

Application of Forensic DNA Testing in the Legal System

Dragan Primorac, Moses S. Schanfield¹, Damir Primorac²

Laboratory for Clinical and Forensic Genetics, Split University Hospital and School of Medicine, Split, Croatia;

¹Analytical Genetic Testing Center, Denver, Col, USA; and ²Municipal Court, Split, Croatia

DNA technology has taken an irreplaceable position in the field of the forensic sciences. Since 1985, when Peter Gill and Alex Jeffreys first applied DNA technology to forensic problems, to the present, more than 50,000 cases worldwide have been solved through the use of DNA based technology. Although the development of DNA typing in forensic science has been extremely rapid, today we are witnessing a new era of DNA technology including automation and miniaturization. In forensic science, DNA analysis has become "the new form of scientific evidence" and has come under public scrutiny and the demand to show competence. More and more courts admit the DNA based evidence. We believe that in the near future this technology will be generally accepted in the legal system. There are two main applications of DNA analysis in forensic medicine: criminal investigation and paternity testing. In this article we present background information on DNA, human genetics, and the application of DNA analysis to legal problems, as well as the commonly applied respective mathematics.

Key words: anthropology, forensic; DNA fingerprinting; forensic medicine; fragment length; jurisprudence; legal medicine; minisatellite repeats; paternity; polymerase chain reaction; polymorphism, restriction

After the discovery of DNA structure in 1953 (1), the technological advancement in biomedical research has allowed the identification, through the Human Genome Project, of more than 150,000 genes coded by approximately 3 billion base pairs of DNA that is contained by the 23 pairs of somatic chromosomes.

For the vast majority of loci that code for a protein there is only one form of a gene. This is because most genes are not tolerant to mutations. Some genes are more tolerant to mutations and may have more than one form of a gene, *alleles* of that particular gene. Loci with alleles that occur relatively frequently are called *polymorphic*, whereas those loci that do not have alleles are called *monomorphic*. The genetic variation in blood groups, serum proteins, and transplantation antigens at the protein level reflects variation at the DNA level. Advances in DNA technology have allowed detection of variation (polymorphism) in specific DNA sequences.

It is known that the DNA sequences at many loci do not code for a translated gene. These loci are highly polymorphic due to the presence of multiple alleles represented by different numbers of repeated segments. These repeated segments are referred to as variable number tandem repeats or VNTRs (2-5). At any given locus they can differ by as few as one repeat or have as many as a hundred repeats. The most common method used to analyze VNTRs is amplified fragment length polymorphism (AFLP) (6,7). In some cases VNTR regions can cause genetic diseases. Sometimes the repeat in a VNTR is

three base pairs long, and codes for an amino acid in the coding region of a protein, which is referred to as a codon. If the number of repeats becomes too large it can distort the protein, causing a genetic disease. Two of these trinucleotide repeat diseases are X-linked mental retardation (Fragile X) and Huntington's disease. Most VNTR loci are in the regions of anonymous DNA, which do not code for proteins or are in the introns, non-coding parts of a gene. It is these VNTR loci that are used in forensic analysis of samples. During forensic analysis one or more loci are the particular interest of examination. For most forensic DNA testing, the majority of loci tested are within the 44 autosomes.

Analysis of the sex chromosomes is important if gender determination is needed. For example, the Amelogenin (AMEL) gene has two forms. The X chromosome form of the gene contains a small deletion (6bp) in a non-translated region of DNA, which produces a shorter product when amplified by polymerase chain reaction (PCR) than the same gene on the Y (male) chromosome.

Y chromosome can provide important information if there is a question about identifying lineages from a specific male. This is possible because the Y chromosome contains highly polymorphic regions (8). Human Y chromosome is present in normal males as a single copy that is paternally inherited and cannot recombine or rearrange because there is only one Y chromosome. In a forensic analysis, the Y chromosome can play an important

role in sexual assault cases where more than one male has contributed to a mixture (9). These markers can also be very useful in paternity cases where the child is male.

Nomenclature

Prior to 1985, various groups of scientists met to standardize the use of reagents and nomenclature. One of the more noteworthy attempts are the regular meetings of the transplantation community to standardize tissue typing reagents for purposes of transplantation. With the advent of the Human Genome Project and the wide scale search and mapping of human genes it was necessary to develop guidelines for how new genes would be identified and how old systems would be incorporated. The results from this discussion became the framework for meetings to follow on human genetic nomenclature. The system now referred to as the International System of Gene Nomenclature or ISGN (1987) was the beginning (10). This was followed by specific recommendations of the DNA Commission of the International Society for Forensic Haemogenetics relating to the use of restriction fragment length polymorphism (RFLP) and polymerase chain reaction (PCR) (11,12). In general, all loci have unique identifiers of 2-7 characters in length that are all in capital letters, for example, AMEL for the Amelogenin locus. Phenotypes (i.e., the results of laboratory tests) are presented as the locus name followed by a space and then the observed results (e.g., GYPA A). A two band pattern is represented by the locus symbol, a space and symbols of the alleles separated by a comma (e.g., GYPA A,B). A genotype can only be determined by combining the laboratory results with the pedigree analysis of inheritance in a family. It is easy to deduce a heterozygote by the presence of two bands or dots. However, due to the presence of mutations that can cause null alleles, a single band (dot) can be determined to be a true homozygote only by family study. Operationally, in most cases a single band pattern is a homozygote and is treated as such. Genotypic results are presented as the locus name followed by an asterisk and the deduced results separated by a slash followed by the second asterisk (e.g., GYPA*A/*B, GYPA*/*A). Genotype is always italicized (*GYPA*A/*B*) or underlined if italics is not available (GYPA*/*A). For anonymous DNA segments, such as D14S1, the following rules hold. The anonymous DNA region is identified as D for a DNA segment. The number represents the chromosome the anonymous DNA is found on. "S" indicates that it is a single copy region that only occurs once in the genome, and "Z" indicates that it has multiple locations. The number indicates the order in which it was reported as a DNA marker for that chromosome. For D14S1, it is a DNA segment on chromosome 14, a single copy, and the first marker found on Chromosome 14. This example is the first reported hypervariable RFLP (13). The previously stated guidelines require the use of phenotypes when presenting DNA results.

For DNA polymorphisms the following standards are in force. Only phenotypes are presented. Alleles or band sizes are always reported from the smallest to the largest, for example: 10, 15, 17 or 2.33, 3.56 kb. Heterozygous phenotypes are always separated by a comma

without a space. Single band patterns, whether for RFLP or PCR based systems, are reported as a single band phenotype, i.e., 3.56 or 12, and not as a genotype.

Mitochondrial DNA

Mitochondrial DNA (mtDNA) is another source of DNA of forensic interest. The mitochondria contain multiple copies of mtDNA, where 22 genes encode transfer RNA (tRNA), two encode ribosomal RNAs (12S, 16S RNA), and 13 encode protein enzymes involved in the electron transport chain of oxidative phosphorylation and ATP production (14). The mtDNA is 16,569 bp long, circular, and without any intron sequences. Two non-coding hypervariable regions are of particular interest for forensic analysis: HV1 and HV2. Unlike nuclear DNA, the mitochondria and their DNA originate in the cytoplasm of the egg that formed the zygote and are therefore of maternal origin. Thus, mtDNA is maternally inherited, and represents the female ancestry of an individual. Further, mtDNA in a given cell may be present in thousands copies, compared with only two copies of nuclear DNA. Since mtDNA molecules replicate independently of one another, unlike nuclear chromosomes which pair before replication, there is no mechanism by which mtDNA can undergo recombination. The only source of variation in mtDNA is mutations that change the base pair sequence of the mtDNA. The fact that mtDNA is maternally inherited precludes an individual from being heterozygous, making it useful in tracing maternal lineages within families and populations. MtDNA is primarily employed in forensic casework to test evidence that does not contain a sufficient quantity or quality of DNA for nuclear DNA testing, such as repatriated remains from graves (15). MtDNA also occurs in non-nuclear tissues such as hair shafts. However, one of the major problems in working with mtDNA in hair is the presence of the two or more subpopulations of mtDNA (heteroplasmy) within an individual. Heteroplasmy is probably the result of a much higher error rate during mtDNA replication than in nuclear DNA. It may be because mtDNA molecules are replicated independently of one another and are not strictly tied to meiotic or mitotic cell division. Since each cell contains a population of mtDNA molecules, a single cell can harbor some molecules that have an mtDNA mutation and others than do not. This phenomena may be responsible for a variable expression of mitochondrial diseases as well. However, it is possible that the number of mutant mtDNA molecules may change through replicative segregation as cells divide and as mitochondria proliferate. For the forensic purposes, heteroplasmy is important since it may strengthen or complicate forensic identity testing. However, heteroplasmy represents an additional level of variation that in most cases can increase the power of mtDNA testing. Further information about this issue can be found in an excellent article by Holland and Parson (16). MtDNA plays an important role in the identification of human remains, particularly skeletal, of decomposed bodies. One of the most famous cases that was solved using this technology was identification of the Tzar Nicholas II, where it was confirmed that he had the same heteroplasmy as the remains of his brother Georgij Romanov, the Grand Duke of Russia (17).

Methods of DNA Analysis Used in Forensic Sciences

The most important step for any forensic analysis is the collection and identification of evidence. If evidence is not properly documented, collected, packaged, and preserved, most likely the final data will not be admissible in a court of law.

Documentation of evidence should be thorough and the person should follow all guidelines specially developed for such purposes. All specimens found should be labeled with an item number, date, time, location, and collector's name. If a case number is assigned, that too would be added. An exhaustive description of the collection and preservation of evidence for DNA testing can be found in Lee et al (18).

Restriction Fragment Length Polymorphisms

The first method to be applied in forensic DNA analysis was restriction fragment length polymorphisms (RFLP), where DNA is digested with restriction endonucleases, the fragments separated by electrophoresis on an agarose medium, transferred to a nylon membrane and detected by hybridization with either a radioactive or chemiluminescent probe (19-21). Two main types of DNA polymorphism are studied using RFLP technology: single base changes and minisatellite VNTRs loci. Analysis of the single-base polymorphism has only limited value in forensic DNA identification since there are limited numbers of alleles with these changes and limited numbers of useful loci. However, since the mid 1980s, the most common method used in forensic DNA analysis has been RFLP analysis of VNTR minisatellite loci. These loci vary in length due to the variation in the number of repeated DNA segments at a specific locus. The VNTR loci used contain repeated sequences that vary between 15 and 70 base pairs. Standardization of the test systems in forensic science has led to the use of the restriction enzyme *HaeIII* in the United States and the enzyme *HinfI* in Europe, with other countries choosing one of the two systems. However, this technique requires a relatively large amount of DNA, it is labor-intensive and lengthy, and in some locations requires the use of radioactivity. An extensive review of the application of RFLP technology and its forensic applications can be found in Wayne (22).

Polymerase Chain Reaction

The polymerase chain reaction (PCR) is a method where a certain region of DNA is copied (amplified), producing at optimum a billion copies that are completely identical to the original. This technology has several advantages over RFLP (21,23). It requires less DNA, it is rapid (results are obtained often within 24 hours), and does not require radioactive isotopes. However, it has two possible disadvantages: the susceptibility to contamination, and the fact that many of the PCR loci have fewer alleles than minisatellite VNTRs. The first forensically used PCR system detected the HLA-DQA1 locus, a region known to be part of the human histocompatibility system. The kit was developed by the Cetus Corporation (Emeryville, CA, USA) and marketed by the Perkin Elmer Corporation (Norwalk, CT, USA). The DQA1 typing system relied on the probes to detect variation in the DNA sequence that was previously de-

tected using antibodies to proteins. The original system had six alleles. When testing evidence, approximately 16% of the time two individuals in a case would be the same type, which lead to a need for additional PCR based typing systems. As in the RFLP system, the original tests were based on sequence variation. Researchers soon found an abundant class of VNTR polymorphisms with small repeats (2 to 9 repeats). This class of VNTR loci is referred to as short tandem repeats (STRs), or microsatellite repeats to differentiate them from the larger VNTRs detected by RFLP and some PCR based typing systems called large tandem repeats (LTRs) or minisatellites. The systems chosen for forensic testing have primarily 4 bp (tetranucleotide) repeats that are repeated from 5 to approximately 50 times, depending on the loci (24,25). These STRs are very small in size (100-400 bp) and are very helpful in analyzing degraded DNA, in comparison with RFLP bands that can vary in length from 500 to 12,000 bps.

In the USA, 13 STR loci were chosen for use in a national database of convicted offenders (26). This database is referred to as the Combined DNA Indexing System, CODIS. It was originally limited to convicted sex offenders but has been expanded in some jurisdictions. The 13 CODIS loci were chosen to overlap other sets of loci chosen by Forensic Science Services in the United Kingdom for their national database and organizations such as INTERPOL.

The new STR based typing systems have several advantages. The first is that multiple loci are amplified simultaneously (multiplex reactions). The second is that these test systems can be directly detected without the use of probes. One primer from each set of primers has a fluorescent label attached to it, such that the PCR multiplexes can be detected using different wavelengths of light. This allows simultaneous electrophoresis and detection of eight to ten loci at a time (27). Further, the use of fluorescence makes the systems amenable to automated detection. Commercial kits are available to go with the different automated detection systems.

A review of basic PCR technology and STR PCR technology and their application to forensic testing can be found in refs. 26 and 28.

Genetics Rules and Statistical Considerations

The distribution (segregation) of parental genotypes in offspring depends on the combination of the alleles in the parents. In each case, they refer to a given gene locus.

Mendelian Inheritance

Mendel's laws state the expected distribution of alleles in the offspring of a specific mating. Mendelian inheritance consists of three laws:

1. *The cross.* If a homozygous A,A father mates with a homozygous B,B mother, all offspring must be heterozygous for alleles A and B; noted as A,B.

2. *The intercross/segregation.* If two heterozygous A,B parents mate, the following genotypes will occur: A,A, A,B and B,B at the ratio of 1:2:1, respectively. This law also states that the two members of a gene pair separate (segregate) from each other during meiosis, with 50% of gametes carrying one allele (A) whereas 50% of gametes carry the other allele (B).

3. *Independent assortment.* When the mating individuals have more than one segregating locus, they are genetically unlinked, i.e., each locus segregates independently. This law indicates that the separation of alleles at one locus during meiosis is independent of the other (Fig. 1).

Rules of Parentage Testing

Mendel's laws are used to create four rules for parentage testing:

1. The child cannot have a genetic marker that is absent in both parents.
2. The child must inherit a pair of genetic markers from each parent.
3. The child cannot have a pair of identical genetic markers unless both parents have the marker.
4. The child must have the genetic marker that is present as an identical pair independently in both parents.

Hardy-Weinberg Equilibrium

Gregor Mendel described the behavior of alleles in a mating. Similarly, the behavior of alleles in a population was independently described by Hardy and Weinberg in 1908. The Hardy-Weinberg equilibrium indicates a predictable relationship between allele frequencies and genotype frequencies at a single locus (29). Thus, the sum (Σ) of 3 allele frequencies ($z_1, z_2,$ and z_3) is:

$$\Sigma = z_1^2 + z_2^2 + z_3^2 + 2z_1z_2 + 2z_1z_3 + 2z_2z_3 = z_1^2 + z_2^2 + z_3^2 + 2z_1z_2 + 2z_1z_3 + 2z_2z_3$$

For example, if we suppose that alleles Z_1, Z_2 and Z_3 have frequencies or proportion of z_1, z_2 and z_3 , respectively, then the frequencies for the homozygous genotypes $Z_1, Z_1; Z_2, Z_2$ and Z_3, Z_3 are, respectively, $z_1 \cdot z_1, z_2 \cdot z_2,$ and $z_3 \cdot z_3$ and the frequency of heterozygous genotypes Z_1, Z_2, Z_1, Z_3 and Z_2, Z_3 are $2 \cdot z_1z_2, 2 \cdot z_1z_3$ and $2 \cdot z_2z_3$.

If the Hardy-Weinberg equilibrium is correct, allele frequencies at a locus in a population are constant over time. However, departures from Hardy-Weinberg equilibrium can occur for several reasons:

1. sampling errors between generations due to a small population size, referred to as random genetic drift;
2. inbreeding caused by the inter-relatedness of parents, referred to as non-random genetic drift and leading to a decrease in heterozygosity;
3. migration of genes from one population to another, which changes the frequencies and affect the Hardy-Weinberg equilibrium until a new equilibrium is reached;

4. mutation or changes in an inherited allele, which affect allele frequencies; and

5. natural selection, which can increase the frequency of an allele over time. Children with this allele have a higher chance of surviving than children with other alleles. Examples of this are G6PD deficiency or thalassemia in areas with malaria.

Linkage Equilibrium

Linkage equilibrium is the concept that alleles on independent loci will be dispersed in a population as a multi-locus Hardy-Weinberg equilibrium. If loci are close to each other on the same chromosome, they are in genetic equilibrium. Although mixing populations can reach HardyWeinberg equilibrium, genes on the same chromosome reach equilibrium based on the genetic distance and the degree of recombination (rearrangement) between the loci. On the other hand, if there is a mix of two populations that have many different loci and their alleles have markedly different frequencies, associated alleles that are not genetically linked can be considered to be in a statistical linkage disequilibrium. In practice, this latter type of disequilibrium is difficult to demonstrate at the population level. However, it is possible that it can be demonstrated in specific individuals that are recently introduced into a population.

DNA Evidence in the Court

In the USA, scientific evidence falls into the realm of the expert witness as opposed to a lay witness. A lay witness can only testify to things they have seen or in some cases heard. An expert witness is allowed to express an opinion, based on the knowledge and experience the expert possesses (30). This is generally considered to be outside the scope of knowledge of a regular or lay person. Expert witnesses can include individuals who received the evidence as part of what is referred to as "chain of custody", performed a specific test, such as a DNA test, or have detailed theoretical knowledge about DNA or drug testing. The fact that an expert did the testing does not mean that evidence is automatically admitted. Normally there is a requirement to offer some proof that the presented evidence is reliable and provides useful information.

In the USA, two major standards exist for deciding if the scientific finding will be admitted into evidence: the "general acceptance" test and "sound methodology" standard. For the first, the theory and methodology used must be generally accepted in the scientific community. How difficult it is to apply this rule is shown by the case of Kelly vs. State, where, prior to admitting an RFLP VNTR profile match, five experts had to answer a question asked by a defense expert "Is radioactive technology too new to be generally accepted in the scientific community?" (31). The "sound methodology standard" allows a much broader framework for deciding whether the proposed testimony has sufficient scientific validity and reliability.

The Council of Europe, Committee of Ministers, on February 10th, 1992, during the 470th meeting, published the Recommendation No. 92 on the use of DNA analysis within the framework of the criminal justice system (32). Some of the most important recommendations are:

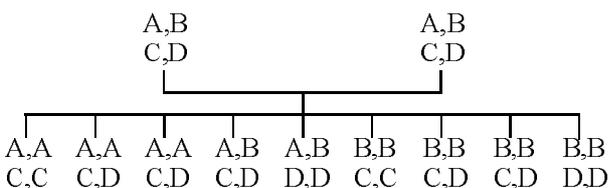


Figure 1. An example of a mating of two parents heterozygous for A and B alleles at locus 1, and C and D at locus 2. Genotypes resulting from this mating are: A C; A C,D; A D; A,B C; A,B C,D; A,B D; B C; B C,D; and B D at the ratio 1:2:1:2:4:2:1:2:1, respectively.

1. Samples collected for DNA analysis and the information derived from such analysis for the purpose of the investigation and prosecution of criminal offences must not be used for other purposes. However, where the individual from whom the samples have been taken so wishes, the information can be given to him/her.

2. Samples collected from living persons for DNA analysis for medical purposes, and the information derived from such samples, may not be used for the purposes of investigation and prosecution of criminal offences unless in circumstances laid down expressly by the domestic law.

3. Biological samples or other body tissues taken from individuals for DNA analysis should not be kept after the rendering of the final decision in the case for which they were used, unless it is necessary for purposes directly linked to those for which they were collected.

Croatian Criminal Law, Article 265, Section 5, states that, during the pretrial phase, the court may request genetic analysis of the samples isolated from the individual who is suspected to have committed a crime in which the form of punishment is prison sentence (33). This law also states that, after analysis, the data collected may be retained for ten years if the person has been convicted of one of the following crimes: serious offences against life, integrity and security of persons, and sexual freedom. Here it is important to emphasize that the Minister of Health determines all procedures for evidence collection, storage, and analysis.

The use of DNA typing to identify human remains in Croatia has led to the application of DNA analysis in paternity and criminal cases. Since DNA technology was introduced to the Split-Dalmatian County Municipal courts in 1994, more than 100 requests for paternity identification have been received, out of which 76 cases have been tested (2 in 1994, 11 in 1995, 17 in 1996, 11 in 1997, 7 in 1998, and 28 in 1999). DNA analysis excluded 11 falsely accused alleged fathers.

In the past 6 years, DNA results have been admitted into evidence in almost all courts in Croatia. Interestingly, the paternity in one case was confirmed by the analysis of DNA isolated from paraffin embedded tissue that belonged to an aborted fetus. In all cases of falsely accused men, the exclusion was made by analyzing more than three alleles.

Thus far, the Laboratory for Clinical and Forensic Genetics at University Hospital Split has analyzed more than 150 samples from criminal cases. One of the first criminal cases in Croatia where DNA technology was successfully used was an explosion accident in military barracks near Zagreb. Split Clinical and Forensic Genetics Laboratory successfully tested more than 30 samples from that event.

Paternity (Parentage) Testing

The basic function of all forensic testing, whether parentage testing or identification, is to exclude the maximum number of individuals possible. This is done in parentage testing by identifying the obligate allele and determining if the alleged or putative father has this allele (for expository purposes the questioned father will be called the alleged father) (34). The obligate allele is

the allele that had to come from the biological father. For example, if a mother is type A,A and the child is type A,B, the biological father must have contributed the B allele to the child. Unless demonstrated to the contrary by the violation of Mendelian inheritance, the mother is assumed to be the mother of a child. If the alleged father has the "B" allele as either a homozygote (two B alleles) or a heterozygote (one B allele), he cannot be excluded. If the alleged father does not have the "B" allele, he could not give it to the child, and is therefore excluded. An exception to this would be if the mutation of A to B is relatively common. If the mother is type A,B and the child is A,B then the obligate alleles are A and B and the biological father could contribute either an A or a B allele. Any alleged father that has A or B cannot be excluded. If, however, the alleged father were type C,C or any other non-A, non-B genotype he would be excluded. In this case there would be two obligate alleles.

Mathematics and Paternity

If an alleged father is not excluded, the weight of this evidence toward paternity can be calculated. The weight of the evidence is the relative chance that the alleged father gave the child the obligate allele when compared to an unrelated individual in the population. The relative chance of the alleged father transmitting the obligate allele versus an unrelated individual in the population is referred to as the *System* or *Paternity Index* (SI or PI). The System or Paternity Index is a "likelihood ratio", i.e., the ratio of the chance that the alleged father versus the population gave a gene. Mendel's laws determine the likelihood or chance that the alleged father can give the obligate allele. If the father is either homozygous for the obligate allele or has both obligate alleles, his likelihood of transmitting the obligate allele is 1.0 (2/2). If the alleged father has only one copy of the obligate allele, or only one of the two obligate alleles, his likelihood of transmitting the obligate allele(s) will be 0.5 (1/2). The SI or PI for a given locus is the likelihood that the alleged father can transmit the allele divided by the frequency of the obligate allele(s). Thus, SI or PI will be either $1/p$ or $0.5/p$, depending on the number of obligate alleles the alleged father has. If there are two obligate alleles as determined by the types of the mother and child, p will be calculated using the formula $p=p_1+p_2$, and SI will be $1/(p_1+p_2)$ or $0.5/(p_1+p_2)$. SI is calculated for each locus that was tested. The rules of probability theory determine how SIs are combined to determine the combined likelihood of paternity. This is called either the *Paternity Index* (PI) if the term SI is used, or the *Combined Paternity Index* (CPI) if the term PI is used.

Paternity Index or Combined Paternity Index

The rules of combining probabilities are quite simple. If one is interested in the combined probability of the characteristics A and B and C, then the individual probability of A and B and C are multiplied. If one is interested in the combined probability of A and B, then the individual probabilities are added. Therefore the likelihood that the alleged father gave locus 1*A, locus 2*C, and locus 3*E to one child is the product of the likelihood for 1*A.2*C.3*E. Therefore, multiplying all of the individual likelihood ratios generates the PI or CPI. For a reasonable battery of tests this number should easily exceed

400 to 1,000. The PI or CPI basically states that the alleged father is 400 or 1,000 times as likely to be the biological father of the child as an unrelated man from the same population. In Germany, a minimum PI or CPI of 1,000 is required, whereas in the United States the required PI or CPI can be less than 100. Since many individuals have a hard time relating to a likelihood ratio and interpreting it, an alternative way of presenting the data is to convert the likelihood ratio to the probability of paternity.

Probability of Paternity

An 18th century mathematician Bayes developed a theorem to estimate the probability of event occurring even when the event cannot be directly measured. This is the basis of Bayesian mathematics that is widely used, especially in fields such as genetic counseling, to estimate the risk of having an affected child. Bayes' formula for the estimation of an event occurring is as follows:

$$\text{Bayes Formula} = \frac{X \cdot p}{X \cdot p + y(1-p)}$$

where X is the likelihood that an event will occur, Y is the likelihood that it will not occur, p is the prior probability that X will occur, and 1-p is the prior probability that it will not occur.

There are several constraints on Bayes' formula. One is that it is exhaustive, i.e., that it must include all of the options in the universe. In the case of paternity testing, X is the combined likelihood that the alleged father transmitted all of the obligate alleles and consists of the product of all of the 0.5 and 1.0 values for each locus. Y is the combined likelihood or chance that some unrelated person in the population is the biological father and the product of all of the obligate allele frequencies. Prior probability is the probability (likelihood) that the event in question could occur without any knowledge of current outcome. In this case it is the likelihood that the alleged father is the biological father before there are any laboratory test results. There are many possible ways to calculate the prior probability that the alleged father is the biological father. One could assume that it is the laboratory inclusion rate that is approximately 70%. One could also assume that it is the number of eligible males in the area of conception. By convention it is assumed that the alleged father is equally likely to be or not to be the biological father. This is the so-called neutral prior probability. Thus the prior probability is 0.5. Some statisticians believe that the prior probability should be calculated by dividing the desired outcome by the total number of possible outcomes, e.g., the prior probability of getting a single dot on a dice (half of a pair of dice) is 1/6, whereas the prior probability of getting a head on a coin is 1/2. Since for parentage testing there are only two possible outcomes – he is or he is not the biological father – the prior probability is 1 over the two possible options or 0.5, indicating that the “neutral prior” may be the correct prior. If the value of 0.5 is chosen for the prior probability, the formula reduces to a formula that uses the PI (CPI):

$$\text{Transmission Probability of Paternity (p=0.5)} =$$

$$\frac{1}{1 + (1/PI[CPI])}$$

The probability of paternity is the probability that the alleged father is the biological father of the child. Therefore, if the PI (CPI) is 400, then the probability that the alleged father is the biological father is 0.9975, or 99.75%. In contrast, the probability that the alleged father is not the biological father is 0.0025 or 0.25%. This latter value is 1 in 400, or the reciprocal of PI.

Genetic Information of the Test or the Random Man Not Excluded (RMNE)

The Paternity Index and the probability of paternity deal with the Mendelian likelihood that the alleged father is the biological father of the child. An alternative assessment of the same information is to ask how much genetic information is present in this mother-child pair; how powerful is the test at preventing the false inclusion of an alleged father. This is similar to the statistical concept of the power of a test. Ideally the test should be as powerful as practical, to exclude all falsely accused alleged fathers. We know that there are falsely accused alleged fathers because approximately 30% of alleged fathers are excluded worldwide (35).

The power of a paternity test can be determined by the calculation of the *Random Man Not Excluded* or RMNE. This statistics is comparable to the population frequencies calculated for forensic identification (see below). The RMNE is the proportion of the population that could contribute all of the obligate alleles and therefore could not be excluded, or would be falsely included. This term is used to describe the frequency of those individuals who cannot be excluded by chance. The formula for the single locus RMNE is: $p^2 + 2p(1-p)$ or, simplified, $1 - (1-p)^2$. The Combined Random Man Not Excluded (CRMNE) is analogous to the PI or CPI, i.e., it is the product of the individual values. The value of the CRMNE is normally quite small, such as 0.0025. It is easier to refer to the reciprocal of this number (1-CRMNE), which is often referred to as the Probability of Exclusion. However, since the term “Probability of Exclusion” can be confused with the “Probability of Paternity”, a less confusing term is the *Exclusionary Power (EP) or Power of Exclusion (PE)*. EP (PE) represents the probability of excluding a falsely accused man. It turns out that if a prior probability of 0.5 is used, the EP (PE) is also the Bayesian Probability of Paternity under what can be called A non-exclusion model. To avoid confusion with the “Probability of Paternity” derived from the inheritance model, this should be referred to as the “Non-exclusion Probability of Paternity” or the Weiner Probability of Paternity, after the scientist who proposed it in 1976 (36). The non-exclusion probability of paternity has some advantages over the standard probability of paternity (37). The biggest advantage of this method is that it always increases with the number of tests performed and is totally analogous to the population frequencies used in forensic identification profiles. The traditional probability of paternity (allele transmission probability of paternity) can decrease if the alleged father is heterozygous and the mother and child are the same heterozygous type, consisting of two common alleles, such that $p_1 + p_2$ is greater than 0.5.

Table 1 presents an example of a paternity examination. The obligate alleles, generated after the examination of the types of the mother and child, are bold faced in the table. A column with all of the obligate alleles is shown, as is the presence of all of the obligate alleles in the alleged father. The column "p" is the allele frequency for each allele, based on a Croatian database (M. Kubat, personal communication). RMNE is calculated using the formula from above. "X" is the likelihood that the AF can transmit the obligate allele and is determined by the father's type. The SI is determined by X/p . For the locus D5S818, the mother and child are the same, creating two obligate alleles and an SI close to one. The combined values are presented at the bottom of the table. For the complete trio, the alleged father is approximately 200,000 times as likely to be the father as an unrelated Croatian male. This translates to a 99.9995% probability of paternity. Looking at the RMNE side, there is a 99.9996% chance of excluding a falsely accused father. Thus, for this case, the probability of paternity based on allele transmission and the probability of paternity based on non-exclusion are identical.

Occasionally it is necessary to do a paternity test in the absence of the mother. In Germany, this is referred to as a *deficient test* since not all parties are present. There is significant loss of information in the absence of the mother. The formula used to calculate the frequency of potential mothers in a given population is:

$$RMNE = \frac{p^2 + q^2 + 2pq + 2p(1-p-q) + 2q(1-p-q)}{(p+q)^2 + 2(p+q)(1-p-q)}$$

Similarly, the SI is calculated differently. In the case of an absent parent there is a specific formula for each child-alleged parent phenotype set. The formulas taken from Brenner (39) are presented in Table 2.

To demonstrate the loss of information and differences in results, the paternity case in Table 1 was recalculated without the data on the mother, using the appropriate formulas from Table 2. The results are presented in Table 3, using the same format as in Table 1. The first thing to note is that the Exclusion Power is decreased approximately 1,700 fold, whereas CPI decreases only by approximately 60 fold, indicating the differences between the two methods of calculating probabilities of paternity. In general, the RMNE method provides a more conservative estimator of relationship. The ratio of the combined likelihoods taken from Table 1 is basically 0.805 (Transmission/RMNE), whereas the same ratio is 6.08 in Table 3, indicating that the transmission method produces results that are approximately 6 times larger than the RMNE method when the mother is missing.

Maternity Identification

Occasionally a child is found abandoned or dead and the question arises who the mother of the child is. Unlike the situation of paternity identification, here there is no knowledge about the mother or father of the child. Therefore, as is done in other forensic evidence, the genetic profile of the child is the evidence. If the child is heterozygous at a given locus, there are two different obligate alleles. In contrast, if the child is homozygous at a given locus, there is only one obligate allele.

The formula used to calculate the RFNE which stands for "Random Female Not Excluded" is the same

formula used for the RMNE when the mother is missing. If the mother does not have either of the two alleles that the child has, the mother is excluded. But if the mother has even a single allele that the child has, she cannot be excluded. If this result holds for all tested loci, then the alleged mother cannot be excluded. The Combined Random Female Not Excluded (CRFNE) is the product of all of the individual RFNEs and represents the percentage of females in the population that could not be excluded by chance, or that could be falsely included. The inverse, or $1/CRMNE$ will generate the number of women that would have to be tested to find a coincidental match. If a Bayesian probability of maternity is needed, $1-CRFNE$ is the Bayesian probability of maternity as above.

Alternatively, it is possible to calculate a maternity index based on allele transmission, similar to the SI or PI seen above, and convert that to a Bayesian probability. The same formulas presented in Table 2 apply to calculating a maternity index in the absence of a father.

Table 4 is an example of maternity identification. The child from the case in Table 1 is used again. Since the child's profile generates the RMNE/RFNE, they are identical in the paternity case without the mother in Table 3 and the maternity case in Table 4. However, due to the different shared alleles, SIs and combined PI are different. In this case the transmission method is approximately 9 times greater than the RFNE method. It should be noted that even with the loss of information in both the motherless (Table 3) and fatherless (Table 4) examples, the probability of parentage, even when the most conservative method is used (e.g., RMNE/RFNE) is greater than 99%, indicating that identification of parents can still be made.

Mixed Populations Calculations

For the calculation of a paternity index, it is customary to use the population frequencies of the alleged father or mother. If the alleged father or mother attributes their background to be a mixture of two diverse populations, such as Croatian and Sub-Saharan African, then the value of p used should be the average of the two population frequencies if they are available. On the other hand, if the questioned parent states that he or she is Italian, and only a Croatian database is available, the effect on the results will be negligible. The RMNE/RFNE is calculated as described above, depending on the case.

In reality, almost all human populations are to some extent mixed with other populations over their history. Alleles specific for Central Asia, Northern Asia, and Africa can be found in European populations due to the historical infusion of genes by Mongols and other Asian visitors, as well as the presence of African slaves in many populations. Special accommodations do not need to be made for these historical events as they will be represented in population samples. If the subject of the investigation is not a native but rather a tourist from, for example, Puerto Rico, it would be useful if a Puerto Rican database could be used. However, if a population specific database is not available, there is an error created by not using the appropriate database, though at the present time the magnitude of errors for STRs has not yet been published. If the structure of the population in question is known, a synthetic frequency could be calculated using the proportional weighted frequencies of the constituent populations.

Table 1. Example of a complete paternity examination^a

Loci	Mother	Child	Obligate allele	Alleged father	p (Y)	RMNE	X	SI
D3S1358	15,16	15, 18	18	16, 18	0.1400	0.2604	0.50	3.571
VWA	16,19	13 ,16	13	13, 16	0.0100	0.0199	0.50	50.000
FGA	20,21	21, 22	22	21, 22	0.1575	0.2902	0.50	3.175
THO1	9	8 ,9	8	7, 8	0.1150	0.2168	0.50	4.348
TPOX	6,7	6, 11	11	11	0.2675	0.4634	1.00	3.738
CSF1PO	11,13	10 ,11	10	10 ,12	0.2750	0.4744	0.50	1.818
D5S818	10,12	10,12	10	11, 12	0.0487 ^{FBI}			
			12		0.3875	0.6821	0.50	1.146
D13S317	11	8 ,11	8	8 ,12	0.1550	0.2860	0.50	3.226
D7S820	8,10	8, 12	12	9, 12	0.1400	0.2604	0.50	3.571
AMEL	X,X	X,Y		X,Y				
Combined						3.64 x 10 ⁻⁶		221,234
Exclusion power						99.99964%		
Likelihood						274,697		
Probability of paternity						99.9996%		99.9995%

^aAbbreviations: p (Y) = allele frequency for each allele or obligate allele based on Croatian or FBI database. FBI indicates frequencies taken from the European FBI database (38). RMNE (random man not excluded) = the proportion of the population that could contribute all of the obligate alleles and therefore could not be excluded. X = likelihood that the alleged father can transmit the obligate allele, determined by the father's type. SI (system index) = likelihood that the alleged father can transmit the allele divided by the frequency of the obligate allele (SI=X/p). Obligate alleles are written in bold.

Identification of Human Remains

A variety of methods are available to identify human remains. Depending on the circumstances and the state of the remains, the most common methods include identification of remains by a living person who knew the deceased victim by direct facial recognition, or recognition of special features, such as scars or marks (tattoos), matching of fingerprints (if *pre mortem* inked prints are available), dentition, and DNA.

In war, when a large number of bodies are often buried in common graves and premortem data are unavailable, the identification is much more difficult (40). The development of genomic and mitochondrial DNA technology has provided the forensic community with a valuable new tool for establishing personal identification. Because of the advanced decomposition of many bodies found in the war or mass disaster situations, forensic sci-

entists use identification by DNA typing from skeletal remains (41). Bone samples are more likely to persist due to their physical durability. It also appears that conditions within bones are relatively favorable for the preservation of DNA (41). Several authors note the importance of DNA extraction methods prior to DNA amplification and analysis to increase the percentage of identification (43-46). Furthermore, it has been suggested that mtDNA is the method of choice when someone is working with highly degraded evidence. However, following the modification of the standard DNA extraction procedure, and repurifying the extracted DNA with NaOH, we found that the percentage of successful identifications using genomic DNA increased up to 85% (47-49). Also, we observed that dental samples yielded enhanced results in 20-30% of cases when compared with results obtained from the long bones (47).

Identification of Remains Using Parents

The material presented in the sections on paternity and maternity have to do with the identification of one or more parents of a living individual. When we have unidentified remains of an individual, it is still possible to identify them if either the parents or the children of the suspected individual are available for testing. There are two sets of circumstances that can lead to the situation where it is necessary to establish identity by reconstruction. The first is that of a missing person, suspected to be dead but no body recovered (e.g., when a blood stain is found in someone's apartment, the person is missing and no body is recovered). Does the bloodstain found at the suspected crime scene belong to the person who lived

Table 2. Formulas for the calculation of the system index (SI) with a parent missing (39)^a

Child's alleles	Parent's alleles	SI
A,B	B,C	1/4b
A,B	A	1/2a
A	A,B	1/2a
A	A	1/a
A,B	A,B	(a+b)/4ab

^aA, B, and C are alleles, and a and b are allele frequencies.

Table 3. Calculation of paternity with the mother missing^a

Loci	Child	Obligate allele	Father	p (Y)	RMNE	SI
D3S1358	15,18	15 18	16,18	0.2463 ^{FBI} 0.1400	0.6234	1.786
VWA	13,16	13 16	13,16	0.0100 0.2015 ^{FBI}	0.3783	26.241
FGA	21,22	21 22	21,22	0.1735 ^{FBI} 0.1575	0.5524	3.028
THO1	8,9	8 9	7,8	0.1150 0.1650 ^{FBI}	0.4816	2.174
TPOX	6,11	6 11	11	0.0112 ^{FBI*} 0.2675	0.4797	1.869
CSF1PO	10,11	10 11	10,12	0.2750 0.3005 ^{FBI}	0.8198	0.909
D5S818	10,12	10 12	11,12	0.0487 ^{FBI} 0.3875	0.6821	0.645
D13S317	8,11	8 11	8,12	0.1550 0.3189 ^{FBI}	0.7232	1.613
D7S820	8,12	8 12	9,12	0.1626 ^{FBI} 0.1400	0.5136	1.786
AMEL	X,Y		X,Y			
Combined					0.0063	974
Exclusion power					99.375%	
Likelihood					160	
Probability of paternity					99.379%	99.897%

^aAbbreviations and symbols: p (Y) = allele frequency for each allele or obligate allele based on Croatian and FBI database. FBI indicates frequencies taken from the European FBI database (38); asterisk indicates "minimum allele frequency" which is a conservative estimate. RMNE (random man not excluded) = the proportion of the population that could contribute all of the obligate alleles and therefore could not be excluded. SI (system index) with a parent missing is calculated according to formulas given in Table 2. A "minimum allele frequency" is a conservative estimate. A discussion of the calculation of minimum allele frequencies is found in ref. 38. Obligate alleles are written in bold.

there and who is the child of two known individuals? The second circumstance occurs when remains are recovered, whether a single homicide or from a mass grave, which cannot be identified by non-genetic methods.

In the process of identifying remains, the calculation is a little bit different when compared with standard paternity identification. Rather than determining the frequency of a single obligate allele in the population of potential fathers or mother (RMNE or RFNE), it is necessary to determine the frequency of one obligate allele in each of two parents. Going back to the rule of combining probabilities or likelihoods, the probability of getting one parent with allele A and a second parent with allele B, is the frequency of possible parents with allele A times the frequency of possible parents with allele B. In practice, this is done by obtaining the RMNE/RFNE for each allele and then multiplying the respective RMNE for the first allele times the RFNE for the second allele to obtain the likelihood of getting two parents that have those alleles. This is the Random Parents Not Excluded (RPNE). Thus the formula for calculating RPNE for any two alleles, whether alike or different would be $(p^2+2p[1-p])x(q^2+2q[1-q])$. For example, using the locus

D3S1358 and a stain of the type D3S1358 15,17, with the frequencies of 0.2425 and 0.1875, respectively, then the likelihood of finding two parents, one with a D3S1358*15 allele and one with a D3S1358*17 allele by chance would be:

$$(0.2425^2+2\cdot 0.2425\cdot 0.7575)\cdot (0.1875^2+2\cdot 0.1875\cdot 0.8125)=0.4262\cdot 0.3398=0.1448.$$

Thus about 14.5% of Croatian parents would not be excluded. On the other hand, 85.5% of Croatian parents would be excluded with this single test. By multiplying all of the individual RPNEs, the combined RPNE or CRPNE is created, generating the relative population frequency or percentage of population of potential parents. Similarly, the likelihood in favor of the non-excluded parents is 1/CRPNE. The Bayesian probability of parenthood, using the non-exclusion method and a prior probability of 0.5 is 1-CRPNE.

Table 5 delineates the identification of remains from Bosnia before STR technology was available (48). In this case, the population of potential parents was calculated for each allele in the RMNE column and multiplied to-

Table 4. Calculation of maternity with the father missing^a

Loci	Child	Obligate allele	Mother	p (Y)	RMNE	SI
D3S1358	15,18	15 18	15,16	0.2463 ^{FBI} 0.1400	0.6234	1.015
VWA	13,16	13 16	16,19	0.0100 0.2015 ^{FBI}	0.3783	1.241
FGA	21,22	21 22	20,21	0.1735 ^{FBI} 0.1575	0.5524	1.441
THO1	8,9	8 9	9	0.1150 0.1650 ^{FBI}	0.4816	3.030
TPOX	6,11	6 11	6,7	0.0112 ^{FBI*} 0.2675	0.4797	22.321
CSF1PO	10,11	10 11	11,13	0.2750 0.3005 ^{FBI}	0.8198	0.832
D5S818	10,12	10 12	10,12	0.0487 ^{FBI} 0.3875	0.6821	5.779
D13S317	8,11	8 11	11	0.1550 0.3189 ^{FBI}	0.7232	1.568
D7S820	8,12	8 12	8,10	0.1626 ^{FBI} 0.1400	0.5136	1.538
AMEL	X,Y		X,X			
Combined					0.0063	1,422
Exclusion power					99.375%	
Likelihood					160	
Probability of maternity					99.3793%	99.93%

^aAbbreviations and symbols: p (Y) = allele frequency for each allele or obligate allele based on Croatian and FBI database. FBI indicates frequencies taken from the European FBI database (38); asterisk indicates "minimum allele frequency" which is a conservative estimate. SI (system index) with a parent missing is calculated according to formulas given in Table 2. Asterisk indicates "minimum allele frequency" used. A "minimum allele frequency" is a conservative estimate. A discussion of the calculation of minimum allele frequencies is found in ref. 38. Obligate alleles are written in bold.

gether to create the RPNE for each locus. The data presented is based on testing for DQA1 and Polymarker, which provides less information than the STR loci (34,49,50). Thus the values are lower than those observed in the previous tables. In this case, only 0.4% of Croatian couples could be the parents of these remains, or 99.62% would be excluded. This converts to a likelihood of approximately 260 to 1 in favor of these parents being the parents of the remains, and that there is a 99.62% certainty of paternity by non-exclusion.

Identification of Remains Using Children

If the missing person is the father of children, a regular paternity examination can be performed as above. In Table 6, NN, whose remains were found, is suspected to be the father of a child. When the mother is also available, this becomes a highly informative situation. The testing of the mother and child indicate that 99.997% of Croatian non-father males would be excluded but the remains NN were not excluded. Calculation of the SI and PI indicates that the remains are approximately 45,000 times as likely to be the boy's father than an unrelated Croatian male, or by transmission, there is a 99.998% probability that this is the child's father. By

non-exclusion there is 32,000 to 1 likelihood and a 99.9997% probability of paternity. Again, the two methods are in good agreement, and it is reasonably certain that these remains represent the father of the child in question.

Additional reading on the application of DNA technology to forensic parentage testing can be found in Schanfield (34).

Parentage Identification versus Forensic Identification

In parentage testing, the frequency of potential parents is determined by the RMNE frequency or the population of potential allele donors. In forensic identity testing the equivalent population statistics is the frequency of potential phenotype (genotype) donors (34). Table 7 compares the results of a paternity identification ($RMNE = (p^2 + 2p[1-p])$) versus a forensic identification ($2pq$). In this case the population of donors (RMNE) is approximately seven times larger than the forensic population frequency. Thus, at one locus there is almost an order of magnitude difference between the forensic identification and the paternity identification.

Table 5. Identification of remains using parents(48)^a

Locus	Mother	Bone	Father	Obligate alleles	p	RMNE	RPNE
LDLR	B	<u>A</u> , B	<u>A</u>	A	0.410	0.6519	0.5423
				B	0.590	0.8319	
GYPA	<u>A</u> , B	<u>A</u> , B	<u>A</u> , B	A	0.560	0.8064	0.5535
				B	0.440	0.6864	
HBGG	A	<u>A</u> , B	<u>B</u>	A	0.530	0.7791	0.5603
				B	0.470	0.7191	
D7S8	A	<u>A</u> , B	<u>A</u> , B	A	0.653	0.8796	0.5045
				B	0.347	0.5736	
GC	<u>A</u> , C	<u>A</u> , C	<u>A</u> , C	A	0.279	0.4802	0.4014
				C	0.595	0.8360	
DQA1	2 , 4	<u>3</u> , 4	<u>2</u> , 3	3	0.105	0.1990	0.1123
				4	0.340	0.5644	
Combined							0.0038
Likelihood of parenthood							261
Probability of parenthood							99.62%

^aAbbreviations and symbols: p = allele frequency for obligate allele based on the Croatian database (49). RMNE (random man not excluded) = the proportion of the population that could contribute all of the obligate alleles and therefore could not be excluded. RPNE (random parents not excluded) = percentage of parents that would not be excluded as a potential parents. Obligate alleles are written in bold, and alleles presumably transmitted by father are underlined.

The estimates in Table 7 are based on a heterozygous phenotype. If the results were homozygous, the Hardy-Weinberg equilibrium estimator would be p^2 . This assumes that there are no significant deviations from Hardy-Weinberg equilibrium due to substructuring or other events. In the USA, the National Research Council recommended that to be conservative a substructuring correction should be added to all populations, even those without documented sub structuring, such as Croatia. The Hardy-Weinberg equilibrium estimator for a homozygous individual then becomes $p^2+p(1-p)(0.01)$. In the example in Table 7, if the evidence is D3S1358 15, the Hardy-Weinberg equilibrium estimator would be $0.2425^2+2 \cdot 0.2425 \cdot 0.7575 \cdot 0.01=0.0606$. The RMNE for a homozygous individual is $p^2+2p(1-p)$. In the example in Table 6 the RMNE is $0.2425^2+2 \cdot 0.2425 \cdot 0.7575=0.4262$. It is apparent that the ability to individualize an item of evidence is much easier than to individualize a parent.

The last step in the process, as in parentage testing, is to combine all of the individual values to obtain the frequency of the multi-locus genotypic array or profile. Multiplying each of the individual locus genotypic frequencies does this.

Forensic Identity Testing

The objective of forensic identity testing is to compare an evidentiary stain (blood, body fluid or tissue) to a victim or suspect. If enough testing is performed, only the biological donor or his/her identical twin will not be excluded. When an item of evidence and a potential donor are consistent with each other, the likelihood of a coincidental match is calculated to provide a weight to the evidence. For example, before DNA testing, ABO blood types were used in sexual assault cases. If the victim is an O secretor, and the evidence collected is from an A secretor, then the donor had to be an A

secretor. A suspect was tested and found to be an A secretor, and thus cannot be excluded. Is this useful information or in the legal sense probative evidence? Since approximately 40% of the European population secrete blood group A in their body fluids, this is not a very useful or probative piece of evidence. With DNA evidence only a minute segment of the population will normally match by chance. In parentage testing, the frequency of donors of an allele is the significant value. For identity testing, the expected population genotype frequencies are the population statistics of interest (51). The Hardy-Weinberg equilibrium as described above determines the expected genotype frequencies.

In Table 8, the remains of NN are treated as an item of forensic evidence rather than a forensic paternity. Using the nine loci, only 1 in 22 billion Croatian males would match by chance. Since there are not 22 billion Croatians or even people on the Earth, there is a reasonable certainty that the sample could only have come from this individual if a forensic comparison had been made. Comparing the population frequency in Table 8 to that in Table 6 indicates that there is approximately 70,000 times more individualization power in forensic identity than parentage testing.

Individualization and Identity

Individualization is the concept that a genetic profile represents a unique individual. This has been indicated in the realm of fingerprint identification for many years, in which examiners will state that a fingerprint originated from a specific individual. Even identical twins do not have identical fingerprint patterns, although they will always share genetic profiles. At the present time, with the number of markers available, it is possible to generate virtually any level of individualization. The FBI recently stated that if the level of individualization

Table 6. Identification of remains using a child of the decedent^a

Loci	Wife	Child	Obligate allele	Remains NN	p	RMNE	X	SI
D3S1358	14,16	15,16	15	15,17	0.2425	0.4262	0.50	2.062
VWA	15,16	16,19	19	19	0.0900	0.1719	1.00	11.111
FGA	21,22	21,22	21	21,22	0.2125			
			22		0.1575	0.6031	1.00	2.703
THO1	7,8	7,8	7	8,9	0.1724 ^{FBI}	0.4922		
			8		0.1150		0.50	1.740
TPOX	11	10,11	10	9,10	0.0625	0.1211	0.50	8.000
CSF1PO	10,12	10	10	10,11	0.2750	0.4744	0.50	1.818
D5S818	11,12	10,12	10	10,12	0.0775	0.1490	0.50	6.452
D13S317	8,12	8,11	11	11,12	0.3350	0.5578	0.50	1.493
D7S820	8,11	8	8	8,10	0.1650	0.3028	0.50	3.030
AMEL	X,X	X,Y		X,Y				
Combined						3.14 x 10 ⁻⁵		45,720
Exclusion power						99.9969 %		
Likelihood						31,812		
Probability of paternity						99.9969%		99.9978%

^aAbbreviations and symbols: p (Y) = allele frequency for each allele or obligate allele based on Croatian and FBI database. FBI indicates frequencies taken from the European FBI database (38). RMNE (random man not excluded) = the proportion of the population that could contribute all of the obligate alleles and therefore could not be excluded. X = likelihood that the alleged father can transmit the obligate allele and it is determined by the father's type. SI (system index) = likelihood that the alleged father can transmit the allele divided by the frequency of the obligate allele (SI=X/p). Obligate alleles are written in bold.

exceeds 360 billion then a state of identity exists, such that they will report that a biological specimen originated from a specific individual or his/her identical twin. In the profile presented in Table 8, 360 billion was not reached but it would have been if additional loci were tested. The CODIS 13 would certainly have reached that level, or even changing that test kit may have reached that level (51). Does that mean that the sample was not individualized to the point of identity? It depends on the definition one uses. The definition by the FBI is to some extent arbitrary. The level of individualization needed to reach identity is something that needs to be determined at a local level, unless there is an international consensus. The idea that the likelihood of a coincidental match is less than 1 in 23 billion Croatsians is strong evidence that the profiles originated from a common source.

Acknowledgments

We acknowledge the assistance of Gregory LaBerge, Denver Police Department, Denver, CO, USA, and Patricia McCarthy, Cardinal Strich University, Milwaukee, WI, USA, presently consulting educational psychologist, for editorial assistance and comments in the preparation of this manuscript. We also thank Dr M. Kubat for allowing us to use a part of his unpublished data on 9 STRs loci for the Croatian population.

References

- Watson JD, Crick FHC. A structure for deoxyribose nucleic acid. *Nature* 1953;171:737-8.
- Inman K, Rudin N, editors. An introduction to forensic DNA analysis. New York (NY): CRC Press; 1997.
- Budowle B, Monson KL, Guisti AL, Brown BL. The assessment of frequency estimates of Hae-III generated VNTR profiles in various reference databases. *J Forensic Sci* 1994;39:319-52.
- Federal Bureau of Investigation. The application of forensic DNA testing to solve violent crimes. Washington, DC: US Department of Justice; 1990.
- Budowle B, Chakraborty R, Guisti AM, Eisenberg AJ, Alen RC. Analysis of the VNTR locus D1S80 by the PCR followed by high resolution PAGE. *Am J Hum Genet* 1991; 48:137-44.
- Latorra D, Stern MC, Schanfield SM. Characterization of human AFLP systems apolipoprotein B, phenylalanine hydroxylase and D1S80. *Genome Research* 1994;3:351-8.
- Budowle B, Baechtel FS, Comey CT. Some considerations for use of AMP-FLPs for identity testing. In: Ritter C, Schneider PM, editors. *Advances in forensic haemogenetics*. New York (NY): Springer Verlag; 1992. p. 11-7.
- Jobling MA, Heyer E, Dieltjes P, de Knijff P. Y-chromosome-specific microsatellite mutation rates re-examined using a minisatellite, MSY1 *Hum Mol Gen* 1999;8:2117-20.
- Prinz M. Experience with using Y chromosome specific STRs in forensic casework. In: Zabit C, editor. *Proceedings of the 10th International Symposium on Human Identification*; 1999 Sep 29–Oct 2; Lake Buena Vista, FL, USA. Madison (WI): Promega Corporation. In press 2000.
- Shows TB, McAlpine PJ, Bouchix C. Guidelines for human gene nomenclature. An international system for human gene nomenclature (ISGN, 1987). *Cytogenet Cell Genet* 1987;46:11-28.
- 1991 Report concerning recommendations of the DNA Commission of the International Society for Forensic Haemogenetics relating to the use of DNA polymorphism. *Vox-Sanguinis* 1992;63:70-3.
- Mayr WR. Recommendations of the DNA Commission of the International Society for Forensic Haemogenetics relating to the use of PCR-based polymorphisms. *Vox Sanguinis* 1993;64:124-6.

Table 7. Comparison of population frequency estimates for a forensic paternity case versus a forensic identity case for the locus D13S1358

	RMNE ^a (paternity)	Forensic evidence
Allele:	15;17	15;17
Frequency:	$(p+q)^2 + 2(p+q)(1-p-q)$	$2pq$
	0.6751	0.0909

^aRMNE (random man not excluded)=the proportion of the population that could contribute all of the obligate alleles and therefore could not be excluded. p is the frequency of allele 15, and q is frequency of allele 17. Frequencies for the D13S1358 15 locus=0.2425, and for D13S1358 17 locus=0.1875.

13 Wyman A, White R. A highly polymorphic locus in human DNA. *Proc Natl Acad Sci USA* 1980;77: 6754-8.

14 Wallace DC. Diseases of the mitochondrial DNA. *Annu Rev Biochem* 1992;61:1179-212.

15 Holland MM, Fisher DL, Mitchell LG, Rodriguez WC, Canik JJ, Merrill CR, et al. Mitochondrial DNA sequence analysis of human skeletal remains: identification of remains from the Vietnam War. *J Forensic Sci* 1993;38:542-53.

16 Holland MM, Parsons TJ. Mitochondrial DNA sequence analysis – validation and use for forensic casework. *Forensic Sci Rev* 1999;11:22-50.

17 Ivanov PL, Wadhams MJ, Roby RK, Holland MM, Weedn VW, Parsons T. Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov, establishes the authenticity of the remains of Tzar Nicholas II. *Nat Genet* 1996;12:417-20.

18 Lee HC, Ladd C, Scherzinger CA, Bourke MT. Forensic applications of DNA typing. Part 2: Collection and preservation of evidence. *Am J Forensic Med Path* 1998;19:10-8.

19 Jeffreys AJ, Thein SL, Wilson V. Individual specific “fingerprints” of human DNA. *Nature* 1985;316:76-9.

20 Jeffreys AJ, Wilson V, Thein SL. Hypervariable “minisatellite” regions in human DNA. *Nature* 1985;314:67-73.

21 National Research Council, National Academy of Sciences. The evaluation of forensic DNA evidence. Washington (DC): National Academy Press; 1996.

22 Wayne J. DNA: RFLP. In: Siegel JA, Saukko PJ, Knupfer GC, editors. *Encyclopedia of forensic sciences*. London: Academic Press. In press 2000.

23 Reynolds R, Sensabaugh G, Blake E. Analysis of genetic markers in forensic DNA samples using the polymerase chain reaction. *Anal Chem* 1991;63:1-15.

24 Hammond HA, Jin L, Hong Y, Caskey CT, Chakraborty R. Evaluation of 13 short tandem repeat loci for use in personal identification applications. *Am J Hum Genet* 1994;55:175-89.

25 Alford RL, Hammond HA, Coto I, Caskey CT. Rapid and efficient resolution of parentage by amplification of short tandem repeats. *Am J Hum Genet* 1994;55:190-5.

26 McEwen JE. Forensic DNA data banking by state crime laboratories. *Am J Hum Genet* 1995;56:1487-92.

27 Buel E, Schwartz MB, LaFountain MJ. Capillary electrophoresis STR analysis: comparison to gel-based systems. *J Forensic Sci* 1998;43:164-70.

28 Schanfield, M. DNA: PCR-STR. In: Siegel JA, Saukko PJ, Knupfer GC, editors. *Encyclopedia of forensic sciences*. London: Academic Press. In press 2000.

29 Weir BS. Population genetics of DNA profiles. *J Forensic Sci* 1993;33:218-25.

30 Title 28, United States Court of Appeals (1979). Federal rules of evidence as amended to February 1, 1979 (1980).

31 Kelly v. State, 824 S.W. 2d 568, (Tex. Crim. App 1992) affirming 792 S.W. 2d 579, (Tex. Ct. App 1990), Houston, TX, USA, 1992.

32 Council of Europe. Committee of Minister’s. Recommendation No. R (92) 1 on the use of analysis of deoxyribonucleic acid (DNA) within the framework of the criminal justice system. 470th meeting of the Minister’s Deputies (1992).

Table 8. Forensic identification using bone NN as example^a

Locus	NN	p	q	Formula	Population frequency
D3S1358	15,17	0.2425	0.21182	$2pq$	0.1027
VWA	19	0.0900		$p^2+0.01p(1-p)$	0.0606
FGA	21,22	0.2125	0.1575	$2pq$	0.0669
THO1	8,9	0.1150	0.1650	$2pq$	0.0380
TPOX	9,10	0.1232	0.0625	$2pq$	0.0154
CSF1PO	10,11	0.2750	0.3005	$2pq$	0.1653
D5S818	10,12	0.0775	0.3538	$2pq$	0.0548
D13S317	11,12	0.335	0.30867	$2pq$	0.2068
D7S820	8,10	0.165	0.29064	$2pq$	0.0959
Amelogenin	X,Y				
Combined					4.38078 X10 ⁻¹¹
Likelihood					22,826,996,000

^ap or q = allele frequency for each allele based on Croatian and FBI database. Frequencies taken from the European FBI database (51) are in bold.

- 33 Pavišić B, Vučković M, Veić P, Radolović A, editors. *Zakon o kaznenom postupku*. Zagreb: AKD Hrvatski Tiskarski Zavod; 1998. p. 248-53.
- 34 Schanfield, M. DNA: parentage testing. In: Siegel JA, Saukko PJ, Knupfer GC, editors. *Encyclopedia of forensic sciences*. London: Academic Press. In press 2000.
- 35 Allen R, editor. AABB Annual Meeting. Paternity testing, specialist interested group. Annual report. San Francisco (CA): AABB; 1999.
- 36 Weiner AS. Likelihood of parentage. In: Sussman LN, editor. *Paternity testing in blood grouping*. Springfield (VA): Charles C. Thomas; 1976. p. 124-31.
- 37 Li CC, Chakravarti A. An expository review of two methods of calculating a paternity probability. *Am J Hum Genet* 1988;43:197-205.
- 38 Budowle B, Moretti TR, Baumstark AL, Defebaugh DA, Keys KM. Population data on the thirteen CODIS core short tandem repeat loci in African Americans, U.S. Caucasians, Hispanics, Bahamians, Jamaicans, and Trinidadians. *J Forensic Sci* 1999; 44:1277-86.
- 39 Brenner C. A note on paternity computation in cases lacking a mother. *Transfusion* 1993;33:51-4.
- 40 Gunby P. Medical team seeks to identify human remains from mass graves of war in former Yugoslavia. *JAMA* 1994;272:1804-6.
- 41 Gaensslen RE, Lee CH. Genetic markers in human bone tissue. *Forensic Sci Rev* 1990;2:126-46.
- 42 Parson JT, Weeden WV. Preservation and recovery of DNA in postmortem specimens and trace samples. In: Haglund DW, Sorg HM, editors. *Forensic taphonomy: the postmortem fate of human remains*. New York-London: CRC Press; 1997. p. 109-38.
- 43 Fisher DL, Holland MM, Mitchell L, Sledzik PS, Wilcox AW, Wadhams M, et al. Extraction, evaluation and amplification of DNA from decalcified and un-decalcified United States civil war bone. *J Forensic Sci* 1993;38:60-8.
- 44 Ortner DJ, Tuross N, Stix AI. New approaches to the study of disease in archeological New World populations. *Hum Biol* 1992;64:337-60.
- 45 Hochmeister MN, Budowle B, Borer UV, Eggmann U, Comey CT, Dirnhofer R. Typing of deoxyribonucleic acid (DNA) extracted from compact bone from human remains. *J Forensic Sci* 1991;36:1649-61.
- 46 Burgi SB, editor. *First European-American Intensive Course in PCR Based Clinical and Forensic Testing*. Laboratory manual. Split, Croatia: Genomyx; 1997.
- 47 Primorac D, Anđelinović Š. Identification of human remains from mass graves found in Croatia and Bosnia and Herzegovina. In: Zabit C, editor. *Proceedings of the 10th International Symposium on Human Identification*; 1999 Sep 29–Oct 2; Lake Buena Vista, FL, USA. Madison, WI: Omega Corporation. In press 2000.
- 48 Primorac D, Anđelinović S, Definis-Gojanović M, Drmić I, Rezić B, Schanfield MS, et al. Identification of war victims from mass graves in Croatia, Bosnia and Herzegovina by the use of standard forensic methods and DNA typing. *J Forensic Sci* 1996;41:891-4.
- 49 Keys KM, Budowle B, Anđelinović S, Definis-Gojanović M, Drmić I, Marcikić M, et al. Northern and Southern Croatian population data on seven PCR-based loci. *Forensic Sci Int* 1996; 81:191-9.
- 50 Gill P, Evett I. Population genetics of short tandem repeat (STR) loci. In: Weir B, editor. *Human identification: the use of DNA markers*. Dordrecht: Kluwer Publ; 1995. p. 69-87.
- 51 Budowle B, Monson KL, Chakraborty R. Estimating minimum allele frequencies for DNA profile frequency estimates for PCR-based loci. *Int J Legal Med* 1996;108:173-6.

Received: December 23, 1999

Accepted: January 21, 2000

Correspondence to:

Dragan Primorac
 Laboratory for Clinical and Forensic Genetics
 Split University Hospital
 Spinčićeva 1
 21000 Split, Croatia
dragan.primorac@st.tel.hr

ESOP European Society of Pharmacovigilance
The 8th Annual Meeting of Pharmacovigilance
September, 21-23 2000
Verona — Italy

For further information please contact:
Professor G.P. Velo

Clinical Pharmacology Research Unit
Policlinico Borgo Roma
University of Verona
37134 Verona, Italy

Phone: +39 045 80986111 / 8074899
Fax: +39 045 581111
e-mail: esop@sfm.univr.it
web site: www.sfm.univr.it/esop
