



Electrophoresis Forensics Lab



Wrongfully Convicted?

Student guide



Background information	P. 12
Today's lab	P. 17
Student lab protocol	P. 18
Pre-lab questions	P. 19
Post-lab questions	P. 22
CER table	P. 23
Extension: Probability and DNA profiles	P. 25



Background information

Forensic DNA analysis

DNA analysis is considered the gold standard in forensic science. But why is this the case? When comparing two people's DNA, more than 99% will be identical. Even with so much similarity, there are more than enough differences to make every person's genetic sequence unique. This makes DNA analysis a powerful tool for identification of individuals. Scientists can isolate DNA from biological crime scene evidence, like a bloodstain or a hair follicle, and use the information in the DNA to match the sample to a specific person.

As forensic approaches have evolved, there have been several methods for matching DNA evidence to individuals. Most forensic DNA analysis looks at regions of the genome that vary in length between individuals.

Short tandem repeats

Most modern forensic DNA analysis focuses on *short tandem repeats* (STRs). STRs are short sequences, typically 2 to 5 DNA bases, that repeat several times in a row (Figure 1). The number of times the bases repeat, however, varies between people. Therefore, the length of DNA where the STRs are located also varies between people. Forensic scientists take advantage of this and measure these differences in length to compare DNA between individuals. There are many locations in the human genome where the number of copies of a specific STR varies between people. Forensics scientists give each STR a complex name that reflects its location in the genome, but for simplicity, we will refer to an example as STR 1. At this location in the human genome, the sequence GATA is repeated anywhere between 5 and 16 times (Figure 1).

STR 1

5 REPEATS GATA GATA GATA GATA GATA

6 REPEATS GATA GATA GATA GATA GATA GATA

7 REPEATS GATA GATA GATA GATA GATA GATA GATA

UP TO 16 REPEATS

Figure 1. Short tandem repeats

STRs are short DNA sequences that repeat several times in a row. For example, STR 1 is a specific location in the human genome where the sequence GATA is repeated between 5 and 16 times.



An *allele* is one of two or more versions of a DNA sequence that appears in the same location in the genome. Forensic scientists refer to STR alleles by the number of tandem repeats they contain. In the example of STR 1, if there are 5 GATA repeats, then forensic scientists would say that it is the “5” allele. Because humans have two copies of each chromosome, one inherited from each biological parent, people have two alleles for each STR region. Scientists use the term *genotype* to refer to different combinations of alleles that a person could have. For instance, someone who is “9, 9” at STR 1 inherited 9 GATA repeats from each parent. On the other hand, someone who is “5, 9” at STR 1 inherited 5 GATA repeats from one parent and 9 GATA repeats from their other parent (Figure 2).

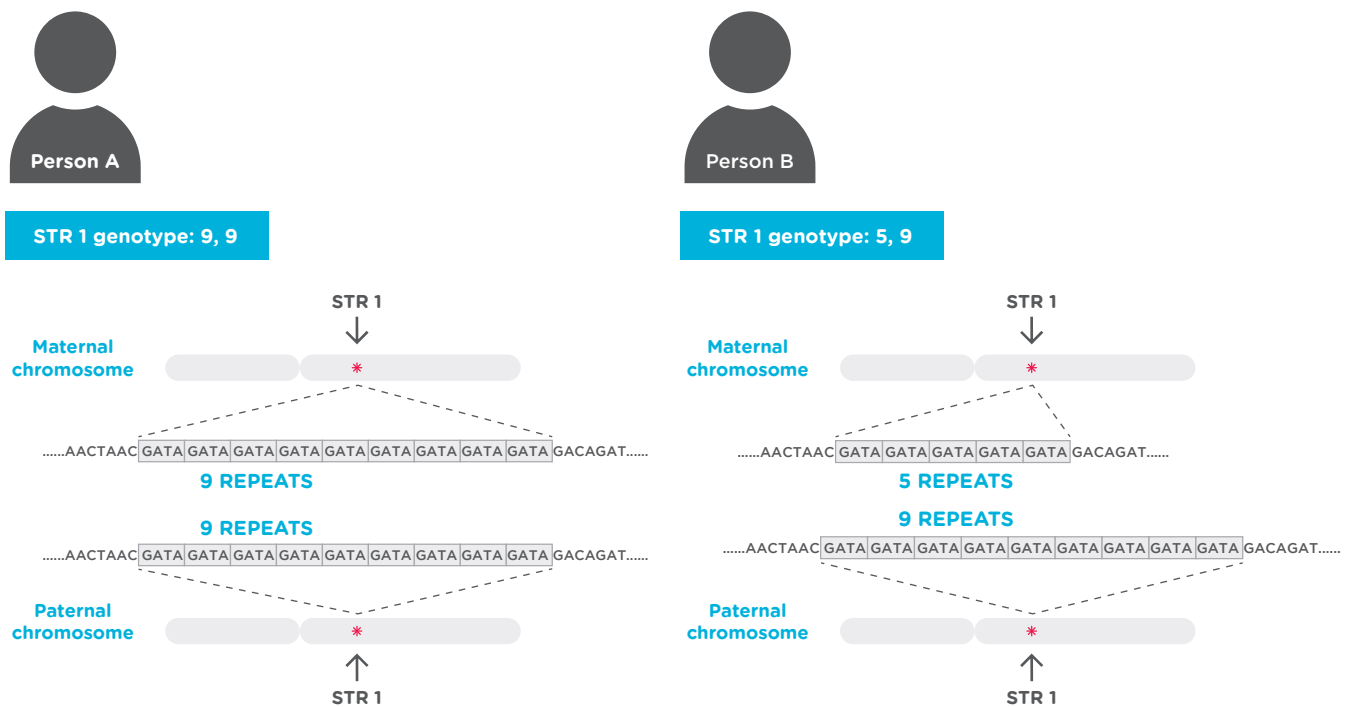


Figure 2. STR genotypes

People have two alleles for each STR region in the genome. The alleles could be the same or they could be different. Person A has two copies of the same allele for STR 1 and their genotype is 9, 9. Person B has two different alleles for STR 1 and their genotype is 5, 9.

Non-coding regions of the genome

When we think of DNA, we often think of genes, which contain the coded instructions to make RNA and proteins. However, the majority of the genome doesn't code for proteins. Scientists refer to this DNA as non-coding DNA. Non-coding DNA can have many functions. For example, non-coding DNA can be involved in gene regulation or have a structural role, like protecting the ends of chromosomes. There are also many regions of non-coding DNA where scientists don't know what the function is, or if it even has a function. Importantly, these non-coding regions are more variable between individuals than coding regions. For this reason, forensic scientists analyze non-coding regions of the genome when creating DNA profiles.



STR profile

In forensic investigations, scientists examine if a person's STR genotypes match the DNA evidence. Comparing the number of repeats at multiple locations in the genome can create a unique DNA profile for genetic identification (Figure 3). The more STR regions compared in a forensic investigation, the more likely it is to find differences between people's *STR profiles*.

Police can compare STR profiles from DNA evidence with DNA from a suspect if they have one. But law enforcement agencies can also compare DNA evidence to databases of DNA profiles. The FBI's Combined DNA Index System (CODIS) is a DNA profile database for 20 standard STR locations (Figure 4). As of 2020, CODIS contains more than 19,000,000 DNA profiles from convicted offenders and individuals who have been arrested, as well as other DNA samples from forensic investigations (FBI, 2020). A centralized DNA database means that local, state, and federal authorities can all compare DNA profiles in the same way. And as of 2020, close to 500,000 criminal investigations have been aided by CODIS in the US (FBI, 2020).

STR location	Evidence genotype	Suspect 1 genotype	Suspect 2 genotype
STR 1	5, 8	5, 8	5, 8
STR 2	12, 13	12, 13	14, 14
STR 3	11, 12	11, 12	12, 12
STR 4	12, 12	12, 12	12, 12
STR 5	12, 18	12, 18	12, 18
Result		Match	No match

Figure 3. Example STR profiles

Forensic scientists compare STR genotypes at multiple locations with variable STRs to determine whether DNA samples could have come from the same person. In this example, the STR genotypes show that suspect 1 could have been the source of the DNA evidence, but suspect 2's DNA does not match the evidence.

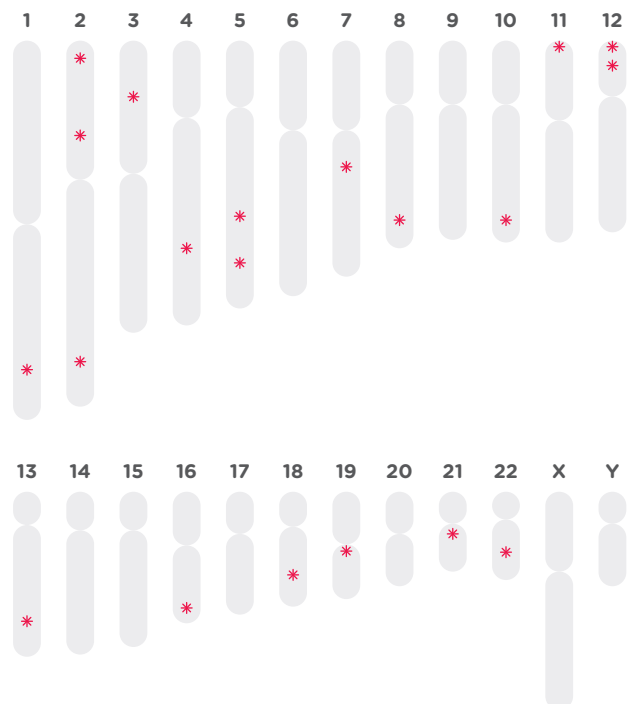


Figure 4. STR locations

Asterisks indicate the chromosomal locations of the 20 standard STR locations used in the FBI CODIS system.

FBI. "CODIS - NDIS Statistics." September 2020. <https://www.fbi.gov/services/laboratory/biometric-analysis/codis/ndis-statistics>. Accessed 11/20/2020.



Generating STR profiles

Forensic DNA analysis starts with a biological sample, such as a hair follicle or blood stain. Scientists break open the cells and extract the DNA for analysis (Figure 5). Luckily, forensic scientists only need a tiny amount of DNA for analysis. For example, crimes have been solved using DNA from a single eyelash or trace saliva on a cigarette butt. The key is a technique called polymerase chain reaction (PCR). PCR allows scientists to make many copies of specific regions of DNA, such as regions with a variable number of STRs.

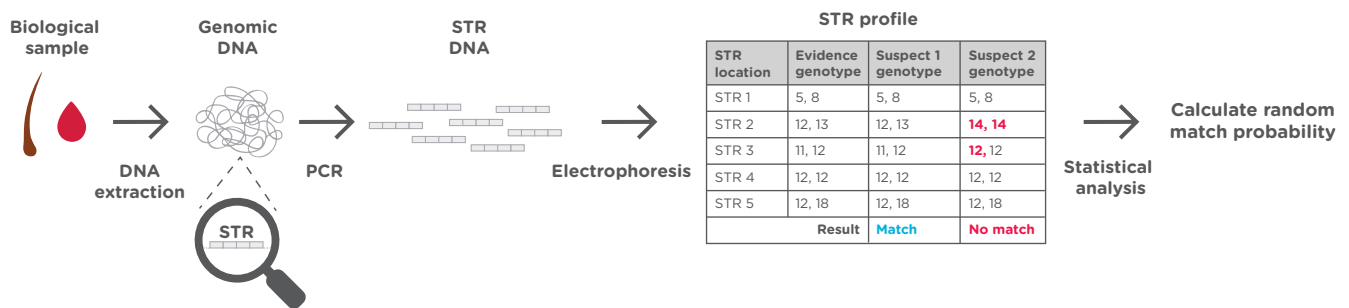


Figure 5. Forensic DNA analysis workflow

DNA is extracted from a biological sample. Then, PCR is used to make many copies of just the locations in the genome with STRs that forensic scientists want to analyze. Electrophoresis techniques separate the copied STR DNA by length and allow forensic scientists to determine the STR genotypes. If there is a DNA match, then forensic scientists perform statistical calculations to determine the strength of the DNA evidence and the likelihood of getting a random match.

Recall that the vast majority of DNA sequences are the same in all humans and that forensic analysis focuses only on regions of DNA that tend to differ between people, like STRs. PCR allows forensic scientists to take a sample like blood that contains all of an individual’s DNA and then copy just the 20 different STR locations they are interested in. After PCR, electrophoresis techniques determine the number of repeats present at each STR region (Figure 5).

Gel electrophoresis allows scientists to separate a mixture of DNA fragments based on length. Scientists use gels made of agarose, a polysaccharide extracted from seaweed, that forms a web-like mesh. An electric field causes the negatively charged DNA fragments to migrate through the gel towards the positive electrode. Smaller DNA fragments move easily through the agarose mesh and migrate quickly through the gel. Longer DNA fragments get caught in the agarose mesh and thus migrate more slowly through the gel.

Gel electrophoresis reveals “bands” of DNA fragments, with the shorter pieces of DNA being located further along the gel (Figure 6). By including a mixture of DNA fragments of known sizes, referred to as a DNA ladder, it is possible to calculate the size of unknown DNA fragments present in the samples.



Remember that STR alleles vary based on the number of repeats, and the DNA from an STR allele with more repeats will be longer than the DNA from an STR allele with fewer repeats (Figure 2). Gel electrophoresis lets scientists visualize these differences in length and differentiate between STR alleles. The gel shown in Figure 6 shows results for a single STR location. We can tell that the evidence DNA sample, suspect 1, and suspect 2 have two different alleles for this STR because the DNA samples show two bands of different sizes. On the other hand, we can tell that suspect 3 has two copies of the same allele for this STR because only one band appears on the gel.

When comparing crime scene DNA with DNA from suspects, the unique band patterns can quickly eliminate suspects whose STR profile does not match the evidence. For instance, in the gel shown in Figure 6, the DNA fragments of suspect 1 and suspect 3 do not match the evidence, so these suspects can be excluded as being the source of the DNA evidence. On the other hand, suspect 2's DNA bands do match the evidence, which means that suspect 2 cannot be excluded as being the source of the DNA evidence.

It is crucial to keep in mind that forensic investigations typically examine at least 20 STR regions, and even if someone's complete STR profile matches the DNA evidence, it does not prove their guilt. If forensic scientists get a DNA match, then the last step of forensic DNA analysis is to calculate how common that particular STR profile is (Figure 5). The *random match probability* represents the chance that two unrelated individuals share the same STR profile. It takes into account the number of STR regions compared, the number of different alleles for each STR location, and how common each allele is in the general population. Modern STR analysis can yield genetic profiles that are statistically quite rare, on the order of one in many trillions. For more information on the statistics related to STR profiles, refer to the extension activity on probability and DNA profiles (page 25).

Remember though, that even if the statistical analysis very strongly suggests a DNA match, that still doesn't prove a suspect's guilt. It only supports that the suspect's DNA was present.



Figure 6. Interpreting gel electrophoresis results
 This gel shows results examining a single STR location in the genome. By comparing the pattern of bands from the DNA evidence to the DNA of the suspects, forensic scientists can rule out suspects 1 and 3. However, suspect 2 matches the DNA evidence and cannot be excluded from further investigation.



Today's lab

Only recently has DNA analysis become standard in criminal investigations. For cases that predate the use of DNA analysis, sometimes DNA can still be collected from physical evidence stored for an extended period. In many real-world law enforcement investigations, new DNA analysis of old evidence has cracked cold cases that were previously unsolvable. In other cases, new DNA analysis from closed cases has cleared wrongly convicted individuals. Today you will use gel electrophoresis to determine if DNA from an incarcerated individual matches newly extracted DNA from old crime scene evidence.

The case

In 1999, a young man was brutally attacked and strangled in an alley. The victim was found unconscious and severely injured, but he survived. The victim told the police that while the perpetrator had been wearing a ski mask, he was very tall and had a chipped front tooth. A ski mask found in a dumpster in the alley was collected as evidence.

Later that week, an officer spotted an individual, J.M., walking down the street near the crime scene. J.M. resembled the description of the perpetrator, and he was brought in for questioning. Although the victim had never seen the perpetrator's face because of the mask, he identified J.M. in a live lineup and testified in court that he was certain that J.M. was the attacker because of his distinctive chipped front tooth. In addition, several hairs were retrieved from the ski mask found in the alley. Microscopic analysis of the hair from the ski mask was said to match a hair sample collected from J.M. Based on this evidence, the jury found J.M. guilty.

J.M. was convicted of attempted homicide and is currently serving a life sentence. J.M. has always maintained his innocence, and after years of appeals, the court has approved his request for DNA analysis of the evidence from his case. You have been sent DNA from four samples: a cheek swab from J.M., a cheek swab from the victim, and DNA from two hairs collected from the ski mask. The DNA you will test today represents the result of PCR amplification at a single STR location in the genome.

It's your job to ensure that the criminal justice system did not fail in this case. Can you prove J.M. to be innocent? Or will the evidence only implicate him further?



Student lab protocol



Protective gloves and eyewear should be worn for the entirety of this experiment.

1. Place the prepared gel into the electrophoresis chamber.
2. Add enough electrophoresis buffer to fill the chamber and just cover the gel.
 - You will need 30 ml of TBE buffer for a blueGel™ or Bandit™ electrophoresis system. Do not overfill the chamber.
 - If using another electrophoresis system, refer to the manufacturer's instructions for the recommended buffer type and volume.
3. Use a micropipette to load samples in the following order. To prevent contamination, use a new tip for each sample.
 - Well 1: 10 µl Fast DNA Ladder 1 (tube L)
 - Well 2: 10 µl Victim DNA (tube V)
 - Well 3: 10 µl J.M. DNA (tube JM)
 - Well 4: 10 µl DNA Evidence 1 (tube E1)
 - Well 5: 10 µl DNA Evidence 2 (tube E2)
4. Run the gel for 15-20 minutes.
 - The blueGel™ and Bandit™ electrophoresis systems run at a fixed voltage.
 - If using another gel electrophoresis system, set the voltage in the 70-90 V range.
5. To visualize the DNA samples, turn on the blue light in your electrophoresis system, or move the gel to a transilluminator.
6. If needed, continue to run the gel until there is sufficient separation between the bands to interpret the results.
7. If desired, take a photo to document the gel electrophoresis results.
8. Compare the bands from the DNA samples to the DNA ladder to obtain size estimates.

Detailed operating instructions for miniPCR electrophoresis systems



blueGel

<https://links.minipcr.com/blueGelRun>



Bandit

<https://links.minipcr.com/BanditViewit>





Pre-lab questions

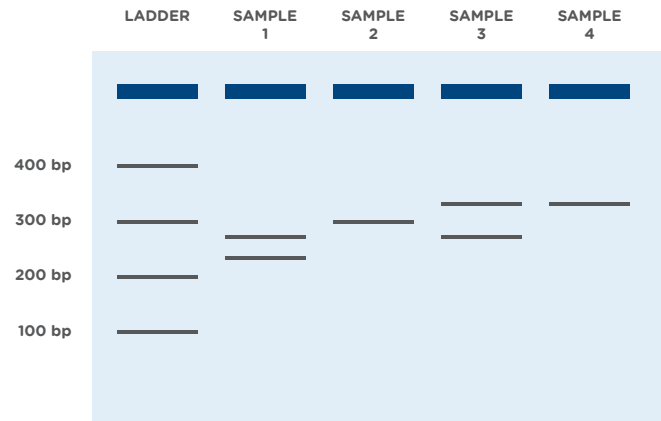
Review

1. Why is it essential to compare variable regions of DNA if you are trying to identify someone?
2. Describe how you could recognize a short tandem repeat within a DNA sequence.
3. How do different alleles for the same STR location vary?
4. Why is gel electrophoresis a good tool to differentiate between alleles for the same STR region?



Critical thinking

5. Gel electrophoresis was used to separate fragments for a single STR location, which we will call “STR 1” for simplicity.
- a. Every person has two alleles for STR 1. Which sample(s) comes from a person where the two alleles are different from each other? Explain how you can tell.



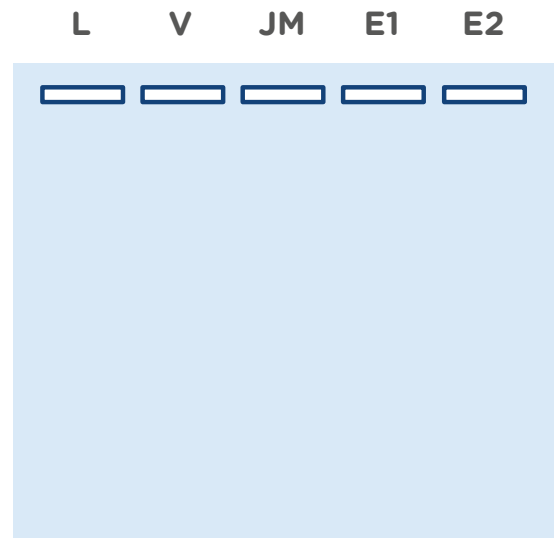
- b. Which sample(s) comes from a person where the two alleles are the same? Explain how you can tell.
- c. Do any of the samples come from people who have the same genotype? Explain how you can tell.
- d. Do any of the samples have alleles that are the same? If so, circle them on the gel and explain how you can tell.



Post-lab questions

Interpreting results

1. Use the image on the right to draw what your gel looks like. For each sample, draw the bands that you see on your actual gel.
2. Compare J.M.'s DNA with the DNA evidence from the crime scene. What can you conclude?



Critical thinking

3. Why do you think it is helpful to analyze the victim's DNA if a sample is available?
4. If you were working on this case, what would your next steps be? Explain your reasoning.



CER table

Fill in the table based on your results from the lab. Use the rubric on the next page to help your answers.

Question:

Based on your results, do you think J.M. is guilty?

Claim

Make a clear statement that answers the above question.

Evidence

Provide data from the lab that supports your claim.

Reasoning

Explain clearly why the data you presented supports your claim. Include the underlying scientific principles that link your evidence to your claim.



Score	4	3	2	1
CLAIM A statement that answers the original question/problem.	Makes a clear, accurate, and complete claim.	Makes an accurate and complete claim.	Makes an accurate but incomplete or vague claim.	Makes a claim that is inaccurate.
EVIDENCE Data from the experiment that supports the claim. Data must be relevant and sufficient to support the claim.	All of the evidence presented is highly relevant and clearly sufficient to support the claim.	Provides evidence that is relevant and sufficient to support the claim.	Provides relevant but insufficient evidence to support the claim. May include some non-relevant evidence.	Only provides evidence that does not support claim.
REASONING Explain why your evidence supports your claim. This must include scientific principles/knowledge that you have about the topic to show why the data counts as evidence.	Provides reasoning that clearly links the evidence to the claim. Relevant scientific principles are well integrated in the reasoning.	Provides reasoning that links the evidence to the claim. Relevant scientific principles are discussed.	Provides reasoning that links the evidence to the claim, but does not include relevant scientific principles or uses them incorrectly.	Provides reasoning that does not link the evidence to the claim. Does not include relevant scientific principles or uses them incorrectly.

We recommend that teachers use the following scale when assessing this assignment using the rubric. Teachers should feel free to adjust this scale to their expectations.

Rubric score	3	4	5	6	7	8	9	10	11	12
Equivalent Grade	55	60	65	70	75	80	85	90	95	100



Extension: Probability and DNA profiles

Using the product rule in forensics

If you flip a coin once, there is a 50% chance that it will land heads up. In fact, any time you flip a coin, the likelihood of getting a “heads” on that specific toss is 50% regardless of whether the coin landed heads or tails on the previous tosses. We can calculate the probability of getting any combination of heads and tails over a series of tosses using the product rule. The product rule calculates the probability of a series of independent events by multiplying each event’s probability. For example, if you want to know the probability of getting 3 “heads” in a row, you multiply 0.5 x 0.5 x 0.5 to get 0.125 or 12.5%.

The product rule can also estimate the probability of a given STR profile in the human population (Figure 1). Once forensic scientists have an STR profile with the genotypes for each STR location, they can calculate the *random match probability*. The random match probability estimates the frequency at which a given STR profile occurs in a population. This is a useful metric because it is the same as the probability that a randomly selected person would share the genetic profile being analyzed. To calculate the random match probability using the product rule, forensic scientists multiply the frequency of each STR genotype in the profile (Figure 1). The genotype frequency represents how prevalent the genotype is in the overall population.

STR location	Genotype	Genotype frequency
STR 1	7, 8	0.0054
STR 2	13, 13	0.0764
STR 3	11, 12	0.1772

$$\begin{aligned}
 \text{Random match probability} &= \text{genotype frequency STR 1} \times \text{genotype frequency STR 2} \times \text{genotype frequency STR 3} \\
 &= 0.0054 \times 0.0764 \times 0.1772 \\
 &= 7.31 \times 10^{-5}
 \end{aligned}$$

Figure 1. Calculating random match probability using the product rule

While you use the product rule for a series of coin tosses and STR profiles, STR genotypes are more complicated than a coin flip because there are more than two alleles for each STR location. Further, some STR alleles are quite common, while others are exceedingly rare. STR genotype frequencies range from more than 20% to less than 0.00003%. That means some STR genotypes are so common they are found in more than one in five people, whereas other genotypes are so rare they are found in fewer than one in thirty million people!



Interpreting random match probabilities

Multiplying the genotype frequencies for all 20 STR locations used in CODIS using the product rule gives random match probabilities that can be quite small, like one in many trillions. However, there are instances where the random match probability can be much larger. For example, if someone has several STR genotypes that are more common in the general population, then the chance of a random match will be more likely, though typically still very rare. Further, sometimes DNA evidence degrades before it is collected. In these cases, it might not be possible to analyze as many STRs. And with fewer STRs, the chance of getting a random match becomes more likely.

It is essential to note that the random match probability only represents the rarity of a specific genetic profile. Unfortunately, random match probabilities have been misinterpreted in many legal settings. For example, given a random match probability of 1 in 5 trillion, a prosecutor might claim that there is only “a 1 in 5 trillion chance that the defendant is innocent.” This error in reasoning is referred to as the prosecutor’s fallacy. It ignores additional factors that might make matching the DNA profile at the crime scene more or less likely. For example, the person may have been at the crime scene for reasons unrelated to the crime.

Naming of STR locations

Before, we referred to STR locations with simplified names like STR 1 and STR 2. In reality, forensic scientists use a complex naming system for regions of the genome with STRs. In the following questions, we will use the actual STR names because you will be using real data on the frequencies of the STR genotypes in the human population.

Most of the STR locations used in forensic analysis are in the regions of the genome that do not contain genes. These STR regions are named using a standardized system (Figure 2). Each STR name starts with a D for “DNA,” followed by a number that tells you which chromosome this STR is on. Then there is an “S” followed by another number. This second number is needed because there are many different STR regions on each chromosome, so each one needs to have a unique identifier.

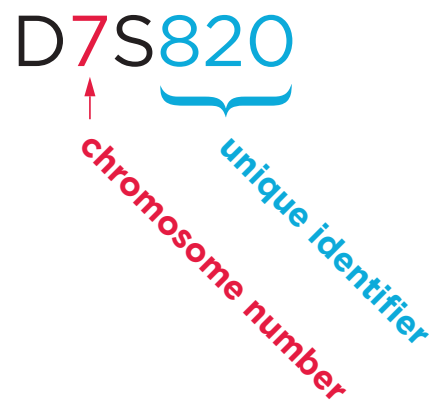


Figure 2. STR naming conventions



Forensic scientists use the frequency of each STR allele to determine the frequency of each STR genotype using the formulas below:

For genotypes with two copies of the same allele
(ex: “5, 5” or “11, 11”)
genotype frequency = p^2
where p = the frequency of the allele

For genotypes with different alleles (ex: “13, 14” or “6, 12”)
genotype frequency = $2pq$
 p = the frequency of 1st allele
 q = the frequency of the 2nd allele

D7S820 alleles	Allele frequency
7 repeats	.0164
8 repeats	.1655
9 repeats	.1216
10 repeats	.2949

Table 1. Frequencies for select D7S820 alleles

National Institute of Standards and Technology. “1036 Revised U.S. Population” July 2017. Dataset. <https://strbase.nist.gov/NISTpop.htm>. Accessed 7/29/2020.

Table 1 shows the allele frequencies for four D7S820 alleles. Scientists have compiled STR profiles from many individuals selected to represent the entire human population and have used this information to calculate each allele’s frequency. With this information, scientists can calculate the estimated frequency of every STR genotype.

2. Use the information in Table 1 to calculate genotype frequencies using the formulas above. Show your work.

a. D7S820 genotype 7, 7

b. D7S820 genotype 7, 9

c. D7S820 genotype 10, 10

d. D7S820 genotype 8, 9



In the previous questions, you focused on a single STR location, but remember that forensic scientists analyze many STR regions. Table 2 shows an STR profile that contains seven STR locations. In this table, the genotype frequencies, like those you calculated in the previous problems, have been calculated for you.

STR location	Genotype	Genotype frequency
D2S441	11, 14	0.1544
D5S818	11, 12	0.2230
D7S820	7, 8	0.0054
D10S1248	13, 13	0.0764
D13S317	11, 12	0.1772
D16S539	8, 14	.00092
D18S51	20, 21	.00069

Table 2. Genotype frequencies for select STR locations

National Institute of Standards and Technology. "1036 Revised U.S. Population" July 2017. Dataset. <https://strbase.nist.gov/NISTpop.htm>. Accessed 7/29/2020.

3. Based on the genotype frequencies in Table 2:

a. For which STR location(s) does this individual have a common genotype shared by more than 20% of the population?

b. For which STR location(s) does this individual have a genotype so rare that less than 0.1% of the population share it?

4. Imagine you are testing a crime scene DNA sample. While it is standard for forensic scientists to examine at least 20 STR locations to create a DNA profile, this has not always been the case.

a. Multiply the genotype frequencies in Table 2 to calculate the random match probability for this STR profile using the product rule. Show your work.



- b. Scientific notation is often used with very small numbers. Compare the random match probability that you just calculated with the scientific notation to the right. What order of magnitude is the random match probability?
- i. On the order of one in a million
 - ii. On the order of one in a billion
 - iii. On the order of one in a trillion

Scientific notation	In fraction form	In words
1E-6	1/1,000,000	one in a million
1E-9	1/1,000,000,000	one in a billion
1E-12	1/1,000,000,000,000	one in a trillion

Table 3. Interpreting scientific notation

- c. How confident would you be in using this DNA sample to convict someone of a crime?
5. Imagine that your DNA evidence was exposed to the elements for an extended time. Because the DNA is degraded, you only get results for the first two STR locations in the table above (D2S441 and D5S818).
- a. What would the random match probability be then? Show your work.
 - b. The current US population is 331,000,000 (331 million). How many individuals in the US are predicted to match your evidence DNA profile? Show your work.

- c. How confident would you be in using this DNA sample to convict someone of a crime?

