ANALYSIS OF POPULATIONS WITHIN THE UAE USING TANDEM REPEAT DNA MARKERS

by

Ahmed Abdulla Ahmed Mohammed

B.Sc.

Thesis submitted for the degree of Doctor of Philosophy in the University of

Glasgow, Department of Forensic Medicine and Science

United Kingdom

March, 2001

Lastly, I would like to thank my wife and my children Saad, Jenan, Afnan and Salman for their support and patient during my study.

CONTENTS

ACKNOWLEDGEMENTS ---- **I**

CONTENTS ---- **III**

ABBREVIATIONS ---- **XI**

ABSTRACT ---- **XIII**

CHAPTER 1: INTRODUCTION ---- **1**

1.1 FORENSIC USE OF VARIATION ---- 1

1.1.1 Protein and Enzyme Polymorphism ---- *1*

1.1.2 Limitation of protein markers ---- *2*

1.2 HUMAN GENOME ---- 4

1.2.1 Deoxyribonucleic Acid (DNA) Polymorphisms ---- *6*

1.2.2 Non-coding DNA ---- *7*

1.2.3 Repetitive DNA ---- *7*

1.2.3.1 Interspersed repeat sequences ---- 8

1.2.3.2 Tandem repetitive DNA sequences ---- 9

1.3 SATELLITE DNA (CLASSICAL SATELLITES) ---- 10

1.4 MINISATELLITE DNA ---- 10

1.4.1 Minisaellite and restriction enzyme ---- *13*

1.4.2 Generation of polymorphisms ---- *13*

1.4.3 Mutation rate (minisatellite instability) ---- *14*

1.4.4 Minisatellite DNA probes ---- *15*

1.4.5 Multi-locus DNA probes ---- *15*

1.4.6 Limitation of MLPs ---- *16*

1.4.7 Single locus DNA probes ---- *16*

1.4.8 Chemiluminescent detection (non-isotopic system) ---- *17*

1.4.9 Forensic applications of minisatellites ---- *17*

1.4.10 Matching and binning criteria ---- *18*

1.4.11 Profile frequency ---- *20*

1.4.12 Debate over the use of VNTRs in forensic world ---- *21*

1.4.13 Minisatellite limitations ---- 23
1.5 POLYMERASE CHAIN REACTION (PCR) ---- 24
1.6 MICROSATELLITE DNA (AUTOSOMAL STRs) ---- 25
1.6.1 Types of STR loci ---- 27
1.6.1.1 Simple STR loci ---- 27
1.6.1.2 Compound STR loci ---- 27
1.6.1.3 Complex STR loci ---- 28
1.6.1.4 Hypervariable STR loci ---- 28
1.6.2 Detection of STR loci ---- 29
1.6.3 STR multiplex systems ---- 30
1.7 ADDITIONAL PCR SYSTEMS ---- 34
1.7.1 HLA-DQα ---- 34
1.7.2 PolyMarker (PM) ---- 35
1.7.3 Amplified fragment length polymorphisms (Amp-FLPs) ---- 35
1.7.4 Y chromosome short tandem repeat (Y-STR) ---- 37
1.8 PROJECT BACKGROUND ---- 38
1.8.1 Introduction ---- 38
1.8.2 Physical and political structure ---- 41
1.8.3 The Climate ---- 42
1.8.4 The historical background ---- 42
1.8.5 Origins of Arabic people ---- 43
1.8.6 Arabs in north Africa ---- 44
1.8.7 The UAE population structure ---- 44
1.8.8 Crime rate ---- 50
1.8.9 Abu Dhabi ---- 50
1.8.10 Sharjah ---- 51
1.8.11 The UAE tribes ---- 51
1.8.12 Marriages ---- 51
1.9 OBJECTIVES OF RESEARCH ---- 52
CHAPTER 2: MATERIALS AND METHODS ---- 54
2.1 MATERIALS ---- 54
2.1.1 Chemicals ---- 54
2.1.2 Restriction endonuclease enzyme ---- 54
2.1.3 K562 DNA ---- 54

2.1.4 DNA molecular ladder ---- 54
2.1.5 Single locus DNA probe and CDP-Star ---- 55
2.2 GENERAL PREPARATORY PROCEDURES ---- 55
2.2.1 pH measurement ---- 55
2.2.2 Autoclaving ---- 55
2.2.3 Glassware ---- 55
2.2.4 Solutions and buffers ---- 55
2.3 DNA SAMPLE COLLECTION AND PACKAGING ---- 56
2.3.1 Source of Blood Samples ---- 56
2.3.1.1 Blood samples from Abu Dhabi City ---- 56
2.3.1.2 Blood samples from Sharjah City ---- 56
2.3.2 Blood sample packaging ---- 56
2.4 ISOLATION OF DNA FROM WHOLE BLOOD ---- 57
2.4.1 Nucleic acid extraction ---- 57
2.5 ELECTROPHORESIS OF DNA ---- 58
2.5.1 Agarose gel electrophoresis ---- 58
2.5.2 Mini-gel electrophoresis ---- 59
2.5.3 Quality and quantity of DNA ---- 60
2.5.4 Restriction digestion ---- 60
2.5.5 Analytical gel electrophoresis ---- 60
2.6 SOUTHERN BLOTTING ---- 61
2.6.1 Depurinating of DNA ---- 61
2.6.2 Denaturing of DNA ---- 62
2.6.3 Neutralising of DNA ---- 62
2.6.4 Capillary transfer ---- 62
2.7 FIXATION OF DNA BLOTS ---- 65
2.7.1 Fixation by UV cross-linking ---- 65
2.7.2 Fixation by oven baking ---- 65
2.8 HYBRIDISATION ANALYSIS ---- 65
2.8.1 Preparation of hybridisation solution ---- 65
2.8.2 Pre-hybridisation of DNA blots ---- 66
2.8.3 Hybridisation of DNA blots ---- 66
2.9 WASHING OF BLOTS ---- 67
2.9.1 Wash solution 1 ---- 67

2.9.2 Wash solution 2 ---- 67

2.10 CHEMILUMINESCENT DETECTION ---- 67

2.10.1 CDP-Star ---- 67

2.10.2 Autoradiography ---- 68

2.10.3 Stripping and re-probing the nylon membrane ---- 68

2.11 POLYMERASE CHAIN REACTION (PCR) ---- 69

2.11.1 Isolation of DNA from whole blood ---- 69

2.11.2 Dilution of DNA ---- 69

2.11.3 Amplification of DNA samples ---- 71

2.11.4 Multiplex PCR ---- 71

2.11.5 Storage conditions ---- 72

2.11.6 PCR amplification ---- 72

2.12 310 GENETIC ANALYSER ---- 72

2.12.1 Matrix file preparation ---- 73

2.12.2 DNA typing ---- 74

2.13 STATISTICAL ANALYSIS SOFTWARES ---- 75

CHAPTER 3: VARIABLE NUMBER TANDEM REPEATS ---- 77

3.1 INTRODUCTION ---- 77

3.2 POPULATION SUBSTRUCTURE ---- 77

3.3 SELECTION OF VNTR LOCI ---- 78

3.4 DESCRIPTION OF THE LOCI AND PROBES ---- 79

3.4.1 D2S44 locus ---- 79

3.4.2 D7S22 locus ---- 80

3.4.3 D1S7 locus ---- 80

3.4.4 D7S21 locus ---- 81

3.4.5 D12S11 locus ---- 81

3.5 SAMPLES COLLECTION ---- 82

3.5.1 DNA extraction ---- 82

3.5.2 Analysis of restricted DNA ---- 84

3.5.3 Analytical gel electrophoresis ---- 86

3.6 ANALYSIS OF VNTR PROFILES AND ESTIMATION OF ALLELE FREQUENCIES ---- 86

3.6.1 Binning approach ---- 90

3.6.2 Fixed bin approach ---- 90

3.6.3 Allele size ranges observed ---- 93

3.7 INTER AND INTRA-GEL VARIATION ---- 94
3.7.1 Inter-gel variation ---- *96*
3.7.2 Intra-gel variation ---- *97*
3.8 MATCH CRITERION ---- 98
3.9 DISCUSSION ---- 102
3.9.1 Samples collection ---- *102*
3.9.2 DNA extraction and digestion ---- *102*
3.9.3 VNTR system ---- *102*
3.9.4 Non-isotopic detection ---- *103*
3.9.5 Inter and intra-gel variation ---- *103*

CHAPTER 4: VNTR STATISTICAL ANALYSIS ---- 105

4.1 INTRODUCTION ---- 105
4.2 ALLELE FREQUENCIES ---- 105
4.3 GENOTYPE FREQUENCIES ---- 119
4.4 POPULATION SUBSTRUCTURE AND INBREEDING ---- 121
4.5 COMPARATIVE STUDIES ---- 122
4.5.1 RxC test ---- *122*
4.5.2 F-statistics ---- *123*
4.5.3 The effect of population substructure ---- *127*
4.6 PHYLOGENTIC RELATIONSHIP BETWEEN POPULATIONS ---- 128
4.7 HARDY-WEINBERG EQUILIBRIUM ---- 129
4.7.1 Chi-square test ---- *131*
4.7.2 Exact test ---- *131*
4.7.3 Independence of loci ---- *133*
4.8 HETEROZYGOSITY ---- 134
4.9 FORENSIC AND PATERNITY PARAMETERS ---- 137
4.9.1 Forensic parameters ---- *137*
4.9.1.1 Matching probability ---- 138
4.9.1.2 Discrimination power ---- 138
4.9.2 Paternity parameters ---- *139*
4.7.2.1 Power of exclusion ---- 139
4.9.2.2 Paternity index ---- 140
4.10 DISCUSSION ---- 143
4.10.1 Hardy-Weinberg equilibrium tests ---- *143*

4.10.2 Heterozygosity ---- *143*

4.10.3 Comparison of the three populations studies ---- *143*

4.10.4 Forensic and paternity parameters ---- *144*

CHAPTER 5: AUTOSOMAL SHORT TANDEM REPEATS ---- 147

5.1 INTRODUCTION ---- 147

5.2 THE GENERAL POPULATION ---- 148

5.3 DNA EXTRACTION AND DILUTION ---- 149

5.4 MULTIPLEX ANALYSES OF STR LOCI ---- 149

5.5 DESCRIPTION OF THE LOCI ---- 151

5.5.1 D5S818 locus ---- *151*

5.5.2 Human vWAFA31/A (vWA) locus ---- *151*

5.5.3 D13S317 locus ---- *152*

5.5.4 Human THO1 locus ---- *152*

5.5.5 D7S820 locus ---- *153*

5.5.6 Human TPOX locus ---- *153*

5.5.7 D16S539 locus ---- *153*

5.5.8 Human CSF1PO locus ---- *154*

5.5.9 Human amelogenin locus ---- *154*

5.5.10 Allele designation and sizing ---- *155*

5.6 RESULTS ---- 158

5.6.1 Co-amplification of STR loci ---- *158*

5.6.2 Allele frequencies ---- *158*

5.7 DISCUSSION ---- 159

5.7.1 Samples collection ---- *159*

5.7.2 DNA extraction ---- *159*

5.7.3 Multiplex PCR ---- *160*

CHAPTER 6: STR STATISTICAL ANALYSIS ---- 161

6.1 INTODUCTION ---- 161

6.2 ALLELE FREQUENCIES ---- 161

6.3 GENOTYPE FREQUENCIES ---- 170

6.4 POPULATION SUBSTRUCTURE ---- 172

6.5 HARDY-WEINBERG EQUILIBRIUM ---- 173

6.5.1 Chi-square test ---- *173*

6.5.2 Exact test ---- *173*

6.5.3 Independence of loci ---- *175*

6.6 HETEROZYGOSITY ---- 177

6.7 COMPARITIVE STUDY ---- 179

6.7.1 RxC test ---- *179*

6.7.2 F-statistics ---- *181*

6.8 PHYLOGENETIC RELATIONSHIP BETWEEN POPULATIONS ---- 184

6.9 FORENSIC AND PATERNITY PARAMETERS ---- 186

6.9.1 Forensic parameters ---- *186*

6.9.1.1 Matching probability ---- 188

6.9.1.2 Discrimination power ---- 188

6.9.2 Paternity parameters ---- *188*

6.9.2.1 Power of exclusion ---- 188

6.9.2.2 Typical paternity index ---- 189

6.10 COMBINING ALL VNTR AND STR LOCI ---- 190

6.11 COMPARISON BETWEEN POPULATIONS ---- 190

6.11.1 Comparison with Arabic populations ---- *191*

6.11.1.1 Allele frequency ---- 191

6.11.1.2 RxC test ---- 195

6.11.1.3 Forensic and paternity parameters comparison ---- 196

6.11.2 Comparison with major world populations ---- *197*

6.11.2.1 RxC test ---- 198

6.11.2.2 Heterozygosity comparison ---- 202

6.11.2.3 Forensic and paternity parameters comparison ---- 203

6.11.3 Indian subcontinent populations ---- *204*

6.12 DISCUSSION ---- 205

6.12.1 Hardy-Weinberg Equilibrium test ---- *205*

6.12.2 Comparison between different UAE populations ---- *206*

6.12.3 Comparison between two UAE Arabic native populations ---- *207*

6.12.4 Comparison between other Arabic populations ---- *207*

6.12.5 Comparison between the UAE and major world populations ---- *207*

6.12.6 Comparison between the Indo-Pakistani populations residing in two different countries ---- *207*

6.12.7 Forensic and paternity parameters ---- *208*

6.12.8 Combination of all STR and VNTR loci ---------- *209*

CHAPTER 7: CONCLUSION AND FUTURE WORK ---------- **212**

7.1 THE ACHIEVEMENT OF THIS STUDY ---------- 212

7.2 THE FUTURE WORK ---------- 215

APPENDIX A: ---------- **218**

APPENDIX B: ---------- **221**

APPENDIX C: ---------- **226**

APPENDIX D: ---------- **236**

APPENDIX E: ---------- **255**

APPENDIX F: ---------- **266**

APPENDIX G: ---------- **268**

APPENDIX H: ---------- **293**

REFERENCES ---------- **308**

ABBREVIATIONS

µg	microgram
µl	microlitre
AABB	American Association of Blood Banks
ADA	Adenosine Deaminase
AK	Adenylate kinase
AmpFLP	Amplified fragment length polymorphism
bp	base pair
CCD	charged-coupled device
cm	centimetre
cM	centimorgan
CODIS	Combined DNA Index System
CSF1PO	c-fms proto-oncogene
CXR	carboxy-X-rhodamine
df	degree of freedom
EAP	erythrocyte acid phosphatase
EDNAP	European DNA Profiling Group
EDTA	ethylenediaminetetra acetic acid
EsD	Esterase D
ESWG	English Speaking Working Group
EtBr	ethidium bromide
FBI	The Federal Bureau of Invistigation
FES/FPS	c-fes/fps proto-oncogene
FGA	alpha fibrinogen gene
F_{ST}	Co-ancestry coefficient
g	gram
g	gravity
Gc	group specific component
GCC	Gulf Co-operation Council
GDA	genetic data analysis
GITAD	Groupo Iberoamericano de Trabajo en Analisis de DNA
GLO	glyoxalase
GYPA	Glycophorin A
HBGG	Haemoglobin G Gammaglobin
HLA	human leukocyte antigen
Hp	haptoglobin
HWE	Hardy-Weinberg equilibrium
IEF	isoelectric focussing
ILS	Internal lane standard
ISFG	International Society for Forensic Genetics
ISFH	International Society for Forensic Haemogenetics
kb	kilo base pair
kg	kilo gram
kV	kilo volt
l	litre
LDLR	Low Density Lipoprotein Recepter
LE	Linkage equilibrium

LINES	long interspersed repeated sequences
ml	millilitre
MLP	multilocus probe
mtDNA	mitochondrial DNA
MVR	Minisatellite variant repeat
NICE	Non-Isotopic Chemiluminescent Enhanced
nm	nanometre
NRC	National Research Council
PCR	polymerase chain reaction
PD	power of discrimination
PE	power of exclusion of paternity
PGM	phosphoglucomutase
PI	paternity index
PM	Polymarker
psi	pounds per square inch
RBC	red blood cell
RFLP	restriction fragment length polymorphism
Rh	Rhesus system
rpm	revolutions per minute
s	second
SD	standard deviation
SDS	sodium dodecyl sulphate
SE	standard error
SGM	Second Generation Multiplex
SINES	short interspersed repeated sequences
SLP	single locus probe
SSC	saline sodium citrate
ssDNA	single stranded DNA
STR	short tandem repeat
SWGDAM	Scientific Working Group for DNA Analysis Method
TBE	Tris-borate EDTA
Tf	Transferrin
TFPGA	tools for population genetic analysis
THO1	tyrosine hydroxylase
TPI	typical paternity index
TPOX	thyroid peroxidase
u	unit
UAE	United Arab Emirates
UPGMA	unweighted pair group method using arithmetic averaging
UV	ultraviolet
V	volt
v/v	volume to volume
VNTR	variable number tandem repeat
vWA	von Willebrand factor A
W	probability of paternity
w/v	weight to volume

ABSTRACT

This research has been carried out to study the UAE population structure, particularly as the UAE has an unusual population composition. Transient workers from Indian sub-continental outnumber the native Arabic population by approximately three (60%) to one (20%). Other Arab populations make up approximately 15% of the population, the majority are from Egypt, Yemen, Palestine and Sudan. The relatively high levels of consanguineous marriages (between close relatives such as first cousin) between UAE native population added more complexity to the population. Therefore, it was important to analyse the population before applying any new tools to forensic analysis and paternity testing in the UAE.

In order to examine the population, blood samples were collected from the three ethnic groups (UAE Arab, Indian and Pakistani) living in two Emirates of the UAE (Sharjah and Abu Dhabi). Additionally, blood samples were collected from El-Minia city in Egypt in order to represent another Arabic population and to compare this to UAE native Arabic population.

Two polymorphic systems were chosen (VNTRs and STRs) to investigate the population substructure. Five single loci VNTR/*Hin*fI probes MS1, MS31, MS43A, YNH24 and G3 were used to profile 173 individuals from the UAE native Arabic population, 154 individuals from Indian and 112 individuals from Pakistani populations. The F_{ST} was calculated by comparing the UAE Arabic population to both Indian and Pakistani populations. The highest value of (0.0062) was observed between the UAE Arabic and Indian populations. No evidence of substructure was observed when Indian population compared to Pakistani population (F_{ST} = -0.003).

In addition, eight simple STR loci D5S818, D7S820, D13S317, D16S539, vWA, THO1, TPOX and CSF1PO (GenePrint™ PowerPlex™ 1.2 System) were used to profile 229 UAE native Arab (100 from Sharjah and 129 from Abu Dhabi), 194 Indian, 197 Pakistani and 121 Egyptian individuals. No evidence of substructure was observed when the two native UAE populations compared to each other (F_{ST} = -0.0004). The greatest substructure value was observed when Indian population compared to Egyptian population (F_{ST} = 0.0129). The value of F_{ST} was low (F_{ST} = 0.003) when the two Arab populations were compared to each other and when the two Indian subcontinent populations were compared to each other (F_{ST} = 0.002). The data were also tested for obedience with Hardy-Weinberg equilibrium using the exact test, Chi-square test and heterozygosity test.

From the phylogenetic tree and RxC test, as expected, there was more similarity between the two Arab populations studied compared to other Indo-Pakistani populations and vice versa. Therefore, by using a very conservative, the upperbound, value of F_{ST} = 0.023 only two databases, Indo-Pakistani and Arabs, might be sufficient to be used in the UAE and covering approximately 95% of the UAE total population.

The data were analysed to estimate a number of forensic and paternity parameters, including matching probability, discrimination power, probability of paternal exclusion and typical paternity index in order to assess the application of these two systems for forensic and paternity tests in the UAE.

The five VNTR loci and eight STR loci together were proved to be very powerful tool for forensic analysis and determining paternity within the UAE.

CHAPTER 1: INTRODUCTION

1.1 FORENSIC USE OF VARIATION

Forensic analysis, in general, is a comparative science. An unknown sample is compared to a known sample to determine whether they match or possibly have a common origin. One of the first method of human identification, latent fingerprint, was discovered by Francis Galton in 1892. Forensic scientist continued to look for additional ways to identify suspects, such as the linking of biological remains (body fluids, hairs, tissues, skeleton remains, etc.) which commonly occur at the scene of crime. Similarly, biological samples can be used in civil cases, such as paternity and immigration analysis. These problems of individual identification can be resolved by analysing a number of genetic polymorphisms or markers.

1.1.1 Protein and enzyme polymorphism

The ABO blood group system was the first genetic polymorphism discovered and documented by Karl Landsteiner in 1900 in Vienna (Landsteiner K., 1900). The first attempt to apply this discovery to forensic cases of dried blood stains was undertaken in 1902 by Max Richter and Karl Landsteiner (Landsteiner K. and Richter M., 1902). Later, Leone Lattes developed a new technique for the ABO grouping in 1916 (reviewed by Gaensslen R., 1983) and subsequently, this concept of identification was extended to be used in paternity cases as well.

The ABO blood group system was the only one used on forensic samples for a long period of time. The next major advancement came with the separation of proteins according to size and charge, which was achieved with the introduction of starch gel electrophoresis (Smithies O., 1955a, b). Another significant breakthrough was achieved in the early 1970's in the field of forensic serology when other protein and enzyme markers, which had been known to be polymorphic for a decade or more, were found to be useful for forensic analysis as well.

The Scotland Yard lab (Metropolitan Police Forensic Science Laboratory) in London found that some of the markers that had previously been used for paternity testing could also be used on bloodstains and other kinds of physiological fluid evidence found at crime scenes. The discovery of red blood cell antigens (ABO, Rh, MNSs, P, Kell, Duffy, and Lewis), red blood cell enzymes and proteins (isozymes) (PGM, EsD, EAP, AK, ADA, GLO) and the serum proteins (Hp, Gc, Tf, Gm and Km system factors), all greatly enhanced the characterisation of human blood (Sensabaugh G., 1982). Later, with the advent of isoelectric focussing (IEF) (Caspers M. *et al.*, 1977), which is a modified of electrophoresis, more subtypes could be detected. For example, only 3 phenotypes can be detected by conventional separation of protein such as PGM, however, a total of 10 phenotypes could be detected by applying IEF system.

1.1.2 Limitation of protein markers

Although these typing systems can provide valuable evidence, the main drawback of studying polymorphism at the protein level is that specific and individual analytical procedures are required for the detection of each polymorphism. In addition, most of these genetic markers show relatively poor levels of individual variation. For example, with ABO blood group system the population can be divided into only four different blood

group types (A, B, AB, and O) type O and A common, where is type AB is less common. Hence, matching of an ABO type between forensic blood sample and a suspect has a weak statistical value.

In paternity testing, the situation is even worse. For example, if a mother is type O and her child is type A, then the father must have given to the child gene A. The father must be group A or AB. Half the European population for example falls within these two groups, and thus 50% of the population would not be excluded by this test. To overcome these problems, forensic serologists used dozens of different markers, which together enhanced individual characterisation.

From the forensic point of view the quantity and the quality of the blood found at the scene of crime is critical. Since a trace or limited amount of biological sample usually can be found at a scene of crime and even if found in high quantity the protein markers are complex molecules that are unstable and frequently deteriorate in forensic samples. This limits the number of markers that can be analysed for a particular stain. In the analysis of semen stains only ABO, GLO, Pep A, Lewis and PGM markers can be analysed, as many of other markers are either not present or not detectable in semen. Furthermore, semen samples are often contaminated with vaginal fluid which it self has enzyme and blood group activity.

Moreover, detection of some protein polymorphisms is difficult after about 26 weeks or even less for some others (Gill P. *et al.*, 1985) and bacterial activity may be responsible for producing erroneous results. However, forensic serology allowed for the first time the principle of individual identification. The identification and characterisation

of biological samples based on DNA analysis was very important to overcome the limitations of conventional markers.

1.2 HUMAN GENOME

The human body consists of 100 trillion (10^{14}) cells. Each nucleated cell contains a full set of chromosomes (46 chromosomes, 23 homologous pairs); 23 inherited from the mother and 23 inherited from the father. The true number of human chromosomes was determined only in 1956 (Tjio H. and Levan A., 1956) and before that it believed to be 48. However, the accurate identification of each individual human chromosome was achieved only in 1970 (reviewed by Fa-Ten Kao, 1985).

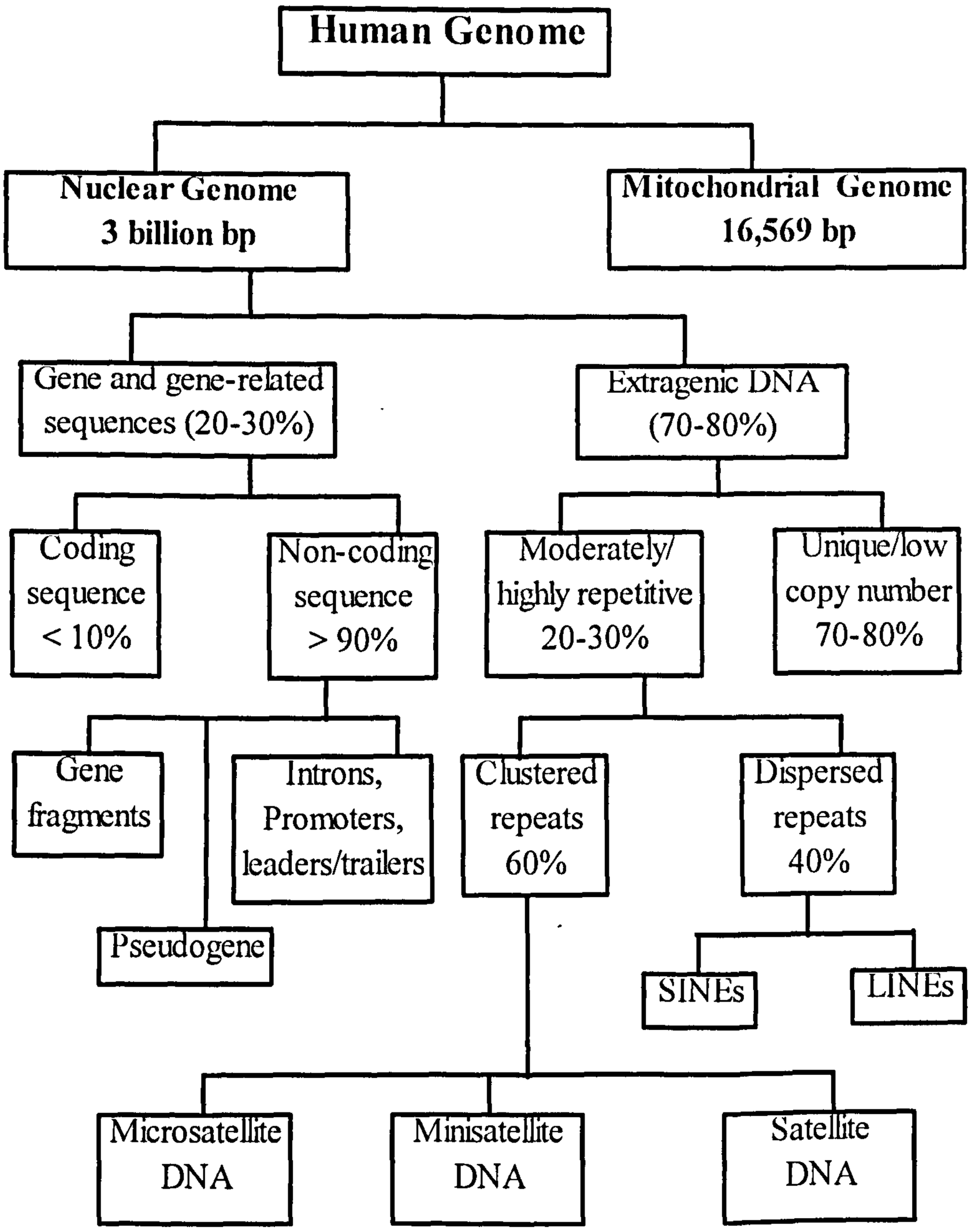
Each haploid cell contains of 3×10^9 base pairs (Krawczak M. and Schmidtke J., 1994). However, approximately 3% of the total human genome involved in the expression of gene products (reviwed by Southern E., 1995). Each individual (with the exception of identical twins) is a genetically unique and that this uniqueness stems from variation in a precise chemical sequence of bases along our DNA molecules. It has been estimated that approximately one out of every 300-1,000 nucleotides varies between two unrelated individuals (Ullrich A. *et al.*, 1980; Cooper D. *et al.*, 1985; Pena S. *et al.*, 1995).

The bulk of the human genome does not code for genes and is either intergenic (between coding regions) or introns (associated with the gene, but not encoding). See figure 1.1.

4A

Figure 1.1 Diagram showing composition of the human genome.

Adapted from Strachan (1992) The Human Genome, BIOS Scientific Publishers.


Human Genome
Nuclear Genome
3 billion bp
Mitochondrial Genome
16,569 bp
Gene and gene-related sequences (20-30%)
Extragenic DNA
(70-80%)
Coding sequence
< 10%
Non-coding sequence
> 90%
Moderately/ highly repetitive
20-30%
Unique/low copy number
70-80%
Gene fragments
Introns, Promoters, leaders/trailers
Pseudogene
Clustered repeats
60%
Dispersed repeats
40%
SINEs
LINEs
Microsatellite DNA
Minisatellite DNA
Satellite DNA

1.2.1 Deoxyribonucleic Acid (DNA) Polymorphisms

Conventional forensic serology testing can detect genetic polymorphisms at the protein level, however, genetic polymorphisms, since the advent of molecular biology, can now be detected directly at the DNA level. The challenge in forensic genetics is to develop DNA typing systems which access as much variability as possible and are simple, robust and reliable. While most of our DNA shows rather modest levels of variation between individuals, there exist regions that can show extreme levels of variability.

There are several advantages of using DNA markers instead of protein markers in both, forensic and paternity analysis. For example, DNA is remarkably stable (Bar W. *et al.*, 1989; Doran G. *et al.*, 1986; Madisen L. *et al.*, 1987) and its composition is relatively immune to environmental alterations (Adams D., 1988; McNally L. *et al.*, 1989a, b). DNA can be obtained from dead and aged material such as in 2,400 year old mummies (Paabo S., 1985). In addition, DNA can be typed from minute forensic specimens (Higuchi R. *et al.*, 1988; Li H. *et al.*, 1988; Hochmeister M. *et al.*, 1991) and can be found in almost all biological specimens with many sites of variation. There are both important considerations as forensic samples are collected in ranging degrees of preservation and quantity.

Further, the devolpment of a differential cell lysis method where it is possible to separate sperm nuclei from vaginal cellular debris from semen-contaminated vaginal swabs allowed the positive identification of rapists. Similarly vaginal DNA can be isolated from extracts of penile swabs (Gill P. *et al.*, 1985; Giusti A. *et al.*, 1986). It is also easy to determine the gender (sex) of an individual by DNA typing system. However, still little information concerning complete physical appearance or age can be discerned by

DNA analysis. In the future, it may be possible to yield information on, for example, ethnicity, age, eye colour and height.

1.2.2 Non-coding DNA

The majority of the genomes in higher eukaryotes, such as humans, (approximately 97%) consist of DNA sequences which do not code for proteins, these include satellite DNA, repetitive sequences, introns and spacer DNA between genes (Britten R. and Davidson E., 1971; Britten R. and Kohne D., 1968). In the past it has been termed 'selfish DNA' or 'junk DNA' (Doolittle W. and Sapienza C., 1980; Orgel L. and Crick F., 1980; Charlesworth B. *et al.*, 1994) though it may have unknown functions. Such DNAs contribute little to the survival of the organism and can accumulate significant amounts of deletions, duplications, or other alterations without causing any noticeable phenotypic effects. Therefore, this part of DNA displays the most extensive genetic variability and is a very important region for forensic study. A high portion of non-coding DNA is found as repetitive DNA and is dispersed throughout the human genome.

1.2.3 Repetitive DNA

The genomes of most eukaryotes consist of unique and repetitive DNA sequences (Charlesworth B. *et al.*, 1994). At least 20 to 30% of the human genome has been estimated to be repetitive DNA (Schmid C. and Jelinek W., 1982; Singer M., 1982; Fa-Ten Kao, 1985). The repeated DNA sequences are either interspersed with unique DNA sequences or exist in clusters of tandem arrays.

1.2.3.1 Interspersed repeat sequences

Large number of sequences is present in multiple copies per haploid chromosome set in the mammalian genome (Britten R. and Kohne D., 1968; Jelinek W. and Schmid C., 1982; Weiner A., 1986). These interspersed repetitive sequences include families of SINES (short interspersed repeated sequences) or LINES (long interspersed repeated sequences) and occupy some 10 to 15% of the human genome (Singer M., 1982; Jelinek W. and Schmid C., 1982).

Some of these repetitive sequences are clustered in discrete locations (Singer M., 1982; Moyzis R. *et al.*, 1987) but Davidson E. *et al.*, 1975; Schmid and Deininger, 1985 have demonstrated that the majority of this repetitive DNA is interspersed throughout higher eukaryotic genomes. The repetitive units of (SINEs) and (LINEs) are scattered throughout the genome as individual units of specific sequences and not as tandem arrays. These interspersed repeated sequences can be divided into two general types.

a) Short interspersed repeated sequences (SINES)

The size of the repeating unit is less than 500 bp, the most abundant human SINEs is the Alu DNA sequence family (Schmid C. and Jelinek W., 1982; Deininger P. *et al.*, 1992). The Alu repeat unit is composed of 300 bp in length with an *Alu*I restriction site that cleaves it into 170 bp and 130 bp fragments. The Alu units are repeated more than 500,000 times (Singer M., 1982; Sutcliffe C. *et al.*, 1982; Sapienza and St-Jacques, 1986; Hwu H. *et al.*, 1986; Deininger P. *et al.*, 1992) representing approximately 5% of the total haploid human genome which would normally occur every 5 to 10 kb (Willard H. and Waye J., 1987). Individual Alu repeats are closely related but not identical in sequences and on the average, they differ from a consensus sequence by about 14% and most of

these variants involve single base pair substitution (Fa-Ten Kao, 1985). Most Alu elements are flanked by short direct repeats of about 10 bp whose sequences are not common to the elements (Fa-Ten Kao, 1985).

b) Long interspersed repeated sequences (LINES)

The size of the repeating unit is longer than 500 bp (more heterogeneous in sequence than Alu family) and the major human LINEs is the *Kpn*I (or L1) DNA sequence family (Shafit-Zagardo B. *et al.*, 1982). The consensus structure is about 6.4 kb long and most copies have an A-rich tail at 3′ end and are flanked by short repeats. This 6.4 kb *Kpn*I family consists of approximately 1-4 x 10^4 copies per haploid genome (Fa-Ten Kao, 1985) making up 1 to 2% of the total human DNA. The mechanisms for the generation of these repetitive DNAs are not well understood.

1.2.3.2 Tandem repetitive DNA sequences

Tandem repeating DNA is characterised by blocks of DNA of some common sequence which is repeated over and over in tandem fashion. These repetitive sequences recombine and replicate, producing arrays of varying sizes by a number of mechanisms although the exact mechanisms are not known. Unequal exchange of repeated sequences between non-homologous molecules generate duplications and deletions. Replication slippage, mispairing between bases within a single DNA molecule followed by replication also gives rise to variation in array size (Dover G., 1982; Moore G., 1983).

These repetitive DNA sequences constitute approximately 10% of the total human genome and it can be divided into three parts, satellite, minisatellite, and microsatellite DNA.

1.3 SATELLITE DNA (CLASSICAL SATELLITES)

Four major classes of tandem repeat DNA have been classified within the human genome, satellite I, II, III, and IV (Singer M., 1982; Fa-Ten Kao, 1985). About 20% of the human genome composed of various satellite DNA families (Milkos G. and Joun B., 1979). Satellite DNAs were originally identified by the separate banding of part of the genomic DNA in equilibrium density gradient centrifugation. Satellite DNAs can be classified according to their genomic localisation, such as centromeric, telomeric or dispersed along the chromosome. Satellite DNA is primarily located near the centromeric regions of the chromosomes and the repeat unit length can extend for several thousand base pairs (Ballantyne J. *et al.*, 1989).

Alphoid class is another major tandemly repeating DNA species, which composed of approximately 2% of the whole genome (Willard H. and Waye J., 1987). The alphoid satellite repetitive DNA family consists of tandem arrays of approximately 171 bp monomers (Manuelidis L., 1978; reviewed in Willard H. and Waye J., 1987). These sequences are localised to the pericentromeric region of each human chromosome (Yang T. *et al.*, 1982; Willard H. *et al.*, 1983; Jabs E. *et al.*, 1984).

1.4 MINISATELLITE DNA

During the last two decades, a number of highly variable regions in the human genome have been detected and characterised. In addition to DNA sequence variants in the form of RFLPs, another significant source of detectable sequence variation are tandemly repeated sequences (minisatellites). In minisatellites the variation arising from differences in the number of repeat units (Bell G. *et al.*, 1982; Jeffreys A. *et al.*, 1985a, b) (see figure

1.2). The length of repeat units ranging from 9 to 80 bp and total length of VNTR fragments ranging from more than a hundred to over 20 kb (Nakamura Y. *et al.*, 1987).

The first hypervariable DNA region (D14S1) in the human genome with fragments more than 15 different lengths observed in 1980 (Wyman A. and White R., 1980). Such a variable region consists of tandem repeats of a short sequence known as Variable Number Tandem Repeats (VNTRs) (Capon D. *et al.*, 1983; Nakamura Y. *et al.*, 1987b). Due to the organisational similarity between these hypervariable regions and satellite DNAs (Brutlag D., 1980), they are also known as "minisatellites" (Jeffreys A. *et al.*, 1985a).

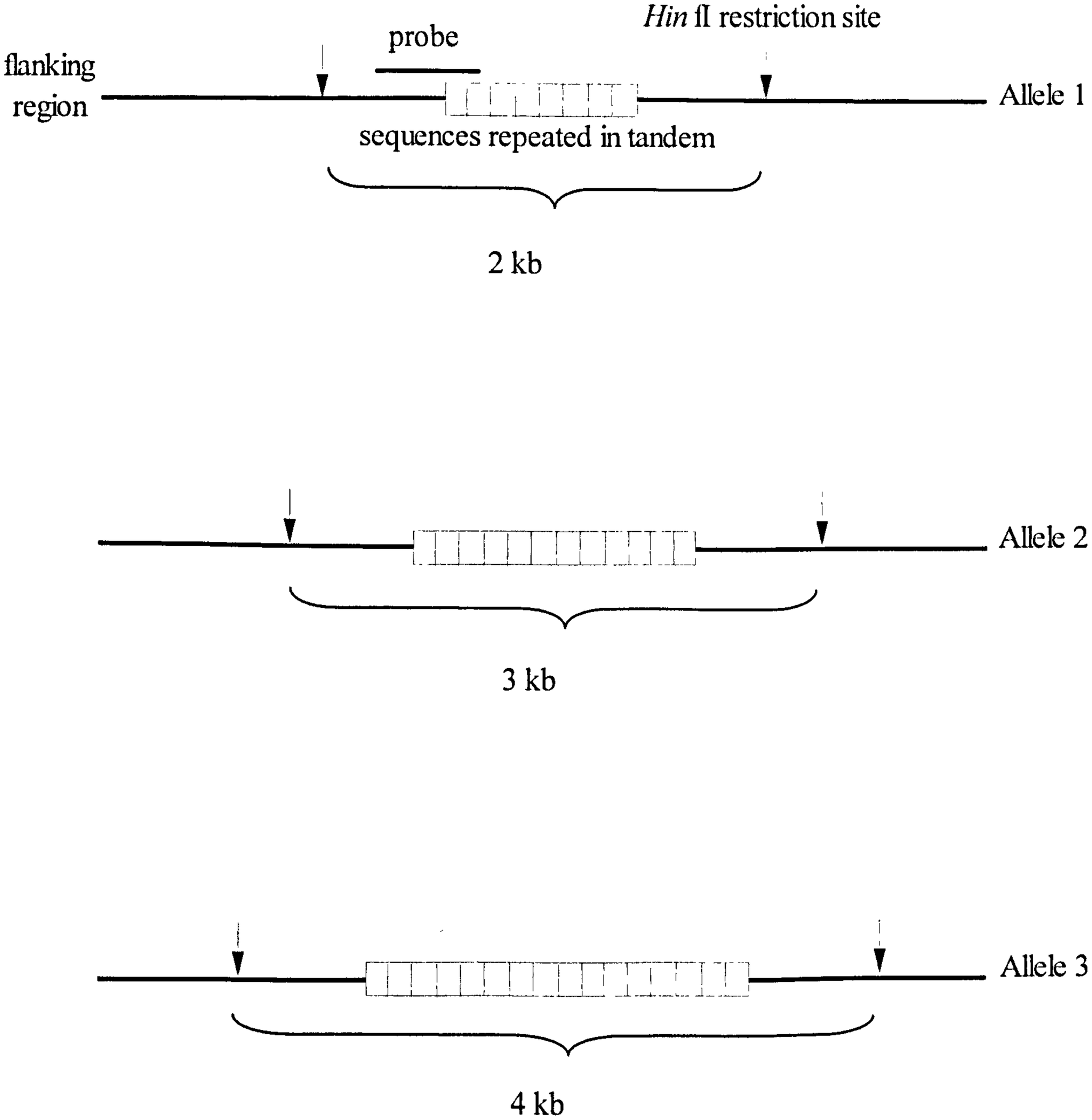
Polymorphisms at these regions is based on allelic differences in the number of short repeated sequences, and is presumably generated by mitotic or meiotic unequal recombination, sister chromatid exchange at mitosis in the germ line or by DNA slippage during replication. It is unlike the typical RFLPs generated by an alteration (gain or loss) of the restriction enzyme cutting sites which produce just two alleles and three genotypes, the number of alleles found at these VNTR loci can be very large. More than 70 alleles at a single locus can be recognised in the population (Wong Z. *et al.*, 1986; Jeffreys A., 1993) and consequently heterozygosity values can reach almost 100% (Krontiris T., 1995) compared to 50% or less (Nakamura Y. *et al.*, 1987a) with conventional RFLPs. As a result, 'Minisatellite' RFLPs based on VNTR loci are more variable and thus more discriminating of individuals than RFLPs from 'classical' satellites (Fowler C. *et al.*, 1988).

11 A

Figure 1.2 A diagram of structure of VNTR alleles in chromosome.

The allelic variation is resulting from a variable number of units repeated in tandem.

The DNA probe is a unique sequence in which detects a restriction fragment for each allele. The length of the fragments depends on the number of repeating units present.


Hin fI restriction site
probe
flanking region
Allele 1
sequences repeated in tandem
2 kb
Allele 2
3 kb
Allele 3
4 kb

1.4.1 Minisaellite and restriction enzyme

Restriction endonuclease enzymes which, were identified first in 1970 (Kelly T. and Smith H., 1970), played a very important role in recombinant DNA techniques and since then hundreds of endonucleases have been isolated from different bacterial species.

To demonstrate the size differences between different minisatellite alleles it is necessary to isolate them from their surrounding DNA into relatively small fragments, but which leaves the minisatellite DNA intact. The distance between the repeating units and the restriction sites in the flanking DNA is the same, if no mutations occur, for all alleles. Therefore, the polymorphisms are due to the variation in the number of repeating units and not in the position of the restriction sites (figure 1.2).

Several restriction enzymes were used in forensic community such as *Hae*III, *Hin*fI, and *Pst*I. However, *Hae*III generates smaller bands and *Pst*I and *Hin*fI generate larger bands. These restriction enzymes will demonstrate a similar degree of polymorphism but different band sizes. *Hae*III is a widely used in north America forensic laboratories and *Hin*fI is widely used in European forensic laboratories.

1.4.2 Generation of polymorphisms

As with other satellite sequences, the mechanism by which minisatellites form and evolve is not yet well understood. However, circumstantial evidence from several studies suggests that a variety of mechanisms are involved. The relative importance of each mechanism will vary with the location, length and complexity of the particular minisatellite.

The hypervariabilty at these minisatellites in the human genome results from changes in the number of repeats, presumably driven either by unequal (sister chromatids) exchange at meiosis (Jefrreys A. *et al.*, 1999) or by slippage at replication forks leading to the gain or loss of repeat units (Schlotterer C. and Tautz D., 1992).

Jeffreys reported high levels of germline instability responsible for deriving minisatellite hypervariability into human populations (Jeffreys A. *et al*, 1988). Therefore, germ-line instability must be taken in to account when using hypervariable loci as genetic markers, particularly in pedigree analysis and parenthood testing.

1.4.3 Mutation rate (minisatellite instability)

In general, the mutation rate varies dramatically from locus to locus and in non coding DNA region mutation rates are much higher than in coding DNA. The high level of variability of minisatellites must reflect a high degree of instability at these loci. The germ line mutation rate increases with heterozygosity. Both gain and loss of minisatellite repeat units have been observed, the length changes are usually small but can occasionally involve substantial number of repeats (Jeffreys A. *et al.*, 1999). The most variable locus (D1S7), detected by MS1 probe, showed the highest germline mutation rate of about 5% per gamete (Jeffreys A. *et al.*, 1988). To use single locus minisatellite probes as genetic markers, the degree of germline and somatic instability at these hypervariable loci must be taken into consideration. Significant mutation rates at marker loci may lead to false exclusion of genuine relationship in, for example, paternity analysis but not for individual identification (Odelberg S. *et al.*, 1989). VNTRs have a very high mutation rate, leading to changes in length. An individual mutation usually changes the length by only one or few repeating units, resulting in a very large number of alleles, no one of which is common.

1.4.4 Minisatellite DNA probes

There are tens of probes now commercially available for forensic community to use which are able to detect tens of polymorphic sites on different chromosomes. These probes can be divided into two categories, multi-locus and single locus probes.

1.4.5 Multi-locus DNA probes

Multi-locus probes (MLPs) hybridise to many different sites in the genome and are thus capable of identifying polymorphism at many loci at a time. Usually this can be achieved by Southern blot hybridisation at low stringency. MLPs such as 33.15 and 33.6 were the most widely used in forensic (Gill P. and Werrett D., 1987) and paternity (Jeffreys A. *et al.*, 1991; Debenham P., 1991) casework. These two probes can detect approximately 60 hypervariable loci simultaneously (Jeffreys A. *et al.*, 1986; Jeffreys A., 1987) where each fragment represent only one allele and only the DNA fragment sizes ranging between > 3 to 20 kb were clearly resolvable by gel electrophoresis (Jeffreys A. *et al.*, 1985a; Jeffreys A., 1987; Brinkmann B., 1991).

On average 36 resolvable fragments (> 3 kb) per individual can be detected using both MLPs 33.15 and 33.6 (Jeffreys A., 1987). The shorter hybridizing DNA fragments are probably derived from short and comparatively less variable minisatellites. Therefore, the probability of shared bands between two randomly selected individuals increases for smaller minisatellite fragments (Jeffreys A. *et al.*, 1985a, b). However, a very unique profile can be obtained when using these two MLPs, even when closely related individuals from isolated and inbreed communities have been tested the chance that two people (other than identical twins) share the same pattern was extremely low (Jeffreys A., 1993). Hence Jeffreys termed it a "DNA fingerprint" (Jeffreys A. *et al.*, 1985a).

1.4.6 Limitation of MLPs

Although MLPs are a very powerful and useful tool in forensic and paternity analysis and give a high level of discrimination, there are some limitations which reduce their full usefulness. For example, many MLPs have been shown to cross-hybridise to polymorphic loci in a wide range of animal, bird and plant species. This technique also requires a good quality DNA (approximately 500 ng) and such material is often not available to the forensic scientist where the DNA is often partially degraded or is recovered in amounts too small for MLP detection. Interpretation and statistical value of a mixed or partially degraded sample is complicated. These limitations have been over come by the development of single locus probes (SLPs) (Nakamura Y. *et al.*, 1987b).

1.4.7 Single locus DNA probes

The next technique to be used was single locus probes (SLPs) such as YNH24, MS1, MS31, MS43A, G3, MS621, MS205, MS8, and others. Each SLP is hybridised at high stringency and detect only two bands (alleles) from the specific locus at each time, one band inherited from the mother and one from the father. An alternative term, DNA profiling, has been suggested because the pattern is not individual specific. Eventually SLP became the probe of choice especially in forensic field. SLPs have several advantages over MLPs. For example, it generates simpler and more easily interpreted results, and technically more robust for use with dirty evidence sample. It is easier to identify the presence of a mixture of biological samples from two or more individuals. It is almost human specific and rarely cross-hybridise with DNA from other species.

In order to obtain high discrimination power a battery of SLPs (typically five different SLPs) can be analysed sequentially (Jeffreys A., 1993). The major disadvantage

is that one SLP will give much lower discrimination power if compared to MLP. It extends the amount of time it takes to complete a test of several SLPs and relatively increases the cost.

1.4.8 Chemiluminescent detection (non-isotopic system)

There are several advantages of using VNTR probes (non-isotopic system) which are chemiluminiscently detected than using the isotopic system which was radioactively (^{32}p) labelled (Schaap A. *et al.*, 1989; Giles A. *et al.*, 1990). The use of non isotopic system enhanced the intensity and clarity of DNA fragments and greatly reduced the exposure time of the X-ray film required for visualisation (from 1-2 weeks instead of 4-6 weeks for a complete analysis) (Benzinger E. *et al.*, 1995; NRC, 1996). Chemiluminescent is simple and safe to be handled. Therefore, it does not require radioactive license and also eliminates storage and disposal concerns (Risen L. *et al.*, 1991). In addition, radioactive probes have a relatively short life span compared to chemiluminiscently lablled probes (Baum H. *et al.*, 1990).

1.4.9 Forensic applications of minisatellites

The detection of DNA polymorphisms by highly polymorphic minisatellite loci (Jeffreys A., *et al.* 1985a, b; Nakamura Y, *et al.* 1987b) represented a major impact on the field of forensic science (Gill P. *et al.*, 1985; 1987; Gill P. and Werrett D., 1997; Budowle B. *et al.*, 1990). In paternity testing (Baird M. *et al.*, 1986; Jeffreys A., 1987; Diamond J., 1987; Smith J. *et al.*, 1990b; Jeffreys A. *et al.*, 1991) and also in immigration disputes (Jeffreys A. *et al.*, 1985c). The presence of a large number of alleles at each locus (Jeffreys A. *et al.*, 1985a; Balazs I. *et al.*, 1989) made VNTR the most variable and informative genetic markers yet discovered. The probability of a matching DNA profile

between unrelated individuals is extremely small, especially at several VNTR loci. Related individuals, in particular identical twins and siblings, have a far greater probability of matching genotypes. In general, analysing several loci produces a DNA profile that is extremely uncommon. Therefore, these hypervariable DNA regions provide the forensic scientist an excellent tool to exclude an individual who has been falsely associated with an evidentiary sample.

The first use of minisatellites was in 1985, which involved an UK immigration dispute and had been satisfactory resolved (Jeffreys A. *et al.*, 1985a). It was applied to a criminal case for the first time early in 1986 in the UK and it was not used in the USA until late 1986. In November 1983 a schoolgirl had been found raped and murdered on wastegrond in a small village near Leicester. In July 1986 a second girl was found in the same area. A man was arrested and charged with the murder of the second girl after making a confession (Gill P. and Werrett D., 1987). The police believed him also to be responsible for the first murder and asked Jefferys to carry out DNA profiling. The suspect's DNA profile, however, did not match that of semen from both the scenes of crime. Therefore he could not have committed either offence. Blood from all the males in the area (a total of 1500 individuals) was analysed, but without success. Towards the end of the exercise a suspect was apprehended by police as a result of a conversation overheard by chance in a public house. The suspect's blood was tested and his profile matched that of the crime material.

1.4.10 Matching and binning criteria

Different fragment sizes can be observed from a population of individuals exhibiting variation in length and each distinct fragment length defines an allele, but any given fragment is subject to measurement error. Therefore VNTR alleles are not discrete but are

continuous. Each locus exhibits substantial variation of fragment size with two alleles per individual. It is important to estimate the frequencies of these alleles before applying the system to the field of forensic and paternity testing. However, the problem arises by the fact that each fragment cannot be measured without error when using agarose gel electrophoresis. Therefore, one would observe phenotypes rather than genotypes and phenotypes of different alleles may be identical.

To overcome this limitation these fragment sizes in population data are assigned to a relatively small number of bins (group of similar size alleles). Two common methods have been adopted, fixed bin and floating bins (Balazs I. *et al.*, 1989), are often used in forensic laboratories. Fixed bins have been developed by Budowle (Budowle B. *et al.*, 1991b, c) from the FBI and it is most widely employed in the USA because this system is more conservative and easier for the average laboratory to use (Monson K. and Budowle B., 1993). Although floating bins are more accurate (it is require searching the whole database for each calculation) but in general the two methods provide a similar statistical value. The two procedures (fixed and floating bins) are recommended by the NRC (NRC, 1996).

In the fixed bin procedures, fragments (alleles) of similar molecular weight are pooled into one of the 31 fixed bins. The boundaries for these bins are determined by 30 known fragment sizes (size standard markers or allelic ladder) which is run on the gel flanking the unknown samples. Then adjacent bins with no or few alleles are combined (rebinned) to produce a grouped frequency distribution for each locus. Each bin is considered as a discrete allele and by this means it is possible to determine the frequency of each allele (bin) within a population. In the floating bin procedure, a window of

measurement uncertainty can be made around the unknown fragment and the frequency of that bin is the total proportion of alleles in the database that are within this window.

The matching process consists of determining whether two fragments (alleles) are close enough to be within the limits of the measurement uncertainty or not. A match between two DNA fragments can be declared by visual examination first then can be confirmed by computer assistance if necessary. When a match is declared then those alleles (bin) frequency from the database can be calculated.

1.4.11 Profile frequency

DNA profiling means to determine the genotype of a person at several loci in the human genome and it is consider as a major advance in individual identification. The technique of sampling a number of individuals chosen randomly in the population is used to help estimate the match probability of a particular DNA profile in the general population. Large databases are preferable to small ones. However, it is generally accepted that 200 characteristics or alleles which can be obtained from 100 people should be sufficient to provide reasonable estimates of the proportions of these characteristics at each VNTR region (ISFH 1991 report, 1992; NRC recommendation, 1992; Eriksen B. and Svensmark O., 1994b).

A standardised set of single locus probes has been developed for Europe, enabling laboratories to compare databases of estimated DNA fragment lengths. In north America, the FBI has similarly developed a somewhat different panel of single locus probes. Since European and north American sets of markers and restriction enzymes are not the same, it is not possible to compare database information between the two continents.

Minisatellite analysis can provide extraordinarily powerful evidence for courts to evaluate, further, biostatistical evaluations of the evidence indicating that the matching pattern between the forensic specimen and the defendant is extremely rare in the population, occurring perhaps in one person in a million to one in a billion (NRC, 1996).

1.4.12 Debate over the use of VNTRs in forensic world

Even a two loci VNTR profile can be very rare. When using the FBI's methods the probability of a chance match of unrelated individuals for five VNTR loci the power of discrimination can reach $<1/10^{12}$ (Risch N. and Devlin B., 1992) assuming that these loci are in Hardy Weinberg equilibrium and observing independence within and between VNTR loci. This estimation has resulted in considerable controversy to the forensic use of VNTR loci especially when dealing with USA population which consists of several major ethnic groups. It has been criticised by some experts, who argue that these loci are not in Hardy-Weinberg equilibrium (Lander E., 1989, 1991a, b; Cohen J., 1990; Cohen J. *et al.*, 1991; Geisser S. and Johnson W., 1993). In addition, they claimed that the excess of homozygotes in forensic databases are due to mixture of subpopulations within each major ethnic group (Caucasian, African and Hispanic) which could indicate lack of allele independence within and between VNTR loci. On the other hand Devlin and Chakraborty (Devlin B. *et al.*, 1990 and 1992; Chakraborty R. *et al.*, 1992a, b ; Chakraborty R. and Jin L., 1992) have shown that an excess of homozygotes not necessarily due to population sub-structuring. It could be a consequence of analytical limitations such as allele coalescence or undetectable alleles (Steinberger E. *et al.*, 1993).

Lewontin and Hartl (Lewontin R. and Hartl D., 1991) argue that the existence of subpopulations will lead to substantial variation in allele frequency within each major racial group at each VNTR locus (Lewontin R., 1972). Therefore, the estimation

probability of a match between two unrelated individuals within a certain group might not be correct or reliable and may be overestimated if one major population database was use for the whole subpopulations (Lewontin R. and Hartl D., 1991). Chakraborty and Kidd on the other hand disagreed on that and declared that the approximations of allele frequency for the existing population data are close enough to the true value and valid estimate can be calculated even if the reference population is a mixed one (Chakraborty R. and Kidd K., 1991). Caskey supported Chacraborty and Kidd's observations as well (Caskey C., 1991). Moreover, Budowle indicated that the allele frequency variation between major ethnic groups is greater than within a group (Budowle B. and Monson K., 1994). Weir and Devlin stated that the databases show very little or no evidence of disequilibrium, either within or between loci (Devlin B. *et al.*, 1990 and 1991; Devlin B. and Risch N., 1992 and 1993a, b; Weir B.,1992a, b, c). Herrin supported Weir and Devlin's observations (Herrin G., 1993).

To resolve this argument among population geneticists the National Research Council (NRC) in the USA released a report (May 1992) and proposed a method to account for population substructure. They suggested using the ceiling principle in forensic applications of VNTR analysis, which entails the study of 100 individuals from each of 15 to 20 relatively homogenous subpopulations representing the racial and ethnic diversity of groups in the USA. For each allele, the maximum frequency among the groups sampled, or 5%, whichever is larger, would be chosen. The product or multiplication rule then would be applied to determine the profile frequency. They also recommended that until the ceiling principle could be implemented, an interim ceiling principle, where the maximum allele probability that occurs in the existing databases for at least three major ethnic groups or 10% whichever is larger should be applied. The NRC claimed that the ceiling principle to be conservative and giving estimates greater than or equal to the actual

genotype frequencies in the appropriate reference population. However, this claim has been widely criticised by many of the scientific community (Cohen J., 1992; Weir B., 1993; Balazs I., 1993; Devlin B. *et al.*, 1993, 1994; Morton N. *et al.*, 1993; Kaye D., 1993; Slimowitz J. and Cohen J., 1993; Evett I. *et al.*, 1993; Collins A. and Morton N., 1994), and supported by others (Lempert R., 1993; Lander E. and Budowle B., 1994). By 1995 more information was available from a number of relevant populations and therefore experts had the ability to use an appropriate database. In summary, the ceiling principle was rejected by most of the experts. Alternatively, there are two conservative methods which generally have been accepted and widely used in forensic field, fixed bin and floating bin methods. For a single band observation, the value 2p was applied instead of p^2 to become more conservative (NRC, 1992).

1.4.13 Minisatellite limitations

Despite the fact that VNTR systems are very informative in the field of forensic and paternity analysis, technical problems have again prevented its full use. Minisatellite fragment lengths can vary in a quasi-continuous fashion and not as a discrete allele, making exact allele size identification impossible. Therefore, a single fragment can be observed due to either true homozygosity or due to a heterozygote with the two fragments similar in size, but not identical, and the distance between the two alleles is very close so that the distinct bands blur together or coalesce. Large fragment pairs with a small repeat unit (such as D1S7 locus) are more likely to coalesce than smaller fragments (Devlin B. *et al.*, 1990). In forensic cases it is possible that the very large alleles may be lost in partially degraded samples and in some other cases small alleles might migrate off the end of the gel or the probe cannot detect these small alleles (Budowle B. *et al.*, 1991b). It is difficult

to distinguish between a blank gel lane caused by two null alleles (very small alleles) and technical difficulties or because of mutation at a single locus.

The two main limitations of VNTRs in forensic science are the requirement for high molecular weight DNA in high quantity (>50 ng) which is not always possible to get in forensic samples. The other main limitation is the relatively long time required to complete an analysis, which might varies from one to two weeks.

1.5 POLYMERASE CHAIN REACTION (PCR)

The development of the polymerase chain reaction (PCR) (Saiki R. *et al.*, 1985; 1988; Mullis K. and Faloona F., 1987) for the *in vitro* amplification of specific DNA sequences by enzyme catalysed reaction has made possible a number of rapid approaches for determining DNA polymorphisms (Sensabaugh G., 1991). First, double stranded DNA is denatured to single strands. These are then hybridised to specific primers situated either side of the sequence of interest. A heat stable DNA polymerase (e.g. *Taq* polymerase) that can withstand the heat denaturation step is used to synthesize complementary DNA using the existing strand as a template. This is followed by re-melting to produce more templates for the next cycle. Therefore, the main purpose of using PCR method in the field of forensic is to amplify or copy specific segments of DNA which are of interest, to provide enough DNA for analysis. Theoretically and practically PCR can be used to amplify genetic material from as little as one cell (Saiki R. *et al.*, 1988; Li H. *et al.*, 1988; Higuchi R. *et al.*, 1988).

There are several systems of DNA polymorphism that can be used in forensic and paternity testing based on PCR amplification of specific target DNA sequences and these

systems are HLA-DQα, polymarker (PM), Amp-FLPs, MVRs, mtDNA (forensic use only), Y-STR, and autosomal STRs. See figure 1.3.

1.6 MICROSATELLITE DNA (AUTOSOMAL STRs)

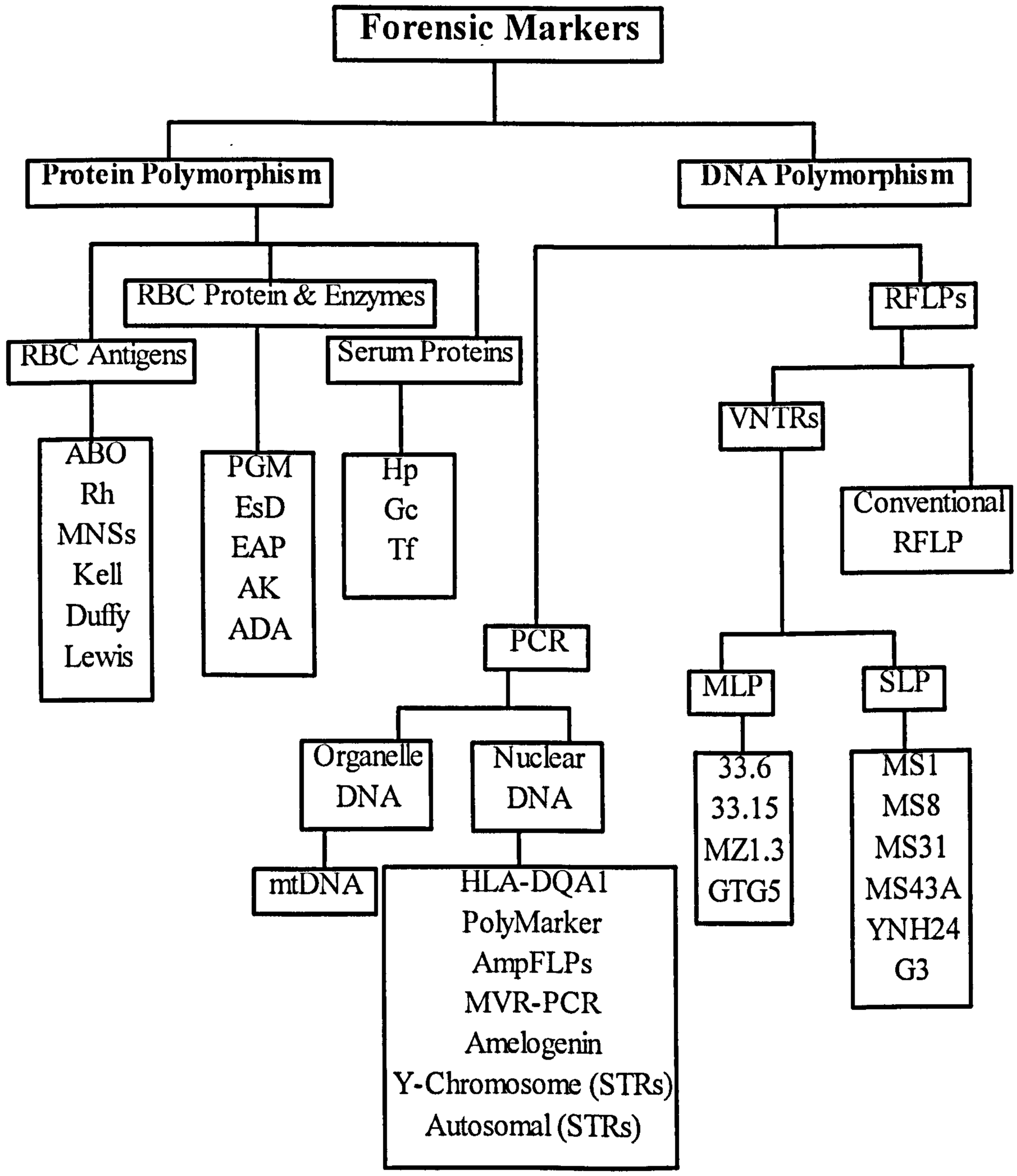
Microsatellite DNA or short tandem repeat (STR) loci, like minisatellites, consist of tandem repeated sequences. However, the repeat unit is shorter and typically 2-7 base pairs long (Edwards, A. *et al.*, 1991a, b and 1992; Litt M. and Luty J., 1989; Weber J. and May P., 1989; Tautz D., 1989; Craig J., 1988). This type of polymorphism is a length polymrphism as in VNTRs and occurs on average every 6-10 kb throughout the human genome (Edwards A. *et al.*, 1991a, b; Edwards A., 1991) with approximately 500,000 STR loci in total (Beckmann J. and Weber J., 1992; Oldroyd N. *et al.*, 1995). The mono and di-nucleotide repeat sequences are the most abundant microsatellites (Beckmann J. and Weber J., 1992). STR loci can be amplified by PCR to yield very short, usually less than 400 bp, but nevertheless variable DNA fragments. STR alleles can be sized with high level of precision, opening the possibility of a precise and accurate definition of matching profiles (Edwards A. *et al.*, 1991a, b; Litt M. and Luty J., 1989; Weber J. and May P., 1989).

Microsatellite loci are extremely useful when analysing samples that contain very small amounts of highly degraded DNA as in mass disaster (Whitaker J. *et al.*, 1995; Goodwin W. *et al.*, 1999) and minute stain analysis (Wiegand P. *et al.*, 1993; Gill P. *et al.*, 1994a; Monaghan B. and Newhall P., 1996). Microsatellites are less variable with less number of alleles than minisatellites and therefore less discriminative, but good individual specificity can be achieved by typing a number of loci. Severely degraded DNA can also

R5A

Figure 1.3 Diagram showing forensic markers used in forensic analysis and paternity testing.

Rh = Rhesus system, PGM = Phosphoglucomutase, Ak = Adenylate kinase, ADA = Adenosine deaminase, EsD = Esterase D, GLO = Glyoxalase I and Tf = Transferrin.


Forensic Markers
Protein Polymorphism
DNA Polymorphism
RBC Protein & Enzymes
RBC Antigens
Serum Proteins
RFLPs
ABO
Rh
MNSs
Kell
Duffy
Lewis
PGM
EsD
EAP
AK
ADA
Hp
Gc
Tf
VNTRs
Conventional RFLP
PCR
MLP
SLP
Organelle DNA
Nuclear DNA
33.6
33.15
MZ1.3
GTG5
MS1
MS8
MS31
MS43A
YNH24
G3
mtDNA
HLA-DQA1
PolyMarker
AmpFLPs
MVR-PCR
Amelogenin
Y-Chromosome (STRs)
Autosomal (STRs)

be typed (Jeffreys A. *et al.*, 1992; Hochmeister M. *et al.*, 1995; Zierdt H. *et al.*, 1996). An example is DNA from the skeletal remains of a murder victim buried for eight years have already been typed by microsatellite system (Hagelberg E. *et al.*, 1990). Furthermore, the possibility to use DNA prepared from buccal swab samples for STR analysis instead of drawing blood and the speed analysis, which can be rapid and cheap compared to other systems, made STRs attractive for paternity and forensic analysis (Schumm J. *et al.*, 1993; Micka K. *et al.*, 1996; Hochmeister M. *et al.*, 1996; Rostedt I. *et al.*, 1996). In addition, the allelic window is small and that greatly reduces the allelic dropout where the larger allele might not be detectable.

1.6.1 Types of STR loci

STR loci can be divided into four categories depending on the uniformity of their repeat sequences (Urquhart A. *et al.*, 1994; Gill P. *et al.*, 1995).

1.6.1.1 Simple STR loci

The different versions or alleles for each locus differ in steps of one complete repeat unit (repeat units of identical length and sequence), with only few exceptions. Examples of this type are HUMTHO1, HUMFES/FPS, D5S818, D13S317, D7S820, D16S539, TPOX and CSF1PO loci. The simple STR can be further subdivided into simple consisting of one repeating sequence (HUMFES/FPS) and simple with non-consensus alleles (HUMTHO1).

1.6.1.2 Compound STR loci

This type of STR loci consists of two or more adjacent simple repeat sequences (HUMGABRB15) or consists of compound with non-consensus alleles

(HUMvWFA31/A). Further examples of this type are D18S51, D8S1179, D2S1338, D3S1358 and D19S433 loci.

1.6.1.3 Complex STR loci

This type of STR loci consists of regular tetranucleotide repeat units with interspersion of dimer, trimer and hexamer invariants. Examples of this type are D21S11 and HUMFIBRA/FGA loci.

1.6.1.4 Hypervariable STR loci

These highly polymorphic loci contain complex compound repeat regions and may display alleles which differ by only 1 bp. This type of STR loci consists of tetranucleotide repeats with different monomer, dimer, trimer and hexamer invariants that scattered throughout the locus. Examples of this type are HUMACTBP2 (SE33) and D11S554 loci.

Of the four groups, simple STRs are generally the least discriminating in terms of identifying one individual from other. However, they are very simple to interpret. Therefore, the majority of STR loci employed in forensic analysis are simple STRs. See table 1.

Table 1 different types of STR loci

The simple repeat consists of one regular 4 bp repeats. The compound type consists of two or more different 4 bp repeats. The complex repeat consists of several regular 4 bp repeat units with interspersion of di, tri and hexamer invariants. The hypervariable repeats consists of regular 4 bp repeats with different mono, di, tri and hexamer invariants scattering throughout the locus.

Type	Locus	Repeat
Simple	D5S818	$(AGAT)_{7-16}$ Allele repeat unit: AGAT
Compound	HUMvWA	$(ATCT)_2(GTCT)_{3-4}(ATCT)_{9-13}$ Allele repeat unit: ATCT
Complex	D21S11	$(TCTA)_{4-6}(TCTG)_{5-6}$ [$(TCTA)_3TA(TCTA)_3TCA$ $(TCTA)_2TCCA$ TA] $(TCTA)_{8-16}(TATCTA)_{0-1}TC$ Allele repeat unit: TCTR (R = A or G)
Hypervariable	ACTBP2	A common repeat structure (AAAG) with different mono, di, tri, tetra and hexamer invariants that are scattered throughout the locus.

1.6.2 Detection of STR loci

The polyacrylamide gel electrophoresis (PAGE) method has been widely used to separate the PCR products. At present another method is in common use such as capillary electrophoresis which is filled with polymer. Amplified products can be detected by direct staining using dyes such as ethidium bromide (EtBr) and silver staining (Allen R. *et al.*, 1989; Walsh P. et al., 1992). Silver staining of polyacrylamide gels was initially used to detect unlabelled primers giving a good sensitivity of detecting and visualising alleles (Bassam B. *et al.*, 1991). Several loci can be detected at a time since STR alleles fall into discrete size categories and therefore the alleles will not overlap. However, when the

primers are labelled with fluorescent dyes (Fregeau C. and Fourney R., 1993; Tully U. *et al.*, 1993) a large number of STR loci can be analysed simultaneously and alleles differing from each other by only one base pair are distinguishable.

There are now five different dyes available which can be used simultaneously, one dye usually is used for size standard and the other four can be used to detect STR loci. Overlapping loci can therefore be distinguished based on the colour of fluorescence that is detected. At present there is already a commercial kit where sixteen loci can be detected plus an internal size standard by using four such dyes. To designate the different alleles a size standard can be run and hence the sizes of each allele can be calculated or by running an allelic ladder so alleles can be designated compared to the ladder.

There are two main advantages of using fluorescent detection over silver staining, more loci can be analysed and that the internal size standard can be run in the same lane as the PCR. Hence alleles can be calculated with higher precision. The other advantage of using fluorescent dye is that it is less time consuming.

1.6.3 STR multiplex systems

Multiplexing or co-amplification, meaning to amplify simultaneously more than one locus in a single tube and under the same conditions, was first introduced by Chamberlain *et al.* in 1988 (Chamberlain J. *et al.*, 1988). There are a number of multiplexing kits which are able to amplify from 3 to 16 loci simultaneously have been reported and available to forensic laboratories to use. These kits have been developed through several agencies such as the UK Forensic Science Service (FSS), Applied Biosystems Division (Perkin Elmer, Foster city, CA, USA) and Promega Corporation (Madison, Wisconsin, USA). The Quad (Quadroplex) was the first standardised STR analysis as a multiplex

amplification of the four simple STR loci (HUMTHO1, HUMvWA31/A, HUMF13A01, and HUMFES/FPS). This kit was introduced in 1994 by the Forensic Science Service in the UK (Kimpton C. *et al.*, 1993; 1994a, b; Lygo J. *et al.*, 1994). Although, the Quad is far less discriminating than the earlier method such as VNTRs it had considerable forensic potential use. A Second Generation Multiplex or SGM system (heptaplex) was introduced in July 1996 by Forensic Science Service which is able to co-amplify 6 STR loci (HUMTHO1, HUMvWAF31/A, HUMFIBRA/FGA, D8S1179, D18S51 and D21S11) plus a sex determining locus simultaneously (Kimpton C. *et al.*, 1996; Evett I. *et al.*, 1997). The SGM system had improved discrimination over the Quad with the similar level of sensitivity (Sparkes R. *et al.*, 1996 a, b).

Applied Biosystems Division has developed several kits which amplify three and four STR loci. However, in 1997 the company designed a kit which is commercially available now and called AmpFlSTR® Profiler Plus™. This kit can detect nine tetranucleotide STR loci (HUMvWA, HUMFGA, D3S135, D5S818, D7S820, D8S1179, D13S31, D18S51 and D21S11) plus gender test. Another kit was reported which was called AmpFlSTR® Cofiler™. This kit is able to detect 6 STR loci (HUMTHO1, HUMTPOX, CSF1PO, D3S135, D7S820 and D16S539) plus amelogenin locus. Using these two kits together can increase dramatically the power of discrimination in forensic application. Two loci of this kit which are duplicated in the Profiler plus and Cofiler can provide a high probability that results from both amplifications have originated from the same specimen. A third kit (SGM plus) was developed which detects 10 STR loci (HUMvWA, HUMTHO1, HUMFGA, D19S433, D3S135, D65477, D8S1179, D16S539, D18S51 and D21S11) plus the amelogenin gender marker. Shortly (June 2001) a kit called Identifier is going to be available which is able to detect 15 STR loci plus amelogenin.

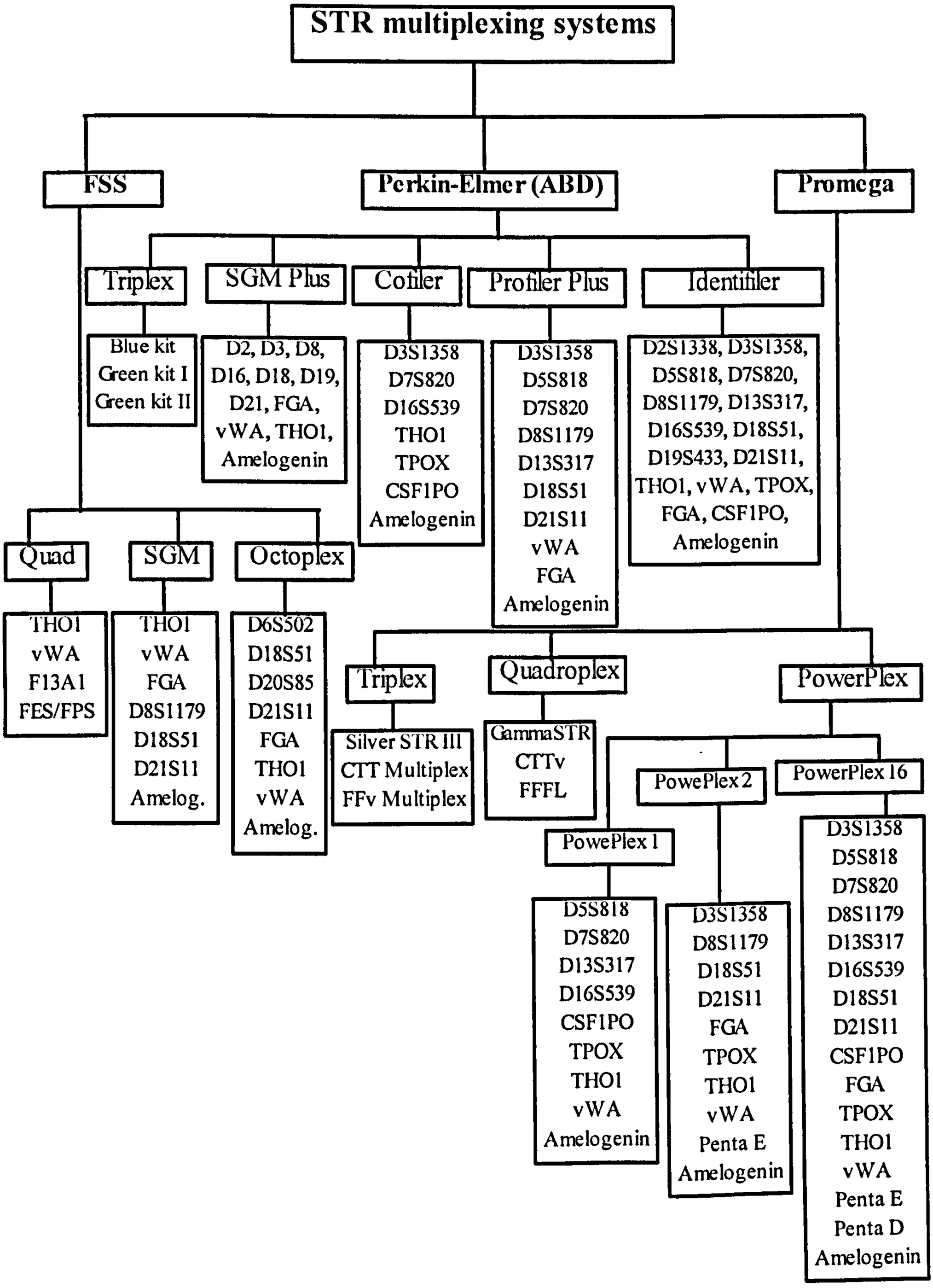
Promega has designed several STR kits as well since 1993 when first STR system was designed using silver staining. The first multiplex system was available commercially in 1994 (CTT system) which detected only three STR loci. Fluorescent STRs were started in 1995 and the first megaplex became available with the introduction of the PowerPlex™ System in 1997. PowerPlex™ 1.2 System became available in 1998 which allowed the amplification of eight STR loci (HUMTHO1, HUMvWA, HUMTPOX, HUMCSF1PO, D5S818, D7S820, D13S317 and D16S539) plus amelogenin locus. In May 2000 the PowerPlex™ 16 System released which detects 15 STR loci (HUMTHO1, HUMvWA, HUMTPOX, HUMCSF1PO, HUMFGA, Penta E, Penta D, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51 and D21S11) plus amelogenin locus using three fluorescent dyes (Sprecher C., *et al.*, 2000). See figure 1.4.

In the USA, the FBI has established the Combined DNA Index System (CODIS). This system uses two kits (ProfilerPlus and Cofler or PowerPlex 1 and 2) in order to amplify thirteen different STR loci (Budowle B. *et al.*, 1999). These STR loci selected as standard sets of loci to be profiled from millions of individuals and stored in centralised searchable databases in order to link suspects to crime scenes through STR profiling.

These new multiplex STR kits had dramatically increased the discrimination power of PCR testing and in the past it has been lower than RFLP testing but now the discriminating power is comparable if not better. Additionally, with the new PowerPlex 16 system it is now possible to amplify all 13 CODIS STR loci in a single reaction.

32A

Figure 1.4 Diagram showing commercial and non-commercial forensic STR multiplexes kits.


STR multiplexing systems
FSS
Perkin-Elmer (ABD)
Promega
Triplex
SGM Plus
Cofiler
Profiler Plus
Identifiler
Blue kit
Green kit I
Green kit II
D2, D3, D8, D16, D18, D19, D21, FGA, vWA, THO1, Amelogenin
D3S1358 D7S820 D16S539 THO1 TPOX CSF1PO Amelogenin
D3S1358 D5S818 D7S820 D8S1179 D13S317 D18S51 D21S11 vWA FGA Amelogenin
D2S1338, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D19S433, D21S11, THO1, vWA, TPOX, FGA, CSF1PO, Amelogenin
Quad
SGM
Octoplex
THO1 vWA F13A1 FES/FPS
THO1 vWA FGA D8S1179 D18S51 D21S11 Amelog.
D6S502 D18S51 D20S85 D21S11 FGA THO1 vWA Amelog.
Triplex
Quadroplex
PowerPlex
Silver STR III CTT Multiplex FFv Multiplex
GammaSTR CTTv FFFL
PowePlex 1
PowePlex 2
PowerPlex 16
D5S818 D7S820 D13S317 D16S539 CSF1PO TPOX THO1 vWA Amelogenin
D3S1358 D8S1179 D18S51 D21S11 FGA TPOX THO1 vWA Penta E Amelogenin
D3S1358 D5S818 D7S820 D8S1179 D13S317 D16S539 D18S51 D21S11 CSF1PO FGA TPOX THO1 vWA Penta E Penta D Amelogenin

1.7 ADDITIONAL PCR SYSTEMS

In addition to allelic variation in repeat number, polymorphisms at minisatellite and microsatellite loci can also be caused by sequence changes in the sequence repeat units. An approach has been developed to detect variation within the repeat units of different minisatellite alleles and it is known as minisatellite variant repeat – polymerase chain reaction (MVR-PCR) (Jeffreys A. *et al.*, 1991; Neil D. and Jeffreys A., 1993; Yamamato T. *et al.*, 1994; Tamaki K. *et al.*, 1995).

1.7.1 HLA-DQα

The human leukocyte antigen HLA-DQα gene (Gyllensten U. and Erlich H., 1988) was the first to be analysed routinely using PCR (Saiki R. *et al.*, 1986 and 1988) for human identification purposes (Hochmeister M. *et al.*, 1991; Reynolds R. *et al.*, 1991; Blake E. *et al.*, 1992; Comey C. *et al.*, 1993). The HLA-DQα kit was produced in 1990 which detect only 6 alleles but later another modified version was designed with 7 alleles. This gene shows substantial variation in its base sequence from individual to individual (Ballantyne J. *et al.*, 1989).

The size of the amplification products is either 242 bp or 239 bp. Although it is a rapid, easy to use, sensitive and potentially useful test for detecting exclusions in forensic analysis it has limited statistical power for positive identification and with relatively high cost.

1.7.2 PolyMarker (PM)

Simultaneous amplification and typing of a number of loci using several sets of primers in a single test tube under the same set of conditions (multiplexing) has enhanced the capabilities for human identification testing cases. AmpliType® PM PCR Amplification and Typing Kit (PE Applied Biosystems) was the second kit to be developed for forensic purposes. Five sequence specific polymorphism loci (LDLR, GYPA, HBGG, D7S8, and Gc) can be detected (Budowle B. *et al.*, 1995) and used in conjunction with the HLA-DQα locus.

The size of the alleles ranges from 138 as in Gc locus up to 214 as in LDLR locus. The addition of these five loci (PolyMarker kit) plus the DQα increased the power of discrimination for this system such as in identity and paternity testing (Bruce B. *et al.*, 1995; Blake E. *et al.*, 1992; Kratzer A. and Bar W., 1996). Disadvantages of these markers are that each locus consists of very limited number of alleles (2 or 3) and mixtures can not easily be detected. The power of discrimination of DQα and polymarker when compared to minisatellites is still poor.

1.7.3 Amplified fragment length polymorphisms (Amp-FLPs)

Amplification of fragment length polymorphisms (Amp-FLPs) is a second class of PCR-based polymorphic loci which contain Variable Number Tandem Repeat (VNTR) "core" units consisting of non-coding sequence. Consequently, Amp-FLP typing systems are based on differences in length between alleles (length polymorphism) rather than sequence differences. First, alleles are amplified by PCR, they can then be separated by size using polyacrylamide gel electrophoresis and subsequently detected by silver staining. These polymorphic markers are a very suitable for genetic characterisation of individuals.

The amplified fragment length polymorphism (AmpFLP) technique analysed loci (Budowle B. *et al.*, 1991a; Sajantila A. *et al.*, 1991) such as the apolipoprotein B gene (3' ApoB). It is situated on chromosome 2 and the fragments observed range from ~ 570 to 900 bp, consisting of 16 alleles (Gene M. *et al.*, 1995; Knott T. *et al.*, 1986; Boerwinkle E. *et al.*, 1989; Ludwig E. *et al.*, 1989). D17S30 (also designated D17S5; Odelberg S. *et al.*, 1989) locus is located on chromosome 17 and detected by probe YNZ22 (Nakamura Y. *et al.*, 1987a; Horn G. *et al.*, 1989). This locus consists of 14 alleles, ranging in length from 170 to 1080 bp (Gecz J., 1991) with a 70 bp core sequence (Wolff R. *et al.*, 1988).

The introduction of commercially available AmpliFLP™ D1S80 PCR Amplification kit by Perkin-Elmer (Applied Biosystems) which released in 1991 helped the widespread use of the D1S80 locus for human identification purposes. The D1S80 locus was the first and the major AmpFLPs system widely used. This highly variable region on the short arm of the first chromosome was first described by Nakamura Y. in 1988 as a restriction length polymorphism detectable by probe pMCT118 (Nakamura Y. *et al.*, 1988). However, PCR amplification and typing of this locus was first described by Kasai R. (Kasai R. *et al.*,1990). The alleles associated with this locus are resolved into discrete entities by using polyacrylamide gel electrophoresis. Subsequently, Budowle B. (Budowle B. *et al.*, 1991a) provided D1S80 population data indicating that this marker would be suitable for forensic applications.

The length of the repeat unit has been determined to be 16 base pairs (Kasai R. *et al.*, 1990) and allele range in size from approximately 350 to 1000 bp and more than 27 alleles have been observed (Rand S. *et al.*, 1992; Skowasch K. *et al.*, 1992; Pinheiro M. *et al.*, 1996). There are two common alleles (18 and 24 repeat units) with combined frequency of more than 50% (Alonso A. *et al.*, 1993; Thymann M. *et al.*, 1993; Alkhayat

A. *et al.*, 1996; Schnee-Griese J. *et al.*, 1993). The allele designations are based on the number of repeats in the amplified fragments (Sajantila A. and Budowle B., 1992). Unknown samples can be compared with the allelic ladder (PE Applied Biosystems) which is comprised of 27 amplified D1S80 alleles ranging in size from 14 (the smallest allele) to 41 (the largest allele) repeat units with the exclusion of allele 15 which is very rare. The system is very sensitive and as little as 100 pg of template DNA can be typed (Baechtel F. *et al.*, 1995; Kloosterman A. *et al.*, 1993)

The main two disadvantages of this marker are that it can not be multiplexed and the allelic dropout (Walsh P. *et al.*, 1992) where a small allele will amplify preferentially in comparison to large allele especially in the case of degraded DNA. Hence, a heterozygote can be read as homozygote. Another disadvantage is that the allele sizes are relatively large (350-1000 bp) compared to STRs where the average allele sizes are less than 400 bp. Therefore, in degraded DNA samples there is more chance to obtain positive result by using STRs rather than D1S80.

1.7.4 Y chromosome short tandem repeat (Y-STR)

Y chromosome can be divided into two parts, pseudoautosomal and non-pseudoautosomal. The non-pseudoautosomal (male-specific) region of the Y chromosome is many times larger than the size of the pseudoautosomal region and it is uniparental inherited (like mtDNA), therefore displaying a haploid inheritance.

The Y chromosome is rich in several classes of repeated DNA sequences, and microsatellite loci are among one of these classes. Therefore, Y chromosome STR-based haplotyping (Rower L. *et al.*, 1992), which can be amplified by PCR with fluorescent detection of the alleles, are very useful in many fields. It is a very useful marker in the

forensic identification of male DNA (Roewer L and Epplen J, 1992; Roewer L., 1998; Caglia A. *et al.*, 1998; Tun Z. *et al.*, 1999) such as in rape cases with male/female or more than one male stain mixtures or with vasectomized males (Prinz M. *et al.*, 1997). Y chromosome STRs have been used in paternity analysis where the alleged father is not available (Chakraborty R. 1985; Pena S. and Chakraporty R., 1994; Weichhold G., *et al.*, 1996). In these cases any patrilinear relative of an alleged father would carries the same Y chromosome. Human evolutionary studies and analysis of migration, settlement of human populations in historic time spans have also been studied using Y chromosome STRs (Jobling M. and Tyler-Smith C., 1995; de Knijff P. *et al.*, 1997; Kayser M. *et al.*, 1997b).

Y chromosome microsatellites have a mutation rate almost equal to that of autosomes (Heyer E. *et al.*, 1997). More than 14 loci of highly polymorphic chromosome Y specific microsatellites (Kayser M. *et al.*, 1997a) are available for forensic and paternity use. Typing of several Y-STR loci in a single test tube (multiplexing) probably will be soon commercially available.

1.8 PROJECT BACKGROUND

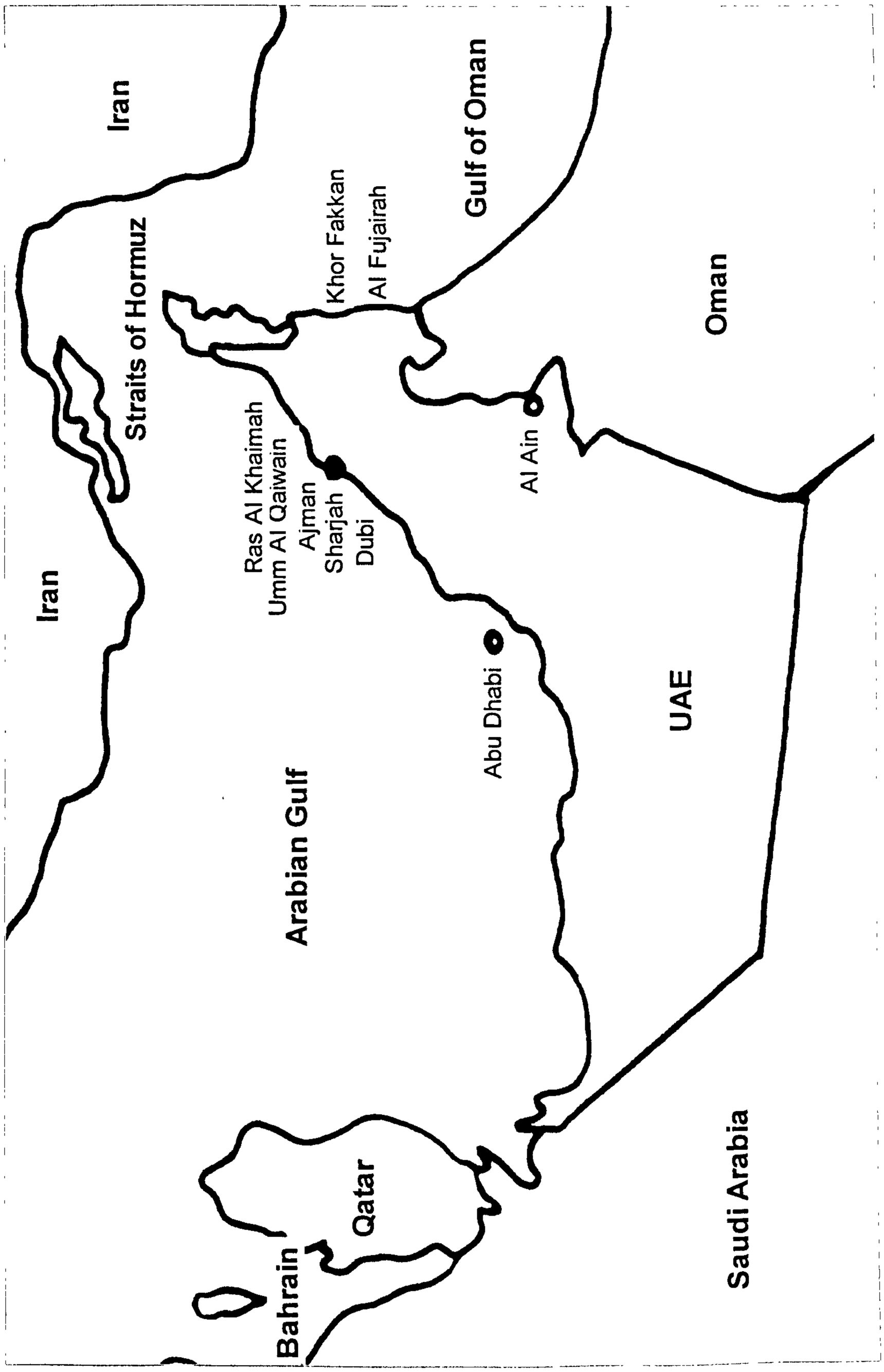
1.8.1 Introduction

The United Arab Emirates (UAE) is an independent federal state made up of seven emirates (Abu Dhabi, Dubai, Sharjah, Ajman, Umm Al-Qaiwain, Ras Al-Khaimah, and Fujairah) (see figure 1.5 and 1.6). In 1958 the population was estimated by the United Nations to be approximately 86,000 (Abdullah M., 1999). The total population now is approximately 3 million people (according to the census of 1999) of whom only about

38A

Figure 1.5 Map of the UAE.

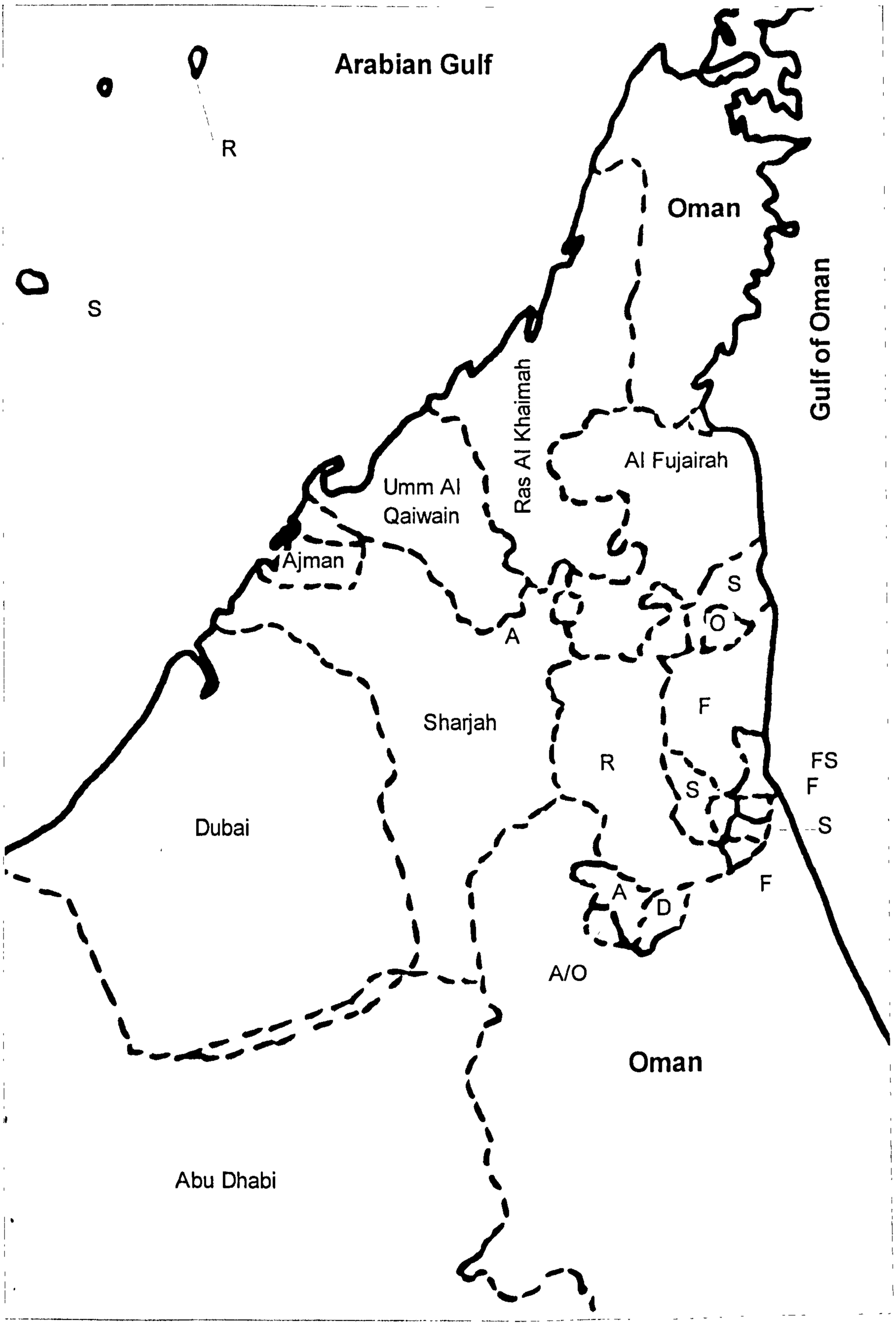
In this map the whole UAE borders can be seen. The seven emirates which the UAE made of is seen on the map (Abu Dhabi, Dubai, Sharjah, Ajman Umm Al Qaiwain, Ras Al Khaima and Al Fujairah). In addition the UAE neighbours countries also observed (Oman, Saudi Arabia, Qatar, Bahrain and Iran).


Iran
Straits of Hormuz
Gulf of Oman
Khor Fakkan
Al Fujairah
Ras Al Khaimah
Umm Al Qaiwain
Ajman
Sharjah
Dubi
Oman
Al Ain
Arabian Gulf
Abu Dhabi
UAE
Qatar
Bahrain
Saudi Arabia

39A

Figure 1.6 Emirates internal boundaries (the northern area of the UAE).

A = Ajman, O = Oman, R = Ras Al Khaimah, S = Sharjah, F = Al Fujairah, D = Dubai.


Arabian Gulf
R
S
Oman
Gulf of Oman
Ras Al Khaimah
Al Fujairah
Umm Al Qaiwain
Ajman
A
S
O
F
Sharjah
R
S
FS
F
S
Dubai
A
D
F
A/O
Oman
Abu Dhabi

one-fifth are UAE citizens. Prior to the creation of the UAE as an independent federal state on 2 December 1971, the Emirates were known as the Trucial States, a name derived from a set of agreements between the rulers (Sheikhs) of the Emirates and Britain in 1820 AC (Al-Sayegh F., 1998).

1.8.2 Physical and political structure

The UAE is a relatively small country (83,600 square kilometres). The Emirate of Abu Dhabi accounts for 80% of the total land while Ajman has an area of only 260 square kilometres (0.3%). The UAE is located in the eastern part of the Arabian Peninsula, astride the tropic of Cancer. Its Gulf coastline of approximately 650 kilometres is more than 40% of the total on the Arabian side of the Gulf. The UAE's territory bisects the territory of Oman south of the Musandam Peninsula, giving a UAE coastline of about 90 kilometres on the Gulf of Oman. The UAE shares land borders with Saudi Arabia on the northwest, west, south, and southeast and with Oman on the southeast and northeast (Ministry of Information and Culture, 1992). See figure 1.5

The UAE consists of two distinct physical regions, an eastern mountain zone and a western desert zone. The mountain zone (with the gravel plains, west of the mountains) is roughly fifty miles from north to south and twenty miles from east to west. The western desert area also consists of a coastal strip and makes up more than two thirds of the UAE.

Before the discovery of oil in 1958, the Emirates carried no political or economic weight, in fact, they began to witness serious economic and social growth only in the aftermath of World War II. The 1960's is considered as one of the most important decades in the modern history of the Gulf Emirates because during this period national

consciousness increased and the Emirates gained world recognition because of oil (Al-Sayegh F., 1998)

1.8.3 The Climate

The temperature varies from mild in winter (15-25 °C) to very hot in summer (40-48 °C), with a very high humidity (90-100%) especially on the coast strip. This gives an indication of how fast the forensic samples can be deteriorated in summer.

1.8.4 The historical background

People probably arrived in the land now known as the UAE around nine or ten thousand years ago. The ancient history of what is now the UAE, and of the Gulf as a whole, remains largely conjectural. What is clear is that a highly developed civilisation existed here but later this civilisation disappeared. The Bahrain culture was part of a network of ancient trading settlements along the Arabian side of the Gulf (Donald H., 1970). Important trading with several centres, including Bahrain culture existed as far back as 3000 BC. Existence of tumuli (mounds), apparently like those on Bahrain, on Umm Al-Nar (a small island just east of Abu Dhabi city) led some archaeologist to believe that Bahrain culture and Umm Al-Nar were part of the same culture (Vine P. and Casey P., 1992).

In the seventh century AD inhabitants of the region converted to Islam. Islam had the effect of uniting Arabia for the first time and of bringing both coasts of the Gulf under Arab Muslim rule. This led to the period of greatest flowering of Gulf commerce with India and China.

Arabian seamen remained the leading traders of the Indian Ocean until the Portuguese invasion into eastern waters in the early sixteenth century. One of the places that Alfonso de Aluquerque captured was Khor Fakkan (see figure 1.5) in 1506 (Frauke H., 1982; Donald H., 1970). Later in 1820 the British occupied the Arab Gulf region till Britain's withdrawal from the region in 1971.

The region traditionally depended on the limited occupations of fishing and pearling as well as small-scale trade, including pearls, with India, east Africa and the Iranian coast. The pearl trade suffered two major setbacks during the 1930s, the development of cheap artificial pearls by the Japanese and the great depression. Therefore, due to decreased trade, UAE national workers flocked in great numbers to the new oil areas; first individual emigrants, then, at later stages, families followed. Most of the immigrants headed for Kuwait where jobs were relatively more abundant whether in the oil sector, government departments, or the private sector. Emigrants continued from the early 1950s, with some of the emigrants staying abroad for as long as twenty years, until economic conditions changed in the Emirates (early 1970s) (Al-Sayegh F., 1998).

1.8.5 Origins of Arabic people

Arab is a name which has been given to the ancient inhabitants of the Arabian Peninsula, however, in present time Arabs (~ 200 million) constitute the majority of the population in the following countries Saudi Arabia, Yemen, Jordan, Syria, Iraq, Egypt, Sudan and the nations of north Africa.

The Arabian Peninsula with a great desert is located in the south west of Asia and it is one of the most sparsely populated areas of the world. It consists of seven countries,

Saudi Arabia (three quarters of the Peninsula), Yemen, Oman, UAE, Kuwait, Qatar, and Bahrain. See figure 1.7.

The earliest known events in Arabian history are migrations from the south west of peninsula what it is now known as Yemen (the land of Arab descent) into neighbouring areas. About 3500 BC, Semitic speaking peoples of Arabian origin migrated into the valleys of the Tigris and Euphrates rivers in Mesopotamia, supplanting the Sumerians, and became the Assyro-Babylonians. Another group of Semites left Arabia about 2500 BC and settled along the eastern shore of the Mediterranean sea, later some of these migrants became the Amorites and Canaanites (Serai H., 1997). See figure 1.8.

1.8.6 Arabs in north Africa

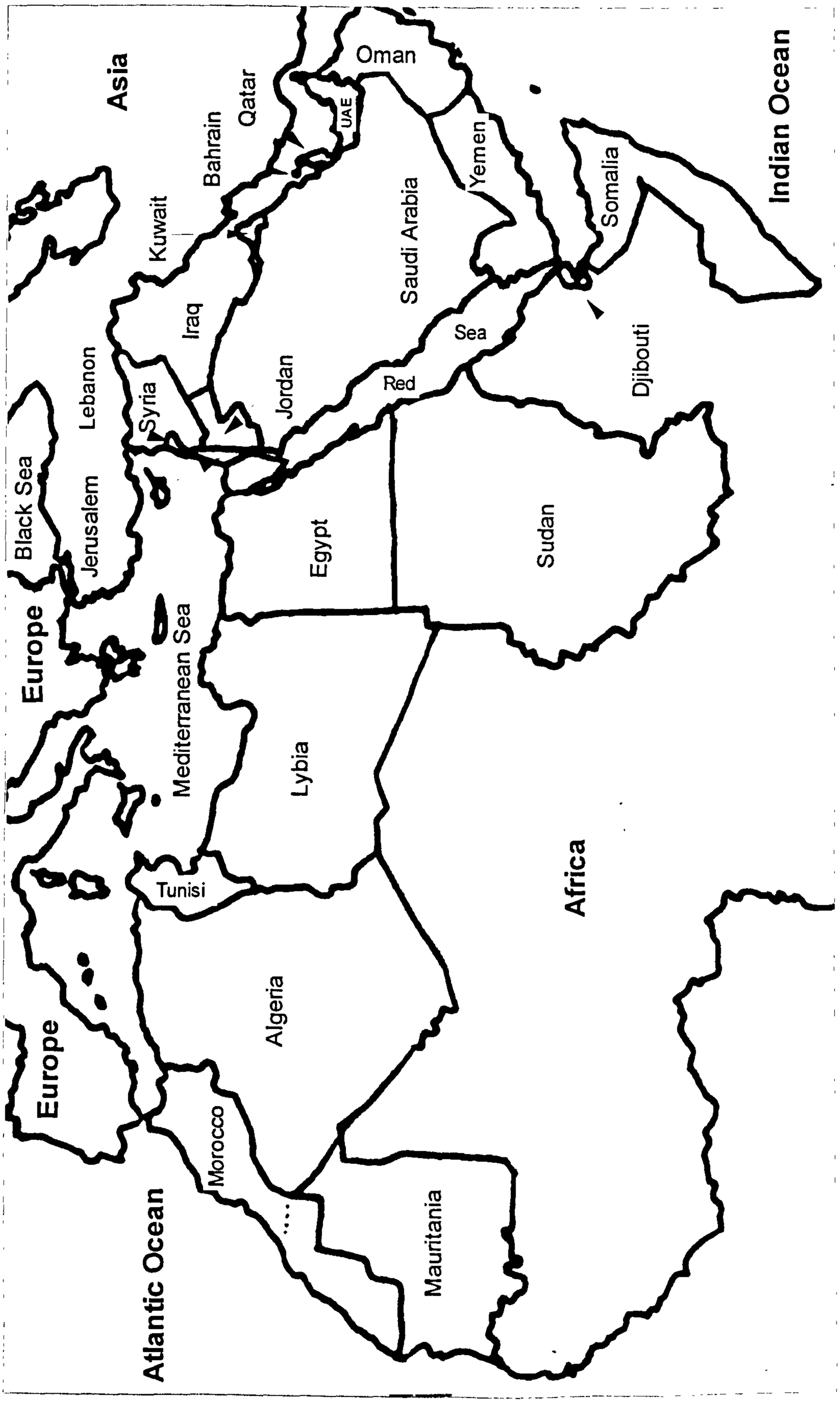
Most Egyptians are descendent from the ancient Egyptians and Arabs who conquered the area in the 7th century AD. Berber peoples are the origin inhabitants of the Arab countries of north Africa since the earliest recorded time around 3000 BC. However, through the centuries Berbers have mixed with many other ethnic groups, notably the Arabs, that they are now identified usually on a linguistic rather than a racial basis, although, some Arab tribes migrated to that area since the 7th century AD, when Arab Muslims conquered north Africa. Berbers constitute about 40% and 30% of Morocco and Algeria population respectively. See figure 1.8.

1.8.7 The UAE population structure

The population of UAE can be divided into three sectors urban (hadhar), nomadic (bedu), and rural. The bedu are the people of the desert who are animal owners and move

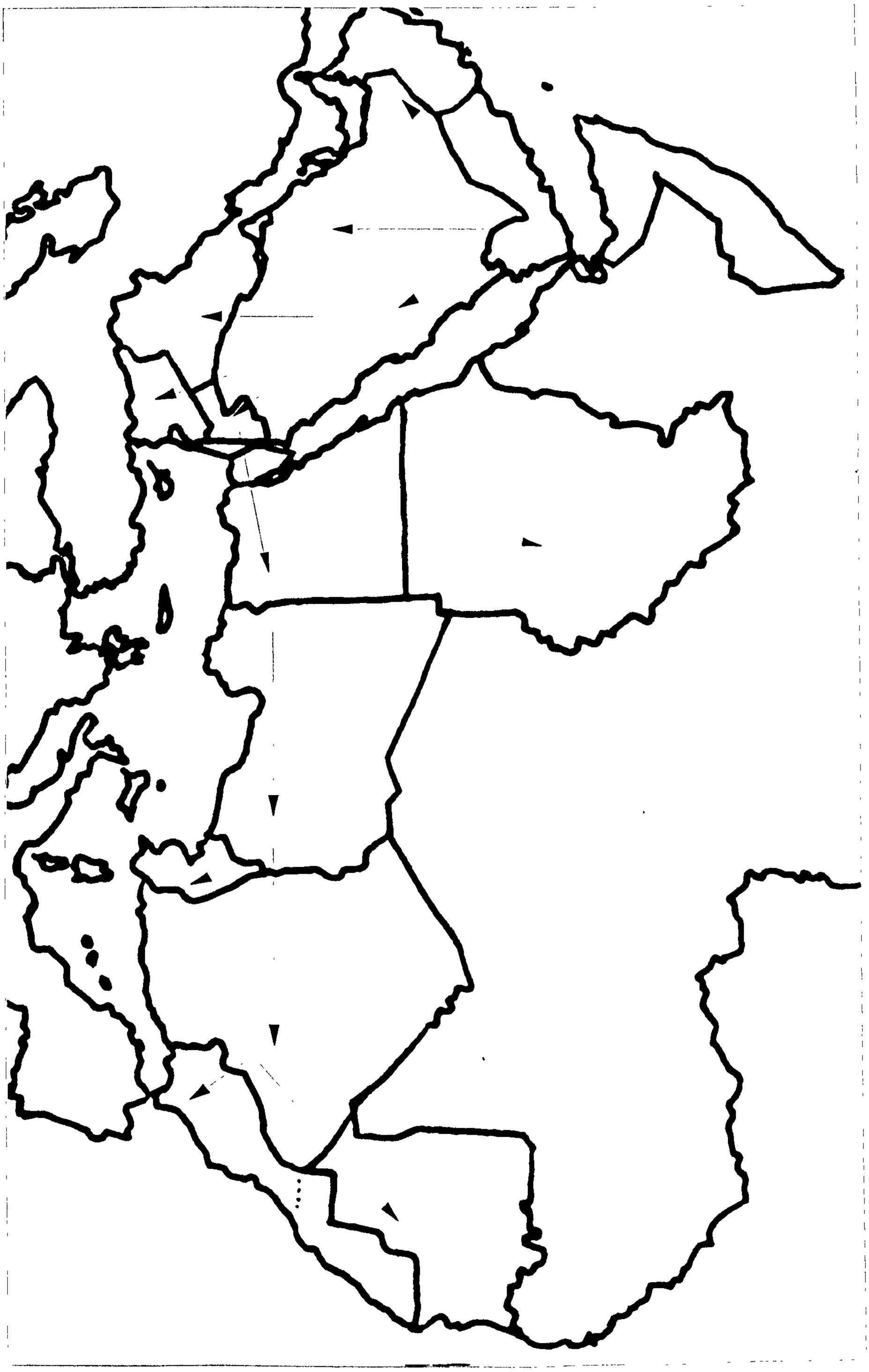
44A

Figure 1.7 Map of the Arab world showing all Arab countries within Asia and Africa continents.


Asia
Oman
Qatar
Bahrain
UAE
Kuwait
Iraq
Saudi Arabia
Yemen
Somalia
Indian Ocean
Black Sea
Europe
Lebanon
Jerusalem
Syria
Jordan
Red
Sea
Djibouti
Egypt
Sudan
Mediterranean Sea
Lybia
Tunisi
Algeria
Europe
Morocco
Mauritania
Atlantic Ocean
Africa

45A

Figure 1.8 Map of the Arab world showing the migration of Arabs.



about with their camels, sheep, and goats in search of grazing, and concentrate round their wells. Some of them used to gather firewood from the scrub and bring it into the coastal towns. Hadhar people made their living from the sea. They lived by fishing and also by the pearling industry. Agriculture used to be the main way of life for rural people in the eastern mountain area and the oases. However, what was an overwhelmingly rural and nomadic population a generation ago is now preponderantly urban (Al-Sayegh F., 1998).

The rate of natural increase for the native population is high, approximately 2.5% annually. However, there was a much higher rate for the first decade after the foundation of the UAE in 1971 due to the granting of citizenship to several Arab tribes from out side the country such as from Saudi Arabia, Oman, Qatar or from north of Iran. See figure 1.9 and 1.10.

The two most immediately striking features about the UAE's population are its rapid growth and the fact that resident foreigners greatly outnumber the native population. Both circumstances are an outgrowth of the massive, extremely rapid modernisation that followed discovery and production of oil in the 1960's. Most of the immigrant workers are Asiatic (more than 40% from India and Pakistan) (Taryam A., 1987). In 1968 a census of the Trucial States (UAE) showed a total population of 180,000 (approximately 114,000 (63%) local population). By 1975 the population was 655,937, an astounding 364% increase in seven years. By 1979 the population had reached an estimated 900,000 and stood at an estimated 1,130,000 as of 1982. In 1995 the population was 2,411,041 which the local population represented only 24% of the total population and at present time there are 3 million or a little more with only 20% or less of native population (Abdullah M., 1999).

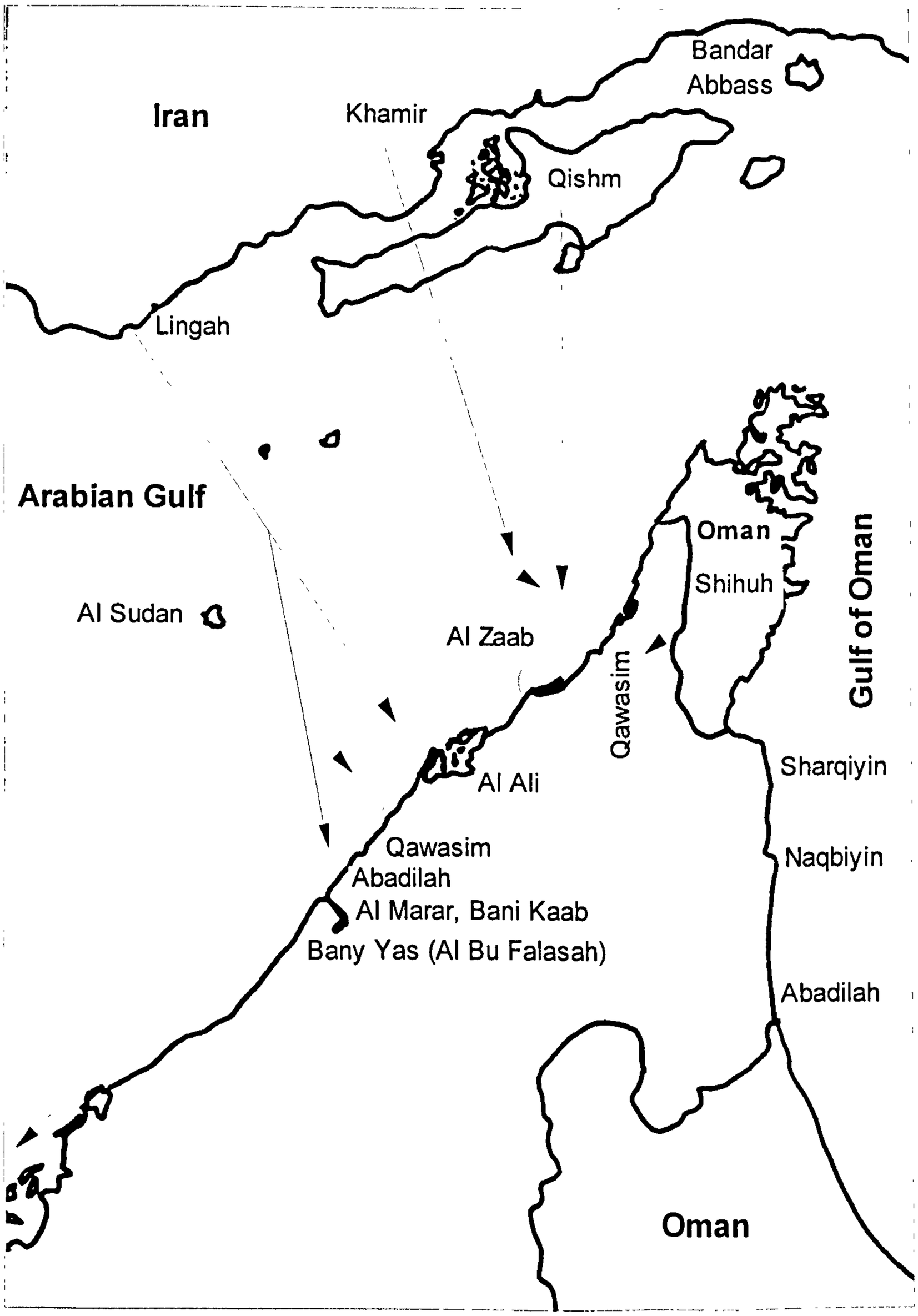
47A

Figure 1.9 Map of northern part of the UAE.

This map showing the migration of different Arab tribes to UAE from Iranian part.

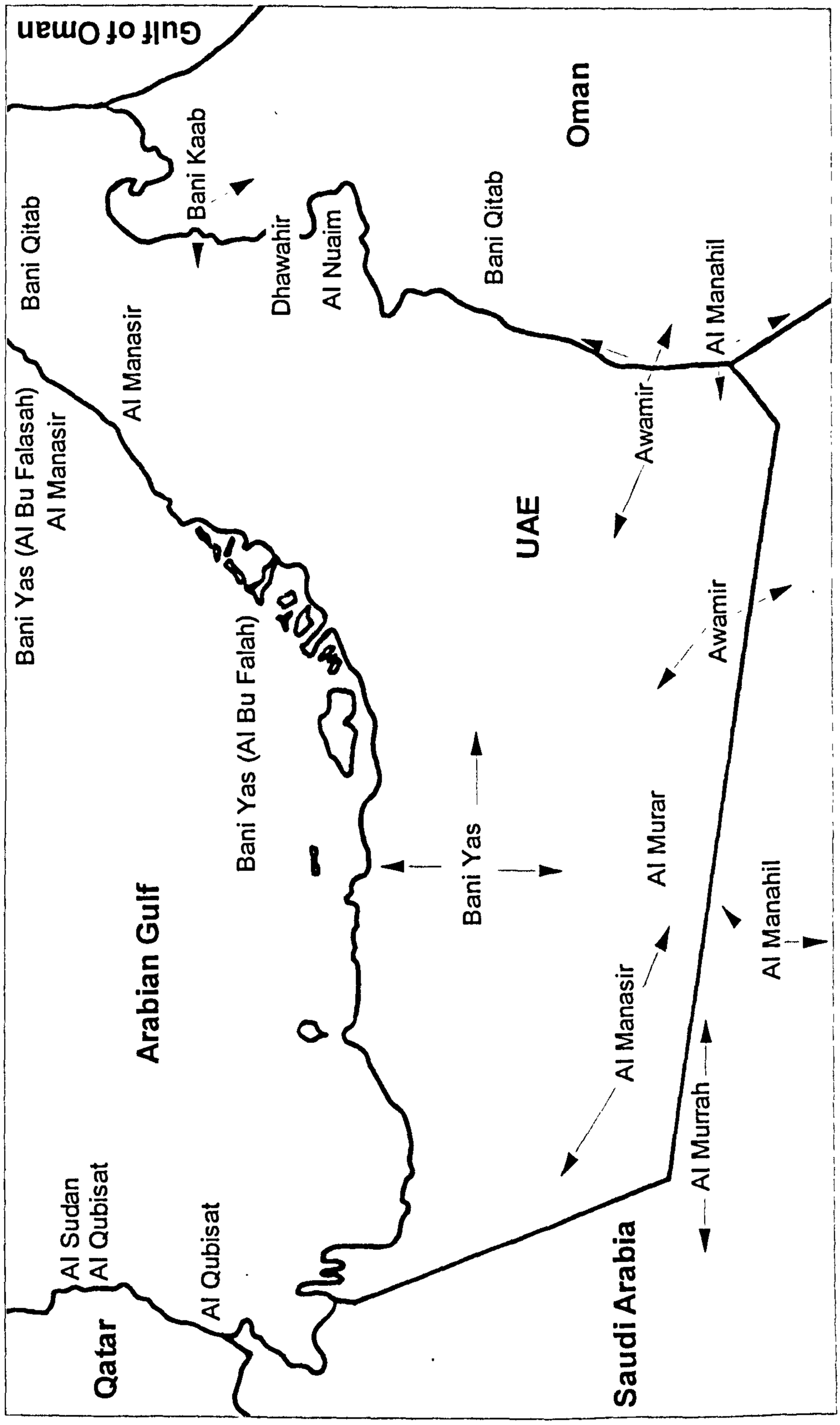
The migration of Al Zaab tribe from Ras Al Khaimah to Abu Dhabi also can be seen.

The name of some other tribes is seen (Al Sudan, Al Ali, Qawasim, Abadilah, Al Marar, Bani Kaab, Bani Yas, Naqbiyin, Sharqiayin and Shihuh).


Bandar
Abbass
Iran
Khamir
Qishm
Lingah
Arabian Gulf
Oman
Shihuh
Gulf of Oman
Al Sudan
Al Zaab
Qawasim
Al Ali
Sharqiyin
Qawasim
Abadilah
Al Marar, Bani Kaab
Bany Yas (Al Bu Falasah)
Naqbiyin
Abadilah
Oman

48A

Figure 1.10 Map of the UAE showing names of tribes and their migrations into the UAE.


Gulf of Oman
Oman
Bani Kaab
Bani Qitab
Dhawahir
Al Nuaim
Bani Qitab
Bani Yas (Al Bu Falasah)
Al Manasir
Al Manasir
Awamir
Al Manahil
Bani Yas (Al Bu Falah)
UAE
Bani Yas
Awamir
Arabian Gulf
Bani Yas
Bani Yas
Al Manasir
Al Murar
Al Murrah
Al Manahil
Al Sudan
Al Qubisat
Qatar
Al Qubisat
Saudi Arabia
Al Mu

The oil rich Emirates of Abu Dhabi, Dubai, and Sharjah had experienced particularly more rapid growth. Because 80% of the UAE's inhabitants are there in response to the country's need for employees, the immigrants give the population a very diverse complexion and a distinctly youthful and male character (more than 70% of the total population was male in 1984 and 66.6% in 1995). Approximately 85% of the UAE population live in the Emirates of Abu Dhabi (39%), Dubai (29%), and Sharjah (17%). The other four Emirates contain only 15% of the total population.

1.8.8 Crime rate

According to the census of 1996 by Ministry of Interior, approximately 52% of total crime was committed by people from Asia (mostly from Pakistan, India, and Iran), 14% local and the rest by other nationalities. 17% of the total crime occurred in Abu Dhabi, 31% in Dubai, and 28% in Sharjah.

Although the actual level of criminal activity, particularly by western standards, appears modest, its scope and nature are new and unsettling to UAE society and create an exaggerated perception and fear of the potential dangers posed by the foreign population (Abdullah M., 1999; Khalil A. and Al-Akubaisi S., 1999).

1.8.9 Abu Dhabi

Abu Dhabi had a very rapid growth from a small nondescript town of about 2,000 in the late 1940's and 46,000 in 1968 to a city with a population that approached 500,000 by the mid-1970s (Malcolm C., 1986). Abu Dhabi is the Capital of UAE and 39% of the total population live there. The second most developed city in Abu Dhabi emirate is Al-Ain, which is located approximately 160 kilometres away from Abu Dhabi City. The

population of Al-Ain City was approximately 13,000 in 1968. The Emirate of Abu Dhabi has the largest proportion of the nomadic population.

1.8.10 Sharjah

Sharjah literally means "the eastern place". The population of Sharjah has been estimated to be 31,500 in 1968 (Donald H., 1970). Kalba, Khor Fakkan, and Dibbah lie on the Gulf of Oman, the north western of Indian part of the Indian Ocean, belong to Sharjah Emirate (Al-Sayegh F., 1998).

1.8.11 The UAE tribes

In the territory of the UAE, as throughout all of Arabia, the basic political units of traditional society were the tribes. There are many different types of tribe in Arabia. They vary from the totally sedentary to the totally nomadic, and there are many which are at some point in between these two extremes (Al-Sayegh F., 1998).

The most important feature of the tribe is kinship. It is the kinship and close intermarriage within the main families, which gives the tribes its cohesiveness. Abu Dhabi has the greatest number of tribal and sub-tribal groups, including a few that are still nomadic (Al-Sayegh F., 1998). For more details on tribes see appendix A.

1.8.12 Marriages

In the UAE the extended family, under the leadership of the oldest male, remains intact. Marriages continue to be carefully arranged within a small social grouping with those in which a young man marries his father's brother's daughter (first cousin) is considered to be the ideal (Bener A. *et al.*, 1996). However, marriages between different

tribes are common as well. For the last decade a few marriages of people from India, Philippine, Iran, Egypt, Europe, and USA to UAE citizen occurred. The average family size in the UAE is approximately six children.

1.9 OBJECTIVES OF RESEARCH

In the UAE the ABO blood group system and some protein markers such as phosphoglocomutase (PGM) and erythrocyte acid phosphotase (EAP) were the only genetic markers used until very recently. All these markers have a low value of power of discrimination and power of exclusion of paternity. Very recently polymerase chain reaction (PCR) system was introduced using HLA-DQα and polymarker (PM) DNA markers in both forensic and paternity cases. Although the PCR system highly increased the power of discrimination and power of exclusion of paternity still there is a need for a system with higher power of discrimination and exclusion.

A non-match can be declared as a definitive proof that two samples had different origins. However, in inclusions it is important to have knowledge of how frequently a match is expected to occur in the general population. Therefore, when introducing a new marker it is necessary to examine the population structure and test for HWE in order to generate databases to be used as references. Usually a random sample is collected representing the whole population and the allele frequencies of a number of markers are determined. It is recommended to generate several databases representing all ethnic groups present in a population and use the most similar database representing a suspect in order to estimate the power of statistical match accurately.

Several objectives of this study can be defined as the following:

a) Investigate the structure of the UAE populations using VNTR and STR loci.

b) Investigate any significant difference between the UAE local people from two different Emirates using STR loci.

c) Investigate any significant difference between the UAE Arab population database and Indo-Pakistani populations using STR and VNTR loci.

d) Investigate any significant difference between the DNA database from the Indian population compared to Pakistani population using STR and VNTR loci.

e) Investigate any significant difference between the DNA databases of two Arab populations (UAE and Egyptian) using STR loci.

f) Examine the usefulness of five VNTR loci and eight STR loci in criminal cases analysis and paternity testing in the UAE.

g) Establish the UAE Arab, Indian and Pakistani databases for 5 VNTR loci and eight STR loci. In addition, establish the Egyptian database for eight STR loci.

CHAPTER 2: MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 Chemicals

The chemicals were obtained from Sigma Chemical Co., UK; Fisher Scientific UK; ICN Biomedical Ltd. and BDH Laboratory supplies (AnalaR® grade; Poole, Dorset) unless otherwise indicated.

2.1.2 Restriction endonuclease enzyme

The *Hin*fI restriction endonuclease enzyme was purchased from Gibco-BRL, Paisly, Strathclyde, UK, together with its reaction buffer (REACT 2), in the 10X concentrate buffer supplied for use with *Hin*fI. The enzyme and reaction buffer were stored at -20 °C.

2.1.3 K562 DNA

The human cell line K562 DNA digested with *Hin*fI was purchased from Promega, Delta House, Southampton, UK, and stored at -20 °C.

2.1.4 DNA molecular ladder

The ACES™ Chemiluminescent Marker contains 30 double-stranded DNA size markers (ladder) ranging from 526 to 22,621 base pairs. The DNA Marker, DNA dilution buffer and Alkaline Phosphatase-Labelled Marker Probe were all purchased from Gibco-BRL and stored at 4 °C. The MW100 ladder probe was obtained from CellMark Diagnostics, Abindon, Oxon, UK and stored at -20 °C.

2.1.5 Single locus DNA probe and CDP-Star

All NICE™ (Non-Isotopic Chemiluminescent Enhanced) single locus probes MS1, MS31, MS43A, YNH24, G3 and the chemiluminescent substrate CDP-Star™ were purchased from CellMark Diagnostics.

2.2 GENERAL PREPARATORY PROCEDURES

2.2.1 pH measurement

The pH of solutions was measured using a Corning pH meter 220 and combination electrode (Corning Incorporation, New York, USA).

2.2.2 Autoclaving

Equipment and solutions were sterilised at 15 psi for 20 min using a Laboratory Thermal Equipment Autoclave 225E. Small batches were sterilised in a Prestige High Dome pressure cooker.

2.2.3 Glassware

Glassware was washed in a glass washing machine and rinsed with Milli-Q water (Millipore Corporation, Bedford, MA, USA) prior to its use.

2.2.4 Solutions and buffers

All solutions and buffers were made up using Milli-Q water. Solutions were then autoclaved and stored at room temperature.

2.3 DNA SAMPLE COLLECTION AND PACKAGING

2.3.1 Source of Blood Samples

Whole blood samples were collected from unrelated individuals within the United Arab Emirates (UAE) from two cities, Abu Dhabi and Sharjah. Samples were collected from three population groups residing in the UAE. 194, 197 and 229 samples from the Indian, Pakistani and UAE native Arabic populations respectively.

2.3.1.1 Blood samples from Abu Dhabi City

Whole blood samples were collected in ethylenediaminetetra acetic acid (EDTA) vacutainer 5 ml tubes from 163, 172, 129 unrelated individuals from the Indian, Pakistani and UAE Arab respectively by the Abu Dhabi Blood Bank Centre.

2.3.1.2 Blood samples from Sharjah City

Whole blood samples were collected in EDTA vacutainer 5ml tubes from 31, 25 and 100 unrelated individuals from the Indian, Pakistani and UAE Arab respectively by the Sharjah Blood Bank Centre.

2.3.2 Blood sample packaging

All blood samples from Abu Dhabi and Sharjah were sent to Abu Dhabi forensic science laboratory. On arrival, samples were aliquoted into 2 or 3 sterilised 1.5 ml microcentrifuge tubes. In addition, 100% cotton fabric swatch stains (4 cm x 5 cm) were prepared and dried at room temperature. All samples were stored at -70 °C. A batch of 150 samples were sent by TNT Express to the Department of Forensic Medicine and

Science, University of Glasgow, packaged with ice. On arrival (two days), blood samples were extracted directly or stored at -20 °C for later extraction.

2.4 ISOLATION OF DNA FROM WHOLE BLOOD

2.4.1 Nucleic acid extraction

DNA was extracted from whole blood by using Puregene™ DNA Isolation Kit (Gentra Systems, Inc. Minneapolis, MN 55447, USA). Frozen whole blood samples were thawed at ambient temperature and 300 µl was added to a 1.5 ml microcentrifuge tubes (SARSTEAD, Germany) which contained 900 µl RBC lysis solution. Mixed by inversion and incubated for 10 min at room temperature, inverted again once during the incubation. It was then centrifuged (Micro centaur, Scotlab) for 20 s at 11,000g, the supernatant was removed with a micropipette leaving behind visible white cell pellet and 20 to 30 µl of residual liquid.

The pellet was re-suspended before adding 300 µl Cell Lysis Solution and pipeted up and down or alternatively the tube was inverted several times to lyse the cells. If cell clumps were visible after mixing the incubation continued for 10-20 min at 37 °C or for 30 min at room temperature until the solution was homogenous.

Samples incubated at 37 °C were allowed to cool to room temperature, and then 100 µl Protein Precipitation Solution was added to the cell lysate and vortexed vigorously for 20 s then centrifuged at 11,000*g* for 3 min. The precipitated proteins formed a tight dark brown pellet.

The supernatant, which contained the DNA, was poured into a clean 1.5 ml microcentrifuge tube containing 300 µl 100% isopropanol. The precipitated protein pellets was discarded. Samples were mixed by inverting gently 50 times until the white threads of DNA formed a visible clump and then centrifuged at 11,000*g* for 1 min. The DNA was visible as a small white pellet. The whole supernatant was removed and left the pellet behind.

300 µl ethanol (70%) was added to each tube and inverted several times to wash the DNA pellet. Samples were centrifuged at 11,000*g* for 1 min and the ethanol was carefully pipeted off leaving the pellet on the side of the tube to air dry for 15 min.

DNA Hydration Solution (100 µl) was added to each sample and then DNA was allowed to rehydrate by mixing first and then incubating the samples in a water bath (Techne Company, Cambridge, UK) at 65 °C for 1 h. After incubation, samples were left overnight at room temperature and pipetted (Gilson-P200) up and down several times before quantifying the DNA or storing at -20 °C for later quantification.

2.5 ELECTROPHORESIS OF DNA

2.5.1 Agarose gel electrophoresis

All agarose in this study (SEAKEM® and I.D.NA®) was supplied by flowgen FMC (FMC Corperation).

2.5.2 Mini-gel electrophoresis

Mini-gels were used to assess restricted DNA and to estimate the quality and quantity of extracted DNA.

The appropriate (0.55 g) weight of agarose (SeaKem 1% w/v) was added to the volume (55 ml) of 1X TBE (0.09 M Tris-borate, 2 mM EDTA) required for the gel being cast. The agarose suspension was heated in a microwave oven until all the agarose had dissolved. The agarose solution was then allowed to cool to around 60 °C at which point 2 µl of ethidium bromide (10 mg/ml) was added to a total volume of 55 ml and mixed by swirling. The gel solution was then poured into the electrophoresis apparatus and allowed to set for 30 min. Sufficient 1X TBE running buffer was then added to just submerge the gel. The DNA samples to be loaded were mixed with one-tenth volume of loading buffer (0.25% bromophenol blue w/v + 40% sucrose w/v) (Sambrook J. *et al.* 1989). Molecular weight standard samples of known quantity and quality run along side the unknown samples.

The gel was run at 100 to 120 V until the bromophenol blue had migrated two-thirds of the way down the gel. The DNA in the gel was visualised by fluorescence of bound ethidium bromide under UV light (transilluminator 312 nm UltraViolet, Specronics Corporation, Westbury, NY, USA). The gel was photographed using a polaroid camera and polaroid kodak film then the amount of DNA recovered from each sample was estimated.

2.5.3 Quality and quantity of DNA

All DNA samples were tested by using mini-gel electrophoresis (1% w/v) for their quality and quantity before proceeding to the digestion of DNA by restriction enzyme.

2.5.4 Restriction digestion

The DNA (1-3 µg) solution was placed in a 0.5 ml eppendorf tube together with 1X REACT 2 buffer (diluted from 10X) and 10-20 units of restriction enzyme (*Hin*fI) were added ensuring that its concentration did not exceed 10% (v/v). The reactions were incubated at 37 °C for 6 to 24 h. An aliquot (2-6 µl) of the reaction was mixed with loading buffer (0.25% bromophenol blue w/v + 40% sucrose w/v) (Sambrook J., *et al.*, 1989), and analysed on an agarose gel (1% w/v). The digested DNA sample was run at 110 V for 20 min (until the bromophenol blue had migrated at least half way down the gel). The gel was then visualised under UV light. A uniform smear indicated a complete digest. The gel was then photographed and documented.

2.5.5 Analytical gel electrophoresis

Analytical or maxi-gels were used to size separate genomic DNA prior to Southern blotting for the DNA profiling described.

Horizontal submarine gel (24 cm X 31 cm) was prepared by adding 0.75% (w/v) agarose (I.D.NA®) to appropriate volume of 1X TBE (0.09 M Tris-borate, 2mM EDTA) required for the gel being cast. The agarose suspension was then heated in a microwave oven until all the agarose dissolved. The agarose solution was then allowed to cool to around 60 °C and during this step, the gel was mixed by swirling periodically and no ethidium bromide was added. The gel solution was then poured into the electrophoresis

apparatus and allowed to set for 55 min. The thickness of the gel was approximately 0.7 cm. Sufficient 1X TBE running buffer was then added to submerge the gel.

Digested DNA samples (0.6 to 1.2 μg) were mixed with one-tenth volume of loading buffer (0.25% (w/v) bromophenol blue + 40% sucrose (w/v), prior to loading samples. 3 known DNA size markers (ladder) [(2.2 μl of solution A mixed with 7.8 μl of solution B) (Life Technologies, Inc)] were always run along side test DNA as a molecular size marker lanes. Two known DNA samples (K562) were also run on the gel as an internal control. The gel was then run over night, around 18 h, at 55 V or until bromophenol blue dye had migrated to the 2 to 3 cm from the end of the gel.

2.6 SOUTHERN BLOTTING

Hybond N^+ (Amersham International plc., Amersham, UK) nylon membrane was used as a solid support matrix and the modified method of Southern 1975 (Southern E., 1975).

2.6.1 Depurinating of DNA

Prior to blotting, the DNA was depurinated by submerging the gel in 0.25 M HCl solution and agitated gently to ensure that the gel could move freely in the buffer, for approximately 30 min depending on the gel thickness. Progress was monitored by the pH change in the gel as indicated by the blue to yellow transition of the bromophenol blue indicator of the tracking dye. After the dye had changed colour to yellow, the gel was left in depurinating solution for an additional 5 to 10 min.

Note: it was important to note that the depurination step was only necessary if DNA fragments greater than 10 kb were to be transferred. Also note that acid depurination is temperature dependent, consequently at elevated temperature the rate of depurination was much greater and may result in a reduced hybridisation signal.

2.6.2 Denaturing of DNA

The gel was rinsed in appropriate volume (approximately 1 litre) of Milli-Q water and then soaked in a denaturing solution (0.5 M NaOH, 1.5 M NaCl) so as to completely covered the gel. The gel was then left for 45 min at room temperature with gently shaking. During this time the bromophenol dye reverted from yellow to blue and the treatment was continued for an additional 15 min after this colour change.

2.6.3 Neutralising of DNA

Denaturing solution was poured off and the gel was rinsed in Milli-Q water. The gel was then placed in appropriate volume of neutralising solution (1 M Tris-Cl (pH 8.0), 1.5 M NaCl). The gel was soaked for 15 min at room temperature by gently agitating and then for another 15 min with freshly made neutralising solution.

2.6.4 Capillary transfer

A tray dish was filled with blotting buffer [10X SSC (1.5 M NaCl, 0.15 M tri-sodium citrate)], pH 7.0. A glass platform was made and covered with a wick made from three sheets of Whatman 3 MM filter paper and saturated with blotting buffer so that the ends of the two Whatman 3 MM filter paper were in the buffer to maintain a flow of transfer buffer. Air bubbles between the paper and the glass platform were removed by gently rolling a pipette over the surface.

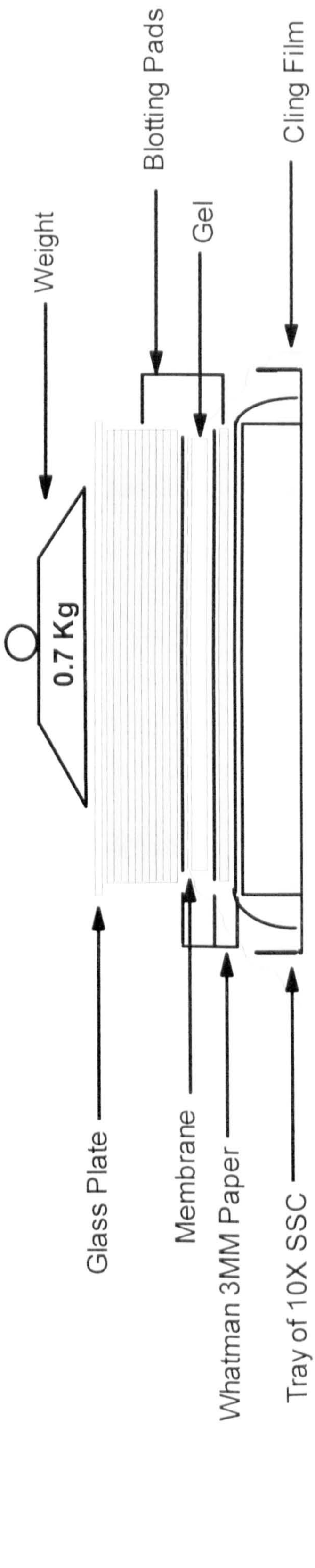
The gel was placed on top of the Whatman 3 MM paper (which act as a wick) with the well side of the gel touching it, ensuring that there were no air bubbles trapped beneath it. A sheet of nylon membrane (Hybond-N+, Amersham International plc. Amersham, UK) was cut to the exact size of the gel and placed on top of the gel, again avoiding air bubbles beneath the membrane.

Two sheets of Whatman 3 MM paper cut to the size of the gel were then wetted with blotting buffer 10X SSC and placed on top of the Hybond-N+ membrane. The area of the wick within a few mm of the gel was covered with cling film to prevent the blotting buffer being absorbed directly into the blotting pads above. Ten absorbent blotting pads (Promega, UK) were then placed on top of the 3 MM paper. Finally a glass plate (300 g) was placed on top of the stack of blotting pads and a 700 to 800 g weight placed on top (the total weight placed on the gel was from 1 to 1.1 kg) of the glass plate as shown in figure 2.1.

Transfer buffer was drawn from the reservoir by capillary action, passing through the gel eluting the ssDNA, which was deposited or trapped onto the surface of the membrane. The blot was left overnight. Prior to removing the membrane from the gel, the membrane was marked with soft graphite pencil to allow later identification and orientation of tracks. The membrane then was rinsed briefly in 2X SSC to remove adhering agarose.

63A

Figure 2.1 Set up for Southern blotting.


Weight
0.7 Kg
Blotting Pads
Gel
Cling Film
Glass Plate
Membrane
Whatman 3MM Paper
Tray of 10X SSC

2.7 FIXATION OF DNA BLOTS

2.7.1 Fixation by UV cross-linking

The UV light box first covered with a piece of cling film then the Hybond-N+ membrane was placed DNA-side down on a transilluminator and exposed to UV light (312 nm) for 2 min.

2.7.2 Fixation by oven baking

After fixation by UV cross-linking the nylon membrane was sandwiched between Whatman 3 MM paper and then the membrane was baked in an oven at 80 °C for two h. The membrane then was either hybridised immediately or stored in a fridge at 4 °C for later hybridisation.

Note: gloves or forceps were always used when handle the membrane.

2.8 HYBRIDISATION ANALYSIS

2.8.1 Preparation of hybridisation solution

99 ml of 0.5M Na_2HPO_4 (pH 7.2), 1 ml 1% SDS and 1g casein hammersten (BDH Laboratory Supplies, UK) were mixed together. After casein was dissolved the solution was centrifuged at 3,600*g* for 30 min to remove undissolved materials and then stored at room temperature until required (usually it had been used within 2-4 weeks).

2.8.2 Pre-hybridisation of DNA blots

Each nylon membrane filter was pre-hybridised by first wetting the membrane in 2X SSC. The membrane was then placed in a clean glass hybridisation cylinder ensuring that DNA was on the exposed surface and there were no air bubbles between the glass and the membrane. 12 ml of hybridisation buffer was placed in a clean falcon 50 ml tube and warmed up and added to the hybridisation cylinder. The cylinder was then incubated in hybridisation oven (Techne Hybridiser HB-1D, Techne Company, UK) at 50 °C with constant rotation (4 rpm) overnight or at least 2 h.

2.8.3 Hybridisation of DNA blots

Pre-hybridisation buffer was poured out of the cylinder. 6-7 µl of NICE™ single locus probe (SLP) and 5 µl of the ladder probe MW100 (CellMark Diagnostics, Abingdon, Dxon, UK) were added to 10 ml pre-warmed fresh pre-hybridisation buffer. The solution was mixed and immediately poured into the hybridisation cylinder and incubated at 50 °C with constant rotation (4 rpm) for 20 min. The hybridisation solution was then poured into another cylinder to hybridise the second membrane. Three membranes were hybridised with the same hybridisation solution at each time. The membranes were not allowed to dry out at any time and were always covered with hybridisation solution during hybridisation.

2.9 WASHING OF BLOTS

2.9.1 Wash solution 1

After hybridisation was completed the hybridisation solution was discarded. The three filters were placed in a plastic box and covered with 500 ml pre-warmed (50 °C) of wash solution 1 (0.01M Na_2HPO_4, 0.1% SDS w/v) and incubated at 50 °C in a water bath by manually shaking for 15 min. The position of the filters was changed during the incubation to ensure proper wash to all three filters. This step was repeated again with fresh pre-warmed (50 °C) wash solution 1 for another 15 min.

2.9.2 Wash solution 2

Wash solution 1 was replaced with 500 ml of wash solution 2 (0.1M malic acid (Sigma), 0.15M NaCl) and incubated by manually shaking at room temperature for 11 min. This step was repeated again with fresh wash solution 2 for another 11 min. The position of the filters was alternated during the incubation to ensure proper wash to all three filters.

2.10 CHEMILUMINESCENT DETECTION

2.10.1 CDP-Star

The membranes were removed from wash solution 2 and rinsed in 250 ml of CDP-Star assay buffer [(10 ml/l 0.1M $MgCl_2$, 10.5 ml/l Diethanolamine (98% solution) pH 9.5 (Sigma Co.)]. The first membrane was rolled up and then was placed in a clean glass hybridisation cylinder ensuring that DNA face out and there were no air bubbles between the glass and the membrane. 12 ml of CDP-Star assay buffer placed in a clean falcon tube

and then 80 µl of CDP-Star™ was added to the solution and mixed. The solution then was poured into the hybridisation cylinder and incubated in a hybridisation oven at room temperature with constant rotation (4 rpm) for 10 min. The solution then was poured into another cylinder and reused for the second and third membrane. After the solution was poured off the membrane was removed and blotted to remove the excess solution. The membrane was then sandwiched between cling film and air bubbles were removed using a pipette.

2.10.2 Autoradiography

After the membrane was marked for orientation ECL X-ray film (Amersham) was exposed to the membrane in a film cassette. The cassette was then incubated in an oven at 30 °C for 80 to 90 min before the film was developed using Kodak X-O mat instrument (Kodak Company). The X-ray film was exposed again to the membrane for a different length of times if required, for example the first exposure was not clear.

The X-ray film images were scanned using an Epson (GT-9000) with resolution of 216 dpi. The images were then analysed using image analysis software (BioMax 1D, Kodak Scientific Imaging System, 4 Science Park, New Haven, CT 06511, Eastman Kodak Company).

2.10.3 Stripping and re-probing the nylon membrane

After hybridisation and autoradiography the bound probe and blocking agents were removed by washing the membrane in alkali solution (0.4M NaOH) and incubated at 55 °C in a water bath for 45 min by manually shaking during the incubation from time to time. The alkali solution was then replaced with appropriate volume of pre-warmed acidic

solution (0.2M Tris-HCl pH 7.5, 0.1% SDS) and incubated at 50 to 55 °C for 35 min by manually shaking during the incubation. The membrane was then removed and rinsed in 2X SSC, then was air dried at room temperature for at least one h. The membrane was then sandwiched in Whatman 3 MM filter paper and baked at 80 °C for 1 to 2 h. The membrane could be stripped up to ten times.

A diagram shows the main steps involved, from obtaining a suitable source of biological material containing DNA, to the eventual visualisation of the hybridisation pattern on an autoradiograph. See figure 2.2.

Note: each membrane was striped prior to the first hybridisation for 10 min in alkali solution (at the same conditions and concentrations mentioned above) and then for another 10 min in Tris-HCl solution, which lowered the background signal.

2.11 POLYMERASE CHAIN REACTION (PCR)

2.11.1 Isolation of DNA from whole blood

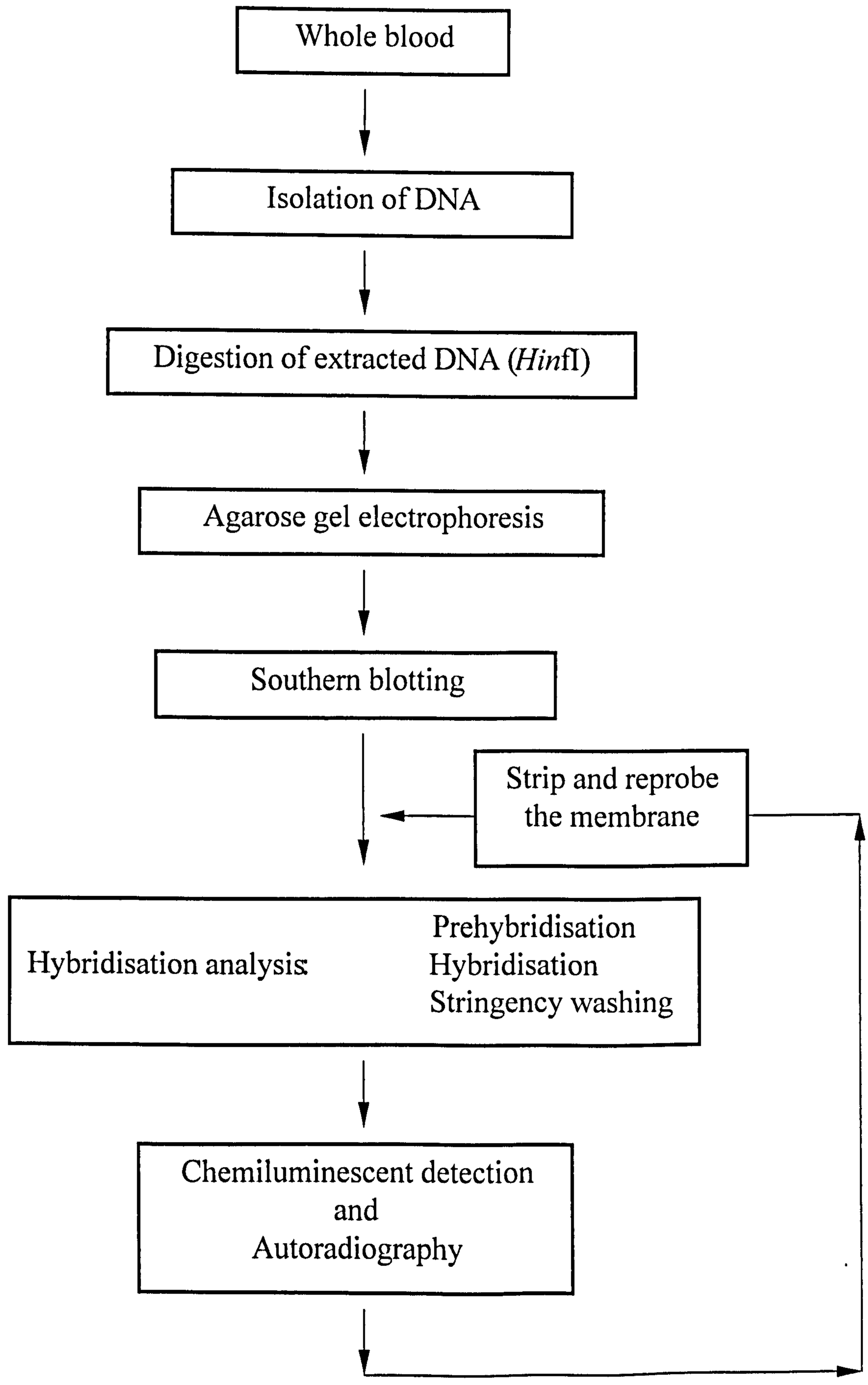
The extracted DNA as previously described in section 2.4 was used for PCR.

2.11.2 Dilution of DNA

The diluted DNA samples were checked on an agarose (I.D.NA 1% w/v) mini-gel. Lambda DNA size standards (31 ng, 20 ng, 10 ng and 5 ng) were purchased from Gibco-BRL and run next to the samples to estimate the quality and quantity of DNA in each sample. The samples were diluted up to 2-10 ng/µl by re-diluting the DNA used

69 A

Figure 2.2 Diagram showing the stages involved in the preparation of VNTR profiling.


Whole blood
Isolation of DNA
Digestion of extracted DNA (*Hin*fI)
Agarose gel electrophoresis
Southern blotting
Strip and reprobe the membrane
Hybridisation analysis:
Prehybridisation
Hybridisation
Stringency washing
Chemiluminescent detection and Autoradiography

previously for VNTR analysis, if required, to ensure that the optimal amount of DNA will be analysed.

2.11.3 Amplification of DNA samples

The GenePrint™ PowerPlex™ 1.2 system kit (Promega Corporation, Madison, WI, USA) which is commercially available was purchased and used to co-amplify eight simple tetrameric short tandem repeat (STR) loci. These loci are as the following D5S818, D13S317, D7S820, D16S539, vWA, THO1, TPOX, and CSF1PO plus a sex determining locus the amelogenin amplification.

2.11.4 Multiplex PCR

The co-amplification of the eight STR loci plus amelogenin locus was performed according to the manufacturer's instructions except that one third of total reaction volume was used. Each sample contained 5.7 µl of Milli-Q water, 0.8 µl Gold STR 10X buffer, 0.8 µl PowerPlex™ 1.2 10X Primer Pair Mix, 0.2 µl (0.8 u) AmpliTaq Gold™ DNA polymerase (Perkin-Elmer Applied Biosystem, USA), and 0.5-2.5 µl (2-5 ng) diluted DNA added to each reaction mixture.

The number of reactions to be set up was determined (usually 24) including positive (5 ng K562) and negative (Milli-Q water) control reactions. Reactions were mixed gently in 1.5 ml amber-coloured microcentrifuge tube (eppendorf, Germany). The reaction mixture then was allocated in 24 of 0.2 ml (thin-walled) thermo pre-labelled tubes (Life Sciences, UK). The DNA template then was added to each 0.2 ml tube and all tubes then placed in the 2400 thermal cycler (Perkin-Elmer, Aplied Biosystems). Aerosol-resistant pipette tips (Life Science International, UK) were used.

Note: Gold STR 10X buffer and PowerPlex™ 1.2 10X Primer Pair Mix was thawed and stored on ice. Each reagent was mixed by vortexing for 5-10 s before each use.

2.11.5 Storage conditions

All components were stored at -20 °C. The PowerPlex™ 1.2 10X primer pair mix, PowerPlex™ 1.2 allelic ladder mix, THO1 allele 9.3, fluorescent ladder (CXR) and matrix standards were stored in the dark. The post amplification components (allelic ladder, THO1 allele 9.3, fluorescent ladder (CXR) and matrix standard) were packaged separately to prevent cross-contamination.

2.11.6 PCR amplification

Perkin-Elmer GeneAmp® PCR System 2400 Thermal Cycler (Perkin-Elmer Applied Biosystems, USA) was used. The Polymerase Chain Reaction (PCR) and the cycling program carried out as the following: 11 min at 95 °C followed by 10 cycles at 96 °C for 1 min, then at 94 °C for 30 s, then at 60 °C hold for 30 s followed by 45 s at 70 °C.

This was followed by 22 cycles of 90 °C for 30 s, then at 60 °C for 30 s, after that at 70 °C for 45 s and finally 60 °C for 30 min.

2.12 310 GENETIC ANALYSER

The ABI PRISM® 310 Genetic Analyser uses a capillary electrophoresis system removing the need to pour and load conventional slab gel. Its highly automated, easy to use format makes it a widely used in many laboratories. DNA fragments can be separated in the capillary tube which filled with Performance Optimised Polymers (POP-4). This

instrument can analyse 48/96 samples each time and yields extremely reproducible result in approximately 30 min per sample.

GeneScan® analysis software which, works with the data collection software (multicolour fluorescent data) can quickly and accurately analyse PCR samples. Using an internal size standard which, is injected with each sample GeneScan software can automatically sizes the PCR products.

There are several different dye sets that can be used in different types of experiments. In this study a set of three different fluorescent dyes were used and these dyes are TMR (yellow), Fluorescein (blue), and CXR (red). The fluorescence from each dye set must be collected using the correct virtual filter set (A, B, C, etc.). Virtual filter set A were selected in this experiment which correspond to this dye set by choosing a module file GS POP4 A for each run.

2.12.1 Matrix file preparation

A matrix file was generated for using filter set A. A set of four standards was run under the same capillary gel conditions as used for samples and allelic ladders. Fluorescein matrix standard, TMR matrix standard and fluorescent ladder (CXR) (included with the kit) were used for the blue, yellow and red standards, respectively. For the fourth colour (green), the HEX matrix standard (PE, Applied Biosystems, USA) was used.

2.12.2 DNA typing

Loading cocktail was prepared by combining and mixing 1.5 µl of the fluorescent ladder containing carboxy-X-rhodamine labelled fragments (CXR), 60-400 bases (included with the kit) as Internal Lane Standard (ILS), with 17 µl of deionised formamide (Fluka, Gillingham-Dorest, UK) in 0.5 ml tube. 2 µl of PCR product were mixed with 18 µl of the loading cocktail. 2 µl of PowerPlex 1.2 allelic ladder mix were also mixed with 18 µl of the loading cocktail.

The samples were denatured at 95 °C for 3 min and immediately chilled on ice for 3 min. All tubes were placed into an autosampler tray that holds 48 tubes (0.5 ml). Capillary electrophoresis (47 cm x 50 µm) was carried out on an ABI PRISM® 310 Genetic Analyser (PE Applied Biosystems, USA). The running buffer, 1X Genetic Analyser Buffer, was used. The polymer was injected into the capillary at the anodic side. Samples were injected electrokinetically at cathodic side for 5 s at 15 kV in Performance Optimised Polymer 4 (POP4™; 1 ml glass syringe). Electrophoresis was performed at a voltage of 15 kV for 24 min and the capillary was kept at 60 °C.

Data were collected (raw data) using the ABI PRISM 310 collection software application, version 1.0.2, with the GeneScan run module GS POP4 A (virtual filter set A). The size of DNA fragments was determined by using the 310 GeneScan analysis software application, version 2.0.2 (PE Applied Biosystems) by comparing them to fragments contained in a size standard. See table 2.1.

Table 2.1 The allelic ladder information.

Fluorescein = 5′-terminal fluorescein label. (blue), TMR = 5′-terminal carboxy-tetramethylrhodamine label. (yellow)

STR locus	Fluorescent Label	Size range of allelic ladder components (bp)	Repeat numbers of allelic ladder components	Repeat numbers of alleles not present in allelic ladder
Amelogenin	TMR	209.5 (X) 215.8 (Y)	—	none
D5S818	Fluorescein	112 – 145	7 – 15	none
D13S317	Fluorescein	168 – 200	7 – 15	none
D7S820	Fluorescein	211 – 243	6 – 14	none
D16S539	Fluorescein	261 – 301	5, 8 – 15	none
vWA	TMR	126 – 166	11, 13 – 21	none
THO1	TMR	176 – 200	5 – 11	8.3, 9.3
TPOX	TMR	222 – 250	6 – 13	none
CSF1PO	TMR	291 – 327	6 – 15	none

2.13 STATISTICAL ANALYSIS SOFTWARES

Several computer programs for the analysis of VNTRs and STRs data that have been distributed by authors as free programs over the Internet were used. These programs are as the following.

1. Tools for Analysis of Population Statistics, PowerStats (Tereba A., 1999). Through Promega Co.

2. Genetic Data Analysis, version 1.0 (d16b) (Lewis P. and Zaykin D.). From the GDA Home Page at http://lewis.eeb.uconn.edu/lewishome/software.html.

3. Tools for population genetic analyses (TFPGA) version 1.3 (Miller M., 1997).

4. RxC: A program for the analysis of contingency tables (Miller M., 1997). From the R x C Home Page at http//herb.bio.nau.edu/~miller/rxc.htm.

CHAPTER 3: VARIABLE NUMBER TANDEM REPEATS

3.1 INTRODUCTION

Variable number tandem repeat (VNTR) loci are extremely polymorphic exhibiting a large number of alleles in a population, which yields a high probability of finding different alleles in different individuals. Although there are more than hundred alleles at some loci only 11-26 are generally resolvable (NRC, 1996). Applications of these highly informative genetic markers have been very useful in human identification in forensic medicine and paternity testing.

This study has been carried out since no data published for the UAE population nor Pakistani or Indian population who resides in the UAE. Before applying DNA methods to identity and paternity testing in the UAE populations, it is necessary to study the population structure of the UAE. The UAE contains a complex population from a lot of different ethnic groups. The majority, approximately 80%, is from the UAE Arab, Indian and Pakistani populations.

3.2 POPULATION SUBSTRUCTURE

In order to determine the value of any marker for individual identification it is desirable to obtain data on the allele/genotype frequencies in the populations of interest so that the forensic scientist will be able to provide an estimate of the rarity of a genetic profile. Lots of studies have been reported VNTR allele frequencies for a number of

ethnic groups, however, when this study was commenced, no published data for UAE, Pakistanis or Indian from the UAE were available.

The main purpose of this study was to examine the suitability of the application of DNA systems to forensic and paternity testing in the UAE. The population of the UAE is composed of different ethnic groups including Arabs, Pakistani, Indian, Bangladeshi, Filipino, Iranian, Afghan, Ceylonese, Russian, American and European. These populations are not fully homogenised. Allele frequencies may vary between different ethnic groups. Therefore, it was important to study these three populations (UAE, Indian and Pakistani) which makes 80% of the whole population in the UAE.

In addition, the high level of marriage between relatives among the local people increases the complex nature of the UAE population. Moreover, the average of a family size is around eight including the parents. These might have effects on HWE and the statistical interpretation of results, in particular applying the product rule. To investigate the effects caused by the presence of different ethnic groups in the UAE and the relatively high level of close marriages with a large size of a family, several VNTR loci were examined for the presence of population substructure.

3.3 SELECTION OF VNTR LOCI

There are a large number of VNTR loci in the human genome and hundreds of polymorphic sites have been identified and mapped to specific chromosomal regions. There are lots of different probes that can be used to detect these highly polymorphic loci. In this study five different human VNTR loci have been analysed, these were D1S7, D7S21, D12S11, D7S22 and D2S44. Both the size range and frequency distribution of

their alleles in the three ethnic groups UAE, Indian, and Pakistani residing in the UAE were determined.

These five loci have been studied extensively and a lot of different population data were published in Europe (Bjerre A., 1997; Holmlund G. *et al.*, 1994; Erikesn B. and Svensmark O., 1994a, b; Morling N. and Hansen H., 1993; Hansen H. and Morling N., 1993a, b; Papiha S. *et al.*, 1998), USA (Budowle B. *et al.*, 1991c) and some other countries (Moura-Neto R. and Budowle B., 1997; Tsui P. and Wong D., 1996; Valverde E. *et al.*, 1993). However, it was not possible to perform a comparative study between data published in the USA from several different ethnic groups and UAE nor Europe because different sets of loci were used in the USA (except D1S7, D2S44), different approaches of measurement techniques to size alleles (mm or bp) and different restriction enzymes were used as well. Also inter-laboratory comparison was not possible because the measurement error was very high (10%) (Schneider P. *et al.*, 1991). However, later the inter-laboratory variation decreased due to optimising and standardising the methods for VNTR typing with SLPs (Mudd J. and Baechtel F., 1994; Duewer D. *et al.*, 1995; Bjerre A., *et al.*, 1997) but was still high.

3.4 DESCRIPTION OF THE LOCI AND PROBES

3.4.1 D2S44 locus

D2S44 locus is located on chromosome 2 and detected by YNH24 probe (Nakamura Y. *et al.*, 1987a, b). This is a very common probe to many laboratories because of its sensitivity (Budowle B. *et al.*, 1988), working well under variety of protocols (Nakamura Y. *et al.*, 1987b; Balazs I. *et al.*, 1989). This locus is highly polymorphic with 97%

observed heterozygosity (Nakamura Y. *et al.*, 1987a, b; Balazs I. *et al.*, 1989) and mutation rate less than 0.2%. The size of core repeat is 32 base pairs (Probe catalogue, Cellmark diagnostics).

3.4.2 D7S22 locus

D7S22 locus is located on chromosome 7 and detected by G3 probe (Wong Z. *et al.*, 1986; 1987). This locus shows a high level of heterozygosity 96% (Smith J. *et al.*, 1990a) with 0.3% observed mutation rate (Jeffreys A. *et al.*, 1988; Smith J. *et al.*, 1990a). The size of core repeat is 37 base pairs (Jeffreys A. *et al.*, 1988).

3.4.3 D1S7 locus

D1S7 locus is located on chromosome 1 and detected by MS1 probe (Wong Z. *et al.*, 1987). This locus has the highest observed mutation rate which is 5.2% per gamete (Jeffreys A. *et al.*, 1988) therefore, displays the least variability between ethnic groups (Buffery C. *et al.*, 1991). This locus exhibits a high level of heterozygosity 99% (Smith J. *et al.*, 1990a), absence of common alleles (Wong Z. *et al.*, 1987; Smith J. *et al.*, 1990 b) and probably less sensitive if compared to other loci to the effects of population substructure (Lander E., 1989; Cohen J., 1990). MS1 is considered to be an extremely powerful forensic probe. However, in paternity testing could lead to the false exclusion of genuine parents because of its high mutation rate. Therefore, care must be taken into consideration when using this probe for paternity analysis. It has a 9 base pairs core repeat unit (Jeffreys A. *et al.*, 1988).

3.4.4 D7S21 locus

D7S21 locus is located on chromosome 7 and detected by MS31 probe (Wong Z. *et al.*, 1987). This locus has a high level of heterozygosity 97% (Smith J. *et al.*, 1990a) with low mutation rate 0.7% (Jeffreys A. *et al.*, 1988). The size of core repeat is 20 base pairs (Jeffreys A. *et al.*, 1988). Although D7S21 and D7S22 loci are located on the same chromosome, no detectable linkage was reported to be observed between either pair of syntenic DNA markers (Wong Z. *et al.*, 1987; Royle N. *et al.*, 1988; Hansen H. and Morling N., 1993b).

3.4.5 D12S11 locus

D12S11 locus is located on chromosome 12 and detected by MS43A probe (Wong Z. *et al.*, 1987). This locus observes a high level of heterozygosity 96% (Smith J. *et al.*, 1990a) with only 0.7% mutation rate (Jeffreys A. *et al.*, 1988; Smith J. *et al.*, 1990a). The size of core repeat is 45 base pairs (Jeffreys A. *et al.*, 1988). See table 3.1.

Table 3.1 Summery information published about the five probes used in this study from Cellmark diagnostics probe catalogue (Caucasian population).

Probe	Locus	Repeat unit (bp)	Mutation rate %	Heterozygosity %
YNH24	**D2S44**	32	0.2	97
G3	**D7S22**	37	0.3	96
MS1	**D1S7**	9	5.2	99
MS31	**D7S21**	20	0.7	97
MS43A	**D12S11**	45	0.7	96

3.5 SAMPLES COLLECTION

Blood samples were collected from three most populous populations including UAE, Pakistani and Indian individuals. According to several scientific bodies such as National Research Council (NRC) and International Society for Forensic Genetics (ISFG) samples should be collected at random for any marker to build up a reference database. Therefore, samples were collected from blood banks from Abu Dhabi and Sharjah in order to make the population as random as possible and to obtain a sufficient quantity of samples for analysis.

In small databases rare alleles may not be accurately represented and their frequency estimate will often be arbitrarily increased. In general, 100 to 200 unrelated individuals (200-400 characteristics or alleles) are usually recommended to be collected for each DNA system to assess allele frequencies (ISFH 1991 report, 1992; NRC recommendations, 1992; 1996). Therefore, based on these recommendations, whole blood samples were obtained from 173 unrelated UAE Arab, 154 unrelated Indian and 112 unrelated Pakistani people. These samples were collected from two cities in order to represent the whole UAE populations (section 2.3.1).

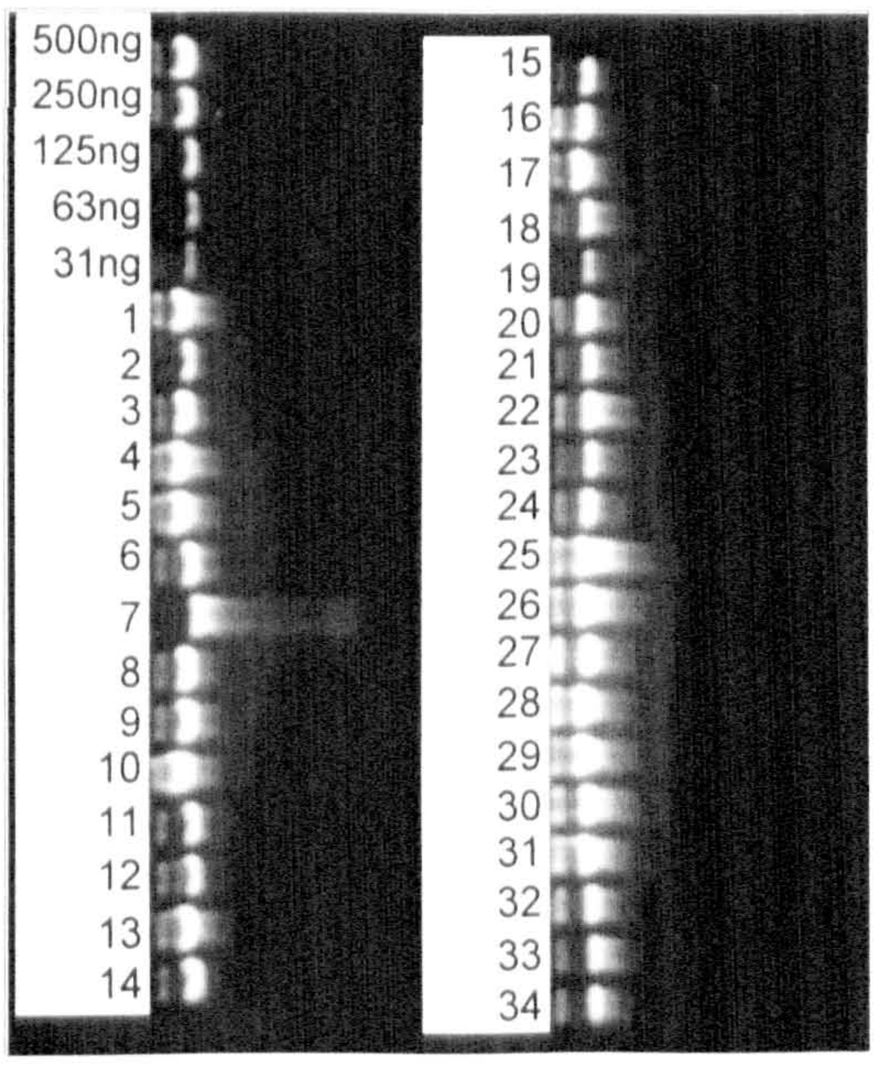
3.5.1 DNA extraction

The DNA was extracted successfully from 439 individuals from the three populations by using Puregene™ DNA Isolation Kit (section 2.4). The quality and the quantity of extracted DNA were estimated by using mini-gel electrophoresis as shown in figure 3.1.

82A

Figure 3.1 showing the yield of DNA from blood samples.

These DNA samples were extracted using Puregene kit. The extracted DNA of unknown quantity was compared to a serial dilution of known quantity of lambda DNA (500 ng, 250 ng, 125 ng, 63 ng and 31 ng). A large quantity of high molecular weight (100 ng/µl) DNA was routinely recovered. Some degradation can be observed as well and seen as a smear of long tail as in line 7.


500ng
250ng
125ng
63ng
31ng
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34

The extracted DNA of unknown quantity was compared to a serial dilution of known quantity of lambda DNA (500 ng, 250 ng, 125 ng, 63 ng and 31 ng) as seen in figure 3.1. A large quantity of high molecular weight (100 ng/µl) DNA was routinely recovered.

3.5.2 Analysis of restricted DNA

To demonstrate the size differences between different minisatellite alleles it is necessary to isolate them from their surrounding DNA. This is done using *Hin*fI restriction enzyme. This enzyme cleaves the surrounding DNA into relatively small fragments, but leaves the minisatellite DNA intact.

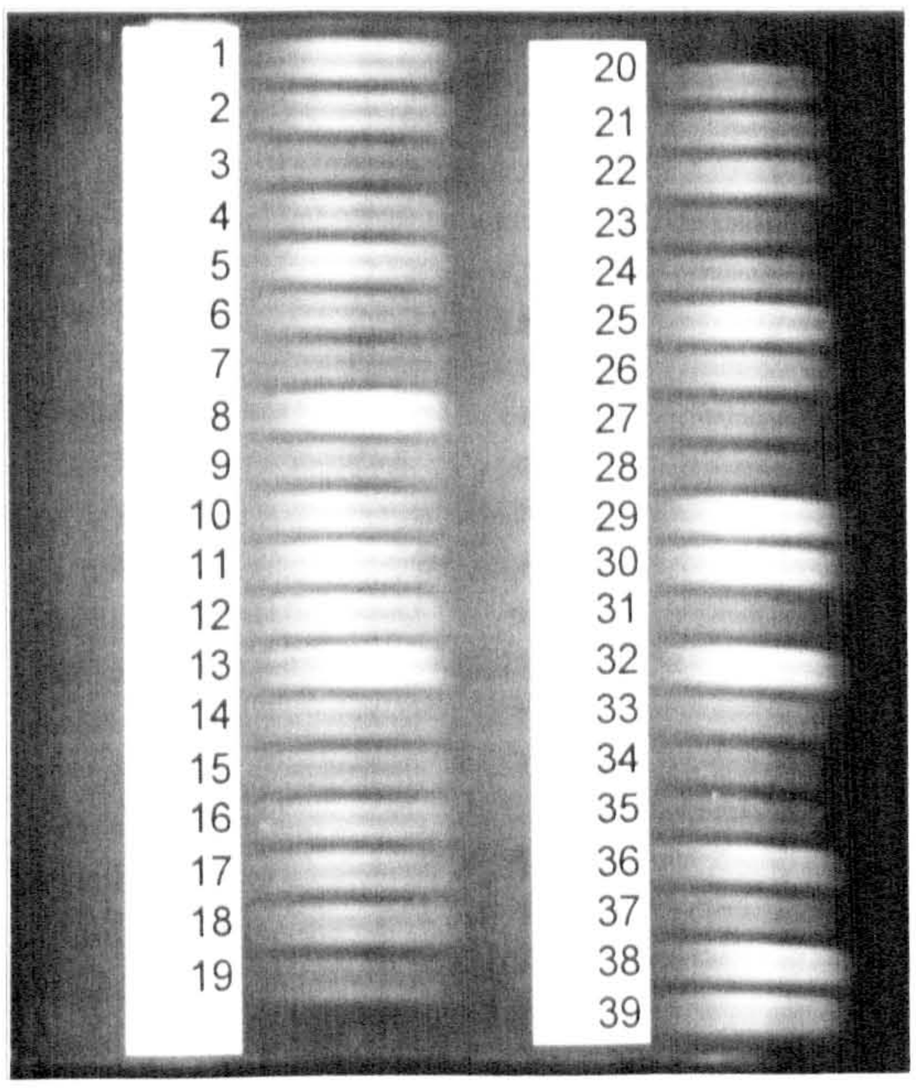
The *Hin*fI enzyme is a four base pair cutter that cleaves double stranded DNA at specific sites, approximately every 256 bp (4^4) in the genome. However, this enzyme will not cleave within the repeat unit (except if mutations have occurred). It is necessary to check for complete digestion because if partially digested DNA was typed a wrong allele size or null allele might be the result.

A complete digest was evidenced by a uniform smear made up of different size fragments running down the length of the lane. A completely digested DNA can be observed as a bright white smear as is shown in figure 3.2. The smear of ethidium bromide stained DNA was viewed under UV light on a transilluminator at 312 nm. Gels were photographed and documented. See figure 3.2.

84 A

Figure 3.2 The DNA extracted from samples was digested using *Hin*fI enzyme.

All DNA samples observes a complete digestion.


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39

3.5.3 Analytical gel electrophoresis

Digested DNA was run on 0.75% agarose/TBE gel. The DNA was blotted by capillary action and transferred onto nylon membrane by the modified method of Southern (Southern E., 1975). The DNA was fixed using UV light and oven baking. A nylon membrane was sequentially hybridised with five single locus probes (SLPs) (YNH24/D2S44, G3/D7S22, MS1/D1S7, MS31/D7S21, and MS43A/D12S11) and detected by chemiluminescent on an X-ray film. See figure 3.3a, b, c, d, e).

3.6 ANALYSIS OF VNTR PROFILES AND ESTIMATION OF ALLELE FREQUENCIES

The size of each allele on the autoradiograph were determined using the image analysis software BioMax 1D (Kodak Scientific Imaging System) for the three ethnic groups. The use of software increased the sizing precision (Bjerre A., *et al.*, 1997). The DNA fragment size markers (ladder) consisted of a mixture of defined fragment lengths ranging from 526 bp to 22,621 bp. The ladder loaded in three positions, two flanking the samples and one central on each gel. The size of unknown fragments were estimated based on their mobility relative to the size marker fragments in order to obtain the most accurate size of the unknown fragments.

Briefly, an image of the autorad is scanned and made adjustable for computer analysis. The program defines both the centres of each band and the relationship between marker fragment sizes and electrophoretic migration. Subsequently, the sizes of VNTR bands are determined relative to the flanking lanes of size marker fragments. The analysis is largely objective, with the operator providing information as to the contents of each lane

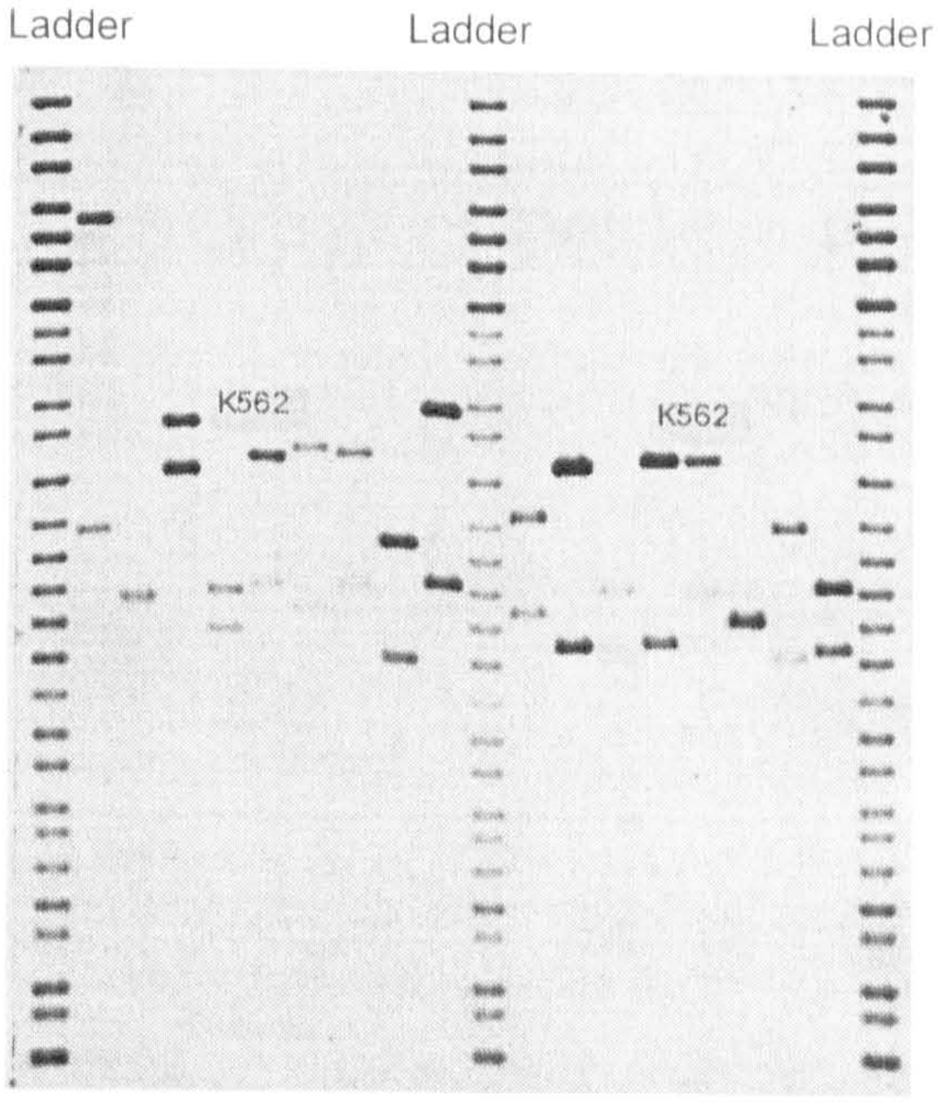
86 A

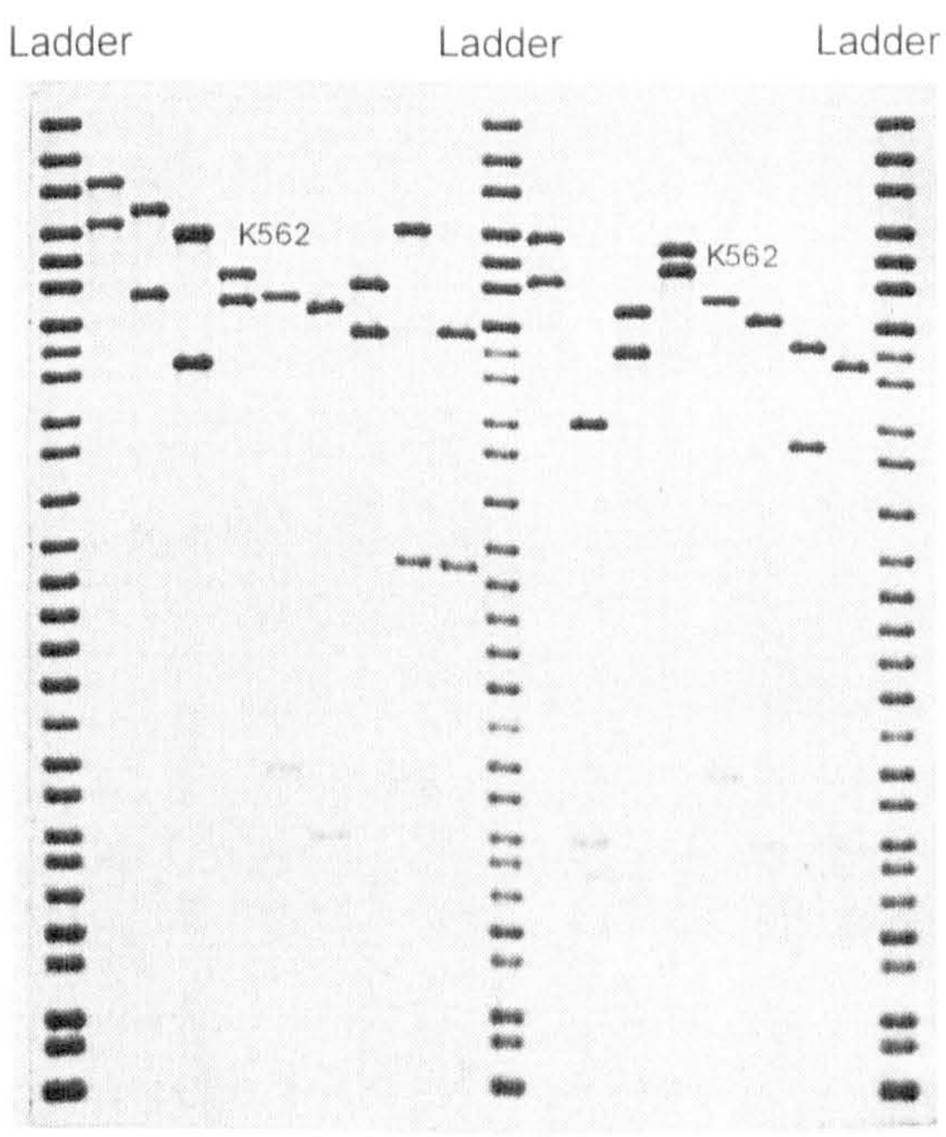
Figure 3.3a Southern blot of DNA samples (*Hin*fI enzyme digested).

This autoradiograph illustrating fragment length variation at D2S44 detected by the probeYNH24. As seen DNA size marker (ladder) was loaded in three positions, two flanking the samples and one central on the gel, two K562 control DNA samples, and fifteen unknown samples were routinely loaded.

Figure 3.3b Southern blot of DNA samples (*Hin*fI enzyme digested).

This autoradiograph illustrating fragment length variation at D7S22 detected by the probe G3. As seen DNA size marker (ladder) was loaded in three positions, two flanking the samples and one central on the gel, two K562 control DNA samples, and fifteen unknown samples were routinely loaded.


Ladder
Ladder
Ladder
K562
K562


Ladder
Ladder
Ladder
K562
K562

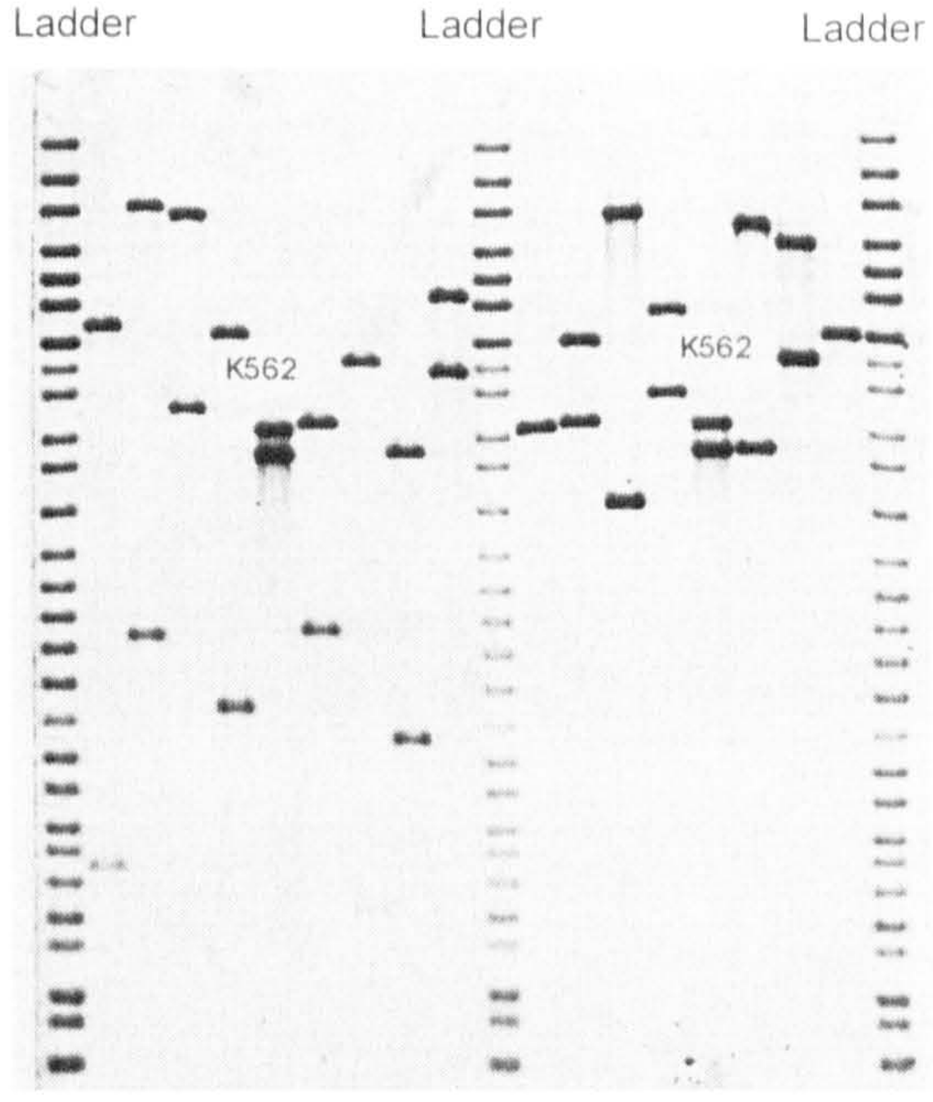
87A

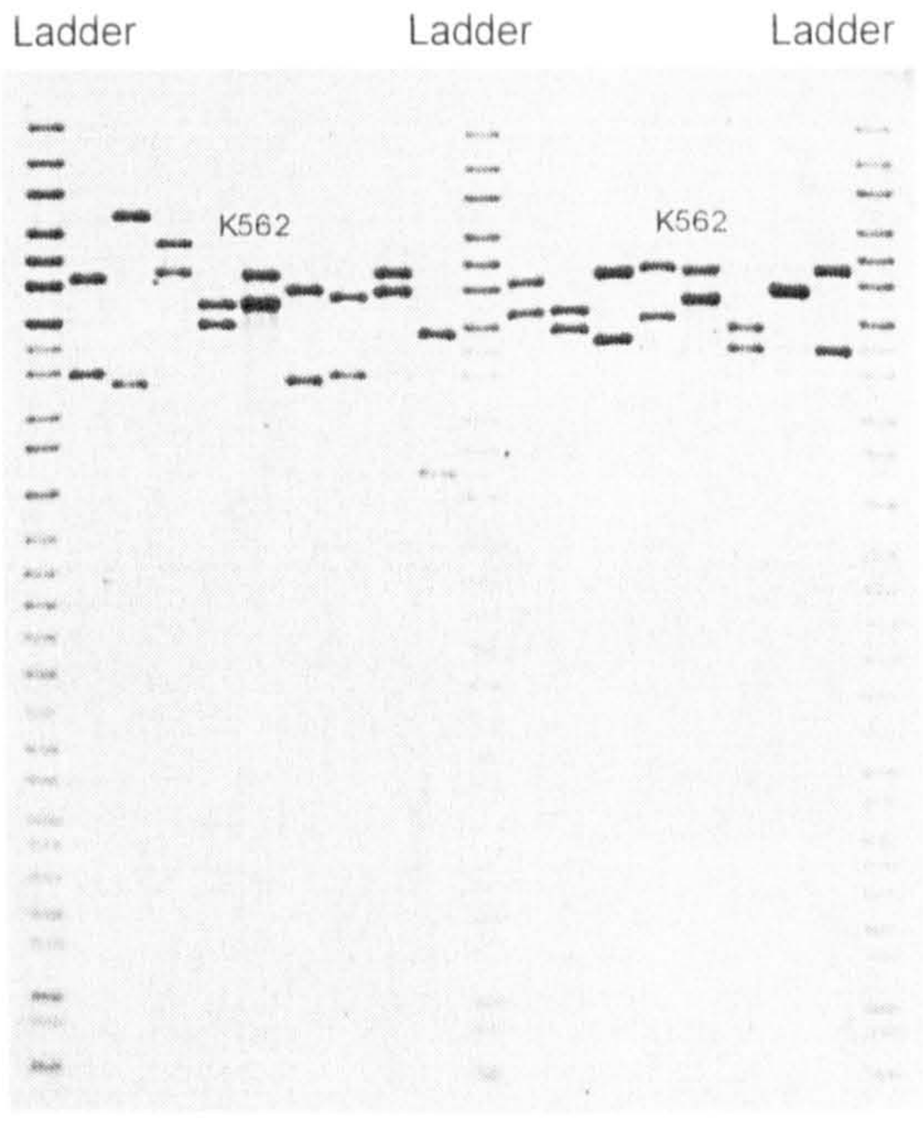
Figure 3.3c Southern blot of DNA samples (*Hin*fI enzyme digested).

This autoradiograph illustrating fragment length variation at D1S7 detected by the probe MS1. As seen DNA size marker (ladder) was loaded in three positions, two flanking the samples and one central on the gel, two K562 control DNA samples, and fifteen unknown samples were routinely loaded.

Figure 3.3d Southern blot of DNA samples (*Hin*fI enzyme digested).

This autoradiograph illustrating fragment length variation at D7S21 detected by the probe MS31. As seen DNA size marker (ladder) was loaded in three positions, two flanking the samples and one central on the gel, two K562 control DNA samples, and fifteen unknown samples were routinely loaded.

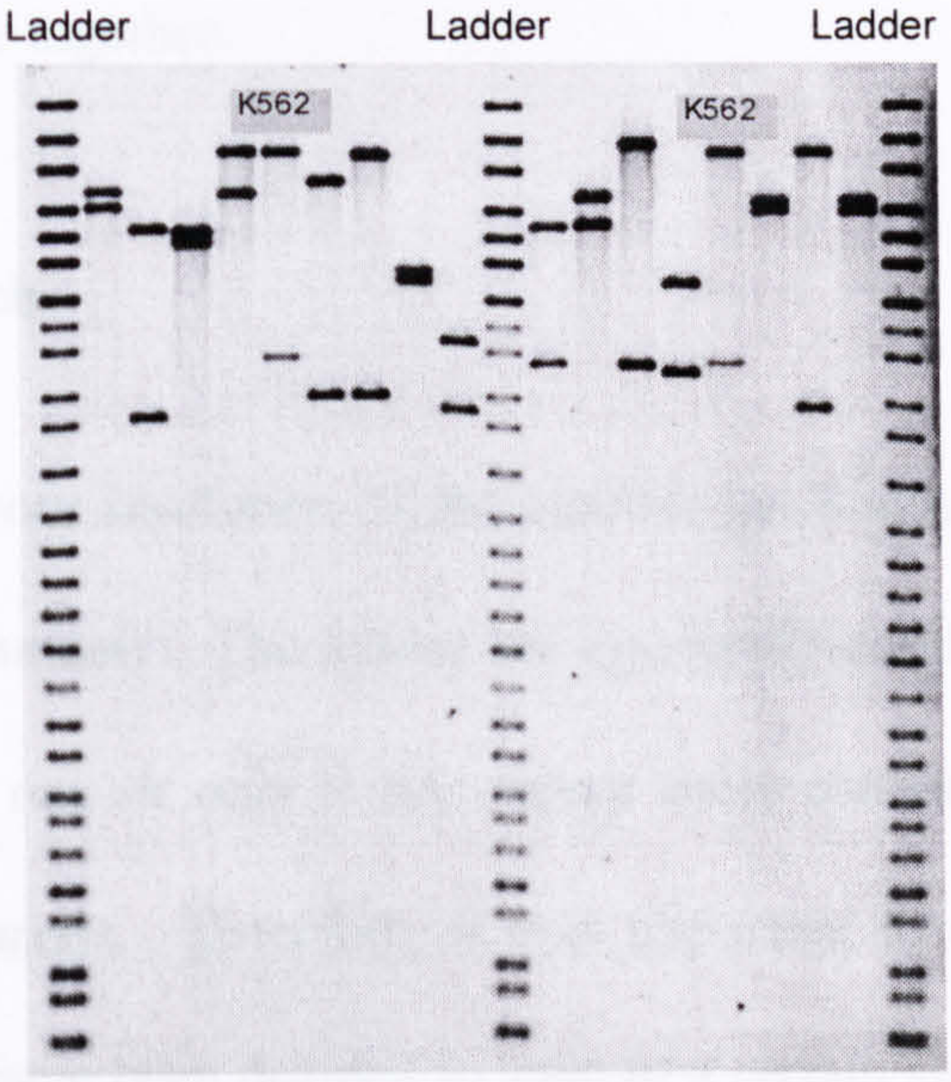

Ladder
Ladder
Ladder
K562
K562


Ladder
Ladder
Ladder
K562
K562

88A

Figure 3.3e Southern blot of DNA samples (*Hin*fI enzyme digested).

This autoradiograph illustrating fragment length variation at D12S11 detected by the probe MS43A. As seen DNA size marker (ladder) was loaded in three positions, two flanking the samples and one central on the gel, two K562 control DNA samples, and fifteen unknown samples were routinely loaded.


Ladder
Ladder
Ladder
K562
K562

(e.g., markers or genomic sample, number of band present). An edit function allows removal of spurious bands, such as those caused by background. Bands may also be edited into the profile if required, for example the machine may fail to distinguish between two bands that are close together.

3.6.1 Binning approach

Because of the poor resolution of the agarose gel Southern blot techniques alleles can not be defined accurately. The alleles are essentially continuous rather than distinct. Alleles that differ by one or only a few repeat units cannot easily be resolved using agarose gel electrophoresis. Therefore, it was important to collect a group of similar fragment sizes which have been decided to span to a certain size in advance and consider it as an allele. Each size ranges called a bin and each bin consider as an allele. Each bin contains a number of similar allele sizes. All alleles in a population data sample have been placed into different bins to simplify the estimate of VNTR allele frequencies. See figure 3.4.

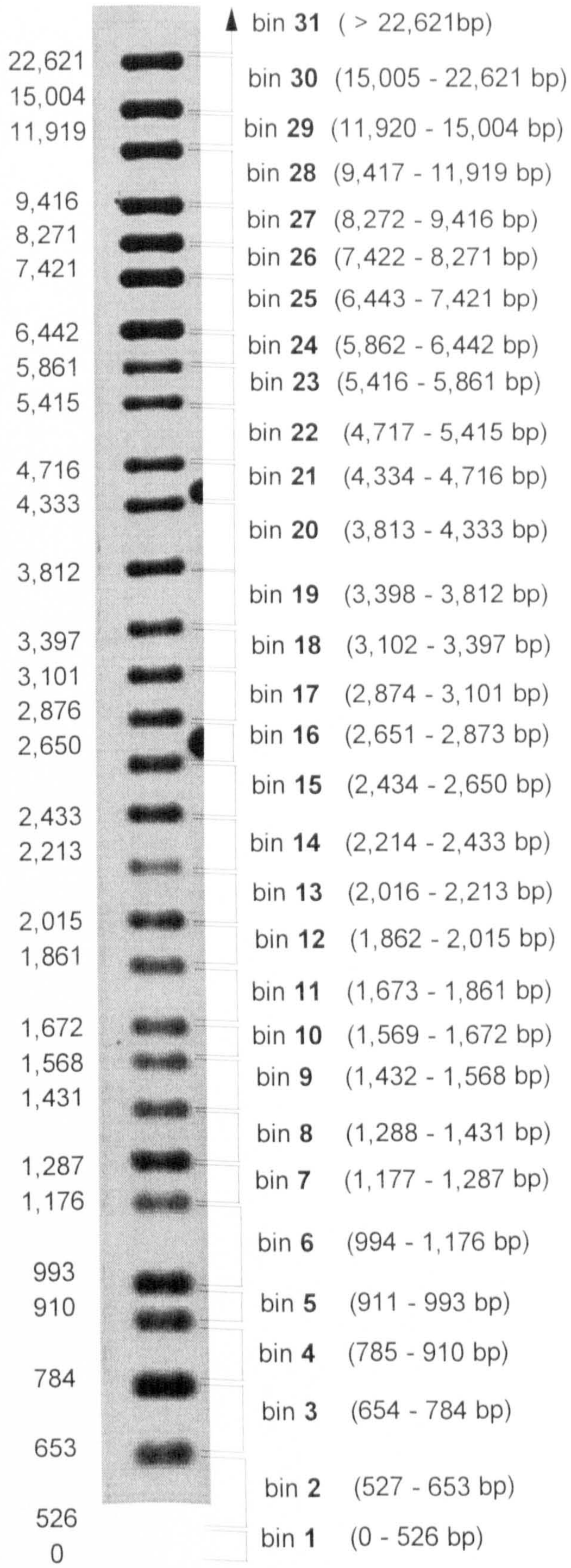
3.6.2 Fixed bin approach

Different statistical criteria for the construction of databases and estimating allele frequencies were proposed (Baird M. *et al.*, 1986; Budowle B. *et al.*, 1991b; Gill P. *et al.*, 1990; Pascali V. *et al.*, 1991). Two analytical procedures are widely used, the fixed bin (Budowle B. *et al.*, 1991b) and the floating bin (Balazs I. *et al.*, 1989) methods. Although the floating bin method is statistically preferable, the fixed bin method is simpler and easier for the average laboratory to use. Moreover, comparison between different populations is not possible by using the floating bin method. Fixed bin allows the

90A

Figure 3.4 The DNA size marker (ladder).

The diagram shows the size marker DNA fragments (ladder) in base pairs. The size marker consists of 30 known DNA fragment sizes. The total number of bins is 31 and each bin considers as an allele. The size range of each bin is different. The smallest fragment (bin) is not seen because it was run off the gel.


bin 31 (> 22,621bp)
22,621
bin 30 (15,005 - 22,621 bp)
15,004
11,919
bin 29 (11,920 - 15,004 bp)
bin 28 (9,417 - 11,919 bp)
9,416
bin 27 (8,272 - 9,416 bp)
8,271
bin 26 (7,422 - 8,271 bp)
7,421
bin 25 (6,443 - 7,421 bp)
6,442
bin 24 (5,862 - 6,442 bp)
5,861
bin 23 (5,416 - 5,861 bp)
5,415
bin 22 (4,717 - 5,415 bp)
4,716
bin 21 (4,334 - 4,716 bp)
4,333
bin 20 (3,813 - 4,333 bp)
3,812
bin 19 (3,398 - 3,812 bp)
3,397
bin 18 (3,102 - 3,397 bp)
3,101
bin 17 (2,874 - 3,101 bp)
2,876
bin 16 (2,651 - 2,873 bp)
2,650
bin 15 (2,434 - 2,650 bp)
2,433
bin 14 (2,214 - 2,433 bp)
2,213
bin 13 (2,016 - 2,213 bp)
2,015
bin 12 (1,862 - 2,015 bp)
1,861
bin 11 (1,673 - 1,861 bp)
1,672
bin 10 (1,569 - 1,672 bp)
1,568
bin 9 (1,432 - 1,568 bp)
1,431
bin 8 (1,288 - 1,431 bp)
1,287
bin 7 (1,177 - 1,287 bp)
1,176
bin 6 (994 - 1,176 bp)
993
bin 5 (911 - 993 bp)
910
bin 4 (785 - 910 bp)
784
bin 3 (654 - 784 bp)
653
bin 2 (527 - 653 bp)
526
bin 1 (0 - 526 bp)
0

population structure to be more easily analysed. Therefore, the fixed bin method was adopted and the data was initially analysed using the fixed bin method in this study.

The strategy was first described by Wainscoat *et al.* (Wainscoat J. *et al.*, 1987) and subsequently refined by Budowle B. *et al.* (Budowle B. *et al.*, 1991b). The procedure simply divides an autoradiograph into 31 sectors. Alleles are operationally defined according to the bin designation and their frequency of occurrence is the sum of the frequencies of each fragment length within the bin. There are a total of 31 different size bins in this system. The bin size dimensions and numbering convention are given in figure 3.4. Hereafter, the bins themselves will be used to operationally define up to 31 different "alleles" for each of the five VNTR loci as in method # 1 (appendix D).

As suggested by Budowle and Monson (1990) (Waye J., 1990), bins that have few observed events are considered to be statistically insignificant. In order to reduce the significance placed on such rare events, adjacent bins were combined (rebinned) to give a total of 2 or more events (~ 1%). The frequency of occurrence of fragment lengths in the "expanded bin" is then used to derive the estimated genotype frequency. The binning concept is used to derive frequencies for VNTR alleles.

In this study (method # 2), each bin should has a minimum of 3 events in the UAE and Indian populations and minimum of 2 in the Pakistani population. As shown in appendix D.

In method # 3 it was rebinned in a different way, the same size ranges for the three populations was used and minimum of 2 events in each bin should be observed in the three populations. The reason for creating this method is for comparison and statistical

analysis between the three populations. The size ranges of a bin for the three populations are the same. See appendix D.

The frequency of occurrence of all alleles in the three populations were estimated for each SLP, which is important to be determined prior to its use in cases and comparison between populations.

Method # 2 is recommended to be used in forensic and paternity testing to estimate the genotype frequency for each individual in the population because it is considered as a conservative method, especially when the databases are small (Budowle B. *et al.*, 1993a). See appendix D.

3.6.3 Allele size ranges observed

The fragment size (allele) ranged from approximately 866 bp (D7S22) to approximately 28,334 bp (D1S7) in the three population tested. See table 3.2 for smallest and largest three alleles observed at the five loci in the three populations.

Table 3.2 This table shows the largest and smallest three alleles for the five loci studied in the three populations.

Locus	UAE		Indian		Pakistani	
	smallest	largest	smallest	largest	Smallest	largest
D2S44	**1,375** 1,391 1,673	7,611 7,413 7,383	1,991 2,013 2,046	9,039 7,388 7,176	1,912 2,003 2,011	**9,760** 7,007 6,075
D7S22	1,517 1,537 1,574	16,463 15,043 12,979	1,430 1,547 1,599	**25,019** 24,561 12,905	**866** 1,524 1,601	14,643 13,637 12641
D1S7	**871** 1,259 1,349	**28,334** 24,159 23,986	1,155 1,218 1,307	26,397 23,233 21,201	1,028 1,073 1,234	26,142 20,588 14,841
D7S21	**1,450** 1,718 1,732	**12,039** 11,726 11,051	3,298 3,704 3,719	11,589 11,500 10,600	2,353 3,137 3,286	11,974 11,174 10,656
D12S11	**1,934** 3,517 3,668	16,226 16,084 15,555	2,322 3,570 4,506	**21,444** 20,345 19,781	3,236 3,681 4,260	20,059 18,774 18,623

The smallest and largest alleles observed in this study for the three populations were at locus D1S7. See figure 3.5a and b.

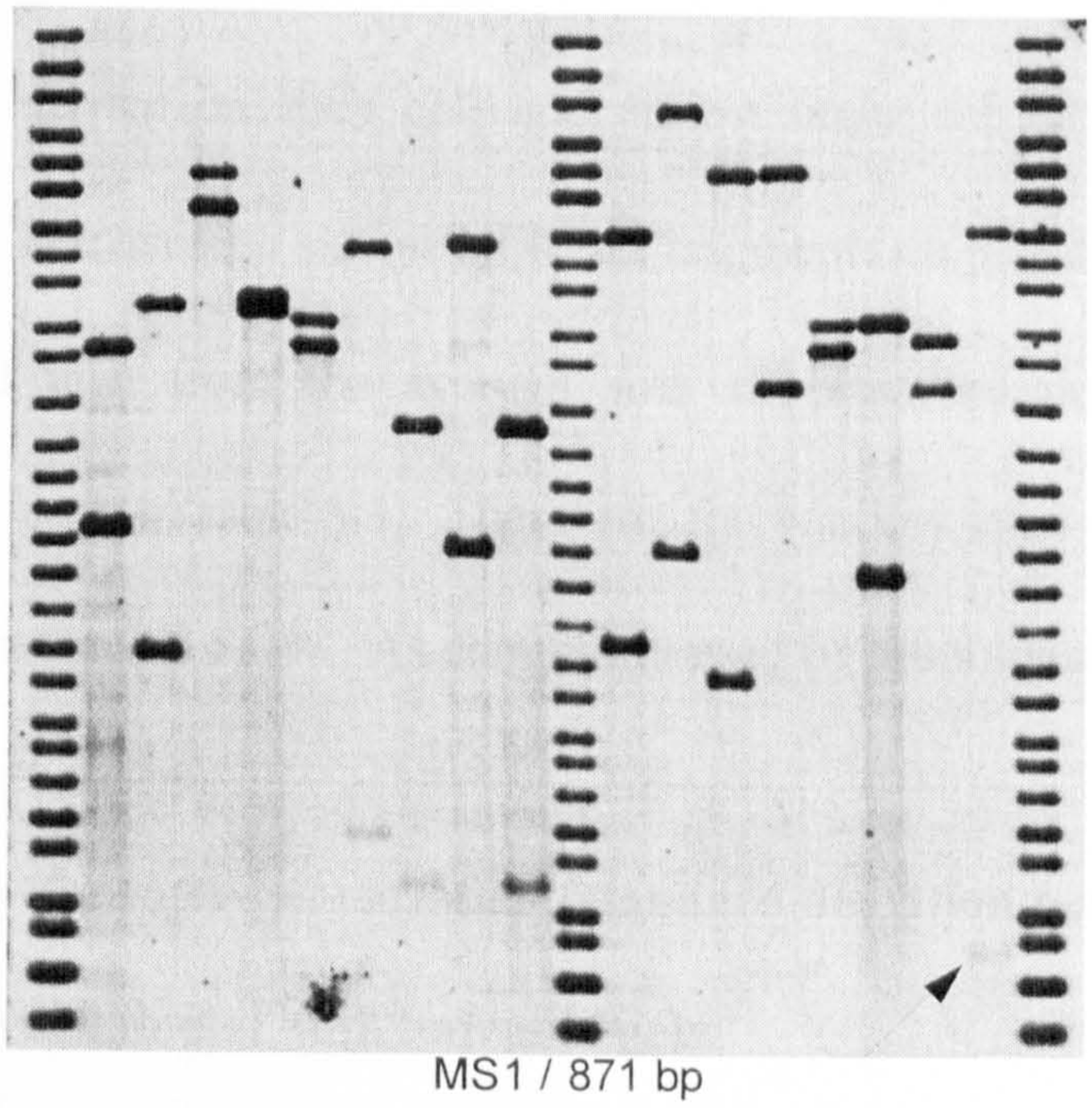
3.7 INTER AND INTRA-GEL VARIATION

The mobility of DNA fragments in agarose gel electrophoresis are not precise and varies from gel to gel in the same laboratory. Even on the same gel at different positions the variation in migration was observed. Therefore, it was necessary to calculate the variation rate within a gel (intra-gel variation) and between different gels (inter-gel variation). The precision can be determined through repeated typing of a single sample.

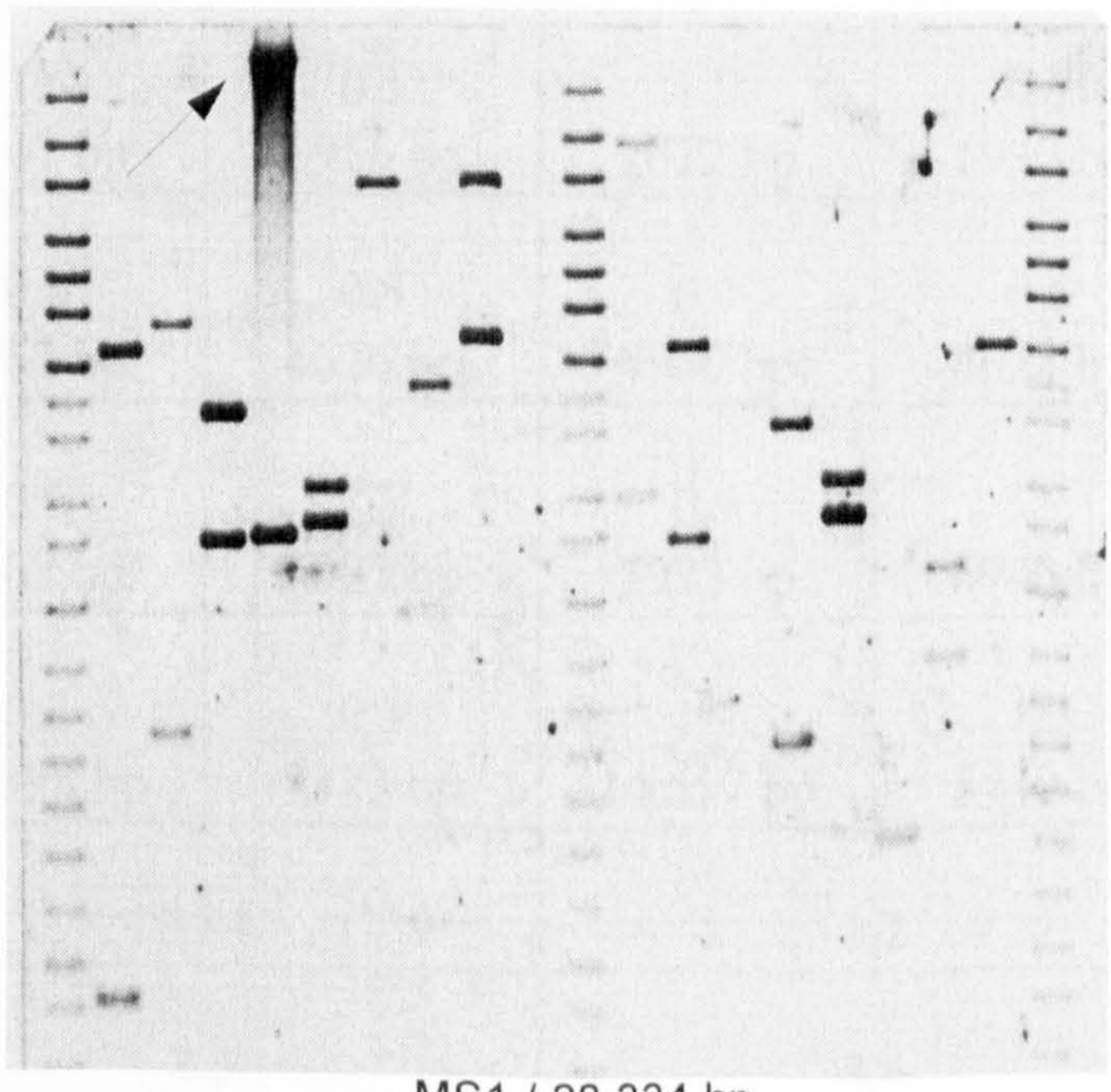
94A

Figure 3.5a An X-ray film showing one of the smallest allele observed in this study for the three populations tested.

Figure 3.5b An X-ray film showing the largest allele observed in this study for the three populations tested.



MS1 / 871 bp



MS1 / 28,334 bp

3.7.1 Inter-gel variation

The variation between different gels was measured by running human cell line (K562) on 32 gels (twice on each gel) and almost under the same conditions. See appendix B and C. The average size of all K562 fragments on 32 gels for the five probes was calculated and then from the average size, the standard deviation (SD%) was determined (1%) as seen in table 3.3. Therefore, the window (size range) for any allele was set to its average size ±3 SD%. As commonly used by the forensic community.

Table 3.3 Summery for inter-gel variation (standard deviation percentage) of human cell line K562 and variation at both halves of gels.

Locus	Left (lane 6)		Right (lane 16)		Average SD%
	SD% of Large fragments	SD% of Small fragments	SD% of Large fragments	SD% of Small fragments	
D2S44	0.9 4171 bp	1 2958 bp	1.1 4154 bp	1.4 2913 bp	1.1
D7S22	1.5 7101 bp	0.5 1956 bp	1.9 7032 bp	0.7 1937 bp	1.1
D1S7	0.4 4952 bp	0.5 4626 bp	0.7 4960 bp	0.6 4623 bp	0.5
D7S21	0.6 7902 bp	1.7 7005 bp	0.9 7923 bp	2 6940 bp	1.3
D12S11	0.8 13,898 bp	0.5 5373 bp	1.3 14,517 bp	1.3 5449 bp	1
SD%	0.84		1.2		1
3 SD%	**2.5**		**3.6**		**3**

3.7.2 Intra-gel variation

To measure the variation within gels, K562 was run twice on each gel in two separated positions, once on each half of the gel. The mean of four K562 fragments on each gel was calculated for the five probes. From the mean, the standard deviation (SD%) was determined (0.8%) as shown in table 3.4. When 3 SD% was calculated the measurement error was 2.4%.

Table 3.4 Summery for intra-gel variation (standard deviation percentage) of human cell line K562.

Probe	K562 Fragments	SD%
D2S44	Large (4163 bp) Small (2936 bp)	0.5 1.1
D7S22	Large (7066 bp) Small (1946 bp)	1.1 0.8
D1S7	Large (4956 bp) Small (4624 bp)	0.5 0.5
D7S21	Large (7912 bp) Small (6972 bp)	0.6 1.2
D12S11	Large (13889 bp) Small (5350 bp)	0.9 0.6
	Average SD%	**0.8**
	3 SD%	**2.4**

3.8 MATCH CRITERION

There is no absolute match criterion in VNTR typing, and the evaluation involves determining whether the two profiles are sufficiently similar such that they could have come from the same source or individual.

The larger the number of loci tested, the greater the chance of exclusion can be obtained. The binning concept is used only to derive frequencies for VNTR alleles and not to determine whether VNTR fragments "match". Once it has been established that VNTR profiles match, the statistical significance of the result must be derived.

When two VNTR fragments have been run having the same length twice in two different lines within a single gel or on two different gels they will migrates at different speed. Variations in measurement error between different gels were calculated and the measurement error window has been estimated to be ±3%. The measurement error window within a gel was estimated to be ±2.4%. In this study a window with boundaries 0.97x to 1.03x for a fragment of estimated length x was adopted (two different gels). If an estimated length is close to the boundary of a fixed bin, it is possible that the true length belongs to the neighbouring bin. If there are two bins having a common boundary lying in this window, the fragment is assigned to the bin with the highest population frequency (NRC, 1996).

For example, a DNA sample of an individual from UAE has been run on two separate gels and the fragments sizes for the five loci were as shown in table 3.5.

Table 3.5 shows an example of two profiles where they match at all five loci. These two samples are from one individual but run on two different gels.

Gel/locus	D2S44	D7S22	D1S7	D7S21	D12S11
Run # 1	4,759	3,275	*6,536*	7,962	10,434
	2,273	2,756	868	5,773	9,223
Run # 2	4,746	3,248	*6,223*	7,966	10,405
	2,244	2,732	854	5,807	9,262

Two fragments of D1S7 locus has been chosen to check whether they are within ±3% or not. The fragment size in the first run was 6536 bp [(-3% = 0.97 x 6536 = 6340 bp) and (+3% = 1.03 x 6536 = 6760 bp)] and the fragment size in the second run was 6223 bp [(-3% = 0.97 x 6223 = 6036 bp) and (+3% = 1.03 x 6223 = 6410 bp)]. Therefore, it can be noticed that these two fragments overlaps each other and since they are within the measurement error ±3% match can be declared.

A match was declared in another sample which was run twice on a single gel (intra-gel variation = ±2.4%) since the two fragments overlaps each other in all five loci (each locus at a time) and since they are within the measurement error ±2.4% (table 3.6).

Table 3.6 shows an example of two profiles where they match at all five loci. These two samples are from one individual and run on a single gel.

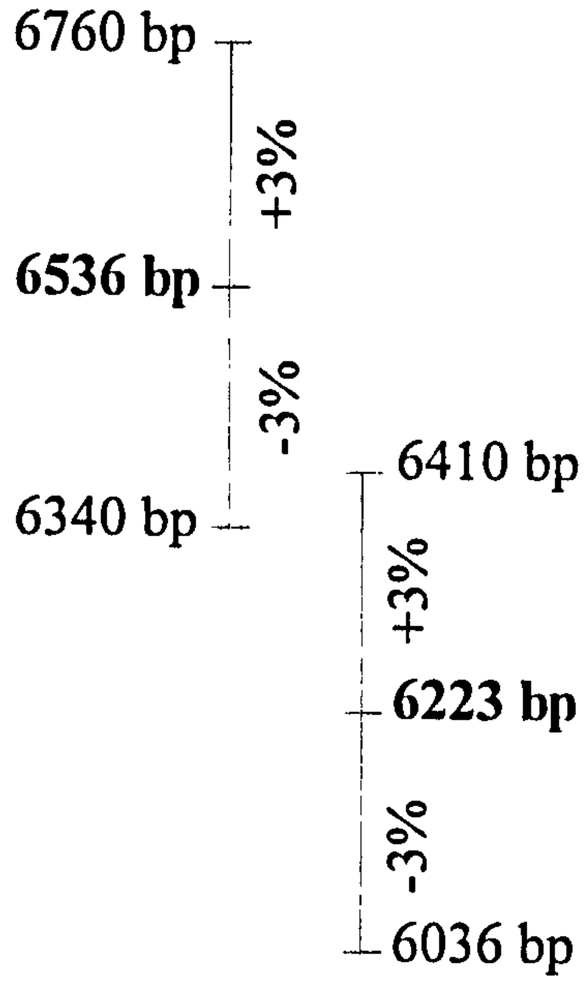
Single Gel/locus	D2S44	D7S22	D1S7	D7S21	D12S11
Position # 1	4,816	3,316	6,454	7,950	10,417
	2,320	***2,809***	888	5,858	9,257
Position # 2	4,743	3,265	6,374	7,871	10,321
	2,287	***2,764***	880	5,830	9,197

Example of calculating uncertainty windows for two fragments at D7S22 was calculated as the following. A single sample was run twice on a single gel in two different positions, 1 and 2. Position # 1 sample band size is measured as 2,809 bp. 2.4% of that is 67 bp (0.024 x 2,809 = 67). The lower limit of the uncertainty window is 2,809 – 67 = 2,742 bp. The upper limit is 2,809 + 67 = 2,876 bp. Position # 2 sample band size is measured as above and gives a range of 2,698 to 2,830 bp. Since the two samples overlap each other the two bands do match (Figure 3.6). In the same way the calculation can be done for all fragment pairs for the five loci.

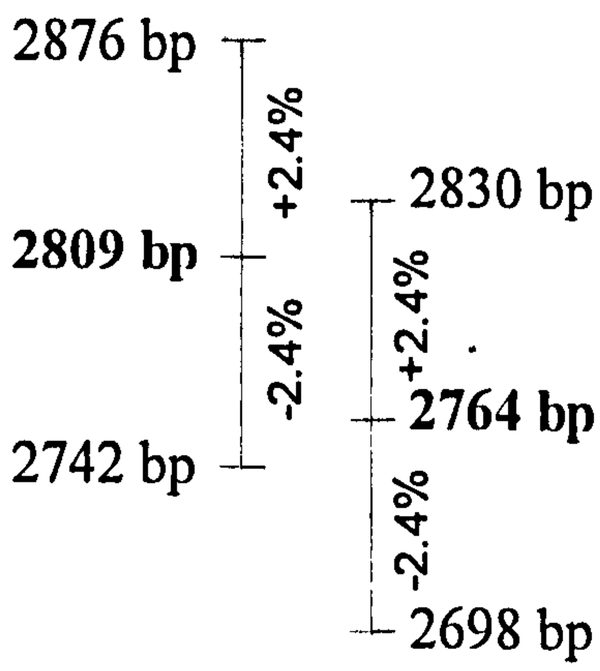
100A

Figure 3.6 The intra-gel and inter-gel variations.

These two diagrams show the extent of uncertainty and how the two windows overlap each other. In this case a match between these two fragments is declared.



Inter-gel variation



Intra-gel variation

3.9 DISCUSSION

3.9.1 Samples collection

This study was performed on 439 samples. All samples were obtained from randomly selected people from three different populations residing within the UAE. These people were chosen because they constitute 80% of the total UAE population. The VNTR analysis requires a high molecular weight. Therefore, whole fresh blood samples, 5 ml/EDTA tubes, were collected.

3.9.2 DNA extraction and digestion

The DNA was extracted from blood samples by using Puregene™ DNA Isolation kit. This kit allowed a rapid isolation and yielded a high molecular weight. This kit also allowed the biohazard of organic extraction solvents to be avoided. The DNA was successfully digested with *Hin*fI restriction enzyme. The *Hin*fI enzyme is widely used in most of the European forensic and commercial laboratories such as Cellmark Diagnostics and Forensic Science Service, UK (Hallenberg C. and Morling N., 2001).

3.9.3 VNTR system

The VNTR markers are extremely polymorphic having a very large number of different alleles. Therefore, these markers are very useful in forensic and paternity analysis. Because of the nature of the agarose gel electrophoresis the alleles will migrate with measurement uncertainty. Even on a single agarose gel electrophoresis the migration of the same DNA sample which run twice will not be the same. The two halves of a gel displayed different level of variability.

DNA size standards were placed in three positions on each gel flanking unknown DNA samples in order to aids precision of measurements. A common positive control, DNA from the cell line K562, which is commercially available, was used to check that the system is working as it should work. Additionally, Genomic DNA, usually from a staff member can be used instead of K562 (Bjerre A. *et al.*, 1997).

3.9.4 Non-isotopic detection

Labelling of probes with non-radioactive systems is very common and the radioactively detection methods is almost replaced by chemiluminescently detection. With the new system it was possible to avoid the biohazards of radioactivity. This system also allowed a faster development of thinner bands, which make a more precise determination of the DNA fragment lengths possible. Using non-isotopic detection multiple development was possible and in a short period of time.

3.9.5 Inter and intra-gel variation

There was 6% inter-gel variation observed. Inter-laboratory comparison of the VNTR result is not practical because of high variation. The match criteria based on standard deviations. Two bands match if they are within 6% of each other if they run onto different gels and 4.8% if they run on a single gel. Allele fragments below 800 bp may not be detectable with this electrophoretic system.

For sizing of DNA fragments, scanners have replaced the video cameras. A match between two samples can be first determined visually and then a fixed match criterion can be used. Half of the European countries use a fixed match criterion (Hallenberg C. and Morling N., 2001).

The data was first binned according to their size ranges then rebinned so that each bin (allele) has a frequency of ~1%. Three different size ranges were conducted (method # 1, method # 2 and method # 3). The three methods are seen in appendix D. In method # 3 it has been rebinned in certain way that the size ranges are the same for all three populations. This helped to compare the allele frequency between the three populations and helped in statistical analysis purposes. The fixed bin method was used because of its ease, conservative and usefulness in statistical analysis (Monson K. and Budowle B., 1993; NRC, 1996).

CHAPTER 4: VNTR STATISTICAL ANALYSIS

4.1 INTRODUCTION

VNTR polymorphisms have been studied in three populations residing in the UAE. In total 439 samples were typed and analysed at five VNTR loci (D1S7, D2S44, D7S21, D7S22, and D12S11).

Population data and genotypes from the three populations were collected because there was a need to analyse this data to test for population substructure to assess the usefulness of these five loci in forensic and paternity testing. Therefore, the allele and genotype frequencies, their variances and standard deviations were calculated.

A standard Chi square goodness of fit, the exact test, heterozygosity tests and variance tests were used to test the data for Hardy Weinberg equilibrium. The RxC contingency test was used to compare allele distributions in different populations in order to test for the level of homogeneity between populations. The F_{ST} (co-ancestry coefficient) was performed to measure the genetic differentiation between the three populations. The F_{ST} value can be used to calculate the allele and genotype frequencies in conservative way.

4.2 ALLELE FREQUENCIES

Before using any marker in the forensic and paternity analysis it is important to calculate the population frequencies of that marker. The allele frequencies for the three populations first were calculated from the genotypes observed at five VNTR loci. The

allele frequencies were calculated by dividing the number of specific allele observed over the total number of alleles (twice the number of individuals).

Three different versions of fixed bin were used to calculate the allele frequencies, method # 1, method # 2 and method # 3, (appendix D). Method # 3 was used in the statistical analysis because the size ranges of alleles were the same for the three populations. This method was used in order to allow the comparison to be made between different populations.

Differences in allele frequencies were calculated for the three populations at the five loci. The standard deviation (SD) was calculated according to the equation given by Evett and Weir (equation 4.1) in order to set the confidence interval of the allele within 95% probability (Evett and Weir, 1998).

$$SD = \pm 1.96\sqrt{p(1-p)/2n} \qquad (4.1)$$

Where (p) is the frequency of the allele and (n) is the number of individual.

Among the five loci tested D1S7 locus had the largest allele number (22 alleles). In the UAE population allele 12 was the most common allele with a frequency of 0.098 and allele 1 had the lowest frequency of 0.009. In the Indian and Pakistani populations allele 14 was the most common allele with a frequency of 0.107 and 0.134 respectively. Allele 3 and 4 had the lowest frequency of 0.013 in the Indian population and allele 4 and 22 had the lowest frequency of 0.009 in the Pakistani population. The D1S7 is the most informative locus in the three populations for forensic purposes but not a useful marker for

paternity analysis because of its high mutation rate reported (5.2%) (Jeffreys A. *et al.*, 1988).

D7S21 locus had the lowest allele number (10 alleles). Allele 7 is being the most common allele in the three populations with a frequency of 0.266, 0.26 and 0.201 in the UAE, Indian and Pakistani populations respectively. Allele 2 had the lowest frequency of 0.026 in the UAE population and allele 1 had the lowest frequency of 0.01, 0.018 in the Indian and Pakistani populations respectively.

D7S22 locus had 16 different alleles. Allele 12 is being the most common allele with a frequency of 0.168 in the UAE population. Allele 2 (~1.7 kb) which is usually the most common allele observed in Caucasian population reported (Probe catalogue, Cellmark; Henke L. *et al.*, 1991), was also the most common in the Indian and Pakistani populations with a frequency of 0.219. However, It was not the most common allele in the UAE population (0.084). Allele 1 had the lowest frequency of 0.006, 0.006 and 0.013 in the UAE, Indian and Pakistani populations respectively.

D2S44 locus had 12 different alleles. Allele 9 is being the most common allele with a frequency of 0.211 in the UAE population but allele 5 with a frequency of 0.153, 0.192 is the most common allele in the Indian and Pakistani populations respectively. Allele 1 had the lowest frequency of 0.026, 0.006 in the UAE and Pakistani populations respectively. Allele 1, 2 and 12 in the Indian population had the lowest frequency of 0.013. The fairly even distribution of alleles makes this locus particularly informative for forensic and paternity analysis purposes.

D12S11 locus had 11 different alleles. The most two common alleles in the three populations are allele 8 and 9 which have a combined frequency of 0.54, 0.47 and 0.53 in the UAE, Indian and Pakistani populations respectively. Allele 11 had the lowest frequency of 0.012 in the UAE population and allele 1 had the lowest frequency of 0.006 and 0.013 in the Indian and Pakistani populations respectively. The allele frequencies and their standard deviations (SD) for the three populations are given in table 4.1a-e, 4.2a-e and 4.3a-e. See figure 4.1 for allele frequency distributions.

Table 4.1a Allele frequencies in 173 UAE individuals for the locus D1S7. The maximum and minimum allele frequencies are shown in bold italic

Allele (bin)	Size range (D1S7)	Observed allele	Allele frequency	SD	Heter.	Heter. frequency
1	0-1431	***3***	***0.0087***	0.0098	3	0.0173
2	1432-1672	6	0.0173	0.0137	6	0.0347
3	1673-1861	5	0.0145	0.0125	5	0.0289
4	1862-2015	7	0.0202	0.0149	7	0.0405
5	2016-2213	11	0.0318	0.0184	9	0.0520
6	2214-2433	16	0.0462	0.0221	16	0.0925
7	2434-2650	14	0.0405	0.0208	12	0.0694
8	2651-2876	13	0.0376	0.02	13	0.0751
9	2877-3101	13	0.0376	0.02	13	0.0751
10	3102-3397	13	0.0376	0.02	13	0.0751
11	3398-3812	23	0.0665	0.0263	23	0.1329
12	3813-4333	***34***	***0.0983***	0.0314	34	0.1965
13	4334-4716	24	0.0694	0.0269	24	0.1387
14	4717-5415	28	0.0809	0.0288	28	0.1618
15	5416-5861	27	0.0780	0.0282	21	0.1214
16	5862-6442	16	0.0462	0.0221	16	0.0925
17	6443-7421	25	0.0723	0.0272	19	0.1098
18	7422-8271	12	0.0347	0.0192	10	0.0578
19	8272-9416	14	0.0405	0.0208	10	0.0578
20	9417-11919	18	0.0520	0.0233	18	0.1040
21	11920-15004	12	0.0347	0.0192	10	0.0578
22	15005-	12	0.0347	0.0192	10	0.0578

Table 4.1b Allele frequencies in 173 UAE individuals for the locus D7S22. The maximum and minimum allele frequencies are shown in bold italic.

Allele (bin)	Size range (D7S22)	Observed allele	Allele frequency	SD	Heter.	Heter. frequency
1	0-1568	***2***	***0.0058***	0.0080	2	0.0116
2	1569-1672	29	0.0838	0.0292	27	0.1561
3	1673-2650	12	0.0347	0.0192	12	0.0694
4	2651-2876	4	0.0116	0.0114	4	0.0231
5	2877-3101	7	0.0202	0.0149	5	0.0289
6	3102-3397	46	0.1329	0.0357	30	0.1734
7	3398-3812	8	0.0231	0.0159	6	0.0347
8	3813-4716	9	0.0260	0.0159	7	0.0405
9	4717-5415	23	0.0665	0.0263	19	0.1098
10	5416-5861	17	0.0491	0.0227	15	0.0867
11	5862-6442	25	0.0723	0.0272	19	0.1098
12	6443-7421	***58***	***0.1676***	0.0394	32	0.1850
13	7422-8271	44	0.1272	0.0351	32	0.1850
14	8272-9416	22	0.0636	0.0257	18	0.1040
15	9417-11919	32	0.0925	0.0306	24	0.1387
16	11920-	8	0.0231	0.0159	8	0.0462

Table 4.1c Allele frequencies in 173 UAE individuals for the locus D7S21. The maximum and minimum allele frequencies are shown in bold italic.

Allele (bin)	Size range (D7S21)	Observed allele	Allele frequency	SD	Heter.	Heter. frequency
1	0-3812	13	0.0376	0.0200	13	0.0751
2	3813-4333	***9***	***0.0260***	0.0169	7	0.0405
3	4334-4716	12	0.0347	0.0192	10	0.0578
4	4717-5415	25	0.0723	0.0272	23	0.1329
5	5416-5861	36	0.1040	0.0321	28	0.1618
6	5862-6442	53	0.1532	0.0380	49	0.2832
7	6443-7421	***92***	***0.2659***	0.0466	70	0.4046
8	7422-8271	36	0.1040	0.0321	26	0.1503
9	8272-9416	50	0.1445	0.037	46	0.2659
10	9417-	20	0.0578	0.0245	16	0.0925

Table 4.1d Allele frequencies in 172 UAE individuals for the locus D12S11. The maximum and minimum allele frequencies are shown in bold italic.

Allele (bin)	Size range (D12S11)	Observed allele	Allele frequency	SD	Heter.	Heter. frequency
1	0-3812	10	0.0291	0.0178	8	0.0465
2	3813-4716	14	0.0407	0.0210	14	0.0814
3	4717-5415	35	0.1017	0.0319	29	0.1686
4	5416-5861	6	0.0174	0.0137	6	0.0349
5	5862-6442	5	0.0145	0.0125	5	0.0291
6	6443-7421	25	0.0727	0.0274	23	0.1337
7	7422-8271	51	0.1483	0.0376	41	0.2384
8	8272-9416	***98***	***0.2449***	0.0455	76	0.4419
9	9417-11919	89	0.2587	0.0463	57	0.3314
10	11920-15004	7	0.0203	0.0149	7	0.0407
11	15005-	***4***	***0.0116***	0.0114	4	0.0233

Table 4.1e Allele frequencies in 173 UAE individuals for the locus D2S44. The maximum and minimum allele frequencies are shown in bold italic.

Allele (bin)	Size range (D2S44)	Observed allele	Allele frequency	SD	Heter.	Heter. frequency
1	0-2015	***9***	***0.0260***	0.0169	9	0.0520
2	2016-2213	10	0.0289	0.0176	10	0.0578
3	2214-2433	23	0.0665	0.0263	17	0.0983
4	2434-2650	31	0.0896	0.0302	27	0.1561
5	2651-2876	45	0.1301	0.0355	43	0.2486
6	2877-3101	30	0.0867	0.0296	24	0.1387
7	3102-3397	22	0.0636	0.0257	20	0.1156
8	3398-3812	31	0.0896	0.0302	27	0.1561
9	3813-4333	***73***	***0.2110***	0.0429	57	0.3295
10	4334-4716	28	0.0809	0.0288	22	0.1272
11	4717-5415	32	0.0925	0.0306	28	0.1618
12	5416-	12	0.0347	0.0192	12	0.0694

Table 4.2a Allele frequencies in 154 Indian individuals for the locus D1S7. The maximum and minimum allele frequencies are shown in bold italic.

Allele (bin)	Size range (D1S7)	Observed allele	Allele frequency	SD	Heter.	Heter. frequency
1	0-1431	5	0.0162	0.0141	5	0.0325
2	1432-1672	6	0.0195	0.0155	6	0.0390
3	1673-1861	***4***	***0.0130***	0.0127	4	0.0260
4	1862-2015	***4***	***0.0130***	0.0127	4	0.0260
5	2016-2213	6	0.0195	0.0155	6	0.0390
6	2214-2433	12	0.0390	0.0216	12	0.0779
7	2434-2650	7	0.0227	0.0167	7	0.0455
8	2651-2876	10	0.0325	0.0198	10	0.0649
9	2877-3101	7	0.0227	0.0167	7	0.0455
10	3102-3397	11	0.0357	0.0208	11	0.0714
11	3398-3812	17	0.0552	0.0176	15	0.0974
12	3813-4333	26	0.0844	0.031	22	0.1429
13	4334-4716	24	0.0779	0.0299	22	0.1429
14	4717-5415	***33***	***0.1071***	0.0137	27	0.1753
15	5416-5861	23	0.0747	0.0292	23	0.1494
16	5862-6442	27	0.0877	0.0316	25	0.1623
17	6443-7421	22	0.0714	0.0288	18	0.1169
18	7422-8271	20	0.0649	0.0274	14	0.0909
19	8272-9416	10	0.0325	0.0198	10	0.0649
20	9417-11919	22	0.0714	0.0288	22	0.1429
21	11920-15004	6	0.0195	0.0155	6	0.0390
22	15005-	6	0.0195	0.0155	6	0.0390

Table 4.2b Allele frequencies in 154 Indian individuals for the locus D7S22. The maximum and minimum allele frequencies are shown in bold italic.

Allele (bin)	Size range (D7S22)	Observed allele	Allele frequency	SD	Heter.	Heter. frequency
1	0-1568	***2***	***0.0065***	0.009	2	0.0130
2	1569-1672	47	0.1526	0.0402	37	0.2403
3	1673-2650	3	0.0097	0.011	3	0.0195
4	2651-2876	4	0.0130	0.0127	4	0.0260
5	2877-3101	11	0.0357	0.0208	11	0.0714
6	3102-3397	21	0.0682	0.0282	17	0.1104
7	3398-3812	9	0.0292	0.0188	7	0.0455
8	3813-4716	18	0.0584	0.0263	16	0.1039
9	4717-5415	13	0.0422	0.0225	13	0.0844
10	5416-5861	30	0.0974	0.0331	26	0.1688
11	5862-6442	33	0.1071	0.0345	31	0.2013
12	6443-7421	***52***	***0.1688***	0.0417	44	0.2857
13	7422-8271	23	0.0747	0.0294	23	0.1494
14	8272-9416	16	0.0519	0.0247	14	0.0909
15	9417-11919	22	0.0714	0.0288	20	0.1299
16	11920-	4	0.0130	0.0127	4	0.0260

Table 4.2c Allele frequencies in 154 Indian individuals for the locus D7S21. The maximum and minimum allele frequencies are shown in bold italic.

Allele (bin)	Size range (D7S21)	Observed allele	Allele frequency	SD	Heter.	Heter. frequency
1	0-3812	***3***	***0.0097***	0.0110	3	0.0195
2	3813-4333	9	0.0292	0.0188	7	0.0455
3	4334-4716	6	0.0195	0.0155	6	0.0390
4	4717-5415	24	0.0779	0.0300	22	0.1429
5	5416-5861	29	0.0942	0.0325	27	0.1753
6	5862-6442	48	0.1558	0.0406	42	0.2727
7	6443-7421	***80***	***0.2597***	0.0490	68	0.4416
8	7422-8271	41	0.1331	0.0380	35	0.2273
9	8272-9416	57	0.1851	0.0433	51	0.3312
10	9417-	11	0.0357	0.0208	11	0.0714

Table 4.2d Allele frequencies in 154 Indian individuals for the locus D12S11. The maximum and minimum allele frequencies are shown in bold italic.

Allele (bin)	Size range (D12S11)	Observed allele	Allele frequency	SD	Heter.	Heter. frequency
1	0-3812	***2***	***0.0065***	0.0090	2	0.0130
2	3813-4716	12	0.0390	0.0216	12	0.0779
3	4717-5415	52	0.1688	0.0417	40	0.2597
4	5416-5861	5	0.0162	0.0141	5	0.0325
5	5862-6442	4	0.0130	0.0127	4	0.0260
6	6443-7421	14	0.0455	0.0233	12	0.0779
7	7422-8271	20	0.0649	0.0274	20	0.1299
8	8272-9416	65	0.2110	0.0455	55	0.3571
9	9417-11919	***80***	***0.2597***	0.0490	56	0.3636
10	11920-15004	44	0.1429	0.0390	38	0.2468
11	15005-	10	0.0325	0.0198	8	0.0519

Table 4.2c Allele frequencies in 154 Indian individuals for the locus D2S44. The maximum and minimum allele frequencies are shown in bold italic.

Allele (bin)	Size range (D2S44)	Observed allele	Allele frequency	SD	Heter.	Heter. frequency
1	0-2015	***2***	***0.0065***	0.0090	2	0.0130
2	2016-2213	6	0.0195	0.0155	6	0.0390
3	2214-2433	28	0.0909	0.0321	26	0.1688
4	2434-2650	44	0.1429	0.0390	36	0.2338
5	2651-2876	***47***	***0.1526***	0.0402	37	0.2403
6	2877-3101	45	0.1461	0.0394	35	0.2273
7	3102-3397	31	0.1006	0.0335	27	0.1753
8	3398-3812	36	0.1169	0.0359	36	0.2338
9	3813-4333	40	0.1299	0.0376	36	0.2338
10	4334-4716	15	0.0487	0.0241	15	0.0974
11	4717-5415	10	0.0325	0.0198	10	0.0649
12	5416-	4	0.0130	0.0127	4	0.0260

Table 4.3a Allele frequencies in 112 Pakistani individuals for the locus D1S7. The maximum and minimum allele frequencies are shown in bold italic.

Allele (bin)	Size range (D1S7)	Observed allele	Allele frequency	SD	Heter.	Heter. frequency
1	0-1431	8	0.0357	0.0243	8	0.0714
2	1432-1672	3	0.0134	0.0151	3	0.0268
3	1673-1861	6	0.0268	0.0212	6	0.0536
4	1862-2015	***2***	***0.0089***	0.0123	2	0.0179
5	2016-2213	7	0.0313	0.0227	7	0.0625
6	2214-2433	8	0.0357	0.0243	8	0.0714
7	2434-2650	11	0.0491	0.0282	11	0.0982
8	2651-2876	13	0.0580	0.0306	13	0.1161
9	2877-3101	6	0.0268	0.0212	6	0.0536
10	3102-3397	8	0.0357	0.0243	6	0.0536
11	3398-3812	13	0.0580	0.0306	13	0.1161
12	3813-4333	19	0.0848	0.0365	19	0.1696
13	4334-4716	9	0.0402	0.0257	9	0.0804
14	4717-5415	***30***	***0.1339***	0.0447	26	0.2321
15	5416-5861	12	0.0536	0.0294	10	0.0893
16	5862-6442	14	0.0625	0.0318	14	0.1250
17	6443-7421	12	0.0536	0.0294	12	0.1071
18	7422-8271	15	0.0670	0.0327	13	0.1161
19	8272-9416	8	0.0357	0.0243	8	0.0714
20	9417-11919	12	0.0536	0.0294	12	0.1071
21	11920-15004	6	0.0268	0.0212	6	0.0536
22	15005-	***2***	***0.0089***	0.0123	2	0.0179

Table 4.3b Allele frequencies in 112 Pakistani individuals for the locus D7S22. The maximum and minimum allele frequencies are shown in bold italic.

Allele (bin)	Size range (D7S22)	Observed allele	Allele frequency	SD	Heter.	Heter. frequency
1	0-1568	3	0.0134	0.0151	3	0.0268
2	1569-1672	***49***	***0.2188***	0.0541	39	0.3482
3	1673-2650	4	0.0179	0.0174	4	0.0357
4	2651-2876	4	0.0179	0.0174	4	0.0357
5	2877-3101	***2***	***0.0089***	0.0123	2	0.0179
6	3102-3397	23	0.1027	0.0398	21	0.1875
7	3398-3812	6	0.0268	0.0212	6	0.0536
8	3813-4716	7	0.0313	0.0227	7	0.0625
9	4717-5415	9	0.0402	0.0257	9	0.0804
10	5416-5861	16	0.0714	0.0337	16	0.1429
11	5862-6442	14	0.0625	0.0318	14	0.1250
12	6443-7421	33	0.1473	0.0465	27	0.2411
13	7422-8271	14	0.0625	0.0318	14	0.1250
14	8272-9416	20	0.0893	0.0374	18	0.1607
15	9417-11919	12	0.0536	0.0294	12	0.1071
16	11920-	8	0.0357	0.0243	8	0.0714

Table 4.3c Allele frequencies in 112 Pakistani individuals for the locus D7S21. The maximum and minimum allele frequencies are shown in bold italic.

Allele (bin)	Size range (D7S21)	Observed allele	Allele frequency	SD	Heter.	Heter. frequency
1	0-3812	***4***	***0.0179***	0.0174	4	0.0357
2	3813-4333	6	0.0268	0.0212	6	0.0536
3	4334-4716	5	0.0223	0.0194	5	0.0446
4	4717-5415	22	0.0982	0.0390	18	0.1607
5	5416-5861	31	0.1384	0.0453	29	0.2589
6	5862-6442	42	0.1875	0.0512	34	0.3036
7	6443-7421	***45***	***0.2009***	0.0525	35	0.3125
8	7422-8271	33	0.1473	0.0465	27	0.2411
9	8272-9416	22	0.0982	0.039	20	0.1786
10	9417-	14	0.0625	0.0318	14	0.1250

Table 4.3d Allele frequencies in 112 Pakistani individuals for the locus D12S11. The maximum and minimum allele frequencies are shown in bold italic.

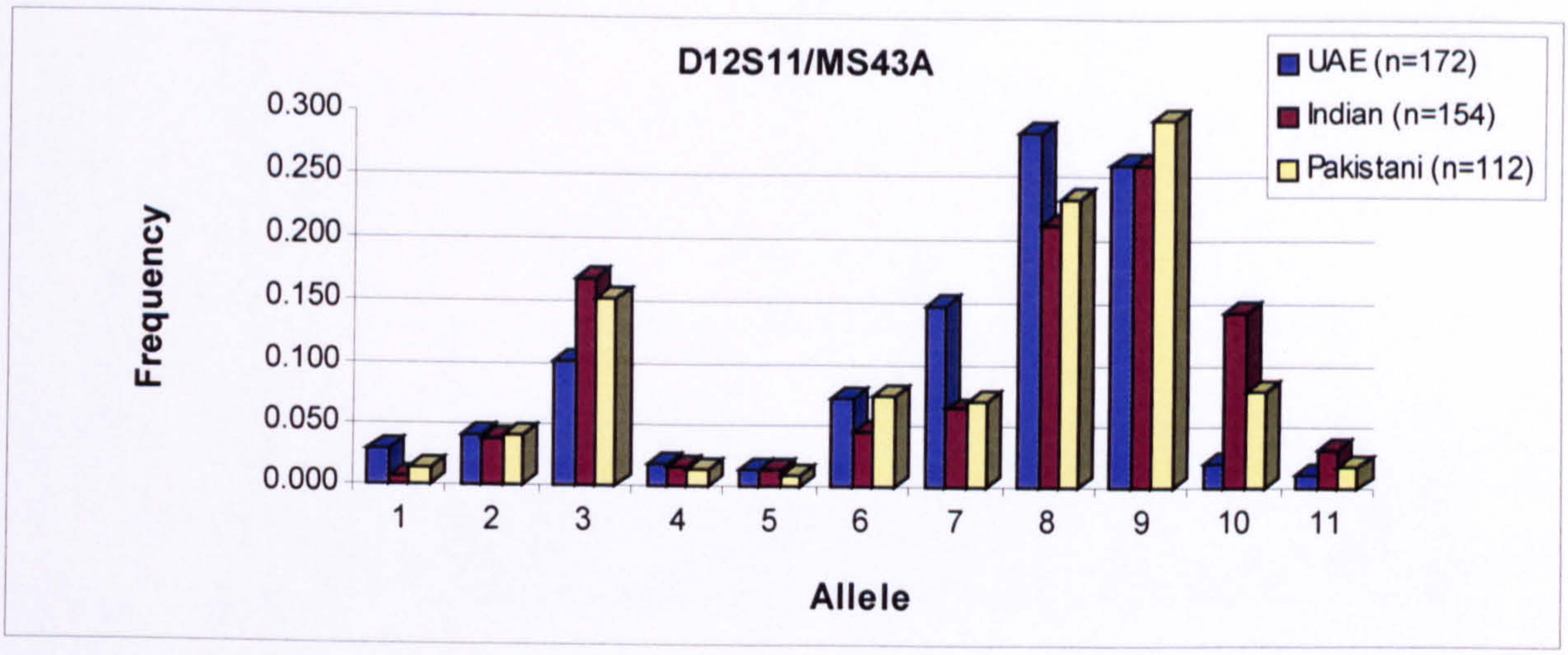
Allele (bin)	Size range (D12S11)	Observed allele	Allele frequency	SD	Heter.	Heter. frequency
1	0-3812	3	0.0134	0.0151	1	0.0089
2	3813-4716	9	0.0402	0.0257	9	0.0804
3	4717-5415	34	0.1518	0.047	20	0.1786
4	5416-5861	3	0.0134	0.0151	3	0.0268
5	5862-6442	***2***	***0.0089***	0.0123	2	0.0179
6	6443-7421	17	0.0759	0.0347	17	0.1518
7	7422-8271	16	0.0714	0.0337	14	0.1250
8	8272-9416	52	0.2321	0.0553	36	0.3214
9	9417-11919	***66***	***0.2946***	0.0598	46	0.4107
10	11920-15004	18	0.0804	0.0357	16	0.1429
11	15005-	4	0.0179	0.0174	4	0.0357

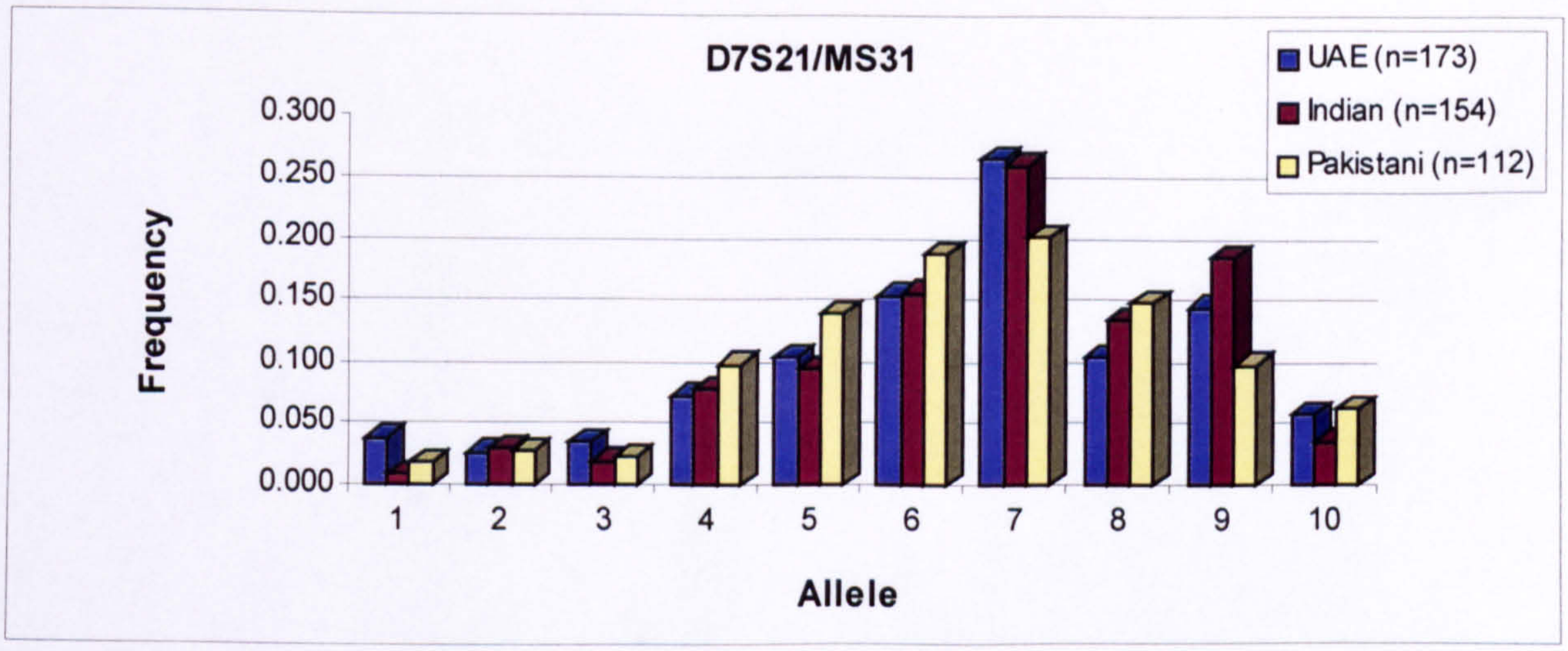
Table 4.3e Allele frequencies in 112 Pakistani individuals for the locus D2S44. The maximum and minimum allele frequencies are shown in bold italic.

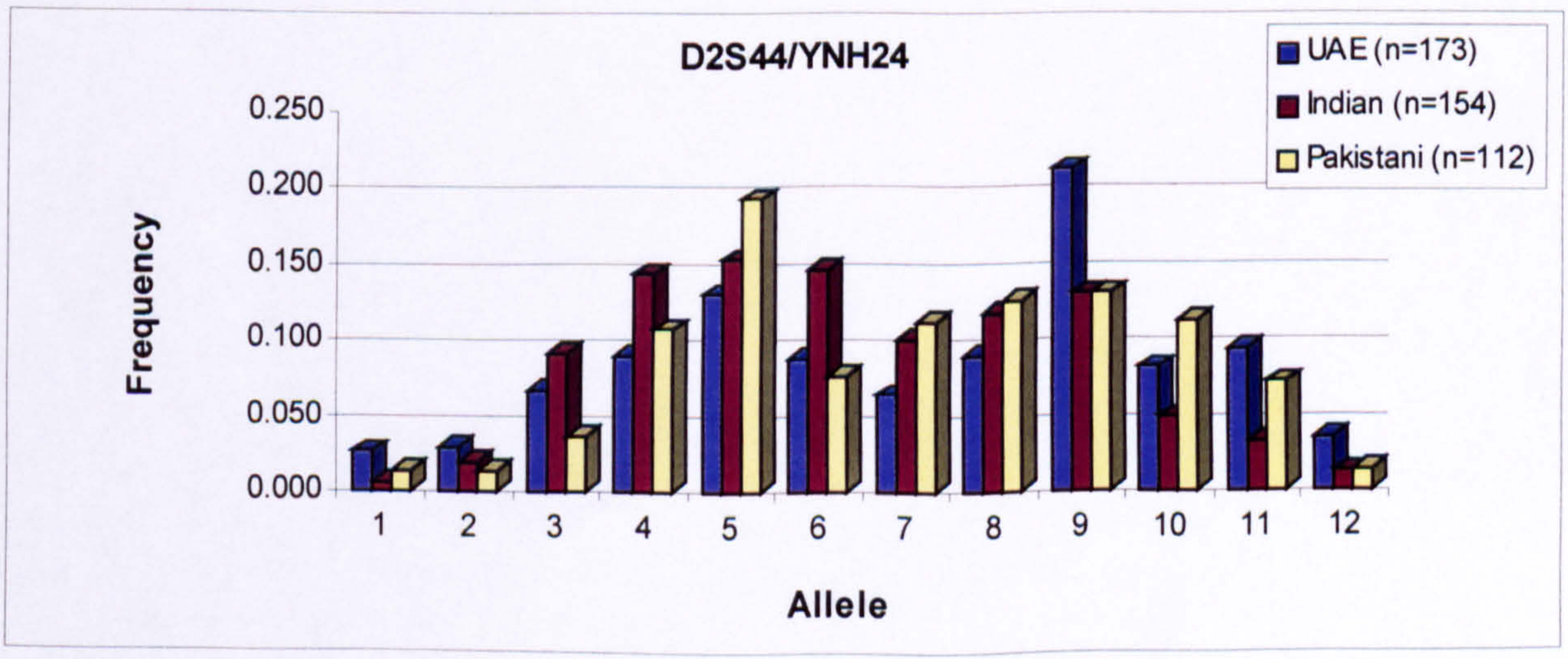
Allele (bin)	Size range (D2S44)	Observed allele	Allele frequency	SD	Heter.	Heter. frequency
1	0-2015	***3***	***0.0134***	0.0151	3	0.0268
2	2016-2213	***3***	***0.0134***	0.0151	3	0.0268
3	2214-2433	8	0.0357	0.0243	4	0.0357
4	2434-2650	24	0.1071	0.0406	18	0.1607
5	2651-2876	***43***	***0.1920***	0.0515	31	0.2768
6	2877-3101	17	0.0759	0.0347	15	0.1339
7	3102-3397	25	0.1116	0.0412	25	0.2232
8	3398-3812	28	0.1250	0.0433	22	0.1964
9	3813-4333	29	0.1295	0.0439	23	0.2054
10	4334-4716	25	0.1116	0.0412	23	0.2054
11	4717-5415	16	0.0714	0.0337	16	0.1429
12	5416-	***3***	***0.0134***	0.0151	3	0.0268

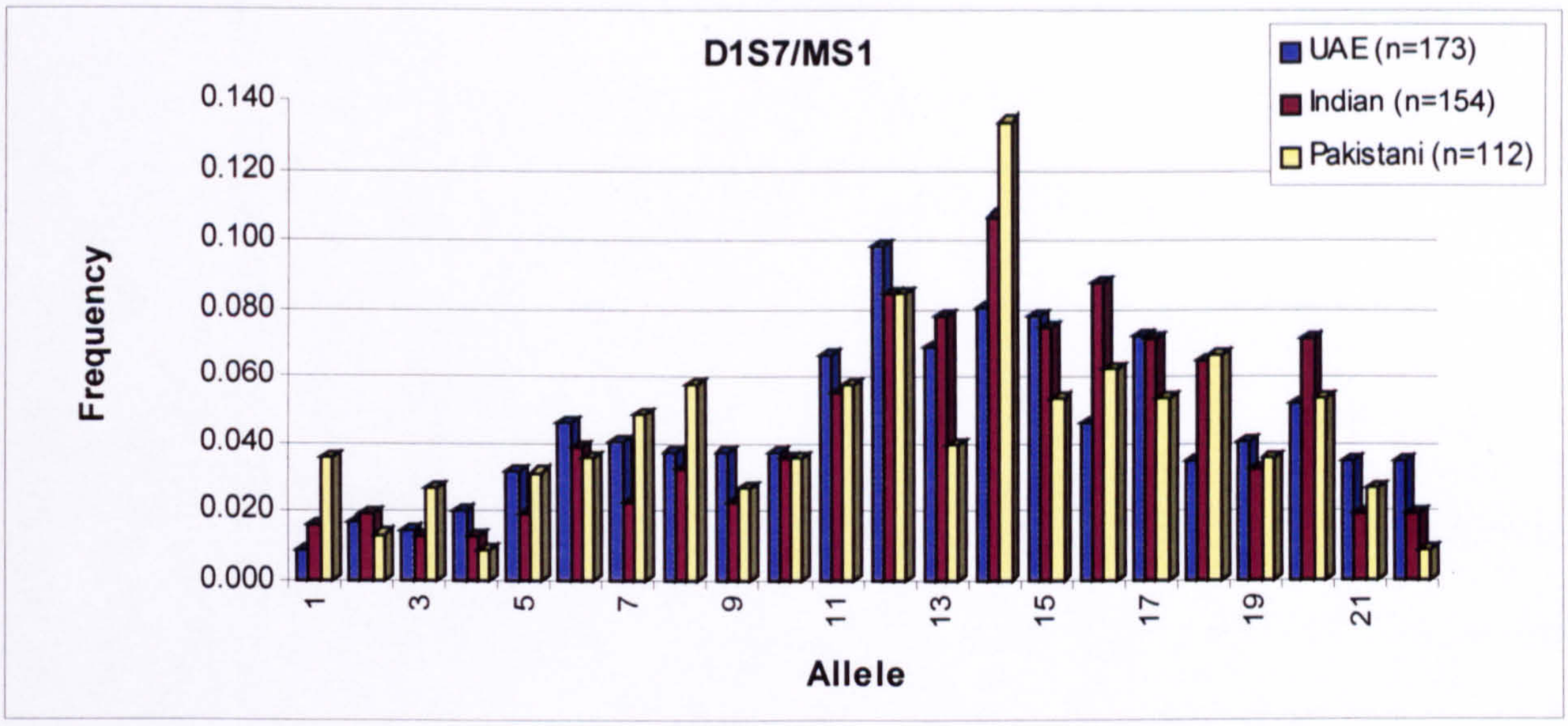
116 A

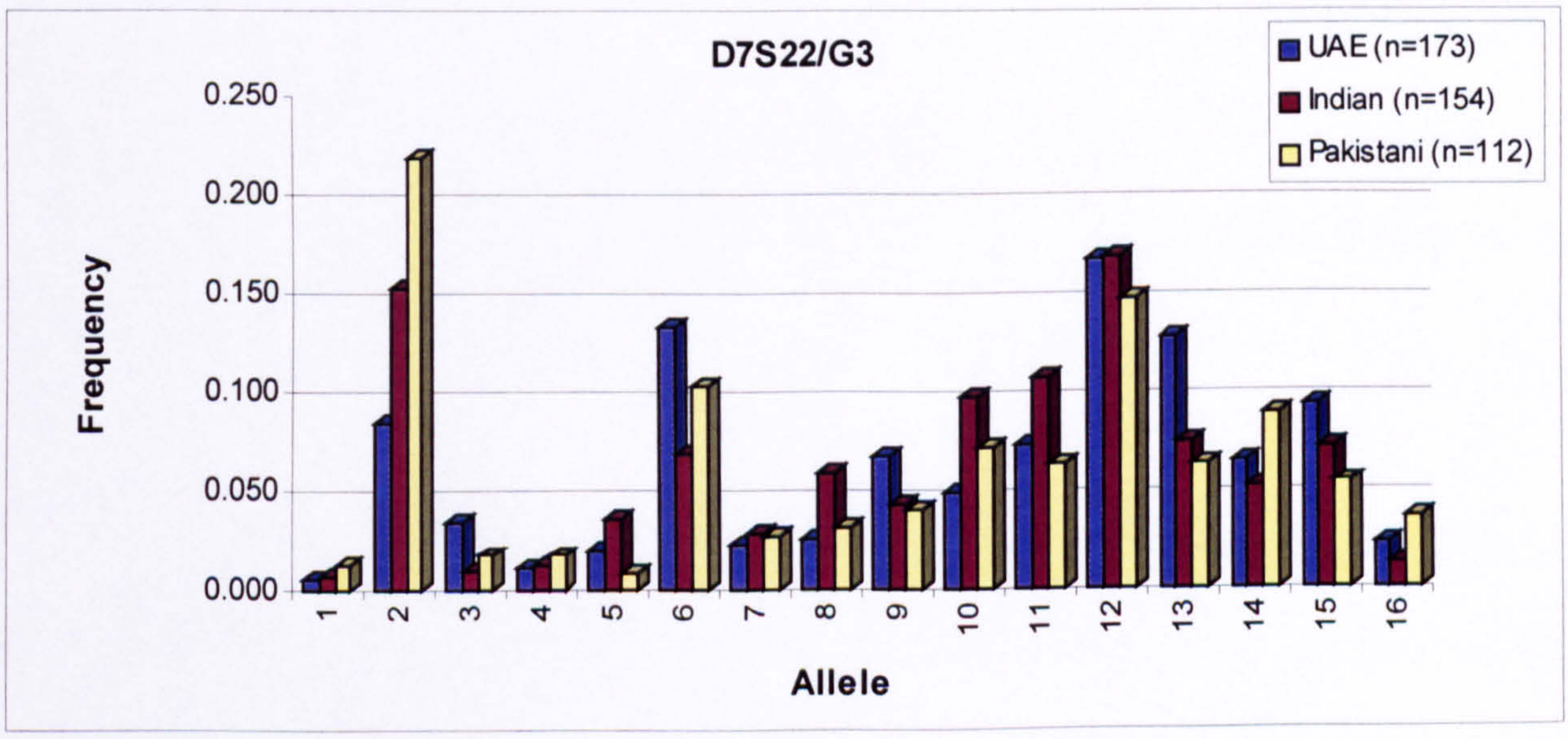
Figure 4.1 Allele frequency distributions for the five VNTR loci in the three populations.


D12S11/MS43A
UAE (n=172)
Indian (n=154)
Pakistani (n=112)
Frequency
0.300
0.250
0.200
0.150
0.100
0.050
0.000
1
2
3
4
5
6
7
8
9
10
11
Allele


D7S21/MS31
UAE (n=173)
Indian (n=154)
Pakistani (n=112)
Frequency
0.300
0.250
0.200
0.150
0.100
0.050
0.000
1
2
3
4
5
6
7
8
9
10
Allele


D2S44/YNH24
UAE (n=173)
Indian (n=154)
Pakistani (n=112)
Frequency
0.250
0.200
0.150
0.100
0.050
0.000
1
2
3
4
5
6
7
8
9
10
11
12
Allele


D1S7/MS1
UAE (n=173)
Indian (n=154)
Pakistani (n=112)
Frequency
0.140
0.120
0.100
0.080
0.060
0.040
0.020
0.000
1
3
5
7
9
11
13
15
17
19
21
Allele


D7S22/G3
UAE (n=173)
Indian (n=154)
Pakistani (n=112)
Frequency
0.250
0.200
0.150
0.100
0.050
0.000
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Allele

4.3 GENOTYPE FREQUENCIES

Each individual's alleles make up their genotype. A genotype frequency is the proportion of individuals in a population that possess a given genotype at s specific locus. The observed genotype frequencies were calculated by dividing a number of specific genotype observed over the total number of individuals. The expected genotype frequencies were calculated according to Hardy-Weinberg equation. In the case of homozygous, the genotype frequency is equal to p^2 or q^2 and in the case of heterozygous, the genotype frequency is equal to 2pq, where p and q are the frequency of allele p and allele q.

The possible number of genotypes can be calculated for any marker from the total number of alleles observed at that marker. The number of genotypes = n (n + 1)/2, where n is the total number of alleles detected from a population. Theoretically, if there are n alleles, there are n homozygous genotypes and n (n – 1)/2 heterozygous ones.

The standard deviation (SD) for the observed genotype frequencies was calculated according to the equation given by Evett and Weir (equation 4.2) in order to set the confidence interval of the genotype within 95% probability (Evett and Weir, 1998).

$$SD = \pm 1.96\sqrt{p(1-p)/n} \qquad (4.2)$$

Where (p) is the frequency of the genotype and (n) is the number of individual. The standard deviation (SD) for the observed genotypes is shown in appendix E.

The total number of possible combination of genotypes (theoretically) available for the five loci analysed in any of the three populations is 9.7 billion (253 x 55 x 136 x 78 x 66 = 9,742,281,120).

There were 39, 37 and 29 observed genotypes out of 66 different possible genotypes at D12S11 locus in the UAE, Indian and Pakistani populations respectively. However, the total number of 51 genotypes was observed in the three populations together. The genotype 8/9 had the highest frequency in the three populations. Occurring in 29 (16.9%) out of 172 individuals in the UAE population, 18 (11.7%) out of 154 individuals in the Indian population and 14 (12.5%) out of 112 individuals in the Pakistani population.

D7S21 locus had 55 possible genotypes, where 48 genotypes observed in the three populations. There were 43, 37 and 35 observed genotypes in the UAE, Indian and Pakistani populations respectively. The genotype 6/7 and 7/9 had the highest frequency in the UAE population, occurring in 16 (9.2%) out of 173 individuals. The genotype 7/9 had the highest frequency in the Indian population and occurring in 19 (12.3%) out of 154 individuals. In the Pakistani population the genotype 7/8 had the highest frequency, occurring in 9 (8%) out of 112 individuals. The homozygote frequency for the allele 9 was the highest in the three populations.

Out of 78 possible genotypes, only 60, 48 and 44 genotypes were observed in the UAE, Indian and Pakistani populations respectively for D2S44 locus. The most common genotypes were 9/11 (6.4%), occurring in 11 out of 173 individuals, 5/8 (6.5%) occurring in 10 out of 154 individuals and 5/7 (8%) occurring in 9 out of 112 individuals in the UAE, Indian and Pakistani populations respectively. Genotype 9/11 which is the most common in the UAE population is not observed in the Indian population.

There were 102 genotypes observed in the three populations out of 136 possible genotypes at D7S22 locus. Only 72, 72 and 56 genotypes were observed in the UAE, Indian and Pakistani populations respectively. The most common genotypes were 12/12 (7.5%), occurring in 13 out of 173 individuals in the UAE population. In the Indian and Pakistani populations the genotype 2/12 were the most common, occurring in 11 (7.1%) out of 154 individuals and 8 (7.1%) out of 112 individuals respectively.

There were 173 genotypes observed in the three populations out of 253 possible genotypes at D1S7 locus. Only 111, 102 and 83 genotypes were observed in the UAE, Indian and Pakistani populations respectively. The genotypes 6/12, 11/12 and 12/13 were the most common and occurring in 5 (2.9%) out of 173 individuals in the UAE population. The most common genotypes were 14/20 and 12/14 & 12/15 occurring in 5 (3.2%) out of 154 individuals and occurring in 4 (3.6%) out of 112 individuals in the Indian and Pakistani populations respectively. The homozygote frequency for all alleles was very low.

4.4 POPULATION SUBSTRUCTURE AND INBREEDING

No population obeys the H-W principal however most are close enough that deviations are not significant. The degree of substructure depends on the size of a population and how people mate (level of inbreeding). Genetic differentiation means differences in allele and genotype frequencies among the sub-populations. Allele frequencies may differ in different populations because of several reasons, such as natural selection, random processes in the transmission of alleles (genetic drift) or chance differences in allele frequency among the first founders of the sub-populations.

If the mating process is not random the degree of inbreeding will start to affect HWE. For example, in an inbred population the average proportion of heterozygous genotypes will be reduced relative to that expected under random mating. Therefore, heterozygote test and exact test may assist to detect an inbred population.

4.5 COMPARATIVE STUDIES

Two tests have been used to detect for population differentiation, RxC and F-statistics. Comparisons of the data results obtained in this study from the three populations (UAE, Indian and Pakistani) were performed for five VNTR loci.

4.5.1 RxC test

The RxC contingency table was used to calculate the p value of the exact test to compare the allele frequencies of the three populations at five VNTR loci. This test can be used to determine if significant differences in allele frequencies exist among groups of individuals. The comparison was performed for the five loci to examine the presence of heterogeneity between the three populations studied.

There were significant differences between the UAE and Indian populations at three loci D2S44, D7S22 and D12S11. The results showed that there were also significant differences between the UAE and Pakistani populations at these three loci. However, there were no significant differences observed between the Indian and Pakistani populations except at one locus D2S44 ($p = 0.0028$). This indicates that these two populations are very similar. Therefore, the database from both populations can be pooled together at these four loci (D1S7, D7S21, D7S22 and D12S11) in order to have only one larger database set for both populations. The results can be seen in table 4.4.

Table 4.4 Showing the p values and standard errors (SE) for the three populations at five VNTR loci. 20 batches of 2,500 replicates per batch performed. The total number of replicates was 50,000.

Locus	UAE/Ind.	SE	UAE/Pak	SE	Ind/Pak.	SE
D1S7	0.6021	0.0229	0.2249	0.0129	0.6746	0.0152
D2S44	***0.0000***	0.0000	***0.0194***	0.0049	***0.0028***	0.0016
D7S21	0.2467	0.0199	0.2189	0.0222	0.0728	0.0086
D7S22	***0.0016***	0.0014	***0.0005***	0.0004	0.0910	0.0109
D12S11	***0.0000***	0.0000	***0.0027***	0.0009	0.4403	0.0267

The locus D1S7 showed little differentiation among major ethnic groups in previous study (Buffery C. *et al.*, 1991). This may be because of a high mutation rate which generates a stationary distribution of allele frequencies. In this study there was no differences between the three populations at this locus which support the previous finding.

4.5.2 F-statistics

The *F*-statistics ($F_{ST} = \theta$) is considered as the most useful measure of population substructure (Hartl D. and Clark A., 1997). Therefore, in order to test the presence of substructure within the UAE populations, the genotypes of the three populations (UAE Arab, Indian and Pakistani) were compared using *F*-statistics (Lewis P. and Zaykin D., 2001).

The F-statistics analysis includes three parameters F_{ST}, F_{IS} and F_{IT}. Each of these accommodates a measure of relationship between a pair of alleles. The parameter F_{ST} (co-ancestry coefficient) measures the genetic variation and differentiation between two sub-populations being compared in terms of differences in their allelic characteristic. The F_{IS}

(inbreeding co-efficient) is the probability that the pair of alleles carried by a male and female gametes that produced it are identical by descent from a common ancestor (this parameter can be estimated with only one population). The F_{IT} indicates the relation between alleles within individuals relative to the total population.

The interpretation of F_{ST} values was suggested by Wright (Wright S., 1978) as the following, the range of 0 to 0.05 considered as indicating little genetic differentiation between tested populations. The range 0.05-0.15 indicates moderate genetic differentiation. The range 0.15-0.25 indicates great genetic differentiation and the values above 0.25 indicate very great genetic differentiation.

The overall value of F_{ST} over the five loci was (- 0.0033) when compared between the Indian and Pakistani populations. This indicates that there is no significant divergence or substructure (genetic differentiation) between these two populations. The result can be seen in table 4.5.

Table 4.5 F-statistics value of UAE/Pakistani, UAE/Indian and Indian/Pakistani populations.

F-statistics parameters for UAE/Pakistani populations	
F_{ST}	0.0054
F_{IT}	0.0518
F_{IS}	0.0466
F-statistics parameters for UAE/Indian populations	
F_{ST}	0.0062
F_{IT}	0.0409
F_{IS}	0.0349
F-statistics parameters for Indian /Pakistani populations	
F_{ST}	-0.0033
F_{IT}	0.0026
F_{IS}	0.0059

The bootstrapping over the five loci was applied in order to generate confidence intervals. The results showed that the inbreeding coefficients F_{IS} and F_{IT} were not statistically significantly different from zero. Because the 95% bootstrap confidence interval for F_{ST} value overlap zero when Indian populations compared to Pakistani population, therefore, the co-ancestry coefficient is not significant and thus a non significant amount of divergence had occurred between the two populations. See table 4.6.

Table 4.6 95% confidence interval bootstrapping over the five VNTR loci for the UAE/Pakistani, UAE/Indian and Indian/Pakistani populations.

UAE/Pakistani	Lower bound	Upper bound
F_{ST}	0.0026	0.0092
F_{IT}	0.0232	0.0841
F_{IS}	0.0206	0.0756
UAE/Indian		
F_{ST}	0.0011	0.0126
F_{IT}	0.0073	0.0781
F_{IS}	0.0051	0.0715
Indian /Pakistani		
F_{ST}	-0.0033	-0.0032
F_{IT}	-0.0255	0.0199
F_{IS}	-0.0223	0.0233

When the UAE Arab population was compared to both Indian and Pakistani populations the F_{ST} values was 0.0054 and 0.0062 respectively. This may indicates that little genetic differentiation exist between the UAE Arab and both Indian and Pakistani populations. For forensic identification purposes the maximum upper bound value of the two populations, UAE/Indian (0.0126), can be applied for statistical recalculation of genotype frequencies. Therefore, this value will compensate to some degree for existing of sub-populations.

The relation between the three F-statistics parameters can be given by the following equation (NRC, 1996; Hartl D. and Clark A., 1997).

$$F_{ST} = (F_{IT} - F_{IS})/(1 - F_{IT})$$

4.5.3 The effect of population substructure

The UAE consists of several different ethnic groups. This can have an effect on the estimation of the DNA profile in forensic analysis. For example, when suspect's DNA matches that obtained from a crime stain it is possible that the suspect left the stain or someone else from the population who has a similar DNA profile as the suspect. In this case the simple product rule routinely used to give an estimate of the profile match probability using a reference database. However, this database might not be representative of the relative offender population and may overstate the strength of the evidence. Therefore, a formula was proposed by Balding and Nichols (Balding D. and Nichols R., 1994) in order to compensate for the general database not being truly representative of the relevant suspect population. This method account for population subdivision by incorporating the genetic correlation (F_{ST} or θ) into the calculation.

The formula for homozygote is $$\frac{[2\theta+(1-\theta)p_i][3\theta+(1-\theta)p_i]}{(1+\theta)(1+2\theta)}$$

The formula for heterozygote is $$\frac{2[\theta+(1-\theta)p_i][\theta+(1-\theta)p_j]}{(1+\theta)(1+2\theta)}$$

Where $\theta = F_{ST}$ and p_i, p_j are the allele frequency.

It can be notice that applying the product rule is equivalent to adopting Balding and Nichols formula with $\theta = 0$ and that happen when no heterogeneity exists in the population. The FSS and most of other European laboratories no longer implement the product rule with $\theta = 0$ in forensic calculations. Instead, the above formula is used with $\theta = 0.01$ and higher (Foreman L. *et al.*, 1998; Foreman L. and Lambert J., 2000).

A randomly selected individual from the UAE database was selected and his genotype frequency at three loci D2S44, D1S7, D12S11 were recalculated by two different methods. In the first method the usual product rule was used and in the second method the Balding and Nichols formula was used. The F_{ST} (θ) value of upper bound 0.0126 was used.

Calculation of genotype frequency for an individual from the UAE database at three loci D2S44, D1S7, D12S11 and the genotypes were 8/9, 11/18 and 8/9 respecively. The use of $F_{ST} = 0.0126$ increased the probability of match and decreased the likelihood ratio by two fold. The result is shown in table 4.7.

Table 4.7 The effect of substructure on the value of likelihood ratio.

Locus	Usual product rule		Balding and Nichols formula	
	Frequency	Likelihood R.	Frequency	Likelihood R.
D2S44	0.038	26	0.042	24
D1S7	0.005	200	0.007	143
D12S11	0.127	8	0.134	7
Total	**0.00002**	**50,000**	**0.00004**	**25,000**

4.6 PHYLOGENTIC RELATIONSHIP BETWEEN POPULATIONS

The phylogenetic relationship between the UAE, Indian and Pakistani populations, based on co-ancestry distance measures for 5 VNTR loci, was assessed by drawing an UPGMA (unweighted pair group method using arithmetic averaging) phylogenetic tree

using GDA software. From the tree (figure 4.2) it can be concluded that the Indian and Pakistani populations have a closer relationship as compared to the UAE Arab population.

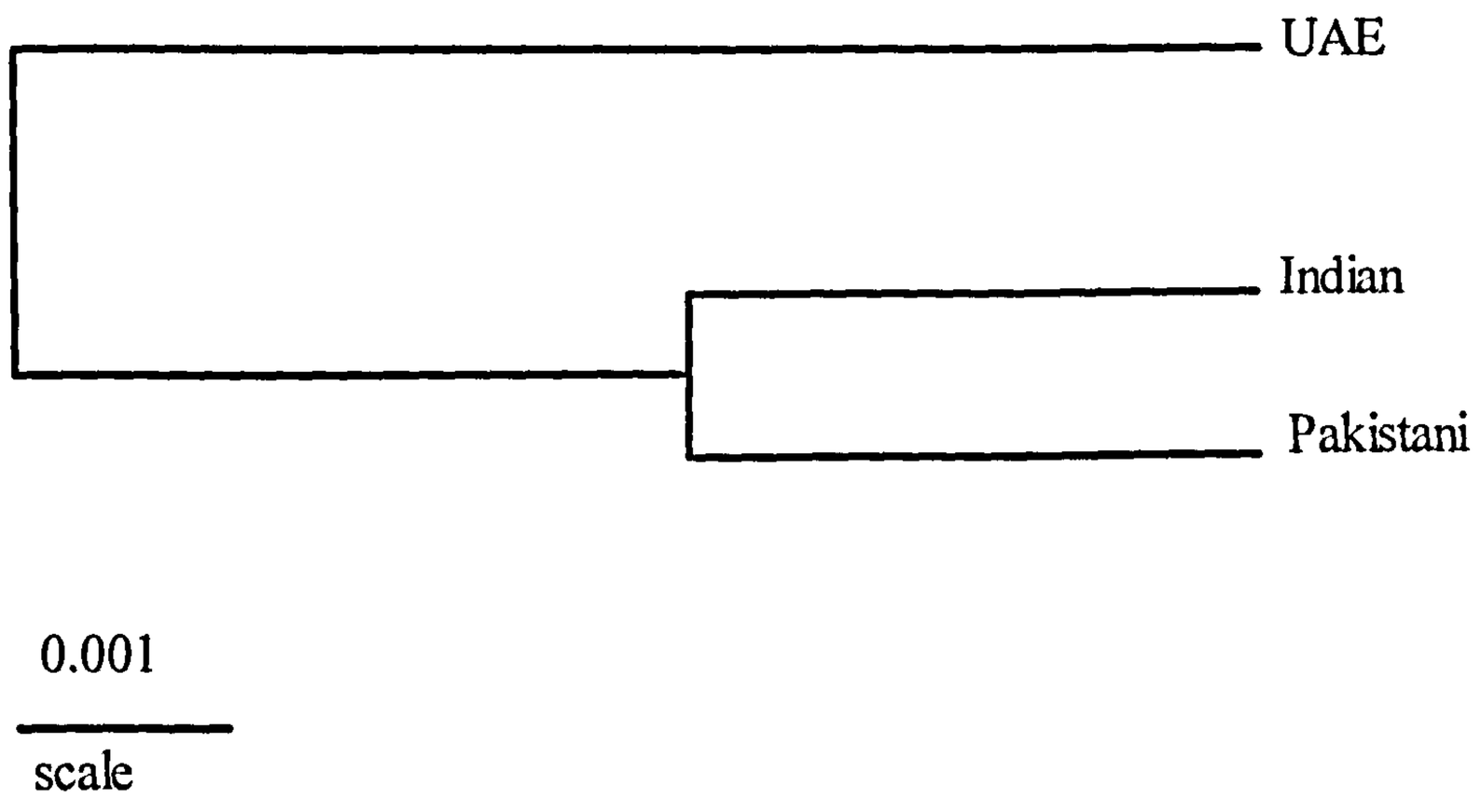
4.7 HARDY-WEINBERG EQUILIBRIUM

In Mendelian genetic analysis it is possible to calculate the expected proportions of different genotypes in the progeny of two parents whose genotypes are known. The same principles can be applied when the progeny come from a large number of inbreeding individuals and not from just one pair of parents.

The method used to calculate the expected proportions of different genotypes in a population was published in 1908 by the British mathematician (Hardy G.) and, independently, by German physician (Weinberg W.). It is now known as the Hardy-Weinberg equation (HWE). In the HWE the probabilities of the genotypes can be given by the equation ($p^2 + 2pq + q^2 = 1$), where one means the whole population. The p^2, 2pq and q^2 are genotype frequencies but p and q are allele frequencies. The p^2 and q^2 are frequency of homozygotes while 2pq is the frequency of heterozygote. The total frequency of all alleles in a population is equal to one ($p + q = 1$). A population to which the HWE is applicable consists of diploid, sexually reproducing individuals. A population should be under HWE and the allele frequencies should be passed to the next generation independently and randomly from a common gene pool. Before using any new markers in the forensic analysis it is desirable to test for Mendelian inheritance and HWE. Based on these observations the product (multiplication) rule can be applied and calculated. HWE depends mainly on the existence of a very large size and randomly mating populations. Although these characteristics are not possible in a natural population (Hammond H. *et al.*, 1994) usually populations do not deviate significantly from HWE. A significant deviation

129A

Figure 4.2 UPGMA phenogram showing the relationships between the three populations, the UAE Arab, Indian and Pakistani population.


UAE
Indian
Pakistani
0.001
scale

from HWE may exist in a very small isolated inbred population.

In VNTR system, a single band does not necessarily mean a homozygote and it might be a heterozygote. Therefore, it has become a standard practice to use 2p instead of p^2 to estimate a genotype frequency. This called a 2p rule, which has been recommended by the NRC in 1992 and 1996 (NRC, 1992; 1996).

4.7.1 Chi-square test

The Chi-square goodness-of-fit test (X^2) can be used to test whether the observed frequencies differ significantly from those which would be expected from the value proposed by the null hypothesis. The Chi-square test can be calculated by the following equation

$$X^2 = \sum \frac{(O-E)^2}{E} \qquad (4.5)$$

Where $\sum$ means "the sum of", O for "observed" and E for "expected".

The probability value (p) above 0.05 indicates that the null hypothesis is true. If p value is greater than 0.05 then the difference between observed and expected results is not significant. However, if the probability value is less than 0.05 indicates rejection of the null hypothesis and there is a significant difference and therefore indicates that the population is not in Hardy-Weinberg equilibrium.

4.7.2 Exact test

This test was introduced by Fisher in 1935 (Fisher R., 1935) which also used to indicate whether a population is in HWE or not. Guo introduced a new method that

allowed hyprervariable alleles to be tested (Guo S. and Thompson E., 1992). A p value of less than 0.05 indicates that the population is not in equilibrium.

Two tests were performed to test the departure of the observed genotype frequencies from those expected under HWE. The exact test for independence of the alleles within individuals was performed for each locus in the three populations using 2,000 permutations of alleles. There was agreement between the observed and expected genotype values under HWE for the five loci in the three populations. However, D1S7 and D7S22 loci in the UAE significantly deviated from HWE. In addition, D12S11 in the Indian and Pakistani populations observed significant departure from HWE. In the Indian population the significant was not high ($p = 0.044$). X^2 value of the Chi square test and p value of the exact test for the three populations is shown in table 4.8.

Table 4.8 Shows the p values of Chi square test and Fisher exact test for the three populations (UAE, Indian and Pakistani) at five VNTR loci.

	Locus	D1S7	D2S44	D7S22	D7S21	D12S11
UAE	X^2	261.1	57.2	196.8	48.5	60.4
	df	231	66	120	45	55
	p (Chi)	0.0775	0.8844	***0.0044***	0.2947	0.2809
	p (Fisher)	***0.0052***	0.6949	***0.0000***	0.4212	0.0881
Indian	X^2	215.4	63.4	94.9	32.7	63.4
	df	231	66	120	45	55
	p (Chi)	0.7610	0.5665	0.9559	0.9137	0.2051
	p (Fisher)	0.5159	0.2501	0.7867	0.8883	***0.0439***
Pakistani	X^2	234.6	93.9	121.3	38.2	120.8
	df	231	66	120	45	55
	p (Chi)	0.4224	0.0622	0.4484	0.7526	***0.0103***
	p (Fisher)	0.4612	0.0694	0.4372	0.6405	***0.0057***

4.7.3 Independence of loci

In order to use the product rule over several loci it is important that these loci are not linked together. Therefore, a test for linkage disequilibrium should be performed. The linkage disequilibrium was tested for all possible pairs of loci in the three populations. The probabilities shown in table 4.9 are estimates obtained using shuffling tests (number of runs = 2,000) of the exact significance level. An exact probabilities less than 0.05 indicates a statistically significant amount of disequilibrium. The UAE population showed significant pairwise disequilibrium between locus D7S22 and D7S21. The Pakistani population showed a significant pairwise disequilibrium between locus D12S11 and D7S22 also between D7S21 and D1S7.

Table 4.9 Association between pair of loci has been investigated for the three populations.

The result showed no evidence for correlation between these pair of loci in Indian population. The UAE and Pakistani populations showed evidence for correlation between one and two pair of loci respectively. The significant departure observed is in bold italic.

Locus combination (VNTRs)	UAE (n=173)	Indian (n=154)	Pakistani (n=112)
D2S44/D12S11	0.75	0.66	0.46
D2S44/D7S22	0.42	0.49	0.82
D2S44/D7S21	0.29	0.69	0.48
D2S44/D1S7	0.88	0.86	0.06
D12S11/ D7S22	0.33	0.18	***0.002***
D12S11/D7S21	0.17	0.47	0.63
D12S11/D1S7	0.24	0.74	0.34
D7S22/D7S21	***0***	0.27	0.34
D7S22/D1S7	0.13	0.11	0.56
D7S21/D1S7	0.34	0.68	***0.01***

4.8 HETEROZYGOSITY

The genetic variation can be measured as a simple method by the level of heterozygosity observed. A significant deficiency in the value of heterozygosity may indicate evidence of population substructure and non-random mating of the examined population. Therefore, it is important to test a data set and determine whether the number of observed heterozygotes is within the predictions of HWE law or not.

The value of the expected and observed heterozygosity was calculated by using TFPGA software (Miller M., 1997). The values were calculated for each population at five loci. The results showed that the expected number of heterozygosity was not significantly different from the observed ones except at locus D7S22 in the UAE population. All the three populations at the five loci showed heterozygosity values equal or higher than 75%. The highest value being observed in D1S7 locus, UAE (92%), Indian (92%) and Pakistani (96%). The lowest value was observed in the UAE population at D7S22 locus (75%) and the lowest value observed in the Indian and Pakistani populations were at D12S11 locus (82% and 75% respectively). See table 4.10.

Table 4.10 Expected and observed heterozygosity for the five loci in the UAE, Indian and Pakistani populations (number of heterozygote alleles is in brackets beneath the frequency). The standard deviation (SD) and the p value from Chi square test are shown. No significant differences were observed ($p > 0.05$) in the three populations at the five loci except locus D7S22 in the UAE population ($p < 0.05$).

Locus		D1S7	D2S44	D7S22	D7S21	D12S11
UAE n = 173	Observed	0.9249 (160)	0.8555 (148)	0.7514 (130)	0.8324 (144)	0.7849 (136)
	Expected	0.9431 (163)	0.8887 (154)	0.9029 (156)	0.8514 (147)	0.8107 (140)
	SD	0.018 (3.1)	0.024 (4.2)	0.023 (4)	0.022 (3.8)	0.03 (5.2)
	X^2	0.055	0.234	4.33	0.061	0.114
	p-value	0.8-0.9	0.1-0.2	***0.02-0.05***	0.8-0.9	0.7-0.8
Indian n = 154	Observed	0.9156 (141)	0.8766 (135)	0.8831 (136)	0.8831 (136)	0.8182 (126)
	Expected	0.9365 (144)	0.8820 (136)	0.9014 (139)	0.8387 (129)	0.8298 (128)
	SD	0.02 (3.1)	0.026 (4)	0.024 (3.7)	0.03 (4.6)	0.03 (4.6)
	X^2	0.062	0.0007	0.065	0.38	0.032
	p-value	0.8-0.9	>0.95	0.7-0.8	0.5-0.7	0.8-0.9
Pakistani n = 112	Observed	0.9554 (107)	0.8304 (93)	0.9107 (102)	0.8571 (96)	0.7500 (84)
	Expected	0.9385 (105)	0.8817 (99)	0.8907 (100)	0.8589 (96)	0.8166 (91)
	SD	0.023 (2.6)	0.03 (3.4)	0.029 (3.2)	0.033 (3.7)	0.036 (4)
	X^2	0.038	0.364	0.04	0	0.54
	p-value	0.8-0.9	0.5-0.7	0.8-0.9	>0.99	0.3-0.5

Additionally, the observed heterozygosity at the five loci in the three populations was calculated manually from X-ray films. Some loci showed very high level of heterozygosity such as D1S7 and D7S21. The result is shown in table 4.11.

Table 4.11 The observed level of heterozygosity at five loci in the three populations. This observation was determined prior to applying the fixed bin system and counted directly from the autoradigraphs (two bands for heterozygotes and single bands for homozygotes). The heterozygosity in all three populations at the five loci was higher than 90% except D7S22 locus in the UAE population (81%).

Population/locus	D2S44	D7S22	D1S7	D7S21	D12S11
UAE	94%	81%	95%	93%	91%
Indian	92%	91%	98%	98%	95%
Pakistani	94%	92%	97%	91%	91%

The expected heterozygosity can be calculated by using the formula $h = [(1 - \sum x^2)(n/n - 1)]$, where x = allele frequency and n = total number of alleles observed in a population sample (Nei M. and Roychoudry A., 1974).

$$\text{Observed heterozygosity (h)} = \frac{n_h}{n} \qquad (4.3)$$

Where n_h is the number of heterozygotes and n is the total number of individuals profiled.

The standard deviation (SD) of the expected heterozygosity was calculated to determine the range of the expected heterozygosity by using the following equation (Edwards A. *et al.*, 1992).

$$SD = \sqrt{h(1-h)/n} \qquad (4.4)$$

Where (h) is the heterozygosity frequency and (n) is the number of population.

4.9 FORENSIC AND PATERNITY PARAMETERS

4.9.1 Forensic parameters

Population databases are used to calculate the frequency of each allele for a given locus. These databases are generally defined by racial group and geographical region because alleles may have different frequencies in different populations. The forensic parameters should be calculated for any new marker to be used in forensic identification. These main parameters are matching probability (MP), also known as probability of match and probability of discrimination (PD), also known as power of discrimination. These two parameters were calculated for the five VNTR loci in the three populations.

Additionally, the combined probability of discrimination (PD) and the combined of matching probability (MP) were calculated for the three populations at the five VNTR loci.

4.9.1.1 Matching probability

Matching probability (MP) can be defined as the number of individuals that may be surveyed before finding the same DNA pattern in a randomly selected individual from a population. For example, in the UAE population at D1S7 locus where MP = 0.013, this value of MP means that there is a 0.013 or 1.3% chance of two individuals selected from a random population having the same genotype.

The combined matching probability for the five loci was 0.00000002 in the UAE population and 0.00000003 in the Indian and Pakistani populations. This can be expressed as the likelihood ratio and therefore the chance of occurrence of the same profile, for example, in the UAE population for the five loci was 1 in 50 million.

$$\text{Matching Probability (MP)} = \sum (\text{expected genotype frequency})^2 \qquad (4.6)$$

The combined probability of match over several loci is the product of the value for all the loci (Jones D., 1972; Fisher R., 1951).

4.9.1.2 Discrimination power

The power of discrimination is the probability of two individuals selected from a population at random having a different genotype. The power of discrimination (PD) value is equal to one minus the value of matching probability. In the same example above where PD = 0.987. This value of PD means that the probability of selecting two individuals at random with different genotypes is 0.987 or 98.7%. Therefore, the higher the power of discrimination the more useful the marker will be in forensic analysis.

$$\text{Power of Discrimination (PD)} = 1 - \text{MP} \qquad (4.7)$$

For several loci $PD_{combined} = 1 - \prod_{i=1}^{n}(1 - p_{di})$ (4.8)

Where $\prod_{i=1}^{n}$ = the combined results of (1-PD) for all loci and i = frequencies for all possible alleles.

4.9.2 Paternity parameters

The paternity parameters should be calculated for any new marker to be used in paternity analysis from allele and genotype frequencies of a desired population. There are two common parameters such as power of exclusion (PE) or paternity exclusion and paternity index (PI). Therefore, the value of these two parameters was calculated for the five loci in the three populations tested.

4.7.2.1 Power of exclusion

Power of exclusion (PE) is the probability that the alleged man from a population can be excluded from being the true biological father when the man did not have the child's paternal allele at some loci or the man had neither of the child's alleles (Brenner C. and Morris J., 1990).

Power of exclusion can be calculated from the following equation (Brenner C. and Morris J., 1990), $PE = h^2 (1–2 * h * H^2)$ (4.9)

Where h is the heterozygosity number and H is the homozygosity number.

For several loci $PE_{combined} = 1 - \prod_{i=1}^{n}(1 - p_{Ei})$ (4.10)

Where $\prod_{i=1}^{n}$ = the combined results of (1-PE) for all loci and i = frequencies for all possible alleles.

The D1S7 had the highest value of power of exclusion (0.909) in the Pakistani population followed by UAE population (0.846) while D12S11 in the Pakistani population had the lowest value (0.510). The Indian and Pakistani populations had the highest value of combined power of exclusion (0.999). The combined power of exclusion in the UAE population was 0.997.

4.9.2.2 Paternity index

The paternity index (PI) reflects how many more times likely it is that an alleged father being tested is the biological father, rather than a randomly selected unrelated person from an ethnic population group (Brenner C. and Morris J., 1990). The typical paternity index for the five loci tested ranged from 11.2 for D1S7 locus to 2.00 for D12S11 locus. The five loci showed a combined typical paternity index of 320, 1208 and 1295 in the UAE, Indian and Pakistani populations respectively.

The typical paternity index (TPI) can be calculated by the following equation

$$TPI = H + h/2H \qquad (4.11)$$

Where H is the homozygosity number and h is the heterozygosity number.

In paternity cases when an individual can not be excluded from being the father of a child, there can be two possibilities, either the alleged father is the biological father or some one else in the population is the father. The following equation is used,

$$PI = \frac{P(Ma)xP(AFb)}{P(Ma)xP(fb)}$$

Where, P(Ma) is probability of the mother passing on allele a, P(AFb) is probability of the alleged father passing on allele b and (fb) is the frequency of allele b in the population. When the mother and the alleged father are both heterozygous the equation can be further simplified to PI = 0.5/fb. The chance of the mother passing on allele a is 0.5 and the chance of the father passing on allele b is 0.5 too. However, in the situation where the alleged father is homozygous the equation is PI = 1/fb.

The probability of paternity can be calculated using the following equation

Probability of paternity (W) = PI/PI+1

Therefore, the probability of paternity for the five loci in the UAE, Indian and Pakistani populations was 99.7%, 99.9% and 99.9% respectively assuming a prior probability of guilt of 50%.

Forensic and paternity parameters for the five loci in the UAE, Indian and Pakistani populations were calculated. The result is given in tables 4.12a, b and c.

Table 4.12a Forensic and paternity parameters for the five loci in the UAE population. MP = Matching Probability, PD = Power of discrimination, TPI = Typical Paternity Index and PE = Paternity of Exclusion.

UAE Pop.	MP	PD	Exp. as 1 in	TPI	PE
D1S7	0.013	0.987	78.1	6.65	0.846
D2S44	0.025	0.975	39.7	3.46	0.706
D7S21	0.042	0.958	23.8	2.98	0.660
D7S22	0.025	0.975	40.7	2.01	0.512
D12S11	0.067	0.933	15.0	2.32	0.571
Combined	0.00000002	0.99999998	1 in $4.5\ 10^7$	320	0.997

Table 4.12b Forensic and paternity parameters for the five loci in the Indian population.

Ind. Pop.	MP	PD	Exp. as 1 in	TPI	PE
D1S7	0.013	0.987	77.5	5.92	0.827
D2S44	0.031	0.969	31.8	4.05	0.748
D7S21	0.051	0.949	19.4	4.28	0.761
D7S22	0.022	0.978	44.6	4.28	0.761
D12S11	0.057	0.943	17.4	2.75	0.633
Combined	0.00000003	0.99999997	1 in $3.7\ 10^7$	1208	0.999

Table 4.12c Forensic and paternity parameters for the five loci in the Pakistani population.

Pak. Pop.	MP	PD	Exp. as 1 in	TPI	PE
D1S7	0.015	0.985	64.7	11.2	0.909
D2S44	0.034	0.966	29.6	2.95	0.657
D7S21	0.040	0.960	24.9	3.50	0.709
D7S22	0.028	0.972	35.2	5.60	0.817
D12S11	0.060	0.940	16.6	2.00	0.510
Combined	0.00000003	0.99999997	1 in $2.8\ 10^7$	1295	0.999

4.10 DISCUSSION

4.10.1 Hardy-Weinberg equilibrium tests

The exact test with 2,000 shufflings was performed to test departure from the HWE. The data showed some deviation from HWE and linkage dis-equilibrium between the five loci in the three populations. More deviations were observed in the UAE and Pakistani populations compared to Indian population. This may be because of a true reflection of the population being out of HWE at some of these loci or sampling effect. Another possibility is because of high consanguineous marriages (between close relatives) (Shami S. and Siddiqui H., 1984; Bener A. *et al.*, 1996) or because of how VNTRs behave in agarose gel electrophoresis. The product rule can be used in forensic and paternity casework analysis to estimate the frequency of a multiple locus DNA profile with appropriate F_{ST} value.

4.10.2 Heterozygosity

The apparent heterozygote deficiency at D7S22 in the UAE population may not be caused by population substructure. One possible explanation may be because of inability to detect extreme sized alleles or distinguish between two closely spaced alleles in Southern gel electrophoresis system (Chakraborty R. *et al.*, 1992).

4.10.3 Comparison of the three populations studies

The RxC contingency table was used to test for the heterogeneity in allele frequencies among the three populations. There was significant difference observed at three loci when the UAE population compared to Indian and Pakistani populations. There was no significant difference between the Indian and the Pakistani population in the allele

frequencies for four loci (D1S7, D7S21, D7S22 and D12S11), so they can be pooled to create a population with an increased number of individuals. However, there was evidence of heterogeneity between the two Indian subcontinent populations for the D2S44 locus as there was significant difference ($p = 0.003$).

When these three populations compared to each other by using F-statistics, there were little genetic differences between the UAE and other two populations (Indian and Pakistani). However there was no genetic differences detected between the two Indian subcontinent populations and therefore these two populations can be considered as one population.

Nichols and Balding (Nichols R. and Balding D., 1991; Balding D. and Nichols R., 1994) recommended a routine use of relatively high F_{ST} values (0.05), while others considered this as a high value (Brookfield J., 1991; Morton N. *et al.*, 1993). Most of the studied populations had F_{ST} value less than 0.01 (Morton N., 1992; NRC, 1996). The NRC recommended (NRC, 1996) the use of 0.01 as a conservative value of F_{ST} between different Caucasian groups and for some small isolated populations a value of 0.03 to be used. The FSS recommended the use of $F_{ST} = 0.03$ in the UK however, the value of 0.05 should be used with the Indo-Pakistani population. In this study the highest value of F_{ST} was 0.006, which is less than 0.01. However, the use of upper bound (0.0126) is found to be more conservative value for forensic VNTR markers in the UAE.

4.10.4 Forensic and paternity parameters

These five VNTR loci were found to be well suitable for forensic analysis and paternity testing in the UAE, Indian and Pakistani populations. D1S7 locus is not a suitable marker in paternity testing because it has a high mutation rate (5.2%) but it is an

excellent marker for individual identification. For forensic identification purposes the combined power of discrimination for these five VNTR loci is 0.99999998, 0.99999997, and 0.99999997 in the UAE, Indian and Pakistani populations respectively. In paternity testing the combined exclusion power is 99.7%, 99.9% and 99.9% in the UAE, Indian and Pakistani populations. For inclusion purpose, the combined paternity index is 320 with the probability of paternity of 99.69% in the UAE population. That means it is 320 times more likely the alleged father is the true biological father compared to an unrelated man in the population. It can be noticed that some power of these markers is lost because of binning system. The combined paternity index is much higher in the Indian (1295) and Pakistani populations (1208) compared to the UAE (320) with the probability of paternity of 99.92% in both populations. This may be because of deficiency in heterozygosity in the UAE population at some loci.

It was recommended that a number of markers to be used in paternity testing should have a power of exclusion of 99% or higher (Hammond *et al.*, 1994). However, this level (99%) considers as low level for paternity exclusion. The power of the combined exclusion of paternity after typing the five loci at the three populations was higher or equal to 99.7%. Therefore, these five loci are able to provide a highly discriminating system for exclusion of paternity testing.

These five loci observed high value of paternity index in the Indian and Pakistani populations and relatively high in the UAE population. For forensic purposes the power of discrimination is very high in the three populations (0.99999998). These values are significantly higher than one obtained from protein markers or DQ-alpha and polymarker which is the system currently used in the UAE for paternity testing and forensic analysis.

CHAPTER 5: AUTOSOMAL SHORT TANDEM REPEATS

5.1 INTRODUCTION

The major advantages for individual differentiation of microsatellites (STRs) in comparison to minisatellites (VNTRs) analysis include discrete allele designation and PCR amplification. Moreover, analysis of STR loci is rapid, sensitive, cost effective and generally amenable to automation. STRs represent a rich source of highly polymorphic markers in the human genome and occur at an estimated frequency of one tri or tetrameric STR every 300-500 kb throughout the human genome (Edwards A. *et al.*, 1991a, b). The analysis of STR loci by automated fluorescence has become the method of choice for use in routine forensic investigations (Ziegle J. *et al.*, 1992; Fregeau and Fourney 1993) and parentage testing.

The use of STR loci is facilitated by the availability of commercially available kits. There are several multiplexing kits available in which amplification of several loci can be achieved in a single-tube reaction, with different dye labels on the primers allowing the identification of alleles at different loci even when they overlap in size. Simultaneous amplification decreases the cost and time of determining genotypes in population studies on several loci. Additionally, less DNA sample is consumed than when analysing each locus independently and less reagents are required.

In this study the PowerPlex 1.2 system (Promega, Madison, Wisconsin, USA) was used which is able to detect eight different tetrameric STR loci plus the amelogenin locus.

5.2 THE GENERAL POPULATION

The determination of allele/genotype frequency data is essential to estimate the frequency of a genetic profile in a population. A lot of studies have been reported for autosomal STR allele frequencies for a number of different ethnic groups, however, published data for Arabs (native UAE), Egyptian, Asian-Indian and Pakistani population from the UAE are limited. Additionally, no published data for any Arab group existed at the beginning of this study. Therefore, it was desirable to determine the extent of genetic variation between the four populations (UAE, Egyptian, Indian and Pakistani), and estimate the power of discrimination and probability of genotype identity by chance based on the analysis of eight STR loci. For that reason it was desirable to collect enough samples from the relevant populations to acquire a suitable databases (NRC report 1996).

Over the course of human evolution, a number of partially differentiated populations have formed as a consequence of geographic, religious and social stratification. Therefore, it is of primary importance to recognise the various ethnic groups represented in the general population. Since allele frequencies are expected to vary between racial groups (Cavali-Sforza L. and Bodmer W., 1971), it is important that databases be established for individual racial groups represented in the general population.

In addition to compiling databases of representative racial groups within the general population, it is also necessary to establish databases for populations that are genetically and geographically isolated from the general population.

5.3 DNA EXTRACTION AND DILUTION

The DNA samples which were extracted for VNTR analysis (see section 3.5.1) were also used for STR analysis. In addition, further DNA samples have been extracted. DNA samples of high quality and quantity were re-diluted between 10 to 50 fold using Milli-Q water to become at range of 2 to 10 ng/µl which is more suitable for PCR amplification. In order to check the quantity of diluted DNA it was compared to a commercial lambda DNA of known concentration ranging from 5 to 30 ng using agarose mini-gels electrophoresis. See figure 5.1.

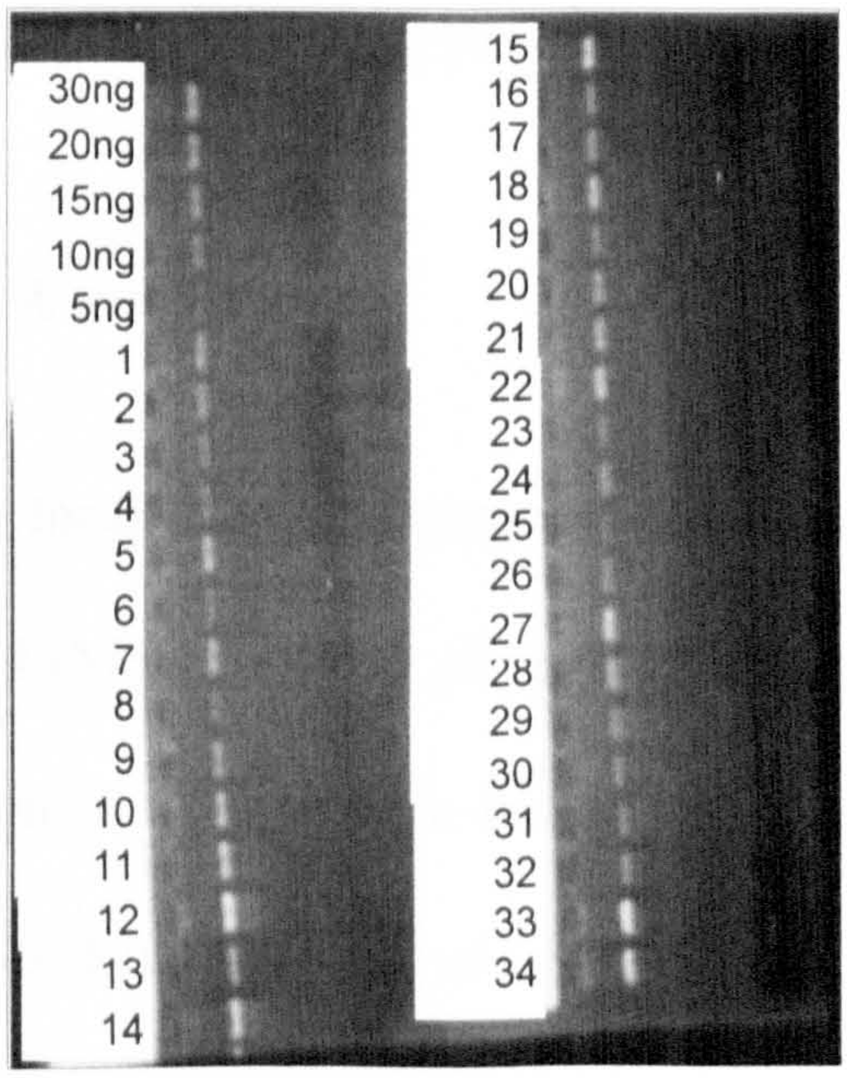
5.4 MULTIPLEX ANALYSES OF STR LOCI

The GenePrint™ PowerPlex™ 1.2 system was used to co-amplify eight STR loci plus amelogenin. All STR loci used in this kit are the tetra nucleotide repeats type. These type of STR loci are preferable over the dinucleotide repeats in forensic and paternity analysis since a wider allele spacing and a low percentage of stutter is observed (Edwards A. *et al.*, 1991a, b; Kimpton C. *et al.*, 1993; Schumm J. *et al.*, 1993). Moreover, the European DNA profiling group (EDNAP) confirmed that simple repeat STR loci are easier to interpret and very useful in comparison studies (Gill P. *et al.*, 1994b). According to the manufacturer the loci included in this kit have been selected because they have a high degree of heterozygosity and display a minimum of artefacts such as repeat slippage, terminal nucleotide addition and genetic artefacts called microvariant alleles.

149A

Figure 5.1 Diluted DNA samples.

The DNA samples where diluted and run on 1% agarose gel electrophoresis for 15 min at 110 voltage.


30ng
20ng
15ng
10ng
5ng
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34

5.5 DESCRIPTION OF THE LOCI

5.5.1 D5S818 locus

D5S818 locus is located on long arm of chromosome 5 at 5q21-q31. The sequence of the repeat unit is AGAT. This locus has the smallest allele sizes compared to other loci in the PowerPlex™ 1.2 kit and ranged from 112 to 145 base pairs (7-15) in this study. The small size of this locus makes it extremely useful in forensic applications especially in highly degraded samples. This locus has been included in some other commercial kits such as Profiler™ and Profiler Plus™ (PE-Applied Biosystems).

5.5.2 Human vWAFA31/A (vWA) locus

The vWA locus is located on chromosome 12 at position 12p12-pter, between nucleotides 1640 and 1794 (Mancuso D. *et al.*, 1989). It is one of three (TCTA) tandem repeat regions within intron 40 of the human von Willebrand factor gene (Peake I. *et al.*, 1990; Ploos V. *et al.*, 1990; Kimpton C. *et al.*, 1992). This region is reported to be the most polymorphic of the three regions (Kimpton C. *et al.*, 1992). There are two different sequences of the repeat units (TCTA and TCTG) between the two flanking regions (Brinkmann B. and Weigand P., 1993; Moller A., *et al.*, 1994). Micro-heterogeneity can be observed in the structure of the alleles (Barber M. *et al.*, 1995). The size of alleles ranged from 126 to 166 base pairs (allele 13 to allele 21) in this study. Applicability of this STR locus in forensic analysis and paternity testing have been widely evaluated (Sajantila A. *et al.*, 1994; Lorente J. *et al.*, 1994; Neuhuber F. and Radacher M., 1997). Studies on families were carried out without any evidence of mutations (Lorente J. *et al.*, 1994; Trabetti E. *et al.*, 1993). This locus has been reported to be reliable and robust and therefore it is recommended by the EDNAP group (Kimpton C. *et al.*, 1995). The vWA

locus is included in almost all commercially and non-commercially kits for human identity purposes such as Cofiler™, Profiler Plus™, SGM Plus (PE-Applied Biosystems) and Quadroplex (FSS).

5.5.3 D13S317 locus

This locus is located on chromosome 13 at position 13q22-q31. The sequence of the repeat unit is AGAT. The size of alleles ranges from 168 to 200 base pairs (7-15). This locus is included in other commercial kits such as Profiler™ and Profiler Plus™ (PE-Applied Biosystems) because of its usefulness in forensic and paternity analysis.

5.5.4 Human THO1 locus

The HUMTHO1 locus is located on chromosome 11 at position 11p15.5 within intron 1 near the tyrosine hydroxylase gene (Polymeropoulos M. *et al.*, 1991; Edwards A. *et al.*, 1991 and 1992). The sequence of the repeat unit is AATG. More than eight alleles, 5-11, have been reported to be observed in this locus (Budowle B. *et al.*, 1997; Brinkmann B., 1998) but only 5 of them are common. The repeat region is regularly spaced with 4 nucleotides in each repeat except one triple repeat or microvariant type (allele 8.3, 9.3 and 10.3) where one base (A) of the AATG is missing (Puers C. *et al.*, 1993).

The HUMTHO1 is a useful marker in forensic analysis and paternity investigations (Brinkmann B. *et al.*, 1991; Lareu M. *et al.*, 1994; von Oorschot R. *et al.*, 1996; Neuhuber F. and Radacher M., 1997). It has high heterozygosity index (Gill P. and Evett I., 1995). This locus is ideal candidate for obtaining reproducible results between laboratories and recommended by the EDNAP group (Kimpton C. *et al.*, 1995). It is included in almost all

commercially and non-commercially human identity kits such as Cofiler™, Profiler Plus™, SGM Plus (PE-Applied Biosystems) and Quadroplex (FSS).

5.5.5 D7S820 locus

This locus is located on chromosome 7 at position 7q. The sequence of the repeat unit is AGAT. Nine alleles have been described, ranging from 215 to 247 bp (6-14) with primer set of Promega. This locus was used in some other commercial kits such as Profiler Plus™ and Cofiler™ (PE-Applied Biosystems) for its usefulness in forensic and paternity analysis.

5.5.6 Human TPOX locus

This locus is located on chromosome 2 at position 2p23-2pter. The sequence of the repeat unit is AATG. It was identified within intron 10 of the human thyroid peroxidase gene (TPOX) (Anker R. *et al.*, 1992). All alleles in this locus differed by one four repeat unit and therefore it can be classified as a simple repeat STR type. Eight alleles have been described, ranging from 224 to 252 bp (6-13) with primer set of Promega Corporation. This locus was described in some other commercial kits such as Cofiler™ (PE-Applied Biosystems) for its usefulness in forensic and paternity analysis. However, this locus is not a very good marker compared to vWA because it has only 3 or 4 common alleles and the rest are very rare.

5.5.7 D16S539 locus

This locus is located on chromosome 16 at position 16q24-qter. The sequence of the repeat unit is AGAT. This locus is a simple STR type where alleles differ by one four bp repeat unit. Nine alleles have been described, ranging from 264 to 304 bp (5-15) with

primer set of Promega Corporation. This locus was described in some other commercial kits such as Cofiler™ (PE-Applied Biosystems) for its usefulness in forensic and paternity analysis. However, this locus is not as good as vWA because it has only 3 or 4 common alleles and the rest are very rare.

5.5.8 Human CSF1PO locus

This locus is located on chromosome 5 at position 5q33.3-34 (human c-fms proto-oncogene for CSF-1 receptor gene). The sequence of the repeat unit is AGAT (Hampe A. *et al.*, 1989). It can be classified as a simple STR. Although 10 alleles have been described, ranging from 291 to 327 bp (6-15) with primer set of Promega Corporation, this locus shows relatively few common alleles, three to four. This locus was described in some other commercial kits such as Profiler™ and Cofiler™ (PE-Applied Biosystems) for its usefulness in forensic and paternity analysis.

5.5.9 Human amelogenin locus

Amelogenin is not an STR, it is a six base pairs deletion at X chromosome, but displays a 212 base X-specific band and a 218 base Y-specific band. This locus is located on chromosome X and Y. It is included in several commercial kits such as Cofiler™, profiler™ and SGM plus (PE-Applied Biosystems). It is included also in the SGM kit which is developed by FSS, UK.

Of the eight STR loci analysed in this study, six are located on different chromosomes. The other two are located on chromosome 5, D5S818 and CSF1PO, separated by approximately 22-45 cM and not linked together (reviewed by Holt C. *et al.*, 2000).

5.5.10 Allele designation and sizing

Allele nomenclature is based on the structure of the alleles. The fragments are designated according to the repeat unit number between two unique flanking regions. Therefore, an allele containing seven contiguous identical copies of the repeat unit is called allele 7 while allele 8 contains eight such repeats. Non-consensus alleles are described by the number of consensus repeats, followed by a decimal point and the number of the partial repeat. For example, the STR locus HumTHO1 contains a non-consensus allele, which is one base pair shorter than ten repeat units and thus designated 9.3 (Pures C. *et al.*, 1993).

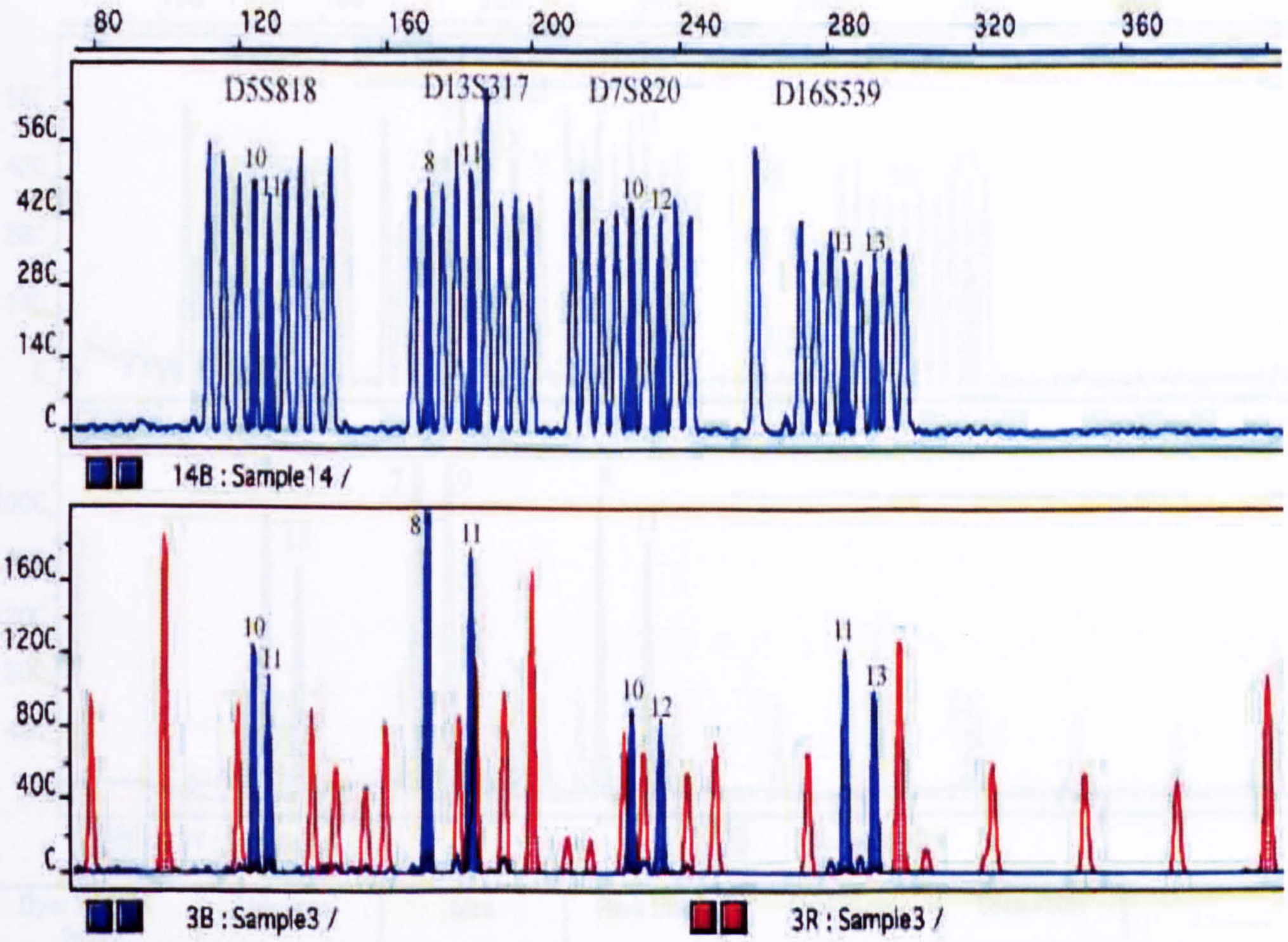
STRs within protein coding genes are named after the sequence of the gene such as the HumvWA and HumTPOX loci (Barber D. *et al.*, 1995). Repetitive sequences without any connection to protein coding genes are named according to their chromosomal location such as D3S1358 and D5S818 loci being on chromosomes 3 and 5 respectively (Bar W. *et al.*, 1997).

Allelic ladders (Puers C. *et al.*, 1994) of all loci were loaded on each capillary electrophoresis run and injected twice, flanking the amplified DNA samples plus positive control (human cell line K562) and negative control. Internal size standard, CXR from Promega Corporation or GS-500 from PE Applied Biosystems, was included in each sample run to determine the size of DNA fragments and helps in eliminating the inter-capillary electrophoresis variation (Kimpton C. *et al.*, 1993). The use of size standard and allelic ladder allows accurate and confident assignment of alleles at these STR loci. The GeneScan software automatically sized the PCR products (DNA fragments). See figure 5.2.

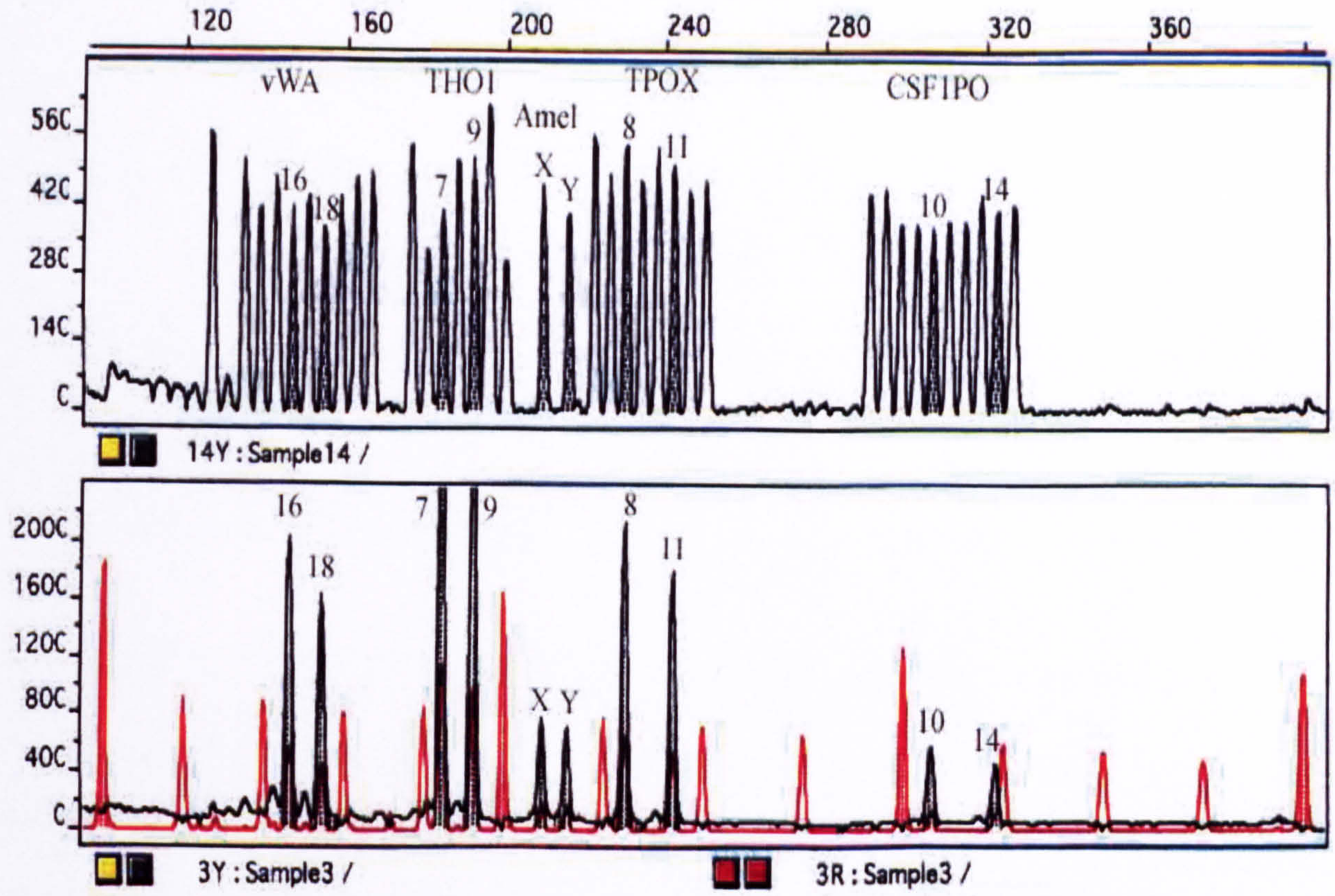
155 A

Figure 5.2 Electrophoretic profile of a single DNA sample.

These two electrophoretic profiles are illustrating a single DNA sample which amplified using the GenePrint PowerPlex 1.2 system and detected with the ABI PRISM 310 Genetic Analyser. The fluorescein-labelled loci (D5S818, D13S317, D7S820 and D16S5339) are displayed in the first profile. The TMR-labelled loci (vWA, THO1, Amelogenin, TPOX and CSF1PO) are displayed in the second profile. The allelic ladder for all loci is included. The fluorescent ladder (CXR), 60-400 bases, is shown in red and is used as the internal size standard.



Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
14B, 67	15.15	124.23	488	4060	4132
14B, 68	15.30	128.37	458	3981	4172
14B, 74	16.81	171.67	489	4277	4584
14B, 77	17.23	183.61	511	4570	4697
14B, 86	18.69	227.11	445	4529	5097
14B, 88	18.96	235.13	426	4252	5169
14B, 95	20.57	285.21	339	3484	5609
14B, 97	20.83	293.28	348	3637	5680
3B, 43	14.70	124.18	1258	9793	4009
3B, 44	14.85	128.37	1095	8615	4048
3B, 48	16.30	171.52	2486	23374	4444
3B, 50	16.69	183.42	1794	17893	4552
3B, 52	18.11	227.08	913	8787	4938
3B, 54	18.37	235.18	843	8177	5008
3B, 57	19.92	285.11	1258	12415	5432
3B, 59	20.17	293.01	1004	10111	5499



Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
14Y, 65	15.94	146.44	419	4219	4345
14Y, 67	16.21	154.48	378	3890	4421
14Y, 73	17.24	184.14	415	4532	4702
14Y, 75	17.52	192.04	520	5626	4776
14Y, 78	18.11	209.47	469	4480	4937
14Y, 79	18.32	215.83	406	3808	4995
14Y, 82	18.80	230.22	543	5554	5125
14Y, 85	19.19	242.20	502	5053	5232
14Y, 92	21.27	307.14	378	4048	5801
14Y, 96	21.78	323.12	412	4685	5938
3Y, 97	15.46	146.32	2062	28747	4214
3Y, 99	15.73	154.40	1650	25240	4288
3Y, 117	16.73	184.41	3303	42064	4561
3Y, 119	16.99	192.29	2887	38574	4632
3Y, 129	17.55	209.52	792	11321	4785
3Y, 131	17.75	215.80	729	10992	4840
3Y, 137	18.21	230.31	2187	27616	4966
3Y, 140	18.59	242.28	1816	22367	5069
3Y, 192	20.60	307.09	588	10025	5617
3Y, 201	21.08	323.03	469	7206	5748

In order to determine the mean of the length of each allele, allelic ladders with almost all ranges of alleles were analysed on four different capillary electrophoresis runs (appendix F). Allelic windows constructed using ±0.5bp relative to the allelic ladders in order to designate the allelic size of unknown samples. The 3 standard deviation was much less than ±0.5bp.

5.6 RESULTS

5.6.1 Co-amplification of STR loci

Eight tetranucleotide repeat STR loci plus amelogenin were amplified successfully as a co-amplification of nine loci in a single test tube via PCR technique. Four different populations and 741 individuals from UAE, Indian, Pakistani and Egyptian were analysed and typed successfully. The UAE Arabic native population was divided into two portions, 100 sample profiled from the Emirate of Sharjah and 129 sample from the Emirate of Abu Dhabi.

5.6.2 Allele frequencies

Capillary electrophoresis was carried out using an ABI Prism 310 Genetic Analyser (Applied Biosystems, Foster, CA, USA) for analysing the raw data. The size of alleles was determined by using the 310 GeneScan analysis software application, version 2.0.2 (Applied Biosystems) by comparing them to fragments contained in a size standard (CXR or GS 500) and the allelic ladder.

The distributions of observed allelic frequencies for the eight STR loci in the four population samples (UAE, Egyptian, Indian and Pakistani) are shown in chapter 6.

5.7 DISCUSSION

5.7.1 Samples collection

This study was performed on 620 whole fresh blood samples that were obtained from randomly selected from three different populations residing within the UAE. Additionally, 121 samples were obtained from Egyptian population living in El-Minia city which is 250 km south of Cairo. This was done in order to compare two Arabic populations, UAE Arab native and Egyptian. Approximately 15% of the total UAE population are Arab from different Arabic countries and approximately 5% are Egyptians.

Blood samples are widely used for DNA extraction. Blood samples were used successfully in protein and DNA analysis (Droog S. *et al.*, 1996). There are other sources of DNA such as semen, saliva, buccal swabs, hair roots and urine (Brinkmann B., *et al.*, 1992). Since PCR requires very little DNA sample, buccal swabs are getting more popular and can be used in paternity testing especially when samples are obtaining from children and venipuncture is not practical (Maha G. *et al.*, 1995; Okazaki K. *et al.*, 1996).

5.7.2 DNA extraction

All the amplifications for the four populations studied were made with DNA extracted from whole blood and by using Puregene™ DNA Isolation Kit. This kit do not contains toxic organic solvents as in organic extraction. This kit allows the rapid isolation of high quality purified DNA suitable for DNA analysis. Since STR analysis requires only 1-2 ng therefore, the extracted DNA has been diluted up to 50 fold to be more suitable for PCR amplification.

5.7.3 Multiplex PCR

The co-amplification of the nine loci was performed using the PowerPlex 1.2 system according to the manufacturer's instructions, except that the PCR was carried out in one third of the recommended total reaction volume containing 2-5 ng template DNA.

The use of PowerPlex 1.2 kit (Promega Corporation) were found to be well suitable for forensic analysis and paternity testing in the UAE populations. All STR loci used in the kit is a simple STR type except vWA which is a compound. Simple STR systems are, in general, the most suitable loci for comparison results between laboratories (Gill P. *et al.*, 1994b; Kimpton C. *et al.*, 1995). Individual STR alleles can be precisely and individually identified.

A very small sizes of these eight STR loci made them more likely to be successful on old or severely degraded material (Gill P. *et al.*, 1994a). Out of eight STR loci six of them (TPOX, THO1, vWA, D5S818, D7S820 and D13S317) plus amelogenin locus have DNA fragment sizes ≤ 250 base pairs which is a very important aspect of forensic casework (Gill P. *et al.*, 1996).

All these eight STR loci are included in Combined DNA Indexing System (CODIS), PowerPlex™ 16 System (Promega Co.) and Identifiler (PE, Applied Biosystems)

CHAPTER 6: STR STATISTICAL ANALYSIS

6.1 INTODUCTION

STR polymorphisms have been studied in three populations residing in the UAE. In total 741 samples were typed and analysed at eight STR loci D5S818, D7S820, D13S317, D16S539, vWA, THO1, TPOX and CSF1PO. The UAE Arabic population samples were collected from two different Emirates, Abu Dhabi (129 samples) and Sharjah (100 samples) where the distance between the two Emirates is about 180 km. Egyptians from El-Minia City (121 samples) were profiled and compared to the UAE Arabic population.

The allele and genotype frequencies were calculated. Comparison of the genotype frequencies for the four populations was made. Comparison between two UAE native Arabic populations, Sharjah and Abu Dhabi, was made. A standard Chi square goodness of fit, heterozygosity and the exact test were used to test the data for Hardy Weinberg equilibrium. The RxC contingency test was used to compare allele distributions for population sample homogeneity. The F_{ST} (co-ancestry coefficient) was performed to measure the genetic variation between the four populations and between two UAE native Arabic sub-populations.

6.2 ALLELE FREQUENCIES

Variances in observed allele frequencies and their standard deviations (equation 4.1) were counted for the four populations at the eight loci. Only one locus (D5S818) is

showing here (tables 6.1 a-d) and the other seven loci can be seen in appendix G (section 1.1).

Table 6.1a Allele frequencies in 229 UAE Arab individuals for the locus D5S818. The maximum and minimum allele frequencies are shown in bold italic.

Allele	Observed (D5S818)	Allele frequency	SD	Heterozygosity	Heterozygosity frequency
8	***6***	***0.0131***	0.0053	6	0.0262
9	18	0.0393	0.0091	16	0.0699
10	57	0.1245	0.0154	49	0.2140
11	109	0.2380	0.0199	71	0.3100
12	***176***	***0.3843***	0.0227	100	0.4367
13	85	0.1856	0.0182	73	0.3188
14	7	0.0153	0.0057	5	0.0218

Table 6.1b Allele frequencies in 192 Indian individuals for the locus D5S818. The maximum and minimum allele frequencies are shown in bold italic.

Allele	Observed (D5S818)	Allele frequency	SD	Heterozygosity	Heterozygosity frequency
8	2	0.0052	0.0037	2	0.0104
9	16	0.0417	0.0102	16	0.0833
10	42	0.1094	0.0159	40	0.2083
11	***125***	***0.3255***	0.0239	85	0.4427
12	115	0.2995	0.0234	83	0.4323
13	79	0.2057	0.0206	55	0.2865
14	4	0.0104	0.0052	4	0.0208
15	***1***	***0.0026***	0.0026	1	0.0052

Table 6.1c Allele frequencies in 197 Pakistani individuals for the locus D5S818. The maximum and minimum allele frequencies are shown in bold italic.

Allele	Observed (D5S818)	Allele frequency	SD	Heterozygosity	Heterozygosity frequency
7	2	0.0051	0.0036	2	0.0102
8	***1***	***0.0025***	0.0025	1	0.0051
9	26	0.0660	0.0125	24	0.1218
10	46	0.1168	0.0162	44	0.2234
11	128	0.3249	0.0236	88	0.4467
12	***131***	***0.3325***	0.0237	95	0.4822
13	55	0.1396	0.0175	41	0.2081
14	5	0.0127	0.0056	5	0.0254

Table 6.1d Allele frequencies in 121 Egyptian individuals for the locus D5S818. The maximum and minimum allele frequencies are shown in bold italic.

Allele	Observed (D5S818)	Allele frequency	SD	Heterozygosity	Heterozygosity frequency
8	5	0.0207	0.0179	5	0.0413
9	17	0.0702	0.0322	17	0.1405
10	34	0.1405	0.0438	24	0.1983
11	42	0.1736	0.0477	36	0.2975
12	***95***	***0.3926***	0.0615	59	0.4876
13	44	0.1818	0.0486	32	0.2645
14	3	0.0124	0.0139	3	0.0248
15	***2***	***0.0083***	0.0114	2	0.0165

Before using any marker in the forensic and paternity analysis it is important to calculate the population frequencies of that marker. The allele frequencies for each population was counted from the genotypes observed at eight STR loci and divided by the total number of all alleles for each population.

Among the eight loci tested vWA and CSF1PO loci had the largest allele number (10 alleles) only six are common in vWA and only 4 are very common in CSF1PO. The vWA locus had 8, 10, 8 and 9 different alleles in the UAE, Indian, Pakistani and Egyptian populations respectively. CSF1PO locus had 9, 7, 7 and 8 different alleles in the UAE, Indian, Pakistani and Egyptian populations respectively. At locus vWA allele 17 was the most common allele in the UAE, Egyptian and Pakistani populations with a frequency of 0.32, 0.26 and 0.28 respectively. Allele 16 was the most common allele with a frequency of 0.24 in the Indian population. Allele 13 had the lowest frequency of 0.002, 0.002 and 0.008 in the UAE, Egyptian and Pakistani populations respectively. While allele 9 observed the lowest frequency of 0.003 in the Indian population.

CSF1PO locus had three most common alleles 10, 11 and 12 with a combined frequency of 0.88, 0.88, 0.88 and 0.87 in the UAE, Egyptian, Indian and Pakistani populations respectively. Allele 7 and 16 had the lowest frequency of 0.002 in the UAE population. Allele 7 and 14 had the lowest frequency of 0.013 in the Egyptians. Allele 15 had the lowest frequency of 0.005 in the Indian population while allele 8 had the lowest frequency of 0.005 in the Pakistani population. Although this locus has 10 different alleles it is not very discriminative because alleles are not evenly distributed over the locus.

D16S539 and THO1 loci had the lowest allele number (8 alleles). D16S539 locus had 8 different alleles in each population. At locus D16S539 allele 11 was the most common allele with a frequency of 0.347, 0.308, 0.352 and 0.299 in the UAE, Egyptian, Indian and Pakistani populations respectively. However, allele 15 had the lowest frequency of 0.002, 0.0042, 0.003 and 0.003 in the UAE, Egyptian, Indian and Pakistani populations respectively.

THO1 locus observed only five common alleles. This locus had 7, 7, 5 and 6 different alleles in the UAE, Indian, Pakistani and Egyptian populations respectively. Allele 6 is being the most common allele with a frequency of 0.28 and 0.27 in the UAE and Pakistani populations respectively. Allele 7 was the most common allele with a frequency of 0.23 in the Egyptians. Allele 9 was the most common allele with a frequency of 0.30 in the Indian population.

D5S818 locus had 7, 8, 8 and 8 different alleles in the UAE, Indian, Pakistani and Egyptian populations respectively. Allele 12 was the most common allele with a frequency of 0.384, 0.39 and 0.332 in the UAE, Egyptian and Pakistani populations respectively. Allele 11 was the most common in Indian population with a frequency of 0.326. However, allele 8 showed the lowest frequency of 0.013 and 0.003 in the UAE and Pakistani populations while Allele 15 had the lowest frequency of 0.008 and 0.003 in the Egyptian and Indian populations respectively.

D7S820 locus had 9, 8, 7 and 7 different alleles in the UAE, Indian, Pakistani and Egyptian populations respectively with only five are common. Allele 10 was the most common allele with a frequency of 0.301, 0.33, 0.253 and 0.254 in the UAE, Egyptian, Indian and Pakistani populations respectively. Allele 8.3 was only observed in the UAE

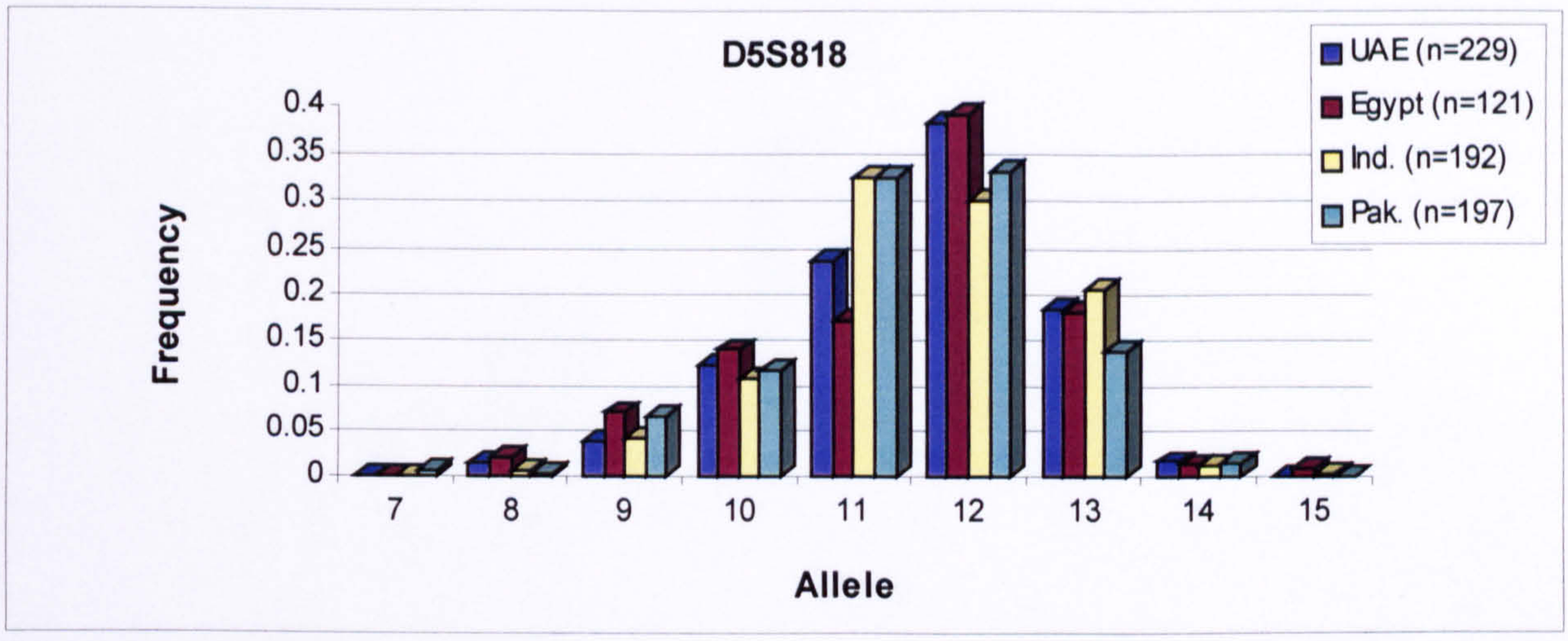
population and had the lowest frequency of 0.004. Allele 13 had the lowest frequency of 0.004 in the Egyptians, while allele 14 had the lowest frequency of 0.003 in the Indian population and in the Pakistani population the lowest frequency of 0.013 was observed in allele 13. Alleles 10 and 11 have a combined frequency of 0.52, 0.61, 0.46 and 0.48 in the UAE, Egyptian, Indian and Pakistani populations respectively.

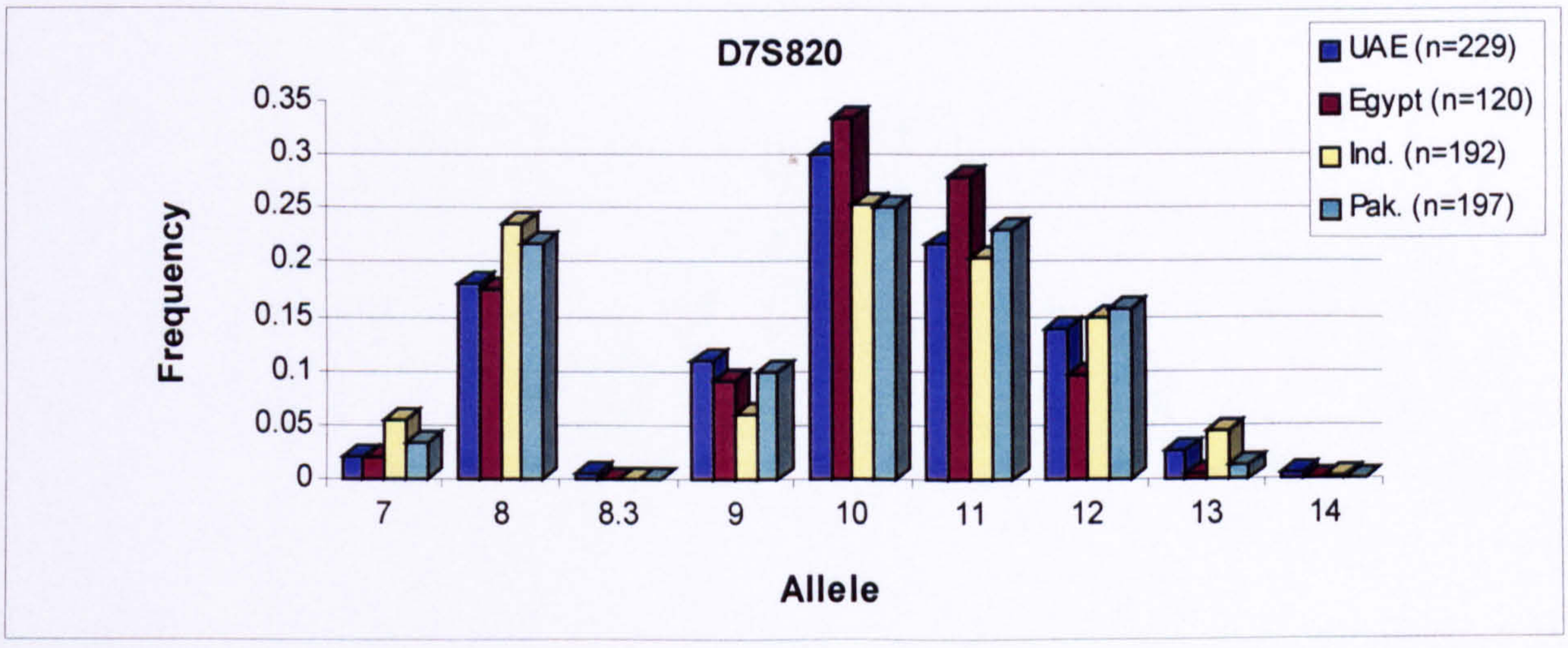
D13S317 locus had 8, 9, 7 and 9 different alleles in the UAE, Indian, Pakistani and Egyptian populations respectively. The most two common alleles are allele 11 and 12 which have a combined frequency of 0.59, 0.57, 0.57 and 0.55 in the UAE, Egyptian, Indian and Pakistani populations respectively. Allele 15 had the lowest frequency of 0.002, 0.004 and 0.003 in the UAE, Egyptian and Indian populations respectively while allele 14 had the lowest frequency of 0.015 in the Pakistani population.

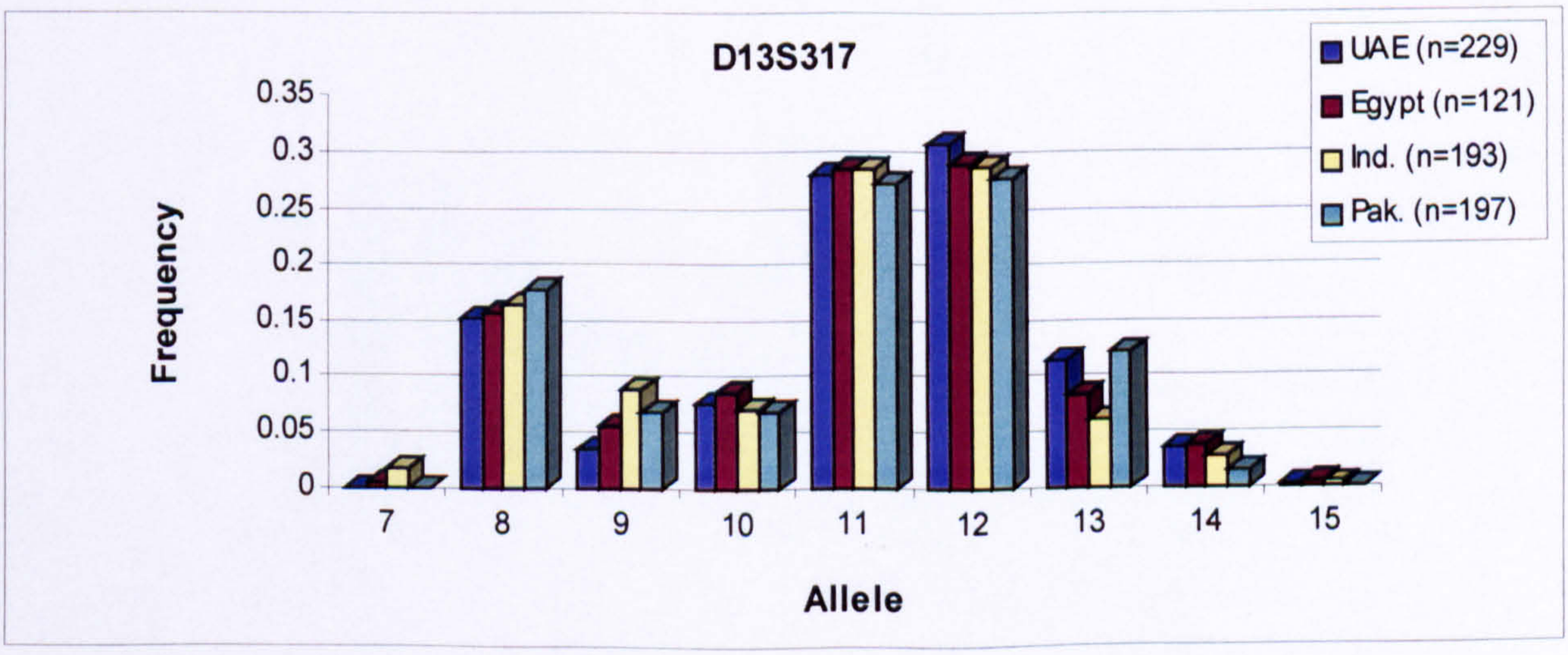
TPOX locus had 7, 7, 5 and 5 different alleles in the UAE, Indian, Pakistani and Egyptian populations respectively in which only four are common. The two most common alleles are allele 8 and 11 which have a combined frequency of 0.72, 0.72, 0.73 and 0.76 in the UAE, Egyptian, Indian and Pakistani populations respectively. Allele 7 had the lowest frequency of 0.002 in the UAE population while allele 14 had the lowest frequency of 0.003 in the Indian population. However, allele 12 had the lowest frequency of 0.028 and 0.025 in the Pakistani and Egyptian populations respectively. Allele 6 and 7 only observed in the UAE population but allele 13 and 14 only observed in the Indian population. Pakistani and Egyptian populations had the lowest number of alleles (only 5). See figure 6.1.

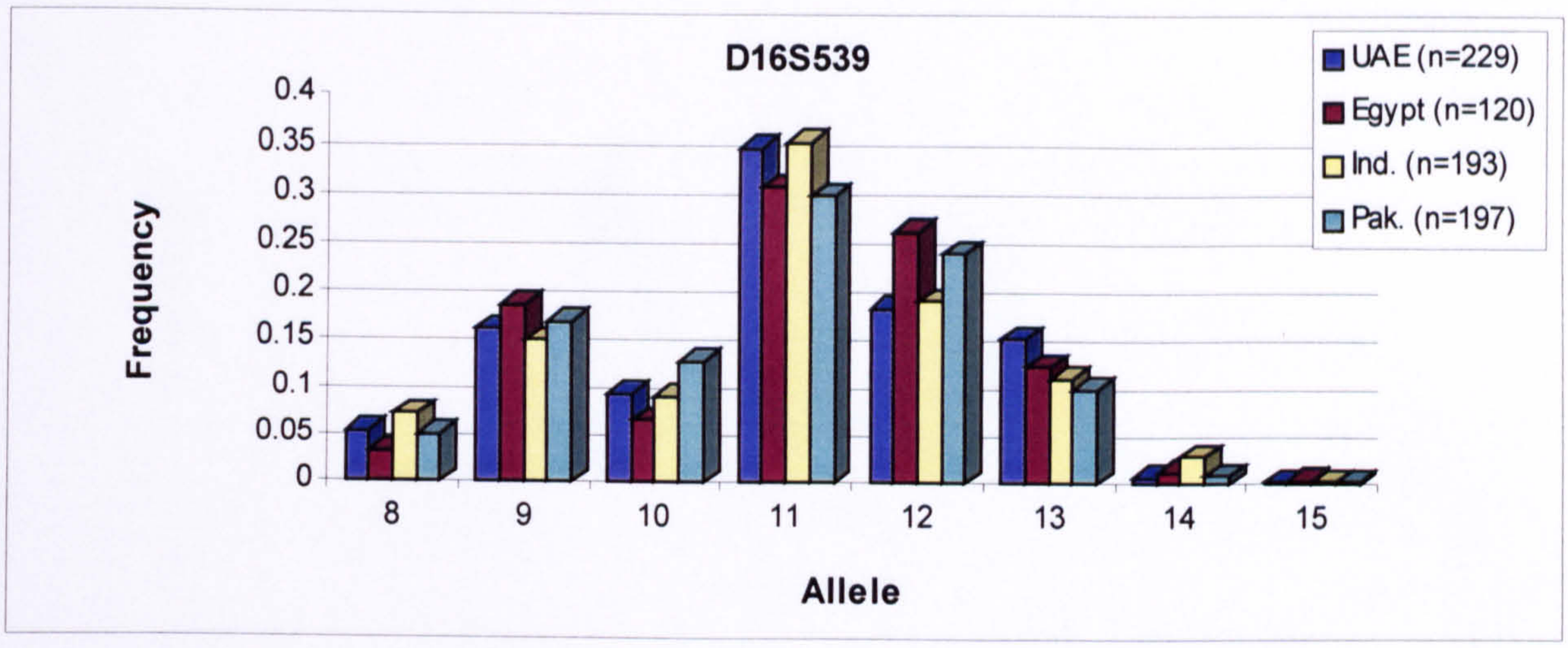
166A

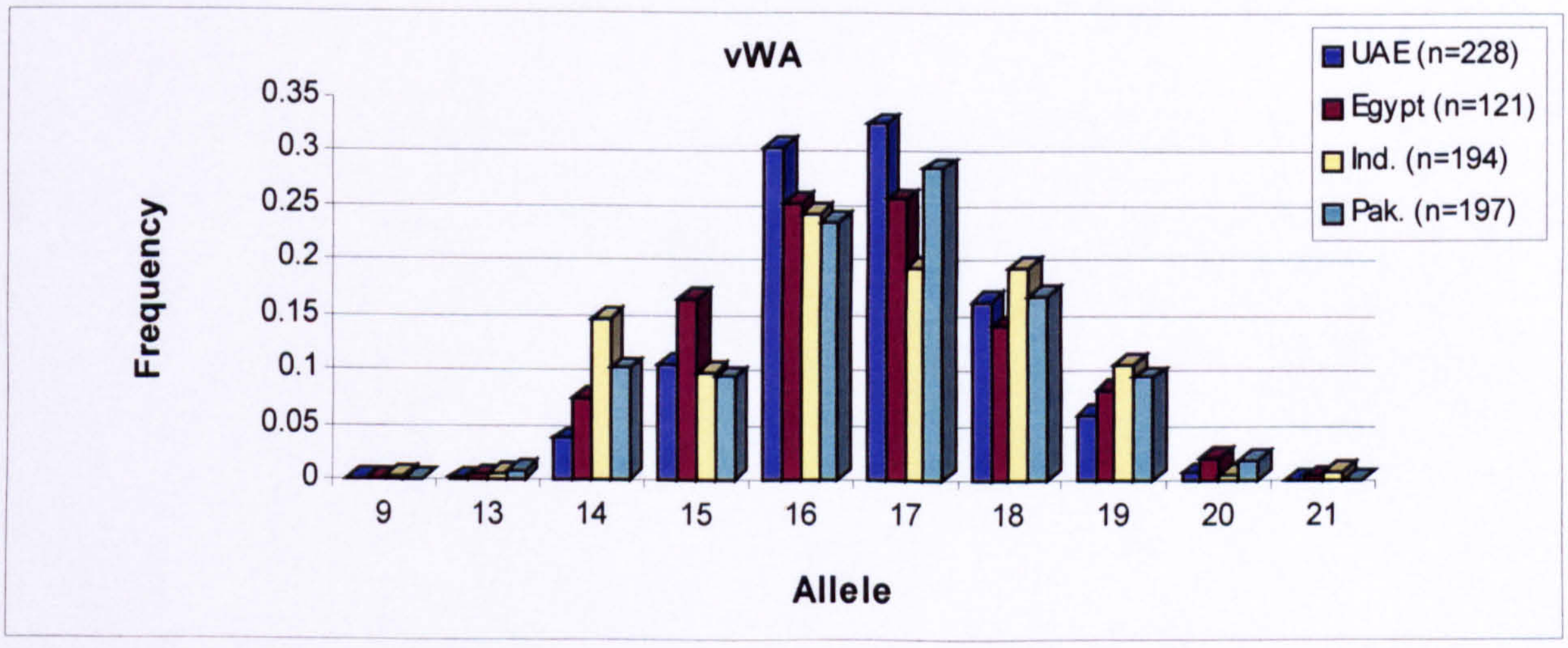
Figure 6.1 Allele frequency distributions for the eight loci in the four populations (UAE, Egyptian, Indian and Pakistani).

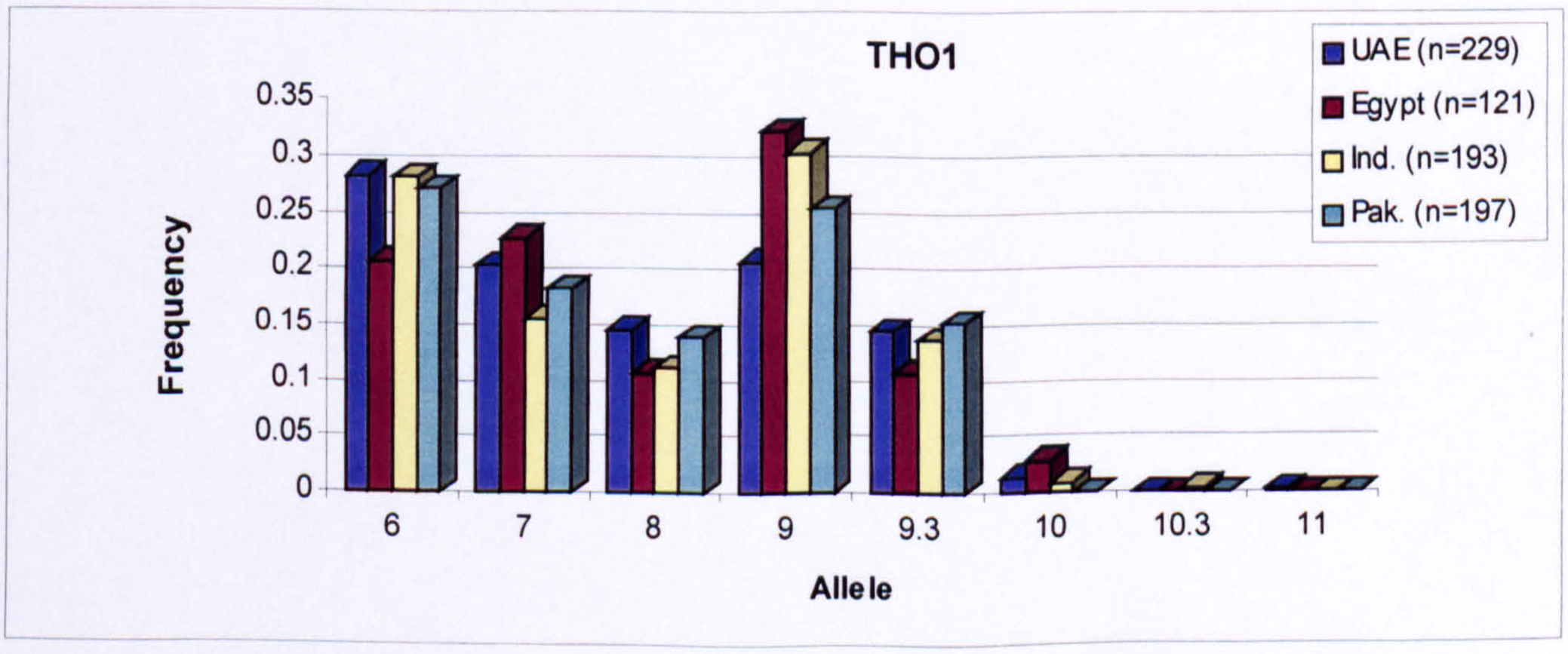

D5S818
UAE (n=229)
Egypt (n=121)
Ind. (n=192)
Pak. (n=197)
Frequency
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
7
8
9
10
11
12
13
14
15
Allele

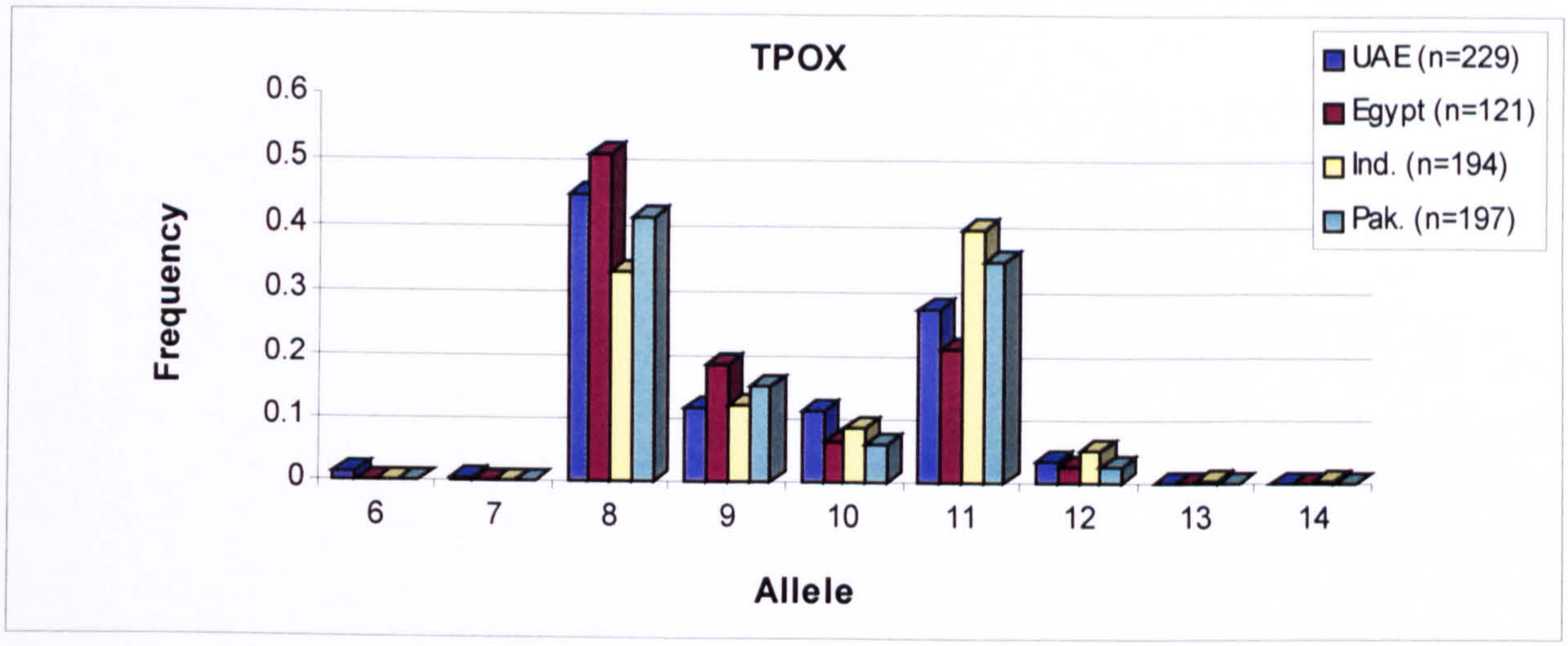

D7S820
UAE (n=229)
Egypt (n=120)
Ind. (n=192)
Pak. (n=197)
Frequency
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
7
8
8.3
9
10
11
12
13
14
Allele

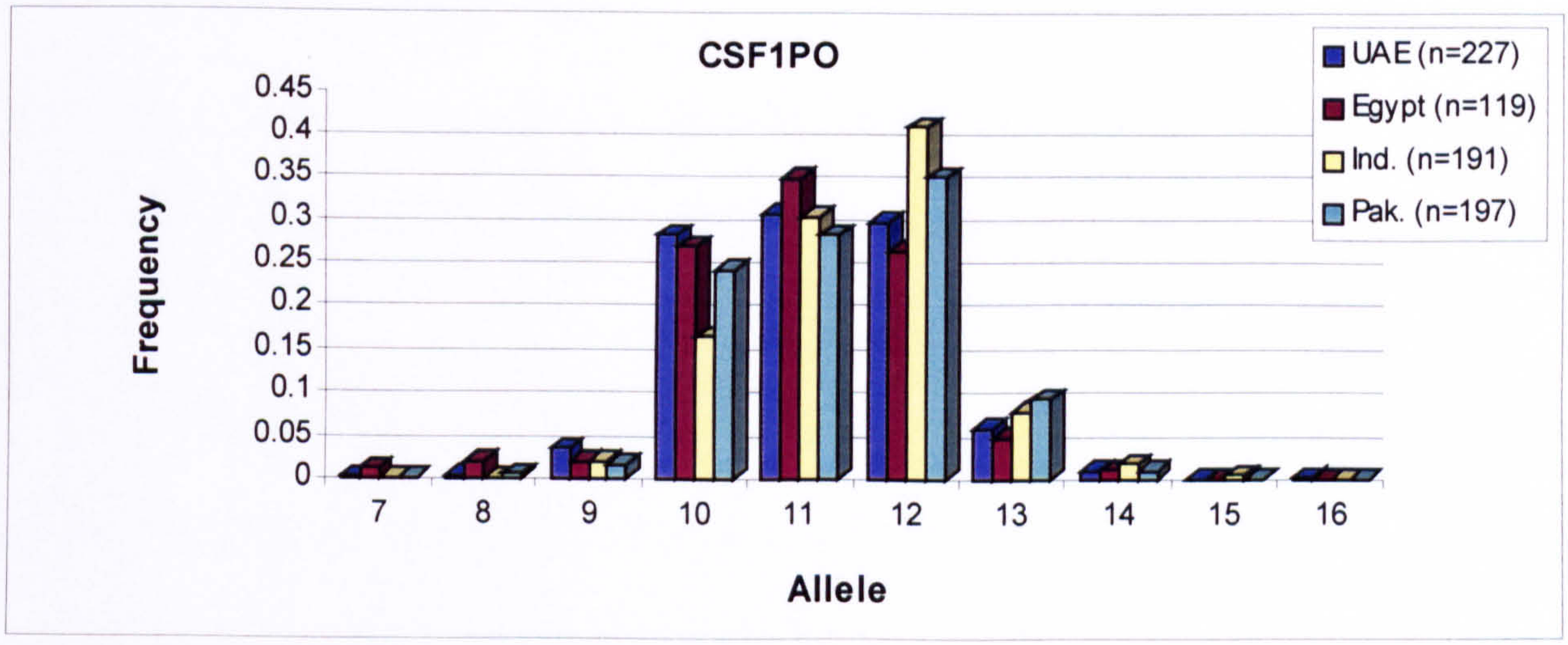

D13S317
UAE (n=229)
Egypt (n=121)
Ind. (n=193)
Pak. (n=197)
Frequency
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
7
8
9
10
11
12
13
14
15
Allele


D16S539
UAE (n=229)
Egypt (n=120)
Ind. (n=193)
Pak. (n=197)
Frequency
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
8
9
10
11
12
13
14
15
Allele


vWA
UAE (n=228)
Egypt (n=121)
Ind. (n=194)
Pak. (n=197)
Frequency
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
9
13
14
15
16
17
18
19
20
21
Allele


THO1
UAE (n=229)
Egypt (n=121)
Ind. (n=193)
Pak. (n=197)
Frequency
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
6
7
8
9
9.3
10
10.3
11
Allele


TPOX
UAE (n=229)
Egypt (n=121)
Ind. (n=194)
Pak. (n=197)
Frequency
0.6
0.5
0.4
0.3
0.2
0.1
0
6
7
8
9
10
11
12
13
14
Allele


CSF1PO
UAE (n=227)
Egypt (n=119)
Ind. (n=191)
Pak. (n=197)
Frequency
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
7
8
9
10
11
12
13
14
15
16
Allele

6.3 GENOTYPE FREQUENCIES

In any genetic marker to be useful for forensic identification, the relative frequencies of the genotypes are important. The total number of possible genotypes expected (theoretically) for the combination of eight loci analysed in the UAE population are 2.07 x 10^{12} (28 x 45 x 36 x 36 x 36 x 28 x 28 x 45). In the Egyptian population there are 8.33 x 10^{11} (36 x 28 x 45 x 36 x 45 x 21 x 15 x 36). In the Indian population there are 2.53 x 10^{12} (36 x 36 x 45 x 36 x 55 x 28 x 28 x 28). In the Pakistani population there are 2.96 x 10^{11} possible genotypes (36 x 36 x 28 x 36 x 36 x 15 x 15 x 28).

The standard deviation (SD) for the observed genotype frequencies was calculated according to the equation (4.2). The standard deviation (SD) for the observed genotypes for all four populations at eight loci is shown in appendix G (section 1.2).

There were 21, 20, 19 and 20 observed genotypes out of 28, 36, 36 and 36 different possible genotypes at D5S818 locus in the UAE, Egyptian, Indian and Pakistani populations respectively. The genotype 12/12 had the highest frequency in the UAE and Egyptian populations and occurred in 38 (16.6%), 18 (14.9%) out of 229 and 121 individuals respectively. The genotype 11/12 had the highest frequency in the Indian and Pakistani populations, occurred in 40 (20.8%) and 46 (23.3%) out of 192 and 197 individuals in the Indian and Pakistani populations respectively.

D7S820 locus had 45, 28, 36 and 36 possible genotypes in the UAE, Egyptian, Indian and Pakistani populations respectively. There were 27, 18, 28 and 22 observed genotypes in the UAE, Egyptian, Indian and Pakistani populations respectively. The genotype 10/11 had the highest frequency in the UAE and Egyptian populations, occurred in 28 (12.2%), 20 (16.7%) out of 229 and 120 individuals respectively. The genotype 8/10

had the highest frequency in the Indian and Pakistani populations, occurred in 25 (13%) and 33 (16.8%) out of 192 and 197 individuals respectively.

Out of 36, 45, 45 and 28 possible genotypes, only 28, 26, 30 and 24 genotypes were observed in the UAE, Egyptian, Indian and Pakistani populations respectively for D13S317 locus. The most common genotypes were 11/12, occurred in 46 (20.1%), 22 (18.2%), 27 (14%) and 37 (18.8%) out of 229, 121, 193 and 197 individuals in the UAE, Egyptian, Indian and Pakistani populations respectively.

D16S539 locus had 36 possible genotypes in the UAE, Egyptian, Indian and Pakistani populations. There were 24, 22, 26 and 23 observed genotypes in the UAE, Egyptian, Indian and Pakistani populations respectively. The genotype 11/11 had the highest frequency in the UAE population, occurred in 31 (13.5%) out of 229 individuals. The genotype 11/12 had the highest frequency in the Egyptian, Indian and Pakistani populations, occurred in 19 (15.8%), 26 (13.5%) and 23 (11.7%) out of 120, 193 and 197 individuals respectively.

There were 25, 25, 26 and 27 genotypes observed out of 36, 45, 55 and 36 possible genotypes in the UAE, Egyptian, Indian and Pakistani populations at vWA locus. The genotypes 16/17, 16/17, 17/18 and 16/17 were the most common and occurred in 42 (18.4%), 14 (11.6%), 20 (10.3%) and 26 (13.2%) out of 227, 121, 194 and 197 individuals in the UAE, Egyptian, Indian and Pakistani populations respectively

THO1 locus had 28, 21, 28 and 15 possible genotypes in the UAE, Egyptian, Indian and Pakistani populations respectively. There were 20, 19, 18 and 15 observed genotypes in the UAE, Egyptian, Indian and Pakistani populations respectively. The genotype 6/6,

7/9, 6/9 and 6/9 had the highest frequency in the UAE, Egyptian, Indian and Pakistani populations, occurred in 25 (10.9%), 20 (16.5%), 26 (13.5%) and 25 (12.7%) out of 229, 121, 193 and 197 individuals respectively.

TPOX locus had 28, 15, 28 and 15 possible genotypes in the UAE, Egyptian, Indian and Pakistani populations respectively. There were 19, 13, 18 and 12 observed genotypes in the UAE, Egyptian, Indian and Pakistani populations respectively. The genotype 8/11 had the highest frequency in the UAE, Indian and Pakistani populations, occurred in 58 (25.3%), 50 (15.8%) and 63 (32%) out of 229, 194 and 197 individuals respectively. However, genotype 8/8 had the highest frequency in the Egyptians occurred in 34 (28.1%) out of 121 individuals.

CSF1PO locus had 45, 36, 28 and 28 possible genotypes in the UAE, Egyptian, Indian and Pakistani populations respectively. There were 18, 22, 17 and 20 observed genotypes in the UAE, Egyptian, Indian and Pakistani populations respectively. The genotype 11/12 had the highest frequency in the UAE, Indian and Pakistani populations, occurred in 49 (21.6%), 41 (21.5%) and 43 (21.8%) out of 227, 191 and 197 individuals respectively. The genotype 10/11 had the highest frequency in the Egyptians and occurred in 25 (21%) out of 119 individuals.

6.4 POPULATION SUBSTRUCTURE

The distribution of allele and genotype frequencies may differ among different sub-population. The UAE consists of several different ethnic groups as mentioned earlier. The UAE native Arab, Indian and Pakistani populations make up approximately 80% of total UAE population. Other 20% consist of people from more than a hundred different

countries. Out of 20% approximately 15% are Arab from several countries. Approximately 5% of Arab are from Egypt. Therefore, Egyptian individuals were chosen in order to represent the Arabic population. These four populations studied only intermix to a limited degree.

6.5 HARDY-WEINBERG EQUILIBRIUM

The Chi square test, exact test and linkage equilibrium was calculated by TFPGA (tools for population genetic analyses) (Miller M., 1997) and GDA (Lewis P. and Zaykin D., 2001) softwares.

6.5.1 Chi-square test

The Chi-square goodness-of-fit test (X^2) was calculated by using equation (4.5). All four populations were tested. This test has been mentioned in chapter 4 (4.5.1).

6.5.2 Exact test

This test has been mentioned in chapter 4 (4.5.2). The exact test was applied on all four populations using 2,000 permutations. There was agreement between the observed and expected genotype values under HWE for the eight loci in the four populations. However, vWA, THO1 and TPOX in the Pakistani population were significantly deviated from HWE using exact test ($p<0.05$). X^2 value of Chi square test and p value of the exact test for the four populations is shown in table 6.2.

Table 6.2 X^2 value of Chi-square test and p value of exact test in the UAE, Egyptian, Indian and Pakistani populations. df = degree of freedom.

UAE	X^2	df	p Chi-square	p Exact test
D5S818	33.4	21	***0.042***	0.201
D7S820	48.1	36	0.086	0.399
D13S317	21.9	28	0.787	0.632
D16S539	23.6	28	0.701	0.327
vWA	32.3	28	0.261	0.089
THO1	35.4	21	***0.025***	0.373
TPOX	24.2	21	0.285	0.231
CSF1PO	48.9	36	0.075	0.211

Indian	X^2	df	p Chi-square	p Exact test
D5S818	18.6	28	0.901	0.689
D7S820	21.1	28	0.819	0.522
D13S317	36	36	0.468	0.229
D16S539	19.2	28	0.893	0.909
vWA	29.9	36	0.754	0.537
THO1	19.6	21	0.544	0.539
TPOX	19.3	21	0.566	0.439
CSF1PO	204.9	21	***0.000***	0.07

Pakistani	X^2	df	p Chi-square	p Exact test
D5S818	18.5	28	0.912	0.867
D7S820	28.9	21	0.116	0.145
D13S317	30.1	21	0.09	0.096
D16S539	26	28	0.573	0.4
vWA	41.5	28	***0.048***	***0.023***
THO1	19.9	10	***0.031***	***0.037***
TPOX	22.3	10	***0.014***	***0.005***
CSF1PO	21.1	21	0.455	0.068

Egyptian	X^2	df	p Chi-square	p Exact test
D5S818	34.4	28	0.188	0.337
D7S820	21.7	21	0.474	0.383
D13S317	41.3	36	0.250	0.115
D16S539	20.2	28	0.857	0.446
vWA	32.8	36	0.62	0.574
THO1	18.5	15	0.238	0.305
TPOX	9.7	10	0.466	0.455
CSF1PO	20.8	28	0.832	0.586

6.5.3 Independence of loci

The linkage disequilibrium was tested using the exact test and 2,000 shuffling for all possible pairs of loci in the four populations. There was no significant ($p>0.05$) pairwise disequilibrium between all loci tested in the Egyptian population. However, the UAE population showed significant pairwise disequilibrium only at three pair of loci (D13S317/CSF1PO, D7S820/D16S539 and D16S539/vWA). The Indian population showed significant pairwise disequilibrium at three pair of loci (D5S818/D7S820, D7S820/CSF1PO and vWA/THO1). The Pakistani population showed maximum number of significant pairwise disequilibrium at 11 pair of loci, probably because of HW disequilibrium at three loci (vWA, THO1 and TPOX). The results can be seen in table 6.3.

Table 6.3 Association between pair of loci has been investigated for the four populations.

The result showed no evidence for correlation between these pair of loci in the Egyptian population. The UAE, Indian and Pakistani populations showed evidence for correlation between 3, 3 and 11 out of 28 pair of loci respectively. The significant departure ($p<0.05$) observed in bold italic.

Locus combination (STRs)	UAE (n=229)	Egyptian (n=121)	Indian (n=194)	Pakistani (n=197)
D5/D13	0.65	0.54	0.58	***0.01***
D5/D7	0.55	0.60	***0.04***	0.63
D5/D16	0.88	0.22	0.05	0.44
D5/vWA	0.10	0.90	0.42	***0.02***
D5/THO1	0.33	0.91	0.84	0.48
D5/TPOX	0.15	0.30	0.25	0.70
D5/CSF1PO	0.20	0.99	0.29	0.58
D13/D7	0.06	0.11	0.27	0.73
D13/D16	0.91	0.38	0.25	0.36
D13/vWA	0.23	0.24	0.29	0.45
D13/THO1	0.05	0.40	0.45	0.85
D13/TPOX	0.12	0.92	0.34	***0.01***
D13/CSF1PO	***0.002***	0.82	0.20	***0.01***
D7/D16	***0.001***	0.93	0.38	0.62
D7/vWA	0.40	0.42	0.13	***0.01***
D7/THO1	0.05	0.76	0.56	0.12
D7/TPOX	0.67	0.11	0.33	***0.001***
D7/CSF1PO	0.56	0.20	***0.01***	***0.02***
D16/vWA	***0.02***	0.88	0.37	***0.01***
D16/THO1	0.24	0.70	0.58	***0.01***
D16/TPOX	0.23	0.22	0.05	0.32
D16/CSF1PO	0.44	0.38	0.39	0.06
vWA/THO1	0.24	0.32	***0.003***	***0.01***
vWA/TPOX	0.70	0.55	0.18	0.42
vWA/CSF1PO	0.10	0.36	0.52	0.13
THO1/TPOX	0.29	0.09	0.70	0.57
THO1/CSF1PO	0.16	0.19	0.86	***0.03***
TPOX/CSF1PO	0.51	0.79	0.61	0.57

6.6 HETEROZYGOSITY

The observed and expected heterozygosity was calculated using TFPGA (tools for population genetic analyses) (Miller M., 1997) and GDA (Lewis P. and Zaykin D., 2001) softwares. The results showed that the expected number of heterozygosity was not significantly different from the observed ones (p value is > 0.05). The greatest difference was observed in the UAE population at locus vWA where the observed number of heterozygotes was 69% whereas the expected number was 76%. The greatest difference was observed in the Egyptian population at locus TPOX where the observed number of heterozygotes was 61% whereas the expected number was 66%. In the Indian population the greatest difference was observed at CSF1PO locus where the observed number of heterozygotes was 65% where the expected number was 71%. However, the greatest difference was observed in the Pakistani population at locus THO1 where the observed number of heterozygotes was 70% whereas the expected number was 79%.

All the three populations at the five loci showed observed heterozygosity values equal or higher than 61%. The highest value being observed in the UAE and Egyptian populations at D13S317 (79% and 81% respectively), Indian population at vWA (82%) and Pakistani population at D7S820 (82%). The lowest value was observed in the UAE population at vWA locus (69%) and in the Egyptian population was at TPOX locus (61%). The lowest value observed in the Indian and Pakistani populations were at CSF1PO locus (65%). The standard deviation of the expected heterozygosity was calculated by using equation (4.4). The results are given in table (6.4).

Table 6.4 The observed and expected heterozygosity at eight STR loci in the UAE, Indian, Pakistani and Egyptian populations (number of heterozygote alleles is in brackets next to the frequency). The standard deviation (SD) and the p value from Chi square test were calculated. No significant differences were detected ($p > 0.05$).

	Locus	Observed heterozygosity	Expected heterozygosity	SD
UAE (n = 229)	D5S818	0.6987 **(160)**	0.7438 **(170)**	0.029 **(6.6)**
	D7S820	0.7598 **(174)**	0.7977 **(183)**	0.027 **(6.1)**
	D13S317	0.7860 **(180)**	0.7829 **(179)**	0.027 **(6.2)**
	D16S539	0.7686 **(176)**	0.7848 **(180)**	0.027 **(6.2)**
	VWA (n = 228)	0.6930 **(158)**	0.7607 **(173)**	0.028 **(6.4)**
	THO1	0.7642 **(175)**	0.7928 **(182)**	0.027 **(6.1)**
	TPOX	0.7031 **(161)**	0.6956 **(159)**	0.03 **(7)**
	CSF1PO (n = 227)	0.7577 **(172)**	0.7334 **(166)**	0.029 **(6.7)**
Indian (n = 194)	D5S818 (n = 192)	0.7448 **(143)**	0.7482 **(144)**	0.031 **(6)**
	D7S820 (n = 192)	0.7813 **(150)**	0.8094 **(155)**	0.028 **(5.4)**
	D13S317 (n = 193)	0.7617 **(147)**	0.7915 **(153)**	0.029 **(5.6)**
	D16S539 (n = 193)	0.8083 **(156)**	0.7898 **(152)**	0.029 **(5.7)**
	vWA	0.8196 **(159)**	0.8241 **(160)**	0.027 **(5.3)**
	THO1 (n = 193)	0.7668 **(148)**	0.7729 **(149)**	0.030 **(5.8)**
	TPOX	0.6701 **(130)**	0.7066 **(137)**	0.033 **(6.3)**
	CSF1PO (n = 191)	0.6492 **(124)**	0.7068 **(135)**	0.033 **(6.3)**
Pakistan (n = 197)	D5S818	0.7614 **(150)**	0.7462 **(147)**	0.031 **(6.1)**
	D7S820	0.8223 **(162)**	0.8012 **(158)**	0.028 **(5.6)**
	D13S317	0.7970 **(157)**	0.7937 **(156)**	0.029 **(5.7)**
	D16S539	0.7310 **(144)**	0.7941 **(156)**	0.029 **(5.7)**
	VWA	0.8122 **(160)**	0.8083 **(159)**	0.028 **(5.5)**
	THO1	0.6954 **(137)**	0.7858 **(155)**	0.029 **(5.8)**
	TPOX	0.7513 **(148)**	0.6824 **(134)**	0.033 **(6.5)**
	CSF1PO	0.6497 **(128)**	0.7306 **(144)**	0.032 **(6.2)**
Egyptian (n = 121)	D5S818	0.7355 **(89)**	0.7574 **(92)**	0.039 **(4.7)**
	D7S820 (n = 120)	0.7417 **(89)**	0.7623 **(91)**	0.039 **(4.7)**
	D13S317	0.8099 **(98)**	0.7917 **(96)**	0.037 **(4.5)**
	D16S539 (n = 120)	0.7500 **(90)**	0.7795 **(94)**	0.038 **(4.5)**
	VWA	0.7934 **(96)**	0.8109 **(98)**	0.036 **(4.3)**
	THO1	0.7355 **(89)**	0.7778 **(94)**	0.038 **(4.6)**
	TPOX	0.6116 **(74)**	0.6559 **(79)**	0.043 **(5.2)**
	CSF1PO (n = 119)	0.6891 **(82)**	0.7323 **(87)**	0.041 **(4.8)**

6.7 COMPARITIVE STUDY

The RxC contingency test and F-statistics were used to detect for population differentiation. Comparisons of the data results obtained in this study from the four populations (UAE, Egyptian, Indian and Pakistani) were performed for eight STR loci. The results from the two UAE native Arabic populations (Abu Dhabi, Sharjah) were also compared.

6.7.1 RxC test

The RxC contingency table was used to calculate the p value of the exact test to compare the allele frequencies of the four populations. The comparison was performed for the eight loci to examine the presence of heterogeneity between the four populations studied. There were significant differences between the UAE and Indian populations at all loci except D16S539 locus. The results showed that there were also significant differences between the UAE and Pakistani populations but only at three loci which are D5S818, vWA and TPOX. Other significant differences between the Indian and Pakistani populations were observed at D7S820, D13S317 and TPOX.

There were significant differences between Egyptian and UAE populations at three loci. There were significant differences between Egyptian and Pakistani populations at four loci. There were significant differences between Egyptian and Indian populations at six loci. The RxC results are shown in table 6.5.

Table 6.5 The p values and standard errors of the exact test for the RxC contingency table for the compared allele frequencies in the four populations at eight loci. 20 batches of 2,500 replicates per batch performed. The total number of replicates was 50,000.

Locus	UAE/Pak.	SE	UAE/Ind.	SE	Ind./Pak.	SE
D5S818	***0.0131***	0.0043	***0.0434***	0.0092	0.1691	0.0231
D7S820	0.1931	0.0189	***0.0034***	0.0020	***0.0400***	0.0101
D13S317	0.1659	0.0192	***0.0003***	0.0003	***0.0129***	0.0036
D16S539	0.0606	0.0111	0.1293	0.0181	0.0552	0.0091
vWA	***0.0003***	0.0002	***0.0000***	0.0000	0.0558	0.0125
THO1	0.1485	0.0175	***0.0261***	0.0107	0.1779	0.0219
TPOX	***0.0016***	0.0008	***0.0000***	0.0000	***0.0117***	0.0032
CSF1PO	0.1290	0.0171	***0.0003***	0.0002	0.0826	0.0131

Locus	Egy. /UAE	SE	Egy. /Ind.	SE	Egy. /Pak.	SE
D5S818	0.155	0.016	***0.002***	0.001	***0.001***	0.0003
D7S820	0.171	0.024	***0.002***	0.001	0.055	0.009
D13S317	0.704	0.029	0.548	0.028	0.210	0.028
D16S539	0.136	0.025	0.059	0.009	0.256	0.028
vWA	***0.011***	0.005	***0.004***	0.001	0.207	0.023
THO1	***0.001***	0.001	***0.028***	0.007	***0.0002***	0.0001
TPOX	***0.011***	0.003	***0.000***	0.000	***0.009***	0.004
CSF1PO	0.239	0.028	***0.000***	0.000	***0.008***	0.002

The RxC contingency test was used to compare the allele frequencies of the two UAE native Arabic populations. The comparison was performed for the eight loci to examine the presence of heterogeneity between Sharjah and Abu Dhabi populations. There were no significant differences between the two UAE populations at all loci except THO1 locus. These two populations of the UAE compared to a third UAE population (Al Ain) at three loci (THO1, TPOX and CSF1PO) which was published earlier (Bayoumi R.,

et al., 1997). There was only one significant difference at TPOX locus. These three cities are separated by approximately 160 km. The results are shown in table 6.6.

Table 6.6 The p values and standard errors of the exact test for the RxC contingency table for the compared allele frequencies in three UAE native Arabic populations. 20 batches of 2,500 replicates per batch performed. The total number of replicates was 50,000.

Locus	Sha/Abu	SE	Sha/Al Ain	SE	Abu/Al Ain	SE
D5S818	0.883	0.010	-	-	-	-
D7S820	0.082	0.010	-	-	-	-
D13S317	0.464	0.022	-	-	-	-
D16S539	0.322	0.022	-	-	-	-
vWA	0.144	0.013	-	-	-	-
THO1	***0.004***	0.001	0.211	0.016	0.140	0.018
TPOX	0.552	0.016	***0.0001***	0.0001	***0.044***	0.007
CSF1PO	0.135	0.015	0.057	0.014	0.987	0.002

6.7.2 *F*-statistics

The F-statistics (F_{ST}) is considered to be a useful measure of population substructure (Hartl D. and Clark A., 1997). Therefore, in order to test the presence of substructure between the four populations (UAE, Egyptian, Indian and Pakistani) and within the UAE native populations (Sharjah and Abu Dhabi), the genotypes of all populations were compared using *F*-statistics (Lewis P. and Zaykin D., 2001). The F-statistics analysis has been mentioned in chapter 4 (4.6.2).

The overall value of F_{ST} over the eight loci was (0.0032) when compared between the UAE Arab and Egyptian populations. This indicates that there is very little genetic differentiation between these two Arabic populations. The value of F_{ST} was (0.0029)

when compared between the UAE and Pakistani populations. The value of F_{ST} was (0.0089) when compared between the UAE and Indian populations. The value of F_{ST} was (0.0021) when compared between the Pakistani and Indian populations. The value of F_{ST} was (0.0063) when compared between the Egyptian and Pakistani populations. The value of F_{ST} was (0.013) when compared between the Egyptian and Indian populations.

The maximum genetic differentiation was observed between the UAE and Indian populations (F_{ST} = 0.009) and between Egyptian and Indian populations (F_{ST} = 0.013). The results are shown in table 6.7.

Table 6.7 F-statistics values of all six possible pair of populations, UAE/Egyptian, UAE/Pakistani, UAE/Indian, Indian/Pakistani, Egyptian/Indian and Egyptian/Pakistani populations.

Compared population	F_{ST} value
UAE/Egyptian	0.0031
UAE/Pakistani	0.0029
UAE/Indian	0.0089
Indian/Pakistani	0.0021
Egyptian/Pakistani	0.0063
Egyptian/Indian	0.013

The bootstrapping over the eight STR loci for all pair of different populations was applied in order to generate 95% confidence intervals. The results are shown in table 6.8.

Table 6.8 95% confidence interval bootstrapping over the eight STR loci for the UAE/Egyptian, UAE/Pakistani, UAE/Indian, Indian/Pakistani Egyptian/Indian and Egyptian/Pakistani populations.

Compared population	Lower bound	Upper bound
UAE/Egyptian	0.0006	0.0064
UAE/Pakistani	0.001	0.0049
UAE/Indian	0.0039	0.0149
Indian/Pakistani	0.0009	0.0035
Egyptian/Pakistani	0.0022	0.0110
Egyptian/Indian	0.0054	0.0231

The overall value of F_{ST} over the eight loci was extremely low (- 0.0004) when compared Abu Dhabi to Sharjah. This indicates that there is no significant genetic differentiation between the UAE native populations and also indicates that the mating between people is probably random. Therefore, it is possible to use only one general database for all UAE Arabic native population. The result can be seen in table 6.9.

Table 6.9 F-statistics parameters and value between the UAE native populations, Sharjah and Abu Dhabi.

F-statistics parameters	F-statistics values
F_{ST}	- 0.0004
F_{IT}	0.0279
F_{IS}	0.0283

The bootstrapping over the eight loci was applied and it showed that the inbreeding coefficients F_{IS} and F_{IT} were not statistically significantly different from zero. Because the 95% bootstrap confidence interval for F_{ST} value overlap zero, therefore, the co-ancestry coefficient is not significant and thus a non significant amount of divergence had occurred between the two populations within UAE native population. The results can be seen in table 6.10.

Table 6.10 95% confidence interval bootstrapping over the eight STR loci in the UAE native populations, Sharjah and Abu Dhabi. No significant amount of divergence has occurred between Sharjah and Abu Dhabi populations.

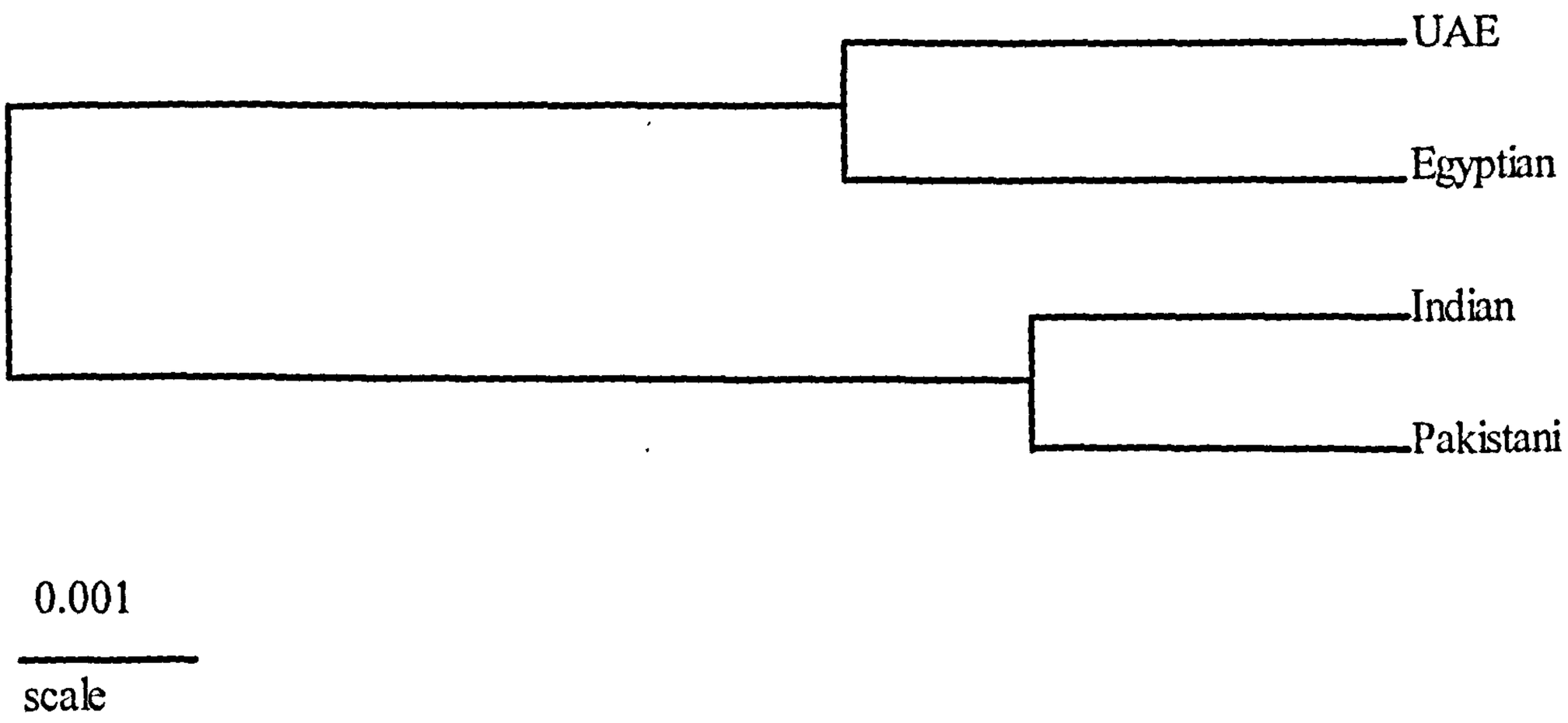
F-statistics parameter	Lower bound	Upper bound
F_{ST}	- 0.0028	0.0024
F_{IT}	0.0014	0.0544
F_{IS}	0.0038	0.0522

6.8 PHYLOGENETIC RELATIONSHIP BETWEEN POPULATIONS

The phylogenetic relationship between the UAE, Egyptian, Indian and Pakistani populations, based on co-ancestry distance measures for 8 STR loci, was assessed by drawing an UPGMA (unweighted pair group method using arithmetic averaging) phylogenetic tree using GDA software. From the tree (figure 6.2) it can be concluded as would be expected that the Indian and Pakistani populations have a closer relationship as compared to other two Arabic populations and vice versa. Similar relationships between the UAE, Indian and Pakistani populations were observed by using 5 VNTR loci in chapter 4 (figure 4.2).

184A

Figure 6.2 The UPGMA phylogenetic tree showing the relationship between the four populations.


UAE
Egyptian
Indian
Pakistani
0.001
scale

6.9 FORENSIC AND PATERNITY PARAMETERS

6.9.1 Forensic parameters

As mentioned in chapter 4 (section 4.9.1). The matching probability (MP) and probability of discrimination (PD) were calculated for the eight STR loci in the four populations. In addition, the combined probability of discrimination (PD) and the combined of matching probability (MP) were calculated. See table 6.11.

Tables 6.11 Forensic and paternity parameters for the eight STR loci in the UAE, Egyptian, Indian and Pakistani populations. MP = Matching Probability, PD = Power of Discrimination, TPI = Typical Paternity Index and PE = Paternity of Exclusion.

UAE Pop.	MP	PD	Exp. as 1 in	TPI	PE
D5S818	0.104	0.896	9.6	1.66	0.426
D7S820	0.069	0.931	14.5	2.08	0.527
D13S317	0.086	0.914	11.6	2.34	0.573
D16S539	0.078	0.922	12.7	2.16	0.542
vWA	0.095	0.905	10.5	1.62	0.415
THO1	0.075	0.925	13.4	2.12	0.534
TPOX	0.140	0.860	7.1	1.68	0.433
CSF1PO	0.126	0.874	8.0	2.06	0.523
Combined	0.000000006	0.999999994	1.64×10^8	207	0.996

Egyp. Pop.	MP	PD	Exp. as 1 in	TPI	PE
D5S818	0.093	0.907	10.8	1.89	0.485
D7S820	0.096	0.904	10.4	1.94	0.496
D13S317	0.087	0.913	11.6	2.63	0.618
D16S539	0.084	0.916	11.9	2.00	0.510
vWA	0.063	0.937	15.8	2.42	0.587
THO1	0.087	0.913	11.5	1.89	0.485
TPOX	0.170	0.830	5.9	1.32	0.315
CSF1PO	0.120	0.880	8.4	1.61	0.412
Combined	0.000000007	0.999999993	1.4×10^8	188	0.996

Ind. Pop.	MP	PD	Exp. as 1 in	TPI	PE
D5S818	0.108	0.892	9.3	1.96	0.501
D7S820	0.066	0.934	15.2	2.29	0.565
D13S317	0.073	0.927	13.7	2.10	0.530
D16S539	0.073	0.927	13.7	2.61	0.615
vWA	0.057	0.943	17.4	2.77	0.636
THO1	0.087	0.913	11.5	2.14	0.539
TPOX	0.136	0.864	7.4	1.52	0.384
CSF1PO	0.130	0.870	7.7	1.43	0.354
Combined	0.000000003	0.999999997	3.03×10^8	317	0.997

Pak. Pop.	MP	PD	Exp. as 1 in	TPI	PE
D5S818	0.109	0.891	9.2	2.10	0.530
D7S820	0.078	0.922	12.8	2.81	0.641
D13S317	0.082	0.918	12.2	2.46	0.593
D16S539	0.070	0.930	14.3	1.86	0.478
vWA	0.070	0.930	14.3	2.66	0.622
THO1	0.079	0.921	12.7	1.64	0.421
TPOX	0.176	0.824	5.7	2.01	0.512
CSF1PO	0.119	0.881	8.4	1.43	0.355
Combined	0.000000006	0.999999994	1.78×10^8	339	0.998

6.9.1.1 Matching probability

The matching probability (calculated by using equation 4.6) of the eight loci tested ranged from 0.069 for D7S317 locus to 0.14 for TPOX locus in the UAE population. In the Egyptian population it is ranged from 0.063 for vWA to 0.17 for TPOX. It is ranged from 0.057 for vWA to 0.136 for TPOX in the Indian population. In the Pakistani population the matching probability is ranged from 0.07 for D16S539 and vWA loci to 0.176 for TPOX locus. The combined matching probability for the eight loci in the UAE population was 0.000000006. This value can be expressed as the likelihood ratio (LR), which equals to one over the matching probability (LR = 1/MP). Therefore, the chance of occurrence of the same profile for the eight loci was 1.64×10^8 in the UAE population. For other three populations see table 6.11.

6.9.1.2 Discrimination power

See section (4.9.1.2). The power of discrimination and the combined power of discrimination were calculated using equations 4.7 and 4.8. Among the four populations tested the Indian population had the highest value of discrimination power. See table 6.11.

6.9.2 Paternity parameters

See section (4.9.2). The power of exclusion (PE) and typical paternity index (TPI) were calculated for the eight loci in the four populations tested. See table 6.11

6.9.2.1 Power of exclusion

Power of exclusion and combined power of exclusion was calculated using equations 4.9 and 4.10. See table 6.11.

The D7S820 locus had the highest value of power of exclusion (0.641) in the Pakistani population while vWA locus had the highest value of power of exclusion (0.636) in the Indian population. The vWA locus had the highest value of power of exclusion (0.573) in the UAE population while D13S317 locus had the highest value of power of exclusion (0.618) in the Egyptian population. The CSF1PO locus in the Indian and Pakistani populations had the lowest value of power of exclusion, 0.354 and 0.355 respectively. The lowest value of power of exclusion was observed in the UAE and Egyptian populations at D5S818 (0.426) and TPOX (0.315) loci respectively.

The Indian and Pakistani populations had the highest value of combined power of exclusion 0.997 and 0.998 respectively. The combined power of exclusion in the UAE and Egyptian populations was 0.996. The combined exclusion power of 99.6% in the UAE population indicates that it is expected that 1 in 250 of non-fathers tested with a mother and child would fail to show at least one mismatch with paternal allele in the child.

6.9.2.2 Typical paternity index

The typical paternity index (equation 4.11) for the eight loci tested ranged from 1.62, 1.32, 1.43 and 1.43 for vWA, TPOX, CSF1PO, CSF1PO loci to 2.34, 2.63, 2.77 and 2.81 for D13S317, D13S317, vWA and D7S820 loci in the UAE, Egyptian, Indian and Pakistani populations respectively. The eight loci showed a combined paternity index of 207, 188, 317 and 339 in the UAE, Egyptian, Indian and Pakistani populations respectively. The probability of paternity for the eight loci was 99.52%, 99.47%, 99.69% and 99.71% assuming a prior probability of guilt of 50% in the UAE, Egyptian, Indian and Pakistani populations respectively.

6.10 COMBINING ALL VNTR AND STR LOCI

The combined matching probability, combined typical paternity index and combined power of exclusion for the five VNTR loci plus the eight STR loci were calculated. The probability of paternity in the UAE, Indian and Pakistani populations is 99.998%, 99.9997% and 99.9998% respectively. See appendix H for examples on simulated forensic and paternity cases. The results are shown in table 6.12.

Table 6.12 Combining the results of all five VNTR loci plus eight STR loci in the UAE, Indian and Pakistani populations.

Population	MP	Exp. as 1 in	TPI	PE
UAE	1.38×10^{-16}	7.4×10^{15}	66,323	0.99999
Indian	3.6×10^{-17}	1.1×10^{16}	382,844	0.999998
Pakistani	1.9×10^{-16}	5×10^{15}	438,488	0.999998

6.11 COMPARISON BETWEEN POPULATIONS

The results of allele frequency distribution, RxC contingency test, forensic and paternity parameters acquired from the UAE Arabic population in this study were compared to other Arab and major world populations.

6.11.1 Comparison with Arabic populations

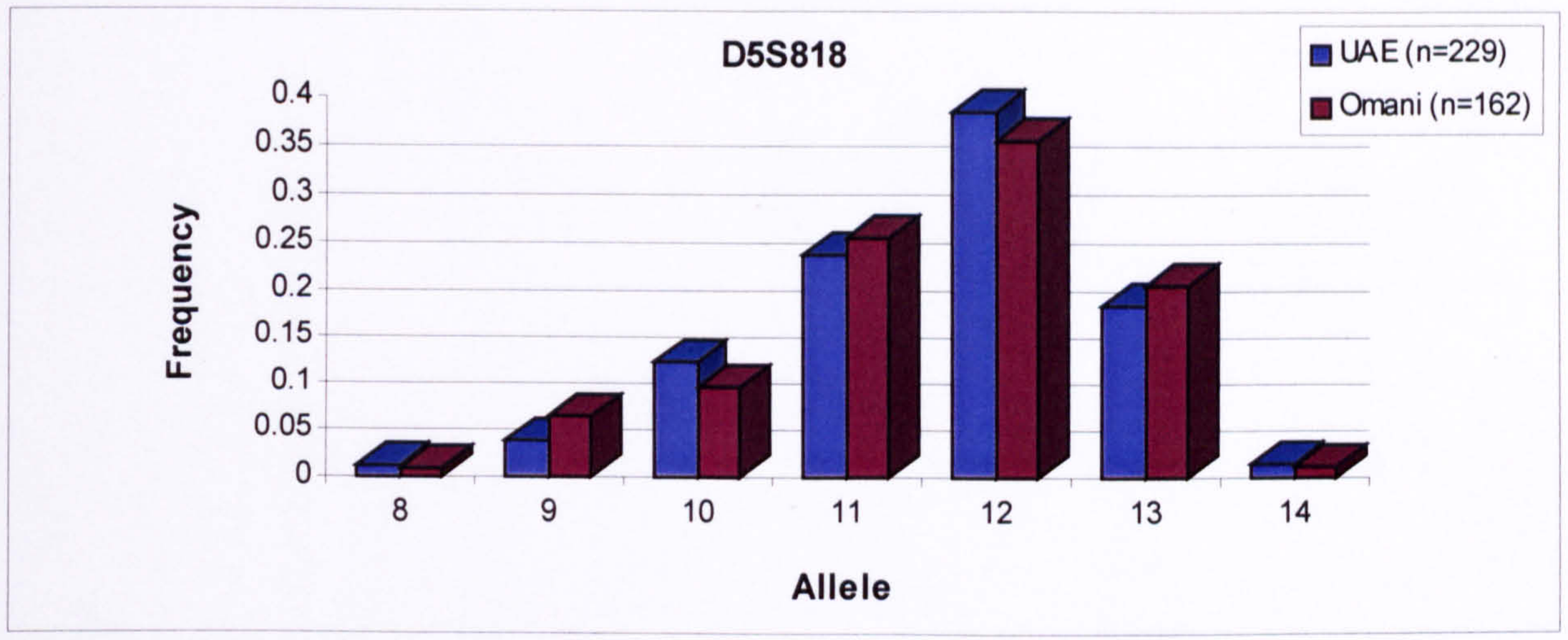
6.11.1.1 Allele frequency

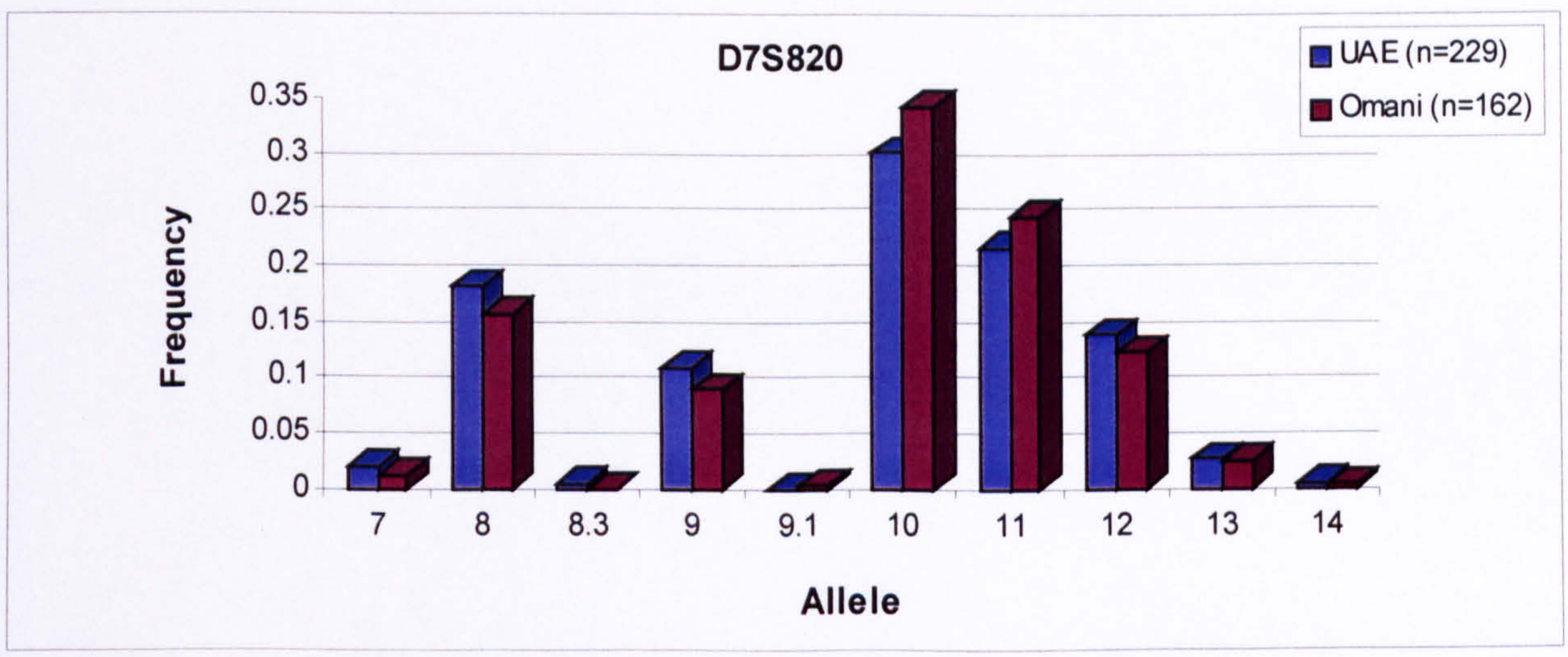
The allele frequency distribution from the UAE population compared to Omani, Saudi, Qatari, Yemeni and Moroccan populations (Ahmed A. *et al.*, 2000; Mohammed A. *et al.*, 2000; Tahir M. *et al.*, 2000; Sinha S. *et al.*, 1999; Sebetan I. and Hajar H., 1998; Klintschar M. *et al.*, 1998a, b; Brandt-Casadevall C. *et al.*, 2000; Takeshita H. *et al.*, 1997).

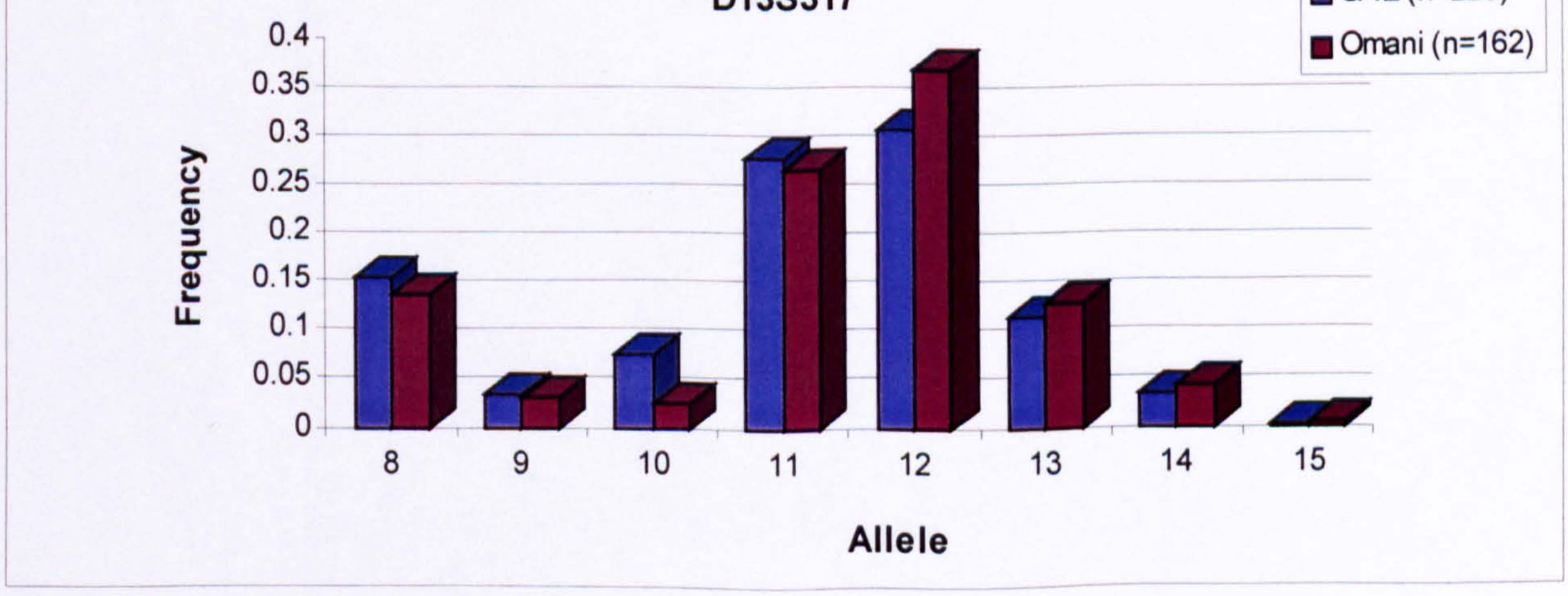
Overall, the allele frequency distribution observed from the UAE samples was similar to almost all other Arabic populations with some interesting features. For example, allele 14 at vWA is observed at approximately three times higher in the Moroccan population (0.102) compared to UAE population (0.037). Allele 6 at THO1 is observed at a lower frequency (half) in the Moroccan population (0.168) compared to Saudi Arabian population (0.335). Other example is allele 15 at vWA which is observed at a more than twice higher frequency in the Saudi population (0.175) compared to Yemeni population (0.075). The results are shown in figure 6.3.

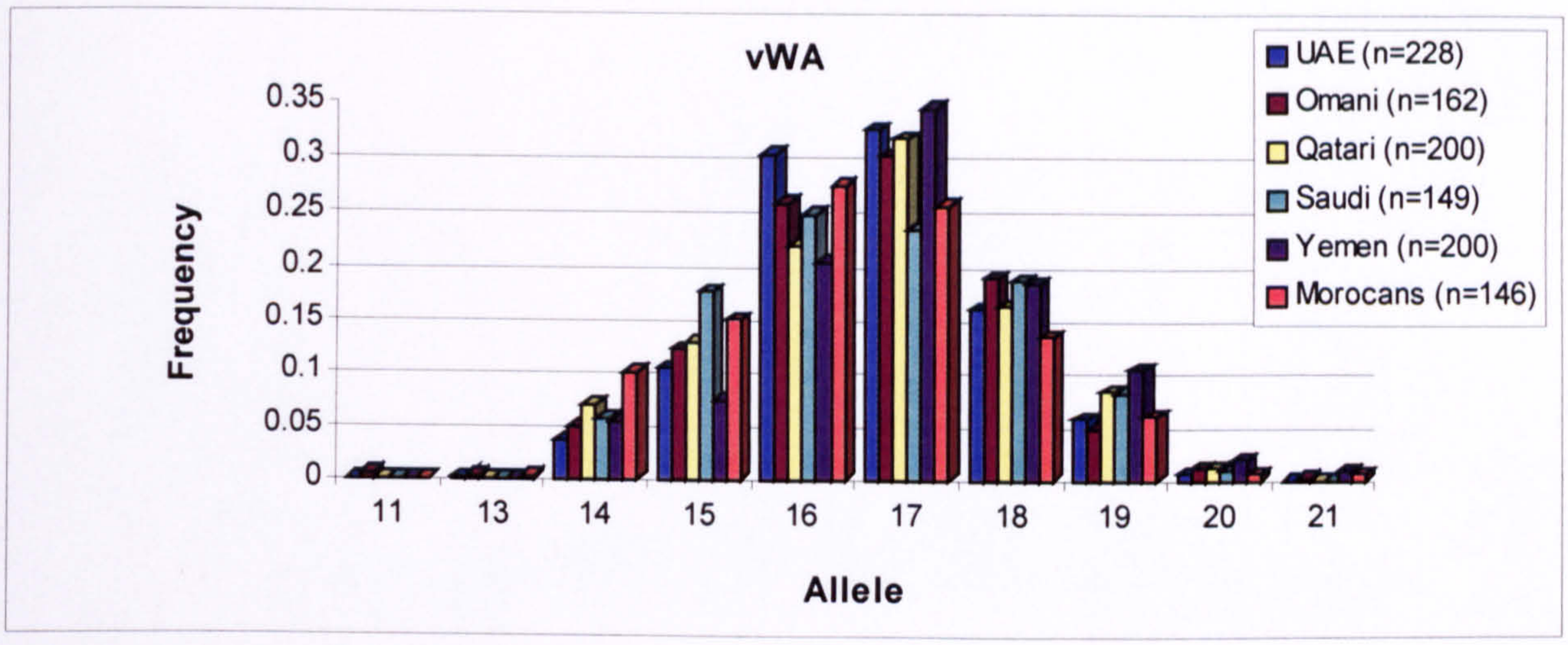
191 A

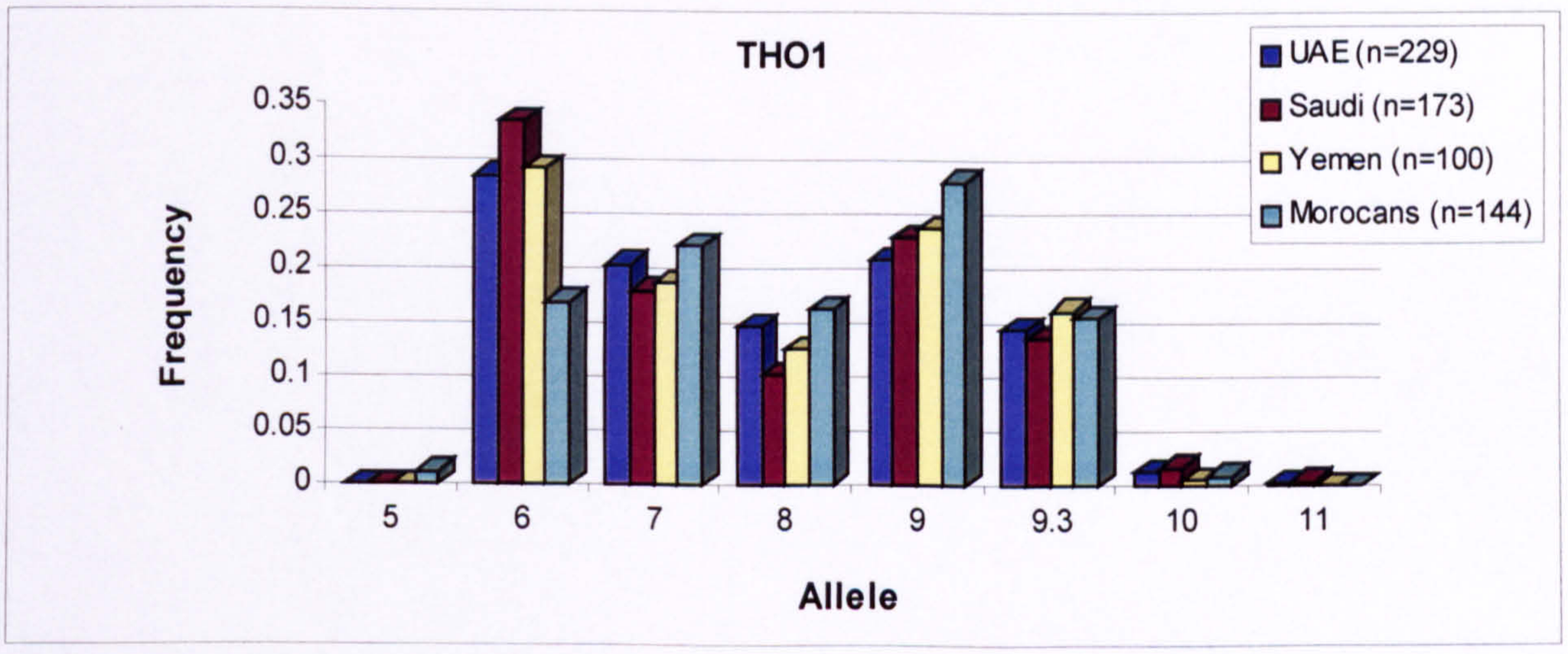
Figure 6.3 Allele frequencies in the UAE population compared to other Arabic populations.

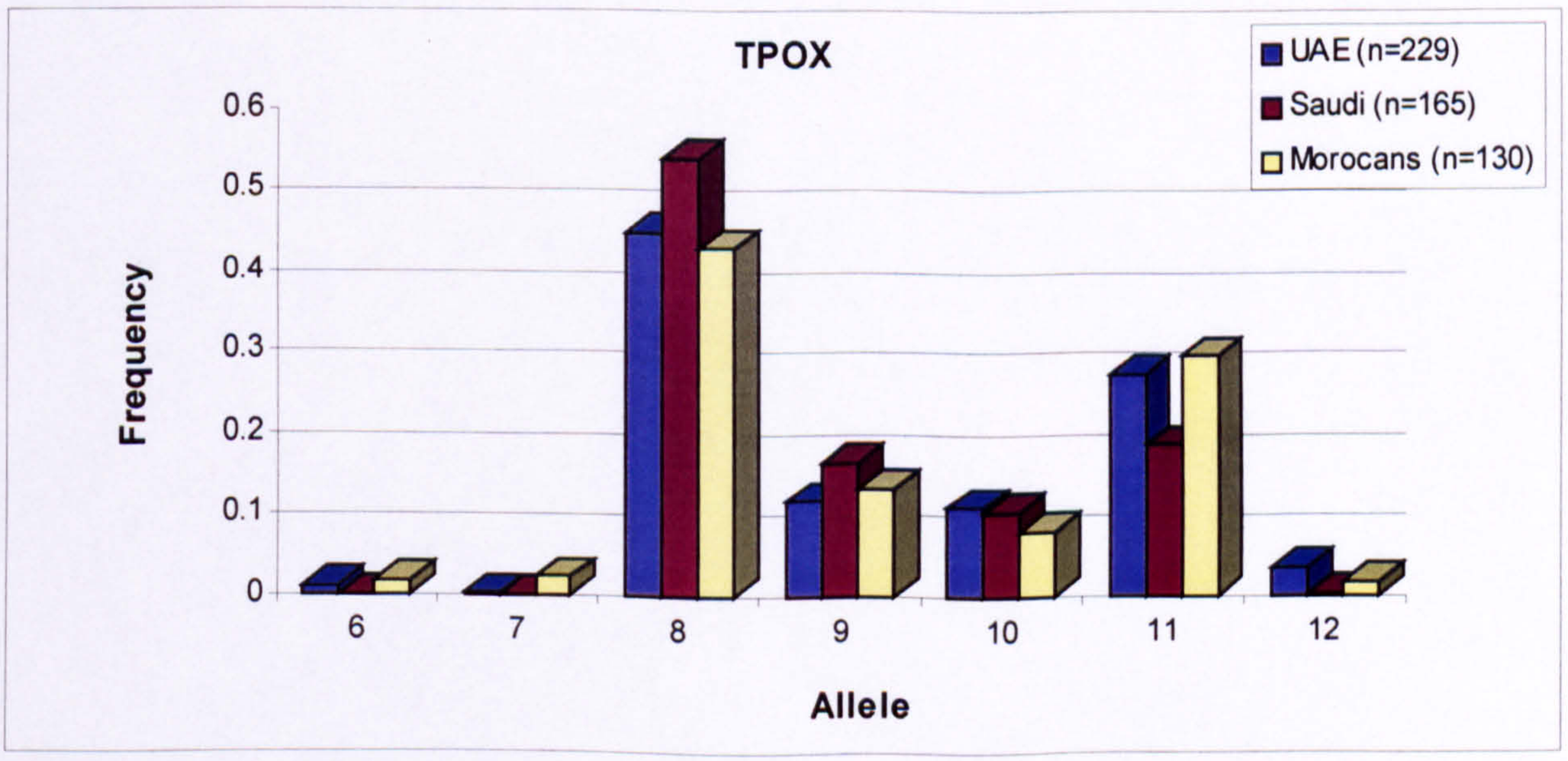

D5S818
UAE (n=229)
Omani (n=162)
Frequency
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
8
9
10
11
12
13
14
Allele

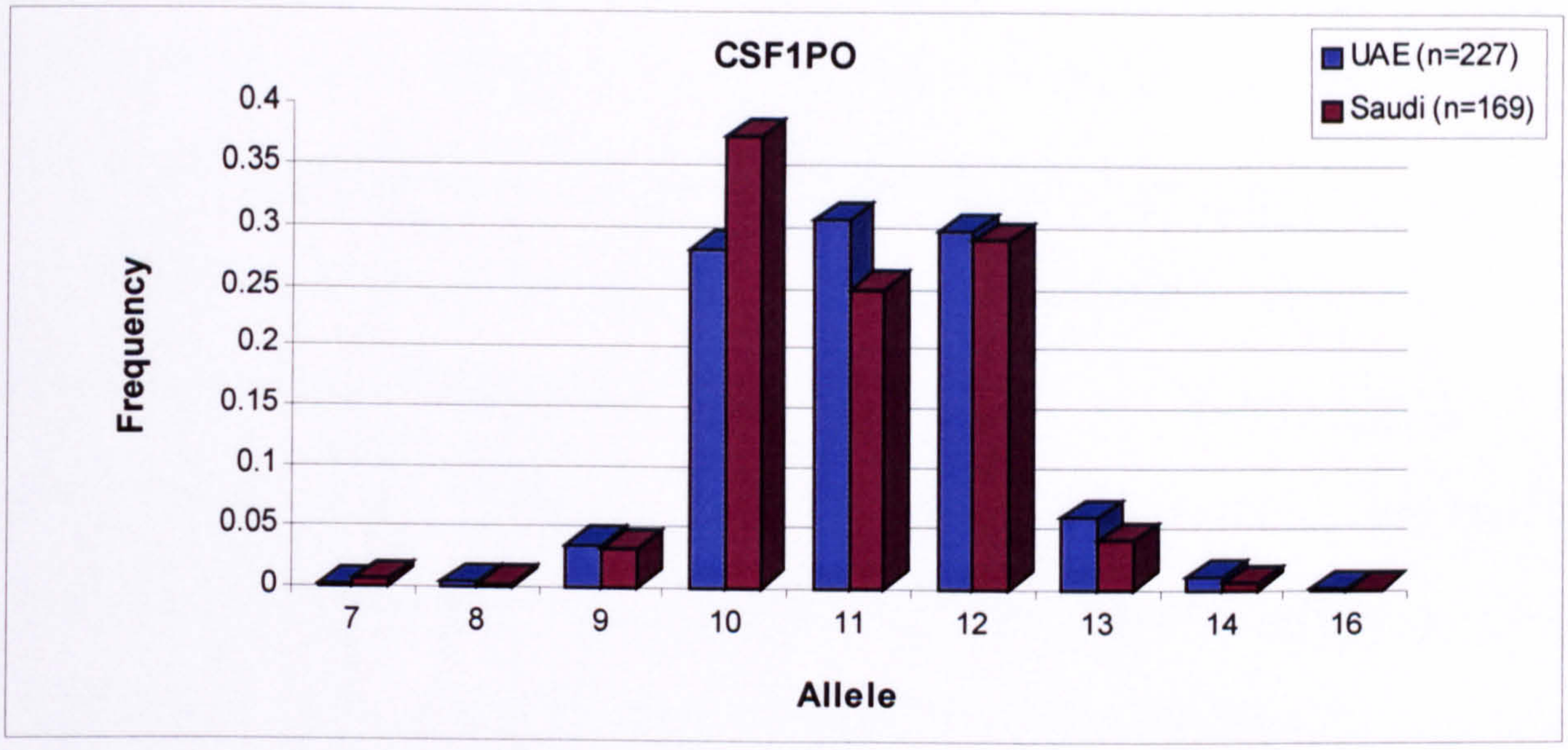

D7S820
UAE (n=229)
Omani (n=162)
Frequency
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
7
8
8.3
9
9.1
10
11
12
13
14
Allele


D13S317
UAE (n=229)
Omani (n=162)
Frequency
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
8
9
10
11
12
13
14
15
Allele


vWA
UAE (n=228)
Omani (n=162)
Qatari (n=200)
Saudi (n=149)
Yemen (n=200)
Morocans (n=146)
Frequency
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
11
13
14
15
16
17
18
19
20
21
Allele


THO1
UAE (n=229)
Saudi (n=173)
Yemen (n=100)
Morocans (n=144)
Frequency
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
5
6
7
8
9
9.3
10
11
Allele


TPOX
UAE (n=229)
Saudi (n=165)
Morocans (n=130)
Frequency
0.6
0.5
0.4
0.3
0.2
0.1
0
6
7
8
9
10
11
12
Allele


CSF1PO
UAE (n=227)
Saudi (n=169)
Frequency
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
7
8
9
10
11
12
13
14
16
Allele

6.11.1.2 RxC test

The RxC contingency test was performed to compare the allele frequencies of the UAE to other Arabic populations. There was no significant different between the UAE and Omani populations at all four loci tested. However, there were significant differences between the UAE and Saudi populations at two loci vWA and TPOX. The results showed that there was also significant different between the UAE and Yemeni populations at vWA. Other significant differences between the UAE and Moroccan populations were observed at vWA and THO1. Significant different observed between the UAE and Qatari populations at vWA. These significant differences may reflect that this test is a sensitive test or may be because of other reason. The results can be seen in table 6.13.

Table 6.13 The RxC test results comparing the UAE population to the Omani, Saudi, Yemeni, Qatari and Moroccan (Mor.) populations. 20 batches of 2,500 replicates per batch performed. The total replicates were 50,000. The p values and standard errors (SE) were calculated and shown in the table.

Locus	UAE/Omani	SE	UAE/Saudi	SE	UAE/Mor.	SE
D5S818	0.517	0.023	-	-	-	-
D7S820	0.711	0.024	-	-	-	-
D13S317	0.050	0.018	-	-	-	-
vWA	0.431	0.036	***0.005***	0.002	***0.004***	0.002
THO1	-	-	0.379	0.032	***0.002***	0.001
TPOX	-	-	***0.003***	0.001	0.061	0.007
CSF1PO	-	-	0.153	0.020	-	-

Locus	UAE/Qatari	SE	UAE/Yemeni	SE
vWA	***0.026***	0.008	***0.014***	0.001
THO1	-	-	0.913	0.008

6.11.1.3 Forensic and paternity parameters comparison

The UAE population was compared to other Arabic populations at different number of loci for the observed heterozygosity, power of exclusion and power of discrimination. The UAE population observed the lowest value of heterozygosity (0.69), power of exclusion (0.415) and power of discrimination (0.905) at vWA locus. The highest value of heterozygosity (0.81) was observed in Qatari population but the highest value of power of exclusion (0.614) and power of discrimination (0.934) was observed in Saudi population at vWA. There are some differences at other loci as shown in table 6.14 a, b and c.

Table 6.14a Observed Heterozygosity in the UAE population compared to other Arabic populations.

Locus	UAE	Omani	Saudi	Qatari	Yemeni
D5S818	0.70	0.76	-	-	-
D7S820	0.76	0.75	-	-	-
D13S317	0.79	0.70	-	-	-
vWA	0.69	0.72	0.77	0.81	0.72
THO1	0.76	-	0.73	-	0.78
TPOX	0.70	-	0.61	-	-
CSF1PO	0.76	-	0.73	-	-

Table 6.14b The power of exclusion in the UAE population compared to other Arabic populations.

Locus	UAE	Omani	Saudi	Qatari	Yemeni
D5S818	0.426	0.529	-	-	-
D7S820	0.527	0.541	-	-	-
D13S317	0.573	0.569	-	-	-
vWA	0.415	0.582	0.614	0.610	0.591
THO1	0.534	-	0.564	-	0.575
TPOX	0.433	-	0.395	-	-
CSF1PO	0.523	-	0.460	-	-

Table 6.14c The power of discrimination in the UAE population compared to other Arabic populations.

Locus	UAE	Omani	Saudi	Qatari	Yemeni
D5S818	0.896	0.899	-	-	-
D7S820	0.931	0.904	-	-	-
D13S317	0.914	0.916	-	-	-
vWA	0.905	0.922	0.934	0.929	0.924
THO1	0.925	-	0.915	-	0.922
TPOX	0.860	-	0.821	-	-
CSF1PO	0.874	-	0.865	-	-

6.11.2 Comparison with major world populations

The results of allele frequency distribution obtained from the UAE population in this study were compared to four major ethnic data populations (African American, Caucasian American, Hispanic American and Mongoloid). These data were previously published in technical manual of the GenePrint™ PowerPlex™ 1.2 system, Promega Co., 1998; Yamamoto T. *et al.*, 1999; Entrala C. *et al.*, 1999; Jarreta B. *et al.*, 1999; Martin P. *et al.*,

1998; Budowle B. *et al.*, 1997; Lorente J. *et al.*, 1994; Rangel-Villalobos H. *et al.*, 1999; Scoular C., 1999; Andres M. *et al.*, 1996; Lorente M. *et al.*, 1997 and Nata M. *et al.*, 1999.

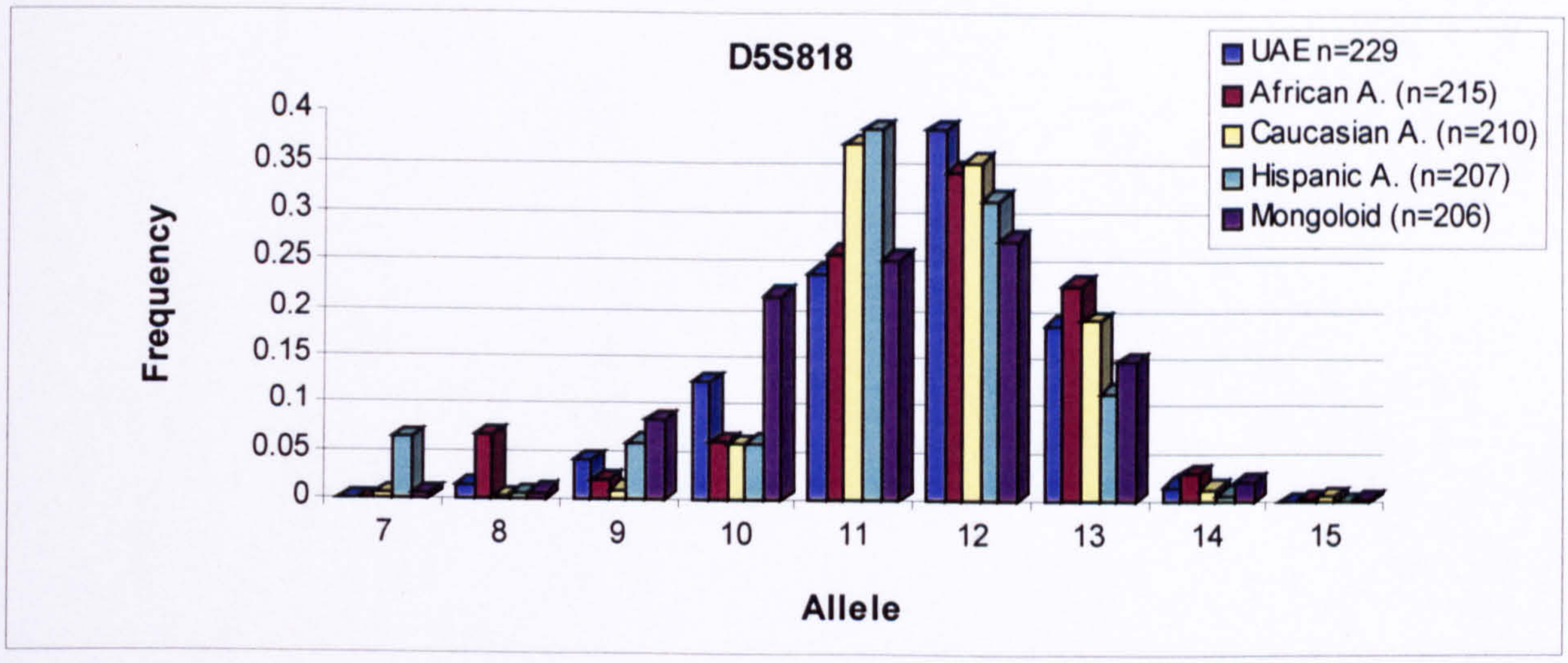
Allele 10 and 11 at D5S818 locus had the most variation between the five populations. Allele 9, 11 and 12 at D7S820 locus had showed the most variation between the five populations. Allele 8, 9, 10 and 12 at D13S317 locus had observed the maximum variation between the five populations. The loci D16S539 and TPOX observed the least variation between the five populations (Mongoloid population is not included in D16S539). Allele 14, 15 and 16 at vWA had showed the most variation while allele 7, 8, 9 and 9.3 observed at THO1 the most variation between the five populations. CSF1PO locus had showed the most variation in the allele 12 however, allele 7 and 8 were extremely rare in four populations tested but were less rare in African American population. The results are shown in figure 6.4.

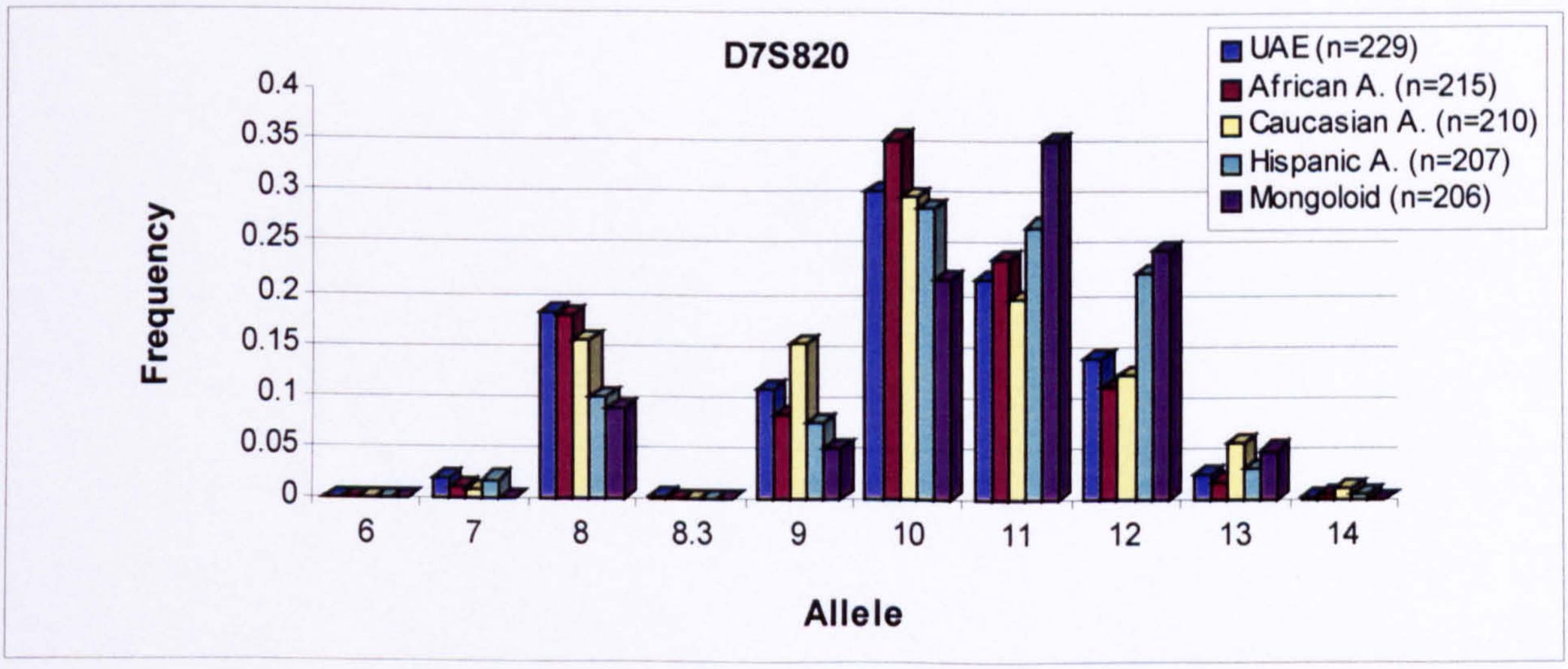
6.11.2.1 RxC test

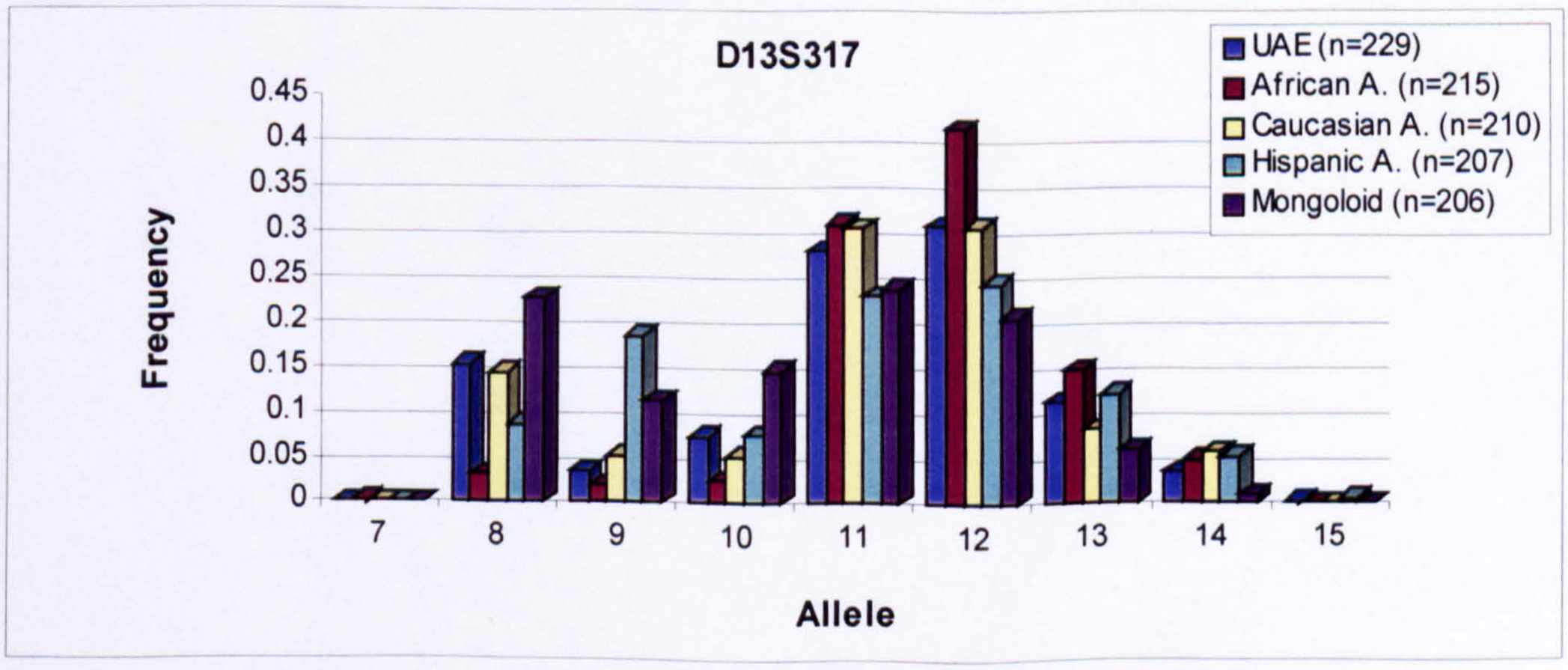
The UAE allele frequency was compared to African American, Caucasian American, Hispanic American, Japanese, Spanish as a southern European Caucasian populations by applying the RxC contingency test. Two Caucasian populations were tested (Caucasian American and Spanish). The results showed that there was no significant difference for four loci when the UAE compared to the Spanish populations. One explanation might be due to a long history of Arabs settlement (approximately eight centuries) in Spain and especially in the south part (Jarreta B. *et al.*, 1999) however, it may be coincidental. The most significant differences were observed between the UAE and Japanese (Mongoloid) populations. The results are seen in table 6.15.

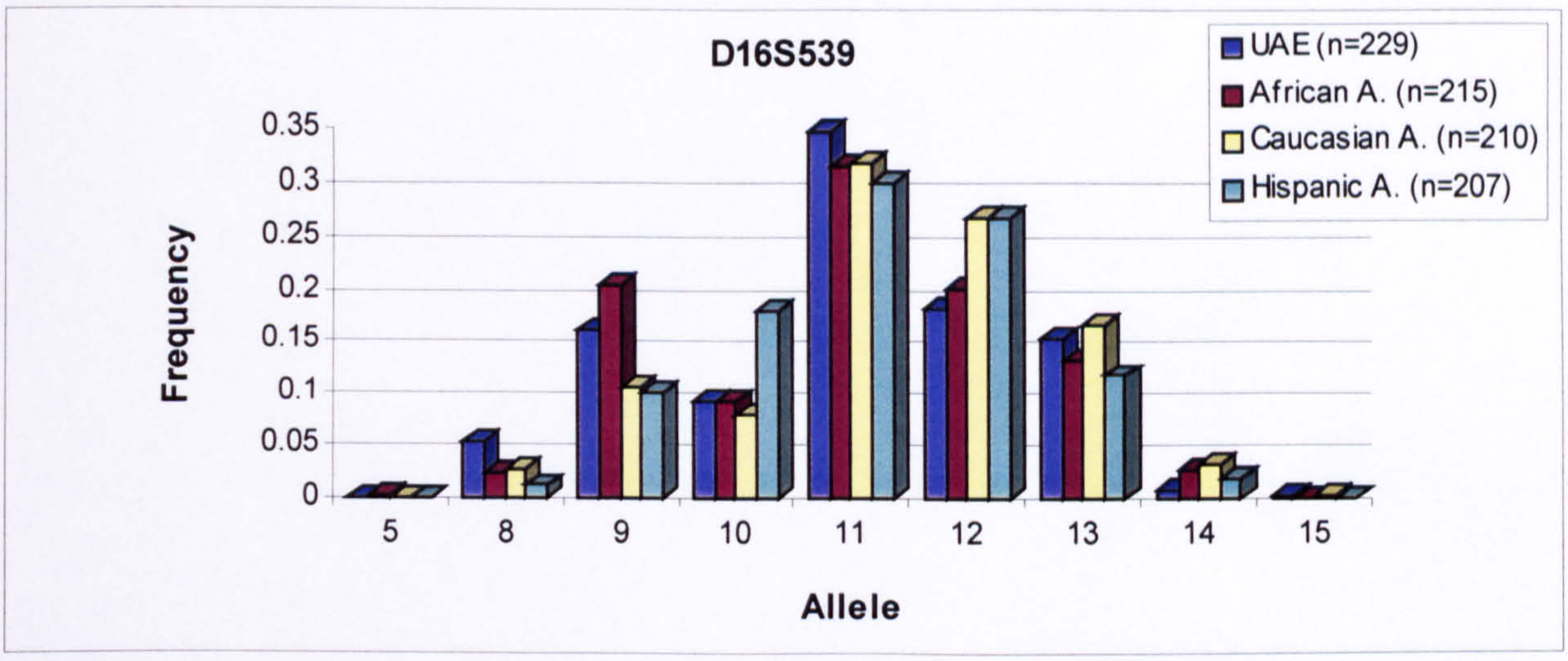
198A

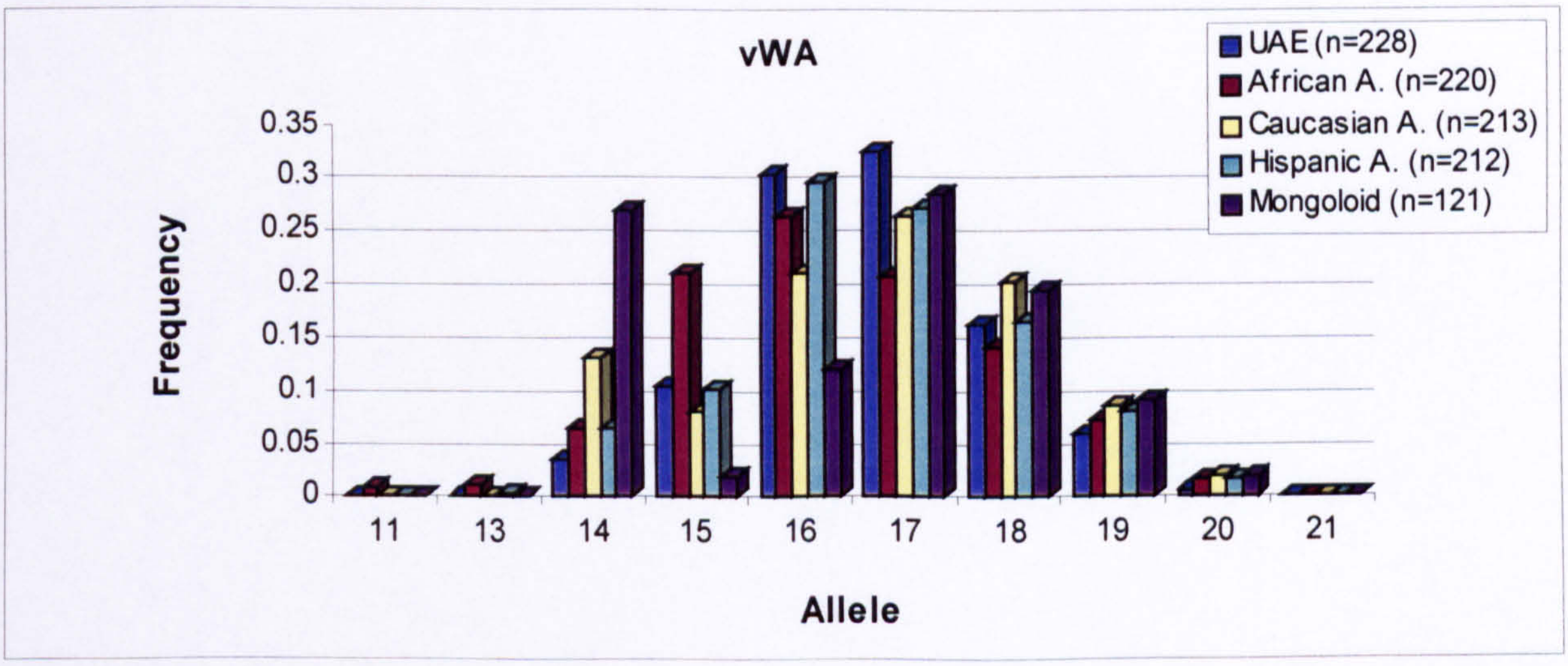
Figure 6.4 The UAE Arab allele frequencies compared to four world major ethnic groups.

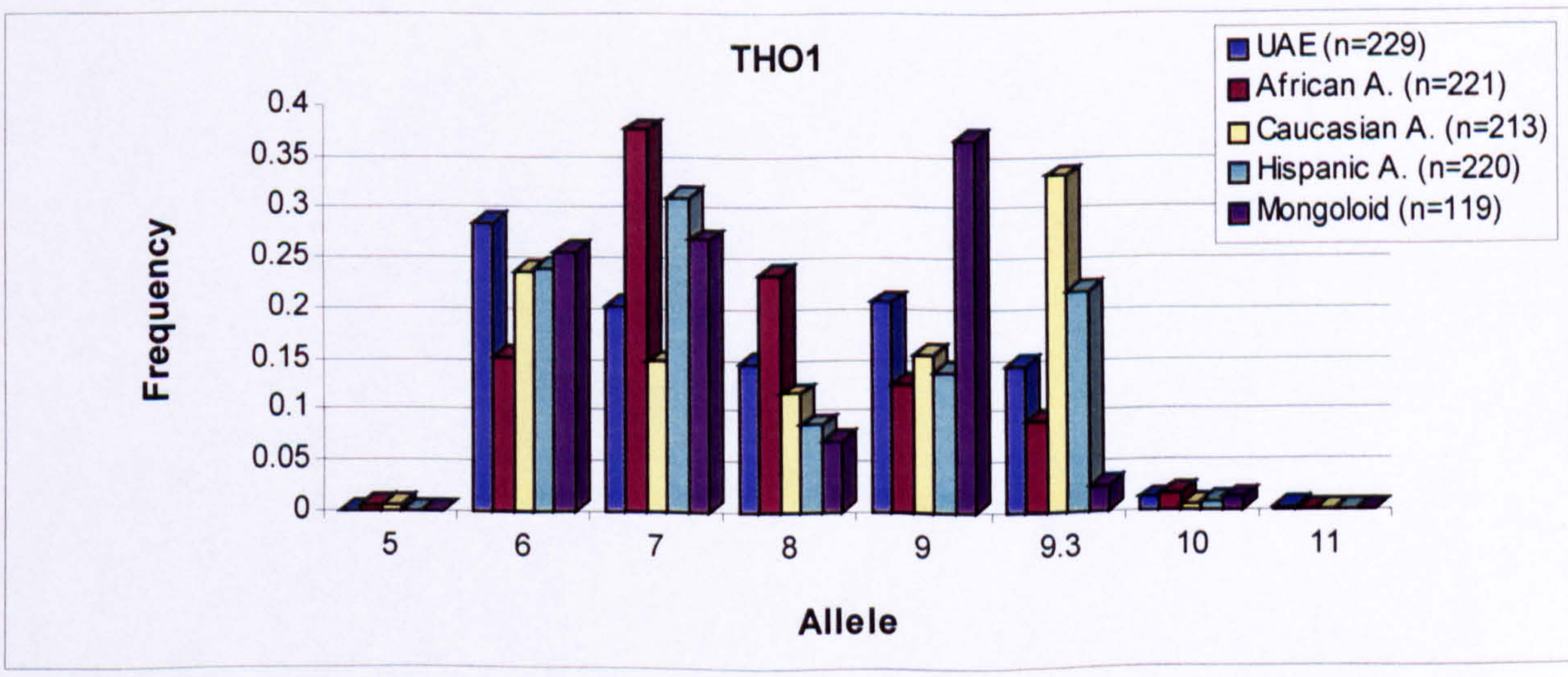

D5S818
UAE n=229
African A. (n=215)
Caucasian A. (n=210)
Hispanic A. (n=207)
Mongoloid (n=206)
Frequency
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
7
8
9
10
11
12
13
14
15
Allele

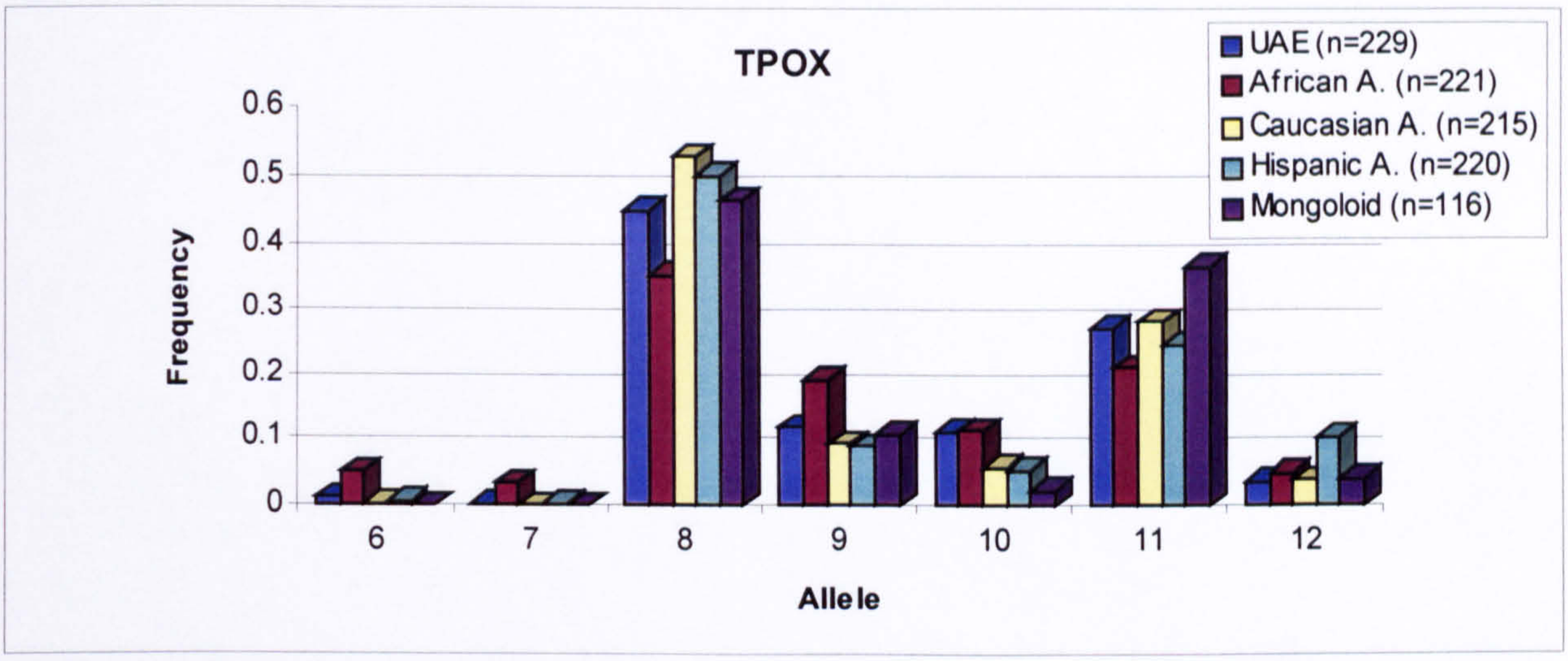

D7S820
UAE (n=229)
African A. (n=215)
Caucasian A. (n=210)
Hispanic A. (n=207)
Mongoloid (n=206)
Frequency
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
6
7
8
8.3
9
10
11
12
13
14
Allele


D13S317
UAE (n=229)
African A. (n=215)
Caucasian A. (n=210)
Hispanic A. (n=207)
Mongoloid (n=206)
Frequency
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
7
8
9
10
11
12
13
14
15
Allele


D16S539
UAE (n=229)
African A. (n=215)
Caucasian A. (n=210)
Hispanic A. (n=207)
Frequency
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
5
8
9
10
11
12
13
14
15
Allele


vWA
UAE (n=228)
African A. (n=220)
Caucasian A. (n=213)
Hispanic A. (n=212)
Mongoloid (n=121)
Frequency
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
11
13
14
15
16
17
18
19
20
21
Allele


THO1
UAE (n=229)
African A. (n=221)
Caucasian A. (n=213)
Hispanic A. (n=220)
Mongoloid (n=119)
Frequency
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
5
6
7
8
9
9.3
10
11
Allele


TPOX
UAE (n=229)
African A. (n=221)
Caucasian A. (n=215)
Hispanic A. (n=220)
Mongoloid (n=116)
Frequency
0.6
0.5
0.4
0.3
0.2
0.1
0
6
7
8
9
10
11
12
Allele

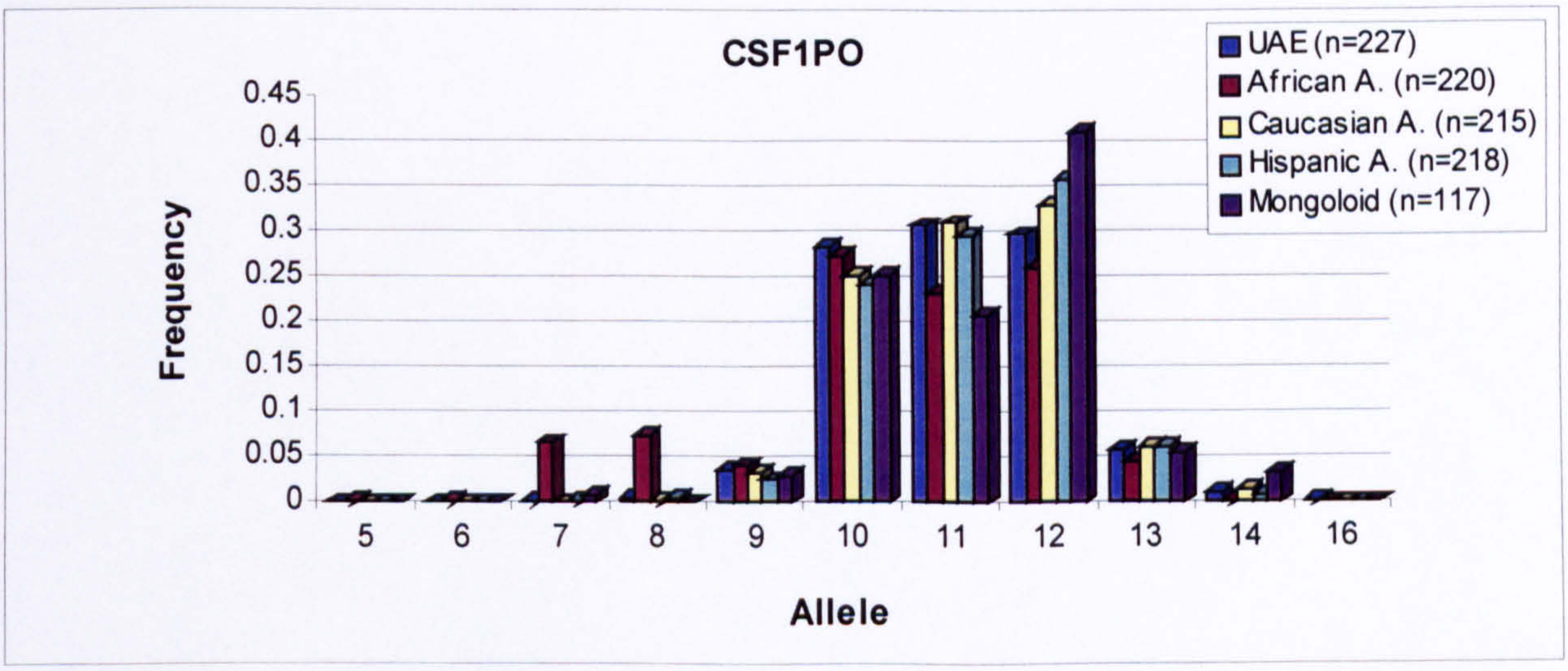

CSF1PO
UAE (n=227)
African A. (n=220)
Caucasian A. (n=215)
Hispanic A. (n=218)
Mongoloid (n=117)
Frequency
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
5
6
7
8
9
10
11
12
13
14
16
Allele

Table 6.15 p values of exact test for RxC contingency table for the UAE allele frequencies compared to four major world populations (African American, Caucasian American, Hispanic American and Mongoloid) and Spanish Caucasian (southern Spain) using RxC contingency test.

Locus	UAE/ African A.	SE	UAE/ Caucasian A.	SE	UAE/ Japanese	SE
D5S818	***0.000***	0.000	***0.000***	0.000	***0.000***	0.000
D7S820	0.406	0.020	0.068	0.014	***0.000***	0.000
D13S317	***0.000***	0.000	0.226	0.022	***0.000***	0.000
D16S539	***0.044***	0.020	***0.001***	0.001	-	-
vWA	***0.000***	0.000	***0.000***	0.000	***0.000***	0.000
THO1	***0.000***	0.000	***0.000***	0.000	***0.000***	0.000
TPOX	***0.000***	0.000	***0.012***	0.006	***0.000***	0.000
CSF1PO	***0.000***	0.000	0.934	0.016	***0.000***	0.000

Locus	UAE/ Hispanic A.	SE	UAE/Spain	SE
D5S818	***0.000***	0.000	***0.001***	0.001
D7S820	***0.000***	0.000	0.555	0.034
D13S317	***0.000***	0.000	0.830	0.016
D16S539	***0.000***	0.000	***0.000***	0.000
vWA	0.178	0.018	***0.000***	0.000
THO1	***0.000***	0.000	***0.001***	0.001
TPOX	***0.000***	0.000	0.066	0.009
CSF1PO	0.427	0.028	0.365	0.034

6.11.2.2 Heterozygosity comparison

The observed heterozygosity frequency from the UAE Arabic population compared to other four major populations. The UAE population observed minimum heterozygosity at D5S818 and vWA. The African American population observed minimum heterozygosity at D13S317. The Caucasian American population showed minimum heyterozygosity at D16S539. The Hispanic American population observed minimum

heterozygosity at CSF1PO. However, Mongoloid population showed minimum heterozygosity at D7S820, THO1 and TPOX. The results are shown in table 6.16.

Table 6.16 Observed Heterozygosity in the UAE Arabic population compared to other four major world ethnic groups.

Locus	UAE	African American	Caucasian American	Mongoloid (Japanese)	Hispanic American
D5S818	0.70	0.78	0.71	0.78	0.77
D7S820	0.76	0.76	0.80	0.74	0.82
D13S317	0.79	0.69	0.71	0.81	0.78
D16S539	0.77	0.78	0.73	-	0.78
vWA	0.69	0.83	0.82	0.77	0.75
THO1	0.76	0.73	0.77	0.61	0.74
TPOX	0.70	0.75	0.65	0.62	0.67
CSF1PO	0.76	0.82	0.78	0.73	0.70

6.11.2.3 Forensic and paternity parameters comparison

The statistical value of power of exclusion and power of discrimination from the UAE population compared to all other four major world populations (African Zimbabwe, Caucasian Scottish, Mongoloid Japanese and Hispanic Mexican). The results are showed in tables 6.17 and 6.18.

Table 6.17 The power of exclusion in the UAE Arabic population compared to other major world populations.

Locus	UAE	African Zimbabwe	Caucasian Scottish	Mongoloid Japanese	Hispanic Mexican
D5S818	0.426	-	0.463	-	-
D7S820	0.527	-	0.553	-	-
D13S317	0.573	-	0.644	-	-
D16S539	0.542	-	0.555	-	-
vWA	0.415	0.629	0.625	0.581	0.604
THO1	0.534	0.476	0.550	0.475	0.544
TPOX	0.433	0.525	0.447	0.367	-
CSF1PO	0.523	0.546	0.497	0.484	0.450

Table 6.18 The power of discrimination in the UAE Arabic population compared to other four major world populations.

Locus	UAE	African Zimbabwe	Caucasian Scottish	Mongoloid Japanese	Hispanic Mexican
D5S818	0.896	-	0.880	0.918	-
D7S820	0.931	-	0.918	0.906	-
D13S317	0.914	-	0.929	0.936	-
D16S539	0.922	-	0.918	-	-
vWA	0.905	0.931	0.943	0.923	0.922
THO1	0.925	0.877	0.879	0.844	0.904
TPOX	0.860	0.897	0.878	0.788	-
CSF1PO	0.874	0.903	0.871	0.861	0.859

6.11.3 Indian subcontinent populations

Indian and Pakistani populations residing in the UAE compared to Indian subcontinent (Indo-Pakistani) populations residing in the UK at THO1 and vWA loci (Evett I. *et al.*, 1997). Pairwise RxC contingency test was used to test for homogeneity in allele frequency distribution. The result showed that there is no significant different between these populations. See table 6.19.

Table 6.19 p values of exact test for the UAE (Indian) and UAE (Pakistani) allele frequencies compared to UK (Indo-Pakistani). There was no significant differences observed (p>0.05).

Locus	UAE (Indian)/ UK (Indo-Pakistani)	SE	UAE (Pakistani)/ UK (Indo-Pakistani)	SE
THO1	0.864	0.011	0.175	0.026
vWA	0.064	0.011	0.426	0.028

6.12 DISCUSSION

The co-amplification of eight STR loci in a single test tube reduced effectively the labour and cost needed and gave much more information in a shorter period of time. From the eight STR loci 63, 64, 55 and 60 different alleles were found in the UAE, Indian, Pakistani and Egyptian populations respectively. The possible number of different combinations of genotypes are 2.07×10^{12}, 8.33×10^{11}, 2.53×10^{12} and 2.96×10^{11} in the UAE, Egyptian, Indian and Pakistani populations respectively.

There were no great differences in allele frequency distributions among the four populations tested. However, there was more similarity between the two Arabic populations and there was more similarity between the Indian and Pakistani populations as would be expected.

6.12.1 Hardy-Weinberg Equilibrium test

The Chi square test and Fisher's exact test with 2,000 shufflings were performed to test for HWE. The exact test is recommended in small size sample (Guo S. and

Thompson E., 1992). The data showed that none of the UAE, Egyptian or Indian loci observed a significant deviation from HWE. However, three loci (vWA, THO1 and TPOX) in the Pakistani population do display significant deviation from HWE. Although it was not highly significant at THO1 locus ($p = 0.037$). Deviation from HWE is possible and this might be a true reflection of the population being out of HWE at these loci or alternatively could be due to sampling error. There are several studies have been reported such as for locus vWA for the British population (Drozd M. *et al.*, 1994) and Russian population (Sajantila A. *et al.*, 1994) where deviations were observed.

There was no significant differences detected between observed and expected heterozygosity among the four data populations analysed based on Chi square test ($p>0.05$).

6.12.2 Comparison between different UAE populations

The RxC contingency table were used to compare the distribution of allele frequencies within UAE native population and between other populations studied in this work. There were significant differences between the UAE and Indian populations at all eight loci tested except D16S539. However, there were significant differences between the UAE and Pakistani populations only at three loci, D5S818, vWA and TPOX. When the Indian population was compared to Pakistani population there were also significant differences at three loci, D7S818, D13S317 and TPOX.

6.12.3 Comparison between two UAE Arabic native populations

Minimum differences in allele frequency distributions were observed within the UAE Arabic native populations. This may indicate that there is little variation within a population (Budowle B. and Monson K., 1994).

6.12.4 Comparison between other Arabic populations

The UAE population was compared to other Arabic populations by using RxC contingency test and large differences between different Arabic populations were not detected.

6.12.5 Comparison between the UAE and major world populations

The differences in the distribution of allele frequencies between the UAE and other world major populations were high. This result supports the previous studies which states that the variation in allele frequency distributions is greater between the major ethnic groups than within a certain ethnic group (Weir B., 1992a; Budowle B. and Monson K., 1994; Budowle B. *et al.*, 1994).

Comparitive studies between different major ethnic groups showed that THO1 and vWA loci might provide useful information on ethnic differentiation studies. Locus D13S317 might be used also for the same reason.

6.12.6 Comparison between the Indo-Pakistani populations residing in two different countries

When Indo-Pakistani population from the UK compared to Indian and Pakistani populations residing in the UAE using the RxC test no significant differernces were

observed. This finding may be very useful where the Indo-Pakistani database from the UAE could reasonably be applied on any Indo-Pakistani population residing in any other country in the world. This might be of especial use in the GCC countries (UAE, Oman, Saudi Arabia, Kuwait, Qatar and Bahrain) where a huge number of Indo-Pakistani workers reside.

6.12.7 Forensic and paternity parameters

The power of discrimination and typical paternity index (TPI) are both lowest in the UAE population, however the match probability (MP) in the UAE population is still highly significant at 1 in 1.64 X 10^8, especially as the population of the UAE is approximately three million. The typical paternity index ranged from 188 in the Egyptian population to 339 in the Pakistani population. This relatively low value may require for more loci to be applied when undertaking paternity testing. To increase this value of paternity index the VNTR loci which has been studied earlier can be added which will rise the paternity index to a suitable value.

The PowerPlex 1.2 system offered improved discrimination over the Quadruplex system (1 in 10,000) (Gill P. *et al.*, 1995). By using this kit the power of discrimination was reached higher or equivalent to the second generation multiplex (SGM) system (1 x 10^8) (Sparkes R. *et al.*, 1996a, b). However, unlike the SGM system this kit does not contains any of hypervariable or compound loci such as FGA and D21S11.

The typical paternity index for the eight STR loci in the four populations tested was higher than the combined paternity index for both CTTv and FFFl multiplexes in Hispanic-American (86.5) (Lins A. *et al.*, 1998).

The combined probability of exclusion and probability of paternity for these eight STR loci was higher than other multiplex as CTTv, GammaSTR and FFFL multiplex in all American populations (Lins A. *et al.*, 1998).

The matching probability of the eight STR loci in all four populations studied was higher than of seven loci [THO1, TPOX, F13A1, F13B, CD4, FES/FPS and D5S373 (0.0000013)] in north-east Italy (Cossutta F. *et al.*, 2000).

It can be concluded that these eight STR loci were found to be well suited for forensic analysis and paternity testing in the UAE, Egyptian, Indian and Pakistani populations. For forensic identification purposes the combined power of discrimination for these eight loci is 0.999999994, 0.999999993, 0.999999997, and 0.999999994 in the UAE, Egyptian, Indian and Pakistani populations respectively. In paternity testing, the combined exclusion power is 99.6%, 99.6%, 99.7% and 99.8% in the UAE, Egyptian, Indian and Pakistani populations. As it can be clearly observes that the powers of discrimination between the four populations are very similar.

6.12.8 Combination of all STR and VNTR loci

The combination of all five VNTR and eight STR loci studied gave a very high value of all forensic and paternity parameters.

The combined typical paternity index for all five VNTR and eight STR loci was 66323, 382844 and 438488 in the UAE, Indian and Pakistani population respectively. It was higher than the comined paternity index in Caucasian-American for both PowerPlex™ 1 System and FFFL Multiplex including 12 STR loci having a value of 3976 (Lins A. *et al.*, 1998).

The combined power of exclusion of paternity for all five VNTR and eight STR loci was 0.99999, 0.999998 and 0.999998 in the UAE, Indian and Pakistani populations respectively. These values were higher than of both PowerPlex™ System and FFFL Multiplexes in the Hispanic-American (0.9997) and Caucasia-American (0.9998) (Lins A et al., 1998). It was also higher than what obtained from the study of 13 STR loci in Japanese population (0.99985) (Yamamoto T. *et al.*, 2000) and higher than AmpFlSTR Profiler Plus in Portugal population (0.99988) (Pinheiro M. *et al.*, 2000).

The matching probability was 1.3×10^{-16}, 3.6×10^{-17} and 1.9×10^{-16} in the UAE, Indian and Pakistani populations respectively. These values were higher than the result obtained from AmpFlSTR Profiler Plus in Japanese population (2.28×10^{-11}) (Nakamura S. *et al.*, 2000).

All the forensic and paternity parameters of the combination of the five VNTR and eight STR loci in the three populations studied were lower than the forensic and paternity parameters of the PowerPlex 16 System™ in all four major world populations studies (Sprecher C. *et al.*, 2000). However, the power of exclusion in the Hispanic-American (0.999998) was equal to the Indian and Pakistani populations (0.999998) (Sprecher C. *et al.*, 2000).

It was recommended for number of markers to be used in paternity analysis to have a power of exclusion of ≥ 99% (it means excluding ≥ 99% of the male population from being the biological father) (Hammond H. *et al.*, 1994). However, the power of exclusion of 99% is relatively weak. The English speaking working group (ESWG) of the ISFG reported that the majority of the European laboratories used values in the range of probability of paternity (W = 99.9-99.999%) (Hallenberg C. and Morling N., 2001). In

this study the probability of paternity would reach higher than or equal to 99.999% when combining the results of all five VNTR and eight STR loci in the UAE, Indian and Pakistani populations.

In conclusion, the result of the combining of all five VNTR and eight STR loci was excellent for forensic analysis and paternity tesing in the UAE population. However, using the PowerPlex™ 16 System (Promega Co.) or the Identifiler (Applied Biosystems) would give better result and in a single reaction, require small amount of sample and fast. Therefore, it is preferable to use the PowerPlex™ 16 System or the Identifiler than using VNTR system as the use of the commercial kits are not problematic in the UAE because of the expence.

CHAPTER 7: CONCLUSION AND FUTURE WORK

7.1 THE ACHIEVEMENT OF THIS STUDY

This project, in general, was based on the study of the UAE population structure and the application of two most powerful genetic systems for forensic and paternity analysis in the UAE. Blood samples were collected and purified using puregene™ kit which provided high quantities and qualities of extracted DNA. Additionally, it is simple, quick and easy method to perform avoiding the biohazard of organic extraction.

The first part of this thesis was the use of five VNTR loci to examine the structure nature of the three most populous populations residing in the UAE. The UAE Arab represents approximately 20%, the Asian-Indian approximately 35% and the Pakistani approximately 25%. The analysis of the data as would expected revealed a little evidence of dis-similarity between the Arab and the other two populations from the Indian subcontinent. However, there was no evidence of substructure when Indian population compared to Pakistani population. Therefore, there is no need to use three databases for the three populations, instead two databases would be sufficient one for the Arab and the other for Indian subcontinent. Additionally, in order to compensate the existence of population substructure the F_{ST} (θ) value of 0.0126 was found to be conservative.

The forensic statistical parameters for these five VNTR loci were sufficient for forensic analysis. However, it is require more loci to be used in paternity testing since only four of these loci can be used. Although these loci are extremely powerful the binning system applied on these loci reduced their power. The VNTR system still was the most powerful tool for forensic analysis and paternity testing when this project started.

However, the VNTR system gradually lost its popularity since the introduction of STR multiplexing system such as PowerPlex 1.2 kit (Promega Co.) and Profiler plus, SGM plus (PE-Applied Biosystems). Recently a commercially kit was released, the PowerPlex 16 system (Promega Co.), which amplifies 16 loci including the amelogenin. A second kit might be released very soon (June 2001), the Identifiler (PE-Applied Biosystems), which also amplifies 16 loci including the amelogenin. These two kits probably will replace the use of VNTRs in forensic and paternity analysis.

The second part of this thesis was the use of eight STR loci using PowerPlex 1.2 kit (Promega Co.) to examine the structure nature of the three most populous populations residing in the UAE as mentioned above. In addition, a fourth population (Egyptian Arab) was tested as well since Egyptians represent approximately 5% of total UAE population. The analysis of the data revealed very little evidence of substructure between the two Arabic populations and between the two Indian subcontinent populations. However, the highest evidence of substructure was observed when Indian population compared to Egyptian population. This supports the findings of VNTR data analysis. Therefore, there is no need to use four databases one for each population, instead two databases would be sufficient one for the Arab and the other for Indian subcontinent. In order to compensate the existence of population substructure the F_{ST} (θ) value of 0.023 was found to be conservative for these STR loci.

Almost all the examined STR and VNTR loci for the four populations studied were in HWE and the loci were in linkage equilibrium (LE). In addition, almost all these loci are on different chromosomes. Therefore, the product rule for combining genotype frequencies across unlinked loci is valid for use at these loci.

The third part of this thesis was to compare the two UAE native populations from two Emirates, Abu Dhabi and Sharjah. The analysis of the data revealed no evidence of substructure between the two Emirates and therefore a single database could be used for the entire UAE native population although this assumes that the other Emirates have a similar population structure. The forensic and paternity statistical parameters for these eight STR loci were sufficient for forensic analysis but require additional loci for paternity testing. Therefore, adding VNTR loci to STR loci would be sufficient for paternity testing.

All the examined STR loci, for the UAE Arabic native population, were compared with different populations such as Arabic and other four world major populations. There were little differences between the UAE Arabic native population compared to other Arabic populations. One of the Arabic population such as Omani observed no significant differences in allele distributions at all four loci tested ($p>0.05$) when compared to the UAE Arabic native population using RxC contingency test. However, there were significant differences when compared the UAE Arabic native population to other world major population groups. These results confirm other studies that there is little variation in allele distributions within a major group and more variation between different major groups (Weir B., 1992a; Budowle B. and Monson K., 1994; Budowle B. *et al.*, 1994). These results also confirm the necessity of establishing a specific database for each population.

When Indian and Pakistani populations residing in the UAE were compared to a population from the Indian subcontinent residing in the UK no significant differences were observed. This result justifies applying a single database for Indian subcontinent no matter in which country they are residing. Therefore, it would be resonable for this

database of Indo-Pakistani population to be used across all the six Arabic Gulf Co-operation Council (GCC) countries (UAE, Qatar, Bahrain, Saudi Arabia, Oman and Kuwait) since a large Indian and Pakistani populations residing in all these countries.

The statistical parameters for all VNTR and STR examined loci were calculated and the practical use was determined by analysis of several different simulated forensic and paternity cases. Their use in forensic analysis and paternity testing yielded powerful evidence of forensic identification and paternity. The use of these loci was also tested in resembling motherless cases. See appendix H.

7.2 THE FUTURE WORK

The two combined DNA systems approved to be informative enough for forensic analysis and paternity testing purposes in the UAE. However, in forensic stain analysis it is sufficient and preferable to use only STR system since minute and severely degraded samples are common. These STR loci are almost of simple type having alleles smaller than 350 base pairs in length and easy to standardise to allow the comparison of DNA profiling between different laboratories.

The UAE population contains other two distinct major populations such as African (Sudanese and Somali) and Mongoloid (Filipino). These populations represent approximately 5% of the total UAE populations. Since the allele frequency distribution may greatly differ between different major groups. Therefore, in the future these two populations should be typed and compared to other populations residing in the UAE in order to measure the amount of substructure and the value of F_{ST}. At the present the two databases (Indo-Pakistani and Arab) and the F_{ST} value of 0.05 would be applied in cases

where any individual from these populations involved the more conservative value should be reported. Alternatively, published databases might be used temporarily.

The UAE population also consists of an isolated population, the shihuh, who lives on border between the UAE and Oman. However, the majority of this population lives in Oman. In order to deal with this population a value of F_{ST} = 0.05 (NRC, 1996) could be applied and the database of the UAE Arabs would be used. In the future, if possible, a database of this population should be generated.

The FBI in the United States has established the combined DNA index system (CODIS) which consists of 13 STR loci. These 13 loci usually typed to give information necessary in matches in order to link suspects to criminal cases (Budowle B. *et al.*, 1999). By using CODIS loci in different countries the search through databases by different laboratories would be possible and DNA profiles can be compared to identify linkage between unresolved cases. Therefore, the future work would be the use of the PowerPlex 16 system (Promega Co.) or the Identifiler (PE Applied Biosystems) in the UAE since these two kits contains all 13 CODIS loci and require a single amplification reaction instead of two. The UAE database then can be stored in the CODIS banking.

In addition, there are several working groups in the USA (SWGDAM), Latin America (GITAD) and Europe (EDNAP). Other two working groups of ISFG called the Spanish and Portuguese working group and English speaking working group. These groups discuss and report the progress of the standardisation between different laboratories across different countries and try to solve technical and statistical problems in forensic DNA analysis. Therefore, there is a need to establish an Arabic working group on DNA in the region. The GCC which consists of six Arabic countries already uses the DNA (HLA

DQ-alpha, PM and D1S80) and almost all forensic laboratories are well funded therefore it would be advisable to use, for example, the PowerPlex 16 system (Promega Co.) or the Identifier (Applied Biosystems) and establish their DNA databases to be used in forensic analysis and paternity testing. Typing DNA from convicted offenders and establishing a centralised searchable databases of people from these countries would be important in order to link suspects to a scene of crime.

Moreover, establishing the UAE database for Y-chromosome STR system would be very helpful in forensic casework and paternity testing. The Y-STR can be very useful also in paternity testing especially in deficiency cases, when the alleged father is deceased, with a male offspring (Chakraborty R., 1985). Any male relative to the allelged father could be tested in his place.

APPENDIX A:

THE UAE TRIBES

1. The Bani-Yas

Discovery of water on Abu Dhabi's island in 1761 made permanent settlement along the coast of what is now the Emirate of Abu Dhabi possible for the first time of the Bani Yas. The Bani Yas came originally from Najd, the central part of the Arabian peninsula and the traditional heartland of Saudi Arabia. The Bani Yas consists of several tribes and these tribes are Al Bu-Falah, Al Bu-Falasah, Al Hawamel, Al Qubaisat, Al Sudan, Al Rumaithat, Al Maharbah, Al Murrar, Al Mazarea, Al Qumzan, Al Bu Muhair.

The ruler of Abu Dhabi is from one family of Al Bu-Falah, which is the Al Nuhayyan. The ruler of Dubai is from one family of Al Bu-Falasah, which is Al Maktum.

The rulers of Abu Dhabi effectively gained the support of various tribal groups in the Buraimi area, especially the Manasir, Dhawahir, and Awamir. In contrast with tribes in some parts of the Middle East, these are not large. By the most generous estimate the Bani Yas were only about 15,000 as of the late nineteenth century. The other tribes generally are composed of less number of individuals, some numbering only in the hundreds.

In 1833 a subsection of the Bani Yas called the Al Bu Falasah (about 800 people) seceded from Abu Dhabi and settled in Dubai.

2. The Qawasim

The origin of the Qawasim is somewhat obscure. However, it has been said that the Qawasim was part of a migratory group of peninsula Arabs called Hawala which from the Arabia verb meaning "to move to a new residence". The Hawala left central Arabia to settle in Persia shortly after the Islamic conquest of that country in the mid-seventh century. Subsequently, the Hawala occupied both shores of the Gulf for many centuries. The Arab Qawasim had ruled Lingah, it used to be the capital of Qawasim state, and the eastern coast of the Gulf as early as 1723. Sharjah and Ras Al Khaimah used to be under one Qasimi ruler as one Emirate till 1910 when Ras Al Khaimah has been separated from Sharjah. At present Qasimi is the family name of the rulers of Sharjah and Ras Al-Khaimah (Al-Sayegh F., 1998).

There are other smaller tribes like Al Nuaim (the ruler of Ajman is from this family), Al mualla (the ruler of Umm Al Qaiwain is from this family), and Al Sharqi (the ruler of Al Fujairah from this family).

3. The Zaab

Their stronghold is the Peninsula of Al-Hamra island in Ras Al Alkhaimah, also called Zaab in the past because almost the entire population belonged to that tribe. However, late in 1968 a large part of the population of Al-Hamra island had migrated to Abu Dhabi and settled there. The Zaab of the UAE have continued to maintain close links with their relatives on the Batinah coast in Oman (Heard-Bey F., 1982).

4. The Shihuh

One other tribal group bears mentioning because it is so strikingly different from the others. This is the Shihuh, living in the rugged and remote Musandam Peninsula in Oman. They live on the border between Oman and UAE (Ras Al Khaimah) with majority in Oman. However, a number of them migrated to the UAE land for better standard of live and better education. Their isolation and unique characteristics (linguistic, social, and other) have led to wild speculation about their origins. Some have thought they were of Persian or even of Portuguese descent (Malcolm P., 1987). In fact, they are essentially Arab (Al-Sayegh F., 1998).

5. Other tribes

A lot of other Arab tribes have been given the nationality of the UAE after their migration to the country specially from the border like Al Marazeeq, Bani Hammad, and others from the northern coast (Persian side) and islands of the Arabian Gulf. Some other tribes like Bani Kaab from border of Oman, other tribes from the border between the UAE and Saudi Arabia as well, and recently some tribes from Yemen. As we can notice here that there was no clear defined border between the UAE and its neighbours. Therefore, some of these tribes used to live at that area but later after the border has been defined they had to belong to one of these countries.

APPENDIX B:

Calculation of inter-gel variation of K562 fragments at five VNTR probes. The largest and smallest fragments are in bold italic.

Table 1.1 Inter-gel variation of K562 fragments at D2S44/YNH24 in 32 agarose gels.

D2S44	Left (lane 6)		Right (lane 16)	
Gel No.	Larger fragment (bp)	Smaller fragment (bp)	Larger fragment (bp)	Smaller fragment (bp)
1	4169	2949	4088	2873
2	4201	2963	4115	2894
3	4189	2980	4184	2936
4	***4218***	2992	4157	2889
5	4159	2982	4171	***2988***
6	4181	2967	4164	2920
7	4178	2977	4196	2930
8	4169	2960	4147	2906
9	***3985***	***2810***	***3940***	***2752***
10	4179	2956	4150	2917
11	4162	2961	4172	2946
12	4208	2952	4163	2892
13	4157	2941	4190	2911
14	4198	2972	4112	2864
15	4192	2956	4152	2899
16	4189	2965	4176	2909
17	4164	2966	4181	2865
18	4193	2964	4136	2907
19	4201	***3002***	4153	2933
20	4187	2967	4184	2951
21	4165	2969	***4197***	2957
22	4179	2943	4168	2923
23	4155	2950	4175	2941
24	4181	2978	4181	2936
25	4174	2957	4174	2929
26	4174	2969	4194	2939
27	4142	2929	4183	2956
28	4158	2936	4182	2931
29	4157	2955	4142	2916
30	4186	2976	4184	2921
31	4191	2963	4109	2907
32	4155	2952	4137	2930

Mean	**4171**	**2958**	**4154**	**2913**
SD	38	31	48	40
SD%	0.9	1	1.1	1.4

Table 1.2 Inter-gel variation of K562 fragments at D7S22/G3 in 32 agarose gels.

D7S22	Left (lane 6)		Right (lane 16)	
Gel No.	Larger fragment (bp)	Smaller fragment (bp)	Larger fragment (bp)	Smaller fragment (bp)
1	7151	1949	6855	1920
2	7173	1952	7101	1913
3	7183	1961	7159	1940
4	***7223***	1965	7103	1913
5	7110	***1989***	7138	1957
6	7160	1961	7143	1945
7	7139	1950	7170	1941
8	7119	1939	7162	1933
9	7097	1951	7052	1914
10	7160	1944	7025	1946
11	7116	1945	7168	1935
12	7184	1961	7151	1927
13	7073	***1939***	***7225***	1949
14	7196	1967	6846	***1908***
15	7163	1959	6943	1933
16	7162	1953	7037	1941
17	7073	1953	6926	1938
18	7174	1960	6847	1938
19	7184	1970	7076	1938
20	7134	1958	7185	1940
21	7111	1950	6858	***1957***
22	7199	1952	6856	1944
23	7076	1957	7142	1941
24	7131	1968	7122	1933
25	6843	1948	6887	1942
26	7097	1955	7151	1932
27	***6813***	1945	6855	1957
28	7103	1948	6932	1942
29	7097	1958	7150	1950
30	7097	1955	7134	1954
31	6863	1967	***6789***	1915
32	6829	1952	6830	1940
Mean	**7101**	**1956**	**7032**	**1937**
SD	109	10	137	13

SD%	1.5	0.5	1.9	0.7

Table 1.3 Inter-gel variation of K562 fragments at D1S7/MS1 in 32 agarose gels.

D1S7	Left (lane 6)		Right (lane 16)	
Gel No.	Larger fragment (bp)	Smaller fragment (bp)	Larger fragment (bp)	Smaller fragment (bp)
1	4961	4631	4905	***4564***
2	4974	4657	***4896***	4570
3	4968	4632	4968	4638
4	4988	4669	4943	4614
5	***4901***	4600	4902	4602
6	4958	4636	4959	4618
7	4925	***4587***	5004	4651
8	4943	4616	4941	4594
9	4929	4614	4945	4601
10	4966	4632	4984	4630
11	4936	4604	4963	4618
12	***4990***	***4678***	4984	4641
13	4927	4602	5017	4671
14	4970	4642	4949	4600
15	4982	4661	4983	4638
16	4972	4645	5010	4651
17	4926	4613	5015	4667
18	4972	4643	4943	4608
19	4963	4637	4904	4591
20	4957	4635	4956	4636
21	4925	4601	4981	4645
22	4966	4633	4974	4634
23	4928	4599	4960	4635
24	4942	4621	4926	4618
25	4979	4641	***5019***	***4671***
26	4929	4609	4960	4641
27	4952	4606	4982	4635
28	4938	4609	4983	4649
29	4934	4602	4940	4601
30	4944	4617	4968	4642
31	4963	4628	4912	4564
32	4967	4624	4940	4599
Mean	**4952**	**4626**	**4960**	**4623**
SD	22	22	34	29
SD%	0.4	0.5	0.7	0.6

Table 1.4 Inter-gel variation of K562 fragments at D7S21/MS31 in 32 agarose gels.

D7S21	Left (lane 6)		Right (lane 16)	
Gel No.	Larger fragment (bp)	Smaller fragment (bp)	Larger fragment (bp)	Smaller fragment (bp)
1	7922	7060	7883	7031
2	7923	7082	7865	6993
3	7935	7112	7940	7089
4	7988	***7129***	7900	7033
5	7902	7031	7913	7058
6	7935	7089	7908	7041
7	7857	7028	7942	7094
8	7894	7053	7906	7063
9	7852	6998	7822	***6697***
10	7931	6763	7953	6967
11	7883	7028	7957	7079
12	***7995***	7110	7947	6819
13	7851	7000	8026	***7144***
14	7960	7112	7817	6784
15	7937	7085	7995	6857
16	7929	7079	***8148***	6956
17	7851	7009	7923	6897
18	7939	7087	7862	6789
19	7964	7105	7844	6993
20	7857	7053	7966	7102
21	7888	7029	7916	6779
22	7967	7111	7908	6788
23	7852	7018	7937	7084
24	7888	7053	7936	7027
25	7899	6751	7942	6818
26	7856	7023	7944	7078
27	7856	6765	7950	6791
28	7856	7019	8001	6869
29	7878	7024	7948	7078
30	***7848***	***6743***	7900	6789
31	7887	6827	***7769***	6716
32	7872	6772	7877	6770
Mean	**7902**	**7005**	**7923**	**6940**
SD	44	121	68	137
SD%	0.6	1.7	0.9	2

Table 1.5 Inter-gel variation of K562 fragments at D12S11/MS43A in 32 agarose gels.

D12S11	Left (lane 6)		Right (lane 16)	
Gel No.	Larger fragment (bp)	Smaller fragment (bp)	Larger fragment (bp)	Smaller fragment (bp)
1	13847	*5286*	13903	*5052*
2	13987	5347	13691	5291
3	*14166*	5385	13897	5297
4	13850	5323	13951	5305
5	*13588*	5302	13633	5306
6	13844	5359	13798	5328
7	13810	5328	13815	5331
8	13890	5372	13645	5334
9	13801	5378	*13558*	5343
10	13784	5333	13988	5344
11	13879	5347	13875	5348
12	13815	5311	13976	5352
13	13888	5351	14113	5352
14	13911	5364	13686	5353
15	13874	5359	13749	5360
16	14093	5356	14044	5362
17	13823	5320	13939	5363
18	13774	5339	13957	5368
19	13888	*5390*	13757	5371
20	14078	5370	13955	5379
21	13743	5319	13930	5380
22	13879	5344	13963	5385
23	13760	5327	13850	5386
24	13987	5373	13810	5390
25	13955	5375	14150	5394
26	13776	5358	14074	5398
27	13893	5368	13941	5400
28	13823	5324	14052	5401
29	13910	5344	13987	5410
30	13689	5318	14145	5422
31	13711	5321	13964	5433
32	13898	5373	*14517*	*5449*
Mean	**13863**	**5346**	**13916**	**5456**
SD	117	26	186	67
SD%	0.8	0.5	1.3	1.3

APPENDIX C:

Calculating the intra-gel variation of K562 fragments at five VNTR loci.

Table 1.1 Intra-gel variation of K562 fragments at D2S44/YNH24 in 32 gels.

Gel No.	Larger fragments (bp)		Mean	SD	SD%
	Left (lane 6)	Right (lane 16)			
1	4169	4088	4128	57	1.4
2	4201	4115	4158	61	1.5
3	4189	4184	4186	4	0.1
4	4218	4157	4188	43	1.0
5	4159	4171	4165	8	0.2
6	4181	4164	4172	12	0.3
7	4178	4196	4187	13	0.3
8	4169	4147	4158	16	0.4
9	3985	3940	3962	32	0.8
10	4179	4150	4164	21	0.5
11	4162	4172	4167	7	0.2
12	4208	4163	4185	32	0.8
13	4157	4190	4173	23	0.6
14	4198	4112	4155	61	1.5
15	4192	4152	4172	28	0.7
16	4189	4176	4182	9	0.2
17	4164	4181	4172	12	0.3
18	4193	4136	4164	40	1.0
19	4201	4153	4177	34	0.8
20	4187	4184	4185	2	0.1
21	4165	4197	4181	23	0.5
22	4179	4168	4173	8	0.2
23	4155	4175	4165	14	0.3
24	4181	4181	4181	0	0.0
25	4174	4174	4174	0	0.0
26	4174	4194	4184	14	0.3
27	4142	4183	4162	29	0.7
28	4158	4182	4170	17	0.4
29	4157	4142	4149	11	0.3
30	4186	4184	4185	1	0.0
31	4191	4109	4150	58	1.4
32	4155	4137	4146	13	0.3
			Average SD%		0.5

Table 1.2 Intra-gel variation of K562 fragments at D2S44/YNH24 in 32 gels.

Gel No.	Smaller fragments (bp)		Mean	SD	SD%
	Left (lane 6)	Right (lane 16)			
1	2949	2873	2911	54	1.8
2	2963	2894	2928	49	1.7
3	2980	2936	2958	31	1.1
4	2992	2889	2940	73	2.5
5	2982	2988	2985	4	0.1
6	2967	2920	2943	33	1.1
7	2977	2930	2953	33	1.1
8	2960	2906	2933	38	1.3
9	2810	2752	2781	41	1.5
10	2956	2917	2936	28	0.9
11	2961	2946	2953	11	0.4
12	2952	2892	2922	42	1.5
13	2941	2911	2926	21	0.7
14	2972	2864	2918	76	2.6
15	2956	2899	2927	40	1.4
16	2965	2909	2937	40	1.3
17	2966	2865	2915	71	2.4
18	2964	2907	2935	40	1.4
19	3002	2933	2967	49	1.6
20	2967	2951	2959	11	0.4
21	2969	2957	2963	8	0.3
22	2943	2923	2933	14	0.5
23	2950	2941	2945	6	0.2
24	2978	2936	2957	30	1.0
25	2957	2929	2943	20	0.7
26	2969	2939	2954	21	0.7
27	2929	2956	2942	19	0.6
28	2936	2931	2933	4	0.1
29	2955	2916	2935	28	0.9
30	2976	2921	2948	39	1.3
31	2963	2907	2935	40	1.3
32	2952	2930	2941	16	0.5
			Average SD%		**1.1**

Table 1.3 Intra-gel variation of K562 fragments at D7S22/G3 in 32 gels.

Gel No.	Larger fragments (bp)		Mean	SD	SD%
	Left (lane 6)	Right (lane 16)			
1	7151	6855	7003	209	3.0
2	7173	7101	7137	51	0.7
3	7183	7159	7171	17	0.2
4	7223	7103	7163	85	1.2
5	7110	7138	7124	20	0.3
6	7160	7143	7151	12	0.2
7	7139	7170	7154	22	0.3
8	7119	7162	7140	30	0.4
9	7097	7052	7074	32	0.4
10	7160	7025	7092	95	1.3
11	7116	7168	7142	37	0.5
12	7184	7151	7167	23	0.3
13	7073	7225	7149	107	1.5
14	7196	6846	7021	247	3.5
15	7163	6943	7053	156	2.2
16	7162	7037	7099	88	1.2
17	7073	6926	6999	104	1.5
18	7174	6847	7010	231	3.3
19	7184	7076	7130	76	1.1
20	7134	7185	7159	36	0.5
21	7111	6858	6984	179	2.6
22	7199	6856	7027	243	3.5
23	7076	7142	7109	47	0.7
24	7131	7122	7126	6	0.1
25	6843	6887	6865	31	0.5
26	7097	7151	7124	38	0.5
27	6813	6855	6834	30	0.4
28	7103	6932	7017	121	1.7
29	7097	7150	7123	37	0.5
30	7097	7134	7115	26	0.4
31	6863	6789	6826	52	0.8
32	6829	6830	6829	1	0.0
			Average SD%		**1.1**

Table 1.4 Intra-gel variation of K562 fragments at D7S22/G3 in 32 gels.

Gel No.	Smaller fragments (bp)		Mean	SD	SD%
	Left (lane 6)	Right (lane 16)			
1	1949	1920	1934	21	1.1
2	1952	1913	1932	28	1.4
3	1961	1940	1950	15	0.8
4	1965	1913	1939	37	1.9
5	1989	1957	1973	23	1.1
6	1961	1945	1953	11	0.6
7	1950	1941	1945	6	0.3
8	1939	1933	1936	4	0.2
9	1951	1914	1932	26	1.4
10	1944	1946	1945	1	0.1
11	1945	1935	1940	7	0.4
12	1961	1927	1944	24	1.2
13	1939	1949	1944	7	0.4
14	1967	1908	1937	42	2.2
15	1959	1933	1946	18	0.9
16	1953	1941	1947	8	0.4
17	1953	1938	1945	11	0.5
18	1960	1938	1949	16	0.8
19	1970	1938	1954	23	1.2
20	1958	1940	1949	13	0.7
21	1950	1957	1953	5	0.3
22	1952	1944	1948	6	0.3
23	1957	1941	1949	11	0.6
24	1968	1933	1950	25	1.3
25	1948	1942	1945	4	0.2
26	1955	1932	1943	16	0.8
27	1945	1957	1951	8	0.4
28	1948	1942	1945	4	0.2
29	1958	1950	1954	6	0.3
30	1955	1954	1954	1	0.0
31	1967	1915	1941	37	1.9
32	1952	1940	1946	8	0.4
			Average SD%		**0.8**

Table 1.5 Intra-gel variation of K562 fragments at D1S7/MS1 in 32 gels.

Gel No.	Larger fragments (bp)		Mean	SD	SD%
	Left (lane 6)	Right (lane 16)			
1	4961	4905	4933	40	0.8
2	4974	4896	4935	55	1.1
3	4968	4968	4968	0	0.0
4	4988	4943	4965	32	0.6
5	4901	4902	4901	1	0.0
6	4958	4959	4958	1	0.0
7	4925	5004	4964	56	1.1
8	4943	4941	4942	1	0.0
9	4929	4945	4937	11	0.2
10	4966	4984	4975	13	0.3
11	4936	4963	4949	19	0.4
12	4990	4984	4987	4	0.1
13	4927	5017	4972	64	1.3
14	4970	4949	4959	15	0.3
15	4982	4983	4982	1	0.0
16	4972	5010	4991	27	0.5
17	4926	5015	4970	63	1.3
18	4972	4943	4957	21	0.4
19	4963	4904	4933	42	0.8
20	4957	4956	4956	1	0.0
21	4925	4981	4953	40	0.8
22	4966	4974	4970	6	0.1
23	4928	4960	4944	23	0.5
24	4942	4926	4934	11	0.2
25	4979	5019	4999	28	0.6
26	4929	4960	4944	22	0.4
27	4952	4982	4967	21	0.4
28	4938	4983	4960	32	0.6
29	4934	4940	4937	4	0.1
30	4944	4968	4956	17	0.3
31	4963	4912	4937	36	0.7
32	4967	4940	4953	19	0.4
			Average SD%		**0.5**

Table 1.6 Intra-gel variation of K562 fragments at D1S7/MS1 in 32 gels.

Gel No.	Smaller fragments (bp)		Mean	SD	SD%
	Left (lane 6)	Right (lane 16)			
1	4631	4564	4597	47	1.0
2	4657	4570	4613	62	1.3
3	4632	4638	4635	4	0.1
4	4669	4614	4641	39	0.8
5	4600	4602	4601	1	0.0
6	4636	4618	4627	13	0.3
7	4587	4651	4619	45	1.0
8	4616	4594	4605	16	0.3
9	4614	4601	4607	9	0.2
10	4632	4630	4631	1	0.0
11	4604	4618	4611	10	0.2
12	4678	4641	4659	26	0.6
13	4602	4671	4636	49	1.1
14	4642	4600	4621	30	0.6
15	4661	4638	4649	16	0.3
16	4645	4651	4648	4	0.1
17	4613	4667	4640	38	0.8
18	4643	4608	4625	25	0.5
19	4637	4591	4614	33	0.7
20	4635	4636	4635	1	0.0
21	4601	4645	4623	31	0.7
22	4633	4634	4633	1	0.0
23	4599	4635	4617	25	0.6
24	4621	4618	4619	2	0.0
25	4641	4671	4656	21	0.5
26	4609	4641	4625	23	0.5
27	4606	4635	4620	21	0.4
28	4609	4649	4629	28	0.6
29	4602	4601	4601	1	0.0
30	4617	4642	4629	18	0.4
31	4628	4564	4596	45	1.0
32	4624	4599	4611	18	0.4
			Average SD%		**0.5**

Table 1.7 Intra-gel variation of K562 fragments at D7S21/MS31 in 32 gels.

Gel No.	Larger fragments (bp)		Mean	SD	SD%
	Left (lane 6)	Right (lane 16)			
1	7922	7883	7902	28	0.3
2	7923	7865	7894	41	0.5
3	7935	7940	7937	4	0.0
4	7988	7900	7944	62	0.8
5	7902	7913	7907	8	0.1
6	7935	7908	7921	19	0.2
7	7857	7942	7899	60	0.8
8	7894	7906	7900	8	0.1
9	7852	7822	7837	21	0.3
10	7931	7953	7942	16	0.2
11	7883	7957	7920	52	0.7
12	7995	7947	7971	34	0.4
13	7851	8026	7938	124	1.6
14	7960	7817	7888	101	1.3
15	7937	7995	7966	41	0.5
16	7929	8148	8038	155	1.9
17	7851	7923	7887	51	0.6
18	7939	7862	7900	54	0.7
19	7964	7844	7904	85	1.1
20	7857	7966	7911	77	1.0
21	7888	7916	7902	20	0.3
22	7967	7908	7937	42	0.5
23	7852	7937	7894	60	0.8
24	7888	7936	7912	34	0.4
25	7899	7942	7920	30	0.4
26	7856	7944	7900	62	0.8
27	7856	7950	7903	66	0.8
28	7856	8001	7928	103	1.3
29	7878	7948	7913	49	0.6
30	7848	7900	7874	37	0.5
31	7887	7769	7828	83	1.1
32	7872	7877	7874	4	0.0
			Average SD%		**0.6**

Table 1.8 Intra-gel variation of K562 fragments at D7S21/MS31 in 32 gels.

Gel No.	Smaller fragments (bp)		Mean	SD	SD%
	Left (lane 6)	Right (lane 16)			
1	7060	7031	7045	21	0.3
2	7082	6993	7037	63	0.9
3	7112	7089	7100	16	0.2
4	7129	7033	7081	68	1.0
5	7031	7058	7044	19	0.3
6	7089	7041	7065	34	0.5
7	7028	7094	7061	47	0.7
8	7053	7063	7058	7	0.1
9	6998	6697	6847	213	3.1
10	6763	6967	6865	144	2.1
11	7028	7079	7053	36	0.5
12	7110	6819	6964	206	3.0
13	7000	7144	7072	102	1.4
14	7112	6784	6948	232	3.3
15	7085	6857	6971	161	2.3
16	7079	6956	7017	87	1.2
17	7009	6897	6953	79	1.1
18	7087	6789	6938	211	3.0
19	7105	6993	7049	79	1.1
20	7053	7102	7077	35	0.5
21	7029	6779	6904	177	2.6
22	7111	6788	6949	228	3.3
23	7018	7084	7051	47	0.7
24	7053	7027	7040	18	0.3
25	6751	6818	6784	47	0.7
26	7023	7078	7050	39	0.6
27	6765	6791	6778	18	0.3
28	7019	6869	6944	106	1.5
29	7024	7078	7051	38	0.5
30	6743	6789	6766	33	0.5
31	6827	6716	6771	78	1.2
32	6772	6770	6771	1	0.0
			Average SD%		**1.2**

Table 1.9 Intra-gel variation of K562 fragments at D12S11/MS43A in 32 gels.

Gel No.	Larger fragments (bp)		Mean	SD	SD%
	Left (lane 6)	Right (lane 16)			
1	13847	13903	13875	40	0.3
2	13987	13691	13839	209	1.5
3	14166	13897	14031	190	1.4
4	13850	13951	13900	71	0.5
5	13588	13633	13610	32	0.2
6	13844	13798	13821	33	0.2
7	13810	13815	13812	4	0.0
8	13890	13645	13767	173	1.3
9	13801	13558	13679	172	1.3
10	13784	13988	13886	144	1.0
11	13879	13875	13877	3	0.0
12	13815	13976	13895	114	0.8
13	13888	14113	14000	159	1.1
14	13911	13686	13798	159	1.2
15	13874	13749	13811	88	0.6
16	14093	14044	14068	35	0.2
17	13823	13939	13881	82	0.6
18	13774	13957	13865	129	0.9
19	13888	13757	13822	93	0.7
20	14078	13955	14016	87	0.6
21	13743	13930	13836	132	1.0
22	13879	13963	13921	59	0.4
23	13760	13850	13805	64	0.5
24	13987	13810	13898	125	0.9
25	13955	14150	14052	138	1.0
26	13776	14074	13925	211	1.5
27	13893	13941	13917	34	0.2
28	13823	14052	13937	162	1.2
29	13910	13987	13948	54	0.4
30	13689	14145	13917	322	2.3
31	13711	13964	13837	179	1.3
32	13898	14517	14207	438	3.1
			Average SD%		0.9

Table 1.10 Intra-gel variation of K562 fragments at D12S11/MS43A in 32 gels.

Gel No.	Smaller fragments (bp)		Mean	SD	SD%
	Left (lane 6)	Right (lane 16)			
1	5286	5052	5169	165	3.2
2	5347	5291	5319	40	0.7
3	5385	5297	5341	62	1.2
4	5323	5305	5314	13	0.2
5	5302	5306	5304	3	0.1
6	5359	5328	5343	22	0.4
7	5328	5331	5329	2	0.0
8	5372	5334	5353	27	0.5
9	5378	5343	5360	25	0.5
10	5333	5344	5338	8	0.1
11	5347	5348	5347	1	0.0
12	5311	5352	5331	29	0.5
13	5351	5352	5351	1	0.0
14	5364	5353	5358	8	0.1
15	5359	5360	5359	1	0.0
16	5356	5362	5359	4	0.1
17	5320	5363	5341	30	0.6
18	5339	5368	5353	21	0.4
19	5390	5371	5380	13	0.2
20	5370	5379	5374	6	0.1
21	5319	5380	5349	43	0.8
22	5344	5385	5364	29	0.5
23	5327	5386	5356	42	0.8
24	5373	5390	5381	12	0.2
25	5375	5394	5384	13	0.2
26	5358	5398	5378	28	0.5
27	5368	5400	5384	23	0.4
28	5324	5401	5362	54	1.0
29	5344	5410	5377	47	0.9
30	5318	5422	5370	74	1.4
31	5321	5433	5377	79	1.5
32	5373	5449	5411	54	1.0
			Average SD%		**0.6**

APPENDIX D:

There are three methods have been created for allele frequency calculation Method # 1, method # 2 and method # 3. In method # 1 there are 31 fixed bins in all five loci. In method # 2 it has been rebinned so minimum of 3 counts observed in the UAE and Indian population and minimum of 2 counts observed in Pakistani population. Therefore, the minimum frequency for each bin is ~ 1%. In method # 3 the size ranges of any bin in the three populations is the same with minimum of 2 counts in each bin.

Table 1.1 Fixed bin (method # 1) for the UAE, Indian and Pakistani population data at locus D2S44 (YNH24).

Allele (bin) (D2S44)	Size range (bp)	Count			Allele frequency		
		UAE	Ind.	Pak.	UAE	Ind.	Pak.
1	0-526	0	0	0	0	0	0
2	527-653	0	0	0	0	0	0
3	654-784	0	0	0	0	0	0
4	785-910	0	0	0	0	0	0
5	911-993	0	0	0	0	0	0
6	994-1176	0	0	0	0	0	0
7	1177-1287	0	0	0	0	0	0
8	1288-1431	2	0	0	0.0058	0	0
9	1432-1568	0	0	0	0	0	0
10	1569-1672	0	0	0	0	0	0
11	1673-1861	5	0	0	0.0145	0	0
12	1862-2015	2	2	3	0.0058	0.0065	0.0134
13	2016-2213	10	6	3	0.0289	0.0195	0.0134
14	2214-2433	23	28	8	0.0665	0.0909	0.0357
15	2434-2650	31	44	24	0.0896	0.1429	0.1071
16	2651-2876	45	47	43	0.1301	0.1526	0.1920
17	2877-3101	30	45	17	0.0867	0.1461	0.0759
18	3102-3397	22	31	25	0.0636	0.1006	0.1116
19	3398-3812	31	36	28	0.0896	0.1169	0.1250
20	3813-4333	73	40	29	0.2110	0.1299	0.1295
21	4334-4716	28	15	25	0.0809	0.0487	0.1116
22	4717-5415	32	10	16	0.0925	0.0325	0.0714
23	5416-5861	3	1	0	0.0087	0.0032	0
24	5862-6442	3	0	1	0.0087	0	0.0045
25	6443-7421	5	2	1	0.0145	0.0065	0.0045
26	7422-8271	1	0	0	0.0029	0	0
27	8272-9416	0	1	0	0	0.0032	0
28	9417-11919	0	0	1	0	0	0.0045
29	11920-15004	0	0	0	0	0	0
30	15005-22621	0	0	0	0	0	0
31	22622-	0	0	0	0	0	0
Total		346	308	224	1	1	1

Table 1.2 Fixed bin (method # 1) for the UAE, Indian and Pakistani population data at locus D7S22 (G3).

Allele (bin) (D7S22)	Size range (bp)	Count			Allele frequency		
		UAE	Ind.	Pak.	UAE	Ind.	Pak.
1	0-526	0	0	0	0	0	0
2	527-653	0	0	0	0	0	0
3	654-784	0	0	0	0	0	0
4	785-910	0	0	1	0	0	0.0045
5	911-993	0	0	0	0	0	0
6	994-1176	0	0	0	0	0	0
7	1177-1287	0	0	0	0	0	0
8	1288-1431	0	1	0	0	0.0032	0
9	1432-1568	2	1	2	0.0058	0.0032	0.0089
10	1569-1672	29	47	49	0.0838	0.1526	0.2188
11	1673-1861	9	1	3	0.0260	0.0032	0.0134
12	1862-2015	1	0	0	0.0029	0	0
13	2016-2213	0	0	0	0	0	0
14	2214-2433	0	0	1	0	0	0.0045
15	2434-2650	2	2	0	0.0058	0.0065	0
16	2651-2876	4	4	4	0.0116	0.0130	0.0179
17	2877-3101	7	11	2	0.0202	0.0357	0.0089
18	3102-3397	46	21	23	0.1329	0.0682	0.1027
19	3398-3812	8	9	6	0.0231	0.0292	0.0268
20	3813-4333	5	9	6	0.0145	0.0292	0.0268
21	4334-4716	4	9	1	0.0116	0.0292	0.0045
22	4717-5415	23	13	9	0.0665	0.0422	0.0402
23	5416-5861	17	30	16	0.0491	0.0974	0.0714
24	5862-6442	25	33	14	0.0723	0.1071	0.0625
25	6443-7421	58	52	33	0.1676	0.1688	0.1473
26	7422-8271	44	23	14	0.1272	0.0747	0.0625
27	8272-9416	22	16	20	0.0636	0.0519	0.0893
28	9417-11919	32	22	12	0.0925	0.0714	0.0536
29	11920-15004	6	2	8	0.0173	0.0065	0.0357
30	15005-22621	2	0	0	0.0058	0	0
31	22622-	0	2	0	0	0.0065	0
Total		346	308	224	1	1	1

Table 1.3 Fixed bin (method # 1) for the UAE, Indian and Pakistani population data at locus D1S7 (MS1).

Allele (bin) D1S7	Size range (bp)	Count			Allele frequency		
		UAE	Ind.	Pak.	UAE	Ind.	Pak.
1	0-526	0	0	0	0	0	0
2	527-653	0	0	0	0	0	0
3	654-784	0	0	0	0	0	0
4	785-910	0	0	0	0	0	0
5	911-993	0	0	0	0	0	0
6	994-1176	0	1	2	0	0.0032	0.0089
7	1177-1287	1	1	2	0.0029	0.0032	0.0089
8	1288-1431	2	3	4	0.0058	0.0097	0.0179
9	1432-1568	3	5	1	0.0087	0.0162	0.0045
10	1569-1672	3	1	2	0.0087	0.0032	0.0089
11	1673-1861	5	4	6	0.0145	0.0130	0.0268
12	1862-2015	7	4	2	0.0202	0.0130	0.0089
13	2016-2213	11	6	7	0.0318	0.0195	0.0313
14	2214-2433	16	12	8	0.0462	0.0390	0.0357
15	2434-2650	14	7	11	0.0405	0.0227	0.0491
16	2651-2876	13	10	13	0.0376	0.0325	0.0580
17	2877-3101	13	7	6	0.0376	0.0227	0.0268
18	3102-3397	13	11	8	0.0376	0.0357	0.0357
19	3398-3812	23	17	13	0.0665	0.0552	0.0580
20	3813-4333	34	26	19	0.0983	0.0844	0.0848
21	4334-4716	24	24	9	0.0694	0.0779	0.0402
22	4717-5415	28	33	30	0.0809	0.1071	0.1339
23	5416-5861	27	23	12	0.0780	0.0747	0.0536
24	5862-6442	16	27	14	0.0462	0.0877	0.0625
25	6443-7421	25	22	12	0.0723	0.0714	0.0536
26	7422-8271	12	20	15	0.0347	0.0649	0.0670
27	8272-9416	14	10	8	0.0405	0.0325	0.0357
28	9417-11919	18	22	12	0.0520	0.0714	0.0536
29	11920-15004	12	6	6	0.0347	0.0195	0.0268
30	15005-22621	9	4	1	0.0260	0.0130	0.0045
31	22622-	3	2	1	0.0087	0.0065	0.0045
Total		**346**	**308**	**224**	**1**	**1**	**1**

Table 1.4 Fixed bin (method # 1) for the UAE, Indian and Pakistani population data at locus D7S21 (MS31).

Allele (bin) (D7S21)	Size range (bp)	Count			Allele frequency		
		UAE	Ind.	Pak.	UAE	Ind.	Pak.
1	0-526	0	0	0	0	0	0
2	527-653	0	0	0	0	0	0
3	654-784	0	0	0	0	0	0
4	785-910	0	0	0	0	0	0
5	911-993	0	0	0	0	0	0
6	994-1176	0	0	0	0	0	0
7	1177-1287	0	0	0	0	0	0
8	1288-1431	0	0	0	0	0	0
9	1432-1568	1	0	0	0.0029	0	0
10	1569-1672	0	0	0	0	0	0
11	1673-1861	5	0	0	0.0145	0	0
12	1862-2015	0	0	0	0	0	0
13	2016-2213	0	0	0	0	0	0
14	2214-2433	0	0	1	0	0	0.0045
15	2434-2650	2	0	0	0.0058	0	0
16	2651-2876	0	0	0	0	0	0
17	2877-3101	0	0	0	0	0	0
18	3102-3397	2	1	2	0.0058	0.0032	0.0089
19	3398-3812	3	2	1	0.0087	0.0065	0.0045
20	3813-4333	9	9	6	0.0260	0.0292	0.0268
21	4334-4716	12	6	5	0.0347	0.0195	0.0223
22	4717-5415	25	24	22	0.0723	0.0779	0.0982
23	5416-5861	36	29	31	0.1040	0.0942	0.1384
24	5862-6442	53	48	42	0.1532	0.1558	0.1875
25	6443-7421	92	80	45	0.2659	0.2597	0.2009
26	7422-8271	36	41	33	0.1040	0.1331	0.1473
27	8272-9416	50	57	22	0.1445	0.1851	0.0982
28	9417-11919	19	11	13	0.0549	0.0357	0.0580
29	11920-15004	1	0	1	0.0029	0	0.0045
30	15005-22621	0	0	0	0	0	0
31	22622-	0	0	0	0	0	0
Total		**346**	**308**	**224**	**1**	**1**	**1**

Table 1.5 Fixed bin (method # 1) for the UAE, Indian and Pakistani population data at locus D12S11 (MS43A).

Allele (bin) (D12S11)	Size range (bp)	Count			Allele frequency		
		UAE	Ind.	Pak.	UAE	Ind.	Pak.
1	0-526	0	0	0	0	0	0
2	527-653	0	0	0	0	0	0
3	654-784	0	0	0	0	0	0
4	785-910	0	0	0	0	0	0
5	911-993	0	0	0	0	0	0
6	994-1176	0	0	0	0	0	0
7	1177-1287	0	0	0	0	0	0
8	1288-1431	0	0	0	0	0	0
9	1432-1568	0	0	0	0	0	0
10	1569-1672	0	0	0	0	0	0
11	1673-1861	0	0	0	0	0	0
12	1862-2015	1	0	0	0.0029	0	0
13	2016-2213	0	0	0	0	0	0
14	2214-2433	0	1	0	0	0.0032	0
15	2434-2650	0	0	0	0	0	0
16	2651-2876	0	0	0	0	0	0
17	2877-3101	0	0	0	0	0	0
18	3102-3397	0	0	1	0	0	0.0045
19	3398-3812	9	1	2	0.0262	0.0032	0.0089
20	3813-4333	0	0	1	0	0	0.0045
21	4334-4716	14	12	8	0.0407	0.0390	0.0357
22	4717-5415	35	52	34	0.1017	0.1688	0.1518
23	5416-5861	6	5	3	0.0174	0.0162	0.0134
24	5862-6442	5	4	2	0.0145	0.0130	0.0089
25	6443-7421	25	14	17	0.0727	0.0455	0.0759
26	7422-8271	51	20	16	0.1483	0.0649	0.0714
27	8272-9416	98	65	52	0.2849	0.2110	0.2321
28	9417-11919	89	80	66	0.2587	0.2597	0.2946
29	11920-15004	7	44	18	0.0203	0.1429	0.0804
30	15005-22621	4	10	4	0.0116	0.0325	0.0179
31	22622-	0	0	0	0	0	0
Total		344	308	224	1	1	1

Table 2.1 Method # 2 the rebinned UAE population data for locus D2S44 (YNH24).

Allele (bin) (D2S44)	Size range (bp)	Count	Allele frequency
1	0-1861	7	0.0202
2	1862-2213	12	0.0347
3	2214-2433	23	0.0665
4	2434-2650	31	0.0896
5	2651-2876	45	0.1301
6	2877-3101	30	0.0867
7	3102-3397	22	0.0636
8	3398-3812	31	0.0896
9	3813-4333	73	0.2110
10	4334-4716	28	0.0809
11	4717-5415	32	0.0925
12	5416-5861	3	0.0087
13	5862-6442	3	0.0087
14	6443-	6	0.0173
Total		**346**	**1**

Table 2.2 Method # 2 the rebinned UAE population data for locus D7S22 (G3).

Allele (bin) (D7S22)	Size range (bp)	Count	Allele frequency
1	0-1672	31	0.0896
2	1673-1861	9	0.0260
3	1862-2650	3	0.0087
4	2651-2876	4	0.0116
5	2877-3101	7	0.0202
6	3102-3397	46	0.1329
7	3398-3812	8	0.0231
8	3813-4333	5	0.0145
9	4334-4716	4	0.0116
10	4717-5415	23	0.0665
11	5416-5861	17	0.0491
12	5862-6442	25	0.0723
13	6443-7421	58	0.1676
14	7422-8271	44	0.1272
15	8272-9416	22	0.0636
16	9417-11919	32	0.0925
17	11920-	8	0.0231
Total		**346**	**1**

Table 2.3 Method # 2 the rebinned UAE population data for locus D1S7 (MS1).

Allele (bin) (D1S7)	Size range (bp)	Count	Allele frequency
1	0-1431	3	0.0087
2	1432-1568	3	0.0087
3	1569-1672	3	0.0087
4	1673-1861	5	0.0145
5	1862-2015	7	0.0202
6	2016-2213	11	0.0318
7	2214-2433	16	0.0462
8	2434-2650	14	0.0405
9	2651-2876	13	0.0376
10	2877-3101	13	0.0376
11	3102-3397	13	0.0376
12	3398-3812	23	0.0665
13	3813-4333	34	0.0983
14	4334-4716	24	0.0694
15	4717-5415	28	0.0809
16	5416-5861	27	0.0780
17	5862-6442	16	0.0462
18	6443-7421	25	0.0723
19	7422-8271	12	0.0347
20	8272-9416	14	0.0405
21	9417-11919	18	0.0520
22	11920-15004	12	0.0347
23	15005-22621	9	0.0260
24	22622-	3	0.0087
Total		**346**	**1**

Table 2.4 Method # 2 the rebinned UAE population data for locus D7S21 (MS31).

Allele (bin) (D7S21)	Size range (bp)	Count	Allele frequency
1	0-1861	6	0.0173
2	1862-3397	4	0.0116
3	3398-3812	3	0.0087
4	3813-4333	9	0.0260
5	4334-4716	12	0.0347
6	4717-5415	25	0.0723
7	5416-5861	36	0.1040
8	5862-6442	53	0.1532
9	6443-7421	92	0.2659
10	7422-8271	36	0.1040
11	8272-9416	50	0.1445
12	9417-	20	0.0578
Total		**346**	**1**

Table 2.5 Method # 2 the rebinned UAE population data for locus D12S11 (MS43A).

Allele (bin) (D12S11)	Size range (bp)	Count	Allele frequency
1	0-3812	10	0.0291
2	3813-4716	14	0.0407
3	4717-5415	35	0.1017
4	5416-5861	6	0.0174
5	5862-6442	5	0.0145
6	6443-7421	25	0.0727
7	7422-8271	51	0.1483
8	8272-9416	98	0.2849
9	9417-11919	89	0.2587
10	11920-15004	7	0.0203
11	15005-	4	0.0116
Total		**344**	**1**

Table 2.6 Method # 2 the rebinned Indian population data for locus D2S44 (YNII24).

Allele (bin) (D2S44)	Size range (bp)	Count	Allele frequency
1	0-2213	8	0.0260
2	2214-2433	28	0.0909
3	2434-2650	44	0.1429
4	2651-2876	47	0.1526
5	2877-3101	45	0.1461
6	3102-3397	31	0.1006
7	3398-3812	36	0.1169
8	3813-4333	40	0.1299
9	4334-4716	15	0.0487
10	4717-5415	10	0.0325
11	5416-	4	0.0130
Total		**308**	**1**

Table 2.7 Method # 2 the rebinned Indian population data for locus D7S22 (G3).

Allele (bin) (D7S22)	Size range (bp)	Count	Allele frequency
1	0-1672	49	0.1591
2	1673-1650	3	0.0097
3	2651-2876	4	0.0130
4	2877-3101	11	0.0357
5	3102-3397	21	0.0682
6	3398-3812	9	0.0292
7	3813-4333	9	0.0292
8	4334-4716	9	0.0292
9	4717-5415	13	0.0422
10	5416-5861	30	0.0974
11	5862-6442	33	0.1071
12	6443-7421	52	0.1688
13	7422-8271	23	0.0747
14	8272-9416	16	0.0519
15	9417-11919	22	0.0714
16	11920-	4	0.0130
Total		**308**	**1**

Table 2.8 Method # 2 the rebinned Indian population data for locus D1S7 (MS1).

Allele (bin) (D1S7)	Size range (bp)	Count	Allele frequency
1	0-1431	5	0.0162
2	1432-1672	6	0.0195
3	1673-1861	4	0.0130
4	1862-2015	4	0.0130
5	2016-2213	6	0.0195
6	2214-2433	12	0.0390
7	2434-2650	7	0.0227
8	2651-2876	10	0.0325
9	2877-3101	7	0.0227
10	3102-3397	11	0.0357
11	3398-3812	17	0.0552
12	3813-4333	26	0.0844
13	4334-4716	24	0.0779
14	4717-5415	33	0.1071
15	5416-5861	23	0.0747
16	5862-6442	27	0.0877
17	6443-7421	22	0.0714
18	7422-8271	20	0.0649
19	8272-9416	10	0.0325
20	9417-11919	22	0.0714
21	11920-15004	6	0.0195
22	15005-	6	0.0195
Total		**308**	**1**

Table 2.9 Method # 2 the rebinned Indian population data for D7S21 (MS31).

Allele (bin) (D7S21)	Size range (bp)	Count	Allele frequency
1	0-3812	3	0.0097
2	3813-4333	9	0.0292
3	4334-4716	6	0.0195
4	4717-5415	24	0.0779
5	5416-5861	29	0.0942
6	5862-6442	48	0.1558
7	6443-7421	80	0.2597
8	7422-8271	41	0.1331
9	8272-9416	57	0.1851
10	9417-	11	0.0357
Total		308	1

Table 2.10 Method # 2 the rebinned Indian population data for locus D12S11 (MS43A).

Allele (bin) D12S11	Size range (bp)	Count	Allele frequency
1	0-4716	14	0.0455
2	4717-5415	52	0.1688
3	5416-5861	5	0.0162
4	5862-6442	4	0.0130
5	6443-7421	14	0.0455
6	7422-8271	20	0.0649
7	8272-9416	65	0.2110
8	9417-11919	80	0.2597
9	11920-15004	44	0.1429
10	15005-	10	0.0325
Total		308	1

Table 2.11 Method # 2 the rebinned Pakistani population data for locus D2S44 (YNH24).

Allele (bin) (D2S44)	Size range (bp)	Count	Allele frequency
1	0-2015	3	0.0134
2	2016-2213	3	0.0134
3	2214-2433	8	0.0357
4	2434-2650	24	0.1071
5	2651-2876	43	0.1920
6	2877-3101	17	0.0759
7	3102-3397	25	0.1116
8	3398-3812	28	0.1250
9	3813-4333	29	0.1295
10	4334-4716	25	0.1116
11	4717-5415	16	0.0714
12	5416-	3	0.0134
Total		**224**	**1**

Table 2.12 Method # 2 the rebinned Pakistani population data for D7S22 (G3).

Allele (bin) (D7S22)	Size range (bp)	Count	Allele frequency
1	0-1568	3	0.0134
2	1569-1672	49	0.2188
3	1673-1861	3	0.0134
4	1862-2876	5	0.0223
5	2877-3101	2	0.0089
6	3102-3397	23	0.1027
7	3398-3812	6	0.0268
8	3813-4716	7	0.0313
9	4717-5415	9	0.0402
10	5416-5861	16	0.0714
11	5862-6442	14	0.0625
12	6443-7421	33	0.1473
13	7422-8271	14	0.0625
14	8272-9416	20	0.0893
15	9417-11919	12	0.0536
16	11920-	8	0.0357
Total		**224**	**1**

Table 2.13 Method # 2 the rebinned Pakistani population data for D1S7 (MS1).

Allele (bin) (D1S7)	Size range (bp)	Count	Allele frequency
1	0-1176	2	0.0089
2	1177-1287	2	0.0089
3	1288-1431	4	0.0179
4	1432-1672	3	0.0134
5	1673-1861	6	0.0268
6	1862-2015	2	0.0089
7	2015-2213	7	0.0313
8	2214-2433	8	0.0357
9	2434-2650	11	0.0491
10	2651-2876	13	0.0580
11	2877-3101	6	0.0268
12	3102-3397	8	0.0357
13	3398-3812	13	0.0580
14	3813-4333	19	0.0848
15	4334-4716	9	0.0402
16	4717-5415	30	0.1339
17	5416-5861	12	0.0536
18	5862-6442	14	0.0625
19	6443-7421	12	0.0536
20	7422-8271	15	0.0670
21	8272-9416	8	0.0357
22	9417-11919	12	0.0536
23	11920-15004	6	0.0268
24	15005-	2	0.0089
Total		**224**	**1**

Table 2.14 Method # 2 the rebinned Pakistani population data for locus D7S21 (MS31).

Allele (bin) (D7S21)	Size range (bp)	Count	Allele frequency
1	0-3812	4	0.0179
2	3813-4333	6	0.0268
3	4334-4716	5	0.0223
4	4717-5415	22	0.0982
5	5416-5861	31	0.1384
6	5862-6442	42	0.1875
7	6443-7421	45	0.2009
8	7422-8271	33	0.1473
9	8272-9416	22	0.0982
10	9417-	14	0.0625
Total		224	1

Table 2.15 Method # 2 the rebinned Pakistani population data for D12S11 (MS43A).

Allele (bin) (D12S11)	Size range (bp)	Count	Allele frequency
1	0-3812	3	0.0134
2	3813-4716	9	0.0402
3	4717-5415	34	0.1518
4	5416-5861	3	0.0134
5	5862-6442	2	0.0089
6	6443-7421	17	0.0759
7	7422-8271	16	0.0714
8	8272-9416	52	0.2321
9	9417-11919	66	0.2946
10	11920-15004	18	0.0804
11	15005-	4	0.0179
Total		224	1

Table 3.1 Method # 3 the rebinned UAE, Indian and Pakistani population data for locus D2S44 (YNH24).

Allele (bin) D2S44	Size range (bp)	Count			Allele frequency		
		UAE	Ind.	Pak.	UAE	Ind.	Pak.
1	0-2015	9	2	3	0.0260	0.0065	0.0134
2	2016-2213	10	6	3	0.0289	0.0195	0.0134
3	2214-2433	23	28	8	0.0665	0.0909	0.0357
4	2434-2650	31	44	24	0.0896	0.1429	0.1071
5	2651-2876	45	47	43	0.1301	0.1526	0.1920
6	2877-3101	30	45	17	0.0867	0.1461	0.0759
7	3102-3397	22	31	25	0.0636	0.1006	0.1116
8	3398-3812	31	36	28	0.0896	0.1169	0.1250
9	3813-4333	73	40	29	0.2110	0.1299	0.1295
10	4334-4716	28	15	25	0.0809	0.0487	0.1116
11	4717-5415	32	10	16	0.0925	0.0325	0.0714
12	5416-	12	4	3	0.0347	0.0130	0.0134
Total		**346**	**308**	**224**	**1**	**1**	**1**

Table 3.2 Method # 3 the rebinned UAE, Indian and Pakistani population data for locus D7S22 (G3).

Allele(bin) D7S22	Size range (bp)	Count			Allele frequency		
		UAE	Ind.	Pak.	UAE	Ind.	Pak.
1	0-1568	2	2	3	0.0058	0.0065	0.0134
2	1569-1672	29	47	49	0.0838	0.1526	0.2188
3	1673-2650	12	3	4	0.0347	0.0097	0.0179
4	2651-2876	4	4	4	0.0116	0.0130	0.0179
5	2877-3101	7	11	2	0.0202	0.0357	0.0089
6	3102-3397	46	21	23	0.1329	0.0682	0.1027
7	3398-3812	8	9	6	0.0231	0.0292	0.0268
8	3813-4716	9	18	7	0.0260	0.0584	0.0313
9	4717-5415	23	13	9	0.0665	0.0422	0.0402
10	5416-5861	17	30	16	0.0491	0.0974	0.0714
11	5862-6442	25	33	14	0.0723	0.1071	0.0625
12	6443-7421	58	52	33	0.1676	0.1688	0.1473
13	7422-8271	44	23	14	0.1272	0.0747	0.0625
14	8272-9416	22	16	20	0.0636	0.0519	0.0893
15	9417-11919	32	22	12	0.0925	0.0714	0.0536
16	11920-	8	4	8	0.0231	0.0130	0.0357
Total		**346**	**308**	**224**	**1**	**1**	**1**

Table 3.3 Method # 3 the rebinned UAE, Indian and Pakistani population data for locus D1S7 (MS1).

Allele (bin) D1S7	Size range (bp)	Count			Allele frequency		
		UAE	Ind.	Pak.	UAE	Ind.	Pak.
1	0-1431	3	5	8	0.0087	0.0162	0.0357
2	1432-1672	6	6	3	0.0173	0.0195	0.0134
3	1673-1861	5	4	6	0.0145	0.0130	0.0268
4	1862-2015	7	4	2	0.0202	0.0130	0.0089
5	2016-2213	11	6	7	0.0318	0.0195	0.0313
6	2214-2433	16	12	8	0.0462	0.0390	0.0357
7	2434-2650	14	7	11	0.0405	0.0227	0.0491
8	2651-2876	13	10	13	0.0376	0.0325	0.0580
9	2877-3101	13	7	6	0.0376	0.0227	0.0268
10	3102-3397	13	11	8	0.0376	0.0357	0.0357
11	3398-3812	23	17	13	0.0665	0.0552	0.0580
12	3813-4333	34	26	19	0.0983	0.0844	0.0848
13	4334-4716	24	24	9	0.0694	0.0779	0.0402
14	4717-5415	28	33	30	0.0809	0.1071	0.1339
15	5416-5861	27	23	12	0.0780	0.0747	0.0536
16	5862-6442	16	27	14	0.0462	0.0877	0.0625
17	6443-7421	25	22	12	0.0723	0.0714	0.0536
18	7422-8271	12	20	15	0.0347	0.0649	0.0670
19	8272-9416	14	10	8	0.0405	0.0325	0.0357
20	9417-11919	18	22	12	0.0520	0.0714	0.0536
21	11920-15004	12	6	6	0.0347	0.0195	0.0268
22	15005-	12	6	2	0.0347	0.0195	0.0089
Total		**346**	**308**	**224**	**1**	**1**	**1**

Table 3.4 Method # 3 the rebinned UAE, Indian and Pakistani population data for locus D7S21 (MS31).

Allele (bin) D7S21	Size range (bp)	Count			Allele frequency		
		UAE	Ind.	Pak.	UAE	Ind.	Pak.
1	0-3812	13	3	4	0.0376	0.0097	0.0179
2	3813-4333	9	9	6	0.0260	0.0292	0.0268
3	4334-4716	12	6	5	0.0347	0.0195	0.0223
4	4717-5415	25	24	22	0.0723	0.0779	0.0982
5	5416-5861	36	29	31	0.1040	0.0942	0.1384
6	5862-6442	53	48	42	0.1532	0.1558	0.1875
7	6443-7421	92	80	45	0.2659	0.2597	0.2009
8	7422-8271	36	41	33	0.1040	0.1331	0.1473
9	8272-9416	50	57	22	0.1445	0.1851	0.0982
10	9417-	20	11	14	0.0578	0.0357	0.0625
Total		346	308	224	1	1	1

Table 3.5 Method # 3 the rebinned UAE, Indian and Pakistani population data for D12S11 (MS43A).

Allele (bin) D12S11	Size range (bp)	Count			Allele frequency		
		UAE	Ind.	Pak.	UAE	Ind.	Pak.
1	0-3812	10	2	3	0.0291	0.0065	0.0134
2	3813-4716	14	12	9	0.0407	0.0390	0.0402
3	4717-5415	35	52	34	0.1017	0.1688	0.1518
4	5416-5861	6	5	3	0.0174	0.0162	0.0134
5	5862-6442	5	4	2	0.0145	0.0130	0.0089
6	6443-7421	25	14	17	0.0727	0.0455	0.0759
7	7422-8271	51	20	16	0.1483	0.0649	0.0714
8	8272-9416	98	65	52	0.2849	0.2110	0.2321
9	9417-11919	89	80	66	0.2587	0.2597	0.2946
10	11920-15004	7	44	18	0.0203	0.1429	0.0804
11	15005-	4	10	4	0.0116	0.0325	0.0179
Total		344	308	224	1	1	1

APPENDIX E:

GENOTYPE FREQUENCIES (VNTRs)

1.1 The standard deviation (SD) of observed genotypes for the three populations, UAE, Indian and Pakistani at D12S11 locus. The maximum frequencies of genotypes observed are in bold italic.

Genotype observed (D12S11)	UAE (n=172)		Indian (n=154)		Pakistani (n=112)	
	Frequency	SD	Frequency	SD	Frequency	SD
1,2	0.0058	0.0113	0	0	0	0
1,4	0	0	0	0	0.0089	0.0174
1,5	0.0116	0.0160	0	0	0	0
1,6	0.0173	0.0194	0.0065	0.0127	0.0268	0.0299
1,7	0.0173	0.0194	0.013	0.0179	0	0
1,8	0.0116	0.0160	0	0	0	0
1,9	0.0116	0.0160	0	0	0	0
2,2	0.0058	0.0113	0.0065	0.0127	0	0
2,5	0	0	0	0	0.0089	0.0174
2,6	0.0173	0.0194	0.013	0.0179	0	0
2,7	0.0116	0.0160	0.026	0.0251	0.0268	0.0299
2,8	0.0058	0.0113	0	0	0.0089	0.0174
2,9	0	0	0.0065	0.0127	0.0089	0.0174
3,3	0.0058	0.0113	0	0	0	0
3,4	0.0058	0.0113	0.0065	0.0127	0.0089	0.0174
3,6	0.0058	0.0113	0.0065	0.0127	0.0179	0.0246
3,7	0.0231	0.0224	0.013	0.0179	0.0089	0.0174
3,8	0	0	0.0065	0.0127	0	0
3,9	0.0116	0.0160	0.0065	0.0127	0.0089	0.0174
3,10	0.0116	0.0160	0	0	0	0
4,4	0.0058	0.0113	0.0065	0.0127	0.0179	0.0246
4,5	0.0116	0.0160	0.0065	0.0127	0.0357	0.0344
4,6	0.0231	0.0224	0.0065	0.0127	0.0268	0.0299
4,7	0.052	0.0331	0.0455	0.0329	0.0536	0.0417
4,8	0.0173	0.0194	0.0195	0.0218	0.0179	0.0246
4,9	0.0231	0.0224	0.0455	0.0329	0.0089	0.0174
4,10	0	0	0.013	0.0179	0	0
5,5	0.0231	0.0224	0.0065	0.0127	0.0089	0.0174
5,6	0.0289	0.0250	0.039	0.0306	0.0446	0.0382
5,7	0.0462	0.0313	0.0649	0.0389	0.0446	0.0382

5,8	0.0058	0.0113	0.039	0.0306	0.0625	0.0448
5,9	0.0462	0.0313	0.026	0.0251	0.0357	0.0344
5,10	0.0116	0.0160	0	0	0.0268	0.0299
6,6	0.0116	0.0160	0.0195	0.0218	0.0357	0.0344
6,7	***0.0925***	***0.0432***	0.0909	0.0454	0.0536	0.0417
6,8	0.0405	0.0294	0.0455	0.0329	0.0446	0.0382
6,9	0.0462	0.0313	0.0519	0.0350	0.0536	0.0417
6,10	0.0116	0.0160	0.013	0.0179	0.0357	0.0344
7,7	0.0636	0.0364	0.039	0.0306	0.0446	0.0382
7,8	0.0405	0.0294	0.0519	0.0350	***0.0804***	***0.0504***
7,9	***0.0925***	***0.0432***	***0.1234***	***0.0519***	0.0357	0.0344
7,10	0.0289	0.0250	0.013	0.0179	0.0089	0.0174
8,8	0.0289	0.0250	0.0195	0.0218	0.0268	0.0299
8,9	0.0173	0.0194	0.0519	0.0350	0	0
8,10	0.0116	0.0160	0.013	0.0179	0.0268	0.0299
9,9	0.0116	0.0160	0.0195	0.0218	0.0089	0.0174
9,10	0.0173	0.0194	0.0195	0.0218	0.0268	0.0299
10,10	0.0116	0.0160	0	0	0	0

1.2 The observed genotype and genotype frequencies for the three populations, UAE, Indian and Pakistani at D7S21, D7S22, D1S7 and D2S44 loci. The maximum frequencies of genotypes observed are in bold italic.

Genotype observed (D7S21)	UAE (n=173)		Indian (n=154)		Pakistani (n=112)	
	Count	Frequency	Count	Frequency	Count	Frequency
1,1	1	0.0058	0	0	1	0.0089
1,2	1	0.0058	0	0	0	0
1,3	1	0.0058	0	0	0	0
1,5	1	0.0058	0	0	0	0
1,6	1	0.0058	0	0	0	0
1,7	1	0.0058	0	0	1	0.0089
1,8	2	0.0116	1	0.0065	0	0
1,9	1	0.0058	0	0	0	0
1,10	0	0	1	0.0065	0	0
2,3	2	0.0116	2	0.013	3	0.0268
2,4	1	0.0058	1	0.0065	1	0.0089
2,5	1	0.0058	0	0	0	0
2,6	0	0	1	0.0065	2	0.0179
2,7	4	0.0231	1	0.0065	1	0.0089
2,8	2	0.0116	2	0.013	0	0
2,9	2	0.0116	4	0.026	2	0.0179
2,10	1	0.0058	1	0.0065	0	0
3,3	3	0.0173	6	0.039	7	0.0625
3,4	0	0	1	0.0065	1	0.0089
3,6	1	0.0058	4	0.026	2	0.0179
3,7	6	0.0347	6	0.039	1	0.0089
3,8	10	0.0578	6	0.039	7	0.0625
3,9	6	0.0347	13	0.0844	6	0.0536
3,10	1	0.0058	7	0.0455	0	0
3,11	2	0.0116	1	0.0065	0	0
4,6	1	0.0058	0	0	0	0
4,7	0	0	2	0.013	0	0
4,8	4	0.0231	1	0.0065	1	0.0089
5,6	1	0.0058	0	0	0	0
5,7	1	0.0058	0	0	0	0
5,8	1	0.0058	1	0.0065	0	0
5,9	0	0	2	0.013	0	0
5,10	0	0	1	0.0065	2	0.0179
6,6	1	0.0058	1	0.0065	0	0
6,7	4	0.0231	2	0.013	0	0
6,8	8	0.0462	4	0.026	4	0.0357
6,9	7	0.0405	1	0.0065	8	0.0714
6,10	0	0	0	0	1	0.0089
7,7	5	0.0289	0	0	1	0.0089

7,8	15	0.0867	5	0.0325	4	0.0357
7,9	10	0.0578	2	0.013	7	0.0625
7,10	0	0	2	0.013	0	0
8,8	11	0.0636	5	0.0325	8	0.0714
8,9	***29***	***0.1676***	***18***	***0.1169***	***14***	***0.125***
8,10	5	0.0289	17	0.1104	6	0.0536
9,9	16	0.0925	12	0.0779	10	0.0893
9,10	0	0	9	0.0584	6	0.0536
9,11	2	0.0116	7	0.0455	3	0.0268
10,10	0	0	3	0.0195	1	0.0089
10,11	0	0	0	0	1	0.0089
11,11	0	0	1	0.0065	0	0

Genotype observed (D7S22)	**UAE (n=173)**		**Indian (n=154)**		**Pakistani (n=112)**	
	Count	**Frequency**	**Count**	**Frequency**	**Count**	**Frequency**
1,2	0	0	0	0	1	0.0089
1,10	0	0	1	0.0065	0	0
1,12	0	0	1	0.0065	1	0.0089
1,15	1	0.0058	0	0	0	0
1,16	1	0.0058	0	0	0	0
2,2	1	0.0058	5	0.0325	5	0.0446
2,3	0	0	0	0	3	0.0268
2,4	1	0.0058	1	0.0065	0	0
2,5	0	0	2	0.013	0	0
2,6	5	0.0289	5	0.0325	5	0.0446
2,7	0	0	1	0.0065	0	0
2,8	0	0	3	0.0195	0	0
2,9	1	0.0058	1	0.0065	0	0
2,10	6	0.0347	5	0.0325	3	0.0268
2,11	2	0.0116	3	0.0195	5	0.0446
2,12	4	0.0231	11	0.0714	8	0.0714
2,13	1	0.0058	2	0.013	2	0.0179
2,14	3	0.0173	1	0.0065	6	0.0536
2,15	4	0.0231	2	0.013	2	0.0179
2,16	0	0	0	0	4	0.0357
3,5	1	0.0058	0	0	0	0
3,6	2	0.0116	0	0	0	0
3,7	1	0.0058	0	0	0	0
3,10	2	0.0116	0	0	0	0
3,11	3	0.0173	0	0	0	0
3,12	1	0.0058	1	0.0065	0	0
3,13	2	0.0116	1	0.0065	0	0
3,14	0	0	0	0	1	0.0089
3,15	0	0	1	0.0065	0	0
4,6	1	0.0058	0	0	0	0
4,10	0	0	0	0	1	0.0089

4,11	0	0	1	0.0065	0	0
4,12	0	0	1	0.0065	1	0.0089
4,13	1	0.0058	1	0.0065	2	0.0179
4,14	1	0.0058	0	0	0	0
5,5	1	0.0058	0	0	0	0
5,6	1	0.0058	0	0	0	0
5,7	0	0	0	0	1	0.0089
5,9	0	0	1	0.0065	1	0.0089
5,10	1	0.0058	2	0.013	0	0
5,12	0	0	3	0.0195	0	0
5,13	1	0.0058	1	0.0065	0	0
5,14	0	0	1	0.0065	0	0
5,15	1	0.0058	1	0.0065	0	0
6,6	8	0.0462	2	0.013	1	0.0089
6,8	0	0	0	0	1	0.0089
6,9	6	0.0347	0	0	1	0.0089
6,10	1	0.0058	0	0	5	0.0446
6,11	2	0.0116	4	0.026	1	0.0089
6,12	2	0.0116	4	0.026	3	0.0268
6,13	4	0.0231	1	0.0065	4	0.0357
6,14	2	0.0116	1	0.0065	1	0.0089
6,15	4	0.0231	2	0.013	0	0
7,7	1	0.0058	1	0.0065	0	0
7,10	0	0	1	0.0065	0	0
7,11	1	0.0058	2	0.013	1	0.0089
7,12	1	0.0058	2	0.013	2	0.0179
7,13	2	0.0116	0	0	1	0.0089
7,14	1	0.0058	1	0.0065	1	0.0089
8,8	1	0.0058	1	0.0065	0	0
8,9	0	0	2	0.013	1	0.0089
8,10	1	0.0058	1	0.0065	0	0
8,11	1	0.0058	3	0.0195	1	0.0089
8,12	3	0.0173	2	0.013	2	0.0179
8,13	2	0.0116	1	0.0065	0	0
8,14	0	0	0	0	1	0.0089
8,15	0	0	2	0.013	1	0.0089
8,16	0	0	2	0.013	0	0
9,9	2	0.0116	0	0	0	0
9,10	1	0.0058	2	0.013	1	0.0089
9,11	2	0.0116	3	0.0195	0	0
9,12	2	0.0116	0	0	2	0.0179
9,13	2	0.0116	1	0.0065	1	0.0089
9,14	4	0.0231	2	0.013	1	0.0089
9,15	0	0	1	0.0065	0	0
9,16	1	0.0058	0	0	1	0.0089
10,10	1	0.0058	2	0.013	0	0
10,11	0	0	2	0.013	1	0.0089
10,12	2	0.0116	2	0.013	1	0.0089
10,13	0	0	4	0.026	2	0.0179

10,14	0	0	2	0.013	0	0
10,15	1	0.0058	3	0.0195	2	0.0179
10,16	0	0	1	0.0065	0	0
11,11	3	0.0173	1	0.0065	0	0
11,12	5	0.0289	7	0.0455	1	0.0089
11,13	0	0	4	0.026	1	0.0089
11,14	1	0.0058	1	0.0065	2	0.0179
11,15	2	0.0116	1	0.0065	1	0.0089
12,12	13	0.0751	4	0.026	3	0.0268
12,13	5	0.0289	3	0.0195	0	0
12,14	3	0.0173	2	0.013	1	0.0089
12,15	2	0.0116	5	0.0325	3	0.0268
12,16	2	0.0116	0	0	1	0.0089
13,13	6	0.0347	0	0	0	0
13,14	1	0.0058	3	0.0195	0	0
13,15	8	0.0462	1	0.0065	1	0.0089
13,16	3	0.0173	0	0	0	0
14,4	2	0.0116	1	0.0065	1	0.0089
14,15	1	0.0058	0	0	2	0.0179
14,16	1	0.0058	0	0	2	0.0179
15,15	4	0.0231	1	0.0065	0	0
15,16	0	0	1	0.0065	0	0

Genotype observed (D1S7)	UAE (n=173)		Indian (n=154)		Pakistani (n=112)	
	Count	Frequency	Count	Frequency	Count	Frequency
1,3	0	0	0	0	1	0.0089
1,6	0	0	0	0	1	0.0089
1,7	1	0.0058	0	0	1	0.0089
1,9	0	0	0	0	1	0.0089
1,11	0	0	0	0	3	0.0268
1,12	1	0.0058	1	0.0065	0	0
1,15	0	0	1	0.0065	0	0
1,16	0	0	1	0.0065	1	0.0089
1,19	0	0	1	0.0065	0	0
1,20	0	0	1	0.0065	0	0
1,21	1	0.0058	0	0	0	0
2,5	1	0.0058	1	0.0065	0	0
2,7	0	0	1	0.0065	0	0
2,10	1	0.0058	0	0	0	0
2,13	1	0.0058	1	0.0065	0	0
2,15	1	0.0058	1	0.0065	0	0
2,16	0	0	0	0	1	0.0089
2,17	0	0	1	0.0065	0	0
2,18	0	0	0	0	1	0.0089
2,19	0	0	0	0	1	0.0089
2,20	1	0.0058	0	0	0	0

2,22	1	0.0058	1	0.0065	0	0
3,4	0	0	0	0	1	0.0089
3,8	0	0	1	0.0065	1	0.0089
3,9	1	0.0058	0	0	0	0
3,10	0	0	0	0	1	0.0089
3,12	1	0.0058	0	0	1	0.0089
3,13	1	0.0058	2	0.013	0	0
3,18	0	0	1	0.0065	0	0
3,20	1	0.0058	0	0	1	0.0089
3,21	1	0.0058	0	0	0	0
4,7	1	0.0058	1	0.0065	0	0
4,8	1	0.0058	0	0	0	0
4,11	1	0.0058	1	0.0065	0	0
4,17	1	0.0058	1	0.0065	0	0
4,18	0	0	0	0	1	0.0089
4,20	3	0.0173	1	0.0065	0	0
5,5	1	0.0058	0	0	0	0
5,6	1	0.0058	1	0.0065	0	0
5,12	0	0	0	0	1	0.0089
5,13	0	0	1	0.0065	0	0
5,14	2	0.0116	0	0	2	0.0179
5,15	1	0.0058	0	0	0	0
5,16	1	0.0058	1	0.0065	3	0.0268
5,17	1	0.0058	1	0.0065	1	0.0089
5,19	1	0.0058	0	0	0	0
5,20	0	0	1	0.0065	0	0
5,22	1	0.0058	0	0	0	0
6,7	0	0	0	0	1	0.0089
6,8	0	0	0	0	1	0.0089
6,9	0	0	2	0.013	0	0
6,12	5	0.0289	1	0.0065	1	0.0089
6,13	1	0.0058	0	0	0	0
6,14	0	0	3	0.0195	0	0
6,15	3	0.0173	2	0.013	0	0
6,16	1	0.0058	1	0.0065	0	0
6,17	2	0.0116	1	0.0065	1	0.0089
6,18	0	0	1	0.0065	0	0
6,19	0	0	0	0	1	0.0089
6,20	2	0.0116	0	0	0	0
6,21	1	0.0058	0	0	1	0.0089
6,22	0	0	0	0	1	0.0089
7,7	1	0.0058	0	0	0	0
7,10	0	0	1	0.0065	0	0
7,11	1	0.0058	0	0	0	0
7,12	1	0.0058	1	0.0065	2	0.0179
7,14	3	0.0173	0	0	3	0.0268
7,15	2	0.0116	0	0	1	0.0089
7,17	0	0	1	0.0065	1	0.0089
7,18	2	0.0116	1	0.0065	1	0.0089

7,20	0	0	1	0.0065	0	0
7,22	1	0.0058	0	0	1	0.0089
8,9	1	0.0058	0	0	0	0
8,10	0	0	1	0.0065	0	0
8,11	1	0.0058	0	0	0	0
8,12	1	0.0058	2	0.013	1	0.0089
8,13	1	0.0058	1	0.0065	1	0.0089
8,14	4	0.0231	2	0.013	3	0.0268
8,15	0	0	1	0.0065	1	0.0089
8,16	1	0.0058	1	0.0065	1	0.0089
8,19	0	0	0	0	1	0.0089
8,20	2	0.0116	0	0	1	0.0089
8,21	0	0	1	0.0065	2	0.0179
8,22	1	0.0058	0	0	0	0
9,10	1	0.0058	1	0.0065	0	0
9,11	1	0.0058	0	0	1	0.0089
9,12	1	0.0058	0	0	0	0
9,13	2	0.0116	2	0.013	0	0
9,14	0	0	0	0	1	0.0089
9,15	2	0.0116	0	0	1	0.0089
9,16	1	0.0058	1	0.0065	1	0.0089
9,17	2	0.0116	0	0	0	0
9,20	1	0.0058	0	0	1	0.0089
9,21	0	0	1	0.0065	0	0
10,10	0	0	0	0	1	0.0089
10,11	1	0.0058	0	0	1	0.0089
10,13	0	0	1	0.0065	0	0
10,14	0	0	1	0.0065	1	0.0089
10,15	3	0.0173	2	0.013	0	0
10,16	1	0.0058	3	0.0195	3	0.0268
10,18	1	0.0058	0	0	0	0
10,19	2	0.0116	0	0	0	0
10,20	2	0.0116	0	0	0	0
10,21	0	0	1	0.0065	0	0
10,22	1	0.0058	0	0	0	0
11,11	0	0	1	0.0065	0	0
11,12	5	0.0289	3	0.0195	1	0.0089
11,13	1	0.0058	2	0.013	1	0.0089
11,14	4	0.0231	3	0.0195	1	0.0089
11,15	2	0.0116	2	0.013	0	0
11,16	0	0	1	0.0065	2	0.0179
11,17	2	0.0116	1	0.0065	1	0.0089
11,18	1	0.0058	1	0.0065	1	0.0089
11,19	1	0.0058	1	0.0065	0	0
11,20	2	0.0116	0	0	1	0.0089
12,12	0	0	2	0.013	0	0
12,13	5	0.0289	0	0	0	0
12,14	2	0.0116	1	0.0065	4	0.0357
12,15	1	0.0058	2	0.013	4	0.0357

12,16	4	0.0231	2	0.013	0	0
12,17	1	0.0058	1	0.0065	0	0
12,18	1	0.0058	1	0.0065	2	0.0179
12,19	1	0.0058	2	0.013	0	0
12,20	1	0.0058	4	0.026	2	0.0179
12,21	3	0.0173	0	0	0	0
12,22	0	0	1	0.0065	0	0
13,13	0	0	1	0.0065	0	0
13,14	1	0.0058	0	0	1	0.0089
13,15	1	0.0058	1	0.0065	0	0
13,16	2	0.0116	4	0.026	1	0.0089
13,17	4	0.0231	1	0.0065	2	0.0179
13,18	1	0.0058	1	0.0065	0	0
13,19	0	0	2	0.013	1	0.0089
13,20	0	0	3	0.0195	1	0.0089
13,21	0	0	0	0	1	0.0089
13,22	3	0.0173	0	0	0	0
14,14	0	0	3	0.0195	2	0.0179
14,15	4	0.0231	3	0.0195	1	0.0089
14,16	2	0.0116	4	0.026	0	0
14,17	3	0.0173	3	0.0195	2	0.0179
14,18	0	0	1	0.0065	3	0.0268
14,19	1	0.0058	1	0.0065	2	0.0179
14,20	0	0	5	0.0325	2	0.0179
14,21	1	0.0058	0	0	0	0
14,22	1	0.0058	0	0	0	0
15,15	3	0.0173	0	0	1	0.0089
15,16	0	0	1	0.0065	0	0
15,17	0	0	2	0.013	1	0.0089
15,18	0	0	1	0.0065	1	0.0089
15,19	1	0.0058	2	0.013	0	0
15,20	0	0	1	0.0065	0	0
15,21	0	0	1	0.0065	0	0
16,16	0	0	1	0.0065	0	0
16,18	0	0	2	0.013	0	0
16,19	1	0.0058	0	0	0	0
16,20	1	0.0058	1	0.0065	0	0
16,21	1	0.0058	1	0.0065	1	0.0089
16,22	0	0	1	0.0065	0	0
17,17	3	0.0173	2	0.013	0	0
17,18	0	0	0	0	1	0.0089
17,19	1	0.0058	1	0.0065	1	0.0089
17,20	1	0.0058	2	0.013	1	0.0089
17,21	1	0.0058	0	0	0	0
17,22	0	0	2	0.013	0	0
18,18	1	0.0058	3	0.0195	1	0.0089
18,19	1	0.0058	0	0	1	0.0089
18,20	1	0.0058	2	0.013	1	0.0089
18,21	1	0.0058	1	0.0065	0	0

18,22	1	0.0058	1	0.0065	0	0
19,19	2	0.0116	0	0	0	0
20,21	0	0	0	0	1	0.0089
21,21	1	0.0058	0	0	0	0
22,22	1	0.0058	0	0	0	0

Genotype observed (D2S44)	**UAE (n=173)**		**Indian (n=154)**		**Pakistani (n=112)**	
	Count	**Frequency**	**Count**	**Frequency**	**Count**	**Frequency**
1,3	0	0	1	0.0065	0	0
1,4	1	0.0058	0	0	0	0
1,5	1	0.0058	1	0.0065	0	0
1,6	0	0	0	0	1	0.0089
1,7	2	0.0116	0	0	2	0.0179
1,9	2	0.0116	0	0	0	0
1,10	2	0.0116	0	0	0	0
1,11	1	0.0058	0	0	0	0
2,3	0	0	1	0.0065	0	0
2,4	1	0.0058	0	0	0	0
2,5	3	0.0173	1	0.0065	1	0.0089
2,6	1	0.0058	1	0.0065	0	0
2,7	1	0.0058	0	0	0	0
2,8	1	0.0058	0	0	2	0.0179
2,9	3	0.0173	2	0.013	0	0
2,11	0	0	1	0.0065	0	0
3,3	3	0.0173	1	0.0065	2	0.0179
3,4	3	0.0173	5	0.0325	0	0
3,5	5	0.0289	4	0.026	0	0
3,6	2	0.0116	2	0.013	0	0
3,7	1	0.0058	4	0.026	0	0
3,8	1	0.0058	2	0.013	1	0.0089
3,9	4	0.0231	2	0.013	2	0.0179
3,10	0	0	5	0.0325	1	0.0089
3,12	1	0.0058	0	0	0	0
4,4	2	0.0116	4	0.026	3	0.0268
4,5	5	0.0289	3	0.0195	4	0.0357
4,6	1	0.0058	5	0.0325	0	0
4,7	1	0.0058	3	0.0195	1	0.0089
4,8	5	0.0289	7	0.0455	2	0.0179
4,9	3	0.0173	9	0.0584	4	0.0357
4,10	3	0.0173	2	0.013	6	0.0536
4,11	2	0.0116	1	0.0065	1	0.0089
4,12	0	0	1	0.0065	0	0
5,5	1	0.0058	5	0.0325	6	0.0536
5,6	5	0.0289	8	0.0519	1	0.0089
5,7	3	0.0173	2	0.013	9	0.0804
5,8	5	0.0289	10	0.0649	3	0.0268

5,9	7	0.0405	6	0.039	3	0.0268
5,10	3	0.0173	2	0.013	7	0.0625
5,11	4	0.0231	0	0	2	0.0179
5,12	2	0.0116	0	0	1	0.0089
6,6	3	0.0173	6	0.039	1	0.0089
6,7	0	0	5	0.0325	3	0.0268
6,8	1	0.0058	2	0.013	4	0.0357
6,9	5	0.0289	5	0.0325	2	0.0179
6,10	2	0.0116	3	0.0195	1	0.0089
6,11	5	0.0289	3	0.0195	3	0.0268
6,12	2	0.0116	0	0	0	0
7,7	1	0.0058	2	0.013	0	0
7,8	2	0.0116	6	0.039	3	0.0268
7,9	5	0.0289	3	0.0195	2	0.0179
7,10	2	0.0116	0	0	2	0.0179
7,11	3	0.0173	2	0.013	2	0.0179
7,12	0	0	1	0.0065	1	0.0089
8,8	2	0.0116	0	0	3	0.0268
8,9	8	0.0462	6	0.039	3	0.0268
8,10	2	0.0116	0	0	3	0.0268
8,11	1	0.0058	2	0.013	1	0.0089
8,12	1	0.0058	1	0.0065	0	0
9,9	8	0.0462	2	0.013	3	0.0268
9,10	6	0.0347	2	0.013	2	0.0179
9,11	11	0.0636	0	0	5	0.0446
9,12	3	0.0173	1	0.0065	0	0
10,10	3	0.0173	0	0	1	0.0089
10,11	1	0.0058	1	0.0065	1	0.0089
10,12	1	0.0058	0	0	0	0
11,11	2	0.0116	0	0	0	0
11,12	0	0	0	0	1	0.0089

APPENDIX F:

Table 1.1 Shows the mean size of the allelic ladder and allele designation in the GenePrint™ PowerPlex™ 1.2 system (Promega, Madison, WI, USA) using capillary electrophoresis.

D5S818		D13S317	
Allele mean size (bp)	**Allele designation**	**Allele mean size (bp)**	**Allele designation**
112.03	7	167.82	7
116.13	8	171.79	8
120.32	9	175.76	9
124.51	10	179.59	10
128.62	11	183.74	11
132.77	12	187.75	12
136.87	13	191.74	13
140.96	14	195.75	14
145.13	15	199.77	15

D7S820		D16S539	
Allele mean size (bp)	**Allele designation**	**Allele mean size (bp)**	**Allele designation**
211.06	6	261.00	5
215.08	7	273.19	8
219.10	8	277.26	9
223.14	9	281.22	10
227.16	10	285.22	11
231.22	11	289.23	12
235.24	12	293.21	13
239.34	13	297.23	14
243.33	14	301.32	15

vWA		THO1	
Allele mean size (bp)	Allele designation	Allele mean size (bp)	Allele designation
126.32	11	176.25	5
134.35	13	180.28	6
138.38	14	184.20	7
142.35	15	188.15	8
146.44	16	192.12	9
150.42	17	195.01	9.3
154.43	18	196.06	10
158.43	19	200.12	11
162.42	20		
166.30	21		

TPOX		CSF1PO	
Allele mean size (bp)	Allele designation	Allele mean size (bp)	Allele designation
222.27	6	291.16	6
226.22	7	295.08	7
230.22	8	299.04	8
234.25	9	303.00	9
238.27	10	306.99	10
242.27	11	310.99	11
246.29	12	315.01	12
250.32	13	318.92	13
		322.98	14
		327.01	15

X		Y	
Allele mean size (bp)	Allele designation	Allele mean size (bp)	Allele designation
209.45	X	215.81	Y

APPENDIX G:

ALLELE AND GENOTYPE FREQUENCIES (STRs)

1.1 Allele frequencies in the UAE Arab, Egyptian, Indian and Pakistani individuals for the loci D7S820, D13S317, D16S539, vWA, THO1, TPOX and CSF1PO. The maximum and minimum allele frequencies are shown in bold italic.

Allele frequencies in 229 UAE Arab individuals for the locus D7S820.

Allele	Observed (D7S820)	Allele frequency	SD	Heterozygosity	Heterozygosity frequency
7	9	0.0197	0.0065	9	0.0393
8	83	0.1812	0.0180	63	0.2751
8.3	*2*	***0.0044***	0.0031	2	0.0087
9	50	0.1092	0.0146	40	0.1747
10	*138*	***0.3013***	0.0214	100	0.4367
11	99	0.2162	0.0192	71	0.3100
12	63	0.1376	0.0161	51	0.2227
13	12	0.0262	0.0075	10	0.0437
14	*2*	***0.0044***	0.0031	2	0.0087

Allele frequencies in 120 Egyptian individuals for the locus D7S820.

Allele	Observed (D7S820)	Allele frequency	SD	Heterozygosity	Heterozygosity frequency
7	5	0.0208	0.0181	5	0.0417
8	42	0.1750	0.0481	36	0.3000
9	22	0.0917	0.0365	16	0.1333
10	*80*	***0.3333***	0.0596	60	0.5000
11	67	0.2792	0.0568	43	0.3583
12	23	0.0958	0.0372	17	0.1417
13	*1*	***0.0042***	0.0082	1	0.0083

Allele frequencies in 192 Indian individuals for the locus D7S820.

Allele	Observed (D7S820)	Allele frequency	SD	Heterozygosity	Heterozygosity frequency
7	21	0.0547	0.0116	19	0.0990
8	90	0.2344	0.0216	64	0.3333
9	23	0.0599	0.0121	19	0.0990
10	***97***	***0.2526***	0.0222	69	0.3594
11	78	0.2031	0.0206	66	0.3438
12	57	0.1484	0.0181	47	0.2448
13	17	0.0443	0.0105	15	0.0781
14	***1***	***0.0026***	0.0026	1	0.0052

Allele frequencies in 197 Pakistani individuals for the locus D7S820.

Allele	Observed (D7S820)	Allele frequency	SD	Heterozygosity	Heterozygosity frequency
7	13	0.0330	0.0090	13	0.0660
8	85	0.2157	0.0207	71	0.3604
9	39	0.0990	0.0150	37	0.1878
10	***99***	***0.2513***	0.0219	79	0.4010
11	91	0.2310	0.0212	67	0.3401
12	62	0.1574	0.0183	52	0.2640
13	***5***	***0.0127***	0.0064	5	0.0254

Allele frequencies in 229 UAE Arab individuals for the locus D13S317.

Allele	Observed (D13S317)	Allele frequency	SD	Heterozygosity	Heterozygosity frequency
8	70	0.1528	0.0168	58	0.2533
9	16	0.0349	0.0086	16	0.0699
10	34	0.0742	0.0122	30	0.1310
11	128	0.2795	0.0210	90	0.3930
12	***141***	***0.3079***	0.0216	107	0.4672
13	52	0.1135	0.0148	44	0.1921
14	16	0.0349	0.0086	14	0.0611
15	***1***	***0.0022***	0.0022	1	0.0044

Allele frequencies in 121 Egyptian individuals for the locus D13S317.

Allele	Observed (D13S317)	Allele frequency	SD	Heterozygosity	Heterozygosity frequency
7	***1***	***0.0041***	0.0081	1	0.0083
8	38	0.1570	0.0458	32	0.2645
9	13	0.0537	0.0284	13	0.1074
10	21	0.0868	0.0355	19	0.1570
11	69	0.2851	0.0569	49	0.4050
12	***70***	***0.2893***	0.0571	60	0.4959
13	20	0.0826	0.0347	14	0.1157
14	9	0.0372	0.0238	7	0.0579
15	***1***	***0.0041***	0.0081	1	0.0083

Allele frequencies in 193 Indian individuals for the locus D13S317.

Allele	Observed (D13S317)	Allele frequency	SD	Heterozygosity	Heterozygosity frequency
7	6	0.0155	0.0063	6	0.0311
8	64	0.1658	0.0189	54	0.2798
9	34	0.0881	0.0144	30	0.1554
10	27	0.0699	0.0130	25	0.1295
11	110	0.2850	0.0230	70	0.3627
12	**111**	**0.2876**	0.0230	75	0.3886
13	23	0.0596	0.0120	23	0.1192
14	10	0.0259	0.0081	10	0.0518
15	***1***	***0.0022***	0.0022	1	0.0052

Allele frequencies in 197 Pakistani individuals for the locus D13S317.

Allele	Observed (D13S317)	Allele frequency	SD	Heterozygosity	Heterozygosity frequency
8	70	0.1777	0.0193	60	0.3046
9	27	0.0685	0.0127	27	0.1371
10	27	0.0685	0.0127	21	0.1066
11	107	0.2716	0.0224	69	0.3503
12	***109***	***0.2766***	0.0225	89	0.4518
13	48	0.1218	0.0165	42	0.2132
14	***6***	***0.0152***	0.0062	6	0.0305

Allele frequencies in 229 UAE Arab individuals for the locus D16S539.

Allele	Observed (D16S539)	Allele frequency	SD	Heterozygosity	Heterozygosity frequency
8	24	0.0524	0.0104	20	0.0873
9	74	0.1616	0.0172	64	0.2795
10	43	0.0939	0.0136	39	0.1703
11	***159***	***0.3472***	0.0222	97	0.4236
12	84	0.1834	0.0181	70	0.3057
13	70	0.1528	0.0168	58	0.2533
14	3	0.0066	0.0038	3	0.0131
15	***1***	***0.0022***	0.0022	1	0.0044

Allele frequencies in 120 Egyptian individuals for the locus D16S359.

Allele	Observed (D16S359)	Allele frequency	SD	Heterozygosity	Heterozygosity frequency
8	8	0.0333	0.0227	8	0.0667
9	45	0.1875	0.0494	35	0.2917
10	16	0.0667	0.0316	14	0.1167
11	***74***	***0.3083***	0.0584	52	0.4333
12	63	0.2625	0.0557	45	0.3750
13	30	0.1250	0.0418	22	0.1833
14	3	0.0125	0.0141	3	0.0250
15	***1***	***0.0042***	0.0082	1	0.0083

Allele frequencies in 193 Indian individuals for the locus D16S539.

Allele	Observed (D16S539)	Allele frequency	SD	Heterozygosity	Heterozygosity frequency
8	28	0.0725	0.0132	28	0.1451
9	58	0.1503	0.0182	48	0.2487
10	35	0.0907	0.0146	33	0.1710
11	***136***	***0.3523***	0.0243	92	0.4767
12	74	0.1917	0.0200	58	0.3005
13	43	0.1114	0.0160	41	0.2124
14	11	0.0285	0.0085	11	0.0570
15	***1***	***0.0026***	0.0026	1	0.0052

Allele frequencies in 197 Pakistani individuals for the locus D16S539.

Allele	Observed (D16S539)	Allele frequency	SD	Heterozygosity	Heterozygosity frequency
8	20	0.0508	0.0111	20	0.1015
9	67	0.1701	0.0189	53	0.2690
10	51	0.1294	0.0169	41	0.2081
11	***118***	***0.2995***	0.0231	78	0.3959
12	95	0.2411	0.0215	61	0.3096
13	39	0.0990	0.0150	31	0.1574
14	3	0.0076	0.0044	3	0.0152
15	***1***	***0.0025***	0.0025	1	0.0051

Allele frequencies in 228 UAE Arab individuals for the locus vWA.

Allele	Observed (vWA)	Allele frequency	SD	Heterozygosity	Heterozygosity frequency
13	***1***	***0.0022***	0.0022	1	0.0044
14	17	0.0373	0.0087	13	0.0570
15	48	0.1053	0.0144	40	0.1754
16	138	0.3026	0.0215	88	0.3860
17	***148***	***0.3246***	0.0219	84	0.3684
18	74	0.1623	0.0173	62	0.2719
19	27	0.0592	0.0111	25	0.1096
20	3	0.0066	0.0038	3	0.0132

Allele frequencies in 121 Egyptian individuals for the locus vWA.

Allele	Observed (vWA)	Allele frequency	SD	Heterozygosity	Heterozygosity frequency
13	***1***	***0.0041***	0.0081	1	0.0083
14	18	0.0744	0.0331	16	0.1322
15	40	0.1653	0.0469	28	0.2314
16	61	0.2521	0.0548	45	0.3719
17	***62***	***0.2562***	0.0550	46	0.3802
18	34	0.1405	0.0438	30	0.2479
19	20	0.0826	0.0347	20	0.1653
20	5	0.0207	0.0179	5	0.0413
21	***1***	***0.0041***	0.0081	1	0.0083

Allele frequencies in 194 Indian individuals for the locus vWA.

Allele	Observed (vWA)	Allele frequency	SD	Heterozygosity	Heterozygosity frequency
9	***1***	***0.0026***	0.0026	1	0.0052
13	2	0.0052	0.0037	2	0.0103
14	57	0.1469	0.0180	45	0.2320
15	38	0.0979	0.0151	36	0.1856
16	***94***	***0.2423***	0.0218	66	0.3402
17	75	0.1933	0.0200	63	0.3247
18	75	0.1933	0.0200	61	0.3144
19	41	0.1057	0.0156	39	0.2010
20	2	0.0052	0.0037	2	0.0103
21	3	0.0077	0.0044	3	0.0155

Allele frequencies in 197 Pakistani individuals for the locus vWA.

Allele	Observed (vWA)	Allele frequency	SD	Heterozygosity	Heterozygosity frequency
13	***3***	***0.0076***	0.0044	3	0.0152
14	40	0.1015	0.0152	36	0.1827
15	37	0.0939	0.0147	27	0.1371
16	92	0.2335	0.0213	66	0.3350
17	***112***	***0.2843***	0.0227	86	0.4365
18	66	0.1675	0.0188	60	0.3046
19	37	0.0939	0.0147	35	0.1777
20	7	0.0178	0.0067	7	0.0355

Allele frequencies in 229 UAE Arab individuals for the locus THO1.

Allele	Observed (THO1)	Allele frequency	SD	Heterozygosity	Heterozygosity frequency
6	***130***	***0.2838***	0.0211	80	0.3493
7	93	0.2031	0.0188	77	0.3362
8	67	0.1463	0.0165	55	0.2402
9	95	0.2074	0.0189	77	0.3362
9.3	66	0.1441	0.0164	56	0.2445
10	6	0.0131	0.0053	4	0.0175
11	***1***	***0.0022***	0.0022	1	0.0044

Allele frequencies in 121 Egyptian individuals for the locus THO1.

Allele	Observed (THO1)	Allele frequency	SD	Heterozygosity	Heterozygosity frequency
6	50	0.2066	0.0510	36	0.2975
7	55	0.2273	0.0528	47	0.3884
8	26	0.1074	0.0390	22	0.1818
9	***78***	***0.3223***	0.0589	50	0.4132
9.3	26	0.1074	0.0390	16	0.1322
10	***7***	***0.0289***	0.0211	7	0.0579

Allele frequencies in 193 Indian individuals for the locus THO1.

Allele	Observed (THO1)	Allele frequency	SD	Heterozygosity	Heterozygosity frequency
6	109	0.2824	0.0229	79	0.4093
7	60	0.1554	0.0184	50	0.2591
8	43	0.1114	0.0160	39	0.2021
9	***117***	***0.3031***	0.0234	77	0.3990
9.3	53	0.1373	0.0175	47	0.2435
10	3	0.0078	0.0045	3	0.0155
10.3	***1***	***0.0026***	0.0026	1	0.0052

Allele frequencies in 197 Pakistani individuals for the locus THO1.

Allele	Observed (THO1)	Allele frequency	SD	Heterozygosity	Heterozygosity frequency
6	***107***	***0.2716***	0.0226	79	0.4010
7	72	0.1827	0.0197	52	0.2640
8	***55***	***0.1396***	0.0176	39	0.1980
9	100	0.2538	0.0222	60	0.3046
9.3	60	0.1523	0.0183	44	0.2234

Allele frequencies in 229 UAE Arab individuals for the locus TPOX.

Allele	Observed (TPOX)	Allele frequency	SD	Heterozygosity	Heterozygosity frequency
6	5	0.0109	0.0049	5	0.0218
7	***1***	***0.0022***	0.0022	1	0.0044
8	***206***	***0.4498***	0.0232	116	0.5066
9	54	0.1179	0.0151	46	0.2006
10	51	0.1114	0.0147	47	0.2052
11	125	0.2729	0.0208	91	0.3974
12	16	0.0349	0.0086	16	0.0699

Allele frequencies in 121 Egyptian individuals for the locus TPOX.

Allele	Observed (TPOX)	Allele frequency	SD	Heterozygosity	Heterozygosity frequency
8	***123***	***0.5083***	0.0630	55	0.4545
9	45	0.1860	0.0490	29	0.2397
10	16	0.0661	0.0313	16	0.1322
11	52	0.2149	0.0518	42	0.3471
12	***6***	***0.0248***	0.0196	6	0.0496

Allele frequencies in 194 Indian individuals for the locus TPOX.

Allele	Observed (TPOX)	Allele frequency	SD	Heterozygosity	Heterozygosity frequency
8	129	0.3325	0.0239	77	0.3969
9	47	0.1211	0.0166	39	0.2010
10	34	0.0876	0.0144	32	0.1649
11	*154*	*0.3969*	0.0248	90	0.4639
12	21	0.0541	0.0115	19	0.0979
13	2	0.0052	0.0037	2	0.0103
14	***1***	***0.0026***	0.0026	1	0.0052

Allele frequencies in 197 Pakistani individuals for the locus TPOX.

Allele	Observed (TPOX)	Allele frequency	SD	Heterozygosity	Heterozygosity frequency
8	***162***	***0.4112***	0.0248	100	0.5076
9	60	0.1523	0.0181	58	0.2944
10	24	0.0609	0.0120	22	0.1117
11	139	0.3477	0.0240	105	0.5330
12	***11***	***0.0279***	0.0083	11	0.0558

Allele frequencies in 227 UAE Arab individuals for the locus CSF1PO.

Allele	Observed (CSF1PO)	Allele frequency	SD	Heterozygosity	Heterozygosity frequency
7	***1***	***0.0022***	0.0022	1	0.0044
8	2	0.0044	0.0031	2	0.0088
9	16	0.0352	0.0086	16	0.0705
10	128	0.2819	0.0211	88	0.3877
11	***139***	***0.3062***	0.0216	101	0.4449
12	135	0.2974	0.0215	103	0.4537
13	27	0.0595	0.0111	27	0.1189
14	5	0.0110	0.0049	5	0.0220
16	***1***	***0.0022***	0.0022	1	0.0044

Allele frequencies in 119 Egyptian individuals for the locus CSF1PO.

Allele	Observed (CSF1PO)	Allele frequency	SD	Heterozygosity	Heterozygosity frequency
7	***3***	***0.0126***	0.0141	3	0.0252
8	5	0.0210	0.0182	5	0.0420
9	5	0.0210	0.0182	5	0.0420
10	64	0.2689	0.0563	48	0.4034
11	***83***	***0.3487***	0.0605	51	0.4286
12	63	0.2647	0.0561	41	0.3445
13	12	0.0504	0.0278	8	0.0672
14	***3***	***0.0126***	0.0141	3	0.0252

Allele frequencies in 191 Indian individuals for the locus CSF1PO.

Allele	Observed (CSF1PO)	Allele frequency	SD	Heterozygosity	Heterozygosity frequency
9	8	0.0209	0.0073	8	0.0419
10	63	0.1649	0.1377	49	0.2565
11	116	0.3037	0.0235	76	0.3979
12	***156***	***0.4084***	0.0251	80	0.4188
13	30	0.0785	0.0138	28	0.1466
14	7	0.0183	0.0069	7	0.0366
15	***2***	***0.0052***	0.0037	0	0

Allele frequencies in 197 Pakistani individuals for the locus CSF1PO.

Allele	Observed (CSF1PO)	Allele frequency	SD	Heterozygosity	Heterozygosity frequency
8	***2***	***0.0051***	0.0036	2	0.0102
9	7	0.0178	0.0067	7	0.0355
10	95	0.2411	0.0215	53	0.2690
11	111	0.2817	0.0227	75	0.3807
12	***138***	***0.3503***	0.0240	82	0.4162
13	37	0.0939	0.0147	33	0.1675
14	4	0.0102	0.0051	4	0.0203

1.2 Genotype frequencies in the UAE Arab, Egyptian, Indian and Pakistani individuals for the loci D5S818, D7S820, D13S317, D16S539, vWA, THO1, TPOX and CSF1PO.

Genotype observed (D5S818)	UAE (n=229)			Egyptian (n=121)		
	Count	Frequency	SD	Count	Frequency	SD
8,10	1	0.0044	0.0086	0	0	0
8,11	2	0.0087	0.0120	2	0.0165	0.0227
8,12	3	0.0131	0.0147	2	0.0165	0.0227
8,14	0	0	0	1	0.0083	0.0162
9,9	1	0.0044	0.0086	0	0	0
9,10	3	0.0131	0.0147	2	0.0165	0.0227
9,11	4	0.0175	0.0170	5	0.0413	0.0355
9,12	4	0.0175	0.0170	8	0.0661	0.0443
9,13	5	0.0218	0.0189	2	0.0165	0.0227
10,10	4	0.0175	0.0170	5	0.0413	0.0355
10,11	11	0.048	0.0277	5	0.0413	0.0355
10,12	23	0.1004	0.0389	14	0.1157	0.0570
10,13	11	0.048	0.0277	2	0.0165	0.0227
10,15	0	0	0	1	0.0083	0.0162
11,11	19	0.083	0.0357	3	0.0248	0.0277
11,12	34	0.1485	0.0461	15	0.124	0.0587
11,13	20	0.0873	0.0366	9	0.0744	0.0468
12,12	38	0.1659	0.0482	18	0.1488	0.0634
12,13	34	0.1485	0.0461	18	0.1488	0.0634
12,14	2	0.0087	0.0120	2	0.0165	0.0227
13,13	6	0.0262	0.0207	6	0.0496	0.0387
13,14	3	0.0131	0.0147	0	0	0
13,15	0	0	0	1	0.0083	0.0162
14,14	1	0.0044	0.0086	0	0	0

Genotype observed (D7S820)	UAE (n=229)			Egyptian (n=120)		
	Count	Frequency	SD	Count	Frequency	SD
7,8	2	0.0087	0.0120	0	0	0
7,8.3	1	0.0044	0.0086	0	0	0
7,10	2	0.0087	0.0120	4	0.0333	0.0321
7,11	3	0.0131	0.0147	1	0.0083	0.0162
7,12	1	0.0044	0.0086	0	0	0
8,8	10	0.0437	0.0265	3	0.025	0.0279
8,8.3	1	0.0044	0.0086	0	0	0
8,9	9	0.0393	0.0252	3	0.025	0.0279
8,10	24	0.1048	0.0397	19	0.1583	0.0653
8,11	15	0.0655	0.0320	12	0.1	0.0537
8,12	11	0.048	0.0277	2	0.0167	0.0279
8,13	1	0.0044	0.0086	0	0	0
9,9	5	0.0218	0.0189	3	0.025	0.0279
9,10	19	0.083	0.0357	7	0.0583	0.0419
9,11	7	0.0306	0.0223	5	0.0417	0.0358
9,12	5	0.0218	0.0189	1	0.0083	0.0162
10,10	19	0.083	0.0357	10	0.0833	0.0494
10,11	28	0.1223	0.0424	20	0.1667	0.0667
10,12	21	0.0917	0.0374	9	0.075	0.0471
10,13	5	0.0218	0.0189	1	0.0083	0.0162
10,14	1	0.0044	0.0086	0	0	0
11,11	14	0.0611	0.0310	12	0.1	0.0537
11,12	13	0.0568	0.0299	5	0.0417	0.0358
11,13	4	0.0175	0.0170	0	0	0
11,14	1	0.0044	0.0086	0	0	0
12,12	6	0.0262	0.0207	3	0.025	0.0279
13,13	1	0.0044	0.0086	0	0	0

Genotype observed (D13S317)	UAE (n=229)			Egyptian (n=121)		
	Count	Frequency	SD	Count	Frequency	SD
7,11	0	0	0	1	0.0083	0.0162
8,8	6	0.0262	0.0207	3	0.0248	0.0277
8,9	2	0.0087	0.0120	1	0.0083	0.0162
8,10	6	0.0262	0.0207	1	0.0083	0.0162
8,11	17	0.0742	0.0339	12	0.0992	0.0533
8,12	24	0.1048	0.0397	14	0.1157	0.0570
8,13	8	0.0349	0.0238	4	0.0331	0.0319
8,14	1	0.0044	0.0086	0	0	0
9,10	1	0.0044	0.0086	1	0.0083	0.0162
9,11	4	0.0175	0.0170	4	0.0331	0.0319
9,12	7	0.0306	0.0223	5	0.0413	0.0355
9,13	2	0.0087	0.0120	1	0.0083	0.0162
9,14	1	0.0044	0.0086	1	0.0083	0.0162
10,10	2	0.0087	0.0120	1	0.0083	0.0162
10,11	10	0.0437	0.0265	5	0.0413	0.0355
10,12	8	0.0349	0.0238	11	0.0909	0.0512
10,13	4	0.0175	0.0170	0	0	0
10,14	1	0.0044	0.0086	0	0	0
10,15	0	0	0	1	0.0083	0.0162
11,11	19	0.083	0.0357	10	0.0826	0.0490
11,12	46	0.2009	0.0519	22	0.1818	0.0687
11,13	8	0.0349	0.0238	4	0.0331	0.0319
11,14	5	0.0218	0.0189	1	0.0083	0.0162
12,12	17	0.0742	0.0339	5	0.0413	0.0355
12,13	19	0.083	0.0357	4	0.0331	0.0319
12,14	3	0.0131	0.0147	4	0.0331	0.0319
13,13	4	0.0175	0.0170	3	0.0248	0.0277
13,14	2	0.0087	0.0120	1	0.0083	0.0162
13,15	1	0.0044	0.0086	0	0	0
14,14	1	0.0044	0.0086	1	0.0083	0.0162

Genotype observed (D16S539)	UAE (n=229)			Egyptian (n=120)		
	Count	Frequency	SD	Count	Frequency	SD
8,8	2	0.0087	0.0120	0	0	0
8,9	5	0.0218	0.0189	0	0	0
8,10	2	0.0087	0.0120	0	0	0
8,11	8	0.0349	0.0238	5	0.0417	0.0358
8,12	5	0.0218	0.0189	2	0.0167	0.0279
8,13	0	0	0	1	0.0083	0.0162
9,9	5	0.0218	0.0189	5	0.0417	0.0358
9,10	11	0.048	0.0277	2	0.0167	0.0279
9,11	19	0.083	0.0357	15	0.125	0.0592
9,12	15	0.0655	0.0320	12	0.1	0.0537
9,13	14	0.0611	0.0310	5	0.0417	0.0358
9,14	0	0	0	1	0.0083	0.0162
10,10	2	0.0087	0.0120	1	0.0083	0.0162
10,11	11	0.048	0.0277	8	0.0667	0.0446
10,12	9	0.0393	0.0252	1	0.0083	0.0162
10,13	5	0.0218	0.0189	3	0.025	0.0279
10,14	1	0.0044	0.0086	0	0	0
11,11	31	0.1354	0.0443	11	0.0917	0.0516
11,12	30	0.131	0.0437	19	0.1583	0.0653
11,13	28	0.1223	0.0424	4	0.0333	0.0321
11,14	1	0.0044	0.0086	1	0.0083	0.0162
12,12	7	0.0306	0.0223	9	0.075	0.0471
12,13	10	0.0437	0.0265	9	0.075	0.0471
12,14	0	0	0	1	0.0083	0.0162
12,15	1	0.0044	0.0086	1	0.0083	0.0162
13,13	6	0.0262	0.0207	4	0.0333	0.0321
13,14	1	0.0044	0.0086	0	0	0

Genotype observed (vWA)	UAE (n=228)			Egyptian (n=121)		
	Count	Frequency	SD	Count	Frequency	SD
13,16	0	0	0	1	0.0083	0.0162
13,18	1	0.0044	0.0086	0	0	0
14,14	2	0.0088	0.0121	1	0.0083	0.0162
14,15	2	0.0088	0.0121	1	0.0083	0.0162
14,16	4	0.0175	0.0170	4	0.0331	0.0319
14,17	4	0.0175	0.0170	8	0.0661	0.0443
14,18	2	0.0088	0.0121	1	0.0083	0.0162
14,19	1	0.0044	0.0086	2	0.0165	0.0227
15,15	4	0.0175	0.0170	6	0.0496	0.0387
15,16	12	0.0526	0.0290	7	0.0579	0.0416
15,17	10	0.0439	0.0266	9	0.0744	0.0468
15,18	13	0.057	0.0301	8	0.0661	0.0443
15,19	2	0.0088	0.0121	3	0.0248	0.0277
15,20	1	0.0044	0.0086	0	0	0
16,16	25	0.1096	0.0406	8	0.0661	0.0443
16,17	42	0.1842	0.0503	14	0.1157	0.0570
16,18	23	0.1009	0.0391	10	0.0826	0.0490
16,19	6	0.0263	0.0208	8	0.0661	0.0443
16,20	1	0.0044	0.0086	1	0.0083	0.0162
17,17	32	0.1404	0.0451	8	0.0661	0.0443
17,18	17	0.0746	0.0341	9	0.0744	0.0468
17,19	9	0.0395	0.0253	3	0.0248	0.0277
17,20	0	0	0	3	0.0248	0.0277
18,18	6	0.0263	0.0208	2	0.0165	0.0227
18,19	6	0.0263	0.0208	2	0.0165	0.0227
18,20	1	0.0044	0.0086	0	0	0
19,19	1	0.0044	0.0086	0	0	0
19,20	0	0	0	1	0.0083	0.0162
19,21	0	0	0	1	0.0083	0.0162

Genotype observed (THO1)	UAE (n=229)			Egyptian (n=121)		
	Count	Frequency	SD	Count	Frequency	SD
6,6	25	0.1092	0.0404	7	0.0579	0.0416
6,7	22	0.0961	0.0382	13	0.1074	0.0552
6,8	17	0.0742	0.0339	5	0.0413	0.0355
6,9	23	0.1004	0.0389	13	0.1074	0.0552
6,9.3	17	0.0742	0.0339	3	0.0248	0.0277
6,10	0	0	0	2	0.0165	0.0227
6,11	1	0.0044	0.0086	0	0	0
7,7	8	0.0349	0.0238	4	0.0331	0.0319
7,8	14	0.0611	0.0310	7	0.0579	0.0416
7,9	23	0.1004	0.0389	20	0.1653	0.0662
7,9.3	16	0.0699	0.0330	4	0.0331	0.0319
7,10	2	0.0087	0.0120	3	0.0248	0.0277
8,8	6	0.0262	0.0207	2	0.0165	0.0227
8,9	15	0.0655	0.0320	8	0.0661	0.0443
8,9.3	8	0.0349	0.0238	1	0.0083	0.0162
8,10	1	0.0044	0.0086	1	0.0083	0.0162
9,9	9	0.0393	0.0252	14	0.1157	0.0570
9,9.3	15	0.0655	0.0320	8	0.0661	0.0443
9,10	1	0.0044	0.0086	1	0.0083	0.0162
9.3,9.3	5	0.0218	0.0189	5	0.0413	0.0355
10,10	1	0.0044	0.0086	0	0	0

Genotype observed (TPOX)	UAE (n=229)			Egyptian (n=121)		
	Count	Frequency	SD	Count	Frequency	SD
6,8	1	0.0044	0.0086	0	0	0
6,10	1	0.0044	0.0086	0	0	0
6,11	2	0.0087	0.0120	0	0	0
6,12	1	0.0044	0.0086	0	0	0
7,9	1	0.0044	0.0086	0	0	0
8,8	45	0.1965	0.0515	34	0.281	0.0801
8,9	22	0.0961	0.0382	18	0.1488	0.0634
8,10	21	0.0917	0.0374	6	0.0496	0.0387
8,11	58	0.2533	0.0563	28	0.2314	0.075
8,12	13	0.0568	0.0299	3	0.0248	0.0277
9,9	4	0.0175	0.0170	8	0.0661	0.0443
9,10	8	0.0349	0.0238	3	0.0248	0.0277
9,11	15	0.0655	0.0320	7	0.0579	0.0416
9,12	0	0	0	1	0.0083	0.0162
10,10	2	0.0087	0.0120	0	0	0
10,11	16	0.0699	0.0330	6	0.0496	0.0387
10,12	1	0.0044	0.0086	1	0.0083	0.0162
11,11	17	0.0742	0.0339	5	0.0413	0.0355
11,12	1	0.0044	0.0086	1	0.0083	0.0162
12,12	1	0.0044	0.0086	0	0	0

Genotype observed (CSF1PO)	UAE (n=227)			Egyptian (n=119)		
	Count	Frequency	SD	Count	Frequency	SD
7,10	0	0	0	1	0.0084	0.0164
7,11	0	0	0	1	0.0084	0.0164
7,12	1	0.0044	0.0086	1	0.0084	0.0164
8,10	0	0	0	1	0.0084	0.0164
8,11	0	0	0	2	0.0168	0.0231
8,12	0	0	0	1	0.0084	0.0164
8,13	2	0.0088	0.0122	1	0.0084	0.0164
9,10	2	0.0088	0.0122	1	0.0084	0.0164
9,11	7	0.0308	0.0225	3	0.0252	0.0282
9,12	6	0.0264	0.0209	1	0.0084	0.0164
9,13	1	0.0044	0.0086	0	0	0
10,10	20	0.0881	0.0369	8	0.0672	0.0450
10,11	38	0.1674	0.0486	25	0.2101	0.0732
10,12	36	0.1586	0.0475	16	0.1345	0.0613
10,13	9	0.0396	0.0254	2	0.0168	0.0231
10,14	3	0.0132	0.0149	2	0.0168	0.0231
11,11	19	0.0837	0.0360	16	0.1345	0.0613
11,12	49	0.2159	0.0535	18	0.1513	0.0644
11,13	5	0.022	0.0191	2	0.0168	0.0231
11,14	2	0.0088	0.0122	0	0	0
12,12	16	0.0705	0.0333	11	0.0924	0.0520
12,13	10	0.0441	0.0267	3	0.0252	0.0282
12,14	0	0	0	1	0.0084	0.0164
12,16	1	0.0044	0.0086	0	0	0
13,13	0	0	0	2	0.0168	0.0231

Genotype observed (D5S818)	Indian (n=192)			Pakistani (n=197)		
	Count	Frequency	SD	Count	Frequency	SD
7,11	0	0	0	1	0.0051	0.0099
7,12	0	0	0	1	0.0051	0.0099
8,10	0	0	0	1	0.0051	0.0099
8,11	1	0.0052	0.0102	0	0	0
8,12	1	0.0052	0.0102	0	0	0
9,9	0	0	0	1	0.0051	0.0099
9,10	2	0.0104	0.0144	3	0.0152	0.0171
9,11	5	0.026	0.0225	9	0.0457	0.0292
9,12	6	0.0313	0.0246	10	0.0508	0.0307
9,13	3	0.0156	0.0175	2	0.0102	0.0140
10,10	1	0.0052	0.0102	1	0.0051	0.0099
10,11	12	0.0625	0.0342	14	0.0711	0.0359
10,12	16	0.0833	0.0391	19	0.0964	0.0412
10,13	8	0.0417	0.0283	7	0.0355	0.0258
10,14	2	0.0104	0.0144	0	0	0
11,11	20	0.1042	0.0432	20	0.1015	0.0422
11,12	40	0.2083	0.0574	46	0.2335	0.0591
11,13	27	0.1406	0.0492	16	0.0812	0.0381
11,14	0	0	0	2	0.0102	0.0140
12,12	16	0.0833	0.0391	18	0.0914	0.0402
12,13	17	0.0885	0.0402	16	0.0812	0.0381
12,14	2	0.0104	0.0144	3	0.0152	0.0171
12,15	1	0.0052	0.0102	0	0	0
13,13	12	0.0625	0.0342	7	0.0355	0.0258

Genotype observed (D7S820)	Indian (n=192)			Pakistani (n=197)		
	Count	Frequency	SD	Count	Frequency	SD
7,7	1	0.0052	0.0102	0	0	0
7,8	2	0.0104	0.0144	3	0.0152	0.0171
7,9	2	0.0104	0.0144	4	0.0203	0.0197
7,10	4	0.0208	0.0202	3	0.0152	0.0171
7,11	8	0.0417	0.0283	3	0.0152	0.0171
7,12	2	0.0104	0.0144	0	0	0
7,13	1	0.0052	0.0102	0	0	0
8,8	13	0.0677	0.0355	7	0.0355	0.0258
8,9	5	0.026	0.0225	5	0.0254	0.0220
8,10	25	0.1302	0.0476	33	0.1675	0.0522
8,11	16	0.0833	0.0391	15	0.0761	0.0370
8,12	12	0.0625	0.0342	14	0.0711	0.0359
8,13	4	0.0208	0.0202	1	0.0051	0.0099
9,9	2	0.0104	0.0144	1	0.0051	0.0099
9,10	2	0.0104	0.0144	7	0.0355	0.0258
9,11	7	0.0365	0.0265	10	0.0508	0.0307
9,12	3	0.0156	0.0175	11	0.0558	0.0321
10,10	14	0.0729	0.0368	10	0.0508	0.0307
10,11	20	0.1042	0.0432	22	0.1117	0.0440
10,12	14	0.0729	0.0368	12	0.0609	0.0334
10,13	3	0.0156	0.0175	3	0.0152	0.0171
10,14	1	0.0052	0.0102	0	0	0
11,11	6	0.0313	0.0246	12	0.0609	0.0334
11,12	12	0.0625	0.0342	15	0.0761	0.0370
11,13	3	0.0156	0.0175	1	0.0051	0.0099
12,12	5	0.026	0.0225	5	0.0254	0.0220
12,13	4	0.0208	0.0202	0	0	0
13,13	1	0.0052	0.0102	0	0	0

Genotype observed (D13S317)	Indian (n=193)			Pakistani (n=197)		
	Count	Frequency	SD	Count	Frequency	SD
7,8	1	0.0052	0.0102	0	0	0
7,10	1	0.0052	0.0102	0	0	0
7,11	3	0.0155	0.0174	0	0	0
7,13	1	0.0052	0.0102	0	0	0
8,8	5	0.0259	0.0224	5	0.0254	0.0220
8,9	6	0.0311	0.0245	4	0.0203	0.0197
8,10	6	0.0311	0.0245	8	0.0406	0.0276
8,11	16	0.0829	0.0389	13	0.066	0.0347
8,12	21	0.1088	0.0439	19	0.0964	0.0412
8,13	0	0	0	13	0.066	0.0347
8,14	4	0.0207	0.0201	3	0.0152	0.0171
9,9	2	0.0104	0.0143	0	0	0
9,10	2	0.0104	0.0143	1	0.0051	0.0099
9,11	10	0.0518	0.0173	6	0.0305	0.0240
9,12	9	0.0466	0.0297	12	0.0609	0.0334
9,13	1	0.0052	0.0102	4	0.0203	0.0197
9,14	1	0.0052	0.0102	0	0	0
9,15	1	0.0052	0.0102	0	0	0
10,10	1	0.0052	0.0102	3	0.0152	0.0171
10,11	6	0.0311	0.0245	4	0.0203	0.0197
10,12	7	0.0363	0.0264	6	0.0305	0.0240
10,13	2	0.0104	0.0143	2	0.0102	0.0140
10,14	1	0.0052	0.0102	0	0	0
11,11	20	0.1036	0.0430	19	0.0964	0.0412
11,12	27	0.1399	0.0489	37	0.1878	0.0545
11,13	7	0.0363	0.0264	8	0.0406	0.0276
11,14	1	0.0052	0.0102	1	0.0051	0.0099
12,12	18	0.0933	0.0410	10	0.0508	0.0307
12,13	10	0.0518	0.0173	14	0.0711	0.0359
12,14	1	0.0052	0.0102	1	0.0051	0.0099
13,13	0	0	0	3	0.0152	0.0171
13,14	2	0.0104	0.0143	1	0.0051	0.0099

Genotype observed (D16S539)	Indian (n=193)			Pakistani (n=197)		
	Count	Frequency	SD	Count	Frequency	SD
8,9	4	0.0207	0.0201	3	0.0152	0.0171
8,10	6	0.0311	0.0245	5	0.0254	0.0220
8,11	11	0.057	0.0327	5	0.0254	0.0220
8,12	4	0.0207	0.0201	5	0.0254	0.0220
8,13	3	0.0155	0.0174	2	0.0102	0.0140
9,9	5	0.0259	0.0224	7	0.0355	0.0258
9,10	3	0.0155	0.0174	9	0.0457	0.0292
9,11	24	0.1244	0.0466	21	0.1066	0.0431
9,12	11	0.057	0.0327	13	0.066	0.0347
9,13	5	0.0259	0.0224	7	0.0355	0.0258
9,14	1	0.0052	0.0102	0	0	0
10,10	1	0.0052	0.0102	5	0.0254	0.0220
10,11	10	0.0518	0.0173	14	0.0711	0.0359
10,12	6	0.0311	0.0245	11	0.0558	0.0321
10,13	6	0.0311	0.0245	1	0.0051	0.0099
10,14	2	0.0104	0.0143	0	0	0
10,15	0	0	0	1	0.0051	0.0099
11,11	22	0.114	0.0448	20	0.1015	0.0422
11,12	26	0.1347	0.0482	23	0.1168	0.0449
11,13	16	0.0829	0.0389	14	0.0711	0.0359
11,14	5	0.0259	0.0224	1	0.0051	0.0099
12,12	8	0.0415	0.0281	17	0.0863	0.0392
12,13	9	0.0466	0.0297	7	0.0355	0.0258
12,14	1	0.0052	0.0102	2	0.0102	0.0140
12,15	1	0.0052	0.0102	0	0	0
13,13	1	0.0052	0.0102	4	0.0203	0.0197
13,14	2	0.0104	0.0143	0	0	0

Genotype observed (vWA)	Indian (n=194)			Pakistani (n=197)		
	Count	Frequency	SD	Count	Frequency	SD
9,20	1	0.0052	0.0101	0	0	0
13,14	0	0	0	1	0.0051	0.0099
13,17	1	0.0052	0.0101	0	0	0
13,18	1	0.0052	0.0101	2	0.0102	0.0140
14,14	6	0.0309	0.0244	2	0.0102	0.0140
14,15	2	0.0103	0.0142	4	0.0203	0.0197
14,16	14	0.0722	0.0364	12	0.0609	0.0334
14,17	9	0.0464	0.0296	12	0.0609	0.0334
14,18	12	0.0619	0.0339	3	0.0152	0.0171
14,19	6	0.0309	0.0244	2	0.0102	0.0140
14,20	0	0	0	2	0.0102	0.0140
15,15	1	0.0052	0.0101	5	0.0254	0.0220
15,16	14	0.0722	0.0364	4	0.0203	0.0197
15,17	7	0.0361	0.0263	7	0.0355	0.0258
15,18	7	0.0361	0.0263	9	0.0457	0.0292
15,19	6	0.0309	0.0244	2	0.0102	0.0140
15,20	0	0	0	1	0.0051	0.0099
16,16	14	0.0722	0.0364	13	0.066	0.0347
16,17	15	0.0773	0.0376	*26*	*0.132*	0.0473
16,18	13	0.067	0.0352	16	0.0812	0.0381
16,19	9	0.0464	0.0296	5	0.0254	0.0220
16,20	1	0.0052	0.0101	3	0.0152	0.0171
17,17	16	0.0309	0.0244	13	0.066	0.0347
17,18	*20*	*0.1031*	0.0428	22	0.1117	0.0440
17,19	11	0.0567	0.0325	19	0.0964	0.0412
18,18	7	0.0361	0.0263	3	0.0152	0.0171
18,19	7	0.0361	0.0263	7	0.0355	0.0258
18,20	0	0	0	1	0.0051	0.0099
18,21	1	0.0052	0.0101	0	0	0
19,19	1	0.0052	0.0101	1	0.0051	0.0099

Genotype observed (THO1)	Indian (n=193)			Pakistani (n=197)		
	Count	Frequency	SD	Count	Frequency	SD
6,6	15	0.0777	0.0378	14	0.0711	0.0359
6,7	22	0.114	0.0448	21	0.1066	0.0431
6,8	14	0.0725	0.0366	18	0.0914	0.0402
6,9	26	0.1347	0.0482	25	0.1269	0.0465
6,9.3	16	0.0829	0.0389	15	0.0761	0.0370
6,10	1	0.0052	0.0102	0	0	0
7,7	5	0.0259	0.0224	10	0.0508	0.0307
7,8	3	0.0155	0.0174	6	0.0305	0.0240
7,9	20	0.1036	0.0430	14	0.0711	0.0359
7,9.3	5	0.0259	0.0224	11	0.0558	0.0321
8,8	2	0.0104	0.0143	8	0.0406	0.0276
8,9	12	0.0622	0.0341	9	0.0457	0.0292
8,9.3	9	0.0466	0.0297	6	0.0305	0.0240
8,10.3	1	0.0052	0.0102	0	0	0
9,9	20	0.1036	0.0430	20	0.1015	0.0422
9,9.3	17	0.0881	0.0400	12	0.0609	0.0334
9,10	2	0.0104	0.0143	0	0	0
9.3,9.3	3	0.0155	0.0174	8	0.0406	0.0276

Genotype observed (TPOX)	Indian (n=194)			Pakistani (n=197)		
	Count	Frequency	SD	Count	Frequency	SD
8,8	26	0.1340	0.0479	30	0.1523	0.0502
8,9	11	0.0567	0.0325	23	0.1168	0.0449
8,10	7	0.0361	0.0263	6	0.0305	0.0240
8,11	50	0.2577	0.0616	63	0.3198	0.0651
8,12	8	0.0412	0.0280	8	0.0406	0.0276
8,13	1	0.0052	0.0101	0	0	0
9,9	4	0.0206	0.0199	1	0.0051	0.00995
9,10	7	0.0361	0.0263	3	0.0152	0.0171
9,11	18	0.0928	0.0408	29	0.1472	0.0495
9,12	3	0.0155	0.0174	3	0.0152	0.0171
10,10	1	0.0052	0.0101	1	0.0051	0.0099
10,11	15	0.0773	0.0376	13	0.066	0.0347
10,12	2	0.0103	0.0142	0	0	0
10,14	1	0.0052	0.0101	0	0	0
11,11	32	0.1649	0.0522	17	0.0863	0.0392
11,12	6	0.0309	0.0244	0	0	0
11,13	1	0.0052	0.0101	0	0	0
12,12	1	0.0052	0.0101	0	0	0

Genotype observed (CSF1PO)	Indian (n=191)			Pakistani (n=197)		
	Count	Frequency	SD	Count	Frequency	SD
8,10	0	0	0	1	0.0051	0.0099
8,12	0	0	0	1	0.0051	0.0099
9,10	1	0.0052	0.0102	2	0.0102	0.0140
9,11	1	0.0052	0.0102	3	0.0152	0.0171
9,12	6	0.0314	0.0247	1	0.0051	0.0099
9,13	0	0	0	1	0.0051	0.0099
10,10	7	0.0366	0.0266	21	0.1066	0.0431
10,11	24	0.1257	0.0470	20	0.1015	0.0422
10,12	16	0.0838	0.0393	21	0.1066	0.0431
10,13	6	0.0314	0.0247	8	0.0406	0.0276
10,14	2	0.0105	0.0145	1	0.0051	0.0099
11,11	20	0.1047	0.0434	18	0.0914	0.0402
11,12	41	0.2147	0.0582	43	0.2183	0.0577
11,13	8	0.0419	0.0284	8	0.0406	0.0276
11,14	2	0.0105	0.0145	1	0.0051	0.0099
12,12	38	0.199	0.0566	28	0.1421	0.0488
12,13	14	0.0733	0.0370	15	0.0761	0.0370
12,14	3	0.0157	0.0176	1	0.0051	0.0099
13,13	1	0.0052	0.0102	2	0.0102	0.0140
13,14	0	0	0	1	0.0051	0.0099
15,15	1	0.0052	0.0102	0	0	0

APPENDIX H:

FORENSIC AND PATERNITY SIMULATED CASES

1.1 Forensic case work

The main purpose of DNA profiling is to test whether a suspect is the source of an item of biological evidence which left at seen of crime or not. In the case of exclusion the DNA profiling from the evidence will be different and therefore must have originated from some one else in the population other than the suspect. If the DNA profiling from the suspect is similar (match) to the biological evidence this might be because they have originated from the same source. However, there is a possibility that someone else in the population have the same DNA profiling of the suspect. Therefore, it is important to calculate the rarity of suspect's DNA profiling. The product rule is used in this calculation and the result usually expressed 1 in number of people or likelihood ratio (LR) (Gill P. and Evett I., 1995).

1.1.1 VNTR system

In this simulated cases study five VNTR loci were used to calculate the rarity of an individual DNA profile in each population (UAE, Indian and Pakistani) using the databases developed in this study and by applying the product rule. In all cases the match were assumed between a crime stain and the suspect. The LR calculated using 4 methods, which are based on different assumptions. Method (1) the suspect was not related to the perpetrator, method (2) the suspect and the perpetrator belonged to an isolated population,

method (3) the suspect and the perpetrator belonged to the same sub-population and method (4) where the suspect was a close relative to the perpetrator.

It was found that there was evidence of substructure nature in the UAE population (F_{ST} = 0.0062), the upper bound was (F_{ST} = 0.0126). Therefore, in the second method the two F_{ST} values where used A (F_{ST} = 0.0062) and B (F_{ST} = 0.0126). In the third assumption when the suspect and the perpetrator are from isolated population the F_{ST} = 0.05 was used. In the fourth assumption the F_{ST} = 0.25 (brother-brother) was used in the calculation of LR considering the closest relationship between the suspect and the perpetrator. The full description can be seen in the example below.

An example for using the equation and how to calculate the probability of match and LR in forensic cases. These are the genotypes of the suspect from case 1 (UAE population).

The formula for homozygote is $$\frac{[2\theta+(1-\theta)p_i][3\theta+(1-\theta)p_i]}{(1+\theta)(1+2\theta)}$$

The formula for heterozygote is $$\frac{2[\theta+(1-\theta)p_i][\theta+(1-\theta)p_j]}{(1+\theta)(1+2\theta)}$$

Where $\theta = F_{ST}$ and p_i, p_j are the allele frequency.

Case 1 (UAE)	D2S44	D1S7	D7S21	D7S22	D12S11
Suspect	4,8	2,15	9,16	8,8	2,8

The calculation of the probability of match (MP) and likelihood ratio (LR) was calculated for D2S44 locus from the UAE Arab database. p_i (frequency of allele 4) = 0.0896, p_j (frequency of allele 8) = 0.0896.

Method (1) $2\,pq = 2\,(0.0896 \times 0.0896) = 0.0161$ (LR = 62)

Method (2) A $$\frac{2[\theta+(1-\theta)p_i][\theta+(1-\theta)p_j]}{(1+\theta)(1+2\theta)}$$

$$\frac{2[0.0062+(1-0.0062)x0.0896][0.0062+(1-0.0062)x0.0896]}{(1+0.0062)(1+2x0.0062)} = 0.0178 \text{ (LR} = 56)$$

Method (2) B $$\frac{2[\theta+(1-\theta)p_i][\theta+(1-\theta)p_j]}{(1+\theta)(1+2\theta)}$$

$$\frac{2[0.0126+(1-0.0126)x0.0896][0.0126+(1-0.0126)x0.0896]}{(1+0.0126)(1+2x0.0126)} = 0.0197 \text{ (LR} = 51)$$

Method (3) $$\frac{2[\theta+(1-\theta)p_i][\theta+(1-\theta)p_j]}{(1+\theta)(1+2\theta)}$$

$$\frac{2[0.05+(1-0.05)x0.0896][0.05+(1-0.05)x0.0896]}{(1+0.05)(1+2x0.05)} = 0.0316 \quad (\text{LR} = 32)$$

Method (4) $$\frac{2[\theta+(1-\theta)p_i][\theta+(1-\theta)p_j]}{(1+\theta)(1+2\theta)}$$

$$\frac{2[0.25+(1-0.25)x0.0896][0.25+(1-0.25)x0.0896]}{(1+0.25)(1+2x0.25)} = 0.1073 \quad (\mathrm{LR}=9)$$

Table 1.1 Two examples of forensic cases, in case 1 the suspects is from the UAE population and in case 2 the suspect from Indian population. The probability of match (MP) and likelihood ratio (LR) were calculated using 4 methods, which are based on different assumptions. Method (1) the suspect was not related to the perpetrator. Method (2) A: the suspect and the perpetrator belonged to a substructured population (F_{ST} = 0.0062) and method (2) B: where the upper bound was used (F_{ST} = 0.0126). Method (3) the suspect and the perpetrator belonged to an isolated population and Method (4) where the suspect was a close relative to the perpetrator (brother-brother) (F_{ST} = 0.25).

Case 1 (UAE)	D2S44	D7S22	D1S7	D7S21	D12S11	Combined
Fragment size (bp)	2,504	1,598	3,045	7,555	4,665	
	3,578	8,471	5,954	-	9,053	
Suspect	4/8	2/15	9/16	8/8	2/8	
MP (1)	0.0161	0.0155	0.0035	0.208	0.0199	3.6×10^{-9}
LR (1)	62.1	64.5	285.7	4.8	50.3	276,293,570
MP (2) A	0.0178	0.0172	0.0045	0.0139	0.0228	4.4×10^{-10}
LR (2) A	56.2	58.1	222.2	71.9	43.9	2,290,076,094
MP (2) B	0.0197	0.0191	0.0056	0.0173	0.0259	9.4×10^{-10}
LR (2) B	50.8	52.4	178.6	57.8	38.6	1,060,697,626
MP (3)	0.0316	0.0309	0.0139	0.0428	0.0434	2.5×10^{-8}
LR (3)	31.6	32.4	71.9	23.4	23	39,619,106
MP (4)	0.1073	0.1066	0.0845	0.2552	0.1297	3.2×10^{-5}
LR (4)	9.3	9.3	11.8	3.9	7.7	30,648

Case 2 (Indian)	D2S44	D7S22	D1S7	D7S21	D12S11	Combined
Fragment size (bp)	2,819	2,988	2,277	8,000	4,799	
	3,605	5,269	2,990	8,795	4,866	
Suspect	5/8	5/10	6/8	8/9	3/3	
MP (1)	0.0357	0.0069	0.0025	0.0493	0.0570	1.7×10^{-9}
LR (1)	28	144.9	400	20.3	17.5	576,528,120
MP (2) A	0.0379	0.0084	0.0034	0.0517	0.0594	3.3×10^{-9}
LR (2) A	26.4	119	294.1	19.3	16.8	299,579,784
MP (2) B	0.0409	0.0100	0.0044	0.0542	0.0619	6×10^{-9}
LR (2) B	24.4	100	227.3	18.5	16.2	166,217,216
MP (3)	0.0545	0.0207	0.0122	0.0688	0.0766	7.3×10^{-8}
LR (3)	18.3	48.3	82	14.5	13.1	13,767,382
MP (4)	0.1313	0.0954	0.0817	0.1451	0.1513	2.2×10^{-5}
LR (4)	7.6	10.5	12.2	6.9	6.6	44,336

1.1.2 STR system

The same two suspects in cases 1 and 2 were profiled using eight STR loci. In method (2) the $F_{ST} = 0.013$ in A and $F_{ST} = 0.023$ in B and the rest are same as in VNTR system. The results are shown in table 1.2.

Table 1.2 These two forensic cases examples are the same as previously profiled for VNTRs. Eight STR loci were added to five VNTR loci and the probability of match (MP) decreased dramatically and the likelihood ratio LR increased dramatically.

Case 1 (UAE)	D5S818	D7S820	D13S317	D16S539	vWA
Suspect	13/13	10/11	11/11	11/11	16/20
MP (1)	0.0344	0.1303	0.0781	0.1205	0.004
LR (1)	29.1	7.7	12.8	8.3	250
MP (2) A	0.0447	0.1352	0.0915	0.1354	0.0117
LR (2) A	22.4	7.4	10.9	7.4	85.6
MP (2) B	0.0532	0.1390	0.1020	0.1572	0.0175
LR (2) B	18.8	7.2	9.8	6.4	57.1
MP (3)	0.0780	0.1487	0.1317	0.1786	0.0329
LR (3)	12.8	6.7	7.6	5.6	30.4
MP (4)	0.3026	0.2089	0.3626	0.4091	0.1295
LR (4)	3.3	4.8	2.8	2.4	7.7

Case 1 (UAE)	THO1	TPOX	CSF1PO	Combined STRs	Combined VNTRs + STRs
Suspect	6/6	8/8	10/10		
MP (1)	0.0805	0.2023	0.0795	2.2×10^{-10}	7.9×10^{-19}
LR (1)	12.4	4.9	12.6	4,556,171,836	1.3×10^{18}
MP (2) A	0.0940	0.2184	0.0929	1.7×10^{-9}	7.3×10^{-19}
LR (2) A	10.6	4.6	10.8	602,697,199	1.4×10^{18}
MP (2) B	0.1046	0.2307	0.1035	5.2×10^{-9}	4.9×10^{-18}
LR (2) B	9.6	4.3	9.7	194,108,220	2.1×10^{17}
MP (3)	0.1343	0.2636	0.1330	$4.2\ 10^{-8}$	1.1×10^{-15}
LR (3)	7.4	3.8	7.5	23,401,115	9.3×10^{14}
MP (4)	0.3656	0.4848	0.3642	7.8×10^{-5}	2.5×10^{-9}
LR (4)	2.7	2.1	2.7	12,548	384,571,104

Case 2 (Indian)	D5S818	D7S820	D13S317	D16S539	vWA
Suspect	11/11	8/8	10/11	9/11	15/17
MP (1)	0.1060	0.0549	0.0398	0.1059	0.0378
LR (1)	9.4	18.2	25.1	9.4	26.5
MP (2) A	0.1204	0.0687	0.0464	0.1120	0.0430
LR (2) A	8.3	14.6	21.6	8.9	23.3
MP (2) B	0.1317	0.0766	0.0514	0.1165	0.0470
LR (2) B	7.6	13.1	19.5	8.6	21.3
MP (3)	0.1626	0.1041	0.0646	0.1284	0.0579
LR (3)	6.2	9.6	15.5	7.8	17.3
MP (4)	0.3938	0.3331	0.1493	0.1986	0.1360
LR (4)	2.5	3	6.7	5	7.4

Case 2 (Indian)	THO1	TPOX	CSF1PO	Combined STRs	Combined VNTRs + STRs
Suspect	6/7	11/11	12/13		
MP (1)	0.0878	0.1575	0.0641	8.2×10^{-10}	1.4×10^{-18}
LR (1)	11.4	6.3	15.6	1,198,441,087	6.9×10^{17}
MP (2) A	0.0934	0.1731	0.0724	2.2×10^{-9}	7.2×10^{-18}
LR (2) A	10.7	5.8	13.8	464,859,258	1.4×10^{17}
MP (2) B	0.0977	0.1852	0.0786	4×10^{-9}	2.4×10^{-17}
LR (2) B	10.2	5.4	12.7	248,768,396	4.1×10^{16}
MP (3)	0.1089	0.2177	0.0945	1.8×10^{-8}	1.3×10^{-15}
LR (3)	9.2	4.6	10.6	55,845,329	7.7×10^{14}
MP (4)	0.1803	0.4450	0.1830	7.8×10^{-6}	1.7×10^{-10}
LR (4)	5.5	2.2	5.5	123,733	5,485,820,566

1.2 Paternity case work

In paternity testing it is important to either exclude an alleged father or provide a high likelihood of paternity in the case of an inclusion (Strom C. *et al.*, 1996). In each paternity case a child must have one allele from the true biological father and the other

allele is from the mother. The American Association of blood Banks (AABB) recommended that the exclusion of the alleged father is based on the presence of exclusionary events at two or more loci (Chakraborty R. and Stivers D., 1996) because of mutation occurrence. In this case study the D1S7 was not used because of its high mutation rate (5.2%) (Jeffreys A. *et al.*, 1988) and therefore only four VNTRs were used. In the case of inclusion the probability that the alleged father is the true biological father should be determined by comparing the observed alleles to their distribution in the general population and expressed as either paternity index (PI) or the probability of paternity (w) (Strom C. *et al.*, 1996).

In this simulated study two systems were used, VNTRs and STRs. Four VNTR loci and eight STR loci were used to show how these systems can be applied and calculated in paternity cases. The alleles of the child were compared to the mother and the alleged father. The paternity index (PI) and the probability of paternity (w) was calculated for each locus and then combined for all VNTR and STR loci. The paternity index was calculated using allele frequency databases generated in this study. The combined paternity index was then calculated by using the product rule. The probability of paternity was calculated using the combined paternity index assuming a prior probability of 50% (Gjertson D. *et al.*, 1988).

The paternity index was calculated using 3 methods. In method (1) the alleged father (AF) was not related to the biological father, in method (2) the alleged father, the biological father and the mother (M) belonged to the same sub-population. In method (3) the alleged father was a relative to the biological father.

Example of a motherless case also investigated by ignoring the genotype of the mother. The same three methods were used in the calculation of paternity index in the motherless case. The probability of paternity was calculated for the trio family and for resembling case.

There was evidence of substructure nature in the UAE general population. The highest variation was observed between the Egyptian and Indian populations ($F_{ST} = 0.013$) for combined eight STR loci. The upper bound was $F_{ST} = 0.023$. However, only three populations were compared in five VNTR loci and the highest variation were observed between the UAE Arab and Indian populations ($F_{ST} = 0.0062$). The upper bound was $F_{ST} = 0.0126$. Therefore, in method (2) for VNTR loci $F_{ST} = 0.0126$ was used and for STR loci $F_{ST} = 0.023$ was used. In method (3) the $F_{ST} = 0.25$ was used considering the closest relationship between the alleged father and the biological father (brother-brother or father-child). All equations are seen in appendix H (section 2.1).

1.2.1 VNTR system

In this paternity case example only four VNTR loci were used. The D1S7 locus was not used because of its high mutation rate. Since the parents from the UAE Arab therefore the UAE database was used. The result is shown in table 1.3.

Table 1.3 An example of simulated paternity case from the UAE Arab population. The paternity index (PI) was calculated for each VNTR locus by the three methods. The standard method (1), which assumes that the alleged father was unrelated to the biological father. In method (2) it was assumes that the biological father and the mother belong to the substructure using F_{ST} value. In the method (3) it was assumed that the alleged father was relative to the true biological father. The three methods were used in trio case where the genotype of the mother was available and in motherless case by ignoring the genotype of the mother.

Case 1 (UAE)	D2S44	D7S22	D7S21	D12S11	CPI	W%
Mother	2,858 (5)	9,759 (15)	7,257 (7)	6,932 (6)		
	3,454 (8)	10,526 (15)	7,474 (8)	9,762 (9)		
Child	(8)	(15)	(4)	(9)		
	(10)	(15)	(7)	(6)		
Father	2,828 (5)	3,325 (6)	5,050 (4)	8,313 (8)		
	4,582 (10)	10,352 (15)	5,645 (5)	9,662 (9)		
PI Trio (1)	6.18	5.41	6.92	1.68	389	99.74%
(2)	5.61	4	6.18	1.56	216	99.54%
(3)	1.72	1.69	1.75	1.25	6	85.71%
Mother less (1)	3.09	5.41	3.46	0.97	56	98.25%
(2)	2.97	4.93	3.3	0.98	47	97.92%
(3)	1.51	0.92	1.55	0.98	2	67.8%

An example for the use of the paternity index calculations. The genotypes of the mother, child and the alleged father were used for the D2S44 locus from table 1.3.

The calculation of the paternity index (PI) in the trio case

$$\text{Method (1) PI} = \frac{1}{2p_j} = \frac{1}{2x0.0809} = 6.18$$

$$\text{Method (2) PI} = \frac{1+3\theta}{2[\theta+(1-\theta)p_j]} = \frac{1+3x0.016}{2[0.0126+(1-0.0126)x0.0809} = 5.41$$

$$\text{Method (3) PI} = \frac{1}{2p_j(1-2\theta_{AT})+2\theta_{AT}} = \frac{1}{2x0.0809(1-2x0.25)+2x0.25} = 1.72$$

The calculation of the paternity index (PI) in motherless case

$$\text{Method (1) PI} = \frac{1}{4p_i} = \frac{1}{4x0.0809} = 3.09$$

$$\text{Method (2) PI} = \frac{1+2\theta}{2[2(1-\theta)p_i+\theta]} = \frac{1+2x0.0126}{2[2(1-0.0126)x0.0809+0.0126} = 5.95$$

$$\text{Method (3) PI} = \frac{1}{2[2p_i(1-2\theta_{AT})+\theta_{AT}} = \frac{1}{2[2x0.0809(1-2x0.25)+0.25} = 1.52$$

1.2.2 STR system

The same example case used for VNTR also used for typing 8 STR loci. The combined paternity index (CPI) for all these loci was calculated. By adding these all loci together the CPI increased dramatically. The result is shown in table 1.4.

Table 1.4 An example of simulated paternity case from the UAE Arab population. The paternity index (PI) and probability of paternity (W%) were calculated for each STR locus by the three methods. The combined paternity index (CPI) and the probability of paternity for all 12 VNTR and STR loci were calculated.

Case 1 (UAE)	D5S818	D7S820	D13S317	D16S539	vWA	THO1
Mother	12	8	10	10	15	9
	13	11	11	11	16	9.3
Child	11	8	11	10	16	9
	12	12	12	11	17	9
Father	11	10	12	10	16	6
	11	12	12	14	17	9
PI Trio (1)	4.20	3.63	3.25	1.13	1.65	2.41
(2)	3.84	3.39	3.08	1.07	1.56	2.15
(3)	1.62	1.57	1.53	1.06	1.25	1.41
Mother less (1)	2.1	1.82	1.62	2.66	1.60	2.41
(2)	2.05	1.79	1.61	2.53	1.59	2.32
(3)	1.35	1.29	1.24	1.45	0.4	0.83

Case 1 (UAE)	TPOX	CSF1PO	CPI	W%	CPI VNTRs+STRs	W% VNTRs+STRs
Mother	8	12				
	10	12				
Child	8	12				
	11	13				
Father	8	9				
	11	13				
PI Trio (1)	1.83	8.40	3,423	99.97%	1,331,547	99.99992%
(2)	1.71	6.59	1621	99.94%	350,136	99.9997%
(3)	1.29	1.79	17	94.44%	102	99%
Mother less (1)	1.47	4.20	392	99.75%	21,952	99.995%
(2)	1.47	3.76	305	99.67%	14,335	99.993%
(3)	0.40	1.62	1	50%	2	67.8%

2.1 Equations for paternity

2.1.1 Paternity indexes calculations using three methods, (1), (2) and (3) for trio cases. (Lee H. *et al.*, 2000). C = child, M = mother and AF = alleged father.

C	M	AF	Method (1)	Method (2)	Method (3)
A_iA_i	A_iA_i	A_iA_i	$\frac{1}{p_i}$	$\frac{1+3\theta}{4\theta+(1-\theta)p_i}$	$\frac{1}{p_i(1-2\theta_{AT})+2\theta_{AT}}$
A_iA_i	A_iA_i	A_iA_j (i≠j)	$\frac{1}{2p_i}$	$\frac{1+3\theta}{2[3\theta+(1-\theta)p_i]}$	$\frac{1}{2p_i(1-2\theta_{AT})+2\theta_{AT}}$
A_iA_i	A_iA_j (i≠j)	A_iA_i	$\frac{1}{p_i}$	$\frac{1+3\theta}{3\theta+(1-\theta)p_i}$	$\frac{1}{p_i(1-2\theta_{AT})+2\theta_{AT}}$
A_iA_i	A_iA_j (i≠j)	A_iA_j (i≠j)	$\frac{1}{2p_i}$	$\frac{1+3\theta}{2[2\theta+(1-\theta)p_i]}$	$\frac{1}{2p_i(1-2\theta_{AT})+2\theta_{AT}}$
A_iA_j (i≠j)	A_iA_i	A_jA_j	$\frac{1}{p_j}$	$\frac{1+3\theta}{2\theta+(1-\theta)p_j}$	$\frac{1}{p_j(1-2\theta_{AT})+2\theta_{AT}}$
A_iA_j (i≠j)	A_iA_i	A_iA_j	$\frac{1}{2p_j}$	$\frac{1+3\theta}{2[2\theta+(1-\theta)p_j]}$	$\frac{1}{2p_j(1-2\theta_{AT})+2\theta_{AT}}$
A_iA_j (i≠j)	A_iA_i	A_jA_k (k≠i,j)	$\frac{1}{2p_j}$	$\frac{1+3\theta}{2[\theta+(1-\theta)p_j]}$	$\frac{1}{2p_j(1-2\theta_{AT})+2\theta_{AT}}$
A_iA_j (i≠j)	A_iA_j (i≠j)	A_iA_i	$\frac{1}{p_i+p_j}$	$\frac{1+3\theta}{4\theta+(1-\theta)(p_i+p_j)}$	$\frac{1}{(p_i+p_j)(1-2\theta_{AT})+2\theta_{AT}}$
A_iA_j (i≠j)	A_iA_j (i≠j)	A_iA_j	$\frac{1}{p_i+p_j}$	$\frac{1+3\theta}{4\theta+(1-\theta)(p_i+p_j)}$	$\frac{1}{(p_i+p_j)(1-2\theta_{AT})+2\theta_{AT}}$
A_iA_j (i≠j)	A_iA_j (i≠j)	A_jA_k (k≠i,j)	$\frac{1}{2(p_i+p_j)}$	$\frac{1+3\theta}{2[3\theta+(1-\theta)(p_i+p_j)]}$	$\frac{1}{2(p_i+p_j)(1-2\theta_{AT})+2\theta_{AT}}$
A_iA_j (i≠j)	A_jA_k (k≠i,j)	A_jA_j	$\frac{1}{p_j}$	$\frac{1+3\theta}{2\theta+(1-\theta)p_j}$	$\frac{1}{p_j(1-2\theta_{AT})+2\theta_{AT}}$
A_iA_j (i≠j)	A_jA_k (k≠i,j)	A_jA_l (l≠j)	$\frac{1}{2p_j}$	$\frac{1+3\theta}{2[\theta+(1-\theta)p_j]}$	$\frac{1}{2p_j(1-2\theta_{AT})+2\theta_{AT}}$

2.1.2 Paternity indexes calculations using three methods, (1), (2) and (3) for motherless cases (Lee H. *et al.*, 2000).

C	AF	Method (1)	Method (2)	Method (3)
A_iA_i	A_iA_i	$\frac{1}{p_i}$	$\frac{1+2\theta}{(1-\theta)p_i+3\theta}$	$\frac{1}{p_i(1-2\theta_{AT})+2\theta_{AT}}$
A_iA_i	A_iA_j (i≠j)	$\frac{1}{2p_i}$	$\frac{1+2\theta}{2[(1-\theta)p_i+\theta]}$	$\frac{1}{2[p_i(1-2\theta_{AT})+2\theta_{AT}]}$
A_iA_j (i≠j)	A_iA_i	$\frac{1}{2p_i}$	$\frac{1+2\theta}{2(1-\theta)p_i+2\theta}$	$\frac{1}{2[p_i(1-2\theta_{AT})+\theta_{AT}]}$
A_iA_j (i≠j)	A_iA_j (i≠j)	$\frac{1}{4p_i}+\frac{1}{4p_j}$	$\frac{(p_i+p_j)(1+2\theta)}{2[2(1-\theta)p_ip_j+(p_i+p_j)\theta]}$	$\frac{(p_ip_j)}{2[2p_ip_j(1-2\theta_{AT})+(p_ip_j)\theta_{AT}}$
A_iA_j (i≠j)	A_jA_k (k≠i,j)	$\frac{1}{4p_i}$	$\frac{1+2\theta}{2[2(1-\theta)p_i+\theta]}$	$\frac{1}{2[2p_i(1-2\theta_{AT})+\theta_{AT}]}$

REFERENCES

Abdullah, M. (1999). Imbalance of the UAE population structure and solution. Al-Khaleej press, Sharjah, UAE. (in Arabic)

Adams, D. (1988). Validation of the FBI procedure for DNA analysis. Crime Laboratory Digest 15, 106 -108

Ahmed, A., Linacre, A., Mohammed, A., Vanezis, P. and Goodwin, W. (2001). STR population data for 10 STR loci including the GenePrint™ PowerPlex™ 1.2 kit from El-Minia (Central Egypt). Forensic Sci. Int. 117, 233-234.

Alkhayat, A., Alshamali, F. and Budowle, B. (1996). Population data on the PCR-based loci LDLR, GYPA, HBGG, D7S8, Gc, HLA-DQA1, and D1S80 from Arabs from Dubai. Forensic Sci. Int. 81, 29-34.

Allen, R., Graves, G. and Budowle, B. (1989). Polymerase chain reaction amplification products separated on rehydratable polyacrylamide gels and stained with silver. BioTechniques 7, 736-744.

Alonso, A. and Martin, P. (1993). Amplified fragment length polymorphism analysis of the VNTR locus D1S80 in central Spain. Int. J. Leg. Med. 105, 311-314.

Al-Sayegh, F. (1998). UAE from tribe to country. Al-Khaleej centre for books, Dubai, UAE. (in Arabic).

Andres, M., Prieto, V., Flores, I., Sanz, P. (1996). Population genetics of three STRs: THO1, CSF1PO and TPOX in Southern Spain, in Carracedo A., Brinkmann B. and Bar W. (Eds), Advances in Forensic Haemogenetics 6, Springer, Berlin pp. 489-491.

Anker, R., Steinbrueck, T. and Donis-Keller H. (1992). Tetranucleotide repeat polymorphism at the human thyroid peroxidase (hTPO) locus. Hum. Mol. Genet. 1, 137.

Baetchel, F., Presley, K. and Smerick, J. (1995). D1S80 typing of DNA from simulated forensic specimens. J. Forensic Sci. 40, 524-533

Baird, M., Balazs, I., Giusti, A., Milyazaki, L., Nicholas, L., Wexler, K., Kanter, E., Glassberg, J., Allen, F., Rubinstein, P. and Sussman, L. (1986). Allele frequency distribution of two highly polymorphic DNA sequences in three ethnic groups and its application to the determination of paternity. Am. J. Hum. Genet. 39, 489-501

Balazs, I. (1993). Population genetics of 14 ethnic groups using phenotypic data from VNTR loci. In DNA fingerprinting: state of the science, in Pena, S., Chakraborty, R., Epplen J. and Jeffreys A. (Eds), Basel, Switzerland, Birkhauser Verlag, pp. 193-210.

Balazs, I., Baird, M., Clyne M. and Meade, E. (1989). Human population genetic studies of five hypervariable DNA loci. Am. J. Hum. Genet. 44, 182-190

Balding, D. and Nichols, R. (1994). DNA profile match probability calculation, how to allow for population stratification, relatedness, database selection and single bands. Forensic Sci. Int. 64, 124-140

Ballantyne, J., Sensabaugh G. and Witkowski, J. (1989). DNA Technology and Forensic Science. 32 Banbury report, USA.

Bar, W., Brinkmann, B., Budowle, B., Carracedo, A., Gill, P., Lincoln, P., Mayr, W. and Olaisen, B. (1997). DNA recommendations further report of the DNA commission of the ISFH regarding the use of short tandem repeat systems. Int. J. Leg. Med. 110, 175-176.

Bar, W., Kratzer, A., Machler, M. and Schmid W. (1989). Postmortem stability of DNA. Forensic Sci. Int. 39, 59-70.

Barber, M., Piercy, R., Andersen, J. and Parkin B. (1995). Structural variation of novel alleles at the Hum vWA and Hum FES/FPS short tandem repeat loci. Int. J. Leg. Med. 108, 31-35.

Bassam, B., Caetano-Anolles, G. and Gresshoff, P. (1991). Fast and sensitive silver staining of DNA in polyacrylamide gels. Anal. Biochem. 196, 80-83.

Baum, H., Fitz-Charles, H. and Mckee, R. (1990). A non-isotopic DNA detection system with the sensitivity of ^{32}P: applications for paternity and forensic identifications. In Polesky, H. and Mayr, W. (Eds), Advances in Forensic Haemogenetics 3, Springer-Verlag, Berlin, pp. 37-39.

Bayomi, R., Al-Gazali, L., Jaffer, U., Nur-E-Kamal, M., Dawodu, A., Bener, A., Eapen V. and Budowle B. (1997). United Arab Emirates population data on three DNA tetrameric short tandem repeat loci: HUMTHO1, TPOX and CSF1PO – derived using multiplex polymerase chain reaction and manual typing. Electrophoresis 18, 1637-1640.

Bechmann, J. and Weber, J. (1992). Survey of human and rat microsatellites. Genomics 12, 627-631

Bell, G., Selby, M. and Rutter, W. (1982). The highly polymorphic region near the human insulin gene is composed of simple tandemly repeating sequences. Nature 295, 31-35

Bener, A., Abdulrazzaq, Y., Al-Gazali, L., Micallef, R., Al-Khayat, A. and Gaber, T. (1996). Consanguinity and associated socio-demographic factors in the United Arab Emirates. Hum Hered. 46, 256-264.

Benzinger, E., Riech, A, Shirley, R. and Kucharik, K. (1995). Evaluation of the specificity of alkaline phosphatase-conjugated oligonucleotide probes for forensic DNA analysis. Appl. Theor. Electrophor. 4, 161-165.

Bjerre, A., Court, S., Lincoln, P., Morling, N. (1997). A report of the 1995 and 1996 paternity testing workshops of the English speaking working group of the international society for forensic haemogenetics. Forensic Sci. Int. 90, 41-55.

Blake, E., Mihalovich J., Higuchi, R., Walsh, S. and Erlich, H. (1992). Polymerase chain reaction (PCR) amplification and human leukocyte antigen (HLA)-DQα oligonucleotide typing on biological evidence samples: casework experience. J. Forensic Sci. 37, 700-726.

Boerwinkle, E., Xiong W., Fourest, E. and Chan, L. (1989). Rapid typing of tandemly hypervariable loci by the polymerase chain reaction: application of the apolipoprotein B 3` hypervariable region. Proc. Natl. Acad. Sci. USA 86, 212 – 216

Brandt-Casadevall, C., Taroni, F., Dimo-Simonin, N., Baala, L., Sefiani, A. and Mangin P. (2000). Moroccan population alleles frequency on 9 PCR-based loci, in Sensabaugh G., Lincoln, P. Olaisen, B. (Eds), Progress in Forensic Genetics 8, Excerpta Medica, Amsterdam pp. 221-223.

Brenner, C. and Morris, J. (1990). Paternity index calculations in single locus hypervariable DNA probes: validation and other studies. In proceedings for the International Symposium on Human Identification (1989) pp. 21-53. Promega Corporation, Madison, WI, USA.

Brinkmann, B. (1998). Overview of PCR-based systems in identity testing, in Lincoln, P. and Thomson, J. (Eds), Methods in Molecular Biology: Forensic DNA Profiling Protocols, 98,105-119. Humana Press, Totowa.

Brinkmann, B. and Wiegand, P. (1993). Medicolegal implications of PCR based VNTRs. Proceedings from the Fourth International Symposium on Human Identification (1993) pp. 149-160.

Brinkmann, B., Rand, S. and Bajanoski, T. (1992). Forensic identification of urine samples. Int. J. Legal Med. 105, 59-61.

Brinkmann, B., Rand, S. and Wiegand, P. (1991). Population and family data of RFLP's using selected single- and multi-locus systems. Int. J. Leg. Med. 104, 81-86.

Britten, R. and Davidson, E. (1971). Repetitive and non-repetitive sequences and a speculation on the origins of evolutionary novelty. Quarterly Review of Biology 46, 111-133.

Britten, R. and Kohne, D. (1968). Repeated sequences in DNA. Science 161, 529-540.

Brookfield, J. (1991). The effect of population subdivision on estimates of the likelihood ratio in criminal cases using single-locus DNA probes. Heredity 69, 97-100.

Brutlag D. (1980). Molecular arrangement and evolution of heterochromatic DNA. Ann. Rev. Genet. 14, 121-144.

Budowle, B. and Monson, K. (1994). Greater differences in forensic DNA profile frequencies estimated from racial groups than from ethnic subgroups. Clin. Chim. Acta 228, 3-18.

Budowle, B. Chakraborty, R., Giusti, A., Eisenberg, A. and Allens, R. (1991a). Analysis of the VNTR locus D1S80 by the PCR followed by high-resolution PAGE. Am. J. Hum. Genet. 48, 137-144.

Budowle, B., Giusti, A., Waye, J., Baechtel, F., Fourney, R., Adams, D., Presley, L., Deadman, H. and Monson, K. (1991b). Fixed-bin analysis for statistical evaluation of continuous distributions of allelic data from VNTR loci, for use in forensic comparisons. Am. J. Hum. Genet. 48, 841-855.

Budowle, B., Lindesy, J., DeCou J., Koons B., Giusti A. and Comey C. (1995). Validation and population studies of the loci LDLR, GYPA, HBGG, D7S8, and Gc (PM loci), and HLA-DQα using a multiplex amplification and typing procedure. J. Forensic Sci. 40, 45-54.

Budowle, B., Monson, K., Anoe, K., Baechtel, F., Bergman, D., Buel, E., Campbell, P., Clement, M., Corey, H., Davis, L., Dixon, A., Fish, P., Giusti, A., Grant, T., Gronert, T., Hoover, D., Jankowski, L., Kilgore, A., Kimoto, W., Landrum, W., Leone, H., Longwell, C., MacLaren, D., Medlin, L., Narveson, S., Pierson, M., Pollock, J., Raquel, R., Reznicek, J., Rogers, G., Smerick, J. and Thompson, R. (1991c). A preliminary report on binned general population data on six VNTR loci in Caucasians, Blacks and Hispanics from the United States. Crime Lab. Dig. 18, 9-25.

Budowle, B., Monson, K., Giusti, A. and Brown B. (1993). The assessment of frequency estimates of HaeIII-generated VNTR profiles in various reference databases. J. Forensic Sci.

Budowle, B., Monson, K., Giusti, A. and Brown, B. (1994). Evaluation of HinfI-generated VNTR profile frequencies determined using various ethnic databases. J. Forensic Sci. 39, 988-1008.

Budowle, B., Moretti, T., Baumstark, A., Defenbaugh, D. and Keys K. (1999). 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. 44, 1277-1286.

Budowle, B., Nhari, L., Moretti, T., Kanoyangwa, S., Masuka, E., Defenbaugh, D. and Smerick J. (1997). Zimbabwe black population data on the six short tandem repeat loci – CSF1PO, TPOX, THO1, D3S1358, vWA and FGA. Forensic Sci. Int. 90, 215-221.

Budowle, B., Smerick, J., Keys, K. and Moretti, T. (1997). United States population data on the multiplex short tandem repeat loci – HUMTO1, TPOX, and CSF1PO – and the variable number tandem repeat locus D1S80. J. Forensic. Sci. 42, 846-849.

Buffery, C., Burridge, F., Greenhalgh, M., Jones, S. and Willott, G. (1991). Allele frequency distributions of four variable number tandem repeat (VNTR) loci in the London area. Forensic Sci. Int. 52, 53-64.

Caglia, A., Dobosz, M., Boschi, I., d'Aloja, E. and Pascali, V. (1998). Increased forensic efficiency of a STR-based Y-specific haplotype by addition of the highly polymorphic DYS385. Int. J. Leg. Med. 111, 142-146.

Capon, D., Chen, E., Levenson, A., Seeburg, P. and Goeddel, D. (1983). Complete nucleotide sequences of the T24 human bladder oncogene and its normal homologue. Nature 302, 33-37.

Caspers, M., Posey, Y. and Brown, R. (1977). Separator isoelectric focusing, an improved method of protein analysis and purification. Anal. Biochem. 79, 166-180.

Cavalli-Sforza, L. and Bodmer, W. (1971). In: the genetics of human populations. WH Freeman and Company, San Fransisco, CA, USA.

Chakraborty, R. (1985). Paternity testing with genetic markers: are Y-linked genes more effiecient than autosomal ones. Am. J. Med Genet. 21, 298-305.

Chakraborty, R. and Jin, L. (1992). Heterozygote deficiency, population substructure and their implications in DNA fingerprinting. Hum. Genet. 88, 267-272.

Chakraborty, R. and Kidd, K. (1991). The utility of DNA typing in forensic work. Science 254, 1735-1379.

Chakraborty, R. and Stivers, D. (1996). Paternity exclusion by DNA markers, effects of paternal mutations. J. Forensic Sci. 41, 671-677.

Chakraborty, R., De Andrade, M., Daiger, S. and Brdowle, B. (1992a). Apparent heterozygote deficiencies observed in DNA typing data and their implications in forensic applications. Ann. Hum. Genet. 56, 45-57.

Chakraborty, R., Sirinivasan, M., Jin, L. and DeAndrade, M. (1992b). Effects of population subdivision and allele frequency differences on interpretation of DNA typing for human identification. Proceedings of the International Symposium on Human Identification (1992) pp. 205-222.

Chamberlain, J., Gibbs, R., Ranier, J., Nguyen, P. and Caskey, C. (1988). Deletion screening of the Duchenne muscular dystrophy locus via multiplex DNA amplification. Nucleic. Acids Res. 16, 11141-11156.

Charlesworth, B., Sniegowski, P. and Stephan, W. (1994). The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 371, 215-220.

Cohen J., (1992). The ceiling principle is not always conservative in assigning genotype frequencies for forensic DNA testing. Am. J. Hum. Genet. 51, 1165-1168.

Cohen, J. (1990). DNA fingerprinting for forensic identification: potential effects on data interpritation of subpopulation heterogeneity and band number variability. Am. J. Hum. Genet. 46, 358-368.

Cohen, J., Green, P. and Devlin, B. (1991). Forensic DNA tests and Hardy-Weinberg equilibrium. Science 243, 1037-1038.

Collins, A. and Morton, N. (1994). Likelihood ratios for DNA identification. Proc. Natl. Acad. Sci. USA 91, 6007-6011.

Comey, C., Budowle, B., Adams, D., Baumstark, A., Lindsey, J. and Presley, L. (1993). PCR amplification and typing of the HLA DQα gene in forensic samples. J. Forensic Sci. 38, 239-249.

Cooper, D., Smith, B., Cooke, H., Niemann, S. and Schmidtke, J. (1985). An estimate of unique DNA sequence heterozygosity in the human genome. Hum. Genet. 69, 201-205.

Cossutta, F., Perossa, R., Frassanito, F., Altamura, B. and Fattorini, P. (2000). Population genetics of fourteen STR loci in north-east Italy, in Olaisen B., Birnkmann, B. and Lincoln, P. (Eds), Progress in Forensic Genetics 8, Excerpta Medica, Amsterdam pp. 178-180.

Craig, J., Fowler, S., Burgoyne, A., Scott, C. and Hadrin, J. (1988). Repetitive deoxyribonucleic acid (DNA) and human genome variation, A concise review relevant to forensic biological. J. Forensic Sci. 33, 1111-1126.

Davidson, E., Hough, B., Klein, W. and Britten, R. (1975). Structural genes adjacent to interspersed repetitive DNA sequences. Cell 4, 217-238.

de Knijff, P., Kayser, M., Caglia, A., Corach, D., Fretwell, N., Gehrig, C., Heidorn, F., Herrmann, S., Herzog, B., Hidding, M., Honda, K., Jobling, M., Krawczak, M., Leim, K., Meuser, S., Meyer, E., Oesterreich, W., Pandya, A., Parson, W., Piccinini, A., Perez-Lezaun, A., Prinz, M., Schneider, P. M., Szibor, R., Teifel-Greding, J., Weichhold, G. and Roewer, L. (1997). Chromosome Y microsatellite: population genetic and evolutionary aspects. Int. J. Leg. Med. 110, 134-149.

Debenham, P. (1991). The value of multi-locus probes for parentage testing. Proceedings from the Second International Symposium on Human Identification (1991) pp. 155-161.

Deininger, P., Batzer, M., Hutchinson, C., Edgell, M., (1992). Master genes in mammalian repetitive DNA amplification. Trends Genet. 8, 307-312.

Devlin, B. (1991). Forensic DNA tests and Hardy-Weinberg equilibrium. Science 243, 1039-1041.

Devlin, B. and Risch, N. (1992). A note on Hardy-Weinberg equilibrium of VNTR data by using the Federal Bureau of Investigation's fixed-bin method. Am. J. Hum. Genet. 51, 549-553.

Devlin, B. and Risch, N. (1993a). Physical properties of VNTR data, and their impact on a test of allelic independence. Am. J. Hum. Genet. 53, 324-329.

Devlin, B., Risch, N. and Roeder, K. (1990). No excess of homozygosity at DNA fingerprint loci. Science 249, 1416-1420.

Devlin, B., Risch, N. and Roeder, K. (1993b). Statistical evaluation of DNA fingerprinting: a critique of the NRC's report. Science 259, 748-749, 837.

Devlin, B., Risch, N. and Roeder, K. (1994). Comments on the statistical aspects of the NRC's report on DNA typing. J. Forensic. Sci. 39, 28-40.

Diamond, J. (1987). Abducted orphans identified by grandpaternity testing. Nature 327, 552-553.

Doolittle, W. and Sapienza, C. (1980). Selfish genes, the phynotype the paradigm and genome evolution. Nature 284, 601-603.

Doran, G., Dickel, D., Ballinger, W., Agee, O., Laipis, P. and Hauswirth, W. (1986). Anatomical, cellular and molecular analysis of 8000-yr-old human brain tissue from Windover archaeological site. Nature 323: 803 – 806

Dover, G. (1982). Molecular drive: a cohesive mode of species evolution. Nature 299, 111-117.

Droog, S., Lakenberg, N., Meulenbelt, I., de Matt, M., Huisman, L., Jie, A. and Slagboom, P. (1996). Isolation and storage of DNA for population studies. Fibrinolysis 2, 29-30.

Drozd, M., Archard, L., Lincoln, P., Morling, N., Nellemann, L., Phillips, C., Soteriou, B. and Court, D. (1994). An investigation of the HUMvWA31A locus in British Caucasians. Forensic Sci. Int. 69, 161-170.

Duewer, D., Currie, L., Reeder, D., Leigh, S., Liu, H. and Mudd, J. (1995). Interlaboratory comparison of autoradiographic DNA profilling measurements. 2. Measurement uncertainty and its propagation. Anal. Chem. 67, 1220-1231.

Dykes, D. (1990). DNA analysis of paternity testing, in Roberston, J., Ross, A.M. and Burgoyne, L.A. (Eds), DNA in Forensic Science Theory, techniques and application, Ellis Horwood, England pp. 141-155.

Edwards, A. (1991). Strategy for identification and isolation of short tandem repeats. Proceedings from the Second International Symposium on Human Identification (1991) pp. 309-310.

Edwards, A., Civitello, A., Hammond, H. and Caskey, C. (1991a). DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Am. J. Hum. Genet. 49, 746-756.

Edwards, A., Hammond, H., Chakraborty, R. and Caskey, C. (1991b). DNA typing with trimeric and tetrameric tandem repeats: polymorphic loci, detection systems, and population genetics. Proceedings from the Second International Symposium on Human Identification (1991) pp.31-51.

Edwards, A., Hammond, H., Jin, L., Caskey, C. and Chakraborty, R. (1992). Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics 12, 241-253.

Entrala, C., Lorente, J., Lorente, M., Alvarez, J., Budowle, B. and Villanueva, E. (1999). Spanish population data on the loci D13S317, D7S820, and D16S539 generated using silver staining (SilverSTR III™ Multiplex). J. Forensic Sci. 44, 1032-1034.

Eriksen, B. and Svensmark O. (1994). DNA polymorphism in Greenland. Allele and profile frequencies in a Greenland population sample using the VNTR probes MS1, MS31, MS43a and YNH24. Int. J. Leg. Med. 106, 254-257.

Eriksen, B. and Svensmark, O. (1994a). Analysis of a Danish Caucasian population sample of single locus DNA-profiles: Allele frequencies, frequencies of DNA-profiles and heterozygosity. Forensic Sci. Int. 59, 119-129.

Eriksen, B. and Svensmark, O. (1994b). The effect of sample size on the estimation of the frequency of DNA-profiles in RFLP-analysis. Forensic Sci. Int. 65, 195-205.

Evett, I. and Weir, B. (1998). Interpreting DNA evidence: Statistical genetics for forensic scientists. Sinauer Associates, INC. Publisher, Sunderland, Massachusetts, USA pp. 64-66.

Evett, I., Gill, P., Lambert, J., Oldroyd, N., Frazier, R., Watson, S., Panchal, S., connolly, A. and Kimpton, C. (1997). Statistical analysis of data for three British ethnic groups from a new STR multiplex. Int. J. Leg. Med. 110, 5-9.

Evett, I., Scranage, J. and Pinchin, R. (1993). An illustration of the advantages of efficient statistical methods for RFLP analysis in forensic science. Am. J. Hum. Genet. 52, 498-505.

Fa-Ten Kao, (1985). Human Genome Structure. International Review of Cytology 96, 51-88.

Fisher, R. (1935). The logic of scientific inference. J. Roy. Stat. Soc. 98, 39-54.

Fisher, R. (1951). Standard calculations for evaluating a blood group system. Heredity 5, 95-102.

Foreman, L. and Lambert, J. (2000). Genetic differentiation within and between four UK ethnic groups. Forensic Sci. Int. 114, 7-20.

Foreman, L., Lambert, J. and Evett, I. (1998). Regional genetic variation in Caucasians. Forensic Sci. Int. 95, 27-37.

Fowler, C., Drinkwater, R., Skinner, J. and Burgoyne, L. (1988). Human satellite-III DNA: an example of a "macrosatellite" polymorphism. Hum. Genet. 79, 265-272.

Fregeau, C. and Fourney, R. (1993). DNA typing with fluorescently tagged short tandem repeats: a sensitive and accurate approach to human identification. Biotechniques 15, 100-119.

Gaensslen, R. (1983). Sourcebook in forensic serology, immunology, and biochemistry. National Institute of justice USA.

Gecz, J. (1991). PCR amplification of large VNTR alleles of D17S5 (YNZ22) locus. Nucleic Acids Res. 19, 5806.

Geisser, S. and Johnson, W. (1993). Testing independence of fragment lengths within VNTR loci. Am. J. Hum. Genet. 53, 1103-1106.

Gene, M., Huguet, E., Sanchez-Garcia, C., Moreno, P., Corbella, J. and Mezquita, J. (1995). Study of the 3′ -ApoB minisatellite performed by PCR in the population of Catalonia (northeast Spain). Hum. Hered. 45, 70-74.

GenePrint™ PowerPlix™ 1.2 System: technical manual (1998). Promega Co., Madison, WI, USA.

Giles, A., Booth, J., Parker, J., Garman, A., Garrick, D., Akhavan, H. and Schaap, A. (1990). Rapid, simple, non-isotopic probing of Southern blots for DNA finerprinting, in Polesky H. and Mayr W. (Eds), Advances in Forensic Haemogenetics 3, Springer-Verlag, Berlin pp.37-39.

Gill, P. and Evett, I. (1995). Population genetics of short tandem repeat (STR) loci. Genetica 96, 69-87.

Gill, P. and Werrett, D. (1987). Exclusion of a man charged with murder by DNA fingerprinting. Forensic Sci. Int. 35, 145-148.

Gill, P., Ivanov, P., Kimpton, C., Piercy, R., Benson, N., Tully, G., Evett, I., Hagelberg, E. and Sullivan, K. (1994a). Identification of the remains of the Romanov family by DNA analysis. Nat. Genet. 6, 130-135.

Gill, P., Jeffreys, A. and Werrett, D. (1985). Forensic applications of DNA 'fingerprints'. Nature 318, 577-579.

Gill, P., Kimpton, C., D'Aloja, E., Anderson, J., Bar, W., Brinkmann, B., Holgersson, S., Johnsson, V., Kloosterman, A., Lareu, M., Nellemann, L., Pfitzinger, H., Phillips, C., Schmitter, H., Schneider, P. and Stenersen, M. (1994b). Report of the European DNA profiling group (EDNAP) – towards standardisation of short tandem repeat (STR) loci. Forensic Sci. Int. 65, 51-59.

Gill, P., Kimpton, C., Urquhart, A., Oldroyd, N., Millican, E., Watson, S. and Downes, T. (1995). Automated short tandem repeat (STR) analysis in forensic casework – a strategy for the future. Electrophoresis 16, 1543-1552.

Gill, P., Lygo, J., Fowler, S. and Werrett, D. (1987). An evaluation of DNA fingerprinting for forensic purposes. Electrophoresis 8, 38-44.

Gill, P., Sullivan, K. and Werret, D. (1990). The analysis of hypervariable DNA profiles: problem associated with the objective determination of the probability of a match. Hum. Genet. 85, 75-79.

Gill, P., Urquhart, A., Millican, E., Oldroyd, N., Watson, S., Sparkes R. and Kimpton, C. (1996). Criminal intelligencedatabases and interpretation of STRs, in Carracedo, A., Brinkmann, B. and Bar, W. (Eds), Advances In Forensic Haemogenetics 6, pp. 235-242.

Giusti, A., Baird, M., et al. (1986). Application of deoxyribonucleic acid (DNA) polymorphisms to the analysis of DNA recovered from sperm. J. Forens. Sci. 31, 409-417.

Gjertson, D., Mickey, M., Hopfield, J., Takenouchi, T.and Terasaki, P. (1988). Calculation of probability of paternity, using DNA sequencies. Am. J. Hum. Genet. 43, 860-869.

Goodwin, W., Linacre, A. and Vanezis, P. (1999). The use of mitochondrial DNA and short tandem repeat typing in the identification of air crash victims. Electrophoresis 20, 1707-1711.

Guo, S. and Thompson, E. (1992). Performing the Exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 48, 361-372.

Gyllensten, U. and Erlich H. (1988). Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQα locus. Proc. Natl Acad. Sci., USA 85, 7652-7656.

Hagelberg, E., Sykes B. and Hedges, R. (1989). Ancient bone DNA amplified. Nature 342, 485.

Hallenberg, C. and Morling, N. (2001). A report of the 1997, 1998 and 1999 paternity testing workshops of the English speaking working group of the international society for forensic genetics. Forensic Sci. Int. 116, 23-33.

Hammond, H., Jin, L., Zhong, Y., Caskey, C. and Chakraborty, R. (1994). Evaluation of 13 short tandem repeat loci for use in personal identification applications. Am. J. Hum. Gnet. 55, 175-189.

Hampe, A., shamoon, B., Gobet, M., Sherr, C. and Galibert, F. (1989). Nucleotide sequence and structural organisation of the human FMS proto-oncogene. Oncogene Res. 4, 9-17.

Hansen, H. and Morling, N. (1993a). Genetic investigations in immigration cases and frequencies of DNA fragments of the VNTR systems D2S44, D5S43, D7S22, and D12S11 in Turks. Forensic Sci. Int. 60: 23-35.

Hansen, H. and Morling, N. (1993b). Paternity testing with VNTR DNA systems. II. Evaluation of 271 cases of disputed paternity with the VNTR systems D2S44, D5S43, D7S21, D7S22 and D12S11. Int. J. Leg. Med. 105, 197-202.

Hartl, D. and Clark, A. (1997). Principles of population genetics. Third edition. Sinauer Associates, Inc. publisher, Sunderland, Massachusetts, USA.

Hawley, D. (1970). The Trucial States.

Heard-Bey, F. (1982). From Trucial States to United Arab Emirates.

Henke, L., Cleef, S.; Zakrzewska, M. and Henke, J. (1991). Population genetic data determined for five different single locus minisatellite probes, in Burke, T., Dolf G., Jeffreys, A. and Wolff, R. (Eds), DNA fingerprints: Approaches and Applications pp. 144-153.

Herrin, G. (1993). Probability of matching RFLP patterns from unrelated individuals. Am. J. Hum. Genet. 52, 491-497.

Heyer, E., Puymirat, J., Dieltjes, P., Bakker, E. and Knijff, P. (1997). Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. Hum. Mol. Genet. 6, 799-803.

Higuchi, R., von Beroldingen, C., Sensabaugh, G. and Erlich, H. (1988). DNA typing from single hairs. Nature 332, 543-546.

Hochmeister, M., Budowle, B., Borer, U., Eggmann, U., Comey, C. and Dirnhofer, R. (1991). Typing of deoxyribonucleic acid (DNA) extracted from compact bone from human remains. J. Forensic Sci. 36, 1649-1661.

Hochmeister, M., Budowle, B., Borer, U., Rudin, O., Bohnert, M. and Dirnhofer, R. (1995). Confirmation of the identity of human skeleton remains using multiplex PCR amplification and typing kits. J. Forensic. Sci. 40, 701-705.

Hochmeister, M., Budowle, B., Eisenberg, A., Borer, U. and Dirnhofer, R. (1996). Using multiplex PCR amplification and typing kits for the analysis of DNA evidence in a serial killer case. J. Forensic Sci. 41, 155-162.

Hochmeister, M., Budowle, B., Jung, J., Borer, U., Comey, C. and Dirnhofer, R. (1991). PCR-based typing of DNA extracted from cigarette butts. Int. J. Leg. Med. 104, 229-233.

Holmlund, G., Wahlgren, B. and Lindblom, B. (1994). A D12S11 (MS43A) allele with an internal HinfI restriction site. Forensic Sci. Int. 66, 105-109.

Holt, C., Stauffer, C., Willian, J., Lazaruk, K., Nguyen, T., Budowle, B. and Walsh, P. (2000). Practical applications of genotypic surveys for forensic STR testing. Forensic Sci. Int. 112, 91-109.

Horn, G., Bugawan T., Long, C. and Erlich, H. (1988). Allelic sequence variation of the HLA-DQ loci: relationship to serology and insulin-dependent diabetes susceptibility. Proc. Natl. Acad. Sci. USA 85, 6012-6016.

Horn, G., Richards, B. and Klinger, K. (1989). Amplification of a highly polymorphic VNTR segment by the polymerase chain reaction. Nucleic Acids Res. 17, 2140.

Hwu, H., Roberts, J., Davidson, E. and Britten, R. (1986). Insertion and/or deletion of many repeated DNA sequences in human and higher ape evolution. Proc. Natl. Acad. Sci. USA 83, 3875-3879.

ISFH report (1992). 1991 report concerning recommendations of the DNA commission of the international society for forensic haemogenetics relating to the use of DNA polymorphisms. Forensic Sci. Int. 52, 125-130.

Jabs, E., Wolf, S. and Migeon, B. (1984). Characterization of a cloned DNA sequence that is present at centromeres of all human autosomes and the X chromosome and shows polymorphic variation. Proc. Natl. Acad. Sci. USA 81: 4884-4888.

Jarreta, B., Roche, P. and Abecia, E. (1999). Genetic variation at six STR loci (HUMTHO1, HUMTPOX, HUMCSF1PO, HUMF13AO1, HUMFES/FPS, HUMvWFA31) in Argon (Norht Spain). Forensic Sci. Int. 100, 87-92.

Jeffreys, A. (1987a). Genetic typing and confirmation of parenthood. Lancet 3, 806.

Jeffreys, A. (1987b). Highly variable minisatellites and DNA fingerprints. Biochem. Soc. Trans. 15, 309-317.

Jeffreys, A. (1993). DNA typing: approaches and applications. J. Forensic Sci. Soc. 33, 204-211.

Jeffreys, A., Allen, M., Hagelberg, E. and Sonnberg, A. (1992). Identification of the skeletal remains of Josff Mengele by DNA analysis. Forensic Sci. Int. 56, 65-76.

Jeffreys, A., Barber, R., Bios, P., Buard, J., Dubrova, Y., Grant, G., Hollies, C., May, C., Neumann, R., Panayi, M., Ritchie, A., Shone, A., Signer, E., Stead, J. and Tamaki K. (1999). Human minisatellites, repeat DNA instability and meiotic recombination. Electrophoresis 20, 1665-1675.

Jeffreys, A., Brookfield, J. and Semeonoff, R. (1985c). Positive identification of an immigration test-case using human DNA fingerprints. Nature 317, 818-819.

Jeffreys, A., MacLeod, A., Tamaki, K., Neil, D. and Monckton, D. (1991). Minisatellite repeat coding as a digital approach to DNA typing. Nature 354, 204-209.

Jeffreys, A., Royle, N., Wilson, V. and Wong, Z. (1988). Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA. Nature 332, 278-281.

Jeffreys, A., Turner, M. and Debenham, P. (1991). The efficiency of multilocus DNA fingerprint probes for individualisation and establishment of family relationships, determined from extensive casework. Am. J. Hum. Genet. 48, 824-840.

Jeffreys, A., Wilson, V. and Thein, S. (1985a). Hypervariable "minisatellite" regions in human DNA. Nature 314, 67-73.

Jeffreys, A., Wilson, V. and Thein, S. (1985b). Individual-specific "fingerprints" of human DNA. Nature 316, 76-79.

Jeffreys, A., Wilson, V., Thein, S., Weatherall, D. and Ponder, B. (1986). DNA "fingerprints" and segregation analysis of multiple markers in human pedigrees. Am. J. Hum. Genet. 39, 11-24.

Jelinek, W. and Schmid, C. (1982). Repetitive sequences in eukaryotic DNA and their expression. Ann. Rev. Biochem. 51, 813-844.

Jobling, M. and Tyler-Smith, C. (1995). Fathers and sons: the Y chromosome and human evolution. Trends Genet. 11, 449-456.

Jones, D. (1972). Blood samples, probabilities and discriminations. J. Forensic Sci. Soc. 12, 355-359.

Kasai, K., Nakamura, Y. and White, R. (1990). Amplification of a variable number of tandem repeats (VNTR) locus (pMCT118) by the polymerase chain reaction (PCR) and its application to forensic science. J. Forensic Sci. 35, 1196-1200.

Kaye, D. (1993). DNA evidence: probability, population genetics, and the courts. Harvard J. Law Technol. 7, 101-172.

Kayser, M., Caglia, A., Fretwell, N., Gehring, C., Graziosi G., Heidorn, F., Herrmann, S., Herzog, B., Hidding, M., Honda, K., Jobling, E., Krawczak, M., Leim, K., Meuser, S., Meyer, E., Oesterreich, W., Pandya, A., Parson, W., Penacino, G., Perez-Lezaun, Piccinini, A., Prinz, M., Schmitt, C., Schneider, P., Szibor, R., Teifel-Greding, J., Weichhold, G., de Knijff, P. and Roewer, L. (1997a). Evaluation of Y-chromosomal STRs: a multicenter study. Int. J. Legal Med. 110, 125-133.

Kayser, M., Knijff, P., Dieltjes, P., Krawczak, M., Nagy, M., Zerjal, T., Pandya, A., Tyler-Smith, C. and Roewer L. (1997b). Applications of microsatellite-based Y chromosome haplotyping. Electrophoresis 18, 1602-1607.

Kelly, T. and Smith, H. (1970). A restriction enzyme from Hemophilus influenzae II. Base sequence of the recognition site. J. Mol. Biol. 51, 393-409.

Khalil, A. and Al-Kubaisi, S. (1999). Imbalance population in the UAE within 30 years. Al-Khaleej press, Sharjah, UAE. (in Arabic).

Kimpton, C., Fischer, D., Watson, S., Adams, M., Urquhart, A., Lygo, J. and Gill, P. (1994a). Evaluation of an automated DNA profiling system employing multiplex amplification of four tetrameric STR loci. Int. J. Leg. Med. 107, 1-10.

Kimpton, C., Gill, P., D'Aloja, E., Anderson, J., Bar, W., Holgersson, S., Jacobsen, S., Johnsson, V., Kloosterman, A., Lareu, M., Nellemann, L., Pfitzinger, H., Phillips, C.,

Rand, S., Schmitter, H., Scheinder, P., Sternersen, M. and Vide, M. (1995). Report on the second EDNAP collaborative STR exercise. Forensic Sci. Int. 71, 137-152.

Kimpton, C., Gill, P., Urquhart, A., Walton, A., Adams, M., Watson, S. and Millican, E. (1994b). Automated DNA profiling employing multiplex amplification of short tandem repeat loci, in Bar, W., Fiori, A. and Rossi, U. (Eds), Advances in Forensic Haemogenetics 5, Springer-Verlag, Berlin pp. 309-311.

Kimpton, C., Gill, P., Walton, A., Urquhart, A., Millican, E. and Adams, M. (1993). Automated DNA profiling employing multiplex amplification of short tandem repeat loci. PCR Methods Applic. 3, 13-22.

Kimpton, C., Oldroyd, N., Watson, S., Frazier, R., Johnson, P., Millican, E., Urquhart, A., Sparkes, R. and Gill, P. (1996). Validation of highly discriminating short tandem repeat amplification systems for individual identification. Electrophoresis 17, 1283-1293.

Kimpton, C., Walton, A. and Gill, P. (1992). A further tetranucleotide repeat polymorphism in the vWA gene. Hum. Mol. Genet. 1, 287.

Klintschar, M., Al Hammadi, N., Lux, T. and Reichenpfader, B. (1998a). Genetic variation at the short tandem repeat loci HumvWA, HumFXIIIB, and HumFES/FPS in the Egyptian and Yemenian. J. Forensic Sci. 43, 850-353.

Klintschar, M., Kozma, Z., Al Hammadi, N., Abdull Fatah, M. and Nohammer, C. (1998b). A study on the short tandem repeat systems HumCD4, HumTHO1 and HumFIBRA in population samples from Yemen and Egypt. Int. J. Leg. Med 111, 107-109.

Kloosterman, A., Budowle, B. and Daselaar, P. (1993). PCR-amplification and detection of the human D1S80 VNTR locus: Amplification conditions, population genetics, and application in forensic analysis. Int. J. Leg. Med. 105, 257-264.

Knott, T., Wallis, S., Pease, R., Powell, L. and Scott, J. (1986). A hypervariable region 3` to the human apolipoprotein B gene. Nucleic Acids Res. 14, 9215-9216.

Kratzer, A. and Bar, W., (1996). Efficiency of 6 STR systems, HLA DQα and the polymarker systems (PM) in paternity testing, in Carracedo, A., Brinkmann, B. and Bar, W. (Eds), Advances in Forensic Haemogenetics 6, Springer, Berlin pp. 560-562.

Krawczak, M. and Schmidtke, J. (1994). DNA fingerprinting. BIOS scientific publishers, Oxford, UK.

Krontiris, T. (1995). Minisatellites and human disease. Science 269, 1682-1683.

Lander, E. (1989). DNA fingerprinting on trial. Nature 339, 501-505.

Lander, E. (1991a). Lander reply. Am. J. Hum. Genet. 49, 899-903.

Lander, E. (1991b). Research on DNA typing catching up with courtroom application. Am. J. Hum. Genet. 48, 819-823.

Lander, E. and Budowle, B. (1994). DNA fingerprinting dispute laid to rest. Nature 371, 735-738.

Landsteiner, K. (1900). Zur kenntnis der antifermentativen, lyteschen und agglutinierenden wirkungen des blustrums und der lymphe. Zentralbl Bakteriol 27, 357-362.

Landsteiner, K. and Richter, M. (1902). Uber die verwertbarket individueller blutdifferezen fur die forensische praxis. Medizinisch Beamte 16: 85 – 89

Lareu, M., Phillips, C., Carracedo, A., Lincoln, P., Syndercombe, Court, D. and Thompson, J. (1994). Investigation of the STR locus HUMTHO1 using PCR and two electrophoresis formats: UK and Galician Caucasian population surveys and usefulness in paternity investigations. Forensic Sci. Int. 66, 41-52.

Lee, H., Lee, J., Han, G. and Hwang, J. (2000). Motherless case in paternity testing. Forensic Sci. Int. 114, 57-65.

Lempert, R. (1993). DNA, science, and the law: two cheers for the ceiling principle. Jurimetrics J. 34, 41-57.

Lewis, P. and Zaykin, D. (2001). Genetic Data Analysis (GDA): Computer program for the analysis of allelic data. Version 1.0 (d16b). Free program distributed by the authors over the internet from the GDA Home Page at http://lewis.eeb.uconn.edu/lewishome/software.html

Lewontin, R. (1972). The apportionment of human diversity. Evol. Biol. 6, 381-398.

Lewontin, R. and Hartl, D. (1991). Population genetics in forensic DNA typing. Science 254, 1745-1750.

Li, H., Gyllensten, U., Cui, X., Saiki, R., Erlich, H. and Arnheim, N. (1988). Amplification and analysis of DNA sequences in single human sperm and diploid cells. Nature 335, 414-417.

Lins, A., Micka, K., Sprecher, C., Taylor, J., Bacher, J., Rabbach, D., Bever, R., Creacy, S. and Schumm, J. (1998). Development and population study of an eight locus shrot tandem repeat (STR) multiplex system. J. Forensic Sci. 43, 1168-1180.

Litt, M and Luty, J. (1989). A hypervariable microsatellite revealed by *in vitro* amplification amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am. J. Hum. Genet. 44, 397-401.

Lorente, J., Lorente, M., Budowle, B., Wilson, M. and Villanueva, E. (1994). Analysis of short tandem repeat (STR) HUMVWA in the Spanish population. Forensic Sci. Int. 65, 169-175.

Lorente, M., Lorente, J., Wilson, M., Budowle, B. and Villanueva, E. (1997). Spanish population data on seven loci: D1S80, D17S5, HUMTHO1, HUMvWA, ACTBP2, D21S11 and HLA-DQA1. Forensic Sci. Int. 86, 163-171.

Ludwig, E., Friedl, W. and McCarthy, B. (1989). High resolution analysis of a hypervariable region in the human apolipoprotein B gene. Am. J. Hum. Genet. 45, 458-464.

Lygo, J., Johnson, P., Holdaway, D., Woodroffe, S., Whitaker, J., Clayton, T., Kimpton, C. and Gill, P. (1994). The validation of short tandem repeat (STR) loci for use in forensic casework. Int. J. Leg. Med. 107, 77-89.

Madisen, L., Hoar, D., Holroyd, C., Crispe, M. and Hodes, M. (1987). DNA banking, the effects of storage of blood and isolated DNA on the integrity of DNA. Am. J. Med. Genet. 27, 379-390.

Mancuso, D., Tuley, E., Westifield, L., Worall, N., Shelton-Inloes, B., Sorace, J., Alevy, Y. and Sadler, J. (1989). Structure of the gene for human von Willebrand factor. J. Biol. Chem. 264, 19514-19527.

Manuelidis, L. (1978). Complex and simple sequences in human repeated DNAs. Chromosoma 66, 1-21.

Martin, P., Garcia, O., Albarran, C., Sancho, M. and Alonso, A. (1998). A Spanish population study of the STR loci HumLPL, D5S818, D7S820 and D13S317. Int. J. Leg. Med. 112, 70-71.

McNally, L., Shaler, R., Baird, M., Balazs, I., DeForest, P. and Kobilinsky, L. (1989a). Evaluation of deoxyribonucleic acid (DNA) isolated from human blood stains exposed to ultraviolet light, heat, humidity, and soil contamination. J. Forensic Sci. 34, 1059-1069.

McNally, L., Shaler, R., Baird, M., Balazs, I., Kobilinsky, L. and DeForest, P. (1989b). The effects of environment and substrata on deoxyribonucleic acid (DNA): the use of casework samples from New York. J. Forensic Sci. 34, 1074-1077.

Micka, K., Sprecher, C., Lins, C., Theisen, A., Comey, C., Koons, B., Crouse, C., Endean, D., Pirelli, K., Lee, S., Duda, N., Ma, M. and Schumm, J. (1996). Validation of multiplex polymorphic STR amplification sets developed for personal identification applications. J. Forensic Sci. 41, 582-590.

Milkos, G. and John, B. (1979). Heterochromatin and satellite DNA in man: properties and prospects. Am. J. Hum Genet. 31, 264-280.

Miller, M. (1997). RxC: A program for the analysis of contingency tables. Free computer software program distributed by authors over the internet from the R x C Home Page at http//herb.bio.nau.edu/~miller/rxc.htm.

Miller, M. (1997). Tools for population genetic analyses (TFPGA) version 1.3: A windows program for the analysis of allozyme and molecular population genetic data. Computer software distributed by author.

Ministry of Information and Culture (1992). The United Arab Emirates, Abu Dhabi, UAE.

Mohammed, A., Linacre, A., Vanezis, P. and Goodwin, W. (2000). STR data for the GenePrint PowerPlex 1.2 system loci from three United Arab Emirates populations. Forensic Sci. Int. 3013, 1-2. (in press)

Moller, A., Meyer, E. and Brinkmann, B. (1994). Different types of structural variation in STRs: HumFES/FPS, HumVWA and HumD21S11. Int. J. Leg. Med. 106, 319-323.

Monaghan, B. and Newhall, P. (1996). STR results for both epithelial and spermatozoal fractions obtained from 13-year old microscope slides of vaginal swab material. Can. Soc. Forensic Sci. J. 29, 13-19.

Monson, K. and Budowle, B. (1993). A comparison of the fixed bin method with the floating bin and direct count methods: effect of VNTR profile frequency estimation and refference population. J. Forensic Sci. 38, 1037-1050.

Moore, G. (1983). Slipped-mispairing and the evolution of introns. Trends in Biochemical Sciences 8, 411-414.

Morling, N. and Hansen, H. (1993). Paternity testing with VNTR DNA ssytems. I. Matching criteria and population frequencies of the VNTR systems D2S44, D5S43, D7S21, D7S22 and D12S11 in Danes. Int. J. Leg. Med. 105, 189-196.

Morton, N. (1992). Genetic structure of forensic applications. Proc. Natl. Acad. Sci. USA 89, 2556-2560.

Morton, N., Collins, A. and Balazs, I. (1993). Kinship bioassay on hypervariable loci in blacks and Caucasians. Proc. Natl. Acad. Sci. USA 90, 1892-1899.

Moura-Neto, R. and Budowle, B. (1997). Fixed bin population data for the VNTR loci D1S7, D2S44, D4S139, D5S110, D10S28 and D14S13 in population sample from Rio De Janeiro, Brazil. J. Forensic Sci. 42, 926-928.

Moyzis R., Albright, K., Bartholdi, M., Cram, L., Deaven, L., Hildebrand, C., Joste, N., Longmire, J., Meyne, J. and Schwarzacher-Robinson, T. (1987). Human chromosome-specific repetitive DNA sequences: novel markers for genetic analysis. Chromosoma 95, 375-386.

Mudd, J. and Baechtel, F. (1994). Interlaboratory comparison of autoradiographic DNA profilling measurements. 1. Data and and summary statistics. Anal. Chem. 66, 3303-3317.

Mullis, K. and Faloona, F. (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods in Enzymology. 155, 335-350.

Nakamura, S., Nagai, T., Matsui, H., Tabuchi, M., Sugie, H., Furukawa, M., Imai, H. and Kurihara, K. (2000). Genetic analysis of AmpFlSTR Profiler Plus loci in Japanese, in Olaisen B., Brinkmann B. and Lincoln P. (Eds), Progress in Forensic Genetics 8, Excerpta Medica, Amesterdam pp. 236-238.

Nakamura, Y. Gillian, S., O'Connell, P., Leppert, M., Lathrop, G., Lalouel, J. and White, R. (1987b). Isolation and mapping of a polymorphic DNA sequence pYNH24 on chromosome 2 (D2S44). Nucleic Acids Res. 15, 10073.

Nakamura, Y., Carlson, M., Krapcho, K. and White, R. (1988). Isolation and mapping of a polymorphic DNA sequence (pMCT118) on chromosome 1p (D1S80). Nucleic Acids Res. 16, 9364.

Nakamura, Y., Leppert M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E. and White, R. (1987a). Variable number of tandem repeat (VNTR) markers for human gene mapping. Science 235, 1616-1622.

Nata, M., Kimura, T., Hashiyada, M., He, P., Yan, W., Li, X., Funayama, M. and Sagisaka, K. (1999). Allele frequencies of eight STRs in Japanese and Chinese. Int. J. Leg. Med. 112, 396-399.

National Research Council (NRC), (1992). DNA technology in forensic science. National Academy Press, Washington, D.C., USA.

National Research Council (NRC), (1996). The evaluation of forensic DNA evidence. National Academy Press, Washington, D.C., USA.

Nei, M. and Roychoudry, A. (1974). Sampling variances of heterozygosity and genetic distance. Genetics 76, 379-390.

Neil, D. and Jeffreys, A. (1993). Digital DNA typing at a second hypervariable locus by minisatellite variant repeat mapping. Hum. Mol. Genet. 2, 1129-1135.

Neuhuber, F. and Radacher, M. (1997). A genetic study of the short tandem repeat systems vWA and THO1 in an Austrian population. Forensic Sci. Int. 87, 211-217.

Nichols, R. and Balding, D. (1991). Effects of population structure on DNA fingerprint analysis in forensic science. Heredity 66, 297-302.

Odelberg, S., Plaetke, R., Eldridge, J., Ballard, L., O'Connell, P., Nakamura, Y., Leppert, M., Lalouel, J. and White, R. (1989). Characterisation of eight VNTR loci by agarose gel electrophoresis. Genomics 5, 915-924.

Okazaki, K., Wang, W., Nakamura, M. and Tamaki, Y. (1996). STR typing of buccal swabs for paternity testing with reference to Japanese population data on the D20S85, D14S118 and D14S543 loci. Japanese J. Leg. Med. 50, 404-411.

Oldroyd, N., Urquart, A., Kimpton, C., Millican, E., Watson, S., Downes, T. and Gill, P. (1995). A highly discriminating octoplex short tandem repeat polymerase chain reaction system suitable for human identification. Electrophoresis 16, 334-337.

Orgel, L. and Crick, F. (1980). Selfish DNA: the ultimate parasite. Nature 284, 604-607.

Paabo, S. (1985). Molecular cloning of ancient Egyptian mummy DNA. Nature 314, 644-645.

Papiha, S., Galdern, R., Sertedaki, A., Pena, J., Zhong, Y. and chakraborty, R. (1998). Study of three hypervariable DNA loci (D1S7, D2S22 and D12S11) in three European populations. Ann. Hum. Biol. 25: 29-41

Pascali, V., d'Aloja, E., Dobosz, M. and Pescarmona, M. (1991). Estimating allele frequencies of hypervariable DNA systems. Forensic Sci. Int. 51, 273-280.

Peake, I., Bowen, D., Bignell, P., Liddell, M., Sadler, J., Standen, G. and Bloom, A. (1990). Family studies and prenatal diagnosis in severe von Willebrand disease by polymerase chain reaction amplification of a variable number tandem repeat region of the von Willebrand factor gene. Blood 76, 555-561.

Pech, M. (1987). The United Arab Emirates A Venture in Unity.

Pena, S. and Chakraborty, R. (1994). Paternity testing in the DNA era. Trends Genet. 10, 190-194.

Pena, S., Prado, V. and Epplen, J. (1995). DNA diagnosis of human genetic individuality. J. Mol. Med. 73, 555-564.

Pinheiro, M., Pontes, M., Abrantes, D., Castro, A., Fernandez-Fernandez, I. and de Pancorbo, M. (2000). Northe Portugal population genetic data for nine STRs loci, in Olaisen B., Brinkmann, B. and Lincoln P. (Eds), Progress in Forensic Genetics 8, Excerpta Medica, Amsterdam pp. 205-207.

Pinheiro, M., Pontes, M., Huguet, M., Gene, M., da Casta, J. and Moreno, P. (1996). Study of three AMPFLPs (D1S80, 3` ApoB and YNZ22) in the population of the north of Portugal. Forensic Sci. Int. 79, 23-29.

Ploos V. and Reitsema R. (1990). Tetranucleotide repeat polymorphisms in the vWF gene. Nucleic Acids Res. 18, 4957.

Polymeropoulos, M., Xiao, H., Rath, D. and Merril, C. (1991). Tetranucleotide repeat polymorphism at the human tyrosine hydroxylase gene (TH). Nucleic Acids Res. 19, 3753.

Prinz, M., Boll K., Baum, H. and Shaler, B. (1997). Multiplexing of Y chromosome specific STRs and performance for mixed samples. Forensic. Sci. Int. 85, 209-218.

Puers, C., Hammond, H., Caskey, C., Lins, A., Sprecher, C., Brinkmann, B. and Schumm, J. (1994). Allelic ladder characterisation of the short tandem repeat polymorphism located in the 5' flanking region to the human coagulation factor XIII A subunit gene. Genomics 23, 260-264.

Puers, C., Hammond, H., Jin, L., Caskey, C. and Schumm, J. (1993). Identification of repeat sequence heterogeneity at the polymerphic short tandem repeat locus HUMTHO1 $[AATG]_n$ and reassignment of alleles in population analysis by using a locus-specific allelic ladder. Am. J. Hum. Genet. 53: 953-958.

Rand, S., Puers, C., Skowasch, K., Wiegand, P., Budowle, B. and Brinkmann, B. (1992). Population genetics and forensic efficiency data of 4 AmpFLPs. Int. J. Leg. Med. 104, 329-333.

Rangel-Villalobos, H., Rivas, F., Torres-Rodriguez, M., Jaloma-Cruz, A., Gallegos-Arreola, M., Lopez-Satow, J., Cantu, J. and Figuera, L. (1999). Allele frequency distributions of six Amp-FLPs (D1S80, APO-B, vWA, THO1, CSF1PO and HPRTB) in a Mexican population. Forensic Sci. Int. 105, 125-129.

Reynolds, R., Sensabaugh, G. and Blake, E. (1991). Analysis of genetic markers in forensic DNA samples using the polymerase chain reaction. Analytical Chem. 63, 1-14.

Risch, N. and Devlin, B. (1992). On the probability of matching DNA fingerprints. Science 255: 717-720.

Risen, L., Ibanez, E., Podell, S. and Roszak, D. (1991). Chemiluminescent DNA analysis employing SNAP® alkaline phosphatase-labeled probes. Proceedings of the International Symposium on Human Identification, Promega Corporation pp. 318.

Roewer, L. (1998). The Y chromosome: forensic applications and evolutionary aspects, in Olaisen B., Brinkmann B., Lincoln P. (Eds), Progress in Forensic Genetics 7, Excerpta Medica, Amsterdam pp. 243-245.

Roewer, L. and Epplen, J. (1992). Rapid and sensitive typing of forensic stains by PCR amplification of polymorphic simple repeat sequences in case work. Forensic Sci. Int. 53, 163-171.

Roewer, L., Arnemann, J., Spurr, N., Grzeschik, K. and Epplen, J. (1992). Simple repeat sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts. Hum. Genet. 89, 389-394.

Rostedt, I., Lalu, K., Lukka, M. and Sajantila, A. (1996). Genotyping of five short tandem repeat loci via triplex and duplex PCR. Forensic Sci. Int. 82, 217-226

Royle, N., Clarkson, R., Wong, Z. and Jeffreys, A. (1988). Clustering of hypervariable minisatellites in the proterminal regions of human autosomes. Genomics 3, 352-360.

Saiki, R., Bugawan, T., Horn, G., Mullis, K. and Erlich, H. (1986). Analysis of enzymatically amplified β-globin and HLA-DQα DNA with allele-specific oligonucleotide probes. Nature 324, 163-166.

Saiki, R., Gelfand, D., Stoffel, S., Scharf, S., Higuchi, R., Horn, C., Mullis, K. and Erlich, H. (1988). Primer directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239, 487-491.

Saiki, R., Scharf, S., Faloona, F., Mullis, K., Horn, G., Erlich, H. and Arnheim, N. (1985). Enzymatic amplification of B-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 230, 1350-1354.

Saiki, R., Walsh, P., Levenson, C. and Erlich, H. (1989). Genetic analysis of amplified DNA with immobilized sequence specific oligonucleotide probes. Proc Natl Acad Sci USA 86, 6230-6234.

Sajantila A., Strom, M., Budowle, B., Karhunen, P. and Peltonen, L. (1991). The polymerase chain reaction and post-mortem forensic identity testing: application of amplified D1S80 and HLA-DQα loci to the identification of fire victims. Forensic Sci. Int. 51, 23-34.

Sajantila, A., and Budowle, B. (1992). PCR amplification of alleles at the D1S80 locus: comparison of a Finnish and a north American Caucasian population sample, and forensic casework evaluation. Am. J. Hum. Genet. 50, 816-825.

Sajantila, A., Pacek, P., Lukka, M., Syvanen, A., Nokelainen, P., Sistonen, P., Peltonen, L. and Budowle, B. (1994). A microsatellite polymorphism in the von Willebrand Factor gene: comparison of allele frequencies in different population samples and evaluation for forensic medicine. Forensic Sci. Int. 68, 91-102.

Sambrook, J., Fritsch, E. and Maniatis, T. (1989) Molecular Clonong: A laboratory manual. Cold Spring HarbourLaboratory Press, New York, USA.

Sapienza, C. and St-Jacques, B. (1986). 'Brain-specific' transcription and evolution of the identifier sequence. Nature 319, 418-420.

Schaap, A., Akhavan, H. and Romano, L. (1989). Chemiluminescent substrates for alkaline phosphotase: application to ultrasensitive enzyme linked immunoassays and DNA probes. Clin. Chem. 35, 1863-1864.

Schlotterer, C. and Tautz, D. (1992). Slippage synthesis of simple sequence DNA. Nucleic Acids Res. 20, 211-215.

Schmid, C. and Jelinek, W. (1982). The Alu family of dispersed repetitive sequences. Science 216, 1065-1070.

Schnee-Griese, J., Blab, G., Herrmann, S., Schneider, H., Forster, R., Babler, G. and Pflug, W. (1993). Frequency distribution of D1S80 alleles in the German population. Forensic Sci. Int. 59, 131-136.

Schneider, P., Fimmers, R., Woodroffe, S., Werrett, D., Bar, W., Brinkmann, B., Eriksen, B. Jones, S., Kloosterman, A., Mevag, B., Pascali, V., Rittner, C., Schmitter, H., Thompson, J. and Gill, P. (1991). Report of a European collaborative exercise comparing DNA typing results using a single locus VNTR probe. Forensic Sci. Int. 49, 1-15.

Schumm, J., Lins, A., Puers, C. and Sprecher, C. (1993). Development of nonisotopic multiplex amplification sets for analysis of polymorphic STR loci, in Proceedings from the Fourth International Symposium on Human Identification (1993), pp. 177-181.

Scoular, C. (1999). STR analysis of a Scottish population. Forensic science MSc, University of Strthclyde, UK.

Sebetan, I. and Hajar, H. (1998). Analysis of the short tandem repeat (STR) locus HumvWA in a Qatari population. Forensic Sci. Int. 95, 169-171.

Sensabaugh, G. (1982). Isozymes in forensic science. Isozymes Curr. Top. Biol. Med. Res. 6, 247-282.

Sensabaugh, G. (1991). Forensic application of the polymerase chain reaction. J. Forensic Sci. Soc. 31, 201-204.

Serai, H. (1997). History of old Arabian Peninsula. Al-Khaleej cetre for books, Dubai, UAE. (in Arabic).

Shafit-Zagardo, B., Maio, J. and Brown, F. (1982). KpnI families of long, interspersed repetitive DNAs in human and other primate genomes. Nucleic Acids Res. 10, 3175-3193.

Shami, S. and Siddiqui, H. (1984). The effects of parental consanguinity in Rawalpindi city (Punjab), Pakistan. Biologica 30, 189-200.

Siebert P. and Fukuda M. (1987). Molecular cloning of a human glycophorin A. Proc. Natl. Acad. Sci., USA 84, 6735-6739.

Singer, M. (1982). Highly repeated sequences in mammalian genomes. Int. Rev. Cytol. 76: 67-112.

Sinha, S., Amjad, M., Rogers, C., Hamby, J., Tahir, U., Balamurugan, K., Al-Kubaidan, N., Choudhry, A., Budowle, B. and Tahir, M. (1999). Typing of eight short tandem repeat (STR) loci in a Saudi Arabian population. For. Sci. Int. 104, 143-146.

Skowasch, K., Weigand, P. and Brinkmann, B. (1992). PMCT118 (D1S80): a new allelic ladder and improved electrophoretic separation lead to the demonstration of 28 alleles. Int. J. Leg. Med. 105, 165-168.

Sensabaugh, G. (1991). Forensic application of the polymerase chain reaction. J. Forensic Sci. Soc. 31, 201-204.

Serai, H. (1997). History of old Arabian Peninsula. Al-Khaleej cetre for books, Dubai, UAE. (in Arabic).

Shafit-Zagardo, B., Maio, J. and Brown, F. (1982). KpnI families of long, interspersed repetitive DNAs in human and other primate genomes. Nucleic Acids Res. 10, 3175-3193.

Shami, S. and Siddiqui, H. (1984). The effects of parental consanguinity in Rawalpindi city (Punjab), Pakistan. Biologica 30, 189-200.

Siebert P. and Fukuda M. (1987). Molecular cloning of a human glycophorin A. Proc. Natl. Acad. Sci., USA 84, 6735-6739.

Singer, M. (1982). Highly repeated sequences in mammalian genomes. Int. Rev. Cytol. 76: 67-112.

Sinha, S., Amjad, M., Rogers, C., Hamby, J., Tahir, U., Balamurugan, K., Al-Kubaidan, N., Choudhry, A., Budowle, B. and Tahir, M. (1999). Typing of eight short tandem repeat (STR) loci in a Saudi Arabian population. For. Sci. Int. 104, 143-146.

Skowasch, K., Weigand, P. and Brinkmann, B. (1992). PMCT118 (D1S80): a new allelic ladder and improved electrophoretic separation lead to the demonstration of 28 alleles. Int. J. Leg. Med. 105, 165-168.

Slightom, J., Blechl, A. and Smithies, O. (1980). Human fetal $^{G}\gamma$- and $^{A}\gamma$- globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. Cell 21: 627-638.

Slimowitz, J. and Cohen, J. (1993). Violations of the ceiling principle: exact conditions and statistical evidence. Am. J. Hum. Genet. 53, 314-323.

Smith, J., Anwar, R., Riley, J., Jenner, D., Markham, A. and Jeffreys, A. (1990a). Highly polymorphic minisatellite sequences: allele frequencies and mutation rates for five locus-specific probes in a Caucasian population. J. Forensic Sci. Soc. 30, 19-32.

Smith, J., Newton, C., Alves, A., Anwar, R., Jenner, D. and Markham, A. (1990b). Highly polymorphic minisatellite DNA probes. Further evaluation for individual identification and paternity testing. J. Forensic Sci. Soc. 30, 3-8.

Smithies, O. (1955a). Grouped variation in the occurrence of new protein components in normal human serum. Nature 175, 307-308.

Smithies, O. (1955b). Zone electrophoresis in starch gels: Group variation in the serum proteins of normal human adults. Biochem. J. 61, 629-641.

Southern, E. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98, 503-517.

Southern, E. (1995). DAN fingerprinting by hybridisation to oligonucleotide arrays. Electrophoresis 16, 1539-1542.

Sparkes, R., Kimpton, C., Gillbard, S., Carne, P., Anderson, J., Oldroyd, N., Thomas, D., Urquhart, A. and Gill, P. (1996a). The validation of a 7-locus multiplex STR test for use in forensic casework: (II) artefacts, casework studies and success rates. Int. J. Leg. Med. 109, 195-204.

Sparkes, R., Kimpton, C., Watson, S., Oldroyd, N., Clayton, T., Barnett, L., Arnold, J., Thompson, C., Hale, R., Chapman, J., Urquhart, A., Gill, P. (1996b). The validation of a 7-locus multiplex STR test for use in forensic casework: (I) mixtures, ageing, degradation and species studies. Int. J. Leg. Med. 109, 186-194.

Sprecher, C., Krenke, B., Amiott, B. Rabbach, D. and Kelly, G. (2000). The PowerPlex™ 16 system. Profiles in DNA 4 (1), 3-6.

Steinberger, E., Thompson, L. and Hartmann, J. (1993). On the use of excess homozygosity for subpopulation detection. Am. J. Hum. Genet. 52, 1275-1277.

Strom, C., Rechiskey, S., Ginsberg, N., Verlinsky, O. and Verlinsky, Y. (1996). Prenatal paternity testing with deoxyribonucleic acid techniques. Am. J. Obstet. Gynacol. 174, 1849-1853.

Sutcliffe, J., Milner, R., Bloom, F. and Lerner, R. (1982). Common 82-nucleotide sequence unique to brain RNA. Proc. Natl. Acad. Sci. USA 79, 4942-4946.

Tahir, M. Balamurugan, K., Tahir, U., Amjad, M., Ba Awin, M., Chaudhary, O. Hamby, J., Budowle, B. and Herrera, R. (2000). Allelic distribution of nine short tandem repeat (STR), HLA-DQA1, and polymarker loci in an Omani sample population. Forensic Sci. Int. 109, 81-85.

Takeshita, H., Meyer, E. and Brinkmann, B. (1997). The STR loci HumTPO and HumLPL: population genetic data in eight populations. Int. J. Leg. Med. 110, 331-333.

Tamaki, K., Huang, X., Yamamoto, T., Uchihi, R., Nozawa, H. and Katsumata, Y. (1995). Applications of minisatellite variant repeat (MVR) mapping for maternal identification from remains of an infant and placenta. J. Forensic Sci. 40, 695-700.

Taryam, A. (1987). The establishment of the United Arab Emirates 1950-1985, USA.

Tautz D. (1989). Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res. 17, 6463-6471.

Tereba A., (1999). Tools for Analysis of Population Statistics PowerStats. Profiles in DNA 2 (3), 14-16.

Thymann M., Nellennann, L., Masumba, G., Irgens-Moller, L. and Morling, N. (1993). Analysis of the locus D1S80 by amplified fragment length polymorphism technique (AMP-FLP), Frequency distribution in Danes, Intra and inter laboratory reproducibility of the technique. Forensic Sci. Int. 60, 47-56.

Tjio, H. and Levan, A. (1956). The chromosome numbers of man. Hereditas 42, 1-6.

Trabetti, E., Galavotti, R. and Pignatti, P. (1993). Genetic variation in the Italian population at five tandem repeat loci amplified in vitro: use in paternity testing. Mol. Cell. Probes 7, 81-87.

Tsui, P. and Wong, D. (1996). Allele frequencies o four VNTR loci in the Chinese population in Hong Kong. Forensic sci. Int. 79, 175-185.

Tully, U., Sullivan, K. and Gill, P. (1993). Analysis of 6 VNTR loci by 'multiplex' PCR and automated fluorescent detection. Hum. Genet. 92, 554-562.

Tun, Z., Honda, K., Nakatome, M., Nakamura, M., Shimada, S., Ogura, Y., Kuroki, H., Yamazaki, M., Terada, M., Matoba, R., (1999). Simultaneous detection of multiple STR loci on sex chromosomes for forensic testing of sex and identity. J. Forensic Sci. 44, 772-777.

Ullrich, A., Dull, T., Gray, A. and Sures, I. (1980). Genetic variation in the human insulin gene. Science 209, 612-615.

Urquhart, A., Kimpton, C., Downs, T. and Gill, P. (1994). Variation in short tandem repeat sequences – a survey of twelve microsatellite loci for use as forensic identification markers. Int. J. Leg. Med. 107, 13-20.

Valverde, E., Cabrero, C., Cao, R., Rodriguez-Calvo, M., Diez, A., Barros, F., Alemany, J. and Carracedo, A. (1993). Population genetics of three VNTR polymorphisms in two different Spanish populations. Int. J. Leg. Med. 105, 251-256.

van Oorschot, R., Gutowski, S., Robinson, S., Hedley, J. and Andrew, I. (1996). HUMTHO1 validation studies: effect of substrate, environment and mixtures. J. Forensic Sci. 41, 142-145.

Vine, P. and Casey, P. (1992). United Arab Emirates: Heritage & Modern Development. Immel Publishing Limmited, London.

Wainscoat, J., Pilkington, S., Peto, T. and Higgs, D. (1987). Allele-specific DNA identity patterns. Hum. Genet. 75, 384-387.

Walsh, P., Erlich, H. and Higuchi, R. (1992). Preferential PCR amplification of alleles: Mechanisms and solutions. PCR Methods and Applications 1, 241-250.

Waye, J. (1990). Forensic analysis of restriction fragment length polymorphism: theoretical and practical considerations of design and implementation. Proceedings for the International Symposium on Human Identification (1989) pp. 117-152. Promega Co., Madison, WI, USA.

Weber, J. and May, P. (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am. J. Hum. Genet. 44, 388-396.

Weichhold, G. Keil, W. and Bayer, B. (1996). The Y-linked locus Y27H39 (DYS19), frequency distribution in south bavarian and application to paternity testing, in Carracedo A., Brinkmann B. and Bar W. (Eds), Advances in Forensic Haemogenetics 6, Springer, Berlin pp. 641-643.

Weiner, A. (1986). Nonviral retroposons: Genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Annals Review of biochemistry 55, 631-661.

Weir, B. (1992a). Population genetics in the forensic DNA debate. Proc. Natl. Acad. Sci. USA 89, 11654-11659.

Weir, B. (1992b). Independence of VNTR alleles defined as fixed bins. Genetics 130, 873-887.

Weir, B. (1992c). Independence of VNTR alleles defined as floating bins. Am. J. Hum. Genet. 51, 992-997.

Weir, B. (1993). Forensic population genetics and the National Research Council (NRC). Am. J. Hum. Genet. 52: 437-440.

Weir, B. (1996). Genetic data analysis II, Methods for discrete population genetic data. Sinauer Associates, Inc. publishers, Sunderland, Massachusetts, USA.

Whitaker, J., Clayton, T., Urquhart, A., Millican, E., Downs, T., Kimpton, C. and Gill, P. (1995). Short tandem repeat typing of bodies from a mass disaster: high success rate and characteristic amplification patterns in highly degraded samples. Biotechniques 18, 670-677.

Whman, A. and White, R. (1980). A highly polimorphic locus in human DNA. Proc. Natl. Acad. Sci. USA. 77, 6754-6758.

Wiegand, P., Budowle, B., Rand, S. and Brinkmann, B. (1993). Forensic validation of the STR systems SE33 and TC11. Int. J. Leg. Med. 105, 315-320.

Willard, H. and Waye, J. (1987). Hierarchical order in chromosome-specific human alpha satellite DNA. Trends in Genet. 3, 192-198.

Willard, H., Smith, K. and Sutherland, J. (1983). Isolation and characterization of a major tandem repeat family from the human X chromosome. Nucleic Acids Res. 11, 2017-2033.

Wolff, R., Nakamura, Y., and White, R. (1988). Molecular characterisation of a spontaneously generated new allele at a VNTR locus: no exchange of flanking DNA sequence. Genomics 3, 347-351.

Wong, Z., Wilson, V., Jeffreys, A. and Thein, S. (1986). Cloning a selected fragment from a human DNA "fingerprint": isolation of an extremely polymorphic minisatellite. Nucleic Acids Res. 14: 4605 – 4616.

Wong, Z., Wilson, V., Patel, I., Povey, S. and Jeffreys, A. (1987). Characterisation of a panel of highly variable minisatellites cloned from human DNA. Annals of Hum. Genet. 51, 269-288.

Wright, S. (1978). Evolution and the Genetics of populations. Volume 4. Variability within and among natural populations. University of Chicago Press, Chicago, USA.

Yamamoto, T., Davis, C., Brown, M., Schneider, W., Casey, M., Goldstein, J. Russell, D. (1984). The human LDL receptor: a cystein-rich protein with multiple Alu sequences in its mRNA. Cell 39, 27-38.

Yamamoto, T., Tamaki, K., Kojima, T., Uchihi, R., Katsumata, Y. and Jeffreys, A., (1994). DNA typing of the D1S8 (MS32) locus by rapid detection minisatellite variant

repeat (MVR) mapping using polymerase chain reaction (PCR) assay. Forensic Sci. Int. 66, 69-75.

Yamamoto, T., Uchihi, R., Nozawa, H., Huang, X., Leong, Y., Tanaka, M., Mizutani, M., Tamaki, K. and Katsumata, Y. (1999). Allele distribution at nine STR loci – D3S1358, vWA, FGA, THO1, TPOX, CSF1PO, D5S818, D13S317 and D7S820 – in the Japanese population by multiple PCR and capillary electrophoresis. J. Forensic Sci. 44, 167-170.

Yamamoto, T., Uchihi, R.,Mizutani, M., Ohtaki, H., Tanaka, M. and Katsumata, Y. (2000). Allele frequencies at 13 STR loci in the Burmese population, in Olaisen, B., Birnkmann, B. and Lincoln, P. (Eds), Progress in Forensic Genetics 8, Excerpta Madica, Amsterdam pp. 239-241.

Yang, T. Hansen, S., Oishi, K., Ryder, O. and Hamkalo A. (1982). Characterization of a cloned repetitive DNA sequence concentrated on the human X chromosome. Proc. Natl. Acad. Sci. USA 79, 6593-6597.

Zeigle, J., Su, Y., Kevin, P., Corcoran, L., Mayrand, P., Hoff, L., McBride, L., et al., (1992). Application of automated DNA sizing technology for genotyping microsatelite loci. Genomics 14, 1026-1031.

Zierdt, H., Hummel, S. and Herrmann, B. (1996). Amplification of human short tandem repeats from medieval teeth and bone samples. Hum. Biology 68, 185-199.