DNA PROFILING USING STRs

INTRODUCTION

Most individuals of the same species, whether African elephants, portobello mushrooms, white oak trees, or humans, have nearly identical DNA. But the DNA sequence at certain locations, or **loci**, throughout the genome varies among individuals. These variations can be used to distinguish one individual from another of the same species. The process of analyzing these DNA variations for the purpose of identification is known as **DNA profiling**, or genetic fingerprinting.

DNA profiling techniques have been used for a variety of reasons, including forensic science (matching a suspect's or victim's DNA with samples found at the scene of a crime or catastrophe), paternity testing, historical investigations, missing-person investigations, identifying victims of accidents and disasters, and cataloging convicted offenders in a database. In this activity, you will learn about DNA profiling and how to apply it to solve different cases. Think what you learn in the classroom has no practical use? Think again.

PART 1: DNA PROFILING BASICS

While most of the genome is identical among individuals of the same species, differences do exist. DNA profiling takes advantage of these differences. Variations occur throughout the genome, and in particular, in regions of noncoding DNA, which is DNA that is not transcribed and translated into a protein. Variations in noncoding regions are less likely to affect an individual's phenotype, and therefore changes in these regions are less likely to be eliminated by natural selection.

DNA profiling uses a category of DNA variations called **short tandem repeats**. STRs are comprised of units of bases, typically two to five bases long, that repeat multiple times. The repeat units are found at different locations, or loci, throughout the genome.

Every STR has multiple **alleles**, or variants, each defined by the number of repeat units present or by the length of the sequence. They are surrounded by nonvariable segments of DNA known as **flanking regions**.

For example, the STR allele in **Figure 1** could be designated as "6" because the repeat unit (GATA) repeats six times, or as 70 bp (where bp stands for base pairs) because it is 70 bp in length, including the flanking regions. A different allele of this same STR would have a different number of GATA repeat units but the same flanking regions.

Flanking regions are important because knowing their sequences enables geneticists to isolate the STR using polymerase chain reaction, or PCR, amplification.
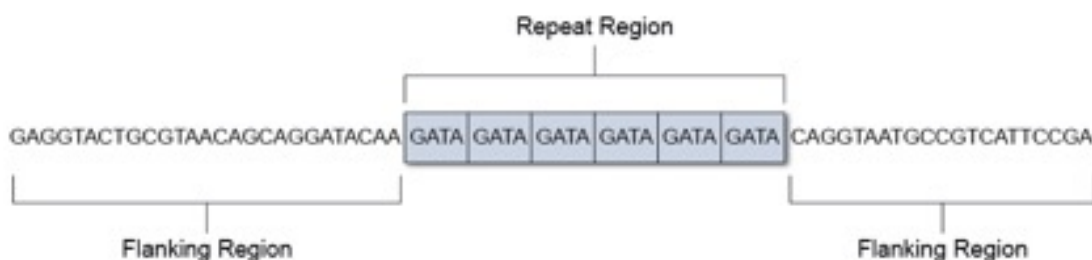
Figure 1. Each one of the rectangles above represents a repeat unit. In this example, the STR is comprised of six repeats of the four-base unit GATA. On either side of the STR is a flanking region of DNA.
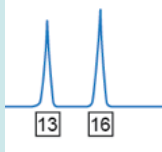
If you were to write out the STR sequence in Figure 1, it would be GATAGATAGATAGATAGATAGATA. For STRs with many repeat units, writing out the sequence can get very unwieldy, so geneticists use shorthand. The repeat unit is placed in brackets with a subscript indicating the number of times it repeats. The shorthand for the STR in the example above would be $[GATA]_6$.

1. Identify the flanking sequences and the number of repeat units [GAAT] in the following STR, known as TPOX, on human chromosome 2:

    CCACACAGGTAATGAATGAATGAATGAATGAATGCCTAAGTGCC

    a. partial flanking sequences: _____ and _____

    b. number of repeat units: _____

2. Write out the STR shown above using genetic shorthand. _____

STRs are inherited just like any gene or segment of DNA. Every individual has two alleles per STR, one inherited from each parent. However, many different alleles are often present within a population. If the inherited alleles for a given STR in an individual are identical (i.e., contain the same number of repeat units), the individual is **homozygous** for that STR. If the individual has inherited two different alleles for a given STR, then he or she is **heterozygous** for that STR. **Figure 2** shows a simple model of STR inheritance.
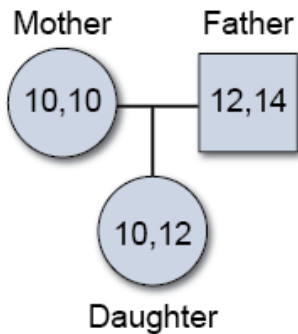

Figure 2. Model of inheritance for a single STR. The numbers refer to the allele number. Here, the mother is homozygous for an allele with 10 repeat units. The father is heterozygous at this STR locus, with one allele having 12 repeats and the other having 14. Their daughter is heterozygous, having inherited allele 10 from her mother and allele 12 from her father.

3. Is there any way that the daughter in Figure 3 could have been homozygous for the STR shown? Explain your answer.

4. Using the diagram on the right, identify all of the possible STR allele combinations that the son could have inherited at this particular locus.

    _____
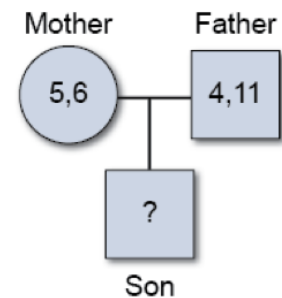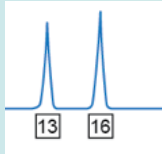
PART 2: HOW DNA PROFILING IS DONE

So how do scientists analyze DNA to determine an individual's STR alleles?

First, DNA is extracted and isolated from the other materials in an individual's cells. Once extracted, the target STRs are amplified by PCR. Different STRs are tagged using a different fluorescent marker so that they can easily be detected and differentiated. For efficiency, many STRs are analyzed at once during genetic profiling.

You may be familiar with gel electrophoresis as a technique for separating DNA fragments by size. Researchers analyzing many fragments from many samples use a variation of the technique called **capillary electrophoresis**, which is faster and more easily automated.

Figure 3 shows a model of capillary electrophoresis. As DNA samples go through the system, the results appear as an electropherogram, or a graph showing the quantity of light at specific wavelengths detected over time. DNA fragments appear as different-colored peaks that can easily be compared across samples and to the DNA ladder.
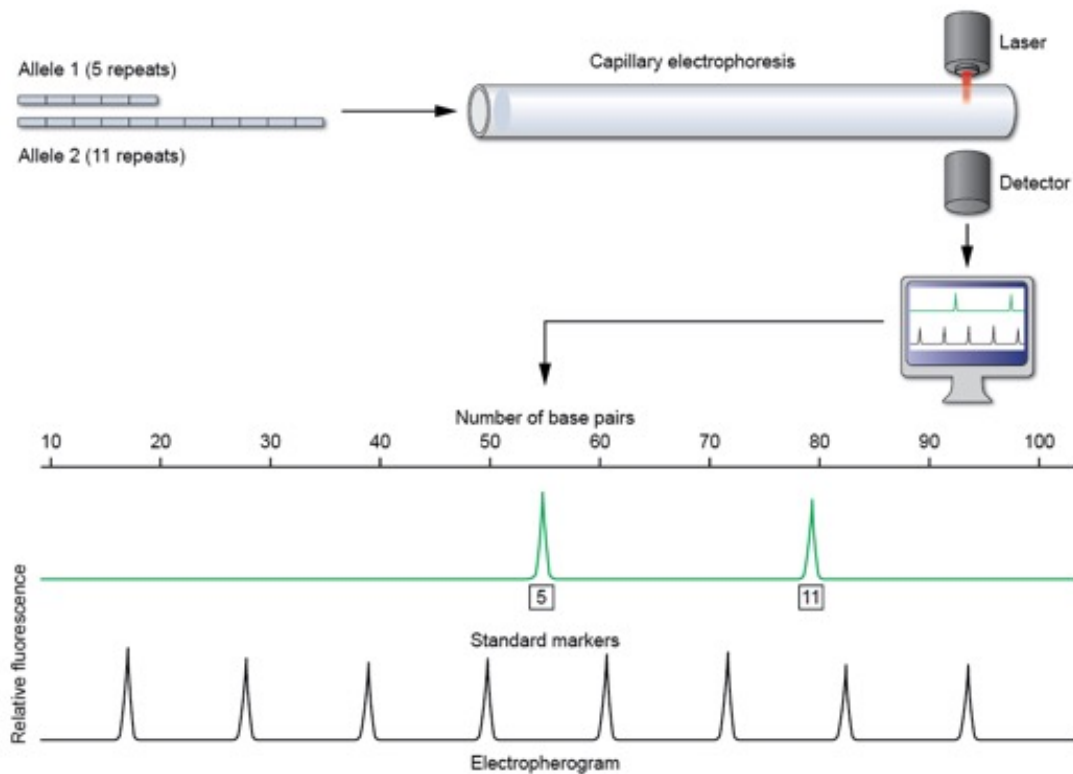
Figure 3. A simplified schematic of how an electropherogram for a single STR is generated. In this example, the individual has one allele with five repeats and another with 11 repeats. During PCR, the DNA fragments are amplified and a fluorescent tag is applied. After PCR, the DNA samples are run through ultrathin capillary tubes filled with a polymer. An electrical field is then applied, causing the negatively charged DNA fragments to move toward a positively charged electrode, with larger DNA fragments moving more slowly than smaller fragments. As the sample is drawn through the capillary tube, a laser excites the tags and the emitted fluorescence is measured by an electronic camera (the "detector"). Fragments of the same length pass the detector at the same time. Running fragments of known DNA sizes (known as a DNA ladder or DNA standards) through the system allows scientists to determine the sizes of alleles in the DNA samples. The alleles are shown in the electropherogram as labeled peaks. This technique also allows scientists to construct a graph that relates fragment sizes (measured as number of base pairs) versus time. Computer software uses data from the detector to determine the sizes of alleles in the DNA samples, in this example 55 bp (allele 5) and 79 bp (allele 11).

Example of a DNA Profile
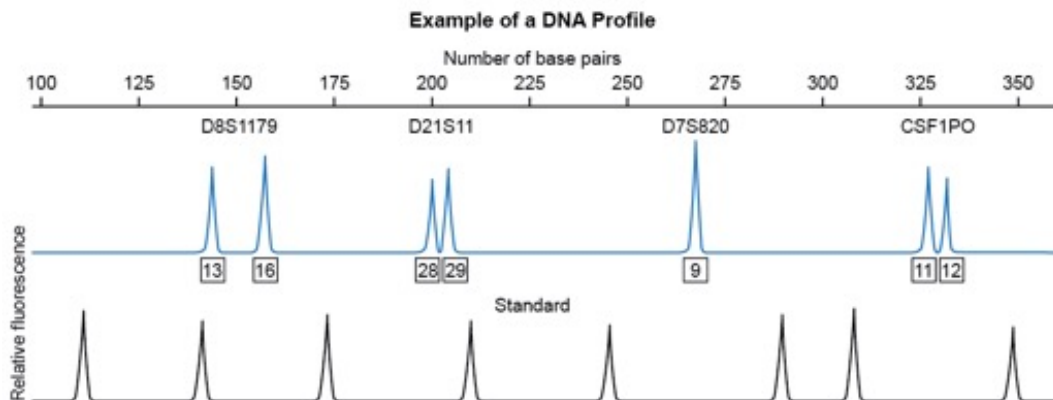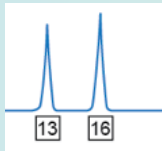
Figure 4. An electropherogram from an individual, showing the alleles at four STR loci: D8S1179, D21S11, D7S820, and CSF1PO. Each locus shows the alleles below it as peaks in the electropherogram. The number of repeats for each allele is indicated.

5. Examine the electropherogram in Figure 4. Describe how you determine whether an individual is heterozygous or homozygous for a particular STR.

6. List the STR locus or loci at which this individual is homozygous. _____

7. Which locus has the longest DNA fragments? _____ How do you know?

Different individuals have unique sets, or profiles, of STR alleles. For many forensic analyses, researchers use a core set of 13 STR loci established in 1996 by the FBI Laboratory. They are CSF1PO, FGA, TH01, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, and D21S11. In addition, a gene marker called AMEL is used to determine an individual's gender and is analyzed along with the 13 STR loci as part of that individual's genetic fingerprint.

When all 13 loci are analyzed, the chance of two people who aren't identical twins having the same *exact* set of alleles for each of these 13 STRs is extremely low—some put it in the neighborhood of 1 in 10,000,000,000,000. In other words, the chance of a random perfect match at all 13 loci is about one in 10 trillion.

PART 3: BUILD A DNA PROFILE AND SOLVE A CRIME

Today's technology has made it easier to quickly and accurately generate DNA profiles. In this part of the activity, you will model the process yourself to solve a crime. Good luck, detective!

Crime Report: A thief has stolen a priceless collection of jewels from the Museum of Precious Jewels. Forensic technicians obtained skin cells from a forehead print left on the glass enclosure of the jewel exhibit. DNA has been isolated and PCR amplified for some of the standard STR loci. A partial genetic profile generated from the collected DNA is shown in **Figure 5**.

**Figure 5.** The DNA profile of the forehead print from the scene of the crime. Each colored line shows the alleles for one of four of the core CODIS STR loci (D5S818, CSF1PO, D7S820, D8S1179).

A suspect was identified in the case. Her DNA was collected, and data for the four STR loci that were included in the analysis of the forehead print were obtained. The suspect's profile data is shown on the next page.

Your task: Compare the suspect's data below to the electropherogram in Figure 5. Analyze each STR locus by locating the repeats for each locus and counting the number of repeats, and indicate whether the suspect is homozygous or heterozygous at each locus. Record your answers on the next page.

SUSPECT'S DNA PROFILE DATA

D5S818

allele 1: 5'–CAATCATAGCCACAAGATAGATAGATAGATAGATAGATAGATAGATACCAAAGAG–3'

allele 2: 5'–CAATCATAGCCACAAGATAGATAGATAGATAGATAGATAGATAGATAGATACCAAAGAG–3'

CSF1PO

allele 1: 5'–GGCCATCTTCAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATGCTAGTCC–3'

allele 2: 5'–GGCCATCTTCAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATGCTAGTCC–3'

D7S820

allele 1: 5'–CCTCATTGACGATAGATAGATAGATAGATAGATAGATACATAGTCAG–3'

allele 2: 5'–CCTCATTGACGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATACATAGTCAG–3'

D8S1179

allele 1: 5'–GTTCATTTTCATCTATCTATCTATCTATCTATCTATCTATCTATCTACGAATGTACA–3'

allele 2: 5'–GTTCATTTTCATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTACGAATGTACA–3'

Locus	Repeat unit	# of repeats Allele 1	# of repeats Allele 2	Homozygous or Heterozygous
STR D5S818 on Chromosome 5	AGAT			
STR CSF1PO on Chromosome 5	TAGA			
STR D7S820 on Chromosome 7	GATA			
STR D8S1179 on Chromosome 8	TCTA			

Now, use the data to fill in the suspect's electropherogram and answer the analysis questions.

DNA Profile of Suspect
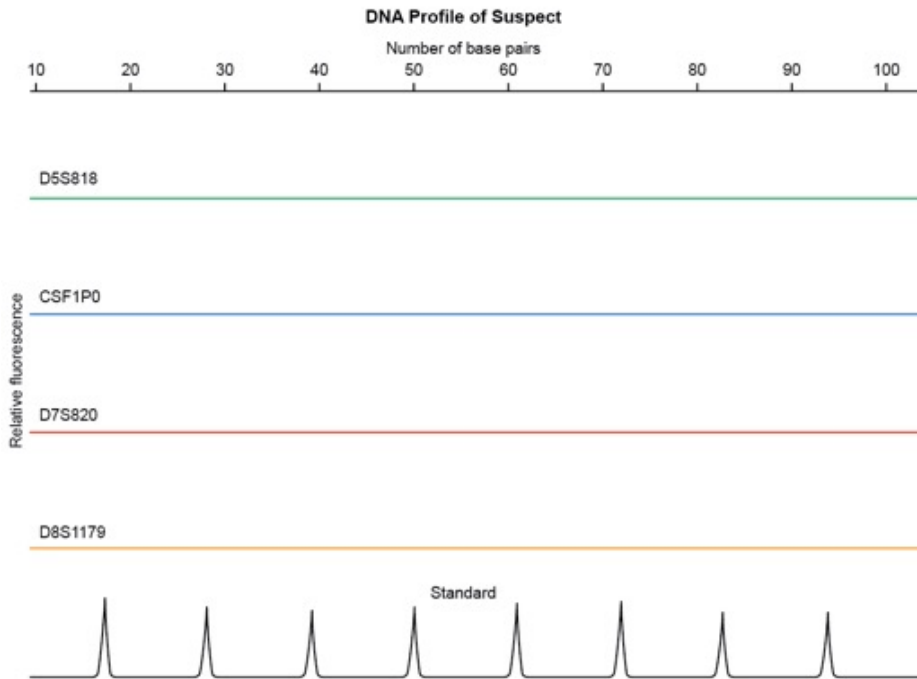
Analysis Questions:

1. Compare your DNA profile to the one generated by the suspect's DNA. Do they match?

2. Make a claim about this suspect's guilt or innocence based on this evidence. How confident are you that your claim is correct?
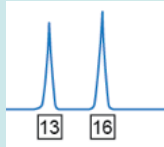
EXTENSION ACTIVITY

Allele frequencies tell you how common a given allele is in a population. An allele frequency of 0.03 means that 3% of the population can be expected to have that particular allele. The frequency in the U.S. population of each STR allele from the suspect's profile has been provided in the table below.

STR allele from suspect's DNA profile	Frequency in the U.S. population
D5S818 allele 1	0.0106
D5S818 allele 2	0.0198
CSF1PO allele 1	0.0656
CSF1PO allele 2	0.0656
D7S820 allele 1	0.0005
D7S820 allele 2	0.1361
D8S1179 allele 1	0.0787
D8S1179 allele 2	0.0787

Table 1: Frequencies for suspect alleles. (*Data source:* Promega allele frequencies https://www.promega.com/products/pm/genetic-identity/population-statistics/allele-frequencies/)

To calculate the probability of a given genotype, we can use the formulas below:

Heterozygous genotype frequency = 2*pq*, where *p* is the frequency of the first allele and *q* is the frequency of the second allele

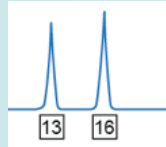**Homozygous genotype frequency = *p*²**, where *p* is the frequency of the allele

3. Using the U.S. frequency data, calculate the probability of having the given genotype at each locus. Show your work.

   a. Probability of genotype for **D5S818:** _____

   b. Probability of genotype for **CSF1PO:** _____

   c. Probability of genotype for **D7S820:** _____

   d. Probability of genotype for **D8S1179:** _____

Remember that the alleles above are inherited independently of one another. To determine the probability of a locus being inherited with another locus, we use the product rule for combining probabilities. To calculate the probability of having a given genetic fingerprint, the probabilities of having each STR genotype are multiplied by one another.

4. Calculate the probability of someone else having a DNA profile identical to that of the suspect. Show your work.

The probability calculated in Question 4 represents the likelihood that someone other than the suspect left the DNA found at the crime scene.

5. Based on your calculations, explain to the members of the jury why they should feel confident that the suspect was at the scene of the crime.