DNA Technology and Genomics

I. DNA cloning permits production of many copies of a specific gene or other DNA segment.
   A. To study a particular gene, scientists needed to develop methods to isolate the small, well-defined portion of a chromosome containing the gene of interest. Techniques for **gene cloning** enable scientists to prepare identical copies of gene-sized pieces of DNA.
   B. One basic cloning technique begins with the insertion of a foreign gene into a bacterial plasmid.
      1. First, a foreign gene is inserted into a bacterial plasmid to produce a recombinant DNA molecule.
      2. The plasmid is returned to a bacterial cell, producing a *recombinant bacterium,* which reproduces to form a **clone** of identical cells. Every time the bacterium reproduces, the recombinant plasmid is replicated as well.
      3. The bacterial clone will make the protein encoded by the foreign gene. The potential uses of cloned genes fall into two general categories.
         a. To produce a protein product. For example, bacteria carrying the gene for human growth hormone can produce large quantities of the hormone.
         b. To prepare many copies of the gene itself so that the gene's nucleotide sequence can be learned or to provide an organism with a new metabolic capability by transferring a gene from another organism.

II. Gene cloning and genetic engineering were made possible by the discovery of **restriction enzymes** that cut DNA molecules at specific locations.
   A. In nature, bacteria use restriction enzymes to cut foreign DNA, to protect themselves against phages or other bacteria. They work by cutting up the foreign DNA, a process called *restriction.*
   B. Restriction enzymes are very specific, recognizing short DNA nucleotide sequences and cutting at specific points in these sequences. Bacteria protect their own DNA by methylating the sequences recognized by these enzymes.
   C. Each restriction enzyme cleaves a specific sequence of bases or **restriction site.** These are often a symmetrical series of four to eight bases on both strands running in opposite directions.
   D. Because the target sequence usually occurs (by chance) many times on a long DNA molecule, an enzyme will make many cuts. Copies of a DNA molecule will always yield the same set of **restriction fragments** when exposed to a specific enzyme.
   E. Restriction enzymes cut covalent sugar-phosphate backbones of both strands, often in a staggered way that creates single-stranded **sticky ends.**
      1. These extensions can form hydrogen-bonded base pairs with complementary single-stranded stretches (sticky ends) on other DNA molecules cut with the same restriction enzyme.
      2. These DNA fusions can be made permanent by **DNA ligase,** which seals the strand by catalyzing the formation of covalent bonds to close up the sugar-phosphate backbone.
      3. Restriction enzymes and DNA ligase can be used to make a stable recombinant DNA molecule, with DNA that has been spliced together from two different organisms.

III. Eukaryotic genes can be cloned in bacterial plasmids.
   A. Recombinant plasmids are produced by splicing restriction fragments from foreign DNA into plasmids.
   B. The original plasmid used to produce recombinant DNA is called a **cloning vector,** defined as a DNA molecule that can carry foreign DNA into a cell and replicate there.

C. The process of cloning a human gene in a bacterial plasmid can be divided into six steps.

1. The first step is the isolation of vector and source DNA.
    a. The source DNA comes from human tissue cells grown in lab culture. The source of the plasmid is typically *E. coli*.
    b. This plasmid carries two useful genes, *amp*^{R}, conferring resistance to the antibiotic ampicillin and *lacZ*, encoding the enzyme ß-galactosidase that catalyzes the hydrolysis of sugar.
    c. The plasmid has a single recognition sequence, within the *lacZ* gene, for the restriction enzyme used.
2. DNA is inserted into the vector.
    a. Both the plasmid and human DNA are digested with the same restriction enzyme. The enzyme cuts the plasmid DNA at its single restriction site within the *lacZ* gene. It cuts the human DNA at many sites, generating thousands of fragments. One fragment carries the human gene of interest. All the fragments, bacterial and human, have complementary sticky ends.
3. The human DNA fragments are mixed with the cut plasmids, and base-pairing takes place between complementary sticky ends. DNA ligase is added to permanently join the base-paired fragments. Some of the resulting recombinant plasmids contain human DNA fragments.
4. The recombinant plasmids are mixed with bacteria that are lacZ−, unable to hydrolyze lactose. This results in some bacteria that have taken up the desired recombinant plasmid DNA, and other bacteria that have not.
5. The transformed bacteria are plated on a solid nutrient medium containing ampicillin and a molecular mimic of lactose called X-gal.
    a. Only bacteria that have the ampicillin-resistance (*amp*^{R}) plasmid will grow. Each reproducing bacterium forms a *colony* of cells on the agar.
    b. The lactose mimic in the medium is used to identify plasmids that carry foreign DNA.
        (1) Bacteria with plasmids lacking foreign DNA stain blue when ß-galactosidase from the intact *lacZ* gene hydrolyzes X-gal.
        (2) Bacteria with plasmids containing foreign DNA inserted into the *lacZ* gene are white because they lack ß-galactosidase.
6. In the final step, thousands of bacterial colonies with foreign DNA must be sorted through to find those containing the gene of interest.
    a. One technique, **nucleic acid hybridization,** depends on base-pairing between the gene and a complementary sequence, a **nucleic acid probe,** on another nucleic acid molecule.
        (1) A radioactive or fluorescent tag is used to label the probe.
        (2) The probe will hydrogen-bond specifically to complementary single strands of the desired gene.
        (3) After **denaturating** (separating) the DNA strands in the bacterium, the probe will bind with its complementary sequence, tagging colonies with the targeted gene.

IV. Cloned genes are stored in DNA libraries.

A. A complete set of recombinant plasmid clones, each carrying copies of a particular segment from the initial genome, forms a **genomic library.**

B. The library can be saved and used as a source of other genes or for gene mapping.

1. In addition to plasmids, certain bacteriophages are also common cloning vectors for making genomic libraries.
2. Fragments of foreign DNA can be spliced into a phage genome using a restriction

enzyme and DNA ligase. An advantage of using phage as vectors is that phage can carry larger DNA inserts than plasmids can.

3. The recombinant phage DNA is packaged in a capsid *in vitro* and allowed to infect a bacterial cell.
4. Infected bacteria produce new phage particles, each with the foreign DNA.

C. A more limited kind of gene library can be developed by starting with mRNA extracted from cells.

1. The enzyme reverse transcriptase is used to make single-stranded DNA transcripts of the mRNA molecules.
2. The mRNA is enzymatically digested, and a second DNA strand complementary to the first is synthesized by DNA polymerase.
3. This double-stranded DNA, called **complementary DNA (cDNA),** is modified by the addition of restriction sites at each end.
4. Finally, the cDNA is inserted into vector DNA.
5. A **cDNA library** represents that part of a cell's genome that was transcribed in the starting cells.
6. This is an advantage if a researcher wants to study the genes responsible for specialized functions of a particular kind of cell.
7. By making cDNA libraries from cells of the same type at different times in the life of an organism, one can trace changes in the patterns of gene expression.

V. Eukaryote genes can be expressed in prokaryotic host cells.

A. A cloned eukaryotic gene can be made to function in a prokaryotic host by inserting an **expression vector,** a cloning vector containing a highly active prokaryotic promoter. The prokaryotic host will then recognize the promoter and proceed to express the foreign gene that has been linked to it.

B. The presence of long noncoding introns in eukaryotic genes may prevent correct expression of these genes in prokaryotes, which lack RNA-splicing machinery.

C. This problem can be avoided by using a cDNA form of the gene.

D. Molecular biologists can avoid incompatibility problems by using eukaryotic cells as hosts for cloning and expressing eukaryotic genes.

1. Yeast cells are as easy to grow as bacteria and, unlike most eukaryotes, have plasmids.
2. Scientists have constructed **yeast artificial chromosomes (YACs)** that combine the essentials of a eukaryotic chromosome (an origin site for replication, a centromere, and two telomeres) with foreign DNA.
3. These chromosome-like vectors behave normally in mitosis and can carry more DNA than a plasmid.
4. Another advantage of eukaryotic hosts is that they are capable of providing the posttranslational modifications that many proteins require.

VI. The polymerase chain reaction (PCR) amplifies DNA *in vitro*.

A. DNA cloning is the best method for preparing large quantities of a particular gene or other DNA sequence.

B. When the source of DNA is small or impure, the **polymerase chain reaction (PCR)** is quicker and more selective.

C. This technique can quickly amplify any piece of DNA without using cells.

1. The DNA is incubated in a test tube with special DNA polymerase, a supply of nucleotides, and short pieces of single-stranded DNA as a primer.
2. PCR can make billions of copies of a targeted DNA segment in a few hours. This is faster than cloning via recombinant bacteria.
3. In PCR, a three-step cycle—heating, cooling, and replication—brings about a chain

reaction that produces an exponentially growing population of identical DNA molecules.

a. The reaction mixture is heated to denature the DNA strands.

b. The mixture is cooled to allow hydrogen-bonding of short, single-stranded DNA primers complementary to sequences on opposite sides at each end of the target sequence.

c. A heat-stable DNA polymerase extends the primers in the 5’ C 3’ direction. The DNA polymerase used was isolated from prokaryotes living in hot springs.

4. By being complementary to sequences bracketing the targeted sequence, the primers determine the DNA sequence that is amplified.

5. PCR is so specific and powerful that only minute amounts of partially degraded DNA need be present in the starting material.

6. Occasional errors during PCR replication impose limits to the number of good copies that can be made when large amounts of a gene are needed.

7. Increasingly, PCR is used to make enough of a specific DNA fragment to clone it merely by inserting it into a vector.

VII. One indirect method of rapidly analyzing and comparing genomes is **gel electrophoresis.**

A. Gel electrophoresis separates nucleic acids or proteins on the basis of their rate of movement through a gel in an electrical field. Rate of movement depends mostly on size electrical charge of the macromolecules.

B. Restriction fragment analysis detects DNA differences that affect restriction sites. In restriction fragment analysis, the DNA fragments produced by restriction enzyme digestion of a DNA molecule are sorted by gel electrophoresis.

1. When the mixture of restriction fragments from a particular DNA molecule undergoes electrophoresis, it yields a band pattern characteristic of the starting molecule and the restriction enzyme used.

2. The separated fragments can be recovered undamaged from gels, providing pure samples of individual fragments.

C. We can use restriction fragment analysis to compare two different DNA molecules representing, for example, different alleles of a gene.

1. Because the two alleles differ slightly in DNA sequence, they may differ in one or more restriction sites.

2. If they do differ in restriction sites, each will produce different-sized fragments when digested by the same restriction enzyme.

3. In gel electrophoresis, the restriction fragments from the two alleles will produce different band patterns, allowing us to distinguish the two alleles.

4. Restriction fragment analysis is sensitive enough to distinguish between two alleles of a gene that differ by only one base pair in a restriction site.

D. A technique called **Southern blotting** combines gel electrophoresis with nucleic acid hybridization.

1. Although electrophoresis will yield too many bands to distinguish individually, we can use nucleic acid hybridization with a specific probe to label discrete bands that derive from our gene of interest.

2. The probe is a radioactive single-stranded DNA molecule that is complementary to the gene of interest.

3. Southern blotting reveals not only whether a particular sequence is present in the sample of DNA, but also the size of the restriction fragments that contain the sequence.

4. One of its many applications is to identify heterozygous carriers of mutant alleles associated with genetic disease.
5. In the example below, we compare genomic DNA samples from three individuals: an individual who is homozygous for the normal ß-globin allele, a homozygote for sickle-cell allele, and a heterozygote.
    a. We combine several molecular techniques to compare DNA samples from three individuals.
    b. We start by adding the same restriction enzyme to each of the three samples to produce restriction fragments.
    c. We then separate the fragments by gel electrophoresis.
    d. We transfer the DNA fragments from the gel to a sheet of nitrocellulose paper, still separated by size. This also denatures the DNA fragments.
    e. Bathing the sheet in a solution containing a radioactively labeled probe allows the probe to attach by base-pairing to the DNA sequence of interest. We can visualize bands containing the label with autoradiography.
    f. The band pattern for the heterozygous individual will be a combination of the patterns for the two homozygotes.

VIII. Restriction fragment length differences are useful as genetic markers and can be used to examine differences in *noncoding* DNA as well.
- A. Differences in DNA sequence on homologous chromosomes that produce different restriction fragment patterns are scattered abundantly throughout genomes, including the human genome.
- B. A **restriction fragment length polymorphism** (**RFLP**) can serve as a genetic marker for a particular location (locus) in the genome.
- C. RFLPs are detected and analyzed by Southern blotting, frequently using the entire genome as the DNA starting material.
- D. The probe is complementary to the sequence under consideration.

IX. **Genomics,** the study of genomes and their interactions, is yielding new insights into fundamental questions about genome organization, the regulation of gene expression, growth and development, and evolution.
- A. Starting with a long DNA sequence, how does a researcher recognize genes and determine their function?
    1. DNA sequences are collected in computer data banks that are available via the Internet to researchers everywhere.
    2. Special software scans the sequences for the telltale signs of protein-coding genes, looking for start and stop signals, RNA-splicing sites, and other features.
    3. The software also looks for *expressed sequence tags (ESTs),* sequences similar to those in known genes.
    4. From these clues, researchers collect a list of gene candidates.
- B. Genes account for only a small fraction of the human genome. Much of the enormous amount of noncoding DNA in the human genome consists of repetitive DNA and unusually long introns.
- C. The typical human gene specifies several different polypeptides by using different combinations of exons.
- D. Nearly all human genes contain several exons, and an estimated 75% of these multiexon genes are alternatively spliced. Along with this is additional polypeptide diversity via posttranslational processing. There are a much greater number of possible interactions between gene products as a result of greater polypeptide diversity.
- E. To determine what the others are and what they may do, scientists compare the sequences of new gene candidates with those of known genes.

1. In some cases, the sequence of a new gene candidate will be similar in part to that of a known gene, suggesting similar function.
2. In other cases, the new sequences will be similar to a sequence encountered before, but of unknown function.
3. In still other cases, the sequence is entirely unlike anything ever seen before.

F. One way scientists determine the function of new genes identified by genome sequencing is by disabling the gene and observing the consequences.
   1. Using ***in vitro*** **mutagenesis,** specific mutations are introduced into a cloned gene, altering or destroying its function.
   2. When the mutated gene is returned to the cell, it may be possible to determine the function of the normal gene by examining the phenotype of the mutant.
   3. Researchers may put a mutated gene into cells from the early embryo of an organism to study the role of the gene in development and functioning of the whole organisms.
   4. In nonmammalian organisms, a simpler and faster method, **RNA interference (RNAi),** has been applied to silence the expression of selected genes.
      a. This method uses synthetic double-stranded RNA molecules matching the sequences of a particular gene to trigger breakdown of the gene's mRNA.
      b. The RNAi technique has had limited success in mammalian cells but has been valuable in analyzing the functions of genes in nematodes and fruit flies.

G. A major goal of genomics is to learn how genes act together to produce a functioning organism.
   1. By looking for groups of genes that are expressed in a coordinated pattern and which genes are transcribed in certain cells, we can learn how genes interact
   2. This will reveal which genes are active at different developmental stages, in different tissues, or in tissues in different states of health.
   3. Scientists can detect and measure the expression of thousands of genes at one time using **DNA microarray assays.**
      a. Tiny amounts of a large number of single-stranded DNA fragments representing different genes are fixed on a glass slide in a tightly spaced grid (array) called a *DNA chip.*
      b. The fragments, sometimes representing all the genes of an organism, are tested for hybridization with various samples of fluorescently labeled cDNA molecules.
      c. Spots where any of the cDNA hybridizes fluoresce with an intensity indicating the relative amount of the mRNA that was in the tissue.

H. Comparisons of genome sequences from different species allow us to determine the evolutionary relationships even between distantly related organisms.
   1. The more similar the nucleotide sequences between two species, the more closely related these species are in their evolutionary history.
   2. The genomes of two closely related species are likely to be similarly organized.

I. The next step after mapping and sequencing genomes is **proteomics,** the systematic study of full protein sets *(proteomes)* encoded by genomes.

J. Because we are all probably descended from a small population living in Africa 150,000 to 200,000 years ago, the amount of DNA variation in humans is small. Most of our diversity is in the form of **single nucleotide polymorphisms (SNPs),** single base-pair variations.
   1. In humans, SNPs occur about once in 1,000 bases, meaning that any two humans are 99.9% identical.
   2. The locations of the human SNP sites will provide useful markers for studying

human evolution, the differences between human populations, and the migratory routes of human populations throughout history.

3. SNPs and other polymorphisms will be valuable markers for identifying disease genes and genes that influence our susceptibility to diseases, toxins, or drugs.

X. Practical applications of DNA technology

A. DNA technology is reshaping medicine and the pharmaceutical industry.

B. The identification of genes whose mutations are responsible for genetic diseases may lead to ways to diagnose, treat, or even prevent these conditions.

C. Diseases of all sorts involve changes in gene expression within the affected genes and within the patient's immune system. DNA technology can identify these changes and lead to the development of targets for prevention or therapy.

D. PCR and labeled nucleic acid probes can track down the pathogens responsible for infectious diseases.

E. Medical scientists can use DNA technology to identify individuals with genetic diseases before the onset of symptoms, even before birth.

F. Genetic disorders are diagnosed by using PCR and primers corresponding to cloned disease genes, and then sequencing the amplified product to look for the disease-causing mutation.

G. It is even possible to identify symptomless carriers of these diseases.

H. It is possible to detect abnormal allelic forms of genes, even in cases in which the gene has not yet been cloned.

1. The presence of an abnormal allele can be diagnosed with reasonable accuracy if a closely linked RFLP marker has been found.

2. The closeness of the marker to the gene makes crossing over between them unlikely, and the marker and gene will almost always be inherited together.

I. Techniques for gene manipulation hold great potential for treating disease by **gene therapy,** the alteration of an afflicted individual's genes.

1. A normal allele is inserted into somatic cells of a tissue affected by a genetic disorder.

2. For gene therapy of somatic cells to be permanent, the cells that receive the normal allele must be ones that multiply throughout the patient's life.

3. Bone marrow cells, which include the stem cells that give rise to blood and immune system cells, are prime candidates for gene therapy.

a. A normal allele can be inserted by a retroviral vector into bone marrow cells removed from the patient.

b. If the procedure succeeds, the returned modified cells will multiply throughout the patient's life and express the normal gene, providing missing proteins.

4. Gene therapy poses many technical questions.

a. These include regulation of the activity of the transferred gene to produce the appropriate amount of the gene product at the right time and place.

b. In addition, the insertion of the therapeutic gene must not harm other necessary cell functions.

5. Gene therapy raises some difficult ethical and social questions.

a. Some critics suggest that tampering with human genes, even for those with life-threatening diseases, is wrong.

b. They argue that this will lead to the practice of eugenics, a deliberate effort to control the genetic makeup of human populations.

c. The most difficult ethical question is whether we should treat human germ-line cells to correct the defect in future generations.

6. From a biological perspective, the elimination of unwanted alleles from the gene pool could backfire.
    a. Genetic variation is a necessary ingredient for the survival of a species as environmental conditions change with time.
    b. Genes that are damaging under some conditions could be advantageous under other conditions, as in the example of the sickle-cell allele.

J. DNA technology has been used to create many useful pharmaceuticals, mostly proteins.
    1. By transferring the gene for a protein into a host that is easily grown in culture, one can produce large quantities of normally rare proteins.
    2. New pharmaceutical products are responsible for novel ways of fighting diseases that do not respond to traditional drug treatments.
        a. One approach is to use genetically engineered proteins that either block or mimic surface receptors on cell membranes to prevent viral infection.
    3. DNA technology can also be used to produce vaccines, which stimulate the immune system to defend against specific pathogens.

XI. DNA technology offers forensic, environmental, and agricultural applications.

A. In violent crimes, blood, semen, or traces of other tissues may be left at the scene or on the clothes or other possessions of the victim or assailant.
    1. DNA testing can identify the guilty individual with a much higher degree of certainty, because the DNA sequence of every person is unique (except for identical twins).
    2. The probability that two people who are not identical twins have the same DNA fingerprint is very small.
    3. In practice, forensic DNA tests focus on only about five tiny regions of the genome. The probability that two people will have identical DNA fingerprints in these highly variable regions is typically between one in 100,000 and one in a billion. The exact figure depends on the number of markers and the frequency of those markers in the population.

B. Increasingly, genetic engineering is being applied to environmental work.
    1. Scientists are engineering the metabolism of microorganisms to help cope with some environmental problems.
        a. For example, genetically engineered microbes that can extract heavy metals from their environments
        b. In addition to the normal microbes that participate in sewage treatment, new microbes that can degrade other harmful compounds are being engineered.
        c. Bacterial strains have been developed that can degrade some of the chemicals released during oil spills.

C. For many years, scientists have been using DNA technology to improve agricultural productivity.
    1. DNA technology is now routinely used to make vaccines and growth hormones for farm animals.
    2. **Transgenic organisms** are made by introducing genes from one species into the genome of another organism.
        a. An egg cell is removed from a female animal and fertilized *in vitro*.
        b. Meanwhile, the desired gene is obtained from another organism and cloned.
        c. The cloned DNA is injected directly into the nuclei of the fertilized egg.
        d. The engineered embryos are surgically implanted in a surrogate mother.
    3. Agricultural scientists have engineered a number of crop plants with genes for desirable traits. These include delayed ripening and resistance to spoilage and disease.

XII. DNA technology raises important safety and ethical questions; mostly, that recombinant DNA technology might create hazardous new pathogens.

A. Strict laboratory procedures are designed to protect researchers from infection by engineered microbes and to prevent their accidental release.

B. Some strains of microorganisms used in recombinant DNA experiments are genetically crippled to ensure that they cannot survive outside the laboratory.

C. Much public concern centers on **genetically modified (GM) organisms** used in agriculture.

1. GM organisms have acquired one or more genes (perhaps from another species) by artificial means.
2. Advocates of a cautious approach fear that GM crops might somehow be hazardous to human health or cause ecological harm.
3. In particular, transgenic plants might pass their new genes to close relatives in nearby wild areas through pollen transfer.