



DNA Markers Evaluation to Preparing a Genetic Database for Identification in Crimes and Incidents

Ali Mohammadi^{1*}, Maryam Ramezani¹

¹Human Genetics Research Center, Baqiyatallah University of Medical Sciences Tehran, Iran

Corresponding Author: Ali Mohammadi, Human Genetics Research Center, Baqiyatallah University of Medical Sciences Tehran, Iran. Tell: +98-2186034956, Email: Sepehre3293@gmail.com

Received April 2, 2018; Accepted April 30, 2018; Online Published June 25, 2018

Abstract

Natural disasters and man-made incidents take the lives of many individuals and leave them unable to be identified. In such cases, the only possible way to detect an unidentified body or its parts is through genetic markers. This technology is based on the theory that inter-individual differences result from differences in the genetic information held in their DNA. A genetic database covering a target population can be used whenever necessary to identify otherwise unidentifiable individuals. A total of 99% of the DNA of every human is the same; in fact, only a relatively small amount is different. Genetic tests can be carried out to distinguish individuals from each other. One percent of the variable genomic regions that frequently and sporadically emerge are genetic markers. Several types of sequences, including variable number of tandem repeats (VNTRs), short tandem repeats (STRs), single nucleotide polymorphisms (SNPs), Y chromosomes, and mitochondrial DNAs (mtDNAs) at the DNA level, can be used to identify individuals. Each of these sequences has its own capabilities and limitations and can be used by anyone depending on the type of database application. With a genetic database of individuals working in high-risk occupations such as the armed forces and the software-based management of this information, it will be possible to identify genetically the bodies of unidentifiable victims of accidents and incidents. This study introduces and evaluates the sequences that can be used in genetic databases.

Keywords: DNA Marker, VNTRs, STRs, SNPs, Y Chromosomes, mtDNA

Citation: Mohammadi A, Ramezani M. DNA markers evaluation to preparing a genetic database for identification in crimes and incidents. Int J Med Rev. 2018;5(2):50-54. doi:10.29252/IJMR-050202.

Introduction

Natural disasters like floods and earthquakes as well as man-made ones like wars, accidents, and plane crashes, may take the lives of many people and leave their bodies identifiable; genetic techniques can be used to identify them. Genetic identification is also one way to identify a criminal from crime scenes. Unidentified bodies can be identified using their genetic information and matching it with the genetic database or with the victims' parents, children, brothers, or sisters. This technology is based on the theory that inter-individual differences result from differences in the genetic information held in their DNA. Therefore, a genetic database for a target population would prevent non-identification of missing persons whenever necessary.¹⁻³ A total of 99% of every person's DNA is the same; in fact, only one nucleotide per 1000 nucleotides is different in each individual. Considering this difference, genetic tests can be performed to distinguish individuals from each other.³⁻⁶ Variable genomic regions that appear frequently and sporadically are introduced as genetic markers in various forms. Several types of these DNA

markers can be used for identification purposes, and the most important examples of them are discussed below.

Variable Number of Tandem Repeats

Variable number of tandem repeats (VNTRs) are DNA fragments comprised of repetitive units made up of 7-25 nucleotides which are interconnected. Their central nucleus is repeated 50 times and is 1 kb in length. This marker gets its name from the difference in the number of repetitive units. Technically, this marker is suitable for artificial probes, the application of radioactive materials, and the southern blot. The most important advantage of using VNTRs for identification purposes is the high polymorphism of these regions, which has resulted in the formation of multiple alleles in a single locus (4-7). VNTR polymorphisms are often located near telomeres, and their repetitive nucleus consists of 6 to 100 bp. The nucleus sequence in several of these alleles is repeated thousands of times. The variable number of repetitions creates alleles ranging in size from 500 to 30000 bp. VNTRs were the first polymorphisms used to create DNA profiles.

After some time, however, their use was limited by the need for large amounts of DNA with high molecular weight and the difficulty of interpreting their results. For these reasons, short tandem repeats (STRs) are now used instead of VNTRs. (Figure 1)^{5,6}

Mitochondrial DNA

Unlike nuclear DNA which is passed in equal parts to a child from both parents, mitochondria are inherited solely from mothers, because the sperm cell has no mitochondrion. Thus, the mitochondria are inherited only through the egg to the egg cell. Therefore, mitochondrial DNA (mtDNA) is investigated only to study the maternal inheritance. Since mtDNA is transmitted unchanged (except in the case of mutations) from mothers to children and from daughters to the next generation, one can determine the genetic relationship between a grandmother and a granddaughter without a mother.² Currently, mtDNA is widely used in identification science for two reasons. First, the D-Loop region is highly polymorphic and very suitable for use in identification tests. Secondly, although mtDNA accounts for less than 1% of the cell DNA, its genes have a large number of copies. If there are approximately 200 mitochondria in each cell, there are more than 400 copies of the mitochondrial gene for each nuclear gene that may have 1 or 2 copies. Thus, considering the low levels of nuclear DNA in cases such as hair and sperm as well as bones and teeth specimens where the nuclear DNA is often degenerated, the use of mtDNA often produces good results. The mtDNA sequence is also used in archeological studies (Figure 2).^{3,4}

Single Nucleotide Polymorphism

Variations that occur due to differences in a nucleotide position are called single nucleotide polymorphisms (SNPs). They are found abundantly in the human genome; there is almost one SNP per every 400 nucleotides.² SNPs are advantageous for identification purposes. First, they have higher frequency in the genome (millions) than the STR loci (1 or 2 copies). Secondly, due to their smaller size (less than 100 BP), SNPs produce better results than STRs in studying crushed DNA specimens (such as old tissue). However, their big disadvantage is that, unlike STRs that have a lot of alleles in each locus, most SNPs are made up of two alleles. Therefore, they exhibit poor identification potential at each location, and about 30-40 SNPs should be analyzed when being used for identification purposes.^{2,7,8} SNPs are the simplest type of DNA polymorphism and are, in fact, considered as a difference in only one nucleotide sequence in the DNA sequence.^{9,10} The structure of a typical SNP is shown in Figure 3.

This figure shows two alleles that differ only in the one nucleotide marked with a star. The fourth positions in alleles G and A are guanine and adenine, respectively. In most cases, mutation in a particular locus, which causes the formation of an SNP, leads to the formation of two alleles.

SNPs are formed in cells that undergo meiosis due to the occurrence of mutations during DNA replication.¹¹ Certain genomic regions have more SNPs than other regions. For example, chromosome 1 has an average of one SNP in a

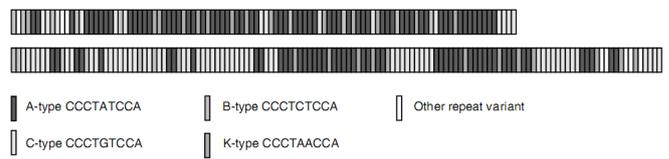


Figure 1. Structure of 2 Alleles at VNTR Locus D1S7. Both of these alleles are relatively short and have 104 and 136 repeats, respectively. The alleles in this locus can have up to 2000 repeats. The central nucleus of the alleles, which is repeated frequently, is 9 bp in length. The image above was taken from the book "An Introduction of Forensic Genetics".²

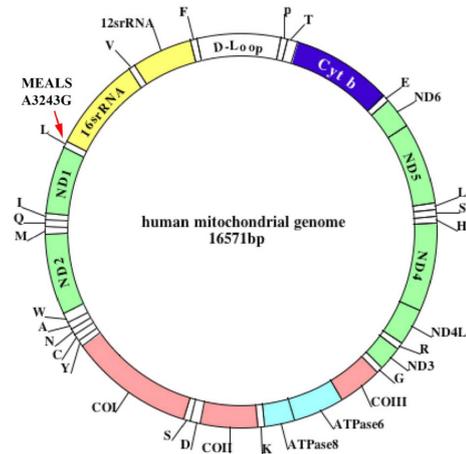


Figure 2. Structure of the Mitochondrial DNA and its D-Loop Sequence.³

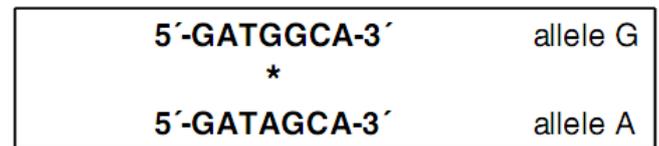


Figure 3. Single Nucleotide Polymorphism. Taken from the book "An Introduction of Forensic Genetics".²

1.45 kb region, while there is one SNP in a 2.18 kb region of chromosome 19.¹² SNPs normally have two alleles, e.g., a guanine allele and an adenine allele. Therefore, SNP is not considered to be a highly informative polymorphism, and therefore does not appear to be suitable for forensic analysis. However, SNPs are found abundantly throughout the genome, and theoretically, it is possible to identify hundreds of them, which in turn leads to a significant increase in the total resolution power. It seems that achieving a resolution of 10 STR requires 50 to 80 SNP, which is far more difficult than the analysis of 10 STR.^{2,13,14}

The Significance of SNPs in Forensic Genetics

One constant challenge in forensic work is the analysis of a severely degraded DNA sample, which should be extracted from bone and teeth specimens of individuals who have long been dead. In such cases, the use of STRs is not possible, because they have a long amplicon and do not produce good results, and alternative methods should be used, one of

which is mtDNA typing.²⁻⁵ Although the use of mtDNA has advantages, its sequence variation is lower than that of STRs and maternal inheritance limits some genetic relationships, such as the father-daughter relationship. Also, the mtDNA study is a time-consuming process.¹⁵ One way to reduce the amplicon length is to redesign the primers of the current known STRs, generally referred to as mini-STRs.¹⁶ Although STR sequences are widely used in forensic laboratories worldwide and there are many commercial kits available for such purpose, there is still a growing interest in SNPs, which could be due to their lower mutation rate and the possibility of analyzing crushed DNA specimens.^{17,18}

Short Tandem Repeats

The polymerase chain reaction (PCR)-based analysis of STRs is one of the most important methods used in identification labs. The proliferation of STRs on genomic DNA, now known as STR loci, is an integral part of research and analyses performed in identification labs. STRs are similar to VNTRs in terms of central repetitive interconnected units, but the length of these repetitive units in STRs is 2-6 bp.^{2,7,9,19}

These central repetitive units may be repeated in a locus up to 40 times, but the most common STRs are those in which repetitive units are repeated 7 to 15 times. STRs are now the most commonly analyzed genetic polymorphisms in forensic genetics. These sequences were first introduced to forensic genetics in the mid-1990s, and they are now used as the main tool in forensic laboratories.²⁰

There are thousands of STRs that can be used in forensic analysis. These sequences are scattered throughout the genome, namely, 22 autosomal chromosomes as well as chromosomes X and Y. Their nucleus consists of 1 to 6 nucleotides, and their number of repeats is generally between 50 and 300 bp. Major loci used in forensic genetics are tetranucleotide repeats.^{21,22} STRs meet the need for a forensic marker entirely. A wide range of biological materials can be analyzed using STRs, and the results obtained from their analysis in different laboratories are well comparable. STRs have very high resolution, especially when several of them are applied simultaneously. They also have high sensitivity, and a small number of cells are sufficient to ensure a successful analysis. Creating STRs profiles is easy and cost-efficient. On the other hand, there are a lot of STRs that are not under any selective pressure throughout the genome.^{2,23,24} Although a large number of STRs are known, only about 20 of them are used in forensic work.^{2,9,19,25} STR markers widely used in forensic genetics consist of nuclei of 4 or 5 bp.^{2,26}

The reparative nucleus may consist of 1 to 6 nucleotides. This example illustrates the structure of two alleles of the D8S1179 locus. The alleles are named based on the number of their repeats.

One of the essential characteristics of any STR used in forensic analysis is that it can be evaluated and analyzed in a multiplex reaction with other STRs.^{2,27,29} In the United Kingdom, the Forensic Science Service (FSS) designed the first system for determining STR type in forensic analysis. Subsequently, this system was replaced by second generation

multiplex (SGM). Two commercial companies, Applied Biosystems and Promega, produce a series of multiplexes that are already used in many laboratories. The AmpF/STR[®] SGM Plus, produced by Applied Biosystems, has now replaced SGM in the United Kingdom. In the United States, STR technology was introduced to the forensic field by selecting 13 loci as the CODIS locus (Combined DNA Index System).^{19,30} Each of these sequences has capabilities and limitations and can be used by anyone depending on the database application type. Collecting genetic information for the preparation of a genetic database includes the stages of DNA extraction, quality control, PCR, profiling, and database creation.

DNA Extraction

The first step for the preparation of a genetic database is to carry out the sampling process on the target population and perform the DNA purification steps on the obtained blood specimens. Different methods have been proposed for the purification of blood specimens, including: (1) Organic purification, (2) DNA purification devices, (3) FTA cards, (4) salting out method, (5) phenol-chloroform.²⁻¹⁰

Quality Control

Specimens should undergo quantitative and qualitative control after performing the purification and before the PCR test. Nano Photometer can be used to perform quantitative and qualitative control. This device can detect very low levels of DNA, RNA, and protein by ultraviolet light spectroscopy. Another quantitative DNA testing method is to use the real time PCR method. The quantitative control can also be accomplished using 1% agarose gel electrophoresis.²⁻²⁰

Polymerase Chain Reaction

The number of intended sequences (here, STR) is increased after the PCR-based quality control step is performed. Many kits are produced by various companies for such purpose, including the AmpFLSTR Identifier PCR Amplification kit.²⁰⁻³⁰

Profiling: Specimens are run on the genetic device analyzer. This system is based on the capillary electrophoresis and argon laser. The device is capable of detecting STR loci using an argon laser with maximum excitation of 488 and 514 nm.²⁰⁻²⁵

Database

The last step includes the preparation of the genetic database from specimen profiles. It is composed of 2 parts. The first part includes the genetic database of known specimens prepared from individuals working in high-risk occupations. The second part relates to specimens from unidentified bodies that are taken from accident scenes. The software is able to match the unknown specimens' profiles with known ones and identify the specimens collected from the incident scene.²⁵⁻³⁰

Conclusions

Based on the foregoing, it is recommended that genetic

information from blood specimens obtained from a target population be extracted and stored as a genetic database. Then, in case of accidents, the genetic information of the unknown specimens should be extracted and matched with the genetic database. This method is currently being implemented for the entire population in some countries around the world. In Iran, there is no genetic database. Thus, DNA specimens extracted from an unidentified specimen must be matched with specimens of all families of missing individuals, and this is a very time-consuming process. Moreover, all family members may have died along with the unidentified individual; consequently, the specimen cannot be identified at all. Therefore, preparing a genetic database for individuals working in high-risk occupations, such as those in the armed forces, seems necessary so as to prevent non-identification. To this end, several types of genetic markers can be used, the most significant of which include VNTRs, STRs, SNPs, mtDNAs, and Y chromosomes. Considering limitations and capabilities as well as the type of database application, one can use each of these markers.

The advantages of a genetic database include:

1. A small amount of tissue specimen is sufficient for determining an individual's identity.
2. Using this method, bone specimens discovered long ago can be used for identification purposes, and this information will not be lost over time.
3. Identification is fast.
4. The genetic information of any person is always assessable and cannot be forged; thus, it is considered to be the best identification method in criminal cases.

Genetic identification is the best and most accurate method for identifying unidentified individuals; thus, the preparation of genetic databases for high-risk occupations seems to be essential.

Authors' Contributions

AM: Conceived the original idea and wrote the manuscript;
MR: Cooperation in wrote the manuscript.

Conflict of Interest Disclosures

The authors declare they have no conflicts of interest.

References

1. Prinz M, Carracedo A, Mayr WR, et al. DNA Commission of the International Society for Forensic Genetics (ISFG): recommendations regarding the role of forensic genetics for disaster victim identification (DVI). *Forensic Sci Int Genet.* 2007;1(1):3-12. doi:10.1016/j.fsigen.2006.10.003.
2. Goodwin W, Linacre A, Hadi S. *An Introduction to Forensic Genetics.* Wiley; 2011.
3. Spelbrink JN. Functional organization of mammalian mitochondrial DNA in nucleoids: history, recent developments, and future challenges. *IUBMB Life.* 2010;62(1):19-32. doi:10.1002/iub.282.
4. Nelson K, Melton T. Forensic mitochondrial DNA analysis of 116 casework skeletal samples. *J Forensic Sci.* 2007;52(3):557-561. doi:10.1111/j.1556-4029.2007.00407.x.
5. Putkonen MT, Palo JU, Cano JM, Hedman M, Sajantila A. Factors affecting the STR amplification success in poorly preserved bone samples. *Investig Genet.* 2010;1(1):9. doi:10.1186/2041-2223-1-9.
6. Loreille OM, Diegoli TM, Irwin JA, Coble MD, Parsons TJ. High efficiency DNA extraction from bone by total demineralization. *Forensic Sci Int Genet.* 2007;1(2):191-195. doi:10.1016/j.fsigen.2007.02.006.
7. Milos A, Selmanovic A, Smajlovic L, et al. Success rates of nuclear short tandem repeat typing from different skeletal elements. *Croat Med J.* 2007;48(4):486-493.
8. Marjanovic D, Durmic-Pasic A, Bakal N, et al. DNA identification of skeletal remains from the World War II mass graves uncovered in Slovenia. *Croat Med J.* 2007;48(4):513-519.
9. Vanek D, Saskova L, Koch H. Kinship and Y-chromosome analysis of 7th century human remains: novel DNA extraction and typing procedure for ancient material. *Croat Med J.* 2009;50(3):286-295. doi:10.3325/cmj.2009.50.286.
10. Kalmar T, Bachrati CZ, Marcsik A, Rasko I. A simple and efficient method for PCR amplifiable DNA extraction from ancient bones. *Nucleic Acids Res.* 2000;28(12):E67. doi:10.1093/nar/28.12.e67.
11. Loreille OM, Parr RL, McGregor KA, et al. Integrated DNA and fingerprint analyses in the identification of 60-year-old mummified human remains discovered in an Alaskan glacier. *J Forensic Sci.* 2010;55(3):813-818. doi:10.1111/j.1556-4029.2010.01356.x.
12. Wood AR, Esko T, Yang J, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet.* 2014;46(11):1173-1186. doi:10.1038/ng.3097.
13. Ehret GB, Munroe PB, Rice KM, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature.* 2011;478(7367):103-109. doi:10.1038/nature10405.
14. Estrada K, Styrkarsdottir U, Evangelou E, et al. Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. *Nat Genet.* 2012;44(5):491-501. doi:10.1038/ng.2249.
15. Dehghan A, Dupuis J, Barbalic M, et al. Meta-analysis of genome-wide association studies in >80 000 subjects identifies multiple loci for C-reactive protein levels. *Circulation.* 2011;123(7):731-738. doi:10.1161/CIRCULATIONAHA.110.948570.
16. Loth DW, Soler Artigas M, Gharib SA, et al. Genome-wide association analysis identifies six new loci associated with forced vital capacity. *Nat Genet.* 2014;46(7):669-677. doi:10.1038/ng.3011.
17. Abecasis GR, Auton A, Brooks LD, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature.* 2012;491(7422):56-65. doi:10.1038/nature11632.
18. Csete K, Beer Z, Varga T. Prenatal and newborn paternity testing with DNA analysis. *Forensic Sci Int.* 2005;147 Suppl:S57-60. doi:10.1016/j.forsciint.2004.09.101.
19. Geada H, Brito RM, Ribeiro T, Espinheira R. Portuguese population and paternity investigation studies with a multiplex PCR--the AmpFISTR Profiler Plus. *Forensic Sci Int.* 2000;108(1):31-37. doi:10.1016/S0379-0738(99)00091-2.
20. Kemp BM, Smith DG. Use of bleach to eliminate contaminating DNA from the surface of bones and teeth. *Forensic Sci Int.* 2005;154(1):53-61. doi:10.1016/j.forsciint.2004.11.017.
21. Bogdanowicz W, Allen M, Branicki W, Lembring M, Gajewska M, Kupiec T. Genetic identification of putative remains of the famous astronomer Nicolaus Copernicus. *Proc Natl Acad Sci U S A.* 2009;106(30):12279-12282. doi:10.1073/pnas.0901848106.
22. The International HapMap Project. *Nature.* 2003;426(6968):789-796. doi:10.1038/nature02168.
23. Willer CJ, Schmidt EM, Sengupta S, et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet.* 2013;45(11):1274-1283. doi:10.1038/ng.2797.
24. Smith NL, Huffman JE, Strachan DP, et al. Genetic predictors of fibrin D-dimer levels in healthy adults. *Circulation.* 2011;123(17):1864-1872. doi:10.1161/CIRCULATIONAHA.110.009480.
25. Locke AE, Kahali B, Berndt SI, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature.* 2015;518(7538):197-206. doi:10.1038/nature14177.

26. Huang J, Sabater-Lleal M, Asselbergs FW, et al. Genome-wide association study for circulating levels of PAI-1 provides novel insights into its regulation. *Blood*. 2012;120(24):4873-4881. doi:[10.1182/blood-2012-06-436188](https://doi.org/10.1182/blood-2012-06-436188).
27. Huang J, Huffman JE, Yamakuchi M, et al. Genome-wide association study for circulating tissue plasminogen activator levels and functional follow-up implicates endothelial STXBP5 and STX2. *Arterioscler Thromb Vasc Biol*. 2014;34(5):1093-1101. doi:[10.1161/atvbaha.113.302088](https://doi.org/10.1161/atvbaha.113.302088).
28. Deloukas P, Kanoni S, Willenborg C, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet*. 2013;45(1):25-33. doi:[10.1038/ng.2480](https://doi.org/10.1038/ng.2480).
29. Smith NL, Chen MH, Dehghan A, et al. Novel associations of multiple genetic loci with plasma levels of factor VII, factor VIII, and von Willebrand factor: The CHARGE (Cohorts for Heart and Aging Research in Genome Epidemiology) Consortium. *Circulation*. 2010;121(12):1382-1392. doi:[10.1161/CIRCULATIONAHA.109.869156](https://doi.org/10.1161/CIRCULATIONAHA.109.869156).
30. Zuk O, Schaffner SF, Samocha K, et al. Searching for missing heritability: designing rare variant association studies. *Proc Natl Acad Sci U S A*. 2014;111(4):E455-E464. doi:[10.1073/pnas.1322563111](https://doi.org/10.1073/pnas.1322563111).