Chapter 20

DNA Technology and Genomics

Lecture Outline

Overview

- One of the great achievements of modern science has been the sequencing of the human genome, which was largely completed by 2003.
- Progress began with the development of techniques for making recombinant DNA, in which genes from two different sources—and often different species—are combined in vitro into the same molecule.
- The methods for making recombinant DNA are central to genetic engineering, the direct manipulation of genes for practical purposes.
    - ° Applications include the introduction of a desired gene into the DNA of a host that will produce the desired protein.
- DNA technology has launched a revolution in **biotechnology,** the manipulation of organisms or their components to make useful products.
    - ° Practices that go back centuries, such as the use of microbes to make wine and cheese and the selective breeding of livestock, are examples of biotechnology.
        - ▪ These techniques exploit naturally occurring mutations and genetic recombination.
    - ° Biotechnology based on the manipulation of DNA *in vitro* differs from earlier practices by enabling scientists to modify specific genes and move them between organisms as distinct as bacteria, plants, and animals.
- DNA technology is now applied in areas ranging from agriculture to criminal law, but its most important achievements are in basic research.

A. DNA Cloning

- 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 multiple identical copies of gene-sized pieces of DNA.

1. DNA cloning permits production of multiple copies of a specific gene or other DNA segment.

- One basic cloning technique begins with the insertion of a foreign gene into a bacterial plasmid.
    - ° *E. coli* and its plasmids are commonly used.
    - ° First, a foreign gene is inserted into a bacterial plasmid to produce a recombinant DNA molecule.

- ° 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.
- ° Under suitable conditions, the bacterial clone will make the protein encoded by the foreign gene.

- The potential uses of cloned genes fall into two general categories.
  - ° First, the goal may be to produce a protein product.
    - ▪ For example, bacteria carrying the gene for human growth hormone can produce large quantities of the hormone.
  - ° Alternatively, the goal may be to prepare many copies of the gene itself.
    - ▪ This may enable scientists to determine the gene's nucleotide sequence or provide an organism with a new metabolic capability by transferring a gene from another organism.
  - ° Most protein-coding genes exist in only one copy per genome, so the ability to clone rare DNA fragments is very valuable.

2. Restriction enzymes are used to make recombinant DNA.

- Gene cloning and genetic engineering were made possible by the discovery of **restriction enzymes** that cut DNA molecules at specific locations.

- 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.*

- Most 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.

- 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.
    - ▪ If the restriction site on one strand is 3'-CTTAAG-5', the complementary strand is 5'-GAATTC-3'.

- 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.

- Restriction enzymes cut covalent sugar-phosphate backbones of both strands, often in a staggered way that creates single-stranded **sticky ends.**
  - ° 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.

- 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.

- 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.

3. Eukaryotic genes can be cloned in bacterial plasmids.

- Recombinant plasmids are produced by splicing restriction fragments from foreign DNA into plasmids.
  - ° 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.

- Bacterial plasmids are widely used as cloning vectors for several reasons.
  - ° They can be easily isolated from bacteria, manipulated to form recombinant plasmids by *in vitro* insertion of foreign DNA, and then reintroduced into bacterial cells.

- Bacterial cells carrying the recombinant plasmid reproduce rapidly, replicating the inserted foreign DNA.

- 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 gene-source DNA.
  - ° The source DNA comes from human tissue cells grown in lab culture.
  - ° The source of the plasmid is typically *E. coli*.
  - ° 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.
  - ° The plasmid has a single recognition sequence, within the *lacZ* gene, for the restriction enzyme used.
  2. DNA is inserted into the vector.
  - ° 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 creates a diverse pool of bacteria: some bacteria that have taken up the desired recombinant plasmid DNA, and other bacteria that have taken up other DNA, both recombinant and nonrecombinant.
  5. The transformed bacteria are plated on a solid nutrient medium containing ampicillin and a molecular mimic of lactose called X-gal.
  - ° Only bacteria that have the ampicillin-resistance (amp^R) plasmid will grow.
    - ▪ Each reproducing bacterium forms a clone by repeating cell divisions, generating a *colony* of cells on the agar.
  - ° The lactose mimic in the medium is used to identify plasmids that carry foreign DNA.
    - ▪ Bacteria with plasmids lacking foreign DNA stain blue when ß-galactosidase from the intact *lacZ* gene hydrolyzes X-gal.
    - ▪ Bacteria with plasmids containing foreign DNA inserted into the *lacZ* gene are white because they lack ß-galactosidase.
  6. Cell clones with the right gene are identified.

- ° In the final step, thousands of bacterial colonies with foreign DNA must be sorted through to find those containing the gene of interest.
- ° 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.
  - ▪ The sequence of the RNA or DNA probe depends on knowledge of at least part of the sequence of the gene of interest.
  - ▪ A radioactive or fluorescent tag is used to label the probe.
  - ▪ The probe will hydrogen-bond specifically to complementary single strands of the desired gene.
  - ▪ After **denaturing** (separating) the DNA strands in the bacterium, the probe will bind with its complementary sequence, tagging colonies with the targeted gene.

4. Cloned genes are stored in DNA libraries.

- In the "shotgun" cloning approach described above, a mixture of fragments from the entire genome is included in thousands of different recombinant plasmids.

- A complete set of recombinant plasmid clones, each carrying copies of a particular segment from the initial genome, forms a **genomic library.**
  - ° The library can be saved and used as a source of other genes or for gene mapping.

- In addition to plasmids, certain bacteriophages are also common cloning vectors for making genomic libraries.
  - ° 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.
  - ° The recombinant phage DNA is packaged in a capsid *in vitro* and allowed to infect a bacterial cell.
  - ° Infected bacteria produce new phage particles, each with the foreign DNA.

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

- The enzyme reverse transcriptase is used to make single-stranded DNA transcripts of the mRNA molecules.

- The mRNA is enzymatically digested, and a second DNA strand complementary to the first is synthesized by DNA polymerase.
  - ° This double-stranded DNA, called **complementary DNA (cDNA),** is modified by the addition of restriction sites at each end.
  - ° Finally, the cDNA is inserted into vector DNA.
  - ° A **cDNA library** represents that part of a cell's genome that was transcribed in the starting cells.
    - ▪ This is an advantage if a researcher wants to study the genes responsible for specialized functions of a particular kind of cell.
    - ▪ 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.

- If a researcher wants to clone a gene but is unsure in what cell type it is expressed or unable to obtain that cell type, a genomic library will likely contain the gene.

- A researcher interested in the regulatory sequences or introns associated with a gene will need to obtain the gene from a genomic library.
  - ° These sequences are missing from the processed mRNAs used in making a cDNA library.

5. *Eukaryote genes can be expressed in prokaryotic host cells.*

- A clone can sometimes be screened for a desired gene based on detection of its encoded protein.
- Inducing a cloned eukaryotic gene to function in a prokaryotic host can be difficult.
  - ° One way around this is to insert an **expression vector,** a cloning vector containing a highly active prokaryotic promoter, upstream of the restriction site.
  - ° The prokaryotic host will then recognize the promoter and proceed to express the foreign gene that has been linked to it.
  - ° Such expression vectors allow the synthesis of many eukaryotic proteins in prokaryotic cells.
- The presence of long noncoding introns in eukaryotic genes may prevent correct expression of these genes in prokaryotes, which lack RNA-splicing machinery.
  - ° This problem can be surmounted by using a cDNA form of the gene inserted in a vector containing a bacterial promoter.
- Molecular biologists can avoid incompatibility problems by using eukaryotic cells as hosts for cloning and expressing eukaryotic genes.
  - ° Yeast cells, single-celled fungi, are as easy to grow as bacteria and, unlike most eukaryotes, have plasmids.
- 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.
  - ° These chromosome-like vectors behave normally in mitosis and can carry more DNA than a plasmid.
- Another advantage of eukaryotic hosts is that they are capable of providing the posttranslational modifications that many proteins require.
  - ° Such modifications may include adding carbohydrates or lipids.
  - ° For some mammalian proteins, the host must be an animal cell to perform the necessary modifications.
- Many eukaryotic cells can take up DNA from their surroundings, but inefficiently.
- Several techniques facilitate entry of foreign DNA into eukaryotic cells.
  - ° In **electroporation,** brief electrical pulses create a temporary hole in the plasma membrane through which DNA can enter.
  - ° Alternatively, scientists can inject DNA into individual cells using microscopically thin needles.
  - ° Once inside the cell, the DNA is incorporated into the cell's DNA by natural genetic recombination.

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

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

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

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

- 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.

- PCR can make billions of copies of a targeted DNA segment in a few hours.
  - This is faster than cloning via recombinant bacteria.

- 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.
  - The reaction mixture is heated to denature the DNA strands.
  - 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.
  - A heat-stable DNA polymerase extends the primers in the 5' $\rightarrow$ 3' direction.

- If a standard DNA polymerase were used, the protein would be denatured along with the DNA during the heating step.

- The key to easy PCR automation was the discovery of an unusual DNA polymerase, isolated from prokaryotes living in hot springs, which can withstand the heat needed to separate the DNA strands at the start of each cycle.

- PCR is very specific.

- By their complementarity to sequences bracketing the targeted sequence, the primers determine the DNA sequence that is amplified.
  - PCR can make many copies of a specific gene before cloning in cells, simplifying the task of finding a clone with that gene.
  - PCR is so specific and powerful that only minute amounts of partially degraded DNA need be present in the starting material.

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

- Devised in 1985, PCR has had a major impact on biological research and technology.
  - PCR has amplified DNA from a variety of sources:
    - Fragments of ancient DNA from a 40,000-year-old frozen woolly mammoth.
    - DNA from footprints or tiny amounts of blood or semen found at the scenes of violent crimes.
    - DNA from single embryonic cells for rapid prenatal diagnosis of genetic disorders.
    - DNA of viral genes from cells infected with HIV.

7. Restriction fragment analysis detects DNA differences that affect restriction sites.

- Once we have prepared homogeneous samples of DNA, each containing a large number of identical segments, we can begin to ask some interesting questions about specific genes and their functions.
  - Does a particular gene differ from person to person?
  - Are certain alleles associated with a hereditary disorder?

- ° Where in the body and when during development is a gene expressed?
- ° What is the location of a gene in the genome?
- ° Is expression of a particular gene related to expression of other genes?
- ° How has a gene evolved, as revealed by interspecific comparisons?

- To answer these questions, we need to know the nucleotide sequence of the gene and its counterparts in other individuals and species, as well as its expression pattern.

- One indirect method of rapidly analyzing and comparing genomes is **gel electrophoresis.**
  - ° Gel electrophoresis separates macromolecules—nucleic acids or proteins—on the basis of their rate of movement through a gel in an electrical field.
    - ▪ Rate of movement depends on size, electrical charge, and other physical properties of the macromolecules.

- In restriction fragment analysis, the DNA fragments produced by restriction enzyme digestion of a DNA molecule are sorted by gel electrophoresis.
  - ° 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.
  - ° The relatively small DNA molecules of viruses and plasmids can be identified simply by their restriction fragment patterns.
  - ° The separated fragments can be recovered undamaged from gels, providing pure samples of individual fragments.

- We can use restriction fragment analysis to compare two different DNA molecules representing, for example, different alleles of a gene.
  - ° Because the two alleles differ slightly in DNA sequence, they may differ in one or more restriction sites.
  - ° If they do differ in restriction sites, each will produce different-sized fragments when digested by the same restriction enzyme.
  - ° In gel electrophoresis, the restriction fragments from the two alleles will produce different band patterns, allowing us to distinguish the two alleles.

- 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.

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

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

- One of its many applications is to identify heterozygous carriers of mutant alleles associated with genetic disease.

- 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.

- We combine several molecular techniques to compare DNA samples from three individuals.
  1. We start by adding the same restriction enzyme to each of the three samples to produce restriction fragments.
  2. We then separate the fragments by gel electrophoresis.
  3. We transfer the DNA fragments from the gel to a sheet of nitrocellulose paper, still separated by size.
  ° This also denatures the DNA fragments.
  4. 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.
  5. We can visualize bands containing the label with autoradiography.
- The band pattern for the heterozygous individual will be a combination of the patterns for the two homozygotes.

8. Restriction fragment length differences are useful as genetic markers.

- Restriction fragment analysis can be used to examine differences in *noncoding* DNA as well.
- Differences in DNA sequence on homologous chromosomes that produce different restriction fragment patterns are scattered abundantly throughout genomes, including the human genome.
- A **restriction fragment length polymorphism** (**RFLP** or Rif-lip) can serve as a genetic marker for a particular location (locus) in the genome.
- RFLPs are detected and analyzed by Southern blotting, frequently using the entire genome as the DNA starting material.
  ° The probe is complementary to the sequence under consideration.
- Because RFLP markers are inherited in a Mendelian fashion, they can serve as genetic markers for making linkage maps.
  ° The frequency with which two RFPL markers—or an RFLP marker and a certain allele for a gene—are inherited together is a measure of the closeness of the two loci on a chromosome.

B. DNA Analysis and Genomics

- The field of **genomics** is based on comparisons among whole sets of genes and their interactions.

1. Entire genomes can be mapped at the DNA level.

- As early as 1980, Daniel Botstein and his colleagues proposed that the DNA variations reflected in RFLPs could serve as the basis of an extremely detailed map of the entire human genome.
  ° Since then, researchers have used such markers in conjunction with the tools and techniques of DNA technology to develop detailed maps of the genomes of a number of species.
- The most ambitious research project made possible by DNA technology has been the sequencing of the human genome, officially begun as the **Human Genome Project** in 1990.
  ° This effort was largely completed in 2003 when the nucleotide sequence of the vast majority of DNA in the human genome was obtained.
  ° An international, publicly funded consortium of researchers at universities and research institutes has taken this project through three stages that provided progressively more

detailed views of the human genome: genetic (linkage) mapping, physical mapping, and DNA sequencing.

- In addition to mapping human DNA, the genomes of other organisms important to biological research are also being mapped.
  - ° Completed sequences include those of *E. coli* and other prokaryotes, *Saccharomyces cerevisiae* (yeast), *Drosophila melanogaster* (fruit fly), *Mus musculus* (mouse), and others.

- These genomes are providing important insights of general biological significance.

- In mapping a large genome, cytogenetic maps based on karyotyping and fluorescence hybridization provide a starting point for more detailed mapping.
  - ° The first stage is to construct a **linkage map** of several thousand markers spaced throughout the chromosomes.
  - ° The order of the markers and the relative distances between them on such a map are based on recombination frequencies.
  - ° The markers can be genes or any other identifiable sequences in DNA, such as RFLPs or simple sequence DNA.

- The human map with 5,000 genetic markers enabled researchers to locate other markers, including genes, by testing for genetic linkage with the known markers.

- The next step was converting the relative distances to some physical measure, usually the number of nucleotides along the DNA.

- For whole-genome mapping, a **physical map** is made by cutting the DNA of each chromosome into identifiable restriction fragments and then determining the original order of the fragments.
  - ° The key is to make fragments that overlap and then use probes or automated nucleotide sequencing of the ends to find the overlaps.

- When working with large genomes, researchers carry out several rounds of DNA cutting, cloning, and physical mapping.
  - ° The first cloning vector is often a yeast artificial chromosome (YAC), which can carry inserted fragments up to a million base pairs long, or a **bacterial artificial chromosome (BAC),** which can carry inserts of 100,000 to 500,000 base pairs.
  - ° After the order of these long fragments has been determined, each fragment is cut into pieces that are cloned in plasmids or phages, ordered, and finally sequenced.

- The complete nucleotide sequence of a genome is the ultimate map.
  - ° Starting with a pure preparation of many copies of a relatively short DNA fragment, the nucleotide sequence of the fragment can be determined by a sequencing machine.
  - ° The usual sequencing technique combines DNA labeling, DNA synthesis with special chain-terminating nucleotides, and high-resolution gel electrophoresis.
  - ° A major thrust of the Human Genome Project has been the development of technology for faster sequencing and more sophisticated computer software for analyzing and assembling the partial sequences.

- One common method of sequencing DNA, the Sanger or *dideoxyribonucleotide chain-termination method,* is similar to PCR.
  - ° Inclusion of special dideoxyribonucleotides in the reaction mix ensures that rather than copying the whole template, fragments of various lengths will be synthesized.

- ° These dideoxyribonucleotides, marked radioactively or fluorescently, terminate elongation when they are incorporated randomly into the growing strand because they lack a 3'-OH to attach the next nucleotide.
- The order of these fragments via gel electrophoresis can be interpreted as the nucleotide sequence.
- While the public consortium followed a hierarchical, three-stage approach for sequencing an entire genome, J. Craig Venter decided in 1992 to try a whole-genome shotgun approach.
  - ° This used powerful computers to assemble sequences from random fragments, skipping the first two steps.
- The worth of his approach was demonstrated in 1995 when he and colleagues reported the complete sequence of a bacterium.
- His private company, Celera Genomics, finished the sequence of *Drosophila melanogaster* in 2000.
- In February 2001, Celera and the public consortium separately announced sequencing more than 90% of the human genome.
- Sequencing of the human genome is now virtually complete, although some gaps remain to be mapped.
  - ° Areas with repetitive DNA and certain parts of the chromosomes of multicellular organisms resist detailed mapping by the usual methods.
- On one level, genome sequences of humans and other organisms are simply lists of nucleotide bases.
  - ° On another level, analyses of these sequences and comparisons between species are leading to exciting discoveries.

2. Genome sequences provide clues to important biological questions.

- **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.
- Rather than inferring genotype from phenotype as classical geneticists did, molecular geneticists can study genes directly.
  - ° This approach poses the challenge of determining phenotype from genotype.
  - ° Starting with a long DNA sequence, how does a researcher recognize genes and determine their function?
- DNA sequences are collected in computer data banks that are available via the Internet to researchers everywhere.
- 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.
- The software also looks for *expressed sequence tags (ESTs),* sequences similar to those in known genes.
  - ° From these clues, researchers collect a list of gene candidates.
- Although genome size increases from prokaryotes to eukaryotes, it does not always correlate with biological complexity among eukaryotes.
  - ° One flowering plant has a genome 40 times the size of the human genome.
- An organism may have fewer genes than expected from the size of its genome.

- ° The estimated number of human genes is 25,000 or fewer, only about one-and-a-half times the number found in the fruit fly.
- ° This is surprising, given the great diversity of cell types in humans.
- 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.
- By doing more mixing and matching of modular elements, humans—and vertebrates in general—reach greater complexity than flies or worms.
  - ° Gene expression is regulated in more subtle and complicated ways in vertebrates than in other organisms.
  - ° The typical human gene specifies several different polypeptides by using different combinations of exons.
    - ▪ Nearly all human genes contain multiple 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.
- About half of the human genes were already known before the Human Genome Project.
- To determine what the others are and what they may do, scientists compare the sequences of new gene candidates with those of known genes.
  - ° In some cases, the sequence of a new gene candidate will be similar in part to that of a known gene, suggesting similar function.
  - ° In other cases, the new sequences will be similar to a sequence encountered before, but of unknown function.
  - ° In still other cases, the sequence is entirely unlike anything ever seen before.
    - ▪ About 30% of the *E. coli* genes are new to us.
- How can scientists determine the function of new genes identified by genome sequencing and comparative analysis?
- One way to determine their function is to disable the gene and observe the consequences.
  - ° Using *in vitro* **mutagenesis,** specific mutations are introduced into a cloned gene, altering or destroying its function.
  - ° 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.
  - ° 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.
- In nonmammalian organisms, a simpler and faster method, **RNA interference (RNAi),** has been applied to silence the expression of selected genes.
  - ° This method uses synthetic double-stranded RNA molecules matching the sequences of a particular gene to trigger breakdown of the gene's mRNA.
  - ° 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.
  - ° In one study, RNAi was used to prevent expression of 86% of the genes in early nematode embryos, one gene at a time.

- ° Analysis of the phenotypes of the worms that developed from these embryos allowed the researchers to group most of the genes into functional groups.
- A major goal of genomics is to learn how genes act together to produce a functioning organism.
  - ° Part of the explanation for how humans get along with so few genes probably lies in the unusual complexity of networks of interactions among genes and their products.
- As the sequences of entire genomes of several organisms neared completion, some researchers began to investigate which genes are transcribed under different situations.
- They also looked for groups of genes that are expressed in a coordinated pattern to identify global patterns or networks of expression.
- The basic strategy in global expression is to isolate mRNAs from particular cells and use the mRNA as a template to build cDNA by reverse transcription.
  - ° Each cDNA can be compared to other collections of DNA by hybridization.
  - ° This will reveal which genes are active at different developmental stages, in different tissues, or in tissues in different states of health.
- Automation has allowed scientists to detect and measure the expression of thousands of genes at one time using **DNA microarray assays.**
  - ° 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).
    - ▪ The array is called a *DNA chip.*
  - ° The fragments, sometimes representing all the genes of an organism, are tested for hybridization with various samples of fluorescently labeled cDNA molecules.
- Spots where any of the cDNA hybridizes fluoresce with an intensity indicating the relative amount of the mRNA that was in the tissue.
- Ultimately, information from microarray assays should provide us a grander view: how ensembles of genes interact to form a living organism.
  - ° DNA microarray assays are being used to compare cancerous versus noncancerous tissues.
    - ▪ This may lead to new diagnostic techniques and biochemically targeted treatments, as well as a fuller understanding of cancer.
- The genomes of about 150 species have been completely or almost completely sequenced by the spring of 2004, with many more in progress.
  - ° Most of these are prokaryotes, including 20 archaean genomes.
  - ° Among the 20 eukaryotic species are vertebrates, invertebrates, and plants.
- Comparisons of genome sequences from different species allow us to determine the evolutionary relationships even between distantly related organisms.
- The more similar the nucleotide sequences between two species, the more closely related these species are in their evolutionary history.
- Comparisons of the complete genome sequences of bacteria, archaea, and Eukarya support the theory that these are the three fundamental domains of life.
- Comparative genome studies confirm the relevance of research on simpler organisms to our understanding of human biology.
  - ° The yeast genome is proving useful in helping us to understand the human genome.

- Comparisons of noncoding sequences in the human genome to those in the much smaller yeast genome revealed regions with highly conserved sequences that are important regulatory sequences in both species.
- Several yeast protein-coding genes are so similar to certain human disease genes that researchers have figured out the functions of the disease genes by studying their normal yeast counterparts.

- The genomes of two closely related species are likely to be similarly organized.
  ° Once the sequence and organization of one genome is known, it can greatly accelerate the mapping of a related genome.
    - For example, the mouse genome can be mapped quickly, with the human genome serving as a guide.

- The small number of gene differences between closely related species makes it easier to correlate phenotypic differences between species with particular genetic differences.
  ° One gene that is clearly different in chimps and humans appears to function in speech.
  ° Researchers may determine what a human disease gene does by studying its normal counterpart in mice, who share 80% of our genes.

- The next step after mapping and sequencing genomes is **proteomics,** the systematic study of full protein sets *(proteomes)* encoded by genomes.
  ° One challenge is the sheer number of proteins in humans and our close relatives because of alternative RNA splicing and posttranslational modifications.
  ° Collecting all the proteins produced by an organism will be difficult because a cell's proteins differ with cell type and its state.
  ° Unlike DNA, proteins are extremely varied in structure and chemical and physical properties.
  ° Because proteins are the molecules that actually carry out cell activities, we must study them to learn how cells and organisms function.

- Complete catalogs of genes and proteins will change the discipline of biology dramatically.
  ° With such catalogs in hand, researchers are turning their attention to the functional integration of individual components in biological systems.

- Advances in **bioinformatics,** the application of computer science and mathematics to genetic and other biological information, will play a crucial role in dealing with the enormous mass of data.

- These analyses will provide understanding of the spectrum of genetic variation in humans.
  ° 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.
    - In humans, SNPs occur about once in 1,000 bases, meaning that any two humans are 99.9% identical.
  ° 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.
  ° SNPs and other polymorphisms will be valuable markers for identifying disease genes and genes that influence our susceptibility to diseases, toxins, or drugs.
    - This will change the practice of 21st-century medicine.

C. Practical Applications of DNA Technology

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

- Modern biotechnology is making enormous contributions both to the diagnosis of diseases and in the development of pharmaceutical products.
  - ° The identification of genes whose mutations are responsible for genetic diseases may lead to ways to diagnose, treat, or even prevent these conditions.
  - ° Susceptibility to many "nongenetic" diseases, from arthritis to AIDS, is influenced by a person's genes.
  - ° 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.
- PCR and labeled nucleic acid probes can track down the pathogens responsible for infectious diseases.
  - ° For example, PCR can amplify and thus detect HIV DNA in blood and tissue samples, detecting an otherwise elusive infection.
  - ° RNA cannot be directly amplified by PCR.
  - ° The RNA genome is first converted to double-stranded cDNA by a technique called *RT-PCR,* using a probe specific for one of the HIV genes.
- Medical scientists can use DNA technology to identify individuals with genetic diseases before the onset of symptoms, even before birth.
  - ° 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.
    - ▪ Cloned disease genes include those for sickle-cell disease, hemophilia, cystic fibrosis, Huntington's disease, and Duchenne muscular dystrophy.
    - ▪ It is even possible to identify symptomless carriers of these diseases.
- It is possible to detect abnormal allelic forms of genes, even in cases in which the gene has not yet been cloned.
  - ° The presence of an abnormal allele can be diagnosed with reasonable accuracy if a closely linked RFLP marker has been found.
  - ° 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.
- Techniques for gene manipulation hold great potential for treating disease by **gene therapy,** the alteration of an afflicted individual's genes.
  - ° A normal allele is inserted into somatic cells of a tissue affected by a genetic disorder.
  - ° 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.
- Bone marrow cells, which include the stem cells that give rise to blood and immune system cells, are prime candidates for gene therapy.
  - ° A normal allele can be inserted by a retroviral vector into bone marrow cells removed from the patient.
  - ° If the procedure succeeds, the returned modified cells will multiply throughout the patient's life and express the normal gene, providing missing proteins.

- This procedure was used in a 2000 trial involving ten young children with SCID (severe combined immunodeficiency disease), a genetic disease in which bone marrow cells do not produce a vital enzyme because of a single defective gene.
  - ° Nine of the children showed significant improvement after two years.
  - ° However, two of the children developed leukemia.
    - ▪ It was discovered that the retroviral vector used to carry the normal allele into bone marrow cells had inserted near a gene involved in proliferation and development of blood cells, causing leukemia.
    - ▪ The trial has been suspended until researchers learn how to control the location of insertion of the retroviral vectors.
- Gene therapy poses many technical questions.
  - ° 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.
  - ° In addition, the insertion of the therapeutic gene must not harm other necessary cell functions.
- Gene therapy raises some difficult ethical and social questions.
  - ° Some critics suggest that tampering with human genes, even for those with life-threatening diseases, is wrong.
  - ° They argue that this will lead to the practice of eugenics, a deliberate effort to control the genetic makeup of human populations.
- The most difficult ethical question is whether we should treat human germ-line cells to correct the defect in future generations.
  - ° In laboratory mice, transferring foreign genes into egg cells is now a routine procedure.
  - ° Once technical problems relating to similar genetic engineering in humans are solved, we will have to face the question of whether it is advisable, under any circumstances, to alter the genomes of human germ lines or embryos.
  - ° Should we interfere with evolution in this way?
- From a biological perspective, the elimination of unwanted alleles from the gene pool could backfire.
  - ° Genetic variation is a necessary ingredient for the survival of a species as environmental conditions change with time.
  - ° Genes that are damaging under some conditions could be advantageous under other conditions, as in the example of the sickle-cell allele.
- DNA technology has been used to create many useful pharmaceuticals, mostly proteins.
- 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.
  - ° By including highly active promoters (and other control elements) into vector DNA, the host cell can be induced to make large amounts of the product of a gene.
  - ° In addition, host cells can be engineered to secrete a protein, simplifying the task of purification.
- One of the first practical applications of gene splicing was the production of mammalian hormones and other mammalian regulatory proteins in bacteria.
  - ° These include human insulin, human growth factor (HGF), and tissue plasminogen activator.

- ° Human insulin, produced by bacteria, is superior for the control of diabetes to the older treatment of pig or cattle insulin.
- ° Human growth hormone benefits children with hypopituitarism, a form of dwarfism.
- ° Tissue plasminogen activator (TPA) helps dissolve blood clots and reduce the risk of future heart attacks.
  - ▪ Like many such drugs, it is expensive.
- New pharmaceutical products are responsible for novel ways of fighting diseases that do not respond to traditional drug treatments.
  - ° One approach is to use genetically engineered proteins that either block or mimic surface receptors on cell membranes.
  - ° For example, one experimental drug mimics a receptor protein that HIV bonds to when entering white blood cells. HIV binds to the drug instead and fails to enter the blood cells.
- DNA technology can also be used to produce vaccines, which stimulate the immune system to defend against specific pathogens.
  - ° A vaccine is a harmless variant or derivative of a pathogen that stimulates the immune system.
  - ° Traditional vaccines are either killed microbes or attenuated microbes that do not cause disease.
  - ° Both are similar enough to the active pathogen to trigger an immune response.
- Recombinant DNA techniques can generate large amounts of a specific protein molecule normally found on the pathogen's surface.
  - ° If this protein triggers an immune response against the intact pathogen, then it can be used as a vaccine.
  - ° Alternatively, genetic engineering can modify the genome of the pathogen to attenuate it.
    - ▪ These attenuated microbes are often more effective than a protein vaccine because they usually trigger a greater response by the immune system.
    - ▪ Pathogens attenuated by gene-splicing techniques may be safer than the natural mutants traditionally used.

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

- 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.
- If enough tissue is available, forensic laboratories can determine blood type or tissue type by using antibodies for specific cell surface proteins.
  - ° However, these tests require relatively large amounts of fresh tissue.
  - ° Also, this approach can only exclude a suspect.
- 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).
  - ° RFPL analysis by Southern blotting can detect similarities and differences in DNA samples and requires only a tiny amount of blood or other tissue.
  - ° Radioactive probes mark electrophoresis bands that contain certain RFLP markers.
  - ° As few as five markers from an individual can be used to create a **DNA fingerprint.**
  - ° The probability that two people who are not identical twins have the same DNA fingerprint is very small.

- DNA fingerprints can be used forensically to present evidence to juries in murder trials.
  - ° An autoradiograph of RFLP bands of samples from a murder victim, the defendant, and the defendant's clothes may be consistent with the conclusion that the blood on the clothes is from the victim, not the defendant.
- The forensic use of DNA fingerprinting extends beyond violent crimes.
  - ° For instance, DNA fingerprinting can be used to settle conclusively questions of paternity.
  - ° DNA fingerprinting recently provided strong evidence that Thomas Jefferson fathered at lease one of the children of his slave Sally Hemings.
  - ° These techniques can also be used to identify the remains of individuals killed in natural or man-made disasters.
- Variations in the lengths of repeated base sequences are increasingly used as markers in DNA fingerprinting.
  - ° Such polymorphic genetic loci have repeating units of only a few base pairs and are highly variable from person to person.
  - ° Individuals may vary in the numbers of *simple tandem repeats (STRs)* at a locus.
  - ° Restriction fragments with STRs vary in size among individuals because of differences in STR lengths.
  - ° PCR is often used to amplify selectively particular STRs or other markers before electrophoresis, especially if the DNA is poor or in minute quantities.
- The DNA fingerprint of an individual would be truly unique if it were feasible to perform restriction fragment analysis on the entire genome.
  - ° 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.
  - ° Despite problems that might arise from insufficient statistical data, human error, or flawed evidence, DNA fingerprinting is now accepted as compelling evidence.
- Increasingly, genetic engineering is being applied to environmental work.
- Scientists are engineering the metabolism of microorganisms to help cope with some environmental problems.
  - ° For example, genetically engineered microbes that can extract heavy metals from their environments and incorporate the metals into recoverable compounds may become important both in mining materials and cleaning up highly toxic mining wastes.
  - ° In addition to the normal microbes that participate in sewage treatment, new microbes that can degrade other harmful compounds are being engineered.
  - ° Bacterial strains have been developed that can degrade some of the chemicals released during oil spills.
- For many years, scientists have been using DNA technology to improve agricultural productivity.
  - ° DNA technology is now routinely used to make vaccines and growth hormones for farm animals.
- **Transgenic organisms** are made by introducing genes from one species into the genome of another organism.

- ° An egg cell is removed from a female animal and fertilized *in vitro*.
  - ° Meanwhile, the desired gene is obtained from another organism and cloned.
  - ° The cloned DNA is injected directly into the nuclei of the fertilized egg.
  - ° Some of the cells integrate the *transgene* into their genomes and express the foreign gene.
  - ° The engineered embryos are surgically implanted in a surrogate mother.

- Transgenic animals may be created to exploit the attributes of new genes (for example, genes for faster growth or larger muscles).

- Other transgenic organisms are pharmaceutical "factories"—producers of large amounts of otherwise rare substances for medical use.
  - ° Transgenic farm mammals may secrete the gene product of interest in their milk.
  - ° Researchers have engineered transgenic chickens that express large quantities of the transgene's product in their eggs.

- The human proteins produced by farm animals may or may not be structurally identical to natural human proteins.
  - ° Therefore, they have to be tested very carefully to ensure that they will not cause allergic reactions or other adverse effects in patients receiving them.
  - ° In addition, the health and welfare of transgenic farm animals are important issues, as they often suffer from lower fertility or increased susceptibility to disease.

- Agricultural scientists have engineered a number of crop plants with genes for desirable traits.
  - ° These include delayed ripening and resistance to spoilage and disease.
  - ° Because a single transgenic plant cell can be grown in culture to generate an adult plant, plants are easier to engineer than most animals.

- The **Ti plasmid,** from the soil bacterium *Agrobacterium tumefaciens*, is often used to introduce new genes into plant cells.
  - ° The Ti plasmid normally integrates a segment of its DNA into its host plant and induces tumors.

- Foreign genes can be inserted into the Ti plasmid (a version that does not cause disease) using recombinant DNA techniques.
  - ° The recombinant plasmid can be put back into *Agrobacterium*, which then infects plant cells, or introduced directly into plant cells.

- Genetic engineering is quickly replacing traditional plant-breeding programs, especially for useful traits determined by one or a few genes, like herbicide or pest resistance.
  - ° Use of genetically modified crops has reduced the need for chemical insecticides.

- Scientists are using gene transfer to improve the nutritional value of crop plants.
  - ° For example, a transgenic rice plant has been developed that produces yellow grains containing beta-carotene, which our bodies use to make vitamin A.
    - ▪ Large numbers of young people in southeast Asia are deficient in vitamin A, leading to vision impairment and increased disease rates.

- DNA technology has led to new alliances between the pharmaceutical industry and agriculture.
  - ° Plants can be engineered to produce human proteins for medical use and viral proteins for use as vaccines.

- ° Several such "pharm" products are in clinical trials, including vaccines for hepatitis B and an antibody that blocks the bacteria that cause tooth decay.
- ° The advantage of pharm plants is that large amounts of proteins might be made more economically by plants than by cultured cells.

3. DNA technology raises important safety and ethical questions.

- The power of DNA technology has led to worries about potential dangers.
  - ° Early concerns focused on the possibility that recombinant DNA technology might create hazardous new pathogens.

- In response, scientists developed a set of guidelines that have become formal government regulations in the United States and some other countries.
  - ° Strict laboratory procedures are designed to protect researchers from infection by engineered microbes and to prevent their accidental release.
  - ° Some strains of microorganisms used in recombinant DNA experiments are genetically crippled to ensure that they cannot survive outside the laboratory.
  - ° Finally, certain obviously dangerous experiments have been banned.

- Today, most public concern centers on **genetically modified (GM) organisms** used in agriculture.
  - ° GM organisms have acquired one or more genes (perhaps from another species) by artificial means.
  - ° Salmon have been genetically modified by addition of a more active salmon growth hormone gene.
  - ° However, the majority of GM organisms in our food supply are not animals but crop plants.

- In 1999, the European Union suspended the introduction of new GM crops pending new legislation.
  - ° Early in 2000, negotiators from 130 countries, including the United States, agreed on a Biosafety Protocol that requires exporters to identify GM organisms present in bulk food shipments.

- Advocates of a cautious approach fear that GM crops might somehow be hazardous to human health or cause ecological harm.
  - ° In particular, transgenic plants might pass their new genes to close relatives in nearby wild areas through pollen transfer.
  - ° Transference of genes for resistance to herbicides, diseases, or insect pests may lead to the development of wild "superweeds" that would be difficult to control.

- To date there is little good data either for or against any special health or environmental risks posed by genetically modified crops.

- Today, governments and regulatory agencies are grappling with how to facilitate the use of biotechnology in agriculture, industry, and medicine while ensuring that new products and procedures are safe.
  - ° In the United States, all projects are evaluated for potential risks by various regulatory agencies, including the Food and Drug Administration, Environmental Protection Agency, the National Institutes of Health, and the Department of Agriculture.
  - ° These agencies are under increasing pressures from some consumer groups.

- As with all new technologies, developments in DNA technology have ethical overtones.

- ° Who should have the right to examine someone else's genes?
- ° How should that information be used?
- ° Should a person's genome be a factor in suitability for a job or eligibility for life insurance?

- The power of DNA technology and genetic engineering demands that we proceed with humility and caution.