Whole-exome sequencing is a lower-cost alternative to whole genome sequencing. In exome sequencing, only the coding, exon-producing regions of the DNA are sequenced. In 2010, whole-exome sequencing was used to save a young boy whose intestines had multiple mysterious abscesses. The child had several colon operations with no relief. Finally, whole-exome sequencing was performed, which revealed a defect in a pathway that controls apoptosis (programmed cell death). A bone-marrow transplant was used to overcome this genetic disorder, leading to a cure for the boy. He was the first person to be successfully treated based on a diagnosis made by whole-exome sequencing.

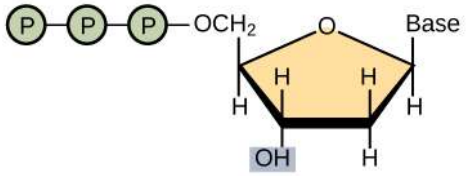
The Science Practice Challenge Questions contain additional test questions related to the material in this section that will help you prepare for the AP exam. These questions address the following standards:
[APLO 2.23][APLO 3.5][APLO 3.20][APLO 3.21]

Strategies Used in Sequencing Projects

The basic sequencing technique used in all modern day sequencing projects is the chain termination method (also known as the dideoxy method), which was developed by Fred Sanger in the 1970s. The chain termination method involves DNA replication of a single-stranded template with the use of a primer and a regular **deoxynucleotide** (dNTP), which is a monomer, or a single unit, of DNA. The primer and dNTP are mixed with a small proportion of fluorescently labeled **dideoxynucleotides** (ddNTPs). The ddNTPs are monomers that are missing a hydroxyl group (–OH) at the site at which another nucleotide usually attaches to form a chain (**Figure 17.12**). Each ddNTP is labeled with a different color of fluorophore. Every time a ddNTP is incorporated in the growing complementary strand, it terminates the process of DNA replication, which results in multiple short strands of replicated DNA that are each terminated at a different point during replication. When the reaction mixture is processed by gel electrophoresis after being separated into single strands, the multiple newly replicated DNA strands form a ladder because of the differing sizes. Because the ddNTPs are fluorescently labeled, each band on the gel reflects the size of the DNA strand and the ddNTP that terminated the reaction. The different colors of the fluorophore-labeled ddNTPs help identify the ddNTP incorporated at that position. Reading the gel on the basis of the color of each band on the ladder produces the sequence of the template strand (**Figure 17.13**).



Dideoxynucleotide (ddNTP)



Deoxynucleotide (dNTP)

Figure 17.12 A dideoxynucleotide is similar in structure to a deoxynucleotide, but is missing the 3' hydroxyl group (indicated by the box). When a dideoxynucleotide is incorporated into a DNA strand, DNA synthesis stops.

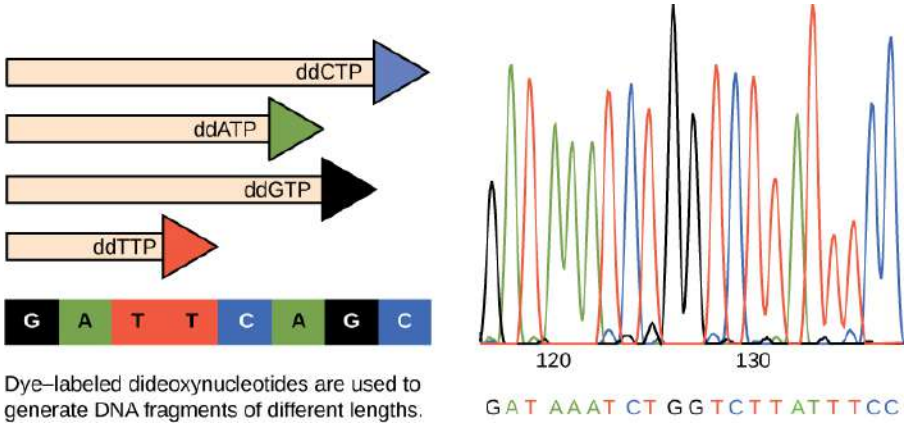


Figure 17.13 Frederick Sanger's dideoxy chain termination method is illustrated. Using dideoxynucleotides, the DNA fragment can be terminated at different points. The DNA is separated on the basis of size, and these bands, based on the size of the fragments, can be read.

Early Strategies: Shotgun Sequencing and Pair-Wise End Sequencing

In **shotgun sequencing** method, several copies of a DNA fragment are cut randomly into many smaller pieces (somewhat like what happens to a round shot cartridge when fired from a shotgun). All of the segments are then sequenced using the chain-sequencing method. Then, with the help of a computer, the fragments are analyzed to see where their sequences overlap. By matching up overlapping sequences at the end of each fragment, the entire DNA sequence can be reformed. A larger sequence that is assembled from overlapping shorter sequences is called a **contig**. As an analogy, consider that someone has four copies of a landscape photograph that you have never seen before and know nothing about how it should appear. The person then rips up each photograph with their hands, so that different size pieces are present from each copy. The person then mixes all of the pieces together and asks you to reconstruct the photograph. In one of the smaller pieces you see a mountain. In a larger piece, you see that the same mountain is behind a lake. A third fragment shows only the lake, but it reveals that there is a cabin on the shore of the lake. Therefore, from looking at the overlapping information in these three fragments, you know that the picture contains a mountain behind a lake that has a cabin on its shore. This is the principle behind reconstructing entire DNA sequences using shotgun sequencing.

Originally, shotgun sequencing only analyzed one end of each fragment for overlaps. This was sufficient for sequencing small genomes. However, the desire to sequence larger genomes, such as that of a human, led to the development of double-barrel shotgun sequencing, more formally known as **pairwise-end sequencing**. In pairwise-end sequencing, both ends of each fragment are analyzed for overlap. Pairwise-end sequencing is, therefore, more cumbersome than shotgun sequencing, but it is easier to reconstruct the sequence because there is more available information.

Next-generation Sequencing

Since 2005, automated sequencing techniques used by laboratories are under the umbrella of **next-generation sequencing**, which is a group of automated techniques used for rapid DNA sequencing. These automated low-cost sequencers can generate sequences of hundreds of thousands or millions of short fragments (25 to 500 base pairs) in the span of one day. These sequencers use sophisticated software to get through the cumbersome process of putting all the fragments in order.

evolution CONNECTION

Comparing Sequences

A sequence alignment is an arrangement of proteins, DNA, or RNA; it is used to identify regions of similarity between cell types or species, which may indicate conservation of function or structures. Sequence alignments may be used to construct phylogenetic trees. The following website uses a software program called **BLAST (basic local alignment search tool) (http://openstaxcollege.org/l/32blast)** .

Under “Basic Blast,” click “Nucleotide Blast.” Input the following sequence into the large "query sequence" box: ATTGCTTCGATTGCA. Below the box, locate the "Species" field and type "human" or "Homo sapiens". Then click “BLAST” to compare the inputted sequence against known sequences of the human genome. The result is that this sequence occurs in over a hundred places in the human genome. Scroll down below the graphic with the horizontal bars and you will see short description of each of the matching hits. Pick one of the hits near the top of the list and click on "Graphics". This will bring you to a page that shows where the sequence is found within the entire human genome. You can move the slider that looks like a green flag back and forth to view the sequences immediately around the selected gene. You can then return to your selected sequence by clicking the "ATG" button.

Cytochrome c oxidase is a highly conserved protein found in bacteria and in the mitochondria of eukaryotes. Based on your knowledge of evolutionary relationships, which of the following statements would you expect to be true of the cytochrome c oxidase protein sequence?

a. The bacterial protein will be more similar to the human protein than the yeast protein.
b. The yeast protein will be more similar to the human protein than the bacterial protein.
c. The bacterial protein will be more similar to the yeast protein than the human protein.
d. The bacterial and yeast protein will share a similar sequence, but the human protein will be unrelated.

Use of Whole-Genome Sequences of Model Organisms

The first genome to be completely sequenced was of a bacterial virus, the bacteriophage fx174 (5368 base pairs); this was accomplished by Fred Sanger using shotgun sequencing. Several other organelle and viral genomes were later sequenced. The first organism whose genome was sequenced was the bacterium *Haemophilus influenzae*; this was accomplished by Craig Venter in the 1980s. Approximately 74 different laboratories collaborated on the sequencing of the genome of the yeast *Saccharomyces cerevisiae*, which began in 1989 and was completed in 1996, because it was 60 times bigger than any other genome that had been sequenced. By 1997, the genome sequences of two important model organisms were available: the bacterium *Escherichia coli* K12 and the yeast *Saccharomyces cerevisiae*. Genomes of other model organisms, such as the mouse *Mus musculus*, the fruit fly *Drosophila melanogaster*, the nematode *Caenorhabditis. elegans*, and humans *Homo sapiens* are now known. A lot of basic research is performed in model organisms because the information can be applied to genetically similar organisms. A **model organism** is a species that is studied as a model to understand the biological processes in other species represented by the model organism. Having entire genomes sequenced helps with the research efforts in these model organisms. The process of attaching biological information to gene sequences is called **genome annotation**. Annotation of gene sequences helps with basic experiments in molecular biology, such as designing PCR primers and RNA targets.

Click through each step of genome sequencing at this **site (http://openstaxcollege.org/l/DNA_sequence)** .

How is deep sequencing an improvement on Sanger sequencing?

a. Deep sequencing allows for much faster sequencing of short strands of DNA as compared to Sanger sequencing which, reads only short sequences of DNA at a slow rate. Also, there is a high risk of chain termination and problems with separation during Sanger sequencing.

b. Sequence coverage is higher in Sanger sequencing as compared to deep sequencing.

c. Sanger sequencing is suitable when there is only one nucleotide different between chains whereas deep sequencing is suitable when there is one or more than one nucleotide different between chains.

d. Sanger sequencing reads and sequences a genome multiple times whereas deep sequencing accurately sequences the whole genome in a single time.

Uses of Genome Sequences

DNA microarrays are methods used to detect gene expression by analyzing an array of DNA fragments that are fixed to a glass slide or a silicon chip to identify active genes and identify sequences. Almost one million genotypic abnormalities can be discovered using microarrays, whereas whole-genome sequencing can provide information about all six billion base pairs in the human genome. Although the study of medical applications of genome sequencing is interesting, this discipline tends to dwell on abnormal gene function. Knowledge of the entire genome will allow future onset diseases and other genetic disorders to be discovered early, which will allow for more informed decisions to be made about lifestyle, medication, and having children. Genomics is still in its infancy, although someday it may become routine to use whole-genome sequencing to screen every newborn to detect genetic abnormalities.

In addition to disease and medicine, genomics can contribute to the development of novel enzymes that convert biomass to biofuel, which results in higher crop and fuel production, and lower cost to the consumer. This knowledge should allow better methods of control over the microbes that are used in the production of biofuels. Genomics could also improve the methods used to monitor the impact of pollutants on ecosystems and help clean up environmental contaminants. Genomics has allowed for the development of agrochemicals and pharmaceuticals that could benefit medical science and agriculture.

It sounds great to have all the knowledge we can get from whole-genome sequencing; however, humans have a responsibility to use this knowledge wisely. Otherwise, it could be easy to misuse the power of such knowledge, leading to discrimination based on a person's genetics, human genetic engineering, and other ethical concerns. This information could also lead to legal issues regarding health and privacy.

17.4 | Applying Genomics

In this section, you will explore the following questions:

- What is pharmacogenomics?
- What is an example of a polygenic human disease?

Connection for AP® Courses

Information presented in section is not in scope for AP®. However, you can study information in the section as optional or illustrative material.

The introduction of DNA sequencing and whole genome sequencing projects, particularly the Human Genome project, has expanded the applicability of DNA sequence information. Genomics is now being used in a wide variety of fields, such as metagenomics, pharmacogenomics, and mitochondrial genomics. The most commonly known application of genomics is to understand and find cures for diseases.

Predicting Disease Risk at the Individual Level

Predicting the risk of disease involves screening currently healthy individuals by genome analysis at the individual level. Intervention with lifestyle changes and drugs can be recommended before disease onset. However, this approach is most applicable when the problem resides within a single gene defect. Such defects only account for approximately 5 percent of diseases in developed countries. Most of the common diseases, such as heart disease, are multi-factored or **polygenic**, which is a phenotypic characteristic that involves two or more genes, and also involve environmental factors such as diet. In April 2010, scientists at Stanford University published the genome analysis of a healthy individual (Stephen Quake, a scientist at Stanford University, who had his genome sequenced); the analysis predicted his propensity to acquire various diseases. A risk assessment was performed to analyze Quake's percentage of risk for 55 different medical conditions. A rare genetic mutation was found, which showed him to be at risk for sudden heart attack. He was also predicted to have a 23 percent risk of developing prostate cancer and a 1.4 percent risk of developing Alzheimer's. The scientists used databases and several publications to analyze the genomic data. Even though genomic sequencing is becoming more affordable and analytical tools are becoming more reliable, ethical issues surrounding genomic analysis at a population level remain to be addressed.

visual CONNECTION

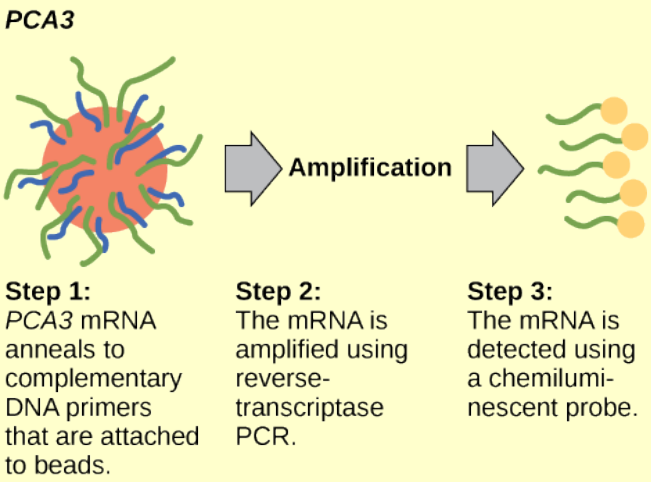


Figure 17.14 *PCA3* is a gene that is expressed in prostate epithelial cells and overexpressed in cancerous cells. A high concentration of *PCA3* in urine is indicative of prostate cancer. The *PCA3* test is considered to be a better indicator of cancer than the more well know PSA test, which measures the level of PSA (prostate-specific antigen) in the blood.

In 2011, the United States Preventative Services Task Force recommended against using the PSA test to screen healthy men for prostate cancer. Their recommendation is based on evidence that screening does not reduce the risk of death from prostate cancer. Prostate cancer often develops very slowly and does not cause problems, while the cancer treatment can have severe side effects. The PCA3 test is considered to be more accurate, but screening may still result in men who would not have been harmed by the cancer itself suffering side effects from treatment. What do you think? Should all healthy men be screened for prostate cancer using the PCA3 or PSA test? Should people in general be screened to find out if they have a genetic risk for cancer or other diseases?

a. In general, all men should be screened as it is necessary to take the risk.
b. In general, all men should be given treatment irrespective of the presence or absence of cancer symptoms.
c. In general, only men suspecting cancer should be screened.
d. There is no requirement of any screening, as cancer shows no problems.

Pharmacogenomics and Toxicogenomics

Pharmacogenomics, also called toxicogenomics, involves evaluating the effectiveness and safety of drugs on the basis of information from an individual's genomic sequence. Genomic responses to drugs can be studied using experimental animals (such as laboratory rats or mice) or live cells in the laboratory before embarking on studies with humans. Studying changes in gene expression could provide information about the transcription profile in the presence of the drug, which can be used as an early indicator of the potential for toxic effects. For example, genes involved in cellular growth and controlled cell death, when disturbed, could lead to the growth of cancerous cells. Genome-wide studies can also help to find new genes involved in drug toxicity. Personal genome sequence information can be used to prescribe medications that will be most effective and least toxic on the basis of the individual patient's genotype. The gene signatures may not be completely accurate, but can be tested further before pathologic symptoms arise.

Microbial Genomics: Metagenomics

Traditionally, microbiology has been taught with the view that microorganisms are best studied under **pure culture** conditions, which involves isolating a single type of cell and culturing it in the laboratory. Because microorganisms can go through several generations in a matter of hours, their gene expression profiles adapt to the new laboratory environment very quickly. In addition, the vast majority of bacterial species resist being cultured in isolation. Most microorganisms do not live as isolated entities, but in microbial communities known as biofilms. For all of these reasons, pure culture is not always the best way to study microorganisms. **Metagenomics** is the study of the collective genomes of multiple species that grow and interact in an environmental niche. Metagenomics can be used to identify new species more rapidly and to analyze the effect of pollutants on the environment (Figure 17.15).

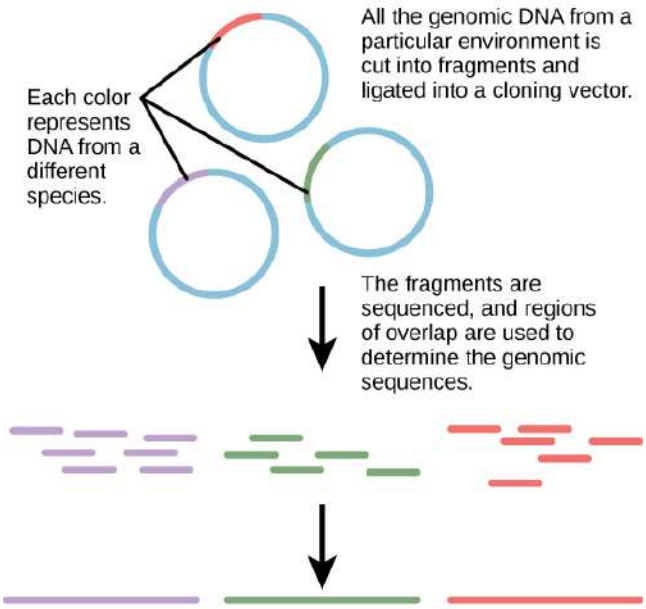


Figure 17.15 Metagenomics involves isolating DNA from multiple species within an environmental niche.

Microbial Genomics: Creation of New Biofuels

Knowledge of the genomics of microorganisms is being used to find better ways to harness biofuels from algae and cyanobacteria. The primary sources of fuel today are coal, oil, wood, and other plant products, such as ethanol. Although plants are renewable resources, there is still a need to find more alternative renewable sources of energy to meet our population's energy demands. The microbial world is one of the largest resources for genes that encode new enzymes and produce new organic compounds, and it remains largely untapped. Microorganisms are used to create products, such as enzymes that are used in research, antibiotics, and other anti-microbial mechanisms. Microbial genomics is helping to develop diagnostic tools, improved vaccines, new disease treatments, and advanced environmental cleanup techniques.

Mitochondrial Genomics

Mitochondria are intracellular organelles that contain their own DNA. Mitochondrial DNA mutates at a rapid rate and is often used to study evolutionary relationships. Another feature that makes studying the mitochondrial genome interesting is that the mitochondrial DNA in most multicellular organisms is passed on from the mother during the process of fertilization.

For this reason, mitochondrial genomics is often used to trace genealogy.

Information and clues obtained from DNA samples found at crime scenes have been used as evidence in court cases, and genetic markers have been used in forensic analysis. Genomic analysis has also become useful in this field. In 2001, the first use of genomics in forensics was published. It was a collaborative attempt between academic research institutions and the FBI to solve the mysterious cases of anthrax communicated via the US Postal Service. Using microbial genomics, researchers determined that a specific strain of anthrax was used in all the mailings.

Genomics in Agriculture

Genomics can reduce the trials and failures involved in scientific research to a certain extent, which could improve the quality and quantity of crop yields in agriculture. Linking traits to genes or gene signatures helps to improve crop breeding to generate hybrids with the most desirable qualities. Scientists use genomic data to identify desirable traits, and then transfer those traits to a different organism. Scientists are discovering how genomics can improve the quality and quantity of agricultural production. For example, scientists could use desirable traits to create a useful product or enhance an existing product, such as making a drought-sensitive crop more tolerant of the dry season.

17.5 | Genomics and Proteomics

In this section, you will explore the following questions:

- What is a proteome?
- What is a protein signature and what is its relevance to cancer screening?

Connection for AP® Courses

Information presented in section is not in scope for AP®. However, you can study information in the section as optional or illustrative material.

Proteins are the final products of genes, which help perform the function encoded by the gene. Proteins are composed of amino acids and play important roles in the cell. All enzymes (except ribozymes) are proteins that act as catalysts to affect the rate of reactions. Proteins are also regulatory molecules, and some are hormones. Transport proteins, such as hemoglobin, help transport oxygen to various organs. Antibodies that defend against foreign particles are also proteins. In the diseased state, protein function can be impaired because of changes at the genetic level or because of direct impact on a specific protein.

A **proteome** is the entire set of proteins produced by a cell type. Proteomes can be studied using the knowledge of genomes because genes code for mRNAs, and the mRNAs encode proteins. Although mRNA analysis is a step in the right direction, not all mRNAs are translated into proteins. The study of the function of proteomes is called **proteomics**. Proteomics complements genomics and is useful when scientists want to test their hypotheses that were based on genes. Even though all cells of a multicellular organism have the same set of genes, the set of proteins produced in different tissues is different and dependent on gene expression. Thus, the genome is constant, but the proteome varies and is dynamic within an organism. In addition, RNAs can be alternately spliced (cut and pasted to create novel combinations and novel proteins) and many proteins are modified after translation by processes such as proteolytic cleavage, phosphorylation, glycosylation, and ubiquitination. There are also protein-protein interactions, which complicate the study of proteomes. Although the genome provides a blueprint, the final architecture depends on several factors that can change the progression of events that generate the proteome.

Metabolomics is related to genomics and proteomics. **Metabolomics** involves the study of small molecule metabolites found in an organism. The **metabolome** is the complete set of metabolites that are related to the genetic makeup of an organism. Metabolomics offers an opportunity to compare genetic makeup and physical characteristics, as well as genetic makeup and environmental factors. The goal of metabolome research is to identify, quantify, and catalogue all of the metabolites that are found in the tissues and fluids of living organisms.

Basic Techniques in Protein Analysis

The ultimate goal of proteomics is to identify or compare the proteins expressed from a given genome under specific conditions, study the interactions between the proteins, and use the information to predict cell behavior or develop drug targets. Just as the genome is analyzed using the basic technique of DNA sequencing, proteomics requires techniques

for protein analysis. The basic technique for protein analysis, analogous to DNA sequencing, is mass spectrometry. Mass spectrometry is used to identify and determine the characteristics of a molecule. Advances in spectrometry have allowed researchers to analyze very small samples of protein. X-ray crystallography, for example, enables scientists to determine the three-dimensional structure of a protein crystal at atomic resolution. Another protein imaging technique, nuclear magnetic resonance (NMR), uses the magnetic properties of atoms to determine the three-dimensional structure of proteins in aqueous solution. Protein microarrays have also been used to study interactions between proteins. Large-scale adaptations of the basic two-hybrid screen (**Figure 17.16**) have provided the basis for protein microarrays. Computer software is used to analyze the vast amount of data generated for proteomic analysis.

Genomic- and proteomic-scale analyses are part of systems biology. **Systems biology** is the study of whole biological systems (genomes and proteomes) based on interactions within the system. The European Bioinformatics Institute and the Human Proteome Organization (HUPO) are developing and establishing effective tools to sort through the enormous pile of systems biology data. Because proteins are the direct products of genes and reflect activity at the genomic level, it is natural to use proteomes to compare the protein profiles of different cells to identify proteins and genes involved in disease processes. Most pharmaceutical drug trials target proteins. Information obtained from proteomics is being used to identify novel drugs and understand their mechanisms of action.

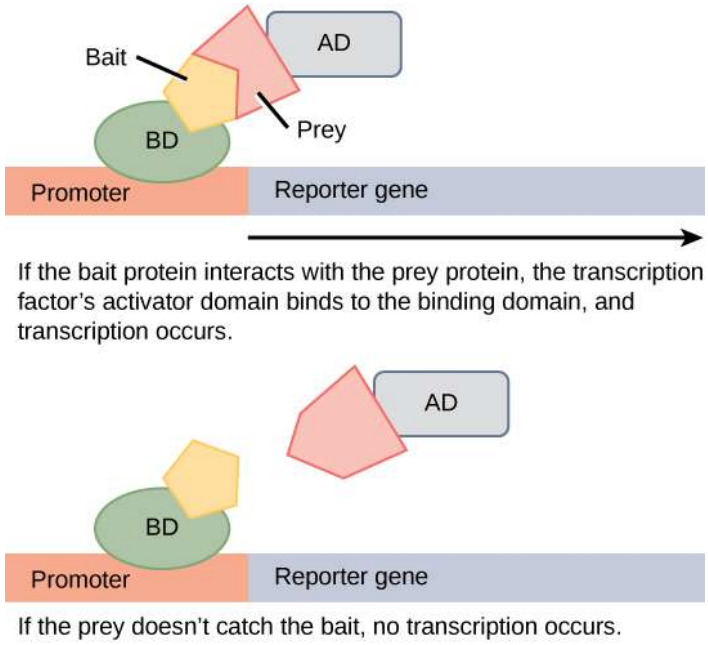


Figure 17.16 Two-hybrid screening is used to determine whether two proteins interact. In this method, a transcription factor is split into a DNA-binding domain (BD) and an activator domain (AD). The binding domain is able to bind the promoter in the absence of the activator domain, but it does not turn on transcription. A protein called the bait is attached to the BD, and a protein called the prey is attached to the AD. Transcription occurs only if the prey "catches" the bait.

The challenge of techniques used for proteomic analyses is the difficulty in detecting small quantities of proteins. Although mass spectrometry is good for detecting small amounts of proteins, variations in protein expression in diseased states can be difficult to discern. Proteins are naturally unstable molecules, which makes proteomic analysis much more difficult than genomic analysis.

Cancer Proteomics

Genomes and proteomes of patients suffering from specific diseases are being studied to understand the genetic basis of the disease. The most prominent disease being studied with proteomic approaches is cancer. Proteomic approaches are being used to improve screening and early detection of cancer; this is achieved by identifying proteins whose expression is affected by the disease process. An individual protein is called a **biomarker**, whereas a set of proteins with altered expression levels is called a **protein signature**. For a biomarker or protein signature to be useful as a candidate for early screening and detection of a cancer, it must be secreted in body fluids, such as sweat, blood, or urine, such that large-scale screenings can be performed in a non-invasive fashion. The current problem with using biomarkers for the early detection of cancer is the high rate of false-negative results. A **false negative** is an incorrect test result that should have been positive. In other words, many cases of cancer go undetected, which makes biomarkers unreliable. Some examples of

protein biomarkers used in cancer detection are CA-125 for ovarian cancer and PSA for prostate cancer. Protein signatures may be more reliable than biomarkers to detect cancer cells. Proteomics is also being used to develop individualized treatment plans, which involves the prediction of whether or not an individual will respond to specific drugs and the side effects that the individual may experience. Proteomics is also being used to predict the possibility of disease recurrence.

The National Cancer Institute has developed programs to improve the detection and treatment of cancer. The Clinical Proteomic Technologies for Cancer and the Early Detection Research Network are efforts to identify protein signatures specific to different types of cancers. The Biomedical Proteomics Program is designed to identify protein signatures and design effective therapies for cancer patients.

KEY TERMS

antibiotic resistance ability of an organism to be unaffected by the actions of an antibiotic

biomarker individual protein that is uniquely produced in a diseased state

biotechnology use of biological agents for technological advancement

cDNA library collection of cloned cDNA sequences

cellular cloning production of identical cell populations by binary fission

chain termination method method of DNA sequencing using labeled dideoxynucleotides to terminate DNA replication; it is also called the dideoxy method or the Sanger method

clone exact replica

contig larger sequence of DNA assembled from overlapping shorter sequences

cytogenetic mapping technique that uses a microscope to create a map from stained chromosomes

deoxynucleotide individual monomer (single unit) of DNA

dideoxynucleotide individual monomer of DNA that is missing a hydroxyl group (–OH)

DNA microarray method used to detect gene expression by analyzing an array of DNA fragments that are fixed to a glass slide or a silicon chip to identify active genes and identify sequences

expressed sequence tag (EST) short STS that is identified with cDNA

false negative incorrect test result that should have been positive

foreign DNA DNA that belongs to a different species or DNA that is artificially synthesized

gel electrophoresis technique used to separate molecules on the basis of size using electric charge

gene targeting method for altering the sequence of a specific gene by introducing the modified version on a vector

gene therapy technique used to cure inheritable diseases by replacing mutant genes with good genes

genetic diagnosis diagnosis of the potential for disease development by analyzing disease-causing genes

genetic engineering alteration of the genetic makeup of an organism

genetic map outline of genes and their location on a chromosome

genetic marker gene or sequence on a chromosome with a known location that is associated with a specific trait

genetic recombination exchange of DNA between homologous pairs of chromosomes

genetic testing process of testing for the presence of disease-causing genes

genetically modified organism (GMO) organism whose genome has been artificially changed

genome annotation process of attaching biological information to gene sequences

genome mapping process of finding the location of genes on each chromosome

genomic library collection of cloned DNA which represents all of the sequences and fragments from a genome

genomics study of entire genomes including the complete set of genes, their nucleotide sequence and organization, and their interactions within a species and with other species

host DNA DNA that is present in the genome of the organism of interest

linkage analysis procedure that analyzes the recombination of genes to determine if they are linked

lysis buffer solution used to break the cell membrane and release cell contents

metabolome complete set of metabolites which are related to the genetic makeup of an organism

metabolomics study of small molecule metabolites found in an organism

metagenomics study of the collective genomes of multiple species that grow and interact in an environmental niche

microsatellite polymorphism variation between individuals in the sequence and number of repeats of microsatellite DNA

model organism species that is studied and used as a model to understand the biological processes in other species represented by the model organism

molecular cloning cloning of DNA fragments

multiple cloning site (MCS) site that can be recognized by multiple restriction endonucleases

next-generation sequencing group of automated techniques used for rapid DNA sequencing

northern blotting transfer of RNA from a gel to a nylon membrane

pharmacogenomics study of drug interactions with the genome or proteome; also called toxicogenomics

physical map representation of the physical distance between genes or genetic markers

polygenic phenotypic characteristic caused by two or more genes

polymerase chain reaction (PCR) technique used to amplify DNA

probe small DNA fragment used to determine if the complementary sequence is present in a DNA sample

protease enzyme that breaks down proteins

protein signature set of uniquely expressed proteins in the diseased state

proteome entire set of proteins produced by a cell type

proteomics study of the function of proteomes

pure culture growth of a single type of cell in the laboratory

radiation hybrid mapping information obtained by fragmenting the chromosome with x-rays

recombinant DNA combination of DNA fragments generated by molecular cloning that does not exist in nature; also known as a chimeric molecule

recombinant protein protein product of a gene derived by molecular cloning

reproductive cloning cloning of entire organisms

restriction endonuclease enzyme that can recognize and cleave specific DNA sequences

restriction fragment length polymorphism (RFLP) variation between individuals in the length of DNA fragments generated by restriction endonucleases

reverse genetics method of determining the function of a gene by starting with the gene itself instead of starting with the gene product

reverse transcriptase PCR (RT-PCR) PCR technique that involves converting RNA to DNA by reverse transcriptase

ribonuclease enzyme that breaks down RNA

sequence mapping mapping information obtained after DNA sequencing

shotgun sequencing method used to sequence multiple DNA fragments to generate the sequence of a large piece of DNA

single nucleotide polymorphism (SNP) variation between individuals in a single nucleotide

Southern blotting transfer of DNA from a gel to a nylon membrane

systems biology study of whole biological systems (genomes and proteomes) based on interactions within the system

Ti plasmid plasmid system derived from *Agrobacterium tumifaciens* that has been used by scientists to introduce foreign DNA into plant cells

transgenic organism that receives DNA from a different species

variable number of tandem repeats (VNTRs) variation in the number of tandem repeats between individuals in the population

whole-genome sequencing process that determines the DNA sequence of an entire genome

CHAPTER SUMMARY

17.1 Biotechnology

Nucleic acids can be isolated from cells for the purposes of further analysis by breaking open the cells and enzymatically destroying all other major macromolecules. Fragmented or whole chromosomes can be separated on the basis of size by gel electrophoresis. Short stretches of DNA or RNA can be amplified by PCR. Southern and northern blotting can be used to detect the presence of specific short sequences in a DNA or RNA sample. The term “cloning” may refer to cloning small DNA fragments (molecular cloning), cloning cell populations (cellular cloning), or cloning entire organisms (reproductive cloning). Genetic testing is performed to identify disease-causing genes, and gene therapy is used to cure an inheritable disease.

Transgenic organisms possess DNA from a different species, usually generated by molecular cloning techniques. Vaccines, antibiotics, and hormones are examples of products obtained by recombinant DNA technology. Transgenic plants are usually created to improve characteristics of crop plants.

17.2 Mapping Genomes

Genome mapping is similar to solving a big, complicated puzzle with pieces of information coming from laboratories all over the world. Genetic maps provide an outline for the location of genes within a genome, and they estimate the distance between genes and genetic markers on the basis of recombination frequencies during meiosis. Physical maps provide detailed information about the physical distance between the genes. The most detailed information is available through sequence mapping. Information from all mapping and sequencing sources is combined to study an entire genome.

17.3 Whole-Genome Sequencing

Whole-genome sequencing is the latest available resource to treat genetic diseases. Some doctors are using whole-genome sequencing to save lives. Genomics has many industrial applications including biofuel development, agriculture, pharmaceuticals, and pollution control. The basic principle of all modern-day sequencing strategies involves the chain termination method of sequencing.

Although the human genome sequences provide key insights to medical professionals, researchers use whole-genome sequences of model organisms to better understand the genome of the species. Automation and the decreased cost of whole-genome sequencing may lead to personalized medicine in the future.

17.4 Applying Genomics

Imagination is the only barrier to the applicability of genomics. Genomics is being applied to most fields of biology; it is being used for personalized medicine, prediction of disease risks at an individual level, the study of drug interactions before the conduct of clinical trials, and the study of microorganisms in the environment as opposed to the laboratory. It is also being applied to developments such as the generation of new biofuels, genealogical assessment using mitochondria,

advances in forensic science, and improvements in agriculture.

17.5 Genomics and Proteomics

Proteomics is the study of the entire set of proteins expressed by a given type of cell under certain environmental conditions. In a multicellular organism, different cell types will have different proteomes, and these will vary with changes in the environment. Unlike a genome, a proteome is dynamic and in constant flux, which makes it both more complicated and more useful than the knowledge of genomes alone.

Proteomics approaches rely on protein analysis; these techniques are constantly being upgraded. Proteomics has been used to study different types of cancer. Different biomarkers and protein signatures are being used to analyze each type of cancer. The future goal is to have a personalized treatment plan for each individual.

REVIEW QUESTIONS

1. How are GMOs created?

a. introducing recombinant DNA into an organism by any means

b. in vitro fertilization methods

c. mutagenesis

d. plant breeding techniques

2. Which technique used to manipulate genetic material results in a significant increase in DNA or RNA fragments?

a. gel electrophoresis

b. nucleic acid extraction

c. nuclear hybridization

d. polymerase chain reaction (PCR)

3. What is the role of the plasmid in molecular cloning?

a. They are used to create clones.

b. They are used as vectors to insert genes into bacteria.

c. They are a functional part of binary fission.

d. They contain the circular chromosome of prokaryotic organisms.

4. What is meant by a recombinant DNA molecule?

a. chimeric molecules

b. bacteria transformed into another species

c. molecules that have been through the PCR process

d. the result of crossing over during cell reproduction

5. Bt toxin is considered to be what?

a. a gene for modifying insect DNA

b. an organic insecticide produced by bacteria

c. useful for humans to fight against insects

d. a recombinant protein

6. What is one trait of the Flavr Savr Tomato?

a. has a better shelf life

b. is not a variety of vine-ripened tomato in the supermarket

c. was not created to have better flavor

d. undergoes soft rot

7. What is the first step in isolating DNA?

a. generating genomic DNA fragments with restriction endonucleases

b. introducing recombinant DNA into an organism by any means

c. overexpressing proteins in *E. coli*

d. lysing the cells in the sample

8. What is genomics?

a. Genomics is the study of entire genomes, including the complete set of genes, their nucleotide sequence and organization, and their interactions within a species and with other species.

b. Genomics is the process of finding the locations of genes on each chromosome.

c. Genomics is an illustration that lists genes and their location on a chromosome.

d. Genomics is a genetic marker, a gene or sequence on a chromosome that co-segregates (shows genetic linkage) with a specific trait.

9. What is required in addition to a genetic linkage map to build a complete picture of the genome?

a. a genetic marker

b. a physical map

c. linkage analysis of chromosomes

d. plasmids

10. Genetic recombination occurs by which process?

a. crossing over
b. chromosome segregation
c. independent assortment
d. sister chromatids

11. Individual genetic maps in a given species are _____.

a. genetically similar
b. genetically identical
c. genetically dissimilar
d. not useful in species analysis

12. Information obtained by microscopic analysis of stained chromosomes is used in what procedure?

a. cytogenetic mapping
b. radiation hybrid mapping
c. RFLP mapping
d. sequence mapping

13. Which of the following is true about linkage analysis?

a. It is used to create a physical map.
b. It is based on the natural recombination process.
c. It involves the breaking and re-joining of DNA artificially.
d. It requires radiation hybrid mapping.

14. The chain termination method of sequencing uses what?

a. labeled ddNTPs
b. only dideoxynucleotides
c. only deoxynucleotides
d. labeled dNTPs

15. What sequencing technique is used to identify regions of similarity between cell types or species?

a. dideoxy chain termination
b. proteins, DNA, or RNA sequence alignment
c. shotgun sequencing
d. whole-exome sequencing

16. Whole-genome sequencing can be used for advances in what field?

a. bioinformatics
b. iron industry
c. multimedia
d. the medical field

17. Sequencing an individual person's genome _____.

a. is currently impossible
b. helps in predicting faulty genes in diseases
c. will not lead to legal issues regarding discrimination and privacy
d. will not help make informed choices about medical treatment

18. Genomics can be used in agriculture to do what?

a. generate new hybrid strains
b. improve disease resistance
c. improve yield
d. improve yield and resistance and generate hybrids

19. What are the uses of metagenomics?

a. identification of biofuels
b. testing for multiple drug susceptibility in a population
c. use in increasing agricultural yields
d. identifying new species more rapidly and analyzing the effect of pollutants on the environment

20. Genomics can be used on a personal level to do what?

a. determine the risks of genetic diseases for an individual's children
b. increase transplant rejection
c. predict protein profile of a person
d. produce antibodies for an antigen

21. What is the percentage of single gene defects causing disease in developed countries?

a. 0.05
b. 0.1
c. 0.2
d. 0.4

22. The rapid identification of new species and the analysis of the effect of pollutants on the environment is a function of what?

a. metagenomics
b. linkage analysis
c. genomics
d. shotgun sequencing

23. The risks of genetic diseases for an individual's children can be determined through _____.

a. metagenomics
b. linkage analysis
c. genomics
d. shotgun sequencing

24. What is a biomarker?

a. the color coding of different genes
b. a protein uniquely produced in a diseased state
c. a molecule in the genome or proteome
d. a marker that is genetically inherited

25. What is a metabolome?

a. a provisional listing of the genome of a species
b. a unique metabolite used to identify an individual
c. a method used for protein analysis
d. the complete set of metabolites related to the genetic makeup of an organism

26. How would you describe a set of proteins with altered expression levels?

a. a group of biomarkers
b. a protein signature
c. the result of a defect in mRNA transcription
d. the results of crossing over during cell replication

27. What is a protein signature?

a. a protein expressed on the cell surface
b. a unique set of proteins present in a diseased state
c. the path followed by a protein after it is synthesized in the nucleus
d. the path followed by a protein in the cytoplasm

28. What describes a protein that is uniquely produced in a diseased state?

a. a genomic protein
b. a genetic defect
c. a chimeric molecule
d. a biomarker

29. The metabolites that results from the anabolic and catabolic reactions of an organisms is called what?

a. genetic metabolic profile
b. metabolic signature
c. metabolome
d. metagenomics

CRITICAL THINKING QUESTIONS

30. Describe the process of Southern blotting.

a. Southern Blotting is used to find DNA sequences. Fragments are separated on gel, incubated with probes to check for the sequence of interest, and transferred to a nylon membrane.
b. Southern blotting is used to find DNA sequences. Fragments are separated on gel, transferred to a nylon membrane, and incubated with probes to check for the sequence of interest.
c. When RNA is used, the process is called Northern blotting.
d. Southern blotting is used to find RNA sequences. Fragments are separated on gel, incubated with probes to check for the sequence of interest, and transferred to a nylon membrane.

31. A researcher wants to study cancer cells from a patient with breast cancer. Is cloning the cancer cells an option?

a. The cancer cells should be cloned along with a biomarker for better detection and study.
b. The cells should be screened first in order to assure their carcinogenic nature.
c. The cancer cells, being clones of each other already, should directly be grown in a culture media and then studied.
d. The cancer cells should be extracted using the specific antibodies.

32. Discuss the uses of genome mapping.

a. Genome mapping is useful in identifying human disease-causing genes, developing microbes to clean up pollutants, and increasing crop yield.
b. Genome mapping is directly required to produce recombinants, in FISH detection, and detecting the methylated parts of genetic material.
c. Genome mapping is useful for knowing the pedigree of diseases in humans and tracing the movement of transposons in plants.
d. Genome mapping identifies human disease-causing genes only.

33. If you had a chance to get your genome sequenced, what are some questions you might be able to have answered about yourself?

a. One can determine the drugs that can rectify a disease, symptoms of the disease and its severity.

b. One can determine the ancestry and genetic origin of diseases and their susceptibility to drugs.

c. One can predict the symptoms of a disease, the vectors to be used in gene therapy and the causal organism of the disease.

d. One can determine the pedigree of a disease, produce recombinants and detect the presence of extracellular genes using FISH.

34. Describe an example of a genomic mapping method

a. The radiation mapping method is an example which uses radiation to break the DNA and is affected by changes in recombination frequency.

b. Cytogenetic mapping obtains information from microscopic analysis of stained chromosomes. It can estimate the approximate distance between markers.

c. In restriction mapping, the DNA fragments are cut by using the restriction enzymes and then stained fragments are viewed on gel.

d. Cytogenetic mapping obtains information from microscopic analysis of stained chromosomes. It can estimate the exact base pair distance between markers.

35. Describe three methods of gene sequencing.

a. Chain termination method - automated sequencers are used to generate sequences of short fragments; Shotgun sequencing method - incorporation of ddNTP during DNA replication; Next-generation sequencing - cutting DNA into random fragments, sequencing using chain termination, and assembling overlapping sequences

b. Chain termination method - incorporation of ddNTP during DNA replication; Shotgun sequencing method - cutting DNA into random fragments, sequencing using chain termination, and assembling overlapping sequences; Next-generation sequencing - automated sequencers are used to generate sequences of short fragments

c. Chain termination method - incorporation of ddNTP during DNA replication; Shotgun sequencing method - automated sequencers are used to generate sequences of short fragments; Next-generation sequencing - cutting DNA into random fragments, sequencing using chain termination, and assembling overlapping sequences

d. Chain termination method - automated sequencers are used to generate sequences of short fragments; Shotgun sequencing method - cutting DNA into random fragments, sequencing using chain termination, and assembling overlapping sequences; Next-generation sequencing - incorporation of ddNTP during DNA replication

36. What is the greatest challenge facing genome sequencing?

a. the lack of resources and use of chemicals for the sequencing of the DNA fragments

b. the ethical issues such as discrimination based on person’s genetics

c. the use of chemicals during the sequencing methods that could incorporate mutations

d. the scientific issues, like conserving the human genome sequences

37. How is shotgun sequencing performed?

a. The DNA is cut into fragments, sequencing is done using the chain termination method, fragments are analyzed to see the overlapping sequences, and the entire fragment is reformed.

b. The DNA is cut into fragments, overlapping sequences are analyzed using a computer, sequencing is done using the chain termination method, and the DNA fragment is reformed.

c. The DNA is cut into fragments, stained with fluorescent dye, sequenced using the chain termination method, fragments are analyzed to see the overlapping sequences, and the entire DNA fragment is reformed.

d. The DNA is cut into fragments, sequencing is done using the chain termination method, the DNA is stained with fluorescent dye, and a computer is used to analyze and reform the entire DNA fragment.

38. Coumadin is a drug frequently given to prevent excessive blood clotting in stroke or heart attack patients, which could lead to another stroke or heart attack. Administration of the drug also can result in an overdose in some patients, depending on the liver function of a patient. How could pharmacogenomics be used to assist these patients?

a. Pharmacogenomics will be able to provide a counter-acting drug to decrease the effect of Coumadin.

b. Pharmacogenomics will test every patient for their sensitivity to the drug.

c. Pharmacogenomics will not be able to provide any help to patients highly sensitive to the drug.

d. Pharmacogenomics will provide an overdose to each patient to test for the symptoms of the drug.

39. Why is so much effort being poured into genome mapping applications?

a. Genome mapping is necessary to know the base pair difference between the markers.

b. The mapping would help scientists understand the role of proteins in specific organelles.

c. The mapping technique identifies the role of transposons.

d. Genome mapping helps identify faulty alleles, which could cause diseases.

40. What is the reason for studying mitochondrial genomics that is most directly important for humans?

a. Mitochondria evolved from bacteria; therefore, their genome is important to study.

b. Mitochondria undergo rapid mutation and it is essential that this pattern be studied.

c. Mitochondria contain DNA, and it is passed on from mother to offspring, which renders it helpful in tracing genealogy.

d. Mitochondria are the only ATP-producing organelles of the cell, thus their genome is important.

41. How can proteomics complement genomics?

a. The genes are responsible to produce proteins and this implies that proteomics complements genomics.

b. Genomics is responsible to decide the structure of the proteins, and, thereby, the result of proteomic studies.

c. The genome is constant but the proteome is dynamic as different tissues possess the same genes but express different genes, thereby complementing genomics.

d. The study of genes is incomplete without the study of their respective proteins and thus they complement each other.

42. How could a proteomic map of the human genome help find a cure for cancer?

a. A genetic map could help in identifying genes that could counteract the cause of cancer.

b. Metabolomics can be used to study the genes producing metabolites during cancer.

c. Proteomics detects biomarkers whose expression is affected by the disease process.

d. The mapping helps in analyzing the inheritance of cancer-causing genes.

43. What contributions have been made through the use of microbial genomics?

a. Microbial genomics has provided various tools to study the psychological behaviors of organisms.

b. Microbial genomics has been useful in producing antibiotics, enzymes, improved vaccines, disease treatments and advanced cleanup techniques.

c. Microbial genomics has contributed resistance in other bacteria by horizontal and lateral gene transfer mechanisms.

d. Microbial genomics has contributed to fighting global warming.

TEST PREP FOR AP® COURSES

44. In separating DNA for genomic analysis, it is important to consider RNA contaminating the sample during the cell lysis step of a DNA extraction. Which is likely to cause what to occur?

a. DNA separates into the supernatant.

b. DNA is destroyed by the protease.

c. DNA is unaffected by the RNA.

d. DNA combines with the RNA.

45. There are many techniques for investigating human genomic disorders. Western blotting looks for protein, Eastern blotting looks for post-translational changes, Northern blotting looks at mRNA, and Southern blotting looks at DNA. If you were to look at sickle cell anemia, a disorder affecting hemoglobin produced in red blood cells, which technique would be the most useful in detecting polymorphism in a sample?

a. Northern blotting

b. Southern blotting

c. Western blotting

d. Eastern blotting

46. A population of insects were formally distinguished by a mix of colors on their thorax and legs. This population now appears to be split into 2 sub-groups, purple and orange-legged. Researchers hypothesize that the purple-legged group may be increasingly resistant to the Bt (*Bacillus thuringiensis*) toxin. Which idea supports this observation?

a. transgenesis

b. natural selection

c. hybridization

d. recombination

47. Describe the process of molecular cloning.

a. The foreign DNA and plasmid are cut with the same restriction enzyme and DNA is inserted within the lacZ gene, whose product metabolizes lactose. The foreign DNA and vector are allowed to anneal. The vector is transferred to a bacterial host that is ampicillin sensitive and those with a disrupted lacZ gene show inability to metabolize X-gal.

b. The foreign DNA and plasmid are denatured using high heat, and DNA is inserted within the lacZ gene, whose product metabolizes glucose. The foreign DNA and vector are allowed to anneal. The vector is transferred to a bacterial host that is ampicillin sensitive and disrupted lacZ gene will metabolize X-gal

c. The foreign DNA and plasmid are cut with the same restriction enzyme and DNA is inserted randomly in the plasmid. The foreign DNA and vector are allowed to anneal. The vector is transferred to a bacterial host that is ampicillin sensitive and the disrupted lacZ gene shows inability to synthesize X-gal.

d. The foreign DNA and plasmid are cut with the same restriction enzyme and DNA is inserted within the lacZ gene, whose product metabolizes lactose. The foreign DNA and vector are allowed to anneal. The vector is transformed into a viral host that is ampicillin sensitive and the disrupted lacZ gene show inability to synthesize X-gal.

48. There are three methods of creating maps to evaluate genomes: cytogenetic (staining chromosomes); radiation hybrid maps (fragments with x-rays); and sequence maps (comparing DNA sequences). Which of the following accurately describes the three methods?

a. Cytogenetic mapping - stained sections of chromosomes are analyzed using microscope, the distance between genetic markers can be found; Radiation hybrid mapping - breaks DNA using radiation and is affected by recombination frequency; Sequence mapping - DNA sequencing technology used to create physical maps.

b. Cytogenetic mapping - stained sections of chromosomes are analyzed using microscope, the approximate distance between genetic markers can be found; Radiation hybrid mapping - breaks DNA using radiation and is unaffected by recombination frequency; Sequence mapping - DNA sequencing technology used to create physical maps.

c. Cytogenetic mapping - stained sections of chromosomes are analyzed using microscope, the distance in base pairs between genetic markers can be found; Radiation hybrid mapping - breaks DNA using radiation and is unaffected by recombination frequency; Sequence mapping - DNA sequencing technology used to create physical maps.

d. Cytogenetic mapping - stained sections of chromosomes are analyzed using a telescope, the distance between genetic markers can be found; Radiation hybrid mapping - breaks DNA using radiation and is affected by recombination frequency; Sequence mapping - DNA sequencing technology used to create physical maps.

49. How many cells with different genetic variations are possible after a single round of meiosis?

a. two

b. three

c. four

d. eight

SCIENCE PRACTICE CHALLENGE QUESTIONS

50. Prokaryotes have an adaptive strategy to identify and respond to viral infections. This strategy uses segments of the cyclic DNA called CRISPRs and genes encoding CRISPR-associated (cas) proteins. When a virus enters the cell, a strand of viral DNA is excised by a cas protein and inserted into the bacterial DNA in a CRISPR region. When the same viral DNA is encountered subsequently, this foreign DNA is targeted by cas proteins that carry RNA markers transcribed from the inserted segment. The cas proteins cleave the viral DNA. The bacteria “remember” the infectious agent, providing a form of immunity.

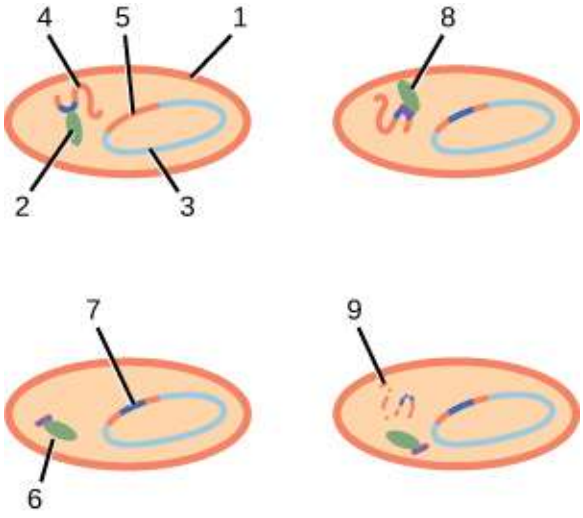


Figure 17.17

A. Use the diagram above to identify the components of a transcript-based response of bacteria to the presence of

viral DNA by placing the corresponding number next to each feature of the diagram:

___ viral DNA ___ degraded viral DNA ___ cell membrane

___ cellular DNA ___ cas protein ___ stored viral DNA template

___ excised viral DNA ___ cas protein-RNA complex

___ cas protein-RNA-viral DNA complex

The CRISPR system was discovered in cultures of yogurt in 2002. Subsequently, researchers developed a technology based on the manipulation of this system. The code for the prokaryotic CRISPR/cas system is highly conserved and is found in the human genome. DNA sequences are known to encode proteins responsible for many heritable diseases. CRISPR/cas is a technology that allows DNA to be cleaved at the boundaries of a nucleotide sequence, making the protein dysfunctional. The break in the strand is then recognized and replaced with the code for the functional protein. If the editing is done with zygote-forming cells, the change is inherited. Not only the patient, but all progeny of the patient, are cured. This technology is the first to easily make genomic modifications of a germ line. In the words of a prominent molecular biologist, this technology, which was recognized as the *Breakthrough of 2015* in the journal *Science*, "democratizes genetic engineering." Just as PCR became a standard, widely used tool, any molecular biology lab is now able to apply this technology.

B. **Pose three questions**—whose pursuit would require an understanding of genetics—regarding the ethical and social issues that accompany the use of this medical technology.

C. **Explain** the value of genetic variation within a population. **Predict** a possible effect that this technology could have, if unregulated, on human genetic variation.

51. Gel electrophoresis of polymers and polymer fragments is an important element in many investigations. Samples of a solution are pipetted into the wells of a gel. The gel is placed in a solution that maintains a constant pH, and an electric field is applied over the length of the gel. Separated components are transferred to a substrate where they can be visualized and identified by comparison with samples of standards. Application of this method to DNA is called a Southern blot, named for the inventor of the technology. The method's application to RNA is called a northern blot, another demonstration that biologist have fun (there are also western, eastern, and far-eastern blots, but these techniques are *not* named for their inventors).

A. Consider the three amino acids shown below and **explain** how, when pipetted into a gel and subjected to an electric field, the amino acids move; how the amino acids are separated as they move; and which amino acid moves furthest.

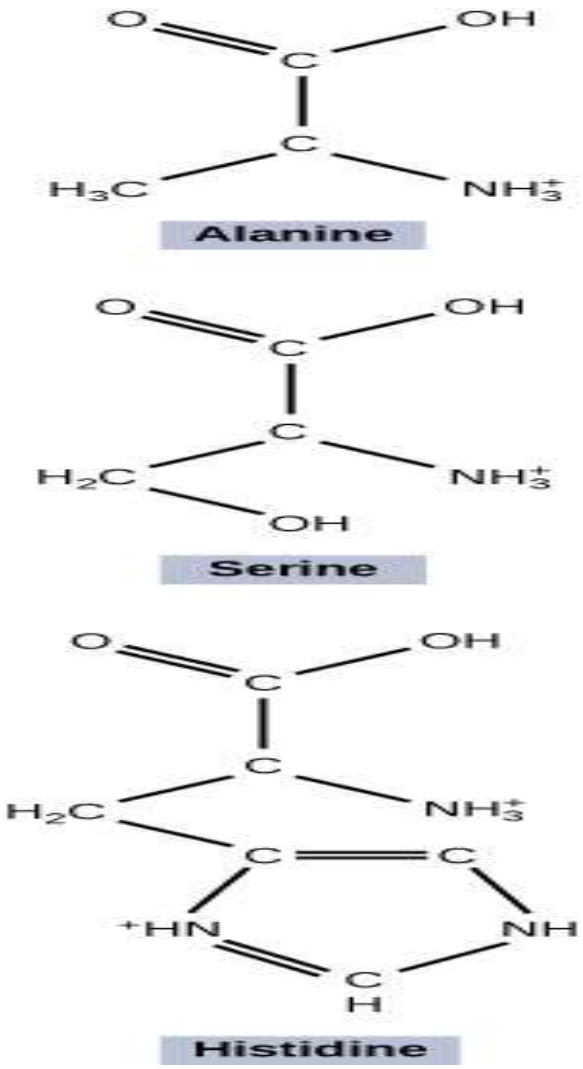


Figure 17.18

B. A biologist wants to determine whether a new protocol is successful in constructing and amplifying a molecular clone of a segment of DNA introduced as a plasmid. After the procedure is complete, the bacterial cells containing the plasmid with the inserted segment are lysed, and a gel is run into which samples of the lysate and the sequences to be cloned have been pipetted. Use the data displayed in the developed gel shown below to **evaluate the question** of whether the protocol was successful.

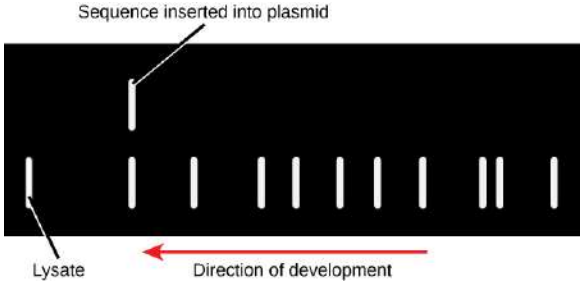


Figure 17.19

C. **Design a plan** to answer the question of whether the new DNA has been incorporated into the DNA of the host organism.

52. Genetic engineering can be applied to heritable information to produce what is referred to as a "knockdown organism." Biotechnology also can be applied to produce nonheritable changes in a "knockdown gene." Post-transcriptional strategies target the mRNA product of a gene. One such strategy uses the conserved genes that encode RNA interference (RNAi) proteins for the regulation of levels of mRNA transcription.

Some viral RNA is double stranded (dsRNA). A cell responds to the presence of double-stranded RNA by the attachment of the enzyme DICER, which cuts dsRNA into short fragments. One strand of the fragment is transferred to the RNA-induced silencing complex (RISC), which searches for an mRNA with a sequence matching that of the fragment strand. When detected, this mRNA is degraded.

A. Common in cancer cells is a mutation of the gene that encodes the protein p53, whose role is to detect and repair errors in DNA; if repairs cannot be made, p53 initiates apoptosis. **Create a visual representation** to **explain** how the DICER-RISC system within the cell can be used to suppress the translation of a mutated form of the gene encoding p53, potentially destroying a tumor.

B. Whole-genome sequences provide a library of potentially expressed proteins, but they do not provide information on the functions of each protein. In an approach called reverse genetics, investigations attempt to determine the function of the gene, often by silencing the gene using RNAi technology. Assume that you have the ability to synthesize dsRNA from a DNA segment taken from an organism whose whole genome has been determined. **Design a plan** for collecting data that could be used to assign a function to the protein encoded by this sequence. (Hint: Don't worry about the number of experiments that might need to be conducted to implement your plan. An automated technique called high-throughput screening robotically supports thousands of simultaneous experiments.)