Forensic DNA Testing

Maj Gen (R) Suhaib Ahmed, HI (M)

DNA from no two individuals except identical twins is alike. The person to person differences in DNA can be discovered by PCR amplification and genomic sequencing. The main advantages of DNA typing are the universality of application, an almost unlimited power to discriminate, extreme sensitivity, and reasonable resistance to degradation by environmental factors. The typing of DNA can be used in:

- Linking a suspect to a crime
- Excluding a falsely accused person
- Recognizing serial crimes
- Resolving parentage disputes
- Identification of the remains of victims

The differences in DNA are either in the form of variable number of tandem repeats (VNTR) or single nucleotide polymorphisms (SNP). Approximately 20% of the human genome comprises tandem repeat sequences. These are called micro-satellites or short tandem repeats (STR) when the repeats are 2-6bp in length. When the repeat sequences are 7-80bp in length these are called mini-satellite or variable number of tandem repeats (VNTR). The number of STR repeat units tends to vary between individuals. This variation (polymorphism) makes them extremely useful in applications like human identification and linkage analysis for diagnosis of genetic disorders.

STR analysis:

There are over a million sites (loci) in the human genome that have STRs of which over 20,000 have been characterized. Considering the highly polymorphic nature and the ease with witch the STRs can be analyzed these provide an extremely powerful tool for human identification. The discrimination power of STRs increases with increase in the number of loci tested. Therefore it is usual to use a battery of STRs. The repeat sequence of STRs range from 2-6bp. Due to technical reasons four base pair (tetra-nucleotide) repeats are best suited for forensic case work.

STR loci

The STRs are mostly located in the non-coding DNA between the genes (inter-genic DNA). Some STRs are also present in the intervening sequences (introns) of known genes. The STRs found in the inter-genic DNA are named according to the chromosome number and the site. For example in the name “D5S818” “D” stands for DNA, “5” is the chromosome number, “S” stands for single copy sequence and “818” is the locus number. The STRs in the introns of the known genes are identified by their location e.g. TH01 is present in the intron-1 of human tyrosine hydroxylase gene and TPOX is located in the thyroid peroxidase gene.

In a typical tetra-nucleotide STR the repeat units comprise four base pairs. For example an STR with "GATA" repeat would look like:

GTATCCTTATGTAATATTTTGAA**GATAGATAGATAGATAGATAGATAGATA**
GATAGATAGATAGATAGGTAGATAGAGGTATAAATAAGGATACAGATATAG

Autosomal STRs

In most forensic DNA applications STRs located on autosomes are used. A list of 15 core STR loci and their chromosomal locations are shown in Table 16.1 (Short Tandem Repeat DNA Internet DataBase).

Table. 16.1. Core STR loci used in forensic DNA testing.

Name	Chromosome	Location	Type	GenBank
TPOX	2p23-2pter	Intron 10 of human thyroid peroxidase gene	Simple	M68651
D3S1358	3p21	-	Complex	11449919
FGA	4q28	Intron 3 of alpha fibrinogen gene	Complex	M64982
D5S818	5q21-q31	-	Simple	G08446 AC008512
CSF1PO	5q33.3-34	c-fms proto-oncogene for CSF-1 receptor gene	Simple	X14720
D7S820	7q	-	Simple	G08616 AC004848
D8S1179	8q24.1-24.2	-	Simple	GO8710 AF216671
Th01	11p15-15.5	Intron 1 of human tyrosine hydroxylase gene	Complex	D00269
vWA	12p12-pter	Gene for von Willebrand antigen.	Complex	M25858
D13S317	13q22-q31	-	Simple	G09017 AL353628.2
D16S539	16q22-24	-	Simple	G07925 AC024591.3
D18S51	18q21.3	-	Simple	X91254 AP001534
D21S11	21q21.1	-	Complex	M84567 AP000433
D2S1338	2q35	-	Complex	AC010136 G08202
D19S433	19q12	-	Complex	G08036 AC008507.6

Y-Chromosome STRs

Several STRs have also been identified on Y-chromosome. These are useful in tracing male DNA in investigations like sexual assault. These may also be used in investigation of paternal lineage inheritance. A list of 15 Y-STR markers available as a commercial kit (Applied Biosystems) is shown in Table 16.2 (Short Tandem Repeat DNA Internet DataBase).

Table. 16.2. Fifteen Y-STRs available as a commercial kit (Applied Biosystems, USA)

Locus	Repeat Numbers	Repeat Motif	GenBank Accession	Reference Allele
DYS19	10-19	TAGA	AC017019	15
DYS385 a/b	7-28	GAAA	AC022486	11
DYS389 I DYS389 II	DYS389I: 9-17 DYS389II:24-34	(TCTG) (TCTA) (TCTG) (TCTA)	AC004617	12, 29
DYS390	17-28	(TCTA) (TCTG)	AC011289	24
DYS391	6-14	TCTA	AC011302	11
DYS392	6-17	TAT	AC011745	13
DYS393	9-17	AGAT	AC006152	12
DYS437	13-17	TCTA	AC002992	16
DYS438	6-14	TTTTC	AC002531	10
DYS439	9-14	AGAT	AC002992	13
DYS448	20-26	AGAGAT	AC025227	22
DYS456	13-18	AGAT	AC010106	15
DYS458	13-20	GAAA	AC010902	16
DYS635	17-27	TSTA compound	AC004772	23

STR allele nomenclature

The STR alleles are named according to the number of repeat units it contains e.g. 7, 8, 9, 10, 11 etc. Some STRs are more complex than simple repeats. The complexity may be present both in the sequence as well as the number of bases in the repeat unit. The variation in sequence of the repeats can be found only by genomic sequencing. However, the variation in number of bases in a repeat results in different sizes that can be picked on gel electrophoresis. For example at the TH01 locus allele 9 has nine simple repeats i.e. $[AATG]_9$. Another allele at the same locus has an additional triplet i.e. $[AATG]_6ATG[AATG]_3$. The resulting allele is 3bp longer and is written as 9.3. The D21S11 locus contains numerous complex alleles like 32.1, 32.2 etc. The complex loci being more polymorphic are more informative.

The STRs are inherited in a simple Mendelian fashion. An individual inherits an allele each from its father and the mother. A person may be homozygous (the same allele on the maternal and the paternal chromosomes) or compound heterozygous (different alleles on the two chromosomes). A typical STR profile comprising genotypes at several loci is shown in Table 16.3.

Table 16.3. The STR profile at various loci in individuals A, B, and C.

Individuals	D3S1358	D5S818	D7S820	D8S117	D21S11	TH01	TPOX
A.	15,18	12,12	9,10	12,16	28,29	7,9	9,11
B.	14,14	10,12	8,10	11,15	30,32.2	7,9.3	8,11
C.	15,17	10,10	9,11	13,15	29,32.2	8,10	8,11

Allele frequencies

In a given population the STR allele frequencies are calculated by simple counting. Each individual has two alleles at each locus. Genotyping of 100 individuals from a population would mean examination of 200 chromosomes (alleles). Each allele of a compound heterozygote is counted as one and homozygotes are counted as two. For example if 20/100 people have allele 8, including 2 homozygotes (8,8), the frequency of allele 8 would be 18+2+2 = 22/200 i.e. 11% or 0.11.

Minimum allele threshold

In a population survey the uncommon or the rare alleles are expected to have an under-representation. In order to overcome this problem it has been recommended to inflate the frequency of rare alleles (<5 counts) to 5. The 5/2N formula is used for this purpose where N is the number of individuals examined. The N is doubled because each individual has two chromosomes (alleles). For example if allele 12 is observed in 2/100 people its frequency by the conventional method would be 2/200 = 1% (0.01). However, by the 5/2N formula its frequency would be 2.5% (0.025).

Genotype frequencies

The allele combination in an individual at a locus is called its genotype e.g. “7,10” or “7,7” etc. The number of possible genotypes increases with an increase in the number of alleles. The possible genotypes can be calculated by the formula [n(n+1)/2] where n is the number of different alleles found in the population. The observed genotype frequencies are calculated by simple counting. The expected genotype frequencies can be calculated from the observed allele frequencies by using Hardy Weinberg equation ($p^2 + q^2 + 2pq = 1$). The homozygotes of two alleles with frequencies of p and q would be equal to p^2 and q^2 respectively while the heterozygotes (compound heterozygotes) would be equal to $2pq$.

Example

Table 16.4 describes an example of calculation of expected genotype frequencies of alleles 10 and 11 with observed frequencies of p and q respectively.

Table 16.4. The expected genotypes of two alleles at D5S818 locus.

D5S818	Frequency	p^2	q^2	$2pq$
Allele 10	p = 0.108	0.0117	-	0.0687
Allele 11	q = 0.318	-	0.101	0.0687

Profile frequencies

The STR profile of an individual is the combination of genotypes at several loci. Larger the number of loci tested rarer would be the combination in the population. The frequency of a profile is calculated by the multiplication rule. The individual genotype frequencies are multiplied to get the combined frequency (see example). The combined frequency of 15 core STR loci in the US Caucasian population is 1 in 2.46 quadrillion (10^{15}).

Example

Table 16.5 gives an example of how a profile frequency can be calculated from various allele and genotype frequencies.

Table 16.5. Example of calculation of profile frequency from the genotype frequencies in a given population.

Loci	**Alleles**	**Allele frequency**	p^2	q^2	$2pq$
D3S1358	15	p = 0.299	0.0894	-	0.0466
	18	q = 0.078	-	0.0061	0.0466
D5S818	10	p = 0.108	0.0117	-	0.0687
	11	q = 0.318	-	0.1011	0.0687
D8S1179	12	p = 0.103	0.0107	-	0.0132
	16	q = 0.064	-	0.0041	0.0132
D21S11	30.2	p = 0.059	0.0035	-	-
	30.2	-	-	-	-
Profile		Profile frequency			
D3 & D5		0.0466 X 0.0687 = 0.0032 (1 in 312)			
D3, D5 & D8		0.0466 X 0.0687 X 0.0132 = 0.000042 (1 in 23,809)			
D3, D5, D8 & D21*		0.0466 X 0.0687 X 0.0132 X 0.0035 = 0.00000015 (1 in 6666,666)			

*additional loci can also be added to this calculation

Forensic calculations and consanguineous marriage

Hardy-Weinberg equation for the calculation of expected genotype frequencies is for populations where mating is random. In a population where consanguineous marriage and marriage between tribe members is very common, the population has several substructures (smaller groups). Therefore Hardy-Weinberg equation would not be applicable as such.

The main genetic effect of consanguineous marriage is an increase in the proportion of homozygotes and a corresponding reduction in the hetrozygotes. The increase in homozygotes as compared to Hardy-Weinberg proportions is by an amount Fpq, while heterozygotes are reduced by $2Fpq$ where F is the inbreeding coefficient, and p and q are the frequencies of the alleles under consideration.

Homozygotes = $p^2 + Fpq$ *or* $q^2 + Fpq$

Heterozygotes = *2pq (1 - F)*

The coefficient of inbreeding is the probability that an individual receives at a given locus two genes that are identical by descent (copies of a single gene carried by a common ancestor). The value of *F* for a first cousin marriage is 0.0625 i.e. 6.25% of the genes are identical by descent. In 1½ cousin and 2nd cousin marriage *F* is 0.0313 and 0.0156 respectively. The average coefficient of inbreeding in the Pakistani population is 0.0280 that may be as high as 0.0350 in selected populations/tribes.

Example

The correction for consanguineous marriage applied to the genotype and profile frequency on the example shown in table 16.5 is presented in table 16.6.

Table 16.6. Example of correction for consanguineous marriage applied to the genotype and profile frequency of the example presented in Table 16.5.

Loci	**Alleles**	**Allele frequency**	$p^2 + Fpq$	$q^2 + Fpq$	$2pq\ (1 - F)$
D3S1358	15	$p = 0.299$	0.0900	-	0.0453
	18	$q = 0.078$	-	0.0068	0.0453
D5S818	10	$p = 0.108$	0.0127	-	0.0668
	11	$q = 0.318$	-	0.1021	0.0668
D8S1179	12	$p = 0.103$	0.0109	-	0.0128
	16	$q = 0.064$	-	0.0059	0.0128
D21S11	30.2	$p = 0.059$	0.0036*	-	-
	30.2	-	-	-	-
Profile		Profile frequency			
D3 & D5		0.0453 X 0.0668 = 0.00303 (1 in 330)			
D3, D5 & D8		0.0453 X 0.0668 X 0.0128 = 0.000039 (1 in 25,641)			
D3, D5, D8 & D21		0.0453 X 0.0668 X 0.0128 X 0.0036 = 0.0000014 (1 in 7142,857)			

* $p^2 + Fp(1-p)$

Mutations in STRs

Genomic DNA is liable to develop spontaneous mutations with the passage of time. In fact the existence of highly polymorphic STRs in the genome is thought to be due to spontaneous mutations that is a fairly regular event. The STR mutations become significant if these are encountered in solving a case of inheritance.

The rate of spontaneous mutations at the STR loci ranges from 1 in 500 to 1 in 1000. The rates of mutations at the core STR loci used in forensic DNA testing are given in Table 16.7. A mutation may be suspected if a disagreement is found between alleles of the parents and the offspring at one odd locus out of the several tested. The mutation can be confirmed by genomic sequencing. It may be pointed out that matching between DNA samples from a crime scene and a suspect would not be affected by mutations in the STR loci.

Table 16.7 Mutation rates at the STR loci commonly used in forensic DNA testing.

STR Loci	Mutation Rate	STR Loci	Mutation Rate	STR Loci	Mutation Rate
CSF1PO	0.16%	D3S1358	0.12%	D16S539	0.11%
FGA	0.28%	D5S818	0.11%	D18S51	0.22%
TH01	0.01%	D7S820	0.10%	D21S11	0.19%
TPOX	0.01%	D8S1179	0.14%	D2S1338	0.12%
VWA	0.17%	D13S317	0.14%	D19S433	0.11%

Core STR loci used in human identification

The core STR loci are sets of DNA markers that are globally accepted for human identification (Table 16.8). The uniformity is adopted to share and compare genetic information between different labs and the legal systems. The loci have been carefully selected to avoid their linkage with any physical character or genetic disease. In addition to the STRs a marker for sex determination, usually Amelogenin, is also included in the profile.

Table 16.9 Core STR loci required by various countries and the legal systems.

Countries	Core STR Loci
US	CSF1PO, FGA, TH01, TPOX, vWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, Amelogenin
UK/European	FGA, TH01, vWA, D2S1338, D3S1358, D8S1179, D16S539, D18S51, D19S433, D21S11, Amelogenin Recommended Loci: D1S1656, D2S441, D10S1248, D12S391, D22S1045, TPOX
Interpol	FGA, TH01, vWA, D3S1358, D8S1179, D18S51, D21S11, Amelogenin

COmbined DNA Index System (CODIS)

United States Federal Bureau of Investigation has created a database, called CODIS, that stores the DNA profiles of convicted offenders and the biological material found at crime scenes. The database contains DNA profiles comprising 15 STR loci including CSF1PO, D2S1338, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D19S433, D21S11, FGA, THO1, TPOX, vWA and Amelogenin to determine sex. Inclusion of profiles in the CODIS database is authorized by the DNA Identification Act of 1994. At present over 5 million entries exist in the database making it the largest in the world. It has helped in over 93,000 criminal investigations.

Single Nucleotide Polymorphisms (SNP)

SNPs are DNA sequence variations that are distributed throughout the genome and are found after every 300 to 500 base pairs. Almost all SNPs have two alleles. By convention the base change is called polymorphism when the frequency of its alleles in a population is more than 1%. SNPs are present in the non-coding as well as the coding sequences of DNA.

The SNPs are identified by restriction enzyme analysis and gel electrophoresis (Chapter 8). More recently real time PCR is being used to analyze multiple SNPs in a single tube by multiplex allele specific probes. The automated SNP genotyping by micro arrays also allow analysis of very large number of SNPs on a single gene chip. The later has potential for human identification applications like investigation of mass disasters.

Mitochondrial DNA

Each mitochondrion contains 2-10 copies of a circular piece of DNA, 16,569 base pairs in length. It is inherited from the mother. Mitochondrial DNA contains two hyper variable regions (HV1 and HV2) that contain many SNPs. The regions can be amplified and sequenced. Since mitochondrial DNA is inherited from the mother it can be useful in tracing maternal inheritance. Mitochondrial DNA can be extracted from structures like hair shafts, old bones and teeth etc.

PCR for sex determination

Determination of sex is an essential part of human identification by DNA test. Most commonly it is done by amplifying a sequence from the amelogenin gene whose length varies between male and the female.

Degraded DNA and "Mini STRs"

The DNA is a very large molecule that can be easily broken to smaller fragments by shearing force or bacterial enzymes (Chapter 2). Such degraded DNA samples are poorly amplified. The problem is most marked in highly degraded DNA samples. In precious and trace forensic samples degradation can completely jeopardize the analysis.

The problem of degraded DNA can be overcome to some extent by using "Mini STRs". In a mini STR analysis the forward and the reverse amplification primers flanking the repeat units are brought to the nearest possible distance from the repeat units. The resultant amplified fragments are of smallest possible size. Many degraded DNA samples can be analyzed to provide sufficient information.

Limitations of DNA test for human identification

Some of the limitations of forensic DNA test are:

1. It can not tell the age of the person.
2. In some situations it may provide information about predisposition to disease, color of eyes, height or hair color.

Applications

Matching suspect with evidence

The profiles of the two or more DNA samples are aligned to see if there is any difference or not. There could be three possible outcomes:

1. Match: When the two or more samples have the same genotypes. The statistical significance of the match is calculated as described in a subsequent section.
2. Exclusion: When the two samples originate from two different sources. This does not require prior knowledge of the allele frequencies in the population.
3. Inconclusive: Neither of the above two outcomes.

Example

The STR profiles of four DNA samples are shown in Table 16.10. Sample 1 was picked from crime scene whereas samples 2-4 are of the suspects. The samples 1 and 2 are completely matching whereas the samples 3 and 4 do not match and are therefore excluded.

Table 16.10 Comparison of the STR profile of a DNA sample collected from a crime scene (serial 1) and three suspects (serial 2-4). The profile of sample at serial 2 completely matches with that of the crime scene DNA whereas the samples at serial 2 & 3 are excluded from the match.

Sample	**D2S1338**	**D3S1358**	**D5S818**	**D7S820**	**D8S1179**	**D13S317**	**D16S539**	**D18S51**
1	22,25	17,17	11,11	8,11	10,16	9,11	12,13	13,13
2	22,25	17,17	11,11	8,11	10,16	9,11	12,13	13,13
3	24,24	18,18	12,14	9,11	8,13	11,11	11,13	14,14
4	20,23	15,15	9,11	8,12	13,16	11,12	10,10	14,17
Sample	**D19S433**	**D21S11**	**FGA**	**TH01**	**TPOX**	**vWA**	**CSF1PO**	**Amel**
1	13,15	29,30	21,25	6,9	8,9	16,16	11,11	XY
2	13,15	29,30	21,25	6,9	8,9	16,16	11,11	XY
3	13,14	28,31.2	21,23	7,9	8,9	21,21	11,12	XY
4	12,14	28,32.2	24,24	6,8	8,11	16,16	8,11	XY

Probability of match

The example shown in Table 16.10 shows a complete match between the samples at serials 1 & 2. However, there is a remote probability that the match could be by chance. The probability of match by chances is inversely proportional to the number of loci tested. Its calculation in a given population requires comprehensive knowledge of the allele frequencies of all the loci tested. A worked example of the profile frequency (probability of match by chance) was discussed in Table 16.5.

Sexual assault

In a sexual assault the DNA testing is done to demonstrate male DNA from the perpetrator in a sample collected from the victim. In a vaginal swab the victim's own DNA is seen mixed with the DNA of the perpetrator. In addition the DNA from the husband may also be present. Less often there may be more than one perpetrator and in that case complex mixtures might be found. If the perpetrator has used condom his DNA would not be found in the vaginal secretions/swab. In that case the evidence material may have to be collected from other places like seminal stains on the victim's clothing or the other objects.

Example

The STR profiles of DNA extracted from high vaginal swab of a victim of sexual assault (sample 1) and the suspects (sample 2 & 3) are shown in Table 16.11. The vaginal swab shows mixture of two DNAs. The minor component has a male genotype and matches exactly with that of the suspect 2 whereas the suspect 3 is excluded.

Table 16.11 DNA profiles of victim (sample 1) and two suspects (samples 2 &3) in a case of sexual assault. The vaginal swab shows mixture of two DNAs. The minor component has a male genotype and matches exactly with that of the suspect 2 whereas the suspect 3 is excluded.

Sample	**D3S1358**	**D5S818**	**D7S820**	**D8S1179**	**D13S317**	**D18S51**
1	16,17(16,18)	11,13(10,11)	11,11(10,12)	11,14(13,15)	11,11(11,11)	12,18(13,15)
2	16,18	10,11	10,12	13,15	11,11	13,15
3	15,16	10,13	11,13	13,15	10,13	12,15
Sample	**D21S11**	**TPOX**	**FGA**	**Th01**	**Amgl XY**	**-**
1	29,30(29,32.2)	8,8(8,8)	23,24(22,23)	9.3,9.3(6,9)	XX(XY)	-
2	29,32.2	8,8	22,23	6,9	XY	-
3	28,31.2	8,8	21,22	6,9.3	XY	-

Resolving mixtures of DNA

Mixture of DNA from more than one source is typically encountered in investigation of sexual assault.

1. Simple mixtures may be resolved on gel electrophoresis.
2. Complex mixtures are best resolved on genetic analyzer.
3. In a mixture the major component is usually of the victim itself.

4. The minor component may be from one or more individuals.
5. Identify all possible alleles.
6. A difficult issue is to differentiate between stutter products and the minor component of the mixture.
    a. The stutter products are usually of smaller size than the main PCR product and therefore would be seen moving ahead of the main product.
    b. The height of the stutter peak is usually less than 10% of the main peak.
    c. The mixture allele peak when present in the stutter peak area is considered significant when its height is more than 15% of the main peak.
    d. An allele peak or band seen in the larger size area is unlikely to be a stutter peak/band.
    e. A peak height less than 10% present in an area where stutter peaks are not expected are considered significant.
7. The alleles once identified are sorted out.
8. The victim's alleles are identified by aligning/comparing them with her own DNA extracted from blood.
9. The DNA samples of the suspect(s) are run and the results are aligned/compared to find any match.
10. In a sexual assault male DNA can also be demonstrated by Y-STR profiling.

Paternity testing

At any autosomal locus an individual inherits one allele each from the biological parents (mother and father). The child inherits mitochondrial DNA only from the mother while Y-chromosome is transmitted from the father to the son. The exceptions to the rules are development of spontaneous mutations or chromosomal aneuploidies (trisomy or monosomy). Most parentage disputes are of paternal in origin. However, occasionally maternity may also be questioned e.g. exchanged babies in a labour room.

In solving a paternity dispute a battery of STR markers are used. The step wise procedure includes:

1. Profiling of the alleged father and the child or the products of conception is done. The mother's profile is usually not required.
2. Of the two paternal alleles at each locus the allele transmitted to the child, called the "obligate allele", is selected.
3. If none of the alleles of the alleged father are present in the child at any of the loci tested the paternity is excluded. For example if the child has genotype 14,18 and the father has 13,15 the paternity is excluded. However, keeping the possibility of spontaneous mutations in mind it is advisable to consider more than one loci before excluding paternity.

4. Frequencies of the obligate paternal alleles are noted from the table of allele frequencies in the reference population.
5. The Paternity Index (PI) is calculated by dividing the prior probability with the frequency of the allele in question. The prior probability is the chance of transmitting the obligate allele by the alleged father to the child. If the alleged father is homozygous for the allele the prior probability is 1.0 and it is 0.5 if he is heterozygous.
6. Combined Paternity Index (CPI) is calculated by multiplying the PI values calculated at each locus.
7. The Probability of Paternity (POP) is calculated by the formula:

   CPI /CPI + (1-prior probability) x 100

Example

An example of calculation of paternity index and the combined paternity index is shown in Table 16.12.

Table 16.12 Calculation of Paternity Index and Probability of Paternity.

Locus	Genotype			Obligate Allele	Frequency	Paternity Index (PI) = Prior probability*/frequency
	Mother	Child	Al/Father			
D3S1358	14,15	14,18	15,18	18	0.078	0.5/0.078 = 6.41
D5S818	9,12	9,9	9,9	9	0.049	1.0/0.049 = 20.41
D7S820	8,11	8,9	9,11	9	0.108	0.5/0.108 = 4.63
D8S1179	10,12	10,15	11,15	15	0.211	0.5/0.211 = 2.37
D21S11	29,32.2	29,31.2	28,31.2	31.2	0.123	0.5/0.123 = 4.06
TPOX	9,11	9,11	8,11	11	0.333	0.5/0.333 = 1.50
TH01	8,9.3	9,9.3	6,9	9	0.250	0.5/0.250 = 2.00
FGA	21,21	20,21	20,24	20	0.093	0.5/0.093 = 5.38
Combined Paternity Index (CPI)			6.41 X 20.41 X 4.63 X 2.37 X 4.06 X 1.50 X 2.00 X 5.38 = 94,072 or 1 in 94,072			
Probability of Paternity			CPI/CPI + (1-prior probability) X 100 94072/(94072 + 0.5) X 100 = 99.999 %			

*If the father is heterozygous for the allele Prior Probability = 0.5
*If the father is homozygous for the allele Prior Probability = 1.0

Identification in mass disasters

In mass disasters the DNA testing is primarily done to identify bodies that are beyond recognition. It may also be required to put the pieces of bodies together.

The process is done as follows:

1. The DNA profiles of the dead bodies or their remains are entered in a computer database.

2. The DNA profiles of the bodies is matched with those of the parents or the offsprings. If these are not available then DNA may be matched with that recovered from the personal effects of the victims.
3. The matched profiles are put together after sorting.

Table. 16.13. Comparison of the DNA profiles of three dead bodies from a mass disaster and a father. The DNA profile of the "father" shares at least one allele at all fifteen loci with the DNA profile of the body number 7 and is therefore a proof of identity.

Sample	D2S1338	D3S1358	D5S818	D7S820	D8S1179	D13S317	D16S539	D18S51
3	20,25	14,15	8,10	8,12	14,16	10,11	10,10	13,16
7	20,23	15,15	9,11	8,12	13,16	11,12	10,10	14,17
16	22,23	14,14	10,11	8,11	13,16	10,12	10,12	15,17
Father	21,23	15,16	9,11	7,12	13,14	12,14	10,11	13,17

Sample	D19S433	D21S11	FGA	TH01	TPOX	vWA	CSF1PO	Amel
3	11,14	29,32.2	23,23	6,9.3	8,11	15,16	8,12	XY
7	12,14	28,32.2	24,24	6,8	8,11	16,16	8,11	XY
16	13,14	30,32.2	22,24	6,9	8,10	14,14	8,11	XY
Father	13,14	32.2,32.2	21,24	7,8	8,10	15,16	8,11	XY

Determining the ethnic origin of a person

Can DNA testing provide information on the ethnic origin of a person? The question is asked more than often especially when no clue is available to identify the perpetrator. There are some STR loci that have more significant differences in allele frequencies between major world populations like Caucasians, Blacks, Hispanics, Asians and East Asians etc. The differences are more marked in the frequency of less common alleles than the more common ones. An important requirement would be to have representative samples of the target ethnic groups with ethnically pure individuals and not the ones with self declared ethnicity. In this context analysis of SNP is considered more informative than the STR markers.

In tribal populations with high frequency of consanguinity and marriage within the same tribe founder effect and genetic drift might cause an unexpectedly higher or lower frequency of alleles than in the rest of the population.

OmniPop 200.1 software is available on the internet free of cost. It is basically a database of the STR allele frequencies published in over 200 studies in the major world populations. By entering the CODIS STR profile of an individual it provides statistical information on finding a similar profile in various world populations. Similar software can be developed for one's own requirements if comprehensive knowledge of the allele frequencies in the smaller target groups and populations is available.

Collection storage and dispatch of samples for forensic DNA testing

1. Collection and storage of samples

    a. DNA can be extracted from any biological material that contains sufficient number of nucleated cells. Since DNA is susceptible to degradation by mal handling and nucleases derived from cells and bacterial contamination the sample should remain as clean and free of bacterial contamination as is possible.

    b. DNA may be extracted from a wide variety of samples. The usual samples are whole or dried blood (stains), buccal smear, wet or dried semen and other body secretions, hair with roots, soft tissues, fresh and dried bones etc.

    c. The samples of stains like blood or body secretions should be air dried and kept in a paper envelop rather than plastic bag.

    d. Fresh blood: 2-3 ml of venous blood should be collected in EDTA. The sample can be stored before dispatch at $4^{o}C$ for 48 hours.

    e. Blood stains on clothes or other objects should be dried at room temperature before dispatch in a paper bag.

    f. Buccal smear obtained on a clean throat swab contains mucosal cells to yield sufficient quantity of good quality DNA. The swab can also be used to obtain DNA from a dead body. The swab should be rubbed several times over clean part of buccal mucosa. It may be air dried before storage or dispatch.

    g. Semen and other body fluids containing nucleated cells are a good source of DNA. The stained clothes or objects should be air dried and treated as blood stains.

    h. Hair shafts do not contain nuclear DNA. Hairs that are plucked from the body and come out with roots can be used to extract DNA. Sufficient DNA can be extracted from 2-3 hairs with roots.

    i. Soft tissues are a very rich source of DNA. Skin is an easily accessible tissue that can be used to collect DNA from a dead body. A full thickness piece of skin measuring 2x2 cm from a clean part of the body or its remains can be taken. If skin is not available then any other available soft tissue should be collected. The soft tissue samples provide an excellent medium for bacterial growth. In a putrefied or heavily contaminated soft tissue sample the yield as well as the quality of DNA can be very poor. The soft tissue samples can be stored as such at $-20^{o}C$ in a suitable container for several days. The sample may be transported in normal saline. Do not put the samples for DNA testing in formalin at any stage.

    j. Bones can also be used as a source of DNA. But the extraction of DNA from a bone is difficult therefore the yield and quality of DNA is also

variable. DNA is best collected from compact bones like humerus or femur. The spongy bones with thin cortex are usually heavily contaminated with mud etc. that may inhibit PCR.

2. **Dispatch of samples**
    a. The samples should be properly labeled and sealed.
    b. The request form should contain all available details of the individual(s) to be tested along with a brief summary of the incident and what exactly is required to be solved by the DNA test.
    c. Unnecessary delay in transport can adversely affect the quality of the sample. There is no special requirement of transporting the samples in ice etc. However, avoid exposing the sample to extreme heat or direct sunlight.
3. **Chain of custody**
    a. Forensic DNA testing is done for medico-legal purpose therefore it is essential to maintain the chain of custody.
    b. A record of the individuals receiving and handing over the samples must be maintained at all steps as they may be called by the court as a witness.

Bibliography

1. Kaiser L and Sever G (1983) Paternity testing: I. Calculation of paternity indexes. Am J Med Genet 15: 323-329.
2. Elston RC (1986) Probability and paternity testing. Am J Hum Genet 39: 112-122.
3. Smith RN (1995) Accurate size comparison of short tandem repeat alleles amplified by PCR. *Biotechniques* 18: 122-128.
4. Brinkmann, B., Klintschar, M., Neuhuber, F., Huhne, J., and Rolf, B. (1998). Mutation rate in human microsatellites: Influence of the structure and length of the tandem repeat. Am J Hum Genet 62: 1408-1415.
5. Short Tandem Repeat DNA Internet DataBase, http://www.cstl.nist.gov/strbase/