Student Manual

Crime Scene Investigator PCR Basics Kit

You are about to conduct real world forensic DNA profiling. As a crime scene investigator, you will use the polymerase chain reaction (PCR) and agarose gel electrophoresis to analyze the DNA samples obtained from a hypothetical crime scene and four suspects. Your job is to identify the perpetrator. In this analysis, a genotype is the particular set of genetic markers, or alleles, in a DNA sample. Every person's genotype is their own uniquely personal genetic barcode. In this experiment, you'll be revealing the genetic barcodes of several individuals, and looking for a match.

How can DNA evidence solve crimes?

DNA profiling refers to the use of molecular genetic methods used to determine the genotype of a DNA sample. This powerful tool is routinely used around the world for investigations of crime scenes, missing persons, mass disasters, human rights violations, and paternity. Crime scenes often contain biological evidence (such as blood, semen, hairs, saliva, bones, pieces of skin) from which DNA can be extracted. If the DNA profile obtained from evidence discovered at the scene of a crime matches the DNA profile of a suspect, this person is included as a potentially guilty person; if the two DNA profiles do not match, the individual is excluded from the suspect pool.

A Brief History of Forensic Analysis

Forensic sciences describe the boundary between science and the law. Forensic science can as easily convict someone of a crime as free someone wrongly convicted. The earliest uses of forensic science for criminal investigations involved the use of photographs to document crime scenes. Fingerprint evidence has been in use for the past 100 or so years. The first genetic evidence to be collected for investigative work involved the use of blood group typing. The 1980's saw the first use of a DNA-based forensic test, restriction fragment length polymorphism analysis, or RFLP. Although RFLP analysis has its limitations, it has been the workhorse of forensic analysis for nearly 20 years. Only with the recent advent of PCR has this aspect of the criminal justice system become truly modernized. Modern forensic DNA profiling makes it possible to distinguish any two people on the planet (with the exception of identical twins), living or dead.

PCR is DNA replication gone crazy in a test tube

PCR produces large amounts of a specific piece of DNA from trace amounts of starting material (template). The template can be any form of double-stranded DNA. A researcher can take trace amounts of DNA from a drop of blood, a single hair follicle, or a cheek cell and use PCR to generate millions of copies of a desired DNA fragment. In theory, only a single template strand is needed to generate millions of new DNA molecules. Prior to PCR, it would have been impossible to do forensic or genetic studies with this small amount of DNA. The ability to amplify the precise sequence of DNA that a researcher wishes to study or manipulate is the true power of PCR.

One of the main reasons PCR is such a powerful tool is its simplicity and specificity. The specificity of PCR is its ability to target and amplify one specific segment of DNA a few hundred base pairs in length out of a complete genome of over 3 billion base pairs. In addition, all that is required for PCR is at least one DNA template strand, DNA polymerase, two DNA primers, and the four nucleotide building block subunits of DNA – A, G, T, and C – otherwise known as the deoxynucleotide triphosphates of adenine, guanine, thymine, cytosine, and reaction buffer.

PCR allows forensic scientists to reveal personal details about an individual's genetic makeup and to determine the most subtle differences in the DNA of individuals - from the tiniest amount

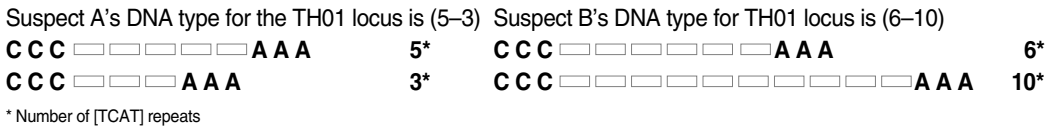
of biological material. The fact that millions of exact copies of a particular DNA sequence can be produced easily and quickly using PCR is the basis for modern forensic DNA testing.

What kinds of human DNA sequences are used in crime scene investigations? There are ~3 billion basepairs in the human genome – greater than 99.5% do not vary between different human beings. However, a small percentage of the human DNA sequence (<0.5%) does differ, and these are the special ***polymorphic*** ("many forms") sequences used in forensic applications. By universal agreement, DNA sequences used for forensic profiling are "anonymous"; that is, they come from regions of our chromosomes (also called ***loci***) that do not control any known traits and have no known functions. Loci are basically genetic addresses or locations. A single ***locus*** may have different forms or types; these different forms are called ***alleles***. A locus may be bi-allelic, having only two different forms, or it may be polymorphic, as described above.

The DNA sequences used in forensic labs are non-coding regions that contain segments of ***Short Tandem Repeats*** or ***STR*s**. STRs are very short DNA sequences that are repeated in direct head-to-tail fashion. The example below shows a locus (known as TH01) found on chromosome 11; its specific DNA sequence contains four repeats of [TCAT].

..C C C T C A T T C A T T C A T T C A T T C A..

For the TH01 STR locus, there are many alternate polymorphic alleles that differ from each other by the number of [TCAT] repeats present in the sequence. Although more than 20 different alleles of TH01 have been discovered in people worldwide, each of us still has only two of these, one inherited from our mother and one inherited from our father. For example as shown in figure 9, suspect A has one allele with 6 repeats, and one allele with 3 repeats, giving a DNA profile for the TH01 locus of 6-3.



* Number of [TCAT] repeats

Fig. 9. Two sample TH01 genotypes.

How are STR alleles detected?

The key to DNA profiling is amplification of the copies present in the small amounts of evidentiary DNA by ***polymerase chain reaction (PCR)***. Using primers specific to the DNA sequences on either side of the [TCAT] STR, billions of copies of each of the two original TH01 alleles in any one person's DNA type are synthesized in the reaction. These copies contain the same number of STRs present in the original DNA copies and can be visualized using agarose gel electrophoresis. By comparison with a DNA size standard, or allele ladder, that corresponds to the known sizes of TH01 alleles, the exact sizes of the PCR products from the sample DNAs can be determined and compared.

A diagram of the results for TH01 typing of Suspect A and Suspect B is shown in figure 10. In this cartoon example, PCR has been performed on DNA from 2 suspects using primers specific for the TH01 locus. Following gel electrophoresis which separates the PCR products according to their size, the pattern of bands is compared to the Allele Ladder to identify the alleles present in the original samples.

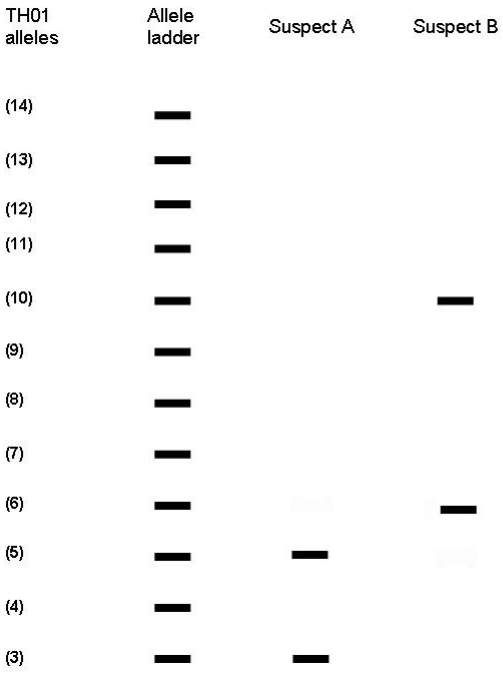


Fig. 10. Illustration of sample TH01 genotypes following gel electrophoresis.

Imagine a scenario in which Suspect A and Suspect B are accused of being involved in a love triangle and committing the murder of a third person in the Highway Motel; the person who actually pulled the trigger is unknown (for more information on this scenario, see the next page). In addition to DNA samples from the crime scene, the forensic specialist will isolate DNA from suspects, victims, and any others present to genotype as controls. Using PCR-based analysis, the samples will be examined at 13 different genetic locations, or loci, using software to interpret the results from the amplification products. In real crime scene analysis, DNA profiling is performed at many loci to improve the ***power of discrimination*** of the testing. In simple terms, the power of discrimination is the ability of the profiling to tell the genetic difference between different individuals. The larger the number of loci profiled, the more powerful the ability to discriminate.

Imagine the following scenario:

Scene: The Highway Motel, #1 Dark Highway, Nowhere
Setting: Room #13.

The motel manager hears loud voices, a woman screams, and a shot rings out. The manager runs to the window in time to see the receding lights of a car leaving in a hurry. The door to room # 13 hangs open. The manager runs to the open door, to see a man lying face down in a pool of blood. He calls 911. The police arrive, and begin to examine the crime scene. An apparent homicide, but with no obvious clues as to who committed the crime. Or…?

A forensic specialist is called in to examine the crime scene and collect evidence. Even though it looks like the people involved left no evidence behind, the specialist can use laboratory tests that can tell who was at the crime scene from a single drop of blood or a lone hair. Is this a science fiction story, or reality?

Very much a reality. Testing is routinely done in forensic testing labs across the US and in many other parts of the world from only a single cell, and sometimes from samples that are decades old. The reason this is possible is because of DNA. To be able to perform laboratory tests, the specialist needs biological material to work with. Often, there is very little material left at the scene of a crime, and not in quantities that will allow analysis. To get around this problem, the specialist takes advantage of a process that each and every cell in your body uses to divide.

The most important part of any cell's life is when it commits to reproducing itself and dividing. The basic result of any cell division is the creation of two identical daughter cells from one original cell. To ensure that this happens, DNA replication must have a high degree of specificity and accuracy, that is, it must copy DNA exactly. To do so, the enzymes involved in DNA replication use the information already contained in the existing strands to make new DNA copies. This basic idea - the exact copying of DNA from a template - is the basis for a new technology that has revolutionized many areas of science, medicine, and the courts.

PCR allows the forensic specialist to specifically amplify, or copy, any region of DNA that he or she is interested in. PCR is the basis for DNA testing that is currently used in nearly all forensic analysis.

In this experiment, you will perform PCR analysis on a single locus, the BXP007 locus, using template DNAs obtained from a simulated crime scene and a victim. Following PCR, you will run an agarose gel to separate the PCR products, visualize the PCR products, compare them to a simulated ladder of possible alleles for this locus, and assign a genotype for the templates. You will then look to see if any of the suspects' genotype match the crime scene, and see whether you can determine whodunit!

Let's examine the DNA evidence and find out who pulled the trigger.

Student Questions – Introduction

1. What kinds of materials obtained from a crime scene might contain DNA?

2. Why do you need to perform PCR on DNA obtained from a Crime Scene?

3. What might you see if you ran a DNA sample extracted from evidence on a gel before PCR?

4. What is a genotype?

5. What is the difference between an allele and a locus?

6. Why do forensic labs analyse non-coding DNA and not genes?

Lesson 1 PCR Amplification

PCR amplification is DNA replication in a test tube. The portion of the DNA you want to make copies of is called the target sequence. The sample of DNA obtained at a crime scene and the suspect's DNA samples contain the target sequence.

PCR relies on three principles of molecular biology

1. Denaturation - melting double stranded DNA template into single stands
2. Annealing - complementary DNA strand hybridization via DNA primers
3. Extension - DNA strand synthesis via DNA polymerase

Denaturation. Before new DNA synthesis can begin the double stranded DNA template must be unwound and separated into single strands. In cells this is carried out by a family of enzymes. In PCR, heat is used to melt apart – or **denature** – the double stranded DNA template.

Annealing. Before a target region of DNA can be amplified, one must determine short sequences of DNA upstream (at the 5' end) and downstream (at the 3' end) of the target loci region of interest. These areas are then used to make short pieces of DNA, called primers or oligonucleotides, which are complementary to regions upstream and downstream of the target loci region (Figure 11). Primers serve as start and stop points for amplifying the target region of the DNA to be copied.

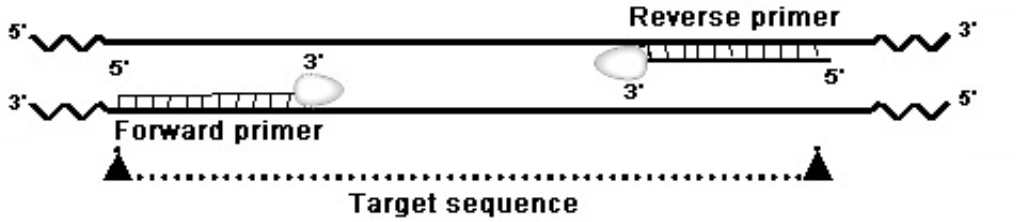


Fig. 11. Primers annealed to a target DNA sequence during PCR.

In PCR, complementary strand hybridization takes place when oligonucleotide primers anneal, or bind, to their respective complementary base pair sequences on the template. Hybridization is the process that describes the binding of the oligonucleotide primer to the template DNA. The two strands anneal to each other, forming a 'hybrid'. Like bookends, the two primers are designed and synthesized in the laboratory with a specific sequence of nucleotides so they will anneal at the opposite ends and on the opposite strands bracketing the target stretch of double-stranded DNA (template strand) to be amplified. Therefore, the target sequence is determined by the location that the primers anneal to.

Extension. Primers are needed because the **DNA polymerase** requires an already existing nucleotide chain to bind and add nucleotides to one at a time. Once the polymerase locates and binds to template DNA and the primer, it initiates the addition of nucleotides and synthesizes new copies of the double stranded template DNA by adding nucleotides onto the primer and extending it. Therefore, primers provide a starting point for the DNA polymerase.

These 3 steps – denaturation, annealing, and extension together make up one PCR cycle. A complete PCR reaction involves many repetitious of a single PCR cycle. In this experiment, your PCR reactions will cycle 35 times.

The enzyme used in PCR – **DNA polymerase** – must be thermally stable because PCR cycles between temperatures of 52°C and 94°C. The thermostable DNA polymerase that performs the polymerization was isolated from a thermophilic bacterium, *Thermus aquaticus* (*Taq*), which lives in high-temperature steam vents such as those found in Yellowstone National Park.

Two template strands are created from the original template after each complete cycle of the strand synthesis reaction – denaturation, annealing, and extension. It is called the polymerase chain reaction because exponential growth of the number of template molecules occurs after each cycle is complete, i.e., the number of DNA copies doubles at each cycle. Therefore, after 35 cycles there will be 2^{35} times more copies than at the beginning. After 35 cycles, the DNA of interest has been amplified sufficiently to be visualized using gel electrophoresis and DNA stains. This allows researchers to determine the presence or absence of the desired PCR products.

In order for PCR to happen efficiently, several components are needed. In addition to the template, the oligonucleotide primers, and the enzyme (*Taq* DNA polymerase), a special reaction buffer is also required, called a **master mix**. The master mix contains all of the components for PCR to occur, including the individual building blocks of DNA (nucleotides, or dNTPs), a special buffer to maintain optimum pH, salts, and $MgCl_2$. Salts and magnesium ions (also known as cofactors) are needed for the *Taq* DNA polymerase to perform optimally. In this experiment, your instructor will provide you with a master mix that comes prepared with all of the ingredients listed above, but also includes colored primers and *Taq* polymerase mixed in. For this reason, it's important to keep the master mix cold before use, so that the enzyme doesn't start to work before you add your DNA templates.

In this part of the experiment, you will obtain DNA samples which have been collected from a crime scene and four individuals suspected of being involved in the crime. Your task is to amplify the region of interest (the BXP007 locus, a polymorphic allele) from the DNA samples. Once complete, you will analyze your PCR products using gel electrophoresis to determine the genotypes of the samples at the BXP007 locus and match the crime scene DNA to one of the suspects.

Student Questions: Lesson 1

PCR Student Questions

1. What does PCR allow you to do with DNA?

2. What components do you need to perform PCR?

3. What is in the master mix and why do you need each component?

4. Why do you need to perform PCR on DNA evidence from a crime scene?

5. What steps make up a PCR cycle, and what happens at each step?

Student Protocol – Lesson One

Student Workstations

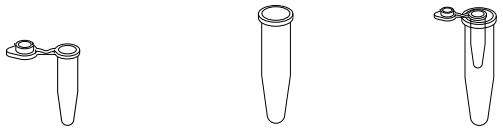
Material	Quantity
Ice bath containing tubes of DNA (as below)	1
Master Mix + primers (MMP, blue liquid)	1
Crime Scene and Suspect A - D DNAs	5 (one of each tube indicated)
PCR tubes	5
PCR adaptors	5
Marking pen	1
2–20 µl adjustable micropipet or fixed volume 20 µl micropipet	1
2–20 µl pipet tips, aerosol barrier	1 rack

1. You will have 6 tubes on ice, and 5 x 0.2 ml PCR tubes in a rack at your workstation.

   In the ice, you should have -
   One yellow tube labeled MMP containing blue liquid.
   5 tubes labeled CS (purple tube), A (green tube), B (blue tube), C (orange tube), and D (pink tube).

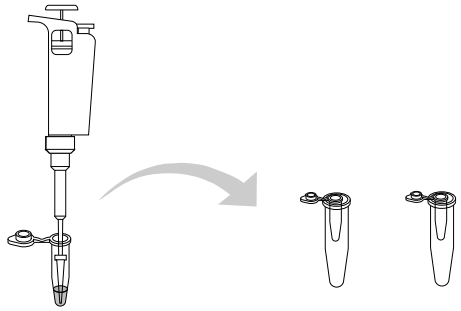
Label PCR tubes CS, A, B, C, and D and include your group name or initials as well. The labels correspond to the following tube contents:

PCR tubes labelled	DNA templates	Master mix + primers (blue liquid)
CS + your initials	20 µl Crime Scene DNA (purple tube)	20 µl MMP (yellow tube)
A + your initials	20 µl Suspect A DNA (green tube)	20 µl MMP (yellow tube)
B + your initials	20 µl Suspect B DNA (blue tube)	20 µl MMP (yellow tube)
C + your initials	20 µl Suspect C DNA (orange tube)	20 µl MMP (yellow tube)
D + your initials	20 µl Suspect D DNA (pink tube)	20 µl MMP (yellow tube)



PCR tube **Capless tube**

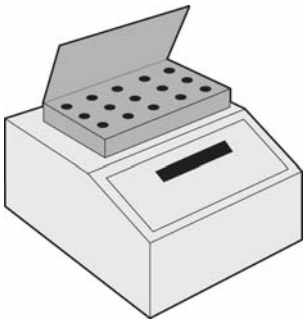
2. Keep tubes on ice during the procedure.
3. Using aerosol barrier pipet tips and either an adjustable micropipet set to 20 µl, or a fixed-volume 20 µl micropipet, add 20 µl DNA to each tube as indicated in the table above. For example, for the Crime Scene DNA, transfer 20 µl of the template into your 'CS' labeled PCR tube. **Important: use a fresh pipet tip for each DNA**.
4. Using aerosol barrier pipet tips and either an adjustable micropipet set to 20 µl, or a fixed-volume 20 µl micropipet, add 20 µl of the Master Mix + primers to each tube as indicated in the table above. Mix the contents of your PCR tubes by gently pipetting up and down. **Important: use a fresh pipet tip each time**. Once you've added MMP to a tube, close the cap. The solution in your PCR tubes should be blue. If it's not blue, talk to your instructor.



Master mix + primers (MMP)

5. Place your capped PCR tubes in their adaptors on ice.

6. When instructed to do so, place your tubes in the thermal cycler.



Lesson Two Electrophoresis of PCR Products

You have completed your PCR amplification. However, at this point, you can't actually tell whether or not you have PCR products. To do this, you must sort your PCR products using gel electrophoresis and then visualize them using a DNA stain. Since DNA is negatively charged, it can be separated using an electric current. In fact, electrophoresis means "carry with current". In agarose gel electrophoresis, DNA is placed in solidified agarose, which forms sieves containing pores that vary in size depending on the concentration of the agarose. The higher the concentration of agarose, the smaller the pore size, and the longer it takes for larger molecules to move through. This is particularly useful when you want to compare DNA molecules of different sizes contained in the same sample. Movement through the gel occurs when an electric current is applied across the gel. Since the gel is immersed in buffer, the current will travel through the buffer and gel, carrying the negatively charged DNA with it toward the positive anode.

In addition to your PCR products, you will also be running a DNA Allele Ladder that represents all of the possible alleles at the BXP007 locus. This is a reference, or marker, that you can compare your PCR reactions to so you can judge their relative sizes and their identities. In the following drawing of a gel, the samples, or bands, seen in the first track, or lane, all come from the BXP007 Allele Ladder. These are the standard sizes of all the alleles know to occur at this locus. There are 8 possible alleles, with the largest at the top of the gel and the smallest at the bottom. The sizes are, from top to bottom, 1500, 1000, 700, 500, 400, 300, 200, and 100 base pairs (bps). Allele names are indicated in the figure. In the next several lanes, we see PCR products that come from DNA samples that have been tested for what alleles they carry at this particular locus. As shown in figure 12, the sample in the lane next to the Allele Ladder, the Crime Scene Sample (CS) has a genotype that corresponds to alleles 5 and 2 on the allele ladder. We would say that the genotype for this sample is 5-2. For the next sample, the genotype would be 7-4, and so on.

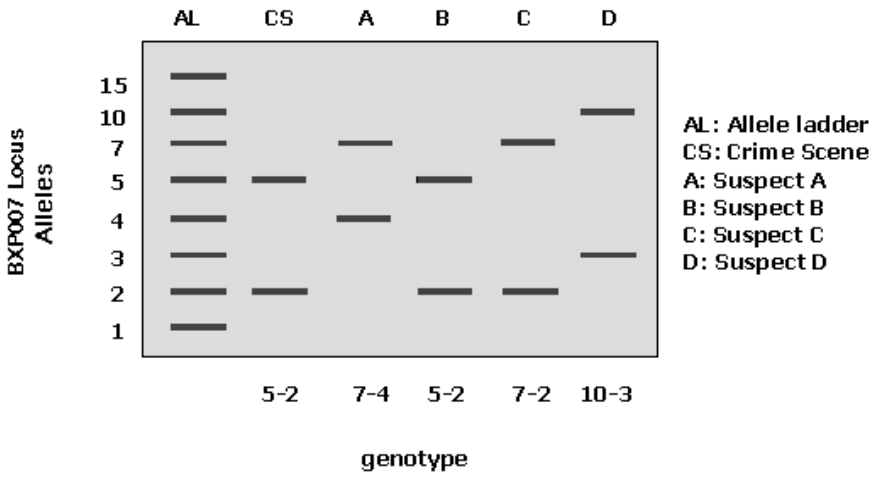


Fig. 12. A cartoon of potential Crime Scene Investigator PCR Basics kit results at the BXP007 locus.

Student Questions – Lesson Two

1. Why does DNA move through an agarose gel?

2. What are the two techniques used to create a DNA profile in this experiment? What function does each perform?

3. What is an Allele Ladder? What is its function in DNA profiling?

4. What is required to visualize DNA following electrophoresis?

Student Protocol – Lesson Two

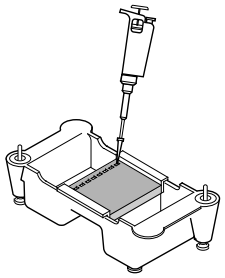
Student Workstation

Material	Quantity
3% agarose gel	1
PCR Samples from previous lab	5
1X TAE running buffer	300–350 ml
Orange G loading dye (LD; orange liquid)	60 µl
Crime Scene Investigator Allele Ladder (orange liquid)	25 µl
Note: Do not confuse these two tubes of orange liquid. They contain different compounds.	
2–20 µl adjustable volume pipet or fixed volume 20 µl micropipet	1
1–20 µl pipet tips, aerosol barrier	1 rack
Gel electrophoresis chamber (may be shared by 2 workstations)	1
Power supply (may be shared by multiple workstations)	1
Fast Blast DNA stain (at common workstation)	1
Gel staining tray	1

Protocol

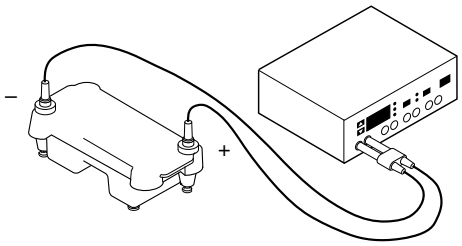
1. Set up your gel electrophoresis apparatus as instructed.
2. Obtain your 5 PCR reactions from the previous lesson, place them into a capless tube adaptor and into a rack.
3. Using aerosol barrier pipet tips and either an adjustable micropipet set to 10 µl, or a fixed-volume 10 µl micropipet, add 10 µl of Orange G loading dye (from the tube labeled 'LD') to each PCR reaction tube and mix well. **Important: use a fresh tip each time**.
4. Using the table below as a guide, load 20 µl of the allele ladder and 20 µl each sample into your gel in the order indicated below.

Lane	Sample	Load volume
1	Allele Ladder	20 µl
2	Crime Scene	20 µl
3	Suspect A	20 µl
4	Suspect B	20 µl
5	Suspect C	20 µl
6	Suspect D	20 µl



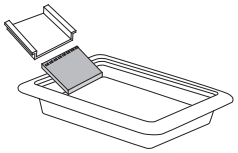
5. Run your gel at 100 V for 30 minutes. Do not let the orange dye front migrate out of the gel.

6. Stain in Fast Blast DNA stain as directed by your instructor and described below.



Staining of Agarose Gels

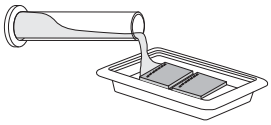
1. When electrophoresis is complete, turn off the power and remove the lid from the gel box. Carefully remove the gel tray and the gel from the gel box. Be careful, **the gel is very slippery**. Nudge the gel off the gel tray with your thumb and carefully slide it into a container for staining.



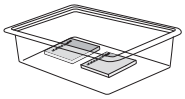
2. There are two protocols for staining your gel. Your instructor will inform you which one you will use.

Protocol 1: Overnight staining (Recommended)

a. Add 120 ml of 1x Fast Blast DNA stain to your staining tray (2 gels per tray).

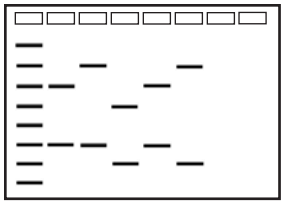


b. Let the gels stain for approximately 4–24 hours with gentle shaking for best results. No destaining is required.



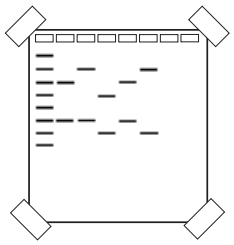
c. The next day, pour off the stain into a waste beaker.

d. Place the gel on a light background and record your result. With a permanent marker, trace the wells and band patterns onto a clear sheet of plastic or acetate sheet.



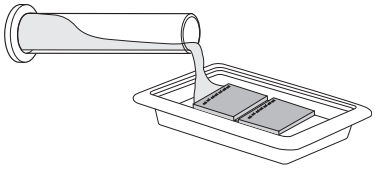
f. Trim away any empty lanes of the gel with a knife or razor blade.

g. To obtain a permanent record, air-dry the gel between cellophane sheets (your instructor will show you how). Tape the dried gel into your lab notebook. Avoid exposure of the stained gel to direct light, since it will cause the bands to fade.



Protocol 2: Quick staining (requires approximately 20 minutes) – This method will allow you to see bands quickly (within 15 minutes), but may require extensive destaining to obtain optimal band-to-background intensity. **Note**: it is important to use warm tap water for the destaining steps of this protocol.

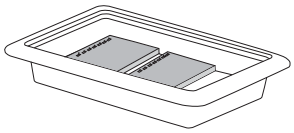
a. Immerse your gel in 100x Fast Blast.



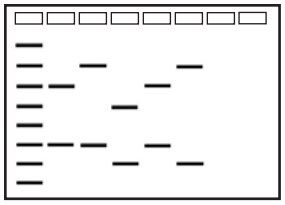
b. Stain the gel for 5 minutes with gentle agitation. Save the used stain for future use. Stain can be reused at least 7 times.

c. Transfer the gels into a large washing container and rinse with **warm** (40–55°) tap water for approximately 10 seconds.

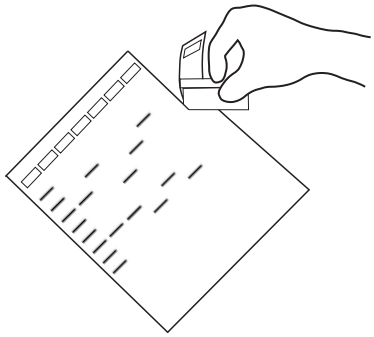
d. Destain by washing three times in **warm** tap water for 5 minutes each, with gentle shaking for best results. You should be able to see bands after 10 minutes with light coming up through the bottom of the staining tray. If necessary, keep destaining in warm water until the desired contrast is reached.



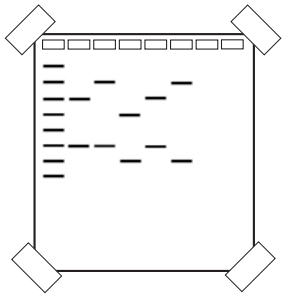
e. Place the gel on a light background and record your result. With a permanent marker, trace the wells and band patterns onto a clear sheet of plastic or acetate sheet.



g. Trim away any empty lanes of the gel with a knife or razor blade.



h. To obtain a permanent record, air-dry the gel on cellophane sheets (your instructor will show you how). Tape the dried gel onto your lab notebook. Avoid exposure of the stained gel to direct light, since it will cause the bands to fade.



Lesson 3: Drying Gels and Analysis of Results

For a permanent record of the experiment, gels can be dried between cellophane sheets and then inserted into lab notebooks. To analyze the wet gels, gels can be scanned, photocopied (a yellow backing improves contrast) or traced onto acetate film. There are 2 methods for drying gels. Please note that agarose gel support film does not work effectively for 3% gels.

GelAir drying frame method

Materials needed for drying gels using the GelAir dryer	Quantity
Cellophane sheets	4
GelAir assembly table	1
GelAir drying frame	2
GelAir clamps	16
GelAir Dryer (optional)	1
Distilled water	500 ml

Procedure

Step 1: Prewet 2 sheets of cellophane in a container of water for 15–20 seconds.

Step 2: Place a plastic drying frame on the GelAir assembly table. Center one sheet of cellophane on the assembly table.

Step 3: Carefully lay your gel on the cellophane, positioning it to accommodate other gels (up to 6 total). If there are bubbles between the gel and the cellophane, gently push them out with your gloved finger.

Step 4: Flood the gels with water and lay the second sheet of cellophane on top of them. If you are drying polyacrylamide gels, try not to trap any bubbles in the sandwich since bubbles will cause cracks in the gel during drying. If there are any bubbles, gently push them out with a gloved finger. You cannot avoid bubbles at the edges of agarose gels since they are so thick, but avoid bubbles between the cellophane and the face of the gel.

Step 5: Place the square metal frame on top of the cellophane sandwich. Secure the eight clamps onto the frame, two on each side. If you are not using a GelAir dryer oven, place the frames upright in a well-ventilated area for 12–36 hours. If you have a GelAir dryer, place up to four drying frames into the oven, turn the heater switch on and set the dial to 3 hours. The dryer will shut off automatically.

Step 6: When the gels are completely dry, they will be flat. Remove the clamps and take the gel/cellophane sandwich from the frame. Trim the excess cellophane surrounding your dried gel with scissors.

Cellophane sandwich and plastic container method

Materials needed for drying gels using plastic containers	Quantity per student workstation
Cellophane sheets	2
Plastic container–Tupperware-type (minimum 6 x 6” opening)	1
Rubber bands	2
Distilled water	500 ml

Procedure

Step 1: Wet two pieces of cellophane in a large volume of water, approximately 500 ml.

Step 2: Place one sheet of cellophane over a plastic container. Pull the cellophane taut so that it makes a flat surface over the top of the container, and use a rubber band to secure the sheet in place.

Step 3: Place a gel onto the cellophane. Remove any air bubbles that are under or around the gel. Flooding the surface of the cellophane around the gel with water will aid in the removal of bubbles.

Step 4: Place the second sheet of wetted cellophane over the gel, being careful not to trap any bubbles. Secure the second sheet of cellophane to the box with a second rubber band.

Step 5: Allow gel to dry for several days in a well-ventilated area.

Analysis of results

Although it is possible to analyze the dried gels, be aware that higher percentage agarose gels may be difficult to dry flat. In addition, exposure of the stained gel to light will cause the stain to fade. It is therefore recommended that analysis and determination of genotypes be done on wet gels.

Once the gels have been stained with Fast Blast DNA stain, it is time to determine the alleles present in each sample, and assign a DNA profile (genotype). For each PCR reaction, compare the bands obtained in each lane to the Allele Ladder run in lane #1. See page 40 for representative results, sizes of the ladder bands, and labeling of the alleles in the ladder. Assign each band in each PCR reaction with an allele assignment according to the band of corresponding size in the allele ladder. The bands in the allele ladder are numbered from top to bottom starting with the largest allele, #15, at the top. The sizes of the bands are indicated in the table below. In the example shown in figure 8 (page 26), the allele assignment for the sample in lane 2 is 3-7, since there is one allele 7 and one allele 3 in that lane. Write down the genotype for each of your samples in the chart below.

Lane	Sample	Number of Bands?	What BXP007 alleles are present?
1	BXP007 Allele Ladder		
2	Crime Scene DNA		
3	Suspect A DNA		
4	Suspect B DNA		
5	Suspect C DNA		
6	Suspect D DNA		

1. Did your samples all generate PCR products? If not, give reasons to explain why.

2. What is the genotype of each of your samples?

3. Does the Crime Scene DNA sample have a genotype that matches any of the suspects? If so, which one matches?

4. What does this result tell you about which suspects are included in the investigation? excluded? Explain your answer.

5. Imagine that each allele at the BXP007 locus is found at exactly the same frequency in a population. Since there are 8 possible alleles at the BXP007 locus, what is the frequency of any one allele from this locus in this population?

6. Given Mendel's Law of Independent Assortment and the assumption above, what is the frequency of the genotype of the Crime Scene sample?

7. If you had a pool of 13 suspects, and only one suspect had a genotype that matched the BXP007 locus found at the crime scene, would you be satisfied that you had identified the perpetrator based only on the genotype frequency calculated for the BXP007 locus? Why or why not? Explain your answer.

Appendix A
DNA and PCR in detail

DNA: A Detailed Look

A DNA molecule is a long polymer consisting of four different components called **nucleotides**. It is the various combinations of these four bases or nucleotides that create a unique DNA code or sequence (also genotype, gene, and allele).

Nucleotides are comprised of three different components:

- **Nitrogen base**
- **Deoxyribose sugar**
- **Phosphate group**

Each nucleotide contains the same ribose sugar and the phosphate group. What makes each nucleotide unique is its nitrogen base. There are four nitrogen bases:

Adenine (A)
Thymine (T)
Guanine (G)
Cytosine (C)

A DNA nucleotide chain is created by the connection of the phosphate group to the ribose sugar of the next nucleotide. This connection creates the “backbone” of the DNA molecule. To designate the different ends of this single-stranded chain, biochemistry terminology is used, in which the carbons on any sugar group are numbered. The sugar of a nucleotide contains 5 carbons. The phosphate group (PO_4) of a given nucleotide is connected to the 5' carbon of the sugar. A hydroxyl group (OH) is attached to the 3' carbon of the sugar, and this 3' OH group connects to the phosphate group of the next nucleotide in the chain.

Thus, the end of a single-strand DNA molecule that has a free phosphate group (i.e., not attached to another nucleotide) is called the 5' end, and the end of the DNA molecule with a free hydroxyl group (with no subsequent nucleotide attached) is called the 3' end (see Figures 13 and 14).

It has become standard that a single-stranded DNA molecule is written with the 5' end on the left and the 3' end on the right. Therefore, a single-stranded DNA chain's sequence is represented from left to right, starting on the left with the 5' nucleotide and moving to the right until the 3' nucleotide is last. Most DNA sequences are read 5' to 3'.

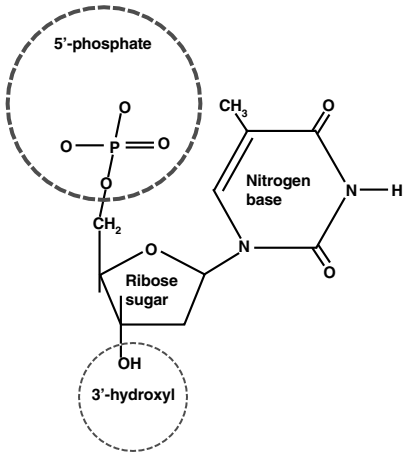


Fig. 13. Structure of one nucleotide of deoxyribonucleic acid.

The long DNA molecules or chains that comprise the chromosomes are not single-stranded molecules. From X-ray crystallography patterns of DNA generated by Rosalind Franklin, and some imaginative molecular model building, Watson and Crick deduced that DNA is in fact a **double-stranded** molecule with the two single strands of DNA held together by **hydrogen bonds** between the nitrogen bases (A, T, G, and C). This double-stranded molecule is often called a duplex (Figures 14 and 15). There are several important properties of double-stranded DNA molecules:

- Chromosomal (also called genomic) DNA is double-stranded
- The overall structure is that of a helix with two strands intertwined
- The structure can be viewed as a twisted ladder
- The phosphate-deoxyribose backbones are the sides of the ladder
- The nitrogen bases (A, T, G, and C) hydrogen bonded to each other are the rungs
- Only the nitrogen bases A & T and C & G can form hydrogen bonds to each other. When A binds to T or C binds to G this is considered **base pairing**. Neither C and T, nor A and G form hydrogen bonds
- The two strands are antiparallel; that is, the strands are oriented in opposite directions. This means that the ladder runs 5' to 3' in one direction for one strand and 5' to 3' in the opposite direction for the other strand

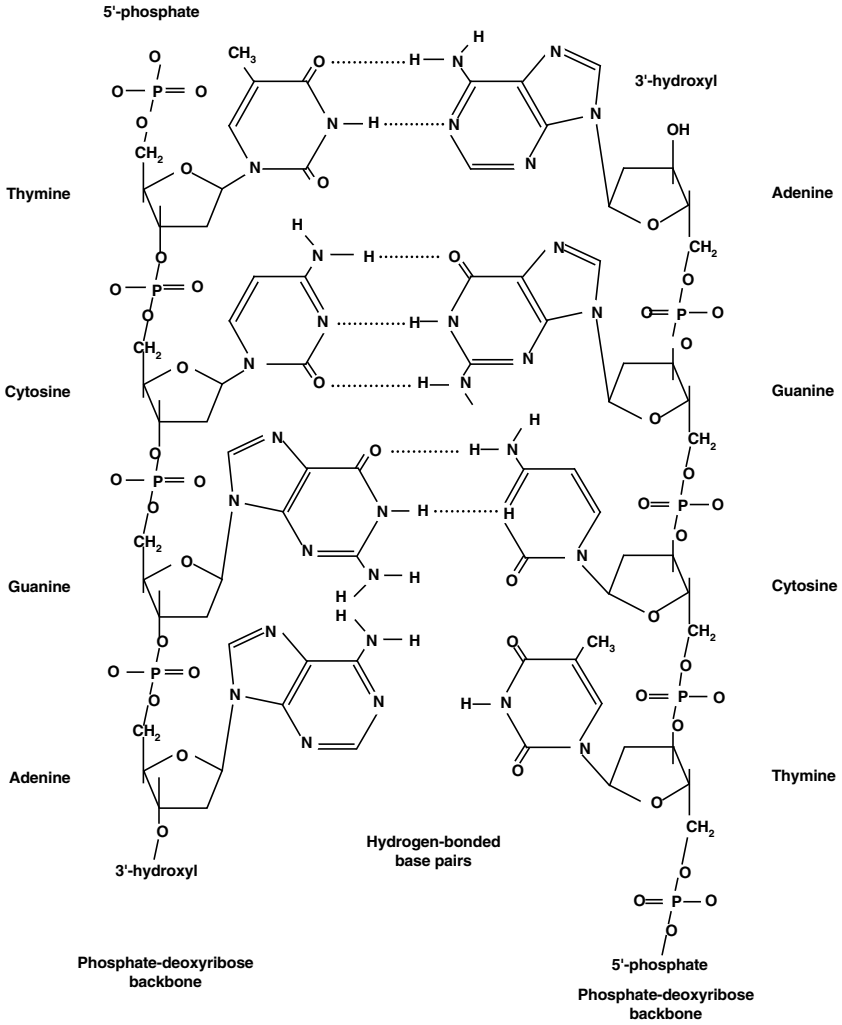


Fig. 14. Molecular structure of a portion of a double-stranded DNA molecule.

DNA Structure Conclusions

- Because A only binds to T, and G only binds to C, the two strands will have exactly the opposite, or complementary, sequence running in opposite directions (one strand 5' to 3' in one direction, the other 5' to 3' in the other direction).
- These two complementary strands anneal or hybridize to each other through hydrogen bonds between the bases.
- A new strand of DNA can be synthesized using its complementary strand as the template for new synthesis.
- Each strand carries the potential to deliver and code for information.

The length of any double-stranded DNA molecule is given in terms of base pairs (bp). If a DNA strand contains over a thousand base pairs, the unit of measure is kilobases (1 kb = 1,000 bp). If there are over one million base pairs in a strand the unit of measure is megabases (1 Mb = 1,000 kb).

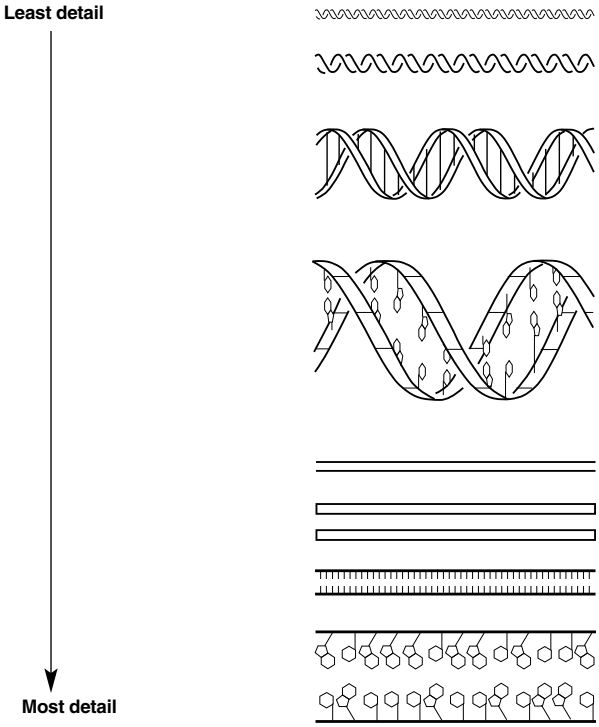


Fig. 15. DNA (deoxyribonucleic acid) — A long chainlike molecule that stores genetic information. DNA is graphically represented in a number of different ways, depending on the amount of detail desired.

DNA Replication — Strand Synthesis

New strands are synthesized by enzymes called **DNA polymerases**. New strands are always synthesized in the 5' to 3' direction. For a new strand of DNA to be synthesized, another single strand is necessary. The single strand of DNA that will be used to synthesize its complementary strand is called the **template strand**.

In order for DNA polymerase to start synthesizing a new complementary strand, a short stretch of nucleotides (approximately 20 base pairs long) called a primer (sometimes also called an oligonucleotide, for "few nucleotides") must be present for the polymerase to start the reaction. This primer is complementary to the template where synthesis will begin. The primer must have a free 3' hydroxyl group (OH) for DNA polymerase to attach the 5' phosphate group of the next nucleotide.

The DNA polymerase grabs free (single) nucleotides from the surrounding environment and joins the 5' phosphate of the new nucleotide to the 3' hydroxyl group (OH) of the new complementary strand. This 5' to 3' joining process creates the backbone of the new DNA strand.

The newly synthesized strand maintains its complementarity with the template strand because the DNA polymerase only joins two nucleotides during new strand synthesis if the new nucleotide has its complement on the template strand. For example, the DNA polymerase will only join a G to the 3' end of the newly synthesized strand if there is the C counterpart on the template strand. Guanine will not be joined to the new strand if A, T, or G is the opposite nucleotide on the template strand.

DNA polymerase and strand synthesis allow DNA to replicate during mitosis. Both new DNA strands are synthesized simultaneously from the two original DNA template strands during mitotic DNA replication.

DNA, RNA, and proteins are closely tied to each other. DNA directs the expression of RNA, which directs the expression of proteins, which carry out most of the biological functions that make up an organism. Thus, you can begin to understand why researchers, in an attempt to understand the mechanisms behind the various life processes, must study nucleotides to get complete answers about the flow of information carried in the genetic code. In the last 20 years, many gains in the areas of nucleic acid techniques have finally allowed researchers the means to study the roles of nucleic acids in life processes.

Individual discoveries by many scientists have contributed the pieces that have begun to solve one of the most mysterious puzzles of life — understanding the hereditary code. In 1985, enough pieces of the puzzle were in place for a major breakthrough to occur. This understanding of how the necessary molecular components interact to faithfully replicate DNA within living cells led to the development of a technique for creating DNA in a test tube. This technique is called the **polymerase chain reaction**, or **PCR**.

PCR: A closer look

PCR Makes Use of Two Basic Processes in Molecular Genetics

1. **Complementary DNA strand hybridization**
2. **DNA strand synthesis via DNA polymerase**

In PCR, complementary strand hybridization takes place when two different **primers** anneal to each of their respective complementary base pair sequences on the template. The two primers are designed and synthesized in the laboratory with a specific sequence of nucleotides such that they can anneal at the opposite ends and on the opposite strands of the stretch of double-stranded DNA (template strand) to be amplified.

Before a region of DNA can be amplified, one must identify and determine the sequence of a piece of DNA upstream and downstream of the region of interest. These areas are then used to determine the sequence of oligonucleotide primers that will be synthesized and used as starting points for DNA replication. Primers are complementary to the up- and downstream regions of the sequence to be amplified, so they stick, or anneal, to those regions. Primers are needed because DNA polymerases can only add nucleotides to the end of a preexisting chain.

The DNA polymerase used in PCR must be a thermally stable polymerase because the polymerase chain reaction cycles between temperatures of 52°C and 94°C. The thermostable DNA polymerase used in PCR was isolated from a thermophilic bacterium, *Thermus aquaticus* (*Taq*), which lives in high-temperature steam vents such as those found in Yellowstone National Park.

Two new template strands are created from the original double-stranded template on each complete cycle of the strand synthesis reaction. This causes exponential growth of the number of template molecules, i.e., the number of DNA strands doubles at each cycle. Therefore, after 35 cycles there will be 3.4×10^{10}, or over 30 billion, times more copies than at the beginning. Once the template has been sufficiently amplified, it can be visualized. This allows researchers to determine the presence or absence of the desired PCR products and determine the similarities and differences between the DNA of individuals. Depending on the DNA sequence analyzed, differences among individuals can be as great as hundreds of base pairs or as small as a single base pair or single point mutation.

PCR Step by Step

PCR involves a repetitive series of cycles, each of which consists of template denaturation, primer annealing, and extension of the annealed primer by *Taq* DNA polymerase. Before beginning DNA amplification, template DNA is prepared from evidence.

Following sample preparation, the template DNA, oligonucleotide primers, thermostable DNA polymerase (*Taq*), the four deoxynucleotides (A, T, G, C), and reaction buffer are mixed in a single micro test tube. The tube is placed into the thermal cycler. Thermal cyclers contain an aluminum block that holds the samples and can be rapidly heated and cooled across extreme temperature differences. The rapid heating and cooling of this thermal block is called **temperature cycling** or **thermal cycling**.

The first step of the PCR temperature cycling procedure involves heating the sample to 94°C. At this high temperature, the template strands separate (denature). This is called the **denaturation step**.

The thermal cycler then rapidly cools to 52°C to allow the primers to anneal to the separated template strands. This is called the annealing step. The two original template strands may reanneal to each other or compete with the primers for the primers' complementary binding sites. However, the primers are added in excess such that the primers actually out-compete the original DNA strands for the primers' complementary binding sites.

Lastly, the thermal cycler heats the sample to 72°C for *Taq* DNA polymerase to extend the primers and make complete copies of each template DNA strand. This is called the extension step. *Taq* polymerase works most efficiently at this temperature. Two new copies of each complementary strand are created. There are now two sets of double-stranded DNA (dsDNA). These two sets of dsDNA can now be used as templates for another cycle and subsequent strand synthesis.

At this stage, a complete temperature cycle (thermal cycle) has been completed (Figure 16).

Temperature cycle = denaturation step + annealing step + extension step

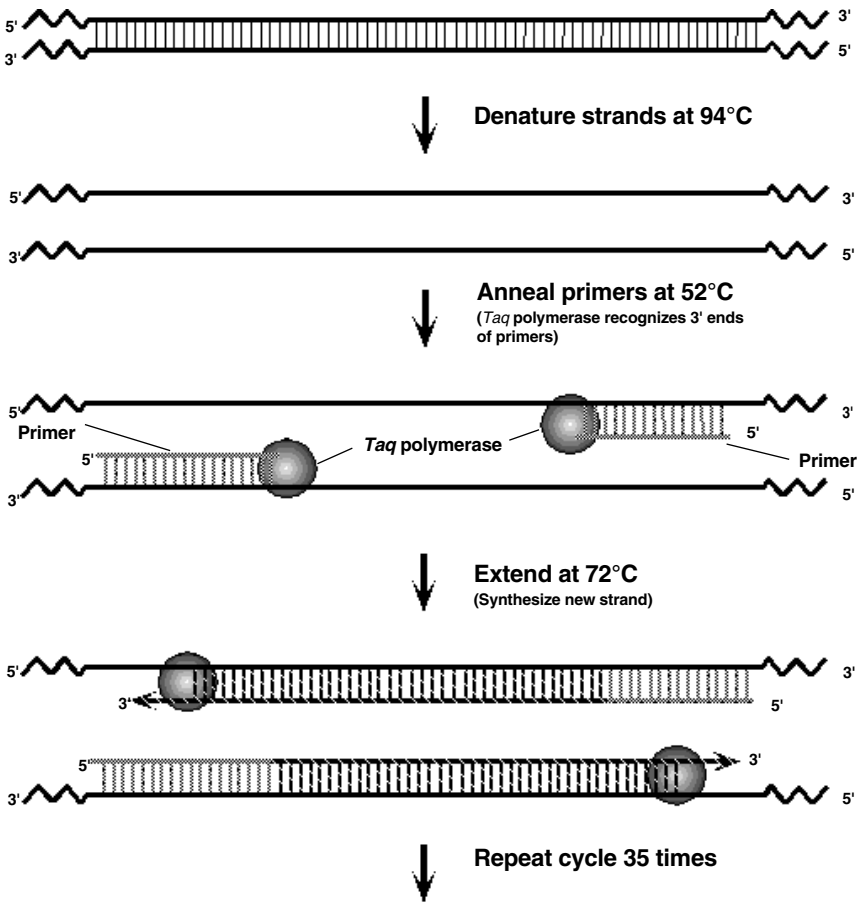


Fig. 16. A complete cycle of PCR.

Usually, thermal cycling continues for about 35 cycles. After each thermal cycle, the number of template strands doubles, resulting in an exponential increase in the number of template DNA strands. After 35 cycles there will be 3.4×10^{10} more copies of the original number of template DNA molecules.

PCR generates DNA of a precise length and sequence. On the first cycle, the two primers anneal to the original genomic template DNA strands at opposite ends and on opposite strands. After the first complete temperature cycle, two new strands are generated that are shorter than the original template strands but still longer than the length of the DNA that the researcher wants to amplify. It isn't until the third thermal cycle that fragments of the precise length are generated (Figure 17).

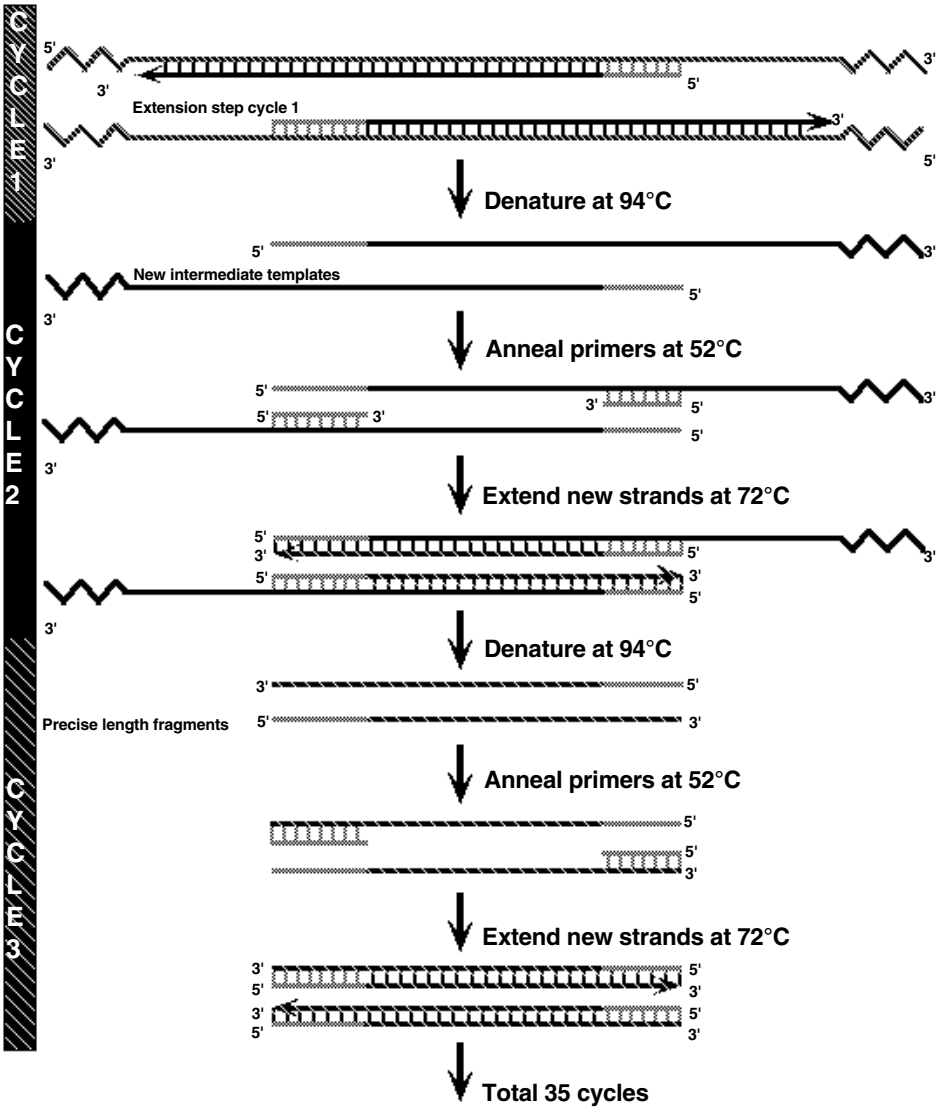


Fig. 17. Generation of precise-length fragments.

It is the template strands of the precise length that are amplified exponentially (X^n, where X = the number of original template strands and n = the number of cycles). There is always one set of original long-template DNA molecules that are never fully duplicated. After each thermal cycle, two intermediate-length strands are produced, but because they can only be generated from the original template strands, the intermediate strands are not exponentially amplified. It is the precise-length strands generated from the intermediate strands that are amplified exponentially at each cycle. Therefore, if 20 thermal cycles were conducted from one double-stranded DNA molecule, there would be 1 set of original genomic template DNA strands, 20 sets of intermediate template strands, and 1,048,576 sets of precise-length template strands (Figure 18). After 35 cycles, there would be 1 set of original genomic template DNA strands, 35 sets of intermediate template strands, and 3.4×10^{10} sets of precise-length template strands.

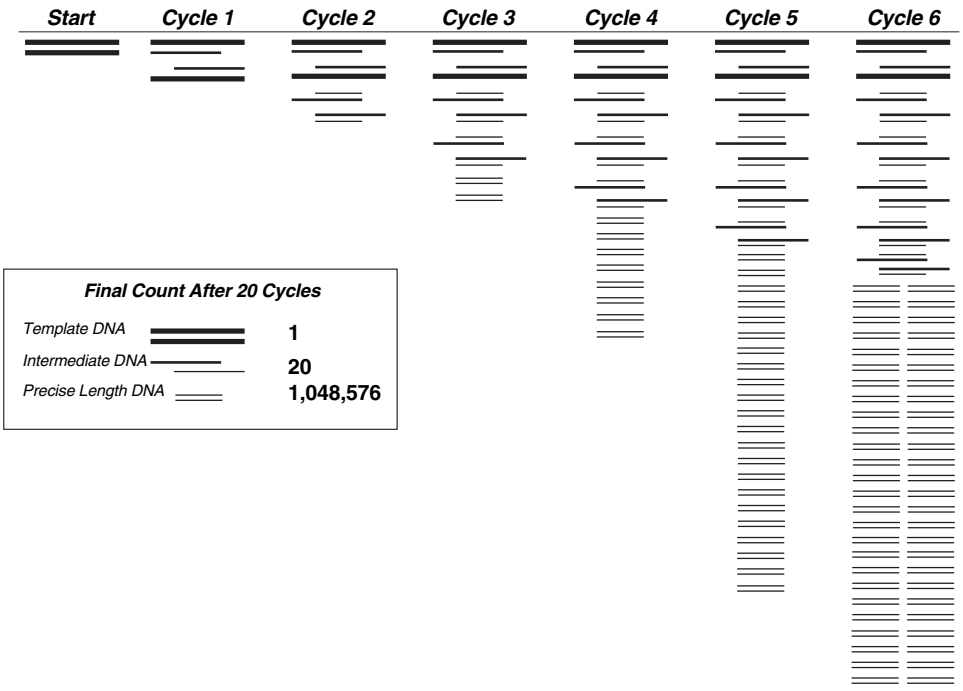


Fig. 18. Schematic of PCR amplification of DNA fragments.

Appendix B

A Brief History of Forensic Testing

Testing Speed vs. Power of Discrimination

Long before DNA evidence took center stage as the preferred means for identifying individuals, forensic laboratories still had the task of unambiguously linking suspects to crime scenes. They often did this indirectly, for example showing that the muddy shoe found in the suspect's car had a shoe print identical to that left at the scene of a crime. Or, showing that a plastic bag found in the suspect's house matched those found in the victim's house. On their own, these pieces of evidence are only suggestive – they don't actually prove that the suspect was at the scene of a crime. Something from the suspect must be left to be able to directly link that individual to the scene. Biological evidence is the key.

Blood typing

The first genetic tests were based not on DNA at all, but on blood types. There are 4 blood groups, or types – A, B, AB, and O. The test for blood groups is very fast and straight-forward. However, 40% of the population is type O, so it's not particularly useful if it turns out that several of your suspects are type O. Blood group determination is more useful to exclude potential suspects. If you know that someone left O-type blood at the scene of the crime, and the victim isn't type O, then you can exclude any suspects that have A, B, or AB blood.

RFLP

Restriction Fragment Length Polymorphism, or RFLP, has been the workhorse of forensic DNA profiling for many years. First described by English geneticist Alec Jeffries in 1985, RFLP makes use of the fact that certain regions in our genomes have DNA sequences that are repeated over and over again (called Variable Number of Tandem Repeats, or VNTRs) next to each other. Jeffries found that the number of repeats at any particular location differs from person to person. Using restriction enzymes that digest highly specific patterns in DNA, Jeffries was able to cut out the VNTRs and compare their sizes directly. He found that the locus (or allele marker) where repeats are found carried a pattern of repeats that comes either from the mother, or the father. In fact, the first legal application of RFLP was used to solve a British immigration case involving family relationships.

Although RFLP is a very powerful tool and has great discriminating potential, it is laborious, cannot easily be automated, and is time-consuming. As it depends on DNA sequences of relatively great size (a minimum of several hundreds of base pairs), DNA must be of at least reasonable quality, and not degraded. Older DNA samples may not be suitable. Perhaps its greatest drawback is that it often requires large amounts of DNA (tens of micrograms). This makes it unsuitable for many crime scenes where there may be only a single hair implicating a suspect.

STR

Short tandem repeats (STRs) are similar to VNTRs, since they too describe regions of DNA where you see DNA sequences repeated over and over. In this case, the repeat sizes are smaller – sometimes only 2 nucleotides are repeated. Like VNTRs, everyone inherits their STR pattern from their parents – for each locus, one STR allele comes from their mother, and one from their father.

STRs are much easier to analyze than RFLPs. First, because they are small they are easy to amplify using PCR. Second, because they can be amplified by PCR, STR tests can be automated, with several tests (>10) usually being run at the same time. Each PCR test

runs about 4 hours, compared to several days for RFLP. Third, even degraded DNA can generally give meaningful results. Degraded DNA may be broken into small fragments, which in most cases is acceptable for PCR, but not for RFLP. And finally, even a single piece of hair can yield enough genetic material to result in successful genetic analysis. While all these attributes are compelling, what makes STR analysis so powerful is that several loci containing STRs can be analyzed at the same time. Individual tests can, at best, distinguish 1 out of every 2000 or so people. In combination, as few as 13 loci can discriminate between any two people in the world (with the exception of identical twins), living or dead.

Of interest is the fact that STRs exist in all species tested to date. This has caused many groups to use STR analysis to trace bloodlines in certain species (such as dogs and race horses) where lineage tracking is big business.

DNA Profiling: What's next?

Several new areas are under development. Mitchondrial DNA (mtDNA) is one hot spot. mtDNA is unusual – it's non-nuclear, and so doesn't follow the same rules of inheritance as chromosomal DNA. In fact, it is generally accepted that mtDNA is maternally inherited. Of relevance to DNA profiling is the fact that mtDNA is a double-stranded circle. It contains many of the genes involved in the Krebs cycle (respiration). These genes function mainly inside mitochondria. Perhaps because it is circular, mitochondrial DNA tends to be much more stable than chromosomal DNA. Forensic analysts often refer to it as the "last resort", but may often check its quality first, before doing any other analyses; if mtDNA is degraded, then it's very likely that the chromosomal DNA is as well.

The reason that mtDNA is so useful for forensic purposes is its high copy number. Whereas a nuclear gene will have only two copies per cell (one from each parent), there are hundreds of copies of cytoplasmic mtDNA per cell, providing a much better opportunity for molecular analysis, even when material is limited.

It had been assumed that because the genes for mtDNA are so essential for metabolism, there would be little variation among people. However, recent efforts to compare mitochondrial DNA sequences in different human populations has shown that there is a 'hypervariable' region on mtDNA that differs significantly between individuals. These regions that have been validated for use in several crime labs, and there are a number of companies that specialize in mtDNA analysis for forensic cases. mtDNA has been especially useful in the identification of remains of missing persons, especially from mass graves and other disaster sites. In the 2004 Indian Ocean tsunami disaster, identification of victims involved a combination of STR analyses and mtDNA analyses.

Amplified fragment length polymorphism (AFLP) is a PCR-based variation of RFLP in which sequences are selectively amplified using primers. It is reliable and efficient method of detecting molecular markers. DNA is cut with restriction enzymes to generate specific sequences, which are then amplified using PCR. AFLP can evaluate more loci than with RFLP. AFLP is also capable of determining a large number of polymorphisms. AFLP-based assays are cost-effective and can be automated.

Finally, there is much excitement over the potential for new forensic analysis using genomic sequences or patterns called single nucleotide polymorphisms, or SNPs. Single nucleotide polymorphisms are essentially single base pair polymorphisms that exist in all DNA. Several classes of SNPs exist. The one of most interest to forensic analyst are SSNPs, or silent single nucleotide polymorphisms, that don't actually change the outcome of any gene product or gene regulation. SNPs are good markers for a number of reasons – there are many more of them (one estimate suggests > 10 million) compared to other types

of markers, they are randomly distributed across the genome, they are bi-allelic (as compared to STRs, which have multiple possible states), and their distribution is fairly uniform. Presently, there are several drawbacks to using SNPs in forensic analysis. First, much more DNA is needed to perform SNP analysis than other methods currently in use in forensic case work. Another disadvantage of SNP typing for forensic applications is that a much larger number of the biallelic SNPs will need to be typed to acheive the same power of discrimination as the 13 multiple allelic STR loci. At this point in time, the cost of SNP typing is substantially higher than STR typing; this, plus the fact that crime labs are fully operational and validated for STR typing means that SNP typing is unlikely to be used for routine casework for some time.

In forensic DNA analysis, there is always a balance between speed, cost effectiveness, and the power of discrimination (Figure 19). STR analysis is the current method of choice as it is highly discriminating, and can be performed in a matter of hours. The other methods are either lacking in their ability to discriminate between two DNA profiles, or are too laborious or lengthy. That is not to say that they aren't used in forensic analysis. In addition, with improvements in technology and increasing information about the genome and mitochondria, it's likely that other methods may become favored over STR profiling.

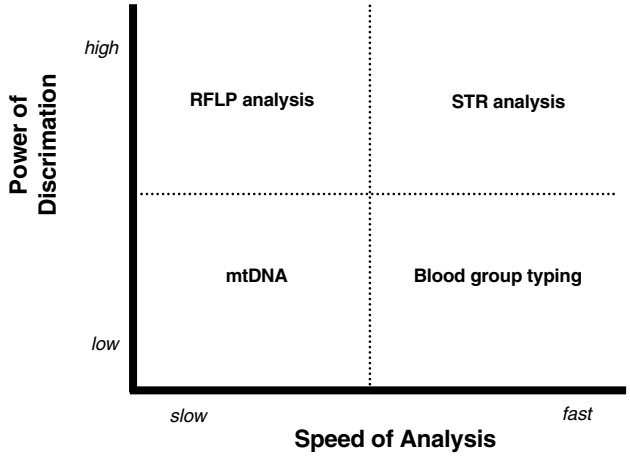


Fig. 19. Comparisons of forensic tests by speed and discrimination power.

DNA Sample Collection and the OJ Simpson case

As illustrated in the figure below, once crime scene samples have been collected, they are taken back to the lab so that DNA may be extracted, quantitated, and analyzed. PCR is then performed to make copies of the specific region or regions that are under analysis. In the CODIS system (**CO**mbined **D**NA **I**ndex **S**ystem; for more information, see information about the CODIS database in this Appendix), PCR primers have been designed to amplify 13 different regions of the human genome. Each set of primers amplifies only one region. Each separate region, or loci, contains an STR that numbers anywhere from 8 to 32 repeats, depending on the loci. To be able to maximally analyze the greatest number of samples in one reaction, the companies that developed these tests did two things – first, they labeled the primers used for PCR with different fluorescent tags. Second, they made the primers in such a way that the PCR products could be distinguished not only by fluorescent color but also by size. In other words, no two over-lapping PCR products would have the same color.

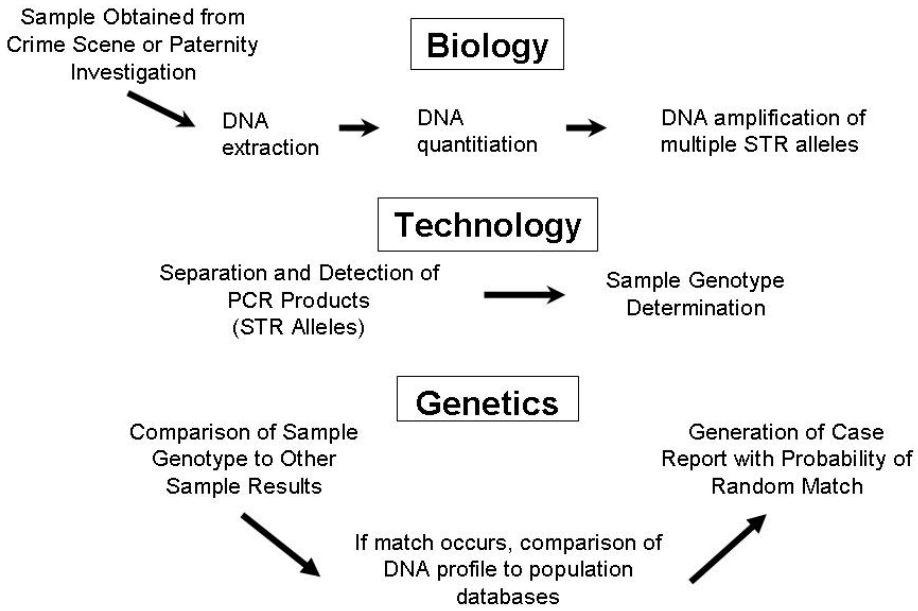


Fig. 20. Steps in forensic DNA sample processing (7).

After completing the PCR reaction, PCR products are separated by size on acrylamide gels and the fluorescent products detected. A typical forensic DNA profiling gel is shown in figure 21 for 7 tested loci. Since the PCR products to be detected are very small (e.g. <50 bp), the better resolving power of acrylamide is necessary for optimum resolution. Once the fluorescent products have been detected, a program automatically compares the products to allele ladders (size markers), and identifies each product based on its size and color match to these allele ladders. The program then assigns a genotype, or DNA profile, for each individual sample for each locus tested.

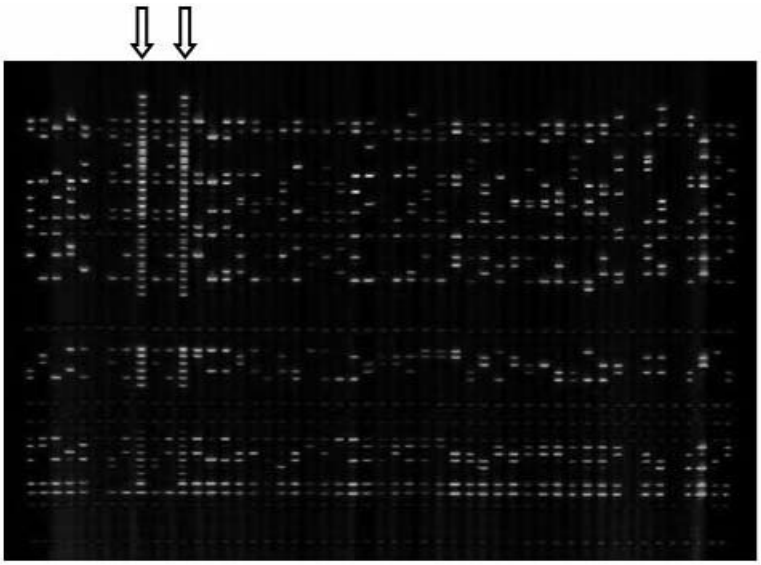


Fig. 21. A typical STR gel. Four different fluorescent tags have been used to identify 7 amplified loci. Allele ladders, or markers, are indicated by arrows.

This DNA profile is then compared to other samples. In forensic studies, this might mean that a sample from a crime scene is compared to that of many suspects. If the samples don't match, then those suspects can be excluded from the crime scene. If there is a match, then those suspects are included in the investigation. At this point, the included DNA profiles may then be compared to a database (such as can be found on STRbase www.strbase.org) to understand whether the genotypes are more prevalent in particular ethnic groups to give a better estimate of the random match probability. In general terms, the random match probability tells you the likelihood of whether someone else from that particular population will have exactly the same genotype. To read more about this topic, see Appendix C.

As DNA testing technologies continue to improve and become able to analyze ever smaller amounts of starting material, it will become even more important that samples are collected appropriately. If you happened to sneeze while collecting evidence from a crime scene, you could find yourself the chief suspect in a murder investigation. Suspect collection and sample processing are the most vulnerable part of any investigation, because that is the point that evidence is most prone to contamination. Contamination means that a sample is impure. It's worth noting that crime scenes often produce mixed samples. Imagining the motel room scenario, many people may have stayed in the room before the crime was committed, and certainly left DNA evidence behind. For DNA evidence, contamination means that more than one DNA type may be present in the same sample (possible, depending on the crime), or that the DNA has been exposed to a substance that interferes with analysis. Either situation can have a big outcome on analysis. Inconclusive DNA evidence can have as big an effect in the courtroom as conclusive evidence. Because of this, police take special precautions when collecting evidence for DNA profiling:

1. Always wear gloves. Skin is an excellent source of cells containing DNA, and everyone sheds skin cells constantly. Only a few cells are needed as the basis for DNA profiling. Gloves can help prevent contamination of suspect collection.
2. Tie hair back or cover it. As we've already mentioned, a single hair is all it takes to implicate someone at a particular place.
3. Sneezing or coughing can expel cells all over evidence, so a mask is often worn.
4. Crime scenes, and evidentiary samples, must be treated in such a way that biological materials are not accidentally transferred from one place to another, thereby providing misleading associations. In one famous crime scene, the body of the victim was covered with a blanket from the household – this mistake made any hairs found on the victim of no value since it wasn't possible to determine whether they were evidence of contact with the person who committed the crime, or were simply transferred to the victim by accident.

As an example of how these procedures can affect the outcome of a criminal investigation, let's take a look at the OJ Simpson trial:

In 1995, Nicole Brown Simpson and a friend, Ron Goldman, were found murdered at the home of Ms. Simpson. Nicole's ex-husband, OJ Simpson, became the primary suspect. A substantial amount of evidence linked Mr. Simpson to the scene of the crime, and DNA profiling (both RFLP- and STR-based) was used to create DNA profiles of the many bloodstains found at the homes of Ms. Simpson and Mr. Simpson. Of 45 bloodstains discussed at trial, all were identified as coming either from the two victims, or from Mr. Simpson (9). However, despite the weight of evidence, Mr. Simpson was acquitted. A prominent criminologist, Dr. Henry Lee, planted doubt in the jury's mind by suggesting that crime scene samples had been mishandled. In addition, this was one of the first, and most prominent, cases involving DNA evidence, so the prosecution devoted a lot of court time to

the complex procedures involved in DNA profiling. The defense was able to plant sufficient doubt in the minds of the jurors that the initial sample collection was either sloppy, or that someone intentionally mishandled evidence. Much of the defense's effort was focused on a police officer that they claimed was so biased that he deliberately planted evidence implicating the defendant. The possibility that a police officer could have mishandled evidence was enough for the jury to acquit.

Despite this apparent setback to DNA profiling, it continues to be used in countless court cases. In 1983, Calvin Johnson was accused and convicted of rape and burglary. Today, he is free because DNA evidence showed he was not the man who committed the crime. DNA profiling has been used outside of the courtroom to settle parental issues, and to identify individuals involved in many disasters. The table below mentions some of the milestones and achievements in forensic analysis.

Milestones in forensic DNA analysis

Year	Event
1985	Alec Jeffries develops multi-locus RFLP probes
1988	FBI starts work using single locus RFLP
1990	PCR analysis using single locus STR begins
1992	FBI initiates STR work using DQA1 locus
1994	DNA Identification Act: provides funding for national DNA database
1995	OJ Simpson trial focuses public attention on DNA evidence
1996	FBI starts to test mtDNA; first multiplex STR kits available
1997	13 core STR loci described
1998	FBI starts CODIS database; Swissair disaster – all remains identified using STR DNA profiling
2001	World Trade Center disaster in NYC – many remains identified using a combination of DNA profiling approaches (8)
2004	California proposition 69: State DNA Database Funding Initiative. State initiative passed with 62% of vote. (14)
2004	Indian Ocean tsunami; Interpol and other world agencies to use DNA profiling to identify victims
Today	Trace your Genetic Genealogy; commercially available packages can trace paternal/maternal ancestry

The CODIS system and DNA Databases

The collection and stockpiling of DNA evidence has the potential to be of great help to law enforcement. CODIS, or COmbined DNA Index System, is a federally maintained database of DNA obtained from crime scenes and convicted violent offenders. CODIS works on federal, state, and local levels to obtain and maintain DNA profiles (13). All DNA profiles originate at the local level, and then migrate to the state and federal levels.

Although CODIS is administered at the federal level, states have the power to legislate local DNA evidence collection. Although CODIS was originally designed to collect information about violent criminals, many states have now enacted legislation that allows collection of DNA evidence even if that person is not convicted of a crime.

In California, proposition 69 requires the collection of DNA from anyone convicted of any felony offense, or any violent sexual assault (14). It also requires DNA collection from anyone arrested or charged with various violent crimes or felony offenses. In favor of this, proponents say that too many violent crimes go unsolved because California does not have a comprehensive DNA database. Further, they point out that the tests do not reveal any medical conditions about individuals, so medical privacy would not be violated.

However, opponents object to DNA collection from people who have not been convicted of a crime. Further, they feel that privacy safeguards are not adequate, and that the state is not compelled to respond to requests to have innocent individuals' information removed. Many people also feel that even one corrupt official could compromise the privacy rights of countless individuals. This is a controversial area, and is continually being examined and developed.

How does CODIS actually work?

CODIS examines 13 loci, or markers, that are uniformly distributed across the human genome. The loci used, and their relative positions, are listed below in Figure 22.

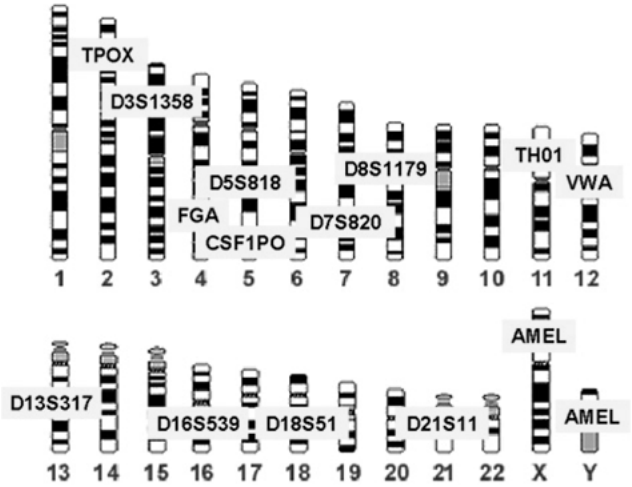


Fig. 22. The 13 core CODIS loci and their genetic locations (7).

In addition to the 13 loci used for STR profiling, forensic analysts also analyze the amelogenin locus. PCR products at this locus produce X chromosome and Y chromosome specific PCR products of different sizes. The amelogenin locus therefore provides gender information about a particular DNA sample.

One important feature of these 13 loci is that they have been carefully chosen so that they reveal no medical or health information about the individual being profiled. That is, these loci come from regions of the human genome that are not known to be associated with any disease or condition. For that reason, they are called "anonymous" markers.

Why choose 13 loci?

The ability to distinguish between any two individual DNA profiles increases with the number of loci tested. If only one locus was examined, many people would likely have the same genotype, and so telling the difference between any two people would be very difficult. In addition, allele frequencies have been shown to vary between ethnic groups. Depending on the ethnic group under study, the power of discrimination at any one locus may only be 1 out of every 200 people. With the addition of more loci, the ability to discriminate between two profiles increases. Take a look at the following fictitious DNA profile. The alleles identified in this person's STR DNA profile and the frequencies in Caucasians for those alleles (3) are listed. The final row lists the Random Match Probability (RMP; described in more detail in Appendix C) for the combined genotype. The RMP tells you how likely it is for anyone else to have the same genotype. For this particular genotype, there is about a one in 2.7 trillion chance that another Caucasian has the same genotype. Since there are only ~ 6.5 billion people alive on Earth today, that's a highly discriminating number!

Fictitious DNA Profile: Random Match Probability Calculation

STR locus	Identified allele	Allele frequency in Caucasians (from database)	Locus frequency in Caucasians	
			Formula	Locus frequency
TPOX	8 12	$p = 0.535$ $q = 0.041$	$2pq$	0.044
TH01	10 10	$p = 0.008$	p^2	0.000064
D3S1358	16 17	$p = 0.253$ $q = 0.215$	$2pq$	0.109
FGA	21 23	$p = 0.127$ $q = 0.134$	$2pq$	0.034
CSF1PO	11 13	$p = 0.301$ $q = 0.096$	$2pq$	0.058
D8S51	14 19	$p = 0.137$ $q = 0.038$	$2pq$	0.010
D21S11	28 29	$p = 0.159$ $q = 0.195$	$2pq$	0.062
			Combined genotype frequency	
Total RMP for this genotype = frequency of TPOX locus [***f*** (TPOX)] x frequency of TH01 locus [***f*** (TH01)] x frequency of D3S1358 locus [***f*** (D3S1358)] etc.			***f*** (TPOX) x ***f*** (TH01) x ***f*** (D3S1358), etc.	3.75×10^{-13} or ~1 in 2.7 trillion

Fig. 23. Fictitious DNA Profile – Random Match Probabilities. In this imaginary Caucasian DNA profile, 7 loci have been examined. One locus – TH01– is homozygous, so only one allele has been identified. A locus frequency is indicated for each combination of alleles at a particular locus, and the total RMP for the combined genotype is also shown.

To calculate the genotype frequency at any partuclar locus, you need to take into account the possibility of inheriting the combination of allelles present at that particular locus from each parent. Allele frequencies have been shown to vary between different ethnic groups, and these frequencies have been published (3). As an example, let's look at the TPOX locus. In Caucasians, the frequency of the 8 allele (let's call this frequency '*p*') is p = 0.535. This means there's about a 53.5% chance that any Caucasian TPOX allele typed would be an 8. Similarly, there's a q = 0.041 chance that a random TPOX allele would be typed as a 12. The chance that this person got the 8 allele from his mother and the 12 allele from his father is represented as *pq*, and the opposite – that he got the 8 from his father and the 12 from his mother – is also *pq*, so the locus frequency at any heterozygous locus can be thought of as *pq* + *pq* or 2*pq*. So, 2*(0.535)(0.041) = 0.044, or 4.4%, of Caucasians have the 8, 12 genotype at the TPOX locus.

At the THO1 locus, since both alleles are the same, the frequency is simply *pp*, or p^2, which is the combined chance of inheriting allele 10 from each parent. So , about 0.0064% of Caucasians have this particular genotype at the TH01 locus.

The Hardy-Weinberg theory is the principle behind the formulae for calculating genotype frequencies at any locus (i.e. p^2, 2*pq*). In essence, Hardy-Weinberg describes the probable genotype frequencies in a population and tracks their changes from one generation to another. In the case of STR calculations, it allows geneticists to take observed allele frequencies and calculate a genotype frequency as described above in the table.

In our example in the table above, the chance for any Caucasian to have this particular combined genotype (TPOX 8, 12; TH01 10,10) is 0.044 x 0.000064 which is ~3 x 10^{-6} or ~3 in 100,000 of all Caucasians screened at these two loci (TPOX and TH01). Mendel's Law of Independent Assortment tells us that alleles are inherited independently, so the Product Rule can be applied to make this calculation. The Product Rule says the combined genotype frequency is the product of all of the separate loci frequencies (represented generically as '***f***'; see the table under combined genotype frequency), as described above for the TH01 and TPOX alleles. The genotype frequency may also be described as the Random Match Probability, or RMP. RMP is described in more detail in Appendix C.

In the US, 13 loci are used for analysis. The combined RMP using 13 loci provides enough discrimination power to tell the difference between any 2 people in the world, with the exception of identical twins. In the UK, law enforcement has elected to use only 10 loci, which could, in some cases, lead to a situation where more than than one person has the same identified genotype.

Plant and Insect Genotyping in Forensic Investigations

The vast majority of DNA profiling associated with criminal investigations involves profiling of the people implicated in the crime. However, genetic profiling using many of the methods described is also performed in some instances, where the particular genotype of a plant or an insect can help associate a piece of evidence with a known location or time. For example, smuggling of endangered species of plants and animals is still prevalent in many parts of the world. To be able to identify animals and plants specifically, many of the same tests described in this Appendix are employed.

Forensic entomology, or the application of the study of insects to criminal cases, is the field devoted to these studies in homicide and wildlife poaching investigations. Insects colonize remains shortly after death and develop in a predictable way. Identification of insects using the methods described here as well as other methods allow an entomologist to estimate elapsed time since death, as well as other factors such as position of wound sites, and whether the body has been moved or disturbed.

Appendix C
Exercises in STR Allele Frequencies and Random Match Probabilities

Exercise 1: Simulation of Inheritance of STR Allele and Power of Discrimination

The Crime Scene Investigator PCR Basics kit allows students to simulate a genotyping at one of the loci commonly used in forensic typing. In real crime scene applications, DNA profiling is performed at a number of different loci to improve the power of discrimination of the testing. In simple terms, the power of discrimination is the ability of the profiling to discriminate between different individuals. The larger the number of loci profiled, the more powerful the ability to discriminate.

This concept can be illustrated in the classroom with a very simple exercise. All students are asked to stand, and they now form a pool of possible suspects for a hypothetical crime. There is an eyewitness who saw the criminal run from the crime scene and has provided a description. As will be apparent, the more bits of identifying information provided by the witness, the greater the number of persons excluded from the suspect pool, and the smaller number of suspects included in the suspect pool. The teacher provides bits of identifying information, asking those who are no longer suspects to sit. The teacher repeats this process (selecting from the suggested list below, but in an order that continues to eliminate students from the pool of suspects) until only one student remains standing – there is now a suspect pool of one! Some of the types of information provided by an eyewitness might include the following:

The criminal wore blue denim jeans.
The criminal wore a T-shirt (or sweatshirt, depending on the season) with letters on it.
The criminal wore glasses.
The criminal had ____ hair color.
The criminal had very short or long hair.
The criminal was male or female.

The question to ask the students is: Does this prove that [name] committed the crime? Why or why not?

The correct answer is that the last student standing could have committed the crime, but an important consideration is whether any other persons who fit this same description were also present. In other words, what is the possibility that another person has this exact set of features? This same consideration comes into play with DNA profiling. What is the chance that a randomly selected individual will have the same identical DNA type of the suspect? This random match probability is an important component of using DNA evidence to solve crimes.

Adding more pieces of observation to the physical description of the escaping criminal made it more likely to identify the correct person as guilty (or innocent). In exactly the same way, adding more genetic loci to the DNA profiling profile makes it a much more powerful tool for solving crimes.

In the US, 13 STR loci have been chosen for forensic typing and inclusion in the national database called CODIS. The average random match probability when all 13 are typed is less than one in a trillion. Since the total world population is about 6.5 billion people, that means that the CODIS system can in theory tell the difference between any two people, with the exception of identical twins.

The next part of the exercise demonstrates the inheritance of STR alleles at four loci and shows how even siblings will have different profiles. The four STR loci we will model are actually used in forensic typing:

Locus name	**Chromosome**	**Allele Range (# repeats)**
VWA (blue)	12	10–24
D8 (green)	8	8–19
D5 (yellow)	5	7–16
TH01 (red)	11	3–14

Materials required for each class:

- 8 small paper bags
- Blocks or small squares of poster board in red, blue, green, yellow (total number = 2 of each color per student)
- Student worksheets, transparency or board copy of the table

Before class, the teacher will

- Label ¼ of each of the blue blocks with one allele (13) for VWA, ¼ with a second different allele (18), ¼ with a third, different allele (16), and ¼ with a fourth, different allele (20) with a permanent marker. Keep the piles of labeled blocks separate. Place two sets of alleles in one paper bag labeled "mom, VWA"; put the remaining two sets into another paper bag labeled "Dad, VWA"
- Repeat with green blocks using D8 alleles (8, 12, 9, 13). Place two sets in a bag labeled "mom, D8" and two sets in a bag labeled "Dad, D8"
- Repeat with yellow blocks using D5 alleles (7, 11, 10, 12). Place two sets in a bag labeled "mom, D5" and two sets in a bag labeled "Dad, D5"
- Repeat with red blocks using TH01 alleles (7, 11, 10, 12). Place two sets in a bag labeled "mom, TH01" and two sets in a bag labeled "Dad,TH01"

In class:

1. One student is named the "mom" of the family and one student the "dad". The mom will take the four bags marked "mom" and the dad will take the four bags marked "dad". They will determine their genotypes at each of the four STR loci; all students will enter these into their data sheets.
2. Each student now "inherits" his or her STR genotype by selecting at random one allele from each of the Mom's bags and one allele from each of the Dad's bags. Each student enters his or her data on the blackboard or transparency master sheet and all students transcribe the data onto their worksheets. By repeating for each student in class, a large "family" of children with the same mother and father have been generated.

Table I: STR Inheritance & Typing Simulation Student Worksheet

Family genotypes: Fill in the names and genotypes of the parents and all the children

Name	VWA Alleles		D8 Alleles		D5 Alleles		Th01 Alleles	
	Maternal	Paternal	Maternal	Paternal	Maternal	Paternal	Maternal	Paternal
Mom								
Dad								

Question 1: Considering only the VWA alleles, how many of your siblings have the same genotype as you do?

Question 2: Considering VWA ***and*** D8, how many of your siblings have the same genotype as you do?

Question 3: Considering VWA, D8, ***and*** D5, how many of your siblings have the same genotype as you do?

Question 4: Considering VWA, D8, D5, ***and*** TH01, how many of your siblings have the same genotype as you do?

Question 5: How do your results demonstrate the principle of increasing power of discrimination used in forensic DNA profiling?

Question 6: If more than one child shared the same genotype, what possible explanations are there?

Exercise 2: Random Match Probabilities

For PowerPoint presentation, lecture information, and figures for STR profiling, instructors are referred to: http://www.cstl.nist.gov/biotech/strbase/

Because each of the loci used in forensic DNA profiling is on a different chromosome, they are each inherited independently of each other (Mendel's Law of Independent Assortment of Chromosomes is the underlying genetic principle). This fact allows the forensic scientist to use the **product rule** to calculate the frequency of any given DNA profile by multiplying individual allele frequencies together. In other words, this is the probability that another person, chosen at random from a population, will have exactly the same genotype, and is also known as the **random match probability (RMP)**.

RMP = f(VWA-1) X f(VWA-2) X f(D8-1) X f(D8-2) X f(D5-1) X f(D5-2) X f(TH01-1) X f(TH01-2)
Where f(…) is the frequency of that allele in the population.

Allele frequency is basically a measure of the relative abundance of a specific allele in a given population. Allele frequencies for the 13 CODIS STR loci are available in many public databases. For this portion of the exercise, visit STRBase on the internet (http://www.cstl.nist.gov/biotech/strbase) and select "Data from NIST US Population Samples". Next click on "Allele Frequencies published in the Journal of Forensic Science" and open the article. Table 1 shows allele frequencies for a Caucasian population, table 2 for an African American population, and table 3 for a Hispanic population.

Use these tables to complete the following chart by writing the allele frequency for each of your alleles from each population:

	VWA alleles		**D8 alleles**		**D5 alleles**		**TH01 alleles**	
My genotype (from Exercise 1)								
Frequency:								
Caucasian								
African American								
Hispanic								

Question 1: What do you notice about allele frequencies among populations? Is there any specific trend?

Question 2: Forensic laboratory genotyping results often report RMPs for specific populations. Use the data in your chart to explain why this might be important. **Hint**: remember that the match probability is used to provide some indication about the "pool" of potential people with the same genotype as a suspect.

Question 3: Use the data from the chart to calculate an RMP for your own genotype for each of the populations. Insert the frequencies for your own alleles into the RMP formula and calculate.

Write the RMP formula with your alleles inserted:

Calculation based on Caucasian population:

Calculation based on African American population:

Calculation based on Hispanic population:

Discussion Questions

1. Imagine that blood, known to come from a criminal, was left at the scene of a crime, collected, and typed for the 13 CODIS loci. No suspect has been arrested, and there are no good investigative leads. Do you think that genotypes at the 13 CODIS loci should be used to make conclusions about the race of any potential suspect? Use what you have learned from the STRBase tables to support your position.

2. What are some of the difficulties in using population studies based on race?

Appendix D
PCR Amplification and Sterile Technique

PCR is a powerful and sensitive technique that allows researchers to make large amounts of DNA from very small amounts of starting material. Because of this sensitivity, contamination of PCR reactions with unwanted, extraneous DNA is always a possibility. Therefore, great care must be taken to prevent cross-contamination of samples. Steps to prevent contamination and failed experiments include:

1. **Filter-type pipet tips**. The end of the barrel of micropipets can easily become contaminated with aerosolized DNA molecules. Pipet (or aerosol barrier) tips that contain a filter can prevent aerosol contamination from micropipets. DNA molecules within the micropipet cannot pass through the filter and cannot contaminate PCR reactions. Xcluda™ aerosol barrier pipet tips (catalog #211-2006EDU and 211-2016EDU) are ideal pipet tips to use in PCR reactions.
2. **Aliquot reagents**. Sharing of reagents and multiple pipettings into the same reagent tube will likely introduce contaminants into your PCR reactions. When possible, aliquot reagents into small portions for each team, or for each student. If an aliquotted reagent tube does become contaminated, then only a minimal number of PCR reactions will become contaminated and fail.
3. **Change pipet tips**. Always change pipet tips. If a pipet tip is used repeatedly, contaminating DNA molecules on the outside of the tip will be passed into other solutions, resulting in contaminated PCR reactions. If you are unsure if your pipet tip is clean, discard the tip and get a new one. The price of a few extra tips is a lot smaller than the time, effort, and cost of failed reactions.
4. **Use good sterile technique**. When opening, aliquotting, or pipetting reagents, leave the tube open for as little time as possible. Tubes that are open and exposed to the air can easily become contaminated by DNA molecules that are aerosolized. Go into reagent tubes efficiently, and close them when you are finished pipetting. Also, try not to pick tubes up by the rim or cap as you can easily introduce contaminating DNA molecules from your fingertips.
5. **Sterilize your equipment and work area**. 10% bleach destroys DNA; wiping down surfaces and rinsing pipet barrels with 10% bleach can get rid of any DNA contamination that may arise.

Appendix E
Glossary of Terms

Allele – A version of a genetic marker, or locus.

Aliquot – The division of a quantity of material into smaller, parts.

Annealing – Binding of oligonucleotide primers to complementary sequences on the template DNA strands.

CODIS – **CO**mbined **D**NA Index **S**ystem is a federally maintained database of DNA obtained from crime scenes and convicted violent offenders.

Cofactors – Ions or small molecules needed by an enzyme to function properly. For example, *Taq* DNA polymerase needs Mg^{2+} in order to function properly. Mg^{2+} would therefore be considered a cofactor.

Denature – The process of melting apart two complementary DNA strands. In vivo denaturation is accomplished by enzymes; in PCR, denaturation is accomplished by heat.

dNTPs – Commonly used abbreviation for all four deoxynucleotide triphosphates (dATP, dTTP, dGTP, dCTP) used in synthesizing DNA.

Ethidium bromide – A fluorescent dye molecule that intercalates between DNA base pairs and fluoresces when exposed to ultraviolet light.

Eukaryotes – Organisms that are made up of cells containing a membrane-bound nucleus that contains the genetic material (DNA).

Exon – The region of a transcribed messenger RNA molecule that gets spliced together and leaves the nucleus for translation into protein sequence.

Extension – This refers to the process of *Taq* polymerase adding dNTPs (deoxynucleotide triphosphates — dATP, dTTP, dCTP, or dGTP) onto the ends of oligonucleotide primers. Extension follows the base pairing rule and proceeds in the 5' to 3' direction.

Genome – A person's complete nuclear genetic make-up. The blueprint to make exactly one particular person, tree, dog, or anything else that relies on DNA.

Genomic DNA – The sum total of the DNA that is found within the nucleus of a cell.

Genotype – The set of markers (alleles) present in a DNA sample.

Intron – The region of a transcribed messenger RNA that is spliced out of the mRNA and is not translated into protein sequence.

Locus – A genetic marker. A locus refers to a position on a chromosome, and may or may not be linked to a gene. (plural, loci)

Lysis – The process of rupturing a cell to release its components.

Master Mix – The main solution of a PCR reaction which contains all of the necessary components (dNTPs, primer, buffer, salts, polymerase, magnesium) of the reaction except the template DNA.

Nucleotides – The fundamental unit of DNA or RNA. They consist of a sugar (deoxyribose or ribose), phosphate, and nitrogenous base (adenine, thymine, cytosine, or guanine and uracil in place of thymine in RNA).

Oligonucleotide – A DNA or RNA molecule usually composed of a small number of nucleotides; see also primer.

PCR – Polymerase chain reaction. The process of amplifying or synthesizing DNA within a test tube.

Polymorphism – Literally translates as "many forms". Polymorphisms refer to genetic differences at a particular locus. A single locus may be polymorphic in different individuals, having several different alleles.

Power of Discrimination – The ability to discriminate between any two genotypes. The power of discrimination increases as more loci are analyzed.

Primers – A small series of nucleotides (usually 3-30 bases in length) that bind to a particular sequence of nucleotides on the target DNA sequence. Primers for PCR are usually synthesized in a laboratory. See also oligonucleotides.

Reagents – Materials needed to conduct an experiment. They are usually solutions or mixtures of various solutions.

Restriction Fragment Length Polymorphism (or RFLP) – A DNA test used to distinguish VNTRs. DNA is digested with enzymes, and particular sequences examined by use of a probe that binds to the DNA region of interest.

STR – Short Tandem Repeat, very small repeated DNA sequences. Repeats may be only 2 to 4 nucleotides in length. STRs are inherited, and vary from person to person, and from locus to locus. STRs are the basis for a commonly used PCR-based DNA test.

***Taq* DNA polymerase** – Heat stable DNA polymerase that was isolated from the heat tolerant bacterium *Thermus aquaticus*. This DNA polymerase is commonly used in PCR reactions.

Template – The strand of DNA that contains the target sequences of the oligonucleotide primers and that will be copied into its complementary strand.

Variable Number of Tandem Repeats (or VNTRs) – DNA sequences that are made up of large, repeated elements. The repeated DNA elements may be many kilobases in length. VNTRs are inherited, and vary from person to person, and from locus to locus.