

DNA PROFILING USING STRs

OVERVIEW

This lesson is designed to give students a firm understanding of genetic profiling using short tandem repeats (STRs), which is a process used by forensics labs around the world. It can be used as an extension activity to the Click and Learn “CSI Wildlife” (<https://www.hhmi.org/biointeractive/csi-wildlife>), which explains the key biological concepts in more detail.

In Part 1 of this lesson, students learn the basics of DNA profiling, including the structure and inheritance of STRs. In Part 2, students learn how DNA profiles are compiled with STRs that are typically used in forensic investigations. In Part 3, they work through a case study involving a robbery and build a DNA profile that can be compared to one constructed from a DNA sample left by a suspect at the scene of the crime. Throughout, analysis questions walk students through calculations on allele frequency and probability (using real data from national databases), providing opportunities for formative assessments on students’ understanding of DNA fingerprinting applications.

This lesson can be used on its own or followed by the accompanying case studies in which students apply what they have learned to solve four cases: 1) twins switched at birth, 2) revisiting evidence from a crime for which the accused was sentenced to life in prison, 3) identifying a missing person, and 4) identifying victims of an earthquake.

KEY CONCEPTS AND LEARNING OBJECTIVES

- A. Regions of highly variable, noncoding, nonregulatory DNA known as short tandem repeats (STRs) are used to build genetic profiles, which can be used in forensic investigations.
- B. STRs are found across the genome. The more STR loci used to build a genetic profile, the more confident investigators can be of a positive match between samples.

Students will be able to

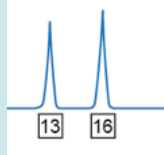
- Interpret electrophoresis results by distinguishing DNA fragments by length and determining whether individuals are homozygous or heterozygous at different STR loci.
- Calculate allele frequencies and the probability of generating a match, at random, at one or more loci using allele frequency data.

CURRICULUM CONNECTIONS

Curriculum	Standards
NGSS (2013)	HS-LS3-3
AP Biology (2012–2013)	3.A.1, 3.A.3
IB Biology (2016)	2.7, 3.5, 7.1

KEY TERMS

allele, DNA profile, electrophoresis, flanking sequence, genetic fingerprint, heterozygous, homozygous, locus (plural: loci), primer, polymerase chain reaction (PCR), repeat unit, short tandem repeat (STR)



LESSON

DNA Profiling Using STRs

TIME REQUIREMENTS

Completing this lesson and all four accompanying case studies will require up to three 50-minute class periods. However, some portions can be assigned for homework or skipped.

SUGGESTED AUDIENCE

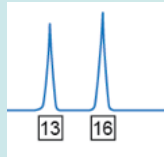
This lesson is appropriate for advanced high school biology (honors, AP, and IB) and introductory college biology.

PRIOR KNOWLEDGE

- Students should be familiar with the concept of genetic inheritance, that offspring inherit half of their DNA from each parent.
- Students will benefit from a working knowledge of how to calculate and interpret allele frequency. (The HHMI BioInteractive “CSI Wildlife” provides background information on how to do these calculations.)
- Students will benefit from prior knowledge of PCR and electrophoresis, but it is not required.

TEACHING TIPS

- You may ask your students to complete the HHMI BioInteractive “CSI Wildlife” (<http://www.hhmi.org/biointeractive/csi-wildlife>) in order to learn more about STRs and how to calculate allele frequencies and probabilities. This interactive may be assigned as homework before conducting this lesson. CSI Wildlife applies genetic profiling to elephant conservation rather than human forensic analysis and uses agarose gel electrophoresis rather than capillary electrophoresis. Capillary electrophoresis is the standard technique in forensic analysis because of the speed and ease of analysis, but the basic principles are the same as in gel electrophoresis.
- Students may wonder why there are so many STR alleles in a population. New alleles arise via mutation. Because STRs consist of repeated DNA sequences, during the process of DNA replication, the DNA polymerase can make an error and generate extra copies of the repeat unit or produce fewer ones. Mutation rates in STR repeat numbers can be up to 100,000 times higher than the rate for point mutations. Because STRs tend to be found in noncoding, nonregulatory regions of DNA, these extra copies are likely to be inconsequential to the overall phenotype and health of an individual. As a result, they avoid being eliminated from the population by natural selection.
- Students may be familiar with another type of variation called VNTR, which like STR is a sequence of DNA that consists of a repeating unit of nucleotides. VNTRs, or variable number tandem repeats, were discovered before STRs. In 1980, a young English geneticist named Alec Jeffreys was studying the evolution of a family of genes known as globins. While analyzing one particular globin gene, he found a region of the sequence that was comprised of the same set of nucleotides repeated over and over again. When he later found another region of repeats within a different globin gene, he was inspired to look throughout the genome and found dozens more. Today, more than 1000 are known. VNTRs were the first polymorphisms used in DNA profiling, and they were successfully used in forensic casework for many years. A key difference between STRs and VNTRs is that the core repeat sequence in a VNTR can range in size from six to 100 base pairs. These repeats can be represented in some alleles thousands of times, creating VNTR alleles that range in size from 500 base pairs to over 30,000 base pairs. STRs, on the other hand, have a core unit of between one and six base pairs and the repeats typically range from 50 to 300 base pairs, making them easier to analyze. The use of VNTRs is limited by the need for a relatively large amount of DNA to interpret the results. As a result, they’ve been replaced by STRs in DNA profiling.



LESSON

DNA Profiling Using STRs

- Another type of variation that students may know about are single-nucleotide polymorphisms (SNPs). SNPs normally have just two alleles—for example one allele with a guanine and one with an adenine—and therefore are not highly polymorphic like VNTRs or STRs. However, SNPs are much more abundant throughout the genome. As of 2008, over 12.8 million SNPs had been identified among humans. It is theoretically possible to type hundreds of them to generate a DNA profile. However, scientists estimate that to achieve the same discriminatory power that is achieved using 10 STRs, 50 to 80 SNPs would have to be analyzed. As a result, STRs are still the primary source of variation in DNA profiling.
- After completing this lesson, you may want to discuss with students how DNA evidence is often used to establish that a suspect was present at a crime scene, but it does not necessarily indicate that a suspect is guilty of a crime. Investigators need additional evidence to eliminate alternative explanations for the presence of a suspect's DNA at a crime scene.
- This lesson can be followed by case studies in which students apply what they learn to four different scenarios. You may have all your students do just one or all of the case studies, or divide your class into groups and assign a different case study to each group. Students can then present their findings to the rest of the class. Each case study emphasizes slightly different concepts and deals with different scenarios; read each one before assigning them to your students to make sure both the subject matter and difficulty level are appropriate.
- For more information on the Switched at Birth and Innocence Project case studies, see:
 - Switched at Birth: <http://www.nytimes.com/2015/07/12/magazine/the-mixed-up-brothers-of-bogota.html>
 - Innocence Project: <http://www.innocenceproject.org/eyewitness-recants-identification-of-andre-bryant-putting-his-conviction-into-question/>

ANSWER KEY

Part 1: DNA Profiling Basics

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

CCACACAGGTAATGAATGAATGAATGAATGCCTAAGTGCC

 - a. partial flanking sequences: CCACACAGGTAAT and GCCTAAGTGCC
 - b. number of repeat units: 5
2. *Write out the STR shown above using genetic shorthand. [GAAT]₅*
3. *Is there any way that the daughter in Figure 2 could have been homozygous for the STR shown? Explain your answer.*

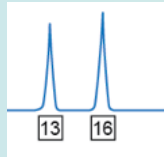
No, because the mother and father do not share any alleles.
4. *Using the diagram on the right, identify all of the possible STR allele combinations that the son could have inherited at this particular locus.*

5,4; 5,11; 6,4; 6,11

Part 2: How DNA Profiling Is Done

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

If there are two distinct peaks, the individual is heterozygous at that locus. If there is just one peak, the



LESSON
DNA Profiling Using STRs

individual is homozygous at that locus. In fact, the quantity of light absorbed can confirm that an individual is homozygous for an allele because there will be twice as many PCR fragments with the fluorescent tag, though this is not discussed in the student activity.

- List the STR locus or loci at which this individual is homozygous. D7S820
- Which locus has the longest DNA fragments? CSF1PO. How do you know?

Examining the size of the alleles by comparing the location of the peaks in the electropherogram to the known number of nucleotides in the DNA ladder shows that CSF1PO has the longest fragments: about 326 bp and 330 bp.

Part 3: Build a DNA Profile and Solve a Crime

Locus	Repeat unit	# of repeats Allele 1	# of repeats Allele 2	Homozygous or Heterozygous
STR D5S818 on Chromosome 5	AGAT	7	8	Heterozygous
STR CSF1PO on Chromosome 5	TAGA	13	13	Homozygous
STR D7S820 on Chromosome 7	GATA	6	12	Heterozygous
STR D8S1179 on Chromosome 8	TCTA	10	10	Homozygous

DNA Fingerprint:

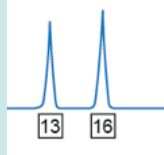
Students should draw an electropherogram identical to the one in Figure 5.

Part 3 Analysis Questions

- Compare your DNA profile to the one generated by the suspect's DNA. Do they match? Yes
- Make a claim about this suspect's guilt or innocence based on this evidence. How confident are you that your claim is correct? Most students will suggest that because the DNA profile from the suspect and the forehead print match, the suspect is guilty of the crime. However, the match in the samples only confirms that the suspect was present at the scene of the crime. To establish guilt, investigators would need to eliminate alternative explanations for the presence of the suspect's DNA at the crime scene. For example, perhaps the suspect had visited the museum earlier in the day.

Extension Activity

- Using the U.S. frequency data, calculate the probability of having the given genotype at each locus. Show your work.
 - Probability of genotype for **D5S818**: $2(0.0106)(0.0198) = 4.20 \times 10^{-4}$
 - Probability of genotype for **CSF1PO**: $(0.0656)^2 = 4.30 \times 10^{-3}$
 - Probability of genotype for **D7S820**: $2(0.0005)(0.1361) = 1.36 \times 10^{-4}$
 - Probability of genotype for **D8S1179**: $(0.0787)^2 = 6.19 \times 10^{-3}$



LESSON
DNA Profiling Using STRs

4. Calculate the probability of someone else having a DNA profile identical to that of the suspect. Show your work. Note: Answers will vary due to rounding.

$$(4.20 \times 10^{-4})(4.30 \times 10^{-3})(1.36 \times 10^{-4})(6.19 \times 10^{-3}) = 1.52 \times 10^{-12}$$

There is less than a one in 100 billion ($1/10^{11}$) chance that the DNA was left by someone other than the suspect (the actual value is one out of 657 billion)—that's a very low chance. Plus, the actual forensic profile would involve more STRs than this one, which had only four.

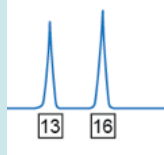
5. Based on your calculations, explain to the members of the jury why they should feel confident that the suspect was at the scene of the crime.

There is a very low probability that another person shares the same DNA profile as the suspect. Because the DNA profile left at the crime scene matches that of the suspect, the members of the jury can have a very high degree of confidence (far beyond a reasonable doubt) that the suspect was indeed at the crime scene recently.

ANSWER KEY FOR CASE STUDIES

Switched at Birth?

- Look at the data tables. Do you see any matches? Explain your findings below.
Examining the DNA profiles suggests that Carlos and William have identical profiles, as do Jorge and Wilber.
- Follow the steps and use the formulas below to calculate the probability of a resident of Bogotá having the partial genetic fingerprints of two of the twins (Carlos and Jorge).
 - Probability of having Carlos's genotype for the individual loci:
D8S1179: $(0.029)^2 = 8.41 \times 10^{-4}$; D18S51: $2(0.116)(0.048) = 1.11 \times 10^{-2}$; TPOX: $(0.036)^2 = 1.30 \times 10^{-3}$
 - Probability of having Jorge's genotype for the individual loci:
D8S1179: $2(0.346)(0.105) = 7.27 \times 10^{-2}$; D18S51: $2(0.116)(0.112) = 2.60 \times 10^{-2}$; TPOX: $(0.008)^2 = 6.40 \times 10^{-5}$
 - Calculate the probability of having Carlos's partial profile. Show your work.
 $(8.41 \times 10^{-4})(1.11 \times 10^{-2})(1.30 \times 10^{-3}) = 1.21 \times 10^{-8}$
 - Calculate the probability of having Jorge's partial profile. Show your work.
 $(7.27 \times 10^{-2})(2.60 \times 10^{-2})(6.40 \times 10^{-5}) = 1.21 \times 10^{-7}$
- Were the sets of twins switched at birth? Explain your answer using evidence from their partial genetic fingerprints and calculations.
Yes; Carlos and William and Jorge and Wilber have identical partial fingerprints, and the odds of this happening by chance—unless they are identical twins—is very low, as shown by the calculations.
- Explain why DNA fingerprints are a more reliable method of determining family relationships than blood typing. Hint: How many different blood types are there? How many different possible DNA fingerprints?
There are far more possible genetic fingerprints than there are blood types. Each STR has many alleles, but the primary blood type gene has only three (A, B, O). Adding in the Rh factor (+ or–) only brings the total number of possible combinations to 6. Because of this, it's far less likely that two individuals would have the same genetic fingerprint by chance than two individuals would have the same blood type by chance.



LESSON
DNA Profiling Using STRs

5. Can you think of a way that this discovery of identical twins switched at birth would provide scientists with an opportunity to see how certain traits are affected by the environment?

Answers will vary, but students should mention that this case study provides an opportunity to study two sets of identical twins (who have identical genetic makeup) who were raised in different environments (city versus country).

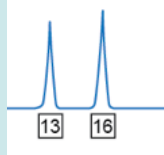
Innocence Project

1. Complete Table 1 using mathematical reasoning. Remember that the total fragment length is equal to the length of the STR plus the length of the flanking sequences. Further, flanking sequences will always be identical for each allele of a given STR.

Completed table:

STR allele from suspect's DNA profile	Repeat structure	Length of flanking sequences (bp)	Total fragment length (bp)
TPOX allele 1	[GAAT] ₉	82	118
TPOX allele 2	[GAAT] ₁₄	82	138
D18S51 allele 1	[AGAA] ₈	234	266
D18S51 allele 2	[AGAA] ₁₅	234	294
D7S820 allele 1	[GATA] ₆	233	257
D7S820 allele 2	[GATA] ₈	233	265
CSF1PO allele 1	[TAGA] ₆	267	291
CSF1PO allele 2	[TAGA] ₆	267	291

2. Draw Bryant's partial genetic fingerprint using the data from Table 1. Students should draw in peaks on the electrophoresis template at locations TPOX 118, 138; D18S51 266, 294; D7S820 257, 265; and CSF1PO 291.
3. Does the evidence suggest that Bryant is guilty or innocent of the crime? Explain your answer. The data suggests that Bryant may be innocent. If he had committed the crime, it is reasonable to expect that his DNA would have been recovered from underneath the victim's fingernails. However, the partial profiles are a match at the TPOX loci, which suggests that someone else may have committed the crime, possibly a male individual whose DNA profile matches that of the DNA found under the victim's fingernails. (Note that this evidence alone is not proof that Bryant is innocent, but suggests he may be innocent.)
4. Suggest one or two reasons why DNA fingerprinting is more reliable than an eyewitness account. Eyewitness accounts can be affected by bias, environmental conditions (e.g., visibility), memory, and other factors.



LESSON

DNA Profiling Using STRs

- Identify one or two reasons why DNA fingerprinting isn't always possible.
There might not have been any "clean" (uncontaminated and legally acquired) DNA available to sample nor a reference to which it could be compared.

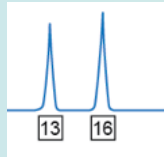
Earthquake Victims

- Based on the data in Figure 1, can you make a claim about whether one of the victims is the son of the parents in Miami? Use evidence from the figure to support your claim.
Based only on these data, Victim 2 could be the missing son: He has alleles at each of the four loci that could have been inherited from the mother and father.
- Now look at the additional data provided in Table 1. Using all the available data, can you make a claim about whether one of the victims is the son of the parents in Miami? Provide at least two pieces of evidence to support your claim.
Neither victim is the son. Of the additional STRs included in the table, none of Victim 1's loci, and just two of Victim 2's loci (D8S1179 and D21S11), have genotypes that could have been inherited from the mother and father. If the child were the missing son, then each locus would have a genotype that is a combination of the parents' genotypes.
- The cousin was on the maternal side, which means that the mother of the missing son is the aunt of the missing cousin. Is it possible that one of the victims is the nephew? Provide evidence to support your claim. (Think about what percentage of DNA a nephew shares with an aunt.)
An aunt and a nephew share approximately 25% of their DNA. Victim 2 could be the missing cousin. At four (D8S1179, CSF1PO, D21S11, and FGA) of the nine loci included in the table and four of the four loci included in Figure 1, Victim 2 and the mother have a common allele. Victim 1 only shares alleles at four of the 13 loci.
- Explain why forensic scientists typically use 13 STRs when making claims of identification rather than just a few.
It's possible for STR matches to occur by chance. By just looking at four loci in Figure 1, Victim 2 has alleles at all four loci that could have been inherited by the two worried parents. Only by looking at more loci was it possible to determine that Victim 2 was likely the cousin, not the son.

A Hotel Fire

- Multiply the probabilities of each genotype to get the probability of a person in the United States having this exact genetic profile.

$$(4.494 \times 10^{-4}) \times (5.472 \times 10^{-4}) \times (1.000 \times 10^{-6}) \times (7.903 \times 10^{-3}) \times (1.626 \times 10^{-4}) \times (4.205 \times 10^{-2}) \times (2.500 \times 10^{-7}) \times (1.400 \times 10^{-5}) \times (1.564 \times 10^{-2}) \times (3.772 \times 10^{-4}) \times (4.170 \times 10^{-2}) \times (4.028 \times 10^{-5}) \times (5.727 \times 10^{-2}) = 2.639 \times 10^{-44}$$
- The population of the United States in 2006 was about 298.4 million (2.98×10^8). Based on this information, explain why the medical examiner was confident in telling the victim's family that the recovered remains were those of their son.
If the chance of a person other than John Doe matching the remains is about one in 10^{37} , it means that statistically, in a sample of 10^{37} Americans, there would only be one, on average, with this exact fingerprint. But there were only 298 million Americans in 2006, far fewer than would be required to find another match. Therefore, the chances of the remains being someone other than John Doe are so infinitesimally small as to be essentially zero.



LESSON

DNA Profiling Using STRs

- If DNA from John Doe's toothbrush had not been available, positive identifications could also have been made by comparing the victim's DNA profile with that of a presumed parent, sibling, and/or other blood relative. Explain why such a comparison could work.

Blood relatives should share a predictable percentage of their DNA, including STRs. A parent and child or two full siblings should share about 50% of their DNA, for example; a grandparent and grandchild, uncle/aunt and niece/nephew about 25%; two first cousins about 12.5%.
- Not all STR alleles have a regular, repeating pattern. For example, one of the FGA alleles present in the profile is designated 16.2, which indicates an irregular structure. The structure of FGA is more complex than that of some of the other STRs. Instead of a simple repeat unit, it consists of the following structure: $[TTTC]_3 TTTT TTCT [CTTT]_n CTCC [TTCC]_2$, where n is a variable number of repeats. An FGA allele with a whole number designation, such as allele 16, will follow this pattern exactly and only vary in the number of variable repeats. Irregular alleles, such as 16.2, will deviate from this pattern.

Compare the structures of alleles 16 and 16.2 below:

Allele 16: $[TTTC]_3 TTTT TTCT [CTTT]_8 CTCC [TTCC]_2$

Allele 16.2: $[TTTC]_3 TTTT TT [CTTT]_9 CTCC [TTCC]_2$

a. What are the differences between the two alleles?

Two nucleotides (CT) are missing in the set right before the nine repeats of [CTTT]. Additionally, allele 16 has eight repeats of CTTT and 16.2 has nine repeats.

b. How do you think these differences arose?

These differences would have been caused by random mutations that occurred during DNA replication.

REFERENCES

Hill, C.R., Duewer, D.L., Kline, M.C., Coble, M.D., Butler, J.M. 2013. "U.S. population data for 29 autosomal STR loci." *Forensic Science International: Genetics* 7(3): e82–e83.

Jeffreys Alec J. 2005. "Genetic Fingerprinting." *Nature Medicine* 11(10): 1035–1039.

"Promega Allele Frequencies," last modified April 9, 2016: <https://www.promega.com/products/pm/genetic-identity/population-statistics/allele-frequencies/>

AUTHORS

Stephanie Keep, consultant; Melissa Csikari and Laura Bonetta, PhD, HHMI

Reviewed by Joan Bienvenue, PhD, MBA, University of Virginia; and Paul Beardsley, PhD, Cal Poly Pomona.

Copyedited by Linda Felaco