

Bioinformatics: Life Sciences on Your Computer

DNA Fingerprinting

This is supplemental reading for the video lecture on DNA Fingerprinting.

Forensics

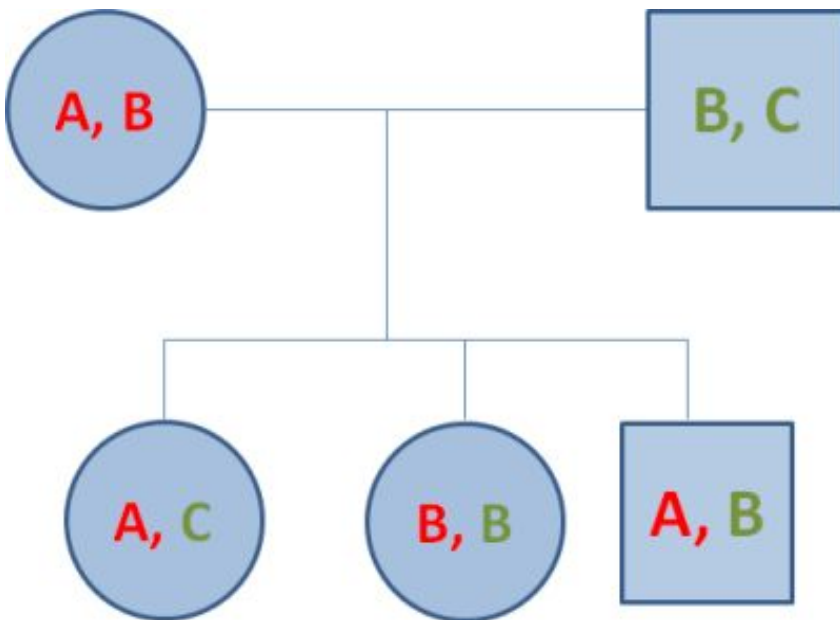
DNA fingerprinting is used to help identify crime suspects. It has been used to exonerate people who had been falsely committed of crimes. The idea is to match genotypes at a specified number of locations. A suspect who is not an exact match, at all locations, is not the person whose DNA is in the collected DNA sample. That does not answer the innocence/guilt question, however, the results can provide evidence that is used to help determine innocence or guilt.

Paternity Testing

The basis for paternity testing is the same as in forensics, however, in forensics, an exact DNA match is required. With paternity, it's more of a puzzle that involves the mother's genotype and the child's genotype. The laws of Mendelian inheritance are used to determine whether an individual is likely the father. Again, a specif number of genetic locations are used.

Autosomal Inheritance

Autosomal chromosomes are every chromosome except for the X and Y chromosomes. Humans have 22 autosomal pairs. For each pair, every individual inherited one chromosome from the mother and the other from the father. An "allele" is a genetic variant at a specific location. So, the general rule is that we get one allele from the mother and one from the father at any particular genetic location. In the diagram below, the circles represent females and the squares represent males.



Note that each child (bottom row) gets one allele (red) from mom and another allele (green) from dad. Any combination is possible. So if a child had an A, A genotype, you could conclude that the child does not have that individual as a father. The "father" has no A allele.

Y-Chromosomal inheritance

For boys, an alternate method of determining paternity is with the Y chromosome. Mothers are XX and

fathers are XY. Boys can only inherit a Y chromosome from the father. Several markers should be used, to avoid mutation. It would not be good for distinguishing paternity between two brothers. Those two brothers each got the same Y chromosome from (presumably) the same father.

Mitochondrial DNA

Mitochondria contain DNA and self-replicate sort of like bacteria. They are found in the cytoplasm of cells. Because eggs have a lot of cytoplasm and sperm have virtually none, mitochondrial DNA is maternally inherited. Usually, paternity is more in question than maternity, but in the cases of lost children of hospital swapped babies, that can be an identifier.

Shared DNA among relatives

How much DNA is shared between relatives. Well, there is some luck involved. Think of two brothers. They each get one of 22 different autosome from their mother. That's like flipping a coin 22 times. Heads, they get the same chromosome; tails, they get different ones. If you flip a coin 22 times, on average, you should expect 11 heads and 11 tails, but it doesn't always work that way. There is some luck involved. So some siblings share more than 50% DNA; some less. On average, here are some percentages.

Identical twins: 100%

Siblings: 50%

Parent/Child: 50%

Aunt/Uncle: 25%

Grandparent or grandchild: 25%

First cousins: 12.5%

Beyond first cousins, relatives share very little DNA on average.

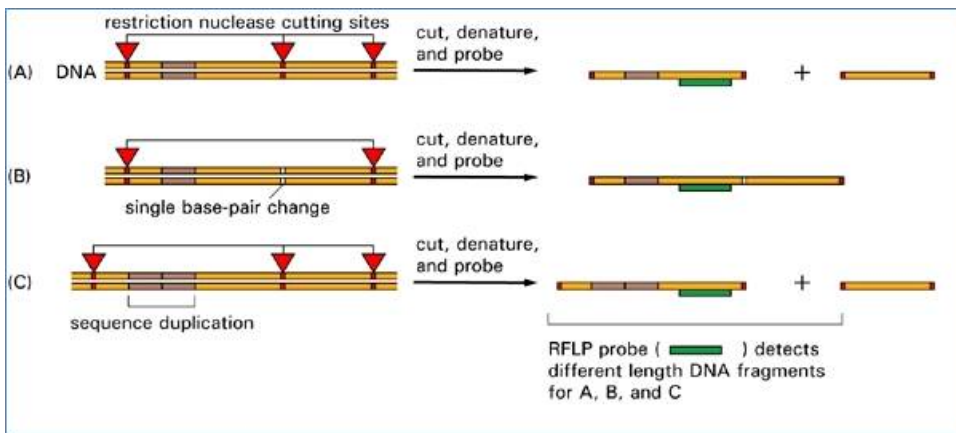
Restriction fragments, Southern blotting

The testing historically was done using restriction enzymes, that cut at very specific sequences of DNA. For instance, the restriction enzyme EcoRI from *E. coli*, cuts DNA every time the exact sequence GAATTC appears. That occurs on average every 4,096 base pairs. Of course that's an average--sometimes the space between sites is longer and sometimes shorter. If you cut a human genome, there will be many fragments and finding the right one is difficult. Running the fragments on a gel will separate fragments by size, but there would be too many to distinguish.

In the mid-1970s, Ed Southern developed a technique to transfer DNA from a gel to a blot. The Southern blot could be re-used and probed with labeled DNA. If the sequence in the labeled DNA is unique enough, then only one fragment would be labeled and visible.

RFLP

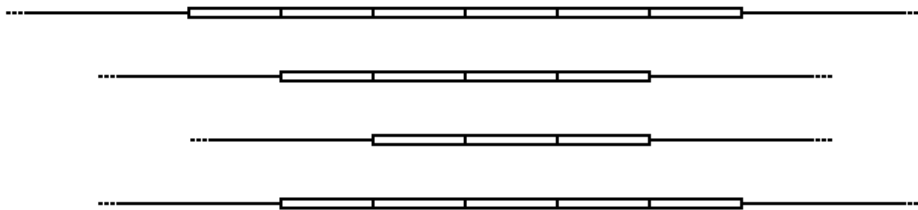
Most human genome fragments would be the same size, regardless of individual. Very few vary between individuals. Variance can come from repeated DNA or from a mutation (SNP) that eliminates a restriction site, or creates a new one.



Note that the duplication causes a larger labeled fragment and the elimination of a restriction site creates a much longer one.

VNTR

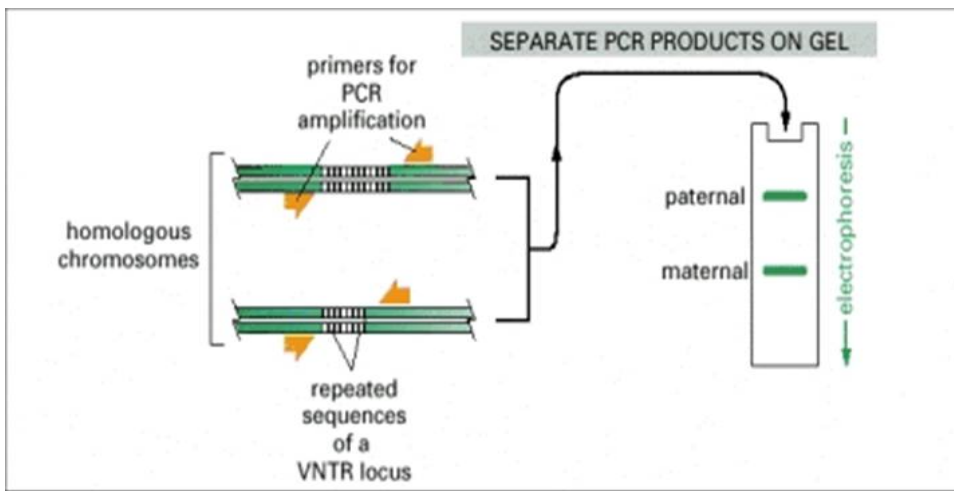
The easiest genomic regions to test for size differences have variable number tandem repeats (VNTR). These are short stretches of DNA that are back to back and repeated. This allows for many possible fragment lengths within a population.



-Image from en.wikipedia.org

PCR

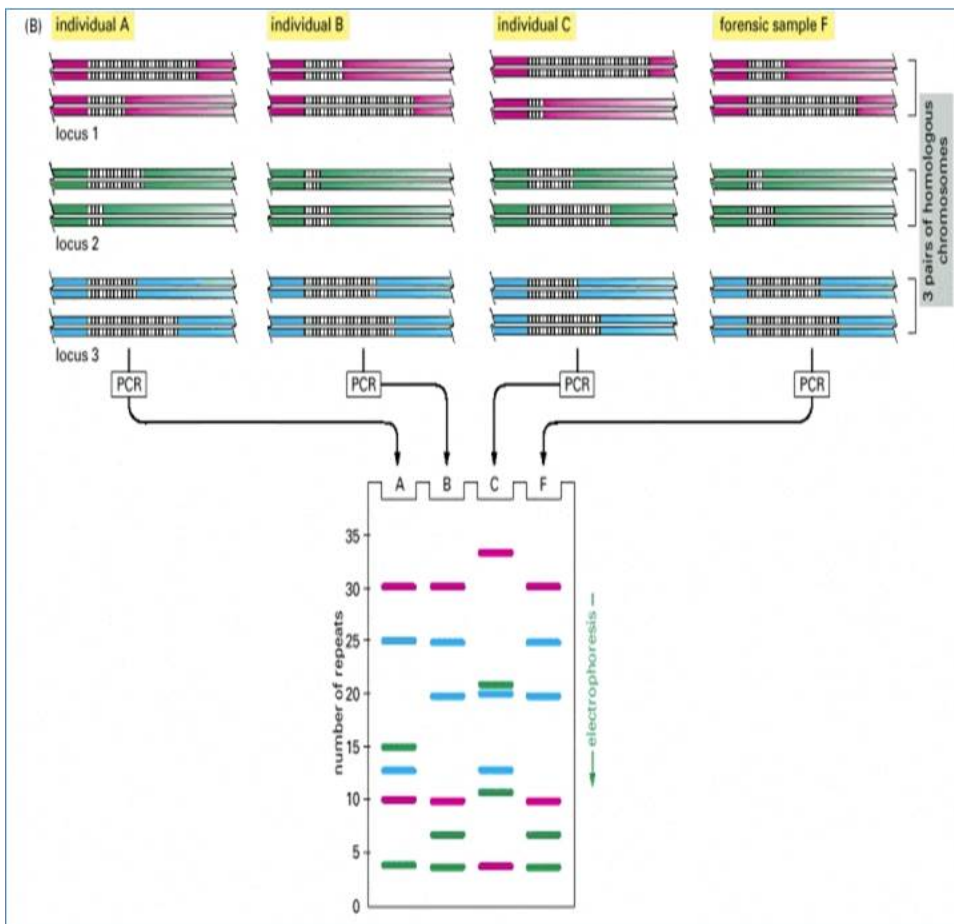
RFLP is a tedious process and labeling Southern blots takes time. The [Polymerase Chain Reaction](#) amplifies DNA from a very specific region by using flanking primer sequences. The previous link is to a tutorial that explains the process. When PCR is finished, the amplified DNA dominates the DNA concentration on any mixture. This allows the investigator to quickly focus on narrow regions of DNA.



-image from Alberts B, Johnson A, Lewis J, et al. *Molecular Biology of the Cell*. 4th edition. New York: Garland Science; 2002. Figure 8-41, How PCR is used in forensic science. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK26837/figure/A1592/>

VNTR Analysis

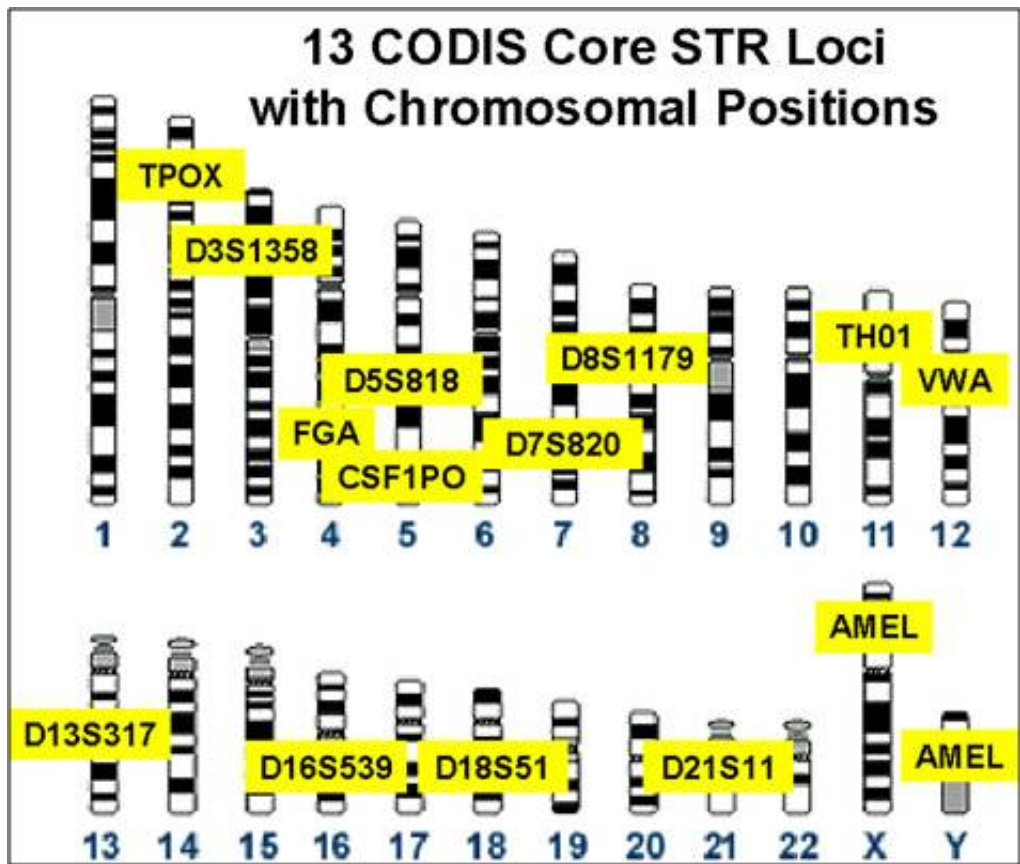
In the diagram below, The forensic sample is on the right. Individual A, B and C are suspects. The gel is on the bottom. Which one is a match? Note that this is at just three locations, not 13. You can see how 13 locations would give you a pretty good fingerprint. There is a lot of variability in genotype at each location.



CODIS

The FBI Combined DNA Index System is called [CODIS](#). There are 13 autosomal locations with known

frequencies, broken down by race, in the U.S. population. The known frequencies and number of repeats can be downloaded from <http://www.cstl.nist.gov/div831/strbase/fbicore.htm>. Here are the 13 locations.



Paternity exercise

The table below shows five CODIS loci. Can paternity be ruled out?

DNA Marker Mother Child Alleged father

D21S11	28, 30	28, 31	29, 31
D7S820	9, 10	10, 11	11, 12
TH01	14, 15	14, 16	15, 16
D13S317	7, 8	7, 9	8, 9
D19S433	14, 16.2	14, 15	15, 17

--from en.wikipedia.org/wiki/Dna_fingerprinting

Let's look at D21S11 in the first row. The child has a 28 allele and a 31 allele. The mother has no 31 allele, so the child must have gotten the 28 allele from the mother. That means 31 came from the father. The father has a 31 allele, so he could be the father based only on that location. For D7S820 (row 2), the 10 allele came from the mother, so 11 came from the father. The alleged father has an 11 allele. Now, TH01 in row three. Child must have gotten 14 from the mother, so 16 came from the father and the alleged father has a 16. For D13S317, the 7 allele was inherited from the mother; 9 came from the father. The father does have a 9. Finally, for D19S433, 14 was from the mother so 15 was inherited from the father. The alleged father has a 15 allele.

You would want to check the other eight loci to confirm paternity, but paternity is certainly possible based on just these five loci. Sometimes, more than 13 loci are used. Each locus is assigned a paternity index (PI). If the alleles are rare in the population, there is more certainty. If a lot of common alleles show up in the testing, additional loci are used to confirm paternity. Paternity is never really proved, but rather shown to be correct beyond reasonable doubt.

Discussion topic: Any questions on the paternity exercise?