1 Biological Identification

What is Identification?

Forensic DNA typing was developed to improve our ability to conclusively identify an individual and distinguish that person from all others. Current DNA profiling techniques yield incredibly rare types, but definitive identification of one and only one individual using a DNA profile remains impossible. This fact may surprise you, as there is a popular misconception that a DNA profile is unique to an individual, with the exception of identical twins. You may be the only person in the world with your DNA profile, but we cannot know this short of typing everyone. What we can do is calculate probabilities. The result of a DNA profile translates into the probability that a person selected at random will have that same profile. In most cases, this probability is astonishingly tiny. Unfortunately, this probability is easily misinterpreted, a situation we will see and discuss many times in the coming chapters.

The drive to identify individuals is as old as humanity and is not limited to forensic applications. Your signature is a form of identification, as are biomarkers such as fingerprints and facial features. Your fingerprint or face can identify you for purposes of unlocking your phone, but neither method is infallible. The same is true of DNA profiling. Any forensic identification method aims to reduce the number of people with given characteristics to the absolute minimum and express the result as a probability. Accordingly, we will approach identification through the lens of probability, as this is the best and proper way to interpret it.

In part due to advances in DNA typing methods, the concept of identification has expanded. Before the development and widespread use of DNA typing

methods, biological identification was more a process of elimination than specific identification. For example, testing a bloodstain could reveal that it came from a person with type A blood, eliminating anyone with type B as a source. This finding is helpful but not definitive, as the proportion of persons worldwide with type B blood is roughly 11%, leaving a large percentage of the population as potential sources of the stain. DNA typing methods typically yield types expected to occur in one in billions or fewer. This finding is much closer to the ideal of individual identification. In later chapters, we will explore how identification in DNA terms expands to include identification of relatives and identification of ancestors.

Biological Identification

In forensic science, the goal of biological testing is to identify an individual as a possible source of biological evidence such as blood, semen, or saliva. Evidence, for example, can range from a fingerprint at a crime scene, or a bloodstain on clothing, to a discarded weapon. Testing is designed to link such evidence to a person of interest (POI) in a crime. There are other areas in which human identification is critical, including:

- Paternity testing where the identity of the father of a child is in question
- Mass disasters in which human remains (often fragmentary) need to be linked to a specific person
- Identification of military casualties and remains from current and past conflicts
- Missing person cases
- Human trafficking
- Historical investigations
- Archaeological investigations

We will touch upon all of these, but our emphasis will be on forensic applications. The methods used in biological identification and DNA profiling are similar for forensic, historical, and archaeological testing. The key differences are usually sample type and the timeframe involved. Forensic cases are contemporary or in the recent past, such as cold cases, which are unsolved criminal investigations that remain open pending discovery of new evidence. Such cases occur or have occurred in the recent past, measured in decades at

most. The oldest cold cases are typically from the 1940s or 1950s, but in those cases it is rare to have testable evidence.

Archaeological cases arise from studies focusing on individuals from eras ranging from hundreds to thousands of years ago. For example, DNA has been extracted and tested from Egyptian mummies and archaic humans such as Neanderthals. Such testing relies on bones or preserved tissues and has limited capabilities compared to what can be accomplished with fresh whole blood samples.

Historical applications relate to times that fall between archaeological and contemporary eras. In Chapter 7, we explore the identification of the last Russian Tsar and his family, killed in 1918.

Biological Methods of Identification

Biological identification is based on traits that are under genetic control. This type of evidence is called *biological evidence*, and most examples are bodily fluids. Blood is the most obvious source of biological evidence. Others include saliva (oral fluid), semen (seminal fluid), vaginal fluid, urine, and feces. All are potential biological and DNA evidence sources, but they must first be located and identified as biological evidence before additional analysis can occur. The type of evidence determines what testing can be done and what information can be obtained. Terms used to describe the evaluation of biological evidence are, appropriately enough, *forensic biology* and *forensic serology*. Blood, saliva, vaginal fluid, and seminal fluid are the most exploited types of biological evidence.

Notice we described these techniques by adding the word "forensic." For use in forensic applications, these methods of identification are adapted from established biological techniques rather than independently developed by the forensic community. Blood typing for ABO groups arose from research into deaths associated with blood transfusions. The techniques were adapted for forensic use. Forensic DNA methods evolved from research in molecular biology. Advances in the field parallel, but often lag, those used in fields such as medicine, pharmacy, and genetics. We revisit this process and its consequences several times in the coming chapters.

Characterization

The first step toward exploiting biological evidence is finding it. Biological evidence can be challenging to locate and identify, particularly with small quantities on a surface containing many other materials. Suppose a reddish stain is collected from a crime scene using a moistened swab. It may appear to be blood, but it is critical to establish this identification before investing additional time and effort in the examination. There is no point in attempting further biological testing on rust, ketchup, or red paint.

Similarly, soiled bedding or underwear associated with a sexual assault can benefit from characterization before further analysis. The testing flow moves from screening tests (also referred to as *presumptive tests*) through to *confirmatory tests* if needed. Most presumptive tests target proteins characteristic of the biological fluid. If we find something that looks like blood, we first perform the presumptive test to establish that it is indeed blood before trying to extract DNA. Ideally, these tests should not have any impact on subsequent DNA testing.

Presumptive tests for blood target hemoglobin, the iron-containing protein responsible for the red color of blood. Many chemical reagents react with hemoglobin to cause a distinct color change. Among standard tests is the phenolphthalein test, which produces a pink color in the presence of hemoglobin, and luminol, in which bright light is emitted because of a chemical reaction. Hemastix, a commercial test strip used to detect blood in urine, is also employed for this task. All these tests react with small amounts of hemoglobin. They can also produce occasional false-positive (where the test incorrectly indicates a substance is present) and false-negative (where the test incorrectly indicates a substance is absent) reactions, so they are used for screening rather than definitive identification. Some laboratories conduct additional testing to confirm the result and to determine whether the blood is human. Current DNA methods are specific to humans, but the time and cost involved in the analysis are significant. Thus, performing these additional testing steps can save time and money.

Stains from semen, vaginal fluid, urine, and saliva can become visible when illuminated by alternative light sources (ALS). A typical ALS system consists of lighting sources and filters. A light source is pointed toward the surface where

a stain may be present. Different light/filter combinations make residues of blood, urine, semen, and other bodily fluids easier to see. Seminal fluid contains a fluid component, called seminal plasma, and sperm cells. Test reagents detect selected enzymes found in abundance in this plasma. Another option is testing for prostate-specific antigen, or PSA (p30), using a small test device. The prostate is a small gland that sits below the bladder in males. A technique known as "Christmas tree staining," due to the colors produced, is utilized along with microscopy to find the sperm cells. The heads of the sperm cells are dyed red, and the tails green. No sperm will be present if the man has had a vasectomy, but the other screening tests will still work. Vaginal fluid is more challenging to identify as it lacks a unique protein to target. Epithelial cells from the vaginal tract are shed into vaginal fluid, and these can be detected microscopically. Finally, saliva contains high levels of a specialized digestive enzyme, amylase, which can be targeted in presumptive testing.

Successive Classification

The flow of analysis of biological evidence utilizes successive classification. Each test in a testing sequence reduces the size of the group (called a *population*) from which the sample might have come. Suppose that a red stain is found on a wall at a crime scene. It appears to be blood, so the crime-scene investigator collects it and sends it to the lab for analysis. The population from which this substance might have come includes anything that resembles dry blood, such as ketchup or red paint. The laboratory characterizes the sample, and the results suggest that it is blood, reducing the size of the potential source population. Next, the laboratory performs a species test that indicates human blood. Still a large population, but much reduced in size from the initial group. Simple blood typing shows the blood to be type A, which represents approximately 40% of people globally. In this way, successive testing allows us to reduce the number of possibilities to ever smaller numbers of potential stain sources.

Q vs. K Comparisons

Many forensic applications of biological typing involve comparing an unknown sample such as the crime-scene stain (the *evidentiary sample*) with

a *reference sample* such as one obtained from a person of interest (POI). This process is referred to as a questioned (Q) sample to known (K) sample comparison. Often there are multiple references or known samples involved. Q vs. K comparisons lead to one of *three possible outcomes*. For the sake of this example, we will assume that the questioned sample is the crime-scene bloodstain from the previous paragraph, and the known is a sample collected from a POI in the case. While the Q evidence sample is characterized by screening tests, there is no need for a serological characterization of the K reference sample since it is collected directly from the POI, typically as a buccal mouth swab or a blood draw. DNA testing is then performed separately on the Q and K samples.

Suppose in the first case that the DNA profile from the crime scene (Q) was unambiguously different from that of the POI (K). This finding results in an *exclusion*, meaning that the POI could not have been the source of the crime-scene stain. The second possible outcome is an *inclusion*, which occurs if the two DNA profiles match in all respects with no unexplainable differences. Additional statistical analysis and statements would follow, as we will discuss in later chapters. Finally, test results may be *inconclusive*. This finding might arise if insufficient information, such as a partial Q DNA profile, exists to support any conclusion.

Genetics and Heredity

Human identification using biological testing rests upon genetics and hereditary control of selected characteristics. DNA profiling is possible because everyone's genetic makeup is unique except for twins arising from the same fertilized egg. As we will see in Chapter 8, tools are emerging to address this situation. Furthermore, your genetic makeup is inherited from your parents through known and predictable processes. Current DNA profiling methods do not target genes (a common misconception), but they target variable regions of DNA that follow standard rules of heredity. DNA targeted in DNA profiling comes from our cells. Figure 1.1 illustrates the key points and features.

The top frame of Figure 1.1 shows a cell with a nucleus and a structure called the mitochondrion. Both structures contain typable DNA. We consider mitochondrial DNA (mtDNA) in Chapter 7. The nucleus contains the chromosomes

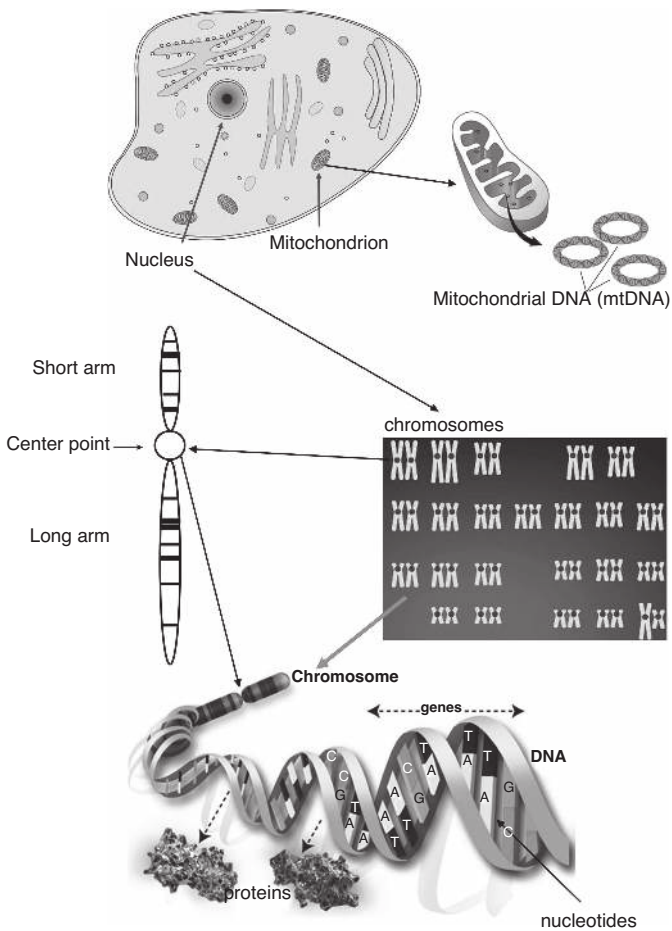

Figure 1.1 An overview of the sources and types of deoxyribonucleic acid (DNA) utilized in forensic DNA typing. DNA is found in two locations within the cell – the mitochondria (mtDNA) and the nucleus. Nuclear DNA is found in 23 pairs of chromosomes. Each chromosome is made up of strands of DNA, which organizes itself into a double-helix shape. The building blocks of DNA are nucleotides, which are illustrated in the next figure. A small portion of DNA corresponds to genes that code for proteins. The remainder of the DNA is referred to as non-coding.

(23 pairs in humans), as illustrated in the middle frame of the figure. The chromosomes have different sizes and are divided into two segments. The dark dot in the image in the middle right shows the dividing point (center point in the diagram at left). This point is essential in cell division and replication. Reproductive cells (eggs and sperm) contain 23 chromosomes, one member of each pair. These combine to form the complete chromosome set of a child. The sex-determining chromosomes are shown in the lower right of the chromosome array. Males have one X chromosome and one Y, while females are XX. DNA profiling targets these chromosomes and allows for the determination of biological sex. Another term we will use in coming chapters is *autosomal DNA*, which refers to DNA that comes from chromosomes other than the X and Y. We will also explore X and Y DNA applications for the identification and study of lineage and ancestry.

Chromosomes are made of DNA, as illustrated in the lower frame of the figure. DNA has a ladder-like configuration that is tightly folded into a double-helix shape. This shape arises from how components along the two strands bond to each other. Genes consist of long sequences of DNA. Each gene provides instructions for building a protein that has a specific function within our bodies. These proteins are large molecules capable of forming complex folded shapes, as shown in the illustration.

A closeup of the DNA structure is shown in Figure 1.2. The top frame shows the double-helix structure. The four essential compounds that link the two DNA strands are called bases – adenine (A), thymine (T), cytosine (C), and guanine (G). Their chemical structure is such that A binds with T (two bonds, as shown) while G bonds with C (with three bonds). A bonded pair such as A-T is called a *base pair*, with one base on one DNA strand and the other base on the other DNA strand. Because of the unique pair bonding of A-T and G-C, their relationship is complementary. The bonds between paired bases can be broken to allow DNA to unzip and then zip closed again in the same way since A binds to T and G to C. This ability to open and close the double-strands of DNA is central to cell replication and DNA typing.

The bases are attached to the DNA backbone, which is constructed of sugar and phosphate groups. These linked groups are the framework of the DNA molecule, with the bases facing toward the interior (the rungs of the ladder). The combination of phosphate, sugar, and base is called a *nucleotide*. One of

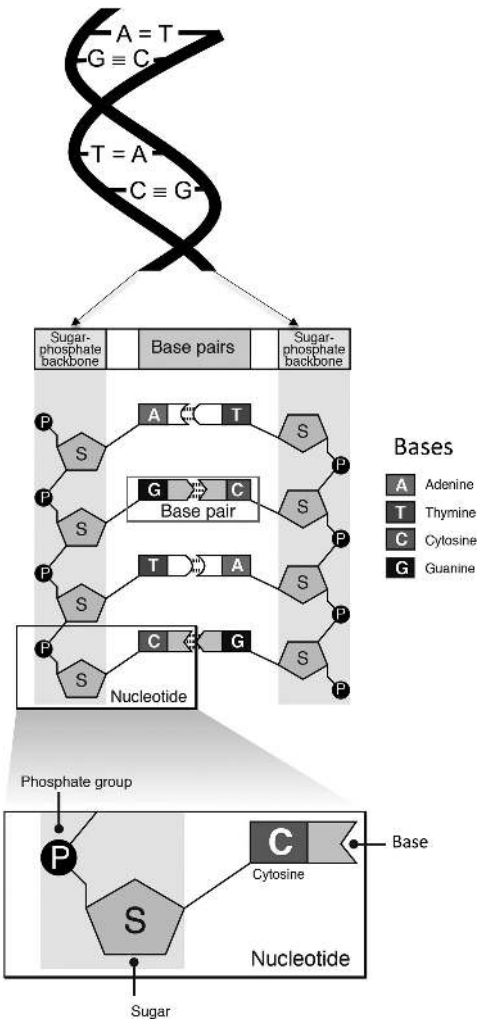

Figure 1.2 Expanded view of DNA structure. The double helix arises from a ladder-like structure with a sugar–phosphate backbone supporting rungs made of bases. The bases pair selectively (A with T and C with G) to link the strands.

the steps in DNA profiling is *amplification*, in which the existing DNA strands are copied; this is accomplished by unzipping a portion of the DNA molecule and adding nucleotides to create two copies of the DNA. We will discuss this step in more detail in Chapter 4.

Rules of Heredity

Variability in base sequences is what makes each of us biologically unique. A portion of the DNA contains information that results in protein synthesis. These proteins dictate our hereditary characteristics, such as eye color and blood type. Thus, rather than an alphabet of 26 letters, as in English, DNA information is communicated with four letters: A, T, G, and C.

Other portions of the DNA, such as those exploited in DNA profiling, do not code for proteins but still follow basic rules of heredity. Figure 1.3

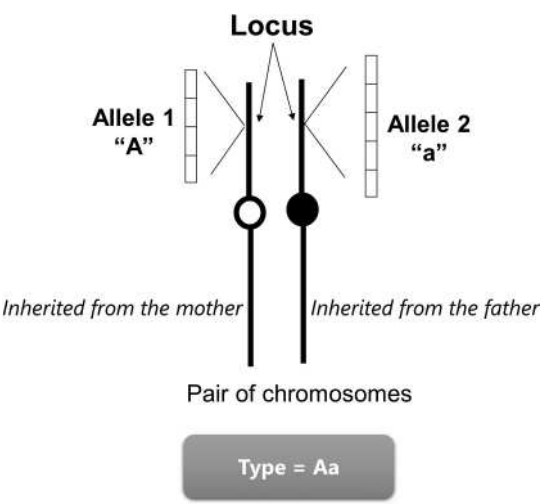

Figure 1.3 A typable location on a pair of chromosomes. The location or locus has two alleles – one from the mother and one from the father. In this example, the allele from the mother is the *A* variant and the allele from the father is the *a* variant. The person with this pattern at this locus is type *Aa*.