May 2006

A Study of the Molecular Basis of Microvariants at the FGA and D21S11 Loci Used in Forensic DNA Testing

Molly McBeth
University of Connecticut

A Study of the Molecular Basis of Microvariants at the FGA and D21S11 Loci Used in Forensic DNA Testing

Molly McBeth
Molecular and Cell Biology
University of Connecticut
Honors Thesis

Thesis Advisor: Dr. Linda Strausbaugh
Honors Advisor: Dr. David Knecht

Abstract

Microvariant alleles, defined as alleles that contain an incomplete repeat unit, often complicate the process of DNA analysis. Understanding the molecular basis of microvariants would help to catalogue results and improve upon the analytical process involved in DNA testing. The first step is to determine the sequence/cause of a microvariant. This was done by sequencing samples that were determined to have a microvariant at the FGA or D21S11 loci. The results indicate that a .2 microvariant at the D21S11 locus is caused by a –TA– dinucleotide partial repeat before the last full TCTA repeat. The .2 microvariant at the FGA locus is caused by a –TT– dinucleotide partial repeat after the fifth full repeat and before the variable CTTT repeat motif. There are several possibilities for the reason the .2 microvariants are all the same at a locus, each of which carry implications on the forensic community. The first possibility is that the microvariants are identical by descent, which means that the microvariant is an old allele that has been passed down through the generations. The second possibility is that the microvariants are identical by state, which would mean that there is a mechanism selecting for these microvariants. Future research studying the flanking regions of these microvariants is proposed to determine which of these possibilities is the actual cause and to learn more about the molecular basis of microvariants.

Acknowledgments

Dr. Linda Strausbaugh

Craig O'Connor

Cara Statz

Joshua Suhl

Bo Pietraszkiewicz

Heather Nelson

Daniel Renstrom

Mary Lajoy

Melanie Ktorides

Table of Contents

List of Tables

List of Figures

Introduction

Humans share approximately 99.7% of their DNA[1]. The 0.3% of our genome that varies between individuals is what makes each person unique. Although we all have the same genes, such as the gene that determines eye color, we all have different variations of these genes, which gives us blue, green and brown eyes. These variations of a gene are called alleles. Each human being has two alleles for every gene, one from our mother and one from our father. Together, these two alleles make up a genotype for that particular gene, also called a locus. A genotype simply describes the two alleles present for a particular individual at a locus. By taking the genotypes at several different loci we can create a DNA profile which describes the alleles present at those loci for an individual. If the DNA profile contains enough loci then that profile would become a unique representation of a specific individual.

The idea behind a DNA profile can be thought of as the game Guess Who. In the game, you have to guess who your partner is looking at by narrowing down the possibilities. This is done by getting more and more specific about how the person looks; blond hair, glasses, blue eyes, until out of all the faces there is only one possibility. A DNA profile does the same thing but rather than narrowing down the possibilities by physical characteristics, a DNA profile uses genetic characteristics. Just like in the game, when enough of these genetic characteristics are determined a unique profile specific to one individual is determined, a genetic fingerprint. This DNA fingerprint can be used for identification purposes and is the basis of modern forensic DNA testing.

The process of DNA fingerprinting was first described by Dr. Alec Jeffreys in 1985[1]. A year later, in 1986, DNA testing was used in England to solve the brutal murder

of two women. It was not until 1995, however, during the O.J. Simpson trial that DNA testing gained public exposure and popularity in the United States[1]. Since then DNA testing technology has grown by leaps and bounds and is now common practice among the forensic community.

The technology used for DNA testing today is the analysis of short tandem repeats (STR). STRs are found within the noncoding region of DNA and contain a 2-6 basepair DNA sequence repeated one after the other up to sixty times. The number of times that the sequence repeats, however, is variable between individuals. Thus, alleles are designated by the number of times the DNA sequence is repeated. For example, in Figure 1 the person represented would have an allele 7 and an allele 8 at that STR locus. This individual is heterozygote and has two different alleles at that locus. If he or she had been homozygous both parents would have contributed the same allele.

There are many advantages to using STRs, the first being that the variability in the number of repeats at any locus makes them useful for identification purposes. Each STR locus has a wide range of possible alleles, giving STRs a high power of discrimination[1]. Also, the STRs have a relatively small size repeat unit compared to earlier technologies. This allows STRs to be used with degraded DNA, common among forensic samples, and still obtain a complete DNA profile. Finally, the largest advantage of STRs is that they are easily amplified using the polymerase chain reaction, or PCR. This allows minute amounts of DNA to be used in forensic testing. This feature of STRs is crucial when considering that forensic samples may be as small as a drop of blood or a single hair.

The process of DNA testing begins at the crime scene. Sources of DNA in forensic samples can include blood, semen, hair, saliva, cigarette butts, postage stamps, dandruff and even fingerprints[1]. With PCR a quantity as small as 0.2 nanograms of DNA can be used to create a complete DNA profile. With such a plethora of different sources of DNA it is important for investigators to determine which samples at a crime scene warrant further testing and may be beneficial to the case.

Crime scene investigators have many tools at their fingertips to help in identifying potential evidence at a crime scene. There are several screening tests which can be used including the Kastle-Myer test used to screen for blood, the Acid phophatase test used to screen for semen, the Phadebas test used to screen for saliva, the Jaffe test used to screen for urine, and UV light which can be used to detect many biological materials[5]. However, even these tests can result in false positives. Gathering too much evidence could overwhelm the analysts and result in a lot of redundant and irrelevant items being tested. Gathering too little evidence may lead to missing that vital piece that could lead to a conviction. Thus having an experienced investigator is crucial in collecting the appropriate evidence[2].

Along with evidence collection, documenting and preserving the evidence is equally important. At a crime scene any potential evidence must first be photographed in the original position before collection. The location and condition of each sample must be documented and collected with the utmost of care. Due to the high sensitivity of PCR any contamination can prove disastrous, potentially leading to inconclusive results. Evidentiary samples are usually collected in paper bags or envelopes to prevent bacteria growth, and kept cool to avoid degradation. All evidentiary samples need to be labeled

and chain of custody maintained to ensure that all samples can be admissible in a court of law[2].

Once the samples are collected at the scene, they are brought to a forensic lab where they will be subjected to DNA testing. The DNA typing process begins with the extraction of the DNA from the source material (Figure 2). The amount of DNA is then tested for quality and quantity before specific regions are amplified using PCR. The PCR products are then separated by size using electrophoresis and detected, determining the specific allele for each locus. The resulting genotypes together create a DNA profile for that sample which can then be compared to other samples from the case or to a database. If a match occurs between a DNA profile from an evidentiary sample and a profile from a suspect then a random match probability (RMP) for that DNA profile is calculated. RMP is the likelihood that the same profile can be found at random in the population. These are the numbers that are used when DNA evidence is presented in court to prove with virtual certainty that the donor of that DNA profile, the suspect, was at the scene of the crime.

Before DNA results ever make it to court, however, each of the steps prior must be done with complete accuracy. This begins with the extraction of the DNA from other cellular materials present within a forensic sample. DNA is packaged inside the cell with proteins that may inhibit the amplification process so it is necessary to remove these proteins in order to obtain only the DNA molecules[1]. There are four main methods used in forensics to extract the genetic material: organic extraction, Chelex extraction, FTA paper extraction and differential extraction[4]. Each method has its own advantages and disadvantages.

Organic extraction, also known as phenol chloroform extraction, is the best at obtaining high molecular weight DNA which can be used for the current method of STR analysis or with the older method of RFLP analysis. However, organic extraction requires transferring the sample several times increasing the risk of contamination[4]. Also, some of the reagents used in organic extraction are PCR inhibitors and therefore must be completely removed for the PCR process to work effectively.

Chelex extraction has less chance for contamination due to fewer steps, also making it faster than organic extraction. Chelex extraction also has the added advantage in that it removes any PCR inhibitors present in the forensic sample. However, Chelex extraction results in only single stranded DNA which means that it can only be used with STR analysis using PCR and there is no possibility of using RFLP[4].

FTA paper involves placing a spot of blood on chemically treated paper that lyses the cells and traps the DNA in the matrix of the paper. This method does not require a quantification step. Also, the FTA paper is stable for long periods of time and can be reused. However, static often causes these paper particles to "jump" between tubes in a sample tray which may cause contamination and complicate the analysis process[4].

Finally, the method of differential extraction for samples that contain both sperm and female epithelial as would be the case for evidence from a sexual assault victim, is very useful in forensics. Differential extraction is similar to organic extraction for isolating DNA but also separates the epithelial cells from the sperm cells. This method allows investigators to isolate the male fraction of a mixture allowing them to obtain a profile for the male perpetrator. However, differential extraction does not work with azoospermic semen and therefore is not beneficial in all cases[4].

The DNA analyst would choose the method of extraction that is most appropriate for each sample. With any of the methods chosen however, it is still necessary to take every precaution to avoid any contamination. Contamination at the extraction stage could be detrimental to a case, completely destroying a sample and obliterating the evidence.

After the DNA is extracted from a sample it needs to be quantified. The reason for this step is that most PCR reactions have an optimal range of DNA concentrations that provide the best results. If too much DNA is used the resulting electropherogram, which is the final product, may have split peaks or be completely off-scale. If too little DNA is used allele dropout may occur[1]. Both cases would result in an inaccurate DNA profile. There are several methods used to quantify the DNA including agarose yield gels with ethidium bromide, slot blot quantification, real-time PCR approaches and several others. Each method is used to estimate the amount of DNA in the sample. Using this information the sample is then diluted to a quantity of DNA that will work optimally with PCR, usually 1-10 nanograms.

PCR is the next step in DNA testing. It is a method used to generate large quantities of a specific sequence of DNA, like genetic Xeroxing[1]. The four major components of PCR are the template DNA, deoxyribonucleotides (dNTPs), the building blocks of the new strands, a polymerase, which is an enzyme that creates the new strands, and primers that bind to the specific region of DNA to be amplified.

The process of PCR contains three steps repeated a number of times (Figure 3). In the first step, the sample is heated to 95°C. This denatures or separates the two strands of DNA. The temperature is then lowered to between 50°C and 60°C. At this

temperature the primers, seen in red, anneal or bind to the complementary regions of the target DNA. The temperature is then raised to approximately 72°C for the extension process in which the DNA polymerase adds the dNTPs to make a new strand of the target DNA. This cycle is repeated twenty to forty times, expontentially making more copies of the target DNA, resulting in billions of copies of a specific region of DNA.

The primers are the most important component of the PCR reaction because they target the region of DNA that will be amplified. Primers are short DNA sequences, usually 15-20 basepairs that bind to the DNA regions flanking the STRs. These flanking regions are constant meaning that they are virtually the same in every individual allowing the same set of primers to work for everyone. Each pair of primers contains a forward primer complementary to the 5' region of the DNA sequence before the STR region, and a reverse primer complementary to the 3' region of the DNA sequence after the STR region. When amplification occurs, only the region between the two primers will be amplified resulting in billions of copies of the target DNA sequence.

Several factors must be considered when designing the primers used to amplify the STR regions. First, the annealing temperatures of the primers must be similar so that both primers will bind to the template DNA. The primers must also be specific to the target region. If the primers are compatible with other regions of DNA then nonspecific product will be amplified rather than the desired STR region. The primers also must not interact with each other and form what is referred to as "primer dimer." This phenomenon appears when the primers bind to each other rather than the template DNA, resulting in no amplification[1]. A good pair of primers will satisfy all these requirements.

An advantage of using STR technology is that many loci can be amplified at once by using several primer sets in a single PCR reaction. This is known as a multiplex. Along with the requirements for a single pair of primers, several other important factors need to be addressed in order for a multiplex to work properly. The first is that it is necessary for the primers to be adjusted so that different STR loci have different size amplification products. This is crucial in being able to identify the alleles that represent each STR loci. In some cases, overlapping size ranges cannot be avoided. In these instances fluorescent dyes are used to label primers that have overlapping size ranges, so that each locus can be distinguished. All of the primers must have similar annealing temperatures and work at the same efficiency. This is necessary to obtain good peak height balance in the electropherograms. Several companies produce multiplex STR kits used by forensic labs. Applied Biosystems uses a combination of two amplification kits known as Profiler Plus™ and COfiler™, to obtain a total of thirteen STR loci and the sex-marker Amelogenin (Figure 4). Amelogenin, a gene that encodes tooth enamel proteins, is found on both the X and Y chromosome[1]. However, on the X homologue the gene contains a six base pair deletion allowing it to be distinguished from the Amelogenin gene on the Y. Therefore, a male is represented by two separate peaks with a six base pair differenece, an X and a Y, while a female only has a single peak for the X.

Separation of the PCR products is the next step. PCR products are separated by size because the alleles for each STR are based on how many times the DNA sequence is repeated. A larger allele corresponds to a larger DNA fragment which would move more slowly through the gel or polymer solution. This allows the DNA fragments, or alleles, to be distinguished and detected.

An internal size standard is used in the separation step. This size standard has fragments of DNA of known length. It is run with the sample and used as a reference to determine the size of each PCR product in basepairs. An allelic ladder which contains all of the possible alleles for each locus is run alongside the samples. The sample peaks are compared to the ladder peaks in order to assign an allele call (Figure 5).

Separation of PCR products is currently performed using capillary electrophoresis. The sample is mixed with a denaturant solution that separates the strands of DNA. They are then heated and loaded into the instrument. A small voltage is applied which draws a tiny portion of each sample into a capillary, which is a thin tube filled with polymer. A separation voltage is then applied and the DNA molecules begin to migrate through the polymer solution. The smaller pieces migrate faster through the polymer resulting in size separation of the PCR products. Near the end of the capillary the DNA molecule passes a tiny window where a camera detects the fluorescence from the primer and collects time, color and intensity data[1]. This data is then converted to a basepair size for the PCR product based on the internal size standard. Finally, the basepair size of the PCR product is compared to the allelic ladder which allows an allele call to be assigned. The allele calls made at each locus together make up the DNA profile.

Currently, there are thirteen loci used when determining the DNA profile of a forensic sample. These loci, chosen by the FBI Laboratory after an evaluation of many STR loci, are the core set used in the national DNA database called the Combined DNA Index System (CODIS). CODIS contains over 1.5 million profiles and is linked to all fifty states[1]. This database serves many purposes, the first being that it contains a DNA profile from every convicted felon. The reason that this is so helpful is because most of

the crimes committed are by repeat offenders. Having the DNA profiles of convicted criminals makes it more likely that the perpetrator will be caught. DNA databases also serve a second and equally important purpose which is the ability to link seemingly unrelated crimes that may have occurred in different states as well as in no-suspect cases[1]. DNA databases can also be used to determine the statistical frequencies of DNA profiles. Allele frequencies, calculated by using DNA database information, are crucial in calculating a random match probability. When all thirteen loci are used the RMP can be as rare as one in a trillion[1].

The thirteen core loci used in CODIS are CSF1PO, FGA, TH01, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51 and D21S11[1]. These thirteen loci cover twelve of the twenty-two autosomal chromosomes (Figure 6). CODIS also contains the sex-typing marker, Amelogenin. The thirteen STR markers used in CODIS were all chosen because they have a wide range of possible alleles, allowing for more discriminating allele frequencies.

Several of the loci used in CODIS have rare alleles which contain a partial repeat. These alleles are referred to as microvariants[1]. The STRs used in forensics have a 4 basepair DNA sequence called the repeat motif. A microvariant occurs when one of the repeat units contains only one, two or three bases of the repeat motif. The microvariant allele is called by a whole number indicating the number of full repeats and a decimal representing how many bases occur in the partial repeat. For example, TH01 has an AATG repeat motif. A common microvariant at this locus is the 9.3 caused by a missing adenine in the seventh repeat resulting in an ATG partial repeat (Figure 7)[1]. Although

microvariants are most commonly seen at the more polymorphic STR loci, they also appear in the simple repeats.

Microvariants can complicate the analytical process and lead to inconclusive results. This is a concern to the forensic community because inconclusive results carry no information useful in court and cannot be used to identify any potential suspects. The cause of the complication is that microvariant alleles can differ from a normal allele by a single base pair making it very difficult to distinguish the alleles. This would not be such a problem if microvariants were a rare occurrence but this is not the case. Some microvariant alleles are prevalent among populations, such as the TH01 9.3 allele which occurs in 33% of the US Caucasian population[1].

Microvariant alleles are also common at the FGA and D21S11 loci which are the focus of this study. The FGA locus is located in the third intron of the human alpha fibrinogen gene on the big arm of chromosome 4[8]. FGA is classified as a complex tetranuclueotide repeat because it contains several repeat motifs, the main one being a "CTTT" repeat (Figure 8). The ".1", ".2" and the ".3" microvariants have all been observed at the FGA locus.

The D21S11 locus is located on chromosome 21[8]. Similar to FGA, D21S11 has more than one tetranucleotide repeat motif (Figure 9). Each of the microvariants, the ".1", ".2" and ".3", have been observed at this locus. However, the ".2" microvariants were the focus of this study due to their prevalence in the sample population.

Other than the cataloguing of the various alleles, very little has been studied about these microvariant alleles. The purpose of this study was to begin an investigation into the molecular basis of these microvariants. The first step of this process was to determine

the sequence of these microvariants which may hint at their cause. The information obtained from this would help to catalogue results and lead to ways to improve the entire analytical process.

Materials and Methods

DNA was collected using buccal swabs and extracted using the standard phenol-chloroform method. The samples were then quantified on a 1% agarose gel with comparisons being made to known standards. Samples were amplified using using *AmpFlSTR*® Profiler Plus®, COfiler® and Identifiler® STR amplification kits following the manufacturer's protocol using 1-2ng of DNA. The samples were then separated using ABI 377 Prism® polyacrylamide gel electrophoresis or ABI 3130® capillary electrophoresis. Allele designations were made by Applied Biosystem's Genotyper ® 2.5 and GeneMapper ID® 3.1 software.

Samples that were assigned microvariant alleles at the FGA and D21S11 loci were then further amplified using primers specific for each STR region.

FGA STR primer:
FWD 5' – GCC CCA TAG GTT TTG AAC TCA – 3'
RVS 5' – TGA TTT GTC TGT AAT TGC CAG – 3'

D21S11 STR primer:
FWD 5' – ATA TGT GAC TCA ATT CCC CAA – 3'
RVS 5' – TGT ATT AGT CAA TGT TCT CCA – 3'

PCR amplification used 2ng of the DNA samples in a final volume of 25 µL containing 12.5 picomoles of primer, 0.2mM of each nucleotide, 2.5mM $MgCl_2$, 1.25U Taq and 1x PCR buffer (500mM KCl, 100mM Tris-HCl, pH 8.3, 15mM $MgCl_2$). The parameters for the PCR amplification, for both FGA and D21S11, were 95ºC for 5 minutes, followed by a cycle of 94ºC for 1 minute, 60ºC for 1 minute and 72ºC for one minute, repeated thirty-five times, and another 5 minutes at 72ºC before a 4ºC hold. The samples were then run on a 1% agarose gel to determine the presence of PCR product.

The PCR samples were then cleaned using the QIAquick® PCR Purification Kit (250) following the manufacturer's directions. The samples were then sequenced using Big Dye® Terminator v3.1 Cycle Sequencing Kit following the manufacturers' protocols. Sequencing was then performed by ABI 3130 capillary electrophoresis with base calls made with the Sequence Analysis® software.

Results

To date 905 samples have been collected and profiled from individuals ranging in age from 8 to 96 years old. Of these, 559 were female and 346 were male. The ethnic distribution of the samples was 675 Caucasians, 50 African Americans, 117 Asians, 30 Hispanics and 35 of other ethnic backgrounds. 180 of the samples were collected from individuals born outside of the United States. Of the 725 born in the United State 514 were from New England, 75 from New York, and 25 from New Jersey. From these samples, 19 containing a microvariant at the FGA locus were sequenced as well as 28 containing a microvariant at the D21S11 locus.

Sequencing of samples with a microvariant at the FGA locus showed interesting results. The ".2" partial repeat caused by a dinucleotide "TT" is seen after the 5th full repeat of "TTTC" and is seen in each allele with a ".2" microvariant (Table 1). The D21S11 locus had a similar pattern, with a dinucleotide "TA" partial repeat seen before the last "TCTA" repeat unit in each of the ".2" microvariants (Table 2).

Discussion

19 samples containing a microvariant at the FGA locus were sequenced and contained a "TT" dinucleotide partial repeat. This partial repeat is always seen after the 5th full repeat. Similarly, the sequenced D21S11 microvariants showed a "TA" dinucleotide partial repeat always seen before the very last full repeat. These results bring up several important questions about the molecular basis of the microvariants. Are these microvariants many different alleles or is it one single allele that has been inherited from a common ancestor? These two alternatives are known as identical by state and identical by descent[3].

By definition, identical by state means that the alleles, in this case the microvariants, look the same but originate from multiple origins. An example of identical by state would be two siblings each with blood group AB born to parents also with blood group AB (Figure 10). Although each sibling has a copy of the A allele and a copy of the B allele, they may not have the same copy of each allele depending on which parent it was inherited from. The first sibling inherited the A allele from the mother and the B allele from the father, whereas the second sibling inherited the A allele from the father and the B allele from the mother. Therefore, although the alleles look the same, they are not, in fact, the same allele.

Identical by descent is when the alleles originate from a single allele that has been inherited from a common ancestor. For instance, a mother with O blood and a father with AB blood have two children each with A blood (Figure 11). Both children would have had to inherit the A allele from their father, and would thus have the same A allele. Similarly, if the microvariant alleles are identical by descent this would mean that the

alleles originate from the same common ancestor and have been fixed in the population either by drift or selection.

Further research into the flanking regions of the microvariants must be done in order to determine which of these alternatives is the molecular basis of the microvariants. Primers will be designed that amplify approximately an 800 bp region flanking the repeat region in both the 3' and the 5' direction. These regions will then be cloned, sequenced and compared to each other using NCBI BLAST.

Due to a mutation rate of approximately 1 in 400 meioses, we would expect to see polymorphisms in these flanking regions of the microvariant alleles[8]. If the alleles are identical by state then we would expect to see many random polymorphisms that produce different haplotypes. A haplotype refers to a set of polymorphisms that are associated with one another and are usually inherited together on a single chromatid[11]. In the case that the alleles are identical by state, this would mean that the microvariant alleles originated from multiple origins and therefore would produce their own unique haplotypes different from each other. However, if the microvariant alleles are identical by descent, then the set of polymorphisms should produce similar haplotypes.

If this should be the case, that the alleles are identical by descent and therefore is the same allele reoccurring in the population, it would have heavy implications on the forensic community. The STR loci chosen for forensic DNA testing were selected on the basis that these loci are not parts of functional genes, but rather on the non-coding sequences, and furthermore are not correlated to any physical or behavioral characteristic that may be selected for[4]. This fact would allow the Hardy-Weinberg Equilibrium formula to be applied to these loci enabling the forensic scientist to determine the

frequency of alleles and genotypes within a population. It is this information that is used to calculate the random match probabilities based on the product rule that are used in court[4]. Should the microvariant alleles be identical by descent, this would indicate that there is some kind of selection or drift occurring at these loci which violates the rules of the Hardy-Weinberg equilibrium, voiding the application of the product rule and theoretically eliminating any use of random match probabilities.

Literature Cited

1. Butler, John M.. Forensic DNA Typing, Academic Press, 2001.
2. Lee, Henry C., Ladd, Carll (2001). "Preservation and Collection of Biological Evidence." Croatian Medical Journal 42(3): 225-228.
3. Strachan, Tom, Read, Andrew P.. Human Molecular Genetics 3, Garland Publishing, 2004.
4. Ladd, Carll, Bourke, Michael. "Forensic Applications of DNA Science." Molecular and Cell Biology 290. University of Connecticut, September 5th, 2005.
5. Ladd, Carll, Bourke, Michael. "Identification of Evidence." Molecular and Cell Biology 290. University of Connecticut, September 12th, 2005.
6. Ladd, Carll, Bourke, Michael. "Short Tandem Repeats and the DNA Typing Process." Molecular and Cell Biology 290. University of Connecticut, September 12th, 2005.
7. Ladd, Carll, Bourke, Michael. "STR Analysis." Molecular and Cell Biology 290. University of Connecticut, September 19th, 2005.
8. Butler, John M., Reeder, Dennis J.. Short Tandem Repeat DNA Internet Database, http://www.cstl.nist.gov/div831/strbase/index.htm.
9. Barts and The London, School of Medicine and Dentistry (2000), Nonparametric Methods of Linkage Analysis 2, http://www.smd.qmul.ac.uk/statgen/dcurtis/lectures/nonpar2.html.
10. Genome Canada (2000), Glossary: Identical by Descent, http://www.genomecanada.ca/GCglossaire/glossaire/index.asp?alpha=i&l=e.
11. Wikipedia contributors (2006). Haplotype. Wikipedia, The Free Encyclopedia. Retrieved February 24, 2006 from http://en.wikipedia.org/w/index.php?title=Haplotype&oldid=41442472.

Table 1: Breakdown of the repeat sequence of the FGA locus. The "TT" partial repeat is consistently seen after the 5th full repeat in each of the alleles designated with a ".2" microvariant.

Sample	Allele	[TTTC]	[TTTT]	[TTCT]	[TT]	[CTTT]	[CTCC]	[TTCC]
7	21	3	1	1	0	13	1	2
40	24	3	1	1	0	16	1	2
71	24	3	1	1	0	16	1	2
73	21	3	1	1	0	13	1	2
96	21	3	1	1	0	13	1	2
96	29	3	1	1	0	21	1	2
97	27	3	1	1	0	19	1	2
112	23	3	1	1	0	15	1	2
115	29	3	1	1	0	21	1	2
276	22	3	1	1	0	14	1	2
278	23	3	1	1	0	15	1	2
324	22	3	1	1	0	14	1	2
439	19	3	1	1	0	11	1	2
452	24	3	1	1	0	16	1	2
456	19	3	1	1	0	11	1	2
505	19	3	1	1	0	11	1	2
510	23	3	1	1	0	15	1	2
512	21	3	1	1	0	13	1	2
514	22	3	1	1	0	14	1	2
557	24	3	1	1	0	16	1	2
618	25	3	1	1	0	17	1	2
7	24.2	3	1	1	1	16	1	2
40	21.2	3	1	1	1	13	1	2
71	21.2	3	1	1	1	13	1	2
73	21.2	3	1	1	1	13	1	2
97	23.2	3	1	1	1	15	1	2
112	21.2	3	1	1	1	13	1	2
115	18.2	3	1	1	1	10	1	2
276	21.2	3	1	1	1	13	1	2
278	21.2	3	1	1	1	13	1	2
324	24.2	3	1	1	1	16	1	2
439	21.2	3	1	1	1	13	1	2
452	18.2	3	1	1	1	10	1	2
456	23.2	3	1	1	1	15	1	2
505	22.2	3	1	1	1	14	1	2
510	22.2	3	1	1	1	14	1	2
512	18.2	3	1	1	1	10	1	2
514	22.2	3	1	1	1	14	1	2
557	24.2	3	1	1	1	16	1	2
618	20.2	3	1	1	1	12	1	2

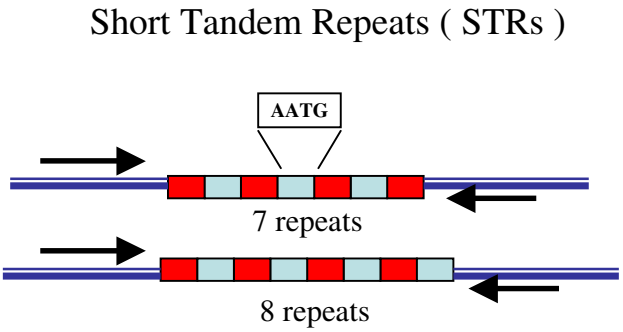
Table 2: Breakdown of the repeat sequence of the D21S11 locus. The "TA" partial repeat is consistently seen before the last full repeat in each of the alleles designated as a ".2" microvariant.

Sample	Allele	[TCTA]	[TCTG]	[TCTA]	[ta]	[TCTA]	[tca]	[TCTA]	[tccata]	[TCTA]	[TA]	[TCTA]	[tc]
4	30	6	5	3	1	3	1	2	1	10	0	1	1
6	30	6	5	3	1	3	1	2	1	10	0	1	1
9	31	5	6	3	1	3	1	2	1	11	0	1	1
16	30	5	6	3	1	3	1	2	1	10	0	1	1
17	29	4	6	3	1	3	1	2	1	10	0	1	1
72	29	4	6	3	1	3	1	2	1	10	0	1	1
82	29	4	6	3	1	3	1	2	1	10	0	1	1
83	29	4	6	3	1	3	1	2	1	10	0	1	1
167	29	4	6	3	1	3	1	2	1	10	0	1	1
192	30	6	5	3	1	3	1	2	1	10	0	1	1
206	29	4	6	3	1	3	1	2	1	10	0	1	1
243	29	4	6	3	1	3	1	2	1	10	0	1	1
262	29	4	6	3	1	3	1	2	1	10	0	1	1
263	29	4	7	3	1	3	1	2	1	9	0	1	1
325	29	4	6	3	1	3	1	2	1	10	0	1	1
429	29	4	6	3	1	3	1	2	1	10	0	1	1
439	29	5	6	3	1	3	1	2	1	9	0	1	1
505	30	4	6	3	1	3	1	2	1	11	0	1	1
581	29	6	5	3	1	3	1	2	1	9	0	1	1
629	29	4	6	3	1	3	1	2	1	10	0	1	1
4	31.2	5	6	3	1	3	1	2	1	11	1	1	1
6	31.2	5	6	3	1	3	1	2	1	11	1	1	1
9	30.2	4	6	3	1	3	1	2	1	11	1	1	1
16	32.2	5	6	3	1	3	1	2	1	12	1	1	1
17	32.2	5	6	3	1	3	1	2	1	12	1	1	1
72	32.2	5	6	3	1	3	1	2	1	12	1	1	1
82	32.2	5	6	3	1	3	1	2	1	12	1	1	1
83	33.2	5	6	3	1	3	1	2	1	13	1	1	1
92	32.2	7	4	3	1	3	1	2	1	12	1	1	1

Sample	Allele	[TCTA]	[TCTG]	[TCTA]	[ta]	[TCTA]	[tca]	[TCTA]	[tccata]	[TCTA]	[TA]	[TCTA]	[tc]
113	31.2	5	6	3	1	3	1	2	1	11	1	1	1
147	32.2	5	6	3	1	3	1	2	1	12	1	1	1
159	30.2	5	6	3	1	3	1	2	1	10	1	1	1
161	31.2	5	6	3	1	3	1	2	1	11	1	1	1
163	32.2	5	6	3	1	3	1	2	1	12	1	1	1
167	33.2	5	6	3	1	3	1	2	1	13	1	1	1
172	30.2	5	6	3	1	3	1	2	1	10	1	1	1
176	30.2	5	6	3	1	3	1	2	1	10	1	1	1
192	30.2	5	6	3	1	3	1	2	1	10	1	1	1
197	32.2	5	6	3	1	3	1	2	1	12	1	1	1
243	32.2	5	6	3	1	3	1	2	1	12	1	1	1
262	33.2	5	6	3	1	3	1	2	1	13	1	1	1
263	32.3	6	5	3	1	3	1	2	1	12	1	1	1
325	32.2	5	6	3	1	3	1	2	1	12	1	1	1
429	33.2	5	6	3	1	3	1	2	1	13	1	1	1
439	32.2	6	5	3	1	3	1	2	1	12	1	1	1
505	30.2	5	6	3	1	3	1	2	1	10	1	1	1
581	33.3	5	6	3	1	3	1	2	1	13	1	1	1
629	34.2	5	6	3	1	3	1	2	1	14	1	1	1

Figures

Figure 1: An example of a short tandem repeat[6]



the repeat region is variable between samples while the flanking regions where PCR primers bind are constant

Homozygote = both alleles are the same length
Heterozygote = alleles differ and can be resolved from one another

Figure 2: Summary of the process of DNA typing with STRs[1]

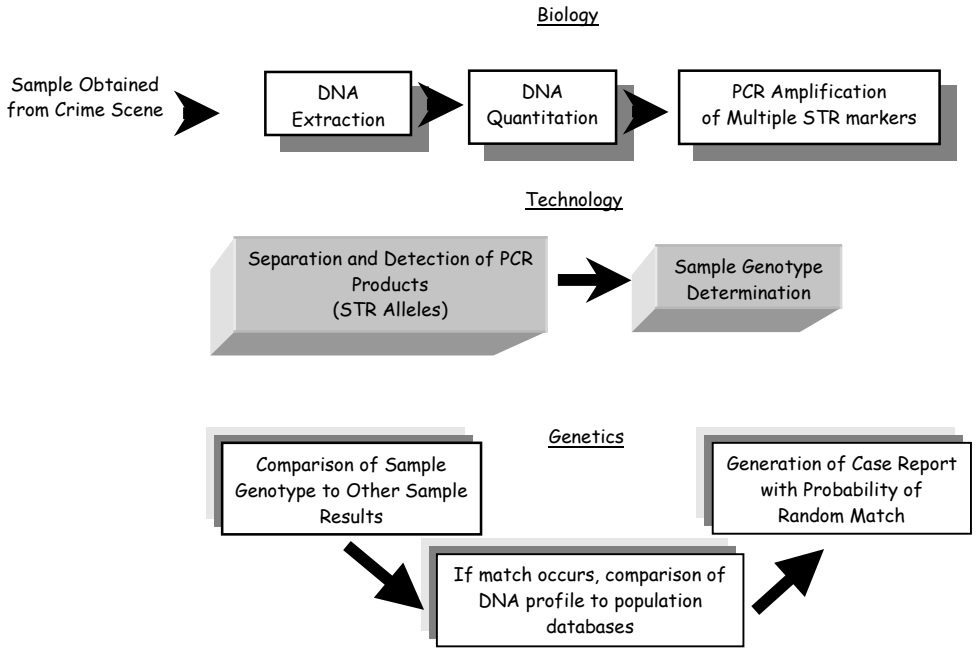


Figure 3: Diagram of the PCR amplification process[4]

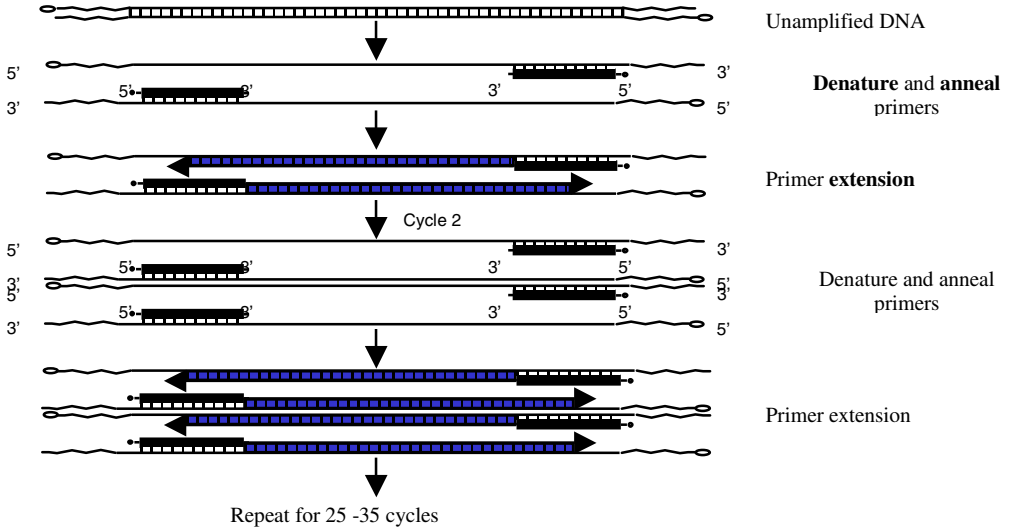


Figure 4: Size ranges of the loci used in the Applied Biosystems amplification kits Profiler Plus™ and COfiler™ [8]

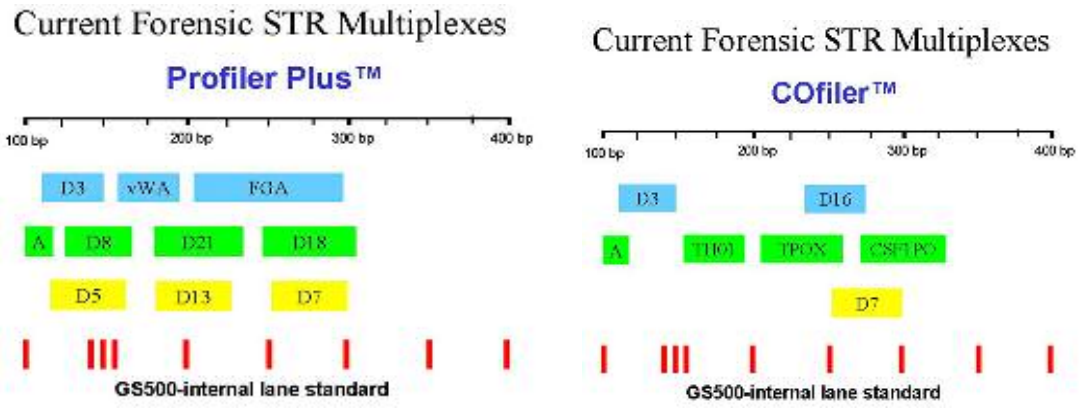


Figure 5: STR genotyping is performed by comparison of sample data to allelic ladders[6]

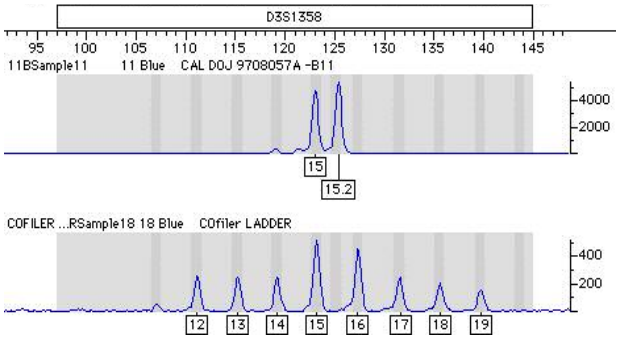


Figure 6: 13 CODIS core STR loci with chromosomal positions[7]

Figure 7: TH01 9.3 microvariant caused by a missing adenine in the seventh repeat

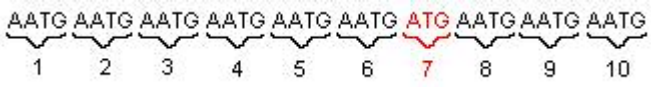


Figure 8: GenBank sequence of FGA including the repeat region and primer sequences[8]

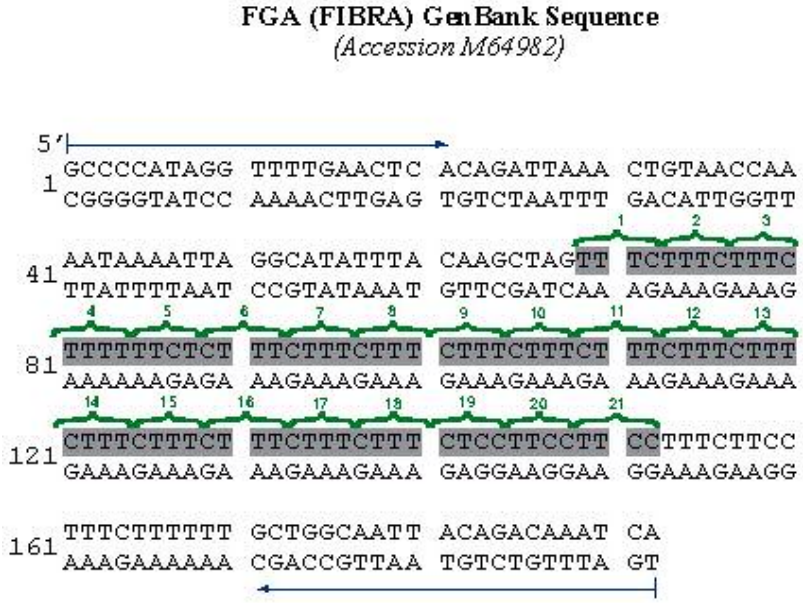


Figure 9: GenBank sequence of the D21S11 locus including the repeat region and primer sequences[8]

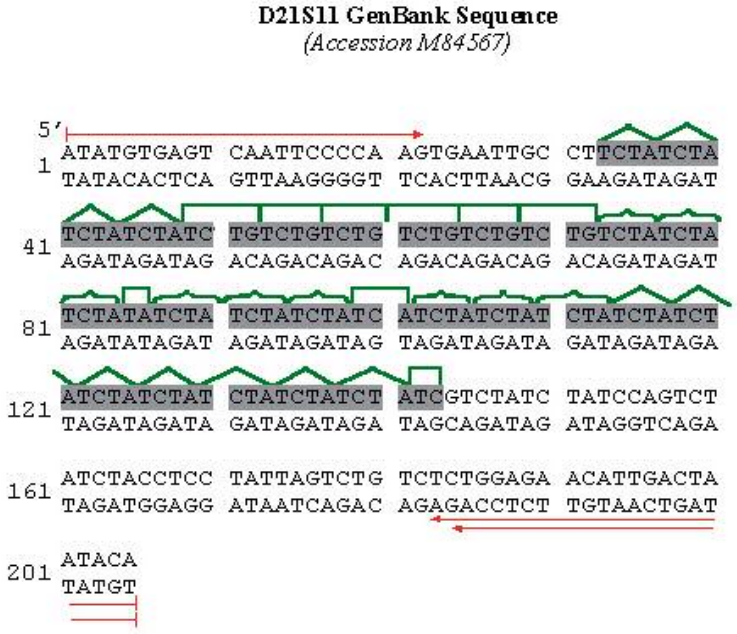


Figure 10: An example of alleles that are identical by state[9]

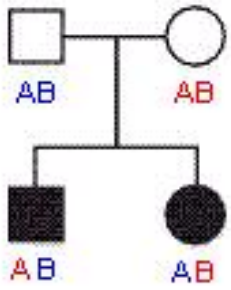


Figure 11: An example of alleles that are identical by descent[9,10]