5.3. Forensic Applications of DNA

Forensic DNA profiling, sometimes referred to as DNA fingerprinting, testing, or typing, relies upon the uniquely individual nature of the “non-coding” or hypervariable regions of DNA. Looking at this region allows us to distinguish a DNA-containing sample of one person from another and has often provided the “smoking gun” evidence long sought for in judicial proceedings.

The first thing to understand in forensic DNA typing is that the sequence of "letters" (nucleotides) in the DNA sample is not actually read. This would be a prohibitively difficult, expensive, and time-consuming task. Instead, we look at just the regions of DNA that contain the differences observed between people. Second, DNA evidence is very good at ruling out a potential suspect as contributing a particular biological sample found at a crime scene. Third, DNA evidence cannot prove that a suspect did contribute a sample found at a crime scene - but it can tell us that we cannot exclude a suspect based upon DNA evidence and that there is a particular probability of an "accidental" random match in the population.

Two primary methods of DNA typing have emerged since it was first introduced into the forensic toolbox in 1985. These are referred to as the RFLP (restriction fragment length polymorphism) and PCR–based (polymerase chain reaction) methods of DNA testing. Both of these two techniques have their advantages and disadvantages, although the PCR method is used almost exclusively in current human forensic investigations. The RFLP method requires a much larger sample size than the PCR technique but is a more direct technique. Older or even partially degraded DNA samples are, however, often rendered unsuitable for RFLP methods. In contrast, the PCR method can use very small samples, even badly damaged and degraded samples, but is an indirect method that is very sensitive to handling and contamination problems. Techniques have now been developed that allow for the rapid analysis of forensic DNA samples, allowing such evidence to become ubiquitous in the courtroom. Many variations on these two primary techniques, including combinations of the two, have been developed for a wide range of investigations beyond human sample identification to now include plant, animal, and microbe testing.

Each of these major techniques is described in more detail in the following sections. The RFLP method is presented here both for historical reasons and since it is related to several currently used methods, especially in plant and animal DNA testing. The RFLP method, however, has essentially been replaced by the PCR-based method in the realm of human DNA analysis for courtroom proceedings.

### DNA Typing: Restriction Fragment Length Polymorphism (RFLP)

The fundamental concept behind the RFLP (restriction fragment length polymorphism) method is that a strand of DNA can be readily cut into small segments of varying sizes by severing the chain at very specific places along the DNA by special enzymes. These smaller pieces of DNA can then be arranged by size and the resulting size patterns are then compared between different samples to find similarities and differences. The key idea here is that individual people differ from one another not in the genetically coding DNA regions but in the hypervariable regions between the genetic information. Using the differences in the hypervariable regions that exist from person to person, size profiles of cut DNA samples collected from different places, such as from the crime scene or from a suspect, can be readily compared.

In our DNA, we all have very similar

---

**BRIEF ON RFLP DNA TYPING**

1. Uses restriction enzymes to cut DNA into different size segments at specific places (loci).
2. Segments are separated by gel electrophoresis based upon their sizes.
3. Smaller DNA fragments move faster in gel.
4. Fragment distributions are compared to population distributions or between samples.
repeated sections of DNA in the non-coding regions of the strands. What is different from person to person, however, is the number of repeated sequences in these sections. The RFLP method begins, after isolation and separation of the DNA in the sample, by cutting the DNA into many smaller segments. The chain is not cut randomly, however, but is instead cut at very specific locations (loci) along the DNA strand. In essence, this process cuts at the locations in DNA that have these repeated sequences. The cutting is performed by a special enzyme (called a restriction endonuclease enzyme) that recognizes just one very specific nucleotide sequence. The enzyme recognizes a certain sequence of DNA "letters", just as we recognize words from a sequence of letters, and then cuts the DNA at this place. For example, one particular restriction enzyme, called “hae III”, recognizes the “GGCC” sequence. When it finds this sequence along the DNA chain, it cuts it between the “G” and the “C” nucleotides of each “GGCC” sequence it finds – and nowhere else. Another restriction enzyme, called “apa I”, recognizes the “GGGCC” sequence (and no others) and cuts between the first and second “C” nucleotides in the sequence. The action of a restriction enzyme on DNA is shown schematically in Figure 5.3.1. Over the years, a large number of these restriction enzymes have been developed, each with a different specific DNA sequence that it recognizes and cuts along the strand.

![Figure 5.3.1. Schematic of the cutting action of the EcoRI restriction enzyme on a portion of DNA it recognizes.](http://bioinquiry.biol.vt.edu/bioinquiry/BioTech1/biotechpaid/biotechhtmls/technic5.html)

![Figure 5.3.3. Experimental electrophoresis setup.](www.science.fau.edu/chemistry/Mari/biochemlab/manual.html)

The DNA from a relatively large number of cells, typically several thousand, are required for the RFLP analysis but because the DNA from all cells from an individual are identical, all the DNA in the sample will be cut in exactly the same places along the chain. When the cutting process is complete, the original DNA molecule has been cut into many smaller lengths. Because of the variability in the DNA non-coding regions from person to person, the array of lengths of DNA fragments can be unique to a particular person. For example, in Figure 5.3.2, two strands of DNA (A and A’) are shown. Each strand has a number of “ACCT” repeated units but the total number of these sections differs between the two strands. When the strands are cut by a restriction enzyme at the “ACCT” sequence (between the two “C”'s), fragments of different lengths will be formed. The lengths of the cut fragments in reality can be quite long, up to thousands of base pairs (kilo base
pairs) with many different sized fragments generated. In the example shown in Figure 5.3.2, chromosome A is cut by the enzyme into four pieces with sizes of 2.5, 4, 5, and 6 kbp. Chromosome A’, however, is cut into only three pieces of 4, 6 and 7.5 kbp lengths. When the sizes are separated and compared, as shown at right, clear differences and similarities between the two strands are seen. The difference between the two sequences in the figure arises solely from the presence of one additional “ACCT” unit on chromosome A that is not present on chromosome A’. This simple change, however, leads to striking and clearly identifiable differences in the fragment length plot.

In large populations, there are many possibilities for the number of times that a particular nucleotide sequence, such as “ACCT”, is repeated within the DNA code. It is often quite desirable from a forensic perspective to know how often any given fragment size is found within a population, especially because we’d like to know what are the odds that a person selected from the population at random would show the same fragment length pattern as the forensic sample. It is quite different to say that there is a one-in-a-hundred chance that two randomly chosen people will have the same pattern of fragments than a one-in-a-trillion chance. To help provide an estimate of these odds, databases obtained from large numbers of people have been built that show how common a particular DNA fragment size generated by a specific cutting enzyme is within a population. How often a particular fragment is observed within the database is called the population frequency of the fragment. For example, if the frequency of two fragment sizes in a population is each one-in-four (25% or 0.25 chance), then the probability of a random match of finding someone in the population with both of these two fragment sizes is \( \frac{1}{4} \times \frac{1}{4} = \frac{1}{16} \) or one-in-sixteen. The chance of a random match goes down very rapidly as more bands are considered and can be figured out by multiplying the population frequency of each band together or, in this example, it is equal to \( (0.25)^n \) (where \( n \) is the number of bands considered). The probability of matching 10 bands in our example is \( (0.25)^{10} \) or a 1-in-1,048,576 chance. In other words, the probability of picking someone at random that had the same 10 matching bands (each with a 25% population frequency) would be slightly less than one-in-a-million. Increasing the number of probes from ten to just sixteen means that statistically you would not expect to find another random match on Earth (about a 1-in-4.3 trillion chance of two randomly chosen people having the same 16-probe profile).

The success of our entire DNA typing process depends upon our ability to separate, sort by size, and visualize the fragments after they have been cut. Luckily, this can be readily done using a common technique called gel electrophoresis (Figure 3.3.3).

Gel electrophoresis is a technique that is used for separating large molecules by their charge and size and is especially good at separating DNA fragments formed by the restriction enzymes. The basic idea behind electrophoresis is that a sample mixture of different sized fragments can be size sorted by placing the mixture on a gel and applying an electrical current through the gel. The gel

![Figure 5.3.2](image-url)
used in this analysis is usually quite similar to the gelatin used in many common food products. The molecules of the mixture move through the gel at differing rates depending upon their relative sizes and charges, with smaller and higher charged components moving most rapidly through the gel (Figure 5.3.3). The size separation occurs because frictional forces slow the larger molecules as they move through the pores in the gel much more than the smaller molecules. The entire process is shown schematically in Figure 5.3.4. The DNA fragments in the buffer solution used during the electrophoresis have an overall negative charge, due to the presence of their charged phosphate backbones. The charged fragments move through the gel when a current is applied depending upon their lengths. In DNA, the longest fragments are found nearest to the starting point while the shortest fragments are found furthest from the start. One analogy that has been used is that gel electrophoresis is like a stream containing many nets with holes in them placed across the stream (the stream’s current is analogous to the electrical charge running through the gel). The smallest fish in the stream pass through the nets most easily and will emerge downstream the fastest. Larger fish travel through the netting obstacles more slowly and, therefore, emerge at a later time. In a similar fashion, the smallest fragments pass through the gel fastest and travel the farthest. Larger fragments travel more slowly and are found closer to the starting point. Many bands are often observed and display a fingerprint-like pattern that may be unique to a particular person (Figure 5.3.5).

After the fragments have been sorted by size, something must be done in order to see the location of each fraction. This is done through a process known as blotting and hybridization. Once the fragments have moved down the gel, they are transferred to a nylon sheet (blotting). The fragments on the sheet are then typically treated with small pieces of DNA that are complimentary to the repeated section of the fragment and contain a radioactive marker. This process of hooking the radioactive markers to the DNA fragments is called hybridization. When the radioactively labeled nylon sheet is finally placed near photographic film, it exposes the film. When the film is developed, it shows the location of each fragment along the gel. Since the fragment bands have essentially taken their own pictures, this photographic is called an autoradiograph (or “autorad”).
A typical result from the gel electrophoresis of a number of samples from a forensic investigation is shown in Figure 5.3.6. In this autoradiograph, the center band came from a bloodstain found at the crime scene that was not from the victim. A DNA analysis of the bloodstain and the blood from seven suspects were run and the results are shown in the Figure. It is clear from the DNA data that all of the suspects, except suspect number 3, can be eliminated from further investigation. It does not, however, prove that the blood came from suspect number 3. In order to say that the blood sample collected from the crime scene came from suspect number 3 “beyond the shadow of a doubt”, more data would be required (different DNA probes with a complete statistical population frequency analysis).

In an actual RFLP DNA analysis, the DNA sample is first digested with the restriction enzyme to selectively slice up the strand. The mixture containing all of the different sized fragments is put into a buffer solution and is then placed in a porous gel of a gel electrophoresis instrument and, finally, the electrical current is turned on. The DNA fragments move toward the positive pole of the chamber with the relative distance that each fragment moves dependent primarily upon the size of the fragment (number of nucleotides). Finally, the distribution of the fragments along the gel is visualized and compared to other samples being considered. Comparison of the sorted fragments between two samples can then readily show their level of similarity.

Besides simply comparing DNA samples between a suspect and a sample collected from a crime scene, other kinds of forensically useful information can be gained by RFLP DNA typing. One example is tracing a person’s family tree for inheritance, immigration, or paternity claims. In this application, molecular probes are usually employed that bind to only one place in a person’s non-coding DNA makeup. The probe will attach itself to DNA fragments that have been cut at different places by the restriction enzyme. These cutting sites are inherited by people in the same way genes are inherited – some from the mother and some from the father. In Figure 5.3.6, the RFLP of one family’s inheritance through three generations is shown. A total of eight variations in the DNA code (numbered 1 thru 8) at a particular locus on a chromosome are seen in the plot. While all the children have one marker from each parent, they may have an overall composition different from one another (e.g., 2,5; 1,2; etc.). Multiple probes and suitable samples readily allow this technique to track ancestries back for generations.

**DNA Typing: Polymerase Chain Reaction Methods**

One of the most problematic features of the RFLP method is that it requires samples containing relatively large amounts of intact DNA. Often, however, samples from crime scenes are either very
small or are partially, or even badly, degraded. To overcome these difficulties, a new technique involving a process known as the polymerase chain reaction (PCR) was developed that can take extremely small quantities of incomplete DNA strands and prepare enough duplicates of the original DNA to allow for an accurate analysis. This duplication process, known as amplification, makes many, many copies of specified pieces of DNA molecules. In fact, within a few hours, an automated PCR instrument can theoretically make millions of copies of just the DNA target region from a single source molecule!

The PCR process is very similar to what happens naturally in cells when DNA is copied in the transcription process. The PCR technique is not a DNA forensic test in itself, however, but is rather a process that provides enough copies of duplicate DNA pieces to allow for typing to occur. PCR is actually used in association with a number of DNA methodologies.

As mentioned before, the PCR process employs several simple steps that beautifully mimic the natural duplication of DNA. It is important to note that it is not necessary to copy the entire DNA

---

**Case History: The Bacterial Trail**

DNA typing can also be used for tracing crops (including for advertising and breeder’s rights purposes), animals, and bacteria, beside people. Millions of dollars can be at stake from a determination of where a biological sample originated. For example, an article from the Wall Street Journal (excerpted below) describes the use of DNA typing to trace the source of an individual bacterial infection that led to legal proceedings (by Leila Abboud, Jan 21, 2003© Wall Street Journal).

“When 68-year-old Patricia Pfoutz died in July after a week of diarrhea and kidney failure, her doctors at a Whitehall, Ohio, hospital were mystified.”

“In August, public-health detectives at the Centers for Disease Control and Prevention made a startling discovery: Mrs. Pfoutz’s death was caused by a strain of E. coli bacteria with a DNA fingerprint that matched a strain found 1,300 miles away in a Greeley, Colo., meat plant.”

“…..For epidemiologists, the genetic match was a powerful illustration of the role DNA fingerprinting can play in food safety and public health. Plaintiffs’ lawyers, meanwhile, say the technique has become a formidable legal weapon in product-liability cases against food companies.”

“…. the implications of using genetic fingerprinting are huge. ‘From a burden-of-proof standpoint, it's the difference between the Wright brothers’ first flight and a trip to the moon”’.

“Before genetic testing, it was hard for public-health officials to determine where an outbreak of bacterial contamination had occurred. Epidemiologists searching for an outbreak's source would rely primarily on dietary questionnaires filled out by sick people and others in an area. Now, with DNA analysis, there is a more objective way -- although not a foolproof one -- to link each sick person directly with a particular food or product. … the [Pfoutz] suit was settled for a multimillion-dollar sum.”

---

**Brief on PCR**

1. Generates many, many duplicates of selected short regions of DNA (STRs).
2. 1st step separates DNA double strand into two single strands (denaturation).
3. 2nd step adds markers (primers) to show where to start copying DNA strands (hybridization).
4. 3rd step completes the strand by filling in complimentary bases starting at the primers (extension and ligation).
5. The cycle (steps 1 thru 3) is repeated until many copies of targeted DNA sequences have been made.
6. Every cycle doubles the amount of DNA.
in a sample for forensic purposes. Luckily, it is sufficient just to copy only the small portions of DNA that are necessary to establish a person’s identity uniquely (just the parts that are different from person to person – parts of the polymorphic hypervariable or non-coding regions).

So how does this “molecular copy machine” work? Like the paper copy machine, the PCR process involves repeated cycles of a small set of steps, or chemical reactions. Each cycle copies only a selected target region of the DNA, much as our paper copy machine copies only one page of a book rather than copying the entire book. As long as the copied page has the information we need, we’re in great shape. Similarly, with the right biochemical tools, we can copy just the DNA information we need for the analysis. Unlike copy machines, however, each PCR cycle doubles the amount of the original DNA sequence that was present at the start.

There are only four main steps in each cycle of our PCR “biological copy machine”. The first step in the PCR process, called **denaturation**, involves unraveling the DNA double helix into its two separate strands. This can be quite simply done by heating the DNA to 94° C for about one minute. Once separated into individual strands, each strand can then serve as a template upon which to rebuild a complete double strand of DNA through base pairing (just like in transcription described before). After the separation of the strands is complete, the temperature is lowered to about 60° C and small, specially constructed pieces of DNA called **primers** are added. These primers are carefully designed to mark the boundaries of the region of DNA that will be duplicated, much in the way that paper clips might be used to mark the exact pages to be copied from a book. The primers work by binding to the beginning and end of the DNA region that will be copied. These chemical markers are used to signal enzymes where to begin to match base pairs to convert the single strands into new double strands. The overall process is shown schematically in Figure 5.3.7. Let’s say that we want to make many copies of a short region of DNA labeled in the Figure as the “Short Tandem Repeat Sequence” (STR). At the first line of the Figure (labeled “1” at the right hand side), we start with a complete double strand of DNA containing a target STR region, in this case an “ACCT” sequence that is repeated four times [“ACCTACCTACCTACCT”]. The strands are then separated to give two single strands. At step two in the Figure, two different primers [“CGCAGGTT” and

![Figure 5.3.7. Steps in the first cycle of PCR duplication (amplification of a DNA target sequence.](image-url)

**Figure 5.3.7.** Steps in the first cycle of PCR duplication (amplification of a DNA target sequence.**
“CTTGCCA”) are added that have been chosen to exactly “fit” near the beginning and end of target STR sequence (like paper clip markers). For example, Primer 1 is exactly the same sequence of bases as a little bit of the DNA sequence that is just to the left of the target STR region to be copied. When Primer 1 is added to the mixture of separated strands, it will bind (a process called annealing) to the complimentary DNA strand just before the STR target region on the lower strand as shown. Primer 2, HOWEVER, is made up of a complimentary (“negative”) version of the sequence that occurs just to the right of the target STR region on the other single strand. So when Primer 2 is added to the separated strands, it will bind just to the right of the STR repeats in the top strand. By adding these two primers, we have marked out the beginning and the end of the STR section that we want to copy. In “real life”, the primers used are up to about thirty nucleotides long. In the final steps of this first cycle, another enzyme (called a DNA polymerase) first finds the primers and then fills out the double strands of the two DNA single strands in a complimentary fashion (places an “A” when it finds a “T”, and so on). This process is called extension. Lastly, the newly added and sequenced nucleotides are hooked together (like knitting) in a process called ligation to form the two fully complete double strands of DNA. This entire process is shown schematically in Figure 5.3.8 for many cycles.

There are several important things to recognize about this PCR process. First, after completing a cycle (denaturation, annealing, extension, and ligation), we now have taken one double strand and have made two identical DNA double strands that can serve as templates for the next cycle of the PCR process. In short, we have doubled the amount of DNA we had when we began by selectively

Figure 5.3.8. A schematic drawing showing how the polymerase chain reaction works to produce copies of selected regions of the DNA strand. “At right a sample of double-stranded DNA is shown unwound from one of the chromosomes of a cell nucleus. The two strands of the sample DNA are split apart (right box) and DNA primers (blue and green) added. The area to be amplified (orange and yellow) lies between the primers, and is filled in by supplying the chemical building blocks of DNA with the enzyme DNA polymerase. Repeating the process produces millions of identical copies (lower left).” (G210/540; www.sciencephoto.com).
copying the target STR region. When the cycle is run for a second time, we end up with four DNA double strands, each of which can again serve as starting templates. After cycle three, we have eight strands, after four cycles we have sixteen and exponentially so on. Since each cycle can be completed in a matter of just a few minutes, within hours we can produce millions and millions of copies of the targeted DNA regions. For example, running 32 cycles could theoretically create over one billion copies of a targeted DNA region in about three hours! An important feature to note is that not all of the DNA copies made in the PCR process are the same size, as shown in Figure 5.3.9. For example, after three cycles there are four copies of just the target STR region and four larger sections. As more cycles are completed, the relative amount of only the targeted STR region becomes very large. This is particularly nice because we are ultimately copying just the portions of DNA that we need for a forensic analysis.

![Figure 5.3.9](image)

**Figure 5.3.9.** Production of target STR region copies by PCR (adapted by JTS from Figure 1 from “DNA Typing and Protocols” see bibliography).

The PCR method differs from the naturally occurring DNA transcription process in cells in that it can copy only selected small regions of DNA, typically fewer than 1000 nucleotides, rather than duplicating the large regions necessary to code for a protein. This effectively restricts current PCR methods, however, from being applied to VNTR (and RFLP) techniques because VNTR/RFLP uses relatively large segments of DNA.

The PCR technique is readily run in the laboratory. Usually, the DNA sample to be amplified is added to a “soup” of biochemical reactants including the chosen primers, nucleotides (A, T, C, and G nucleotides), buffer, and several needed enzymes. The steps are then carried out simply by cycling the temperatures of the mixture – heat to 92° for denaturation, then cool to 60° for annealing, and finally warm to 72 for extension. Every time the temperature is cycled, another cycle of the PCR amplification is completed. Experimentally, it’s almost as simple as pushing the copy button on the copy machine!
The PCR method allows us to prepare the relatively large amounts of DNA that are needed for accurate typing from very tiny samples. It must be coupled, however, with another technique in order to complete the DNA typing analysis. This is most commonly done by looking more closely at the STR fragments that have been amplified in the PCR process.

**DNA STR Typing**

The RFLP method described earlier probes repeated sections of DNA known as VNTR, repeated consecutive sequences all connected together. These VNTR sections usually consist rather large repeated segments of DNA. The problem with using VNTR analysis, therefore, is that the fragments are quite long, up to 1000 nucleotides, and are difficult to copy using PCR methods. This means that we need a relatively large sample in order to get enough DNA to complete the analysis. Besides VNTR, another type of highly repetitious DNA is known and works quite well with PCR amplification methods. These shorter repeated sections are called short tandem repeats or STRs and are made up of short sequences (only 2 to 6 nucleotides – the FBI uses only those with 4 nucleotides) that are repeated usually between seven and twenty times (as opposed to the maximum of about fifty repeats in VNTRs). Since STRs are much shorter than VNTRs, they can be easily duplicated quickly using PCR methods. This allows both very small samples (1 to 2 nanogram) and badly degraded and broken samples of DNA that only contain short intact pieces to be successfully analyzed.

A PCR-STR analysis begins with a PCR amplification where the DNA is first denatured and the chosen primers are added. The “forward” primer is, however, attached to a small molecular dye that fluoresces (gives off light) when excited by an external beam of suitable wavelength light. The reason for the need for this fluorescent “tag” is that it allows us to very easily visualize where the different sized STR pieces are when separated without having to use the radioactive labels that were described in gel electrophoresis for the RFLP method. When we shine light on the sample, the dye molecules fluoresce and we can find the locations of the STR fragments to which they are attached by looking for the telltale light they emit back. The PCR process provides a huge numbers of copies of the targeted STR region with the fluorescent dye molecules attached needed for the analysis.

**BRIEF ON STR**

1. STRs (short tandem repeats) are small, repeated sections of DNA that are connected one to another.
2. They are duplicated and tagged to small molecules (fluorescent dyes) that give off light using the PCR technique.
3. The STRs are separated by capillary electrophoresis and their sizes determined with the aid of the fluorescent dyes.
4. The probability of a random match is determined from the relative frequency of each STR type in the population.
5. Multiple STR sequences (loci) are used in the analysis.

![Figure 5.3.10. Schematic of capillary electrophoresis (from www.bio.davidson.edu/courses/genomics/method/Capillary.html).](image)
Once copied and labeled with dye, the STRs can then be separated using a variation of the gel electrophoresis technique called **capillary electrophoresis**. This technique usually is used to separate the STRs by length (number of repeats) just like gel electrophoresis except that it uses a very narrow charged glass capillary tube instead of a flat gel plate. The different sized STR fragments are carried down the tube by a buffered solvent at rates that depend upon their charges and sizes - with smaller segments with the same charges moving faster. The smallest fragments will, therefore, reach the end of the capillary tube first with the largest fragments coming out last. Over the length of the capillary, the fragments are separated into distinct bands that contain DNA segments of the same size. As each band (corresponding to a fraction of like-sized STR pieces) moves to the end of the capillary tube, it goes past a light source and a detector that records the fluorescence light emitted by the small dye molecules attached to the STR fragments. In this way, it can be determined when each of the STR fragments passes through the capillary tube. A typical capillary electrophoresis setup is shown schematically in Figure 5.3.10. Using one set of primers, the number of sequence repeats at a particular location (locus) on the DNA can be examined. The key piece of information is how many repeats of the particular STR sequence are present at that locus on a person’s DNA. For example, using one DNA locus that the FBI has decided to examine (called locus D13S317), there are nine repeat number possibilities that have been observed in the world consisting of 5, 7, 8, 9, 10, 11, 12, 13, 14 and 15 repeats of the “TATC” sequence. Scientists have studied the occurrences of these repeated sections in many populations throughout the world. The frequency of the occurrences of these STR possibilities (each possibility is called an allele) are shown in Table 5.3.1 for a few select populations. For instance, the 5-repeat pattern is essentially unknown in the United States but does occur is about 0.5% of people living in China. In contrast, the 15-repeat pattern occurs in about 3.3% of Americans but is essentially not found in the German population (North Bavaria). The most common patterns in all of these populations are the 11 and 12-repeat versions (alleles), accounting for about a third of the population for each.

### Table 5.3.1. World distribution of the D13S317 STR fragments
(from www.uni-duesseldorf.de/WWW/MedFak/Serology/DNA-Systeme/d13s317.htm)

<table>
<thead>
<tr>
<th>Population</th>
<th>5</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>0.0000</td>
<td>0.0016</td>
<td>0.1433</td>
<td>0.0609</td>
<td>0.0556</td>
<td>0.3211</td>
<td>0.2938</td>
<td>0.0883</td>
<td>0.0359</td>
<td>0.0016</td>
</tr>
<tr>
<td>Germany</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.1500</td>
<td>0.0620</td>
<td>0.0650</td>
<td>0.2810</td>
<td>0.2920</td>
<td>0.0960</td>
<td>0.0540</td>
<td>0.0000</td>
</tr>
<tr>
<td>US</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.1026</td>
<td>0.0762</td>
<td>0.0662</td>
<td>0.3377</td>
<td>0.2682</td>
<td>0.1093</td>
<td>0.0364</td>
<td>0.0330</td>
</tr>
<tr>
<td>China</td>
<td>0.0053</td>
<td>0.0020</td>
<td>0.2713</td>
<td>0.1572</td>
<td>0.1372</td>
<td>0.2341</td>
<td>0.1556</td>
<td>0.0314</td>
<td>0.0059</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

The success of the STR analysis relies upon comparing an observed repeat pattern (how many times a sequence is repeated at a particular locus) from a forensic sample with how often it randomly occurs in a population. Figuring out the odds of a random DNA STR match with a forensic sample is much like figuring out the odds when playing cards. For example, if you were dealt one card from a full deck of cards, the chance that it would be a red card would be 50% (or 0.50) – on average you’d get a red card half of the time. The odds that another dealt card would be red would again be 50% (0.50) for that one event. But, the odds that you would be randomly dealt two reds cards in a row would be 0.50 times 0.50 = 0.25 or 25%. Thus, your odds of being dealt two sequential red cards would be 1 in 4. Using this same idea, what would be the chances of getting five red cards dealt in a row? The answer would be the chance of each separate event multiplied together or 0.50 x 0.50 x 0.50 x 0.50 x 0.50 or (0.50)⁵ = 0.031 (or 1 in about 32 times). Taking this one final step, what would be the odds of being dealt a four of hearts? The probability of this happening is calculated in the same way as being dealt five red cards – multiplying the probability of each event together. In this case, the chance of getting dealt a heart card is one in four (0.2500) and being dealt a four is one in
thirteen (0.0769) so the probability of being dealt the four of hearts is (0.2500) x (0.0769) = 0.0192 or one in fifty-two.

The probability of any particular random set of DNA STR fragments is determined in exactly the same way as calculating card probabilities. Suppose, for instance, that one of your chromosomes at the D13S317 locus (Table 5.3.1) has 9 “TATC” repeats and the other chromosome at the same locus has 11 “TATC” repeats (remember you have a pair of each chromosome so you have two possible repeat numbers at a particular site). Using the data in Table 5.3.1, it can be determined that the probability of someone chosen at random from the population of the United

![STR loci from one reaction mixture. Scale at the top is in base pairs (bp) (from www.biology.arizona.edu/human_bio/activities/blackett2/str_analysis.html)](image)

![Multiplex for 15 STR loci and the Amelogenin locus. DNA fragments are labeled in blue, green, yellow (depicted in black), and red dyes. The GeneScan size standard is labeled with orange dye. (from](image)
States having this same 9 and 11 repeat pattern would be 0.0762 times 0.3377 or 0.0257 (2.57%). As more and more STR regions on different chromosomes in the sample are used in the analysis, the probability of a random match drops very quickly. For this reason, we usually look at thirteen or more STR loci when running the DNA analysis. For example, using the thirteen STR regions employed in the O.J. Simpson DNA analysis, it was determined that the odds of someone randomly matching his DNA profile would be about one in 7 trillion.

It is important to note that the odds of a random match depends both upon what repeat numbers we use and which population we choose. So for the above example of D13S317, suppose we chose the 8 and 9 repeat variants and compare the Chinese and Australian populations. For the Chinese population, a random match would be 0.2713 (for the 8-repeat version) times 0.1572 (for the 9-repeat version) for a combined random match probability of 0.0426 or 4.3%. For Australia, it would be 0.1433 times 0.0609 for a combined probability of 0.0087 or 0.9%. Thus, it would be almost five times more likely to find a random match between a forensic sample with the 8,9 repeat versions in China as it would be in Australia. Also, in Australia the 11,12 combination (9.4% of the population) is nearly 37,000 times more common than the 7,15 version in the same population (0.0003%).

The ideal for a forensic case would be to decrease the odds of a random match to an infinitesimally small value by looking at many STR sequences. Fortunately, it is relatively easy to look at many STR locations all at once for a DNA sample. This is done by first carefully choosing and using a set of primers that will copy different non-overlapping STR regions so that we can duplicate different STR regions simultaneously through PCR amplification. This process, called **multiplexing**, is illustrated in Figure 5.3.11. In this example, three locations on the DNA strand (labeled D3S1358, vWA, and FGA) are amplified simultaneously and tagged with a blue fluorescent dye. At the top of Figure 5.3.11 is shown a set of three bars depicting the different sizes of the three non-overlapping STR location. The horizontal axis in this figure shows the number of nucleotides in the STR fragment, given as a size of the DNA fragment. The middle plot shows all of the possibilities of repeat numbers for each STR location. So, the D3S1358 locus has 8 possibilities (12 thru 19 repeats), vWA has 11 possibilities (11 thru 21 repeats) and the FGA locus has 14 possibilities (18 thru 30 repeats). The bottom plot shows an example of a “real” STR fragment size plot for these three loci for one person. It is important to note that there can be either one or two fragment sizes (alleles) seen for any given STR locus – one arising from each member of a pair of matched chromosomes. This arises because we inherit one allele (number of repeats) from our mother and one from our father. If both the mother and father “donate” chromosomes which contain the same number of repeats at the STR locus (homozygous), then only one band is seen. If, on the other hand, the inherited STRs are different from the mother and father (heterozygous), then two bands are observed. The person shown in Figure 5.3.11 is homozygous in the D3S1358 STR region (one band observed) but heterozygous in the vWA and FGA regions (two bands observed for each).

One way to look at many STR regions at a time is to use different colored fluorescent dyes attached to the primers. By combining these two ideas – different primers targeting non-overlapping regions coupled with different colored dyes – we can look at many STR regions simultaneously. This is shown in Figure 5.3.12 using five different colored dyes. By employing these concepts, the odds of a random match between a forensic sample and a population can often be reduced to one-in-a-trillion chance – often good enough to convince a jury of the connection between two biological samples.

To summarize the two main DNA typing systems discussed, RFLP and STR methods, Table 5.3.2 compares some of the major features and differences for the two types of analysis.
### Table 5.3.2. Comparison between the RFLP and STR DNA typing methods.

<table>
<thead>
<tr>
<th>Sequence Size</th>
<th>RFLP</th>
<th>STR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Employs variable nucleotide tandem repeats (VNTR) with seven to twenty-five nucleotides connected in up to fifty repeats.</td>
<td>Employs short tandem repeats (STR) with 2 to 6 nucleotides repeated consecutively 7 – 15 times.</td>
<td></td>
</tr>
<tr>
<td>Sample Size</td>
<td>Relatively large amounts of DNA are required (ca. 100m ng or the DNA from several thousand cells).</td>
<td>Relatively small amounts of DNA are required (ca. 1 ng from about 20-50 cells).</td>
</tr>
<tr>
<td>Sample Quality</td>
<td>Requires good quality (undegraded) DNA samples.</td>
<td>Works well on fragmented or partially degraded DNA.</td>
</tr>
<tr>
<td>Reliability</td>
<td>The advantage of using VNTR sequences is that there is a great deal of variability in the number of repeats possible for each DNA locus.</td>
<td>Based upon probabilities of random match in a population. Depends upon comparison with proper population and availability of statistical data.</td>
</tr>
<tr>
<td>Speed</td>
<td>Requires relatively large amounts of DNA. Relatively insensitive to impurities.</td>
<td>Requires very small DNA samples (can work on degraded samples. Very sensitive to contamination. Very large number of STR probe loci possible. Technique lends itself readily to multiplexing and automation.</td>
</tr>
<tr>
<td>Other Advantages and Problems</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Mini-STR and SNP DNA Profiling

Unfortunately, biological samples of great interest are often badly degraded from age, exposure, bleaching, or other factors. Using a relatively new technique, useful DNA forensic evidence can still sometimes be obtained from these samples. This technique employs much smaller fragments of DNA for the analysis called mini-STRs. The main difference in using mini-STRs occurs during the PCR amplification steps in which the primers (the small DNA fragments that tell the enzymes to begin or end copying the DNA strand) are moved much closer to the repeated STR regions of interest. Recently, a complete set of mini-STR primers has been developed that allows for the closest approach of the primers to the STR loci for all 13 CODIS STR (see below). This technique has been especially useful in the identification of remains from the World Trade Center attacks.

Current work is now moving beyond even the mini-STR techniques to look at micro-STR and SNP technologies. Another technique that has gained some forensic use involves exploring DNA differences as appearing as single nucleotide polymorphs (SNP, usually referred to as "SNiPs"). SNPs are variations in the DNA sequence that occur when one base pair is different from what is typically found in a population, as illustrated in Figure 5.3.13. Forensic uses of SNPs are significantly different from the other DNA typing techniques presented so far in that the SNP

---

**Figure 5.3.13.** Single nucleotide polymorphism (SNP). ([www.nlm.nih.gov/visibleproofs/education/dna/snp.pdf](http://www.nlm.nih.gov/visibleproofs/education/dna/snp.pdf))
The Lightest Touch: Touch DNA

One of the most recent advances in the forensic application of DNA involves the use of a technique known as “touch DNA”. In July of 2008, this technique proved crucial in clearing JonBenet Ramsey's family in her death and is now gaining more widespread use, especially in the field of “cold cases.” The success of touch method in both this and other cases has led to numerous requests for similar analyses in other difficult or cold cases. In response to this, DNA laboratories have reported up to 20% increases in the use of the touch DNA technique.

In the “touch” DNA technique, tiny samples are obtained from surfaces that the suspect contacts, including food, utensils, clothing, tables, glasses and the like, in order to isolate to get enough DNA containing material for the analysis. Importantly, the “touch” DNA has been used to recover viable DNA samples from surfaces that do not show a visible stain.

Since the results of the “touch” DNA analysis in the JonBenet Ramsey case were made public, these DNA tests have been used in many other "major cases." In many of these cases, the analysis has helped rule out possible suspects while in others new leads for potential suspects were developed.

In the JonBenet Ramsey case, her parents lived under suspicion for years after her death. Initially, it took police investigators seven years to send the DNA sample found in JonBenet’s undergarments out for analysis, primarily due to problems with the DNA sample that was recovered. This sample had been deemed to lack the quality needed to enter it into the law enforcement data banks for comparison. Using the new “touch” DNA method, this sample was identified as belonging to an unknown male, thus exonerating the Ramsey’s of culpability.

The DNA sample analyzed at that time was from a small drop of blood found on JonBenet. The ‘touch DNA’ technique was used to analyze the clothing that JonBenet was wearing. Analysts have matched the DNA from in skin cells found on the waistband of the murdered child’s long underwear to the DNA in the blood sample from her underpants. The ‘touch DNA’ analysis points to the presence of an unknown male, putting an end to a twelve-year nightmare for John Ramsey. Unfortunately, the mother, Patsy Ramsey, died two years earlier of ovarian cancer with the cloud of suspicion cast by the media and the police still over her head.

The technique actually looks directly at the DNA sequence itself, rather than focusing on DNA fragment lengths or number of repeats. For example, one person might have a sequence “AATCGGGACC” while someone else has a “AAACGGGACC” at the same locus. There is a SNP at the third position where a “T” in the first person has been replaced by an “A” in the second person. These SNP point changes are actually rather common, typically found every 100 to 300 base pairs along the DNA strand. Because of the high frequency of SNPs, forensic analysis requires examination of many, many SNP variations and may be very difficult and time-consuming. While it appears that SNPs will not soon replace STR-based analysis in the near future, SNPs are playing a valuable role in some forensic applications such as mitochondrial DNA (mtDNA) testing, ancestry informative markers (AIMs), Y-SNPs as lineage markers and other potential applications.