

The Application of Short Tandem Repeats to Paternity Testing in Egypt

Thesis Submitted in Accordance with the Requirements of the
University of Glasgow for the Degree of Doctor of Philosophy

by

Amany Mahmoud Ahmed

M.B.Ch.B., MSc

Department of Forensic Medicine & Science

November 2000

ProQuest Number: 13818792

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 13818792

Published by ProQuest LLC (2018). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

GLASGOW
UNIVERSITY
LIBRARY

12108-Copy 1

ACKNOWLEDGEMENT

In The Name of God, The Rahman, The Merciful. All thanks to God who taught man what he did not know.

During my studies in Glasgow I have met many people at the University of Glasgow, Department of Forensic Medicine & Science. I would like to thank them for their advice on scientific matters and for their friendliness to me. I would like to acknowledge the following people individually.

Prof. Peter Vanezis and Dr. William Goodwin, my supervisors. They gave me the opportunity and the facilities to conduct this work. They showed complete understanding and patience in their dealings with me, especially towards my difficulties with the English language. Dr. Will was always on hand to listen, discuss results, offer many comments of improvements, explain concepts and give me his time.

I would like to thank Dr. Adrian Linacre Strathclyde University, Forensic Science Unit. Despite his heavy routine workload he always found time to listen and offer advice.

Thanks to Prof. John Craft, the head of department of biological science, Glasgow Caledonian University, he gave me the opportunity to do my cloning work in his laboratory.

I would like to thank Prof. Dupuy B.M. (Institute of Forensic Medicine, University of Oslo, Oslo, Norway) for providing me the sequenced samples.

Thanks to all my friends who supported me during my work, Ahmed, Hadi, Ghada, Glasgow University, Hanan and Heba Glasgow Caledonian University.

I would also like to thank the technician John Harley for ordering the materials, which I needed, Georgi Riboldi-Tunncliffe for running my samples on the ABI 373 automated sequencer.

Lastly I acknowledge the role of my father, mother, and my sons Khaled and Nafie who supported me during my study. My special thanks to my late husband to whom I owe all my success, I will never forget him and I always remember his words, which encourage me to complete my work and to carry out my responsibilities.

CONTENTS

ACKNOWLEDGEMENT.....	I
CONTENTS	III
ABBREVIATIONS	XI
SUMMARY	XII
CHAPTER 1: INTRODUCTION	1
1.1. GENERAL BACKGROUND	1
1.2. HUMAN GENOME	3
1.2.1. <i>Variable Number Tandem Repeats (VNTRs)</i>	4
1.2.2. <i>Short Tandem Repeats (STRs)</i>	6
1.2.2.1. STRs Types	7
1.2.2.2. STRs Nomenclature	10
1.2.2.3. STRs Analysis.....	11
1.2.2.3.1. Forensic STR Selection Criteria.....	12
1.2.2.3.2. STR Multiplexes	12
1.2.2.3.3. The Use of Size Markers and Allelic Ladders in STR Analysis ...	13
1.2.2.3.4. Resolving and Detection STR Alleles.....	14
1.2.2.3.4.1. Polyacrylamide Gel Electrophoresis (PAGE).....	14
1.2.2.3.4.2. Native and Denaturing Separation Systems.....	14
1.2.2.3.4.3. Silver Staining	15
1.2.2.3.4.4. Fluorescent Labelling	15
1.2.2.3.4.5. Automated Detection System.....	15
1.2.2.3.5. Contamination.....	16
1.2.2.4. Validation of STR Systems	16
1.2.2.4.1. Population Studies	17
1.2.2.4.2. Species Specificity	17

1.2.2.4.3. Mixtures	18
1.2.2.4.4. Null Alleles	18
1.2.2.4.5. Non-template Addition of Nucleotides	19
1.2.2.4.6. Stutter Bands.....	19
1.2.2.4.7. Validation of STRs for Paternity Testing.....	20
1.2.3. <i>Y-Chromosome</i>	21
1.2.3.1. Y-Chromosome Polymorphisms	22
1.2.3.1.1. Indels Polymorphisms	22
1.2.3.1.2. Single Nucleotide Polymorphisms (SNPs)	22
1.2.3.1.3. Minisatellites and Microsatellites.....	23
1.2.3.2. Population Distribution Haplotypes	25
1.2.3.3. Forensic Applications	26
1.2.3.4. Paternity Applications.....	26
1.3. PATERNITY TESTING INTERPRETATION.....	27
1.3.1. <i>Paternity Index (PI)</i>	27
1.3.2. <i>Probability of Paternity</i>	28
1.4. PATERNITY TESTING APPLICATIONS	29
1.4.1. <i>Prenatal and Newborn Paternity Testing</i>	29
1.4.2. <i>Deceased Alleged Father</i>	29
1.4.3. <i>Double Paternity in Dizygous Twins</i>	30
1.4.4. <i>Father-Daughter Incest</i>	30
1.4.5. <i>Determination of Sibship</i>	30
1.5. ETHNIC GROUPES OF THE EGYPTIAN POPULATION.....	31
1.6. AIM OF THE PROJECT	ERROR! BOOKMARK NOT DEFINED.
CHAPTER 2: MATERIALS AND METHODS.....	34
2.1. MATERIALS	34
2.1.1. <i>Chemicals</i>	34
2.2. GENERAL PREPARATORY PROCEDURES	34
2.2.1. <i>pH measurement</i>	34
2.2.2. <i>Autoclaving</i>	34
2.2.3. <i>Glassware</i>	35

2.2.4. <i>Solutions and Buffers</i>	35
2.2.4.1. Buffers	35
2.2.4.2. Cloning Reagents	35
2.2.5. <i>Gels</i>	36
2.2.5.1. 1% Agarose Gel	36
2.2.6. <i>Molecular Weight Markers</i>	36
2.2.7. <i>Polymerase Chain Reaction</i>	37
2.2.8. <i>Sequenced Samples for the Ladders</i>	38
2.3. DNA SAMPLE COLLECTION	38
2.3.1. <i>Source of DNA Samples</i>	38
2.3.2. <i>Blood Sampling</i>	39
2.3.3. <i>Buccal swabs</i>	39
2.4. ISOLATION OF DNA FROM WHOLE BLOOD	40
2.4.1. <i>Nucleic Acid Extraction using Puregene™ kit</i>	40
2.5. ISOLATION OF DNA FROM BUCCAL SWABS	41
2.5.1. <i>Nucleic Acid Extraction using Puregene™ kit</i>	41
2.6. QUALIFICATION AND QUANTIFICATION OF DNA	42
2.6.1. <i>Agarose Gel Electrophoresis</i>	42
2.7. POLYMERASE CHAIN REACTION	43
2.7.1. <i>DIS80 Amplification</i>	43
2.7.1.1. PCR Protocol	43
2.7.2. <i>Autosomal Short Tandem Repeats</i>	43
2.7.2.1. PCR Protocol	44
2.7.3. <i>Amplification of Hypervariable Loci</i>	45
2.7.3.1. PCR Protocol	45
2.7.4. <i>Amplification of Y Chromosome STRs</i>	46
2.7.4.1. PCR Protocol	47
2.8. CONSTRUCTION OF LADDERS	48
2.8.1. <i>Construction of Ladders for Autosomal Simple STRs</i>	48
2.8.2. <i>Construction of Ladders for Hypervariable Autosomal STRs</i>	48
2.8.2.1. Cloning of Sequenced DNA Samples	49

2.8.2.1.1. DNA Purification from Gel Bands.....	49
2.8.2.1.2. Ligation of DNA.....	50
2.8.2.1.3. Making Electrocompetent Cells.....	50
2.8.2.1.4. Transformation.....	51
2.8.2.1.5. PCR Screen for Recombinants.....	51
2.8.2.1.6. Plasmid Miniprep.....	52
2.8.2.1.7. Enzymatic Restriction Digestion.....	53
2.8.2.1.8. Amplification of Plasmids.....	53
2.9. ANALYSIS OF PCR PRODUCT.....	53
2.9.1. <i>Non-denaturing Polyacrylamide Gel Electrophoresis</i>	53
2.9.1.1. Treatment of the Plates.....	54
2.9.1.2. Polyacrylamide Electrophoresis.....	54
2.9.2. <i>Agarose Gel Electrophoresis</i>	55
2.9.3. <i>Denaturing Polyacrylamide Gel Electrophoresis</i>	55
2.10. DIRECT SEQUENCING OF PCR PRODUCTS.....	55
CHAPTER 3: SAMPLE COLLECTION.....	57
3.1. INTRODUCTION.....	57
3.2. IDENTIFICATION OF A RELEVANT POPULATION.....	58
3.3. ASSESSMENT OF METHODS FOR PRESERVING SAMPLES.....	60
3.3.1. <i>Blood and Buccal Samples</i>	60
3.3.2. <i>Quantity and Quality of Extracted DNA</i>	61
3.3.3. <i>Amplification of Preserved Blood Samples</i>	62
3.3.3.1. D1S80 Locus.....	68
3.4. SAMPLING PROCEDURE.....	72
3.4.1. <i>Blood Samples</i>	72
3.4.2. <i>Buccal Samples</i>	72
3.5. EXTRACTION OF SAMPLES.....	73
3.6. DISCUSSION.....	73
3.6.1. <i>Sample Collection</i>	73
3.6.2. <i>Preservation of Samples</i>	75
3.6.3. <i>DNA Extraction</i>	76

3.6.4. <i>Quantity of Extracted DNA</i>	77
3.6.4.1. Blood Samples.....	77
3.6.4.2. Buccal Swabs.....	77
CHAPTER 4: AUTOSOMAL STR ANALYSIS.....	80
4.1. INTRODUCTION.....	80
4.2. SELECTION OF STR LOCI.....	80
4.3. DESCRIPTION OF THE LOCI.....	83
4.3.1. <i>D3S1358 Locus</i>	83
4.3.2. <i>Human vWA Locus</i>	83
4.3.3. <i>Human FGA Locus</i>	84
4.4. SELECTION OF PRIMERS.....	84
4.5. OPTIMIZATION OF THE PCR SYSTEM.....	85
4.5.1. <i>Single Locus PCR</i>	85
4.5.1.1. Annealing Temperature.....	85
4.5.1.2. Magnesium Chloride.....	86
4.5.1.3. Primer Concentration.....	86
4.5.1.4. <i>Taq</i> Polymerase.....	86
4.5.1.5. Extension Time and Temperature.....	87
4.5.2. <i>Multiplex PCR</i>	87
4.6. SIZING AND DESIGNATION OF ALLELES.....	89
4.7. STATISTICAL ANALYSIS.....	100
4.7.1. <i>Genotype Frequencies</i>	100
4.7.2. <i>Allele Frequencies</i>	101
4.7.3. <i>Heterozygosity</i>	108
4.7.4. <i>Paternity Parameters</i>	109
4.7.4.1. Paternity Index.....	110
4.7.4.2. Power of Exclusion.....	111
4.7.5. <i>Forensic Parameters</i>	111
4.7.5.1. Matching Probability.....	111
4.7.5.2. Discrimination Power.....	112
4.7.6. <i>Hardy-Weinberg Equilibrium</i>	112

4.7.6.1. Exact Test.....	113
4.7.6.2. Chi-Square Test.....	113
4.7.7. <i>Independence of Loci</i>	115
4.8. COMPARISON BETWEEN POPULATIONS.....	116
4.8.1. <i>Introduction</i>	116
4.8.2. <i>Comparison with Cairo Population</i>	117
4.8.2.1. F-Statistics.....	119
4.8.3. <i>Comparison with Other Populations</i>	120
4.9. DISCUSSION.....	124
4.9.1. <i>PCR Optimisation</i>	124
4.9.2. <i>Sizing of the Allele</i>	125
4.9.3. <i>Paternity and Forensic Parameters</i>	126
4.9.4. <i>HWE Tests</i>	128
4.9.5. <i>Comparison of the Two Egyptians Studies</i>	128
4.9.6. <i>Comparison with Other Populations</i>	129
CHAPTER 5: HYPERVARIABLE LOCI.....	131
5.1. INTRODUCTION.....	131
5.2. DESCRIPTION OF LOCI	132
5.2.1. <i>D11S554 Locus</i>	132
5.2.2. <i>HUMACTBP2 (SE33) Locus</i>	133
5.2.3. <i>HUMAPOAII Locus</i>	134
5.3. PRIMER DESIGN	135
5.4. OPTIMIZATION OF THE PCR SYSTEM.....	138
5.5. CONSTRUCTION OF ALLELIC LADDERS	138
5.5.1. <i>Cloning of the Ladders</i>	141
5.5.1.1. DNA Purification from Gel Bands	141
5.5.1.2. Ligation and Transformation	142
5.5.1.3. Plasmid Miniprep.....	144
5.5.2. <i>Amplification of the Plasmid</i>	144
5.6. SIZING AND DESIGNATION OF ALLELES.....	149
5.6.1. <i>Allelic Designation</i>	149

5.6.1.1. Intergel Variation	149
5.6.1.2. Intragel Variation	151
5.7. GENOTYPE OF THE EGYPTIAN SAMPLES	151
5.8. NOMENCLATURE	154
5.9. STATISTICAL ANALYSIS	156
5.9.1. Genotype Frequency	156
5.9.2. Allele Frequency	156
5.9.3. Paternity Parameters	157
5.9.3.1. Paternity Index	157
5.9.3.2. Exclusion of Paternity	162
5.9.4. Forensic Parameters	162
5.9.4.1. Matching Probability	162
5.9.4.2. Discrimination Power	162
5.10. HARDY-WEINBERG EQUILIBRIUM.....	164
5.11. HETEROZYGOSITY TEST	164
5.12. COMPARISON WITH OTHER POPULATIONS	166
5.13. COMBINING ALL SIX LOCI	166
5.14. DISCUSSION	168
5.14.1. PCR Optimisation	168
5.14.2. Sizing of the Alleles	168
5.14.3. Construction of the Ladders	169
5.14.4. Paternity and Forensic Parameters	169
5.14.5. HWE	170
5.14.6. Combination of Six Loci.....	171
CHAPTER 6: Y-CHROMOSOME STR	173
6.1. INTRODUCTION.....	173
6.2. SELECTION OF Y CHROMOSOME STR LOCI.....	174
6.3. OPTIMISATION OF THE AMPLIFICATION CONDITIONS.....	174
6.4. CONSTRUCTION OF THE ALLELIC LADDERS.....	176
6.5. IDENTIFICATION OF A NEW ALLELE	180
6.6. STATISTICAL ANALYSIS	180

6.7. COMPARISON WITH OTHER POPULATIONS	191
6.8. DISCUSSION	191
CHAPTER 7: PATERNITY CASE WORK.....	195
7.1. INTRODUCTION.....	195
7.2. PATERNITY CASES	196
7.3. EXCLUSION POWER ANALYSIS	213
7.4. DISCUSSION	216
CHAPTER 8: CONCLUSION AND FUTURE WORK.....	219
8.1. WHAT HAS BEEN ACHIEVED?.....	219
8.2. FUTURE WORK.....	221
APPENDIX 1.....	223
APPENDIX 2.....	228
APPENDIX 3.....	230
REFERENCES	235

ABBREVIATIONS

AABB	American Association of Blood Banks
ACTBP2	beta-actin related pseudogene H-beta-Ac-psi-2
bp	base pair
DNA	deoxyribonucleic acid
EDNAP	European DNA Profiling Group
EtBr	ethidium bromide
FES/FPS	c-fes/fps Proto-oncogene
FGA	alpha fibrinogen gene
Gc	group specific component
GDA	genetic data analysis
HLA	human leukocyte antigen
HP	haptoglobin
HWE	Hardy-Weinberg equilibrium
IPTG	isopropylthiogalactoside
ISFH	International Society of Forensic Haematogenetics
Kb	kilobase
Mb	megabase
MRCA	most recent common ancestor
MP	matching probability
NRC	National Research Council
PCR	polymerase chain reaction
PD	power of discrimination
PE	power of exclusion of paternity
PGM	phosphoglucomutase
PI	paternity index
psi	pounds per square inch
Rh	Rhesus system
SD	standard deviation
SNPs	single nucleotide polymorphisms
STR	short tandem repeat
TEMED	N,N,N,N,-tetramethylethylene diamine
TFPGA	tool for population genetic analysis
TPI	typical paternity index
V	volt
VNTR	variable number tandem repeat
vWA	von Willebrand factor A
W	probability of paternity
X-Gal	5-bromo-4-chloro-3-indolyl β -D-galactopyranoside
YAP	Y-chromosome alu polymorphism

SUMMARY

The project has involved assessing new tools to apply to paternity testing within Egypt. Before examining different genetic markers it was very important to study the Egyptian population for substructure, particularly as the Egyptian population is composed of several ethnic groups and has relatively high levels of consanguineous marriages. In order to examine this 300 blood and buccal samples were collected from unrelated Egyptians living in El-Minia city. Before this could be done a protocol for collecting and preserving biological samples in harsh environmental conditions was developed.

Three simple STR loci D3S1358, HUMvWA and HUMFGA were chosen to investigate the population structure, new primers were designed for all three loci and after optimisation 300 samples were typed. A number of parameters were examined in order to assess the degree of population substructure. No deviation from the Hardy Weinberg equilibrium (HWE) was detected at any of the three loci using the exact and heterozygosity tests. The coancestry coefficient (F_{ST}) was calculated by comparing the El-Minia data to a Cairo population; the low value (0.001) provided further evidence that there is not a high degree of substructuring in the Egyptian population.

The three simple loci alone were of limited value as a tool for paternity testing, therefore three hypervariable STR loci D11S554, ACTBP2 and APOAI1 were analysed. A new primer was designed for the APOAI1 locus and the three loci were

amplified as a triplex. After optimisation of the triplex the 300 Egyptian samples were profiled. As with the simple loci, all three were in HWE and there was no evidence for linkage between loci.

Allelic ladders were constructed for all six loci examined to ensure accurate and precise typing. The hypervariable ladders were cloned into plasmids in order to create an inexhaustible stock.

The six loci used together were a powerful tool for determining paternity issues, the theoretical value of the loci was compared to the actual value after undertaking 30 cases from Egypt. All the cases were inclusion with high paternity indexes and probabilities of paternity. The hypervariable loci were particularly valuable when assessing motherless cases.

To have a complete system for paternity testing, which covers as many case scenarios as possible, as in case of deceased father, a database of seven Y-chromosome STR loci was established using 2 multiplexes. The loci included in multiplex I were DYS19, DYS389I, DYS389II and DYS390 and the loci included in multiplex II were DYS391, DYS392 and DYS393. The study of these loci showed a high power of exclusion of paternity in the Egyptian population. Also an as yet unpublished allele was detected at the DYS392 locus.

CHAPTER 1: INTRODUCTION

1.1. GENERAL BACKGROUND

Paternity testing is required legally in some paternity and immigration cases (Jeffreys *et al.*, 1985), socially in private disputed paternity cases and medically in prenatal diagnosis, genetic counselling and transplantation (Macintyre and Sooman, 1991).

Most cases of doubtful parentage involve a dispute over the biological father; it is a rare event that there is a question over the mother. The development of reliable methods for determining paternity has progressed slowly (Silver, 1989). All methods depend on identifying genes that a child inherits from the father and the mother (Elston, 1986). The most significant developments in resolving paternity cases occurred in the 20th century starting with Landsteiner's description of the ABO blood group system in 1900 (Silver, 1989). In the ABO marker, the population is classified into only four types; group A, B, AB and O. There is a relative high frequency for each type. ABO markers can exclude the man from being the biological father. However, ABO typing excludes only 17% of falsely accused putative fathers (Polesky and Krause, 1977). In 1927 Landsteiner and Levins discovered the next blood system used for paternity testing, which was the MN system. It was followed by the discovery of the Rhesus system (Rh) in 1939 (Silver, 1989). When Rh and MNs are added to ABO typing, the

power of exclusion increased to 53% (Polesky and Krause, 1977). In between 1940 and 1970, other markers such as Kell, Duffy and Kidd were discovered. The maximal chance of an exclusion of paternity using all these systems together was 70% (Silver, 1989).

Further major advances occurred in the 1970's when the human leukocyte antigens (HLA) were utilized in paternity testing and were used in more than 85% of the cases in each year in some labs (Jerome and Gottschall, 1991). The HLA system has a high power of exclusion of approximately 90%, which could be combined with the red cell antigens to produce an overall power of exclusion ranging from 91% to 97% (Houtz *et al.*, 1981). At the same time that the HLA system was recognised, some red cell enzymes and serum proteins such as Phosphoglucomutase (PGM), Haptoglobin (HP) and Group specific component (Gc) were discovered (Dykes and Polesky, 1976). The probability of exclusion of paternity for the combination of all the systems was 99% (Strom *et al.*, 1996).

Despite the classical serological markers having a relatively high exclusion power, their use in cases of inclusion of paternity is limited because of the relatively high chance for a random man in the population having the same marker phenotype as the true father for most of these markers (Chimera *et al.*, 1989).

DNA markers have largely replaced the standard serological markers for resolving paternity cases (Dodd, 1985) as DNA markers have a number of advantages over the serological markers. DNA markers provide higher exclusion power than the

serological markers (Markowicz *et al.*, 1990). In contrast to the conventional blood group analysis, which depends on the availability of blood or body fluid, the results of DNA profiling do not depend so much on the nature of material because the same genetic information is contained in every nucleated somatic cell of an individual. Thus the same DNA type may be obtained from any tissue e.g., blood, saliva, skin, bone, and urine (Jeffreys, 1993). The DNA molecule is more stable and can be used to resolve disputed paternity from different samples such as dried blood and semen, hair, tissues embedded in paraffin and even exhumed remains in cases of deceased parents (Kanter, *et al.*, 1986; Giusti *et al.*, 1986; Alford *et al.*, 1994 and Chakraborty and Stivers, 1996). Additionally, DNA analysis is less time consuming than the analysis of a large number of markers of the serological systems (Strom *et al.*, 1996). DNA analysis can be used in prenatal paternity testing (Strom *et al.*, 1996). The use of DNA markers offers the ability to solve motherless paternity cases (Brenner, 1993), and also in cases of male children where the father is deceased or not available, by typing the Y-chromosome from any of the father's male relatives if possible (Kayser *et al.*, 1998). Therefore, DNA profiling has become a standard technique in the field of paternity testing and also in forensic crime analysis (Baird *et al.*, 1986).

1.2. HUMAN GENOME

The human genome is composed of approximately 3 billion base pairs (bp). The genome has coding and non-coding regions. The coding regions contain DNA sequences (genes), which determine the amino acid sequences of the proteins (Craig *et al.*, 1988). The non-coding regions contain repetitive sequences of different length.

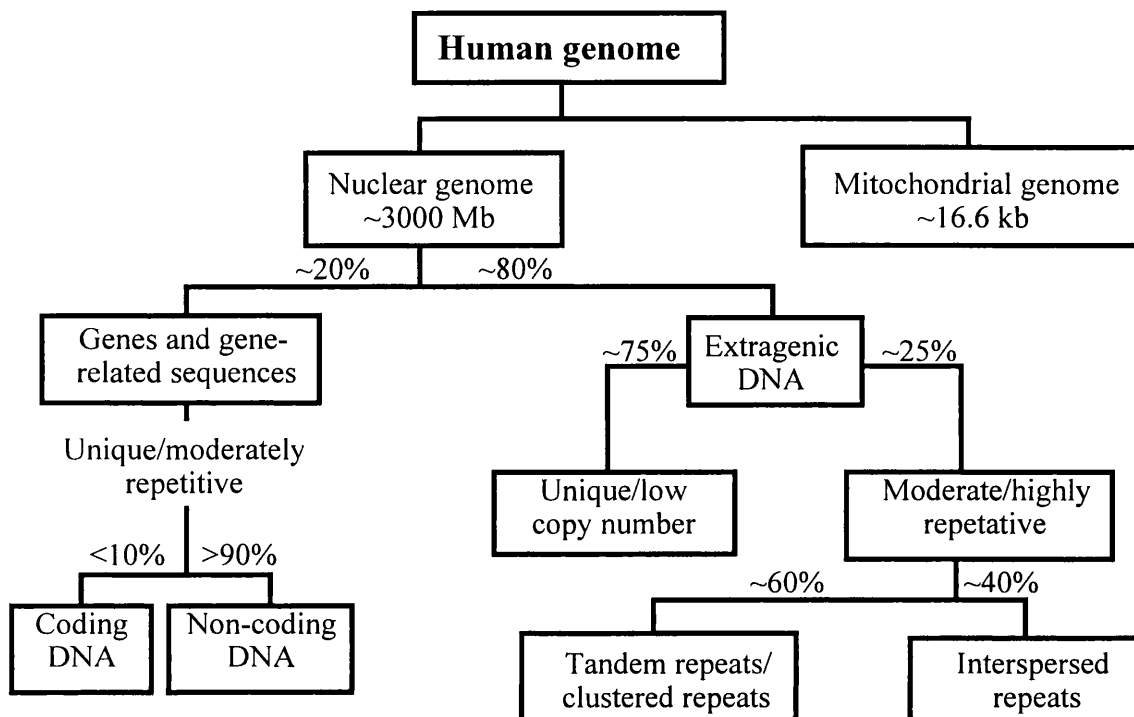
The function of the DNA sequences of repetitive regions is unknown (Hardman, 1986). There are several types of repetitive DNA. Tandemly-repeated DNA has been utilised for forensic studies, this type of DNA can be divided into VNTRs and STRs. Schematic representation of the human genome is shown in figure 1.1.

1.2.1. Variable Number Tandem Repeats (VNTRs)

VNTR loci are situated in the non-coding regions and range in size from approximately 500 bp to 20 kb (Nakamura *et al.*, 1987). They are the most variable loci consisting of core tandemly-repeated units. The length of the repeat units ranges from 9 to 64 bp (Jarman *et al.*, 1986). In the tandem repeat units, there are similarities in the DNA sequence, which define the minisatellite core sequence (Jeffreys *et al.*, 1985b). The number of repeats varies widely from one individual to another and the size of the core repeats differs between loci (Lee *et al.*, 1994).

VNTR loci generally have a large number of alleles, each of which occurs at a low frequency. Because of the high degree of informativeness, VNTR systems became the first powerful tools available for paternity testing (Pena and Chakraborty, 1994). Several loci can be studied simultaneously by using multi-locus probes, which will detect and hybridize at several loci on the genome to produce a multiple-band DNA profile that has been referred to as a DNA fingerprint, although this term has been replaced by DNA profile. The most useful multi-locus probes were named 33.6, 33.15 (Jeffreys *et al.*, 1985b). Often twenty to thirty readable bands could be compared between the mother, child and alleged father.

Figure 1.1 Schematic representation of the human genome



Exclusion is detected by the presence of more than one band in the child not seen in the mother or alleged father (Pena and Chakraborty, 1994). Mutations can occur so exclusion cannot be based upon a single band mismatch as this occurs in about 10% of true families (Dykes, 1990). It is therefore unclear how many bands are required to confidently exclude a father. The multi-locus probes have good resolving power but there are difficulties in the interpretation of the data, which have restricted their use (Dykes, 1990).

In 1986, the single locus probe technique was introduced. It detects two alleles only in heterozygous individuals, one inherited from the mother and one from the father (Kanter *et al.*, 1986). Single-locus probes provide genotype information as they identify several alleles at a specific locus. The comparison of the bands from a single locus probe is easier than that using a multi-locus probe (Smith *et al.*, 1990). However, there are drawbacks in the use of single locus probes method, as the Southern hybridisation technique requires high molecular weight DNA. This technique is also expensive and is time consuming.

1.2.2. Short Tandem Repeats (STRs)

Short tandem repeats (STRs) are simple tandemly-repeated sequences of two to seven base pair in length (Craig *et al.*, 1988). STRs are abundant throughout the human genome and situated mainly in non-coding introns, which represent 90% of the structure of the human genes and they also situated in flanking sequences and within

coding regions (Edwards *et al.*, 1991). STR loci are also present within the 3' end of Alu repeats (Econonou *et al.*, 1990).

STR alleles are short in length typically between 100-500 bp, therefore they are easily amplified by PCR and separated on polyacrylamide gels. Like VNTRs, STRs exhibit a high degree of length polymorphism due to variation in the number of repeated core units (Edwards *et al.*, 1991).

More than 1,300 STRs system have been identified, some of these loci have a heterozygosity of more than 70%. STRs constitute a large source of highly polymorphic alleles, which are suitable for forensic DNA analysis (Edwards *et al.*, 1991).

1.2.2.1. STRs Types

The most common class of STR loci in the human genome contains a (CA)_n (TG)_n dimer and is known as CA repeat. The presence of CA repeats in high copy number and their occurrence throughout the genomes of all eukaryotes was first discovered more than ten years ago (Miesfeld *et al.*, 1981 and Hamada *et al.*, 1982). In the human genome there are 50,000-100,000 interspersed (CA)_n blocks, with a range of n being roughly 15-30 (Braaten *et al.*, 1988). In 1989 three groups simultaneously reported the detection of variability in these microsatellites. They found the repeats were highly polymorphic in CA or GT repeat copy number with alleles in the 70-300 base pair range including primer and flanking sequences (Weber and May, 1989; Litt and Luty, 1989 and Tautz, 1989).

Another class of microsatellites involves tri and tetra nucleotide repeats. The first reported group of simple repeats was the GATA/GACA sequences, which were originally identified in and isolated from snake satellite DNA (Epplen *et al.*, 1982). In 1989 Dryja *et al.*, used PCR to amplify a (CTTT)_n repeat linked to retinoblastoma gene in human chromosome 13. Similar results showed another polymorphic tetranucleotide GATA repeat in the intron 40 of the von Willebrand factor gene (Peake *et al.*, 1990). Trimeric and tetrameric tandem repeats were shown to occur every 300 to 500 kb in the human X chromosome and appear to be interspersed at this frequency throughout the genome (Edwards *et al.*, 1991). Tetranucleotide loci are attractive for forensic and paternity analysis because of the low rate of artifacts generated by repeat slippage, which are common features with dinucleotide repeat loci (Beckman and Weber, 1992).

Tetranucleotide STRs have been classified into four different types (Urquhart *et al.*, 1994). The first type is the simple repeat, which consists of one identical repeating sequence e.g. c-fes/fps Proto-oncogene (HUMFES/FPS locus) (tumor gene precursor). The second is the compound type consists of two or more different repeat sequences e.g. von Willebrand factor A (HUMvWA) locus (blood coagulation). The third type is the complex repeat consists of several regular repeat units with interspersions of dimer, trimer and hexamer invariants e.g. D21S11 locus. The fourth type is the hypervariable repeats, which consists of tetranucleotide repeats with different monomer, dimer, trimer and hexamer invariants that are scattered throughout the locus e.g. Human beta-actin related pseudogene H-beta-Ac-psi-2 (ACTBP2 locus). The different types are shown in table 1.1.

Tabl 1.1 Different types STR loci

The simple repeat, which consists of one identical repeating sequence. The compound type consists of two or more different repeat sequences. The complex repeat consists of several regular repeat units with interspersed dimer, trimer and hexamer invariants. The fourth type is the hypervariable repeats, which consists of tetranucleotide repeats with different monomer, dimer, trimer and hexamer invariants scattered throughout the locus.

Type	Locus	Repeat
Simple	HUMFES	(ATTT) _n
Compound	HUMvWA	(TCTG) ₃₋₄ (TCTA) _n
Complex	D21S11	Class 1: (TCTA) _n (TCTG) _n (TCTA) ₃ TA(TCTA) ₃ TCA(TCTA) ₂ TCCATAT(TCTA) _n TC Class 2: (TCTA) _n (TCTG) _n (TCTA) ₃ TA(TCTA) ₃ TCATCTA) ₂ TCCATAT(TCTA) _n TA.TCTA.TC
Hypervariable	ACTBP2	(AAAG) with different monomer, dimer, trimer, tetramer and hexamer invariants that are scattered throughout the locus.

1.2.2.2. STRs Nomenclature

In October 1993, the DNA commission of the International Society of Forensic Haematogenetics (ISFH) recommended basic principles for the nomenclature of the STR system, which is commonly used today. STRs within protein coding genes are named after the sequence of the gene such as the vWA locus and Hum FIBRA (FGA) locus (Barber *et al.*, 1995, 1996). The same is applied to the pseudogenes such as the ACTBP2 locus (Moos and Gallwitz, 1983).

Repetitive sequences without any connection to protein coding genes are named according to their chromosomal location such as D3S1358 and D11S554 loci being on chromosomes 3 and 11 respectively (Bar *et al.*, 1997).

The number of repeats which STR loci contain is generally used to name alleles. 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 HUMTH01 contains a non-consensus allele, which is one base pair shorter than ten repeats units and is thus designated 9.3 (Puers *et al.*, 1993).

For complex repeats system, although D21S11 is described as a tetrameric, it behaves the same as a dimeric repeat so there are two different nomenclatures. The first one is based upon the number of complete tetramers ignoring the invariant non-tetramers repeats (Moller *et al.*, 1994). The second method defines alleles by the number of dinucleotides present including the invariant trimeric as one repeat (Urquharts *et al.*, 1994). With the hypervariable loci, the repetitive structure can be

very complex and alleles are designated by their size in bp in comparison with a sequenced ladder (Schneider *et al.*, 1998).

1.2.2.3. STRs Analysis

PCR-based STRs have several advantages over the conventional Southern blotting techniques of VNTRs. PCR has greater sensitivity than Southern blotting (Lorente *et al.*, 1994). Additionally, STR analysis has lower cost, greater speed, and needs smaller quantities of DNA than Southern blotting (Edwards *et al.*, 1991). Successful investigations were performed from degraded DNA such as skeletal remains (Hochmeister *et al.*, 1995), human faeces (Mannucci *et al.*, 1995), urine (Benecke *et al.*, 1997) and body parts after a mass disaster (Clayton *et al.*, 1995).

Amplification of VNTRs with PCR did produce more precise allele determination (Boerwinkle *et al.*, 1989). However, if the two alleles of a heterozygote vary by a large number of repeats or the repeat unit itself is long, the smaller of the two alleles will be preferentially amplified possibly leading to allelic drop out and the person would be considered to be a homozygote (Walsh *et al.*, 1992). In STR systems, this phenomenon does not happen frequently because the repeat units are only four or five base pairs long and the alleles are relatively short (Wilson *et al.*, 1993). Therefore, the introduction of STR typing into the paternity testing has become widespread (Rostedt *et al.*, 1996).

1.2.2.3.1. Forensic STR Selection Criteria

STRs, which demonstrate a high degree of variability within the population, are normally selected for forensic applications. Loci with a low frequency of stutter band such as tetranucleotide repeats are preferable (Urquhart *et al.*, 1994). STR loci having small sizes are easy to amplify and allow very small amounts of DNA template to be amplified. This gives a low chance of allelic drop out particularly in degraded samples (Gill *et al.*, 1995). Complex STR loci have a large number of alleles and allele frequencies are generally less than 10%. These loci are difficult to standardise between different laboratories, however, they are significantly more efficient in paternity testing and forensic analysis because the alleles occur at a low frequency (Schneider, 1997). Different chromosomal locations are chosen to ensure exclusion of closely linked loci and absence of linkage. Agreement with Hardy-Weinberg equilibrium is important in forensic systems (Urquhart *et al.*, 1996). In paternity analysis, the mutation rate of the loci chosen must be extremely low to avoid false exclusion of suspected fathers. In all cases, the reliability and reproducibility of the results are also essential (Gill *et al.*, 1995).

1.2.2.3.2. STR Multiplexes

STRs are less variable than VNTRs but several STRs loci can be amplified in the same PCR reaction to increase the speed of the analysis and therefore decreasing the cost. The selection of individual STR loci, each with a limited size range of known alleles, meant that several STR systems could be detected simultaneously in defined

region of the same lane on a gel. This would increase both the power of discrimination and the power of exclusion of paternity (Kimpton *et al.*, 1993).

To construct a multiplex system, primers must be chosen so that annealing temperatures are similar. Affinity, either to each other or to regions of the DNA outside the specific target template, has to be low. There are several different fluorescent dyes which facilitate the development of STR multiplex by allowing the labelling of loci that have overlapping allele size ranges with different fluorescent dyes (Ziegle *et al.*, 1992).

1.2.2.3.3. The Use of Size Markers and Allelic Ladders in STR Analysis

Size markers in DNA typing have been used since the beginning of VNTR analysis (Wyman and White, 1980). With the development of automated systems, the fluorescence internal size markers labelled for example with carboxy-X- rhodamine, can be included within each lane. This reduces differences in electrophoretic mobility between gel lanes and allows automatic sizing of STR products with GENESCAN 672 analysis software (Gill *et al.*, 1995). For the majority of loci, direct comparison with lambda *Pst* I restriction digests (ABD GS350 or GS2500) achieved reliable sizing (Gill *et al.*, 1996).

However, DNA migration in gel electrophoresis does not depend only on the length of the DNA, but also on its nucleotide sequence (Frank and Koster, 1979). If the sequences of marker and DNA sample are not the same, there may be a shift in the migration of DNA sample relative to marker fragments due to change in environmental

parameters, such as acrylamide percentage, the amount of salt in the buffer and the electrophoresis voltage. Therefore, it is recommended to use allelic ladders containing a range of alleles, which have the same sequences and length as the STR locus to avoid the environmental changes and in order to accurately designate the alleles in any given sample (Puers *et al.*, 1993).

1.2.2.3.4. Resolving and Detection STR Alleles

1.2.2.3.4.1. Polyacrylamide Gel Electrophoresis (PAGE)

DNA possesses a constant mass-to-charge ratio. Some form of separation matrix is needed to separate DNA fragments by their molecular size. The requirement for a sieving matrix is met with polyacrylamide or agarose gels. The movement of larger DNA fragments is impeded relative to that of the smaller DNA fragments as the molecules migrate through the gel under influence of an electric current (Heller, 1999).

White and Kusakawa (1997) demonstrated that Metaphor agarose gel could have sufficient resolving power to type dinucleotide and tetranucleotide repeats using SYBR Green to detect the alleles. However, high resolution, denaturing polyacrylamide gels are capable of single base resolution and these gels have become the most popular method for separating STR alleles (Hammond *et al.*, 1994).

1.2.2.3.4.2. Native and Denaturing Separation Systems

In native gel systems, DNA fragments are separated as double stranded fragments. Single base resolution is not achieved under native conditions (Fregeau and Fourney, 1993).

A denaturing gel system is produced by adding high levels of urea and/or formamide to the gel. With a denaturing system, resolution of single base pairs is achieved (Robertson, 1994).

1.2.2.3.4.3. Silver Staining

After electrophoretic separation, DNA fragments may be stained with silver nitrate. Silver staining is less expensive than fluorescence detection techniques because it does not require expensive equipment. Double banding pattern is detected when high-resolution denaturing gels are used. The double bands can make the interpretation of mixtures more difficult with all of the extra bands (Bassam *et al.*, 1991).

1.2.2.3.4.4. Fluorescent Labelling

A fluorescent dye is added on the 5' end of either the forward or the reverse PCR primer. In a denaturing gel, the two strands of DNA are separated during electrophoresis and only the fluorescently labelled strand is detected. This method is commonly used with automated fluorescent detection systems, such as the ABI 373 or 377 (Sullivan *et al.*, 1992).

1.2.2.3.4.5. Automated Detection System

The Applied Biosystems 370A automated sequencer was introduced in 1987. The principle of the machine is that DNA could be detected as it passed through the electrophoretic gel. The detection system used a fluorescent primer, which is excited by a laser and then detected by the camera at the lower end of the gel (Smith *et al.*, 1986). Now various automated fluorescence detection systems have been used for

separation, detection and typing of STR alleles amplified by PCR using fluorescent primer. There are a number of different dyes that can be attached to one of the primers. These dyes are known as FAM, HEX, JOE, TET, NED, ROX and TAMARA. The use of fluorescent dyes allows an internal lane standard to be run with a differently labelled dye so that each sample may be compared to an internal size standard (Roy *et al.*, 1996).

1.2.2.3.5. Contamination

One disadvantage of PCR based methods is that they are prone to contamination. Therefore, it necessary to separate laboratory procedures, such as DNA extraction, PCR amplification and post PCR detection into designated working areas. The most common sources of DNA contamination include: sample contamination with genomic DNA from the environment, contamination between samples during preparation, and contamination of a sample with amplified DNA from previous PCR (Lygo *et al.*, 1994). The contamination process can be overcome by setting up the PCR in the flow-hood.

1.2.2.4. Validation of STR Systems

Before a new STR system may be routinely employed in human identity testing it should be validated to ensure reliability of results (von Oorschot *et al.*, 1994; Urquhart *et al.*, 1995; Micka *et al.*, 1996; Sparkes *et al.*, 1996, 1997 and Wallin *et al.*, 1998).

1.2.2.4.1. Population Studies

200 unrelated individuals and/or 400 meioses are commonly acceptable to be investigated for each STR system to study allele frequencies, the Mendelian inheritance and existence of mutations (Brinkmann, 1998). STR systems that have a low frequency, a high discriminating index and a low mutation rate are suitable for forensic science applications (Brinkmann, 1996).

1.2.2.4.2. Species Specificity

Crouse and Schumm (1995) examined the species specificity of nine STR systems (CSF1PO, TPOX, TH01, HPRTB, FES/FPS, vWA, F13A1, and CSF1PO/TPOX/TH01 and HPRTB/FESFPS/vWA triplex). They reported that no STR PCR products were obtained for 17 of the 23 species tested. Most of the STR PCR products migrated outside of the human allelic ladder fragment range and could not be typed with the PAGE silver stain method employed.

The British Forensic Science Service studied six STR loci and the amelogenin sex-test. No amplification for STR loci was detected among the non-primate while amplification occurred among primates especially the chimpanzees. Although, the size of the allele fell outside the size recorded for humans (Urquhart *et al.*, 1994). No detectable PCR products were observed in the non-primate DNA samples for the AmpF ℓ STR kit (Wallin *et al.*, 1998).

Microbial DNA templates from 30 micro-organisms that may be found in forensic samples were examined with the STR systems HUMTH01, TPOX, CSF1PO and no PCR products were observed at these loci (Fernandez *et al.*, 1996). Another study demonstrated that there were no PCR products for different micro-organisms examined for the AmpF ℓ STR SGM plus kit (Cotton *et al.*, 2000). This type of study has been done for all the commercial kits.

1.2.2.4.3. *Mixtures*

DNA mixtures originating from multiple donors may be present in forensic samples, especially from rape cases where DNA from both the perpetrator and the victim are mixed. A mixture is detected by the presence of 3 or more bands at one or more STR loci (Lygo *et al.*, 1994). Examination of the mixtures of DNA in defined ratios and examination of the peak height ratios between heterozygous alleles at an STR locus are preferable to recognize the mixtures especially in the case where the components of the mixture have the same alleles at a particular locus (Sparkes *et al.*, 1996 and Wallin *et al.*, 1998).

1.2.2.4.4. *Null Alleles*

A null allele is a term for the failure to amplify DNA sequences that are present in a sample. Hammond *et al.*, (1994) discovered the presence of null alleles in STR loci. The most likely cause of this phenomenon is a polymorphism in the primer binding site causing possible failure to anneal during amplification (Gusmao *et al.*, 1996). The incidence of null alleles can be investigated by two different methods.

Firstly by using two different sets of non-overlapping primers where a null allele is identified as a homozygote with one primer set and a heterozygote with the other set. Secondly by parentage analysis where homozygote children with a null allele can be tested (Gill *et al.*, 1995). The problem of null allele occurrence can be resolved by redesigning the primer sets to avoid the polymorphism in the former primer-annealing region (Urquhart *et al.*, 1995).

1.2.2.4.5. Non-template Addition of Nucleotides

An extra adenine is often added to the 3' end of a PCR product resulting in what is commonly referred to as non-template addition. It was found that addition of an extra base by *Taq* DNA polymerase resulted in a PCR product one base pair larger than its predicted length (Oldroyd *et al.*, 1995). This addition of adenine nucleotide is locus specific and for each allele two peaks differ by one base may appear, the larger of the two peaks will be called as the allele (Oldroyd *et al.*, 1995). For STRs system with alleles that can differ by one bp it can be problematic. The problem is largely overcome by adding a final 10 min extension period at 72 °C (Kimpton *et al.*, 1993) or 60 °C for 45 min (AmpF ℓ STR, user manual).

1.2.2.4.6. Stutter Bands

Stutter bands or shadow bands are defined as PCR peaks, which appear one repeat unit shorter than the main allele peak and the major mechanism for generation of these shadow bands is slipped strand mispairing during PCR (Hauge and Litt, 1993). A stutter peak is always lower in peak area than its associated allele (Gill *et al.*, 1997). In

the case of dinucleotide repeat loci, the main stutter band is two bases shorter than the main allele with additional stutter bands four and six bases shorter also visible. This makes interpretation of dinucleotide STRs complicated especially for DNA samples from a mixture of two or more persons or when two alleles from a single person are close in size (Hauge and Litt, 1993).

Results from tetranucleotide repeat loci are easier to interpret because one single stutter band of four bases shorter than each allele is observed (Walsh *et al.*, 1996). Stutter bands occur more frequently with multiplex PCR amplification of STR loci as a result of amplification reactions that are not optimal for all the constituent loci (Clayton *et al.*, 1998).

1.2.2.4.7. Validation of STRs for Paternity Testing

To validate the STR systems for the paternity testing, the American Association of Blood Banks Paternity Committee (AABB) conditioned certain criteria. These criteria of (AABB) standard are as follow: the DNA loci should exhibit Mendelian inheritance and have a low frequency of mutation (Dykes, 1990). The allele frequencies of any population must meet the expectation of the Hardy-Weinberg so they can be used to calculate the genotype frequencies and the probability of paternity. The loci must be unlinked for the product rule to be used to accumulate statistical information across the loci. It is better if the loci have low frequency in the population to achieve a high rate of power of exclusion. The results of the DNA typing have to be

the same from different tissues of the same individual. Data of the different studies showed that these criteria are present in STRs system (Hammond *et al.*, 1994).

In the case where the mother and the father of a deceased alleged father are available, autosomal STRs are useful, as the possible constructions of genotype of the alleged father can be known. It was approved that inclusion of Y-STR loci can increase the probability of paternity and also increase the chance of detecting non-paternity of a male offspring (Kayser *et al.*, 1998). Conversely, in cases where the father is deceased and only one male relative to the alleged father is available, autosomal STR markers are not usually conclusive therefore, Y-linked loci analysis should be included (Trabetti *et al.*, 1994).

1.2.3. Y-Chromosome

The Y-chromosome is present in males and it is the second smallest human chromosome. The average size of the human Y-chromosome is 60 Mb and it consists of a heterochromatic and euchromatic region. The heterochromatic region is more than half the length of the Y-chromosome and it is situated on the distal long arm of the chromosome. The euchromatic region is 28 Mb and contains the short arm, centromere and proximal long arm (Roewer, 1998). The Y-chromosome determines maleness by specifying the development of the testis during embryogenesis. Usually the gender identification of a given DNA sample is determined by amplification of the amelogenin gene (Mannucci *et al.*, 1995). Recently attention has focused on the largest non-recombining region in the Y human chromosome, which is uniparentally inherited. Because it is haploid and, therefore does not recombine the Y-chromosome is passed

down from father to son unchanged except by accumulation of mutations (Jobling *et al.*, 1997). This results in every Y-chromosome containing a record of the mutational events that happened on all previous ancestral Y-chromosomes (Roewer, 1998).

1.2.3.1. Y-Chromosome Polymorphisms

The human Y-chromosome contains many different types of polymorphisms including insertions/deletions, base substitution polymorphisms, major satellites, minisatellites and microsatellites (Roewer, 1998).

1.2.3.1.1. Indels Polymorphisms

Indels are insertions or deletions of the DNA at specific positions on the chromosome. Y-chromosome alu polymorphism (YAP) is one type of insertion, which has a useful role in population studies. Alu is a sequence of approximately 300 bp (Hammer, 1994).

1.2.3.1.2. Single Nucleotide Polymorphisms (SNPs)

SNPs are single nucleotide polymorphisms in which a certain nucleotide such as A is changed to another type of nucleotide such as G. Base substitutions on the Y chromosome are rare enough to be considered as unique events and so ancestral origin can be determined (Jobling and Tyler-Smith, 1995).

1.2.3.1.3. Minisatellites and Microsatellites

The first Y-linked restriction fragment length polymorphisms were identified in 1985 (Casanova *et al.*, 1985). There is one minisatellite, MSY1, which has a mutation rate of over 1% per generation (Jobline *et al.*, 1998). Other searches for further minisatellites have not been successful (Fretwell *et al.*, 1995).

In the 1990s, forensic identification was greatly affected by the development and establishment of STR amplification. Many autosomal STRs have been used in forensic studies with great success (Urquhart *et al.*, 1994). The development of a Y-chromosome-linked STRs system for male identification took a long time to develop. One hypothesis is that the Y-chromosome STR loci occur in a ratio of 1:4 to the autosomal STRs (Malaspina *et al.*, 1990).

In the first half of the decade only the DYS19 system had been tested for forensic application. Many worldwide populations studies for the DYS19 system revealed obvious differences in allele distribution in the major ethnic groups (Santos *et al.*, 1996). Similarities in frequency pattern of DYS19 locus were observed in geographically neighbouring populations. This observation reflects the inherent property of non-recombination of the Y chromosome (Jobling *et al.*, 1997). Now there are more than 14 Y-specific STR markers, which are simple and sensitive markers for male identification (Kayser *et al.*, 1997). These Y-chromosome STR loci are shown in table 1.2.

Table 1.2 Characteristics of Y-chromosomal STRs

The characters of the 14 Y-chromosomal STR loci, which include dinucleotide repeats e.g. DYS288 locus, trinucleotide repeat e.g. DYS388 locus, tetranucleotide repeat e.g. DYS19 locus and pentanucleotide repeat DXYS156Y locus.

Locus	Repeat motifs	Number of repeats	Number of known alleles	Size range
DYS19	(CTAC/C) _n	10-19	10	174-210
DYS288	(CA) _n	- ^c	2	119-121
DYS385 ^a	(GAAA) _n	9-22	49 ^b	360-412
DYS388	(ATA) _n	- ^c	5	126-138
DYS389I	(CTG/AT) _n	7-13	7	239-263
DYS389II	(CTG/AT) _n	23-31	9	353-385
DYS390	(CTG/AT) _n	18-27	10	191-227
DYS391	(CTAT) _n	8-13	6	140-160
DYS392	(ATT) _n	7-16	8	236-263
DYS393	(GATA) _n	9-15	6	108-132
YCA I ^a	(CA) _n	- ^c	3 ^b	124-132
YCA II ^a	(CA) _n	- ^c	28 ^b	147-165
YCA III ^a	(CA) _n	19-25	15 ^b	192-204
DXYS156Y ^d	(TAAAA) _n	- ^c	3	160-170

^aOne set primer amplifies two Y-chromosomal PCR products. ^bClasses of alleles are proposed for those loci where the two Y-specific PCR fragments cannot be unambiguously assigned to either of the two loci. ^cAlleles not yet completely sequenced. ^dPCR primers also amplify chromosome X-specific products.

1.2.3.2. *Population Distribution Haplotypes*

Due to lack of recombination, the alleles of different markers can be typed and haplotypes may be constructed and differences between them evaluated. Using reasonable assumptions about the rates of mutation, the date of the most recent common ancestor (MRCA) of any two or more Y-chromosomes can be estimated. Whitfield *et al.*, (1995) showed that data from sequencing 18,300 bases obtained from five ethnically diverse humans gave an MRCA time of between 37,000 and 49,000 years before present. In other words those five men had a common grandfather 40,000 to 50,000 years ago.

When a significant number of individuals in a population is analysed, the frequency of occurrence of different haplotypes can be used to differentiate between populations and to shed light on the substructures within a population. The distribution of the different Y-chromosome haplotypes between different populations shows certain characteristics (Underhill *et al.*, 1996).

Different types of Y-chromosome STRs have been used to construct highly informative Y-chromosome haplotypes (Cooper *et al.*, 1996 and Roewer *et al.*, 1996). The discrimination rate was between 74% and 90% in the local European population (Kayser *et al.*, 1997). The study of STRs haplotype variation on the Y-chromosome has also facilitated the research of paternal migration at the inter-continental level (Deka *et al.*, 1996 and Karafet *et al.*, 1999). The study of Y-chromosomes has also

demonstrated the demographic histories of a single population and the genetic affinities of closely related groups (Roewer *et al.*, 1996).

1.2.3.3. Forensic Applications

Y-chromosome STRs are the markers of choice in forensic studies due to their high degrees of diversity. Y-chromosome STRs can identify the male component in male/female cell mixtures in body fluids where differential lysis was unsuccessful because of very degraded samples or very low sperm counts (Roewer, 1998). Y-chromosome STRs can also detect the male epithelial cells in ejaculate of vasectomized individuals (Prinz *et al.*, 1997). In multiple rape cases, Y-chromosome STRs analysis is used to determine the number of semen contributors (Kayser *et al.*, 1997).

1.2.3.4. Paternity Applications

Because of the inheritance along male lineages, more than one difference at Y-chromosome STR loci will exclude paternity. Thus Y-linked polymorphic traits have a much higher exclusion power than analogous variable autosomal loci (Chakraborty, 1985). The occurrence of mutations must be considered when applying Y markers. The mutation can be excluded when more than one haplotype difference is detected between the haplotypes of the child and the alleged father (Roewer, 1998). The rate of mutation of Y-chromosome STRs is 0.2% (Heyer *et al.*, 1997).

When the result of an analysis shows an inclusion case, an identical haplotype will exist in almost all males related to the alleged father. On the other hand there is a

special condition where Y-chromosome polymorphisms are useful, namely in deficiency cases, when the alleged father is deceased and relatives such as an uncle can be tested in his place. Any male relative of the alleged father will produce complete information on his Y-chromosomal pattern (Pena and Chakraborty, 1994).

1.3. PATERNITY TESTING INTERPRETATION

For any DNA markers, the outcome of the paternity test will be either negative or positive. In the case of a negative result there is no match between the child and the alleged father, who can therefore be excluded as the father. At least two loci exclusions are required to minimise the chance of a mutation. In the case of a positive result (a match between the putative father and the child) paternity is not proven as another man may have the same genetic profile. Statistical calculation of the paternity index and the probability of paternity must be calculated (Kaye *et al.*, 1991).

1.3.1. Paternity Index (PI)

In any inclusion case, there are two possibilities. Either the alleged father is the biological father (H_1) or another unrelated man is the father (H_0). The DNA profiles for child, mother and alleged father are determined. The Paternity index (PI) is calculated as the ratio of (H_1) possibility over H_0 possibility ($PI = \frac{H_1}{H_0}$). The paternity index tells how many times more likely it is that the alleged father is the real father, compared to a random man being the real father. The calculation is based on the frequency of the

father's DNA profile in the general population. If the father is heterozygous for the DNA marker, this means that there is 0.5 chance that the man passed one of his alleles to the child and the paternity index is calculated by dividing 0.5 by the frequency of the allele passed from the father to the child.

If the father is homozygous for the DNA marker, he must pass this allele to the child and the paternity index is calculated by dividing 1 over the frequency of this allele in the population. The equations for the calculation of PI are shown in appendix 3.

1.3.2. Probability of Paternity

The probability of paternity (W) is calculated simply by converting the PI to a percentage. The calculation of probability of paternity is based on Bayesian rules where the prior probability is multiplied by the paternity index to calculate the posterior probability, which is the probability of paternity (Elston, 1986). The prior probability takes into account the non-genetic evidence. It is difficult to determine the prior probability, therefore the courts generally assume that the alleged father has 50% prior probability of being the biological father. Therefore a value of 0.5 is normally used (Kaye *et al.*, 1991). The probability of paternity is equal to the division of PI over the

PI plus one ($W = \frac{PI}{PI+1} \times 100$).

The paternity index and the probability of paternity represent a comparison of the alleged father against a random man. If there is a close relative who also could be the father he also should be tested. If he is unavailable, this possibility has to be taken into

account in the calculation of the paternity index and the probability of paternity (Lee *et al.*, 2000).

1.4. PATERNITY TESTING APPLICATIONS

In addition to the standard paternity testing, DNA analysis is used to solve many different paternity conditions.

1.4.1. Prenatal and Newborn Paternity Testing

In the past it was impossible to test babies by the conventional serological tests until after the first six months of postnatal life because of transfer of proteins from the mother to the foetus (Strom *et al.*, 1996). The introduction of DNA analysis has allowed prenatal paternity testing by analysis of the amniotic fluid cells. This has a forensic application in the occurrence of pregnancy after a rape (Nata *et al.*, 1993). Information of the test allows the mother to make her decision about termination. It is also useful for the child regarding the presence of heritable diseases (Strom *et al.*, 1996).

1.4.2. Deceased Alleged Father

This type of DNA testing is required to obtain insurance benefits and to investigate an inheritance dispute. There are two ways to solve these cases. Firstly a postmortem sample from the father's DNA is required. Secondly the genotype of the

father needs to be reconstructed based on the surviving family members. If the child is a boy Y-chromosome polymorphism is useful to solve such cases (Kayser *et al.*, 1998).

1.4.3. Double Paternity in Dizygous Twins

Dizygous twins can arise by superfecundation, in which two ova are fertilized in the same menstrual cycle (Verma *et al.*, 1992). If the woman has sexual intercourse with two men at the same menstrual cycle, the resulting twins could have different fathers. Paternity testing using DNA markers is able to resolve questions of this type of disputed paternity (Lu *et al.*, 1994).

1.4.4. Father-Daughter Incest

The DNA analysis is used in case of father-daughter incest. However the probability of exclusion of paternity is lower when the parents are close relatives of each other than when the parents are not relatives. To obtain a high exclusion capability, it is necessary to analyze several independent loci in which both the mother and the child have different genotypes (Cifuentes and Jorquera, 1997).

1.4.5. Determination of Sibship

The most difficult kinship analysis is to determine if two individuals are full siblings when those individuals are the only ones available for DNA testing. This type of analysis requires loci with high heterozygosity. Therefore highly polymorphic STR

analysis can be discriminative in most sibship determinations, and the sharing of two alleles per locus is informative in indicating sibship (Tzeng *et al.*, 2000).

1.5. ETHNIC GROUPEES OF THE EGYPTIAN POPULATION

The Egyptian population is composed of different ethnic groups including Copts, Arabs, Bedouins and Nubians. The Copts are the modern descendants of the ancient Egyptians. The term Copts describes members of a group of people; their language and religion are called Coptic. Many of these people converted to Christianity during the Roman period in Egypt. The Copts constitute less than 10% of the total Egyptian population.

Arab is a name used to describe people from the Arabian peninsula who invaded Egypt by force in 641 A.D. They migrated throughout the middle east and north Africa. Not all Arabs are Moslems as there are large numbers of Arab Christians. The Arabs represent the majority of the Egyptian population.

The Bedouins are nomadic Arab people who migrate through the Egyptian deserts to search for their livestock and water for their families. There are a number of Bedouin tribes that are divided into family units.

Nubians are the people who originally lived in the southern parts of Egypt in the area that is near to the Sudan. The ancient Egyptians had interaction with the Nubians

from the earliest historical period. The Nubians were assimilated into the Egyptian culture and many of them intermarried.

There are also people composed from all mixed ethnic groups and they called either fallahin or city dwellers. Fallahin are the people who inhabit the rural villages along the Nile while city dwellers are the people who live in large cities such as Cairo and Alexandria. Additionally, Romans, Greeks, more recently the Turks, French and English have settled and intermarried with the Egyptian people. Moreover, cultural customs encourage consanguineous marriage especially in rural areas adding further to the complex nature of the Egyptian population. Because of the diversity of the population, there is a possibility of deviation from the Hardy-Weinberg Equilibrium (HWE) in the Egyptian population.

1.6. AIM OF THE PROJECT

In Egypt the ABO blood group system was the only genetic marker used in paternity testing for a long time. Some additional markers are also used now including phosphoglucomutase (PGM) and haptoglobin (HP). All these markers have a low value of the power of exclusion of paternity.

In contrast DNA typing can reduce the number of possible fathers to the minimum. Mathematical estimation of the likelihood of paternity has found legal acceptance in most of the countries of the world. Therefore, the introduction of DNA use in Egypt should not be problematic.

DNA exclusions are easy to interpret; a non-match is a definitive proof that two samples have different origins. However, DNA inclusions cannot be interpreted without knowledge of how frequently a match might be expected to occur in the general population. Any DNA system or marker cannot be used with regard to inclusion until the population frequencies of the patterns have been established.

Therefore, before introducing any DNA marker as a scientific tool into a particular legal system it is important to investigate the structure of the population and to establish the DNA databases to be used as references. A standard way is to estimate the allele frequencies of the DNA markers by counting their occurrence in a random sample. Therefore, in order to estimate the power of a statistical match accurately, it is preferable to use a database from each local population where the DNA system is to be used. Therefore, initially defined objectives of this study were to:

1-Evaluate a protocol for preservation of different blood and buccal samples and use it in the collection of samples from Egypt.

2-Investigate the structure of the Egyptian population using simple STR loci.

3-Detect any significant difference between the DNA database of the Egyptian population and other populations.

4-Establish the Egyptian DNA database for different autosomal and Y-chromosome STR loci and to use them in paternity cases.

CHAPTER 2: MATERIALS AND METHODS

2.1. MATERIALS

2.1.1. Chemicals

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

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 by using the Laboratory Thermal Equipment Autoclave 225E. Small batches could be 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 prior to its use.

2.2.4. Solutions and Buffers

All solutions and buffers were made by using Milli-Q water (Millipore Corporation, Bedford, MA 01730, USA). Solutions were then autoclaved and stored at room temperature.

2.2.4.1. Buffers

10X TBE buffer 108 g Tris-base 45 g Boric Acid 9.3 g EDTA To 1 l by adding Milli-Q water	Sigma Sigma Sigma
--	-------------------------

Loading buffer 0.25% bromophenol blue 0.25% xylene cyanol FF 30% glycerol	BDH Sigma Sigma
--	-----------------------

2.2.4.2. Cloning Reagents

IPTG 24 mg IPTG 1 ml Milli-Q water	NBL
--	-----

Filtered sterilise.

X-Gal 50 mg X-Gal 1 ml N,N'dimethyl-formamide	NBL GIBCO BRL
---	------------------

Stored in a glass bottle/tube, covered with silver foil in freezer.

Miller's LB Agar-1 litre 37 g Miller's LB Agar To 1 l by adding Milli-Q water 1 ml Ampicillin (50 mg/ml)	GibcoBRL SIGMA
---	-----------------------

Autoclaved for 20 min, cooled to 50°C before ampicillin was added.

Miller's LB broth base-1 litre 25 g Miller's LB broth base To 1 l with Milli-Q water	GibcoBRL
--	----------

Autoclaved for 20 min and stored at room temperature.

2.2.5. Gels

2.2.5.1. 1% Agarose Gel

1 g agarose 100 ml 1 X TBE buffer 10 ul Ethidium bromide 20 mg	Sigma
--	-------

2.2.6. Molecular Weight Markers

1Kb DNA Ladder :1 ug/lane Hyperladder 1: 5 ul/lane Hyperladder IV: 5 ul/lane	GIBCO BRL Bioline (16 the Edge buisness center Humber Rd London NW2 6EW,UK) Bioline
--	--

2.2.7. Polymerase Chain Reaction

<p><i>Taq</i> polymerase DNA polymerase 250 units per vial, 5 U/ul</p> <p>10X PCR buffer MgCl₂ solution <u>dNTPs (deoxynucleosidetriphosphates)</u> 100 mM dGTP 100 mM dATP 100 mM dTTP 100 mM dCTP</p>	<p>Promega (Delta House, Chilworth Research Centre, Southampton, UK)</p> <p>Promega Promega Promega</p>
<p>Reddyload PCR mix (1.5mM MgCl₂)</p> <p><i>Taq</i> DNA polymerase 1.25 U Tris-HCl (pH 8.8 at 25 °C) 75 mM (NH₄)₂SO₄ 20 mM MgCl₂ 1.5 mM Tween 20, 0.01% (v/v) dATP, dCTP, dGTP and dTTP, each 0.2 mM</p>	<p>Advanced biotechnologies.Ltd., (Blenheim Road, Epsom, Surrey, UK).</p>

2.2.8. Sequenced Samples for the Ladders

10 sequenced samples were sent by Prof. Dupuy B.M. (Institute of Forensic Medicine, University of Oslo, Oslo, Norway) to construct the ladders for D11S554, SE33 and APOAI1, which were used for precious allele length measurement.

Sample no.	SE33	SE33	D11S554	D11S554	APOAI1	APOAI1
	Allele 1	Allele 2	Allele 1	Allele 2	Allele 1	Allele 2
	Length bp	Length bp	Length bp	Length bp	Length bp	Length bp
1	241 bp	303 bp	204 bp	212 bp	267 bp	283 bp
2	263 bp	333 bp	216 bp	224 bp	279 bp	293 bp
3	247 bp	283 bp	220 bp	254 bp	256 bp	287 bp
4	243 bp	251 bp	196 bp	208 bp	256 bp	283 bp
5	255 bp	259 bp	176 bp	216 bp	291 bp	291 bp
6	255 bp	287 bp	176 bp	228 bp	283 bp	295 bp
7	291 bp	295 bp	212 bp	229 bp	291 bp	291 bp
8	279 bp	295 bp	221 bp	233 bp	267 bp	267 bp
9	251 bp	255 bp	208 bp	250 bp	271 bp	271 bp
10	287 bp	299 bp	250 bp	250 bp	286 bp	286 bp

2.3. DNA SAMPLE COLLECTION

2.3.1. Source of DNA Samples

Whole blood samples were collected from El-Minia, a city that is located in the middle part of Egypt. Samples were collected from 300 unrelated individuals.

Buccal swabs were collected from 20 unrelated Egyptian nationals living in Glasgow. Also buccal swabs were collected from 30 Egyptian families living in El-Minia.

2.3.2. Blood Sampling

A qualified medical person collected whole blood samples from 300 unrelated individuals using venipuncture under sterile conditions. 3 ml of blood was collected in 5 ml EDTA tubes. Each tube was labelled with the date, person's name, age and sex. Blood samples were extracted in Egypt by using Puregene™ kit up to the cell lysis stage and stored at room temperature. Aliquots of blood samples without cell lysis were frozen at $-80\text{ }^{\circ}\text{C}$ and transported to Glasgow. On arrival blood samples were extracted directly or stored at $-20\text{ }^{\circ}\text{C}$ until required.

2.3.3. Buccal swabs

Buccal swabs were collected by twirling a sterile swab (C.E.P swab, Life technologies) on the inner cheek for 30 s. Then the swab was placed in cell lysis solution of Puregene™ DNA Isolation Kit (Gentra Systems, Inc. Minneapolis, MN 55447, USA) and stored at room temperature.

2.4. ISOLATION OF DNA FROM WHOLE BLOOD

2.4.1. Nucleic Acid Extraction using Puregene™ kit

DNA was extracted from whole blood by using Puregene™ DNA Isolation Kit. 300 µl of fresh whole blood was added to a 1.5 ml microcentrifuge tube (SARSTEAD, Germany), which contained 900 µl Red Blood Cell lysis solution. It was inverted to mix and incubated for 10 min at room temperature. It was inverted once again during the incubation and centrifuged (Micro centaur, Scotlab) for 2 min (13,000g). The supernatant was removed except for approximately 30 µl that was used to resuspend the white cell pellet by whirlimixing.

300 µl Cell Lysis Solution was then added to the tube containing the resuspended white cells and mixed by inverting the tube several times. If cell clumps were visible after mixing, the blood samples were incubated for 30 min at room temperature until the solution was homogeneous.

100 µl Protein Precipitation Solution was then added to the cell lysate and vortexed for 20 s and centrifuged at high speed for 3 min. The supernatant was transferred into clean 1.5 ml microcentrifuge tubes, which contained 300 µl 100% isopropanol and the precipitated protein pellets were discarded. The samples were inverted gently 50 times and centrifuged at high speed for 1 min. The whole supernatant was discarded and the DNA pellet washed with 150 µl of 70% EtOH and the tube was inverted several times. Samples were centrifuged at high speed for 2 min and the EtOH carefully pipetted off. The pellets were left to air dry for 15 min.

100 µl DNA Hydration Solution was added to each sample and DNA was allowed to rehydrate by mixing and incubating the sample in a water bath (Techne, UK) at 65 °C for 1 h. After incubation samples were left overnight at room temperature.

2.5. ISOLATION OF DNA FROM BUCCAL SWABS

2.5.1. Nucleic Acid Extraction using Puregene™ kit

DNA was extracted from buccal swabs by using Puregene™ DNA Isolation Kit

A swab was kept in a 1.5 ml microfuge tube containing 300 µl cell lysis solution. The samples were incubated at 65 °C for 1 h. The swab then was taken off from the tube and 1.5 µl of RNase A Solution was added to the cell lysate and mixed by inverting the tube 25 times. The tube was incubated at 37 °C for 15 min. The samples were allowed to cool to room temperature. 100 µl of Protein precipitation Solution was added to the cell lysate and vortexed at high speed for 20 s. The tubes were incubated on ice for 5 min and centrifuged at 13,000g for 3 min.

The supernatant containing the DNA was poured into a 1.5 ml centrifuged tube containing 300 µl isopropanol and 0.5 µl glycogen (20 mg/ml). The tube was inverted gently 50 times and incubated at room temperature for 10 min. The tube was centrifuged for 5 min. The supernatant was discarded and 300 µl of 70% EtOH was added to wash the DNA pellet by inverting the tube several times. DNA was centrifuged for 1 min and the EtOH was carefully poured off. The DNA pellet was left

to air dry for 15 min. 20 μ l of DNA Hydration Solution was added and the DNA was allowed to rehydrate overnight at room temperature.

2.6. QUALIFICATION AND QUANTIFICATION OF DNA

All DNA samples were tested for their quality and quantity.

2.6.1. Agarose Gel Electrophoresis

All agarose (SEAKEM[®] and I.D.NA[®]) was supplied by Flowgen FMC (FMC Corporation).

0.5 g of agarose (SeaKem 1% w/v) was added to 50 ml of 1X TBE (0.09 M Tris-borate, 2 mM EDTA). The agarose suspension was heated in a microwave oven until all the agarose had dissolved. The agarose solution was left to cool to approximately 60 °C. 2 μ l of EtBr (10 mg/ml) was added to the solution and mixed. The gel solution was poured into the gel tray, the comb was placed and the gel was left for 30 min. The gel was placed into the gel tank, which filled with sufficient 1X TBE running buffer. 5 μ l of the DNA samples were mixed with 2 μ l of loading buffer (0.25% bromophenol blue (w/v)). Molecular weight standard samples of known quantity and quality were run along side the unknown samples to provide a comparison. The gels were run at 80 to 100 V until the bromophenol blue had migrated two-thirds of the way down the gel. The DNA in the gels were visualised under UV light (transilluminator 312 nm UltraViolet, Specronics Corporation, Westbury, NY, USA). Gels were photographed

using a Polaroid camera and Polaroid Kodak film and the amount of DNA could be estimated by comparison with the molecular weight standard.

2.7. POLYMERASE CHAIN REACTION (PCR)

PCR is a vitro amplification of a specific target of DNA by using primers, which are specific to the flanking regions. All samples were amplified by using a Techne PCR machine (Techne Cambridge Ltd, Duxford Cambridge England, UK).

2.7.1. D1S80 Amplification

DNA was amplified from blood and buccal samples by using the D1S80 kit (Applied Biosystem; CA, USA).

2.7.1.1. PCR Protocol

20 µl of master mix (dNTPs, forward & reverse primers for D1S80, *Taq* polymorase enzyme), 10 µl of MgCl₂ (5 mM), 20 µl of DNA template and one drop of mineral oil (Sigma). The PCR reaction consisted of 95 °C for 5 min; 94 °C for 1 min, then 65 °C for 1 min, then 72 °C for 1 min, 29 cycles; and then incubation at 72 °C for 10 min.

2.7.2. Autosomal Short Tandem Repeats

All blood and buccal swabs from families' samples were amplified with 3 loci, D3S1358, vWA, FGA. All forward primers were labelled with the fluorescent dye

marker FAM (5-Carboxyfluoresce). The primers were made by (Oswel DNA Service, Medical and Biological Science Building, University of Southampton, Boldrewood, Bassett Crescent East, Southampton SO16 7PX).

Locus	Repeat	Location	Primer Sequence
D3S1358	AGAT	3p	Forward 5' ACTGCAGTCCAATCTGGG Reverse 5' ATGAAATCAACAGAGGCTTG
HumvWA	TCTA	12p12 - p12.2	Forward 5' CCCTAGTGGATGATAAGAATAAT Reverse 5' GGACAGATGATAAATACATAGGAT
HumFGA	TCTT	4q28	Forward 5' CCATAGGTTTTGAACTCACAG Reverse 5' CTTCTCAGATCCTCTGACACTC

2.7.2.1. PCR Protocol

D3S1358 and FGA were amplified as a duplex: template DNA, 0.625U *Taq* polymerase (Promega), 0.2 μ M each primer of D3S1358 and 0.1 μ M each primer of FGA, 200 μ M each nucleotide (Promega), 1X buffer and 2.5 μ M MgCl₂ (Promega). The mixture was diluted to a final volume of 25 μ l and overlaid with mineral oil (Sigma).

HumvWA was amplified as a singleplex: template DNA, 0.625U *Taq* polymerase (Promega), 0.2 μ M each primer, 200 μ M each nucleotide (Promega), 1X buffer and 2 μ M MgCl₂ (Promega). Autoclaved water was added to a final volume of 25 μ l and overlaid with mineral oil (Sigma).

The PCR amplification parameters were as follow, denaturation at 95 °C for 5 min, then 30 cycles of denaturation at 94 °C for 1 min, annealing at 58 °C for 1 min for

the duplex PCR and at 60 °C for 1 min for the vWA locus and extension at 72 °C for 1 min, followed by incubation at 72 °C for 10 min

2.7.3. Amplification of Hypervariable Loci

All blood and buccal swabs from families' samples were amplified with the three hypervariable loci, D11S554, SE33 and APOAI1 as a multiplex. The forward primers for D11S554 and APOAI1 were labelled with the fluorescent dye marker JOE and the forward primer for SE33 was labelled with the fluorescent dye marker FAM (5-Carboxyfluoresce). The primers were made by (Oswel DNA Service, Medical and Biological Science Building, University of Southampton, Boldrewood, Bassett Crescent East, Southampton SO16 7PX).

Locus	Repeat	Location	Primer Sequence
D11S554	AAAG	11P11.2	Forward 5'GGTAGCAGAGCAAGACTGTC Reverse 5'CACCTTCATCCTAAGGCAGC
SE33	AAAG	6	Forward 5'AATCTGGGCGACAAGAGTGA Reverse 5'ACATCTCCCCTACCGCTATA
APOAI1	AAAG	11q23qter	Forward 5'GAGACGGAGGTTGCAGTGA Reverse 5'GTTGAGACTGCATTCTCCC

2.7.3.1. PCR Protocol

Template DNA, 0.15 µM of each primer for D11S554, 0.25 µM of each primer for SE33, 0.3 µM each of primer for APOAI1 and 22.5 µl of ReddyLoad PCR Mix (Advanced Biotechnologies Ltd, Blenheim Road, Surry UK) containing 1.25 U of *Taq* DNA Polymerase, 75 mM Tris-HCl (pH 8.8 at 25 °C), 20 mM (NH₄)₂ SO₄, 1.5 mM

MgCl₂, 0.01% (v/v) Tween 20, 200 μM of each nucleotide and red dye for electrophoresis.

The PCR amplification condition was as follow, denaturation at 95 °C for 5 min, then 29 cycles of denaturation at 94 °C for 1 min, annealing at 60 °C for 1 min and extension at 72 °C for 1 min, followed by 72 °C for 30 min.

2.7.4. Amplification of Y Chromosome STRs

All male blood samples were amplified with the Y chromosome STR DYS19, DYS389 and DYS390 loci as a multiplex I and DYS391, DYS392 and DYS393 as a multiplex II. In multiplex I the forward primers for DYS390 and DYS389 were labelled with the fluorescent dye marker FAM (5- Carboxyfluoresce) and the forward primer for DYS19 was labelled with the fluorescent dye marker JOE. In multiplex II the forward primers for DYS391 and DYS393 were labelled with the fluorescent dye marker FAM (5- Carboxyfluoresce) and the forward primer for DYS392 was labelled with the fluorescent dye marker JOE. The primers were made by (Oswel DNA Service, Medical and Biological Science Building, University of Southampton, Boldrewood, Bassett Crescent East, Southampton SO16 7PX).

Locus	Repeat	Primer Sequence
DYS19	GATA	Forward 5' CTA CTGAGTTTCTGTTATAGT Reverse 5' ATGGCATGTAGTGAGGACA
DYS389I/II	GATA GACA	Forward 5' CCAACTCTCATCTGTATTATCTAT Reverse 5' TCTTATCTCCACCCACCAGA
DYS390	GATA GACA	Forward 5' TATATTTTACACATTTTGGGCC Reverse 5' TGACAGTAAAATGAACACATTGC

Locus	Repeat	Primer Sequence
DYS391	CTGT	Forward 5' CTATTCATTCAATCATACACCCA Reverse 5' CTGGAATAAAAATCTCCCTGGTTGCAAG
DYS392	ATT	Forward 5' TCATTAATCTAGCTTTTAAAAACAA Reverse 5' AGACCCAGTTGATGCAATGT
DYS393	GATA	Forward 5' GTGGTCTTCTACTTGTGTCAATAC Reverse 5' AACTCAAGTCCAAAAAATGAGG

2.7.4.1. PCR Protocol

For multiplex I: template DNA, 0.25 μM of each primer for DYS19, 0.25 μM each of primer for DYS389, 0.12 μM of each primer for DYS390 and 22.5 μl of ReddyLoading PCR Mix (Advanced Biotechnologies Ltd, Blenheim Road, Surry UK). It contains 1.25 U of *Taq* DNA Polymerase, 75 mM Tris-HCl (pH 8.8 at 25 °C), 20 mM $(\text{NH}_4)_2\text{SO}_4$, 2.5 mM MgCl_2 , 0.01% (v/v) Tween 20, 200 μM of each nucleotide and red dye for electrophoresis.

For multiplex II: template DNA, 0.2 μM of each primer for DYS391, 0.2 μM each of primer for DYS393, 0.25 μM of each primer for DYS392 and 22.5 μl of ReddyLoading PCR Mix (Advanced Biotechnologies Ltd, Blenheim Road, Surry UK). It contains 1.25 U of *Taq* DNA Polymerase, 75 mM Tris-HCl (pH 8.8 at 25 °C), 20

mM $(\text{NH}_4)_2\text{SO}_4$, 1.5 mM MgCl_2 , 0.01% (v/v) Tween 20, 200 μM of each nucleotide and red dye for electrophoresis.

The PCR amplification condition was the same for both multiplexes. It is as follow, denaturation at 95 °C for 3 min, then 30 cycles of denaturation at 94 °C for 1 min, annealing at 56 °C for 1 min and extension at 72 °C for 1 min, followed by 65 °C for 10 min.

2.8. CONSTRUCTION OF LADDERS

2.8.1. Construction of Ladders for Autosomal Simple STRs

The samples, which collectively represented all of alleles present, were amplified as a singleplex for each locus as described. The same samples were amplified using the commercial kit (AmpF ℓ STR blue kit) (Applied Biosystem; CA, USA) to calibrate the ladders and designate the alleles. The samples were then mixed together to construct the ladder for each locus.

2.8.2. Construction of Ladders for Hypervariable Autosomal STRs

Construction of ladders for the highly hypervariable loci was performed by cloning of sequenced samples.

2.8.2.1. Cloning of Sequenced DNA Samples

2.8.2.1.1. DNA Purification from Gel Bands

Purification of samples from gel band was performed by using GFX™ PCR DNA and Gel Band Purification Kit (Pharmacia Biotech, Milton Keynes, UK).

Amplification of sequenced samples was performed for D11S554, SE33 and APOI1 as a singleplex. Then the samples were run on 2% agarose to separate the DNA of interest from any other contaminating DNA. The DNA fragments required were identified by illumination of the gel UV light and excised from the gel with a clean scalpel blade. The gel slice was placed in a pre-weighed 1.5 ml microcentrifuge tube. The tube was re-weighed to calculate the weight of agarose slice. 10 µl of Capture Buffer for each 10 mg of gel slice was added (maximum column capacity is 300 µl of Capture Buffer added to a 300 mg gel slice). Mixing by vortexing vigorously and the tubes were incubated at 60 °C until the agarose was completely dissolved. After that centrifugation was performed and the sample was loaded onto a GFX Column resting in a 2 ml Collection Tube and incubated at room temperature for 1 min. Centrifugation at full speed for 30 s, then discardation of the flow through by emptying the collecting tubes. 500 µl of Wash Buffer were added to the column followed by centrifugation for 30 s. Then discardation of the Collection Tube and the GFX Column was transferred to a clean 1.5 ml microcentrifuge tube. 50 µl of elution buffer (10 mM Tris-HCl pH 8.0, TE pH 8.0) was added directly to the centre of the GFX Column and incubated at room temperature for 1 min. Centrifugation at full speed for 1 min was performed to

recover the purified DNA. 5 µl of DNA was run on 1% agarose to evaluate the recovered DNA.

2.8.2.1.2. Ligation of DNA

Ligation of purified DNA was performed using PGMT Vector Promega (Delta House, Chilworth Research Centre, Southampton, UK).

Ligation with the PCR vector required a molar ratio of 1:1 to 1:3 of vector:PCR product. To a 0.5 ml sterile microcentrifuge tube 2X ligation buffer (5 µl), PGMT vector (1 µl), PCR product (3 µl) and T4 DNA ligase enzyme (1 µl) were added. The contents were mixed and ligation allowed to be completed at 4 °C overnight. Routine positive and negative control ligation reactions were incorporated into every ligation experiment.

2.8.2.1.3. Making Electrocompetent Cells

4 ml of Luria Broth with one colony of DH5α were incubated at 37 °C shaking incubator over night. Then 1 ml from this mixture was added to 200 ml of Luria Broth and the mixture was incubated in shaking incubator for 2 h. The culture poured into chilled 50 ml centrifuge tube and the cells were harvest at 3000g for 20 min at 4 °C. Pour of the supernatant and re-suspend the pellets in 50 ml of 10% cold glycerol and spin for 20 min. The last step was repeated again then, the supernatant was poured off with re-suspension of the pellets in a few drops of excess fluid. The cells were aliquated into 50 µl and stored at -80 °C until needed.

2.8.2.1.4. Transformation

The microcentrifuge tube containing the ligation reaction was centrifuged briefly. To a chilled tube, 50 μ l of electrocompetent cells that had been defrosted on ice were added to 1 μ l of ligation reaction and to 1 μ l of pUC19 plasmid. The mixture was kept on ice for 10 min then transferred to a chilled electroporation cuvette, which pulsed at 2.5 KV. 450 μ l of ice-cold SOC was added immediately with gentle mixing. The cuvette was then placed in a 37 °C shaking incubator (220 rpm) for 1 h. 100 μ l then spread onto Amp-X-Gal-IPTG treated agar plates and incubated at 37 °C for 16 h. Transformation with the pUC19 was done to estimate the efficiency of competent cells.

2.8.2.1.5. PCR Screen for Recombinants

Using 'colony PCR', transformants were screened for the presence or absence of inserts. The PCR mixture consisted of 22.5 μ l of ReadyLoad PCR Mix, 0.5 μ l of each primer for each locus of D11S554, SE33 and APOI1 as a singleplex was dispensed into PCR tubes. Using a sterile pipette tip a white colony was picked and streaked onto a labelled gridded LB-Amp agar plate. The tip was then placed into the corresponding labelled PCR mix and shaken vigorously. The tip was discarded and a fresh one used for each subsequent colony. Then the gridded plate was inverted and incubated at 37 °C overnight. The PCR performed on the tubes using profile denaturation at 95 °C for 5 min, then 29 cycles of denaturation at 94 °C for 1 min, annealing at 60 °C for 1 min, extension at 72 °C for 1 min, followed by 72 °C for 10 min. 5 μ l of the PCR product was electrophoresed through a 2% agarose gel containing EtBr (10 mg/ml) to

allow visualisation by UV light. Alongside the PCR product a hyperladder 1 or IV (Bioline) was run.

2.8.2.1.6. *Plasmid Miniprep*

Plasmid minipreps were performed by using GFX™ Micro Plasmid Prep Kit (Pharmacia Biotech, Milton Keynes, UK).

One single colony was picked from the corresponding square on the grid plate for each clone of interest, and was cultured overnight at 37 °C in LB-Amp broth (4 ml) in a shaking incubator. 1.5 ml of the culture was transferred to a microcentrifuge tube and centrifuged for 30 s and the supernatant poured off. Another 1.5 ml of the culture was added to the tube and centrifuged for another 30 s and the supernatant was poured off. The remaining pelleted bacterial cells were resuspended in 300 µl of solution I then 300 µl of solution II was added with mixing by inverting the tubes 10-15 times. After that 600 µl of solution III was added and mixed by inverting the tube until a flocculent precipitate was evenly dispersed. Centrifugation for 5 min to pellet cell debris and the entire supernatant was transferred to a fresh-labelled microcentrifuge tube. Half of the supernatant was loaded onto a GFX Column resting in a 2 ml collection tube and incubated at room temperature for 1 min and then was centrifuged at full speed for 30 s, then discarding of the flow through by emptying the collecting tubes. The same step was repeated for the other half of the supernatant. 400 µl of Wash Buffer was added to the column and centrifugation for 30 s. Then discarding of the collection tube and the GFX Column was transferred to a clean 1.5 ml-microcentrifuge tube. 50 µl of elution

buffer (10 mM Tris-HCl pH 8.0, TE pH 8.0) was added directly to the centre of the GFX Column and incubated at room temperature for 1 min. Centrifugation at full speed for 1 min was performed to recover the plasmid DNA. 5 µl of the plasmid run on 1% of agarose gel containing EtBr (10 mg/ml) to allow visualisation by UV light.

2.8.2.1.7. *Enzymatic Restriction Digestion*

DNA cloned into a known vector can be released using restriction enzymes. The PGMT vector contained sequence recognised by *Sac* II and *Spe* I restriction enzymes. To digest out the insert from the PGMT vector, 7 µl from each sample, 1 µl from *Sac* II and 1 µl from *Spe* I restriction enzymes and 1 µl from buffer were added together. The tubes were incubated at 37 °C for 2 h and then at 4 °C overnight. To the digestion mix, 2 µl of loading buffer was added and the products resolved on 1% agarose containing EtBr (10 mg/ml). Alongside the restriction digests a Hyperladder VI was run.

2.8.2.1.8. *Amplification of Plasmids*

The recovered plasmids that contained the insert were diluted and amplified for D1S554, SE33 and APOA1 as a singleplex. The PCR products were mixed together to construct a specific ladder for each locus.

2.9. ANALYSIS OF PCR PRODUCT

2.9.1. Non-denaturing Polyacrylamide Gel Electrophoresis

This method was used to detect D1S80 locus.

2.9.1.1. Treatment of the Plates.

The front and back plates were washed with 5% decon and then with 100% ethanol. In the fume cupboard 5ml of Dimethyldichlorosilane was poured over the back plate and left to dry for 30 min. Then the plate was washed again with 100% ethanol. 5ml of ethanol, 1.25ml 10% glacial acetic acid and 14 μ l of Silane bind was mixed and applied to the back surface of the front plate by using blue roll. The solution was allowed to dry for at least 5min, then the plate was washed with 100% ethanol. Spacers were placed on the back plate and the short front plate was placed on the top of the spacers. The plates were clamped together by using bulldog clips.

2.9.1.2. Polyacrylamide Electrophoresis

100ml of Acrylamide-Bisacrylamide (19:1) solution was made. 700 μ l 10% ammonium persulphate and 15 μ l of TEMED (Life technologies) were added. Gently swirling mixed the solution. The gel was poured between the treated plates with insertion of a comb and allowed to polymerise for 1 h at room temperature. The appropriate amount of 1X TBE was poured into the upper and lower chambers of the apparatus and the comb was removed. Each well was loaded with a mixture of PCR product (5 μ l) and 6X loading buffer (2 μ l). After electrophoresis was run at 100 V for 3-4 h. The gel was stained for 30 min in 1X TBE buffer containing ethidium bromide (1mg/ml). The gel was washed in dH₂O. The gel was visualised under UV light and photographed using a Polaroid camera.

2.9.2. Agarose Gel Electrophoresis

This method was used to detect PCR products for all amplified STR loci before running them in the 373XL automated sequencer. The method was the same as described in 2.6.1.

2.9.3. Denaturing Polyacrylamide Gel Electrophoresis

Amplified sample product (1 μ l) was added to 2.5 μ l of loading buffer (formamide, dextran blue, GeneScan ROX350 internal standard in case of autosomal STR and ROX500 for highly hypervariable loci and Y-chromosomal STRS). Samples were heated to 95 °C for 4 min and cooled on ice prior to loading on to a 6% polyacrylamide gel. Electrophoresis was carried out on an automated DNA sequencer ABI 373XL (Applied Biosystem; CA, USA). Amplified products were sized automatically using GeneScan analysis 2.1 software.

2.10. DIRECT SEQUENCING OF PCR PRODUCTS

The PCR products for highly hypervariable loci and Y-chromosomal STR were sequenced using the ABI 377 automated sequencer. The ABI PRISM™ BigDye Terminator Cycle Sequencing Reaction Kit (Applied Biosystem; CA, USA) was used which is suitable for performing fluorescence-based cycle sequencing reactions on single or double-stranded DNA templates. First the PCR was performed for each locus according to its specific PCR conditions described previously then run on 1% agarose gel. The PCR products were cut from the gel and purified using GFX™ PCR DNA and

Gel Band Purification Kit (Pharmacia Biotech). Some homozygous samples and Y-chromosomal samples were purified directly from the solution using the same kit. The purified DNA run on 2% agarose gel to approximately quantify the amount of purified DNA. The purified DNA was used for the sequence with the use of the reverse primer for each locus. The gel used was 6% denaturing polyacrylamide gel.

CHAPTER 3: SAMPLE COLLECTION

3.1. INTRODUCTION

Paternity testing and personal identification are the two main areas of forensic DNA analysis. Since 1985, DNA typing of biological materials has become one of the most powerful tools for personal identification due to the high statistical power of matches. It also became important in determining kinship. From the medico-legal point and as a part of the validation process, it is important to examine the population for substructure before introducing DNA markers as a scientific tool into a particular legal system as the allele frequencies vary between the substructure populations (Sajantila *et al.*, 1994). Therefore, in order to accurately estimate the power of a statistical match, it is preferable to use a database from each local population where the DNA system is to be used (Lewontin and Hatl, 1991). At the beginning of this project there was no study into the structure of the Egyptian population or any published data for the Egyptian population or other northern African populations for any STR system; the majority of work had been carried out on populations of European origin.

In order to apply DNA methods to paternity testing in Egypt, investigation of the structure and a suitable database were desirable. Therefore, before the system could be introduced, relevant samples had to be selected.

3.2. IDENTIFICATION OF A RELEVANT POPULATION

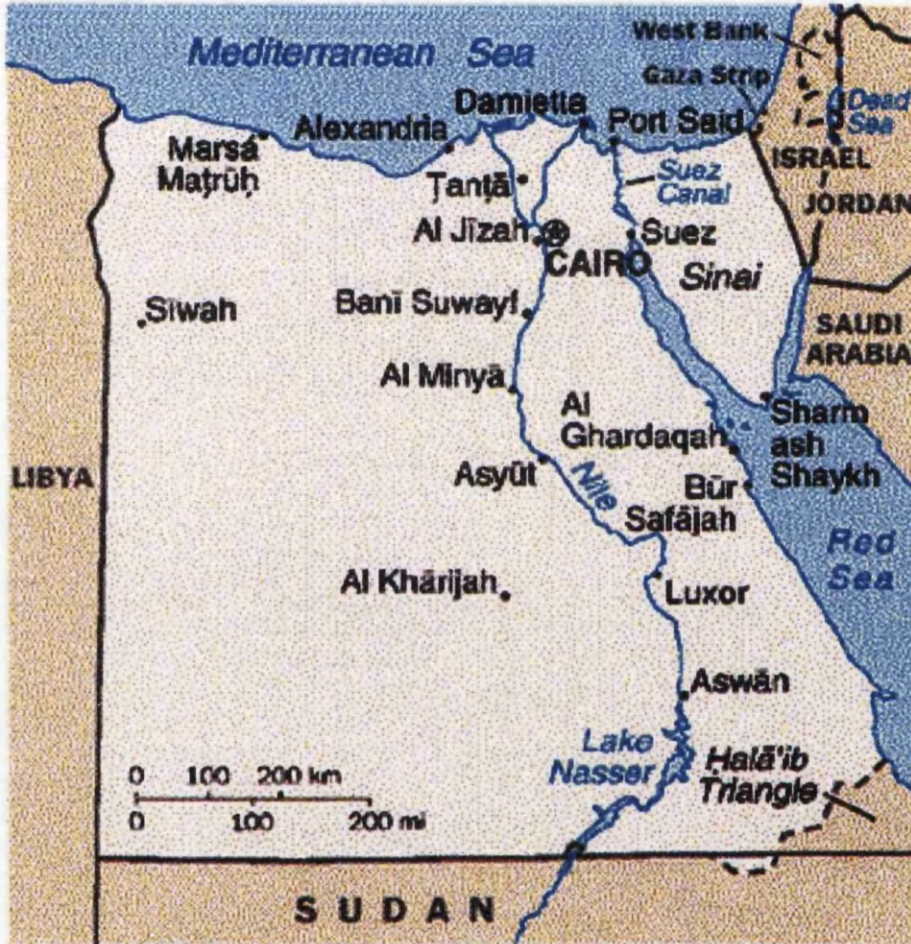
According to the recommendation of the National Research Council, the samples that are used to build up the reference data for any DNA marker should be chosen randomly. However, it is accepted that in practise this may be very difficult and other alternatives are acceptable (NRC, 1996). Regarding the number of samples, 200 unrelated individuals are commonly accepted to be sufficient for each DNA system to assess allele frequencies (Brinkmann, 1998). There were two possible Egyptian populations from whom samples could be collected. Firstly from Egyptian nationals living in Glasgow and secondly from Egyptians within Egypt. It was not possible to collect the required number of random samples from Egyptians living in Glasgow. Therefore, samples had to be collected from Egypt in order to make the population as random as possible and also in order to get a sufficient number of samples. El-Minia City, which is located in the middle part of Egypt (Figure 3.1) was chosen mainly due to the availability of facilities to collect the required number of random samples of the relevant population.

Collection of random samples from Egypt did present a potential problem, as DNA preservation would be problematic due to the high temperatures. Therefore, before collecting samples a protocol was developed to maximise the preservation of the samples.

58A

Figure 3.1 Map of Egypt

Map of Egypt shows the location of El-Minia city from which the blood and buccal samples were collected.



3.3. ASSESSMENT OF METHODS FOR PRESERVING SAMPLES

Different types of DNA samples have been used in genetic analysis, including finger stick blood samples, hair roots, urine samples, cheek scrapings and oral saline rinses (Richards *et al.*, 1993). Blood is one of the richest sources of DNA. On the other hand buccal cells are easy to collect especially from children. Thus blood and buccal samples were chosen to be collected as sources of DNA in this study.

3.3.1. Blood and Buccal Samples

Tests were performed to establish the best technique to preserve the blood and the buccal samples. Blood Samples were collected from some Egyptians volunteers living in Glasgow. Fresh whole blood (300 μ l) was added to 1.5 ml microcentrifuge tubes, which contained 900 μ l RBC lysis solution from the Puregene™ DNA extraction kit, inverted to mix and incubated for 10 min at room temperature then centrifuged for 2 min. The supernatant was removed except for 30 μ l. Then 300 μ l of Cell Lysis Solution was added to each tube and a number of tubes were stored at either -80 °C, 4 °C, room temperature or at 37 °C for one month. The cell lysis solution caused lysis of the white blood cells as it contains an anionic detergent, which solubilizes the cellular components (Puregene™ kit protocol). The remained of the blood was stored directly at -80 °C for the same period of time.

The buccal swabs, which were collected from Egyptian peoples living in Glasgow, were subjected to the same experimental work. Half of the buccal samples were stored in cell lysis solution of Puregene™ kit at either -80 °C, 4 °C, room

temperature or at 37 °C for one month. The other half of each sample was stored at the same temperatures without cell lysis solution.

After the above treatment the extraction was completed for all blood and buccal samples using Puregene™ DNA Isolation Kit as described in the “Material and Methods” section.

3.3.2. Quantity and Quality of Extracted DNA

The quantification of DNA is important for improving the quality of results when performing DNA typing especially for the PCR. Determination of the initial DNA template quantities for PCR can reduce unnecessary consumption of reagents and the biological samples (Budowle *et al.*, 1995).

DNA extracted by using Puregene™ kit is double stranded DNA so agarose gel electrophoresis was suitable to monitor the quantity and quality of DNA. The quantity and quality of the extracted DNA were visualised by comparing the isolated DNA with the commercial lambda DNA of known concentration ranging from 31 ng/6 µl to 500 ng/6 µl (GibcoBRL).

All DNA extracted from blood samples stored in cell lysis solution of Puregene™ kit for one month at different temperatures yielded high molecular weight DNA without significant degradation. The concentration of all blood samples preserved in cell lysis solution was more than 100 ng/µl. There was very little difference between the quantities of extracted DNA stored at different temperatures

(Table 3.1 and Figure 3.2). Blood samples stored at -80°C without cell lysis solution of Puregene™ kit for the same period of time gave lower quantity of DNA with marked degradation (Figure 3.3).

Buccal samples stored in cell lysis solution of Puregene™ kit gave a high yield of DNA without significant degradation whatever temperature was used for the incubation. Slightly less DNA was obtained from samples stored at 37°C (Figure 3.4).

Buccal swabs stored without cell lysis solution yielded a low quantity of DNA with a high level of degradation (Figure 3.5). Based on these findings preservation of both blood and buccal samples in cell lysis solution of Puregene™ kit was identified as a suitable method of preservation for the collection of samples for this study.

3.3.3. Amplification of Preserved Blood Samples

Although DNA yield is one essential measure of the effectiveness of a particular DNA extraction method, the suitability of the DNA for amplification is the most important measure for the extraction where PCR based methods are being used to analyse the DNA.

Table 3.1 The quantity of DNA obtained after storage at different temperatures in cell lysis solution of Puregene™ kit and without it.

The concentration of all blood samples preserved in cell lysis solution at different temperatures was more than 100 ng/μl. Blood samples stored at -80 °C without cell lysis solution of Puregene™ kit for the same period of time gave lower quantity of DNA with marked degradation. (+ referred to the quantity of extracted DNA). (+ very low, ++ little, +++ average and ++++ high).

Temperature	Sample stored in cell lysis solution		Sample stored without cell lysis solution	
	Blood	Buccal	Blood	Buccal
-80 °C	++++	+++	++ with degradation	+ with degradation
4 °C	++++	+++		+ with degradation
Room temp.	++++	+++		+ with degradation
37 °C	+++	++		+ with degradation

Figure 3.2 Quantity and quality of extracted DNA from blood samples stored in cell lysis solution.

4 White Blood Cells samples were stored in cell lysis solution of Puregene™ kit for one month. The samples were stored at either $-80\text{ }^{\circ}\text{C}$, $4\text{ }^{\circ}\text{C}$, room temperature or $37\text{ }^{\circ}\text{C}$. After incubation, DNA extractions were completed and $5\text{ }\mu\text{l}$ representing 5% of the total extraction were analysed using 1% agarose gel containing ethidium bromide.

Lane 1	500 ng of λ marker
Lane 2	5 μl DNA extracted from WBC of sample 1 stored at $-80\text{ }^{\circ}\text{C}$.
Lane 3	5 μl DNA extracted from WBC of sample 2 stored at $-80\text{ }^{\circ}\text{C}$.
Lane 4	5 μl DNA extracted from WBC of sample 3 stored at $-80\text{ }^{\circ}\text{C}$.
Lane 5	5 μl DNA extracted from WBC of sample 4 stored at $-80\text{ }^{\circ}\text{C}$.
Lane 6	5 μl DNA extracted from WBC of sample 1 stored at $4\text{ }^{\circ}\text{C}$.
Lane 7	5 μl DNA extracted from WBC of sample 2 stored at $4\text{ }^{\circ}\text{C}$.
Lane 8	5 μl DNA extracted from WBC of sample 3 stored at $4\text{ }^{\circ}\text{C}$.
Lane 9	5 μl DNA extracted from WBC of sample 4 stored at $4\text{ }^{\circ}\text{C}$.
Lane 10	5 μl DNA extracted from WBC of sample 1 stored at RT.
Lane 11	5 μl DNA extracted from WBC of sample 2 stored at RT.
Lane 12	5 μl DNA extracted from WBC of sample 3 stored at RT.
Lane 13	5 μl DNA extracted from WBC of sample 4 stored at RT.
Lane 14	5 μl DNA extracted from WBC of sample 1 stored at $37\text{ }^{\circ}\text{C}$.
Lane 15	5 μl DNA extracted from WBC of sample 2 stored at $37\text{ }^{\circ}\text{C}$.
Lane 16	5 μl DNA extracted from WBC of sample 3 stored at $37\text{ }^{\circ}\text{C}$.
Lane 17	5 μl DNA extracted from WBC of sample 4 stored at $37\text{ }^{\circ}\text{C}$.

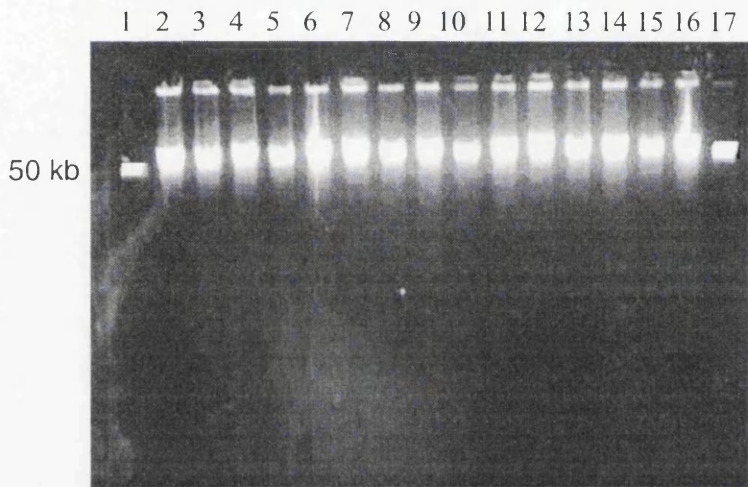


Figure 3.3 Quantity and quality of extracted DNA from blood stored at -80°C .

7 blood samples were stored at -80°C for one month without cell lysis solution of Puregene™ kit. After one month, DNA extractions were completed and 5 μl representing 5% of the total extraction were analysed using 1% agarose gel containing ethidium bromide.

- Lane 1 500 ng λ marker
- Lane 2 5 μl of DNA extracted from blood of sample 1 stored at -80°C .
- Lane 3 5 μl of DNA extracted from blood of sample 2 stored at -80°C .
- Lane 4 5 μl of DNA extracted from blood of sample 3 stored at -80°C .
- Lane 5 5 μl of DNA extracted from blood of sample 4 stored at -80°C .
- Lane 6 5 μl of DNA extracted from blood of sample 5 stored at -80°C .
- Lane 7 5 μl of DNA extracted from blood of sample 6 stored at -80°C .
- Lane 8 5 μl of DNA extracted from blood of sample 7 stored at -80°C .

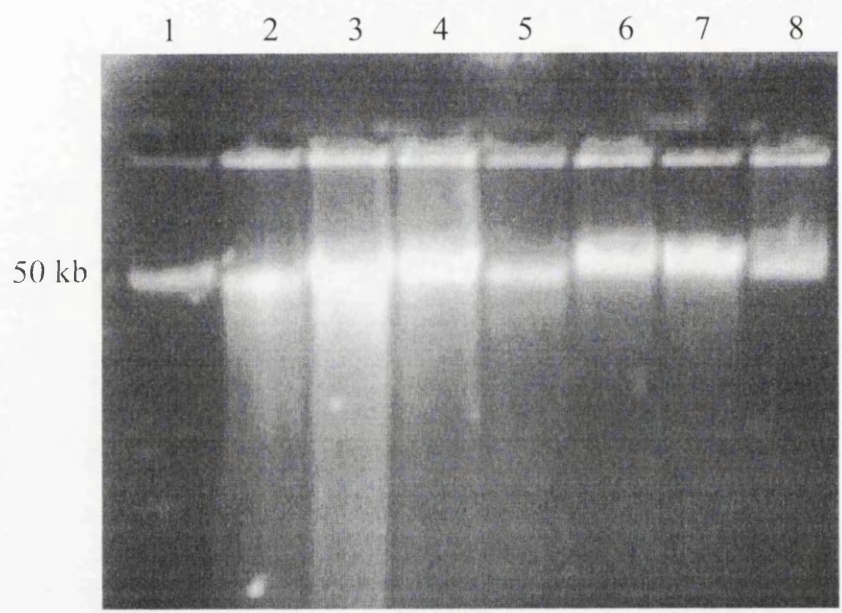


Figure 3.4 Quantity and quality of extracted DNA from buccal samples.

Buccal cells samples were stored in cell lysis solution of Puregene™ kit for one month. The samples were stored at either $-80\text{ }^{\circ}\text{C}$, $4\text{ }^{\circ}\text{C}$, room temperature or $37\text{ }^{\circ}\text{C}$. After incubation, extractions were completed and $5\text{ }\mu\text{l}$ representing 25% of the total extraction were analysed using 1% agarose gel containing ethidium bromide.

Lane 1 500 ng λ marker.

Lane 2 5 μl of DNA extracted from buccal cell stored in cell lysis solution at $-80\text{ }^{\circ}\text{C}$.

Lane 3 5 μl of DNA extracted from buccal cell stored in cell lysis solution at $-80\text{ }^{\circ}\text{C}$.

Lane 4 5 μl of DNA extracted from buccal cell stored in cell lysis solution at $-80\text{ }^{\circ}\text{C}$.

Lane 5 5 μl of DNA extracted from buccal cell stored in cell lysis solution at $4\text{ }^{\circ}\text{C}$.

Lane 6 5 μl of DNA extracted from buccal cell stored in cell lysis solution at $4\text{ }^{\circ}\text{C}$.

Lane 7 5 μl of DNA extracted from buccal cell stored in cell lysis solution at $4\text{ }^{\circ}\text{C}$.

Lane 8 5 μl of DNA extracted from buccal cell stored in cell lysis solution at RT.

Lane 9 5 μl of DNA extracted from buccal cell stored in cell lysis solution at RT.

Lane 10 5 μl of DNA extracted from buccal cell stored in cell lysis solution at RT.

Lane 11 5 μl of DNA extracted from buccal cell stored in cell lysis solution at $37\text{ }^{\circ}\text{C}$.

Lane 12 5 μl of DNA extracted from buccal cell stored in cell lysis solution at $37\text{ }^{\circ}\text{C}$.

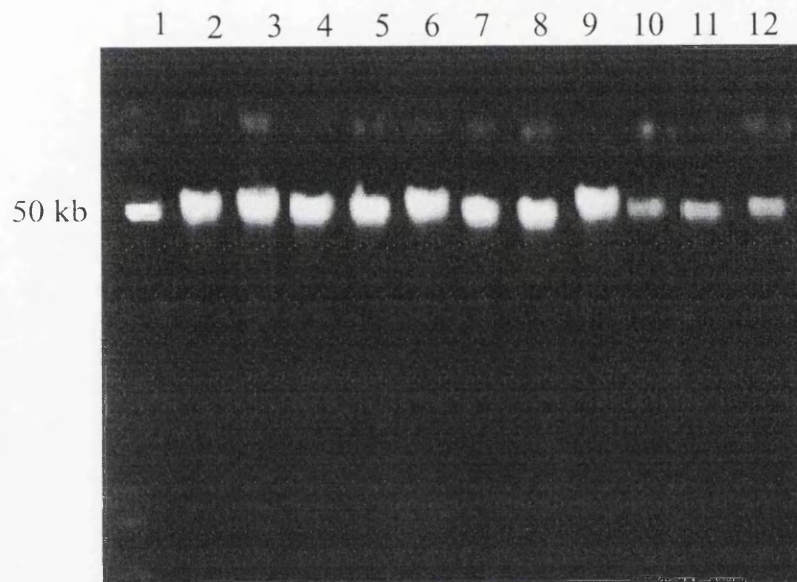
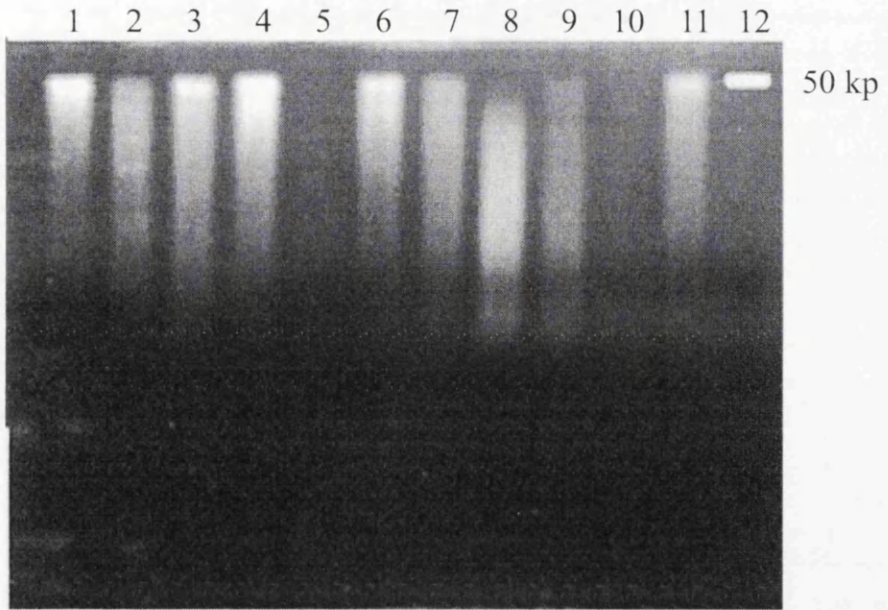


Figure 3.5 Quantity and quality of extracted DNA from buccal samples stored at different temperatures.

11 Buccal swabs samples were stored at either -80°C , 4°C , room temperature or 37°C for one month without cell solution. After incubation, DNA extractions were completed using the Puregene™ kit. $5\ \mu\text{l}$ representing 25% of the total extraction were analysed using 1% agarose gel containing ethidium bromide.

Lane 1	$5\ \mu\text{l}$ of DNA extracted from buccal cell stored at -80°C .
Lane 2	$5\ \mu\text{l}$ of DNA extracted from buccal cell stored at -80°C .
Lane 3	$5\ \mu\text{l}$ of DNA extracted from buccal cell stored at -80°C .
Lane 4	$5\ \mu\text{l}$ of DNA extracted from buccal cell stored at 4°C .
Lane 5	$5\ \mu\text{l}$ of DNA extracted from buccal cell stored at 4°C .
Lane 6	$5\ \mu\text{l}$ of DNA extracted from buccal cell stored at 4°C .
Lane 7	$5\ \mu\text{l}$ of DNA extracted from buccal cell stored at RT.
Lane 8	$5\ \mu\text{l}$ of DNA extracted from buccal cell stored at RT.
Lane 9	$5\ \mu\text{l}$ of DNA extracted from buccal cell stored at RT.
Lane 10	$5\ \mu\text{l}$ of DNA extracted from buccal cell stored at 37°C .
Lane 11	$5\ \mu\text{l}$ of DNA extracted from buccal cell stored at 37°C .
Lane 12	500 ng λ marker.



3.3.3.1. *D1S80 Locus*

D1S80 is a tandem repeat locus; it is classified as an amplified fragment length polymorphism marker with a 16 bp tandem repeat that can be used for the analysis of biological samples (Kasai *et al.*, 1990).

D1S80 locus has an allelic window from between approximately 350 bp to 1000 bp with a high rate of heterozygosity (Cosso *et al.*, 1995).

To determine the success of the protocol developed for the preservation of samples, amplification of D1S80 locus was performed using extracted DNA from blood and buccal samples. The size of D1S80 alleles found was between 350-800 bp. Consequently the samples, which were amplified at D1S80 locus, were expected to amplify successfully with STRs system because the STRs used in this study are smaller in size.

Successful amplification for D1S80 was obtained from all the blood and buccal samples, which were stored in cell lysis solution of Puregene™ kit. The extracted blood and buccal samples which were stored without cell lysis solution failed in 40% of cases. Both alleles of heterozygous samples were clearly visible and more intense using Puregene™ kit extracts after storage in cell lysis than when using extracts of samples stored without storage in cell lysis solution (Table 3.2 and Figure 3.6 and 3.7).

Table 3.2 PCR amplification at D1S80 locus using DNA extracted by using Puregene™ kit.

Some blood and buccal samples extracted by using Puregene™ kit after storage in cell lysis solution and some without storage in cell lysis solution were amplified at D1S80 locus. All samples, which were stored in cell lysis were amplified successfully.

	Puregene™ kit extracted DNA after storage in cell lysis solution		Puregene™ kit extracted DNA after storage without cell lysis solution	
	Blood samples	Buccal samples	Blood samples	Buccal samples
D1S80	Success in all 7 samples	Success in all 7 samples	Success in 4 from 7 samples	Success in 5 from 7 samples

Figure 3.6 Blood and buccal samples extracted by using Puregene™ kit and amplified for D1S80 locus.

PCR Amplification of D1S80 from some blood and buccal samples extracted by using Puregene™ kit after storage in cell lysis solution at -80°C , 4°C , RT and 37°C for one month.

- Lane 1 Hyperladder 1
- Lane 2 Successful PCR of DNA extracted from blood sample stored at -80°C .
- Lane 3 Successful PCR of DNA extracted from blood sample stored at -80°C .
- Lane 4 Successful PCR of DNA extracted from blood sample stored at 4°C .
- Lane 5 Successful PCR of DNA extracted from blood sample stored at 4°C .
- Lane 6 Successful PCR of DNA extracted from blood sample stored at RT.
- Lane 7 Successful PCR of DNA extracted from blood sample stored at RT.
- Lane 8 Successful PCR of DNA extracted from blood sample stored at 37°C .
- Lane 9 Successful PCR of DNA extracted from buccal sample stored at -80°C .
- Lane 10 Successful PCR of DNA extracted from buccal sample stored at -80°C .
- Lane 11 Successful PCR of DNA extracted from buccal sample stored at 4°C .
- Lane 12 Successful PCR of DNA extracted from buccal sample stored at 4°C .
- Lane 13 Successful PCR of DNA extracted from buccal sample stored at RT.
- Lane 14 Successful PCR of DNA extracted from buccalsample stored at RT.
- Lane 15 Successful PCR of DNA extracted from buccal sample stored at 37°C .

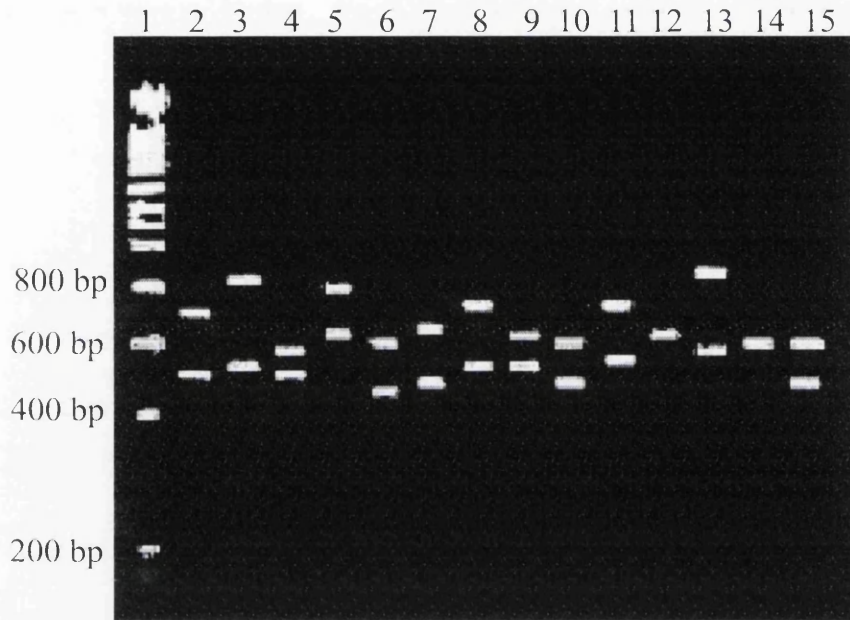
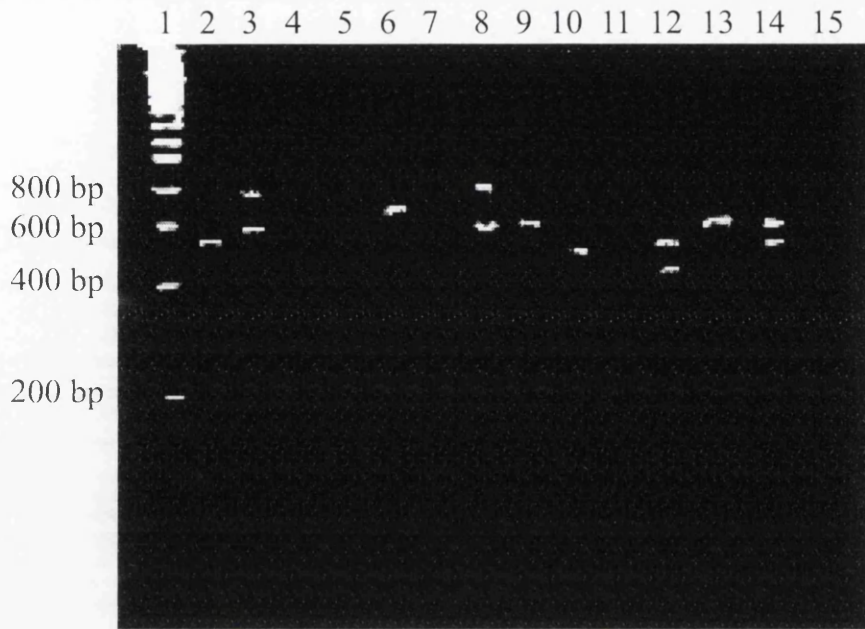


Figure 3.7 Blood and buccal samples extracted by using Puregene™ kit and amplified for D1S80 locus without storage of samples in the cell lysis solution.

PCR Amplification of D1S80 from some blood and buccal samples extracted by using Puregene™ kit after storage of blood samples at -80°C and buccal swabs samples at -80°C , 4°C , RT and 37°C for one month without cell lysis solution.

Lane 1	Hyperladder 1
Lane 2	Successful PCR of DNA extracted from blood sample stored at -80°C .
Lane 3	Successful PCR of DNA extracted from blood sample stored at -80°C .
Lane 4	Failed PCR of DNA extracted from blood sample stored at -80°C .
Lane 5	Failed PCR of DNA extracted from blood sample stored at -80°C .
Lane 6	Successful PCR of DNA extracted from blood sample stored at -80°C .
Lane 7	Failed PCR of DNA extracted from blood sample stored at -80°C .
Lane 8	Successful PCR of DNA extracted from blood sample stored at -80°C .
Lane 9	Successful PCR of DNA extracted from buccal sample stored at -80°C .
Lane 10	Successful PCR of DNA extracted from buccal sample stored at -80°C .
Lane 11	Failed PCR of DNA extracted from blood sample stored at 4°C .
Lane 12	Successful PCR of DNA extracted from buccal sample stored at 4°C .
Lane 13	Successful PCR of DNA extracted from buccal sample stored at RT.
Lane 14	Successful PCR of DNA extracted from buccal sample stored at RT.
Lane 15	Failed PCR of DNA extracted from blood sample stored at 37°C .



3.4. SAMPLING PROCEDURE

After establishing the technique to preserve the samples, 300 unrelated Egyptian blood samples were collected from El-Minia University Hospital. Although 200 samples are generally deemed to be enough, the larger the number of the samples in the databases, the more accurate the data will be (NRC, 1996). Additionally, buccal swabs from 30 families were obtained from the same place.

3.4.1. Blood Samples

Fresh blood samples (3 ml) were collected in an EDTA haematology tubes. White blood cells were separated from the blood samples and stored in the cell lysis solution of Puregene™ kit at ambient temperature around 37 °C for one month. These samples were easily transported, as they did not need any special handling or cooling devices for storage.

3.4.2. Buccal Samples

Buccal samples were collected from 30 different families from Egypt; each swab was stored in cell lysis solution at room temperature, around 37 °C for one month. All buccal swab samples were transported in simple packing without any need for special storage.

3.5. EXTRACTION OF SAMPLES

White blood cell samples and buccal swabs stored in the cell lysis solution were sent to the Human Identification Centre in Glasgow University, where all the samples were extracted using the Puregene™ kit. Samples were checked for the quantity of extracted DNA by using agarose gel electrophoresis. The samples were extracted successfully with 100% success rate. The samples yielded high quantities of DNA (Figure 3.8).

3.6. DISCUSSION

3.6.1. Sample Collection

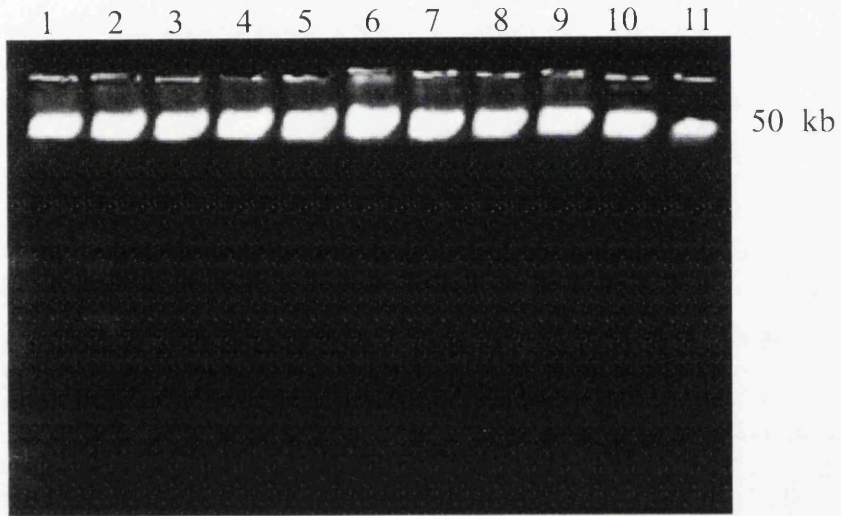
The study was conducted on 300 whole fresh blood samples that were collected from randomly selected, healthy adult Egyptian males and females. In addition, buccal swabs were collected from 30 Egyptian families. All samples were collected from Egyptian people living in El-Minia city, which is 250 km away from Cairo.

The blood sample is one of the richest sources of DNA that can be used in different types of analysis including RFLPs and PCR dependent analysis (Droog *et al.*, 1996). With the widening of the use of PCR, which requires much less DNA than traditional Southern blotting, other alternatives sources of DNA have been used for genetics testing including buccal cells, hair roots, saliva, and urine (Jeffreys, 1993).

Figure 3.8 Blood samples extracted by Puregene™ kit after storage in cell lysis solution at room temperature.

Blood samples stored in cell lysis solution of Puregene™ kit at room temperature in Egypt after lysis of Red Blood Cells. The samples were transported to Glasgow where DNA extractions were completed. 5 µl representing 5% of the total extraction was analysed using 1% agarose gel to quantify the amount of extracted DNA.

Lane 1	5 µl of DNA extracted from blood of sample 1 stored at RT.
Lane 2	5 µl of DNA extracted from blood of sample 1 stored at RT.
Lane 3	5 µl of DNA extracted from blood of sample 2 stored at RT.
Lane 4	5 µl of DNA extracted from blood of sample 3 stored at RT.
Lane 5	5 µl of DNA extracted from blood of sample 4 stored at RT.
Lane 6	5 µl of DNA extracted from blood of sample 5 stored at RT.
Lane 7	5 µl of DNA extracted from blood of sample 6 stored at RT.
Lane 8	5 µl of DNA extracted from blood of sample 7 stored at RT.
Lane 9	5 µl of DNA extracted from blood of sample 8 stored at RT.
Lane 10	5 µl of DNA extracted from blood of sample 9 stored at RT.
Lane 11	500 ng λ marker.



However, some of these sources may be problematic; hair roots yield only small quantities of DNA (Pilkington *et al.*, 1987) and in addition, hair roots may be absent in children under 2 years old and elderly people (Pancorbo *et al.*, 1998). Saliva and urine were impractical for collection and also preservation. By contrast buccal swabs collection was simple and obviated the need of venipuncture. Thomson *et al.*, (1992) recommended the use of buccal cells for children in the paternity analysis. In the United States, the use of buccal swabs has become the trend to test DNA instead of venipuncture for children in paternity testing (Maha *et al.*, 1995). STR typing of buccal swabs in paternity cases is recommended where a bone marrow transplant recipient is involved. In this case typing of DNA from leukocytes yields the genotypes of the donor. In contrast, DNA from buccal cells remains unchanged after bone marrow transplantation (Okazaki *et al.*, 1996).

3.6.2. Preservation of Samples

Any biological substance once out of the body is in a foreign environment and degradation of DNA begins to occur. Degradation has an adverse effect on the ability to obtain a result from DNA typing particularly at loci with bigger size of products where allele drop out occurred in high frequency (Walsh *et al.*, 1992). Therefore, the ability to perform successful analysis on biological samples depends mainly on the first step, which is how the biological samples are collected and preserved. Experiments have been done in this study to achieve a reliable simple method for preservation of blood and buccal samples taken in hostile environmental conditions. Both blood and buccal samples were stored in cell lysis solution of Puregene™ Kit at different

temperatures without any significant effect on the quantity and quality of extracted DNA. Preservation of both blood and buccal samples in cell lysis solution was performed immediately after collection, thus preventing degradation of DNA after extraction and subsequently yielding enough DNA for different types of analysis. In other reports DNA isolation was difficult if blood samples were poorly handled. Two causes were identified for the problem, incorrect blood/EDTA ratio or due to poor mixing of the blood/EDTA before storage in freezer (Droog *et al.*, 1996).

Moreover, the samples preserved by this method can be stored at room temperature, thus there is no need for the cooling equipment, at the place of collection and during the shipment of the samples removing potential problems such as cooling being interrupted for any reason during the transport of samples to the analysing laboratory.

3.6.3. DNA Extraction

A variety of DNA extraction techniques have been used for forensic DNA analysis procedures. This study evaluated one extraction method for its effectiveness in yielding DNA suitable for analysis. Factors that should be considered when evaluating the extraction method include DNA yield and the suitability of DNA for amplification. The amplification of the extracted DNA at D1S80 locus was used to evaluate this type of extraction. Each DNA sample was subject to amplification at the locus D1S80 followed by typing of the locus by electrophoresis on polyacrylamide gel and ethidium bromide staining.

Puregene™ extraction proved to be effective in yielding sufficient DNA suitable for the amplification from both blood and buccal samples. Significant increase in the yield of product from the amplification of D1S80 was observed with samples extracted by using Puregene™ kit after storage of samples in cell lysis solution.

3.6.4. Quantity of Extracted DNA

3.6.4.1. *Blood Samples*

It has been found that storage of white blood cells in cell lysis at different temperatures for one month did not have a marked effect on the ability to isolate high molecular weight DNA. In contrast DNA extracted from blood samples stored directly at -80°C for one month showed degradation of extracted DNA. Madisen *et al.*, (1987) reported that storage of liquid blood at different temperatures for varying periods of time have a marked effect on the extracted DNA yielding low molecular weight DNA. Another study reported that blood stored at -70°C may yield less high molecular weight DNA than expected (Kobilinsky, 1992).

3.6.4.2. *Buccal Swabs*

Buccal swabs, stored in cell lysis solution at different temperatures yielded sufficient quantities of DNA suitable for the amplification of D1S80 locus.

Although the technique used in the study of DNA was PCR dependent, the DNA needed to be of high quality and quantity because several tests would be undertaken.

Moreover, degraded samples may cause drop out of the alleles and incorrect interpretation with heterozygote samples possibly being called as homozygotes. Also as the samples had to be collected from Egypt, it was important that the samples were of high quality, as the time and expense in getting additional samples would be prohibitive. The results of this study demonstrated that incubation of the samples in cell lysis solution of Puregene™ kit is a valid method for preservation of samples especially if the samples are to be transported across long distances. In addition, the use of buccal swabs as a source for DNA can be introduced in Egypt especially for paternity testing.

CHAPTER 4: AUTOSOMAL STR ANALYSIS

4.1. INTRODUCTION

Several methods of DNA typing have been used in forensic applications including analysis of VNTRs, AmpFLPs and STRs. STRs have proven to be the markers of choice in the field of paternity testing and in human identification (Rostedt *et al.*, 1996). There are several advantages to utilizing STRs over the other systems. STRs analysis unlike VNTRs analysis uses PCR amplification and therefore requires less DNA (Lorente *et al.*, 1994). Alleles of STRs are discrete and can be sized precisely so complex interpretation as used in VNTR systems is not needed, making interpretation of results easier (Gill *et al.*, 1990). In addition STRs analysis has lower costs, and greater speed. Moreover, multiplexing of several loci can be performed to increase the power of discrimination and the power of exclusion of paternity of the STR system. The application of STRs typing to paternity testing has also become preferable, as STRs analysis requires a small amount of blood or the use of other sample types such as buccal cells (Thomson *et al.*, 1999). Based on the DNA technologies available, the analysis of STRs was chosen for this study.

4 2. SELECTION OF STR LOCI

There are approximately 400 million STR loci in the human genome including di, tri and tetranucleotide loci (Edwards *et al.*, 1991). Tetranucleotide loci are attractive

for forensic and paternity analysis because of the low rate of artifacts generated by repeat slippage, which are common features with dinucleotide repeat loci (Beckman and Weber, 1992). To examine the suitability of loci for forensic purposes, the European DNA Profiling Group (EDNAP) carried out an inter-laboratory experiment and confirmed that simple repeat STR loci are easier to interpret. They are also the most suitable loci when results are to be shown between labs, or comparison made between different studies (Gill *et al.*, 1994).

The aim of this study was to examine the suitability of the STR system for paternity testing in Egypt. The Egyptian population is composed of different ethnic groups including Copts, Arabs, Bedouins and Nubians. Romans, Greeks, more recently the Turks, and even French and English have also intermarried with the Egyptians. Additionally, cultural practises encourage marriage between relatives especially in rural areas adding further to the complex nature of the Egyptian population. These marriages could have an effect on the HWE and therefore on the statistical interpretation of results. Therefore, to examine the effects caused by the presence of different ethnic groups in the Egyptian populations and the relatively high level of consanguineous marriage, some simple loci were examined.

Initially after looking at published STR population studies, three simple loci were chosen for analysis. These loci were D3S1358, vWA and FGA. Candidate loci were selected for the development because they have the criteria that are recommended by the European DNA Profiling Group (EDNAP) (Kimpton *et al.*, 1995). The chosen loci have four base pair tandem repeat units, high degrees of polymorphisms in all

published data and amplification produces reproducible results with minimal artifact bands, which allows automation and multiplexing (Lezaum *et al.*, 1997). The loci are located on different chromosomes and the absence of linkage and agreement with Hardy-Weinberg equilibrium has been established in other populations (Cabrero *et al.*, 1995). Previous studies have shown that the products from alleles at these loci can be analysed as discretely sized bands and with further sequencing studies, allelic designations have been based on the number of repeat units present (Szibor *et al.*, 1998 and Barber *et al.*, 1995, 1996). A fluorescent STR multiplex for D3S1358, vWA, and FGA has been described (AmpF ℓ STR Blue kit) (Applied Biosystem; CA, USA), although information of the sequence of the primers was not available.

Moreover, many different population data for these loci were published that allowed comparison to be performed between the Egyptian population and other published data (Andres *et al.*, 1998; Budowle *et al.*, 1997; Yamamoto *et al.*, 1998 and Zehner, 1998) to assess the genetic profile of the Egyptian population.

The application of these three loci in paternity testing had been studied (Benecke *et al.*, 1997). Paternity testing should utilise loci with low levels of mutation. D3S1358 locus was used to determine paternity testing for 780 meioses with only two mutations found (0.25% mutation rate) (Szibor *et al.*, 1998). The application of the vWA locus in paternity and forensic cases has also been examined by Sajantila *et al.*, (1994). No evidence of mutation was found in 24 and 90 families examined in separate studies (Lorente *et al.*, 1994 and Trabetti *et al.*, 1993). The FGA locus was found to be very useful for paternity testing and analysis of forensic cases (Neuhuber *et al.*, 1998).

In addition, many studies confirmed the usefulness of AmpF ℓ STR Blue multiplex system in both paternity testing and routine forensic cases (Benecke *et al.*, 1997 and Miscicka-Sliwka *et al.*, 1999). There was also no mutation at any of the AmpF ℓ STR Blue loci in the 82 meiotic events tested (Wallin *et al.*, 1998).

4.3. DESCRIPTION OF THE LOCI

4.3.1. D3S1358 Locus

D3S1358 locus is located on chromosome 3. The sequence of the repeat region is (TCTG)₂₋₃ and (TCTA)_n (Szibor *et al.*, 1998). The small size of D3S1358 alleles makes this locus useful for forensic applications as the locus can be amplified by PCR from severely degraded samples (Santos *et al.*, 1996).

4.3.2. Human vWA Locus

Human vWA locus is one of three polymorphic regions within intron 40 of the von Willebrand factor gene (Kimpton *et al.*, 1992). vWA is located at chromosome 12p12-12pter, between nucleotides 1640 and 1794 (Mancuso *et al.*, 1989). The repeat motif of the vWA locus is (TCTR)_n where R is either A or G (Moller *et al.*, 1994). The vWA locus has been described as a compound repeat with non-consensus repeat alleles (Urquhart *et al.*, 1994).

4.3.3. Human FGA Locus

The FGA locus is located in the third intron of the human alpha fibrinogen gene at the chromosomal position 4q28 (Mills *et al.*, 1992). The FGA locus has been considered as a complex tetranucleotide repeat (Mills *et al.*, 1992). The structure of the repeat region is [TTTC]₃ TTTTTTCT [CTTT]_n CTCC [TTCC]₂. The alleles differ in size by one repeat unit in most cases. However, there are microvariant alleles, which differ by just two bp (Barber *et al.*, 1996).

4.4. SELECTION OF PRIMERS

New primers were designed to amplify D3S1358, vWA and FGA loci. Using the accession number, the sequence of vWA and FGA loci was obtained from GeneBank. The sequence of the D3S1358 locus was not available during the designation of the primer. Therefore, the primer for D3S1358 locus was a modified primer of Li *et al.*, (1993). For PCR primer pair selection, a target range of the template sequence surrounding the repeat unit of the three loci was chosen. Primers selection followed the standard simple rules for the design of the primers (Chou *et al.*, 1992). All primers did not have to be the same length but did not exceed 25 bp, the CG content was not more than 60% and was similar between each primer pair. Primer dimer and hairpin loops were avoided by designing primer sequence, which were non-complementary with sequences in other primers or within the primer itself. The melting temperature TM was calculated as sum of 4 °C for G/C and 2 °C for A/T and 5 °C was subtracted from the total value. The affinity of the primer to each other or to DNA sequences outside the

target template was tested with the help of the computer programs, Primer Pick (Rozen and Skaletsky, 1996 and 1997) and Williamstone (www.Williamstone.com). The sequences of the three loci with the attachment of the primers are shown in appendix 1 section 1.1.

4.5. OPTIMIZATION OF THE PCR SYSTEM

The determination of the concentration of each critical component of PCR and the parameters of thermocycling was required to achieve the highest degree of specificity of the PCR products.

4.5.1. Single Locus PCR

In order to achieve a singleplex PCR optimisation, several experiments were performed to examine the effect of several changes of the PCR parameters. One parameter was tested at a time while all other parameters were set at the default condition which was 1.5 mM MgCl₂, 0.2 μM primer concentration, 1.25 U *Taq* enzyme, 200 μM dNTPs, 1X PCR buffer, 30 cycles. Once a single component of the PCR was optimised, the others were tested sequentially until each component was at the optimal concentration relative to the other in the reaction mix.

4.5.1.1. Annealing Temperature

The annealing temperature is one of the most critical parameters in the PCR. Different annealing temperatures ranging from 56 °C to 64 °C were tested. The yield

of PCR products was most specific at 62 °C for D3S1358, 60 °C for the vWA locus and 61 °C for the FGA locus.

4.5.1.2. Magnesium Chloride

Once the annealing temperature was optimised, the effect of different MgCl₂ concentrations on the PCR product was tested at this temperature. Concentration of each dNTP was kept at 200 µM each and gradually increased the concentration of MgCl₂ from 1.5 mM to 3 mM. The PCR products displayed the greatest degree of specificity at 2 mM of MgCl₂ for vWA and D3S1358 loci and at 2.5 mM MgCl₂ for the FGA locus.

4.5.1.3. Primer Concentration

After optimisation of both annealing temperature and MgCl₂, different primer concentrations were tested ranging from 0.1 µM to 0.3 µM. The optimised products of the PCR were obtained at different concentrations, the optimised yield for D3S1358 locus was achieved at 0.25 µM and 0.2 µM was optimum for both vWA and FGA loci.

4.5.1.4. Taq Polymerase

Different concentrations of *Taq* polymerase were tested ranging from 0.625 U to 2 U per reaction. The optimised enzyme concentrations were 0.625 U/25 µl reaction volumes for vWA and FGA loci and 1.25 U/25 µl for D3S1358 locus. A greater concentration of enzyme resulted in non-specific products. Two different polymerase

enzymes were tested to compare between the results, Ampli *Taq* gold (Applied Biosystem; CA, USA) and *Taq* polymerase (Promega). There was no significant difference between the results of the two enzymes.

4.5.1.5. Extension Time and Temperature

Different extension times ranging from 30 sec to 2 min were tested with two PCR programmes, one at 65 °C and the other at 72 °C. The highest yield of PCR product was obtained at 72 °C for 1 min for the three loci. Final incubation at 65 °C for 10 min improved the amount of the PCR products. Once the conditions had been optimised, the last step was to re-test the optimised parameters several times to achieve reproducible results (Table 4.1).

4.5.2. Multiplex PCR

One advantage of using STR systems is that more than one STR locus can be amplified in one reaction. While the initial aim of this work was not to produce a multiplex, it would be advantageous to analyse the STRs together in a multiplex. The products of vWA and D3S1358 overlapped, therefore it was not possible to amplify these loci as a triplex. If a triplex PCR was advantageous, a different fluorescent dye could be used to label the forward primer for vWA locus but as the primary aim was to investigate the structure of the Egyptian population and to obtain the allele frequencies this was not necessary. The PCR was optimised as a duplex for D3S1358 and the FGA and as a singleplex for the vWA locus.

Table 4.1 The optimised singleplex PCR parameters

The annealing temperature was different for each locus. Two loci may have the same concentration in the other parameters such as MgCl₂ and primer concentration and all the three loci had the same extension temperature.

PCR PARAMETERS	D3S1358	vWA	FGA
Annealing temperature	62 °C	60 °C	61 °C
MgCl ₂	2 mM	2 mM	2.5 mM
Primer concentration	0.25 µM	0.2 µM	0.2 µM
<i>Taq</i> polymerase	1.25 U	0.652 U	0.652 U
Extension temperature	72 °C	72 °C	72 °C

Duplex PCR was performed by combining 0.2 μ M equimolar amounts of each primer for D3S1358 and FGA loci using the singleplex PCR programme with the lower annealing temperature. There was unequal amplification of loci as D3S1358 products were weak compared to the FGA products. Decreasing the concentration of the FGA primer to 0.1 μ M yielded approximately equal products for the two loci (Figure 4.1). Once the primer concentration was determined, the other PCR parameters needed to be changed by testing a range of annealing temperature, $MgCl_2$, *Taq* polymerase enzyme, and extension temperature in that order. All parameters of optimised PCR for the duplex PCR are shown in table 4.2.

4.6. SIZING AND DESIGNATION OF ALLELES

The results were going to be compared to different populations; therefore it was important to correctly designate each allele. The achievement of accurate sizing of the alleles was carried out by construction of allelic ladders from most of the alleles found for the three loci. Sufficient numbers of the samples were initially analysed without ladders to find the samples that represented all of the alleles present. These samples were amplified as a singleplex for each locus and then mixed together to construct a specific ladder for each locus.

88A

Figure 4.1 Optimisation of duplex PCR (D3S1358 and FGA).

1-The upper electrophoretogram shows preferential amplification of FGA locus in the duplex PCR when the same concentration ($0.2 \mu\text{M}$) of the primers of D3S1358 and FGA loci was used in the PCR.

2-The lower electrophoretogram shows optimised PCR products of the duplex PCR obtained by using $0.1 \mu\text{M}$ of the primer of the FGA locus and $0.2 \mu\text{M}$ of the primer of the D3S1358 locus.

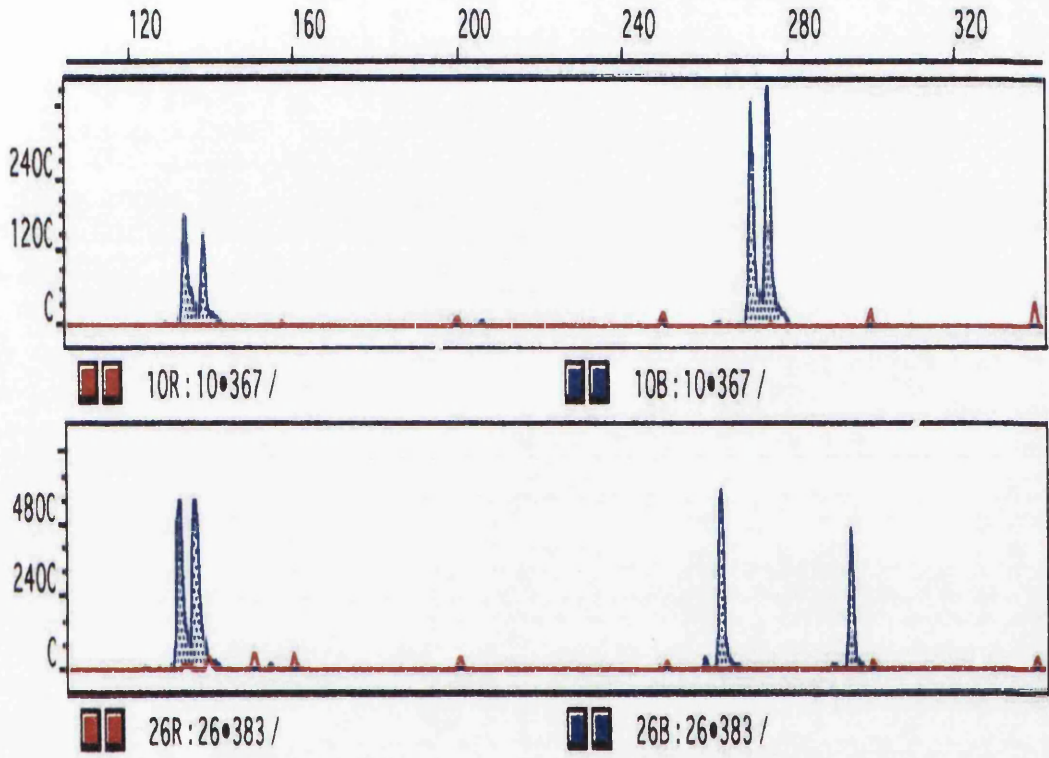


Table 4.2 Optimised parameters for a duplex PCR.

The concentration of the FGA primer was decreased to 0.1 μM to yield approximately equal products of D3S1358 locus.

PCR PARAMETER	D3S1358 & FGA
Annealing temp	58 °C
MgCl ₂	2.5 mM
Primer conc	0.1 μM FGA 0.2 μM D3S1358
<i>Taq</i> polymerase	0.625 U
Extension temp.	72 °C

In order to determine the mean size of each allele, allelic ladders with a range of alleles were reanalysed on three gels (Table 4.3). The standard deviation for D3S1358 ranged from 0.0577-0.1528 bp while for the vWA locus the standard deviation ranged from 0.0577-0.1477 bp and the standard deviation for the FGA locus ranged from 0.0608-0.1877 bp. Allelic windows were constructed using ± 3 SD relative to the allelic ladders in order to designate the allelic size of unknown samples. This precision allowed clear distinction between alleles with full tetrameric repeats and FGA alleles with incomplete 2 bp repeats. As a double check all samples were run in comparison with the allelic ladders to achieve accurate measurement of all alleles (Figures 4.2, 4.3 and 4.4).

Due to the use of new primers, the amplified bands (alleles) produced by the three loci were of different size compared to those using some previously reported primers. Designation of the alleles was performed by amplification of a range of samples with both these primers and the commercial kit (AmpF ℓ STR Blue kit) (Applied Biosystem; CA, USA). Comparison between the sizes of the alleles of the new primers with the results of the amplification with the commercial kit allowed the “in house” ladders to be calibrated (Table 4.4 and Figure 4.5).

Table 4.3 The mean length of the alleles for three loci.

Allelic ladders with most of all ranges of alleles were reanalysed on three different gels in order to determine the mean size of each allele. The SD was calculated for each locus. The FGA locus had the highest SD.

Allele	D3S1358 Mean size (bp) ± SD.	vWA Mean size (bp) ± SD.	FGA Mean size (bp) ± SD.
13	120.07 ± 0.0618	136.14 ± 0.0577	
14	124.14 ± 0.0577	140.12 ± 0.0931	
15	128.05 ± 0.1000	143.63 ± 0.1437	
16	131.89 ± 0.1155	148.13 ± 0.1155	
17	135.78 ± 0.0931	152.19 ± 0.1178	256.25 ± 0.0608
18	140.30 ± 0.1069	156.17 ± 0.0577	260.13 ± 0.1437
19	144.14 ± 0.1528	160.01 ± 0.1121	264.02 ± 0.1124
20		163.96 ± 0.0525	267.96 ± 0.0838
21		167.87 ± 0.1437	271.79 ± 0.1121
22		171.84 ± 0.0618	275.62 ± 0.0904
23			279.40 ± 0.1069
24			283.39 ± 0.1877
25			287.15 ± 0.0931
26			291.07 ± 0.1178
27			294.92 ± 0.1437
28			298.69 ± 0.0932
29			302.61 ± 0.1092
30			306.52 ± 0.1437
Average	0.0983	0.0956	0.1178

92A

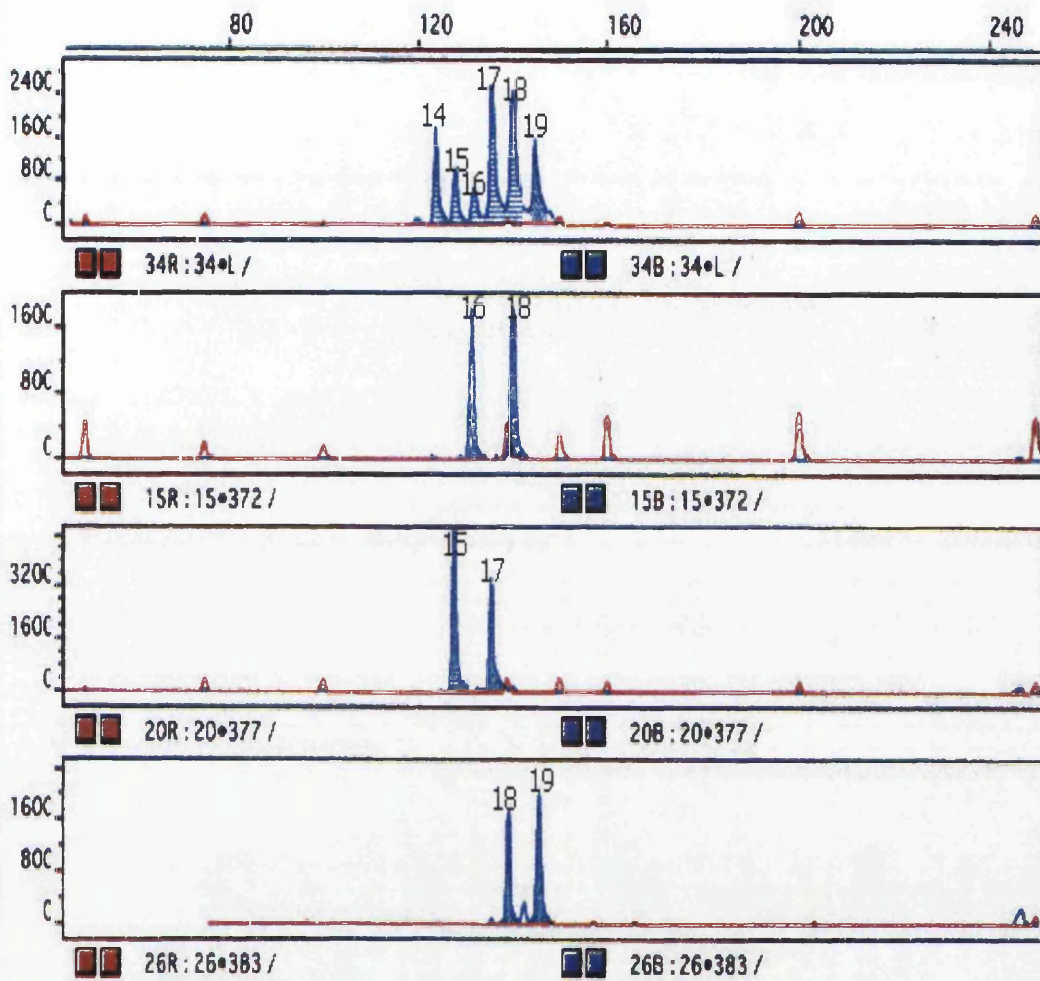
Figure 4.2 Electrophoretogram shows the D3S1358 allelic ladder

The ladder contained alleles from allele 14 to allele 19. The D3S1358 ladder was run in the same gel with the samples. The differences of the size between the ladder and the samples were less than 0.5 bp.

1-Panel one shows the D3S1358 allelic ladder.

2-Panels 2-4 show three samples in comparison to the allelic ladder.

3-The bottom table shows the sizes of the alleles of the ladder and the samples.



Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
34B, 2	98.50	124.12	1827	10861	985
34B, 3	100.80	128.06	1045	7459	1008
34B, 4	103.10	131.99	714	6314	1031
34B, 5	105.30	135.75	2642	22970	1053
34B, 6	107.90	140.33	2506	25035	1079
34B, 7	110.30	144.85	588	8136	1103
15B, 3	104.10	131.60	1866	10612	1041
15B, 5	109.10	140.19	1909	11684	1091
20B, 4	101.40	127.72	5520	29080	1014
20B, 5	106.10	135.62	3523	21124	1061
26B, 2	101.10	140.24	1761	10145	1011
26B, 4	105.50	144.64	2098	14270	1055

93A

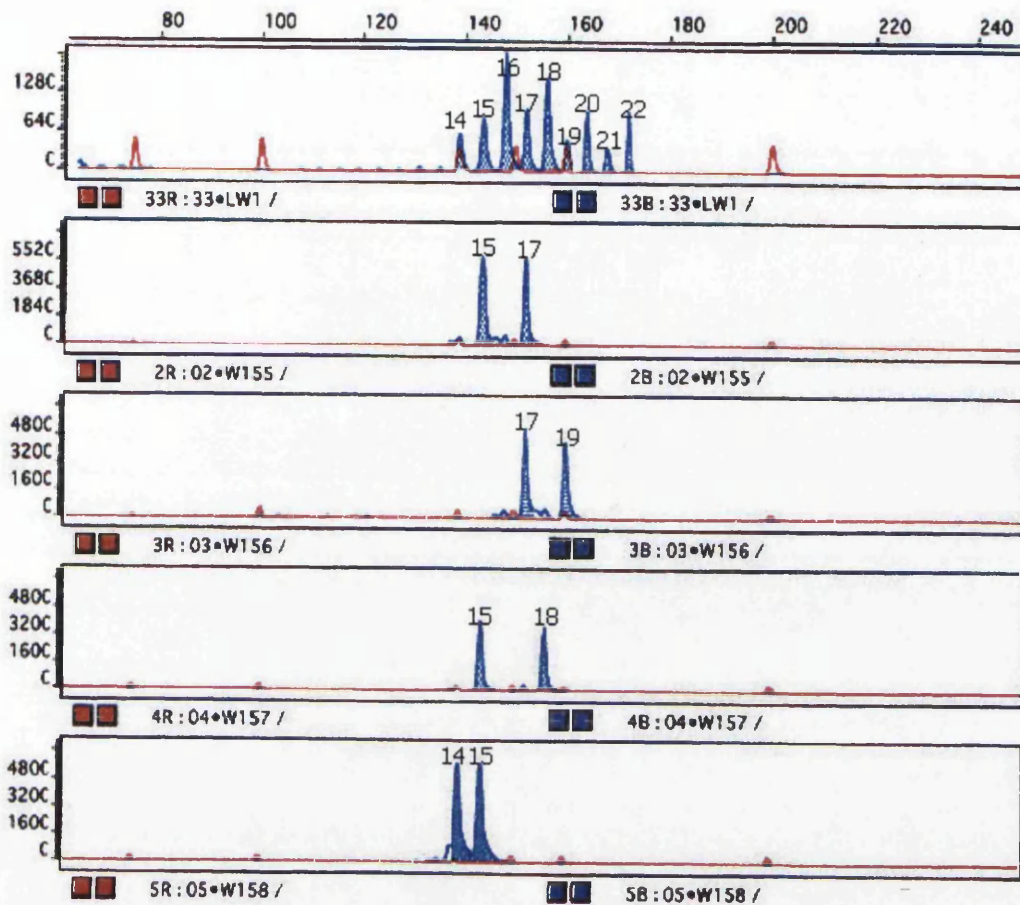
Figure 4.3 Electrophoretogram shows the vWA allelic ladder

The ladder contained alleles from allele 14 to allele 22. The vWA ladder was run in the same gel with the samples. The differences of the size between the ladder and the samples were less than 0.5 bp.

1-Panel one shows the vWA allelic ladder.

2-Panels 2-5 show three samples in comparison to the allelic ladder.

3-The bottom table shows the sizes of the alleles of the ladder and the samples.



Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
33B, 2	146.40	139.14	649	4574	1464
33B, 3	149.60	143.61	928	7340	1496
33B, 4	152.90	148.12	1976	16623	1529
33B, 5	156.10	152.21	1063	9320	1561
33B, 6	159.40	156.16	1619	13611	1594
33B, 7	162.70	160.00	530	4092	1627
33B, 8	166.10	163.99	974	13407	1661
33B, 9	169.40	167.86	392	2853	1694
33B, 10	172.80	171.83	1176	11238	1728
2B, 3	151.20	143.69	5874	42734	1512
2B, 6	157.70	152.20	5950	43634	1577
3B, 13	158.00	152.33	5371	43101	1580
3B, 16	164.60	160.12	4729	39368	1646
4B, 3	151.40	143.69	4296	28046	1514
4B, 6	161.30	156.27	3905	27856	1613
5B, 4	148.20	139.41	5880	63041	1482
5B, 5	151.50	143.90	5874	65579	1515

94A

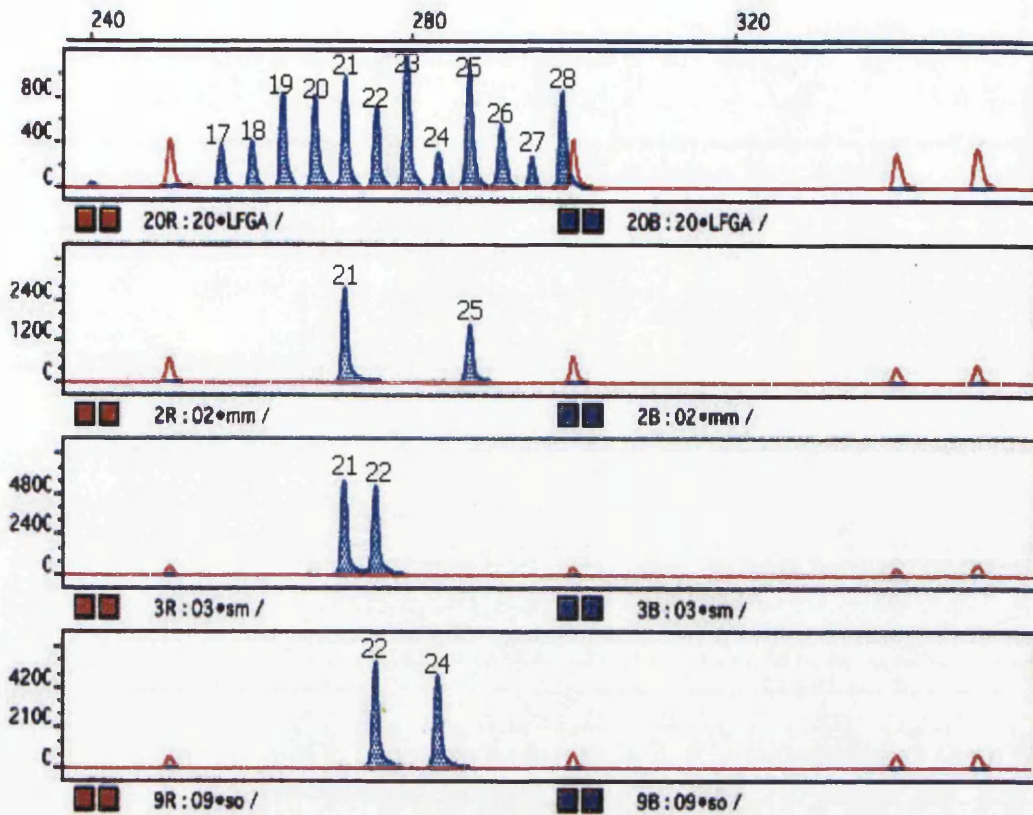
Figure 4.4 Electrophoretogram shows the FGA allelic ladder

The ladder contained alleles from allele 17 to allele 28. The FGA ladder was run in the same gel with the samples. The differences of the size between the ladder and the samples were less than 0.5 bp.

1-Panel one shows the FGA allelic ladder.

2-Panels 2-4 show three samples in comparison to the allelic ladder.

3-The bottom table shows the sizes of the alleles of the ladder and the samples.



Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
20B, 1	266.10	256.24	392	2802	2661
20B, 2	269.90	260.12	424	3085	2699
20B, 3	273.70	264.00	857	6536	2737
20B, 4	277.60	267.97	823	7766	2776
20B, 5	281.30	271.74	1016	8176	2813
20B, 6	285.10	275.61	722	5749	2851
20B, 7	288.80	279.38	1191	10427	2888
20B, 8	292.70	283.35	326	2357	2927
20B, 9	296.40	287.11	1109	9295	2964
20B, 10	300.30	291.07	568	5445	3003
20B, 11	304.10	294.93	285	2050	3041
20B, 12	307.80	298.68	878	7420	3078
2B, 2	282.80	271.76	2805	24682	2828
2B, 6	297.90	287.16	1736	15945	2979
3B, 2	282.90	271.72	5842	51196	2829
3B, 3	286.70	275.59	5511	54642	2867
9B, 2	287.40	275.59	5696	52424	2874
9B, 4	295.00	283.29	5025	47208	2950

Table 4.4 Designation of the alleles of the three loci.

Designation of the alleles was performed by amplification of a range of samples with both the new primers and the commercial kit (AmpF ℓ STR Blue kit) (Applied Biosystem; CA, USA).

D3S1358	New primer		Commercial kit		
Sample no.	Allele 1	Allele 2	Allele 1	Allele 2	Allele calling
1	128.05	128.05	123.57	123.57	15,15
2	135.75	140.30	131.58	135.88	17,18
3	131.88	135.78	127.50	131.58	16,17
4	131.89	131.89	127.50	127.50	16,16
5	135.76	135.76	131.58	131.58	17,17
6	120.06	128.04	115.50	123.57	13,15
7	124.12	131.88	119.54	127.50	14,16

vWA	New primer		Commercial kit		
Sample no.	Allele 1	Allele 2	Allele 1	Allele 2	Allele calling
1	140.11	148.13	167.46	175.55	14,16
2	148.14	163.96	175.55	191.34	16,20
3	143.62	156.16	171.34	183.46	15,18
4	152.19	160.01	179.38	187.12	17,19
5	148.14	152.17	175.55	179.38	16,17
6	143.63	160.02	171.34	187.12	15,19

FGA	New primer		Commercial kit		
Sample no.	Allele 1	Allele 2	Allele 1	Allele 2	Allele calling
1	267.96	272.01	224.49	228.56	20,21
2	272.03	287.15	228.54	244.78	21,25
3	264.02	283.39	220.56	240.71	19,24
4	267.94	283.38	224.45	240.67	20,24
5	279.40	294.91	236.65	252.70	23,27
6	260.13	298.67	216.45	256.90	18,28
7	287.17	291.08	244.72	248.80	25,26
8	264.03	283.37	220.52	240.67	19,24

9/0A

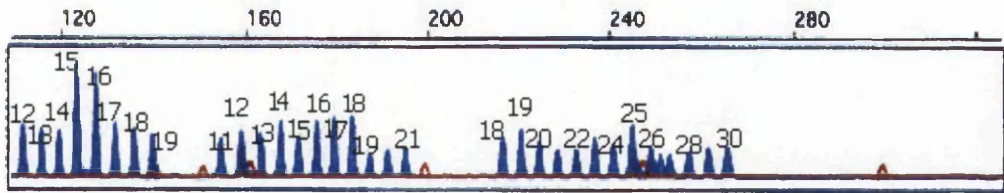
Figure 4.5 Electrophoretogram shows amplification of one sample by using both the new “in house” primers and the AmpF ℓ STR Blue Kit (Applied Biosystem; CA, USA).

1-The first electrophoretogram shows the ladder of the commercial kit (AmpF ℓ STR Blue Kit).

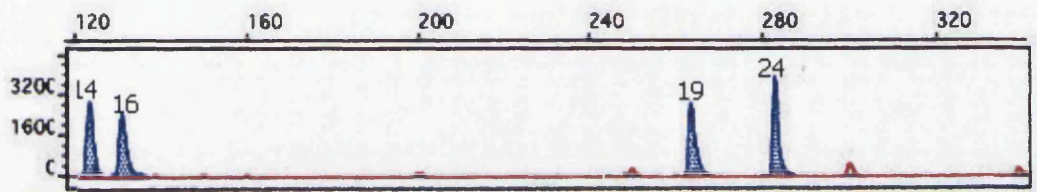
2-The second electrophoretogram shows one sample amplified with the commercial kit.

3-The third electrophoretogram shows same sample amplified as a duplex PCR at D3S1358 and FGA loci using the new primers of this study.

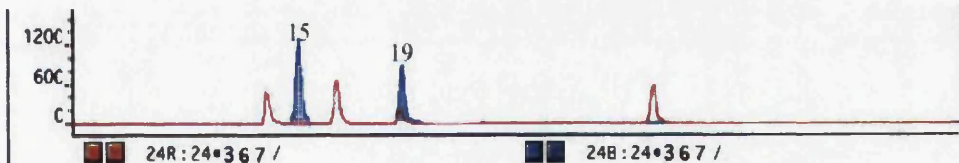
4-The fourth electrophoretogram shows the same sample amplified as a singleplex PCR at vWA locus using the new primer of this study.



7R : Sample 19 /			7B : Sample 19 /		
7B, 14	15.18	119.57	6104	55184	4140
7B, 16	15.48	127.70	6820	46895	4221
7B, 18	16.65	171.57	6560	52188	4540
7B, 22	17.20	187.43	5968	49616	4691
7B, 25	18.36	220.62	6205	53562	5007
7B, 27	19.04	240.76	5817	49291	5192



10R : 10*367 /			10B : 10*367 /		
Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
10B, 3	100.30	123.78	3699	38292	1003
10B, 5	104.90	131.37	3106	33402	1049
10B, 11	189.60	263.58	3525	41071	1896
10B, 13	203.20	283.96	4791	45447	2032



24R : 24*367 /			24B : 24*367 /		
Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
24B, 9	111.80	144.00	1378	6691	1118
24B, 13	121.30	160.32	924	5802	1213

4.7. STATISTICAL ANALYSIS

Once the PCR had been optimised, 300 unrelated Egyptian individuals from El-Minia city were typed for the three loci.

The statistical analysis was then performed using GDA (Lewis and Zaykin, 1999), TFPGA (Miller, 1998) and PowerStats (Tereba, 1999) software programmes.

4.7.1. Genotype Frequencies

A genotype frequency is the proportion of a population that has one genotype relative to all genotypes at a specific locus. The observed genotype frequencies were calculated by dividing the number of occurrence of each genotype over the total number of individuals.

The expected genotype frequencies were calculated according to HWE equation. In the case of homozygous genotype, the genotype frequency is equal to the square of the frequency of the allele (p^2 or q^2). In case of heterozygous genotype, the genotype frequency is equal to two multiplied by the frequencies of the two alleles involved in the genotype ($2pq$).

The standard deviation (SD) for the expected genotype frequencies was calculated according to the equation (Evetts and Weir, 1998).

$$SD = \sqrt{P(1 - P) / n}$$

where p is the frequency of the genotype and n is the number of population. The standard deviation of the expected genotypes are shown in appendix 1 section 1.2.

There were 23 genotypes produced from the possible 28 different genotypes in the Egyptian population at D3S1358 locus. The 15,16 genotype had the highest frequency (16.3%) occurring in 49 out of 300 individuals.

The vWA locus had 55 possible genotypes, from them 41 genotypes were observed. The genotype 16,17 (14%) was the commonest one occurring in 41 from 300 individuals.

From a possible 136 genotypes, 69 were observed for FGA locus. The genotype 22,23 was the most common one seen in 22 from 300 individuals (7.3%). The complete genotypes observed in the three loci are shown in table 4.5.

4.7.2. Allele Frequencies

Any new DNA typing method or marker cannot be fully utilised until the population frequencies of the DNA marker have been established. The allele frequency is the proportion of one allele relative to all alleles at the locus in the population.

Table 4.5 Genotypes observed in 300 Egyptian individuals for the three loci.

There were 23 genotypes in the Egyptian population at D3S1358 locus. The 15,16 genotype had the high frequency (16.3%).

D3S1358 Allele	13	14	15	16	17	18	19
13							
14	1	2					
15	1	17	15				
16	1	15	49	18			
17	2	9	37	39	15		
18		5	19	23	20	3	
19			2	3	2	2	

There were 41 genotypes in the Egyptian population at the vWA locus. The genotype 16,17 (14%) was the commonest one.

vWA Allele	13	14	15	16	17	18	19	20	21	22
13										
14		3								
15		13	6							
16	1	17	19	12						
17		11	25	41	13					
18		6	16	20	27	7				
19		4	6	18	9	7	2			
20			1	4	6	2	1			
21					1	1	1			
22					1					

There were 69 genotypes in the Egyptian population at the FGA locus. The genotype 22,23 had the highest frequency (7.3%).

FGA Allele	17	18	19	19.2	20	21	21.2	22	23	24	25	26	27	28	29	30
17																
18		1														
19			1													
19.2																
20	1	1	2		3											
21	1	1	2	1	11	4										
21.2																
22	1	3	4		12	18		5								
23	1	3	8		6	7	1	22	8							
24	1	2	9		11	15	1	8	8	7						
25		1	3		4	7		6	15	9	3					
26		1	1		3	5		5	7	3	5	1				
27			1		1	3		2	2	2	1					
28			2		1	2		2	2	5	2					
29						1				1						
30									1							

The allele frequencies for the Egyptian population were calculated from the genotypes observed for three loci and divided by the total number of all alleles. The standard deviation was calculated according to the equation (Evetts and Weir, 1998).

$$SD = \sqrt{p(1-p) / 2n}$$

where p is the frequency of the allele and n is the number of population.

Seven alleles were detected at the D3S1358 locus ranging from 120-144 bp. Alleles ranged from allele 13 to allele 19, with allele 16 being the most common allele at a frequency of 0.276. Allele 13 had the lowest frequency of 0.0083.

Ten alleles were detected at the vWA locus, their sizes ranging from 136-172 bp. Alleles ranged from allele 13 to allele 22, among them allele 17 and 16 were relatively common, with a frequency of 0.245 and 0.2383 respectively. Alleles 13 and 22 had the lowest frequency of 0.0017.

The FGA locus had 16 alleles ranging from 256-308 bp. Alleles ranged from allele 17 to allele 30, all are based on the tetrameric repeat unit except two microvariants, which were alleles 19.2 and 21.2. Three alleles of FGA were relatively common, which were alleles 22, 23 and 24 with frequencies of 0.1550, 0.165 and 0.1483 respectively. Alleles 19.2 and 30 had the lowest frequencies of 0.0017. The alleles frequencies are shown in table 4.6 and figure 4.6.

Table 4.6 Allele frequencies in 300 Egyptian individuals for the three loci.

D3S1358 locus had 7 alleles ranging from allele 13 to allele 19. Allele 16 had the highest frequency (0.276) and allele 13 had the lowest frequency (0.0083).

D3S1358 ALLELE	OBSERVED NO	ALLELE FREQUENCY	SD
13	5	0.0083	0.0037
14	51	0.0850	0.0114
15	155	0.2583	0.0179
16	166	0.2767	0.0183
17	139	0.2317	0.0172
18	75	0.1250	0.0135
19	9	0.0150	0.0050

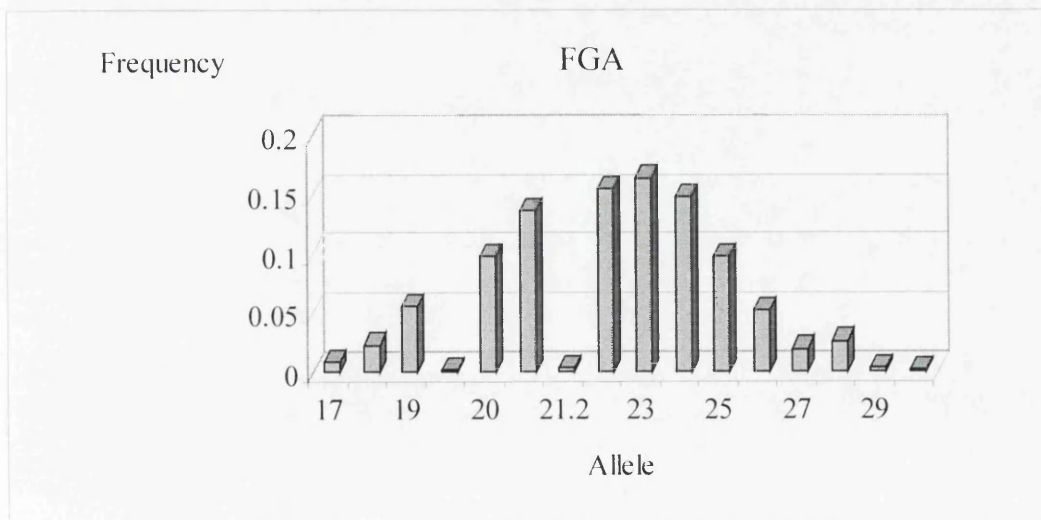
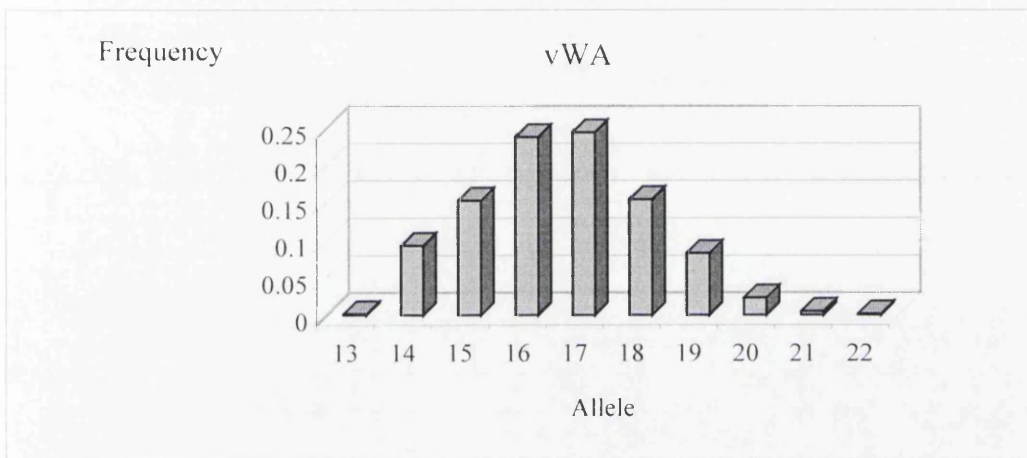
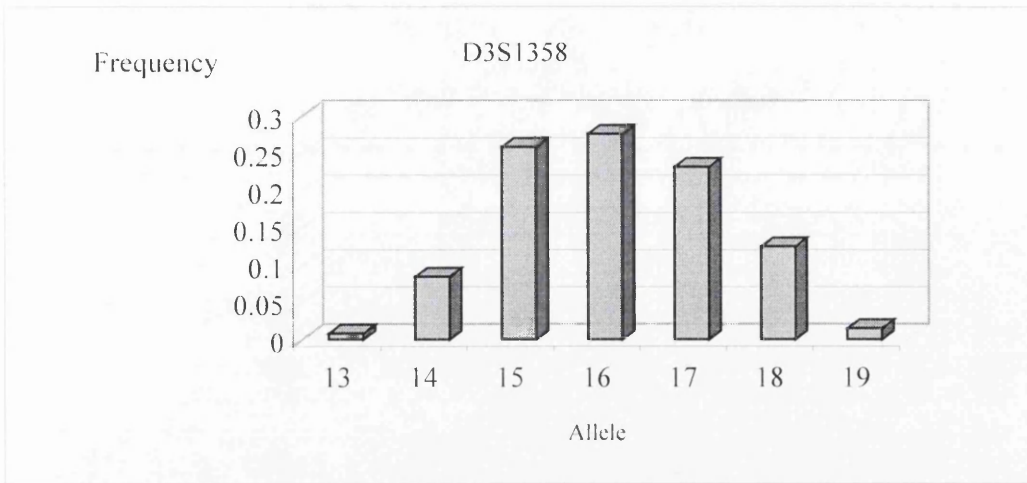
The vWA locus had 10 alleles ranging from allele 13 to allele 22. Allele 17 had the highest frequency (0.245). Alleles 13 and 22 had the lowest frequency (0.0017).

vWA ALLELE	OBSERVED NO	ALLELE FREQUENCY	SD
13	1	0.0017	0.0017
14	56	0.0933	0.0119
15	92	0.1533	0.0147
16	143	0.2383	0.0174
17	147	0.2450	0.0176
18	93	0.155	0.0161
19	50	0.0833	0.0113
20	14	0.0233	0.0062
21	3	0.0050	0.0029
22	1	0.0017	0.0017

The FGA locus had 16 alleles ranging from allele 17 to allele 30. There were two microvariants (19.2 and 21.2). Alleles 23 had the highest frequency (0.165) and alleles 19.2 and 30 had the lowest frequencies (0.0017).

FGA ALLELE	OBSERVED NO	ALLELE FREQUENCY	SD
17	5	0.0083	0.0037
18	14	0.0233	0.0062
19	34	0.0567	0.0094
19.2	1	0.0017	0.0017
20	59	0.0983	0.0122
21	82	0.1367	0.0140
21.2	2	0.0033	0.0023
22	93	0.1550	0.0148
23	99	0.1650	0.0152
24	89	0.1483	0.0145
25	59	0.0983	0.0122
26	32	0.0533	0.0092
27	12	0.0200	0.0057
28	16	0.0267	0.0066
29	2	0.0033	0.0023
30	1	0.0017	0.0017

Figure 4.6 Allele frequencies for the three loci.



4.7.3. Heterozygosity

The level of heterozygosity is a simple method to measure genetic variation. A marked decrease in the percent of heterozygosity indicates non-random mating of the examined population. All the three loci showed heterozygosity values > 80%, the highest value being observed in FGA locus (89%) followed by vWA locus (85.7%) and D3S1358 had a value of 82.3%. The expected heterozygosity was calculated for the three loci using GDA software. The observed heterozygosity was higher than the expected heterozygosity for the three loci. The standard deviation of the expected heterozygosity was calculated to determine the range of the expected heterozygosity using the equation (Edward *et al.*, 1992).

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

Where h is the heterozygosity frequency and n is the number of population. The difference between the observed and the expected heterozygosity was investigated by chi-square test showing no significant difference, as the p-value was 0.946. The results are shown in table 4.7.

4.7.4. Paternity Parameters

It was desirable to calculate the paternity parameters before implanting DNA markers into any legal system to determine the power of these markers in the paternity testing.

Table 4.7 Comparison between the observed and expected heterozygosity

The observed heterozygosity was higher than the expected heterozygosity for the three loci. The standard deviation of the expected heterozygosity was calculated to determine the range of the expected heterozygosity. There was no significant difference between the observed and the expected heterozygosity ($P = 0.946$).

LOCUS	OBSERVED	EXPECTED	SD
D3S1358	247	235	7.1
vWA	257	246	6.7
FGA	267	264	5.6

4.7.4.1. Paternity Index

Assuming that the man could not be excluded, the paternity index tells how many times more likely the alleged father is the true biological father compared to an unrelated man in the population of similar ethnic background (Brenner and Morris, 1990). The paternity index for the three loci ranged from 2.84 for D3S1358 locus to 4.55 for FGA locus. The three loci showed a combined paternity index of 44.78.

The probability of paternity can be calculated using the combined paternity index value. The probability of paternity is equal to the combined paternity index divided by the combined paternity index plus 1. The probability of paternity for the three loci was 97.82% assuming a prior probability of guilt of 50%.

4.7.4.2. Power of Exclusion

Power of exclusion is the probability that the alleged man can be excluded from being the 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 and Morris, 1990).

The FGA locus had the highest value of probability of exclusion (0.775) followed by vWA locus (0.708) while D3S1358 locus had the lowest value (0.642). The combined probability of exclusion was 0.9765.

4.7.5. Forensic Parameters

4.7.5.1. Matching Probability

A match probability is equivalent to the probability that a random person from the population, other than the suspect, will have the same DNA profiles. It is equal to the sum of the square of genotype frequencies of all alleles observed in the database (Jones, 1972).

The matching probability of the three loci ranged from 0.027 for the FGA locus to 0.089 for the D3S1358 locus. The combined matching probability for the three loci was 0.000151. The match probability can be expressed as the likelihood ratio, which is the reciprocal of matching probability, therefore the chance of occurrence of the same profile for the three loci was 1 in 6623 of the local population.

4.7.5.2. Discrimination Power

Power of discrimination is the probability that two individuals selected randomly from the population will differ from the allele in question (Jones, 1972). It is equal to one minus the matching probability. The higher the value of the power of discrimination, the better the locus used for individual identification.

All loci were highly polymorphic. The FGA system was the most informative one followed by the vWA locus, giving a probability of discrimination of 0.97 and 0.93 respectively. The probability of discrimination for the D3S1358 locus was 0.89 and the

combined discrimination power for the three loci was 0.9998. All the paternity and forensic parameters investigated are described in table 4.8.

4.7.6. Hardy-Weinberg Equilibrium

It is necessary to investigate that the DNA forensic markers follow Mendelian inheritance and do not deviate from Hardy-Weinberg equilibrium. HWE depends mainly on the existence of randomly mating populations of infinite size. These characteristics are never available in natural populations (Hammond *et al.*, 1994), however large populations do not tend to deviate significantly from the HWE. If independence of alleles is assumed both within and between loci in the sampled population, HWE can be used to estimate the expected genotype frequencies from the allele frequencies.

4.7.6.1. Exact Test

Exact test (probability test) was introduced by Fisher (1935). This test assumes that the null hypothesis is true and calculates the probability of the observed outcome. A probability value less than 5% means rejection of the null hypothesis and indicates that the population is not in equilibrium.

4.7.6.2. Chi-Square Test

The chi-square goodness-of-fit test compares the observed and the expected values in all categories, squares the differences and divides by the expected value to give the largest difference between them.

Table 4.8 Paternity and Forensic Parameters for the Three Loci.

The FGA locus had the highest value in all paternity and forensic parameters while the D3S1358 locus had the lowest values. Paternity Index (PI), Probability of Paternity (PE), Matching Probability (MP) and Power of Discrimination (PD).

	PI	PE	MP	PD
D3S1358	2.820	0.642	0.089	0.911
vWA	3.49	0.708	0.063	0.937
FGA	4.55	0.775	0.027	0.973
Combined	44.78	0.9765	0.000151	0.9998

Large departures from expectation in both directions cause rejection of the null hypothesis and indicate deviation from equilibrium.

Both tests were performed to test the deviation of the observed genotype frequencies from those expected under HWE.

There was agreement between the observed and expected genotype values under Hardy-Weinberg equilibrium for the three loci. The exact test for independence of the alleles within individuals was performed for each locus, with the significant levels ($p < 0.05$) and 2000 permutations of alleles (Guo and Thompson *et al.*, 1992). X^2 value of Chi square test and p value of exact test are shown in tables 4.9 and 4.10.

4.7.7. Independence of Loci

The product rule used to calculate the significance of matched DNA profiles is based on the assumption that the allele frequencies that each individual's profile is composed of statistically independent random alleles from a common gene pool.

Association between frequencies at two loci have been investigated by comparing two-locus genotype frequencies to the products of the corresponding allele frequencies, assuming independence of alleles both within and between loci. The result showed no evidence for correlation between the alleles at any of the pair of loci, thus multiplication of an individual locus profile frequency could be carried out to accumulate statistical information across loci. These p-values are shown in table 4.10.

Table 4.9 X² value of Chi-square test

The difference between the expected and the observed genotypes for the three loci was tested by using Chi-square test and there was no significant deviation from HWE for the three loci ($p > 0.05$).

Locus	X ²	df	p
D3S1358	14.6982	21	0.8378
vWA	34.8186	45	0.5248
FGA	76.3363	120	0.9993

Table 4.10 p-value of exact test

Deviation from HWE was tested by using exact and there was no significant deviation observed at any locus of the three loci. Association between frequencies at two loci has been also investigated. The result showed no evidence for correlation between the alleles at any of the pair of loci ($p > 0.05$).

	D3S138	vWA	FGA	D3S138 & vWA	D3S138 & FGA	vWA & FGA	D3S1358 & vWA & FGA
P	0.7150	0.1605	0.7015	0.2915	0.4795	0.6915	0.316

4.8. COMPARISON BETWEEN POPULATIONS

4.8.1. Introduction

There was no detectable deviation from the HWE for the three loci in the El-Minia population. Therefore comparisons were performed with other populations to investigate the possibility for using published data for other loci in Egypt.

4.8.2. Comparison with Cairo Population

Comparison was performed for the three loci to examine the heterogeneity between the Egyptians studied in El-Minia and the Egyptians from Cairo (Klitschar *et al.*, 1999). The allele frequencies of both populations are shown in table 4.11. Most of the allele frequencies for Cairo population were within the range of the standard deviation of allele frequencies for El-Minia population.

To detect the presence of significant difference between the compared allele frequencies and to examine if it was reasonable to pool both of the results of the two populations for the three loci, the two-way RxC contingency table test was performed to calculate X^2 value (Power Instat 2.03 programme). The results showed that there was no significant difference for D3S1358 and vWA loci. However, there was a significant difference between the two populations at the FGA locus ($P = 0.0024$) (Table 4.12).

Table 4.11 Allele frequencies of the two Egyptian populations for the three loci.

The number of individuals from Cairo was 140 individuals and 300 individuals from El-Minia city. The most common allele was different between the two Egyptian populations for both D3S1358 and FGA loci while allele 17 was the most common allele at the vWA locus in both populations.

Allele	D3S1358		VWA		FGA	
	Cairo	Minia	Cairo	Minia	Cairo	Minia
13	0.007	0.008		0.002		
14	0.046	0.085	0.082	0.093		
15	0.296	0.259	0.096	0.153		
16	0.243	0.277	0.261	0.238		
17	0.236	0.232	0.271	0.245		0.008
18	0.161	0.125	0.179	0.155	0.018	0.023
19	0.011	0.015	0.082	0.083	0.093	0.057
19.2						0.002
20			0.025	0.023	0.095	0.098
20.2					0.011	
21			0.004	0.005	0.164	0.137
21.2					0.016	0.003
22				0.002	0.132	0.155
22.5					0.004	
23					0.161	0.1615
24					0.179	0.148
25					0.057	0.098
25.2					0.004	
26					0.025	0.053
27					0.011	0.020
28					0.011	0.027
29						0.033
30						0.002
31					0.004	

Table 4.12. X^2 value for the compared allele frequencies for Cairo and Minia populations.

The comparison of the allele frequencies revealed a significant difference for the FGA locus only ($p < 0.05$).

	X^2	df	p
D3S1358	7.756	6	0.2565
vWA	7.065	8	0.5296
FGA	40.987	19	0.0024*

4.8.2.1. *F-Statistics*

To investigate the presence of substructure within the Egyptian population, the genotypes of both populations were compared using F-stats test, which includes F_{ST} (co-ancestry, which is the correlation of genes between individuals within populations, F (the overall inbreeding coefficient, which is the relation between genes within individuals) and f (degree of inbreeding within individuals within populations).

The result produced low F-stats values indicating that there was no significant substructure in the Egyptian populations and that mating was also effectively random. The results are shown in table 4.13. When bootstrapping over loci was performed, it showed that inbreeding coefficients f and F were not statistically significantly different from zero. Since the 95% bootstrap confidence interval for F_{ST} was close to zero, and the co-ancestry coefficient was not significant, the conclusion is that a non-significant amount of divergence had occurred between the two populations (Table 4.14).

4.8.3. Comparison with Other Populations

Comparisons of the results obtained in this study with other data for the three loci were performed including some Arabic populations such as a Moroccan population for vWA and FGA loci (Brandt-Casadevall *et al.*, 2000 and Rolf *et al.*, 1998), Yemenian population for vWA and FGA loci (Klitschar and Al Hammadi, 1998) and Qatari population for the vWA locus only (Sebetan and Hajar, 1998).

Table 4.13 F-stats values.

Comparison was performed between the two Egyptian populations (El-Minia and Cairo) by using F-Statistics. The F-stats values indicating that there was no significant substructure in the Egyptian populations and that mating were also effectively random.

EGYPTIANS	
F_{ST}	0.0011
f	-0.0046
F	-0.0058

Table 4.14 Bootstrapping over the three loci

Bootstrapping over loci was performed with 95% confidence interval. The F_{ST} was close to zero, thus the coancestry coefficient was not significant and thus a non-significant amount of divergence had occurred between the two Egyptian populations.

	LOWER BOUND	UPPER BOUND
F_{ST}	0.00048	0.0019
f	-0.0367	0.0207
F	-0.0356	0.0225

In addition, one population from each major ethnic group (Caucasian, Mongoloid, and Negroid) was chosen and compared with the results obtained in this study. The populations that were chosen for the comparison were the Germany population representing the Caucasian population (Zehner, 1998), a Japanese population representing Mongoloid population (Yamamoto *et al.*, 1998), a Zimbabwean population representing Negroid population (Budowle *et al.*, 1997). The Spanish population was also examined as a southern European Caucasian (Andres *et al.*, 1998). The statistical values of heterozygosity, power of discrimination and probability of exclusion for these populations were compared (Table 4.15).

The FGA locus had the highest rate of heterozygosity, the power of exclusion of paternity and the power of discrimination in all compared countries except Spain where the vWA locus had the highest rate in all categories and Germany where the vWA locus had the highest rate in the power of discrimination.

To detect the degree of variation of the DNA profile frequencies of the Egyptian population was compared to the other mentioned population databases. The two-way RxC contingency table test was performed to calculate X^2 value (Power Instat 2.03) for the compared allele frequencies. There were significant differences between the Egyptian and Moroccan population at the FGA locus and with Yemenian population at the vWA locus. There were also significant differences between the Egyptian and other populations regarding the three loci except with the Spanish population. The X^2 values are shown in table 4.16.

Table 4.15 Comparison of the statistical parameters between the Egyptian population and other populations.

The FGA locus had the highest rate of heterozygosity, the power of exclusion of paternity and the power of discrimination in all compared counties except Spain where vWA locus had the highest rate in all categories.

	Heterozygosity			Power of Exclusion			Power of Discrimination		
	D3	vWA	FGA	D3	vWA	FGA	D3	vWA	FGA
Egyptian (Minia)	0.82	0.89	0.857	0.64	0.708	0.775	0.91	0.937	0.973
Moroccan	-	0.82	0.84	-	0.60	0.72	-	0.92	-
Yemenian	-	0.72	0.83	-	0.591	0.704	-	0.924	0.956
Qatari	-	0.81	-	-	0.601	-	-	0.929	-
Germany	0.86	0.81	0.86	0.61	0.63	0.72	0.93	0.94	0.93
Spain	0.78	0.88	0.80	0.59	0.77	0.62	0.92	0.96	0.940
Japan	0.68	0.77	0.814	-	-	-	0.85	.0916	0.96
Zimbabwe	0.71	0.80	0.91	0.50	0.63	0.77	0.89	0.94	0.98

Table 4.16 X² value for the compared allele frequencies.

There was a significant difference between the Egyptian and the Moroccan population at the FGA locus (0.0388*). Another significant difference occurred between the Egyptian and the Yemen population at the vWA locus (0.0170*). There were also significant differences between the Egyptian and other populations regarding the three loci except the Spanish population. (*There was a significant difference).

D3S1358	X2	df	p
Egypt and Germany	21.413	8	0.0061*
Egypt and Spain	11.018	7	0.1379
Egypt and Japan	36.418	6	< 0.0001*
Egypt and Zimbabwe	177.42	8	< 0.0001*

vWA	X2	df	p
Egypt and Moroccan	8.2875	7	0.5055
Egypt and Yemenian	18.617	8	0.0170*
Egypt and Qatar	11.933	8	0.1542
Egypt and Germany	16.191	8	0.0397*
Egypt and Spain	4.240	8	0.8349
Egypt and Japan	78.582	8	< 0.0001*
Egypt and Zimbabwe	25.897	9	0.0021*

FGA	X2	df	p
Egypt and Moroccan	28.561	17	0.0388*
Egypt and Yemenian	21.575	16	0.1574
Egypt and Germany	51.538	17	< 0.0001*
Egypt and Spain	26.583	18	0.1156
Egypt and Japan	33.672	17	0.0117*
Egypt and Zimbabwe	36.823	17	0.0036*

4.9. DISCUSSION

4.9.1. PCR Optimisation

Three tetranucleotide repeat STR loci were optimised and subsequently 300 Egyptian DNA samples were typed to obtain the allele and genotype frequencies. The analysis of the three loci was used to examine the presence of any deviation in the Egyptian population from the HWE and investigate the usefulness of these loci for paternity testing and forensic cases. The primer for the D3S1358 locus was a modified primer of Li *et al.*, (1993) and the length of the published alleles ranged from 97 bp to 121 bp. However, the products of the vWA and D3S1358 loci overlapped. This is explained by a wrong sizing of the allele by the manual typing used by Li *et al.*, (1993). Therefore the vWA locus was amplified as a singleplex PCR and D3S1358 and FGA loci were amplified as a duplex. PCR component concentrations and time-temperature profiles were optimised for each locus individually and thereafter for the duplex PCR. Co-amplification for D3S1358 and FGA loci in a duplex PCR was found to be successful. There was 100% success in amplification of all three loci.

4.9.2. Sizing of the Allele

The maximum SD for D3S1358 and vWA was 0.1528 that is below 0.5 bp when using ± 3 SD, thus allele sizing by using ROX internal standard was sufficiently accurate. The FGA locus had the largest standard deviation (0.188 bp). This range of the standard deviation for the FGA locus was higher than the values obtained by (Ovington *et al.*, 1997). An allelic ladder was constructed for each locus. Allelic

ladders were run on the same electrophoretic gel to allow direct comparison with the alleles of all samples. The system allowed unambiguous alleles typing for the three loci and allowed the distinction between the tetra-repeat alleles and the two variants observed at the FGA locus. A locus specific allelic ladder provides confidence for inter-and intra-laboratory comparison that may have different electrophoretic systems or size standard markers (Sprecher *et al.*, 1996).

The alleles for the three loci identified in this study differed in size from those determined by other published studies as a result of using new PCR primer sequences. In accordance with the recommendation of the DNA Commission of the International Society of Forensic Haematogenetics (1992, 1994), alleles of the loci were designated according to the number of repeat units in each allele. Therefore some samples were amplified using the new primers used in this study and the commercial kit (AmpF ℓ STR Blue Kit) (Applied Biosystem; CA, USA).

4.9.3. Paternity and Forensic Parameters

The three loci studied are one of a series of polymorphic systems that are being promoted for use in identification studies. The combination of the three loci is highly discriminating in Egyptian population, the combined power of discrimination exceeded of 0.9998. The powers of discrimination were similar to those reported in other studies (Miscicka-Sliwka *et al.*, 1999 and Wallin *et al.*, 1998).

It was recommended that any marker used in paternity testing should have a power of exclusion of $\geq 99\%$ that is a match or non-excluding result in paternity testing has to exclude $\geq 99\%$ of the male population from being the biological father (Hammond *et al.*, 1994). The power of the combined exclusion of paternity after typing the three loci was 97.65%. Thus, the three loci did not provide a sufficient power of exclusion for paternity testing.

For inclusion purpose, the combined paternity index is 44.78 with the probability of paternity of 97.82 % that means it is only 44 times more likely the man is the biological father compared to an unrelated man. In studies of other populations, the combined paternity index was higher than the reported one for this study for the same loci (Miscicka-Sliwka *et al.*, 1999 and Wallin *et al.*, 1998). Although, the combined paternity index for these loci was higher than others such as CTTv Multiplex including (CSF1PO, TPOX, TH01 and vWA) that had paternity index ranging from 10.51 in Hispanic-American to 29.39 in African-Americans. It was also higher than GammaSTR™ Multiplex including (D16S539, D7S820, D13S317 and D5S818), which had a highest paternity index (30.40) in Hispanic-Americans (Lins *et al.*, 1998).

Despite the low value of paternity index for the three loci in the Egyptian population, it is significantly higher than one obtained from the ABO blood group, which is the system currently used in Egypt for paternity testing (PI = 6.5 and probability of paternity = 0.866) (Sensabaugh, 1982). In addition, the two systems can be used together to increase the value of the paternity index and the probability of

paternity. However, it was desirable to study other STR loci, which would increase both of the paternity index and the probability of exclusion of paternity.

4.9.4. HWE Tests

The exact test with 2000 permutations was performed to test deviation from the HWE. The data showed no deviation from HWE and linkage equilibrium between the three loci. Previous presentation of data for different populations for the three loci showed similar results as there were only slight deviations or none at all from HWE and linkage equilibrium (Cabrero *et al.*, 1995 and Miscicka-Sliwka *et al.*, 1999). Deviation from HWE has been reported for locus HUMvWA for a British population (Drozd *et al.*, 1994) and a Russian population (Sajantila *et al.*, 1994). The result indicated that neither the mixture of populations nor the relatively high levels of relative intermarriage caused the Egyptian population to deviate from HWE.

Pairwise comparison of multigenotype data of the three loci did not show random association of alleles between any of the loci, therefore the three loci are in linkage equilibrium. This result was expected since the loci are located on different chromosomes. Thus, the product rule for combining genotype frequencies across unlinked loci is valid for use in these loci. The use of the product rule of allele frequencies is preferred as it gives a smaller standard deviation than that from the product of genotype frequencies (Evetts and Weir, 1998).

4.9.5. Comparison of the Two Egyptians Studies

A RxC contingency table test was used to compare the allele frequencies between the Egyptian population from El-Minia and Cairo. There was no significant difference between the allele frequencies for the two Egyptian populations for D3S1358 and vWA, so they can be pooled to create a population with an increased number of individuals. However, there was evidence of heterogeneity between the two Egyptian populations for the FGA locus as there was significant difference ($p=0.0024$). The FGA locus for the Cairo population showed a deviation from HWE. When the data for both populations was pooled together for the FGA locus, there was no deviation from HWE. Therefore the explanation for this difference may be due to statistical artifact due to multiple testing of the same population sample (Klitschar *et al.*, 1999).

Despite the complex composition of the Egyptian population, the presence of substructure was not detected, as the F_{ST} value was very low of (0.0011). Nichols and Balding (1991) and Balding and Nichols (1994) recommended the routine use of relatively high F_{ST} values (0.05), while Brookfield (1991) stated that 0.05 is high and using 0.01-0.02 is more reasonable. Morton *et al.*, (1993) also supported that 0.05 is a high value and this level of inbreeding has only observed in a few isolates and is unrealistic. Morton (1992) reported that most of the studied populations had F_{ST} value less than 0.01. Additional study comparing the Egyptian populations using another five STR loci also showed no significant difference at any locus (Amany press).

4.9.6. Comparison with Other Populations

Significant differences were absent between the Egyptian population and two Arabic populations (Morocco and Qatar for vWA locus). There was, however a significant difference when the El-Minia population was compared with Morocco at the FGA locus. On the other hand there was a significant difference between the Egyptian population and Yemen population for the vWA locus but not for the FGA locus. In a previous study there were no significant differences between different Arabic populations for HLA-DQA1 and D1S80 (Alkhyat *et al.*, 1996 and Hayes *et al.*, 1995), or for FXIII3 and FES loci (Klitschar *et al.*, 1998).

By comparing the Egyptian population and the German (Caucasian) population, significant differences were observed for the three loci; the p value ranging from 0.0397 for vWA to < 0.0001 for the FGA locus. Although, as the p-value was close to 0.05 (0.04), the difference for the vWA locus was not highly significant. Another study reported no significant difference between German and Qatar populations at the vWA locus (Sebetan and Hajar 1998). Highly significant differences were found for the three STRs tested in this study between the Egyptian population and both the Japanese and Zimbabwean populations. These results are consistent with other data and confirm that the presence of variation is greater between ethnic groups than between different populations derived from the same ethnic origins (Weir, 1992).

Interestingly there was no evidence of population heterogeneity between the Egyptian and the Spanish populations for the three loci. The explanation may be due to

the collection of Spanish samples from Andalusia, in southern Spain, which was occupied by Arabs from 700-1400 (Jarreta *et al.*, 1999).

In conclusion, there was neither deviation from HWE nor substructure in the Egyptian population. Therefore no specific reasons to forbid the use of the same established DNA database to the paternity testing in Egypt.

CHAPTER 5: HYPERVARIABLE LOCI

5.1. INTRODUCTION

The analysis of the three simple STR loci demonstrated that there was no substructure in the Egyptian population and the simple loci were under HWE. However, using these loci for paternity testing was limited as the paternity index and the probability of exclusion were relatively low. Additional STR loci were selected for study. Among the known STR loci, the hypervariable loci are the most informative and polymorphic loci. They exhibit high degrees of length as well as sequence polymorphisms (Moller *et al.*, 1995). An Egyptian database for the highly polymorphic STRs would be a valuable tool for paternity testing and other forensic purposes. Three loci (D11S554, ACTBP2 and APOA11) were chosen for this study. Loci were selected based on their predicted high level of discriminating power, as indicated by previously published data. The allele frequency for any locus did not exceed 15% in most previous studies. Moreover, the combined matching probability for the three loci was 5×10^{-7} and the combined paternity exclusion was 99.77% in previous study (Dupuy and Olaisen, 1997).

5.2. DESCRIPTION OF LOCI

5.2.1. D11S554 Locus

The human D11S554 locus is located at 11p11.2-12 immediately to the 3' of an Alu repeat (Phromchotikul et al., 1992). It contains a repeat block of AAAG with an intervening sequence (Adams *et al.*, 1993). The D11S554 locus is highly polymorphic with more than 36 known alleles (Dupuy and Olaisen, 1997). Nagai *et al.*, (1998) reported the presence of 46 alleles in the Japanese population. The D11S554 locus is useful in paternity testing and human identification as the heterozygosity rate is more than 84% (Phromchotikul et al., 1992). Sequencing of different alleles of the D11S554 locus revealed that there are four distinct families of the repeat pattern (Adams *et al.*, 1993). The four types are described in figure 5.1.

5.2.2. HUMACTBP2 (SE33) Locus

The HUMACTBP2 locus is located on chromosome 6 (Polymeropoulos *et al.*, 1992). It is one of the most polymorphic and informative STR loci (Moller et al., 1995). The power of discrimination can be over 99% (Wiegand *et al.*, 1993). In contrast with other STR loci such as HumFGA, HumvWA, HumFES/FPS, D18S51 and D8S1179, which have also structural variation (Barber *et al.*, 1995, 1996; Barber and Parkin 1996), the ACTBP2 locus exhibits higher degrees of both length and sequence polymorphisms (Liu *et al.*, 1997). Sequence variations such as insertions, deletions and transitions in the flanking regions and the repeat units have been observed (Brinkmann et al., 1996).

130A

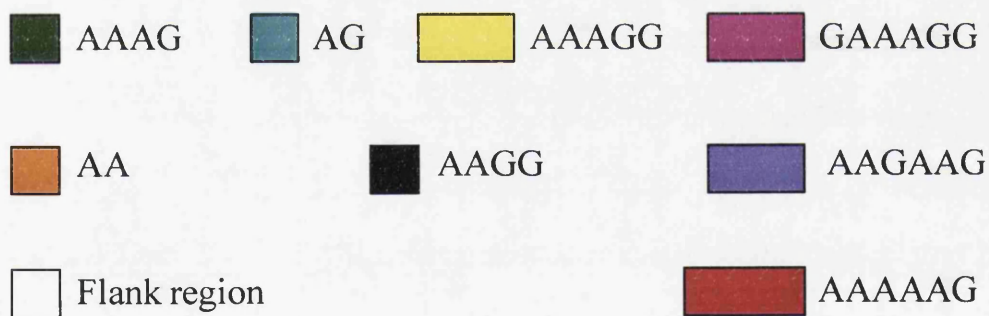
Figure 5.1 Schematic examples for the four different classes of the sequence of the D11S554 alleles.

(IA) This type shows a regular 4 bp repeat AAAG with (AG)₄, AAAGG and a hexamer GAAAGG unit.

(IB) This type contains the same structure as IA except that it does not contain the hexamer unit instead; it contains AA, AAGG and AAGAAG units.

(IIA) This type shows a regular 4 bp repeat AAAG with (AG)₉, and a hexamer AAAAAG unit near the 5' end.

(IIB) This type contains the same structure as IIA except that it does not contain the hexamer unit.



More than 100 alleles have been reported recently, many of which are due to sequence variations (Rolf *et al.*, 1997). Sequencing of more than 250 alleles showed three types of structure variation for the ACTBP2 locus (Moller and Brinkmann, 1994). The three types are described in figure 5.2.

5.2.3. HUMAPOAI1 Locus

The HUMAPOAI1 locus is located at 11q23qter (Kimpton *et al.*, 1993). It is a polymorphic locus with more than 35 alleles. Although, the APOAI1 is a tetranucleotide locus, there are a large number of alleles. Sequencing of 10 alleles of APOAI1 locus showed that all alleles were a relatively regular (AAAG) and 9 of them contained an insertion of (AAGA) sequence (Dupuy and Olaisen, 1997) (Figure 5.3). However, mutations in the central, 3' and 5' flanking regions of the alleles cause size differences of one bp (Dupuy and Olaisen, 1997).

5.3. PRIMER DESIGN

published primers had been used to amplify to D11S554 and APOAI1 as a duplex and SE33 by itself (Dupuy and Olaisen, 1997). Multiplex PCR provides a benefit in conserving samples and time. Therefore, a new primer was designed for the APOAI1 locus to allow amplification of the three loci as a multiplex. The study of the sequence of the three loci obtained via Genebank by the use of accession numbers was performed to identify suitable primer sites. Several primer designs were examined before a new primer was selected for the APOAI1 locus.

132A

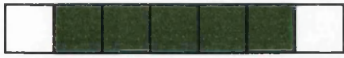
Figure 5.2 Schematic examples for the three different classes of the SE33 alleles.

(A) The lower allelic range shows a common repeat structure with the regular 4 bp repeat AAAG.

(B) The intermediate allelic range that contains an insert hexanucleotide ARAAAG in addition to the regular 4 bp repeat. This irregular unit occurs only once in each fragment but at different positions.

(C) The upper allelic range revealed two hexamer units at different sites within the repeat region leading to a highly complex sequence structure.

A





B



C



 Flank region

 AAAG

 ARAAAG

133A

Figure 5.3 Schematic examples for the two different classes of the APOAI1 alleles.

(A) The first type shows a common repeat structure with the regular 4 bp repeat AAAG.

(B) The second type contains AAGA unit in addition to the regular 4 bp repeat AAAG.



 Flank

 AAAG

 AAGA

The new primer was complementary to the flanking regions far from the published primers causing an increase in the length of the APOAI1 alleles. The melting temperature was calculated as the sum of 4 °C for each G/C and 2 °C for each A/T and was near to the calculated melting temperature for D11S554 and SE33.

The affinity of the primers to each other and to DNA sequences outside the target template was tested with the help of computer programs Primer Pick (Rozen and Skaletsky, 1996 and 1997) and Williamston (www.Williamstone.com). Because of the overlap between the allele sizes of both D11S554 and SE33, two different fluorescent dyes were used. The 5' end of the forward primer for both D11S554 and APOAI1 was labelled with the JOE fluorescent dye and the 5' end of the forward primer for SE33 was labelled with FAM fluorescent dye. The sequences of the three hypervariable loci with the attached primers are shown in appendix 2.

5.4. OPTIMIZATION OF THE PCR SYSTEM

The amplification conditions for a singleplex PCR for each locus was optimised following the same steps as described in 4.5 section. The multiplex PCR was performed by combining equimolar amounts of each primer for the three loci using a singleplex PCR programme with the lower annealing temperature. Preferential amplification occurred of the D11S554 locus and weaker results were obtained for the ACTBP2 and APOAI1 loci. Decreasing the concentration of the D11S554 primer and increasing the concentration of both of ACTBP2 and APOAI1 primers achieved

comparable intensities of amplified products (Figure 5.4). Several experiments were then performed to optimise all the PCR parameters for the multiplex by testing a range of different annealing temperatures, MgCl₂ concentration and extension temperature. Once the optimal condition was determined for a single component, the other conditions were tested sequentially until each component was optimised relative to the other in the reaction mixture. The next step was to re-test each of the optimised parameters several times to ensure reproducibility of the results. The optimised PCR programmes for singleplex and multiplex are described in table 5.1.

5.5. CONSTRUCTION OF ALLELIC LADDERS

The use of a sequenced allelic ladder as a reference for allele designation is important for accurate analysis in forensic studies (Olaisen *et al.*, 1998). The measurement of the size of the alleles for the hypervariable loci against the internal standard ROX500 is not sufficient for confident typing of alleles differing in size by only 1 bp (Dupuy and Olaisen, 1997). The use of a specific sequenced ladder was recommended when the hypervariable loci are to be analysed (Malmgren *et al.*, 1998). These ladders are constructed from the relevant loci in combination with denaturing gels. This helps to decrease electrophoretic mobility shifts in the analysis of hypervariable loci because some variants differ by as little as 1 bp (Bar *et al.*, 1997). Therefore, construction of a specific allelic ladder for each locus of the three hypervariable loci was performed using 10 sequenced samples provided by Professor Dupuy from Norway. The sequenced samples contained a total of 15 different alleles for SE33, 14 different alleles for D11S554 and 10 different alleles for APOA11.

136A

Figure 5.4 Optimisation of multiplex for D11S554, SE33 and APOA11 loci.

1-The upper electrophoretogram shows preferential amplification of D11S554 locus in the multiplex PCR when the same concentration of the primers of D11S554, SE33 and APOA11 loci were used in the PCR.

2-The lower electrophoretogram shows optimised PCR products of the multiplex PCR obtained by using different primer concentrations of the three loci.

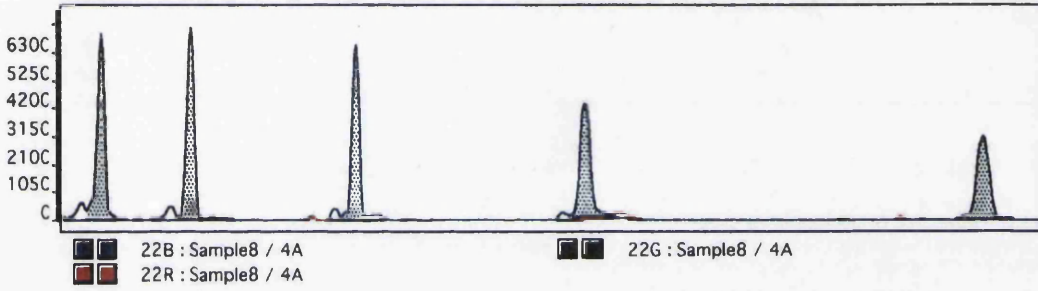
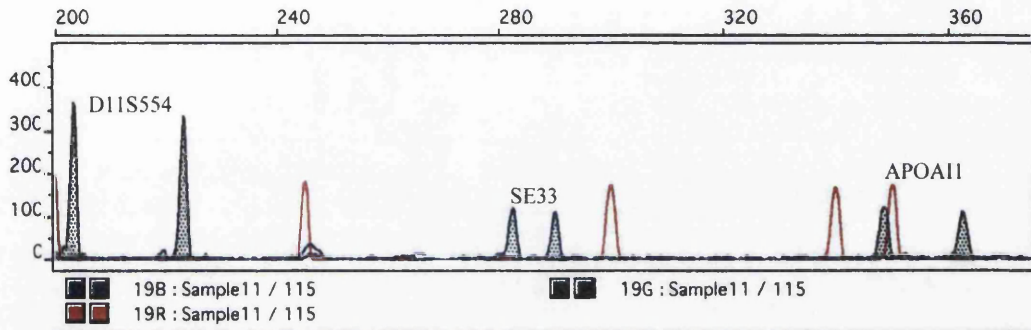


Table 5.1 Optimised parameters for single and multiplex PCR.

The concentration of the each primer of the three loci was optimised to yield approximately equal products all loci.

PARAMETER	D11S554	SE33	APOAI1	MULTIPLEX
Annealing Temp.	62 °C	60 °C	62 °C	60 °C
MgCl ₂	2 mM	1.5 mM	2 mM	1.5 mM
Primer Conc.	0.2 µM	0.25 µM	0.25 µM	0.15 µM D11 0.25 µM SE33 0.3 µM APOAI1
Extension Temp.	72 °C	72 °C	72 °C	72 °C

To obtain permanent stock of the ladders, all alleles for each locus were cloned into a plasmid. This created an inexhaustible stock of template to generate new ladders.

5.5.1. Cloning of the Ladders

5.5.1.1. DNA Purification from Gel Bands

The first step to clone the samples was obtaining a high quantity and quality of DNA from purified PCR products. The samples were amplified as a singleplex for each locus. The PCR products were run on an agarose gel and the alleles were cut and purified using GFX™ PCR DNA and Gel Band Purification Kit (Pharmacia Biotech) as described in the “Material and Methods” section. All DNA samples were successfully purified and the quantity of purified DNA was estimated on a 2% agarose gel in comparison to a standard marker (Figure 5.5). Quantification is important to estimate the amount of purified DNA that has to be added for the ligation reaction. Any sample that failed to yield a high quantity of purified DNA was subjected to another singleplex PCR and purification until all samples yielded enough quantity of DNA for ligation.

5.5.1.2. Ligation and Transformation

Ligation and transformation were performed as described in the “Materials and Methods”. A rapid PCR screen was performed on the resulting white colonies as well as the colonies which had pale blue centres. Electrophoresis of these PCR products through a 2% agarose gel identified the colonies that contained inserts (Figure 5.6).

Figure 5.5 Quantity of the purified PCR products.

Agarose gel electrophoresis of purified PCR products amplified for the highly hypervariable loci for the ten Norway samples. 5 μ l of the purified products were loaded on 2% agarose gel electrophoresis to estimate the quantity of the purified PCR products.

Lane 1	5 μ l of Hyperladder IV.
Lane 2	Failed purification from PCR product of sample 1
Lane 3	5 μ l of purified DNA from PCR product of sample 2.
Lane 4	5 μ l of purified DNA from PCR product of sample 3.
Lane 5	5 μ l of purified DNA from PCR product of sample 4.
Lane 6	5 μ l of purified DNA from PCR product of sample 5.
Lane 7	5 μ l of purified DNA from PCR product of sample 6.
Lane 8	5 μ l of purified DNA from PCR product of sample 7.
Lane 9	5 μ l of purified DNA from PCR product of sample 8.
Lane 10	5 μ l of purified DNA from PCR product of sample 9.
Lane 11	5 μ l of purified DNA from PCR product of sample 10.

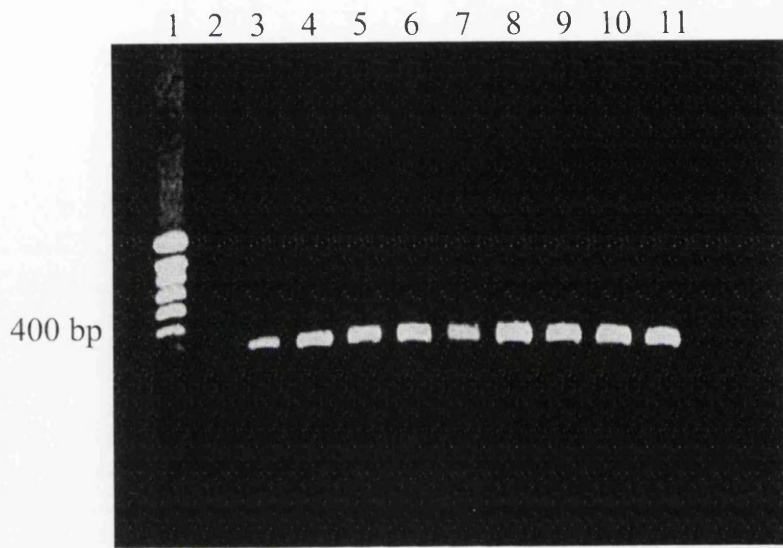
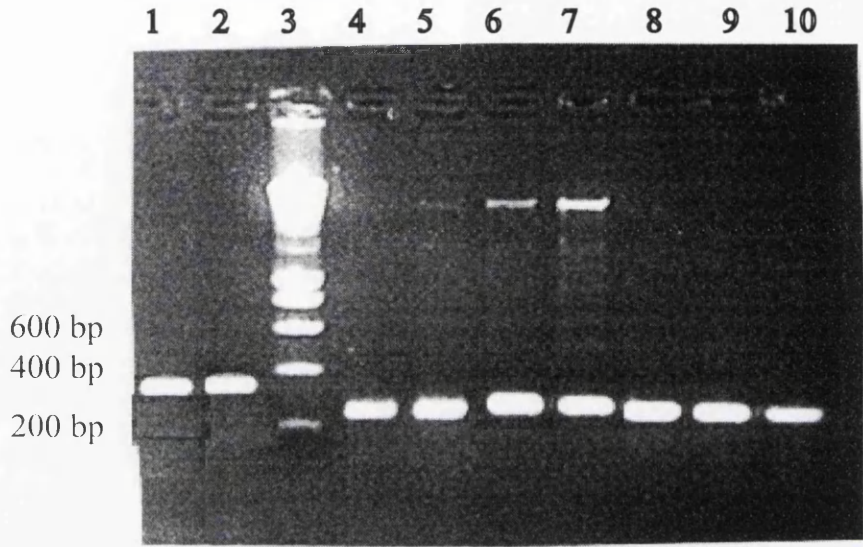


Figure 5.6 Agarose gel electrophoresis of the PCR screen

Cells from 9 white colonies were used in the PCR screen. Each colony was amplified for the same hypervariable locus of the insert. Aliquots of the PCR reactions were analysed through a 2% agarose gel containing ethidium bromide (lanes six and seven contained non-specific bands).

- | | |
|---------|--|
| Lane 1 | One colony contained plasmid expected to carry the APOAI1 insert. |
| Lane 2 | One colony contained plasmid expected to carry the APOAI1 insert. |
| Lane 3 | 5 μ l of Hyperladder IV. |
| Lane 4 | One colony contained plasmid expected to carry the SE33 insert. |
| Lane 5 | One colony contained plasmid expected to carry the SE33 insert. |
| Lane 6 | One colony contained plasmid expected to carry the SE33 insert. |
| Lane 7 | One colony contained plasmid expected to carry the SE33 insert. |
| Lane 8 | One colony contained plasmid expected to carry the D11S554 insert. |
| Lane 9 | One colony contained plasmid expected to carry the D11S554 insert. |
| Lane 10 | One colony contained plasmid expected to carry the D11S554 insert. |



Using this information, cells on the grid plate corresponding to the positive colonies were picked using a sterile pipette tip and incubated overnight in LB-Amp broth.

5.5.1.3. Plasmid Miniprep

Plasmid minipreps were performed on the overnight culture using GFX™ Micro Plasmid Prep Kit as described in the “Material and Methods” section. To confirm the presence of insert, digestion of the plasmid with *Sac* II and *Spe* I restriction enzymes was performed to allow the insert to be digested out of the plasmid and resolved on a 2% agarose gel. The samples show a large band corresponding to the cut PGMT vector and a small band of the expected length of one of the three loci (Figure 5.7). The plasmid was run on a 2% agarose gel to quantify the amount of DNA.

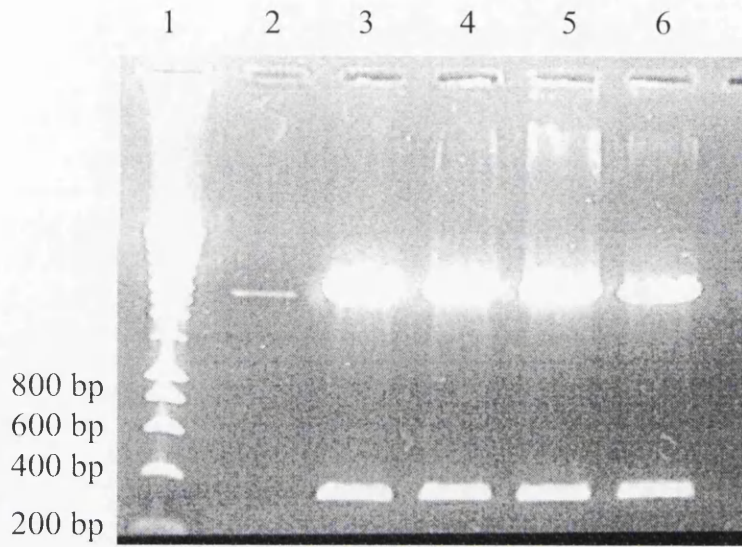
5.5.2. Amplification of the Plasmid

After quantification of the plasmid DNA, the DNA was diluted and amplified as a singleplex for each clone. The PCR products were run on a 373XL automated sequencer to determine that the appropriate alleles had been cloned. This process was repeated until all alleles were cloned except one allele for D11S554 locus and two alleles for the APOA11 locus. The amplified PCR from the plasmids were mixed together to construct a specific ladder for each locus. The allelic ladders were run with all samples (Figures 5.8, 5.9 and 5.10).

142 A
Figure 5.7 Agarose gel electrophoresis of the PCR products from putative PGMT recombinants containing fragments of APOA11 insert.

Recombinant vectors containing putative APOA11 fragments were digested with *Sac* II and *Sap* I restriction enzymes. This allows the insert to be digested out of the plasmid and the products were resolved on a 2% agarose gel containing ethidium bromide. The samples show a large band corresponding to the cut PGMT vector and a small band of the expected length of the APOA11 locus.

- | | |
|--------|---|
| Lane 1 | 5 μ l of Hyperladder IV. |
| Lane 2 | One colony did not contain the APOA11 fragments insert. |
| Lane 3 | One colony contained the APOA11 fragments insert. |
| Lane 4 | One colony contained the APOA11 fragments insert. |
| Lane 5 | One colony contained the APOA11 fragments insert. |
| Lane 6 | One colony contained the APOA11 fragments insert. |



143A

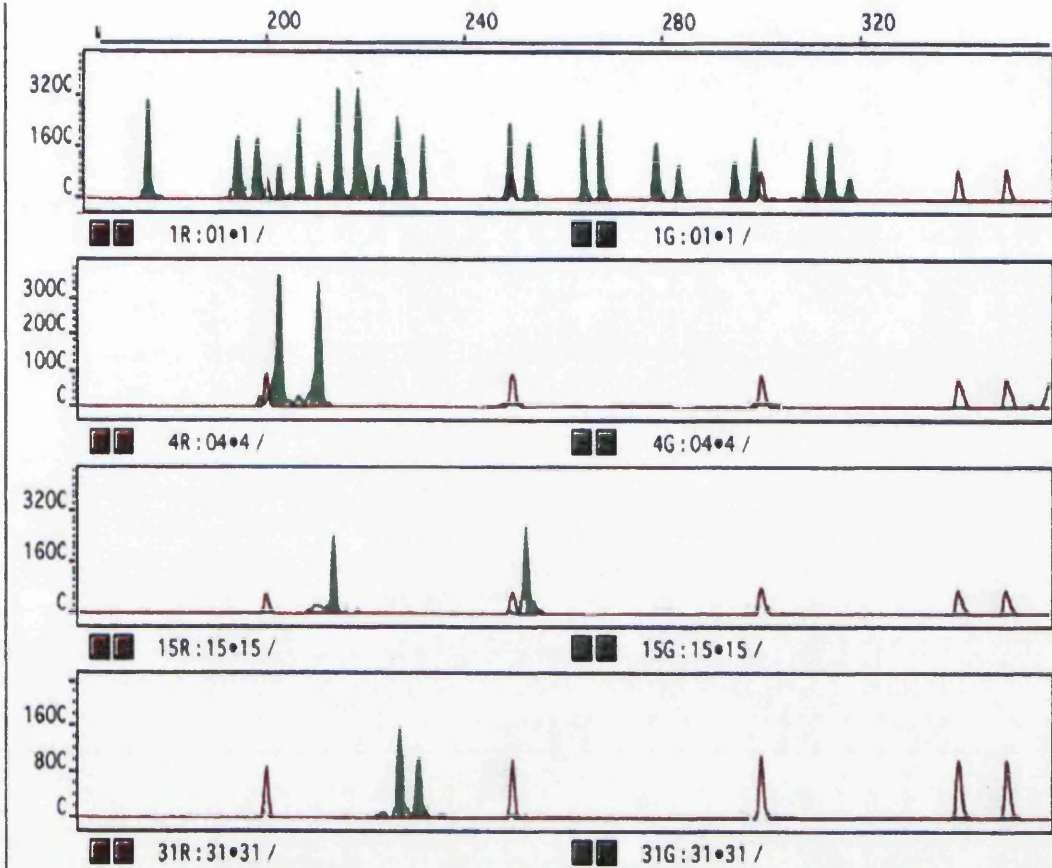
Figure 5.8 Electrophoretogram shows the D11S554 allelic ladder

The ladder of D11S554 locus contained 24 alleles, their sizes ranged from 176 bp to 318 bp. The D11S554 ladder was run in the same gel with the samples. The differences of the size between the ladder and the samples were less than 0.5 bp.

1-Panel one shows the D11S554 allelic ladder.

2-Panels 2-4 show three samples in comparison to the allelic ladder.

3-The bottom table shows the sizes of the alleles of the ladder.



Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
1G, 3	183.90	175.88	1734	11862	1839
1G, 5	199.40	194.95	1047	9481	1994
1G, 6	202.50	198.77	964	9846	2025
1G, 7	205.60	202.67	680	3200	2056
1G, 9	208.70	206.61	1456	9808	2087
1G, 11	211.80	210.54	644	3667	2118
1G, 13	214.90	214.48	1977	12929	2149
1G, 14	218.10	218.53	1885	23268	2181
1G, 15	221.20	222.46	590	5637	2212
1G, 16	224.40	226.51	1496	10341	2244
1G, 17	225.20	227.52	705	4671	2252
1G, 18	228.40	231.57	983	6002	2284
1G, 21	242.60	249.50	1382	10070	2426
1G, 22	245.80	253.35	915	9322	2458
1G, 24	255.40	264.85	1334	10586	2554
1G, 25	258.40	268.45	1354	10615	2584
1G, 26	267.80	279.77	927	9914	2678
1G, 27	270.90	283.52	521	4600	2709
1G, 28	280.50	295.14	555	4740	2805
1G, 29	283.50	298.78	1086	9166	2835
1G, 32	292.50	310.06	1092	8317	2925
1G, 33	295.60	314.02	922	8338	2956
1G, 34	298.60	317.88	377	3105	2986

144A

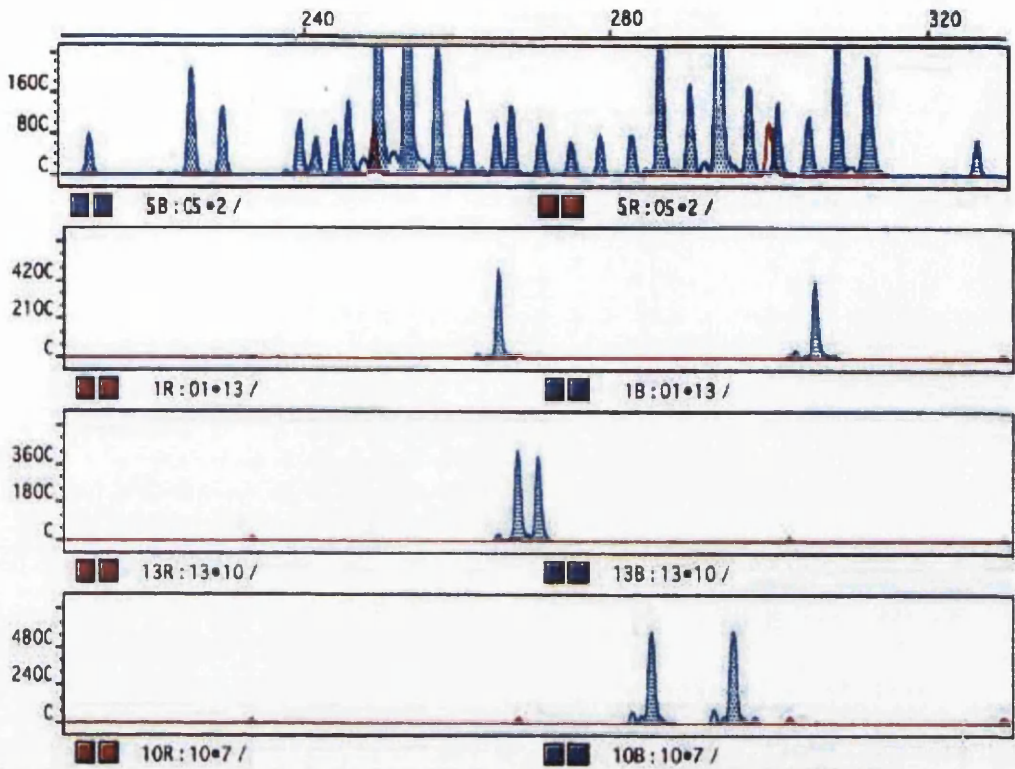
Figure 5.9 Electrophoretogram shows the SE33 allelic ladder

The ladder of the SE33 locus contained 26 alleles, their sizes ranged from 214 bp to 326 bp. The SE33 ladder was run in the same gel with the samples. The differences of the size between the ladder and the samples were less than 0.5 bp.

1-Panel one shows the SE33 allelic ladder.

2-Panels 2-4 show three samples in comparison to the allelic ladder.

3-The bottom table shows the sizes of the alleles of the ladder.



Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
5B_1	214.50	213.61	831	5491	2145
5B_2	224.80	226.69	2170	14299	2248
5B_3	227.90	230.62	1373	8687	2279
5B_4	216.20	240.42	969	6090	2162
5B_5	217.70	242.47	740	4028	2177
5B_6	219.10	244.39	1022	8974	2191
5B_7	220.60	246.44	1744	22017	2206
5B_9	223.40	250.26	5737	45330	2234
5B_11	226.30	254.08	5756	52071	2263
5B_12	229.20	257.90	2968	12594	2292
5B_13	232.10	261.73	1536	6634	2321
5B_15	234.90	265.44	1096	4597	2349
5B_16	257.90	267.67	1388	9864	2579
5B_17	261.00	271.41	1064	7148	2610
5B_19	264.10	275.15	690	4618	2641
5B_22	267.20	278.90	840	4846	2672
5B_23	270.30	282.65	811	5485	2703
5B_25	250.40	286.28	3049	21208	2504
5B_27	253.10	290.04	1865	12602	2531
5B_28	255.90	293.51	5345	40516	2559
5B_29	257.30	297.40	1781	12703	2573
5B_30	258.70	300.95	1466	9788	2587
5B_32	261.40	304.90	1154	8810	2614
5B_34	264.20	308.50	3254	21517	2642
5B_35	266.90	312.45	2405	20776	2669
5B_38	279.00	326.28	704	4371	2790

145A

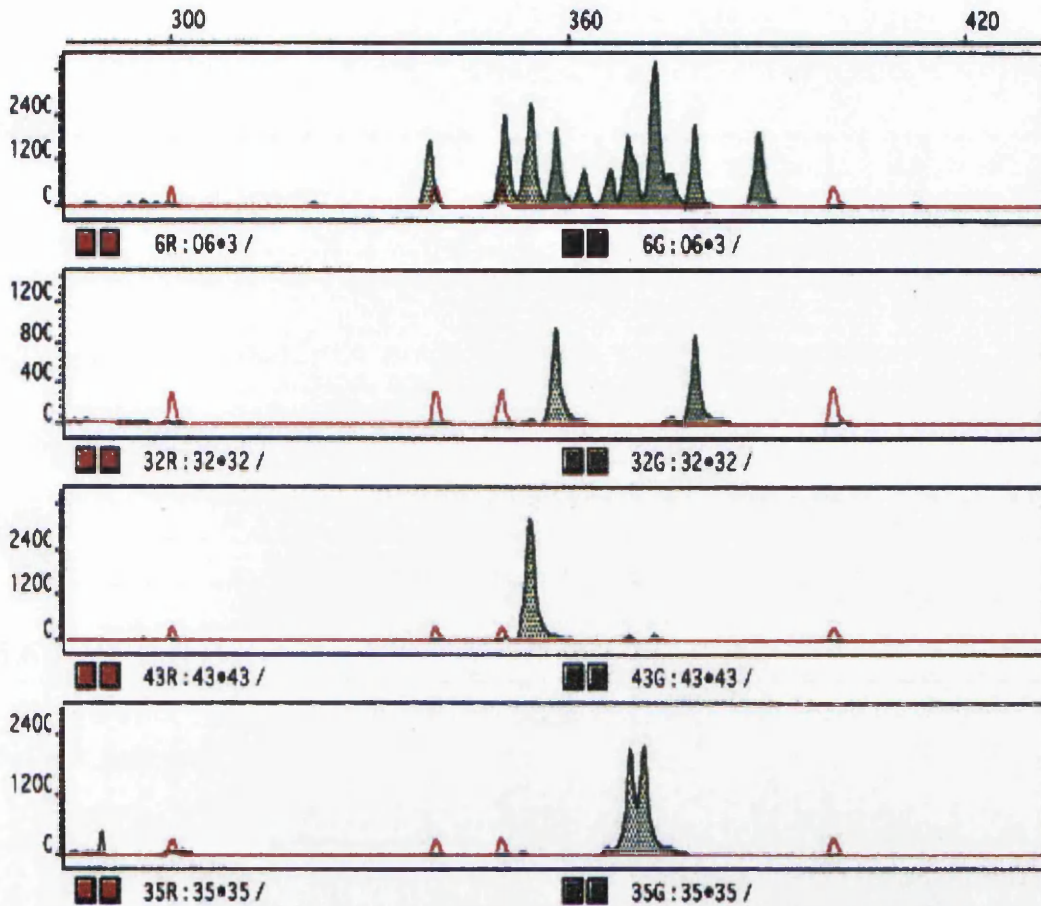
Figure 5.10 Electrophoretogram shows the APOAI1 allelic ladder

The ladder of the APOAI1 locus contained 11 alleles, their sizes ranged from 340 bp to 388 bp. The APOAI1 ladder was run in the same gel with the samples. The differences of the size between the ladder and the samples were less than 0.5 bp.

1-Panel one shows the APOAI1 allelic ladder.

2-Panels 2-4 show three samples in comparison to the allelic ladder.

3-The bottom table shows the sizes of the alleles of the ladder and the samples.



Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
6G, 6	288.10	339.67	1805	14936	2881
6G, 7	295.10	350.61	2495	23526	2951
6G, 8	297.70	354.58	2758	30668	2977
6G, 10	300.10	358.46	2187	20799	3001
6G, 11	302.90	362.46	992	10275	3029
6G, 12	305.60	366.52	1029	9508	3056
6G, 13	307.30	369.08	1906	23678	3073
6G, 14	310.10	373.27	3894	40991	3101
6G, 15	311.70	375.65	906	10450	3117
6G, 16	314.20	379.38	2290	9541	3142
6G, 17	321.00	388.03	2196	23814	3310
32G, 14	300.10	358.26	974	10823	3001
32G, 16	314.20	379.38	890	9541	3142
43G, 17	297.50	354.26	3288	39724	2975
35G, 17	307.70	369.49	2184	21254	3077
35G, 18	309.00	371.44	2227	28229	3090

5.6. SIZING AND DESIGNATION OF ALLELES

Internal standard marker (ROX500) and formamide were added to each sample. All samples were denatured according to the method stated in the “Materials and Methods” section and loaded in a 373XL automated sequencer. The fragments were sized using GeneScan 2.1 software. There were differences between the true sequenced size and the measured size using the ROX500 standard. The true size was 0.5 bp to 8 bp greater than the measured size. The discrepancy between the measured sizes of allele and the true sizes was greater with increasing the size of the allele for the three loci.

5.6.1. Allelic Designation

5.6.1.1. Intergel Variation

The constructed allelic ladders were reanalysed using four 6% denatured polyacrylamide gels to determine the intergel variation by calculating the mean of the length of each allele with the standard deviation in comparison to the internal size marker. The maximum standard deviation for the three allelic ladders using measurements generated by the ROX500 size standard was 0.32. This exceeded the upper limit for confidant typing at the 1 bp difference which exists in these loci (Table 5.2).

Table 5.2 Intergel variations of the alleles of the allelic ladders.

The constructed ladders were reanalysed on four denatured polyacrylamide gels (6%) to determine the intergel variation by calculating the mean of the length of each allele with the standard deviation in comparison to the internal size (ROX500). The maximum standard deviation for the three allelic ladders was 0.32.

D11S554 ± SD	SE33 ± SD	APOA11 ± SD
175.87 ± 0.05	240.45 ± 0.069	339.61 ± 0.109
194.90 ± 0.10	242.45 ± 0.080	350.62 ± 0.118
202.65 ± 0.08	246.49 ± 0.115	354.55 ± 0.116
206.60 ± 0.08	250.22 ± 0.093	358.47 ± 0.158
210.56 ± 0.06	254.11 ± 0.061	362.45 ± 0.164
214.52 ± 0.07	257.93 ± 0.112	366.47 ± 0.179
218.55 ± 0.09	261.75 ± 0.107	369.55 ± 0.220
222.44 ± 0.08	275.16 ± 0.118	373.29 ± 0.319
226.50 ± 0.09	278.95 ± 0.084	
227.56 ± 0.09	282.62 ± 0.157	
231.61 ± 0.15	286.26 ± 0.118	
249.58 ± 0.14	290.02 ± 0.185	
253.40 ± 0.13	293.58 ± 0.164	
	297.51 ± 0.225	
	326.29 ± 0.270	

5.6.1.2. Intragel Variation

Because the SD between gels was too large to allow allele size determination with the required accuracy, intragel variation was also investigated to determine the accurate sizing of the alleles in comparison to a specific ladder for each locus. Samples representing different alleles were run four times in the same gel to calculate the mean of the measurement difference of the alleles that had fragments of equal length in comparison with the specific ladders. The mean differences ranged from 0.05 to 0.14 bp indicating that the difference between the alleles of equal size run on the same gel allowed typing of alleles that differ by 1 bp (Table 5.3).

5.7. GENOTYPE OF THE EGYPTIAN SAMPLES

All the 300 samples were run on the electrophoretic gels along with the allelic ladders. Comparison of the samples with the ladders was performed to calculate the difference in size between the samples and the ladders. Most of the samples gave results that were very close to size of the alleles of the ladders in bp (less than 0.5 bp) as recommended by Gill *et al.*, (1996). Additionally, measurement of the correlation effect was used to confirm the designation of the alleles relative to the allelic ladder (Gill *et al.*, 1996). The correlation effect was calculated according to the equation

$$(C = \delta_1 - \delta_2)$$

Where δ_1 and δ_2 is the deviation of one allele to its respective ladder marker in

Table 5.3 Intragel variations of the alleles of the allelic ladder .

Samples representing different alleles were run four times in the same gel with the specific ladders to calculate the mean of the measurement difference of the alleles that had fragments of equal length in comparison with the specific ladder. The maximum difference was 0.14 bp.

D11S554	Measurement difference (bp)	SD.
175.87	0.05	0.005
194.90	0.05	0.020
202.65	0.06	0.008
206.60	0.08	0.010
210.56	0.06	0.030
214.52	0.07	0.009
218.55	0.09	0.020
222.44	0.08	0.010
226.50	0.09	0.020
227.56	0.10	0.030
231.61	0.12	0.020
249.58	0.11	0.030
253.40	0.14	0.040

SE33	Measurement difference (bp)	SD.
240.45	0.04	0.003
242.45	0.05	0.004
246.49	0.05	0.005
250.22	0.06	0.005
254.11	0.07	0.008
257.93	0.06	0.007
261.75	0.08	0.010
275.16	0.09	0.030
278.95	0.11	0.020
282.62	0.13	0.030
286.26	0.12	0.040
290.02	0.12	0.020
293.58	0.13	0.030
297.51	0.14	0.020
326.29	0.14	0.030

APOAI1	Measurement difference (bp)	SD
339.61	0.06	0.005
350.62	0.07	0.005
354.55	0.09	0.007
358.47	0.12	0.010
362.45	0.14	0.010
366.47	0.13	0.020
369.55	0.12	0.020
373.29	0.14	0.010

a heterozygous sample. If C is less than 0.5 then the allele could be designated. New variants could be calculated by reference to existing ladders and the correlation effect could be calculated. This method was used to type all alleles.

Some samples that gave sizes not included in the ladders were amplified as a singleplex and reanalysed twice, and then the samples added to the ladders to increase the number of alleles included in each locus. Some of these samples were chosen for sequencing to confirm the size and the structure of each allele. The sequence of one sample for the SE33 locus is shown as an example in figure 5.11.

5.8. NOMENCLATURE

The nomenclature of hypervariable STRs based on the number of repeats is difficult due to the complex structure of the repeated regions (Holgersson *et al.*, 1994). The alternative method to designate the alleles is based on their size in comparison with allelic ladders, which will be dependent on the primer used. The designation of each allele should be prefixed by the term “type” in order to confirm that the designation is not as absolute indication of size in base pairs (Gill *et al.*, 1997). It is possible for each laboratory to use a local nomenclature to make it as simple as possible (Schneider *et al.*, 1998). Moreover, because of using a new primer for the APOA11 locus the method of nomenclature used in this study was dependent on the fragment length for each locus as recommended by Rolf *et al.*, (1997). All alleles including microvariants were called beginning with number 1. No specific name was given to the microvariants.

152A

Figure 5.11 Sequence of one sample for the ACTBP2 locus

Sequence of the sample using the reverse primer shows the follow repeats

(TTTC)₂TT(TTTC)₃TC(TTTC)₃TC(TTTC)₁₃(TT)₁(TTTC)₁₈CTTCC(TTTC)₂(TC)₁GT
TC

The sample has 311 bp length.

5.9. STATISTICAL ANALYSIS

After optimisation of the multiplex PCR, 300 samples were typed for the three loci. The statistical analysis was then performed using GDA (Lewis and Zaykin, 1999), TFPGA (Miller, 1998) and PowerStats (Tereba, 1999) software programmes.

5.9.1. Genotype Frequency

For the D11S554 locus 160 genotypes of 861 possible genotypes were observed. The genotype 7,9 (2.7%) was the most common one occurring in 8 from 300 individuals followed by genotype 7,11 having the frequency of 2.3%.

186 genotypes out of 1081 possible genotypes were observed for the SE33 locus. The genotype 17,19 was the most common one with a frequency of 3% as it was seen in 9 from 300 individuals followed by genotype 15,17 with a frequency of 2.7%.

There were 144 out of possible 496 different genotypes in the Egyptian population at the APOA11 locus. The 16,18 genotype had the highest frequency (3.7%) occurring in 11 out of 300 individuals followed by genotype 18,21 with a frequency of 3.3%.

5.9.2. Allele Frequency

The frequency distribution of the observed alleles in the Egyptian population samples for the three loci was calculated from the observed genotypes.

D11S554 locus had 41 alleles ranging from 176 bp to 318 bp. Seven alleles had a single occurrence (0.0017) and allele 11 had the highest frequency (0.1255). The frequency of all alleles was < 10% except for alleles 9 and 11.

The ACTBP2 locus had 46 alleles ranging from 214 bp to 326 bp. All the alleles had frequencies under 10%. Allele 15 was the most common allele (0.0933) followed by alleles 17 and 19 (0.0917). Six different alleles had only one occurrence.

The APOAI1 locus had 31 alleles ranging from 340 bp to 400 bp. Three alleles of APOAI1 were relatively common, allele 16, 18 and 21 with frequencies of 0.1417, 0.1433, and 0.1217 respectively. The last three alleles had the lowest frequency of 0.0033. All except four alleles had a frequency less than 10%.

The standard deviation for each allele frequency was calculated for each locus of the three loci. The results are shown in table 5.4 and figure 5.10.

5.9.3. Paternity Parameters

5.9.3.1. *Paternity Index*

Typical paternity index varied from 4.69 for the APOAI1 locus, 5.17 for the D11S554 locus to 6.82 for the SE33 locus. The combined paternity index of the three loci was 165.47. The probability of paternity was calculated from the paternity index giving a value 99.4% assuming a prior probability of guilt of 50%.

Table 5.4 Allele frequencies in 300 Egyptian individuals for the three loci.

The D11S554 locus had 41 alleles, their size ranging from allele 176 to allele 318. Allele 11 had the highest frequency (0.125). The frequency of all alleles was <10% except alleles 9 and 11 only.

D11S554 ALLELE	SIZE	OBSERVED NO	ALLELE FREQUENCY	SD
1	175.87	1	0.0017	0.0017
2	186.89	2	0.0033	0.0023
3	194.90	9	0.0151	0.0050
4	198.76	31	0.0518	0.0090
5	200.71	7	0.0117	0.0043
6	202.65	34	0.0569	0.0094
7	206.60	57	0.0953	0.0119
8	208.58	3	0.0050	0.0029
9	210.57	62	0.1037	0.0124
10	212.54	2	0.0033	0.0023
11	214.52	75	0.1255	0.0135
12	218.57	33	0.0552	0.0093
13	219.51	26	0.0435	0.0083
14	222.44	29	0.0485	0.0088
15	223.45	27	0.0452	0.0085
16	226.55	15	0.0251	0.0064
17	227.56	36	0.0602	0.0096
18	230.60	3	0.0050	0.0029
19	231.61	31	0.0517	0.0091
20	233.59	2	0.0033	0.0023
21	235.56	12	0.0201	0.0057
22	237.53	3	0.0050	0.0029
23	239.54	6	0.0100	0.0041
24	241.52	8	0.0134	0.0047
25	243.53	2	0.0033	0.0023
26	245.55	12	0.0201	0.0057
27	247.56	1	0.0017	0.0017
28	249.58	29	0.0485	0.0088
29	253.40	16	0.0268	0.0066
30	257.35	6	0.0100	0.0041
31	259.14	1	0.0017	0.0017
32	260.92	3	0.0050	0.0029
33	264.84	2	0.0033	0.0023
34	268.45	1	0.0017	0.0017
35	279.75	1	0.0017	0.0017
36	283.51	1	0.0017	0.0017
37	295.12	3	0.0050	0.0029
38	298.77	2	0.0033	0.0023
39	310.04	2	0.0033	0.0023
40	314.02	1	0.0017	0.0017
41	317.85	3	0.0050	0.0029

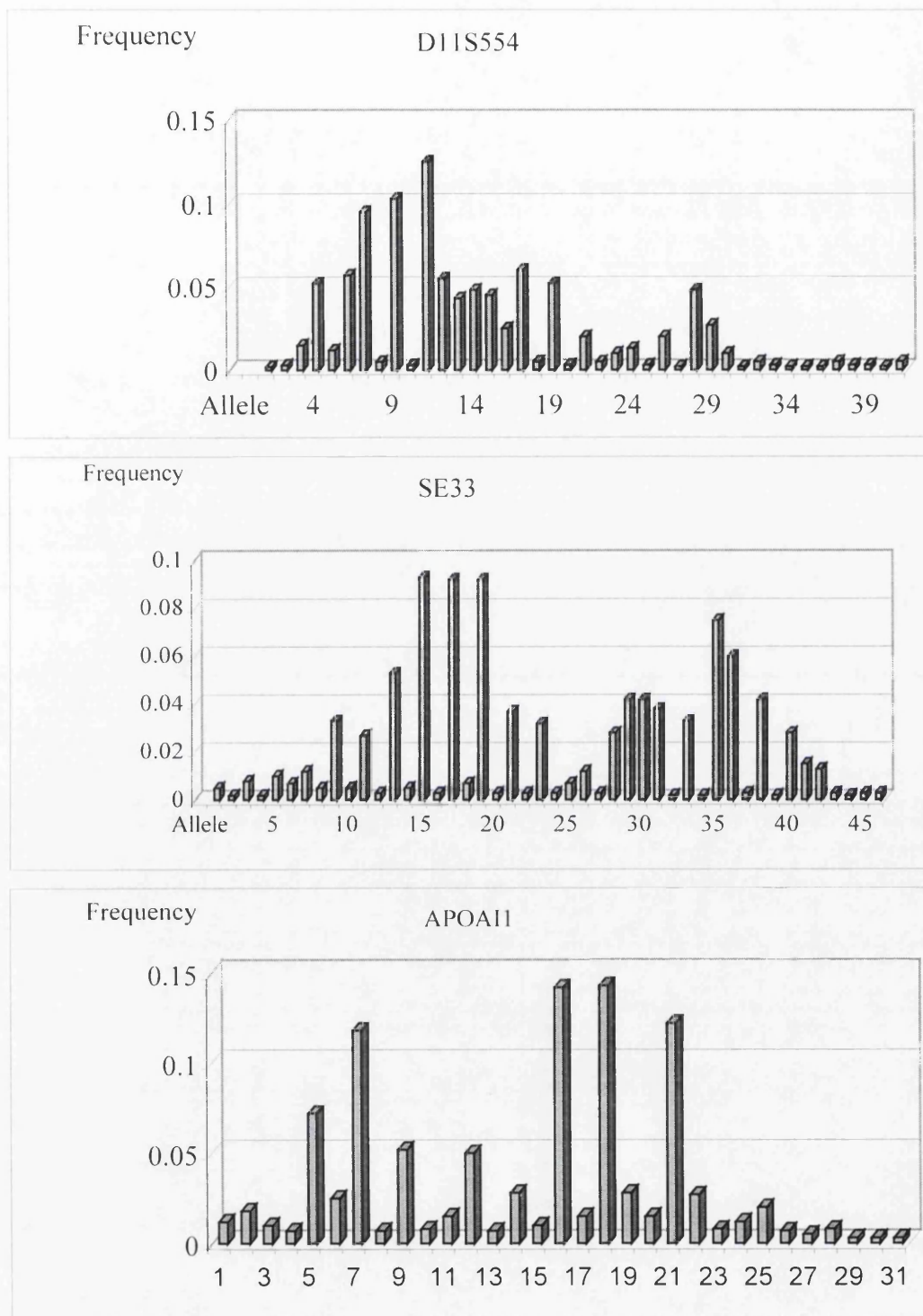
The ACTBP2 locus had 46 alleles ranging from 214 bp to 326 bp. Allele 15 was the most common allele (0.0933) All the alleles had frequencies under 10%.

SE33 ALLELE	SIZE	OBSERVED NO	ALLELE FREQUENCY	SD
1	213.72	3	0.0050	0.0029
2	217.58	1	0.0017	0.0017
3	226.65	5	0.0083	0.0037
4	228.63	1	0.0017	0.0017
5	230.70	6	0.0100	0.0041
6	232.62	4	0.0067	0.0033
7	236.56	7	0.0117	0.0043
8	238.50	3	0.0050	0.0029
9	240.44	20	0.0333	0.0073
10	241.45	3	0.0050	0.0029
11	242.43	16	0.0267	0.0066
12	244.37	2	0.0033	0.0023
13	246.49	32	0.0533	0.0092
14	248.35	3	0.0050	0.0029
15	250.22	56	0.0933	0.0119
16	252.17	2	0.0033	0.0023
17	254.11	55	0.0917	0.0118
18	256.02	4	0.0067	0.0033
19	257.93	55	0.0917	0.0118
20	259.84	2	0.0033	0.0023
21	261.75	22	0.0367	0.0077
22	263.60	2	0.0033	0.0023
23	265.45	19	0.0317	0.0072
24	267.49	2	0.0033	0.0023
25	269.34	4	0.0067	0.0033
26	271.49	7	0.0117	0.0043
27	275.18	2	0.0033	0.0023
28	278.85	17	0.0283	0.0068
29	282.41	25	0.0417	0.0082
30	286.16	25	0.0417	0.0082
31	289.72	23	0.0383	0.0078
32	290.68	1	0.0017	0.0017
33	291.65	21	0.0350	0.0076
34	292.62	1	0.0017	0.0017
35	293.58	44	0.0733	0.0106
36	297.28	36	0.0600	0.0096
37	299.11	2	0.0033	0.0023
38	300.93	25	0.0417	0.0082
39	302.85	1	0.0017	0.0017
40	304.82	17	0.0283	0.0068
41	308.51	9	0.0150	0.0050
42	312.49	8	0.0133	0.0047
43	316.36	2	0.0033	0.0023
44	318.22	1	0.0017	0.0017
45	321.95	2	0.0033	0.0023
46	326.19	2	0.0033	0.0023

The APOAII locus had 31 alleles ranging from 340 bp to 400 bp. Allele 18 had the highest frequency (0.1433). All alleles had a frequency less than 10% except four alleles only.

APOAII ALLELE	SIZE	OBSERVED NO	ALLELE FREQUENCY	SD
1	339.71	7	0.0117	0.0043
2	344.70	11	0.0183	0.0055
3	346.66	6	0.0100	0.0041
4	348.65	4	0.0067	0.0033
5	350.62	43	0.0717	0.0104
6	352.28	15	0.0233	0.0064
7	354.55	71	0.1183	0.0132
8	356.41	3	0.0050	0.0029
9	358.27	31	0.0517	0.0090
10	359.22	5	0.0083	0.0037
11	360.36	9	0.0150	0.0050
12	362.45	30	0.0500	0.0088
13	363.43	4	0.0067	0.0033
14	364.51	17	0.0283	0.0068
15	365.53	6	0.0100	0.0041
16	366.57	85	0.1417	0.0142
17	367.72	9	0.0150	0.0050
18	369.47	86	0.1433	0.0143
19	371.44	17	0.0283	0.0068
20	372.35	9	0.0150	0.0050
21	373.25	73	0.1217	0.0133
22	375.63	16	0.0267	0.0066
23	376.58	5	0.0083	0.0037
24	377.52	12	0.0200	0.0057
25	379.38	7	0.0117	0.0043
26	383.29	4	0.0067	0.0033
27	385.12	4	0.0067	0.0033
28	388.33	5	0.0083	0.0037
29	390.93	2	0.0033	0.0023
30	395.87	2	0.0033	0.0023
31	399.76	2	0.0033	0.0023

Figure 5.12 Allele frequencies for the three hypervariable loci.



5.9.3.2. Exclusion of Paternity

The probability of exclusion of paternity ranged from 0.782 for the APOAI1 locus to 0.850 for the ACTBP2 locus. The D11S554 locus had an exclusion probability of 0.802. The combined probability of exclusion for the three loci was 99.4%.

5.9.4. Forensic Parameters

5.9.4.1. Matching Probability

The matching probability was calculated for the three loci. It ranged from 0.009 for both the D11S554 and the SE33 loci to 0.013 for the APOAI1 locus. The combined matching probability was 0.000001053. Thus the chance of occurrence of the same profile for the three loci was 1 in 949,668 of the local population.

5.9.4.2. Discrimination Power

All loci were highly polymorphic. The ACTBP2 system was as informative as D11S554, giving a probability of discrimination of 0.991. The APOAI1 locus had a power of discrimination of 0.987. The combined discrimination power of the three loci was 0.999998. The paternity index, probability of exclusion, matching probability and the power of discrimination for the loci are shown in table 5.5.

Table 5.5 Paternity and forensic parameters for the hypervariable loci.

The study of the paternity and forensic parameters revealed that the SE33 locus had the highest value and APOAI1 locus had the lowest values in all of these parameters. Paternity Index (PI), Probability of Paternity (PE), Matching Probability (MP) and Power of Discrimination (PD).

LOCUS	PI	PE	MP	PD
D11S554	5.17	0.802	0.009	0.991
SE33	6.82	0.850	0.009	0.991
APOAI1	4.69	0.782	0.013	0.987
Combined	165.366586	0.994	0.000001053	0.999998

5.10. HARDY-WEINBERG EQUILIBRIUM

In hypervariable loci, the expected counts are small because of the low frequency of the alleles. Chi-square test was not performed as it may lead to false rejection of the null hypothesis (Evetts and Weir, 1998), therefore the exact test was performed to test deviation from HWE and for linkage equilibrium. The loci were under the Hardy-Weinberg equilibrium, as there was an agreement between the observed and expected genotype values. The exact test for independence of the alleles within individuals was performed for each locus. The results showed no evidence for correlation between the alleles at any of the pair of loci. The exact test was performed with the significant levels ($p < 0.05$) and 2000 permutations of alleles. The p-values are shown in table 5.6.

5.11. HETEROZYGOSITY TEST

All the three loci showed heterozygosity values $> 89\%$, ranging from 0.892 for the APOA11 locus to 0.927 for the ACTBP2 locus. The expected heterozygosity was calculated for the three loci using GDA software. The observed heterozygosity was less than the expected heterozygosity for the three loci. The standard deviation of the expected heterozygosity was calculated to determine the range of the expected heterozygosity (Edward *et al.*, 1992). The results are shown in table 5.7. The difference between the observed and the expected heterozygosity was investigated by chi-square test which showed no significant difference, the p-value was 0.9904.

Table 5.6 p-value of exact test

Exact test with the significant levels ($p < 0.05$) and 2000 permutations of alleles was performed. The loci were under the Hardy-Weinberg equilibrium. The exact test for independence of the alleles within individuals showed no evidence for correlation between the alleles at any of the pair of loci ($p > 0.05$).

STR Loci	P
D11S554	0.730
SE33	0.134
APOA1	0.999
D11S554 SE33	0.490
D11S554 APOAI1	0.182
SE33 APOAI1	0.439
D11S554 SE33 APOAI1	0.208

Table 5.7 Heterozygosity test for the three hypervariable loci

The SD of the expected heterozygosity was calculated for each locus. The difference between the observed and the expected heterozygosity was investigated by chi-square test showing no significant difference ($p = 0.9904$).

LOCUS	OBSERVED	EXPECTED	SD
D11S554	271	282	4.15
SE33	278	285	3.81
APOAI1	268	275	4.80

5.12. COMPARISON WITH OTHER POPULATIONS

Comparison between the allele frequencies of the Egyptian population and the Norwegian population (300 individuals) (Dupuy and Olaisen, 1997) was performed for the three hypervariable loci. Comparison was also performed with the Japanese population for both of the D11S554 (Nagai *et al.*, 1998) and the SE33 (179 individuals) (Liu *et al.*, 1997) loci. It was difficult to make comparison with other populations, as not many studies have been performed using these loci. The frequency profile comparison was carried out using the two-way RxC contingency table test (Power Instat 2.03). The results demonstrated significant differences between the Egyptian population and the compared populations as shown in table 5.8.

5.13. COMBINING ALL SIX LOCI

Combining the results of the D3S1358, vWA and FGA loci with the results of the hypervariable loci gave a combined matching probability of 16×10^{-11} , a power of discrimination of 0.9999999998, a combined typical paternity index of 7596 and a probability of paternity of 99.985% and a probability of exclusion of paternity of 0.99986.

Table 5.8 X² value of the comparison between Egypt and other populations

The comparison demonstrated significant differences between the Egyptian population and the compared populations ($p < 0.05$). (*There was a significant difference).

D11S554	X ²	df	p
Egypt and Norway	376.53	59	< 0.0001*
Egypt and Japan	589.61	63	< 0.0001*
SE33			
Egypt and Norway	247.36	45	< 0.0001*
Egypt and Japan	118.03	50	< 0.0001*
APOA11			
Egypt and Norway	109.21	37	< 0.0001*

5.14. DISCUSSION

5.14.1. PCR Optimisation

Multiplex reactions were established by designing a new primer for the APOA11 locus to simplify the analysis. The amplification of some samples was repeated after their failure in the first PCR. This led to a 100% success rate in the amplification of samples. The results of allele separation for the multiplex were balanced by adjusting primer concentrations.

5.14.2. Sizing of the Alleles

The average size obtained for the alleles of the three loci differed to their true sequenced size. The same observation was recorded in other studies where there was a deviation up to 10 bp between the actual sequenced size and measured size for the large allele for SE33 locus (Deforce *et al.*, 1998). The cause of this observation may be due to the nucleotide content of the analysed alleles (Dupuy and Olaisen, 1997). Another cause of this difference is the type of fluorescence dye used (Poeltl *et al.*, 1997). The difference increased with increasing in the length of the alleles, which may be explained by the greater number of repetitive units the large allele contains, which may affect electrophoretic mobility (Poeltl *et al.*, 1997).

5.14.3. Construction of the Ladders

The allelic ladders were constructed by cloning sequenced samples. Different alleles, which were not included in the ladders, were amplified as a singleplex and added to ladders to increase the number of alleles included in each ladder.

The use of a specific allelic ladder for each locus allows an accurate and precise designation of alleles. This method is also more accurate than comparison with classical size standards such as bacteriophage or plasmid markers as there is minimal sequence variation between the ladder and different alleles of the same locus (Sprecher *et al.*, 1996). Allelic ladders are also important in the identification of microheterogeneity of different loci such as TH01 locus where allele 9.3 was easily differentiated from allele 10 when the specific allelic ladder was used (Puers *et al.*, 1993). Another study reported the necessity of using a locus specific ladder for each locus to obtain correct allele sizes with a lower SD (Malmgren *et al.*, 1998).

Application of interpretative guidelines such as the ± 0.5 bp rule and the shift calculation (Gill *et al.*, 1996) for detecting intermediate alleles has proved to be useful when assigning alleles. However, some of the samples of the APOA11 locus had to be re-run on a different gel to confirm the results.

5.14.4. Paternity and Forensic Parameters

Typical paternity indices for the highly hypervariable loci were relatively high (165.47). In comparison with Powerplex™ kit including 8 loci, the combined paternity

index was 191 in the Egyptian population (Amany press), which was slightly higher than that of the three hypervariable loci. The typical paternity index was also higher than the combined paternity index for combination of both CTTv and FFFL multiplexes in Hispanic- American population (86.4973) (Lins *et al.*, 1998).

The combined probability of exclusion for the hypervariable loci was 0.994. In another report the value of the probability of exclusion was 0.9977 (Dupuy and Olaisen, 1997). The value was higher than the probability of paternity of the CTTv, GammaSTR™, and FFFL multiplex in all American populations (Lins *et al.*, 1998).

The matching probability of the hypervariable loci was 0.000001053. The value was nearly equal to the matching probability of seven loci (TH01, TPOX, F13A1, F13B, CD4, FES/FPS and D5S373 (0.0000013) in North-East Italy (Cossutta *et al.*, 2000).

5.14.5. HWE

The exact test was used to detect the deviation from HWE. There was no deviation from HWE for the three loci. The same finding was observed for the three loci in another study (Dupuy *et al.*, 1998).

5.14.6. Combination of Six Loci

The combination of all six studied loci studied gave a very high value of all paternity and forensic parameters.

The combined paternity index for the six loci was 7596. It was higher than the value obtained from the study of 4 VNTR loci in an Italian population, which had a value of 614.25 (Pastore *et al.*, 1996). It was also higher than the combined paternity index in Caucasian-American for both PowerPlex™ System and FFFL Multiplex including 12 loci having a value of 3976 (Lins *et al.*, 1998).

The combined power of exclusion for the six loci was 0.99986, which was higher than either of the PowerPlex™ System and FFFL Multiplexes in Hispanic- American (0.99974) and Caucasian-American (0.99981) (Lins *et al.*, 1998). The power of exclusion of paternity was equal to the one obtained from the study of 13 STR loci in Japanese population (0.99985) (Yamamoto *et al.*, 2000) and the power of exclusion of paternity of AmpFℓSTR Profiler Plus in a Portuguese population (0.99988) (Pinheiro *et al.*, 2000).

The matching probability was 16×10^{-11} and expressed as 1 in 6,250,000,000 (1 in 6.25×10^9). It was higher than the result obtained from the PowerPlex™ System having a matching probability of 1 in 2.74×10^8 (Lins *et al.*, 1998). It was also higher than the probability of matching of 9 loci included in the AmpFℓSTR Profiler Plus in a Japanese population (2.28×10^{-11}) (Nakamura *et al.*, 2000). All the paternity and forensic parameters of the combination of the six loci were lower than the paternity and forensic parameters of the Powerplex™ 16 system in all populations' studies (Sprecher *et al.*, 2000).

In conclusion, the result of the combination of all six loci was satisfactory for paternity testing in the Egyptian population as the use of the commercial kits may not be viable for paternity testing in Egypt because of the expense. Whenever the use of the commercial kits is available, the combination of the hypervariable loci with the commercial kits could be valuable particularly where there are deficiency cases.

CHAPTER 6: Y-CHROMOSOME STR

6.1. INTRODUCTION

A number of Y-chromosome polymorphisms have been established recently for paternity testing and casework in stain analysis especially in male lineage cases (Jobling *et al.*, 1997). Y-chromosome polymorphisms include microsatellites (Roewer *et al.*, 1992, 1996 and Mathias *et al.*, 1994), base substitutions (Hammer and Horai, 1995; Whitfield *et al.*, 1995 and Underhill *et al.*, 1996) and alphoid satellite polymorphism (Santos *et al.*, 1995). Y-chromosome STRs are the most informative to study variation within a certain population due to their high levels of polymorphism (Perez-Lezaun *et al.*, 1997 and Pestoni *et al.*, 1998). Y-STR analysis has a considerable role in paternity testing particularly in deficiency cases, when the alleged father is deceased, with a male offspring (Chakraborty, 1985). Any male relative to the alleged father could be tested in his place. Each difference at a Y-STR locus will detect non-paternity. However, with an inclusion a large database of haplotypes is recommended to give statistical significance (Roewer *et al.*, 1996).

Therefore to have a complete system for paternity testing, which covers as many case scenarios as possible, a database for Y-chromosome STR system was established.

6.2. SELECTION OF Y CHROMOSOME STR LOCI

Seven loci were chosen for the study that could be amplified as two multiplexes. The first multiplex included four loci, which were DYS19, DYS389I, DYS389II and DYS390. The second one included DYS391, DYS392 and DYS393 loci. Schneider *et al.*, (1999) recommended the use of these loci plus the DYS385 locus but DYS385 locus could not be amplified in either of the 2 multiplexes. Therefore, to save time only the seven loci were amplified in 2 multiplexes as their use in other populations showed a high power of discrimination of paternity (Kayser *et al.*, 1997). The loci for co-amplification were chosen on the basis of the size ranges of the alleles, the primer sequence and the primer annealing temperature. The reverse primer used to amplify DYS391 locus was a new one designed by Gusmao *et al* (2000). It was designed to reduce the unspecific products amplified using the original primers. The loci are described in table 6.1.

6.3. OPTIMISATION OF THE AMPLIFICATION CONDITIONS

One set of the primers amplifies both loci of DYS389 together as the forward primer anneals in two different places on the same strand. Optimisation of the amplification of each locus in both multiplexes has been previously performed (S. Hadi and M. Lathiqua). Testing of several changes in all parameters of the PCR including different annealing temperature, MgCl₂ concentration and primer concentration was performed first to optimise the singleplex PCR. Reproducible results for each locus were achieved.

Table 6.1 Characters of the chosen seven Y-STR loci

All loci are tetranucleotide except DYS392 locus (trinucleotide). DYS393 locus has the lowest range of allele size and DYS398II locus has the highest range of allele size.

Locus	Repeat motifs	Number of repeats	Number of known alleles	Size range
DYS19	4 bp	10-19	10	174-210
DYS389I	4 bp	7-13	7	239-263
DYS389II	4 bp	23-31	9	353-385
DYS390	4 bp	18-27	10	191-227
DYS391	4 bp	8-13	6	140-160
DYS392	3 bp	7-16	8	236-263
DYS393	4 bp	9-15	6	108-132

To optimize the multiplex PCR, equimolar amounts of each primer for each locus were combined using the same singleplex PCR programme with the lower annealing temperature. Adjustment of individual primer concentration was performed to increase the amount of product for the locus, when the intensity of its amplified product was weak within the multiplex. The amplification cycle conditions for the two multiplexes PCR were identical. The results for each singleplex and multiplex PCR conditions are shown in table 6.2.

6.4. CONSTRUCTION OF THE ALLELIC LADDERS

To determine the sizes of DNA fragments, a specific ladder consisting of all the common alleles of a given locus can be used for comparison on an electrophoretic gel (Puers *et al.*, 1993, 1994). To construct a locus specific allelic ladder for the examined loci, samples with different alleles were selected, then mixed together and amplified to construct a balanced ladder for each locus. The compound ladder was obtained by mixing these single ladders together.

To calibrate the alleles of the ladder, comparison of the alleles of the ladders with two amplified sequenced samples for all loci was performed. These two samples were sent by Gonzalez-Neira (Institute of Legal Medicine, University of Santiago de Compostela, Spain). Allelic nomenclature was used according to de Knijff *et al.*, (1997). Allele designation for the seven Y-STRs of all samples was made by comparison with “in house” allelic ladders (Figures 6.1 and 6.2).

Table 6.2 Optimised PCR Parameters for Single & Multiplex conditions in Multiplex I and II.

The annealing temperature was the same for both multiplexes. The concentration of the each primer was optimised to yield approximately equal products all loci within each multiplex.

Locus	Annealing temperature	MgCL₂	Primer concentration	Extension temperature
DYS19	54 °C	2 mM	0.25 µM	72 °C
DYS389I DYS389II	56 °C	2 mM	0.25 µM	72 °C
DYS390	56 °C	2 mM	0.2 µM	72 °C
Multiplex	56 °C	2.5 mM	0.25 µM DYS19 0.25 µM DYS389 0.12 µM DYS390	72 °C

Locus	Annealing temperature	MgCL₂	Primer concentration	Extension temperature
DYS391	58 °C	1.5 mM	0.25 µM	72 °C
DYS392	54 °C	1.5 mM	0.3 µM	72 °C
DYS393	55 °C	1.5 mM	0.15 µM	72 °C
Multiplex	56 °C	1.5 mM	0.20 µM DYS391 0.25 µM DYS392 0.20 µM DYS393	72 °C

176A

Figure 6.2 Electrophoretogram shows the allelic ladder of the Y-chromosome STR loci of multiplex II

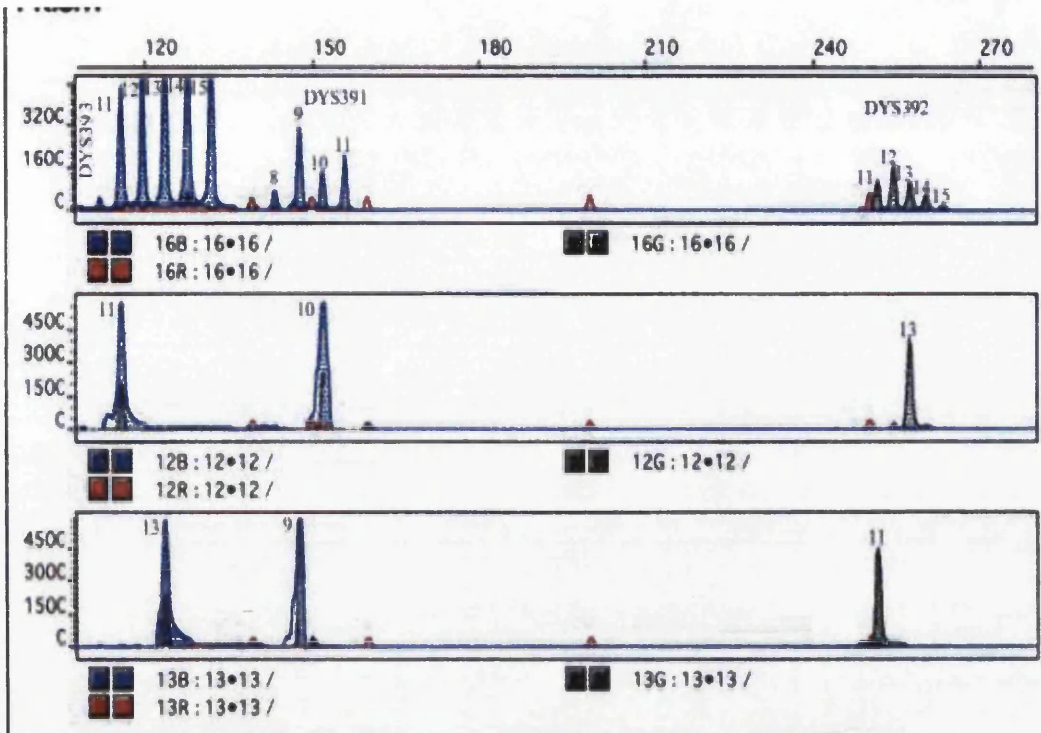
The ladder of the Y-chromosome STR loci of multiplex II contained alleles 11-15 for DYS393 locus, alleles 8-11 for DYS391 locus and alleles 11-15 for DYS392 locus.

The ladder was run in the same gel with the samples. The differences of the size between the ladder and the samples were less than 0.5 bp.

1-Panel one shows the ladder of the Y-chromosome STR loci of multiplex II.

2-Panels 2-3 show two samples in comparison to the allelic ladder.

3-The bottom table shows the sizes of the alleles of the ladder and the samples.



Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
16B, 12	140.40	115.61	4954	31691	1404
16B, 14	143.60	119.52	5542	36459	1436
16B, 16	146.90	123.60	6021	42127	1469
16B, 17	150.20	127.73	5969	49638	1502
16B, 19	153.50	131.91	5860	41525	1535
16B, 21	162.20	143.29	805	4587	1622
16B, 23	165.60	147.78	3238	21925	1656
16B, 24	168.90	151.95	1535	9769	1689
16B, 26	172.30	156.02	2167	13364	1723
16G, 50	253.30	251.46	1160	9063	2535
16G, 51	256.00	254.28	1735	14178	2560
16G, 52	258.60	257.21	1094	9543	2586
16G, 53	261.20	260.14	649	5929	2612
16G, 54	264.00	263.30	194	1754	2640
12B, 19	140.60	115.61	5803	65080	1406
12B, 29	169.00	151.83	5759	116035	1690
12G, 40	258.90	257.09	4152	35774	2589
13B, 22	147.00	123.37	5778	76737	1470
13B, 31	165.80	147.65	5849	87434	1658
13G, 34	253.80	251.35	4498	41286	2538

6.5. IDENTIFICATION OF A NEW ALLELE

There were 2 samples that gave allele length of 237 bp for DYS392 locus. In the sequenced ladder for DYS392 locus, size 240 bp representing allele 7, which is the smallest allele in this locus. Therefore, sequencing of these samples was performed. Purification of the allele direct from 100 μ l PCR solution was performed and the purified DNA was run on 2% agarose gel to estimate the quantity of the purified DNA. The samples were run again on a 373 XL automated sequencer to confirm the length, which was again 237 bp. Sequencing of the sample was performed using the reverse primer as described in "Materials & Methods". The result of the sequence showed 6 repeats only (Figure 6.3). The presence of this allele has been not reported in any of the published data.

6.6. STATISTICAL ANALYSIS

The allele frequencies for the seven loci were calculated for 144 male Egyptian individuals. Several alleles were observed for each locus.

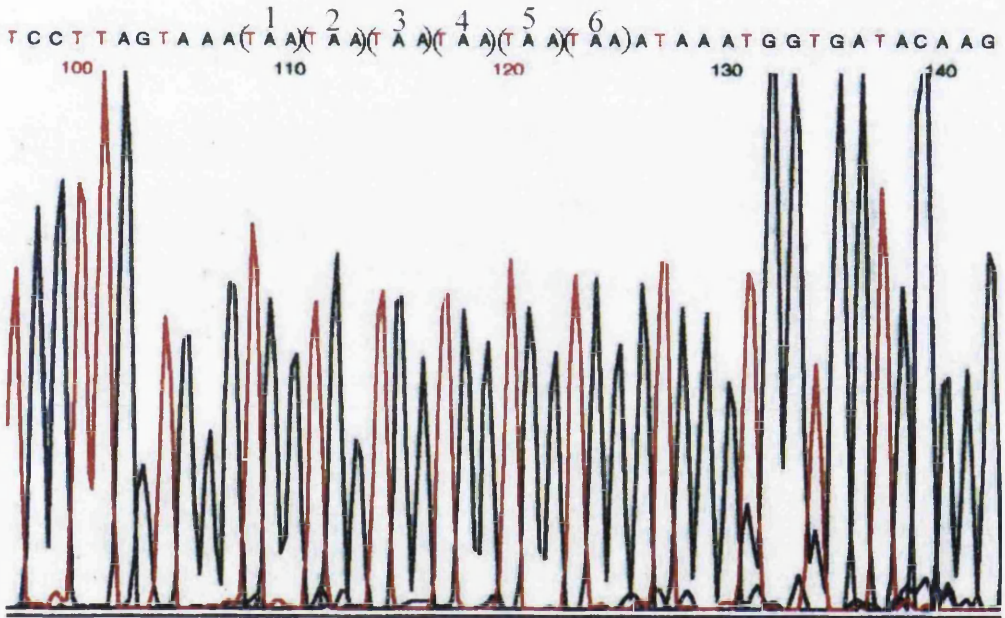
Three and four alleles were observed for the systems DYS391 and DYS389I respectively, the most common alleles having frequencies of 0.722 and 0.632.

Five and six alleles were also observed for the systems DYS393 and DYS19, the most common alleles showing frequencies of 0.409 and 0.611 respectively.

178A

Figure 6.3 Sequence of one sample that gave a new allele at DYS392 locus.

Sequencing of the sample was performed to detect the number and the structure of repeats. The result of the sequence showed 6 repeats only.



Seven alleles were observed at each of DYS398II, DYS390 and DYS392 loci, the most common allele showing a frequency of 0.368, 0.326 and 0.458 respectively. The results are shown in table 6.3 and figure 6.4.

The genetic diversity for each locus ranged from 0.44 at DYS391 locus to 0.776 at DYS390 locus. The genetic diversity for each locus is shown in table 6.4.

Non-pseudoautosomal Y loci are genetically linked, therefore typing of a single Y-STR locus can be extended to more discriminating Y-haplotype analysis (Cooper *et al.*, 1996). By combining the haplotype for each locus, there were 136 different haplotypes among 144 Egyptian individuals. Eight haplotypes occurred twice where the remaining 128 combinations were observed once (Table 6.5).

In Y-chromosome markers, the haplotype diversity value has the same value as the power of discrimination and the chance of exclusion, which is used in paternity calculations (Nei, 1973).

$$\text{Haplotype Diversity} = 1 - \sum p_i^2$$

Where p_i is the allele or haplotype frequencies

The seven Y-STRs showed overall haplotypes diversity of 0.992. This means that the seven loci had a probability of discrimination and probability of exclusion of 0.992.

Table 6.3 Alleles frequency for the examined Y-chromosome STR loci in the Egyptian population.

The number of the observed allele ranged from 3 alleles at DYS391 locus to 7 alleles at DYS398II, DYS390 and DYS392 loci.

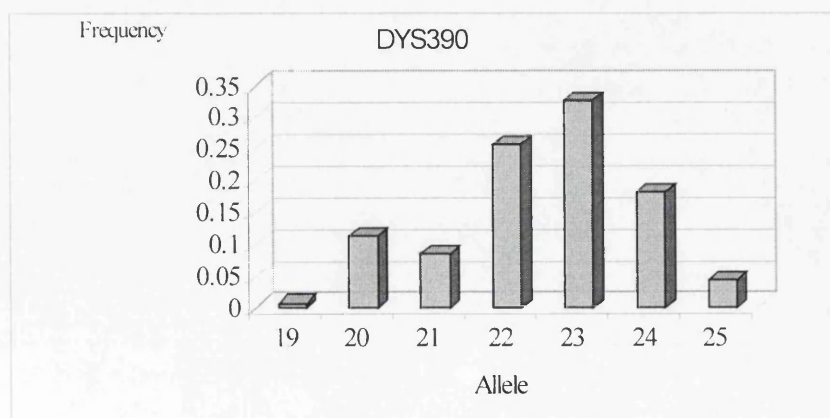
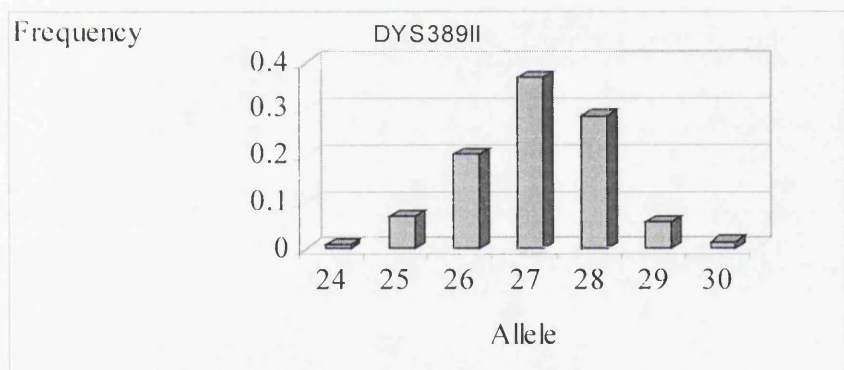
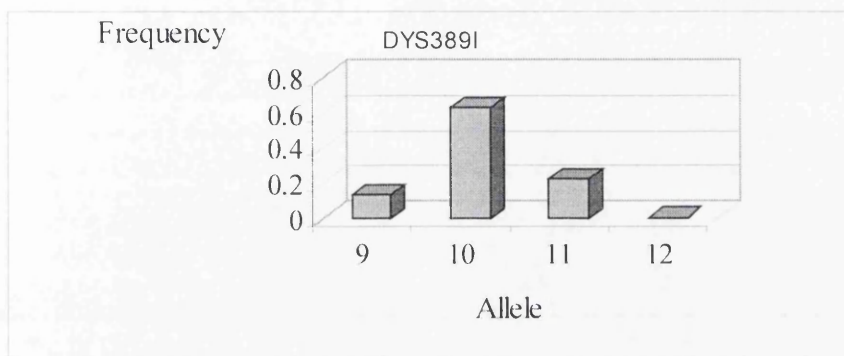
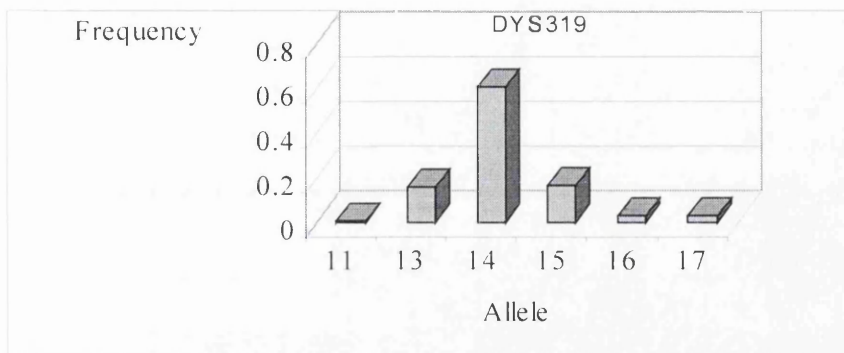
ALLELE	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393
6						0.014	
9		0.132			0.111		
10		0.632			0.722		
11	0.007	0.229			0.167	0.278	0.132
12		0.007				0.458	0.271
13	0.160					0.090	0.409
14	0.611					0.118	0.146
15	0.167					0.035	0.042
16	0.028					0.007	
17	0.028						
19				0.007			
20				0.111			
21				0.083			
22				0.257			
23				0.326			
24			0.007	0.181			
25			0.069	0.044			
26			0.201				
27			0.368				
28			0.285				
29			0.056				
30			0.014				

Table 6.4 Gene Diversity of the Y-Chromosome STRs

The genetic diversity for each locus ranged from 0.44 at DYS391 locus to 0.776 at DYS390 locus.

LOCUS	GENE DIVERSITY
DYS19	0.574
DYS389I	0.533
DYS389II	0.737
DYS890	0.776
DYS391	0.440
DYS392	0.692
DYS393	0.721

Figure 6.4 Allele frequencies for the 7 Y-chromosome STR.



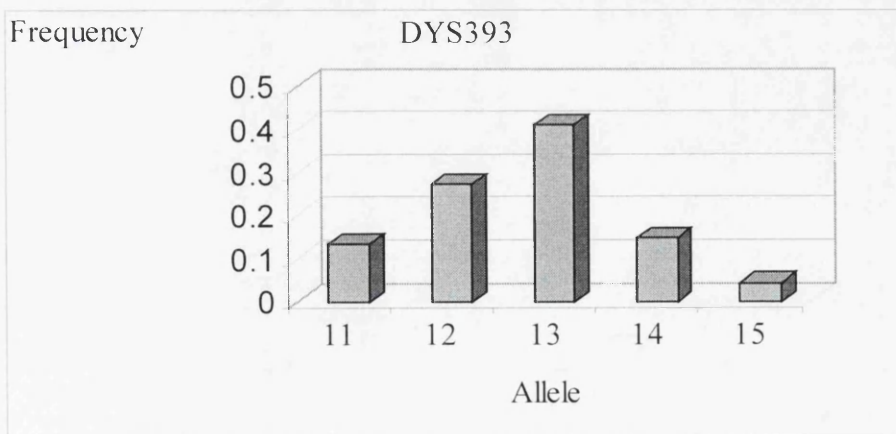
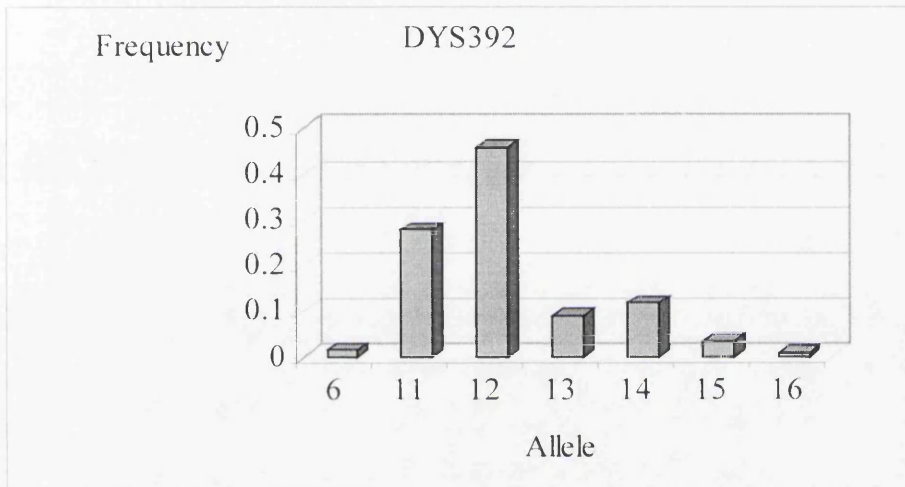
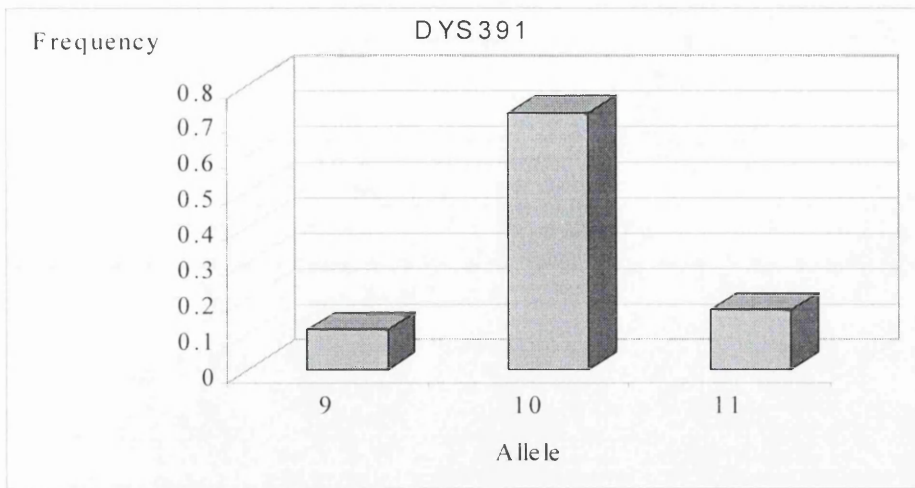


Table 6.5 Haplotype distributions of Y chromosome STRs in Egypt

By combining the haplotype for each locus, there were 136 different haplotypes among 144 Egyptian individuals. Eight haplotypes occurred twice where the remaining 128 combinations were observed once.

Haplotype no.	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	No.
1	11	10	28	22	10	11	12	1
2	13	9	25	23	9	12	12	1
3	13	9	26	22	10	12	13	1
4	13	10	26	20	10	11	13	1
5	13	10	27	21	10	11	13	1
6	13	10	27	23	10	6	13	1
7	13	10	27	23	10	11	13	1
8	13	10	27	24	9	12	12	1
9	13	10	27	24	10	12	13	1
10	13	10	27	24	10	14	13	1
11	13	10	28	23	9	11	13	1
12	13	10	28	23	11	11	11	1
13	13	10	28	25	10	11	11	1
14	13	10	29	24	10	11	11	1
15	13	10	29	25	10	12	13	1
16	13	11	27	22	9	11	13	2
17	13	11	27	22	9	13	13	1
18	13	11	27	22	10	12	12	1
19	13	11	27	22	11	14	13	1
20	13	11	28	24	10	12	13	1
21	13	11	28	25	10	11	13	1
22	13	11	30	21	10	13	11	1
23	13	12	30	23	10	11	15	1
24	14	9	25	22	10	12	13	1
25	14	9	25	23	10	14	14	1
26	14	9	25	23	11	13	13	1
27	14	9	25	23	11	13	14	1
28	14	9	26	22	9	13	12	1
29	14	9	26	23	10	12	12	1
30	14	9	26	23	10	12	13	1
31	14	9	28	21	10	11	12	1
32	14	10	26	20	10	14	12	1
33	14	10	26	22	10	6	14	1
34	14	10	26	22	10	11	11	1
35	14	10	26	22	10	12	13	2
36	14	10	26	22	10	15	13	1

37	14	10	26	22	10	15	14	1
38	14	10	26	22	11	14	13	1
39	14	10	26	23	10	11	11	1
40	14	10	26	23	10	12	13	1
41	14	10	26	23	10	13	12	1
42	14	10	26	23	10	14	12	1
43	14	10	26	23	11	13	12	1
44	14	10	26	24	10	14	13	1
45	14	10	26	24	11	11	13	1
46	14	10	26	25	10	16	13	1
47	14	10	27	19	10	12	13	1
48	14	10	27	20	9	11	12	1
49	14	10	27	20	10	12	12	1
50	14	10	27	21	10	13	13	1
51	14	10	27	22	10	12	12	1
52	14	10	27	22	10	12	14	1
53	14	10	27	22	10	15	13	1
54	14	10	27	22	11	12	13	2
55	14	10	27	23	9	12	14	1
56	14	10	27	23	10	11	12	1
57	14	10	27	23	10	11	13	1
58	14	10	27	23	10	12	13	1
59	14	10	27	23	10	12	14	1
60	14	10	27	24	10	11	14	1
61	14	10	27	24	10	11	15	1
62	14	10	27	24	10	12	12	1
63	14	10	27	24	10	12	13	1
64	14	10	27	24	10	12	15	1
65	14	10	27	24	10	13	12	1
66	14	10	27	24	10	13	14	1
67	14	10	27	24	10	14	14	1
68	14	10	27	24	11	12	13	1
69	14	10	27	24	11	12	14	1
70	14	10	27	24	11	13	14	1
71	14	10	28	20	9	11	11	2
72	14	10	28	20	10	11	11	1
73	14	10	28	20	10	11	13	1
74	14	10	28	21	10	11	11	1
75	14	10	28	21	10	12	14	1
76	14	10	28	22	10	12	12	1
77	14	10	28	22	10	12	13	2
78	14	10	28	22	10	14	14	1
79	14	10	28	23	9	12	13	1
80	14	10	28	23	10	12	12	2
81	14	10	28	23	10	12	14	1

82	14	10	28	23	10	14	13	1
83	14	10	28	23	11	12	12	1
84	14	10	28	23	11	15	13	1
85	14	10	28	24	9	11	13	1
86	14	10	28	24	10	12	13	1
87	14	10	29	20	10	11	10	2
88	14	10	29	20	10	11	15	1
89	14	10	29	20	10	12	12	1
90	14	10	29	24	10	12	11	2
91	14	11	26	21	10	14	15	1
92	14	11	27	22	10	12	13	1
93	14	11	27	22	10	14	12	1
94	14	11	27	22	11	11	12	1
95	14	11	27	22	11	12	14	1
96	14	11	27	23	10	14	12	1
97	14	11	27	23	10	15	12	1
98	14	11	27	23	11	14	12	1
99	14	11	27	24	11	13	11	1
100	14	11	27	25	11	11	12	1
101	14	11	28	20	10	11	12	1
102	14	11	28	23	10	12	12	1
103	14	11	28	23	10	12	13	1
104	14	11	28	23	11	12	11	1
105	15	9	24	21	10	12	13	1
106	15	9	25	21	10	11	12	1
107	15	9	25	23	10	12	13	1
108	15	9	25	23	10	14	11	1
109	15	9	25	23	10	14	12	1
110	15	9	25	24	10	12	13	1
111	15	9	26	21	10	12	12	1
112	15	9	26	22	11	12	13	1
113	15	9	27	21	9	12	13	1
114	15	10	26	21	10	12	14	1
115	15	10	26	23	10	11	12	1
116	15	10	26	23	10	12	12	1
117	15	10	26	23	10	12	13	1
118	15	10	27	22	10	11	12	1
119	15	10	27	22	10	11	13	1
120	15	10	27	22	10	12	14	1
121	15	10	27	22	10	14	12	1
122	15	10	28	20	10	11	13	1
123	15	10	28	23	10	12	13	1
124	15	11	27	23	11	13	14	1
125	15	11	27	24	11	12	11	1
126	15	11	28	20	11	12	12	1

127	15	11	28	22	10	12	13	1
128	15	11	28	23	10	12	14	1
129	16	10	26	20	10	11	15	1
130	16	10	27	22	9	11	11	1
131	16	11	28	23	10	12	12	1
132	16	11	28	24	10	12	13	1
133	17	11	28	22	9	12	13	1
134	17	11	28	23	10	12	11	1
135	17	11	28	23	10	12	13	1
136	17	11	28	23	10	12	14	1

6.7. COMPARISON WITH OTHER POPULATIONS

Comparison was performed for the seven Y-chromosome STR between the allele frequencies of the Egyptian population and the UAE population (M. Lathiqa) as an Arabic population. Comparison was also performed between the Egyptian population and the Italian (Ricci *et al.*, 2000) and Chinese populations (Hou *et al.*, 2000). The frequency profile comparison was carried out using the two-way RxC contingency table test (Power Instat 2.03). The results demonstrated significant differences between the Egyptian population and the compared populations in all the seven loci except for DYS19 and DYS389I loci with the Italian population and DYS393 locus with the UAE population. The results are shown in table 6.6.

6.8. DISCUSSION

The optimised PCR cycling conditions were the same for both multiplexes. The optimised PCR parameters including primer concentrations, MgCl₂, enzyme concentration and extension temperature were differed from another study (Kayser *et al.*, 1997).

According to the recommendations of the International Society of Forensic Haemogenetics (1994), the number of variable repeats had been used to designate the various Y-STR alleles. Confirmation of the designation of the alleles was performed by comparison the allelic ladders.

Table 6.6 X² value of the comparison between the Egyptian and other populations.

There were significant differences between the Egyptian population and the compared populations in all the seven loci except at DYS19 and DYS389I loci with the Italian population and at DYS393 locus with UAE population. (*There was no significant difference).

DYS19	X ²	df	p
Egypt and UAE	15.168	5	0.0097
Egypt and Italy	2.44	5	0.0785*
Egypt and China	72.2	6	< 0.0001
DYS389I			
Egypt and UAE	67.227	4	< 0.0001
Egypt and Italy	3.3158	3	0.345*
Egypt and China	217.58	5	< 0.0001
DYS389II			
Egypt and UAE	46.701	6	< 0.0001
Egypt and Italy	17.428	6	0.0078
Egypt and China	104.91	9	< 0.0001
DYS390			
Egypt and UAE	32.512	7	< 0.0001
Egypt and Italy	76.357	8	< 0.0001
Egypt and China	54.68	7	< 0.0001
DYS391			
Egypt and UAE	10.573	3	0.0143
Egypt and Italy	36.153	3	< 0.0001
Egypt and China	10.6	3	0.0141
DYS392			
Egypt and UAE	46.894	7	< 0.0001
Egypt and Italy	60.317	6	< 0.0001
Egypt and China	63.973	7	< 0.0001
DYS393			
Egypt and UAE	5.1803	5	0.3943*
Egypt and Italy	16.528	4	0.0024
Egypt and China	18.803	4	0.0009

The haplotype diversity and the exclusion probability were 0.992. The haplotype diversity for the seven loci was 0.97 in both Thailand and Japanese populations and 0.98 in both Chinese and Macau populations (Horst *et al.*, 1999 and Gusmao *et al.*, 2000). In another study the mean exclusion probability was 0.99 for the haplotypes of 10 loci as all haplotypes occurred only once (Kayser *et al.*, 1998).

The comparison of the alleles frequency between the Egyptian population and the UAE, Italian and Chinese populations demonstrated the presence of significant differences between the Egyptian population and the compared populations for most of the loci except for DYS19 and DYS389I loci with Italian population and DYS393 locus with UAE population. Similar differences have been observed for a variety of other loci emphasising the need for referring to population-specific databases (Cao *et al.*, 1989).

It was approved that inclusion of Y-STR loci in paternity testing of a male offspring will increase the chance of detecting non-paternity (Kayser *et al.*, 1998).

The application of Y-chromosome STRs to paternity testing in non-exclusions is relatively limited as every member of the male lineage of the alleged father has the same probability of being the biological father. However, it is better in paternity testing to use every available tool of evidence to solve a given case. The Y-chromosomal DNA analysis closes a gap in paternity testing because of the clarity of the exclusion and the high probabilities for non-exclusions in the studied deficiency cases (Kayser *et al.*, 1998).

CHAPTER 7: PATERNITY CASE WORK

7.1. INTRODUCTION

The purpose of the paternity testing is to either exclude an alleged father or provide a high likelihood of paternity in the case of an inclusion (Strom *et al.*, 1996). For each STR tested in a child, one allele must have come from the mother and the other allele must be present in the true father. Because of mutation events 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 and Stivers, 1996). In inclusion cases the probability that the alleged father is in fact the biological father is determined by comparing the observed alleles to their distribution in the general population and expressed as either paternity index or the probability of paternity (Strom *et al.*, 1996).

The use of six STR loci used in this study was investigated in 30 Egyptian paternity cases. For every family the alleles of the child were compared to the mother and the alleged father. Inclusion and exclusion of the alleged father was determined for each locus and then combined for all the six loci. In inclusion cases the paternity index was calculated for each locus using allele frequency databases developed in this study. The combined paternity index was then calculated by using the product rule. The probability of paternity was calculated for each case using the combined paternity index assuming a prior probability of 50% (Gjertson *et al.*, 1988).

7.2. PATERNITY CASES

Inclusion was determined in all 30 cases. The paternity index was calculated using 3 methods, which are based on different assumptions, (1) the alleged father was not related to the biological father, (2) the alleged father, the biological father and the mother belonged to the same subpopulation (3) the alleged father was a relative to the biological father.

The 30 cases were also investigated for motherless cases by ignoring the genotype of the mother. The same three assumptions were used also in the calculation of paternity index in the motherless cases. The probability of paternity was calculated for each case for the complete trio family and for resembling motherless case. The results are shown in table 7.1.

It was found that there was very little evidence of substructure in the Egyptian population ($F_{ST} = 0.0011$). Therefore when the value of F_{ST} was used in the second method assumption, there was no marked difference between the calculated PI in both assumptions (1) and (2) as shown in first 2 cases, therefore the second assumption was not used in the calculation of PI in the remaining cases. In the third assumption the $\theta_{AT} = 0.25$ was used in the calculation of PI considering the closest relationship between the alleged father and the biological father (brother-brother or father-child). The full description of all equations is shown in appendix 3.

Table 7.1 30 Paternity cases

The genotype of mother, child and father were determined by identifying their alleles. The paternity index (PI) was calculated for each locus by three methods as follow: the standard method (1), which assuming that the alleged father was unrelated to the biological father. The second method (2), which assumes that the alleged father, the biological father and the mother belong to the substructure using F_{ST} value and the third method (3) assumes that the alleged father was relative to the biological father. The three methods were used in trio complete case where the genotype of the mother was available and in resembling motherless case by removing the genotype of the mother.

No 1	D3	vWA	FGA	D11	SE33	APOA	CPI	W%
Mother	15,16	16,17	20,22	6,12	15,30	6,22		
Child	15,17	15,17	20,21	11,12	15,29	7,22		
Father	16,17	15,15	21,24	4,11	23,29	7,21		
PI Trio (1)	2.16	6.52	3.66	3.98	11.99	4.23	10626	99.991%
(2)	2.15	6.46	3.64	3.97	11.73	4.21	9912	99.990%
(3)	1.37	1.73	1.57	1.60	1.85	1.62	18	94.737%
Mother less (1)	1.08	3.26	1.83	1.99	5.99	2.12	163	99.390%
(2)	1.08	3.25	1.83	1.99	5.92	2.11	154	99.355%
(3)	1.04	1.53	1.29	1.33	1.71	1.36	6	85.714%

No 2	D3	vWA	FGA	D11	SE33	APOA	CPI	W%
Mother	16,17	17,17	21,22	15,26	11,28	5,9		
Child	16,17	15,17	21,22	9,15	28,35	5,9		
Father	17,17	15,16	22,22	6,9	13,35	9,12		
PI Trio (1)	1.97	3.26	3.43	4.82	6.82	9.62	6966	99.986%
(2)	1.96	3.25	3.39	4.79	6.75	9.46	6605	99.985%
(3)	1.33	1.53	1.55	1.66	1.86	1.9	19	95%
Mother less (1)	2.16	1.63	3.23	2.41	3.41	4.81	450	99.778%
(2)	2.15	1.63	3.21	2.41	3.40	4.77	440	99.773%
(3)	1.34	1.24	1.53	1.41	1.55	1.66	9	90%

No 3	D3	vWA	FGA	D11	SE33	APOA	CPI	W%
Mother	14,16	16,17	19,22	6,17	30,38	16,22		
Child	14,18	15,16	22,23	6,9	30,31	7,16		
Father	15,18	15,19	20,23	7,9	19,31	7,12		
PI Trio (1)	4	3.26	3.03	4.82	13.06	4.23	10520	99.991%
(3)	1.6	1.53	1.50	1.66	1.98	1.62	20	95.238%
Mother less (1)	2	1.63	1.52	2.41	6.53	2.12	165	99.398%
(3)	1.33	1.24	1.20	1.41	1.73	1.36	7	87.5%

No. 4	D3	vWA	FGA	D11	SE33	APOA	CPI	W%
Mother	15,16	14,17	21,22	16,18	26,28	16,16		
Child	15,16	15,17	20,22	12,16	15,28	16,18		
Father	16,18	15,19	18,20	12,26	15,33	18,21		
PI Trio (1)	0.93	3.26	5.09	9.06	5.36	3.49	2615	99.962%
(3)	0.97	1.53	1.67	1.80	1.69	1.55	12	92.308%
Mother less (1)	0.90	1.63	2.55	4.53	2.68	1.75	79	98.75%
(3)	0.95	1.24	1.44	1.64	1.46	1.27	5	83.333%

No. 5	D3	vWA	FGA	D11	SE33	APOA	CPI	W%
Mother	14,17	15,18	21,24	11,12	17,35	1,9		
Child	15,17	15,19	24,25	11,29	31,35	1,18		
Father	15,16	16,19	20,25	1,29	11,31	18,24		
PI Trio (1)	1.94	6.00	5.09	4.82	13.06	3.49	50391	99.992%
(3)	1.32	1.71	1.67	1.66	1.98	1.55	19	95%
Mother less (1)	0.97	3	2.55	2.41	6.53	1.75	204	99.512%
(3)	0.98	1.5	1.44	1.41	1.73	1.27	7	87.5%

No. 6	D3	vWA	FGA	D11	SE33	APOA	CPI	W%
Mother	14,16	15,15	19,22	4,15	17,23	16,16		
Child	14,18	15,16	22,24	4,11	17,38	7,16		
Father	15,18	16,19	20,24	6,11	8,38	7,24		
PI Trio (1)	4	2.09	3.37	3.98	11.99	4.23	5687	99.982%
(3)	1.6	1.35	1.54	1.60	1.85	1.62	16	94.118%
Mother less (1)	2	1.05	1.69	1.99	5.99	2.12	90	98.901%
(3)	1.33	1.02	1.26	1.33	1.71	1.36	5	83.333%

No. 7	D3	vWA	FGA	D11	SE33	APOA	CPI	W%
Mother	14,17	16,19	19,22	6,19	13,19	18,18		
Child	14,16	16,18	22,23	19,19	19,36	5,18		
Father	15,16	15,18	20,23	11,19	15,36	5,21		
PI Trio (1)	1.81	3.23	3.03	9.65	8.33	6.97	9925	99.990%
(3)	1.29	1.53	1.50	1.81	1.79	1.75	17	94.444%
Mother less (1)	0.96	1.62	1.52	9.65	4.17	3.49	332	99.700%
(3)	0.95	1.23	1.20	0.95	1.61	1.55	3	75%

No. 8	D3	vWA	FGA	D11	SE33	APOA	CPI	W%
Mother	17,18	16,18	20,21	1,29	11,31	16,24		
Child	17,18	14,16	20,24	11,29	31,35	16,24		
Father	15,17	14,17	22,24	11,12	17,35	1,16		
PI Trio (1)	1.40	5.36	3.37	3.98	6.82	3.49	2396	99.958%
(3)	1.18	1.69	1.54	1.60	1.86	1.55	14	93.333%
Mother less (1)	1.08	2.68	1.69	1.99	3.41	1.75	58	98.305%
(3)	1.04	1.46	1.26	1.33	1.55	1.27	5	83.333%

No. 9	D3	vWA	FGA	D11	SE33	APOA	CPI	W%
Mother	15,16	16,17	20,21	9,14	13,36	12,16		
Child	14,15	15,16	20,23	9,28	17,36	7,12		
Father	14,17	15,18	23,23	7,28	17,17	7,9		
PI Trio (1)	5.88	3.26	9.38	10.31	10.91	4.23	85550	99.999%
(3)	1.71	1.53	1.72	1.82	1.98	1.62	26	96.296
Mother less (1)	2.94	1.63	4.69	5.16	5.45	2.12	1340	99.925%
(3)	1.49	1.24	1.50	1.66	1.69	1.36	11	91.667%

No. 10	D3	vWA	FGA	D11	SE33	APOA	CPI	W%
Mother	14,16	14,18	24,24	12,29	19,19	2,7		
Child	14,17	16,18	24,28	4,12	19,19	7,21		
Father	17,17	16,19	23,28	4,6	19,28	7,21		
PI Trio (1)	4.32	2.09	18.73	9.65	5.45	4.17	37088	99.997%
(3)	1.62	1.35	1.90	1.81	1.69	1.61	21	95.455%
Mother less (1)	2.16	1.05	9.37	4.83	5.45	4.2	2350	99.957%
(3)	1.37	1.02	1.81	1.66	0.92	0.4	2	66.667%

No. 11	D3	vWA	FGA	D11	SE33	APOA	CPI	W%
Mother	16,16	15,18	18,20	7,11	13,23	12,15		
Child	15,16	16,18	20,21	7,17	13,29	7,12		
Father	15,17	16,17	21,25	17,19	17,29	7,19		
PI Trio (1)	1.94	2.09	3.66	8.31	11.99	4.23	6255	99.984%
(3)	1.32	1.35	1.57	1.79	1.89	1.62	15	93.75%
Mother less (1)	0.97	1.05	1.83	4.16	5.99	2.12	99	99%
(3)	0.98	1.02	1.29	1.61	1.80	1.36	5	83.333%

No. 12	D3	vWA	FGA	D11	SE33	APOA	CPI	W%
Mother	16,17	15,15	20,21	6,13	9,17	7,16		
Child	15,16	15,16	21,24	6,14	17,35	5,7		
Father	15,15	16,18	22,24	9,14	7,35	5,5		
PI Trio (1)	3.87	2.09	3.37	10.31	6.82	13.95	26737	99.996%
(3)	1.59	1.35	1.54	1.82	1.74	1.87	20	95.238%
Mother less (1)	1.94	1.05	1.69	5.16	3.41	6.98	423	99.764%
(3)	1.32	1.02	1.26	1.66	1.55	1.75	8	88.889%

No. 13	D3	vWA	FGA	D11	SE33	APOA	CPI	W%
Mother	15,17	16,18	21,22	7,11	19,31	5,9		
Child	17,17	17,18	21,22	7,11	9,19	9,21		
Father	17,17	17,18	21,21	7,16	9,28	18,21		
PI Trio (1)	4.32	2.04	3.43	2.26	15.15	4.17	4316	99.977%
(3)	1.62	1.34	1.55	1.39	1.88	1.61	14	93.333%
Mother less (1)	4.32	2.64	3.66	2.63	7.58	2.09	1739	99.943%
(3)	1.62	0.4	1.57	1.68	1.76	1.35	4	80%

No. 14	D3	vWA	FGA	D11	SE33	APOA	CPI	W%
Mother	13,16	16,16	23,24	4,7	12,17	3,15		
Child	16,16	16,19	20,23	7,9	15,17	3,18		
Father	16,18	17,19	20,22	9,9	15,29	18,18		
PI Trio (1)	1.8	6	5.09	9.64	5.36	6.98	19826	99.995%
(3)	1.29	1.71	1.67	1.81	1.69	1.75	20	95.238%
Mother less (1)	1.8	3	2.55	4.82	2.68	3.49	621	99.839%
(3)	0.78	1.50	1.44	1.66	1.46	1.55	6	85.714%

No. 15	D3	vWA	FGA	D11	SE33	APOA	CPI	W%
Mother	16,16	17,17	21,23	11,21	17,29	7,18		
Child	15,16	16,17	22,23	17,21	29,36	18,21		
Father	13,15	16,18	22,22	12,17	29,36	9,21		
PI Trio (1)	1.94	2.09	6.45	8.31	8.33	4.17	7579	99.987%
(3)	1.32	1.35	1.73	1.79	1.79	1.61	16	94.118%
Mother less (1)	0.97	1.05	3.23	4.26	10.17	2.09	298	99.666%
(3)	0.98	1.02	1.53	1.61	0.4	1.35	1	50%

No. 16	D3	vWA	FGA	D11	SE33	APOA	CPI	W%
Mother	16,17	17,18	22,26	11,12	19,42	7,7		
Child	16,16	15,17	22,22	11,14	19,35	7,16		
Father	16,16	15,16	20,22	14,16	21,35	16,16		
PI Trio (1)	3.61	3.26	3.23	10.31	6.82	7.14	19084	99.995%
(3)	1.57	1.53	1.53	1.82	1.86	1.75	22	95.652%
Mother less (1)	3.61	1.63	3.23	5.16	3.41	3.57	1194	99.916%
(3)	1.57	1.24	0.87	1.68	1.55	1.56	7	87.5%

No. 17	D3	vWA	FGA	D11	SE33	APOA	CPI	W%
Mother	15,17	16,18	22,23	9,9	13,15	5,21		
Child	17,18	15,16	21,22	9,19	15,19	16,21		
Father	16,18	15,17	19,21	11,19	19,19	16,24		
PI Trio (1)	4	3.26	3.66	9.65	10.91	3.57	17938	99.994%
(3)	1.6	1.53	1.57	1.81	1.83	1.56	20	95.238%
Mother less (1)	2	1.63	1.83	4.83	5.46	1.79	282	99.646%
(3)	1.33	1.24	1.29	1.66	1.69	1.28	8	88.889%

No. 18	D3	vWA	FGA	D11	SE33	APOA	CPI	W%
Mother	14,15	15,16	22,22	6,6	17,30	14,18		
Child	15,17	16,17	22,23	6,28	17,40	7,14		
Father	16,17	17,17	21,23	7,28	15,40	7,19		
PI Trio (1)	2.16	4.08	3.03	10.31	17.67	4.23	20578	99.995%
(3)	1.37	1.61	1.50	1.82	1.89	1.62	18	94.737%
Mother less (1)	1.08	2.04	1.52	5.16	8.84	2.12	324	99.692%
(3)	1.04	1.32	1.20	1.68	1.80	1.36	7	87.5%

No. 19	D3	vWA	FGA	D11	SE33	APOA	CPI	W%
Mother	15,16	17,17	19,24	11,12	11,19	7,16		
Child	15,18	16,17	21,24	12,14	15,19	16,21		
Father	16,18	16,19	21,21	14,28	15,17	21,21		
PI Trio (1)	4	2.09	7.32	10.31	5.36	8.33	28170	99.996%
(3)	1.6	1.35	1.60	1.82	1.69	1.79	19	95%
Mother less (1)	2	1.05	3.66	5.26	2.68	4.17	452	99.779%
(3)	1.33	1.02	1.57	1.68	1.46	1.61	8	88.889%

No. 20	D3	vWA	FGA	D11	SE33	APOA	CPI	W%
Mother	17,18	16,17	20,21	6,7	15,15	2,5		
Child	17,18	16,16	21,23	6,19	15,19	5,24		
Father	17,17	15,16	22,23	19,28	19,26	16,24		
PI Trio (1)	2.8	2.09	3.03	9.65	5.45	25	23314	99.996%
(3)	1.47	1.35	1.50	1.81	1.69	1.92	18	94.737%
Mother less (1)	2.16	2.09	1.52	4.83	2.73	12.5	1131	99.911%
(3)	1.37	0.81	1.20	1.66	1.46	1.85	6	85.714%

No. 21	D3	vWA	FGA	D11	SE33	APOA	CPI	W%
Mother	15,15	17,17	20,21	3,11	17,19	5,20		
Child	15,16	17,18	21,24	11,12	19,19	5,9		
Father	15,16	15,18	24,24	6,12	19,35	9,9		
PI Trio (1)	1.81	3.23	6.74	9.06	5.45	19.23	37415	99.997%
(3)	1.29	1.53	1.74	1.95	1.69	1.9	22	95.652%
Mother less (1)	1.81	1.62	3.37	4.53	5.45	9.62	2438	99.959%
(3)	0.4	1.23	1.54	1.65	0.92	1.81	2	66.667%

No. 22	D3	vWA	FGA	D11	SE33	APOA	CPI	W%
Mother	16,16	16,18	22,22	3,11	19,38	6,16		
Child	16,18	16,19	21,22	7,11	15,19	16,16		
Father	17,18	17,19	21,24	7,21	15,15	5,16		
PI Trio (1)	4	6	3.66	5.25	10.72	3.57	17649	99.994%
(3)	1.6	1.71	1.57	1.68	1.83	1.56	21	95.455%
Mother less (1)	2	3	1.83	2.63	5.36	3.57	553	99.820%
(3)	1.33	1.5	1.29	1.45	1.69	0.88	6	85.714%

No. 23	D3	vWA	FGA	D11	SE33	APOA	CPI	W%
Mother	15,15	16,17	19,22	12,26	9,19	21,21		
Child	15,17	14,17	22,23	9,12	19,38	9,21		
Father	17,18	14,17	22,23	9,15	38,40	9,18		
PI Trio (1)	2.16	5.36	3.03	4.82	11.99	9.62	19504	99.995%
(3)	1.37	1.69	1.50	1.66	1.89	1.81	20	95.238%
Mother less (1)	1.08	3.7	3.14	2.41	5.99	4.81	871	99.885%
(3)	1.04	0.4	0.4	1.41	1.80	1.66	1	50%

No. 24	D3	vWA	FGA	D11	SE33	APOA	CPI	W%
Mother	14,16	15,15	21,21	11,11	22,28	5,12		
Child	16,17	15,17	21,22	11,15	22,29	5,16		
Father	16,17	16,17	22,22	7,15	29,38	16,16		
PI Trio (1)	2.16	2.04	6.45	9.65	11.99	7.14	23480	99.996%
(3)	1.37	1.34	1.73	1.83	1.89	1.75	19	95%
Mother less (1)	2.13	1.02	3.23	4.83	5.99	3.57	725	99.862%
(3)	0.4	1.01	1.53	1.69	1.80	1.56	3	75%

No. 25	D3	vWA	FGA	D11	SE33	APOA	CPI	W%
Mother	15,16	16,16	20,21	4,6	9,23	12,12		
Child	16,17	16,17	20,24	6,11	17,23	12,18		
Father	16,17	17,17	23,24	11,11	15,17	18,18		
PI Trio (1)	2.16	4.08	3.37	7.97	5.45	6.98	9004	99.989%
(3)	1.37	1.61	1.54	1.73	1.69	1.75	17	94.444%
Mother less (1)	2.13	2.04	1.69	3.99	2.73	3.49	279	99.643%
(3)	0.4	1.32	1.26	1.60	1.46	1.55	2	66.667%

No. 26	D3	vWA	FGA	D11	SE33	APOA	CPI	W%
Mother	15,16	17,17	21,21	12,21	15,17	5,7		
Child	15,16	17,19	20,21	4,12	15,17	5,18		
Father	16,17	16,19	20,22	4,9	17,17	18,24		
PI Trio (1)	0.93	6	5.09	9.65	10.91	3.49	10436	99.990%
(3)	1.33	1.71	1.67	1.81	1.83	1.55	20	95.238%
Mother less (1)	0.90	3	2.55	4.83	5.46	1.75	318	99.687%
(3)	0.94	1.5	1.44	1.66	1.69	1.27	7	87.5%

No. 27	D3	vWA	FGA	D11	SE33	APOA	CPI	W%
Mother	16,18	17,17	22,23	11,11	15,38	1,5		
Child	16,17	17,18	21,23	11,17	15,19	1,16		
Father	17,17	15,18	21,21	9,17	19,23	16,22		
PI Trio (1)	4.32	3.23	7.32	8.31	5.45	3.57	16514	99.994%
(3)	1.62	1.53	1.76	1.79	1.69	1.56	21	95.455%
Mother less (1)	2.16	1.62	3.66	4.26	2.73	1.79	267	99.627%
(3)	1.37	1.23	1.57	1.61	1.46	1.28	8	88.889%

No. 28	D3	vWA	FGA	D11	SE33	APOA	CPI	W%
Mother	17,17	17,17	22,23	11,12	15,40	7,26		
Child	16,17	16,17	23,25	9,11	17,40	7,18		
Father	16,16	16,18	21,25	6,9	17,38	18,18		
PI Trio (1)	3.61	2.09	5.09	4.82	11.99	6.98	15492	99.994%
(3)	1.57	1.35	1.67	1.66	1.89	1.75	19	95%
Mother less (1)	1.81	1.05	2.55	2.41	5.99	3.49	244	99.592%
(3)	1.29	1.02	1.44	1.41	1.80	1.55	8	88.889%

No. 29	D3	vWA	FGA	D11	SE33	APOA	CPI	W%
Mother	16,16	15,15	22,24	19,29	17,17	12,18		
Child	16,17	15,16	21,24	7,19	17,30	5,12		
Father	17,18	16,17	21,22	7,26	15,30	5,7		
PI Trio (1)	2.16	2.09	3.66	5.25	11.99	6.97	7249	99.986%
(3)	1.37	1.35	1.57	1.68	1.85	1.75	16	94.118%
Mother less (1)	1.08	1.05	1.83	2.63	5.99	3.49	114	99.130%
(3)	1.04	1.02	1.29	1.45	1.71	1.55	5	83.333%

No. 30	D3	vWA	FGA	D11	SE33	APOA	CPI	W%
Mother	15,16	16,19	22,22	9,11	15,30	7,9		
Child	15,17	16,17	22,23	6,9	30,31	7,16		
Father	16,17	17,18	20,23	6,19	31,35	16,21		
PI Trio (1)	2.16	2.04	3.03	8.79	13.05	3.57	5468	99.982%
(3)	1.37	1.34	1.50	1.80	1.86	1.56	14	93.333%
Mother less (1)	1.08	1.02	1.52	4.39	6.53	1.79	86	98.851%
(3)	1.04	1.01	1.20	1.63	1.73	1.28	5	83.333%

All 30 cases were inclusions. The combined paternity index calculated by method (1) was over 4000 in all trio cases except case number 4 and 8. The probability of paternity for the trio cases was derived from the PI calculated by method (1) showed that 10 cases had a probability of paternity less than 99.991%; 1 case had a probability of paternity more than 99.999% and the remaining 19 cases had a probability of paternity between 99.991% and 99.999%. No mutation was detected for any the six loci.

The average paternity index calculated for both trio and motherless cases using method (1) was compared with the calculated typical paternity index for each locus (Table 7.2). The average PI for the trio cases was relatively the same as the typical PI for the simple examined loci while it was higher in the hypervariable loci than the typical PI. In motherless cases the average PI was $\leq 61\%$ from the typical PI for each locus of the simple three loci, while the average PI was $\geq 74\%$ from the typical PI for each locus of the hypervariable loci. The average PI obtained by using method (3) was low in each locus.

The values of the probability of paternity of the trio cases were higher than those of the motherless cases calculated by both methods but the difference was small using by method (1) than using method (3) (Tables 7.2, 7.3 and Figure 7.1).

Table 7.2 Comparison between the average and typical PI using method (1).

The average PI for the trio cases was relatively the same as the typical PI for the simple examined loci while it was higher in the hypervariable loci than the typical PI. In motherless cases the average PI was $\leq 61\%$ from the typical PI for the simple three loci, while the average PI was $\geq 74\%$ from the typical PI for the hypervariable loci. The difference between the average probability of paternity of trio and motherless cases was small.

Locus	METHOD (1)		Typical PI
	Average PI Trio cases	Average PI Motherless	
D3S1358	2.81	1.72 (61%)	2.82
vWA	3.48	1.83 (52%)	3.49
FGA	4.93	2.69 (59%)	4.55
D11S554	7.58	4.02 (78%)	5.17
SE33	8.49	5.16 (77%)	6.82
APOA11	6.7	3.48 (74%)	4.69
Combined	20787	611	7596
W%	99.995%	99.837%	99.987%

Table 7.3 Comparison between the average and typical PI using method (3).

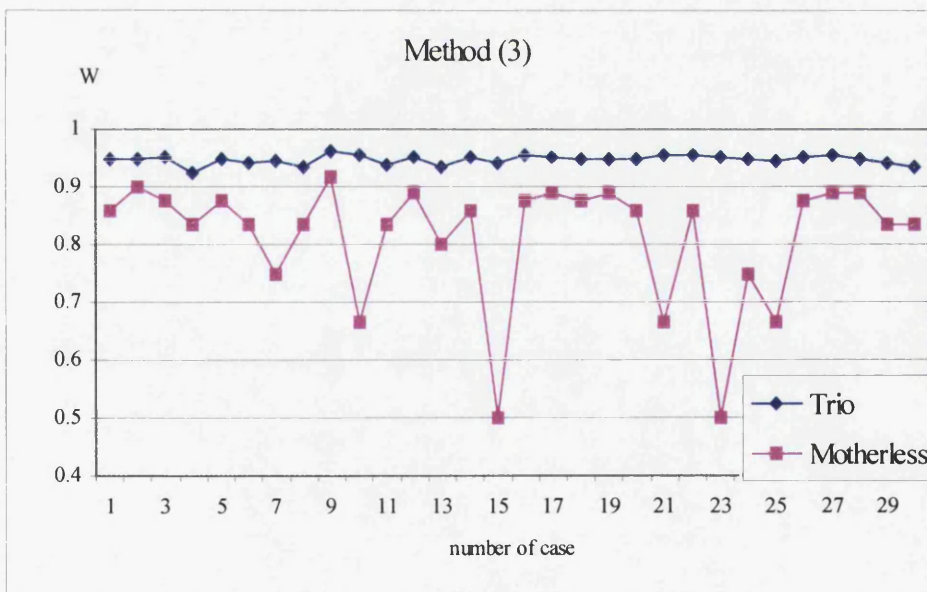
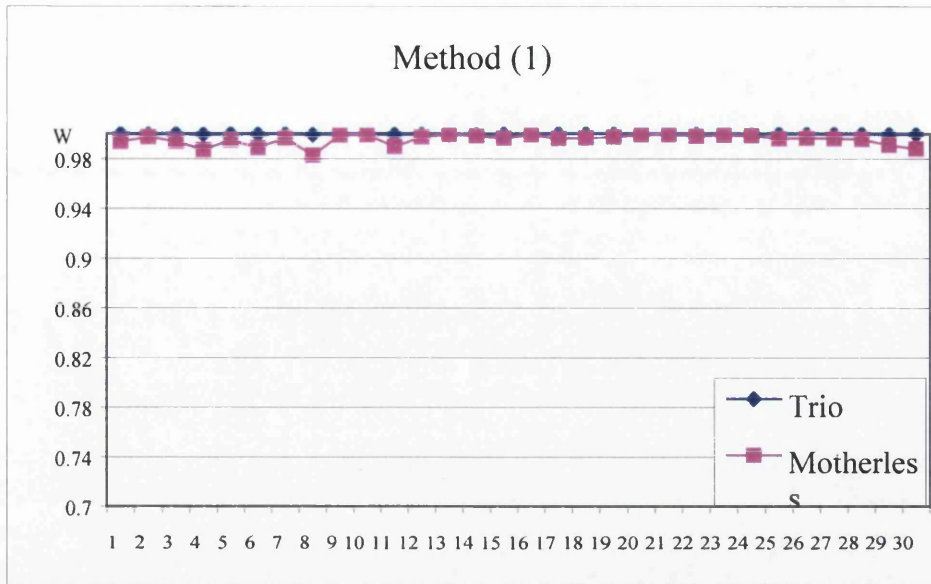
The average PI was low for each locus in both trio and motherless cases. The difference between the average probability of paternity of trio and motherless cases was high.

Locus	METHOD (3)	
	Average PI Trio cases	Average PI Motherless
D3S1358	1.42	1.11
vWA	1.50	1.15
FGA	1.61	1.33
D11S554	1.74	1.55
SE33	1.82	1.56
APOAI1	1.68	1.41
Combined	18	6
W%	94.737%	84.714%

208A

Figure 7.1 The difference of the probability of paternity in trio and motherless cases.

The probabilities of paternity were calculated in trio and motherless cases using methods (1) and (3). The values of the probability of paternity of the trio cases were higher than those of the motherless cases calculated by both methods but the difference was small using method (1) than using method (3).



7.3. EXCLUSION POWER ANALYSIS

The use of the six loci to exclude non-father from paternity was investigated. Each known mother and child of one case was tested with each of the 29 unrelated men from every other case. This assumption method allowed the construction of 870 families. Each locus was examined and the number of excluding loci in each case was calculated. The observed exclusion power of each locus was also calculated based on the results of the exclusion of each locus in the 870 constructed families.

No mismatch was observed and all the assumed 870 family cases showed exclusion. 5 cases (0.57%) showed a single exclusion and the remaining 865 cases (99.43%) showed at least two or more exclusions (Figure 7.2).

The calculated observed exclusion power of each locus was compared with the expected exclusion power (Table 7.4). In simple STR loci the observed exclusion power was less than the expected exclusion power in the three loci. While in hypervariable loci the value of the observed exclusion power was higher than the expected exclusion power in the three hypervariable loci. The combined observed and expected exclusion powers were relatively the same.

RIOA

Figure 7.2 The exclusion power of paternity

The assumed 870 family cases were investigated to determine the exclusion power of paternity. There were 5 cases (0.57%) showed a single exclusion and the remaining 865 cases (99.43%) showed at least two or more exclusions.

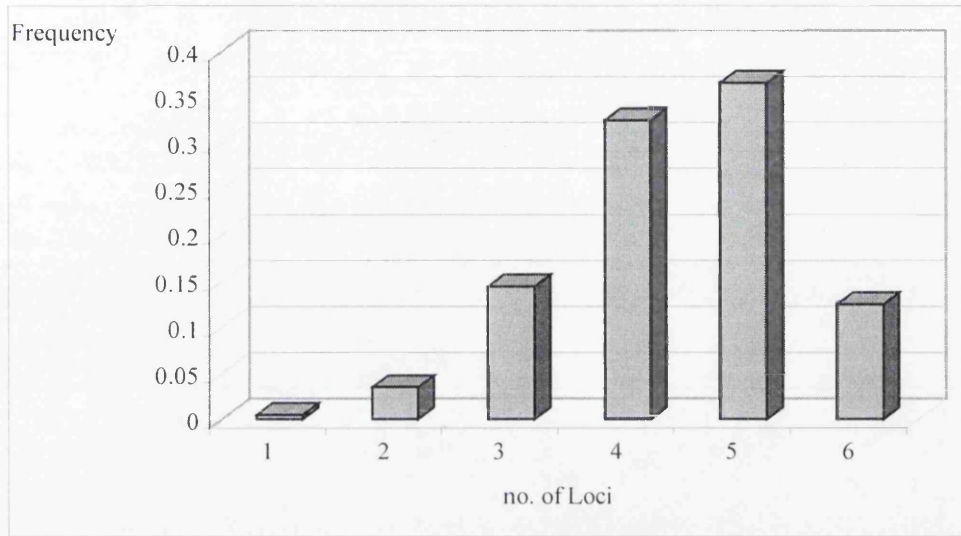


Table 7.4 Comparison of the observed and expected exclusion power for the six loci.

In simple STR loci the observed exclusion power was less than the expected exclusion power in the three loci and in hypervariable loci the value of the observed exclusion power was higher than the expected exclusion power in the three hypervariable loci.

LOCUS	OBSERVED	EXPECTED
D3S1358	0.540	0.642
vWA	0.654	0.708
FGA	0.745	0.775
D11S554	0.850	0.802
ACTBP2	0.853	0.850
APOA11	0.795	0.782
Combined	0.99982	0.99986

7.4. DISCUSSION

In paternity testing the standard established method assumes that the alleged father and the true father are unrelated. Recently the coancestry factor of the population substructure (Balding and Nichols, 1994) and the relatedness of two persons have been incorporated in the interpretation of DNA evidence (Lee *et al.*, 2000). Three methods were used in the calculation of PI including the standard method (1), the use of F_{ST} value in method (2) and assuming the closest relationship between the alleged and the biological father (method 3). It was calculated that the coancestry factor of the Egyptian population substructure was (0.0011). This value is small and had negligible effect in the calculation of PI. Therefore only methods (1 and 3) were used in the calculation of PI. As a conservative method, the $\theta_{AT} = 0.25$ was used in the calculation of PI in method (3) as it is the value for the closest degree of relatedness (Lee *et al.*, 2000).

Using the standard method, the combined paternity index for all except 2 cases was more than 4000. A combined paternity index equal to or greater than 500 is generally considered as a conclusive evidence of paternity (Rand *et al.*, 1992). In another study a combined paternity index more than 1000 is the limit usually accepted by most of the courts in paternity cases (Domenici *et al.*, 1998). Another study recommended that the combined paternity index more than 2000, which means that even if the prior probability is 0.05, the probability of paternity will be over 99% and thus the paternity, be practically proven (Hou *et al.*, 2000). The probability of paternity of all cases was more than 99.95% assuming that the prior probability is 0.5. The

average PI was the same as the typical PI for the simple loci but it was higher in the hypervariable loci.

In motherless cases, the situation is difficult, as the child's paternal allele cannot be determined. The man can be excluded when he has neither of the child's alleles. The maximum mean exclusion chance is usually higher in trio case than that of the motherless case. This value is dependent on the number of alleles of the locus (Lee *et al.*, 2000). In the use of hypervariable loci, where the number of alleles is large and the allele frequency is low, the mean exclusion chance is higher than other loci with a small number of alleles. Therefore, in motherless cases, the use of the hypervariable loci will produce a higher chance of exclusion than with the simple loci. However, no sufficient exclusion can be expected in the motherless cases and a false inclusion can occur in the motherless case (Lee *et al.*, 2000).

When the 30 cases were recalculated ignoring the mother's allele, the average PI value was 611 and the probability of paternity was 99.837%. In three studied cases, the value of the combined PI was higher than the combined PI for 14 loci examined in one motherless case (Lee *et al.*, 2000). The average PI was lower than the typical PI for each locus. However, the drop of the value was more in the simple loci than that of the hypervariable loci. The PI in the motherless case depends only on the frequency of the shared allele between the child and the alleged father (Brenner, 1993). There are a larger number of alleles with a low frequency in the hypervariable loci than in the simple loci. Therefore, the use of the hypervariable loci in motherless cases proved to be advantageous.

The difference between the probability of paternity for trios and motherless cases was small when method (1) was used. The same finding was reported in another study (Lee *et al.*, 2000).

When the method (3) was used, the average PI was low but it was higher in trio cases than the motherless cases. The difference in the probability of paternity between the trio and motherless cases was higher than that of the standard method (1) as the introduction of θ_{AT} causes decrease in the value of the paternity index to be more conservative. The same result was observed in another study (Lee *et al.*, 2000).

In the investigation of the power of exclusion of paternity for the six loci, permutation of cases resulted in 870 families. Among them a single exclusion only was observed in 0.57% of cases and 99.43% of cases showed more than two exclusions. In another study a single exclusion was observed in 0.04% of cases of testing a non-father by using 12 loci of the combined SGM/PowerPlex™1 kits (Thomson *et al.*, 1999). In the simple STR loci the observed exclusion power was less than the expected exclusion power in the three loci. Conversely in the use of the hypervariable loci the observed exclusion power was higher than the expected exclusion power for all loci. The combined exclusion power of the six loci was > 99.998%. The same value was observed at 12 loci of the combination of the SGM plus PowerPlex™1 kits (Thomson *et al.*, 1999).

No detectable mutation was observed in the 30 cases, therefore more family cases are needed to study the mutation rate for each locus. However, in the case of a single

mismatch, the combined paternity index for the loci which are matched has to be calculated. This combined paternity index could be used to evaluate the probability of paternity in such cases. The favoured theory proposes that the mechanism of mutation of STRs is the two phase model (TPM) which allows mutation by one repeat and by more than one repeat (Di Rienzo *et al.*, 1994). Still most of the empirical data has indicated that single repeat mutations are commoner (Brinkmann *et al.*, 1998). Therefore, if the allele length changes by one repeat only this might be strong evidence of mutation. In the case of two mismatches, a large number of loci are required to investigate if the case is a real exclusion or due to mutation.

CHAPTER 8: CONCLUSION AND FUTURE WORK

8.1. WHAT HAS BEEN ACHIEVED?

This project has been on the subject of paternity testing using STR system in Egypt. One area of the research was the development of a method of storage of biological samples. Storage of blood and buccal samples in cell lysis solution of Puregene™ kit proved to be more efficient than the traditional freezing of the samples. Storage in cell lysis solution of Puregene™ provided high quantities and qualities of the extracted DNA. In addition it provided a simple and easy method of transport of samples through long distance. Long-term studies are underway to examine the sensitivity of long-term storage.

The second stage of the experimental part of the thesis dealt with the use of three simple STR loci to examine the structure nature of the Egyptian population using newly designed primers. The loci were the same included in the AmpF ℓ STR Blue kit. The new primers proved to amplify the loci effectively. The analysis of the data revealed no significant evidence of substructure in the Egyptian population. The paternity statistical parameters for these three loci were not sufficient for paternity testing. Therefore, the development of a new multiplex for three of the hypervariable loci was performed. The optimisation of the new multiplex was achieved and yielded reproducible results. Because of the presence of large number of alleles for each locus

of the hypervariable loci, which differ as little as 1 bp, construction of a specific ladder for each locus was essential to achieve accurate sizing of the alleles.

In order to have a complete system for paternity testing, which covers as many case scenarios as possible, the Egyptian database for seven Y-chromosome STR systems was established. There was a new allele for DYS392 locus. The Y-chromosome STR loci provided a high exclusion power in the paternity testing with male child.

All the examined autosomal STR loci were in HWE and the loci were in linkage equilibrium. Therefore, the product rule for combining genotype frequencies across unlinked loci is valid for use in these loci.

All the examined loci were compared with different populations including some Arabic populations. There were significant differences between the Egyptian population and some of the compared populations including some of the Arabic populations. These results confirmed the presence of variations between different populations and the necessity of establishing a specific database for each population or the use of F_{ST} to be more conservative.

The statistical parameters for all examined loci were calculated and the practical use was demonstrated by analysis of 30 different paternity cases. Their use in paternity testing yielded sufficient strong evidence of paternity. The use of these loci was tested in model motherless cases. The use of the hypervariable loci in motherless cases was

more effective than the use of the simple loci because the presence of a large number of alleles with the maximum frequency less than 15%.

8.2. FUTURE WORK

The established STR loci approved to be sufficient in applying them to the paternity testing in Egypt. However, the use of this combination of hypervariable loci may be not suitable for forensic criminal analysis. The hypervariable loci may not be amplified properly in severely degraded samples, which are common in forensic cases. Therefore, there is a possibility of allelic drop out as there are some alleles greater than 400 bp in length. These loci are also difficult to standardise to allow the comparison of typing results between different laboratories (Gill *et al.*, 1994). The FBI has established the Combined DNA Index System (CODIS), which includes 13 core STR loci that should be typed to give information necessary for making matches to locate suspects in criminal cases (Budowle *et al.*, 1999). Several countries have started to collect and type the CODIS loci for millions of individuals and these are stored in centralized searchable databases to enable different laboratories to exchange and compare DNA profiles electronically to identify associations between unresolved cases. Therefore, future work can be done by establishing the database of the CODIS 13 loci for the Egyptian population to store its database in the CODIS banking.

Moreover, there are two working groups of the International Society for Forensic Genetics (the English speaking and the Portuguese and Spanish speaking groups). The aim of these groups is to report the progress of the standardisation between different

laboratories and discussing technical and statistical problems in DNA analysis in paternity testing. It may be desirable to look into setting up an Arabic speaking group, as there are some Arabic countries that have established their DNA databases, which can be used in paternity testing and forensic analysis cases. This would be valuable in the dissemination of knowledge and the standards of DNA analysis particularly as the English and Spanish is not commonly spoken in the Arabic world.

Other future work includes the establishment of the laboratory to provide paternity testing in Egypt. It is preferable that the laboratory be under the qualification of accreditation. This means that the laboratory has to meet the minimum professional standards for insuring a high level of accurate and reliable results in case of disputed paternity. The test results from accredited laboratories have obtained a general acceptance by the courts.

APPENDIX 1

1.1. Sequence of the three simple loci

Sequence of Locus D3S1358 (STRbase-STR DNA Internet Database)

```

5'.....→3'
actgcagtccaatctgggtgacagagcaagaccctgtctcatagatagatagat
agatagatagatagatagatagatagatagatagacagacagatagatacatg
3'←.....5'
caagcctctgttgatttcat
    
```

Sequence of Locus vWA (Genome Bank Database) ACCESSION M25858

```

                    5'.....→3'
1614-aagttgacttggctgagatgtgaaagccctagtggatgataagaataatcag
tatgtgacttggattgatctatctgtctgtctgtctgtctatctatctatctatcta
tctatctatctatctatctatctatctatctatccatctatccatcc
3'←.....5'
atcctatgtatttatcatctgtcctatctct
    
```

Sequence of Locus FGA (Genome Bank Database) ACCESSION M64982

```

                    5'.....→3'
2818-gtcaaatgcccataggttttgaactcacagattaaactgtaaccaaataaa
attagcatatttacaagctagtttctttctttctttttctctttctttctttcttt
ctttctttctttctttctttctttctttctttctctcttcttcttctttcttctt
tcttttttgctggcaattacagacaaatcactcagcagctacttcaataaccatattt
                    3'←.....5'
tcgatttcagaccgtgataatacctacaaccgagtgtcagaggatctgagaagcag
    
```

1.2 The standard deviation of the expected genotypes for D3S1358, vWA and FGA loci.

GENOTYPE D3S1358	OBSERVED FREQUENCY	EXPECTED FREQUENCY	SD
13,14	0.0033	0.0014	0.002
13,15	0.0033	0.0043	0.004
13,16	0.0033	0.0046	0.004
13,17	0.0066	0.0039	0.004
14,14	0.0066	0.0072	0.005
14,15	0.0566	0.0439	0.012
14,16	0.0500	0.0470	0.012
14,17	0.0300	0.0394	0.011
14,18	0.0160	0.0213	0.008
15,15	0.0500	0.0667	0.014
15,16	0.1630	0.1429	0.020
15,17	0.1230	0.1197	0.019
15,18	0.0630	0.0646	0.014
15,19	0.0066	0.0078	0.005
16,16	0.0600	0.0765	0.015
16,17	0.1300	0.1282	0.019
16,18	0.0768	0.0692	0.015
16,19	0.0100	0.0083	0.005
17,17	0.0500	0.0537	0.013
17,18	0.0660	0.0579	0.014
17,19	0.0066	0.0070	0.005
18,18	0.0100	0.0156	0.007
18,19	0.0066	0.0039	0.004

GENOTYPE vWA	OBSERVED FREQUENCY	EXPECTED FREQUENCY	SD
13,14	0.0033	0.0016	0.002
14,14	0.01	0.0087	0.005
14,15	0.043	0.0286	0.010
14,16	0.053	0.0445	0.012
14,17	0.036	0.0457	0.012
14,18	0.02	0.0289	0.010
14,19	0.013	0.0155	0.007
15,15	0.02	0.0235	0.009
15,16	0.063	0.0731	0.015
15,17	0.083	0.0751	0.015
15,18	0.053	0.0475	0.012
15,19	0.02	0.0256	0.009
15,20	0.0033	0.0072	0.005
16,16	0.04	0.0568	0.013
16,17	0.136	0.1168	0.019
16,18	0.067	0.0739	0.015
16,19	0.06	0.0397	0.011
16,20	0.013	0.0111	0.006
17,17	0.0433	0.0600	0.014
17,18	0.09	0.0760	0.015
17,19	0.03	0.0408	0.011
17,20	0.02	0.0114	0.006
17,21	0.0033	0.0002	0.001
17,22	0.0033	0.0008	0.002
18,18	0.0233	0.0240	0.009
18,19	0.0233	0.0258	0.009
18,20	0.0066	0.0072	0.005
18,21	0.0033	0.0016	0.002
19,19	0.0066	0.0069	0.005
19,20	0.0033	0.0039	0.004
19,21	0.0033	0.0008	0.002

GENOTYPE FGA	OBSERVED FREQUENCY	EXPECTED FREQUENCY	SD
17,20	0.0033	0.0016	0.002
17,21	0.0033	0.0023	0.009
17,22	0.0033	0.0026	0.003
17,23	0.0033	0.0028	0.003
17,24	0.0033	0.0025	0.003
18,18	0.0033	0.0005	0.001
18,20	0.0033	0.0046	0.004
18,21	0.0033	0.0064	0.003
18,22	0.0100	0.0072	0.005
18,23	0.0100	0.0077	0.005
18,24	0.0066	0.0069	0.005
18,25	0.0033	0.0046	0.004
18,26	0.0033	0.0025	0.003
19,19	0.0033	0.0032	0.003
19,20	0.0066	0.0111	0.006
19,21	0.0066	0.0155	0.007
19,22	0.0130	0.0176	0.008
19,23	0.0267	0.0187	0.008
19,24	0.0300	0.0168	0.007
19,25	0.0100	0.0111	0.006
19,26	0.0033	0.0060	0.005
19,27	0.0033	0.0023	0.009
19,28	0.0066	0.0030	0.003
19,2,21	0.0033	0.0005	0.001
20,20	0.0100	0.0097	0.006
20,21	0.0360	0.0269	0.009
20,22	0.0400	0.0305	0.010
20,23	0.0200	0.0325	0.010
20,24	0.0360	0.0292	0.010
20,25	0.0130	0.0193	0.008
20,26	0.0100	0.0105	0.006
20,27	0.0033	0.0039	0.004
20,28	0.0033	0.0052	0.004
21,21	0.013	0.0187	0.008
21,22	0.0600	0.0424	0.012
21,23	0.0233	0.0451	0.012

21,24	0.0500	0.0405	0.011
21,25	0.0233	0.0269	0.009
21,26	0.0160	0.0146	0.007
21,27	0.0100	0.0055	0.004
21,28	0.0066	0.0073	0.005
21,29	0.0033	0.0009	0.002
21,2,23	0.0033	0.0011	0.002
21,2,24	0.0033	0.0009	0.002
22,22	0.0160	0.0240	0.009
22,23	0.0733	0.0512	0.013
22,24	0.0267	0.0460	0.012
22,25	0.0200	0.0305	0.010
22,26	0.0160	0.0165	0.007
22,27	0.0066	0.0062	0.005
22,28	0.0066	0.0083	0.005
23,23	0.0267	0.0272	0.009
23,24	0.0267	0.0490	0.013
23,25	0.0500	0.0325	0.010
23,26	0.0233	0.0176	0.008
23,27	0.0066	0.0066	0.005
23,28	0.0066	0.0088	0.005
23,30	0.0033	0.0006	0.001
24,24	0.0233	0.0220	0.009
24,25	0.0300	0.0292	0.010
24,26	0.0100	0.0158	0.007
24,27	0.0066	0.0079	0.005
24,28	0.0160	0.0079	0.005
24,29	0.0033	0.0010	0.002
25,25	0.0100	0.0097	0.006
25,26	0.0160	0.0105	0.006
25,27	0.0033	0.0039	0.004
25,28	0.0066	0.0052	0.004
26,26	0.0033	0.0028	0.003

APPENDIX 2

2.1 Sequence of the hypervariable STR loci with the position of the primers

Sequence of Locus D11S554 (Genome Bank Database) ACCESSION M87277

```
5'.....→3'  
1-gtgatcccgccactgcactccagcctgggtagcagagcaagactgtctcaaaaa  
gaaagaaagagagagagaaaggagaaaggaaagaaagaaagaaaaagaaagaa  
agaaagaaagaaagaaggaaggaaggaaggaaggaaggaaggaaggaaggaag  
gaaggaaggaaggaagaaagagagaaaaatggagcagatatttattgaaatctgcta  
3'←.....5'  
aacnccttctatcctttcttccttagggctgccttaggatgaagggtttgagtc
```

Sequence of Locus SE33 (Genome Bank Database) ACCESSION V00481

```
5'.....→3  
1-ctacagtgagccgaggtcatgccattgcactccaatctgggcgacaagagtgaaa  
ctccgtcaaaagaaagaaagaaagagacaaagagagttagaaagaaagaaagagagag  
agagagaaaggaaggaaggaagaaaaagaaaaagaaagaaagagaaagaaagaa  
agagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaa  
aaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaaggaagga  
3'←.....5'  
aagaaagagcaagttactatagcggtaggggagatggttagaaatatatata
```

Sequence of Locus APOAI1 (Genome Bank Database) ACCESSION J00098

5'.....→3'

6601-gctgagggcaggagaatgggttgaacccgg**gagacggaggttgcagt**gagcc
gagatggcaccactgcactccagcctgggtgacagagggagactccattaaaaaaaa
aagaagaagaaagaaagagagaaaggaagagagggaaagaaagaaaaaggaaagaaa
gaaagaaagagaaagaaagaaaagaaagaaagaaaggaagaaagaaagaaagaaa
gaaagaaagaaagaaagaaagaaagaaagaaagaaagaaaaatagaaagagaa
agaaagggaaaggaaggaatggaaatgagggctaaaacggcgcgccctaggac

3'←.....5'

tgctccgggagaaaggaggacct**gggaggaatgcagtctcaact**ctgtcatctctc

APPENDIX 3

3.1 Equations used in the calculation of paternity and forensic parameters.

$$1- \text{Heterozygosity (h)} = \frac{n_h}{n}$$

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

$$2- \text{Typical Paternity Index (TPI)} = (H+h)/2H$$

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

$$3- \text{Power of Exclusion (PE)} = h^2(1-2*H^2)$$

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

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

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

$$4- \text{ Matching Probability (PM)} = \sum_{i=a}^n \sum_{j \geq 1}^n P_{ij}^2$$

Where i and j represent the frequencies of all possible alleles, n is the number of distinctive genotypes and P_{ij} is the summation of all the genotype frequencies observed in the sample set. The combined probability of match over several loci, is the product of the value for all the loci.

$$5- \text{ Power of Discrimination (PD)} = 1 - \text{PM}$$

Where PM = matching probability

$$\text{For several loci } PD_{\text{comb}} = 1 - \prod_{i=1}^n (1 - P_{di})$$

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

Paternity index calculation in trio cases using the three assumptions (Lee *et al.*, 2000).

Child	Mother	Alleged father	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_j ($i \neq j$)	$\frac{1}{2p_i}$	$\frac{1+3\theta}{2(3\theta+(1-\theta)p_i)}$	$\frac{1}{2p_i(1-2\theta_{AT})+2\theta_{AT}}$
	A_iA_j ($i \neq j$)	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_j ($i \neq j$)	$\frac{1}{2p_i}$	$\frac{1+3\theta}{2(2\theta+(1-\theta)p_i)}$	$\frac{1}{2p_i(1-2\theta_{AT})+2\theta_{AT}}$
A_iA_j ($i \neq j$)	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	$\frac{1}{2p_j}$	$\frac{1+3\theta}{2(2\theta+(1-\theta)p_j)}$	$\frac{1}{2p_j(1-2\theta_{AT})+2\theta_{AT}}$
		A_jA_k ($k \neq i, j$)	$\frac{1}{2p_j}$	$\frac{1+3\theta}{2(\theta+(1-\theta)p_j)}$	$\frac{1}{2p_j(1-2\theta_{AT})+2\theta_{AT}}$
	A_iA_j ($i \neq j$)	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	$\frac{1}{p_i+p_j}$	$\frac{1+3\theta}{4\theta+(1-\theta)p_i+p_j}$	$\frac{1}{(p_i+p_j)(1-2\theta_{AT})+2\theta_{AT}}$
		A_jA_k ($k \neq i, j$)	$\frac{1}{2(p_i+p_j)}$	$\frac{1+3\theta}{2(3\theta+(1-\theta)(p_i+p_j))}$	$\frac{1}{2(p_i+p_j)(1-2\theta_{AT})+2\theta_{AT}}$
	A_jA_k ($k \neq i, j$)	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_jA_i ($i \neq j$)	$\frac{1}{2p_j}$	$\frac{1+3\theta}{2(\theta+(1-\theta)p_j)}$	$\frac{1}{2p_j(1-2\theta_{AT})+2\theta_{AT}}$

Paternity index calculation using the assumption that the alleged father is a relative to the true father in the motherless case Method (3).

Child	Alleged father	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_j ($i \neq j$)	$\frac{1}{2p_i}$	$\frac{1+2\theta}{2((1-\theta)p_i+\theta)}$	$\frac{1}{2(p_i(1-2\theta_{AT})+2\theta_{AT})}$
A_iA_j ($i \neq j$)	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 ($i \neq j$)	$\frac{1}{4p_i} + \frac{1}{4p_j}$	$\frac{(p_i+p_j)(1+2\theta)}{2(2(1-\theta)p_i p_j + (p_i+p_j)\theta)}$	$\frac{(p_i p_j)}{2(2p_i p_j(1-2\theta_{AT})+(p_i p_j)\theta_{AT})}$
	A_iA_k ($k \neq i, j$)	$\frac{1}{4p_i}$	$\frac{1+2\theta}{2(2(1-\theta)p_i+\theta)}$	$\frac{1}{2(2p_i(1-2\theta_{AT})+\theta_{AT})}$

An example for the use of the equations of the paternity index calculation is shown below. These are the genotypes of the mother, child and the alleged father.

No 1	D3	vWA	FGA	D11	SE33	APOA
Mother	15,16	16,17	20,22	6,12	15,30	6,22
Child	15,17	15,17	20,21	11,12	15,29	7,22
Father	16,17	15,15	21,24	4,11	23,29	7,21

The calculation of the PI was calculated for the D3S1358 locus.

First in the trio case

$$\text{Method (1) PI} = \frac{1}{2p_i} = \frac{1}{2 \times 0.2317} = 2.16$$

$$\text{Method (2) PI} = \frac{1}{2(\theta + (1-\theta)p_i)} = \frac{1}{2(0.0011 + (1-0.0011) \times 0.2317)} = 2.15$$

$$\text{Method (3) PI} = \frac{1}{2p_i(1-2\theta_{AT}) + 2\theta_{AT}} = \frac{1}{2 \times 0.2317(1-2 \times 0.25) + 2 \times 0.25} = 1.37$$

Second in motherless case

$$\text{Method (1) PI} = \frac{1}{4p_i} = \frac{1}{4 \times 0.2317} = 1.08$$

$$\text{Method (2) PI} = \frac{1+2\theta}{2(2(1-\theta)p_i + \theta)} = \frac{1+2 \times 0.0011}{2(2(1-0.0011) \times 0.2317 + 0.0011)} = 1.08$$

$$\text{Method (3) PI} = \frac{1}{2(2p_i(1-2\theta_{AT}) + \theta_{AT})} = \frac{1}{2(2 \times 0.2317(1-2 \times 0.25) + 0.25)} = 1.04$$

REFERENCES

- Adams, M., Urquhart, A., Kimpton, C. and Gill, P. (1993). The human D11S554 locus, four distinct families of repeat pattern alleles at one locus. *Hum. Mol. Genet.* 2(7), 1373-1376.
- Alford, R.L., Hammond, H.A., Coto, I. and Caskey, C.T. (1994). Rapid and efficient resolution of parentage by amplification of short tandem repeats. *Am. J. Hum. Genet.* 55, 190-195.
- Alkhyat, 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.
- AmpF ℓ STR BlueTM PCR Amplification Kit: User's Manual (1996). The Perkin-Elmer Corporation.
- Andres, M. I., Prieto, V., Flores, I. and Sanz, P. (1998). Allele frequency distribution of D3S1358, vWA, and FGA loci in a southern Spain population, in Olaisen B., Brinkmann B., Lincoln P. (Eds), *Progress in Forensic Genetics 7*, Excerpta Medica, Amsterdam pp. 243-245.
- Baird, M., Balazs, I., Guisti, A., Miyazaki, L., Nicholas, N.K. and Wexler, K. (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.

Balding, D.J. and Nichols, R.A. (1994). DNA profile match probability calculation, how to allow for population stratification, relatedness, database selection and single bands. *Forensic Sci. Int.* 64, 124-140.

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.

Barber, M.D., McKeown, B.J. and Parkin, B.H. (1996). Structural variation in the alleles of a short tandem repeat system at the human alpha fibrinogen locus. *Int. J. Leg. Med.* 108, 180-185.

Barber, M.D. and Perkin, B.H. (1996). Sequence analysis and allelic designation of the short tandem repeat loci D18S51 and D8S1179. *Int. J. Leg. Med.* 109, 62-65.

Barber, M.D., Piercy, R.C., Andersen, J.F. and Perkin, B.H. (1995). Structural variation of novel alleles at Hum vWA and Hum FES/FPS short tandem repeat loci. *Int. J. Leg. Med.* 108, 31-35.

Bassam, B.J., Caetano Anolles, G. and Gresshoff, P.M. (1991). Fast and sensitive silver staining of DNA in polyacrylamide gels. *Analytical Biochemistry* 196(1), 80-83.

Beckman, J.S. and Weber, J.L. (1992). Survey of human and rat microsatellites. *Genomics* 12, 627-631.

Benecke, M., Schmitt, C. and Staak, M. (1997). Development of a fast STR triplex system for identification of urine samples. *Proceedings from the First European Symposium on Human Identification (1996)* pp. 148.

Boerwinkle, E., Xiong, W., Fourest, E. and Chan, L. (1989). Rapid typing of tandemly repeated hypervariable loci by the polymerase chain reaction, application to the apolipoprotein B 3' hypervariable region. *Proc. Natl. Acad. Sci. USA* 86, 212-216.

Braaten, D.C., Thomas, J.R., Little, R.D., Dickson, K.R., Goldberg, I., Schlessinger, D., Ciccodicola, A. and D'Urso, M. (1988). Locations and contexts of sequences that hybridize to poly dG-dT, dC-dA in mammalian ribosomal DNAs and two X-linked genes. *Nucleic Acids Res.* 16, 865-881.

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 Olaisen B., Brinkmann B., Lincoln P. (Eds), *Progress in Forensic Genetics* 8, Excerpta Medica, Amsterdam pp. 221-223.

Brenner, C.H. (1993). A note on paternity computation in cases lacking a mother. *Transfusion* 33(1), 51-54.

Brenner, C.H. and Morris, J. (1990). Paternity index calculations in single locus hypervariable DNA probes, validation and other studies. Proceedings of the International Symposium on Human Identification, Promega Corporation pp. 21-53.

Brinkmann, B. (1996). The STR approach. Advances in Forensic haemogenetics Volume 6, pp. 41-51.

Brinkmann, B. (1998). Overview of PCR-Based Systems in Identity Testing, in Lincoln, P.J. and Thomson, J. (Eds), Methods in Molecular Biology: Forensic DNA Profiling Protocols, Humana Press, Totowa.

Brinkmann, B., Klintschar, M., Neuhuber, F., Huhne, J. and Rolf, B. (1998). Mutation rate in human microdatellites: Influence of the structure and length of the tandem repeat. Am. J. Hum. Genet. 62(6), 1408-1415.

Brinkmann, B., Sajantila, A., Goedde, H.W., Matsumoto, H., Nishi, K. and Wiegand, P. (1996). Population genetic comparisons among eight populations using allele frequency and sequence data from three microsatellite loci. Eur. J. Hum. Genet. 4, 175-182.

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.

Budowle, B., Baechtel, F.S., Comey, C.T., Giusti, A.M. and Klevan, L. (1995). Simple protocols for typing forensic biological evidence, Chemiluminescent detection for

human DNA quantitation and restriction fragment length polymorphism (RFLP) analyses and manual typing of polymerase chain reaction (PCR) amplified polymorphisms. *Electrophoresis* 16, 1559-1567.

Budowle, B., Moretti, T.R., Baumstark, A.L., Defenbaugh, D.A. and Keys, K.M. (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(6), 1277-1286.

Budowle, B., Nhari, L.T., Moretti, T.R., Kanoyangwa, S.B., Masuka, E., Defenbaugh, D.A. and Smerick, J.B. (1997). Zimbabwe black population data on the six short tandem repeat loci-CSF1PO, TPOX, THO1, D3S1358, VWA and FGA. *Forensic Sci. Int.* 90(3), 215-221.

Cabrero, C., Diez, A., Valverde, E., Carracedo, A. and Alemany, J. (1995). Allele frequency distribution of four PCR-amplified loci in the Spanish population. *Forensic Sci. Int.* 71, 153-164.

Cao, A., Gossens, M. and Prastu, M. (1989). β -Thalassemia mutations in Mediterranean population. *British Journal of Haematology* 71, 309-312.

Casanova, M., Leroy, P., Boucekine, C., Weissenbach, J., Bishop, C., Fellous, M., Purrello, M., Fiori, G. and Siniscalco, M. (1985). A human Y-linked DNA

polymorphism and its potential for estimating genetic and evolutionary distance. *Science* 230(4732), 1403-1406.

Chakraborty, R. (1985). Paternity testing with genetic markers are Y-linked genes more efficient than autosomal ones. *Am. J. Med. Genet.* 21, 297-305.

Chakraborty, R. and Stivers, D.N. (1996). Paternity Exclusion by DNA Markers, Effects of Paternal Mutations. *J. Forensic Sci.* 41(4), 671-677.

Chimera, J.A., Harris, C.R. and Litt, M. (1989). Population genetics of the highly polymorphic locus D16S7 and its use in paternity evaluation. *Am. J. Hum. Genet.* 45(6), 926-931.

Chou, Q., Russell, M., Birch, D.E., Raymond, J. and Bloch, W. (1992). Prevention of pre-PCR mis-priming and primer dimerization improves low-copy number amplifications. *Nucleic Acids Res.* 20, 1717-1723.

Cifuentes, O.L. and Jorquera, G.H. (1997). Paternity analysis in cases of father-daughter incest using multiallelic. *Human Heredity* 47(5), 288-294.

Clayton, T.M., Whitaker, J.P. and Maguire, C.N. (1995). Identification of bodies from the scene of a mass disaster using DNA amplification of short tandem repeat (STR) loci. *Forensic Sci. Int.* 76, 7-15.

Clayton, T.M., Whitaker, J.P., Sparkes, R. and Gill, P. (1998). Analysis and interpretation of mixed forensic stains using DNA STR profiling. *Forensic Sci. Int.* 91, 55-70.

Cooper, G., Amos, W., Hoffman, D. and Rubinsztein, D. (1996). Network analysis of human Y microsatellite haplotypes. *Hum. Mol. Genet.* 5, 1759-1766.

Cosso, S. and Reynolds, R. (1995). Validation of the AmpliFLP D1S80 PCR amplification kit for forensic casework analysis according to TWGDAM guidelines. *J. Forensic Sci.* 40(3), 424-434.

Cossutta, F., Perossa, R., Frassanito, F., Altamura, B.F. and Fattorini, P. (2000). Population genetics of fourteen STR loci in North-East Italy, in Olaisen B., Brinkmann B., Lincoln P. (Eds), *Progress in Forensic Genetics 8*, Excerpta Medica, Amsterdam pp. 178-180.

Cotton, E.A., Allsop, R.F., Guest, J.L., Frazier, R.R.E., Koumi, P., Callow, I.P., Seager, A. and Sparkes, R.L. (2000). Validation of the AMPFLSTR SGM Pluse system for use in forensic casework. *Forensic Sci. Int.* 112, 151-161.

Craig, J., Fowler, S., Burgoyne, L.A., Scott, A.C. and Hardin, J. (1988). Repetitive deoxyribonucleic acid (DNA) and human genome variation, A concise review relevant to forensic biological. *J. Forensic Sci.* 33, 1111-1126.

Crouse, C. and Schumm, J.W. (1995). Investigation of species specificity using nine PCR-based human STR systems. *J. Forensic Sci.* 40, 952-956.

Deforce, D.L.D., Millecamps, R.E.M., Hoofstat, D.V. and Van den Eeckhout, E.G. (1998). Comparison of slab gel electrophoresis and capillary electrophoresis for the detection of the fluorescently labelled polymerase chain reaction products of short tandem repeat fragments. *J. Chromatography A* 806, 149-155.

Deka, R., Jin, L., Shriver, M.D., Yu, L.M., Saha, N., Barrantes, R., Chakraborty, R. and Ferrell, R.E. (1996). Dispersion of human Y chromosome haplotypes based on five microsatellites in global populations. *Genome Res.* 6, 1177-1184.

de Knijff, P., Kayser, M., Caglia, A., Corach, D., Fretwell, N., Gehrig, C., Graziosi, G., Heidorn, F., Herrmann, S., Herzog, B., Hidding, M., Honda, K., Jobling, M., Krawczack, M., Leim, K., Menser, S., Meyer, E., Oesterreich, W., Pandya, A., Parson, W., Penacino, G., Perez-Lezaun, A., Piccinini, A., Prinz, M., Schmitt, C., Schneider, P.M., Szibor, R., Teifel-Greding, J., Weichold, G. and Roewer, L. (1997). Chromosome Y microsatellites, population genetic and evolutionary aspects. *Int. J. Leg. Med.* 110, 134-140.

Di Rienzo, A., Peterson, A.C., Garza, J.C., Valdes, A.M. and Slatkin, M. (1994). Mutational processes of simple sequence repeat loci in human populations. *Proc. Natl. Acad. Sci. USA* 91, 3166-3170.

DNA Recommendations (1992). Report concerning recommendation of the DNA commission of the International Society for Forensic Haemogenetics relating to the use of PCR-based polymorphisms. *Int. J. Leg. Med.* 105, 63-64.

DNA Recommendations (1994). Report concerning further recommendation of the DNA commission of the International Society for Forensic Haemogenetics relating to the use of PCR-based polymorphisms STR systems. *Int. J. Leg. Med.* 107, 159-160.

Dodd, B.E. (1985). DNA fingerprinting in matters of family and crime. *Nature* 318, 506-507.

Domenici, R., Bibbiani, R., Fornaciari, S., Nardone, M.A., Rocchi, I., Spinetti, M., Venturi, M. and Bargagna, M. (1998). Application of conventional and DNA polymorphism to paternity cases, in Olaisen B., Brinkmann B., Lincoln P. (Eds), *Progress in Forensic Genetics 7*, Excerpta Medica, Amsterdam pp. 89-91.

Droog, S., Lakenberg, N., Meulenbelt, I., de Maat, M.P.M., Huisman, L.G.M., Jie, A.F.H. and Slagboom, P.E. (1996). Isolation and storage of DNA for population studies. *Fibrinolysis* 2, 29-30.

Drozd, M.A., Archard, L., Lincoln, P.J., Morling, N., Nellesmann, L.J., Philips, C., Soteriou, B. and Syndercombe, D. (1994). An investigation of the HUMvWA31A locus in British Caucasians. *Forensic Sci. Int.* 69, 161-170.

Dryja, T.P., Mukai, S., Peterson, R., Rapaport, J.M. and Yandell, D.W. (1989). Parental origin of mutations of the retinoblastoma gene. *Nature* 339, 556-558.

Dupuy, M.B., Gedde-Dahl, T. and Olaisen, B. (1998). Comparison of allele frequencies in three hyperpolymorphic STRs between Norwegian Saamis and the main population, in Olaisen, B., Brinkmann, B., Lincoln, P. (Eds), *Progress in Forensic Genetics* 7, Excerpta Medica, Amsterdam pp. 395-397.

Dupuy, M. B. and Olaisen, B. (1997). A dedicated internal standard in fragment length analysis of hyperpolymorphic short tandem repeats. *Forensic Sci. Int.* 86, 207-227.

Dykes, D. (1990). DNA analysis of paternity testing, in Robertson, J., Ross, A.M. and Burgoyne, L.A. (Eds), *DNA in Forensic Science Theory, techniques and application*, Ellis Horwood, England pp. 141-155.

Dykes, D.D. and Polesky, H.F. (1976). The usefulness of serum proteins and erythrocyte polymorphisms in paternity testing. *Am. J. Clin. Pathol.* 65(6), 982-986.

Econonou, E.P., Bergen, A.W., Warren, A.C. and Antonarakis, S.E. (1990). The polydeoxyadenylate trace of Alu repetitive elements is polymorphic in the human genome. *Proc. Natl. Acad. Sci. USA* 87, 2951-2954.

Edwards, A., Civitello, A., Hammond, H.A. and Caskey, C.T. (1991). DNA Typing and Genetic Mapping with Trimeric and Tetrameric Tandem Repeats. *Am. J. Hum. Genet.* 49, 746-756.

Edwards, A., Hammond, H.A., Jin, L., Caskey, C T. and Chakraborty, R. (1992). Genetics Variation at Five Trimeric and Tetrameric Tandem Repeat Loci in Four Human Population Groups. *Genomics* 12, 241-253.

Elston, R.C. (1986). Probability and paternity testing. *Am. J. Hum. Genet.* 39, 112-122.

Epplen, J.T., McCarrey, J.R., Sutou, S. and Ohno, S. (1982). Base sequence of a cloned snake W-chromosome DNA fragment and identification of a male specific putative mRNA in the mouse. *Pro. Natl. Acad. Sci. USA* 79, 3798-3802.

Evvett, I.W. and Weir, B.S. (1998). *Interpreting DNA evidence: Statistical genetics for forensic scientists.* Sinauer Associates, INC. Publisher, Sunderland, Massachusetts.

Fernandez-Rodriguez, A., Alonso, A., Albarran, C., Martin, P., Iturralde, M.J., Montesino, M. and Sancho, M. (1996). Microbial DNA challenge studies of PCR-based systems used in forensic genetics. *Advances in Forensic Haematogenetics* Volume 6, pp. 177-179.

Frank, R. and Koster, H. (1979). DNA chain length markers and the influence of base composition on electrophoretic mobility of oligodeoxyribonucleotides in polyacrylamide gels. *Nucleic Acids Res.* 6, 2069-2087.

Fregeau, C.J. and Fourney, R.M. (1993). DNA typing with fluorescently tagged short tandem repeats: a sensitive and accurate approach to human identification. *BioTechniques* 15, 100-119.

Fretwell, N., Armour, J.A.L. Jeffreys, A.J., Dover, G.A. and Jobling, M.A. (1995). A search for Y-chromosome-specific minisatellites and microsatellites. *Am. J. Hum. Genet.* 57 [suppl], 163.

Gill, P., Brinkmann, B., d'Aloja, E., Andersen, J., Bar, W., Carracedo, A., Dupuy, B., Eriksen, B., Jangblad, M., Johnsson, V., Kloosterman, A.D., Lincoln, P., Morling, N., Rand, S., Sabatier, M., Scheithauer, R., Schneider, P. and Vide, M.C. (1997). Considerations from the European DNA profiling group (EDNAP) concerning STR nomenclature. *Forensic Sci. Int.* 87, 185-192.

Gill, P., Ivanov, P.L., Kimpton, C., Piercy, R., Benson, N., Tully, G., Evett, I., Hagelberg, E. and Sullivan, K. (1994). Identification of the remains of the Romanov family by DNA analysis. *Nat. Genet.* 6, 130-135.

Gill, P., Kimpton, C., D'Aloja, E., Nderssen, J.F., Bar, W., Brinkmann, B., Holgersson, S., Johnsson, V., Kloosterman, A.D., Lareu, M.V., Nellesmann, L., Pfitzinger, H.,

Phillips, C.P., Schmitter, H., Schneider, P.M. and Stenersen, M. (1994). 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.P., Urquhart, A., Oldroyd, N., Millican, E.M., Watson, S.K. and Downes, T.J. (1995). Automated short tandem repeat (STR) analysis in forensic casework-a strategy for the future. *Electrophoresis* 16, 1543-1552.

Gill, P., Sparkes, R. and Kimpton, C. (1997). Development of guidelines to designate alleles using a STR multiplex system. *Forensic Sci. Int.* 89(3), 185-197.

Gill, P., Sullivan, K.M. and Werrett, D.J. (1990). The analysis of hyper-variable DNA profiles, problems 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.P. (1996). A new method of STR interpretation using inferential logic-development of a criminal intelligence database. *Int. J. Leg. Med.* 109, 14-22.

Giusti, A., Baird, M., Pasquale, S., Balazs, I. and Glassberg, J. (1986). Application of deoxyribonucleic acid (DNA) polymorphisms to the analysis of DNA recovered from sperm. *J. Forensic Sci.* 31, 409-417.

Gjertson, D.W., Mickey, M.R., Hopfield, J., Takenouchi, T. and Terasaki, P.I. (1988). Calculation of probability of paternity, using DNA sequences. *Am. J. Hum. Genet.* 43, 860-869.

Guo, S.W. and Thompson, E.A. (1992). Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48, 361-372.

Gusmao, L., Amorim, A., Martins, N., Prata, M.J., Gonzalez-Neira, A., Lareu, M.V. and Carracedo, A. (2000). Y-chromosome STR haplotypes in Macau (SE China), in Olaisen B., Brinkmann B., Lincoln P. (Eds), *Progress in Forensic Genetics* 8, Excerpta Medica, Amsterdam pp. 324-326.

Gusmao, L., Amorim, A., Prata, M.J., Pereira, L., Lareu, M. V. and Carracedo, A. (1996). Failed PCR Amplifications of MBP-STR alleles due to polymorphism in the primer annealing region. *Int. J. Leg. Med.* 108, 313-315.

Gusmao, L., Gonzalez-Neira, A., Sanchez-Diz, P., Lareu, M.V. and Carracedo, A. (2000). Alternative primers for DYS391 typing: advantages of their application to forensic genetics. *Forensic Sci. Int.* 106(3), 163-172.

Hamada, H., Petrino, M.G. and Kakunaga, T. (1982). A novel repeated element with Z-DNA –forming potential is widely found in evolutionarily diverse eukaryotic genomes. *Proc. Natl. Acad. Sci. USA* 79, 6465-6469.

Hammer, M.F. (1994). A recent insertion of an Alu element on the Y chromosome is a useful marker for human population studies. *Mol. Biol. Evol.* 11, 749-761.

Hammer, M.F. and Horai, S. (1995). Y Chromosomal DNA variation and peopling of Japan. *Am. J. Hum. Genet.* 56, 951-962.

Hammond, H.A., Jin, L., Zhong, Y., Caskey, C.T. and Chakraborty, R. (1994). Evaluation of 13 short tandem repeat loci for use in personal identification applications. *Am. J. Hum. Genet.* 55, 175-189.

Hardman, N. (1986). "Structure and Function of Repetitive DNA in Eukaryotes". *Biochemical Journal* 234(1), 1-11.

Hauge, X.Y. and Litt, M. (1993). A study of the region of 'shadow bands' seen when typing dinucleotide repeat polymorphisms by PCR. *Hum. Mol. Genet.* 2, 411-415.

Hayes, J.M., Budowle, B. and Freund, M. (1995). Arab population data on the PCR-based loci, HLA-DQA1, LDLR, GYPA, HBG, D7S8, Gc and D1S80. *J. Forensic Sci.* 40, 888-892.

Heller, C. (1999). Separation of double-stranded and single-stranded DNA in polymer solutions: I. Mobility and separation mechanism. *Electrophoresis* 20(10), 1962-1977.

Heyer, E., Puymirat, J., Dieltjes, P., Bakker, E. and de Knijff, P. (1997). Estimating Y-chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. *Hum. Mol. Genet.* 6, 799-803.

Hochmeister, M.N., Budowle, B., Borer, U.V., Rudin, O., Bohnert, M. and Dirnhofer, R. (1995). Confirmation of the identity of human skeletal remains using multiplex PCR amplification and typing kits. *J. Forensic Sci.* 40, 701-705.

Holgerson, S., Karlsson, J.A., Kihlgren, A., Rosen, B., Savolainen, P. and Gyllensten, U. (1994). Fluorescent-based typing of the two short tandem repeat loci HUMTH01 and HUMACTBP2, reproducibility of size measurements and genetic variation in the Swedish population. *Electrophoresis* 5, 890-895.

Horst, B., Eigel, A., Sanguansermisri, T. and Brinkmann, B. (1999). Human Y-chromosomal STR types in north Thailand. *Int. J. Leg. Med.* 112, 211-212.

Hou, Y., Song, Z., Li, Y., Wu, J. and Zhang, J. (2000). How many STR markers are enough for paternity identification, in Olaisen B., Brinkmann B., Lincoln P. (Eds), *Progress in Forensic Genetics* 8, Excerpta Medica Amsterdam, pp. 384-386

Hou, Y., Wu, J., Li Y., Zhang, J. and Prinz, M (2000). Polymorphisms of chromosome Y specific short tandem repeats in a Chinese population, in Olaisen B., Brinkmann B., Lincoln P. (Eds), *Progress in Forensic Genetics* 8, Excerpta Medica, Amsterdam, pp. 315-317.

Houtz, T.D., Brooks, M.A., Wenk, R.E. and Dawson, R.B. (1981). The utility of HLA and six erythrocyte antigen systems in excluding paternity among 500 disputed cases. *Forensic Sci. Int.* 17, 211-218.

Jarman, A.P., Nicholls, R.D., Weatherall, D.J., Clegg, J.B. and Higgs, D.R. (1986). Molecular characterization of a hypervariable region downstream of the human Alpha-Globin gene complex. *EMBO Journal* 5, 1857-1863.

Jarreta, B. M., Roche, P.D. and Albecia, E. (1999). Genetic variation at six STR loci (HUMTH01, HUMTPOX, HUMCSF1PO, HUMF13A01, HUMFES/FPS, HUMvWAF31) in Aragon (North Spain). *Forensic Sci. Int.* 100, 87-92.

Jeffreys, A.J. (1993). DNA typing, approaches and applications. *J. Forensic Sci. Soc.* 33(4), 204-211.

Jeffreys, A.J., Brookfield, J.F.Y. and Semeonoff, R. (1985). Positive identification of an immigration test case using DNA fingerprint. *Nature* 317, 818-819.

Jeffreys, A.J., Wilson, V. and Thein, S.L. (1985b). Individual-specific 'fingerprints' of human DNA. *Nature* 316, 75-79.

Jerome, L. and Gottschall, M.D. (1991). An Overview of DNA Analysis in Paternity Testing, Current Practices. *The Second International Symposium on Human Identification* pp. 91-102.

Jobling, M.A., Bouzekri, N. and Taylor, P.O. (1998). Hypervariable digital DNA codes for human paternal lineages MVR-PCR at the Y-specific minisatellite, MSY1 (DYF155S1). *Hum. Mol. Genet.* 7(4), 643-653.

Jobling, M.A., Pandya A. and Tyler-Smith C. (1997). The Y chromosome in forensic analysis and paternity testing. *Int. J. Leg. Med.* 110, 118-124.

Jobling, M.A. and Tyler-Smith, C. (1995). Fathers and sons, the Y chromosome and human evolution. *Trends Genetics* 11(11), 449-456.

Jones, D.A. (1972). Blood samples, Probabilities and discriminations. *J. Forensic Sci. Soc.* 12, 355-359.

Kanter, E., Baird, M., Shaler, R. and Balazs, I. (1986). Analysis of restriction fragment length polymorphism in deoxyribonucleic acid (DNA) recovered from dried bloodstains. *J. Forensic Sci.* 31, 403-408.

Karafet, T.M., Zegura, S.L., Posukh, O., Osipova, L., Bergen, A., Long, J., Goldman, D., Klitz, W., Harihara, S., de Knijff, P., Wiebe, V., Griffiths, R.C., Templeton, A.R. and Hammer, M.F. (1999). Ancestral Asian source (s) of new world Y-chromosome founder haplotypes. *Am. J. Hum. Genet.* 64(3), 817-831.

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.H., Vyvial, T.M. and Young, D. (1991). Validation the probability of paternity. *Transfusion* 31, 823-828.

Kayser, M., Caglia, A., Corach, D., Fretwell, N., Gehrig, C., Graziosi, G., 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., Penacino, G., Perez-Lezaun, A., Piccinini, A., Prinz, M., Schmitt, C., Schneider, P.M., Szibor, R., teifel-Greding, J., Weichhold, G., de Knijff, P. and Roewer, L. (1997). Evaluation of Y-chromosomal STRs, a multicenter study. *Int. J. Leg. Med.* 110, 125-133.

Kayser, M., de Knijff, P., Dieltjes, P., Krawczak, M., Nagy, M., Zerjal, T., Pandya, A., Tyler Smith, C. and Roewer L. (1997). The applications of microsatellite-based Y chromosome haplotyping. *Electrophoresis* 18(9), 1602-1607.

Kayser, M., Kruger, C., Nagy, M., Geserick, G., de Knijff, P. and Roewer, L. (1998). Y-chromosomal DNA-analysis in paternity testing: experiences and recommendations, in Olaisen B., Brinkmann B., Lincoln P. (Eds), *Progress in Forensic Genetics* 7, Excerpta Medica, Amsterdam pp. 494-496.

Kimpton, C.P., Gill, P., d'Aloja, E., Anderson, J.F., Bar W., Holgersson, S., Jacobsen, S., Johnsson, V., Kloosterman, A.D., Lareu, M.V., Nellesmann, L., Pfitzinger, H., Phillips, C.P., Rand, S., Schmitter, H., Schneider, P.M., Sternersen, M. and Vide, M.C. (1995). Report on the second EDNAP collaborative STR exercise. *Forensic Sci. Int.* 71, 137-152.

Kimpton, C.P., Gill, P., Walton, A., Urquhart, A., Millican, E.s. and Adams, M. (1993). Automated DNA profiling employing multiplex amplification of short tandem loci. *PCR Methods and Appl.* 3, 13-21.

Kimpton, C.P., Walton, A. and Gill, P. (1992). A further tetranuclotide repeat polymorphisms in the vWF gene. *Hum. Mol. Genet.* 1, 287.

Klitschar, M. and Al Hammadi, N. (1998). A study of five short tandem repeat systems in a Yemenian population sample, in Olaisen B., Brinkmann B., Lincoln P. (Eds), *Progress in Forensic Genetics 7*, Excerpta Medica, Amsterdam pp. 300-302.

Klitschar, M., Al-Hammadi, N. and Reichenpfader, B. (1999). Population genetic studies on the tetrameric short tandem repeat loci D3S1358, VWA, FGA, D8S1179, D21S11, D18S51, D5S818, D13S317 and D7S820 in Egypt. *Forensic Sci. Int.* 104, 23-31.

Klitschar, M., Kozma, Z., Al hammadi, N., abdull Fatah, M. and Nohammer, C. (1998). A study on the short tandem repeat systems HumCD4, HUMTH01, and HUMFIBRA in a Yemenian and an Egyptian population sample. *Int. J. Leg. Med.* 107-109.

Kobilinsky, L. (1992). Recovery and stability of DNA in samples of forensic science significance. *Forensic Sci. Rev.* 4(1), 68-87.

Lee, H.C., Ladd, C., Bourke, M.T., Pagliaro, E. and Tirnady, F. (1994). DNA Typing in Forensic Science 1-Theory and Background. *Am. J. Forensic Med. Pathol.* 15(4), 269-282.

Lee, H.S., Lee, J.W., Han, G.R. and Hwang, J.J. (2000). Motherless case in paternity testing. *Forensic Sci. Int.* 114, 57-65.

Lee, J.W., Lee, H.S., Park, M. and Hwang, J.J. (1999). Paternity probability when a relative of father is an alleged father. *Science Justice* 39(4), 223-230.

Lewis, P.O. and Zaykin, D. (1999). Genetic Data Analysis, computers programme for the analysis of allelic data Version 1.0 (d12). Free programme distributed by the authors over the internet at <http://lewis.eeb.uconn.edu/lewishome/gda.html>

Lewontin, R.C. and Hatl, D.L. (1991). Population genetics in forensic DNA typing. *Science* 254L, 1745-1750.

Lezaum, A.P., Calafell, F., Mateu, E., Comas, D., Bosch, E. and Bertranpetit, J. (1997). Allele Frequencies for 20 Microsatellites in a Worldwide Population Survey. *Human Heredity* 47, 189-196.

Li, H., Schmidt, L., Wei, M., Hustad, T., Lerman, M.I., Zbar, B. and Tory, K. (1993). Three tetranucleotide polymorphisms for loci, D3S11352; D3S1358; D3S1359. *Hum. Mol. Genet.* 2(8), 1327.

Lins, A.M., Micka, K.A., Sprecher, C.J., Taylor, J.A., Bacher, J.W., Rabbach, D.R., Bever, R.A., Creacy, S.D. and Schumm, J.W. (1998). Development and population study of an eight locus short tandem repeat (STR) multiplex system. *J. Forensic Sci.* 43(6), 1168-1180.

Litt, M. and Luty, J.A. (1989). A hypervariable microsatellite revealed by in vitro amplification of a cardiac muscle actin gene. *Am. J. Hum. Genet.* 44, 397-401.

Liu, C., Harashima, N., Katsuyama, Y., Ota, M., Arakura, A. and Fukushima, H. (1997). ACTBP2 gene frequency distribution and sequencing of the allelic ladder and variants in the Japanese and Chinese populations. *Int. J. Leg. Med.* 110, 208-212.

Lorente, J.A., Lorente, M., Budowle, B., Wilson, M.R. and Villanueva, E. (1994). Analysis of short tandem repeat (STR) HUMVWA in the Spanish population. *Forensic Sci. Int.* 65, 169-175.

Lu, H.L., Wang, C.X., Wu, F.Q. and Li, J.J. (1994). Paternity identification in twins with different fathers. *J. Forensic Sci.* 39, 1100-1102.

Lygo, J.E., Johnson, P.E., Holdaway, D.J., Woodroffe, S., Whitaker, J.P., Clayton, T.M., Kimpton, C.P. and Gill, P. (1994). The validation of short tandem repeat (STR) loci for use in forensic cases. *Int. J. Leg. Med.* 107, 77-89.

Macintyre, S. and Sooman, A. (1991). Non paternity and prenatal genetic screening. *Lancet* 338(8771), 869-871.

Madisen, L., Hoar, D.I., Holroyd, C.D., Crispe, M. and Hodes, M.E. (1987). DNA banking, the effects of storage of blood and isolated DNA on the integrity of DNA. *Am. J. Med. Genet.* 27, 379-390.

Maha, G.C., Mason, J.M., Stuhlmiller, G.M. and Heine, U. (1995). Reply to Pena. *Am. J. Hum. Genet.* 56, 1505-1506.

Malaspina, P., Persichetti, F., Novelletto, A., Iodice, C., Terrenato, L., Ferraro, M. and Prantera, G. (1990). The human Y chromosome shows a low level of DNA polymorphism. *Annual Human Genetics* 54, 297-300.

Malmgren, H., Lango, A., Holmlund, G., Engdahl, L., Nordlof, S. and Lindblom, B. (1998). DNA-sequencing of an allelic ladder comprising AAAG repeats and inter-gel variation of allele size, in Olaisen B., Brinkmann B., Lincoln P. (Eds), *Progress in Forensic Genetics* 7, Excerpta Medica, Amsterdam pp. 481-483.

Mancuso, D. J., Tuley, E.A., Westfield, L.A., Worrall, N.K., Shelton-Inloes, B.B., Sorace, J.M., Alevy, Y.G. and Sadler, J.E. (1989). Structure of the gene for human von Willebrand factor. *J. Biological Chemistry* 264, 19512-19527.

Mannucci, A., Casarino, L., Bruni, G., Lomi, A. and De Stefano, F. (1995). Individual identification of flood victims by DNA polymorphisms and autopsy findings. *Int. J. Leg. Med.* 107, 213-215.

Markowicz, K. R., Tonelli, L.A., Anderson, M.B., Green, D.J., Herrin, G.L., Cotton, R.W., Gottschall, J.L. and Garner, D.D. (1990). Use of deoxyribonucleic acid (DNA) fingerprints for identity determination: comparison with traditional paternity testing methods-part II. *J. Forensic Sci.* 35, 1270-1276.

Mathias, N., Bayes, M. and Tyler-Smith, C. (1994). Highly informative compound haplotypes for the human Y Chromosome. *Hum. Mol. Genet.* 3, 115-123.

Micka, K.A., Sprecher, C.J., Lins, A.M., Comey, C.T., Koons, B.W., Crouse, C., Endean, D., Pirelli, K., Lee, S.B., Duda, N., Ma, M. and Schumm, J.W. (1996). Validation of multiplex polymorphic STR amplification sets developed for personal identification applications. *J. Forensic Sci.* 41(4), 582-590.

Miesfeld, R., Krystal, M. and Arnheim, N. (1981). A member of new repeated sequence family, which is conserved throughout eukaryotic evolution is found between the human delta and beta globin genes. *Nucleic Acids Res.* 9, 5931-5947.

Miller, M.P. (1998). Tools for Population Genetic Analysis, computers programme for the analysis of allelic data. Free programme distributed by the authors over the internet from the TFPGA Home Page at <http://www.public.asu.edu/~mmille8/tfpga.htm>

Mills, K.A., Even, D. and Murray, J.C. (1992). Tetranucleotide repeat polymorphism at the human alpha fibrinogen locus (FGA). *Hum. Mol. Genet.* 1, 779.

Miscicka-Sliwka, D., Czarny, J., Berent, J.A., Grzybowski, T. and Wozniak, M. (1999). Population genetics of the STRs vWA, D3S1358 and FGA in the Pomerania-Kujawy region of Poland. *Int. J. Leg. Med.* 112, 391-392.

Moller, A. and Brinkmann, B. (1994). Locus ACTBP2 (SE33)-sequencing data reveal considerable polymorphism. *Int. J. Leg. Med.* 106, 262-267.

Moller, A., Meyer, E. and Brinkmann, B. (1994). Different types of structural variation in STR, HumFES/FPS, HumvWA and HumD21S11. *Int. J. Leg. Med.* 106, 319-323.

Moller, A., Schurenkamp, M. and Brinkmann, B. (1995). Evaluation of an ACTBP2 ladder composed of 26 sequenced alleles. *Int. J. Leg. Med.* 108, 75-78.

Moos, M. and Gallwitz, D. (1983). Structure of two human β -actin-related processed genes one of which is located next to a simple repetitive sequence. *EMBO J.* 2, 757-761.

Morton, N.E. (1992). Genetic structure of forensic populations. *Proc. Natl. Acad. Sci. USA* 89, 2556-2560.

Morton, N.E., Collins, A. and Balazs, I. (1993). Kinship bioassay on hypervariable loci in Blacks and Caucasians. *Proc. Natl. Acad. Sci. USA* 90, 1892-1899.

Nagai, A., Watanabe, Y., Takayama, T., Yamada, S., Bunai, Y., Nakamura, I. and Ohya, I. (1998). Allele frequencies and sequence data of the STR locus D11S554 in a

Japanese population, in Olaisen B., Brinkmann B., Lincoln P. (Eds), Progress in Forensic Genetics 7, Excerpta Medica, Amsterdam pp. 318-320.

Nakamura, S., Nagai, T., Matsui, H., Tabuchi, M., Sugie, H., Furukawa, M., Imai, H. and Kurihara, K. (2000). Genetic analysis of AmpF ℓ STR Profiler Plus loci in Japanese, in Olaisen B., Brinkmann B., Lincoln P. (Eds), Progress in Forensic Genetics 8, Excerpta Medica, Amsterdam pp. 236-238.

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E. and White, R. (1987). Variable number of tandem repeat (VNTR) markers for human gene mapping. Science 235, 1616-1622.

Nata, M., Nagae, M., Aoki, Y., Sagisaka, K. and Uehara, S. (1993). Prenatal paternity testing with DNA analysis. Int. J. Leg. Med. 106(3), 160-162.

National Research Council (1996). The evaluation of Forensic DNA Evidence. National Academy Press, Washington, D.C.

Nei, M. (1973). Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. USA 70, 3321-3323.

Neuhuber, F., Klintschar, M. and Radacher, M. (1998). A collaborative genetic study on the STR system FGA in two Austrian population samples. Forensic Sci. Int. 91, 1-6.

Nichols, R.A. and Balding, D.J. (1991). Effects of population structure on DNA fingerprint analysis in forensic science. *Heredity* 66, 297-302.

Okazaki, K.T., 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(6), 404-411.

Olaisen, B., Bar, W., Brinkmann, B., Budowle, B., Carracedo, A., Gill, P., Lincoln, P., Mayr, W. R. and Rand, S. (1998). DNA Recommendations 1997 of the International Society of Forensic Genetics. *Vox Sanguinis* 74(1), 61-63.

Oldroyd, N.J., Urquhart, A.J., Kimpton, C.P., Millican, E.S., Watson, S.K., Downes, T. and Gill, P.D. (1995). A highly discriminating octoplex short tandem repeat polymerase chain reaction system suitable for human individual identification. *Electrophoresis* 16, 334-337.

Ovington, A., Daselaar, P., Sjerps, M. and Kloosterman, A. (1997). A Dutch population study of the STR loci D21S11 and HUMFIBRA. *Int. J. Leg. Med.* 110, 14-17.

Pancorbo, M.M., Castro, A., Fernandez, I. F. and Orad, A. G. (1998). Population genetics and forensic applications using multiplex PCR (CSF1PO, TPOX, and TH01) loci in the Basque country. *J. Forensic Sci.* 43(6), 1181-1187.

Pastore, L., Vuttariello, E., Sarrantonio, C., Coto, I., Roviello, S., Fortunato, G., Salvatore, F. and Sacchetti, L. (1996). Allele frequency distributions at several variable number of tandem repeat (VNTR) and short tandem repeat (STR) loci in a restricted Caucasian population from South Italy and their evaluation for paternity and forensic use. *Molecular and Cellular Probes* 10, 299-308.

Peake, H.K., Bowen, D., Bignell, P., Liddell, M.B., Sadler, J.E., Standen, G. and Bloom, A.L. (1990). Family studies and prenatal diagnosis in severe von Willebrand disease by polymerase chain reaction amplification of a variable tandem repeat region of the von Willebrand factor gene. *Blood* 76, 555-561.

Pena, S.D.J. and Chakraborty, R. (1994). Paternity testing in the DNA era. *Trends Genetics* 10, 204-209.

Perez-Lezaun, A., Calafell, F., Seielstad, M., Mateu, E., Comas, D. and Bosch, E. (1997). Population genetics of Y-chromosome short tandem repeats in humans. *J. Mol. Evol.* 45, 265-270.

Pestoni, M.L., Cal, M.L., Lareu, M.V., Rodriguez-Calvo, M.S. and Carracedo, A. (1998). Y Chromosome STR haplotypes, genetic and sequencing data of the Galician population (NW Spain). *Int. J. Leg. Med.* 112, 15-21.

Phromchotikul, T., Browne, D. and Litt, M. (1992). Microsatellite polymorphisms at the D11S554 and D11S569. *Hum. Mol. Genet.* 1, 214.

- Pilkington, S., Summers, C., Thein, S.L., O Connor, N.T.J. and Wainscoat, J.S. (1987). Hair root DNA, a source of constitutional DNA in leukaemia. *Lancet* I, 112.
- Pinheiro, M.F., Pontes, M.L., Abrantes, D., Castro, A., Fernandez- Fernandez, I. and de Pancorbo, M.M. (2000). North Portugal population genetic data for nine STRs loci, in Olaisen B., Brinkmann B., Lincoln P. (Eds), *Progress in Forensic Genetics* 8, Excerpta Medica, Amsterdam pp. 205-207.
- Poeltl, R., Luckenbach, C., Fimmers, R. and Ritter, H. (1997). Typing of the short tandem repeat D8S347 locus with different fluorescence markers. *Electrophoresis* 18, 2871-2873.
- Polesky, H.F. and Krause, H.D. (1977). Blood typing in disputed paternity cases, capabilities of American laboratories. *Transfusion* 17, 521-524.
- Polymeropoulos, M.H., Rath, D.S., Xiao, H. and Merrill, C.R. (1992). Tetranucleotide repeat polymorphism at the human beta-actin related pseudogene H-beta Ac-psi-2 (ACTBP2). *Nucleic Acids Res.* 20, 1432.
- Power InStat 2.03 (1995). A statistical programme on a Macintosh computer provided by Jane, N. Glasgow Caledonian University.
- 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.A., Caskey, C.T., Lins, A.M., Sprecher, C.J., Brinkmann, B. and Schumm, J.W. (1994). Allelic ladder characterization 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.A., Jin, L., Caskey, T.C. and Schumm, J.W. (1993). Identification of repeat sequence heterogeneity at the polymorphic short tandem repeat locus HUMTH01 [AATG]_n and reassignment of alleles in population analysis by using a locus specific ladder. *Am. J. Hum. Genet.* 53, 953-958.

Puregene™ DNA Isolation Kit: User's Manual (1992). Genra Systems, Inc. Minneapolis, MN 55447, USA.

Rand, S., Puers, C., Skowasch, K., Wiegand, P., Budowle, B. and Brinkmann, B. (1992). Population genetics and forensic efficiency data of 4 AMPFLP's. *Int. J. Leg. Med.* 104, 329-333.

Ricci, U., Vieri, F., Guarducci, S., Biondi, C., Gori, F., Sereni, A. and Giovannucci Uzielli, M.L. (2000). A survey of nine Y-STRs a forensic identification markers in the population of Tuscany (Central Italy), in Olaisen B., Brinkmann B., Lincoln P. (Eds), *Progress in Forensic Genetics* 8, Excerpta Medica, Amsterdam, pp. 275-277.

Richards, B., Skoletsky, J., Shuber, A.P., Balfour, R., Stern, R.C., Dorkin, H.L. Parad, R.B., Witt, D. and Klinger, K.W. (1993). Multiplex PCR amplification from the CFTR gene using DNA prepared from buccal brushes/swabs. *Hum. Mol. Genet.* 2, 159-163.

Robertson, J.M. (1994). Evaluation of native and denaturing polyacrylamide gel electrophoresis for short tandem repeat analysis. *Advances in Forensic Haemogenetics* Volume 5, pp. 320-322.

Roewer, L. (1998). The Y chromosome, forensic application and evolutionary aspects, in Olaisen B., Brinkmann B., Lincoln P. (Eds), *Progress in Forensic Genetics* 7, Excerpta Medica, Amsterdam pp.407-412.

Roewer, L., Arnemann, J., Spurr, N.K., Grzeschik, K.H. and Eppelen, J.T. (1992). Simple repeat sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts. *Hum. Genet.* 89, 389-394.

Roewer, L., Kayser, M., Dieltjes, P., Nagy, M., Bakker, E., Krawczack, M. and de Knijff, P. (1996). Analysis of molecular variance (AMOVA) of Y Chromosome-specific microsatellites in two closely related human populations. *Hum. Mol. Genet.* 5, 1029-1033.

Rolf, B., Schurenkamp, M., Junge, A. and Brinkmann, B. (1997). Sequence polymorphism at the tetranucleotide repeat of the human beta-actin related pseudogene H-beta-Ac psi-2 (ACTBP2) locus. *Int. J. Leg. Med.* 110, 69-72.

Rolf, B., Waterkamp, K. and Huhne, J. (1998). Allele frequency data for the FGA locus in eight populations. *Int. J. Leg. Med.* 111, 55-56.

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.

Roy, R., Steffens, D.L., Gartside, B., Jang, G.Y. and Brumbaugh, J.A. (1996). Producing STR locus patterns from bloodstains and other forensic samples using an infrared fluorescent automated DNA sequencer. *J. Forensic Sci.* 41, 418-424.

Rozen, S. and Skaletsky, H.J. (1996, 1997). Primer3. Code available at http://www-genome.wi.mit.edu/genome_software/other/primer3.html.

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 the allele frequencies in different population samples and evaluation for forensic medicine. *Forensic Sci. Int.* 68, 91-102.

Santos, F.R., Gerelsaikhon T., Munkhtuja B., Oyunsuren T., Epplen J. and Pena S.D.J. (1996). Geographic differences in the allele frequencies of the human Y-linked tetranucleotide polymorphism DYS19. *Hum. Genet.* 97, 309-313.

Santos, F.R., Pena, S.D.J. and Tyler-Smith, C. (1995). PCR haplotypes for the human Y Chromosome based on alphoid satellite DNA variants and heteroduplex analysis. *Genetics* 165, 191-198.

Santos, S.M.M., Budowle, B, Smerick, J.B., Keys, K. M. and Moretti, T.R. (1996). Portuguese population data on the six short tandem repeat loci-CSF1PO, TPOX, THO1, D3S1358, VWA, FGA. *Forensic Sci. Int.* 83, 229-235.

Schneider, H.R., Rand, S., Schmitter, H. and Weichold, G. (1998). ACTBP2-nomenclature recommendations of GEDNAP. *Int. J. Leg. Med.* 111, 97-100.

Schneider, P.M. (1997). Basic issues in forensic DNA typing. *Forensic Sci. Int.* 88, 17-22.

Schneider, P. M., d'Aloja, E., Dupuy, B. M., Eriksen, B., Jangblad, A., Kloosterman, A. D., Kratzer, A., Lareu, M. V., Pfitzinger, H., Rand, S., Scheithauer, R., Schmitter, H., Skitsa, I., Syndercombe-Court, D. and Vide, M. C. (1999). Results of collaborative study regarding the standardization of the Y-linked STR system DYS385 by the European DNA Profiling (EDNAP) group. *Forensic Sci. Int.* 102(2-3), 159-165.

Sebetan, I.M. and Hajar, H.A. (1998). Analysis of the short tandem repeat (STR) locus HumVWA in a Qatari population. *Forensic Sci. Int.* 95(2), 169-171.

Sensabaugh, G.F. (1982). Isozymes in forensic science. *Isozymes Curr. Top. Biol. Med. Res.* 6, 247-282.

Silver, H. (1989). Paternity testing. *Critical Reviews in Clinical Laboratory Sciences* 27(5), 391-408.

Smith, J.C., Anwar, R., Riley, J., Jenner, D., Markham, A.F. and Jeffreys, A.J (1990). 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, L.M., Sanders, J.Z., Kaiser, R.J., Hughes, P., Dodd, C., Connell, C.R., Heiner, C., Kent, S.B.H. and Hood, L. (1986). Fluorescence detection in automated DNA sequence analysis. *Nature* 321, 674-679.

Sparkes, R., Kimpton, C.P., Gilbard, S., Carne, P., Andersen, J.F., Oldroyd, N.J., Thomas, D., Urquhart, A. and Gill, P. (1996). The validation of a 7-locus multiplex STR test for use in forensic casework. *Int. J. Leg. Med.* 109, 195-204.

Sparkes, R., Kimpton, C.P., Watson, S., Oldroyd, N.J., Clayton, T., Barnett, L., Arnold, J., Thompson, C., Hale, R., Chapman, J., Urquhart, A. and Gill, P. (1996). The validation of a 7-locus multiplex STR test for use in forensic casework (1) Mixtures, ageing, degradation and species studies. *Int. J. Leg. Med.* 109, 186-194.

Sparkes, R., Kimpton, C.P., Watson, S., Oldroyd, N.J., Urquhart, A. and Gill, P. (1997). The forensic validation of a 7-locus multiplex STR test. *Proceedings from the First European Symposium on Human Identification (1996)* pp. 82-89.

Sprecher, C., Krenke, B., Amiott, B., Rabbach, D. and Grooms, K. (2000). The PowerPlex™ 16 system. *Profiles in DNA* 4(1), 3-6.

Sprecher, C., Puers, C., Lins, A.M. and Schumm, J.W. (1996). General approach to analysis of polymorphic short tandem repeat loci. *BioTechniques* 20, 266-276.

Strom, C., Rechisky, S., Ginsberg, N., Verlinsky, O. and Verlinsky, Y. (1996). Prenatal paternity testing with deoxyribonucleic acid techniques. *Am. J. Obstet. Gynecol.* 174(6), 1849-1853.

Sullivan, K.M., Pope, S., Gill, P. and Robertson, J.M. (1992). Automated DNA profiling by fluorescent labelling of PCR products. *PCR Methods and Appl.* 2, 34-40.

Szibor, R., Lautsch, S., Plate, I., Bener, K. and Krause, D. (1998). Population genetic data of the STR HumD3S1358 in two regions of Germany. *Int. J. Leg. Med.* 111, 160-161.

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.

Thomson, D. M., Brown, N.N. and Clague, A.E. (1992). Routine use of hair root or buccal swab specimens for PCR analysis: advantages over using blood. *Clinica Chimica Acta* 207, 169-174.

Thomson, J.A., Pilotti, V., Stevens, P., Ayres, K.L. and Debenham, P.G. (1999). Validation of short tandem repeat analysis for the investigation of cases of disputed paternity. *Forensic Sci. Int.* 100(1-2), 1-16.

Trabetti, E., Casartelli, A. and Pignatti, P.F. (1994). Application of a Y-linked tetranucleotide repeat polymorphism to paternity cases. *Advances in Forensic Haemogenetics Volume 5*, pp. 187-189.

Trabetti, E., Galavotti, R. and Pignatti, P.F. (1993). Genetic variation in the Italian population at five tandem repeat loci amplified in vitro, use in paternity testing. *Molecular Cell Probes* 7, 81-87.

Tzeng, C.H., Lyou, L.Y., Chen, Y.R., Hu, H.Y., Lin, J.S., Wang, S.Y. and Lee, J.C.I. (2000). Determination of sibship by PCR-amplified short tandem repeat. *J. Forensic Sci.* 45(1), 181-183.

Underhill, P.A., Jin, L., Zemans, R., Oefner, P.J. and Cavalli-Sforza, L.L. (1996). A pre-Columbian Y Chromosome-specific transition and its implications for human evolutionary history. *Proc. Natl. Acad. Sci. USA* 93(1), 196-200.

Urquhart, A.J., Chiu, C.T., Clayton, T.M., Downes, T.J., Frazier, R.R.E., Jones, S., Kimpton, C.P., Lareu, M.V., Millican, E.S., Oldroyd, N.J., Thompson, C., Watson, S., Whitaker, J.P. and Gill, P. (1995). Multiplex STR systems with fluorescent detection

as human identification markers. Proceedings from the 5th International Symposium on Human Identification (1994) pp. 73-83.

Urquhart, A.J., Kimpton, C.P., Downes, T.J. 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.

Urquhart, A.J., Oldroyd, N.J., Downes, T.J., Barber, M.D., Alliston-Greiner, R., Kimpton, C.P. and Gill, P. (1996). Selection of STR loci for forensic identification systems. *Advances in Forensic Haemogenetics Volume 6*, pp. 115-117.

Urquhart, A.J., Oldroyd, N.J., Kimptone, C.P. and Gill, P. (1995). Highly discriminating heptaplex short tandem repeat PCR system for forensic identification. *BioTechniques* 18, 116-121.

van Oorschot, R.A.H., Gutowski, S.J. and Robinson, S.L. (1994). HUMTH01: Amplification, species specificity, population genetics and forensic applications. *Int. J. Leg. Med.* 107, 121-126.

Verma, R.C., Luke, S. and Dhawan, P. (1992). Twins with different fathers. *Lancet* 339, 63-64.

Wallin, J.M., Buoncristiani, M.R., Lazaruk, K.D., Fildes, N., Holt, C.L. and Walsh, P.S. (1998). TWGDAM validation of the AmpF ℓ STR blue PCR amplification kit for forensic casework analysis. *J. Forensic Sci.* 43, 854-870.

Walsh, P.S., Erlich, H.A. and Higuchi, R. (1992). "Preferential PCR amplification of alleles, mechanisms and solutions". *PCR Methods and Appl.* 1, 241-250.

Walsh, P.S., Fildes, N.J. and Reynolds, R. (1996). Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA. *Nucleic Acids Res.* 24, 2807-2812.

Weber, J.L. and May, P.E. (1989). Abundant class of human DNA polymorphisms, which can be typed using polymerase chain reaction. *Am. J. Hum. Genet.* 44, 388-396.

Weir, B.S. (1992). Independence of VNTR alleles defined by fixed bins. *Genetics* 130, 873-887.

White, H.W. and Kusukawa, N. (1997). Agarose-based system for separation of short tandem repeat loci. *BioTechniques* 22(5), 976-980.

Whitfield, L.S., Suiston, J.E. and Goodfellow, P.N. (1995). Sequence variation of the human Y Chromosome. *Nature* 378, 379-380.

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.

Wilson, J.W., Williamson, R., Petrou, M., Papaioannou, D. and Parkin, B.H. (1993). Variation of short tandem repeats within and between populations. *Hum. Mol. Genet.* 2(7), 1023-1029.

Wyman, A.R. and White, R. (1980). A highly polymorphic locus in human DNA. Proc. Natl. Acad. Sci. USA 77(11), 6754-6758.

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., Brinkmann B., Lincoln P. (Eds), Progress in Forensic Genetics 8, Excerpta Medica, Amsterdam pp. 239-241.

Yamamoto, T., Uchihi, R., Nozawa, H., Huang, X. L., Leong, Y. K., Tanaka, M., Mizutani, M., Tamaki, K. and Katsuyama, Y. (1998). Polymorphism analysis of nine STR loci on Japanese using multiplex PCR and capillary electrophoresis, in Olaisen B., Brinkmann B., Lincoln P. (Eds), Progress in Forensic Genetics 7, Excerpta Medica, Amsterdam pp. 390-391.

Zehner, R. (1998). Population genetic study using the AmpFI STR Blue kit, in Olaisen B., Brinkmann B., Lincoln P. (Eds), Progress in Forensic Genetics 7, Excerpta Medica, Amsterdam pp. 392-394.

Ziegle, J.S., Su, Y., Corcoran, K.P., Nie, L., Mayrand, E., Hoff, L.B., McBride, L.J., Kronick, M.N. and Diehl, S.R. (1992). Application of automated DNA sizing technology for genotyping microsatellite loci. Genomics 14, 1026-1031.