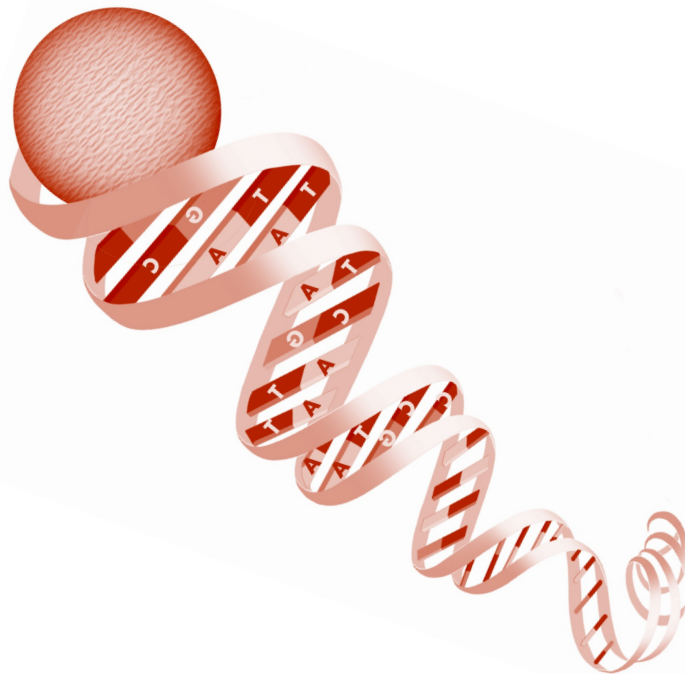




CHROMOSOMAL LABORATORIES, INC.

Introduction to DNA



Setting the Standard for Quality DNA Identification

An Overview of DNA

A single cell is all it takes to create a human being. With the exception of red blood cells, each cell in our body contains a nucleus which holds our genetic blueprints known as DNA (deoxyribonucleic acid). The primary purpose of DNA is to make copies of itself, so that it can turn that single cell into trillions of cells, forming a human being. The genetic information within DNA is passed on through generations with half of the information coming from the mother's side and half from the father's side.

DNA is comprised of 3 key elements; nucleobases (bases), sugar and phosphate. There are 4 bases or nucleotides: A (adenine), T (thymine), C (cytosine) and G (guanine). Each base will attach to a sugar molecule that is attached to a phosphate molecule. The sugar and phosphate form the backbone of DNA while the various combinations of bases attached to sugar are what provide the biological diversity between all living beings.

Each base has a complimentary base which it can pair up with. The base pair rules are such that adenine pairs with thymine and cytosine with guanine. Each base pairs with another base through the use of hydrogen bonds. Two hydrogen bonds comprise the A-T bond while three hydrogen bonds are required for the C-G bond. Due to the increased number of hydrogen bonds between cytosine and guanine, it is more difficult to break apart than the adenine-thymine base pair. This base pairing allows DNA to be composed of two complimentary strands linked together, also known as hybridization.

A DNA sequence has two ends, known as 5' and 3'. The 5' and 3' refers to the position of the carbon atoms in the sugar ring of DNA. The two strands of DNA line up in an anti-parallel fashion, such that one strand is in the 5' to 3' direction while the other is in the 3' to 5' direction. The hybridization of these two strands forms a double-helix structure, which was discovered in 1953, by James Watson and Francis Crick. The pairing of DNA is such that if the DNA sequence of one strand is known, it is easy to determine the sequence of the complimentary strand, due to the base pairing rules of A-T, C-G.

A process known as denaturation can break the double helix structure, separating it back into

two single strands of DNA. Denaturation breaks the hydrogen bonds formed between A-T and C-G through the use of elevated temperatures (near boiling) or by treatment with chemicals such as urea or formamide. The denaturation process is reversible with the use of cooler temperatures. A DNA strand that has been separated by heat will begin to bond back to its complimentary strand as it cools, a process known as reannealing. The denaturation and reannealing processes are key factors in the ability to analyze DNA.

Chromosomes

DNA is formed into dense packets along with proteins called histones which make up chromosomes. Normal human cells are comprised of 23 pairs or 46 chromosomes. Of those 23, 22 are matched pairs of autosomal chromosomes and two are single chromosomes that determine sex. The 22 matched pairs are considered homologous because they are the same size and contain the same genetic structure; one copy is inherited from the mother's side and one from the father's. The two single chromosomes that determine sex are called the "X" and "Y" chromosome. The X chromosome is female and the Y is male, therefore an individual with a sex chromosome pair of XX would be female and XY would be male. Human identity testing is done using the autosomal chromosomes while gender determination is done using the sex chromosomes.

Chromosomes are composed of coding and non-coding regions of DNA. The coding regions, which consist of ~5% of our DNA, are referred to as genes. Genes are made up of exons (protein-coding portions) and introns (intervening sequences); they contain the information needed for a cell to produce proteins. The non-coding regions make up the remaining ~95% and are often referred to as 'junk' DNA because they are not directly related to producing proteins.

DNA Markers

DNA markers are found within the non-coding regions of DNA either between genes or within genes (introns). A locus (plural: loci) is a defined position on a chromosome indicating the location of a gene or DNA marker within the non-coding region. Loci are used to map particular regions of DNA and therefore provide a standard

nomenclature for DNA markers. Each chromosome of a homologous pair, containing a copy of each gene at the same position or locus on the chromosome. However, mutations may occur over time causing the DNA sequences to no longer be identical.

Each locus contains an alternative possibility for a gene which is called an allele. If the alleles at a particular locus on each chromosome are identical, they are referred to as homozygous; if they are different they are referred to as heterozygous. Alleles present at each locus are characterized by genotyping. If a locus has 2 alleles, A and a, then 3 possible genotypes result: AA, Aa and aa. Genotypes AA and aa are considered homozygous because they have identical alleles while, genotype Aa is heterozygous because each allele is different. The genotypes obtained at multiple loci are used to profile an individual's DNA. Because random matches can occur between unrelated individuals, human identity testing or DNA profiling utilizes genotypes at multiple loci in an effort to reduce this possibility.

DNA Marker Nomenclature

A standard nomenclature is used to designate the location of a DNA marker depending on where it falls. If the marker is part of a gene or falls within it, the gene name is used, whereas if the marker falls outside the gene region, the chromosomal position is used. The following are a few examples.

DNA marker TH01 which falls within a gene

TH: Tyrosine Hydroxylase gene
01: Found within intron 1.

Markers D5S818 and DYS19 which fall outside of gene regions

D: Stands for DNA
5: Chromosome number
S: Indicates DNA marker is a single copy sequence
818: Indicates the order in which the marker was discovered; 818th locus
D: Stands for DNA
Y: Y Chromosome
S: Indicates DNA marker is a single copy sequence
19: 19th locus described

Physical Chromosome Location Nomenclature

There are 4 basic regions of a chromosome. The centromere, a short arm, a long arm and telomeres. The centromere is the center region of the chromosome. On either side of the centromere is an arm, one being shorter than the other. The short arm is referred to as 'p' and the longer arm as 'q'. Each arm ends with a telomere. The size of a chromosome determines its number (1-22, largest to smallest).

Before cells divide, the chromosomes change from an unraveled linear form to that of a more compact form. At this point in time the chromosomes can be stained with a dye and observed under a microscope. The chromosomes will appear as a series of light and dark bands dependant upon the different number of A-T, C-G bonds.

The 24 chromosomes (22 autosomal and X and Y) are distinguished from one another by their size (i.e. chromosome number) and banding patterns. This type of analysis is referred to as a karyotype. Bands are classified according to their location (short or long arm) and number. The bands are numbered from the centromere to the telomere, so the larger the number, the closer the band is to the end of the arm. The following is an example of the nomenclature used:
Chromosome location 12p1

12: Chromosome #12
p: Short arm p
1: Band #1

DNA markers that are close to the end of an arm, often use the suffix 'ter' (in place of the band #) indicating the location is close to the terminal end of the chromosome. The following is an example of the nomenclature used:

Chromosome location 15qter

15: Chromosome #15
q: Long arm q
ter: Terminus of the long arm

A chromosome location may be listed as a particular range if the DNA marker is not yet mapped with a high degree of accuracy. The following is an example of the nomenclature used:

Chromosome location 2p23-pter

2: Chromosome #2
 p23-pter: Somewhere on the short arm p
 between band #23 and the terminus

The History of Forensic DNA Analysis

In 1985, an English geneticist named Alec Jeffreys developed a technique known as the RFLP (restriction fragment length polymorphism) method. This method led to the ability to perform human identity testing, also known as DNA typing (profiling), which has revolutionized forensic science and has become the most important investigative tool since the development of fingerprinting and blood group determination.

During his research, Dr. Jeffreys found that particular regions of DNA consisted of several short sequences of base pairs that repeated over and over again. These repeated sequences became known as VNTR's (variable number of tandem repeats). While VNTR's occur in everyone's DNA, it is the number of times they repeat that differ from individual to individual. In order to examine the VNTR region of the DNA, Jeffreys used a restriction enzyme (RFLP) that cut through the DNA on either side of the VNTR, allowing analysis of that specific region of DNA.

DNA technology has come a long way since Jeffreys' discovery 20 years ago. Currently, 2 primary techniques exist that can be used for DNA typing, the RFLP method and the polymerase chain reaction (PCR) method. PCR can be performed in as little as a few hours. This is quite remarkable compared to the 6-8 weeks required by the RFLP method. In addition to more rapid analysis, the sensitivity of the tests has also advanced dramatically. In the past, large blood stains with well-preserved DNA were required to perform DNA profiling with RFLP. These days, using PCR, a useful DNA profile can be determined using very minute samples of blood, and in some cases a single cell can be utilized.

DNA Sample Collection, Extraction and Quantification

DNA samples can be collected from various biological materials, with blood and semen being the most common materials tested in forensic laboratories. Suspects, victims and witnesses can

be linked to crime scenes based on the DNA they transfer to each other or to an object(s). In order to produce DNA profiles that are meaningful and legally accepted in court, the collection of DNA from the crime scene must be performed with utmost care. If the samples are collected, preserved, stored or transported incorrectly prior to analytical testing, the resulting profile will be ambiguous. In order to reduce the rate of bacteriological growth and degradation of the DNA, biological evidence should be stored dry and cold. For more information regarding DNA evidence collection, a brochure entitled "What Every Law Enforcement Officer Should Know About DNA Evidence" is available through the National Institute of Justice.

DNA is degraded by nucleases (enzymes or proteins found in the cell) to recycle the nucleotide components for other uses. In order to work properly, nucleases require magnesium. Therefore, blood is stored at -20C or -80C (for long-term storage) in purple-top tubes containing EDTA, a blood preservative that chelates or binds up all of the free magnesium, thus preventing the nucleases from digesting the DNA.

When a biological sample is collected from a crime scene, more than just DNA is collected; proteins and other cellular materials are also present. Before DNA can be examined in the laboratory it must be isolated from the other cellular material, otherwise it will inhibit the ability to analyze the DNA. While many methods exist for DNA extraction, DNA laboratories traditionally rely upon 3 primary techniques based upon the type of biological material they are working with.

The technique which has been in use for the longest period of time is organic extraction; also known as phenol chloroform extraction. It can be used for either RFLP or PCR typing, with RFLP high molecular weight DNA extracted most effectively. Organic extraction requires the use of several chemicals: sodium dodecyl sulfate (SDS) and proteinase K to break open the cell walls and break down the proteins protecting the DNA and a phenol/chloroform mixture to separate the proteins from the DNA.

A more rapid DNA extraction technique is the Chelex method. Chelex differs from organic extraction in that it has fewer steps, which lends itself to the rapid extraction and also fewer opportunities for contamination between samples. However, it is only useful for PCR-based testing because it produces single-stranded DNA. Chelex is an ion-exchange

resin that binds magnesium, thus protecting DNA. Samples are added to a 5% suspension of Chelex and boiled for several minutes to denature the DNA, disrupt cell membranes, and destroy cell proteins. The sample is then spun in a centrifuge to pull the Chelex to the bottom so that the supernatant can be removed and added to the PCR amplification reaction. One of the PCR advantages of the Chelex extraction method is its ability to remove PCR inhibitors; however the addition of too much blood to the solution can result in some PCR inhibition.

The newest approach to DNA extraction involves the use of FTA paper, an absorbent cellulose-based paper that contains 4 chemical substances. The chemicals in the paper have dual functions, to allow a method for storage of DNA by protecting it from degradation by nucleases and to preserve the paper from bacterial growth. The resulting product is a device that allows storage of DNA at room temperature over several years. To use FTA paper, simply add a spot of blood to the paper and allow it to dry. Upon contact with the paper, the cells are broken open and the DNA is immobilized with the matrix of the paper. A small piece of the paper is removed using a puncher and added to a tube. The paper containing the DNA is then isolated by "washing" it with a solvent that removes inhibitors of the PCR reaction. The newly cleaned paper punch is then added directly to the PCR reaction.

The quality and quantity of DNA present in a sample can be determined once it has been isolated. Determining the quantity of DNA present is essential to PCR-based assays because a narrow concentration range is ideal. Several methods have been developed to quantify DNA, including the slot blot procedure and fluorescence-based microtiter plate assays.

Polymerase Chain Reaction (PCR)

The PCR process is similar to that of a Xerox machine in that it makes copies of specific regions of DNA. The copies are made through a series of ~30 heating and cooling cycles, after which approximately a billion copies of specific DNA sequences are generated. This process is known as amplification and the resulting product is referred to as an 'amplicon'.

To perform the amplification process, a PCR reaction is prepared by mixing several reagents with deionized water until the desired concentration and

volume are achieved. The 2 most important reagents in the PCR reaction are the 2 primers, which are short, ready-made fragments of DNA that will attach to either side of the DNA region to be copied. Different primers will attach to different repeating sequences of base pairs; therefore it is necessary to have some knowledge of the DNA sequence to be copied in order to select the correct primers. Due to the near boiling temperatures used to denature the DNA, thermal stable polymerases must be used so that they do not fall apart. The bacterium *Thermus aquaticus*, which inhabits hot springs, is used to produce *Taq*, the most commonly used thermal stable polymerase.

Quality Control

Quality control procedures are an integral component of the PCR process. Positive and negative controls are used to monitor both the conditions of the PCR run and the technique of the laboratory technician. A negative control is the entire PCR reaction with water or buffer in place of the DNA template, and is used to assess whether the sample was contaminated by extraneous DNA (i.e. from someone in the lab). A positive control is created by using the same PCR primers used on the rest of the samples and therefore verifies that the reaction components and thermal cycling are correctly amplifying a specific region of DNA.

Thermal Cyclers

Thermal cyclers are the instruments used to heat and cool the DNA sample as necessary to perform PCR. Cycling parameters will differ depending on the primer sequence as it is related to various PCR amplification kits. DNA polymerases can exhibit activity below their optimal temperature which can cause primers to bind to each other (called primer dimers) or to anneal non-specifically to the DNA template. Primer dimers are relatively small compared to the PCR products and as a result will be preferentially amplified, leaving the target DNA region to be amplified less efficiently. The resulting lower concentration of target DNA can affect the ability to run the tests. By waiting to introduce the polymerase until after the temperature has been raised above the annealing temperature, the possibility of mispriming can be minimized, a process known as 'hot start' PCR. However hot start PCR does have its disadvantages; it is time consuming, cumbersome and increases the chances of cross-contamination between samples because the tubes must be opened at the thermal cycler to introduce the polymerase.

Another option is the use of Amplitaq Gold; a DNA polymerase that requires heat to become activated and therefore allows for a closed-tube hot start PCR. Because the polymerase is activated just when it is needed, primer dimers and mispriming are minimized.

PCR Primers

PCR yield is directly affected by the annealing characteristics of the primers therefore well designed primers are key to a PCR reaction. Optimal PCR primer design characteristics include: both primers must be specific to the target region, possess similar annealing temperatures, be structurally compatible, and not interact significantly with each other or themselves such that they form primer dimers.

Multiplex PCR

Two or more regions of DNA can be simultaneously amplified by adding more than one primer set to the reaction. With the addition of each new primer the complexity of possible primer interactions increases exponentially, therefore the primer pairs need to be compatible. Compatibility means the primer annealing temperatures should be similar and excessive regions of contiguous bases should be avoided, otherwise the primers will bind to one another instead of the template DNA.

STR Markers

The most popular markers used for DNA typing are short tandem repeat (STR) markers. There are several advantages to using STR markers, these include: the ability to use them in conjunction with PCR, they work well with low quantities or highly degraded DNA, and finally they are highly discriminating between unrelated or closely related individuals. Interpretation of results is also easier with the use of STR markers because computerized DNA databases can be utilized to compare discrete alleles, as opposed to the RFLP-based systems where DNA is grouped simply by size.

Through the commercialization of the STR loci characterized and developed by the Bayer College of Medicine and the Forensic Science Service in England there are now 13 standard STR markers available for DNA analysis. These markers make up the national DNA database known as CODIS (Combined DNA Index System). The 13 CODIS loci are: CSF1PO, FGA, TH01, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539,

D18S51 and D21S11. When all 13 loci are tested, the probability of a random match in unrelated individuals is less than one in a trillion.

Promega Corporation and PE Applied Biosystems are the two primary vendors of STR kits used by the forensic community. In order to obtain information from all 13 STR's, two PCR reactions must be performed by using either Promega's PowerPlex 1.1 and PowerPlex 2.1 or Applied Biosystem's Profiler Plus and COfiler kits. Commercially available kits are more commonly used than private ones for several reasons. STR kits utilize multiplex PCR conditions therefore extensive research must be done to ensure that all the primer pairs are compatible. Most private laboratories do not have the time, resources or funds to decide which markers would be included, design the primers and optimize the PCR multiplexes. The vast use of commercially available kits increases the ability to share data between laboratories because everyone is using the same STR markers.

The 13 CODIS loci are the standard markers used in the United States and will continue to be used far into the future, however additional markers outside the 13 have been used for DNA typing, and more are being evaluated for use in future kits. DNA technology has advanced dramatically over the past 15 years, identifying criminals and exonerating the innocent. Thousands of forensic DNA and paternity cases are conducted every year in both public and private laboratories and the number will only continue to increase as the technology advances further.

© 2005 Chromosomal Laboratories, Inc. All rights reserved. This guide is provided courtesy of Chromosomal Laboratories, Inc., information is general in nature; to learn more, read peer-reviewed literature on these topics.