DRAFT

An Introduction to Forensic Science: The Science of Criminalistics

James T. Spencer, Ph.D.
Professor of Chemistry and Forensic Science
Syracuse University

CHAPTER 5
Biochemical Forensic Analysis I: DNA

Confidential Correspondence

II. Biological Evidence

Chapter 5.1. DNA: The Genetic Record

"With the exception of nuclear DNA analysis, however, no forensic method has been rigorously shown to have the capacity to consistently, and with a high degree of certainty, demonstrate a connection between evidence and a specific individual or source." 2009, NRC Report *Strengthening Forensic Science in the United States: A Path Forward*.

Learning Goals and Objectives

The use of DNA evidence has become the "gold standard" of forensic investigations. In order to understand the wealth of information that can be gained from forensic DNA studies, you will need to demonstrate an understanding of:

- The chemical structure of DNA and how it holds genetic information,
- The transcription and translation processes of DNA,
- What parts of DNA are involved in forensic examinations,
- What are variable number tandem repeats (VNTR) and short tandem repeats (STR),
- What a mutation is,
- How the Restriction Fragment Length Polymorphism (RFLP) method works,
- How the polymerase chain reaction (PCR)/STR method of DNA typing works,
- How frequency of occurrences of STRs in populations is determined and used,
- What is CODIS,
- How mitochondrial DNA can be used in forensic investigations,
- How DNA typing is being used in plants and other living organisms.

Figure 5.1.1. DNA: the Blueprint of life (www.sciencephoto.com).

Introduction. In the search for compelling scientific evidence, the criminal justice system has frequently looked to advances in basic scientific research to lead the way to new forensic techniques. Nowhere has this progress been more dramatic than in the application of modern biochemical techniques to link the evidence gathered from crime scenes with potential suspects. The explosive growth in creative work and new discoveries in biochemistry has lead to the addition of many important new forensic tools. These methods, such as DNA and blood analysis, have become so powerful in providing the critical linking information that they have quickly become the "gold standards" of forensic evidence in courtroom proceedings. Juries have especially come to depend upon DNA evidence to be an integral part of all cases, regardless of what other evidence is available.

The rapid advance of DNA evidence from its first courtroom appearance in 1987 to its commonplace use in forensic investigations today is unprecedented in criminalistics. From the first discovery that biological DNA

Case History:
Colin Pitchfork and the DNA Dragnet

On November 21st of 1983, 15-year old Lynda Mann never returned home from a friend's home in Narborough, England. When her body was found the morning after her disappearance, it was learned that she had been sexually attacked before being murdered. Analysis of biological samples collected from the victim showed that the attacker must have been a young male who belonged to a relatively rare blood type, referred to as A/PGM+. Despite their best efforts, the police quickly came to a dead end in their investigation.

Almost three years later, in 1986, the strikingly similar attack and murder of 15-year old Dawn Ashworth occurred in the same area of Narborough. This time the police had a suspect that they thought could be linked to both crimes, 14-year old Richard Buckland. Buckland. This unusually large, not very intelligent youth, appeared to have more information about the murder than had been released to the public. After two days of incoherent admissions and ramblings, he signed a confession for the murder of Dawn Ashworth, but he maintained his innocence of the Lydia Mann attack. Since police believed that the same person committed both crimes, they turned to Dr. Alec Jeffreys of the University of Leicester for help in trying to connect these two cases with Buckland.

The analysis, based upon Jeffreys' longtime DNA research, confirmed that one man had indeed been responsible for both crimes, but that it wasn't Buckland! This was the first time that an innocent person was exonerated by DNA evidence. Dr Jeffreys was quoted as saying "I have no doubt whatsoever that he would have been found guilty had it not been for DNA evidence". It is now believed that Buckland had witnessed the murder of Ashworth from a distance.

Faced with this dilemma and no other leads, the police launched a massive campaign to screen more than 5,000 local males for blood type and DNA profile. This mass screening, however, came up without a match until a local bakery worker was overheard in a pub saying that he had been paid by a co-worker, Colin Pitchfork, to provide a blood sample for him. The police quickly brought in Pitchfork and obtained an authentic blood sample: it matched. Pitchfork later confessed and was charged with the rape and murder of both Lynda Mann and Dawn Ashworth. He was convicted after a trial that lasted just one day.

This case clearly illustrates the enormous strength of DNA evidence: it both exonerated an innocent suspect and located and help to convict the real attacker.

samples can be uniquely traced to one single human with very high certainty in 1985 to it's first use in a forensic investigation just one year later is truly a monumental achievement.

Mankind has long sought for a way to uniquely identify one person from another - our appearance, signature, and fingerprints have all been used with varying degrees of success. But it is with our DNA that we find the ultimate personal "signature". In fact, each person's DNA is unique among the DNA of all humans that have ever lived - no one (except possibly an identical twin) has ever had the exact same DNA composition as you.

In this chapter, several of the most important forensic bioanalytical techniques are

Figure 5.1.2. New York City, Ground Zero. DNA analysis was critical in the identification of remains from the World Trade Center attack in 2001. (from www.usatoday.com/tech/science/genetics/2006-08-22-dna-partial-matches_x.htm?POE=TECISVA).

described, especially as they apply to courtroom.

So what is biochemistry and molecular biology? Biochemistry and molecular biology are commonly thought of as closely related fields of science that involve the study of chemical compounds and reactions that occur in living systems. Biochemistry is the point at which biology is reduced to its most fundamental level – the level of atoms and molecules. At this level, we are usually most concerned with how molecules are formed and then function to control all the processes that collectively we call life. Simply stated, biochemistry is a study of the "chemistry of life".

In this chapter, we will explore how modern biochemical tools are being applied to forensic problems. How did a seemingly abstract discovery that there are repeated sections in our DNA makeup that does not code for any known purpose lead to the development of one of the most important tools in the forensic toolbox? How can the presence of certain proteins contained in our blood be used to determine ancestry and lineage? How can we obtain uniquely identifying information from minute biological samples collected at the crime scene? These are just a few of the basic questions that will be addressed in this chapter in which we will explore the basic forensic chemistry of DNA, blood and biological fluids.

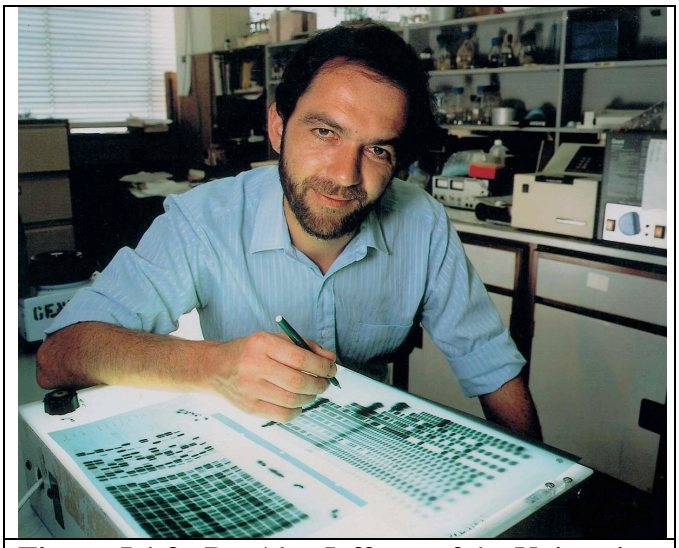


Figure 5.1.3. Dr. Alec Jeffreys of the University of Leicester in England led the 1985 DNA investigation of the Colin Pitchfork Case. (www.nlm.nih.gov/visibleproofs/media/detailed/iii_d_210.jpg)

5.2. DNA: How DNA Works

Introduction

Unraveling the secrets of DNA as the biochemical blueprint in governing the chemistry of life certainly ranks as one of the greatest accomplishments of modern civilization. This one amazing molecule is central to all living things and is arguably the single most important molecule on Earth. It governs our chemical makeup, our biological functions, and even whether we will have brown hair or green eyes. It determines whether an organism will develop into a dog, a tree, or a human being. It can even determine whether an organism will be healthy or carry a biochemical disorder. The more we learn of this one molecule, the more we come to see how fundamentally dominant it really is in all of life's biological processes. Our intimate knowledge of the inner workings of the DNA molecule has changed our lives in both subtle and dramatic ways, from the development of new diagnostic and therapeutic drugs to the designed modification of life forms, such as bacteria, through recombinant techniques.

Using a very simple analogy, DNA can be thought of as the massive biochemical cookbook that contains the instructions for making all of the compounds that we require for life. Inside each one of our more than sixty trillion cells resides one complete and identical copy of our DNA cookbook. Each recipe in our book, called a gene, contains specific directions for making just one biochemical item, such as a protein or enzyme. If we look closely at a typical written cookbook, we see that each recipe contains arrays and sequences of letters that are recognizable by us as words that we can, in turn, understand and translate into an action such as measure, mix, or bake. So, by the careful arrangement and combination of the 26 basic letters of the alphabet on the page, we can completely and fully describe how to prepare just the item we wish to make. On a conceptual level, DNA works in much the same way as a cookbook, except that it uses just four "letters" instead of the 26 alphabetic letters and all of the DNA "words" are just three letters long. But the instructions are just as clear and complete. Our DNA cookbook is unique, however, in that the DNA "recipes" are separated by long stretches of seemingly random letters that are unique from person to person and form the basis of how DNA forensic investigations work.

In this chapter, we begin by describing the basic chemical features of DNA and how it works to direct the chemistry of life. This is followed by the specific uses of DNA in forensic investigations. The field of forensic DNA is heavily laden with highly technical jargon that serves to convey information between specialists both quickly and efficiently. This language is, however, often quite daunting and confusing to those who encounter it for the first time. In the following discussions, however, this highly technical vocabulary of forensic DNA will be kept to the necessary minimum needed to understand the important terminology and concepts of the field.

DNA Background

In recent years, forensic investigations have relied increasingly upon DNA evidence both to convict and exonerate suspects. DNA (deoxyribonucleic acid), the genetic cookbook of life, often provides vital information about crimes and suspects. While the fundamental structure of DNA had been known since 1953, it had no significant place in the courtroom until the discovery by Alec Jeffreys in 1985 that biological samples taken from people could be used for their individual identification. In fact, prior to this work, DNA was thought to be very limited in forensic applications because organisms generally don't tolerate well any deviations from the typical genetic code. In other words, variations in the genetic DNA code usually leads to non-viable offspring. For example, if a cookbook recipe called for one cup of sugar and instead we used one cup of salt, the product of our work would certainly be uneatable. Similarly with DNA, changing one word could

easily render a critical protein biochemically useless. Luckily for forensic science, the regions between DNA "recipes" provides a unique marker that can be used to establish "beyond a reasonable doubt" a connection between a specific person and a biological sample collected at a crime scene. In order to understand how DNA is employed in forensic settings, however, an understanding of some of the basics about DNA itself is first necessary.

DNA, along with carbohydrates and proteins, is one member of a class of important biomolecules referred to as **biopolymers**. These biopolymers are compounds that are formed by linking together small molecular building blocks in a repeating fashion to form much larger molecules. This is conceptually similar to forming a long railroad train by linking together individual cars to produce a far larger assembly. The properties of the resulting biopolymer depends directly upon the identity and chemical properties of the basic building block units that compose the biopolymer. The properties also depend upon how the building blocks are connected together, such as in a straight chain like a train or branched like a tree.

Historically, the discovery of DNA itself has been attributed to Johann Miescher, a nineteenth-century Swiss chemist. Prior to his work, it was believed that cells were made up largely of protein - large biopolymer molecules made up of long chains of small linked amino acid building blocks. Miescher, however, found that certain extracts from the pus cells he was studying could not "belong among any of the protein substances known hitherto". He showed that these fractions were not made up of protein molecules since they were not digested by protease enzymes (enzymes that specifically break down *only* protein molecules). These particular extracts were also shown to arise solely from the nuclei of the cells and were, therefore, named nuclein. Later work on nuclein by another chemist, Albrecht Kossel, showed that nuclein was built of long chains of only four very similar building block molecules (nucleic acid units) and was later renamed a nucleic acid (DNA) by Richard Altmann. The structure we recognize today as the famous double helical arrangement of a long, twisting biopolymer was finally explained in 1953 by James Watson and Francis Crick, work for which they received the 1962 Nobel Prize in Medicine (Figure 5.2.1).

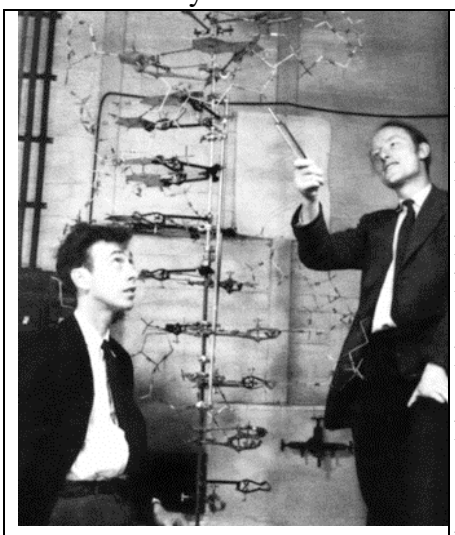


Figure 5.2.1. Watson and Crick with a model of the DNA double helix. (www.sciencephoto.com)

TIMELINE OF DNA

1868	Miescher "discovers" DNA.
1953	Watson and Crick report double helix structure.
1977	First human gene cloned.
1985	Jeffreys reports VNTR DNA sequences.
1985	First report of PCR method.
1986	Jeffreys uses DNA to solve first murder case (Pitchfork case).
1987	First conviction on DNA evidence (Andrews case).
1991	STRs first reported.
1998	FBI starts CODIS database.
2005	2.5 million DNA "finger-prints" in FBI database.

BRIEF ON DNA

(1) DNA is composed of nucleotide building blocks that are connected together into chains.
(2) Nucleotides contain a phosphate, sugar and nitrogen base units
(3) Phosphate and sugar units form the DNA backbone chain and nitrogen bases "hang" from the backbone.
(4) Pairing of nucleotide bases holds two strands of DNA together by hydrogen bonds.
(5) Bases pair such that only adenine (A) pairs with thymine (T) and guanine (G) pairs only with cytosine (C).

DNA, at its most fundamental level, is made up of repeating units called **nucleic acids** (Figure 5.1) connected in a polymeric fashion. The term polymer (from "poly" meaning many and "mer" meaning unit) indicates a large molecule formed by linking together many smaller units (monomers) together into large arrays. In DNA, one nucleic acid unit is linked to another to form a long, linear chain, much in the way that railroad cars or alphabet letters are linked together to form trains or words, respectively These nucleic acid building blocks are themselves composed of three simpler components (Figure 5.2.1): (1) a phosphoric acid residue (PO_4^{2-}), (2) a five-carbon sugar unit (specifically, DNA uses a deoxyribose sugar such as shown at left in Figure 5.2.3), and (3) a nitrogen-containing base. Together, these three components form what is known as a **nucleotide** (just the deoxyribose sugar and the nitrogen base together are called a nucleoside) – make sure you understand what a nucleotide is because it will be used quite often throughout this chapter. The phosphate group and the deoxyribose sugar, when linked together in an alternating fashion, form a

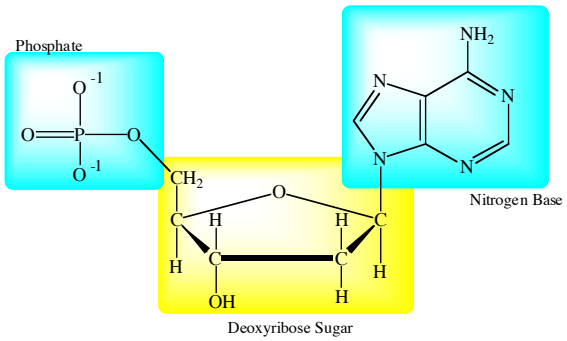


Figure 5.2.2. Nuclei acid (nucleotide) unit.

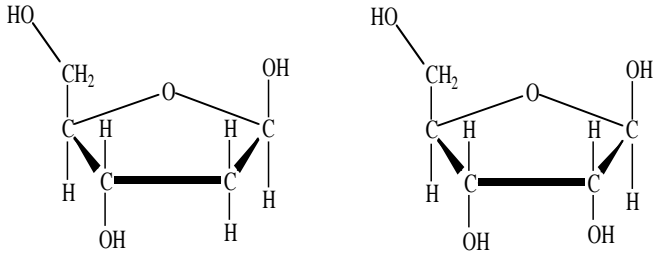


Figure 5.2.3. Deoxyribose (L) and ribose (R) sugars.

straight chain backbone for the DNA polymer. Attached to this backbone are attached nitrogen bases, somewhat similar to the way clothes are hung on a clothesline to dry. In fact, DNA uses only four different nitrogen base units: adenine, guanine, cytosine and thymine (RNA, a close relative of DNA and a molecule important in the transfer of information from DNA into biochemical action, uses uracil instead of thymine). These four bases are shown in Figure 5.2.4. All of the repeating units of the DNA backbone are completely identical (alternating phosphate and sugar units) while the pendant nitrogen bases are chosen from among this small group of four bases (A, T, G, and C). The entire fundamental structure of a strand of DNA thus consists of repeated phosphate-sugar-nitrogen base building blocks that come together to form the observed complete DNA structure as shown in Figure 5.2.5.

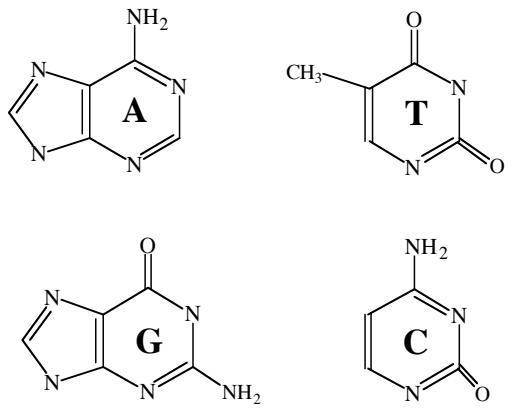


Figure 5.2.4. Nitrogen bases used in DNA (*clockwise from upper left*): adenine, thymine, cytosine, and guanine.

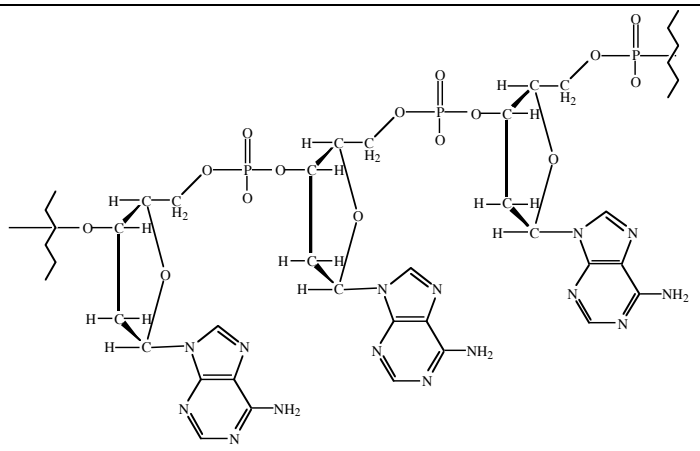


Figure 5.2.5. Structure of DNA polymer – sugar phosphate backbone with pendant adenine nitrogen bases.

The order of the nucleotide units in the DNA chain can be readily specified by simply writing the order of the nitrogen bases in the chain since all the other components are always the same. The

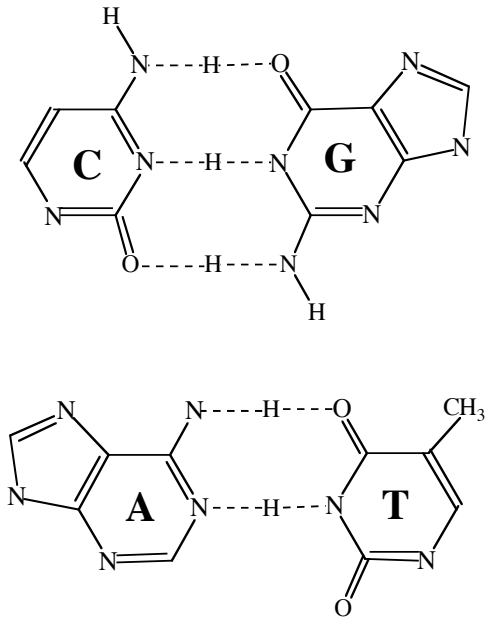


Figure 5.2.6. Hydrogen bonds formed between DNA bases pairs [C = Cytosine, G = Guanine, A = adenine, T = Thymine].

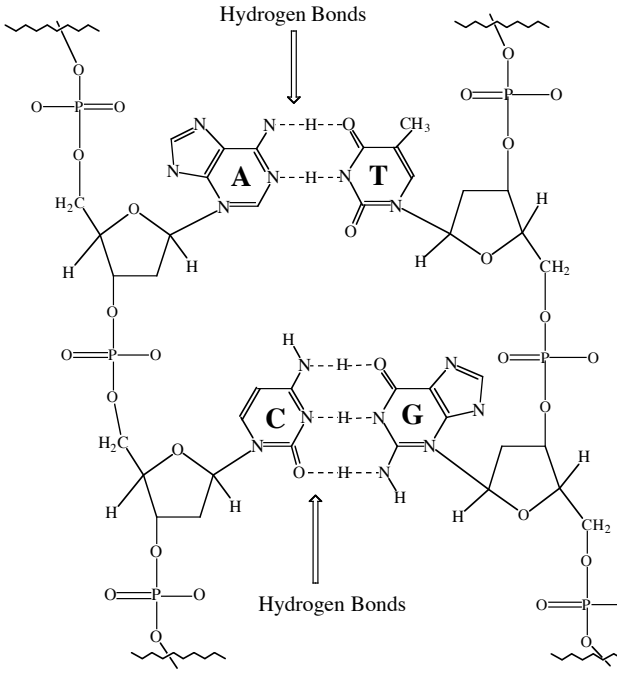


Figure 5.2.7. Hydrogen bonded bases pairs between two strands of DNA[C = Cytosine, G = Guanine, A = adenine, T = Thymine].

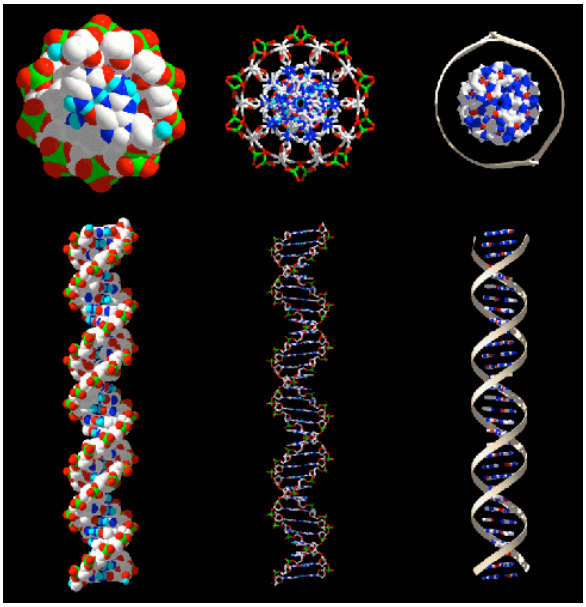


Figure 5.2.8. Several views of the DNA double helix (http://www.psc.edu/~deerfiel/NIH/B-DNA.gif).

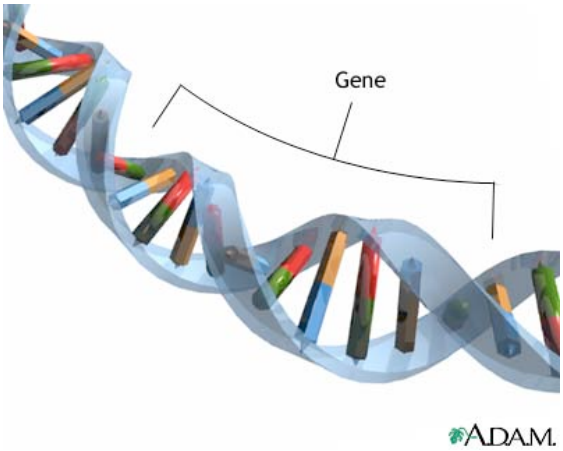


Figure 5.2.9. A gene is a sequence of nuclei acids along the DNA chain (www.dna-sequencing-service.com/dna-sequencing/gene-dna/).

sequence shown in Figure 5.2.5 is "AAA" since the three bases shown are all adenines, while a much larger, 18-nucleotide DNA sequence, could be completely defined as "AAACGTAACGTCGGTAAA". Translated, this means that this latter DNA chain consists of three connected adenine nucleotide units, followed by a cytosine unit, followed by a guanine unit, followed by a thymine unit and so on until the end of the chain is reached.

In a typical strand of the DNA polymer, huge numbers of nucleotide units are strung consecutively together to form a very, very long chain. DNA strands in the nuclei of human cells may contain around 6 billion nucleotide units connected into long chains. These strands of the DNA polymer are usually found in nature not as individual strands but rather as **complementary pairs** of strands, meaning that two DNA chains are required to come together in order to form the observed complete double helical DNA structure. These strands don't just come together in any random fashion but are instead directed together by very specific chemical interactions between the two chains. To form the double strand arrangement, two nitrogen bases, one from each chain, must come together to form a close electrostatic attraction, called a **hydrogen bond**. Only very specific pairs of bases can interact in this fashion as dictated by their chemical structures. In this way, only cytosine and guanine are chemically built to allow a close interaction – close enough to form a hydrogen bond that electrostatically holds together the two bases, and therefore, also holds together the two DNA chains.

Likewise, adenine and thymine can form a hydrogen-bonded unit. These interactions are shown in Figure 5.2.6 for the only two possible combinations. If, however, cytosine and thymine or guanine and adenine are forced to come together, no hydrogen bond can be formed since their structures do not allow for the proper electrostatic "fit". This would be like trying to plug a 3-pronged electrical cord into a 2-pronged outlet; it just doesn't work. When every adenine (A) on one DNA strand is matched with a thymine (T) on the other strand and every cytosine (C) is only matched with a guanine (G), the two strands are said to have a complementary sequence of bases. Having a complementary sequence on the two DNA strands yields a new double chain that is effectively "pinned together" through the hydrogen bonding of the nitrogen base pairs. This is shown schematically in Figure 5.2.7. When these two complementary strands of DNA come together, they form the well-known double helical structure shown in Figure 5.2.8 (like a twisted flight of stairs).

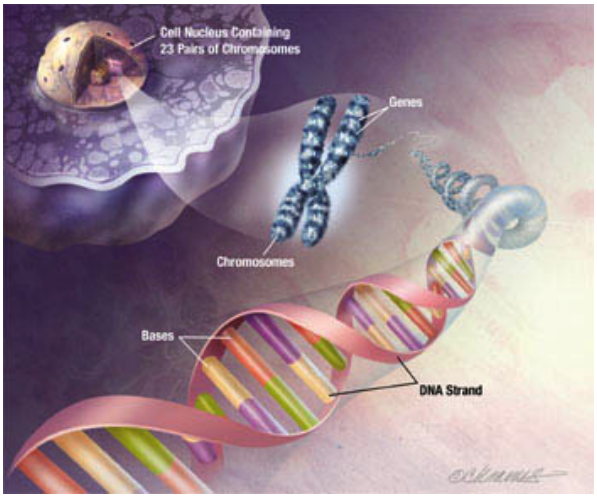


Figure 5.2.10. Nuclear DNA
(http://www.alzheimers.org/images/dna.jpg).

The sequence of nitrogen bases in the DNA polymer is the fundamental basis for controlling all of life's cellular processes and ultimately determining things like what color eyes we have or how tall we might become. The basic DNA unit that determines which characteristics are transmitted to succeeding generations is called a **gene.** A gene is simply a small portion of the DNA strand, a sequence of nucleic acids, that contains the DNA coded message for something inherited (Figure 5.2.9). Each gene regulates the formation of a specific enzyme or protein within the cells of the body, ultimately leading to the direct regulation of all of life's biochemical processes. At their simplest level,

DNA, Gene, Chromosome, Genome?

It may seem confusing, but there's a relatively simple explanation. DNA is a sequence of nuclei acids. A particular length, or sequence, of DNA is called a gene – a "word." The genetic and non-genetic lengths of DNA form form a strand of DNA – a "sentence." A long chain of DNA that is part of the entire DNA compliment of a cell is called a chromosome – a "chapter." All the chromosomes make up the genome – the entire "book."

our genes are really just composed of chemical "address" where we find a message written. The

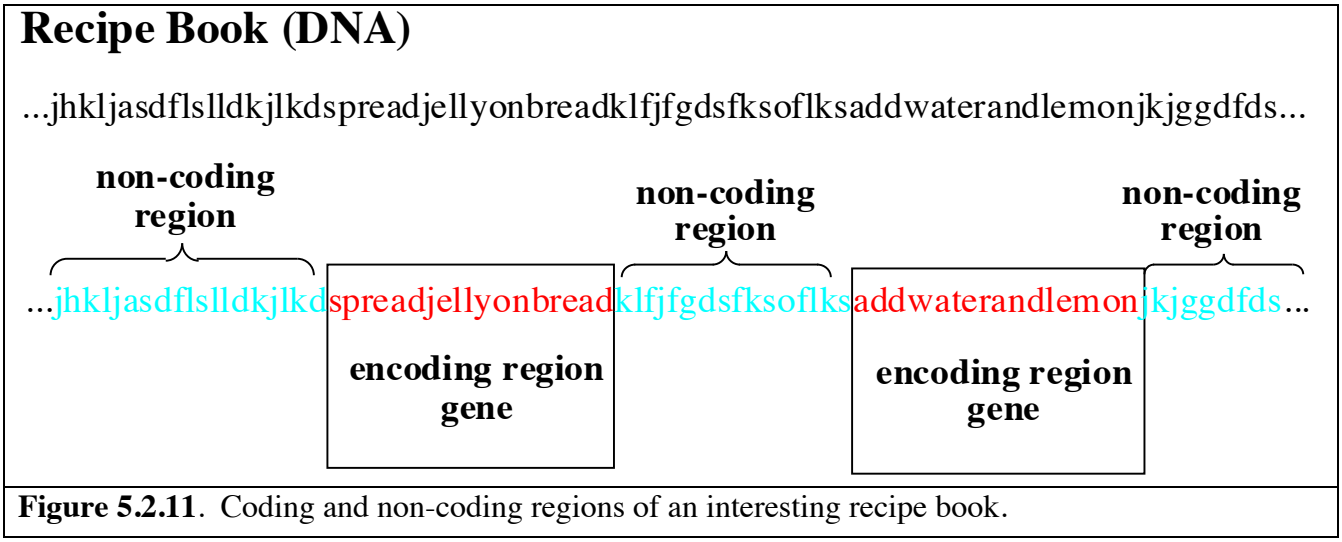


Figure 5.2.11. Coding and non-coding regions of an interesting recipe book.

basic genetic language of this message consists of just four letters (adenine, thymine, guanine, and cytosine) that provides all the chemical information necessary to exactly prepare a specific molecule required by the cell. Thus, a small region of the DNA strand forms a gene that is the blueprint for making a molecule (Figure 5.2.9). The location of a gene on a DNA strand is referred to as its **locus** (the plural is loci) with the genes residing along very specific stretches of the DNA chain (Figure 5.2.10).

A particularly important discovery in 1985 to forensic science was that the relatively short coding regions of DNA (genes) are often found separated from each other by long stretches of DNA that do not encode for anything known. These separating regions are called "**non-coding**" or **hypervariable** regions (sometimes also called "junk" DNA in the popular press) and will be discussed in detail later. The relationship between coding and non-coding DNA regions is illustrated schematically in Figure 5.2.11 using the cookbook analogy. In this analogy, the first part of the strand of letters is meaningless to us until we reach a section that encodes for words we do

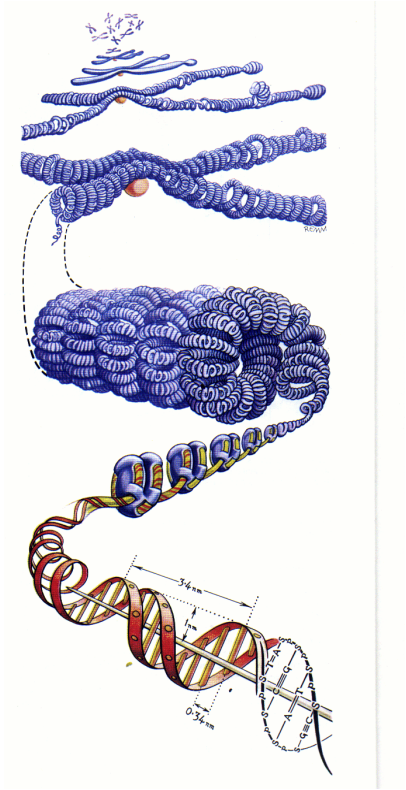


Figure 5.2.13. The coiling of DNA helices to form nuclear chromosomes (http://www.bay13.net/pics/desktop/morepictures/DNA.gif).

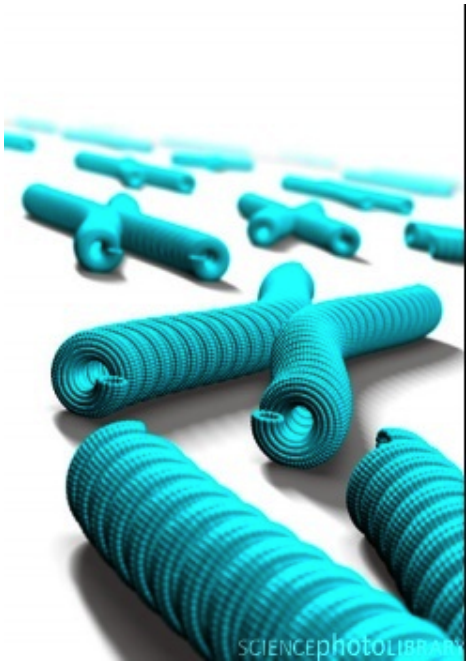


Figure 5.2.12. Human chromosomes showing the 23 pairs in the human genome (Computer artwork) composed of DNA (deoxyribonucleic acid) that and provides the information necessary for the cell to make the chemicals necessary for life (sciencephoto.com; No. P656/199)

understand ("spread jelly on bread"), called the coding region. After this understandable section follows another nonsense region (non-coding region) until we reach another understandable region ("add water and lemon"). In a similar fashion, DNA has long nonsense regions located between smaller genetic regions that code for something our do cells understand - biosynthesis

The DNA of an organism is at times found all coiled up to form chromosomes that is found either in the nucleus (nDNA) or in the mitochondria (mtDNA) of the cell (Figures 5.2.10 and 5.2.12). All cells of an individual organism have identical DNA, so it doesn't matter what cell the DNA is from, they all have exactly the same DNA code. Chromosomes are similar to the chapters of a book – convenient ways to "package" the story into smaller, more readily handled units. In humans, the DNA in most of our cells is arranged into 23 pairs of chromosomes that are identified by their characteristic sizes and shapes. One pair is referred to as the sex chromosome and can either be of a larger variety (the "X" chromosome) or a smaller version ("Y" chromosome). Females have two "X" chromosomes while males have a size mismatched "X" and "Y" pair. One-half of our genetic information comes from our mothers and one half from our fathers, with the "Y' chromosome inheritable only from fathers to sons (Figure 5.1.13).

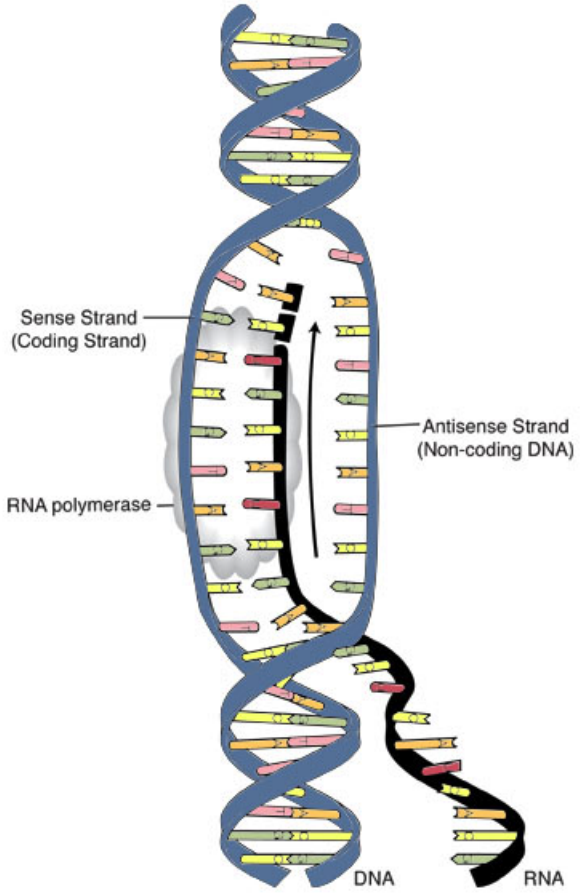


Image adapted from: National Human Genome Research Institute. Talking Glossary of Genetic Terms. Available at: www.genome.gov/Pages/Hyperion//DIR/VIP/Glossary/Illustration/antisense.shtml.

Figure 5.2.14. Schematic drawing showing the transcription of DNA into mRNA. (www.geneticsolutions.com/PageReq?id=1530:1873)

So how do we get from the basic chemical arrangement of DNA just discussed to the color of our eyes? The pathway that leads from DNA "letter" sequences to the synthesis of cellular proteins that regulates life employs two very important processes called transcription and translation. In a very brief summary, DNA cellular control works something like this. In the first process, called **transcription**, a portion of the DNA double helix is first unraveled into single strands in the region that codes for the needed cellular compound (the gene). A brand new strand, called a complimentary strand, forms with A's pairing with U's (since this is RNA that uses uracil in place of DNA's thymine) and G's pairing with C's between the original template and the newly forming strand. The new strand (called messenger RNA or mRNA) is an exact complimentary match to the original template in which every C in the original is matched to a G in the new strand and so on, as shown in Figure 5.2.14. This process can really move along

BRIEF ON DNA TRANSCRIPTION/TRANSLATION

(1) Information in the order of nucleotides in DNA can be transcribed and translated to direct the preparation of proteins in the cell.
(2) Every three nucleotides in DNA codes for one amino acid in the formation of a protein.
(3) The section of DNA coding for a protein is called a gene.
(4) DNA contains very large regions that code for nothing known.

very quickly, with an estimated 90,000 bases added per minute. The mRNA is somewhat like a negative image in photography that is formed from an actual (positive) scene. The negative carries all of the original visual information of the scene but only in reverse. The original positive image can easily be regenerated from the negative by using light and special photographic paper. The newly formed mRNA then moves out of the nucleus of the cell to an organelle in the cell's cytoplasm called the ribosome. On the ribosome, each group of three nitrogen bases in the mRNA strand uniquely designates one specific amino acid in the protein polymer molecule being synthesized. These amino acids are then chemically linked together in a new polymeric chain to form the needed protein in a process known as **translation**. For example, the mRNA code "GGG" translates directly into the command that a glycine amino acid needs to be placed in the growing chain while an "AAA" code specifies that a lysine amino acid be placed next in line. In this manner, the mRNA sequence "AAAAAAGGGAAA" would translate to a small protein composed of two glycines, followed by a lysine and ending with another glycine unit. It's important to see that the order of the bases in the DNA molecule is ultimately translated into a very specific sequence of amino acids that are linked together to form a new protein molecule. This is conceptually quite similar to translating a code into equivalent words in a language. In this analogy, the letters in the word "yes" in English are translated uniquely into the letters of the word "oui" in French. Analogously, "GGG" in DNA is translated into "glycine" in the protein. In DNA, all "words" have just three letters so that exactly three nitrogen bases (chosen from A, T/U, G, and C) are needed to specify one amino acid unit in the growing protein chain. Protein composition, structure, and function (including enzymes, structural proteins, and many others) are, therefore, uniquely dictated by the DNA base ordering. This process is shown schematically in Figure 5.2.15. Since the chemical and physical properties of a protein are determined fundamentally by the order of its linked amino acid units, DNA thus ultimately controls all our protein synthesis and, in turn, cellular processes.

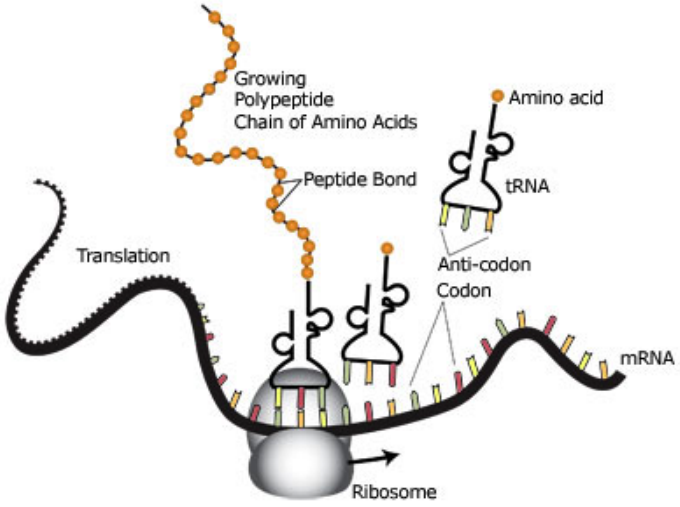


Image adapted from: National Human Genome Research Institute.

Figure 5.2.15. Translation of the sequence of DNA nucleotides into the formation of a protein (from www.geneticsolutions.com/PageReq?id=1530:1873).

In 2003, researchers around the world completed the largest scientific undertaking in history called the Human Genome Project (HGP). This project completely determined the DNA base sequence in all of the approximately 30,000 human genes – in other words, the precise order of nearly 3 billion A, T, G, and C nucleotides in human DNA was figured out – a truly amazing accomplishment! One significant discovery was that we share about 93% of our DNA genetic code with the fruit fly and roundworm. Much more surprising, however, was the finding that there are only about one-third the total number of genes as were originally predicted. The size of individual genes varies quite a bit but an average gene contains about 3000 bases, with the largest known gene having about 2.4 million bases (dystrophin). Also, quite surprisingly, only about 1.5% of the DNA in our bodies actually codes for compounds needed for life! This means that the vast majority of the human DNA polymer does not code for anything known to be biologically important. Our DNA, therefore, contains far

more random "data" than is contained within the coding portions of the DNA. This would be a book with only one sentence out of a hundred with discernable meaning.

Our genes appear to be concentrated in areas along the chromosomes, with vast expanses of non-coding DNA lying between genes (hypervariable, "junk" or "non-coding" regions). Stretches of up to 30,000 repeating C and G bases tend to occur near to concentrations of genes that forms a "barrier" between the genes and the hypervariable DNA, while gene-poor regions are predominantly formed from A and T nucleotides. These non-coding portions may either have had a function in the past but are now not necessary for life or may have arisen by other means such as mutation or viral insertions. It is these "non-coding" repetitive regions that are at the heart of forensic DNA analysis.

BRIEF ON HUMAN GENOME PROJECT

(1) Completed in 2003.
(2) Determined base (nucleotide) sequence of all 30,000 human genes (the order of 3164.7 million nucleotides).
(3) Found humans share ca. 93% of our DNA code with roundworms.
(4) Found only ca. 1.5% of our DNA codes for compounds.

Helpful Reminders

Nucleotide – the building block of DNA, consisting of a sugar and phosphate chain with linked bases.
Gene – Small sequence on nucleotides along the DNA chain.
Chromosome – a long chain of **DNA** containing many genes and spacers (non-coding).
Genome – the entire DNA component of an organism (possibly many chromosomes).
RNA – the analog of DNA that transfers genetic information into the cell.
Ribosome – where the genetic code directs the synthesis of a protein.
Tandem Repeat – a specific pattern of nucleotides in DNA that is repeated back-to-back.

Scientists have discovered that, within the non-coding regions, certain short sequences of nucleotide "letters" (A, T, C, and G) are repeated back-to-back numerous times throughout the DNA chain as shown in Figure 5.2.16. In fact, while we don't yet understand why this occurs, it appears that more than fifty per cent of the human genome is composed of these repeating sequences. Repeated sequences that are arranged consecutively along the DNA chain in the same relative orientation to one another are called **variable number** (or nucleotide) **tandem repeats** (**VNTR**) and are centrally important to one type of DNA forensic test described later, the RFLP method. VNTR sequences usually consist of seven to twenty-five nucleotides connected to one another in up to fifty tandem repeats (tandem here means simply connected one to another as railroad cars connect together to form a train). The VNTR sequences usually consist of seven to twenty-five nucleotides that are connected to one another in up to about fifty tandem repeats (tandem here means simply that they are connected one to another as railroad cars connect together to form a train). These VNTR sequences show a great deal of variability in the number of repeats possible for each place (locus) on the DNA chain. For example, one person may have a particular DNA sequence repeated ten times in

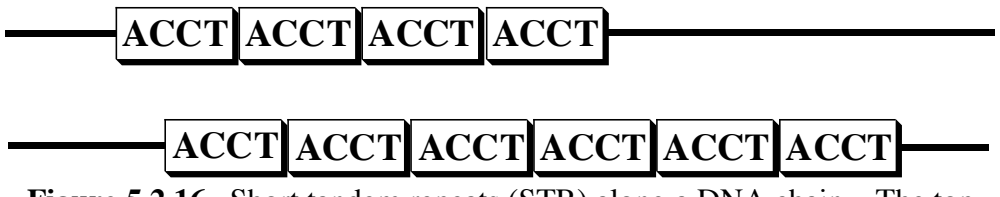


Figure 5.2.16. Short tandem repeats (STR) along a DNA chain. The top sequence consists of four repeated units while the bottom consists of six units.

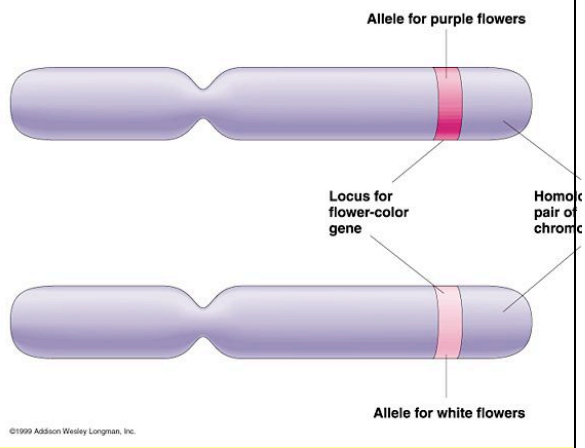
a row while someone else has the same sequence repeated twenty times in a row at the same DNA location (locus). Another type of repeated pattern is very similar to the VNTR's except that it is made up of both *shorter* repeated sections (fewer "letters") and fewer tandem repeats. These shorter patterns are called **short tandem repeats** (**STR**), Figure 5.2.16. The use of STR sequences is the basis of much of the current human forensic DNA testing procedures.

Since protein structure and function derive directly from the ordering of the nitrogen bases in DNA, very small changes in the base ordering in the genetic regions of DNA may cause catastrophic changes in life function. One base unit incorrectly placed can mean the difference between disease or no disease or even life versus death. For example, the substitution of a single nitrogen base for the correct base in the gene regulating hemoglobin production leads to one misplaced amino acid (switching one valine for a glutamic acid) out of the entire 146 amino acids within the hemoglobin molecule. This very simple change, however, may result in a person having sickle cell anemia. Within a species such as humans there is very little variation of the DNA code within genes (more than 99% of our **genetic** DNA is the same from person to person). In contrast, there can be very large differences in our non-genetic (no-coding) DNA since there is no survival difference from changing the code in these regions – they do not code for any biomolecules.

Changes in the order of the nitrogen bases in the DNA chain (called **mutations**), are usually not well tolerated and typically do not result in viable offspring. The genetic (coding) portions of the DNA for each healthy human is essentially the same so that if we were to look at these regions, we'd find very little difference from one person to the next. The genetic coding regions are, therefore, not especially helpful in individual forensic identifications – we're all essentially the same in these regions. Changes in the "nonsense" non-coding DNA regions that separate the genes, however, usually make no difference in the survival of an individual organism since these parts of the DNA code play no role in regulating cells or in developing the traits we observe. Over time, these nonsensical regions have become quite diverse

Allele versus Gene

A gene is a particular sequence of nucleic acids at a particular place along the DNA chain. An allele, is one of several possible variation of a gene.



(http://weloveteaching.com/2011/genetics/mendeliangenetics.html)

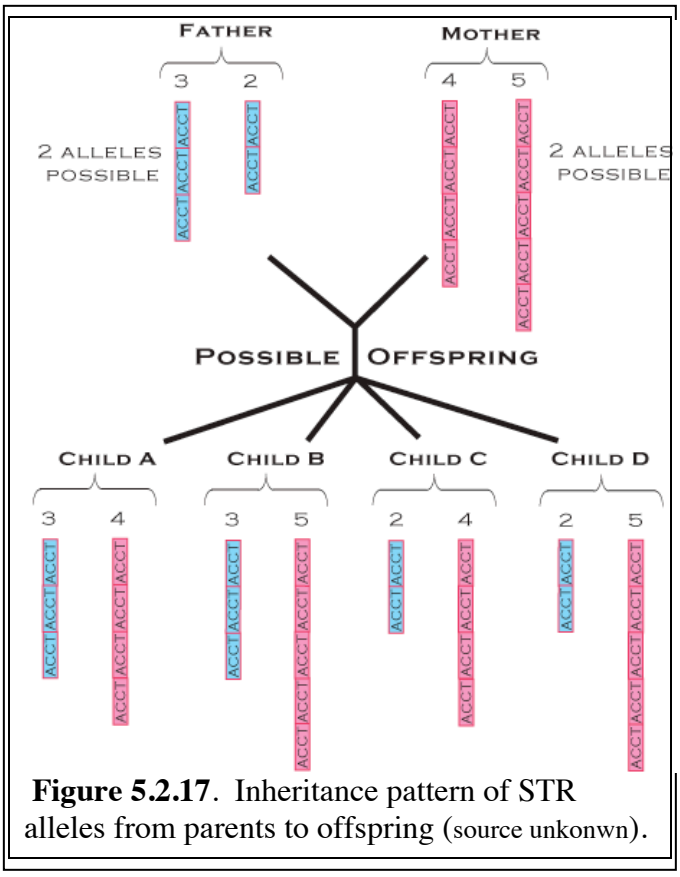


Figure 5.2.17. Inheritance pattern of STR alleles from parents to offspring (source unkonwn).

such that essentially no two people have exactly the same DNA codes in these inter-gene regions (except for identical twins). These differences in our non-coding DNA sequences are called **polymorphisms** (or many forms). The basis, therefore, of forensic DNA testing is to look at the DNA codes in these inter-gene regions since looking at the genetic coding regions would not be able to discriminate one person's DNA from another.

Case History: Chimerism and the Missing Mother

In 2002, Lydia Fairchild was pregnant with her third child when she applied for welfare support. In order to receive the support, however, she was required to provide DNA paternity evidence to show that her estranged partner, Jamie Townsend, was the father of her children. The DNA results came back with some startling news: while Townsend was shown to be the father, Fairchild was also not the children's mother. Fairchild was promptly charged with fraud for claiming welfare benefits using other people's children. When her third child was born, court ordered, administered, and witnessed blood tests were performed on both Fairchild and the newborn in the hospital. These DNA tests also amazingly showed that Fairchild was not the biological mother of the child she had just delivered.

The answer to this strange dilemma came when a similar case in Boston was discovered that had been found to be a case of chimerism. Chimerism arises when a person has two completely separate cell lines within their bodies, each with its own entirely different set of chromosomes. This is thought to arise when two separate embryos are physically mixed, each with its own DNA compliment. During development, different tissues within the body develop from different cell strains.

Further DNA tests on Fairchild showed that samples from her skin did not match her children but samples from her cervical tissue did, indeed, match. It is now thought that natural chimerism, previously believed to be very rare, may be more common than had been presumed. If so, how might this change the role of DNA in legal proceedings?

Sheep/goat chimera. A Chimera contains a mixture of two or more genetically different cell strains. This goat is a chimera made up of white goat fur and grey sheep's wool (www.sciencephoto.com, G285/027).

In a more detailed look at these polymorphisms, many variations are possible at specific loci along the DNA chain. For example, at a particular locus, there may be several variants of the tandem repeat of an "ACCT" nucleotide sequence. This simple STR repeat may occur in a row two times, three times, or more. Each of these possible forms is referred to as an **allele (see box "Allele versus Gene")**. In Figure 5.2.17, the inheritance pattern for a simple STR sequence is shown at a particular DNA locus. If, for instance, the father has the "ACCT" sequence repeated three times on one chromosome and only two times on the paired chromosome, he has two alleles at this specific locus (called heterozygous if the two alleles are different and homozygous if they are the same) (Remember we're talking about the non-coding regions of the DNA). The father's allele composition here can be simply referred to as 3,2. If the mother, however, has four and five repeats on her two chromosomes at the same position, she is 4,5. Four allele combinations are, therefore, possible for any resulting offspring from these parents; (3,4), (3,5), (2,4), and (2,5).

Humans differ in their DNA compositions at many loci in the human genome. Therefore, to measure a person's individuality we need simply to look to differences in their allele compositions at specific loci along their DNA chain.

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.

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.

In our DNA, we all have very similar

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 "GGGCCC" 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.

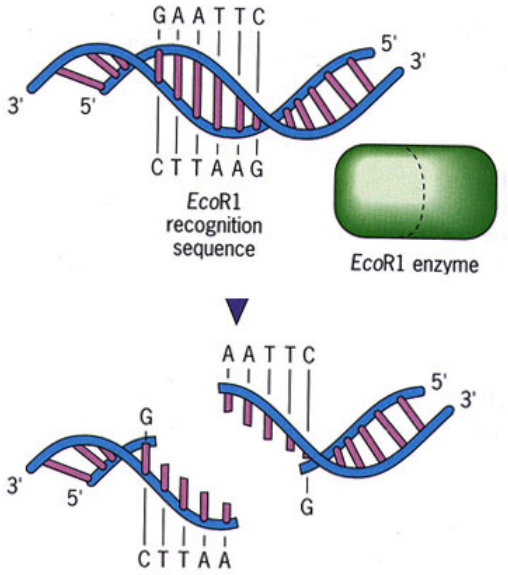


Figure 5.3.1. Schematic of the cutting action of the EcoR1 restriction enzyme on a portion of DNA it recognizes. (http://bioinquiry.biol.vt.edu/bioinquiry/BioTech1/biotechpaid/biotechhtmls/technic5.html)

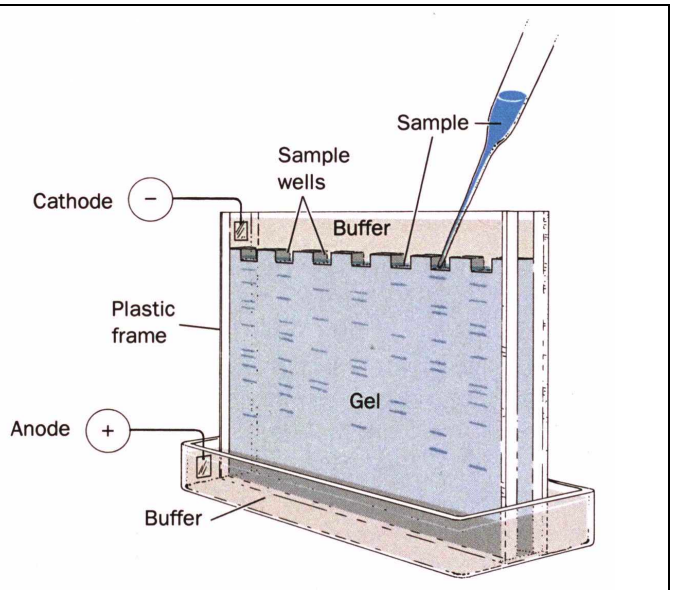


Figure 5.3.3. Experimental electrophoresis setup. (www.science.fau.edu/chemistry/Mari/biochemlab/manual.html)

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

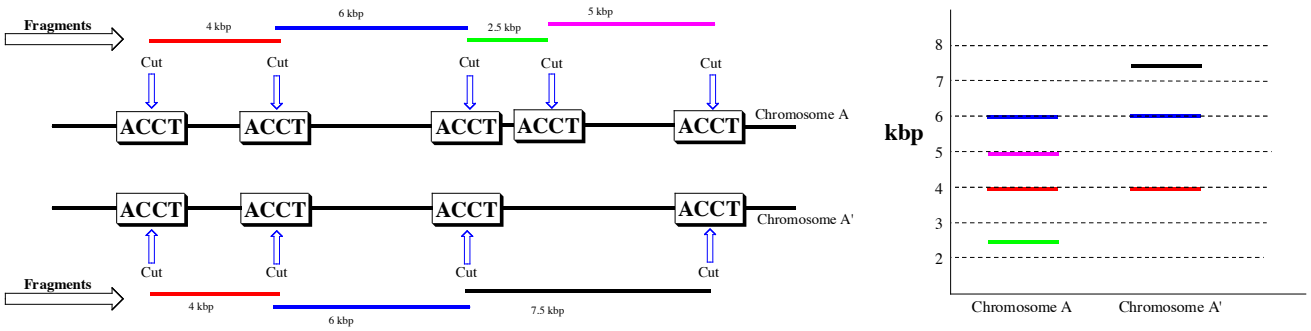


Figure 5.3.2. At left are shown possible "ACCT" repeat patterns and the sites of action for a restriction enzyme on two DNA strands, A and A' (kbp stands for 1000 base pairs). At right is shown a plot of the size distributions arising from the two cut DNA samples.

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 $^1/_4$ x $^1/_4$ 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

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

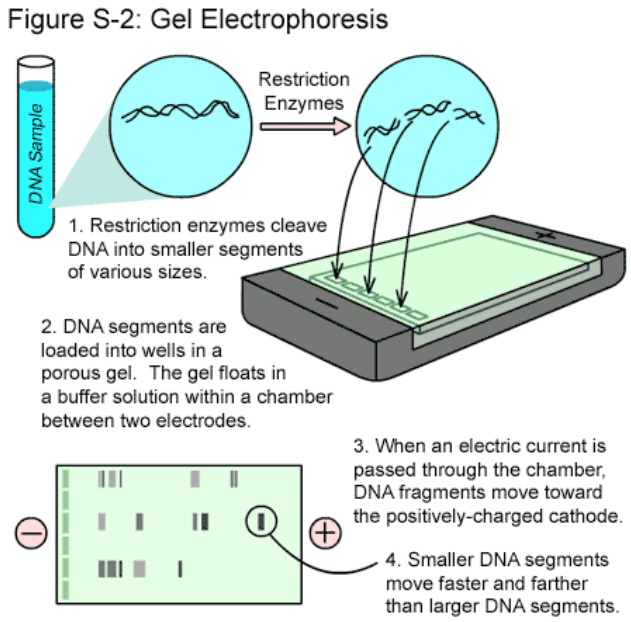


Figure 5.3.4. Gel electrophoresis technique for DNA fingerprinting. (www.stanford.edu/group/hopes/diagnsis/gentest/s7.html)

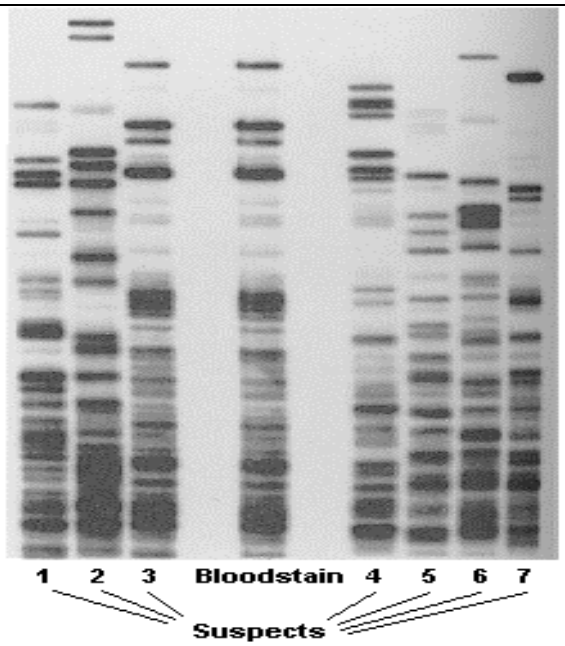


Figure 5.3.5. Autoradiogram from the gel electrophoresis of RFLP samples. (From Prof Bill Shields, ESF)

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 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.

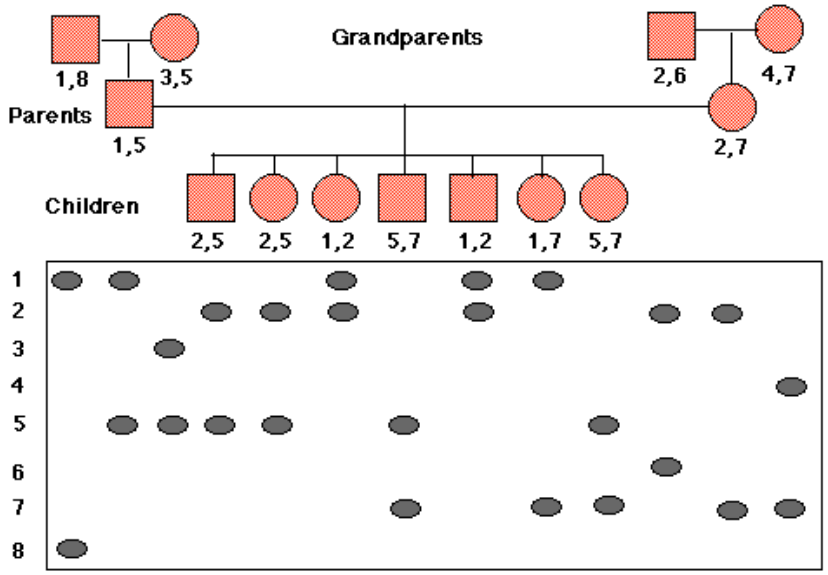


Figure 5.3.6. Inheritance of RFLP marker of three generations of one family. (users.rcn.com/jkimball.ma.ultranet/Biolo gyPages/R/RFLPs.html).

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

Case History: The Bacterial Trail


Patricia Pfoutz

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."

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.

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.

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

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

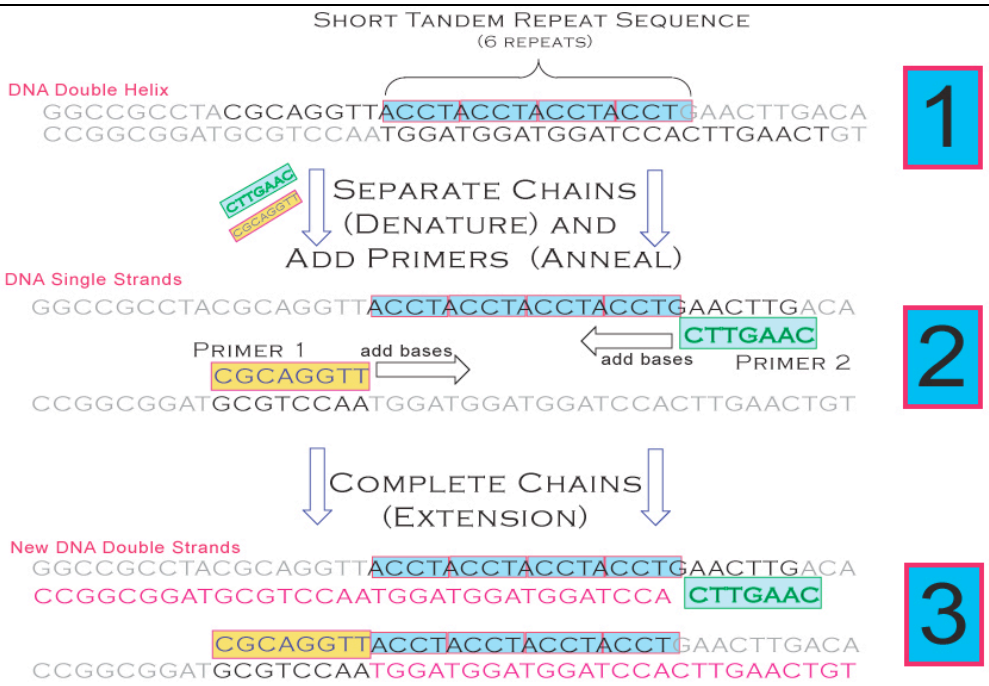


Figure 5.3.7. Steps in the first cycle of PCR duplication (amplification of a DNA target sequence.

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

"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.

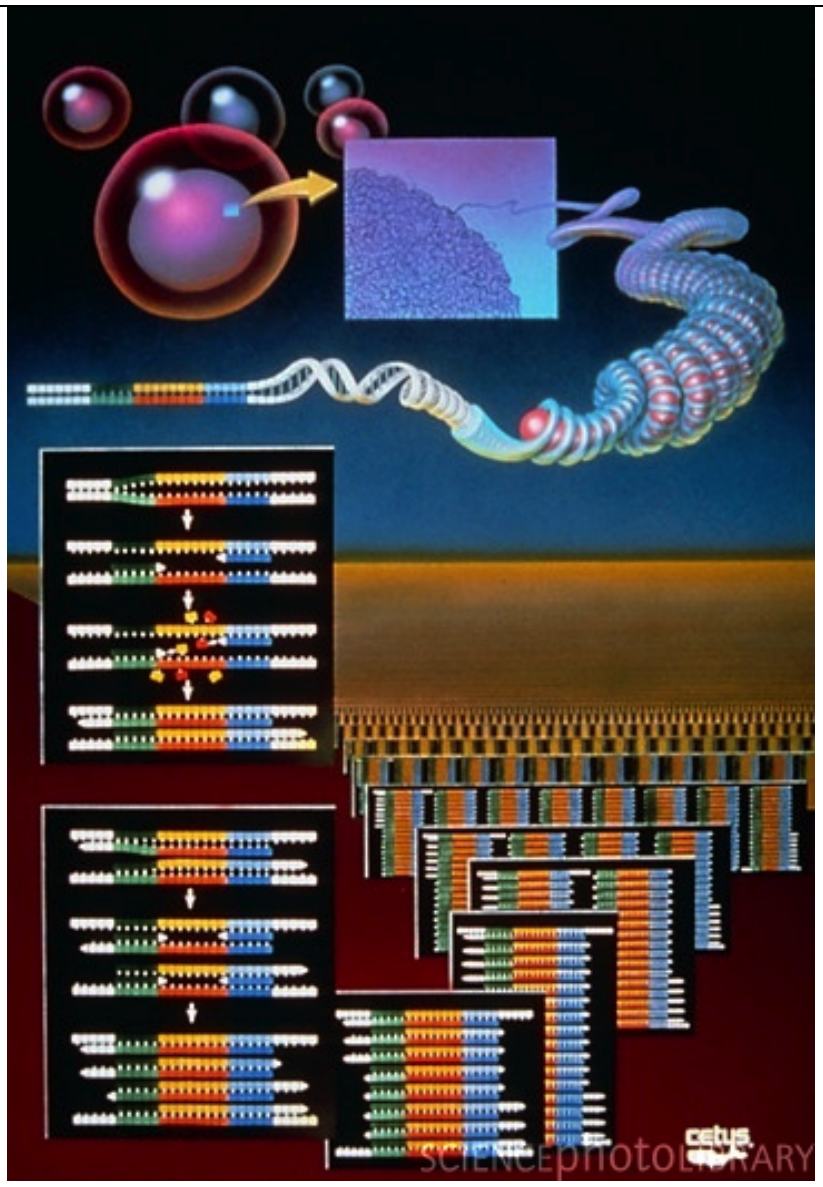


Figure 5.3.8. A schematic drawing showing how the polymerase chain reaction works to produce copies of selected regions of the DNA strand. "At right a sample of double-stranded DNA is shown unwound from one of the chromosomes of a cell nucleus. The two strands of the sample DNA are split apart (right 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).

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

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.

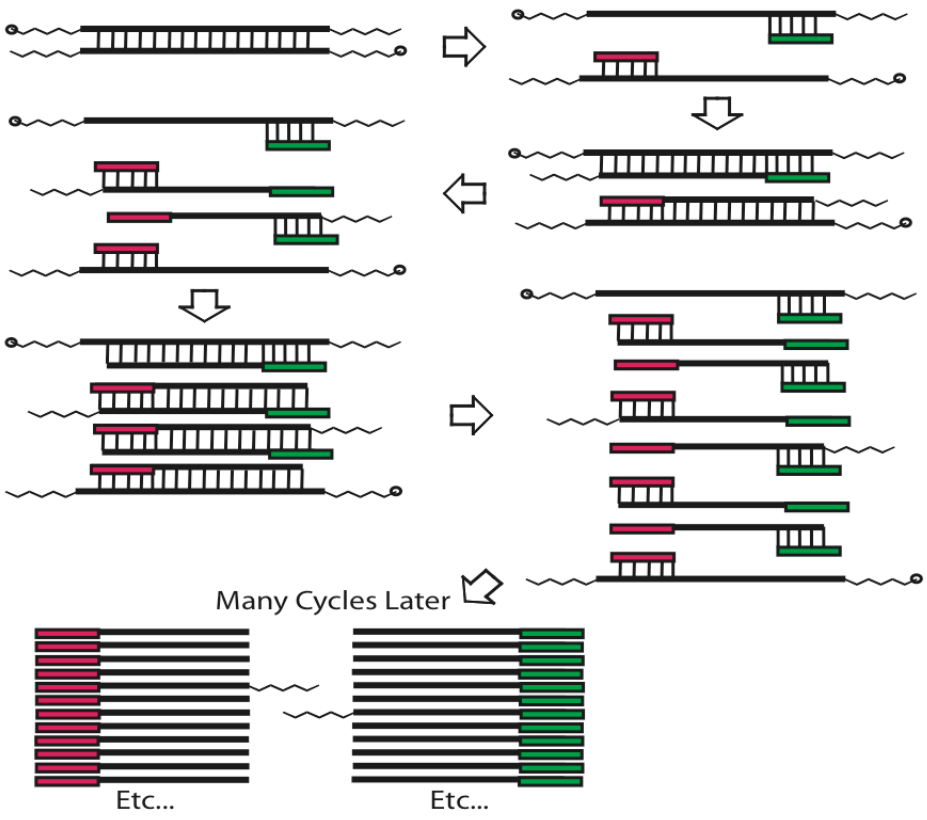


Figure 5.3.9. Production of target STR region copies by PCR (adapted by JTS from Figure 1 from "DNA Typing and Protocols" see bibliography).

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 **STR**s 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.

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.

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.

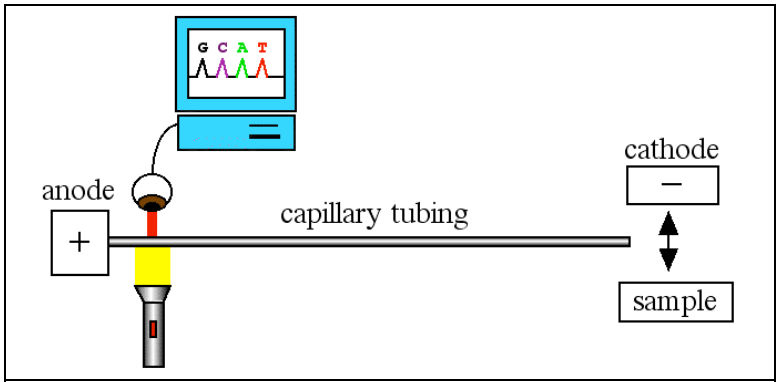


Figure 5.3.10. Schematic of capillary electrophoresis (from www.bio.davidson.edu/courses/genomics/method/Capillary.html).

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)

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

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.

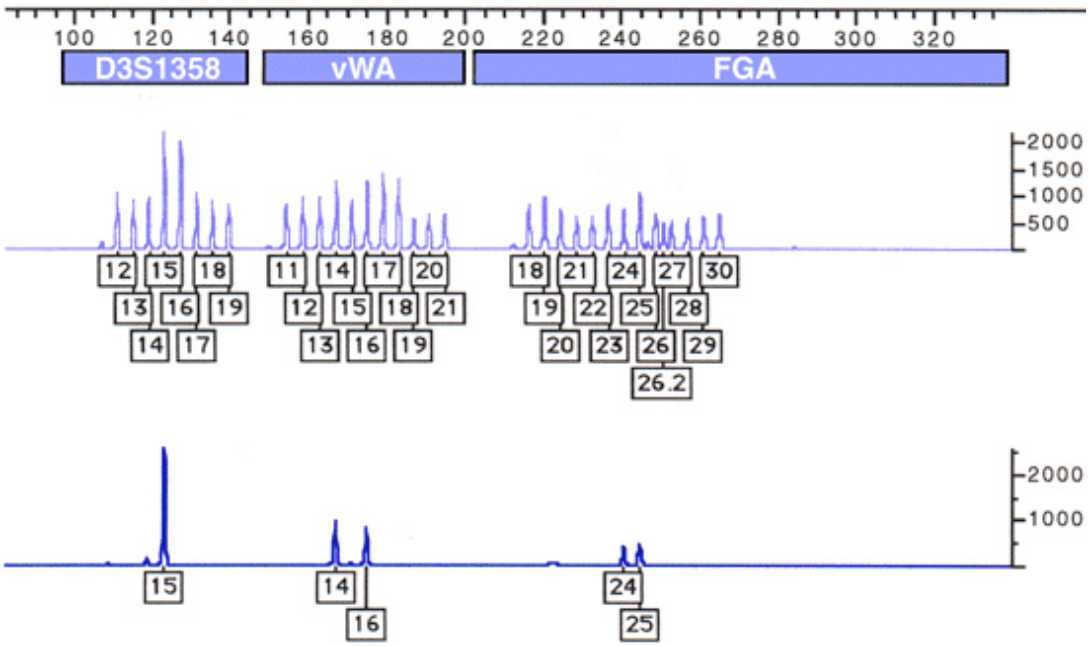


Figure 5.3.11. STR loci from one reaction mixture. Scale at the top is in base pairs (bp) (from www.biology.arizona.edu/human_bio/activities/blackett2/str_analysis.html)

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

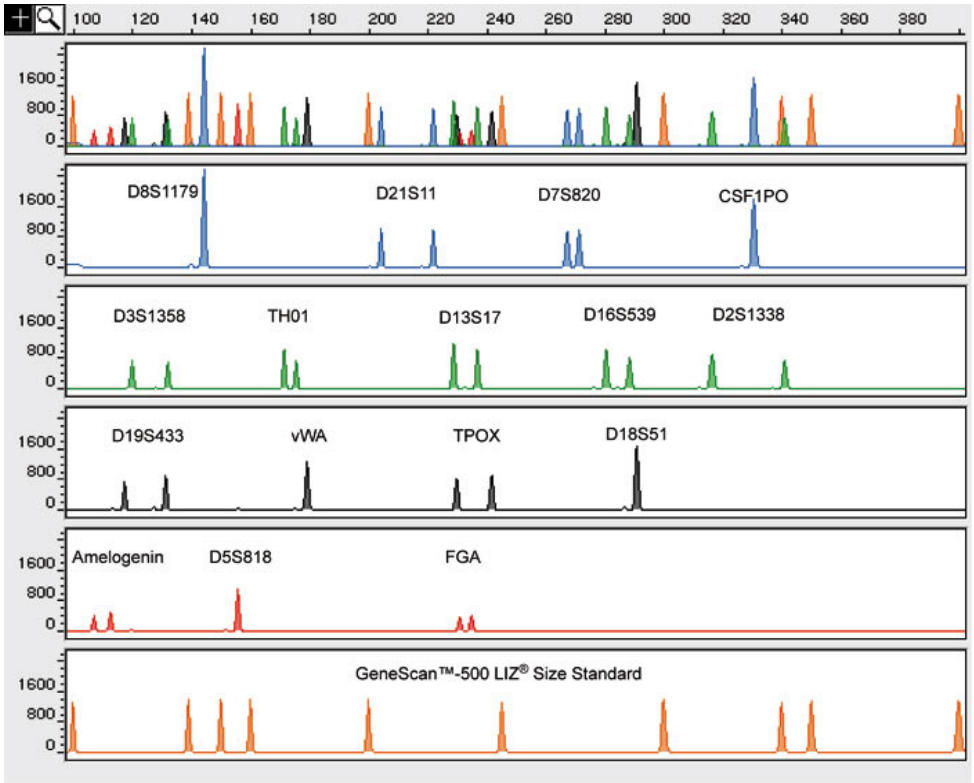


Figure 5.3.12. Multiplex for 15 STR loci and the Amelogenin locus. DNA fragments are labeled in blue, green, yellow (depicted in black), and red dyes. The GeneScan size standard is labeled with orange dye. (from

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 has11 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.

	RFLP	STR
Sequence Size	Employs variable nucleotide tandem repeats (VNTR) with seven to twenty-five nucleotides connected in up to fifty repeats.	Employs short tandem repeats (STR) with 2 to 6 nucleotides repeated consecutively 7 – 15 times.
Sample Size	Relatively large amounts of DNA are required (ca. 100m ng or the DNA from several thousand cells).	Relatively small amounts of DNA are required (ca. 1 ng from about 20-50 cells).
Sample Quality	Requires good quality (undegraded) DNA samples.	Works well on fragmented or partially degraded DNA.
Reliability	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.	Based upon probabilities of random match in a population. Depends upon comparison with proper population and availability of statistical data.
Speed	Relatively slow.	Very rapid analysis.
Other Advantages and Problems	Requires relatively large amounts of DNA. Relatively insensitive to impurities.	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.

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.

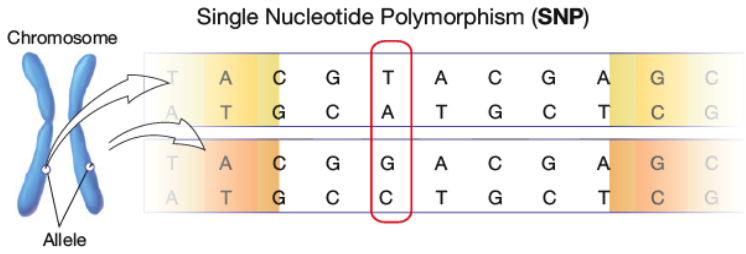
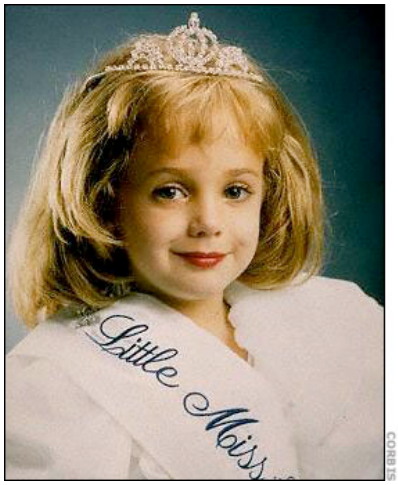


Figure 5.3.13. Single nucleotide polymorphism (SNP). (www.nlm.nih.gov/visibleproofs/education/dna/snp.pdf)

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

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 tool 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.

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.

5.4. Mitochondrial DNA and Y Chromosomal Typing

Mitochondrial DNA. The DNA we've looked at so far is that which is found in the nuclei of cells but a "second genome" exists within each cell. A small, circular portion of DNA is found outside of the nucleus in cellular structures called mitochondria (referred to as mtDNA or mitochondrial DNA). It's thought that at one time, mitochondria were separate bacterial organisms that eventually became permanent residents within cells, providing the cell with a means to convert food stocks and oxygen into useable energy in return for a rich source of nutrients. Typically, each cell has many, many mitochondria, each with it's own identical copy of this small loop of DNA, shown in Figures 5.4.2 and 5.4.2.

Unlike nuclear DNA, the DNA in mitochondria is inherited *only* from mothers - matrilineal inheritance. At fertilization, one-half of the chromosomes in the nucleus comes from the sperm cell and the other half from the egg. The egg cell outside the nucleus may have 100,000 or more mitochondria distributed throughout the cell while the sperm has only a few hundred mitochondria that are located primarily at the base of the flagellum. The flagellum and, essentially, all of the sperm's mitochondria do not penetrate into the egg at fertilization, leaving just the mother's mitochondria in the fertilized cell. This is shown schematically in Figure 5.4.3. The absence of a mixing of paternal and maternal DNA in the mitochondria means that the mtDNA may be inherited

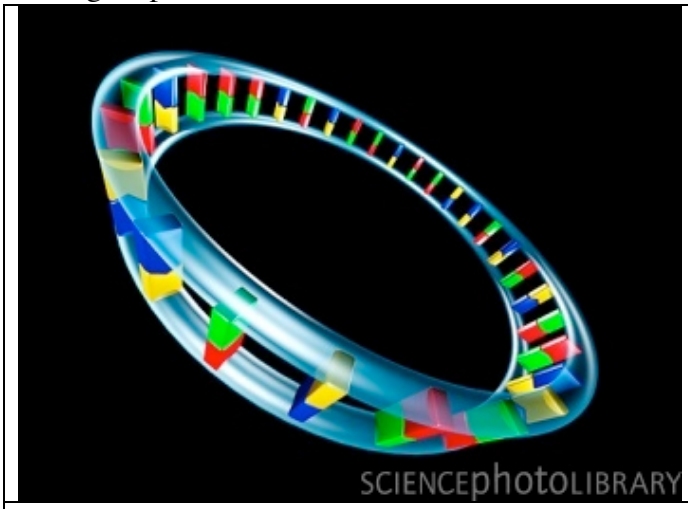


Figure 5.4.1. Circular mitochondrial human DNA inherited from the maternal line (G110/835; www.sciencephoto.com).

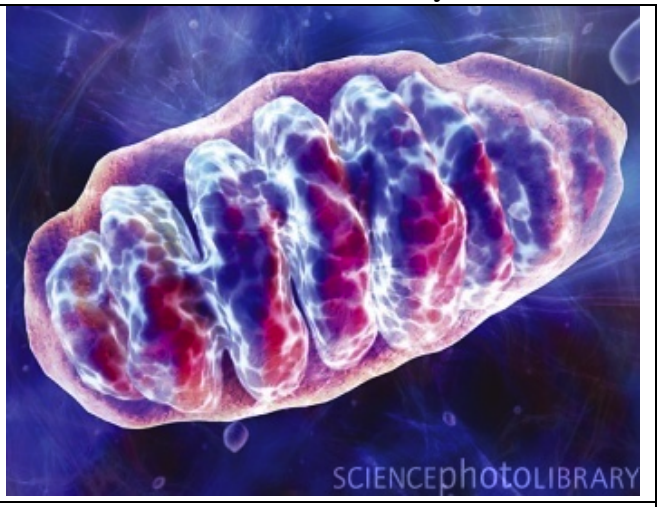


Figure 5.4.2. Artists depiction of a Mitochondrion (G465/065; www.sciencephoto.com).

over thousands of generations without any changes along the mother's bloodline. Thus, mtDNA typing cannot typically distinguish between people along the same matrilineal decent.

In instances where nuclear DNA material is either scarce (e.g., bones, hair, teeth), badly damaged, or entirely absent (such as in our mature red blood cells and hair shafts that do not contain nuclei), it is possible to examine mtDNA for forensic information. While each cell has only one copy of nuclear DNA, it may contain tens of thousands of copies of mtDNA. This means that it might be possible to obtain a mtDNA profile from the DNA found in a single cell.

Most of the mitochondrial DNA codes directly for important proteins needed for respiration and there is comparatively little non-coding or "junk" DNA in mtDNA. For instance, human mtDNA contains a total 16,569 nucleotides (compared with about 3 billion in nDNA) and all but about 1,200 of these code for proteins in mtDNA. Most of these 1,200 non-coding bases are located together in a region know as the control region (Figure 5.4.4).

In forensic mtDNA typing, about 610 nucleotides within the control region are sequenced and compared with a standard sequence. Differences between the standard sequence and that of the

sample are then recorded and analyzed. Currently, mtDNA analysis is difficult, time-consuming, and expensive. Thus its use has been limited to cases in which other DNA methods are not possible or maternal lineage information is particularly needed.

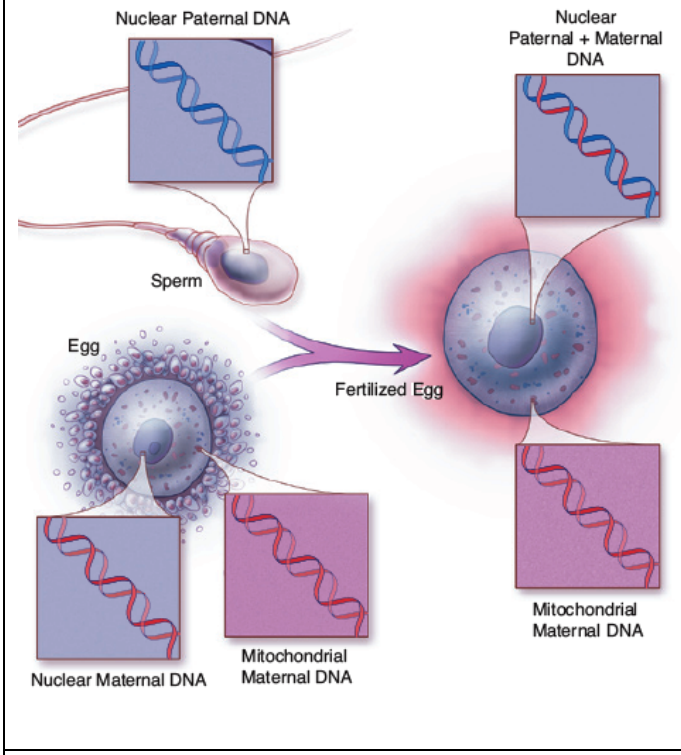


Figure 5.4.3. Inheritance of mtDNA. (www.nlm.nih.gov/visibleproofs/education/dna/mtdna.pdf)

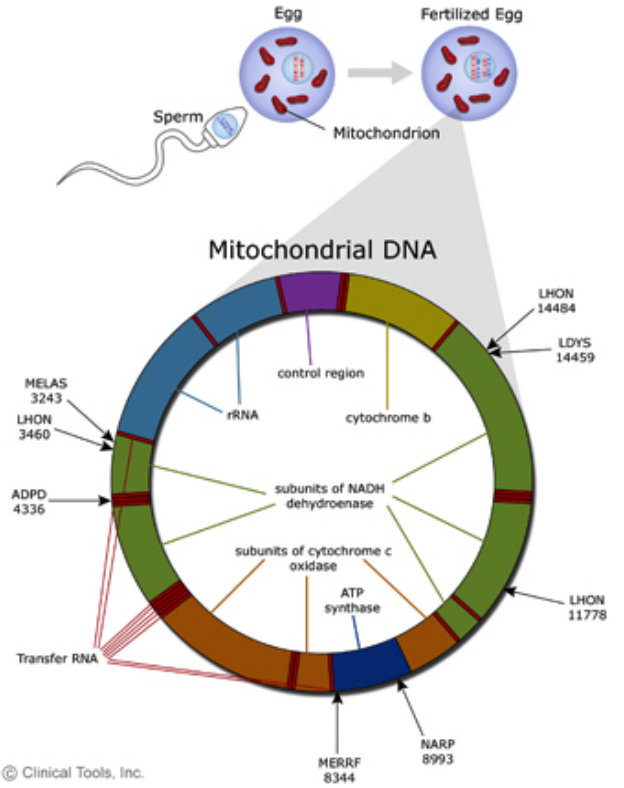


Figure 5.4.4. Mitochondial DNA inheritance and structure (from www.geneticsolutions.com/PageReq?id=1530:1873).

Y Chromosome Typing

Just as mitochondrial DNA is inherited solely from mothers, the Y chromosome in the nucleus is inherited only from fathers. Thus, looking specifically at STRs on the Y chromosome can be used to trace paternal lineage and identify male bodies through paternal relatives. This information is particularly valuable in missing persons, sexual assault, and paternity investigations. The analysis can be very rapid and can often distinguish between single and multiple male donors in sexual assault cases. Additionally, the primers for the Y STRs ignore DNA from females entirely so that separations of male from female cells prior to PCR is not necessary, eliminating a great deal of hard work. Identifying Y STRs from samples taken at different locations at a crime scene can be used to establish where blood samples from males are present.

Y STR analysis has only recently been employed in forensic work since it was originally thought that there was less variability in this chromosome. It has now been established, however, that up to 60% of the Y chromosome is composed of repetitive DNA that lends itself to this type of analysis.

African Lemba Lineage

A southern African tribe believes that they are direct descendants of the Jewish line of Abraham. At the same time, a group of Jews called the Cohanim, wanted to determine if they were direct descendants of Aaron, brother of Moses and direct descendant of Abraham. To answer both of these questions, DNA tests using Y-chromosome markers were performed on people of all faiths as a comparison population. It was found that one particular Y-chromosome marker was found in fifty percent of the Cohanim males, ten percent of all Jewish males, and essentially unknown for those of other faiths. DNA tests of the Lemba tribal leaders showed that fifty percent of these leaders also carried the Y-marker, while the marker was found in ten percent of the remainder of the tribe, similar to the entire world-wide Jewish population and different from non-Jewish populations tested. It seems that the claims of the Lemba tribe are quite reasonable and they are now learning Hebrew and studying to become Rabbis. (photo: www.mindspring.com/~jaypsand/lemba3.jpg)

Plant and Animal DNA typing

DNA is the common thread binding all life forms together. It is as important to plants, animals, bacteria and viruses as it is to humans. It controls all of their biochemical processes down to the simplest molecular level. The fundamental nature of DNA in all life forms, therefore, can be very useful in providing forensically valuable information. Techniques have now been developed, and upheld in court, for obtaining DNA profile data that has proved useful in the identification of non-human subjects. Plant evidence has been used to place suspects at a crime scene, tie a drug dealer to his crop, and detect crop smugglers. Viral and bacterial DNA profiles have been equally successful in catching an attacker who murdered with the HIV (AIDS) virus and to trace soil samples to their local origins through the bacteria it contained. And animal DNA evidence has been used to place a suspect with a victim through pet hairs and animal-based fabric fibers (such as wool, mink, and fox).

Non-human DNA profile data have been used not only in criminal trials but also in civil proceedings. For example, plant breeders often spend a great deal of time and money developing new and improved strains of crops. It has been particularly difficult, however, for these breeders to protect the new strains that they have developed. These issues can be readily resolved using DNA plant profiling techniques by "registering" the DNA profile of a new breed.

5.5. DNA Databanks: CODIS and Beyond

CODIS (Combined DNA Index System)

Many different STR regions are available for the type of analysis described in the previous section. In order to standardize the choices for comparisons, in 1998 the FBI launched a system called CODIS or the Combined DNA Index System. Similar systems are also used around the world.

In the CODIS system, the FBI has chosen a standard set of 13 STR loci that are all "4-letter" sequences. In addition, one locus on the X and Y chromosomes was also chosen to determine gender (the gene for tooth enamel - amelogenin gene). The choice of which of the many STR possibilities to include in the data set was carefully made and arranged such that the STR sequences were as discriminating between people as possible. The locations of these STR sequences on our chromosomes are shown in Figure 5.5.1. As of 2005, over 2.5 million individual DNA CODIS profiles are on record at the FBI, making it the second largest DNA databank in the world behind the National DNA Database of the UK which contains over 3.4 million records.

CODIS contains several databases (called indices) including convicted offender, forensic (samples collected from crime scenes), missing persons and their relatives. There is an increasing push to have more and more people's DNA profiles placed in the CODIS system. For example, military personnel, convicted felons, and many law enforcement offices are currently required by some states to be DNA-profiled. Legislation is also pending in many states to have the DNA profile for anyone arrested to be included in this database. Many states also currently maintain their own DNA databanks that may be used in conjunction with the CODIS system.

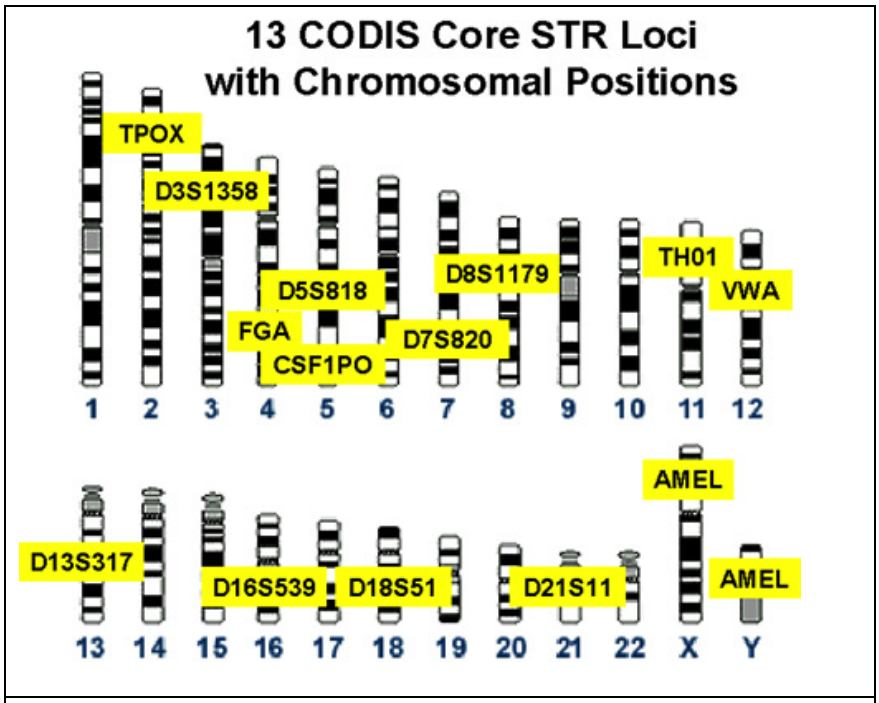


Figure 5.5.1. The chromosomal positions of the 13 STR regions (loci) used by the FBI for the CODIS databank (from www.fbr.org/swksweb/dnajustice.html).

In these computer-based DNA databank systems, the number of repeats of the chosen DNA STRs are easily converted into sets of numbers. A computer-searchable code listing the specific STR followed by the number of repeats found at that location is easily prepared. For example, the code "FGA 21,22" would indicate that someone is heterozygous at the FGA locus with two alleles present - one with 21-repeats and one with 22-repeats. This process is then completed for all 13 STR sites and the amelogenin site on the X or Y chromosome to form the complete CODIS profile. New leads in tens of thousands of "cold cases", often leading to convictions, have now arisen from the use of these powerful DNA databanks. In addition, DNA databases have provided clear links between cases in which no obvious connection had been made before.

Royal Russian Family: Czar Nicholas Romanov

In 1917, the reign of the Russian Czars came to an abrupt end when the Bolsheviks dethroned and imprisoned Czar Nicholas II and his family. In July of 1918, the family was believed to have been unceremoniously murdered and their bodies buried in a shallow swamp outside Yekaterinburg.

In 1991, the bodies believed to be that of Nicholas, Alexandra, and three of their children (along with the royal physician and three servants) were believed found and exhumed. DNA, especially mtDNA, bone samples, and other information was used in 1998 to identify the remains. Samples were compared with known origin samples, including a sample from Prince Philip of England, maternally related to the Czarina, along with samples from maternal relatives of the Czar. Curiously, the remains of the heir, Alexei, and daughter Anastasia, were determined not to be among the remains. In 1998, the royal remains were buried in St. Petersburgh and the Czar and Czarina were canonized in 2000.

Recently, questions have been raised by Alex Knight and Joanna Mountain of Stanford about the validity of the DNA testing in 1994. Their concerns arise from forensic inconsistencies and recent DNA testing of known relatives of the Romanovs that are contradictory to the earlier DNA analysis.

CODIS does not provide any additional information about the sample other than a sample identifier and the coded STR information. Concerns have arisen, however, surrounding the ethical, legal, and social implications of an ever-increasing DNA database. Since DNA dictates our biochemical makeup, DNA information may provide sensitive information about an individual such as susceptibility to disease, behavioral disorders, lineage, and abilities. Additionally, when and how samples are stored or destroyed is an area of legal and ethical debate.

Project Innocence

"Project Innocence" was established in 1992 by lawyers Barry Scheck and Peter Neufeld in an attempt to use DNA evidence to exonerate wrongly convicted suspects. The Project handles cases where post-conviction DNA analysis could possibly lead to a "conclusive proof of innocence". According to the Innocence Project, over 290 people in 32 states have been exonerated through DNA testing through by 2012.

Summary

DNA techniques are among the most powerful tools in modern forensic science. Careful analysis of biological samples can both exonerate and connect suspects with crimes. Recent explorations of these techniques have vastly expanded the utility of these techniques to include essentially all living organisms.

The Case of the Palo Verde Tree

Yellow palo verde tree.
(www.sciencephoto.com; B601/595)

On May 3rd, 1992, the strangled body of Denise Johnson was found underneath a Palo Verde tree (see picture at left) outside Phoenix Arizona. An eyewitness told police that they had seen a white pickup truck speeding away from the area at about the time of the crime. Near the body was found a pager that led investigators to Mark Bogen. Bogen admitted to picking Denise up as a hitchhiker and that, after an argument, he had kicked her out of the car. He also claimed that she had stolen several of his things including his wallet and pager.

Police searched Bogen's car and found, among other items, two seedpods from a Palo Verde tree. At the crime scene, several Palo Verde trees were identified, including one that appeared to have been hit by a vehicle. Police contacted researchers at the University of Arizona to see if it was possible to trace the DNA of the seedpods. A DNA profile was run on the seedpods found in the suspect's truck and on samples obtained from the damaged tree found at the crime scene: they matched. This clearly placed Bogen's vehicle at the crime scene.

The problem was that plant DNA evidence had never been used in court before. In order for this new DNA evidence to be admissible it had to be proved to be statistically meaningful. In other words, can individual plants be traced by their DNA? After a study of other Palo Verde trees, it was easily established that indeed each Palo Verde tree is statistically unique in its DNA profile. Bogen was then convicted of first-degree murder, a verdict upheld by the US Court of Appeals.

INTERESTING DNA CASES FOR FURTHER STUDY:

Snowball the Cat - In 1994, Shirley Duguay was found murdered on Prince Edward Island in Canada. The prime suspect was her estranged husband, Douglas Beamish, who was living nearby in his parent's home. The evidence collected included a leather jacket that was stained with Shirley's blood and contained 27 strands of white hair - cat's hair. The Mounties found a white cat named Snowball that lived in Beamish's parents home. A DNA study connected hair from Snowball with the hair fibers found on the victim.

Green River Killer - In the early 1980s, the Green River Killer was believe to have murdered as many as fifty women near Seattle, Washington. At the time of the murders, DNA technology was not sufficient to provide information in the case. In 2001, Gary Ridgeway was arrested as he was leaving the factory where he worked. DNA evidence had been used to definitively link him to four of the murders. Two years later he pleaded guilty to 48 murders, although estimates place the number of his murders higher yet.

Kenniwick Man - The remains of an ancient man found in the Pacific Northwest had raised a variety of scientific and cultural issues. The remains predated the arrival of European settlers but DNA testing has shown that he is significantly different than Native American populations. The question revolves around who are his decedents and who should control his remains?

Argentina's Children - During a repressive military regime in Argentina in the 1970's, many people were kidnapped and murdered. After the fall of this regime, stories surfaced of children either born in prison or kidnapped from their parents and then given to childless military families. Several years later, when children started showing up to register for school with "suspicious" documents, the "grandmothers of the Placo de Mayo" began to ask difficult questions. mtDNA testing has been used to trace the families of some of these "missing" children and over fifty have been restored to their biological families.

King Louis XVII - After the death of King Louis XVI and Marie Antoinette in 1793, their young son Louis-Charles de France remained in prison until his reported death in 1795. Recently, royalists have speculated that the "Lost Child King" managed to escape from the Revolution. In 1999, a DNA sample from the presumed heart of Louis XVII was removed and an mtDNA profile obtained. This profile was then compared with DNA samples from other known members of the royal family, both living and dead, including a sample from a lock of Marie Antoinette's hair. These analyses confirmed the heart to belong to a person of royal decent.

Additional Cases:

Jefferson/Hemings Paternity
NFL Souvenir Footballs (animal DNA verification)
Jack the Ripper (partial/degraded DNA)
Josef Mengele (DNA confirmation of identity))
Jon Bennet Ramsay, cleared John Mark Karr (exonerated confession)
Tomb of the Unknowns (Vietnam MIA identified in DC tomb)
World Trade Center (remains identification)

Chapter 4 References and Bibliography

Howard C. Coleman and Eric D. Swenson "*DNA in the Courtroom A Trial Watcher's Guide*." **1994**, GeneLex Corporation, Seattle, Washington [ISBN 09644507-0-4].

Bruce Budowle, Tamyra Moretti, Jenifer Smith, and Joseph DiZinno "*DNA Typing Protocols: Molecular Biology and Forensic Analysis*" **2000**, Eaton Publishing, [ISBN 1-881299-23-6].

John Butler "*Forensic DNA Typing: Biology and Technology behind STR Markers*" **2001**, Academic Press [ISBN: 0121479528].

John M. Butler; Yin Shen; and Bruce R. McCord,"The Development of Reduced Size STR Amplicons as Tools for Analysis of Degraded DNA" *J Forensic Sci*. **2003**, *48*, 1054.

David A. Micklos and Greg A. Freyer "*DNA Science: A First Course*" **2003**, Cold Spring Harbor Laboratory Press [ISBN: 0879696362].

Norah Rudin and Keith Inman "An Introduction to Forensic DNA Analysis" **2002**, CRC Press, Boca Raton, Fl. [ISBN: 0849302331].

National Research Council "DNA Technology in Forensic Science" **1992**, NRC Committee on DNA Technology in Forensic Science [ISBN: 0309045878]

GLOSSARY OF TERMS

Allele - one of several possible variants of the genetic code at a specific location along the DNA molecule.

Annealing - To pair nucleotides in complimentary DNA strands together by forming hydrogen bonds.

Base Pairing - when two nitrogen bases in complimentary strands connect to form a hydrogen bond, holding the two DNA strands together. Adenine (A) pairs only with Thymine (T) and Guanine (G) pairs only with Cytosine (C) in DNA.

Chromosome - The strand of DNA and its associated proteins found in the nucleus of a cell that carries its genetic information.

CODIS - Combined DNA Identification System, developed by the FBI to store DNA information in a searchable computer database.

Complimentary DNA - a strand of DNA in which the sequence of bases matches those of another strand of DNA according to pairing rules. Thus, if the original strand contains an "A" nucleotide at a particular location, the complimentary strand would have a "T" nucleotide at the equivalent location.

Denaturation - The process of separating the two strands of DNA into individual strands.

DNA - deoxyribonucleic acid; the basic genetic molecule of living organisms composed of repeating three groups (nucleotides): a phosphate, a deoxyribose sugar, and a nitrogen base.

Enzyme - a protein that catalyzes a chemical reaction; affects the rate of a chemical reaction without being consumed by the reaction.

Extension - Completing a complimentary DNA strand from a template strand.

Gel electrophoresis - the process of separating DNA fragments based upon their charge and size. Charged DNA fragments are placed in a gel bed and moved through the medium by applying an electric current. Shorter fragments move fastest and farthest through the gel while longer fragments move the least.

Gene - Region along the DNA strand that provides coding information for the synthesis of proteins.

Hybridization - the process of joining together complimentary strands of DNA through base-pairing.

Hypervariable DNA - (also called "junk", "non-coding" or "nonsense" DNA) a region of DNA that shows a great degree of variability in a population and do not code for any known protein.

Locus (loci) - a specific position on the DNA strand.

Mitochondrial DNA - mtDNA; a small circle of DNA that resides outside the nucleus in the cellular mitochondria. mtDNA is only inherited from mothers and passed to all offspring.

Mutation - an inheritable change in the base sequence of DNA.

Nucleotide - the basic building block of DNA, consisting of a phosphate, a deoxyribose sugar, and a nitrogen base.

PCR - polymerase chain reaction; laboratory method for the rapid replication of regions of DNA.

Population - a specific group of people defined by geography, race, of other defining features.

Restriction Enzyme - a protein that locates a specific DNA sequence and cuts the DNA strand at that location.

RFLP - restriction fragment length polymorphism.

STR - short tandem repeat; a region of DNA in which a small sequence of nucleotides (2 to 6 nucleotides) is repeated multiple times.

VNTR - Variable number tandem repeat; a region of DNA in which a larger sequence of nucleotides (up to fifty nucleotides) is repeated multiple times.

Y Chromosome - the chromosome inherited only through the father and passed along only to male offspring.

QUESTIONS FOR FURTHER PRACTICE AND MASTERY

5.1. Regions of DNA that do not code for any known protein are called _____.
5.2. The genetic information within our DNA is used to create a(n) _____ molecule in the nucleus which then moves to the ribosome for protein synthesis.
5.3. A permanent alteration or change in the sequence of nucleotides in the DNA of an organism is called a(n) _____.
5.4. In DNA, the base adenine always pairs with _____ while guanine pairs with _____.
5.5. Transcription is process that occurs within the nucleus and forms a _____ molecule from a _____ template
5.6. DNA with different genetic information at the same locus are called
5.7. The sequence of DNA nucleotides that provides the genetic information corresponding to a single characteristic is a _____.
5.8. In the RFLP method of DNA fingerprinting, fragments of DNA created by the restriction enzyme can be separated according to their size by the technique called _____.
5.9. The PCR technique is used for _____.
5.10. A forensic scientist was conducting an genetic investigation and identified a satellite (STR sequence) on a chromosome and constructed the primer CTTGCTTGGAAAATTCCGTGGAC. What is the corresponding template DNA sequence? _____
5.11. What are the limitations in the use of forensic DNA profiling? _____
5.12. A nucleotide is composed of three units:_____, _____, and a _____ unit.
5.13. Given the data below in Table 1, what is the probability that someone from China would have be homozygous for 7 repeats of the D13S317 STR fragment (have the same repeat number in both of their chromosomes)?_____(just the setup would be fine).

Table 1. World distribution of the D13S317 STR fragments
(from .uni-duesseldorf.de/WWW/MedFak/Serology/DNA-Systeme/d13s317.htm)

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

5.14. Why are probability calculations so important in forensic applications of DNA fingerprinting.
5.15. Provide a short definition for each of the following: (a) STR, (b) RFLP Analysis, (c) Mitochondrial DNA.
5.16. The process of copying a portion of nuclear DNA to a mRNA molecule is referred to as _ _____.

5.17. How do the VNTR alleles (variable number tandem repeats) in the hypervariable ("junk") regions of human DNA used for forensic investigations often differ from each other?

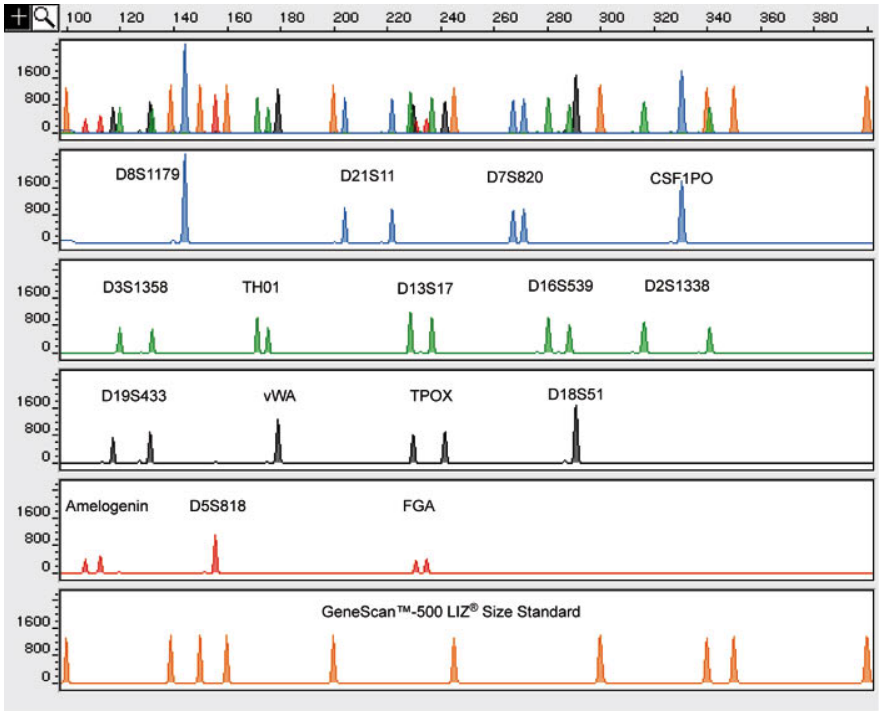
5.18. The basic building block in DNA, consisting of a phosphate, a sugar, and a base, is called a _____.

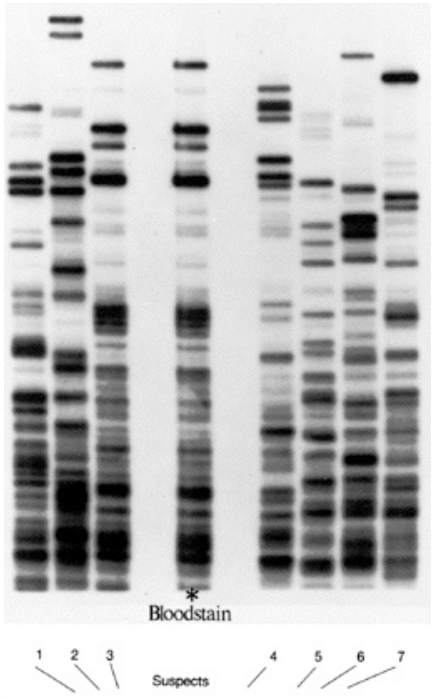
5.19. In PCR, describe each of the following steps: (a) denaturation, (b) annealing, (c) extension, (d) replication, and (e) polymerization.

5.20. DNA is a(n), (a) protein, (b) starch, (c) nucleic acid, (d) enzyme, (e) sugar? _____

5.21. Describe the use of luminol at crime scenes. _____

5.22. Below is shown a multiplex for 15 STR loci and the Amelogenin locus (sex locus). From this information, does the profile come from a male or female? _____.





5.23. Find three loci where the person is homozygous for that allele (use the label shown on the multiplex to indicate the allele below).

1._____
2._____
3._____

5.24. How many different nitrogen bases are used in the makeup of a DNA molecule? _____

5.25. Given the forensic DNA pattern for a sample bloodstain found at a crime scene shown at right, which suspect cannot be eliminated from further consideration and why? _____

_____

5.26. Describe the pattern of inheritance of mitochondrial DNA. _____

5.27 Why is STR analysis has replacing RFLP DNA typing? _____

5.28. A permanent alteration or change in the sequence of nucleotides in the DNA of an organism is called a(n)

a) transmigration
b) transgenic shift
c) mutation
d) transmutation

e) modification

5.29. A forensic scientist was conducting a genetic investigation and identified a satellite (STR sequence) on a chromosome and constructed the primer CTTGCAA. What is the corresponding template DNA sequence?

a) CTTGCAA
b) GAACGTT
c) TCCATGG
d) TCCACAA
e) none of the above

5.30. Write below, side by side, a BRIEF description comparing the differences between nuclear DNA typing and mitochondrial DNA typing (just simply compare the differences between these two techniques).

Nuclear DNA typing	**Mitochondrial DNA typing**

5.31. DNA is a(n)?

(a) starch
(b) enzyme
(c) carbohydrate
(d) nucleic acid
(e) protein

5.32. Use the DNA RFLP pattern of evidence below from a rape investigation to answer the following question.

Blood of Victim	Semen recovered From victim	Blood from Suspect A	Blood from Suspect B	Blood from Suspect C
_____		_____		
	_____			
		_____		
			_____	
				_____
_____				
	_____			
				_____

From the RFLP data above, which suspect(s) cannot be eliminated from consideration?

(a) suspect A
(b) suspect B
(c) suspect C
(d) ALL can be eliminated
(e) NONE can be eliminated

5.33. The method employed for separating out the different lengths of DNA from either a RFLP or PCR/STR most commonly is:

(a) GC/mass spectrometry
(b) thin layer chromatography
(c) infrared spectroscopy
(d) gel electrophoresis
(e) none of the above

5.34. In the PCR process, the first step is to heat the DNA strands. This is to permit the:

a) DNA to coil very tightly in the helical shape.
b) process to take place without DNA degradation.
c) hybridization of DNA to take place.
d) the strands of the DNA double helix to separate.
e) replication of the DNA to take place.

5.35. What are the three components of a nucleotide?
5.36. What are the four nitrogen bases in DNA?
5.37. What are the base pairings for the hydrogen bonds that join the two DNA strands?
5.38. What is a gene?
5.39. About what percent of human DNA "codes"?
5.40. What is VNTR? What is STR?
5.41. Why aren't the coding regions of DNA used to forensically identify an individual?
5.42. What is meant by polymorphism?
5.43. What are the two primary methods of DNA typing?
5.44. List the advantages and disadvantages of RFLP and PCR testing methods.
5.45. What does a restriction enzyme do?
5.46. What is gel electrophoresis?
5.47. What is amplification in the PCR process?
5.48. What is capillary electrophoresis?
5.49. How does the "touch" DNA process differ from other DNA methods?
5.50. How is mitochondrial DNA different from nuclear DNA?
5.51. When would mtDNA analysis be a better choice than nuclear DNA analysis?
5.52. In what types of cases would Y chromosome typing be useful?
5.53. What is CODIS?

EXTENSIVE QUESTIONS

5.54. Explain what "coding" and "noncoding" means with respect to sequencing on genes.
5.55. Explain the translation and transcription processes.
5.56. Explain how gel electrophoresis works?
5.57. Explain the steps in the PCR process.

5.58. Explain how STR repeat patterns in a sample can be used to determine the probability that the sample would match another sample in the population.
5.59. Compare the RFLP and STR DNA typing with respect to sequence size, reliability and speed.