Khalij-Libya Journal of Dental and Medical Research
http://journal.khalijedental.com.ly/ eISSN:2708-888X

Review Article

Overview of Forensic DNA Profiling and Database

Esraa Algeblawi, Moudah Kareem, Sabreen Aboujildah*

Faculty of Biotechnology, Aljafara University, Libya.
Corresponding Email: senyoritabena899@gmail.com

ABSTRACT

Deoxyribonucleic acid (DNA) profiling, has had a tremendous impact on forensic genetics. Before DNA profiling, all forensic genetic casework (e.g., Paternity testing, criminal casework, individual identification) was performed using classical serological genetic markers. Blood groups, human leukocyte antigen (HLA), and polymorphic protein and enzymes were used for solving forensic genetic casework using immunological and electrophoretic methodologies. These genetic markers were nevertheless limited when it was necessary to analyze minimal or degraded material, which is commonly involved in forensic cases. An STR is a region of human DNA containing an array of tandem repeats. Arrays range from only a 10 to about a hundred repeated units. This essay confers the basic concepts of operating of DNA in the criminal investigation. This review primarily summarizes the major tandem repeat markers used in forensic DNA profiling, that assist criminal's conviction, exonerate the inferring individuals, and recognize victims of violence, catastrophes, and armed conflict.

Keywords: DNA fingerprinting, Forensic DNA profiling, Forensic DNA database, STR.

Citation: Algeblawi E, Kareem M, Aboujildah S. Overview of Forensic DNA Profiling and Database. Khalij-Libya J Dent Med Res. 2021;5(1):23-29. https://doi.org/10.47705/kjdmr.215105
Received: 30/12/20; **accepted**: 04/03/21

INTRODUCTION

DNA fingerprinting has revolutionized the forensic science, one of the great advances in the late twentieth century. In 1985 the Eureka shout shook England and was heard around the world when Alec Jeffrey's and his team at the University of Leicester, in the UK, found extraordinarily variable and heritable patterns from repetitive DNA analyzed by multi-locus probes. Not being Holmes, he refrained to call the method after himself, but 'DNA fingerprinting'. Under this name, his invention opened up a new area of science. The technique proved applicable in many biological disciplines, namely in diversity and conservation studies among species, and in clinical and anthropological studies. But the true political and social dimension of genetic fingerprinting became apparent far beyond academic circles when the first applications in civil and criminal cases were published [1-3].

Deoxyribonucleic acid (DNA) profiling, it had a significant impact on forensic genetics. Before DNA profiling, all forensic genetic casework (e.g., Paternity testing, criminal casework, proof of the identity of the person) was performed using conventional serologic genetic markers. Blood groups, human leukocyte antigen (HLA), and polymorphic protein and

enzymes were used for solving forensic genetic casework using immunological and electrophoretic methodologies. However, genetic markers were limited when the minimum or deteriorated matter, usually involved in forensic cases, needed to be examined. It was, besides, difficult to analyze biological material other than blood, and therefore the information obtained from hair, saliva, and even semen in rape cases were rather limited [3].

This review briefly recapitulates the major tandem repeat markers used in forensic DNA analysis, which helps to convict criminals, exonerate the wrongly accused, and identify victims of crime, disasters, and war.

Variable Number Tandem Repeat (VNTR) Profiling

Modern forensic DNA analysis began with a Variable Number of Tandem Repeats (VNTR), or minisatellite techniques. First discovered in 1985 by Sir Alex Jeffrey's, these probes, when hybridized to Southern blot membranes produced highly variable banding patterns that are known as DNA fingerprints. Underlying these complex multi-branded patterns are several forms (alleles) of genetic loci that simultaneously appear in a given individual. The particular combinations of alleles in a given individual are highly specific, yet each is visible because they share a common DNA sequence motif that is recognized by the multiplex molecular probe through complementary base pairing. These multilocus probes are very individualizing, but problematic when it comes to quantifying results.

Some statistics can be calculated with multilocus data, but certain critical calculations cannot be made unless individual-locus genotype data are available. In answer to this need, a series of single-locus VNTR probe systems were developed, and these became standard in U.S. forensic labs from the late 1980s through the early 1990s [4-6].

VNTR Loci for Forensic Testing

The minisatellites are also called variable number tandem repeats (VNTRs) as shown in (Figure 1). The repeat unit length of a VNTR locus can range from 10-30 base pairs (bp), the tandem repeat arrays length can be 10 to 1000 repeat unit long. The numbers of tandem repeat units in some VNTR loci are highly variable, leading to variable lengths of DNA fragments.

A genotype is defined by a particular number of tandem repeat units at a given locus. The chromosomal location of each locus should not be linked. However, loci located far apart on the same chromosome or different chromosomes can be used. The loci selected should be compatible with the restriction endonuclease cleavage sites. Many VNTR loci used for forensic applications are highly polymorphic. Hundreds of different genotypes per locus can be observed among the population. The discriminating power of VNTR loci used for forensic testing can be measured by population match probability (Pm). The lower the Pm, the less likely a match will occur between two randomly chosen individuals (Table 1) [4].

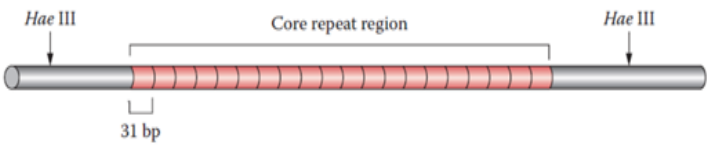

Figure 1: VNTR locus D2S44

Table 1: Commonly used VNTRs used for DNA fingerprinting.

Locus	Chromosome Location	Repeat Unit Length (bp)	HaeIII Fragment Size (kb)
D1S7	1	9	0.5–12
D2S44	2	31	0.7–8.5
D4S139	4	31	2–12
D10S28	10	33	0.4–10
D14S13	14	15	0.7–12
D16S85	16	17	0.2–5
D17S26	17	18	0.7–11
D17S79	17	38	0.5–3

Autosomal Short Tandem Repeat (STR) Profiling.

An STR is a region of human DNA containing an array of tandem repeats. Arrays range from only a 10 to about a hundred repeated units. A repeat unit can be 2 to 6 base pairs (bp) long. STRs are also called microsatellites or simple sequence repeats. The number of repeat units of STR loci can vary greatly among the population. The most commonly used STR loci are 100 to 500 bp in length—shorter than the smallest VNTRs (up to 1000 bp) [3,4].

STR Loci Commonly Used for Forensic DNA Profiling:

Many STR loci have been characterized for forensic DNA profiling (Table 2). The discriminating power of an STR locus used for forensic testing can be measured by a parameter known as population match probability (Pm). The lower the Pm, the less likely a match will occur between two randomly chosen individuals. To achieve a low Pm in forensic STR profiling, many unique characteristics of STR loci are desired. First, the STR loci should be highly variable among the population. Second, if more than one locus is selected; the loci should not be linked. The STR loci employed usually are located on different chromosomes. Loci located in the same chromosome can also be used but should be separated enough to ensure they are not linked (Figure 2 and Figure 3) [3].

Table 1: Some commonly used STR loci

Locus	Repeat Motif	Repeat Category	Chromosome Location
CSF1PO	TAGA	Simple	5q33.1
D2S1338	[TGCC] [TTCC]	Compound	2q35
D21S11	[TCTA] [TCTG]	Complex	21q21.1
Penta D	AAAGA	Simple	21q22.3
TH01	TCAT	Simple	11p15.5
VWA	[TCTG] [TCTA]	Compound	12p13.31
D18S51	AGAA	Simple	18q21.33
D13S317	TATC	Simple	13q31.1

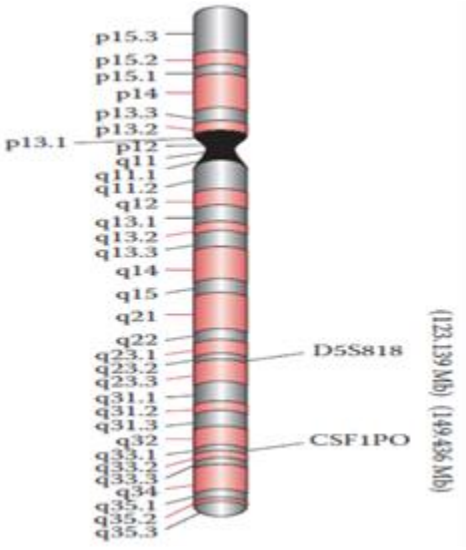

Figure 2: Cytogenetic map showing locations of STR markers on chromosome 5

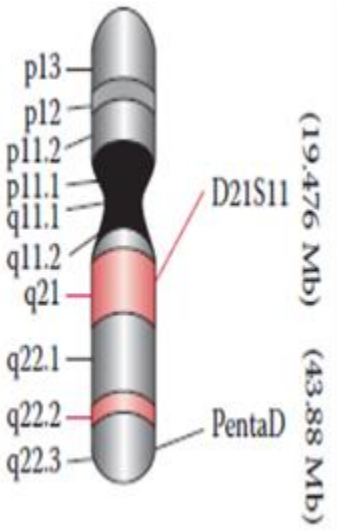

Figure 3: Cytogenetic map showing locations of STR markers on chromosome 21

Y Chromosome Profiling and Gender Typing.

The Y chromosome is inherited from the father and is passed on to all-male offspring (Figure 4).

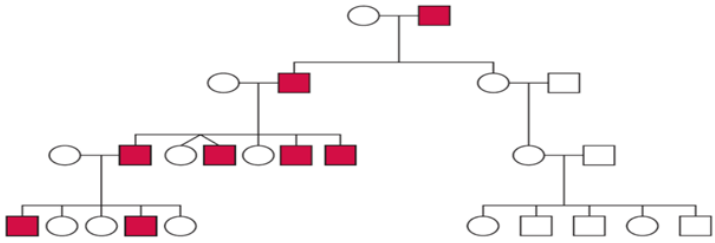

Figure 4: Human family pedigree showing inheritance of Y chromosome

Thus, the Y chromosome is unique to males. The chromosome encodes dozens of genes required for male-specific functions, including sex determination and spermatogenesis. Y chromosome loci are very important for forensic DNA profiling, for instance, the Y chromosome STR (Y-STR) used in forensic DNA testing is male-specific (for humans and certain higher primates) and is thus useful in investigations of sexual assault cases involving male suspects. The evidence gathered in such cases usually consists of mixtures of high levels of female DNA and low levels of male DNA. The Y chromosome-specific loci can be examined without interference from large amounts of female DNA; differential extraction of sperm and non-sperm cells may not be needed. Furthermore, the Y-STR system is useful for determining the numbers of male perpetrators in sexual assault cases involving more than one male. The Y-STR loci used for forensic applications are located in the non-recombining section of the Y chromosome so that paternal lineages can be established. The technique can be used for paternity testing and the identification of missing persons. Finally, data interpretation can be simplified by the use of a single allele per Y-STR locus profile. The major disadvantage of Y chromosome loci is that their discriminating power is low compared to the discriminating power of autosomal loci. Because Y chromosome loci are linked, the product rule for statistical calculations for profile probability does not apply [3].

Human Y Chromosome Genome

The human Y chromosome genome contains approximately 60 million bp and the chromosome can be divided into two regions: the pseudo-autosomal region (PAR) and the male-specific Y (MSY) region [3].

Pseudo-Autosomal Region

Approximately 5% of the Y chromosome sequences are located at the telomeres of the chromosome. In particular, PAR1 is located on the tip of the short arm and PAR2 is located at the tip of the long arm (Figure 5). This region undergoes recombination with homologous regions on the X chromosome during meiosis in males [3].

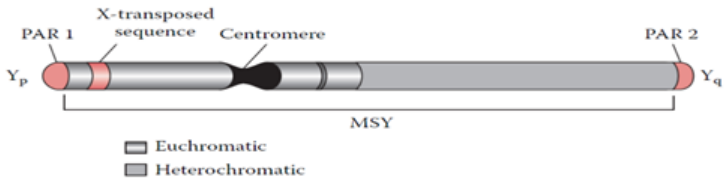

Figure: 5 Human Y-STR chromosome structure

Male-Specific Y Region

The remainder of the Y chromosome is known as the MSY region. It was previously called the non-recombining Y (NRY) region (Figure 5). It does not participate in homologous recombination. However, certain sections involve intrachromosomal gene conversion (3).

Y-STR Profiling Systems

More than 400 STR loci have been identified in the Y chromosome genome. The precise locations of these loci have been sequentially mapped using human genome sequencing data. Most Y-STR loci are located on the long arm of the chromosome; about 22% are located on the short arm and a few are found in the centromeric region.

Y-STRs in the telomeric region have yet to be identified. Only about 5% of Y-STRs are located within 5′ untranslated or intron regions of protein-coding genes. The repeat unit length of identifying Y-STRs has been analyzed. Among the 400 Y-STRs, 6% are dimeric repeats, 39% are trimeric, 45% are tetrameric, 9% are pentameric, and 1% are hexameric (Figure 6). Fewer than half the STRs have been characterized. Some loci are polymorphic and are useful for forensic applications and developing new Y-STR multiplex systems.

The STR loci on the Y chromosome are usually referred to as haplotypes. A haplotype is a collection of alleles that are usually linked (inherited together) since homologous recombination does not occur on the majority of the Y chromosome. The most commonly used Y-STR loci for forensic testing are described below (Table 3) [3].

Table 2: Common Y-STR loci

Locus	EMH	U.S. Haplotype Loci	Repeat Motif
DYS19	Yes	Yes	TAGA
DYS389 I	Yes	Yes	TCTA
DYS389 II	Yes	Yes	[TCTG][TCTA]
DYS390	Yes	Yes	[TCTG][TCTA]
DYS391	Yes	Yes	TCTA
DYS392	Yes	Yes	TAT
DYS438	No	Yes	TTTTC
DYS439	No	Yes	GATA

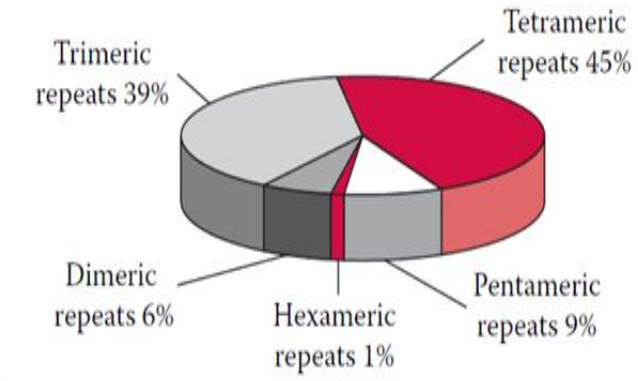

Figure 5: Human Y-STRs with different repeat unit length

Forensic DNA Databases

DNA databases were developed to help solve crimes by creating a network for exchanges of information among law enforcement agencies. More specifically, DNA databases allow forensic laboratories to compare DNA profiles electronically to identify perpetrators. The United Kingdom established the world's first national DNA database (the National DNA Database or NDNAD in 1995. It demonstrated initial success in solving crimes. Three years later, the United States introduced its national COMBINED DNA INDEX SYSTEM or CODIS. By the end of 1998, other countries (Austria, Germany, Netherlands, New Zealand, and Slovenia) had also introduced national DNA databases. Table 4 describes some national DNA databases [3].

Table 4: Characteristics of National DNA Databases

Country	Year Established	Suspect Entry Criteria	Convicted Offender Entry Criteria	Removal Criteria
United Kingdom	1995	Any recordable offense that leads to imprisonment	Entered as suspect	Never removed
New Zealand	1996	No suspects entered	Relevant offense	Never removed

			(>7 years in prison)	unless conviction quashed
Austria	1997	Any recordable offense that leads to imprisonment	Entered as suspect	Only after acquittal
Netherlands	1997	No suspects entered except when suspects DNA is tested for case	Offense leading to >4 years in prison	20 to 30 years after conviction
Germany	1998	Offense leading to> 1 year in prison	After court decision	After acquittal or 5 to 10 years after conviction if prognosis good
Slovenia	1998	Any recordable offense that leads to imprisonment	Entered as suspect	Depends on severity of crime
United States	1998	No suspects entered; under revision	Depends on state law	Depends on state law
Finland	1999	offense leading to > 1 year in prison	Entered as suspect	Only after acquittal
Sweden	2000	No suspects entered	Offense leading to >2 years in prison	10 years after release from prison
Switzerland	2000	Any recordable offense that leads to imprisonment	Entered as suspect	After acquittal or 5 to 30 years after conviction
France	2001	No suspects entered	Sexual assaults and serious crimes	40 years after conviction

The U.K. National Database

NDNAD was established in 1995 in England and Wales. Scotland and Northern Ireland have their databases, but also submit their profiles to NDNAD [3].

Index

Two types of samples are stored in the database: crime scene samples and samples were taken from offenders.

Entries

NDNAD provides that DNA samples can be taken from any individual arrested and detained in police custody in connection with a recordable offense. Most offenses (other than traffic offenses) committed must be recorded in the Police National Computer system as part of an individual's criminal record. Samples can be no intimate (typically mouth swabs) taken without consent or intimate samples (blood) taken with consent and known as criminal justice or CJ samples. The NDNAD contains the world's largest number of DNA profiles in proportion to the population—approximately 5%. As of February 2006, the NDNAD held approximately 3.4 million offender profiles and 290,000 crime scene profiles. Additionally, more than 721,495 investigations were aided with NDNAD. Six agencies in the U.K. are approved to provide DNA profiles from the offender and crime scene samples to the NDNAD. They are accredited both by the United Kingdom Accreditation Service and the Custodian Unit of the Home Office [3].

U.S. Combined DNA Index System (CODIS)

A pilot project started in the United States in 1990 included only 14 state and local agencies. The Congressional DNA Identification Act (1994) authorized the Federal Bureau of Investigation (FBI) to establish a national DNA database. By 1997, 13 STR loci were selected as the core loci for the national database, and CODIS was finally formed in 1998. Currently, all 50 states have authorized the collection of samples from convicted felons for the DNA database. However, state laws vary; all states are required to collect DNA profiles from individuals convicted of felony sex crimes while others authorize collection for additional types of felonies [3].

Infrastructure

CODIS is a database with three hierarchical levels: The Local DNA Index System (LDIS), the State DNA Index System (SDIS), and the National DNA Index System (NDIS). The national tier represents the highest level (Figure 7). The DNA profiles from LDIS are loaded on the state level system. NDIS is the central repository of DNA records submitted by states. Communication is mediated using a secured network with encryption [3].

Index

At all three levels of databases, the data collected in CODIS are separated into two indices: (1) the forensic index containing DNA profiles obtained from evidence collected at crime scenes and (2) the offending index containing DNA profiles of criminals convicted of sex offenses and several other violent crimes [3].

Database Entries

As of April 2007, the NDIS contained more than 4.5 million DNA profiles of convicted felons and 170,000 forensic profiles. Although the proportion of the population represented on the database is only approximately 1.5%, the NDIS contains the largest number of DNA profiles in absolute numbers. Each profile stored in CODIS includes additional information, including specimen identifiers, the sponsoring laboratory that provided the sample, and the DNA profile of the sample. However, it does not include individual criminal history records, dates of birth, social security numbers, or case-related information [3].

DNA Markers

The CODIS software supports the storage and searching of both restriction fragment length polymorphisms (RFLP), and PCR-based DNA profiles (HLA-DQA1, D1S80, and STR).

Ethical Issues

The use of any database involves a balance of individual civil rights and the interests of the criminal justice systems, which may vary from jurisdiction to jurisdiction. The major ethical issues related to criteria for sample entry and Retention.

Declaration of interests.

We declare no competing interests.

REFERENCES

(1) Panneerchelvam S., and Norazmi M.N. (2003). Forensic DNA Profiling and Database. Malaysian Journal of Medical Sciences 10(2): 20–26.
(2) Roewer, L. (2013). DNA fingerprinting in forensics: past, present, future. Investigative Genetics, 4(1), 22. https://doi.org/10.1186/2041-2223-4-22
(3) Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping, and population 485 genetical parameter estimation from sequencing data. Bioinformatics, 27(21):2987–2993.
(4) Frudakis, T. (2008). Molecular Photofitting. https://doi.org/10.1016/B978-0-12-088492-6.X5001-4
(5) Gettings, K. B., Borsuk, L. A., Ballard, D., Bodner, M., Budowle, B., Devesse, L., King, J., Parson, W., Phillips, C., & Vallone, P. M. (2017). Req: A catalog of sequence diversity at human identification Short Tandem Repeat loci. Forensic Science International. Genetics, 31, 111–117. https://doi.org/10.1016/j.fsigen.2017.08.017.
(6) Jeffreys, A. J., Wilson, V., and Thein, S. L. (1985). Hypervariable 'minisatellite' regions in human DNA. Nature, 314 (6006), 67–73.