

A DNA Fingerprinting Exercise for Any Type of Class

by Sandra G. Porter, Ph.D.

"Hands on" laboratory exercises are excellent teaching tools because they have the potential to reach students with all types of learning styles. Unfortunately the large monetary investment required to obtain supplies and equipment and train instructors prevents many schools from providing hands-on learning experiences with DNA technology. Students in non-laboratory courses are even less likely to gain an appreciation for the types of information that can be obtained from DNA analysis. Courses such as mathematics, statistics, community education, and non-major's science courses can include DNA analysis as a topic if this information is presented well in the absence of a lab. Therefore low cost alternatives to traditional laboratory exercises are desirable in order to convey principles of DNA technology to a wider variety of students. The importance of understanding DNA analysis is clear; as the use of DNA information becomes pervasive in society, everyone from a jury member to an individual applying for health insurance will have to make decisions about information derived from DNA analysis.

At Seattle Central Community College, students use an exercise based on a DNA fingerprinting technique, RFLP analysis, to determine the identity of a criminal using evidence obtained from the scene of a crime. This learning exercise can be used in either a laboratory or lecture style course and can accommodate almost any budget, as the cost is limited to providing strips of paper and scissors. Although this exercise was originally designed as part of a genetics course for biotechnology majors, it has also been used successfully for non-science majors.

Applications of DNA analysis

DNA analysis has broad applications. Doctors now use genetic tests to detect specific types of inherited disease such as Huntington's disease or cystic fibrosis. Tests have also been developed to identify an inherited predisposition to certain types of breast cancer and Alzheimer's disease. As more is learned about the information stored in DNA, DNA tests may be used more widely in preventative medicine to help individuals avoid specific foods or certain environmental conditions. DNA analysis is no longer confined to genetic and medical research. Forensic science relies heavily on the ability of DNA to identify the source of biological substances and determine who is most likely to have committed a crime. This ability to identify an individual is enhanced by the variety of substances that contain DNA, including blood, semen, saliva, hair, urine, bone, teeth, feces, and tissues. Recently, the FBI announced that they were able to match DNA samples from letters mailed to relatives by Theodore Kaczinski with DNA obtained from stamps on letters mailed by the Unabomber.

Identification of specimens using DNA has had other benefits, in one third of the cases where this technique has been used, DNA analysis has been able to exonerate people wrongly accused of crimes. Prisoners wrongly accused of rape or murder have been freed on basis of DNA evidence. DNA analysis is now a common tool for establishing paternity, and it has been called on to identify remains after tragedies such as airline accidents and the inferno at the Branch Davidian complex in Waco, Texas. Anthropologists are using DNA analysis to study the migration of human beings across the oceans and historians employ these techniques to identify genetic disease in famous individuals. The variation of DNA sequences between species and individuals has also been useful for wildlife biologists attempting to track endangered species.

Features of DNA that are important for analysis

Deoxyribonucleic acid (DNA) is composed of four different chemical building blocks called "bases". These four bases; adenine (A), guanine (G), thymine (T), and cytosine (C); are joined together in a one strand by strong covalent bonds. These two strands are held together in a double helix because bases with complementary shapes can pair with each other. Adenine is able to pair with thymine and guanine pairs with cytosine. Complementary base pairs are found along the entire length of the DNA duplex. The complementary nature of the two strands provides a basis for copying genetic information and for passing this information on to offspring.

Information is stored in DNA in the sequence of bases just as information can be stored in a book in the sequence of letters. The total amount of DNA in one cell is known as the genome. Each human cell contains approximately 3 billion base pairs of DNA organized in 23 pairs of chromosomes. An analogy could be two sets of encyclopedias with 23 different volumes in each set. Every person inherits one set of 23 chromosomes from the mother and one set of 23 chromosomes from the father. Just as the two sets of encyclopedias would be similar to each other, the two sets of chromosomes from the mother and the father are very similar. For example, the sequence of the bases in chromosome 1 is almost the same in every human, just as the first volume of an encyclopedia would be the same in all copies. However, there are specific regions in the chromosomes where the sequence can vary, on the average 1 base out of every 700-1000 bases will differ between two individuals. Because the genome contains two sets of chromosomes with 3 billion base pairs in each, approximately 2 million bases (or 0.1%) will be different. Only identical twins will share the same DNA sequence at those locations. It is this variability in sequence at specific sites that allows the identification of an individual.

Techniques used for DNA fingerprinting

DNA fingerprinting analysis relies on a combination of several different techniques. DNA must be isolated from different types of samples, digested with enzymes, and DNA fragments must be separated by size using agarose gel electrophoresis. A replica of the gel, containing the DNA fragments, is created by treating the gel with chemicals that cause the DNA to denature (separate into single strands) and then transferring the DNA to a filter. Specific pieces of DNA are detected on the filter by using a process called hybridization. Hybridization capitalizes on the complementary nature of two DNA strands. A piece of DNA called a probe is labeled to allow for detection, boiled, causing it to become single-stranded, and added to the filter. The probe DNA detects specific DNA sequences on the filter because it's only able to bind to fragments that contain the sequence of bases complementary to the probe.

In order to fit this exercise into a lecture period, a detailed discussion of the techniques used for sample isolation, blotting, and hybridization, has been omitted. While these steps are important for an in-depth understanding of DNA analysis, a sufficient understanding of the ability of DNA fingerprinting to identify an individual can be based on a few fundamental concepts. These are: DNA is cut into different sized pieces by enzymes that recognize specific DNA sequences, the sizes of these fragments can be measured and will vary between individuals, each pattern of fragments occurs with a certain frequency in a population, and we can calculate the likelihood of finding a specific combination of patterns within a population.

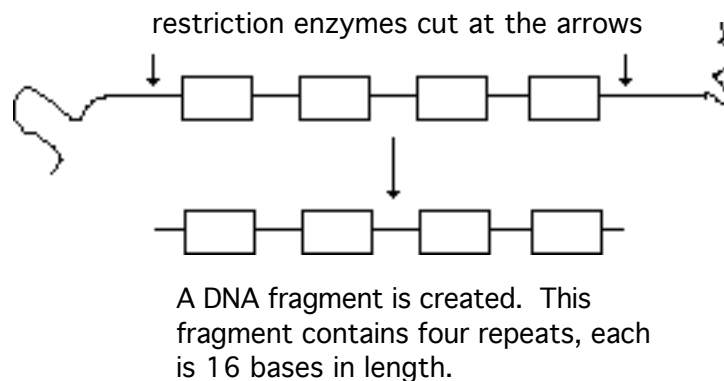
DNA fingerprinting with restriction fragment length polymorphisms (RFLP). Two alternative methods can be employed for DNA fingerprinting, these are RFLP analysis and PCR. Both methods are able to identify patterns of specific DNA sequences from a wide variety of biological samples. Samples that are in good condition and contain

enough DNA can be analyzed by looking for specific restriction fragment length polymorphisms (RFLPs) that occur in highly variable, non-coding, parts of the genome. RFLPs are patterns of DNA fragments of different lengths created when restriction enzymes cut DNA at specific sequences. The fragments are said to be polymorphic because the sizes can vary between individuals.

VNTRs (variable number of tandem repeats) are a type of DNA sequence often used for forensic analysis. VNTRs can vary in length between 1000 to 20,000 bases, within a VNTR there are base sequences of a shorter length (15-75 bases) that are repeated, with the number of repeat units varying between individuals.

If a volume from a set of encyclopedias is used as an analogy for the DNA in one chromosome, a VNTR could be thought of as a paragraph. This paragraph would contain the same sentence repeated over and over again. Every edition of that volume would contain this same paragraph on the same page, but the paragraph would be a different length in each book. The length would vary from edition to edition because the sentence would be repeated in different numbers each time the volume was printed.

An example of a VNTR found in humans is a chromosomal site known as D1S80. The size of the repeating sequence in D1S80 is 16 base pairs long. The D1S80 sequence is located on chromosome 1 and is usually repeated between 14 and 40 times. Restriction enzymes are used to cut DNA on both sides of the repeating sequence, generating a fragment whose length is determined by the number of repeats as shown below.



Use of the polymerase chain reaction (PCR) for DNA fingerprinting. Often DNA samples obtained from crime scenes are too small in quantity or too degraded by sunlight or high temperature to be analyzed by the RFLP method, these samples are subjected to a different fingerprinting technique known as PCR (polymerase chain reaction). PCR is a valuable technique because it provides a method for producing millions of copies of small regions of DNA. Again, in comparing a chromosome to a book, PCR could be thought of as a molecular Xerox machine that would produce several million copies of a single paragraph.

A comparison of the RFLP and PCR techniques. The polymerase chain reaction (PCR) is a far more sensitive technique than RFLP analysis because the reaction only requires a tiny amount of DNA. RFLP analysis requires at least 50 nanograms of DNA where only two nanograms (the amount of DNA in 400 cells) is needed for PCR. PCR is also faster; this technique can be performed in 2-3 days rather than the 4-8 weeks required for RFLP analysis.

The advantage of the RFLP method is that regions of the genome are examined that show more variation than those typically analyzed by PCR. A particular RFLP "fingerprint" might occur in 1 out of every 100,000 to 100 million people, while an individual PCR fingerprint would occur more frequently, on the order of 1 per few thousand people.

DNA fingerprinting in class

This exercise focuses on RFLP analysis using strips of paper with sequences of letters to represent segments of DNA. Students identify specific sequences (restriction sites) on each paper strip and cut (or tear) the paper wherever that restriction site occurs. Instead of using agarose gel electrophoresis, they determine the size of each DNA fragment by counting the letters. The sequences of DNA used for this exercise are much shorter than the DNA fragments analyzed in real-life laboratories, this change has been made in order to minimize the time spent counting letters. After the students find all of the restriction sites and count the letters in each fragment, the data are pooled and used to estimate the frequency of different restriction fragment patterns in a sample population. These results are used to determine if a blood sample from a crime scene contained DNA from the victim or any one of three different suspects. The class also calculates the probability of any other person having a DNA fingerprint identical to the suspect.

Description of the crime. To make the exercise more interesting, crime scenarios can be borrowed from the newspaper (O. J. and Nicole Simpson are an obvious choice) or any mystery novel. Several real-life cases already exist as a matter of public record. Here, the identities of the victim and the suspects are borrowed from the Parker Brother's game, Clue[®]. Miss Scarlet's body was found at midnight in Mrs. Peacock's library by the maid (Mrs. White), apparently the victim of a murderer. A candlestick lay near the body in a pool of blood. The suspects include Mrs. Peacock, Professor Plum, and Mrs. White. Mrs. Peacock, a wealthy middle-aged heiress, was known to hate Miss Scarlet for stealing the attentions of her ex-suitor, Colonel Mustard. Professor Plum, the noted con artist, may have had a reason to silence Miss Scarlet, only she knew his true identity. Perhaps even sweet Mrs. White had her own reasons. She despised Miss Scarlet and blamed Miss Scarlet for the death of her only son.

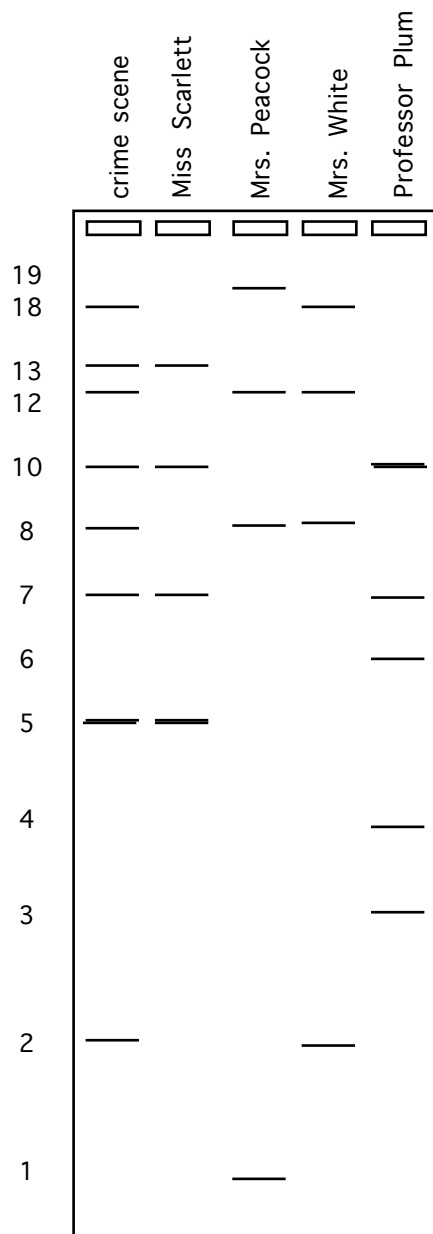
Pieces of paper with DNA sequences (one strand only) are handed out to all members of the class. Examples are given at the end of this paper. Each sample is labeled to indicate the source. A minimum number of DNA sequences would be two strips of paper for each of the suspects, four for the crime scene, two for the victim, and at least ten for the general population (between 18-22, depending on the number of suspects). The size of the general population sample can be increased so that every student gets their own DNA sequence.

The crime scene is described to the students and they are told that the class will work together to identify the potential murderer. A picture of a double-stranded DNA molecule is shown to the class. At this point it can be helpful to ask the class how many different letters they see and explain that each letter of the four letters in their sequence represents a specific chemical structure called a "base". Although DNA has two strands, they've only been given the sequence of one strand because in DNA, the sequence of one strand determines the sequence of the other. As an additional exercise, the students can determine the sequence of the complementary strand. Students are then told that special proteins called "restriction enzymes" are able to find specific base sequences and cut the DNA wherever that sequence occurs.

Digestion of the DNA. In this class, the students will replace the restriction enzymes. They are told to look for the following sequence: 5' GAATTC 3' (the 5' and 3' numbers show the orientation of the DNA sequence); when they find this sequence, they are to act like the restriction enzyme, EcoRI, and cut the sequence after the "G". Then they count the number of letters (bases) in each "fragment". Some DNA sequences have one site for EcoRI, some have multiple sites, the students need to cut this sequence every time it occurs. Students could also draw a line after the "G" with a pen and count the number of bases on each side of the line(s).

Collecting and analyzing the data. After students have counted the number of bases in each EcoRI fragment, the results are tallied and written in a table on the board (or overhead) as shown below. It's best to have the table written out ahead of time so that the numbers can be filled in during class. The students are also asked what conclusions can be drawn from the data at this point. They should notice that the DNA fragments can be different sizes and that the sizes can vary between individuals. A picture of gel is also shown to illustrate how these fragments might look when separated by agarose gel electrophoresis. The movement of DNA fragments through agarose gels is inversely proportional to the log of their molecular weight. In other words, smaller fragments move faster than larger fragments and it's easier to distinguish differences in size between smaller fragments than between larger fragments. It might be helpful to point out too, that real-life gels have less resolving power than the gel in the picture and that it would be impossible to distinguish between DNA fragments that only differ in size by one base using an agarose gel.

Analysis of the pattern of fragments in the gel. The picture of a gel below shows that the differently sized DNA fragments create visual patterns. Using the gel, the students can begin to solve the mystery by trying to identify the source of the blood found at the crime scene.



In this case, the pattern of restriction fragments from blood found at the crime scene is compared to the patterns from the potential suspects and the victim. The table and the gel (below) show that all of the DNA fragments (13, 10, 7, 5) found in Miss Scarlet's blood are present in the sample from crime scene. The two five letter fragments show students what might happen when two DNA fragments are the same or close to the same

size. Professor Plum's DNA fragments don't match any of the fragments from the crime scene DNA, so he can be ruled out as a suspect. Next, the pattern from Mrs. Peacock's blood contains two fragments (12 and 8) that are the same size as DNA fragments in those from the crime scene. However, the other fragments seen in her blood (19 and 1) are missing from the crime scene sample, allowing her to be ruled out as well. All of the fragments seen in Mrs. White's fingerprint can be found in the crime scene sample making her the most likely suspect.

Students are asked to give explanations for the presence of the 12 and 8 letter fragments in samples from both Mrs. Peacock and Mrs. White. This pattern could arise if the two women are related or if this pattern were common in the population.

Statistical analysis. The probability of finding Mrs. White's DNA fingerprint in the general population is then calculated. The probability of finding any one pattern is equal to the number of times that a specific pattern is found divided by the sample size. The sample size will be determined by the number of students holding the "General population" sequences. This will vary for every class, so it's best to draw a table like the one shown below on an overhead transparency and fill in the numbers during the exercise.

The probability of finding a combination of two restriction patterns is determined by multiplying the probability of finding each pattern, this relationship is also known as the product rule. The probability of finding a pattern that matches Mrs. White's, is = $(3/30)(5/30) = 1/60$. One person in a group of sixty people might have the same fingerprint as Mrs. White. Of course the calculation might be a bit misleading because a random group of 60 people wouldn't have been in Mrs. Peacock's living room in the middle of the night. It is important to remember that the circumstances surrounding a crime are important, too.

Identity of samples	fragment sizes	number of students with this pattern	probability of seeing each pattern
crime scene sample	13, 7	1	
	10, 5, 5	1	
	2, 18	1	
	8, 12	1	
Mrs. White	2, 18	1	3/30
	8, 12		5/30
general population	13, 7	2	2/30
(sample size = 30)	10, 5, 5	10	10/30
	2, 18	3	3/30
	8, 12	5	5/30
	6, 4, 10	4	4/30
	3, 10, 7	5	5/30
	1, 19	1	1/30

Analysis of multiple sites in the genome. The results of RFLP analysis become more convincing with data from additional regions of the genome. To illustrate why this is true, data from other portions of the genome can be included in this exercise. For these analysis, it is important that the regions of DNA tested in this process must not be linked. If two sites are linked, then they would tend to occur together and the probability calculations would be invalid. Unlinked sites are often located on separate chromosomes. A typical analysis in criminal case might examine as many as ten different sites.

To save time, students are either given the frequencies for each restriction pattern or asked to determine these numbers outside of class. The probability of obtaining

another matching pattern of fragments is then determined for the combination of all the RFLP patterns tested. In the case of Mrs. White, we might find that the restriction patterns for DNA regions, B, C, D, and E, found in her genome occur in the general population with the following frequencies, 1/20, 1/100, 1/50, 1/200. The probability of having the same set of restriction patterns as Mrs. White at regions A, B, C, D, and E, would equal $(1/60)(1/20)(1/100)(1/50)(1/40) = 1$ out of 2.4×10^8 people. As the population of the United States is close to 2.6×10^8 , this means that only 1 person in the U. S. would be likely to have a restriction pattern identical to Mrs. White.

Sources of error in DNA analysis. At the end of the exercise, drawbacks or potential sources of errors in DNA fingerprinting are discussed. Students are reminded that DNA fingerprinting can identify the most likely source of DNA in a sample, but this technique cannot determine how that DNA got into a sample. Proper sample handling and chain of custody are important aspects in proving a criminal case. Perhaps Professor Plum obtained a sample of Mrs. White's blood and poured it over the candlestick after killing Miss Scarlet. At one time it was thought to be important that the allele frequencies represented the ethnic group of the suspect. If a restriction pattern were to occur more frequently in a population, the probability of finding that pattern would be much higher. Data obtained since that time have revealed that few differences exist between different ethnic groups. The FBI has assembled a national data base of DNA fingerprints, this information has been used to track sex offenders and identify serial killers. The FBI has also created a computer program for determining the sizes of the DNA fragments separated by agarose gel electrophoresis. This program is used to determine the sizes of DNA fragments in a consistent manner and minimize human error. A detailed description of DNA fingerprinting in forensics and paternity testing can be found in the book "DNA in the courtroom" by Howard Coleman and Eric Swenson (1994, GeneLex Press, Seattle, WA).

As noted earlier, we have avoided a discussion of DNA hybridization and in depth descriptions of many of the techniques used in producing a DNA fingerprint. In this case, omitting some of the technical details makes the subject easier to teach in non-genetics courses such as mathematics and statistics. The ability to relate a discussion of probability to forensic analysis could make this subject more interesting for students of mathematics. For genetics students, additional subjects can be introduced and discussed before or after this exercise. RFLP analysis and PCR can be discussed in more detail along with the relative merits and drawbacks to both methods. Through this exercise, students at all levels gain an understanding of the power of DNA fingerprinting as well as the drawbacks.

General population

5' GAATTCCCATACGAGTTCCC 3'

Mrs. Peacock

5' GAATTCCCATACGAGTTCCC 3'

General population

5' GGAATTCCATACGAGTTCCC 3'

crime scene

5' GGAATTCCATACGAGTTCCC 3'

Mrs. White

5' GGAATTCCATACGAGTTCCC 3'

General population

5' GGGGAATTCATACGAATTCCC 3'

Professor Plum

5' GGGGAATTCATACGAATTCCC 3'

General population

5' GGGGAATTATACGGAATTCC
3'

Professor Plum

5' GGGGAATTATACGGAATTCC
3'

General population

5' GGGATTCCATACGAATTCCC 3'

Miss Scarlett

5' GGGATTCCATACGAATTCCC 3'

crime scene

5' GGGATTCCATACGAATTCCC 3'

General population

5' GGGGGAATTCACGAGAATTC
3'

Miss Scarlett

5' GGGGGAATTCACGAGAATTC

3'

crime scene

5' GGGGGAATTCACGAGAATTC

3'

General population

5' GGGATTCATACGAATTCCCC 3'

crime scene

5' GGGATTCATACGAATTCCCC 3'

Mrs. White

5' GGGATTCATACGAATTCCCC 3'

Mrs. Peacock

5' GGGATTCATACGAATTCCCC 3'