Chapter 7: The New Genetics—Techniques for DNA Analysis

Introduction

Before the 1980s, finding the genotype of an individual usually involved various laboratory assays for a gene product—the protein or enzyme. The cases of the ABO and Rhesus blood groups are classic examples of how one infers genotypes from the reaction of gene products with certain chemicals. In the mid 1980s, genetic technology took a great leap forward with the ability to genotype the DNA itself. The geneticist could now examine the DNA directly without going through the laborious process of developing assays to detect individual differences in proteins and enzymes. Direct DNA analysis had the further advantage of being able to identify alleles in sections of DNA that did not code for polypeptide chains. As a result of these new advances, the number of genetic loci that could be detected increased exponentially and soon led to the identification of the genes for disorders that had remained a mystery for the better part of this century.

In this chapter, the major molecular techniques are outlined. The purpose is to provide a quick and understandable reference for the social scientist. The content of this chapter is not something that is required to understand genetics, what genes are, or how they relate to human behavior. Indeed, this chapter may be skipped without any great loss of continuity. Hence, only the essentials are given and the reader interested in the laboratory science behind the techniques is referred to contemporary textbooks on molecular genetics. We begin by defining a series of basic tools and techniques.

Basic Tools and Techniques:

Basic tools: Electrophoresis

Electrophoresis is a technique that separates small biological molecules by their molecular weight. It may be applied to molecules as large as proteins and enzymes as well as to small snippets of DNA and RNA. One begins the procedure by constructing a "gel"—a highly viscous material the actual chemistry of which need not concern us. Purified copies of the biological specimen are then injected into a "starting lane" at one end of the gel. Finally, a weak electric current is passed through the gel for a specified amount of time. Gravity and the electric current cause the biological molecules to migrate to the opposite end of the gel. The extent to which any molecule moves depends upon its electrical charge, molecular weight, the viscosity of the gel, the strength of the current, and the amount of time that the current is applied. With constant charge, viscosity, current, and time, smaller molecules will migrate further through the gel than larger molecules.

[Insert Figure 7.1 about here]

Basic Tools: Cloning

In popular imagination, the term "cloning" is associated with *Brave New World*, science fiction, and a famous sheep called Dolly. To the molecular geneticist, however, the term "cloning" simply means the copying of a desired section of DNA. Historically, cloning began by isolating a small (i.e., several thousand base pairs) section of human DNA and then, after an elaborate series of steps, incorporating this segment into the DNA of another organism. The "other" organism is called a *vector*, and of course, not any type of vector will do. To make many copies of the human DNA segment, one desires a vector that

can reproduce rapidly. Hence, the most common vectors are the smallest organisms with dramatic reproductive potential—plasmids, virus, bacteria, and yeast.1

There are several major reasons for cloning human DNA. One purpose is to obtain large amounts of a human DNA sequence that can then be used as a *probe* in other types of molecular genetic techniques. A second reason is to construct DNA libraries, a topic discussed later in this chapter. Finally, cloning is also used for therapeutic genetic engineering. Here, the hope is that the vector with a cloned copy of the "good" gene missing in a patient might be incorporated into the patient's cells and produce the missing protein or enzyme.2

Basic Tools: Probes

A probe is a segment of single stranded DNA or RNA with a known nucleotide sequence. The probe is either synthesized or cloned in a laboratory in very large amounts and is then placed into physical contact with human DNA that has been treated to become single-stranded.3 Because of complementary base pairing, the probe will bind to the DNA segment that contains the specific nucleotide sequence that complements the probe's sequence.4. An example is given in Figure 7.2.

Basic Tools: Light bulbs

Purified DNA and RNA resemble viscous water. If many small sections of single-stranded DNA are subjected to gel electrophoresis, the sections will migrate to the opposite

---

1 A plasmid is a very small, circular section of DNA that lives inside a bacterium. Generally, the size of the vector places limits on the length of the DNA segment that can be cloned. Plasmids and virus, being the smallest, can be used for segments up to 20 kb. Yeast, the largest, can clone vectors of over one million bases. The procedure for cloning human DNA and inserting the segments into vectors is called *recombinant DNA technology*.

2 The logic of such genetic engineering is straightforward, but there are huge technical issues to surmount before this technique makes it into clinical practice.

3 The process of making double-stranded DNA into single-stranded DNA is termed *denaturing*, and the resultant single-stranded DNA is called *denatured* DNA. The simplest method to denature DNA is to heat it.

4 The process of a single-stranded probe binding to its complementary single-stranded DNA sequence is termed *hybridization* or *annealing*.

side of the gel according to their size, but they will be invisible. If probes are added, they will bind to the appropriate complementary snippets of the single-stranded DNA, forming a double strand. But the probes, being composed of nucleotides, will also be invisible.

To locate the probe after it binds to the DNA, it is necessary to engineer the probe so that it carries the biological equivalent of a light bulb. Two major types of light bulbs are used in molecular genetics. The first type is constructed by labeling the probe with radioactive isotopes.5 The second type of probe uses special fluorescent dyes that allow the probe to be visualized and photographed under specialized lighting conditions. For many types of DNA analyses, radioactive probes are still the method of choice, but fluorescent technology is quickly replacing it.

Basic Tools: Restriction Enzyme

A restriction enzyme6 is an enzyme that recognizes a specific nucleotide sequence and cuts DNA at the sequence. For example, the restriction enzyme EcoRI (for E. Coli Restriction enzyme number I)7 recognizes the sequence `GAATTC` and slices the DNA right after the `G`.

Restriction enzymes are used in a wide variety of techniques. One major advantage is that they can cut DNA (or "digest" DNA, as the microbiologists prefer to call it) into fragments of manageable lengths. Without digestion using restriction enzymes, human DNA segments would simply be too long to allow them to be cloned or subject to many kinds of electrophoresis. Restriction enzymes also play an important role in detecting human polymorphisms.

---

5 The generic name for this procedure is *autoradiography*. Through a complicated series of steps, the molecules in the gel are transferred onto a special paper that is then placed upon X-ray film. The radioactive probe will expose the X-ray film, revealing its location.

6 Aka restriction endonuclease.

7 The peculiar naming of restriction enzymes may be traced to their initial discovery in bacteria. Hence, the enzymes are named after the species in which they were found and numbered roughly in order of their discovery.

Basic Tools: Polymerase Chain Reaction (PCR)

The polymerase chain reaction or PCR is a technique used to "amplify" DNA—i.e., make a sufficient number of copies of a DNA segment to permit it to be used for other types of techniques. Many people are familiar with the forensic application of PCR. When only a tiny drop of blood, semen, or other biological specimen is available at a crime scene, PCR is used to make a sufficient amount of DNA to permit genotyping.

PCR methodology makes use of many of the concepts outlined above, so it will be explained in some detail (see Figure 7.3). The procedure begins with purifying DNA from a biological specimen and then heating it almost to the boiling point of water. The heat separates the double-stranded DNA into two single-strands. Once the DNA has become single-stranded, large amounts of specialized probe8 are added to the mixture along with an enzyme9 and a large number of free nucleotides. The probe binds to the DNA and then the enzyme synthesizes a complementary strand to the DNA beginning with the end of the probe and continuing to the end of the DNA segment. If we began with a single copy of DNA, then the end of this process will result in two copies of the DNA molecule. By repeating the process of heating, bathing the DNA with probes, and adding the enzyme and nucleotides, we would now have four molecules of the desired DNA segment. Continuous repetition results in a geometric progression of DNA copies—8, then 16, then 32—until a sufficient amount of DNA is available for genotyping. Today, PRC is machine automated and is often done in conjunction with a robot that automatically sequences the amplified DNA. The advantage here is that PCR avoids the complicated laboratory procedures

---

8 The type of probe that is added is specialized in two ways. First, two probes are added. The first probe will bind to the "right hand" section of the DNA, while the second probe, an exact complement to the first, binds to the "left hand" section. In this way, both single-strands will have the probe binding in the same place. The other specialized feature of the probe is that one side of it is "open ended" in the sense that it is ready to act as a starting point for the synthesis of another strand of DNA. The term *primer* is used to refer to the type of probe used in PCR analysis.

9 The enzyme is called a *polymerase*. Hence the name *polymerase* chain reaction.

necessary to harvest enough DNA required for other types of genotyping.10 Automation also reduces the time and cost of genotyping. The biggest disadvantage of PCR is that the technique is so sensitive that it is susceptible to contamination from other DNA. Hence, careful laboratory protocol must be followed.

[Insert Figure 7.3 about here]

Types of DNA Polymorphisms:

Polymorphisms 1: Blood groups

In the early days of human genetics, the majority of polymorphisms were those associated with proteins in blood. When you blood is typed, you are informed that you are blood group O+ or AB- or A+, etc. The letter in this blood group gives your phenotype at the ABO locus, and the plus (+) or minus (-) sign denotes your phenotype at the Rhesus locus. A number of other loci such as Kell, Duffy, MN, and Kidd can also be phenotyped from blood. These polymorphisms are still used today to assess suitability of donors and recipients for blood transfusions (ABO locus) and to assess Rhesus incompatibility between a mother and her fetus. However, blood group polymorphisms have given way to other, more sophisticated techniques in modern human genetic research.

Polymorphisms 2: Restriction Fragment Length Polymorphism (RFLP)

At this point, it is helpful to describe a genotyping technique that will use all the tools outlined above even though it is not the preferred method for today's research. *The restriction fragment length polymorphism* or *RFLP* was the technology that began the dramatic explosion in genetic technology by allowing the field to locate genes for Mendelian disorders like Huntington's Disease and cystic fibrosis. The procedure is

---

10 Before PCR, cells called lymphocytes were extracted from blood and then infected with a specialized virus that would in essence immortalize the cells. The cells would then be grown in a culture and when enough of them were available, DNA would be extracted to do a genotype.

illustrated in Figure 7.4. First a probe must be constructed that will bind with a known sequence of DNA. The sequence is the gene that we want to examine. The probe for this example was previously given in Figure 7.2.

[Insert Figure 7.4 about here]

The DNA of an individual—I will use myself as the example—is then purified and the bonds connecting the two strands of the DNA molecule are cut, making the DNA single stranded. I happen to be a heterozygote at the locus at which the probe will bind. The difference in the alleles is subtle, but it appears in the middle nucleotide sequence for alleles 1 and 2 in Figure 7.3. Allele 1 contains the sequence `GACTTC` while allele 2 contains the sequence `GAATTC`. The third nucleotide in this series differs. But this sequence is an important one because it is the one that is recognized by the EcoRI restriction enzyme.

Suppose now that we take my DNA and place it into a solution with this restriction enzyme. Panels (b) and (c) show this for respectively my alleles. Allele 2 contains the necessary sequence for the restriction enzyme to cut the gene in the middle (in addition, of course, to cutting it at the beginning and end of the gene). This allele will now have two fragments. The first will begin with the sequence `AATTC` and will end with the `G` close to the middle. The second will begin with the `AATTC` near the middle and will end with the terminal `G`. For allele 1, on the other hand, the EcoRI enzyme recognizes the initial sequence `GAATTC` and cuts the DNA between the `G` and the `A`. EcoRI will also recognize the last (rightmost) nucleotide sequence and cut the DNA after the `G`. EcoRI does not recognize the middle sequence, so it will not cut it there. Hence, allele 1 will contain one very long DNA fragment.

Now let us subject these fragments to electrophoresis. Allele 1, being quite long, will not move much from the start lane. The two fragments from allele 2, however, are

considerably shorter than the single one for allele 1. These two fragments will move much further along the gel, the smaller of the two migrating more than the larger.

We now have to "light up" the invisible DNA strands by bathing everything in the probe.11 The single stranded probe will bind to all three of my DNA fragments—the long one from allele 1 and two shorter ones from allele 2. When the probe "lights up," all three strands will be revealed—see the panel (d) of Figure 7.3. The laboratory now knows my genotype at this locus.

Polymorphisms 3: Tandem Repeat Polymorphisms

Although RFLPs were the first of the modern molecular methods used to detect polymorphisms, they have given ground to other, more discerning techniques. One generic class of polymorphisms has their origin in DNA nucleotide sequences that are repeated a certain number of times, one right after the other. These "tandem repeats" are highly polymorphic in the sense that a large number of alleles may be found at any given locus. For example, one allele at a locus may have the sequence `CAG` repeated four times, another may have it repeated eight times, while yet a third may have it repeated 20 times. Unfortunately, even though the concept of the tandem repeat is quite simple, the terminology for referring to these polymorphisms is quite confusing to the uninitiated.12 For simplicity's sake, we will lump all these fine distinctions together into the single category of tandem repeat polymorphisms.

---

11 Here, I omit several steps to keep the explanation simple. In actuality, after electrophoresis is done, the molecules in the gel are blotted onto a nitrocellulose filter and the filter is then bathed in a solution containing the probe. Excess probe is washed away and the filter is then placed on top of a sheet of X-ray film. After a suitable time, the radioactive label in the probe will expose the film, revealing the locations of its binding. This whole procedure is called a *Southern blot*.

12 The terms *microsatellite DNA*, *simple sequence repeat polymorphisms* (*SSR* or *SSRP*), and *short tandem repeat polymorphisms* (*STRP*) refer to two, three, or four nucleotides repeated in tandem—e.g., CACACA. The number of repeats is highly variable and some tandem repeats such as the CA repeat may occur at several thousand different areas in human DNA. *Minisatellite DNA* refers to a sequence of 20 or more nucleotides that may be repeated up to 100 times. The *variable number of tandem repeats* (*VNTR*) is a technique for detecting minisatellite DNA repeats that are flanked on either end by restriction sites.

Figure 7.5 illustrates this type of polymorphism. The repetitive nucleotide sequence is GAAC, which is contained eight times in allele 1 but only four times in allele 2. The probe in this case contains the repetitive complement CTTG.

[Insert Figure 7.5 about here]

There are several different ways to genotype for tandem repeats. When the number of nucleotides in a sequence is fairly large (e.g., a series of 25 nucleotides is repeated over and over), then separating alleles by electrophoresis is often done. When the number is small (e.g., repeats of two to four nucleotides), then PCR is commonly used. Here, a special probe for the gene of interest is constructed for the area before the repeated sequence. PCR is then done to amplify the DNA that is then placed into a special sequencing machine that counts the number of repeats.

Polymorphisms 4: Single Nucleotide Polymorphisms (SNP)

A single nucleotide polymorphism or SNP is a sequence of DNA on which humans vary by one and only one nucleotide. Because humans by one nucleotide per every several hundred nucleotides, there are likely to be hundreds of thousands of SNPs scattered throughout the human genome.

The major advantage of SNPs, however, lies in the fact that they can be detected in a highly automated way using DNA chips. This avoids the laborious process of using gel electrophoresis and thus many more people and many more loci can be genotyped within a given amount of time. Genotyping begins with PCR to amplify the region containing a SNP. The resulting DNA is then heated to make it single stranded and placed onto a special chip that contains single-stranded DNA and microcircuitry. Specialized computer

---

Historically, VNTRs were used to construct genetic fingerprints, but in reality any polymorphisms may be used for that purpose.

software can then "read" the specific polymorphism.[13] The potential of SNPs is so great that the Human Genome Project has established a special group, the SNP Consortium, to develop the technology and to locate SNPs.

Polymorphisms 5: Gene Sequencing

Nothing surpasses finding the ordering of the nucleotide sequence for a gene, a technique known as *sequencing*, for the detection of polymorphisms. Although this is the ultimate knowledge for the geneticist and will in time overshadow all other types of polymorphism detection, it is too time consuming and expensive to be used anywhere but in the more advanced research centers.

Here, we eschew explanation of sequencing procedures for one major reason. The Human Genome Project is devoting large amounts of its resources to automating the process of sequencing, so it is difficult to predict the methods of choice by the time these words hit print. Suffice it to say that within the next few decades, the computerized laboratory robots now available to the well-funded genetics laboratory will encounter the dramatic price discounts witnessed by the electronic calculator and later, the personal computer. Every hospital—as well as many laboratories researching individual differences in human behavior—will eventually have a "DNAnalyzer" sitting in the corner of a room spurting out sequence information about genotypes.

Other Molecular Techniques.

In Situ Hybridization

In situ (Latin for "on site") hybridization is a technique used with whole chromosomes to: (1) find which chromosome a gene of known nucleotide sequence is located on; or (2) determine if a section of chromosome has been deleted or duplicated. The

---

[13] The situation is actually more complicated and involves extension of the DNA strand that hybridizes to

technique begins by preparing chromosomes much in the same way that is done for a karyotype (see Chapter 8). The chromosomes are then placed into a solution with a large amount of either radioactively labeled or fluorescent probe.14 Combinations of chemicals and heat are used to split the double-stranded DNA in the chromosomes into single-stranded DNA, permitting the probe to bind with the DNA. Excess probe is washed away and the chromosomes are then stained. If a section of a chromosome "lights up" then we know that it has the nucleotide sequence complementary to that of the probe.

Mutational Screening

Once a probe for a section of a protein-coding gene is identified, the curious geneticist can hardly resist the temptation to see if patients with a particular disorder have an irregularity at the locus. A classic example would be schizophrenia and the genes for dopamine receptors. For a long time, it was known that many drugs that diminish the florid hallucinations and delusions of the schizophrenic had important influences on dopamine receptors. Hence, when the gene for a dopamine receptor was first cloned, it was a natural matter to question if schizophrenics differed from nonschizophrenics in this gene.

Except for direct sequencing, many of the techniques outlined above are not suitable for screening the genotypes of patients. For example, RFLPs will only detect DNA differences at a restriction site, and the chances are quite remote that the aberration in the schizophrenic dopamine receptor allele happens to be at just this site. Several other techniques, collectively known as mutational screening, are preferred in this circumstance.

The two most widely developed mutational screens use souped up versions of electrophoresis to find small and subtle differences in DNA. One technique uses a gel that varies in temperature from the top to the bottom of the gel.15 As the DNA moves through

---

the chip using special color-coded nucleotides that stop DNA synthesis.

14 When a fluorescent probe is used the procedure is called FISH for fluorescence in situ hybridization. FISH is rapidly becoming the method of choice for in situ hybridization.

15 This procedure is called denaturing gradient gel electrophoresis (DGGE).

the gel, the increased temperature will cause it to denature into its single strands and the resolution of the procedure permits detection of small sequence differences between a patient and a control. Once such a difference is detected, then the patient's DNA is sequenced to find the exact nucleotide variation(s).

The second technique is conceptually identical with the exception that the gel does not have a temperature gradient.16 It relies on the fact that small differences in the nucleotide sequence between two strands of DNA can be resolved in different bands on a specialized gel. Again, the patient's DNA is sequenced after a difference is found.

Finding the Gene for a Trait

Having surveyed the various tools and techniques of modern DNA genotyping, we now present the logic of going about the labor-intensive task of actually finding a gene for a trait. The first task is to make certain that the trait has some genetic influence on it. In the case of simple Mendelian disorders, the rarity of the disorder and the risk to different classes of relatives are usually sufficient to implicate a single gene. More complicated methods must be used for complex phenotypes. These complications are treated in later chapters. Here, to keep matters simple, let us assume that we want to locate the gene for a dominant disorder.

The first challenge is to identify the chromosome on which the gene is located. Without knowing either the gene or its product, it is necessary to use a statistical procedure called *linkage analysis* to achieve this goal. Linkage begins with a host of genetic loci called *markers* or *marker loci*. A marker locus is a polymorphic locus with a known location on a chromosome. Linkage does not try to locate the precise position of the disease gene. Rather, it tries to identify which marker loci are close to the disease gene. In this way, a rough area of a particular chromosome will be targeted for further analysis. With current

---

16 This technique is known as single-strand conformation polymorphism (SSCP).

linkage maps, a disease gene can be located to anywhere within 1Mb (million base pairs) to 2Mb of its exact location. The major advantage of new technology is that it has greatly expanded the number of polymorphic loci that can be used as markers in a linkage study. Linkage is treated in greater detail in Chapter 10.

Before the human genome was fully sequenced, positive linkage findings were followed up by a procedure called *positional cloning* or *gene walking* to identify the location of the disease gene.17 Because the actual DNA sequence is now known, contemporary researchers power up their computers and download the nucleotide sequence for the region with the positive linkage. They then try to locate genes within this region by looking for promoter regions, initiation codons, stop codons, etc.

Advanced Topics: How the Human Genome was Sequenced

Earlier descriptions of molecular genetic techniques liberally omitted many important steps in order to focus on conceptual issues. Here, the major technique for sequencing DNA in the human genome project is described in some detail. It can help to give you an appreciation of the laboratory techniques used in modern molecular genetics.

As with all human molecular technique, sequencing begins with biological material, almost always blood. Through a series of steps that need not concern us, the DNA from the blood cells is extracted and purified. A very small sample of this DNA is then used for analysis.

The physical process of extraction and purification breaks the DNA at large numbers of random places. However, very long fragments of DNA—too long to work with—still remain. Hence, the DNA is often bathed with a restriction enzyme to cut it into fragments of more manageable size. PCR is then used to amplify the region to be sequenced.

---

17 The term *reverse genetics* was used in the past to refer to positional cloning.

The first steps in sequencing are the same as those in PCR (Figure 7.6). The DNA is heated so that the double-stranded helix breaks apart into two single strands (a process called *denaturing*). Then a primer—the identical one used in the PCR—is added. This primer is a small section of single-stranded nucleotides that will bind with its complementary section of DNA that was previously amplified in the PCR.

[Insert Figure 7.6 about here]

The next step is to synthesize the rest of the double-stranded DNA starting with the primer. Essentially, this is the same process as DNA replication but here it is performed in the test tube instead of the cell nucleus. The two major ingredients for replication are: (1) a large number of free nucleotides; and (2) a polymerase enzyme that will build the chain.

[Insert Figure 7.7 about here]

Figure 7.7 shows the DNA with its nucleotide "soup." The majority of nucleotides in the soup have the same chemical composition as the nucleotides in ordinary DNA. The clever trick in this step of the process is the addition of a small amount of specially engineered nucleotides that have two key features. The first is that these nucleotides are chain terminating[18]. That is, whenever one of these special nucleotides gets entered into the chain, the process of building the double helix stops dead in its tracks. The second important property is that each special nucleotide is "color-coded" through a chemical tag so that it will fluoresce into a specific color when exposed to the appropriate light conditions. Each type of these special nucleotides is given a different color—e.g., green for adenine, yellow for thymine, etc.

[Insert Figure 7.8 about here]

Next, a large number of polymerase molecules are added (Figure 7.8). The polymerase is a complicated enzyme that ordinarily acts to replicate DNA in the cell after the hydrogen bonds are split and the double helix is separated into two single strands. To

imbue polymerase with a sentience that it really lacks, one could say that the molecule grabs free nucleotides, inspects the next nucleotide in the single-stranded chain, and then "glues" the appropriate nucleotide partner into the other DNA strand. For example, if the next nucleotide on the single strand is A, then the polymerase will place a T on the growing strand.

The next step is simply to wait and let nature take its course. Because there are millions of single-stranded DNA fragments with primer attached to them, millions of polymerase molecules, and several gazillion nucleotides, many thousands of double-stranded DNA molecules will be synthesized. However, these complementary DNA strands will be of different lengths because of the chance incorporation of special nucleotides. Whenever one of these special nucleotides is placed into the chain, further synthesis of the DNA strand stops. The net result is many millions of copies of double-stranded DNA, but all of different lengths (Figure 7.9).

[Insert Figure 7.9 about here]

If the DNA mixture is now heated, the double-stranded DNA will break down into its single strands, giving a large number of single-stranded DNA molecules of various lengths (Figure 7.10). The next step is to load the single-stranded DNA onto an electrophoretic gel. The new forms of electrophoresis are so sensitive that they can detect the difference between two DNA strands that differ in length by only one nucleotide. The result after electrophoresis is completed is depicted in Figure 7.11, where the fragments would have been loaded on the top of the figure.

[Insert Figures 7.0 and 7.11 about here]

When the gel is viewed under the appropriate lighting, the bands will fluoresce. Because the special nucleotides are color-coded, simply reading the sequence of colors gives the nucleotide sequence of the DNA (Figure 7.12).

---

[18] These chain-terminating nucleotides are called *dideoxy* nucleotides.

[Insert Figure 7.12 about here]

This type of sequencing procedure is now highly automated. There are specialized sequencing machines that are essentially computerized robots that perform the automated tasks of timing, heating and cooling, and pipetting mixtures. The newer models also contain a special capillary tube that permits the electrophoresis to be done automatically. Laser lighting and optical scanning allows the colors to be read by a computer instead of a human observer. All the data are processed by specialized software that analyzes the sequence, flags areas of uncertainty, and stores the data.

At this point, we now have thousands of small sections of DNA sequenced. The next step is to put these sections together to get the sequence for the whole human genome. The key here reverts back to the original splitting of the DNA into manageable sizes for sequencing. Think of using many different types of restriction enzymes to cut the DNA. Because one type will cut the DNA at one nucleotide sequence but another will cut the DNA at a different sequence, the resulting sections will have some overlap. By testing which sequences overlap with others, one can reconstruct the whole sequence.

[Insert Figure 7.13 here]

Figure 7.13 illustrates the process. Three separate sections of DNA are depicted in panel (a). Two of the sections contain the nucleotide sequence `ATGCC`, one at the end of its sequence and the other at the beginning of its sequence. Hence these two must overlap. Two of the sections also contain the nucleotide sequence `AATGC`, again one at the beginning and the other at the end. These two sections must also overlap. The result of placing these sections together gives panel (b) in Figure 7.13. The problem, then, is analogous to solving a very large, one-dimensional jig saw puzzle.

In reality, the DNA segments are much larger and the number of overlapping nucleotides is greater than those depicted in Figure 7.13. The DNA nucleotide sequences

for the segments are stored in a computer and powerful search algorithms are used to piece them together.

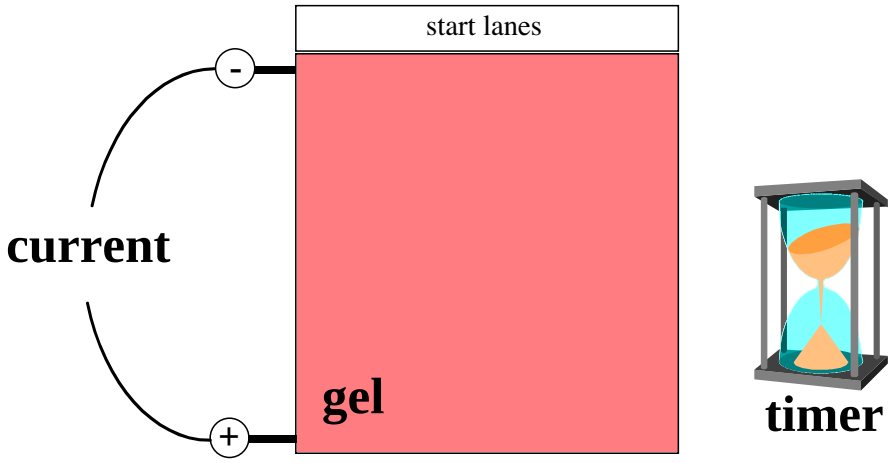


Figure 7.1. Gel electrophoresis. Biological material (DNA, proteins, etc.) is placed in the start lane and an electrical current is turned on for a specified amount of time. Molecules will migrate to the opposite end of the gel according to their molecular weights. Smaller molecules will move further than larger ones.

Figure 7.2. An example of a probe. Probes are manufactured segments of single-stranded DNA that carry a "lightbulb," usually a florescent tag, and bind to its complementary single-stranded DNA sequence.

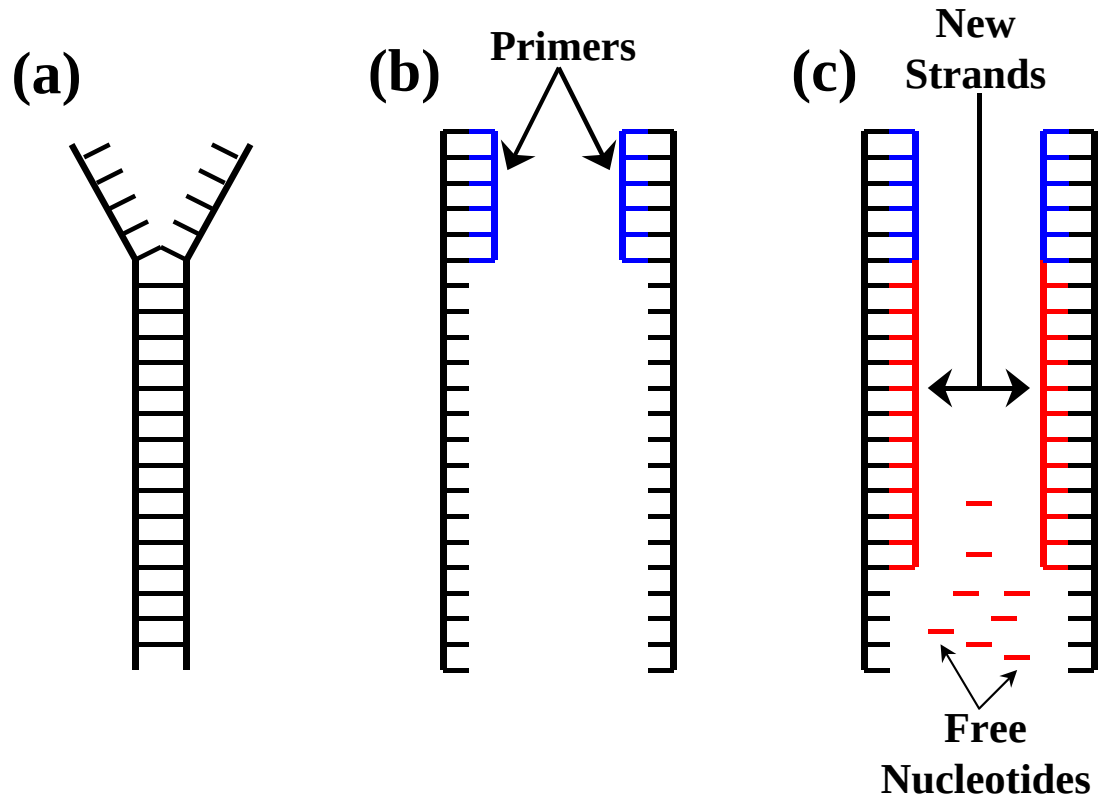


Figure 7.3. PCR: the polymerase chain reaction. (a) With heating DNA becomes single-stranded; (b) a primer section of DNA is added; it will bind with its complementary base pairs on the original DNA; (c) free nucleotides and enzymes are added that will synthesize a new chain starting at the end of the primer.

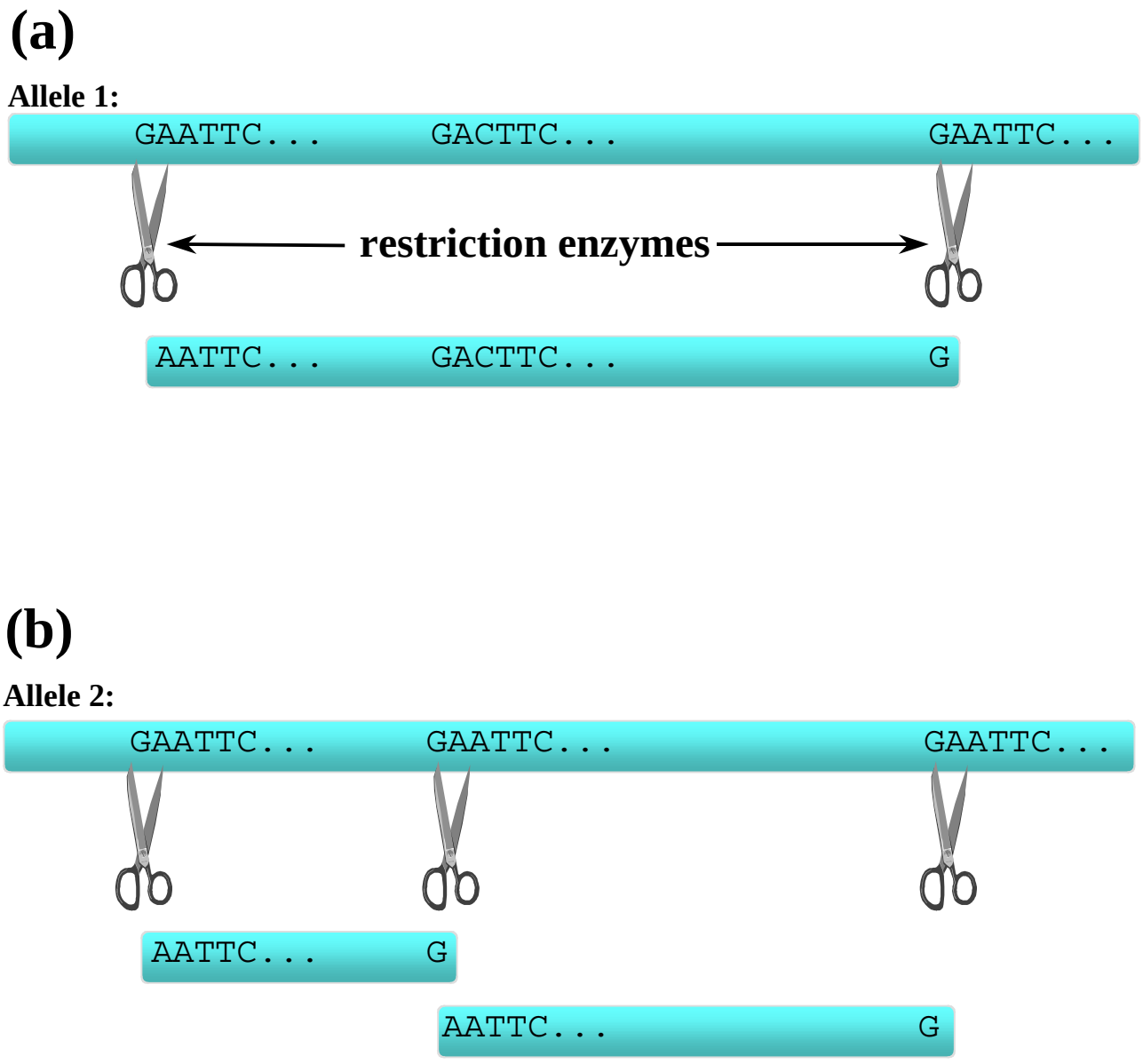
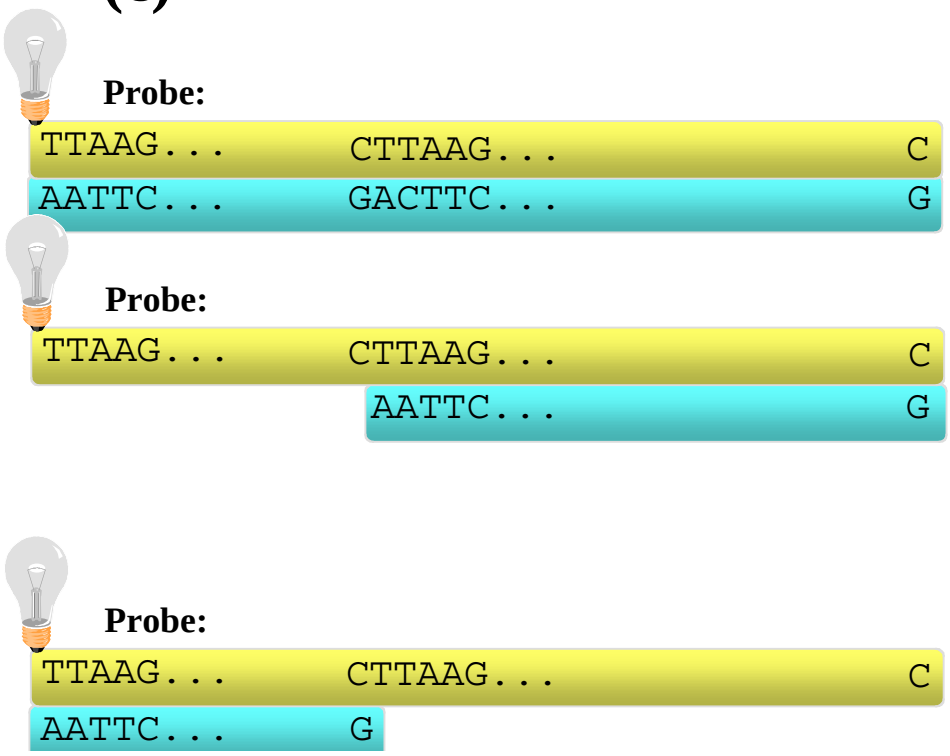


Figure 7.4. The restriction fragment length polymorphism (RFLP). Panel (a): Allele 1 has two restriction sites; incubating this allele in the restriction enzyme will result in one long fragment. Panel (b): Allele 2 has three restriction sites; incubating this allele with the restriction enzyme results in two fragments.

(c)


Probe:
TTAAG... CTTAAG... C
AATTC... GACTTC... G
Probe:
TTAAG... CTTAAG... C
AATTC... G
Probe:
TTAAG... CTTAAG... C
AATTC... G

(d)

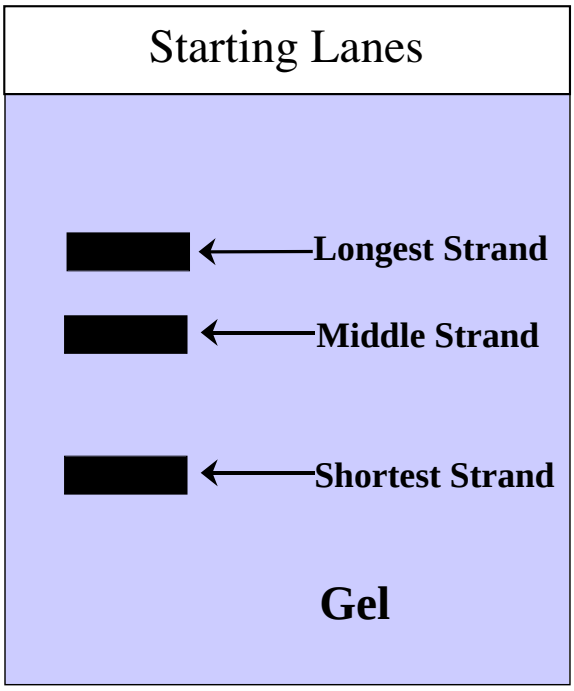

Starting Lanes
Longest Strand
Middle Strand
Shortest Strand
Gel

Figure 7.4. The restriction fragment length polymorphism (RFLP). Panel (c): After the restriction enzyme cuts the DNA, the DNA is subjected to electrophoresis that separates the fragments according to size. A probe is then added to bind to the gene of interest. The binding of the probe after electrophoresis is illustrated in this panel. Panel (d): The gel is then viewed under special lighting (if the probe's light bulb is florescent) or photographic film is placed on it (if the probe is radioactive). The result will be the characteristic bands illustrated in this panel.

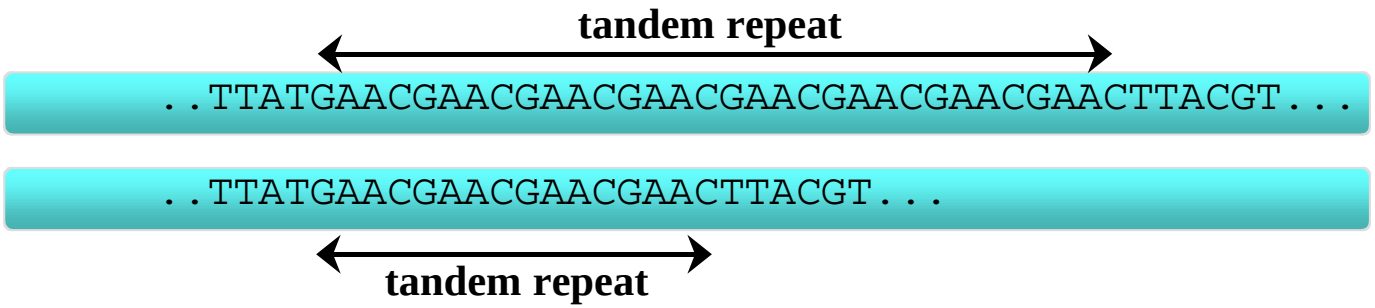


Figure 7.5. Example of a tandem repeat polymorphism. The polymorphism consists in the number of times that a nucleotide sequence is repeated, GAAC in this case.

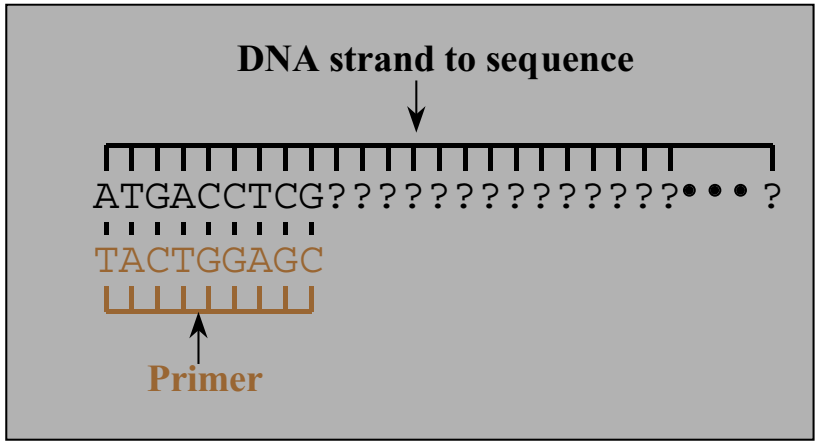


Figure 7.6. Dideoxy Sequencing, Step 1. Heat the DNA to make it single stranded and add a primer. The primer binds to its complementary sequence in the DNA.

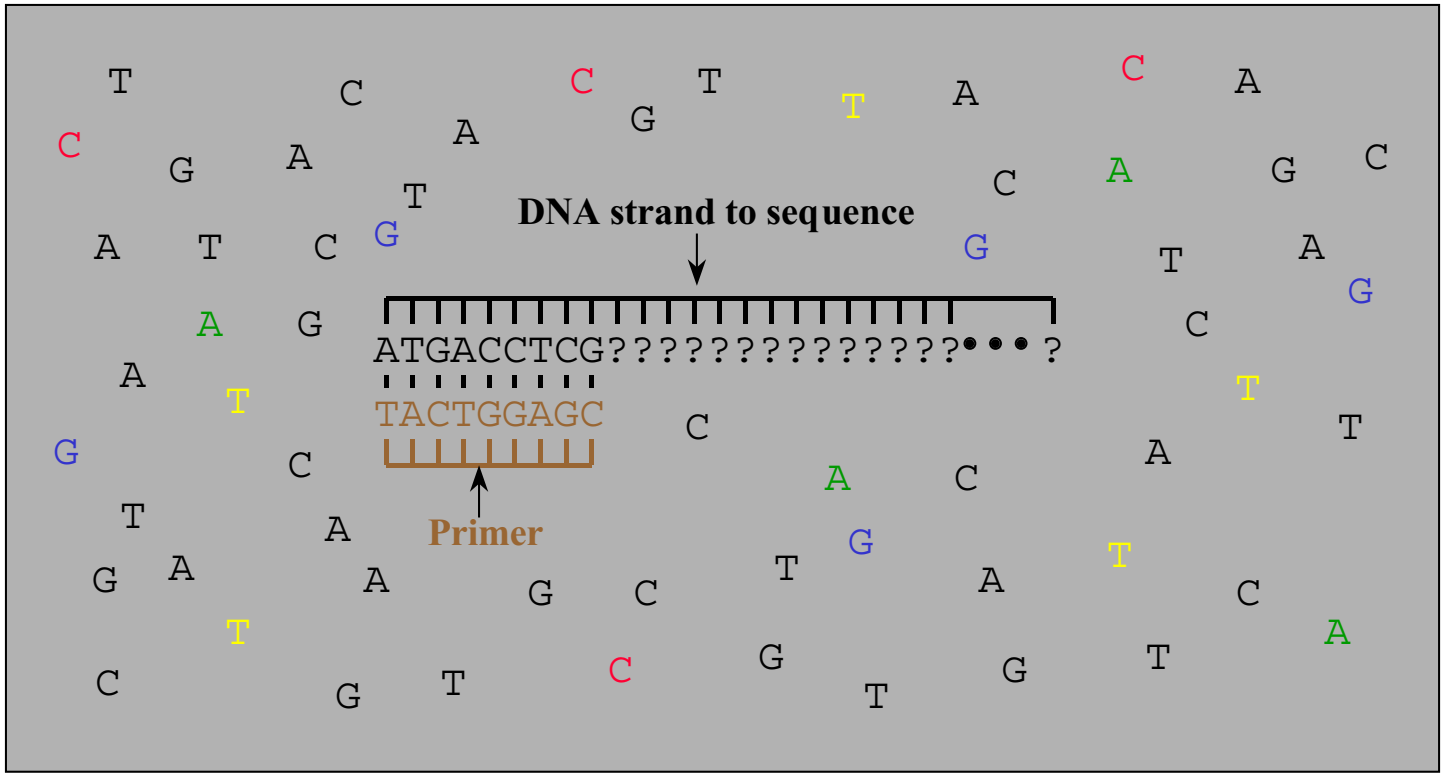


Figure 7.7. Dideoxy Sequencing, Step 2. Add nucleotide alphabet soup. Two types of nucleotides are in the soup. The first (black letters) are ordinary nucleotides. The second (colored letters) are special nucleotides that have two important properties: (1) they will halt the synthesis of the DNA strand whenever they are incorporated into it, and (2) they will fluoresce when viewed under the appropriate lighting.

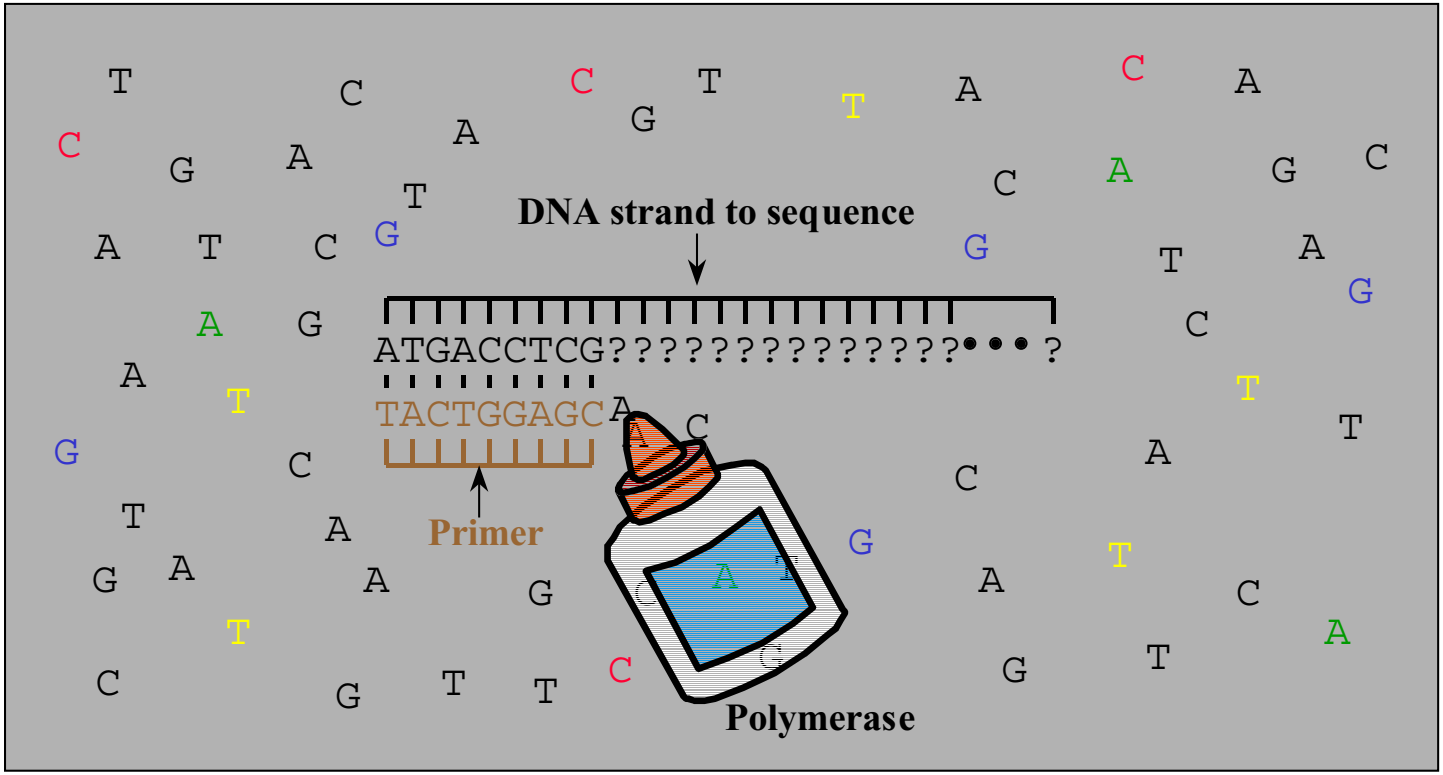


Figure 7.8. Dideoxy Sequencing, Step 3. Add the polymerase (an enzyme that adds free nucleotides to the primer strand). The polymerase will "grab" free nucleotides and add the appropriate one to extend the strand.

Figure 7.9. Dideoxy Sequencing, Step 4. Complementary strands will be synthesized, but they will be of different lengths depending on where the colored nucleotide is incorporated. Eight examples are given below.

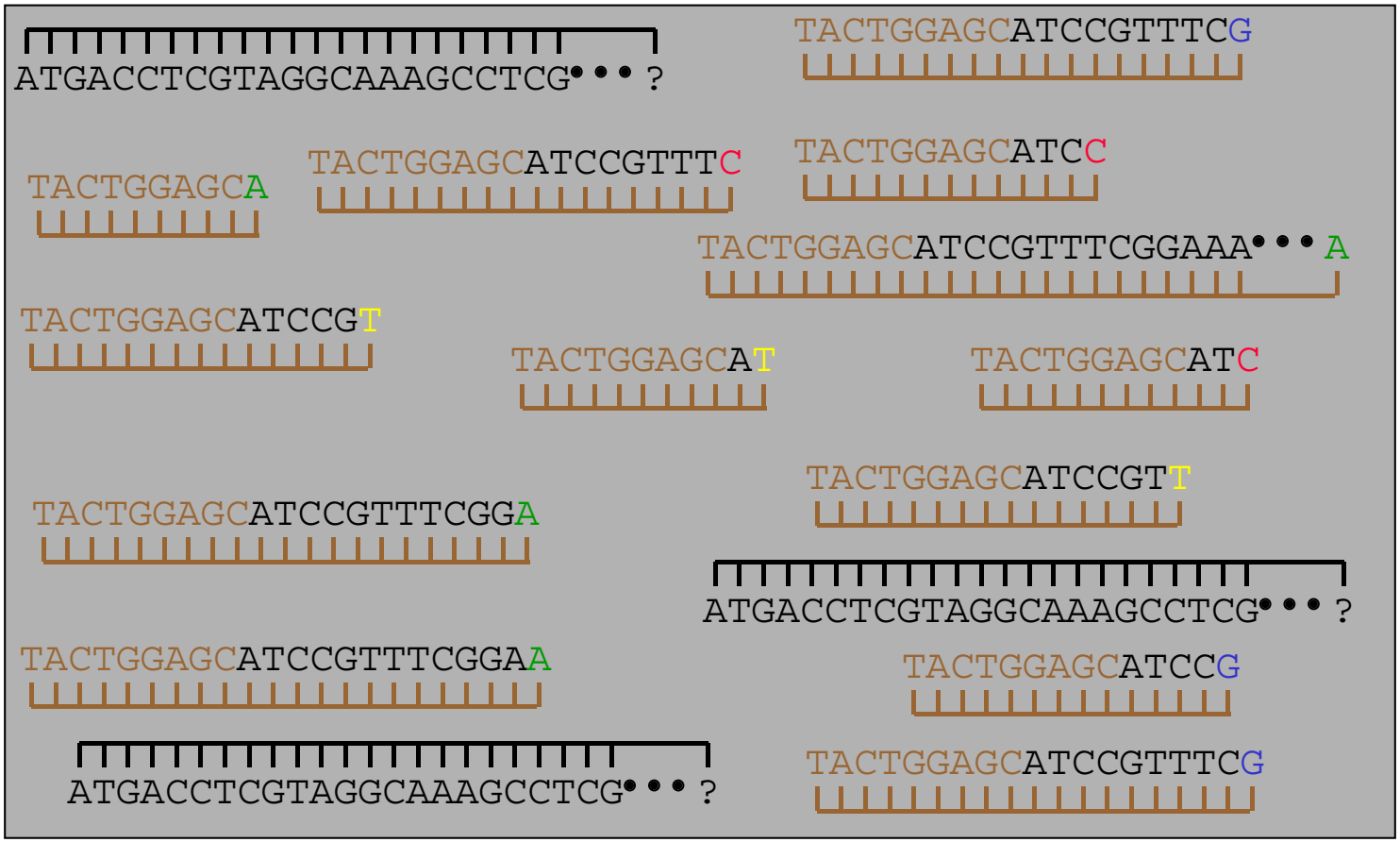


Figure 7.10. Dideoxy Sequencing, Step 5.Heat the DNA to make it single-stranded. There will be many copies of the template strand and also many copies of different length of the synthesized strands.

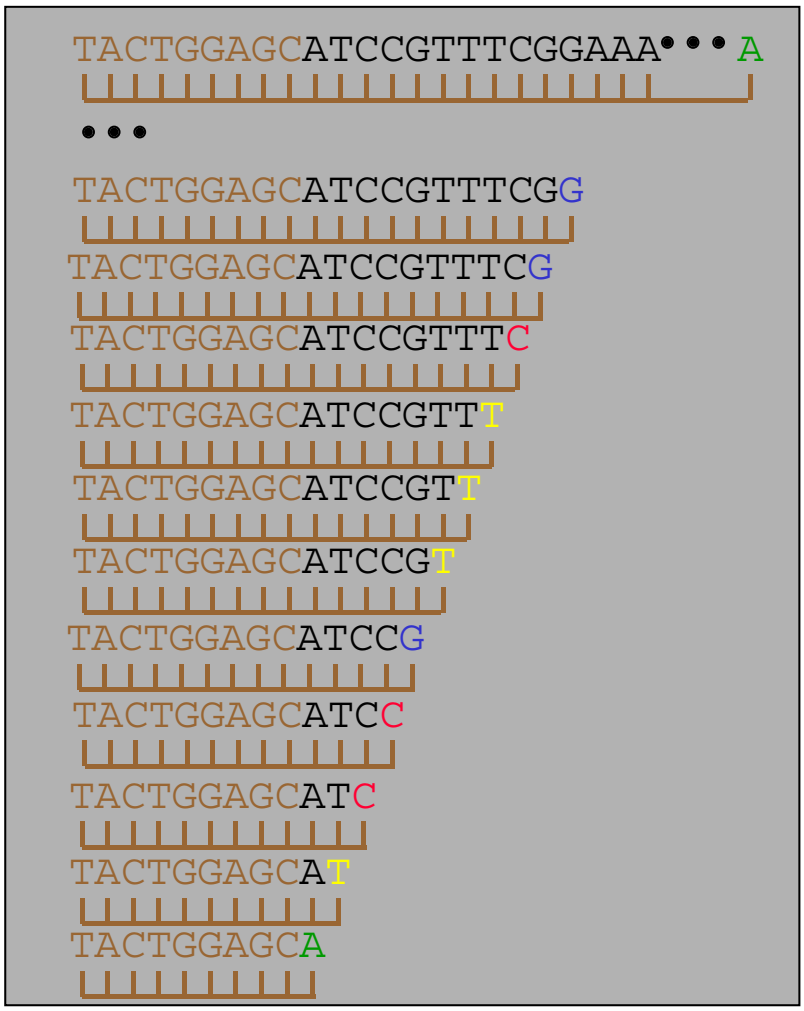


Figure 7.11. Dideoxy Sequencing, Step 6. Use electrophoresis to separate the strands according to size.

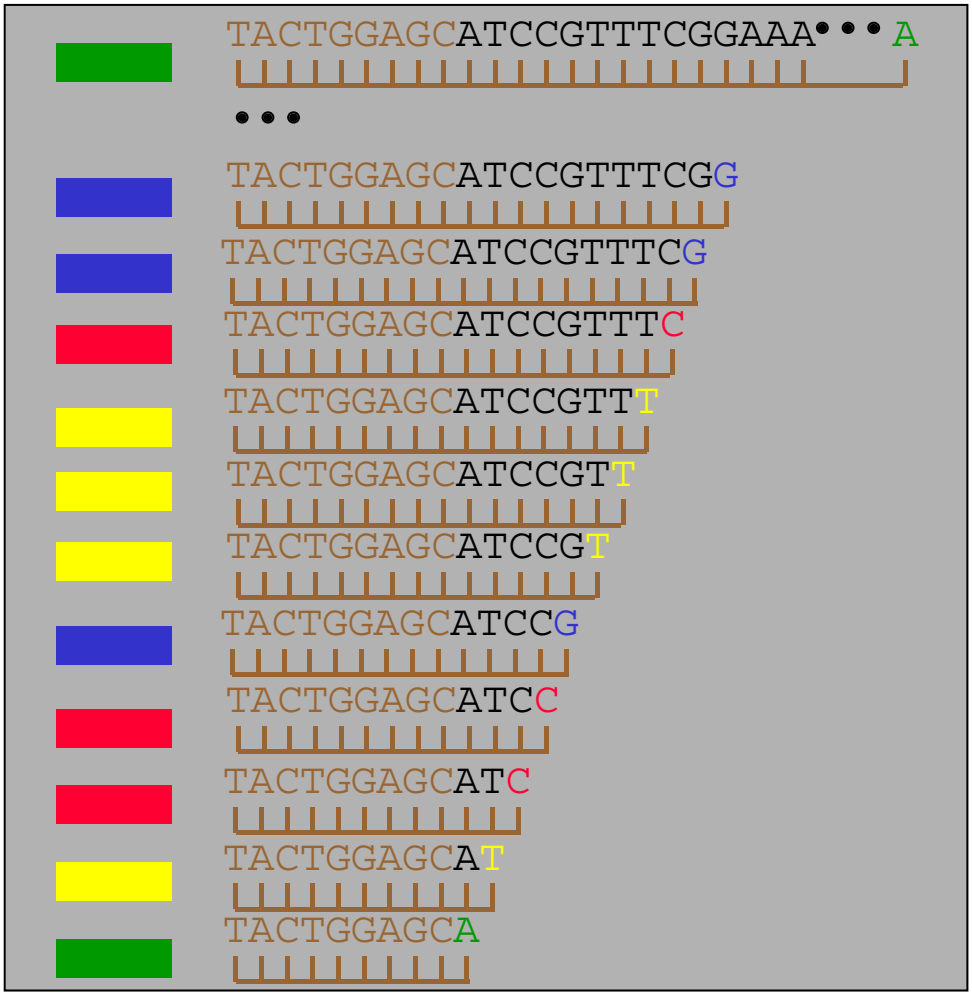


Figure 7.12. Dideoxy Sequencing, Step 7. Viewing the gel under a special light allows the colored nucleotides to fluoresce. This lights up the band. The color-coding permits the DNA sequence to be read.

Figure 7.13. How the human genome was sequenced. (a) Restriction enzymes are used to cut the into a large number of pieces that are then individually sequences; (b) the sequence data for the individual pieces is entered into a computer where powerful software algorithms finds overlapping nucleotide sequences. The actual size of the individual fragments is much larger that those depicted in the figure..