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

Genome Sequence Uses

DNA microarrays are methods that scientists use to detect gene expression by analyzing different DNA fragments that are fixed to a glass slide or a silicon chip to identify active genes and sequences. We can discover almost one million genotypic abnormalities using microarrays; whereas, whole-genome sequencing can provide information about all six billion base pairs in the human genome. Although studying genome sequencing medical applications is interesting, this discipline dwells on abnormal gene function. Knowing about the entire genome will allow researchers to discover future onset diseases and other genetic disorders early. This will allow for more informed decisions 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 developing novel enzymes that convert biomass to biofuel, which results in higher crop and fuel production, and lower consumer cost. This knowledge should allow better methods of control over the microbes that industry uses to produce biofuels. Genomics could also improve monitoring methods that measure the impact of pollutants on ecosystems and help clean up environmental contaminants. Genomics has aided in developing 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

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

- Explain pharmacogenomics
- Define polygenic

Introducing DNA sequencing and whole genome sequencing projects, particularly the Human Genome project, has expanded the applicability of DNA sequence information. Many fields, such as metagenomics, pharmacogenomics, and mitochondrial genomics are using genomics. Understanding and finding cures for diseases is the most common application of genomics.

Predicting Disease Risk at the Individual Level

Predicting disease risk involves screening currently healthy individuals by genome analysis at the individual level. Health care professionals can recommend intervention with lifestyle changes and drugs 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. The medical team performed a risk assessment to analyze Quake’s percentage of risk for 55 different medical conditions. The team found a rare genetic mutation, which showed him to be at risk for sudden heart attack. The results also predicted that Quake had 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, researchers still must address ethical issues surrounding genomic analysis at a population level.

visual CONNECTION

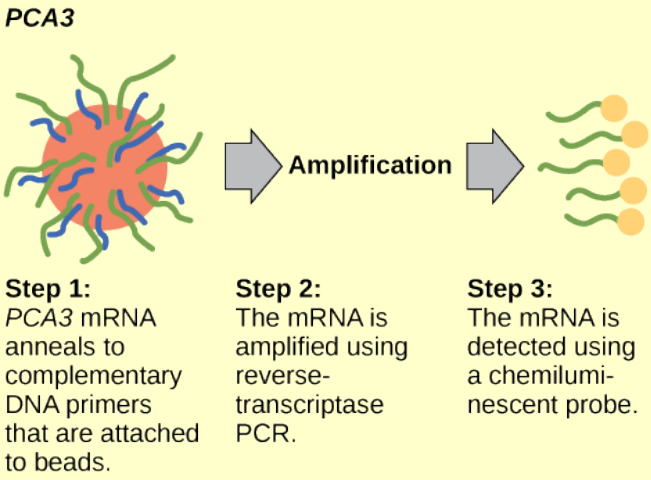

Figure 17.15 *PCA3* is a gene that is expressed in prostate epithelial cells and overexpressed in cancerous cells. A high *PCA3* concentration in urine is indicative of prostate cancer. The *PCA3* test is a better indicator of cancer than the more well known PSA test, which measures the level of PSA (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 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 receive prostate cancer screenings using the *PCA3* or PSA test? Should people in general receive screenings to find out if they have a genetic risk for cancer or other diseases?

Pharmacogenomics and Toxicogenomics

Pharmacogenomics, or toxicogenomics, involves evaluating drug effectiveness and safety on the basis of information from an individual's genomic sequence. We can study genomic responses to drugs 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 drug's presence, which we can use 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 cancerous cell growth. Genome-wide studies can also help to find new genes involved in drug toxicity. Medical professionals can use personal genome sequence information 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 medical professionals can test them further before pathologic symptoms arise.

Microbial Genomics: Metagenomics

Traditionally, scholars have taught microbiology with the view that it is best to study microorganisms under **pure culture** conditions. This involves isolating a single cell type 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 culturing in isolation. Most microorganisms do not live as isolated entities, but in microbial communities or 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 17.16).

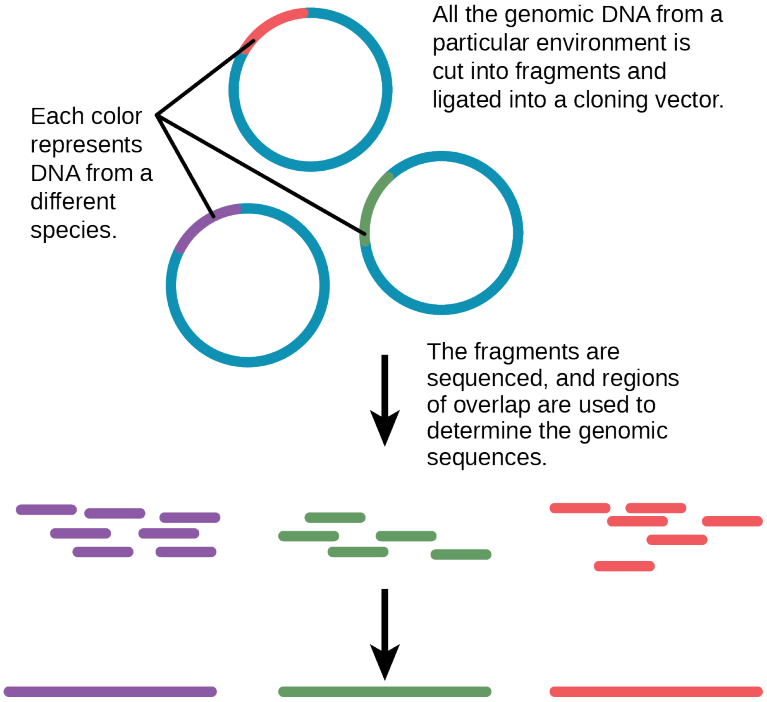

Figure 17.16 Metagenomics involves isolating DNA from multiple species within an environmental niche.

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 antimicrobial 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 scientists often use it to study evolutionary relationships. Another feature that makes studying the mitochondrial genome interesting is that the mitochondrial DNA in most multicellular organisms passes from the mother during the fertilization process. For this reason, scientists often use mitochondrial genomics to trace genealogy.

Experts have used information and clues from DNA samples at crime scenes as evidence in court cases, and they have used genetic markers in forensic analysis. Genomic analysis has also become useful in this field. The first publication showcasing the first use of genomics in forensics came out in 2001. 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 the culprit used a specific anthrax strain 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 agricultural crop yield quality and quantity. Linking traits to genes or gene signatures helps 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. Researchers are discovering how genomics can improve agricultural production's quality and quantity. 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

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

- Explain systems biology
- Describe a proteome
- Define protein signature

Proteins are the final products of genes, which help perform the function that the gene encodes. Amino acids comprise proteins 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 that a cell type produces. We can study proteoms 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. **Proteomics** is the study of proteomes' function. Proteomics complements genomics and is useful when scientists want to test their hypotheses that they based on genes. Even though all multicellular organisms' cells 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 modify themselves after translation by processes such as proteolytic cleavage, phosphorylation, glycosylation, and ubiquitination. There are also protein-protein interactions, which complicate studying 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 studying small molecule metabolites in an organism. The **metabolome** is the complete set of metabolites that are related to an organism's genetic makeup. 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 the metabolites in living organisms' tissues and fluids.

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 scientists analyze the genome using the basic DNA sequencing technique, proteomics requires techniques for protein analysis. The basic technique for protein analysis, analogous to DNA sequencing, is mass spectrometry. Mass spectrometry identifies and determines a molecule's characteristics. Advances in spectrometry have allowed researchers to analyze very small protein samples. X-ray crystallography, for example, enables scientists to determine a protein crystal's three-dimensional structure at atomic resolution. Another protein imaging technique, nuclear magnetic resonance (NMR), uses atoms' magnetic properties to determine the protein's three-dimensional structure in aqueous solution. Scientists have also used protein microarrays to study protein interactions. Large-scale adaptations of the basic two-hybrid screen (**Figure 17.17**) have provided the basis for protein microarrays. Scientists use computer software to analyze the vast amount of data for proteomic analysis.

Genomic- and proteomic-scale analyses are part of **systems biology**, which 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.