DNA

Introduction

Forensic DNA profiling combines scientific disciplines of anatomy, cell and molecular biology, genetics, mathematics, and statistics. This article aims to explain some of the fundamental science underpinning the use of DNA in a forensic context. In particular, it examines the natural attributes of DNA that contribute to its suitability for use in the forensic field and a summary of the practical tools routinely applied to extract the vital information encoded within human DNA.

Fundamental Human Anatomy

The human body is a universe of working parts and functional interactions. We observe the physical manifestations of these interactions all the time, as we walk, talk, breathe, think, or eat. These gross or macroscopic functions of the human body, however, are driven by extremely complex interactions, occurring at the cellular and subcellular level.

Our bodies are made up of trillions of cells. Each cell has a prescribed function relative to its position in the body that is essential to healthy human life. The cells themselves are extremely complex and advanced pieces of biological machinery. A cell is comprised of a cytosol, which is bound by a permeable membrane, and contains a host of miniature organs (organelles) including the nucleus. The nucleus contains DNA – the material that prescribes the cell’s principal functional characteristics.

It may be useful to think of the cell as being like a factory. Membranes enclose the structure and separate different organelles, which can be thought of as departments with specialized functions. The nucleus is the central administration, containing in its DNA a library of information that determines cellular structure and processes. From it instructions are issued for proper regulation of the business of the cell. The mitochondria are the power generators. The cytosol can be thought of as the general work area, where protein machinery (enzymes) carries out the formation of new molecules from imported raw materials. There are special molecular channels in the membranes between compartments and between the cell and its external surroundings. These monitor the flow of molecules in the appropriate directions, similar to personal assistants and receptionists. Like factories, cells tend to specialize in function. For example, many of the cells in higher organisms are largely devoted to the production and export of one or a few molecular products.

Despite the diverse functions of the different types of cells that constitute the human body, each nucleated cell contains an identical copy of a common DNA molecule from which genetic information is read in a linear fashion. Since the amount of information needed to specify the structure and function of a multicellular organism such as a human is immense, the DNA molecule is extremely long. In fact, if the DNA from a single human cell were stretched end to end it would extend approximately 2 m.

Human cells live by a well-defined life cycle, itself, separated into several distinct phases. Similarly, the process of eukaryotic cell division can be divided into principle phases. Cell division can occur by mitosis that produces two diploid cells, or meiosis, which produces four haploid cells. Mitosis results in the production of two daughter cells that are identical to each other and to the parent cell and is the process of somatic cell replication (Figure 1). Our genetic information is carried to the next generation by the haploid cells known as gametes (spermatozoa in the male and the oocyte in the female). These cells are produced by meiosis during gametogenesis (or specifically, spermatogenesis or oogenesis; Figure 2). In either mitosis or meiosis, the DNA (on the chromosomes) must be duplicated and distributed to each daughter cell. An error in DNA replication is called a mutation. If mutation occurs during replication of an adult cell only progeny cells within that person/organism are affected. This is known as somatic mutation. If mutation occurs in the gametes (or germinal cells) there is a possibility that it will be passed onto the offspring of the person/organism. This is referred to as a germline mutation.

This summary of basic cellular biology illustrates the first point as to the applicability of DNA analysis for forensic purposes. A replica DNA molecule exists in all nucleated cells in the human body. This means that cells existing in any individual’s tissues, fluids, or organs carry a copy of the same DNA molecule. This means there are trillions of possible sources of DNA that can be targeted for forensic purposes. It
Parental cell is diploid (resting cell known as gap phase)

Chromosomes are duplicated (in late prophase)

Unpaired chromosomes align at cell equator (during metaphase)

Two daughter cells are diploid, genetically identical to the parent cell

Cytosol divides (during telophase)

Sister chromatids separate (during anaphase)

Figure 1  Diagrammatic representation of mitosis, the process of cell replication for adult or somatic cells

Parental cell is diploid

Chromosomes are duplicated (in late prophase)

Paired homologous chromosomes align at metaphase I then separate at anaphase I

Four daughter cells are haploid, not genetically identical to the parent cell

Sister chromatids separate (during anaphase II)

Figure 2  Diagrammatic representation of meiosis, the process of cell replication for germline cells or gametes
also means that DNA can be cross-compared among various tissue types from the same donor.

The heritable characteristic of DNA is another advantage to its use in the forensic sciences. Familial, and particularly parental, relationships can provide information to the origin of a given DNA profile due to the known presence of inherited characteristics. This application of DNA profiling technology is often used in the identification of human remains or in the investigation of crimes involving disputed paternity.

Fundamental Genetics

A striking attribute of a living cell is its ability to transmit hereditary properties from one cell generation to the next. This power of self-replication is often attributed as the defining difference between the living and the nonliving. Since the beginning of human history, people have wondered how traits are inherited from one generation to the next. Although children tend to resemble one parent more closely than the other, most offspring tend to be a blend of the characteristics of both parents. The idea of a gene, a unit of hereditary information, arose in the mid-nineteenth century from the famous work of an Augustinian monk named Gregor Mendel. Predominantly working in agricultural science, Mendel found that by beginning with “parents” of known genetic background one had a baseline against which the pattern of inheritance could be measured.

The total genetic information carried by a cell, all the DNA in the nucleus, is referred to as the genome. Strictly speaking, this DNA is referred to as nuclear DNA. A small amount of the cells’ total DNA content exists outside the nucleus, in the mitochondria. Mitochondrial DNA (mtDNA) is also widely applied in a forensic context.

The genome of higher species is packaged into bundles known as chromosomes. Most nucleated human cells are diploid, meaning that the overall content of 46 chromosomes actually comprises one pair of each of the 23 chromosomes (two copies of each chromosome). Exceptions include the sex cells or gametes (spermatozoa or ova), which are haploid (a single copy of each chromosome). The human chromosomes are numbered from 1 to 22 according to their size (chromosome 1 being the largest, 2 the second largest, and so on). These 22 chromosomes are referred to as autosomes, as they do not play a role in sex-determination. The 23rd pair is the X and Y chromosomes that play a direct role in sex determination. This chromosomal pair is often referred to as nonautosomes.

During conception, parental haploid gametes fuse and form a progenitor diploid cell that is the origin for embryonic development. Owing to this diploidy, individuals who are DNA profiled show either one or two alleles at each DNA site (or locus). If an individual shows one allele, this implies that there are two copies of the same allele present on each of the chromosomes of that particular pair. This event is referred to as homozygosity, and we refer to the particular individuals’ profile at that DNA site (or locus) as being homozygotic. In such cases, the individual has inherited the same allele from each parent. If an individual shows two alleles at a locus, it implies that there are two different alleles present on each of the chromosomes of that particular pair. This event is referred to as heterozygosity, and we refer to the particular individuals’ profile at that locus as being heterozygotic.

The DNA molecule encodes for the functional activities of the cells in higher organisms with the packages of genetic information on the DNA referred to as genes. The DNA regions containing genes are called coding regions (or exons). Genes have a dedicated role in protein synthesis, with these products being in turn linked to one or a number of biological or biochemical functions. For the most part, these essential biological functions are the same from individual to individual. As such, exonic regions are highly conserved across individuals of the same species – that is they are almost identical genetically. Another reason for DNA sequence conformity in coding regions, or genes themselves, is that alterations to DNA sequences in these areas usually have some deleterious effect on the physical well being of the individual. The health effects reduce the likelihood of survival, mate selection, and reproduction and therefore are directly linked to the propagation of this individual’s particular genotype. In population genetic terms, this link between genotype and ability to survive is referred to as genetic fitness. By this mechanism, a genotype that affects the fitness of an individual will become less prevalent, or even extinct, within a population. This is the notion of selection.

The entire DNA molecule is not made up of genes; in fact, only a little over 5% of the genome codes for
the production of proteins. The remainder of the DNA is made up of noncoding regions, or introns. Despite their prevalence, the prescribed function of noncoding regions is poorly understood. Noncoding regions were initially presumed to be functionless and as a result were commonly, but inappropriately, referred to as *junk DNA*. Recent evidence proposes some functionality for certain noncoding DNA regions [1–3]. Interestingly, large areas of noncoding DNA, many of which are not implicated in regulation, are strongly conserved between species. This may be strong evidence that they too are functional. Although some ambiguity remains as to their exact role, it is foolhardy to assume that noncoding regions are redundant.

However, as noncoding regions appear to have less responsibility when it comes to the messenger characteristics of the DNA, they have not been exposed to the same evolutionary selection pressures as the coding regions. In essence, mutations can occur in the noncoding regions without affecting the genetic fitness of the individual. Over time, this has led to a high degree of polymorphism in the noncoding regions of the human DNA molecule as mutation has been allowed to continue relatively unchecked. Hypervariable loci, such as those used for forensic identification purposes, are common in the intronic regions.

There are three major varieties of polymorphisms that exist on the genome: minisatellites, microsatellites, and single nucleotide polymorphisms (SNPs). Minisatellites (or variable number of tandem repeats (VNTRs)) are large fragments of DNA which are comprised of sequentially aligned homologous units. The individual units are typically between 20 and 100 base pairs (bps) in length and are repeated consecutively up to 100 times. The overall molecular weight (MW) of the locus is therefore determined by the number of times that the units are repeated. This is the polymorphic feature of these loci. Minisatellite loci are highly polymorphic and for that reason are powerful identification markers.

Microsatellites (or short tandem repeats: STRs), as the name suggests, are smaller versions of minisatellites. The same structural conformation exists in principle but, in comparison to minisatellites, the size of the individual units (2–8 bp) and the number of times they repeat (2–20) are reduced. Microsatellites were originally named as such as they were thought to solely be repeats of (cytosine-adenine) CA dinucleotide stretches. Microsatellites have been detected in every studied organism occurring at a higher rate than would be predicted purely on the basis of base composition. Microsatellites can be termed *simple* (an uninterrupted array of homologous repeat motifs), *compound* (repeat motif changes – two or more adjacent simple repeats), or *complex* (array is interrupted or contains several repeats of variable unit length). The overall MW of a microsatellite locus is still determined by the number of times that the units are repeated; however, the number of repeats is typically smaller for STR loci. Both VNTR and STR loci are length-based polymorphisms. This means that polymorphisms can be detected by standard electrophoretic techniques.

SNPs are loci where there is a variance in the individual base that exists at a particular position on the genome. SNPs are classified, and distinguished from single base changes, if the frequency of occurrence of the minor (less frequent) allele exceeds 1%. SNPs are the most common form of polymorphism on the genome, occurring approximately every 1000 bp (in unrelated individuals). SNPs can be bi-, tri-, or tetra-alleleic, meaning that the base that exists at a particular SNP position can vary between two, three, or four possible types. In practical terms, however, biallelic SNP loci are the only form of the polymorphism that are routinely detected and analyzed.

Aside from the polymorphic loci themselves, DNA has an additional level of variability that is introduced through processes such as assortment and recombination. These occur as part of the meiotic division that cells undergo during gametogenesis. Both assortment and recombination have the effect of shuffling genetic material and essentially randomizing the distribution of the diploid genotype into the haploid gametes. In assortment, the order in which chromosomal pairs align at the equator of the cell during metaphase is random. This means that the total number of gametic chromosomal combinations that can be formed from $n$ chromosome pairs is $2^n$ (Figure 3). In humans ($n = 23$), meaning that over eight million combinations are possible among the haploid gametes of any individual.

Recombination, also known as *crossing-over*, involves the physical exchange of genetic material. In recombination, the arms of sister chromatids of a homologous chromosome pair can overlap during prophase I. This contact allows for the physical
Figure 3  An example of the effect of assortment in generating haploid diversity. If there are three pairs of blue chromosomes (I, II, and III) which are duplicated during prophase to give three pink duplicate chromosome pairs. There are four different ways in which these three pairs of pairs can align at the equator of the cell during metaphase (a–d). After the two rounds of meiotic division this produces eight different possible haploid sets (1–8). The number of theoretical possibilities in a diploid organism is equal to $2^n$, where $n$ is the number of chromosomes in the haploid gamete.

exchange of chromosomal segments and the genetic material that they carry (Figure 4). Without recombination the arrangement of alleles on a particular chromosome would remain coupled together. Recombination allows new (and possibly advantageous) combinations to be produced and adds another element of variability to the inheritance of DNA.

Aside from the shuffling effect of assortment and recombination, DNA may be subject to further mutational events during meiotic or mitotic replication. At the outset of cell division the DNA must replicate itself faithfully. Any errors during this replication process will introduce a difference between the parent and progeny cells. Such a mutation occurring during meiosis can lead to a Mendelian inconsistency between the biological parent and their offspring, as the parents’ gametes have mutated to a haplotype that is inconsistent with the parental genotype. These mutations are more commonly observed on the paternal side, i.e., occurring during spermatogenesis, and are an important consideration in parentage or kinship investigations.

Microsatellite loci have a high mutation rate, in comparison with other polymorphic loci such as SNPs. This trend is observed for both autosomal and nonautosomal loci. In fact, microsatellite loci are thought to be located in mutational “hot spots” [4], that is, regions of the genome that are more susceptible to mutation due to their structure or location. A range of mutational models exist that describe the pattern of mutation of a particular locus. The model favored for microsatellite mutation is the model of slipped strand mispairing (SSM) during DNA replication.

Through the basics genetics, we understand that our genetic composition is comprised of equal proportions of DNA from each of our parents. We also understand that there are coding and noncoding segments of the molecule and types of loci that demonstrate considerable variability. We also acknowledge the role that mutation plays in shuffling and altering DNA characteristics, creating additional diversity within populations. The practical suitability of these characteristics is clear; however, it is the physical composition of the DNA molecule itself and its own inherent variation that provides scientists such a suitable template for analysis in a forensic context.

Fundamental Molecular Biology

The realization that DNA is the principle genetic molecule immediately focused attention on its
structure. Likewise, the revelation that the structure of DNA was relatively simple [5] helped scientists to understand that genes had relatively similar three-dimensional structure and that differences between two genes resided in the order and number of the structural building blocks along the molecules and not in the molecules overall shape.

Structurally, DNA comprises two complementary chains of nucleotides twisted about each other in the form of a right-handed double helix. A nucleotide unit consists of a five-carbon sugar (2′-deoxyribose), a phosphate residue, and a nitrogenous base (either a purine or pyrimidine; Figure 5). The linking of the nucleotide units is formed via the phosphate residue through a phosphodiester bond between adjacent sugar groups. This series of bonds forms the sugar–phosphate backbone of the DNA strand and allows for the formation of extremely long polynucleotide chains, containing upward of 3 billion nucleotide units.

The backbone of the DNA is therefore a repetitious structure, which, due to its uniformity, is incapable of encoding information. The importance of the DNA structure is derived from the bases that are attached to each of the sugars of the polynucleotide chain. There are two types of bases which attach
to the sugar group of the nucleotide, the purines (adenine (A) and guanine (G)) and the pyrimidines (cytosine (C) and thymine (T)). In contrast to the regular structure of the sugar phosphate backbone, the order of the purine and pyrimidine bases along the chain is highly irregular. The order of the bases along the polynucleotide chain is referred to as the sequence of the DNA molecule or DNA sequence.

Another vital discovery of the Watson and Crick model is that the two chains of the DNA molecule are made up of complementary pairs of bases, linked together by hydrogen bonds. The pairing of the bases follows simple and strict rules. Adenine is always paired with thymine and guanine is always paired with cytosine. No other pairings are possible due to the physical structure of the bases. The strictness of the pairing rules results in a complementary relationship between the DNA sequences on the two intertwined strands of the double helix. For example, a sequence of 5′-AAGCTG-3′ on one chain, the opposite chain must have the sequence 3′-TTCGAC-5′.

The revelation that the twisted strands of the double helix were always complementary meant that if the strands could be separated, and new DNA synthesized along each, the strict base-pairing relationship would see two identical double-stranded DNA molecules produced. In 1985, Saiki et al. [6] published a paper in which they described for the first time a method to simulate the DNA replication process in a laboratory environment. The technique was called the polymerase chain reaction (PCR). The PCR is essentially a sample preparation step. It allows for the continuous replication of subanalytical quantities of DNA to amplify it to a level such that routine analytical methods can be used for genotyping.

The discovery of the PCR has proven to be the catalyst for the modern biotechnology revolution. By 1988, the PCR had already ranked itself alongside cloning and DNA sequencing as an indispensable tool of molecular biology. The PCR has rapidly advanced the field of molecular biology. If the DNA sequence of a target region is known, the PCR allows for the replication and amplification of that region to the exception of the DNA, which is not required for analysis [7]. The in vitro replication of DNA, via the PCR, models the natural process of DNA replication. Each PCR cycle is one cycle of replication that theoretically doubles the amount of template DNA present in the sample. Each PCR cycle consists of three steps that are mediated by the temperature at which the reaction is proceeding and the relevant chemical and biological additives. The PCR is thermally controlled, readily automated, and is complete in 1–3 h.

The sequence of bases on the entire DNA molecule has now been determined [8]. It is accepted that with the exception of identical twins, no two humans contain identical DNA sequences. Therefore, an individual’s DNA sequence is unique. Although extremely significant in principle, this fact does not have high practical relevance purposes as scientists are not yet able to designate the complete sequence efficiently and cost effectively. It is of more significance that differences exist on the DNA molecule and that this characteristic is another attribute of DNA that aids in its application in the forensic context.

Summary

In forensic DNA profiling, investigators collect biological evidence from crimes primarily with the purpose of assisting in the identification of the donor. The fundamental objective therefore is to differentiate individuals through analysis at the DNA level. In general, therefore, it is areas of difference on the DNA that are of more importance than those that are highly conserved. Both sequence variation and length variation have been utilized in forensic DNA profiling for many years. More recently, sophisticated techniques examining length variation at microsatellite loci are favored as the routine target of analysis.

To be useful in forensic context, the scientific procedure employed must be robust (able to produce a result from compromised samples), sensitive (able to produce a result from small amounts of original material), highly discriminating (able to provide a result which conveys a satisfactory degree of confidence), and accurate (involve minimal subjectivity or risk of error). The forensic community has many analytical tools already that satisfactorily address these requirements, but despite considerable progress the scientific community is still only beginning to understand the vast amounts of information resident on the DNA molecule and to develop technologies to allow its exploitation. Undoubtedly there is much more development to come, but already the implementation of DNA science into the forensic arena has had a tremendous and positive impact. This is
primarily due to the inherent sophistication and informativeness of the DNA molecule itself.

References


Related Articles

**Biological Stains**

**DNA: Sources of**

SIMON J. WALSH