The Islamic University - Gaza

Deanery of Higher Education

Faculty of Science

Master of Biological Sciences

Medical Technology

الجامعة الإسلامية – غزة

عمـــادة الدراسات العليـا

كليـــــــة العلـــــــوم

ماجستير العلـــوم الحياتية

التحـــــاليل الطبيـــــة

Genetic Variation of 15 Autosomal Short Tandem Repeat (STR) Loci in the Palestinian Population of Gaza Strip

Thesis Prepared By
Masood Said Abu Halimah

Supervisor
Prof. Fadel A. Sharif

A Thesis Submitted to the Faculty of Science in Partial Fulfillment of the requirements for the degree of Master of Science in Biological Sciences – Medical Technology

December 20098

DECLARATION

I hereby declare that this submission is my own original work and that, to the best of my knowledge and belief, it contains no materials previously published or written by another person nor materials which to a substantial extent has been accepted for award of any other degree of the university or other institute, except where due acknowledgement has been mad in the text.

Signature	Name	Date
Masoud Abu Halimah	Masoud Said Abu Halimah	Dec.2008

Copyright:
All Rights Reserved: No part of this work can be copied, translated or stored in any kind of a retrieval system, without prior permission of the author.

ABSTRACT

Allele frequencies for 15 autosomal Short Tandem Repeat (STR) loci namely Penta E, D18S51, D21S11, TH01, D3S1358, FGA, TPOX, D8S1179, vWA, Penta D, CSF1PO, D16S539, D7S820, D13S317, D5S818 and Amelogenin, were studied in a sample of 125 unrelated individuals from the Palestinian population of Gaza Strip. PCR amplification was performed with the PowerPlex® 16 PCR amplification system and the amplified products were separated and detected using the ABI 3130 DNA genetic analyzer using reference-sequenced ladder. No significant departure from Hardy Weinberg Equilibrium (HWE) expectations were observed for 11 of the 15 STR loci analyzed (*P*-value >0.05). The exceptions were the Penta E, D5S818, D13S317 and vWA loci but when the Bonferroni procedure was used as a correction for the multiple tests performed on a population sample, none of the previous *P*-values could be considered significant. No linkage disequilibrium was found between the loci located on the same chromosome, and thus they are statistically independent. The F_{IS} value for Gazan population has been calculated as – 0.024. This low F_{IS} value indicates that the population is in random mating with high level of heterozygosity. The combined probability of exclusion, power of discrimination, probability of matching value for all the 15 STR loci were 0.999998595; 0.999999986 and 2.73×10^{-18} , respectively. In the light of the obtained results, we can confidently conclude that these 15 markers are suitable for forensic analyses and paternity testing in the Palestinian population of the Gaza Strip.

Key Words: Short tandem repeat; Allele frequencies; Gaza Strip; Palestine; PowerPlex®16 System

الاختلاف الجيني لـ 15 موقع تكراري ترادفي متنوع في الجينوم البشرية في قطاع غزة

ملخص الدراسة:

الهدف من هذه الدراسة هو دراسة و تحديد الاختلاف الجيني و نسبة البوليمورفية في 15 موقع علي الجينوم البشري ، هذه المواقع تسمى STR (العدد المتابين من المتكررات المترادفة) و هي كالأتي Penta E, D18S51, D21S11, TH01 D3S1358, FGA, TPOX, D8S1179, vWA, Penta D, CSF1PO, D16S539, D7S820, D13S317, D5S818 في عينة مكونة من 125 شخص غير مرتبطين ببعضهم بصلة قرابة في قطاع غزة. و باستخدام جهاز المحلل الجيني من نوع ABI3130 اتضح بأنه يوجد انحرف عن قاعدة Hardy Weinberg Equilibrium (HWE) في أربع مواقع فقط هذه المواقع Penta D,5S818, D13S317 و الموقع vWA و بعد إجراء نظام Bonferroni تبين بأنه لا يوجد أية انحراف بهذه المواقع. تم فحص نسبة ارتباط المواقع التي تقع علي نفس الكرموزومات فتبين بأنها لا ترتبط ببعضها البعض و تنتقل بطريقة مستقلة من جيل لأخر، ثم تم قياس معدل التوالد الداخلي لسكان قطاع غزة فتبين بأنه -0.024 و هذا يشير بان المجتمع في حالة تزاوج عشوائي و نسبة Heterozygosity عالية جداً و ليس له أية دلالة إحصائية.و تم قياس كلاً من نسبة الاستثناء، نسبة التفرقة، و نسبة التوافق لجميع المواقع و كانت 0.999998595، 0.9999999986 و 2.73×10^{-18} علي التوالي. و علي ضوء هذه النتائج يمكن أن نستنتج بان هذه المواقع تستخدم لفحوصات الطب الشرعي و فحوصات اثبات الأبوة في قطاع غزة.

الكلمات المفتاحية: موقع تكراري ترادفي STR ، ترددات الاليل ، قطاع غزة، فلسطين ، نظام PowerPlex®16

DEDICATION

This work is dedicated to:

My great parents, who raised me, supported me, taught me, and loved me.

My brothers and nephews.

Palestine, the Land of the three monotheistic faiths, is where the Palestinian people was born, on which they grew, developed and excelled.

The people of Palestine who suffered and struggled for free Palestine.

ACKNOWLEDGEMENTS

All of this work was carried out at the Islamic University of Gaza – Genetics Laboratory during the period April 2007 to September 2008.

There are lots of people I would like to thank for a huge variety of reasons.

Firstly, I would like to thank my supervisor, *Prof. Fadel Sharif***.** I could not have imagined having a better advisor and mentor for my master.

I am grateful to *Ms. Lennie Pineda Bernal*, Unversity of Zulia - Venezuela for providing much appreciated interest, invaluable guidance and advices.

I am extremely grateful for the assistance, generosity, and advice I received from *Ms. Kim Sturk* and *Ms. Jodi Irwin* at the Armed Forces DNA Identification Laboratory - USA.

I am especially indebted to the outstanding staff of Genetics Laboratory at the Islamic University: *Mr. Nasser Abu Shaban and Mr. Mohammed Ashour* in which both were very valuable in many parts of my work.

It is difficult to overstate my appreciation to my general manger at TransOrint Medical Co, *Eng. Hamam El-Rayyes* for his support.

My especial thanks go to the master degree program coordinators at the Islamic University, especially *Dr. Aboud A Keshawi* and Mr. *Mohammed Awda* who are still struggling for the success of this program.

Finally, I am tempted to individually thank all of my friends, especially *Mr. Ashraf Shaqalaih*, *Mr. Mohammed Laqqan* and *Medhat Abu Wale* for being some of my very best friends.

TABLE OF CONTENTS

Item	Page
Declaration ……	i
Abstract ……	ii
Arabic Abstract ……	iii
Dedication ……	iv
Acknowledgement ……	v
Table of Contents ……	vi
List of Tables ……	ix
List of Figures ……	x
List of Appendixes ……	xi
Abbreviations ……	xii
Chapter One: Introduction ……	1
1.1. Overview ……	1
1.2. Aim of the study ……	4
1.3. Objectives of the study ……	4
Chapter Two: Review of Literature ……	5
2.1. Introduction ……	5
2.2. Repetitive DNA ……	9
2.3. Isolation and Types of STR Markers ……	11
2.4. Characteristics of STRs used in DNA Typing ……	12
2.5. STR Locus Nomenclature ……	15
2.6. Advantages of STRs in DNA Typing ……	16
2.7. DNA Typing Studies in Different Populations ……	16
2.8. DNA Typing ……	17
2.8.1. Multilocus (Minisatellite) Testing ……	17
2.8.2. Single-Locus Probes ……	18
2.8.3. Short Tandem Repeat (STR) Testing ……	18
2.9. Selection of STR loci for DNA Typing ……	20
2.9.1. First Generation Multiplex STR Typing ……	20
2.9.2. Second generation of Multiplex STR Typing ……	20
2.9.3. Expansion of the SGM System ……	21
2.10. Allelic Ladder ……	23
2.11. Factors that Complicate STRs Interpretation ……	26
2.11.1. Repeat Slippage ……	26
2.11.2. Non-Template addition ……	27
2.11.3. Microvariant Alleles ……	27
Chapter Three: Materials and Methods ……	28
3.1. Study Population ……	28
3.2. Ethical Considerations ……	28
3.3. Materials ……	28
3.3.1. Chemicals and Reagents ……	28
3.3.2. Disposables ……	29
3.3.3. Equipment ……	30

3.4. Methods 30
3.4.1. Blood Collection 30
3.4.2. DNA Typing Analysis 31
3.4.2.1. DNA Extraction 31
3.4.2.2. PCR – Amplification of Microsatellite Loci 31
3.4.2.3. Capillary Electrophoresis 33
3.4.2.4. Genotype Determinations 35
3.4.2.5. Controls for Genotyping 37
3.5. Statistical Analysis of Population Samples 38
3.5.1. Expected Heterozygosity 39
3.5.2. Match Probability 39
3.5.3. Power of Discrimination 39
3.5.4. Polymorphism Information Content 40
3.5.5. Power of Exclusion 40
3.5.6. Paternity Index 40
Chapter Four: Results 41
4.1. Genotyping of DNA Samples 41
4.2. Allele Frequencies and Genotypes 43
4.3. Allele Frequency of Amelogenin Locus 43
4.4. Heterozygosity and Homozygosity Calculations 44
4.5. Test for Inbreeding Coefficient (F_{IS}) 44
4.6. Test for Hardy- Weinberg Equilibrium 45
4.7. Test for Linkage Disequilibrium 45
4.8. Variability of Alleles and Genotypes 46
4.9. Most Common Genotypes Frequencies 47
4.10. DNA Profile Frequency Estimates 48
4.11. Forensic Efficiency Parameters 50
4.12. An Example of Parentage Testing 50
4.12.1. Paternity Index 51
4.12.2. Combined Paternity Index 51
4.12.3. Probability of Paternity 51
Chapter Five: Discussion 53
5.1. Distribution of Alleles Frequency 53
5.2. Test for Microsatellite Alleles Variability 54
5.3. Heterozygosity and Homozygosity 55
5.4. Inbreeding Coefficient 56
5.5. Hardy Weinberg Equilibrium 57
5.6. Linkage Disequilibrium 58
5.7. Polymorphism Information Content 59
5.8. Power of Discrimination 60
5.9. Random Matching Probability 61
5.10. Exclusion Probability 61
5.11. Paternity Testing 62
Chapter Six: Conclusions and Recommendations 64
6.1. Conclusions 64
6.2. Recommendations 65

References ………………………………………………………………………... 66
Appendix 1 ………………………………………………………………………... 79
Appendix 2 ………………………………………………………………………... 81

LIST OF TABLES

Table		Page
Table 2.1 :	Types of human repeated DNA sequences…………………………………………	10
Table 2.2 :	Types and examples of STR repeat units………………………………………..	11
Table 2.3 :	Categories for STR Markers………………………………………………………	12
Table 2.4 :	The most polymorphic alleles described for various populations ……………..	17
Table 2.5 :	Information on commercially available STR multiplexes………………………..	22
Table 2.6 :	The PowerPlex® 16 System locus-specific information………………………..	25
Table 3.1 :	Chemicals and reagents used in the present study……………………………..	29
Table 3.2 :	Major disposables used in the present study……………………………………	29
Table 3.3 :	Major equipment used in the present study………………………………………	30
Table 3.4 :	The PowerPlex® 16 System Allelic Ladder information…...............................	37
Table 3.5 :	The PowerPlex® 16 System allele determinations in 9947A standard ………..	38
Table 4.1 :	Genotype array and allele count for the STR locus TH01 from unrelated Gazan samples ………………………………………………………………………..	41
Table 4.2 :	Amelogenin Locus size according to the PowerPlex®16 Allelic Ladder among Gazan population…………………………………………………	43
Table 4.3 :	Observed homozygosity, Observed heterozygosity and Expected heterozygosity for different fifteen STR loci in Gazan population………………	45
Table 4.4 :	Observed numbers of alleles and number of possible genotypes for Gazan population………………………………………………………………………	46
Table 4.5 :	Calculations for the theoretically most common genotype frequencies and profile frequency based on the two most common alleles found in a Gazan population………………………………………………………………………	47
Table 4.6 :	An example for calculating the DNA profile frequency (or combined match probability) using alleles from five STR loci………………………………………	49
Table 4.7 :	Forensic efficiency parameters of the fifteen STR loci in 125 unrelated Gazan individuals………………………………………………………………………	50
Table 4.8 :	An example for parentage test……………………………………………………..	52
Table 5.1 :	The observed heterozygosity in different populations…………………………...	55
Table 5.2 :	The degree of polymorphism using PIC and heterozygosity in different populations. ………………………………………………………………………..	59
Table 5.3 :	The Power of discrimination in some different populations. ……………………	60
Table 5.4 :	The Power of exclusion in different populations…………………………………	62

LIST OF FIGURES

Figure		Page
Figure 2.1 :	Comparison of DNA Typing Technologies …………………	8
Figure 2.2 :	Schematic diagram of minisatellite and microsatellite (STR) DNA markers. …………………………………………	10
Figure 2.3 :	Commonly used STR markers and their distributions over the human genome.	23
Figure 2.4 :	Principle of allelic ladder formation. …………………………	24
Figure 2.5 :	STR alleles shown with stutter products ……………………	26
Figure 2.6 :	Schematic diagram of non-template nucleotide addition …	27
Figure 3.1 :	A diagram illustrating spectral calibration in ABI 3130 Genetic Analyzer ……………………………………………	33
Figure 3.2 :	An electropherogram showing the fragments of the Internal Lane Standard 600. …………………………………	34
Figure 3.3 :	Sizing of DNA fragment peaks using internal size standards …………………………………………………..	36
Figure 4.1 :	An electropherogram illustrating PowerPlex®16 Allelic ladder …………………………………………………………..	41
Figure 4.2 :	An electropherogram illustrating the genotyping result on locus TH01 ………………………………….......................	43
Figure 4.3 :	An electropherogram showing Amelogenin locus and its two possible genotypes ………………………………………	44

LIST OF APPENDIXES

Appendix		Page
Appendix 1:	Allele frequencies for the fifteen STR loci in the Palestinian population of Gaza Strip ………………………	79-80
Appendix 2:	Alleles Genotype for the fifteen STR loci in the Palestinian population of Gaza Strip ……………………………………..	81-85

ABBREVIATION

ABI	Applied Biosystems, Inc.
AF	Alleged Father
al	allele
bp	base pair
CDP	Combined Discrimination Power
CEP	Combined Exclusion Probability
CMP	Combined Matching Probability
CODIS	Combined DNA Index System
CPI	Combined Paternity Index
CSF1PO	Human c-fms Proto-Oncogene for CSF-1 Receptor Gene
CXR	Carboxy- X-Rhodamine
DNA	DeoxyriboNucleic Acid
EDTA	Ethylene Diamine Tetraacetic Acid
FBI	Federal Bureau of Investigations
FGA	Human Fibrinogen Alpha Chain Gene
He	Expected Heterozygosity
HLA	Human Leukocyte Antigen
Ho	Observed Heterozygosity
ht	peak height
HUM	Human
HWE	Hardy Weinberg Equilibrium
ILS	Internal Lane Standard
MLP	Multi-Locus Probe
mtDNA	Mitochondrial DNA
NA	Not Available
NDNAD	National DNA Database
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
PD	Power of Discrimination
PE	Probability of Exclusion
PIC	Polymorphism Information Content
POP	Probability of Paternity
POP-4	Performance Optimized Polymer 4
RFLP	Restriction Fragment Length Polymorphism
SBH	Southern Blot Hybridization
SGM	Second Generation Multiplex
SNP	Single Nucleotide Polymorphism
SSR	Simple Sequence Repeats
STR	Short Tandem Repeats
sz	allele size

TH01	Human Tyrosine Hydroxylase Gene
TPI	Typical Paternity Index
TPOX	Human Thyroid Peroxidase Gene
UK	United Kingdom
USA	United States of America
VNTR	Variable Number of Tandem Repeats
vWA	Human von Willebrand Factor Gene

Chapter One

INTRODUCTION

1.1. Overview

DNA analysis has become one of the most definitive methods for human identification in forensic science, paternity testing, missing persons and mass disaster investigations. Major developments in the field of genetics include the discovery of restriction enzymes, Southern Blot Hybridization (SBH) and Polymerase Chain Reaction (PCR) contributed to the rapid development of DNA typing. In order to rightfully identify an individual, an ideal method would give a guaranteed perfect match in which there is no chance anyone else on this planet could have given the same results (that would be a false positive). Likewise, to exclude an individual, one must look at a trait so unique that the chance for a random match of the suspect and the biological specimen (false positive) does not exist. To eliminate false positive results, the occurrence, or frequency of a certain trait must be as small as possible (Butler, 2001).

The human genome was first explored by using Restriction Fragment Length Polymorphism (RFLP) analysis. In 1985, it was discovered that DNA contains highly polymorphic regions of repeated DNA that could be visualized by “Jeffreys probes”. These probes revealed a minisatellites pattern on unknown loci throughout the genome and were called Multi-Locus Probe (MLPs). Scientists soon applied this discovery to human identification and paternity determination. Lack of discrete alleles made it impossible to set up a reference database to calculate allele frequencies rendering a major drawback of MLPs for legal purposes, since random match probabilities could not be generated. Then in 1988, the Single Locus Probes (SLPs) were introduced to the court system. This method is based on known loci that Vary in Number of Tandem Repeats (VNTRs). Combinations of SLPs were used to increase discriminating power and population databases could be

generated. SLPs were used until 1993, after then a new PCR-based technique became popular. This was based on Short Tandem Repeats (STRs) of 2-6 bp. Whereas RFLP analysis required microgram amounts of intact DNA, PCR allowed the amplification and detection of subnanogram amounts of lower molecular weight DNA in hours rather than days (Kloosterman, 2003).

Both STR and VNTR exhibit a high degree of length polymorphism. However, due to the smaller size of the repeat unit of STR markers (2-6 bp) compared to VNTRs (15-35), the former encompasses more allele frequency in different human populations studied (Kimpton *et al.,* 1993). In recent years, as a result of advances in human genome technologies, a polymorphic marker named Single Nucleotide Polymorphism (SNP) has been identified (Twyman and Primrose, 2003). These markers seem to be the most polymorphic sites in the human genome with approximately one in every 1000 bp, with great genetic stability, which is important in paternity testing. Therefore, it has been proposed that these markers should replace STRs in paternity and individual investigations.

At the moment, STRs are used as markers of choice in most forensic, paternity testing and individual identification studies. The degree of allele frequency for each STR marker has been found to be basically population dependent (Urquhart, 1994; Martinez *et al.,* 2003). The selection of each marker for forensic investigations is based on the previous studies performed on the nature of the allele frequency of STRs in the population of interest. This makes the analysis of allele frequency of known STR markers a prerequisite to forensic study in each population.

The genetic markers forming a DNA profile are found at different frequencies in the population. This collection of genetic markers is known as a population genetic database. If the frequencies of the genetic markers making up a DNA profile are very high and found in half of the population, the

statistical strength or discrimination power of the particular DNA profile is not strong. On the contrary, if the genetic markers of the DNA profile are found only in one in a million individuals, then the statistical discrimination power of the DNA profile is very strong. It is, therefore, important that the data on the genetic markers of a population genetic database are statistically tested before the data are used to determine the occurrence rate of a particular DNA profile in criminal caseworks. Neglecting to do so can exaggerate the strength of the evidence against the defendant (Balding, Greenhalgh and Nichols, 1996).

In the United Kingdom, the National DNA Database (NDNAD) centre uses eleven STR loci or markers and in the United States; the Federal Bureau of Investigations (FBI) uses thirteen STR loci or markers as the basis for its forensic DNA database, Combined DNA Index System (CODIS) [Butler *et al.*, 2004]. These STR loci are polymorphic, which means that the genetic code for people, at the STR loci can be different. Thus, STR locus 1A may have variations b, c, d, e, and f. As we have homologous pairs of chromosomes, we will have two copies of each STR, one from our father and one from our mother. For example, an individual could have 1Ab and 1Af, for the STR locus 1A. If 1Ab occurs in the population at 10%, and 1Af occurs in the population at 5%, then there will be a probability of an individual being 1Ab1Af of 0.5%. By testing the STR variation at each genetic locus, an STR profile is generated that is specific to the individual. Though a possibility exists for another individual having the same STR profile, the statistical matching probability with PowerPlex®16 that examines 15 loci simultaneously is 1.2×10^{-18} (Butler, 2006).

In this study, sixteen loci (fifteen STR loci and Amelogenin), including D18S51, D21S11, TH01, D3S1358, FGA, TPOX, D8S1179, vWA, Amelogenin, Penta D, Penta E, CSF1PO, D16S539, D7S820, D13S317 and D5S818, which have been frequently used for forensic and paternity testing in different populations were tested on 125 unrelated individuals from the

Palestinian population of Gaza Strip. Hitherto, this is the first study of its kind on the analysis of the allele frequency and genotyping of those STR loci in a Palestinian population. Therefore, the obtained data could be used to initiate Palestinian genetic database for the STR polymorphic markers. Moreover, the results of this study can show the utility of those STR markers for paternity testing and individual identifications.

1.2. Aim of the study

The first aim of this study was to determine the genetic structure of Gaza Strip population using fifteen autosomal STR loci namely Penta E, D18S51, D21S11, TH01, D3S1358, FGA, TPOX, D8S1179, vWA, Penta D, CSF1PO, D16S539, D7S820, D13S317, D5S818 and Amelogenin.

The second aim was to evaluate the usefulness of these loci in Gaza Strip population for forensic genetic purposes.

1.3. Objectives of the study

- Get hands on experience in performing and interpreting DNA-typing profile (calculation of forensic efficiency parameters).

- Contribute to the establishment of a DNA database that may improve the justice system in Gaza Strip.

Chapter Two

REVIEW OF LITERATURE

2.1. Introduction

DNA-based genetic analysis has a long-standing and important role in research, medical diagnostics, and DNA profiling in individuals investigations. Characterization or typing of blood, serum, and other body fluid and tissues has been used for forensic and clinical purposes for more than 50 years (Committee on DNA Technology in forensic Science, 1992). In the last decade, methods have become available for DNA typing that is, for showing differences in the genetic material itself. Development in genetic and modern molecular biology have accumulated large amount of knowledge about human genome (Gyapay *et al.,* 1994), especially after the significant progress of the Human Genome Project (Collins and Galas, 1993).

The gamut of DNA typing technologies used over the past 15 years for human identity testing is compared in Figure 2.1. The various DNA markers have been divided into four quadrants based on their power of discrimination, i.e., their ability to discern the difference between individuals, and the speed at which they can be analyzed. The first genetic tool for human identification was the use of ABO blood groups after its discovery at the beginning of the twentieth century. There are only four different blood groups (A, B, AB and O), roughly, 40% of the population is type A, and another 40% is type O (Janeway *et al.,* 1997; Olsen *et al.,* 1997). This method was the first genetic tool used for distinguishing between individuals, can be performed in a few minutes but is not very informative. Thus, while the ABO blood groups are useful for excluding an individual from being the source of a sample, the test is not very useful when an inclusion has been made, especially if the sample is type O. For inclusions, this method is not very useful; the chance for a wrongful inclusion of an individual is high due to the limited variability within

the blood group system and the high frequency of the A and the O alleles. Blood group systems, polymorphic protein and enzyme systems were used in combination to decrease the frequency and thereby adding more confidence to the results. The quality and amount of body fluid available however, sometimes limited this. Despite its imperfections, this was the conventional typing that was used before a major breakthrough in 1985, when the first DNA based human identification method was introduced to the forensic world: the MLP technique used in the UK in the 1986 (Jeffreys *et al.,* 1985). DNA results were visualized as a set of parallel bands on a photographic plate. No statistical results were presented. If a sufficient number of bands from a crime evidence sample matched that of a suspect's DNA sample, the evidence would conclusively identify the suspect as the donor of the DNA evidence. Disadvantages of this method were that it could not be used in cases involving relatives and that a large biological sample was required to give a reliable result. The latter was not always possible in forensic work (Jeffreys *et al.,* 1985).

A few years later, multilocus profiling was replaced by SLP technique, which was the method of choice until 1993 when the first fluorescent STR-marker kit became available. The SLP technique was introduced to forensic casework in the UK in 1988. In the USA, this has always been known as the RFLP technique (Gill *et al.,* 1990). Probes were used to target specific loci on DNA. Each probe produced one or two bands, referred to as a homozygote and heterozygote, respectively, depending on whether the pieces of DNA sequence an individual inherited from its two parents were the same or different at the given locus. First, an assessment was made as to whether two DNA profiles were close enough in size to be considered a match. Secondly, a population database was used to give an estimate of the frequency of the particular DNA profile in the population. This technique provided very powerful evidence with expected DNA profile frequencies of one in many million individuals.

After introduction and development of the PCR method (Saiki *et al,* 1986; Mullis *et al,* 1986; Mullis and Faloona, 1987), the sensitivity of DNA analysis increased considerably as samples containing smaller amounts of DNA could now be analyzed after amplification. Even partially degraded DNA could be amplified to some extent. The introduction of automated sequencing and fragment length analysis techniques also facilitated the analysis of the amplified DNA. One of the first DNA analysis methods in forensic science based on the PCR was analysis of polymorphisms in the HLA-DQA1 gene (a gene encoding a human leukocyte antigen cell surface protein) [Saiki *et al,* 1986; Helmuth *et al,* 1990; Hochmeister *et al,* 1991; Allen *et al,* 1995]. However, this method has mostly been replaced by the so-called STR analysis of “microsatellites”. Microsatellites resemble minisatellites, but are generally shorter. They consist of short sequences ~2-6 bp long, which are tandemly repeated, giving a total length for the loci of ~80-400 bp (Edwards *et al,* 1991). In STR analysis, a PCR is performed using primers on each side of the microsatellite, followed by electrophoresis and detection of fragment lengths. The STRs analyzed in forensic investigations are generally 150-400 bp long regions of tetranucleotide repeats (Gill, 2002).

There are hundreds of these STR regions in the human genome, but 10 to 15 regions are sufficient to give high levels of discrimination between individuals (Hammond *et al.,* 1994). The results of STR profile when fluorescently labeled STR-PCR products analyzed by capillary electrophoresis on a genetic analyzer appear as a series of peaks on a graph; each peak is labeled with a specific number according to its position on the graph. The position of the peak consists out of the number of times the particular core sequence is repeated, as well as the flanking regions. The numerical designations are known as alleles. The advantage of the STR-PCR technique over the SLP technique is that the size of the STR peaks can be measured accurately and unambiguously (Kimpton *et al.,* 1994). STR profiles are represented as a series of alleles, two at each locus inherited from each parent. This has facilitated the construction of intelligence

databases containing profiles from unsolved crimes and known criminals. Several STR regions can be amplified at the same time, which was not possible with the SLP technique (Kimpton *et al.,* 1993). Smaller samples of DNA can be analyzed. This makes recovery of information from degraded DNA specimens possible (Hagelberg *et al.,* 1991; Jeffreys *et al.,* 1992; Wiegland *et al.,* 1993; Gill *et al.,* 1994). There is no limit to the discriminating power of the technique because of the large number of STR regions in the human genome (Gill *et al.,* 1996). This method can accommodate a high throughput of samples as well as a reduced cost of analysis compared to the MLP and SLP techniques (Gill *et al.,* 1995). DNA profiles generated with the STR method are automatically placed on a database and the front end of the DNA analysis process (extraction and PCR amplification) can be automated (Butler, 2001).

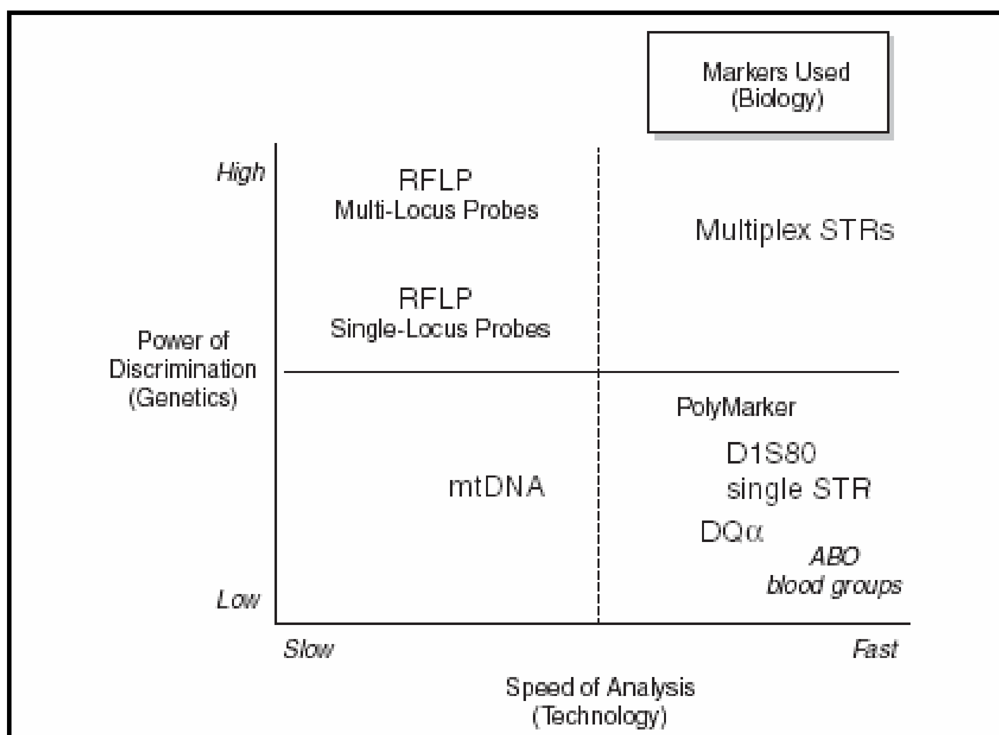

Figure 2.1. Comparison of DNA typing technologies. DNA markers are arbitrarily plotted in relationship to four quadrants defined by the power of discrimination for the genetic system used and the speed at which the analysis for that marker may be performed. (Adapted from Butler, 2005).

Mitochondrial DNA (mtDNA), which is shown in the quadrant with the lowest power of discrimination and longest sample processing time, can be very helpful in forensic cases involving severely degraded DNA samples or when associating maternally related individuals. In many situations, multiple technologies may be used for human identification purposes.

2.2. Repetitive DNA

Eukaryotic genomes are full of repeated DNA sequences (Ellegren, 2004). These repeated DNA sequences come in all types of sizes and are typically designated by the length of the core repeat unit and the number of contiguous repeat units or the overall length of the repeat region. Long repeat units may contain several hundred to several thousand bases in the core repeat. These regions are often referred to as satellite DNA and may be found surrounding the chromosomal centromere. The term satellite arose due to the fact that frequently one or more minor satellite bands were seen in early experiments involving equilibrium density gradient centrifugation (Britten and Kohne 1968; Primrose 1998).

The core repeat unit for a medium length repeat, sometimes referred to as a minisatellite or a VNTR, is in the range of approximately 10–100 bases in length (Tautz, 1993; Chambers and MacAvoy, 2000). The DNA marker D1S80 is a minisatellite with a 16 bp repeat unit and contains alleles spanning the range of 16–41 repeat units (Kasai *et al.,* 1990) as shown in Figure 2.2. DNA regions with repeat units that are 2–6 bp in length are called microsatellites, Simple Sequence Repeats (SSRs), or STRs. Table 2.1 illustrates the types of human repeated DNA sequence. STRs have become popular DNA repeat markers because they are easily amplified by the PCR without the problems of differential amplification. Moreover, the number of repeats in STR markers can be highly variable among individuals, which makes these STRs effective for human identification purposes.

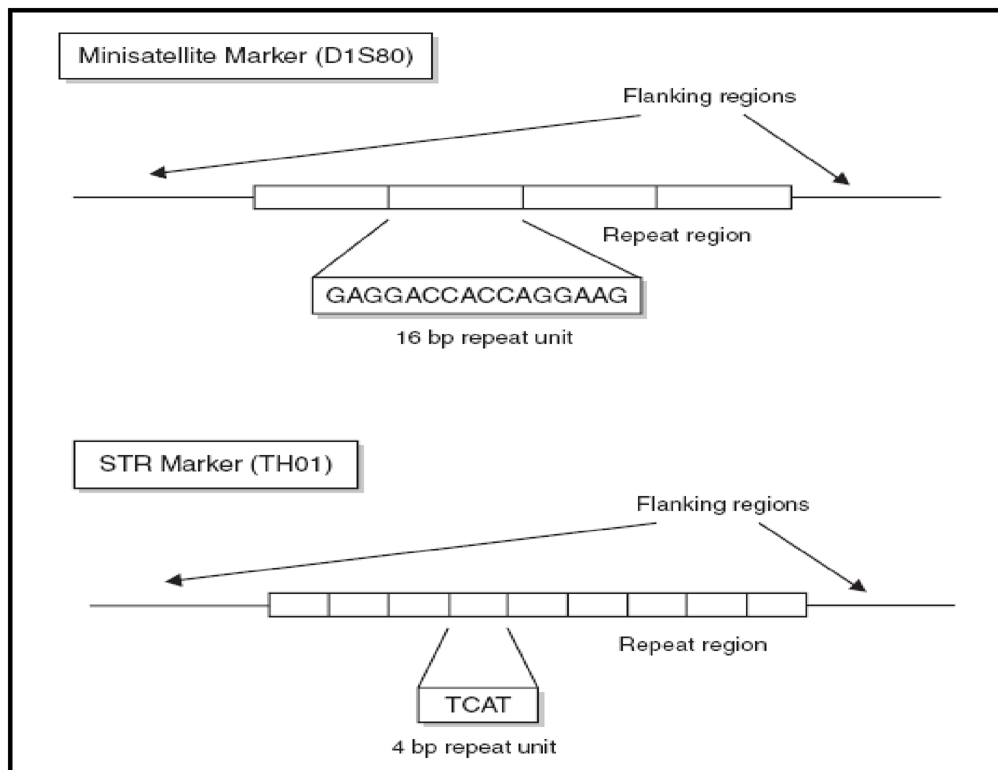

Figure 2.2. Schematic diagram of minisatellite and microsatellite (STR) DNA markers. PCR primers are designed to target invariant flanking sequence regions. The number of tandem repeat units in the repeat regions varies among individuals making them useful markers for human identification (Adapted from Butler, 2005).

Table 2.1. Types of human repeated DNA sequences.

Type of DNA	Description	Length
Satellite DNA	Regions of long stretches of repeated DNA, mostly found along the centromeres.	< 1000 bp
Minisatellite DNA	Type of satellite DNA consisting of medium length repeat units	10-100 bp
VNTR	Term reserved for moderately large repeat, also called “Hypervariable minisatellite DNA”	5-64 bp
Microsatellite DNA	Type of satellite DNA consisting of smaller repeat units the so-called Simple Tandem Repeat, or Short Tandem Repeats (STR)	2-6 bp

2.3. Isolation and types of STR markers

In order to perform analysis on STR markers, the invariant flanking regions surrounding the repeats must be determined. Once the flanking sequences are known then PCR primers can be designed and the repeat region amplified for analysis. New STR markers are usually identified in one of two ways: (1) searching DNA sequence databases such as GenBank for regions with more than six or so contiguous repeat units (Weber and May 1989; Collins *et al.,* 2003, Subramanian *et al.,* 2003); or (2) performing molecular biology isolation methods e.g., cloning and sequencing of randomly selected DNA fragments (Edwards *et al.,* 1991; Chambers and MacAvoy, 2000).

STR repeat sequences are named by the length of the repeat unit. Dinucleotide repeats have two nucleotides repeated next to each other over and over again. Trinucleotides have three nucleotides in the repeat unit, tetranucleotides have four, pentanucleotides have five, and hexanucleotides have six repeat units in the core repeat Table 2.2.

Table 2.2. Types and examples of STR repeat units.

Length of Repeat Unit	Example
Dinucleotide	(CA)(CA)(CA)(CA)
Trinucleotide	(GCC)(GCC)(GCC)
Tetranucleotide	(AATG)(AATG)(AATG)
Pentanucleotide	(AGAAA)(AGAAA)
Hexanucleotide	(AGTACA)(AGTACA)

STR sequences not only vary in the length of the repeat unit and the number of repeats but also in the rigor with which they conform to an incremental repeat pattern. STRs are often divided into several categories based on the repeat pattern (Table 2.3). Simple repeats contain units of identical length and sequence, compound repeats comprise two or more adjacent simple repeats, and complex repeats may contain several repeat

blocks of variable unit length as well as variable intervening sequences (Urquhart *et al.,* 1994). Complex hypervariable repeats also exist with numerous non-consensus alleles that differ in both size and sequence and are therefore challenging to genotype reproducibly (Urquhart *et al.,* 1993; Gill *et al.,* 1994). This last category of STR markers is not as commonly used in forensic DNA typing due to difficulties with allele nomenclature and measurement of variability between laboratories, although two commercial kits now include the complex hypervariable STR locus SE33 (Urquhart *et al.,*1993; Applied Biosystems 2002; Promega Corporation 2002).

Table 2.3. Categories for STR markers.

Category	Example of Repeat Structure	Example (13 CODIS Loci)
Simple repeats	(GATA)(GATA)(GATA)	TPOX, CSF1PO, D5S818, D13S317,D16S539
Simple repeats with non-consensus alleles	(GATA)(GAT-)(GATA)	TH01, D18S51, D7S820
Compound Repeats	(GATA)(GATA)(GACA)	VWA, FGA, D3S1358, D8S1179
Complex repeats	(GATA)(GACA)(CA)(CATA)	D21S11

Not all alleles for a STR locus contain complete repeat units. Even simple repeats can contain non-consensus alleles that fall in between alleles with full repeat units. Microvariants are alleles that contain incomplete repeat units. Perhaps the most common example of a microvariant is the allele 9.3 at the TH01 locus, which contains 9 tetranucleotide repeats and one incomplete repeat of three nucleotides because the seventh repeat is missing a single adenine out of the normal AATG repeat unit (Puers *et al.,* 1993).

2.4. Characteristics of STRs used in DNA typing

For human identification purposes, it is important to have DNA markers that exhibit the highest possible variation or a number of less polymorphic markers that can be combined in order to obtain the ability to discriminate between samples. The small size of STR alleles (~100–400 bp) compared to

minisatellite VNTR alleles (~400–1000 bp) makes the STR markers better candidates for use in forensic applications where degraded DNA is common. PCR amplification of degraded DNA samples can be better accomplished with smaller product sizes because with older technologies such as RFLP, these severely degraded DNA samples would have been very difficult if not impossible to analyze. High molecular weight DNA molecules need to be present in the sample in order to detect large VNTR alleles (e.g., 20,000 bp) with RFLP techniques. Allelic dropout of larger alleles in minisatellite markers caused by preferential amplification of the smaller allele is also a significant problem with minisatellites (Tully *et al.,* 1993). Furthermore, single base resolution of DNA fragments can be obtained more easily with sizes below 500 bp using denaturing polyacrylamide gel electrophoresis. Thus, for both biology and technology reasons the smaller STRs are advantageous compared to the larger minisatellite VNTRs.

Among the various types of STR systems, tetranucleotide repeats have become more popular than di- or trinucleotides. Penta- and hexanucleotide repeats are less common in the human genome (Bacher *et al.,* 1999).

A biological phenomenon known as 'stutter' results when STR alleles are PCR amplified. Stutter products are amplicons that are typically one or more repeat units less in size than the true allele and arise during PCR because of strand slippage (Walsh *et al.,* 1996). Depending on the STR locus, stutter products can be as large as 15% or more of the allele product quantity with tetranucleotide repeats. With di- and trinucleotides, the stutter percentage can be much higher (30% or more) making it difficult to interpret sample mixtures. In addition, the four base spread in alleles with tetranucleotides makes closely spaced heterozygotes easier to resolve with size-based electrophoretic separations compared to alleles that could be two or three bases different in size with dinucleotides and trinucleotide markers, respectively. Thus, to summarize, the advantages of using tetranucleotide

STR loci in DNA typing over VNTR minisatellites or di- and trinucleotide repeat STRs include:

- A narrow allele size range that permits multiplexing;
- A narrow allele size range that reduces allelic dropout from preferential amplification of smaller alleles;
- The capability of generating small PCR product sizes that benefit recovery of information from degraded DNA specimens; and
- Reduced stutter product formation compared to dinucleotide repeats that benefit the interpretation of sample mixtures.

In the past decade, a number of tetranucleotide STRs have been explored for application to human identification. The types of STR markers that have been sought out have included short STRs for typing degraded DNA materials, STRs with low stuttering characteristics for analyzing mixtures, and male-specific Y chromosome STRs for analyzing male-female mixtures from sexual crimes (Carracedo and Lareu, 1998). The selection criteria for candidate STR loci in human identification applications include the following characteristics (Gill *et al.,* 1996, Carracedo and Lareu, 1998):

- High discriminating power, usually > 0.9, with observed heterozygosity > 70%;
- Separate chromosomal locations to ensure that closely linked loci are not chosen;
- Robustness and reproducibility of results when multiplexed with other markers;
- Low stutter characteristics;
- Low mutation rate; and
- Predicted length of alleles that fall in the range of 90–500 bp with smaller sizes better suited for analysis of degraded DNA samples.

In order to take advantage of the product rule, STR markers used in forensic DNA typing are typically chosen from separate chromosomes to avoid any problems with linkage between the markers.

2.5. STR locus momenclature

The nomenclature for DNA markers is fairly straightforward. If a marker is part of a gene or falls within a gene, the gene name is used in the designation. For example, the short tandem repeat (STR) marker TH01 is from the human **t**yrosine **h**ydroxylase gene located on chromosome 11. The ‘01’ portion of TH01 comes from the fact that the repeat region in question is located within intron 1 of the tyrosine hydroxylase gene. Sometimes the prefix HUM- is included at the beginning of a locus name to indicate that it is from the human genome. Thus, the STR locus TH01 would be correctly listed as HUMTH01 (Butler, 2005).

DNA markers that fall outside of gene regions may be designated by their chromosomal position. The STR loci D5S818 and DYS19 are examples of markers that are not found within gene regions. In these cases, the ‘**D**’ stands for DNA. The next character refers to the chromosome number, 5 for chromosome 5 and **Y** for the Y chromosome. The ‘**S**’ refers to the fact that the DNA marker is a single copy sequence. The final number indicates the order in which the marker was discovered and categorized for a particular chromosome. Sequential numbers are used to give uniqueness to each identified DNA marker. Thus, for the DNA marker D3S1358:

D3S1358	
D	DNA
3	Chromosome 3
S	Single Copy Sequence
1358	1358th locus described on chromosome 3

2.6. Advantages of STRs in DNA typing

PCR-based STRs have several advantages over conventional Southern blotting techniques of the larger VNTRs. Determination of discrete alleles allows results to be compared easily between laboratories because actual genotypes for specific markers are generated. In addition, smaller quantities of DNA (0.1-1 ng), including degraded DNA, may be typed using STRs. Due to the smaller size of the repeat arrays, degraded DNA is less of a problem with STR-markers than for the larger VNTR arrays. Thus, the quantity and integrity of the DNA sample is less of an issue with PCR-based typing methods than with conventional RFLP methods. Other advantages are that the method is much faster than RFLP analysis, which can take up to 6-8 weeks using radioactive probes and one week with fluorescent probes. Results of STR typing can be obtained within a day and are also better suited for high-throughput analysis and automation. The discriminating power of multiplexed STR markers is very high compared to RFLP methods, and with proper use of the population databases, estimates of match probability approach 1 in one billion. Additionally typing of multiple loci can be accomplished in a single multiplex reaction (Hochmeister *et al.,* 1991; Lins *et al.,* 1996).

2.7. DNA typing studies in different populations

Several Investigators have employed STR multiplex PCR using various compositions of STRs. Table 2.4 summarizes the results of some of those studies.

Table 2.4. The most polymorphic alleles described for various populations.

Location	Most Polymorphic alleles	Reference
Morocco population	D21S11,D18S51,Penta E, Penta D,FGA	Louai et al., (2003)
Syrian population	D21S11,D18S51,Penta E, Penta D,FGA	Louai et al., (2003)
Jewish population	D12S391, FGA, D8S1179, D21S11, D18S51	Anto`nia et al.,(2002)
Germans population	VWA, D21S11, D18S51, FGA.	Barbara et al.,(2003)
Chinese population	D8S1179, VWA, FGA	Yanmei et al., (2002)
Tuscany (Central Italy) Population	FGA , D7S820, D8S1179, VWA , D21S11,D18S51	Ugo et al., (2002)
Egyptian population	D7S820, VWA, D8S1179, D13S317	Michael et al.,(1999)
Saudi Arabian population	F13A0, THO1,vWA	Sudhir et al.,(1999)
Korean population	vWA , FESFPS , F13A01,CSF1PO, TH01	Myun et al., (2001)
Neuquen population - Argentina	PENTA E , D21S11 , PENTA D,D18S51 , FGA	Ulises et al.,(2003)

2.8. DNA typing

DNA typing has gone through three major stages of technological advancement. Loosely speaking, these were the multilocus, single-locus, and STR stages.

2.8.1. Multilocus (Minisatellite) testing

The first method developed for the DNA profiling was termed multilocus testing. Alec Jeffreys and co-workers pioneered this approach; he discovered tandemly repeated DNA sequences or “minisatellites” that were variable between different individuals. Minisatellites were visualized by digesting the DNA with restriction enzymes to cut it into fragments of differing lengths that ranged between 1 and 20 kb in size. These fragments included relatively long minisatellites and their flanking regions of DNA. The fragments were electrophoresed on a gel that separated them by size and then visualized using multilocus probes that hybridized many minisatellite loci at once. This

made a pattern that looked a bit like a bar code. Jeffreys and co-workers claimed that the probes produced patterns that were specific to an individual and coined the term DNA fingerprints. There are various acronyms used to describe minisatellites, including RFLP, a generic term that refers to any kind of DNA polymorphism that is based on length differences between restriction sites (Jeffreys *et al.,* 1985).

2.8.2. Single-locus probes

The next step in the development of DNA profiling work utilized the same RFLP technology; however, the probes used to visualize the product were altered to target only one locus at a time. These systems were referred to as single-locus probes (in the United Kingdom and New Zealand) and as VNTR systems in the U.S. As expected, most individuals showed one or two alleles at a locus. The use of the PCR for such loci was reported, (Jeffreys *et al.,* 1988) but not extensively implemented into casework. In the early 1990s, statistical interpretation of VNTR profiles was exclusively by use of the product rule (Evett *et al.,* 1990). The alleles were characterized by a measurement of their molecular weight. Each allele was an integer multiple of the repeat sequence plus the flanking DNA. However, the repeat length was small relative to the total fragment length and hence alleles separated by only one or a few repeat units could not be differentiated reliably using the agarose gel-based technology of the time. Although the underlying distribution was discrete, the measurement of molecular weight was essentially continuous. While VNTR loci are still in use in some laboratories, they have largely been replaced by STR loci.

2.8.3. Short tandem repeat (STR) testing

In the mid-1990s, the technology changed to encompass the use of PCR of STR loci (Edwards *et al.,* 1991). STR markers are ideal for population-level studies for a number of reasons. First, they are randomly distributed throughout the genome, commonly occurring in noncoding regions, and are typically selectively neutral. Second, microsatellite loci are often

hypervariable within populations and show much higher mutation rates than other nuclear regions (Weber and Wong, 1993). Variation seen at microsatellite loci arises from differences among alleles in the number of times the basic motif is repeated, with new alleles probably being generated through polymerase slippage and slipped-strand mispairing during DNA replication (Levinson and Gutman, 1987; Kruglyak *et al.,* 1998; Toth *et al*., 2000), which results in the addition or loss of one or a small number of repeats. Third, microsatellite alleles show codominant inheritance, making them relatively easy to score directly. Finally, and most important for field applications, microsatellite marker genotyping requires only miniscule amounts of template DNA, since it is based on PCR (Mullis and Faloona, 1987). Sufficient DNA for microsatellite analyses can be extracted from small pieces of tissue or minute quantities of blood, as well as from single shed hairs or from the epithelial cells sloughed off in urine, feces, or saliva. Once a microsatellite locus has been identified in the genome, oligonucleotide primers can be designed from the DNA sequences upstream and downstream of the microsatellite to amplify that fragment of the genome by PCR. Then microsatellite marker variation can be assayed directly by electrophoresis and visualization of these PCR products in denaturing polyacrylamide gels; because alleles vary in the number of repeats of the microsatellite motif, heterozygous individuals will show two PCR product bands, while homozygotes will only display a single band. Smaller alleles were also more suitable for the PCR reaction as it is more efficient with low molecular weight DNA fragments.

The introduction of PCR-based STR analysis was the major innovation that expanded the utility of DNA profiling. In summary:

- The development of PCR improved the sensitivity of the analysis.
- The time taken per analysis was reduced to less than 24 hours.
- The cost effectiveness of the method was greatly improved due to a reduction in the labor required.

- The shorter STR loci allow the analysis of degraded DNA samples, which are frequently encountered by forensic scientists. This was because these short segments of DNA stood a higher chance of being intact after degradation.
- STR loci can be multiplexed together using several different STR primer pairs to amplify several loci in one reaction. Multiplexing was further facilitated by the development of dye-labeled primers that could be analyzed on automated DNA sequencers.
- The collection of data was automated, and the analysis of data was partially automated.

2.9. Selection of STR loci for DNA typing

STR loci consist of repeated segments of two to eight bases. These are termed dimeric, trimeric, and so on. Dimeric loci are not used for forensic applications because excessive slippage during amplification (termed stuttering) results in a large number of spurious bands that are difficult to interpret. Trimeric, tetrameric, and pentameric loci are less prone to this problem. Early multiplexes were based on a few simple STR loci.

2.9.1. First generation multiplex STR typing

The first multiplex STR typing system is known as Quadruplex and based on 4 STR loci (Kimpton *et al.,* 1993). These loci include TH01, vWA, HUMFES and HUMF13A1. The four Quad loci started to replace the SLP typing system.

2.9.2. Second generation of multiplex STR typing

In 1996, a six-locus STR system combined with the Amelogenin sex test (Sullivan *et al.,* 1993) was introduced (Sparkes *et al.,* 1996). This system, known as the “second-generation multiplex” (SGM) and consisted of two of the Quad minisatellite loci (TH01 and vWA) combined with the STR microsatellite loci D8S1179, D18S51, D21S11 and FGA (Mills *et al.,* 1992) which are highly polymorphic.

2.9.3. Expansion of the SGM-system

In 1999, four additional loci (D3S1358, D16S539, D2S1338 and D19S433) were added to the six SGM-markers, a system now known as AMPFISTR SGM Plus™. A full SGM plus profile could deliver match probabilities of less than one in a million. The SGM plus system was fully accredited for use in forensic cases.

Today both Applied Biosystems and the Promega Corporation have STR kits that address the needs of the DNA typing community and cover a common set of STR loci. Matching probabilities that exceed one in a billion are possible in a single amplification with 1 ng (or less) of DNA sample. Just as impressive is the fact that results can be obtained today in only a few hours compared to the weeks that RFLP methods took just a few years ago. The technology has evolved quickly in the late 1990s for more sensitive, rapid, and accurate measurements of STR alleles. At the same time, the number of STRs that can be simultaneously amplified has increased from three or four with silver-stained systems to over 15 STRs using multiple-color fluorescent tags. A list of commercially available STR multiplexes and when they were released as products is shown in Table 2.5.

Table 2.5. Information on commercially available STR multiplexes.

Name	Source	Release Date	STR Loci Included
TH01, TPOX, CSF1PO monoplexes (silver stain)	Promega	Feb-93	TH01, TPOX, CSF1PO
PowerPlex 1.1	Promega	Sep-98	CSF1PO, TPOX, TH01, VWA, D16S539, D13S317, D7S820, D5S818
AmpFlSTR® Profiler Plus™	ABI	Dec-97	D3S1358, VWA, FGA, Amelogenin, D8S1179,D21S11, D18S51, D5S818, D13S317, D7S820
AmpFlSTR® COfiler™	ABI	May-98	D3S1358, D16S539, Amelogenin, TH01,TPOX, CSF1PO, D7S820
AmpFlSTR® SGM Plus™	ABI	Feb-99	D3S1358, VWA, D16S539, D2S1338, Amelogenin,D8S1179, D21S11, D18S51, D19S433, TH01, FGA
PowerPlex® 16	Promega	May-00	CSF1PO, FGA, TPOX, TH01, VWA, D3S1358,D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, Penta D, Penta E, Amelogenin
AmpFlSTR® Identifiler™	ABI	Jul-01	CSF1PO, FGA, TPOX, TH01, VWA, D3S1358,D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, D2S1338, D19S433, Amelogenin
PowerPlex® ES	Promega	Mar-02	FGA, TH01, VWA, D3S1358, D8S1179, D18S51,D21S11, SE33, Amelogenin
AmpFlSTR® Sefiler™	ABI	Sep-02	FGA, TH01, VWA, D3S1358, D8S1179, D16S539,D18S51, D21S11, D2S1338, D19S433, SE33, Amelogenin

There are literally hundred thousands of STR systems, which have been mapped throughout the human genome; examples are given in Figure 2.3, where chromosomal distribution of commonly used STR markers is shown. An additional sex determination marker is added, Amelogenin. This marker is based on a 6 base pair deletion in a region on the X-chromosome. The Y chromosome does not have this deletion, so its allele is 6 base pairs longer than that of the X-chromosome. A female profile therefore has only one peak for the Amelogenin marker and a male has two, because of the 6 base pair difference between his X and Y chromosome (Sullivan *et al.*, 1993).

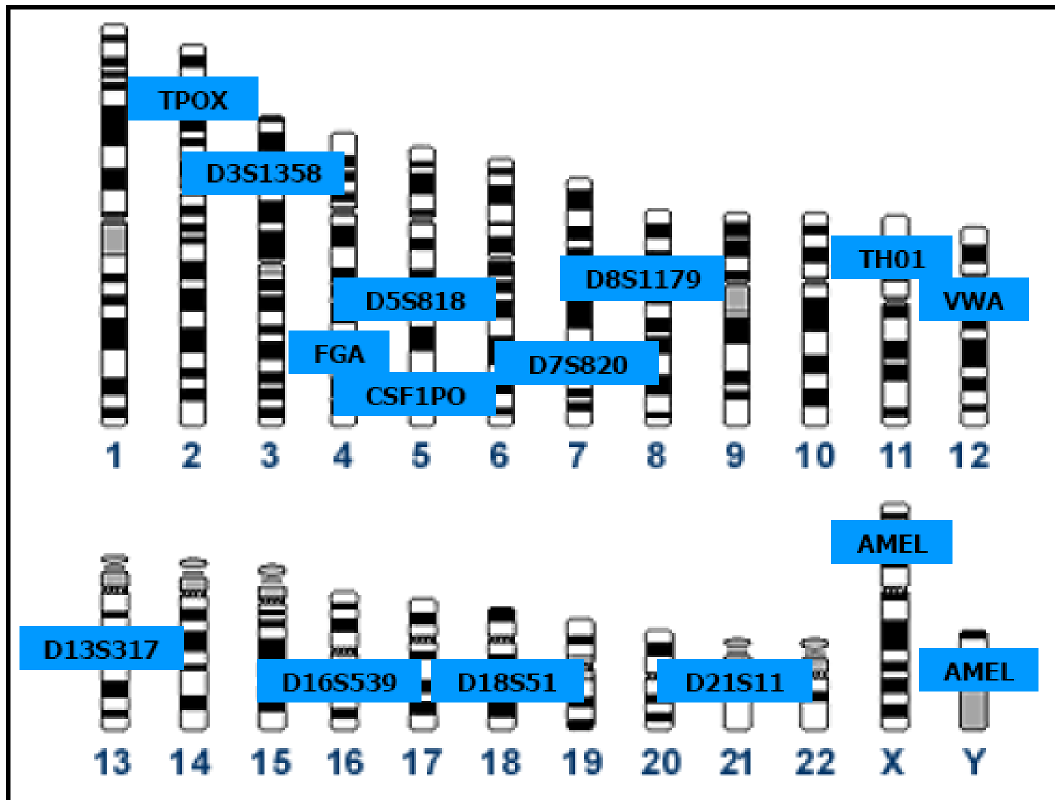

Figure 2.3. Commonly used STR markers and their distributions over the human genome.(Adapted from:http://www.cstl.nist.gov/biotech/strbase/chrom.htm)

2.10. Allelic Ladder

An allelic ladder is an artificial mixture of the common alleles present in the human population for a particular STR marker (Sajantila *et al.,* 1992). They are generated with the same primers as tested samples and thus provide a reference DNA size for each allele included in the ladder. Allelic ladders have been shown to be important for accurate genotype determinations (Smith, 1995). These allelic ladders serve as a standard like a measuring stick for each STR locus. They are necessary to adjust for different sizing measurements obtained from different instruments and conditions used by various laboratories.

Allelic ladders are constructed by combining genomic DNA or locus-specific PCR products from multiple individuals in a population, which possess alleles that are representative of the variation for the particular STR

marker (Sajantila *et al.,* 1992, Baechtel *et al.,* 1993). The samples are then co-amplified to produce an artificial sample containing the common alleles for the STR marker (Figure 2.4). Allele quantities are balanced by adjusting the input amount of each component so that the alleles are fairly equally represented in the ladder.

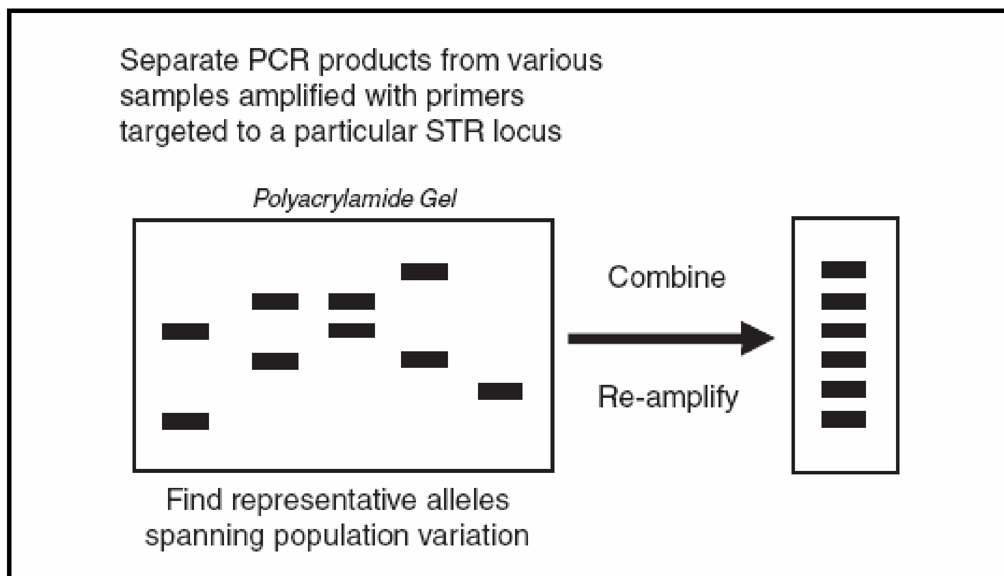

Figure 2.4. Principle of allelic ladder formation. STR alleles from a number of samples are separated on a polyacrylamide gel and compared to one another. Samples representing the common alleles for the locus are combined and re-amplified to generate an allelic ladder. Each allele in the allelic ladder is sequenced since it serves as the reference material for STR genotyping. Allelic ladders are included in commercially available STR kits (Adapted from Butler, 2005).

The loci included in the PowerPlex® 16 System (Tables 2.6) have been selected because they satisfy the needs of several major standardization bodies throughout the world. For example, the United States FBI has selected 13 STR core loci for typing prior to searching or submitting samples in CODIS, the U.S. national database of convicted offender profiles. The PowerPlex® 16 System amplifies all CODIS core loci in a single reaction.

Table 2.6. The PowerPlex® 16 System locus-specific information.

STR Locus	Chromosomal Location	GenBank® Locus and Locus Definition	Repeat Sequence 5′→ 3′
Penta E	15q	NA	AAAGA
D18S51	18q21.3	HUMUT574	AGAA (23)
D21S11	21q11–21q21	HUMD21LOC	TCTA Complex (23)
TH01	11p15.5	HUMTH01, Human Tyrosine Hydroxylase Gene	AATG (23)
D3S1358	3p	NA	TCTA Complex
FGA	4q28	HUMFIBRA, Human Fibrinogen Alpha Chain Gene	TTTC Complex (23)
TPOX	2p23–2pter	HUMTPOX, Human Thyroid Peroxidase Gene	AATG
D8S1179	8q	NA	TCTA Complex (23)
vWA	12p12–pter	HUMVWFA31, Human von Willebrand Factor Gene	TCTA Complex (23)
Amelogenin	Xp22.1–22.3 and Y	HUMAMEL, Human Y Chromosomal Gene for Amelogenin-like Protein	NA
Penta D	21q	NA	AAAGA
CSF1PO	5q33.3–34	HUMCSF1PO, Human c-fms Proto-Oncogene for CSF-1 Receptor Gene	AGAT
D16S539	16q24–qter	NA	GATA
D7S820	7q11.21–22	NA	GATA
D13S317	13q22–q31	NA	TATC
D5S818	5q23.3–32	NA	AGAT

In addition to 13 STR core loci the system contains two low-stutter, highly polymorphic pentanucleotide repeat loci, Penta E and Penta D. These additional loci add significantly to the discrimination power of the system, making the power of exclusion sufficient to resolve between humans. In addition, the extremely low level of stutter seen with Penta E and Penta D makes them ideal loci for evaluation of DNA mixtures often encountered in forensic casework. Finally, the Amelogenin locus is included in the PowerPlex® 16 System to allow gender identification of each sample.

2.11. Factors that complicate STRs interpretation

During PCR amplification of STR alleles, a number of artifacts can arise that may interfere with the clear interpretation and genotyping of the alleles present in the DNA template.

2.11.1. Repeat slippage

Repeat slippage (Levinson and Gutman, 1987; Schlotterer and Tautz, 1992), sometimes called stutter products or shadow bands (Figure 2.5), is due to the loss of a repeat unit during DNA amplification, somatic variation within the DNA in sample material, or both. The amount of this artifact observed depends primarily on the locus and the DNA sequence being amplified. Stutter product formation may be reduced when using STR markers with longer repeat units as Penta D and Penta E (Bacher and Schumm, 1998), and DNA polymerases with faster processivity (Walsh *et al.* 1996).

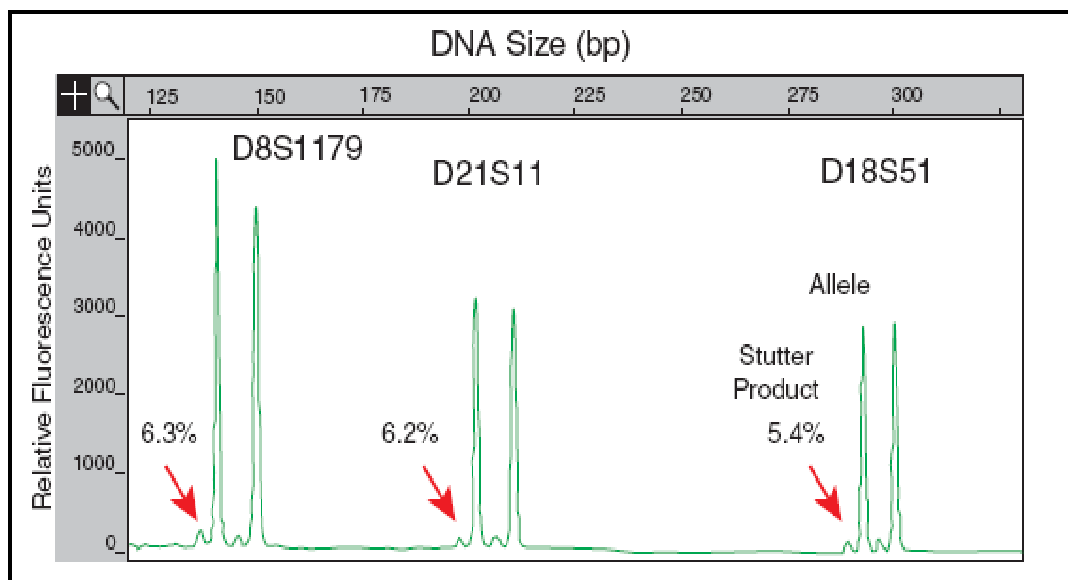

Figure 2.5. STR alleles shown with stutter products (indicated by arrows). Only the stutter percentage for the first allele from each locus is noted (Adapted from Butler, 2005).

2.11.2. Non-template addition

The non-template addition (Smith *et al.,* 1995; Magnuson *et al.,* 1996) occurs when Taq DNA polymerase adds a nucleotide, generally adenine, to the 3´ ends of amplified DNA fragments in a template-independent manner. The efficiency with which this occurs varies with different primer sequences. Thus, an artifact band one base shorter than expected (i.e., missing the terminal addition) is sometimes seen. Non-template addition problem can be reduced by adding the final extension step 60°C or 72°C for 30–45 minutes (Kimpton *et al.,* 1993; Applied Biosystems, 1999) to the amplification protocols to provide conditions for essentially complete terminal nucleotide addition.

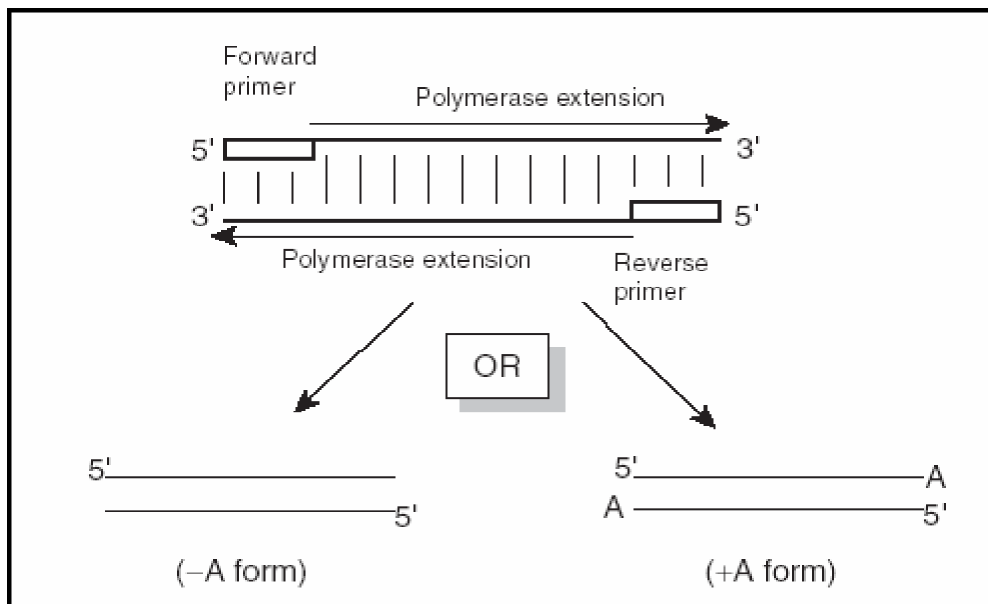

Figure 2.6. Schematic diagram of non-template nucleotide addition (Adapted from Butler, 2005).

2.11.3. Microvariant alleles

The presence of microvariant alleles (alleles differing from one another by lengths other than the repeat length) complicates interpretation and assignment of alleles. There appears to be a correlation between a high degree of polymorphism, a tendency for microvariants and increased mutation rate (Moller *et al.,* 1994; Brinkmann *et al.,* 1995).Thus, FGA and D21S11 display numerous, relatively common microvariants.

Chapter Three

MATERIALS AND METHODES

3.1. Study population

In this descriptive study, 125 apparently healthy unrelated individuals from whom fresh blood samples were collected by members of the local hospitals distributed throughout Gaza Strip. The samples collected were treated as “anonymous unlinked” and single blind for the researcher. This implies that limited information with regard to individuals, who provided the samples, was sent with the samples to the molecular genetics laboratory. No information existed that could have linked a specific sample to a specific individual. This ensures the objectivity and the integrity of the results obtained.

All samples were processed and analyzed in the genetics laboratory at the Islamic University of Gaza. Samples were analyzed according to standard operating procedures. All possible measures were taken to prevent contamination.

3.2. Ethical considerations

Informed consent was obtained from all participants and the researcher explained the objective of the study to them.

3.3. Materials

3.3.1. Chemicals and reagents

Chemicals and reagents used in this study are shown in Table 3.1. All chemicals were of analytical and molecular grade.

Table 3.1. Chemicals and reagents used in the present study.

Reagent	Supplier
PowerPlex® 16 PCR System	Promega, Madison, USA
PowerPlex® Matrix Standards 3100/3130	Promega, Madison, USA
DNA IQ. Extraction System	Promega, Madison, USA
AmpliTaq Gold® DNA polymerase	Applied Biosystems, Foster City, USA
Tissue and Hair Extraction Kit	Promega, Madison, USA
3130 POP-4™ Polymer	Applied Biosystems, Foster City, USA
Capillary Array 36 cm	Applied Biosystems, Foster City, USA
Hi-Di™ Formamide	Applied Biosystems, Foster City, USA
10X Genetic Analyzer Buffer with EDTA	Applied Biosystems, Foster City, USA
Absolute Ethanol Alcohol	Sigma, USA
Isopropyl Alcohol	Sigma, USA
Mineral Oil	Promega, Madison, USA
Nuclease-Free Water	Sigma, USA

3.3.2. Disposables

The major disposables used in this study are listed in Table 3.2.

Table 3.2. Major disposables used in the present study

Item	Supplier
MicroAmp® Optical 96-Well Plate and Septa for the 3130 Genetic Analyzer	Applied Biosystems, Foster City, USA
Optical Adhesive Cover	Applied Biosystems, Foster City, USA
Plate Septa, Retainer and Base	Applied Biosystems, Foster City, USA
Aerosol-Resistant Pipette Filter Tips	Labcon, USA
Micro Tubes, 1.5ml Capacity	Labcon, USA
Micro Tubes, 0.2ml Capacity	Labcon, USA
Disposables Gloves Free Powder	Weihai Sun Genius - China
Vacutainer 5 mL K3-EDTA Tube	Weihai Sun Genius - China

3.3.3. Equipments

All experiments of this research were done at the Islamic university of Gaza Genetics Laboratory. The major equipment that were used are listed in Table 3.3.

Table 3.3. Major equipments used in the present study.

Instrument	Manufacturer
Applied Biosystems 3130 Genetic Analyzer	Applied Biosystems, Foster City, USA
GeneMapper® ID Software, Version 4.0	Applied Biosystems, Foster City, USA
TProfessional Basic Thermocycler	Biometra - Germany
Spectrophotometer	NanoDrop. USA
Safety Cabinet	Heraeus , Germany
Bench top refrigerated microcentrifuge with plate rotor	Centurion Scientific LTD, UK
Freezer -70C	Napco, Czech Republic
Freezer -20C	LG, Korea
Digital reading Micropipette	Socorex - Switzerland
Water Bath	PSelecta - Spain
Vortex Mixer	LW Scientific - USA
MagneSphere® Technology Magnetic Separation Stand	Promega, Madison, USA
Personal Computer	DELL, USA

3.4. Methods

3.4.1. Blood Collection

Approximately 2.5 ml venous blood samples were collected in K3- EDTA tubes. The collected blood samples were received in the same day for processing.

3.4.2. DNA typing analysis

3.4.2.1. DNA extraction

DNA was extracted from blood samples using DNA IQ. Extraction System Kit Extraction was done according to the manufacturer protocol, briefly:

1. Ten-µl blood samples were mixed by vortex, and incubated with 202 µl prepared lyses buffer and resin at room temperature for 5 minutes.
2. The solution was vortexed and placed in the magnetic stand (MagneSphere® Technology Magnetic Separation Stand), the bound DNA-Resin was settled in the bottom of the tube (pellet), and the unbound solution (supernatant) was discarded without disturbing the resin.
3. Bound DNA was washed with 100 µl prepared lyses buffer. Subsequently, the DNA was washed three times with 100 µl 1X wash buffer and air-dried at room temperature.
4. The DNA was eluded by addition of 100-µl elution buffer, and incubated at 65°C for 5 minutes.
5. The tubes were then vortexed and placed in the magnetic stand.
6. Finally, the DNA containing solution was eluted and stored at -20°C until PCR analysis.
7. The quality of the extracted DNA was monitored on ethidium bromide stained 2% agarose gels and the quantity of DNA was measured by spectrophotometer at 260 nm.

3.4.2.2. PCR – amplification of microsatellite loci

Co-amplification of fifteen autosomal STR loci namely Penta E, D18S51, D21S11, TH01, D3S1358, FGA, TPOX, D8S1179, vWA, Penta D, CSF1PO, D16S539, D7S820, D13S317 and D5S818 and a gender marker (Amelogenin) from isolated genomic DNA was performed.

The PowerPlex® 16 System used for the co-amplification, included PowerPlex® 16 10X Primer Pair Mix, Gold STAR 10X Buffer, Positive Control

9947A DNA (10ng/µl), PowerPlex® 16 Allelic Ladder Mix, Internal Lane Standard (ILS) 600. AmpliTaq Gold® DNA polymerase is provided separately.

The PowerPlex®16 amplification Reaction Mix (2.5µl Gold STAR 10X Buffer, 2.5µl PowerPlex® 16 10X Primer Pair Mix, 0.8µl AmpliTaq Gold® DNA polymerase and 18.2 µl nuclease-free water) was added directly to 1.0µl of the isolated genomic DNA (0.5 - 1ng). The fluorescent dye-labeled and unlabeled primer pairs used were: One primer for each of the Penta E, D18S51, D21S11, TH01 and D3S1358 loci is labeled with fluorescein; one primer for each of the FGA, TPOX, D8S1179, vWA and Amelogenin loci is labeled with carboxy-tetra-methyl-rhodamine; and one primer for each of the Penta D, CSF1PO, D16S539, D7S820, D13S317 and D5S818 loci is labeled with 6-carboxy-4´, 5´- dichloro-2´, 7´-dimethoxy-fluorescein. All sixteen loci were amplified simultaneously in a single tube and analyzed in a single injection.

Amplification was performed in a TProfessional Basic Thermocycler using the following protocol: initial 11-minute incubation at 95°C for the activation of AmpliTaq Gold DNA polymerase, then denaturation at 96°C for 2 minutes followed by 10 cycles of 94°C for 1 minute, primer annealing at 60°C for 1 minute and extension at 77°C for 1.5 min and then 22 cycles 90°C for 1 minute for denaturation, primer annealing at 60°C for 1 minute and extension at 77°C for 1.5 min. The last cycle was followed by a final extension step at 60°C for 30 to 45 minutes to promote complete non-template 3'- nucleotide addition.

Following PCR amplification of DNA samples, the resulting fluorescently labeled STR alleles were separated, sized, and genotyped as described below.

3.4.2.3. Capillary electrophoresis

Amplified alleles sizing was performed using the ABI 3130 Genetic Analyzer according to the PowerPlex® 16 System PCR Amplification protocol. Internal Lane Standard (ILS) 600 that is provided in the kit was included with every sample to allow automatic sizing of alleles. Protocols for sample preparation and run conditions for the ABI 3130 Genetic Analyzer were provided in the PowerPlex®16 System PCR Amplification User's Manual (Promega, Madison, USA).

First, the ABI 3130 Genetic Analyzer was calibrated; Spatial and Spectral Calibrations were done and passed with Q value 0.97499 and condition 6.001 for spectral calibration using PowerPlex® Matrix Standards 3100/3130 (Figure 3.1). The purpose of spectral calibration is to generate the matrix that correct the overlapping of fluorescent emission spectra of the dyes.

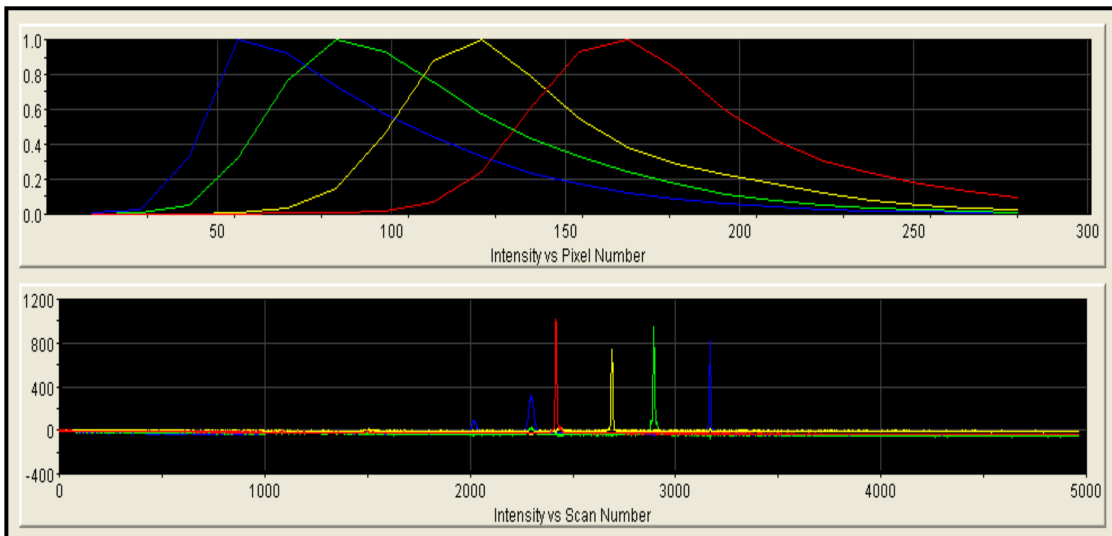

Figure 3.1. A diagram illustrating spectral calibration in ABI 3130 Genetic Analyzer.

Prior to loading the samples onto the ABI 3130 Genetic Analyzer, the loading cocktail was prepared by combining and mixing the internal lane standard and Hi-Di Formamide as follows:

[(0.5µl ILS 600) × (# injections)] + [(9.5µl Hi-Di. formamide) × (# injections)]
Then the cocktail was mixed for 10–15 seconds with a vortex mixer and 10µl of the cocktail were then loaded into MicroAmp® optical 96-well plate, 1µl of amplified sample was added into each well of interest and 1µl of the allelic ladder mix was added into each run. Then the plate was covered with septa, centrifuged to remove air bubbles and loaded into thermal cycler for denaturating. After denaturing the samples for 3 min at 95°C the samples were snap cooled to 0-4°C and immediately loaded into the instrument autosampler.

Internal Lane Standard 600 was used for the PowerPlex®16 amplification kit. ILS 600 contains 22 DNA fragments of 60, 80, 100,120, 140, 160, 180, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 550 and 600 bases in length (Figure 3.2). Each fragment is labeled with carboxy- X-rhodamine (CXR).

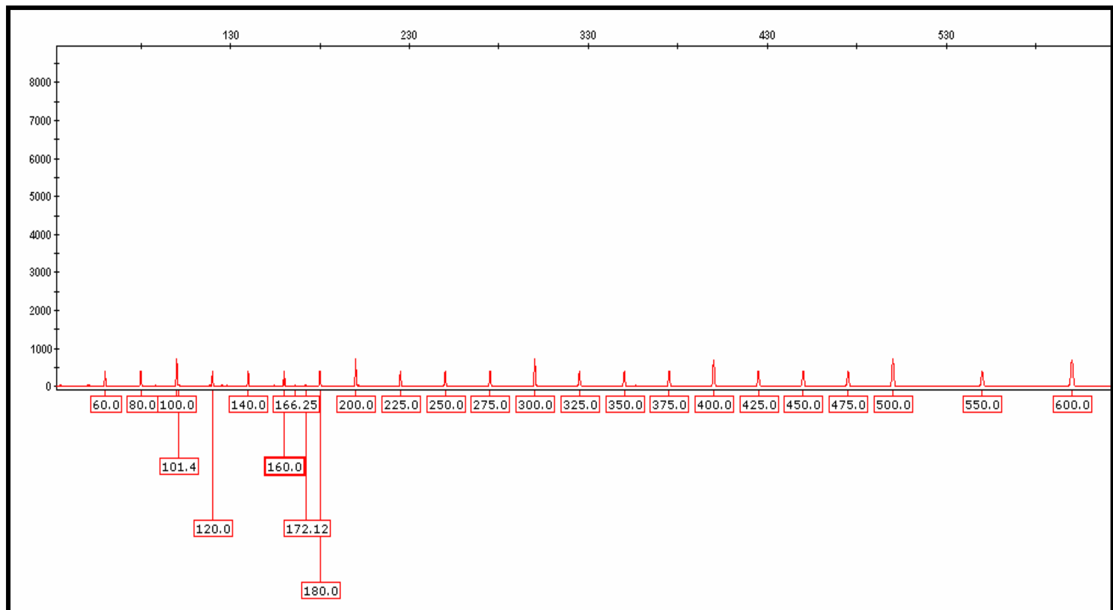

Figure 3.2. An electropherogram showing the fragments of the Internal Lane Standard 600.

Electrophoresis was carried out using a 36 cm capillary array. The separation media was Performance Optimized Polymer 4 (POP-4) and the electrophoresis buffer was a 1:10 dilution of the Genetic Analyzer buffer with

EDTA. Electrophoresis was carried out using the default module "HIDFragmentAnalysis36_POP4"; some modifications were changed and other parameters were confirmed such as the injection time changed to 5 seconds, injection voltage to 3kV and the run time was lengthened to 2,000 seconds. DNA samples are loaded into the capillary with electrokinetic injection. Electrophoresis of fragments took place for approximately 47 to 57 minutes. Data were collected using Data collection Software.

3.4.2.4. Genotype determinations

Samples data collected from the ABI 3130 Genetic Analyzer are represented in the form of peaks that correspond to the various STR alleles amplified from the DNA sample. These peaks are present at various locations in a sample's electropherogram and usually plotted as fluorescent signal intensity versus time passing the detector. The various dye colors were separated and the peaks representing DNA fragments were identified and associated with the appropriate color. The DNA fragments are then sized by comparison to an internal sizing standard.

GeneMapper® Software version 4.0 was used to automatically analyze the collected data; furthermore, fragment sizes were checked manually to avoid any false calling of alleles. Allele sizes were estimated by means of the Local Southern method. The Local Southern Algorithm assigns the allele sizes in relation to the internal size standards; this method uses the size of two peaks on either side of the unknown one being measured in order to make the calculations (Elder and Southern, 1983). The sizing of DNA fragments with internal standards is performed as illustrated in Figure 3.3.

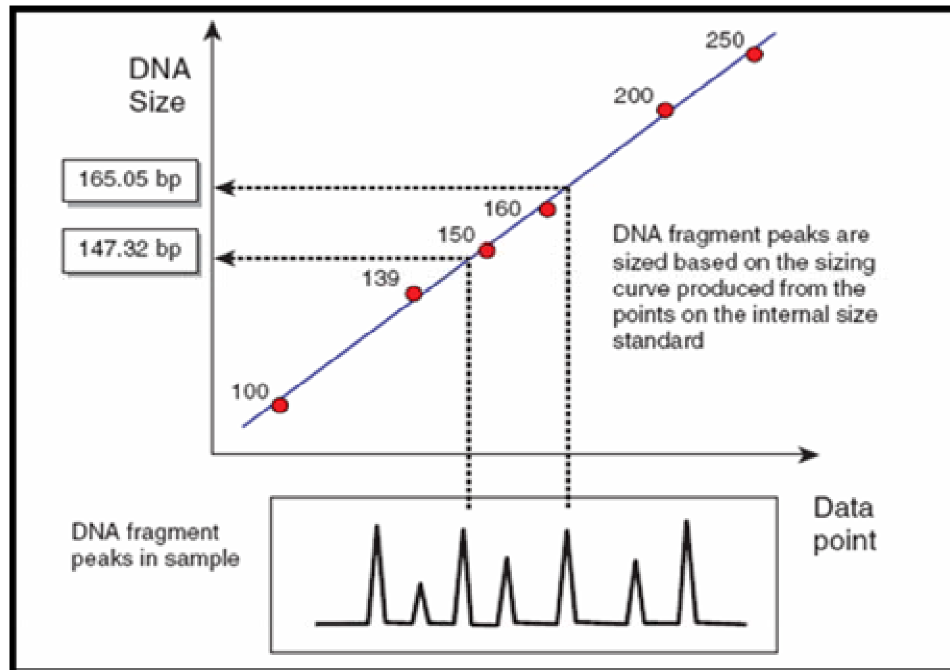

Figure 3.3. Sizing of DNA fragment peaks using internal size standards. Finally, the polymerase chain reaction (PCR) product sizes for the questioned sample are correlated to an allelic ladder that has been sized in a similar fashion with internal standards (Adapted from Butler, 2005).

The allelic ladder is the standard to which STR alleles are compared to obtain the sample genotype. Table 3.4 summarizes the PowerPlex® 16 System Allelic Ladder information. STR genotyping was performed by comparison of the size of a sample's alleles to size of alleles in allelic ladders for the same loci being tested in the sample.

Table 3.4. The PowerPlex® 16 System Allelic Ladder Information.

STR Locus	Label	Size Range of Allelic Ladder	Repeat Numbers of Allelic Ladder
Penta E	FL	379 - 474 bp	5-24
D18S51	FL	290 - 366 bp	8-10, 10.2, 11-13, 13.2, 14-27
D21S11	FL	203 - 259 bp	24, 24.2, 25, 25.2, 26.28, 28.2, 29, 29.2, 30, 30.2, 31, 31.2, 32, 32.2, 33, 33.2, 34, 34.2, 35, 35.2, 36 -38
TH01	FL	156 - 195 bp	4 - 9, 9.3, 10 , 11, 13.3
D3S1358	FL	115 - 147 bp	12-20
FGA	TMR	322 - 444 bp	16-18, 18.2, 19, 19.2, 20, 20.2, 21, 21.2, 22, 22.2, 23, 23.2, 24, 24.2, 25, 25.2, 26-30, 31.2, 43.2, 44.2, 45.2, 46.2
TPOX	TMR	262 - 290 bp	6-13
D8S1179	TMR	203 - 247 bp	7-18
vWA	TMR	123 - 171 bp	10-22
Amelogenin	TMR	106 - 112 bp	X,Y
Penta D	JOE	376 - 449 bp	2.2, 3.2, 5, 7-17
CSF1PO	JOE	321 - 357 bp	6-15
D16S539	JOE	264 - 304 bp	5, 8-15
D7S820	JOE	215 - 247 bp	6-14
D13S317	JOE	176 - 208 bp	7-15
D5S818	JOE	119 - 155 bp	7-16

3.4.2.5. Controls for genotyping

Negative and positive controls were run concurrently with each sample batch, the Positive Control 9947A DNA (10ng/µl) which is provided by Promega was used, and Table 3.5 illustrates the genotype of 9947A DNA positive control. Nuclease-free water instead of template DNA was used as negative control to check for any DNA contamination.

Table 3.5 The PowerPlex® 16 System allele determinations in 9947A standard (Positive Control) DNA Template.

Standard DNA Templates	
STR Locus	**9947A**
Penta E	13,12
D18S51	19,5
D21S11	30,30
TH01	9.3,8
D3S1358	15,14
FGA	24,23
TPOX	8,8
D8S1179	13,13
vWA	18,17
Amelogenin	X,X
Penta D	12,12
CSF1PO	12,10
D16S539	12,11
D7S820	11,10
D13S317	11,11
D5S818	11,11

3.5. Statistical analysis of population samples

The PowerStatsV1.2 (Promega, Madison, USA) was used to calculate the observed heterozygosity (Ho), power of discrimination (PD), probability of exclusion (PE), and polymorphism information content (PIC). Arlequin software program (Schneider *et al.,* 2000) was used to conduct the exact test of population differentiation. In addition, Arlequin software program was used for the expected heterozygosity (He), Hardy Weinberg Equilibrium (HWE) and linkage equilibrium tests as well as for the *F*-statistics. Where test results with *P*-values less than 0.05 were observed the Bonferroni correction had to be applied to the data. The Bonferroni procedure (Weir, 1996) adjusted the rejection level for the smallest *P*-value at an overall level of α = 5 % to 0.05/x, where x is equal to the number of tests conducted on the data. The Ho and He values were calculated by means of the same software program.

Following are the formulas used to compute various parameters for population data analyses:

3.5.1. Expected Heterozygosity

Edwards *et al.* (1992) described the following formula for calculating an unbiased estimate of the expected heterozygosity:

$$H = \frac{n}{(n-1)}\left[1 - \sum_{j=1}^{k}\left(n_j/n\right)^2\right] = \frac{n}{(n-1)}\left[1 - \sum_{j=1}^{k}\left(p_j\right)^2\right]$$

where *n1, n2, …, nk* are the allele counts of *K* alleles at a locus in a sample of *n* genes drawn from the population and *pj* is the allele frequency.

3.5.2. Match Probability

The probability of a match at locus *l*, *PMl*, was first described from genotype data. Fisher (1951) gave

$$PM_l = \sum_i \tilde{G}_{il}^2$$

where $\tilde{G}_i$ is the sample frequency of the *i*th genotype at locus *l*. Jones suggests setting

$$PM_l = \frac{\sum_{i=1}^{n} \tilde{G}_{il}^2 - 1/N_l}{1 - 1/N_l} \approx \sum_{i=1}^{n} \tilde{G}_{il}^2$$

where the first part of this equation is for a sample of size *Nl* at locus *l*. (Jones, 1972).

3.5.3. Power of Discrimination

Brenner and Morris (1990) described the following formula for calculating the power of discrimination:

$$1 - PM$$

3.5.4. Polymorphism Information Content

The PIC was also calculated using marker allelic frequencies using following equation:

$$PIC = 1 - \sum_{i=1}^{n} p_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2p_i^2 p_j^2$$

where n is the number of alleles and pi is the allele probability of the ith allele (Botstein *et al*., 1980).

3.5.5. Power of Exclusion

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

where H = heterozygosity.

3.5.6. Paternity Index

$$\text{Paternity Index (PI)} = \frac{H+h}{2h} = \frac{(1-h)+h}{2h} = \frac{1}{2h} = \frac{1}{2\sum_{i=1}^{n} P_i^2}$$

where Pi is frequency of ith allele in a population of n samples; h = homozygosity; H = heterozygosity.

Chapter Four

RESULTS

Out of the 135 original samples, 10 could not be amplified and were therefore not included for further analysis. The database was thus compiled of the DNA profiles of 125 reference blood samples. The samples were typed with the PowerPlex®16 PCR Amplification kit and 15 STR loci were analyzed.

4.1. Genotyping of DNA samples

After the samples have been collected, DNA extracted and PCR amplified they were genotyped for the 15 STR loci of interest. The genotyping information were then converted into allele frequencies by counting the number of times each allele was observed. Table 4.1 shows an example of allele counting for the locus TH01. The allele frequency is also provided in the table.

Table 4.1. Genotype array and allele count for the STR locus TH01 from unrelated Gazan samples (n = 125 or 250 chromosomes measured).

Genotype Array	6	7	8	9	9.3	10	Allele	Allele Count	Observed Frequency
6	7	-	-	-	-	-	6	64	25.60%
7	11	10	-	-	-	-	7	62	24.80%
8	9	7	1	-	-	-	8	27	10.80%
9	24	19	8	6	-	-	9	70	28.00%
9.3	5	2	-	6	1	-	9.3	18	7.20%
10	1	3	1	1	3	-	10	9	3.60%

The observed alleles for TH01 locus (ranging from 6–10 repeats) are listed across the top and down the left side. Listed at the intersection of the rows and columns are the numbers of observed genotypes. For example, starting in the top left hand corner, the genotype 6/6 was seen seven times in the set of the 125 individuals examined while the genotype 8/10 was seen one time, but the genotype 10/10 was not recorded in the individuals included in this study.

On the right side of Table 4.1, the numbers of observed alleles are counted by summing the columns and rows containing the alleles of interest. Thus, the number of chromosomes containing allele 6 is equal to 64 (from 7 + 11 + 9 + 24 + 5 + 1) for the column containing allele 6 plus seven for the row with the 6/6 genotype. The number of 6/6 genotypes is counted twice since both chromosomes contain an allele 6 at the TH01 marker. The frequency for allele 6 was determined by dividing 64 by the total number of chromosomes, which are 250 since there are two chromosomes for each of the 125 individuals typed. Note that there are 62 alleles 7, 27 alleles 8, 70 alleles 9, 18 alleles 9.3 and nine alleles 10 observed in this study for locus TH01. Genotype array and allele count for the rest of the fifteen STR were calculated in the same manner as TH01 locus. Figure 4.1 represents the electropherogram of the allelic ladder for the TH01 locus. This ladder, which consists of all the known alleles for TH01, was used to verify the alleles in the samples. Figure 4.2 illustrates one representative genotype (8/10) observed in one sample for locus TH01.

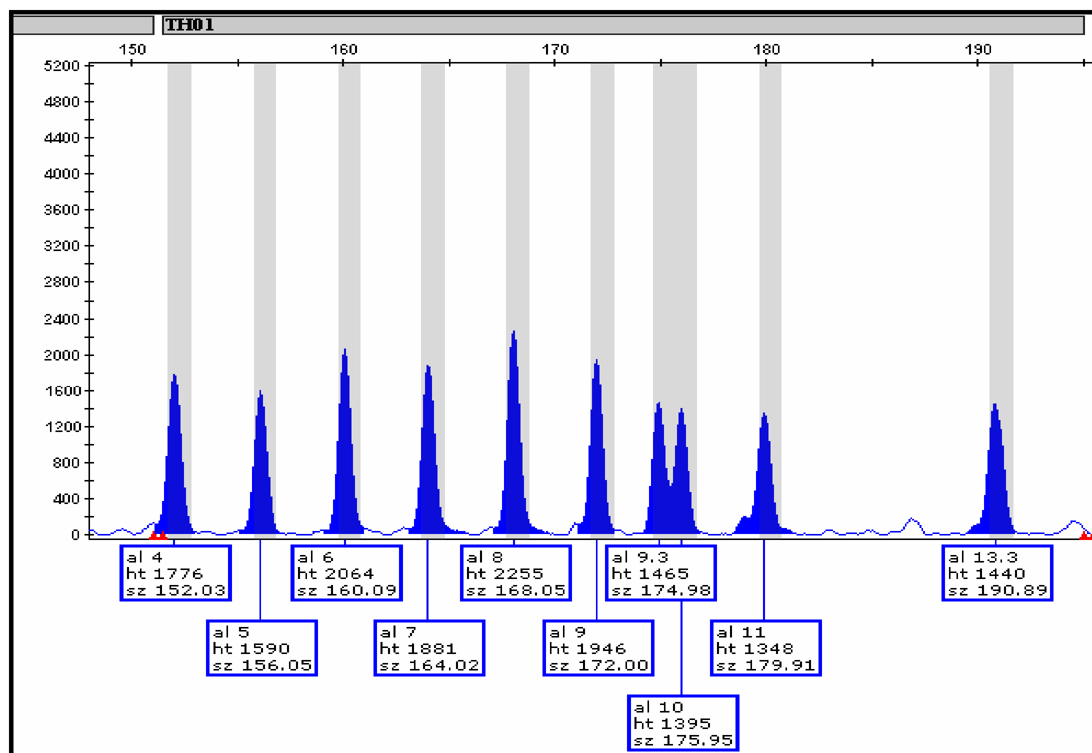

Figure 4.1. An electropherogram illustrating PowerPlex®16 Allelic ladder obtained with GeneMapper version 4.0 indicating possible alleles on locus TH01 (al = allele; ht = peak height; sz = allele size).

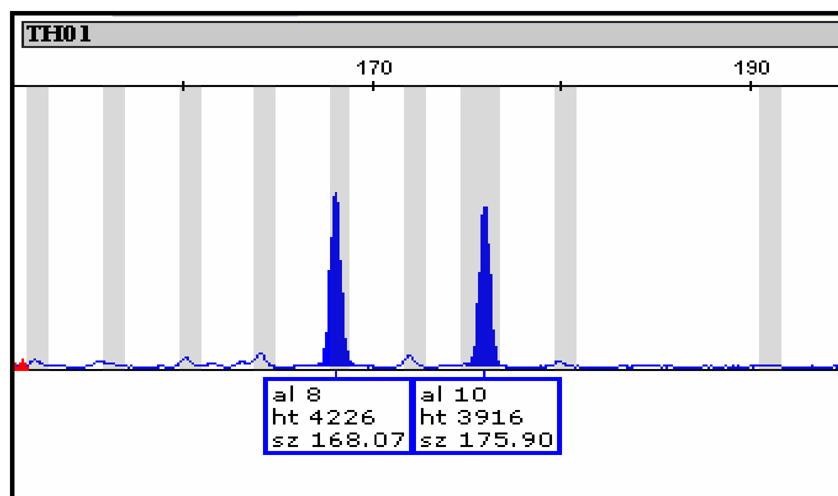

Figure 4.2. An electropherogram indicating the genotyping result for one sample of locus TH01 obtained with PowerPlex®16 amplification and GeneMapper version 4.0.

4.2. Allele frequencies and genotypes

The distributions of observed alleles and genotypes for the fifteen STR loci examined are shown in Appendix 1 and Appendix 2, respectively.

4.3. Allele frequency of amelogenin locus

Along with the fifteen STR loci for identification, Amelogenin locus was used for sex determination (Table 4.2). Amplification of Amelogenin generates different length products from the X and Y-chromosomes. Figure 4.3 shows the possible peaks for both genotypes X/X and X/Y represented as one and two peaks, respectively, in the GeneMapper electropherogram.

Table 4.2. Amelogenin Locus size according to the PowerPlex®16 Allelic Ladder among Gazan population.

Locus	Allele	Base Pair Size	Genotype	Observed Genotype	Frequency
Amelogenin	[Y] Chromosome	109.38 ± 0.5	X/Y	73	58.40%
	[X] Chromosome	103.43 ± 0.5	X/X	52	41.60%

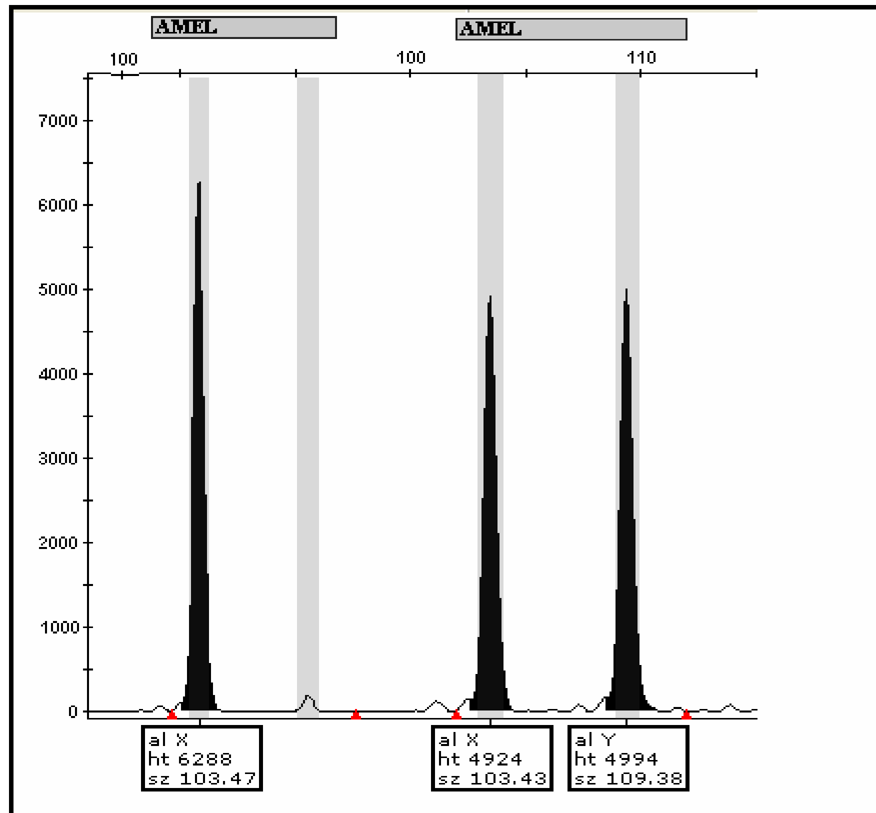

Figure 4.3. An electropherogram showing Amelogenin locus and its two possible genotypes obtained with PowerPlex®16 amplification and GeneMapper version 4.0.

4.4. Heterozygosity and Homozygosity calculations

The results obtained from the tests for heterozygosity (Observed and Expected) and homozygosity for the Gazan population is shown in Table 4.3. The number of homozygotes (h) plus the number of heterozygotes (H) equals 100% of the samples tested. Thus, since h +H= 1, then H = 1 − h and h = 1 − H. A higher heterozygosity means that more allele diversity exists and therefore there is less chance of a random sample matching.

4.5. Test for Inbreeding Coefficient (F_{IS})

F-statistics is a measure of the correlation of alleles within individuals and is related to inbreeding coefficients (NRC II, 1996). *F*-statistics describes the amount of inbreeding-like effects within subpopulations (F_{ST} or θ), among populations (F_{IS} or f), and within the entire population (F_{IT} or F). In the present study, the F_{IS} value for Gazan population has been calculated as –0.024 with non-significant *P*- value of 0.605.

4.6. Test for Hardy- Weinberg Equilibrium

Checking for HWE is performed by taking the observed allele frequencies and calculating the expected genotype frequencies based on the allele frequencies. If the observed genotype frequencies are close to the expected genotype frequencies calculated from the observed allele frequencies, then the population is in Hardy–Weinberg equilibrium and allele combinations are likely to be independent of one another. The results obtained from the tests for HWE for the Gazan population is shown in Table 4.3.

Table 4.3. Observed homozygosity, observed heterozygosity and expected heterozygosity for the different fifteen STR loci in Gazan population.

Locus	Observed Homozygosity	Observed Heterozygosity	Expected Heterozygosity	HWE P-Value
D3S1358	23.20%	76.80%	75.49%	0.14115
TH01	20.00%	80.00%	77.81%	0.15738
D21S11	20.00%	80.00%	84.05%	0.42375
D18S51	8.80%	91.20%	87.34%	0.45463
Penta E	19.20%	80.80%	90.71%	0.00452
D5S818	23.20%	76.80%	76.06%	0.00940
D13S317	32.80%	67.20%	76.61%	0.01900
D7S820	22.40%	77.60%	79.99%	0.79448
D16S539	24.80%	75.20%	79.72%	0.08059
CSF1PO	25.60%	74.40%	71.88%	0.17325
Penta D	18.40%	81.60%	86.44%	0.19407
vWA	28.80%	71.20%	80.86%	0.00625
D8S1179	14.40%	85.60%	84.90%	0.86473
TPOX	34.40%	65.60%	69.83%	0.47437
FGA	23.20%	76.80%	87.42%	0.28489

4.7. Test for Disequilibrium of Linkage

Linkage disequilibrium describes the degree of association between particular alleles at different genetic loci. Linkage disequilibrium can result when polymorphic genetic loci are situated close to each other on the same chromosome. The D21S11 and Penta D loci and the D5S818 and CSF1PO loci are located on chromosomes 21 and 5, respectively. Therefore, we examined the Linkage Disequilibrium between these two pairs of loci, by using two different ways to test for the presence of pairwise linkage

disequilibrium between loci. Exact Test of linkage disequilibrium is an extension of Fisher exact probability test on contingency tables (Slatkin, 1994), and Likelihood ratio test of linkage disequilibrium (Slatkin and Excoffier, 1996). In the present study, all the loci located on the same chromosome are in linkage equilibrium.

4.8. Variability of alleles and genotypes

In terms of the total number of alleles and genotypes, if n alleles exist for a genetic marker, there are n homozygous genotypes, $n(n-1)/2$ heterozygous genotypes and $n(n+1)/2$ total genotypes are possible. Table 4.4 includes the number of possible genotypes for the fifteen STRs loci based on the number of observed alleles that are listed in the same table. For example, D3S1358 locus with seven possible alleles would exhibit seven homozygous possibilities plus $[7\times (7-1)]/2$ heterozygous possibilities or $7 + 21 = 28$ total genotypes. A combination of 10 loci with 10 alleles in each locus would have over 2.5×10^{17} possible genotypes ($55 \times 55 \times 55 \times \ldots$). Based on the total number of alleles seen across all the fifteen STR loci, there are about 8.3167×10^{24} possible genotypes.

Table 4.4. Observed numbers of alleles and number of possible genotypes for Gazan population.

Locus	Number of Observed Alleles	Number of Possible Genotypes
	n	**n(n+1)/2**
D3S1358	7	28
TH01	6	21
D21S11	13	91
D18S51	13	91
Penta E	16	136
D5S818	7	28
D13S317	9	45
D7S820	7	28
D16S539	7	28
CSF1PO	8	36
Penta D	12	78
vWA	7	28
D8S1179	10	55
TPOX	7	28
FGA	14	105
	143 Observed alleles	**8.3167×10^{24}**

4.9. Most common genotypes frequencies

Another measure to reflect the usefulness of a particular set of DNA markers is to examine the frequencies of the most common genotypes, which would therefore be the least powerful in terms of being able to differentiate between two unrelated individuals (Edwards *et al.* 1992). The theoretically most common type can be calculated by considering a sample type that is heterozygous at all loci possessing the two most common alleles at each locus (Foreman and Evett, 2001). In Table 4.5 frequencies from the two most common alleles at each of the fifteen loci were used to estimate a theoretical most common STR profile, which occurs with a combined frequency of approximately one person in 9.09 x 10^{13} (90 Trillion).

Table 4.5. Calculations for the theoretically most common genotype frequencies and profile frequency based on the two most common alleles found in a Gazan population.

Locus	Allele 1	Allele 2	Allele 1 Frequency [p]	Allele 2 Frequency [q]		Most Common Genotype Frequency
Penta E	12	13	0.184	0.14	2pq	0.052
D18S51	12	16	0.188	0.148	2pq	0.056
D21S11	29	30	0.232	0.248	2pq	0.115
TH01	6	9	0.256	0.28	2pq	0.143
D3S1358	16	17	0.304	0.308	2pq	0.187
FGA	22	23	0.164	0.16	2pq	0.052
TPOX	8	11	0.436	0.276	2pq	0.241
D8S1179	13	14	0.220	0.220	2pq	0.097
vWA	16	17	0.212	0.292	2pq	0.124
Penta D	9	11	0.204	0.160	2pq	0.065
CSF1PO	10	11	0.304	0.312	2pq	0.190
D16S539	11	12	0.308	0.224	2pq	0.138
D7S820	10	11	0.300	0.224	2pq	0.134
D13S317	11	12	0.256	0.364	2pq	0.186
D5S818	11	12	0.276	0.348	2pq	0.192
				Profile Frequency 1.10 x 10^{-14}		
				1 in … 9.09 x 10^{13} (90 Trillion)		

4.10. DNA profile frequency estimates

The DNA profile frequency is estimated by first considering the genotype frequency for each locus and then multiplying the frequencies across all calculated loci. In Table 4.6, the DNA profile frequency for five STR loci was calculated using the allele frequencies determined in this study. The DNA profile in question contains the following alleles: 10 and 15 at D8S1179 (heterozygous), 14 and 18 at vWA (heterozygous), 9 at TPOX (homozygous), 2.2 and 13 at Penta D (heterozygous) and 30.2 at D21S11 (homozygous).

Table 4.6 An example for calculating the DNA profile frequency (or combined match probability) using alleles from five STR loci.

DNA Profile		**Determined Allele Frequency**			**Genotype Frequency for Locus**	
Locus	**Allele**	**Times Allele Observed**	**Size of Database**	**Frequency**	**Formula [HWE]**	**Number**
D8S1179	10 15	24 38	250	p = 0.096 q = 0.152	2pq	0.029
vWA	14 18	17 44	250	p = 0.068 q = 0.176	2pq	0.024
TPOX	9 9	46	250	p = 0.184	P^2	0.034
Penta D	2.2 13	8 36	250	p = 0.032 q = 0.144	2pq	0.009
D21S11	30.2 30.2	11	250	p = 0.044	P^2	0.002
				Profile Frequency = 4.25952 x 10^{-10} **Combined Frequency = 2.35 x 10^{9}**		

In our population sample of 250 alleles (125 Gazan individuals), allele 10 for D8S1179 was observed 24 times (24/250) or 9.6% of the time (Table 4.6). The frequency of allele 10 for D8S1179 can therefore be recorded as p= 0.096. In other words, we can assume that there is a 9.6% chance that any particular D8S1179 allele selected at random from an unrelated individual will be allele 10. In the same manner, the chance for observing an allele 15 is q =

0.152 since this allele was seen 38 times in the 250 alleles determined. If the individual with the 10/15 D8S1179 genotype received these alleles at random from each of his parents, then the chance to receive allele 10 from his mother and allele 15 from his father is pq and to receive the 15 from his mother and the 10 from his father is another pq. With either combination possible, the probability to be 10/15 by chance is pq + pq or 2pq.

Substituting the frequency values of p=0.096 and q=0.152 into the formula 2pq (2 × 0.096 × 0.152), results in an estimated genotype frequency of 0.029 or, in other words approximately 3% of people from a Gazan population are expected to have 10/15 genotype at the D8S1179 locus. With the TPOX locus, a homozygous allele 9 was observed (Table 4.6); the combined probability of inheriting allele 9 from both parents is pp or p^2. Since allele 9 was observed 46 times of 250 allele measurements in Gazan population, p= 0.184 and $p^2 = 0.034$.

Multiplication of allele frequencies in the form of the product rule benefits demonstrating that a DNA profile is extremely rare. If the product rule cannot be used, then the power of a genetic test is vastly reduced. Therefore, the assumption that loci and alleles are inherited independently is an important one so that match probabilities can be multiplied. The combination of locus-specific match probabilities is referred to as the product rule. In other words, the match probability for the STR locus D8S1179 can be combined with additional STR loci such as vWA, TPOX, Penta D and D21S11 to decrease the odds of a random match to an unrelated individual (Table 4.6).

In the case of a Gazan population, the chance of a person having the combined genotypes of 10/15 at D8S1179 and 9/9 at TPOX is 0.029 × 0.034 or 0.0986%. Similar calculations for the rest of the loci were determined. The combined profile frequency for these five loci thus becomes 4.26×10^{-10} or about one person in 2.35×10^{9} (2 Billion).

4.11. Forensic efficiency parameters

Statistical analysis is used to interpret DNA results for genetic identity. In order to determine the significance of a match, it is necessary to support DNA typing results with statistical analysis. These analyses assign a value to the results obtained and enable easier resolution of forensic or paternity cases. Across all loci, the values for the matching probability, power of discrimination, power of exclusion, polymorphism information content and typical paternity index for the fifteen STR loci of the Gazan population were determined and are indicated in Table 4.7.

Table 4.7 Forensic efficiency parameters of the fifteen STR loci in 125 unrelated Gazan individuals.

Locus	Matching Probability	Power of Discrimination	Power of Exclusion	Polymorphism Information Content	Typical Paternity Index
D3S1358	0.114	0.886	0.541	0.71	2.16
TH01	0.098	0.902	0.599	0.74	2.50
D21S11	0.049	0.951	0.599	0.82	2.50
D18S51	0.038	0.962	0.820	0.86	5.68
Penta E	0.024	0.976	0.614	0.90	2.60
D5S818	0.118	0.882	0.541	0.72	2.16
D13S317	0.092	0.908	0.686	0.73	1.52
D7S820	0.073	0.927	0.555	0.77	2.23
D16S539	0.077	0.923	0.513	0.76	2.02
CSF1PO	0.147	0.853	0.500	0.66	1.95
Penta D	0.039	0.961	0.629	0.85	2.72
vWA	0.071	0.929	0.447	0.78	1.74
D8S1179	0.047	0.953	0.707	0.83	3.47
TPOX	0.142	0.858	0.364	0.65	1.45
FGA	0.033	0.967	0.541	0.86	2.16
Combined Matching Probability (CMP)				**2.73×10^{-18}**	
Combined Discrimination Power (CDP)				**0.999999986**	
Combined Exclusion Probability (CEP)				**0.999998595**	

4.12. An example of parentage testing

DNA data play an important role in parentage testing. The determination of parentage is made based on whether or not alleles are shared between

the child and the alleged father (AF) when a number of genetic markers are examined. Table 4.8 is an example for paternity test, which was done during this study. The obligate paternal alleles for the child in this example are shown.

4.12.1. Paternity Index

The paternity index (PI) compares the likelihood that a genetic marker (allele) that the alleged father (AF) passed to the child to the probability that a randomly selected unrelated man of similar ethnic background could pass the allele to the child. This is presented in the formula X/Y, where X is the chance that the AF could transmit the obligate allele and Y is the chance that some other man of the same race could have transmitted the allele. X is assigned the value of 1 if the AF is homozygous for the allele of interest and 0.5 if the AF is heterozygous. The potential of a randomly selected man to pass the obligate gene is determined by using a database, which lists the frequency distribution of individual alleles within a given genetic system.

4.12.2. Combined Paternity Index

The combined paternity index (CPI) is determined by multiplying the individual PIs for each locus tested. The CPI is an odds ratio that indicates how many times more likely it is that the alleged father is the biological father than a randomly selected unrelated man of similar ethnic background.

4.12.3. Probability of Paternity

Probability of paternity (POP) is necessary to use the Baysian theorem. This is a formula that tests the hypothesis that the accused is the biological father of the child. For example, a POP of 99% reflects a 99% probability that the hypothesis is correct and a 1% probability that it is not. The CPI is used in the Bayes formula along with another variable called a prior probability (PP). Testing labs typically use a value of 0.5 for the PP assuming this is a neutral, unbiased value. The Bayes' Theorem formula is CPI / CPI + (1 - PP) x 100.

Table 4.8. An example for parentage test.

Locus	Daughter Profile	Father Profile	Mother Profile	Obligate Father Allele	Frequency of Obligate Allele	Likelihood Ratio Formula	Paternity Index
D3S1358	17,18	16,17	16,18	17	0.308	0.5	1.62
TH01	8,9	7,9	6,8	9	0.280	0.5	1.79
D21S11	29,29	29,30	28,29	29	0.232	0.5	2.16
D18S51	14,18	12,18	12,14	18	0.092	0.5	5.43
Penta E	7,18	15,18	5,7	18	0.036	0.5	13.89
D5S818	9,9	9,9	9,9	9	0.096	1.0	10.42
D13S317	12,12	12,12	8,12	12	0.364	1.0	2.75
D7S820	10,11	10,11	10,10	11	0.224	0.5	2.23
D16S539	11,12	12,12	10,11	12	0.224	1.0	4.46
CSF1PO	11,11	11,12	11,12	11	0.312	0.5	1.60
Penta D	9,12	12,13	9,9	12	0.132	0.5	3.79
AMEL	X,X	X,Y	X,X	NA	NA	NA	NA
vWA	19,20	16,19	19,20	19	0.096	0.5	5.21
D8S1179	11,15	13,15	11,13	15	0.152	0.5	3.29
TPOX	11,11	9,11	8,11	11	0.280	0.5	1.79
FGA	19,23	19,20	21.2,23	19	0.068	0.5	7.35

Combined Paternity Index (CPI) = 183642859.9 to 1
Probability of Paternity (POP) = 99.999999 %

Chapter Five

DISCUSSION

The vast majority of DNA molecules (over 99.7 %) are the same between people. It is the small fraction of DNA (0.3 % or about ten million nucleotides) that differ between people and make individuals unique. The amount of variation, however, is not constant throughout the human genome. Regions containing repeated nucleotide sequences such as STR sequences are of interest to the forensic scientist because they show a lot of variation, which make them effective for human identification purposes (Butler, 2001). There are however, similarities if we compare the DNA material of two individuals. An individual's DNA profile increases in uniqueness during its assessment as more loci are typed. The population frequency of the profile is estimated from the frequencies of the alleles of the loci investigated with the aim of answering the following question: "What is the expected frequency of individuals with the same profile in the Gazan population?"

Based on the allelic frequencies, several statistical parameters of genetic and forensic efficiency parameters were estimated (Tables 4.3 and 4.7). These included the observed homozygosity and heterozygosity, expected heterozygosity, the polymorphism information content (PIC), the exact testing for Hardy-Weinberg equilibrium (HWE), the power of discrimination (DP) and the power of exclusion (PE).

5.1. Distribution of alleles frequency

The allelic frequencies among the Gazan population at the fifteen STR loci are illustrated in Appendix 1. At the CSF1P0 locus, a total of 8 alleles were observed, with repeat numbers ranging from 5 –14. The repeat numbers 10, 11 and 12 are the three most predominant alleles in the Gazan as well as in the Moroccan, Syrian and Egyptian populations, respectively. However, allele 11 occurred with a higher frequency (a range of 0.312 to

0.349) in the Gazan and the Egyptian populations, respectively, while allele 12 with a higher frequency (a range of 0.381 to 0.339) in the Syrian and the Moroccan populations, respectively. At the D3S1358, D5S818, D7S820, vWA, TPOX and the D16S539, a total of 7 alleles were found among the 13, 14 to 20, 6 to 13 and 8 to 14, respectively. At vWA, the most predominant alleles 16 and 17 are common in Gazan, Syrian, Moroccan, Egyptian and Jewish populations. Alleles 15 and 18 were also found but with moderate frequency in at least some of them. The frequency of the most predominant allele 17 varied between 0.292 in the Gazan population and 0.339 in the Jewish population (Ahmed *et al.,* 2001; Anto`nia et al., 2002; Louai *et al.,* 2003).

Two of the alleles observed in the present study, allele 21.2 at the FGA locus and allele 35.2 at locus D21S11 were not observed in Syrian, Moroccan, Egyptian and Jewish populations. The repeat structure of those alleles are $[TTTC]_3$TTTTTT $[CTTT]_{14}$CTCC$[TTCC]_2$ and $[TCTA]_5$ $[TCTG]_6$ $[TCTA]_3$ TA $[TCTA]_3$ TCA $[TCTA]_2$ TCCA TA $[TCTA]_{15}$ TA TCTA, respectively (Perkin-Elmer Corporation, 1996; Zhou *et al.,* 1997). Presence of certain alleles in particular populations may be used in the future to define the identity of those populations.

5.2. Test for microsatellite alleles variability

In the present study, the total number of alleles observed across the population was found to be 143 alleles with predicated number of 8.3167×10^{24} possible genotypes (Table 4.4), indicating high level of polymorphism of the selected microsatellites. The highest allele frequencies occurred in the loci D13S317 and TPOX (allele 12 for D13S317 locus and allele 8 for TPOX), while the Penta E locus showed the largest number of different alleles (16 alleles). The TH01 locus represented the smallest number of different alleles (6 alleles). The most common alleles at the fifteen loci were allele 17 for D3S1358 locus, allele 9 for TH01 locus, allele 12 for D5S818 locus, allele 12 for D13S317 locus, allele 10 for D7S820 locus, allele

11 for D16S539 locus, allele 11 for CSF1PO locus, allele 17 for vWA locus, alleles 13 and 14 for D8S1179 locus, allele 8 for TPOX locus, allele 30 for D21S11 locus, allele 12 for D18S51 locus, allele 12 for Penta E locus, allele 9 for Penta D locus and allele 22 for FGA locus. The number of alleles and the expected heterozygosities detected in Gazan population are good indicators of the genetic polymorphism within the breed. Generally the number of alleles is highly dependent on the sample size because of the presence of unique alleles in populations, which occur in low frequencies and also because the number of observed alleles tends to increase with increases in population size. The number of alleles scored for each marker is an invaluable indicator of the future usefulness of the marker for genetic screening. Finding the same number of alleles for certain different loci in various populations (e.g., Gazan, Syrian, Moroccan, and Egyptian populations) may indicate common ancestries (Ahmed *et al.,* 2001; Anto`nia et al., 2002; Louai *et al.,* 2003). The frequency and the number of alleles, however, may be an indication for the degree of inbreeding within each population and thus reflects the homogeneity of the population.

5.3. Heterozygosity and Homozygosity

Observed heterozygosity and expected heterozygosity all over the fifteen loci are presented in Table 4.3, and the observed heterozygosity oscillated between studied populations as illustrated in Table 5.1:

Table 5.1. The observed heterozygosity in different populations.

Locus	**Heterozygosity Frequency**				
	Gazan	**Moroccan**	**Syrian**	**Egyptian**	**Jewish**
TPOX	0.656	-	0.664	0.628	-
D18S51	0.912	0.878	-	-	-
D5S818	-	0.693	-	-	-
Penta E	-	-	0.897	-	-
FGA	-	-	-	0.89	-
D13S317	-	-	-	-	0.724
D21S11	-	-	-	-	0.855
vWA	-	-	-	-	-

The observed heterozygosity in a population relies on the number and the frequency of alleles of each locus. Moreover, the distribution of genotypes in a population sample may deviate from HWE expectation in a number of ways. These include the presence of an excess of homozygotes (and a corresponding lack of heterozygotes) or an excess (or deficiency) of one or more classes of heterozygotes or a combination of those states. There are populations with low heterozygosity, lower than 65% in most tested loci. These populations are small, closed, inbred by cultural or geographical factors one of those populations, the Qatari population where the levels of observed heterozygosity (*Ho)* oscillated, between 0.339 for D19S433 and 0.839 for D2S1338. Interestingly, *Ho* is lower than expected heterozygosity *(He) in* almost all the analyzed loci, with the unique exception of D5S818 locus. This fact was particularly conspicuous for the loci that deviated from the HWE expectations; the departures from HWE expectations detected in the Qatari population seem to be the result of excess of homozygotes over heterozygotes, which is likely to be the consequence of the high consanguinity rates reported for this population, which is 46%. (Ana *et al.,* 2006). On the contrary, to Gazan and other populations such as Egypt, Moroccan, Syrian and Jewish populations in which the available information reveals open or panmictic populations with random mating, and that the tested fifteen loci are neutral and they are not affected by natural selection, so all alleles at each locus have the same opportunity to be selected into gametes.

5.4. Inbreeding coefficient

F-statistics was proposed by Wright (1951) to describe the properties of a subdivided population. Parameters F_{IT} and F_{IS} are the correlation between two uniting gametes with respect to the whole population and to gametes of subpopulations, respectively. Parameter F_{ST} is the correlation between random gametes from different individuals within subpopulations with respect to the total population and is a measure of the differentiation of subpopulations. If there is no significant selective advantage of different

alleles. F_{IS} is called the inbreeding coefficient, and can be interpreted as a measure of inbreeding within the population and F_{IT} can be interpreted as total inbreeding estimates and F_{ST} is the estimate for population differentiation. In the present study, the F_{IS} value for Gazan population has been calculated as – 0.024 with non-significant *P*-value of 0.605. The low F_{IS} value, which is very close to zero, indicates low level of inbreeding within the population. The low inbreeding values can be attributed to random mating. The negative value indicates that level of heterozygosity is high within Gazan population. As compared to other populations such as Qatari population which showed an elevated inbreeding coefficient. The high inbreeding coefficient in the Qatari population might be due to their small population size and or the high rate of consanguineous marriages.

5.5. Hardy Weinberg Equilibrium

The Fisher's exact probability test was conducted to detect significant departures from HW equilibrium and *P*-values were generated by the Markov-chain Monte Carlo method. This test was applied to the data by means of the software produced by Guo and Thompson (1992) [Guo and Thompson, 1992]. There were fifteen within-locus tests conducted on the Gazan population (Table 4.3). No significant departure from HWE expectations were observed (a 5% significance level was taken) in the Gazan population for 11 of the 15 STR loci analyzed (p>0.5). The exceptions were the Penta E (*P*-value =0.00452), D5S818 (*P*-value =0.00940), D13S317 (*P*-value =0.01900) and vWA (*P*-value =0.00625) loci but when the Bonferroni procedure was used as a correction for the multiple tests performed on a population sample, none of the previous *P*-values could be considered significant. The Bonferroni correction lowers the significance level for the entire set of n comparisons by dividing n into the alpha value for each comparison. The adjusted significance level becomes $1-(1-\alpha)^{1/n} \approx \alpha/n$ (Weir, 1996; Perneger, 1998). Thus, a set of 15 comparisons would lower the alpha value from 0.05 to 0.0033 (0.05/15) so only *p*-values below 0.0033 would be considered statistically significant rather than the conventional $p < 0.05$. The

Penta E, D5S818, D13S317 and vWA loci *P*-value did not fail at this newly generated *P*-value of 0.0033. The null hypothesis with HWE testing is that the genotype frequencies are identical to the proportions expected based on the allele frequencies. In other words, independence of the alleles at the measured locus exists in the population under examination. Thus, small *p*-values cast doubt on the validity of the null hypothesis. The Bonferroni correction is often applied to lower the *p*-value and thus make results with multiple comparisons not statistically significant. It is important to keep in mind that since the perfect assumptions for HWE cannot exist in real human populations (random mating, no migration, etc.); therefore, the null hypothesis cannot be true by definition. Deviation from HWE in populations can occur for four principal reasons. These are (1) parents might be related leading to inbreeding and a higher than expected number of homozygotes, (2) population substructure, (3) selection because persons with different genotypes might survive and reproduce at different rates, and the most strongly reason depends on (4) Population size. In large populations, quite small differences can be statistically significant (NRCII, 1996).

5.6. Linkage disequilibrium

The 11 loci in this system exist on different chromosomes and they are in linkage equilibrium, so they will segregate independently during meiosis allowing the genotype frequencies to be multiplied. The D21S11 and Penta D loci, and the D5S818 and CSF1PO, were in linkage equilibrium (*P*-value = 0.72624 and *P*-value=0.44596, respectively). Thus, no linkage disequilibrium was found between the two loci on the same chromosome, and they are statistically independent. The absence of linkage disequilibrium between all tested loci ensures the stability of the calculations concerning genotype frequencies.

5.7. Polymorphism information content

The quality of "informativeness" of a polymorphism as a genetic marker was measured by the heterozygosity or the frequency of heterozygosity in a given population or by a polymorphism information content PIC, which

reflects the probability that a given offspring of a parent carrying a rare allele at a locus will allow deduction of the parental genotype at the locus and is determined by summing the mating frequencies multiplied by the probability that an offspring will be informative (Botstein *et al.* 1980). The ideal PIC value is 1; meaning that the polymorphism is informative in any random mating and the ideal heterozygosity is 100%. As shown in Tables 4.3 and 4.7, the PIC of the 15 loci proved to be more than 0.5, which indicates that the 15 loci are of high polymorphism in the investigated Gazan population. The degree of polymorphism of every marker, expressed in heterozygosity and PIC terms (Table 5.2), shows that the TPOX locus is the least polymorphic marker while Penta E is the most polymorphic marker. Similar degree of polymorphism was found in Syrian and Moroccan populations, where the most polymorphic marker was found to be Penta E, and that the least polymorphic one was TPOX. On the contrary, the Egyptian population showed that the FGA locus is the most polymorphic marker and that the TPOX locus, as observed in our population, is the least polymorphic marker. Even in other populations like the Greece and the Japanese, it was found that the previously mentioned loci have a similar degree of polymorphism (Hashiyada *et al.,* 2003; Leda *et al.,* 2006). The polymorphic nature of microsatellites makes them the markers of choice in characterization and genetic diversity studies. The high PIC values of the selected markers confirm their usefulness for genetic polymorphism studies and linkage mapping programs in human as well. The mean heterozygosity observed, expected and mean PIC values across the fifteen loci in Gazan population were 0.77, 0.81 and 0.78, respectively, indicating high gene diversity.

Table 5.2. The degree of polymorphism using PIC and heterozygosity in different populations.

Polymorphism Information Content Frequency												
Population	**Gazan**		**Moroccan**		**Syrian**		**Egyptian**		**Greece**		**Japanese**	
Locus	**PIC**	**He**	**PIC**	**He**	**PIC**	**He**	**PIC**	**Ho**	**PIC**	**He**	**PIC**	**He**
Penta E	0.90	0.91	0.88	0.89	0.89	0.90	-	-	0.89	0.90	0.91	0.91
TPOX	0.65	0.70	0.65	0.70	0.64	0.69	0.61	0.63	0.57	0.62	0.60	0.66
FGA	-	-	-	-	-	-	0.87	0.89	-	-	-	-

5.8. Power of Discrimination

Power of Discrimination (PD) for all the fifteen loci is presented in Table 4.7. The Penta E and Penta D loci, which were a new pentanucleotide repeat markers, FGA and D18S51 showed a high polymorphic parameter value for PIC and PD for all tested populations, which use PowerPlex®16 in their work. The highest PD observed in some populations is presented in Table 5.3.

Table 5.3. The Power of discrimination in some different populations.

Locus	Power of Discrimination Frequency					
	Gazan	Moroccan	Syrian	Egyptian	Qatari	Jewish
Penta E	0.976	0.973	0.974	-	-	-
Penta D	0.961	0.944	0.951	-	-	-
D12S391	-	-	-	-	-	-
D18S51	0.962	0.962	0.963	-	0.970	0.965
FGA	0.967	0.962	0.957	0.973	0.962	0.963
vWA	-	-	-	0.937	-	-

The Penta E and Penta D loci included in the PowerPlex®16 PCR amplification kits were not typed in the Egyptian, Qatari or Jewish populations because they used different kits in their genotyping studies. Power of discrimination values for all tested loci was above 85% for the CSF1PO and TPOX loci, above 88% for the D5S818 and D3S1358 loci and ranged from 90% to 97%, for the rest of the loci. The PD is the probability that two randomly chosen persons would not have matching DNA profiles (Fisher, 1951). Otherwise stated, this is the probability that an innocent person will be excluded as the donor of an evidence sample. The combination of fifteen STRs proved to be extremely discriminating; The Combined Discrimination Power (CDP) for the Palestinian population of Gaza Strip for the corresponding fifteen STR loci used, has been calculated as 0.999999986 (Table 4.7). These figures mean that those loci can be safely used to establish a DNA-based database for Gaza Strip population.

5.9. Random matching probability

The assumption of random mating in a real-life population is required to estimate genotype or profile frequencies from a reference data set of individual allele frequencies, in order to obtain the Matching Probability (MP). The MP is the probability to find two persons with the same DNA profile if you choose them by random in a population. The MP between unrelated individuals for the fifteen STR loci used in this study was estimated for Gazan population (Table 4.7). In Table 4.5 the frequencies from the two most common alleles at each of the fifteen loci were used to estimate a theoretical most common STR profile and the Combined Matching Probability (CMP) was approximately one person in 9.09×10^{13} (90 Trillion), but in Table 4.6 DNA profile frequencies for five STR loci were determined using allele frequencies from this study, and the CMP was approximately one person in 2.35×10^{9} (2 Billion). These figures mean that good forensic technique should give a small match probability when the suspect and offender are the same person and a large match probability when they are different (Balding, 1999). In the present study, the calculated MP ranged from 0.147 for CSF1PO locus to 0.024 for locus Penta E. The CMP value (for all possible genotypes) for the Palestinian population of Gaza Strip for the corresponding fifteen STR loci used in this study was one person in 3.66×10^{17} (300 Quadrillion). Therefore, the probability that two randomly selected samples will match at all fifteen of the loci is 2.73×10^{-18}, meaning that the chance of finding two individuals with the same genotype in the population is almost null. So if you have determined a DNA profile in criminal evidence and the same profile is found in a suspect, it is impossible to exclude this person as responsible for the crime.

5.10. Exclusion probability

In the case of people who are close relatives such as parents and offspring. The expected conditional probabilities between two relatives can

be calculated either by exclusion probability or by more precisely a likelihood ratio [Typical Paternity Index (TPI), Combined Paternity Index (CPI) and Probability of Paternity (POP) or (Bayes' Theorem)] expressing how much more likely it is that we would see the DNA evidence under the hypothesis that they came from people with a specific relationship as opposed to the hypothesis that they came from two ostensibly unrelated people. The Power of Exclusion (PE) can be calculated to express how rare it would be to find a random man who could not be excluded as the biological father of the child (Fisher, 1951; Chakraborty and Stivers 1996; Butler, 2005). In the present study and from the genotyping data, PE for every locus was calculated and presented in Table 4.7. As expected, the power of exclusion was high for all the microsatellites analyzed it ranged from 0.364 (TPOX) to 0.820 (D18S51), with an average of 0.577. PE for different populations are given in Table 5.4. The combined power of exclusion, which is the exclusion probability considering all fifteen loci, was greater than 99.99%, indicating that these loci are appropriate to determine parentage in Gazan population beyond any reasonable doubt.

Table 5.4. The Power of exclusion in different populations.

Locus	Power of Exclusion Frequency				
	Gazan	Moroccan	Syrian	Egyptian	Jewish
TPOX	0.364	-	-	0.326	-
D18S51	0.820	0.751	-	-	-
D5S818	-	0.418	-	-	0.508
CSF1PO	-	-	0.387	-	-
Penta E	-	-	0.788	-	-
FGA	-	-	-	0.775	0.712

5.11. Paternity Testing

In order to test the utility of the system employed in the present study in paternity evaluation we tested three DNA samples taken from daughter, a mother and a father. Based on the genetic testing results shown in Table 4.8 the father can not be excluded as the biological father of the daughter; the probability of paternity (Bayes' Theorem) and the combined paternity index

have been calculated as 99.999 % and 183,642,859.9, respectively, as compared to an untested random man. It means that is 183,642,859.9 more probable that he is the biological father than he is not. This example presents a strong evidence for the utility of these loci not only in paternity testing but also in forensic genetics.

Chapter Six

CONCLUSIONS AND RECOMMENDATIONS

6.1. Conclusions

The main aim of this study was to determine the genetic structure of Gaza Strip population using 15 autosomal STR loci and to evaluate the usefulness of these loci for forensic genetic purposes. The conclusions that have been drawn from the present study include:

- The number of different alleles observed across the population was found to be 143 alleles with a predicated number of 8.3167×10^{24} possible genotypes.

- The highest allele frequencies occurred in the allele 12 for D13S317 locus (36.4%) and allele 8 for TPOX (43.2%), while the Penta E locus showed the largest number of different alleles (16 alleles) and TH01 locus represented the smallest number of different alleles (6 alleles).

- The heterozygosity of the fifteen STR loci ranged from 65.6 % to 91.2 % (mean value 77.4%), the locus with the highest heterozygosity was D18S51, while locus TPOX has the lowest heterozygosity.

- The power of discrimination values for all tested loci was above 85% for the CSF1PO and TPOX loci, above 88% for the D5S818 and D3S1358 loci and ranged from 90% to 97% for TH01, D13S317, D16S539, D7S820, vWA, D21S11, D8S1179, Penta D, D18S51, FGA, and Penta E.

- The combined probability of exclusion, power of discrimination, probability of matching value for all the 15 STR loci were 0.999998595; 0.999999986 and 2.73×10^{-18} ,respectively.

- The F_{IS} value for Gazan population has been calculated as – 0.024 with non-significant P-value of 0.605. The low F_{IS} value indicates low level of inbreeding within the population and that the population is in random mating with a high level of heterozygosity.

- None of the tested loci showed significant departure for HWE, and they are statistically independent.

- The arrangement of the polymorphic markers in a decreasing order was Penta E, FGA, D18S51, Penta D, D8S1179, D21S11, vWA, D7S820, D16S539, D13S317, TH01, D5S818, D3S1358, CSF1PO, and TPOX.

- According to the statistical parameters, the combined analysis of these 15 STR loci is a powerful tool for forensic identification and paternity testing in the Palestinian population of the Gaza Strip.

6.2. Recommendations

- The results of the current study indicate that the examined fifteen STR loci are useful genetic markers for forensic personal identification and paternity testing in the Palestinian population of the Gaza Strip. Consequently, these loci can be used for establishment of a DNA database that will be beneficial for the population in terms of resolving social and moral disputes and will contribute to improvements in the justice system.

- Further studies are needed to characterize the level of polymorphism in X-chromosomes STRs, Y-chromosomes STRs and mtDNA in the Palestinian population of the Gaza Strip.

REFERENCES

Ahmed A., Linacre A., Mohammed A., Vanezis P. and Goodwin W. (2001). STR population data for 10 STR loci including the GenePrint™ and PowerPlex™ 1.2 kits from El-Minia (Central Egypt). Forensic Science International; 117: 233-234

Allen M., Saldeen T. and Gyllensten U. (1995). Allele-specific HLA-DRB1 amplification of forensic evidence samples with mixed genotypes. BioTechniques; 19(3):454-63.

AmpFISTR Blue™ PCR Amplification Kit: User's Manual (1996). The Perkin-Elmer Corporation.

Ana M. P., Miguel A. A., José A. P. and Rene J. H. (2006). Qatari DNA Variation at a Crossroad of Human Migrations. Human Heredity; 61:67–79

Anto`nia P., Carmen T., Gema J., Jose´ A. Castro M. and Miserico`rdia R. (2002). Jewish population genetic data in 20 polymorphic loci. Forensic Science International; 125:52–58

Bacher J.W., Hennes L.F., Gu T., Tereba A., Micka K.A., et al. (1999). Pentanucleotide Repeats: Highly Polymorphic Genetic Markers Displaying Minimal Stutter Artifact. Proceedings of the Ninth International Symposium on Human Identification, pp. 24–37. Madison, Wisconsin: Promega Corporation.

Baechtel F.S., Smerick J.B., Presley K.W. and Budowle B. (1993). Multigenerational amplification of a reference ladder for alleles at locus D1S80. Journal of Forensic Sciences; 38: 1176–1182.

Balding D.J., Greenhalgh M. and Nichols R.A. (1996). Population genetics of STR loci in Caucasians. International Journal of Legal Medicine; 108:300-305.

Barbara R., Uta I. and Michael K. (2003). Population data on the AmpFlSTR SGM plus PCR amplification kit in Germans and Austrians. Forensic Science International; 132: 84–86

Botstein D., White R. L., Skolnick M. and Davis R.W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphism. American Journal of Human Genetics; 32: 324–331.

Brenner C. and Morris J. (1990).Paternity index calculations in single locus hypervariable DNA probes.Proceedings of the International Symposium on Human Identification: 1989 Promega Corporation. Long Beach: California :21-53.

Brinkmann B., Moller A. and Wiegand P. (1995). Structure of new mutations in 2 STR systems. International Journal of Legal Science; 107, 201.3.

Britten R.J., and Kohne D.E. (1968). Repeated Sequences in DNA. Science; 161: 529-540.

Butler J.M. (2001). Forensic DNA Typing: Biology and Technology behind STR Markers. Academic Press.

Butler J.M. (2005). Forensic DNA Typing. Biology, Technology and Genetic of STR markers. Second Edition. Elsevier Academic Press
Butler J.M. (2006). Genetics and genomics of core STR loci used in human identity testing. Journal of Forensic Sciences*;* in press

Butler J.M., Buel E., Crivellente F. and McCord B. (2004). Forensic DNA typing by capillary electrophoresis using the ABI Prism 310 and 3100 genetic analyzers for STR analysis. Electrophoresis; 25; p.1397.

Carracedo A. and Lareu M.V. (1998). Proceedings from the Ninth International Symposium on Human Identification, pp. 89–107. Madison, Wisconsin: Promega Corporation.

Chakraborty R. and Stivers D. N. (1996) Paternity exclusion by DNA markers: effects of paternal mutations. Journal of Forensic Sciences; 41(4): 671-677.

Chambers G.K. and MacAvoy E.S. (2000). Microsatellites: consensus and controversy. Comparative Biochemistry and Physiology. Part B; 126: 455–476.

Collins F. and Galas D. (1993). A new five-year plan for the U.S. Human Genome Project. Science; 262(5130):11.

Collins J.R., Stephens R.M., Gold B., Long B., Dean M. and Burt S.K. (2003). An exhaustive DNA micro-satellite map of the human genome using high performance computing. Genomics; 82: 10-19.

Committee on DNA Technology in forensic Science, DNA Technology in forensic Science, National Academy Press, Washington, D.C.,1992.

Edwards A., Civitello A., Hammond H.A. and Caskey C.T. (1991). DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. American Journal of Human Genetics; 49: 746 756.

Edwards A., Hammond H.A., Jin L., Caskey C.T. and Chakraborty R. (1992) Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics; 12: 241–53.

Elder J.K. and Southern E.M. (1983). Measurement of DNA length by gel electrophoresis II: Comparison of methods for relating mobility to fragment length. Analytical Biochemistry; 128: 227–231.

Ellegren H. (2004). Microsatellites: simple sequences with complex evolution. Nature Reviews Genetics; 5:435-445.

Evett I, Werrett D, Pinchin R. and Gill P. (1990). Bayesian analysis of single locus DNA profiles. Proceedings of the International Symposium on Human Identification (1989). Madison, Wisconsin: Promega Corporation.

Fisher R.A. (1951). Standard calculations for evaluating a blood group system. Heredity. 5: 95-102.

Foreman L.A and Evett I.V (2001). Statistical Analysis to Support Forensic Interpretation of a New Ten-Locus STR Profiling System. International Journal of Legal Medicine; 114:147-155.

Gill P. (2002). Role of short tandem repeat DNA in forensic casework in the UK – past, present, and future perspectives. BioTechniques; 32 (2), 366-385.

Gill P., Ivanov P.L., Kimpton C., Piercy R., Benson N., et al. (1994). Identification of remains of the Romanov family by DNA analysis. National Genetics; 6:130-135.

Gill P., Kimpton C.P., d'Aloja E., Andersen JF., Bar W., et al. (1994). Report of the European DNA profiling group (EDNAP)--towards standardization of short tandem repeat (STR) loci. Forensic Science International; 65:51-59.
Gill P., Kimpton C.P., Urquhart A., Oldroyd N., Watson T.J and Downes T.J (1995). Automated short tandem repeat (STR) analysis in forensic casework - strategy for the future. Electrophoresis; 16:1543-1552.

Gill P., Sullivan K. and Werrett D.J. (1990). The Analysis of Hypervariable DNA Profiles: Problems Associated with the Objective Determination of the Probability of a Match. Human Genetics; 85:75-79.

Gill P., Urquhart A., Millican E.S., Oldroyd N., Watson T.J., et al. (1996). A new method of STR interpretation using inferential logic - development of a criminal intelligence database. International Journal of Legal Medicine, 109:14-22.

Guo S.W and Thompson E.A (1992). Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics; 48:361-372.

Gyapay G., Morissette J., Vignal A., Dib C., Fizames C., Millasseau P., et al., (1994). The 1993-1994 Genethon human genetic linkage map. Nature Genetics; 7: 246-339.

Hagelberg E., Gray I.C. and Jeffreys A.J. (1991). Identification of the skeletal remains of a murder victim. Nature; 352:427-429.

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. American Journal of Human Genetics; 55:175-89.

Hashiyada M., Itakur Y., Nata M. (2003).PowerPlex 16 analysis in the Japanese population. International Congress Series; 1239: 125– 130

Helmuth R., Fildes N., Blake E., Luce M., Chimera J., et al. (1990). HLA-DQ alpha allele and genotype frequencies in various human populations, determined by using enzymatic amplification and oligonucleotide probes. American Journal of Human Genetics; 47(3):515-23.

Hochmeister M.N., Budowle B., Jung J., Borer U.V., Comey C.T. and Dirnhofer R. (1991). PCR-based typing of DNA extracted from cigarette butts. International Journal of Legal Medicine; 104(4):229-33.

Janeway C.A., Travers P., (1997). The Immune System in Health and Disease. Immuno Biology; Churchill Livingstone press.

Jeffreys A.J., Allen M.J., Hagelberg E. and Sonnberg A. (1992). Identification of the skeletal remains of Josef Mengele by DNA analysis. Forensic Science International, 56:65-76.

Jeffreys A.J., Wilson V. and Thein S.L. (1985) Hypervariable 'Minisatellite' Regions in Human DNA. Nature; 314:67-73.

Jeffreys A.J., Wilson V. and Thein S.L. (1985). Individual Specific 'Fingerprints' of Human DNA. Nature; 316:76-79.

Jeffreys A.J., Wilson V., Neumann R. and Keyte J. (1988). Amplification of human minisatellites by the polymerase chain reaction: towards DNA fingerprinting of single cells. Nucleic Acids Research; 16: 10953−10971.

Jones D.A. (1972). Blood samples: probability of discrimination. Journal of the Forensic Science Society;12: 355–359.

Kasai K., Nakamura Y. and White R. (1990). Amplification of a variable number of tandem repeats (VNTR) locus (pMCT118) by the polymerase chain reaction (PCR) and its application to forensic science. Journal of Forensic Sciences; 35(5) pg. 1196-1200.

Kimpton C.P., Fisher D., Watson S., Adams M., Urquhart A., et al. (1994). Evaluation of an automated DNA profiling system employing multiplex

amplification of four tetrameric STR loci. International Journal of Legal Medicine; 106:302-311.

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 repeat loci. PCR Methods and Applications; 3: 13-22.

Kloosterman A. D. (2003). Current and Future Developments in Forensic DNA Typing. Scriptie Biomedische Wetenschappen van E. Besselink

Kruglyak S., Durrett R.T., Schug M.D. and Aquadro, C.F. (1998). Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. Proceedings of the National Academy of Sciences; 95: 10774-10778.

Leda K., Thomas J., Rebecca S., Jodi A. (2006).Genetic variation for 15 autosomal STR loci (PowerPlex® 16) in a population sample from northern Greece. Forensic Science International; 159: 61–63

Levinson G. and Gutman G.A. (1987). Slipped-strand impairing: A major mechanism for DNA sequence evolution. Molecular Biology and Evolution; 4, 203.21.

Lins A.M., Sprecher C.J., Puers C. and Schumm J.W. (1996). Multiplex sets for amplification of polymorphic short tandem repeat loci-silver stain and fluorescent detection. BioTechniques; 20: 882–889.

Louai A., Ichiroh S., Bernd B. and Carsten H.(2003). Analysis of 15 short tandem repeats reveals significant differences between the Arabian populations from Morocco and Syria. International Journal of Legal Medicine; S150–S155

Magnuson V.L., Ally D.S., Nylund, S.J., Karanjawala Z.E., Rayman J.B., et al. (1996). Substrate nucleotide-determined non-templated addition of adenine by Taq DNA polymerase: Implications for PCR-based genotyping and cloning. BioTechniques; 21, 700.9.

Martinez G., Vazquez E., Schaller C., and Quevedo N. (2003). Genetic data on 11 STRs (CSF1PO, TPOX, TH01, F13A01, FESFPS, vWA, D7S820, F13B, LPL) in an Argentine northeast population. Forensic Science International; 133: 254-255.

Michael K., Nabil A.H. and Barbara R. (1999). Population genetic studies on the tetrameric short tandem repeat loci D3S1358, VWA, FGA, D8S1179, D21S11, D18S51, D5S818, D13S317 and D7S820 in Egypt. Forensic Science International; 104:23–31

Mills K.A., Even D. and Murray J.C. (1992). Tetranucleotide repeat polymorphism at the human alpha fibrinogen locus (FGA). Human Molecular Genetic; 1, 779.

Moller A., Meyer E. and Brinkmann B. (1994). Different types of structural variation in STRs: HumFES/FPS, HumVWA and HumD21S11. International Journal of Legal Science; 106, 319.23.

Mullis K., Faloona F., Scharf S., Saiki R., Horn G. and Erlich H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harbor Symposium on Quantitative Biology; 51 Pt 1:263-73.

Mullis K.B. and Faloona F.A. (1987). Specific synthesis of DNA in vitro via a polymerase catalyzed chain reaction. Methods Enzyme; 155:335-50.

Myun S.H., Pil W.K., Dong H.C., Yang H., Sang K.C. and Wook K. (2001). Genetic variation at eight STR loci in the Korean population. Forensic Science International; 116 35±36.

National Research Council. (1996). The evaluation of Forensic DNA Evidence. Washington, D.C.: National Academy Press.

Olsson M.L., Guerreiro J.F. and Zago M.A. (1997). Molecular analysis of O alleles at the blood group ABO locus in populations of different ethnic origin reveals novel crossing over events and point mutations. Biochemical & Biophysical Research Communications; 234 (3): 779-782.

Perneger T.V. (1998). What's wrong with Bonferroni adjustments. British Medical Journal, 316: 1236–1238.

PowerStatsV12 (1999). A computer program for the analysis of population statistics, Available from: http://www.promega.com/geneticidtools/).

Primrose S.B. (1998). Principles of Genome Analysis: A Guide to Mapping and Sequencing DNA from Different Organisms, Second Edition. Malden, MA: Blackwell Science.

Promega Corporation. (2002). PowerPlex® ES System Technical Manual. Madison, WI.

Puers C., Hammond H.A., Jin L., Caskey C.T. 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 allelic ladder. American Journal of Human Genetics; 53: 953–958.

Saiki R.K., Bugawan T.L., Horn G.T., Mullis K.B. and Erlich H.A. (1986). Analysis of enzymatically amplified beta-globin and HLA-DQ alpha DNA with allele-specific oligonucleotide probes. Nature; 324(6093):163-6.

Sajantila A., Puomilahti S., Johnsson V. and Ehnholm C. (1992). Amplification of reproducible allele markers for amplified fragment length polymorphism analysis. BioTechniques; 12: 16–21.

Schlotterer C. and Tautz D. (1992). Slippage synthesis of simple sequence DNA. Nucleic Acids Research; 20: 211.5.

Schneider S., Roessli D., and Excoffier L. (2000). Arlequin v 3.1: software for population genetics data analysis. Available from: http://anthro.unige.ch/arlequin

Slatkin M. (1994). Linkage disequilibrium in growing and stable populations. Genetics; 137:331-336.

Slatkin M. and Excoffier L. (1996). Testing for linkage disequilibrium in genotypic data using the EM algorithm. Heredity; 76:377-383.

Smith J.R., Carpten J.D., Brownstein M.J., Ghosh S., Magnuson V.L., et al. (1995). Approach to genotyping errors caused by non-templated nucleotide addition by Taq DNA polymerase. Genome Research; 5: 312.7.

Smith R.N. (1995). Accurate size comparison of short tandem repeat alleles amplified by PCR. BioTechniques; 18: 122–128.

Sparkes R., Kimpton C., Gilbard S., Carne P., Andersen J., et al. (1996). The validation of a 7-locus multiplex STR test for use in forensic casework: (II) Artifacts, casework studies and success rates. International Journal of Legal Science; 109(4): 195-204.

Subramanian S., Mishra RK. and Singh L. (2003).Genome wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions. Genome Biology; 4, R13.

Sudhir S., Muhammad A., Carol R., James E., Hamby, et al. (1999). Typing of eight short tandem repeat (STR) loci in a Saudi Arabian population. Forensic Science International; 104:143–146

Sullivan K.M., Mannucci A., Kimpton C.P. and Gill P. (1993). A rapid and quantitative DNA sex test: fluorescence- based PCR analysis of X-Y homologous gene amelogenin. BioTechniques*;* 15, 636–641.

Tautz D. (1993). Notes on definition and nomenclature of tandemly repetitive DNA sequences.

Tully G., Sullivan K.M. and Gill P. (1993). Analysis of 6 VNTR loci by "multiplex" PCR and automated fluorescent detection. Human Genetics; 92: 554–562.

Toth G., Gaspari Z. and Jurka J. (2000). Microsatellites in different eukaryotic genomes: survey and analysis. Genome Research; 10: 967-981.

Twyman R.M. and Primrose S.B. (2003). Techniques patents for SNP genotyping. Pharmacogenomics; 4: 67-79

Ugo R., Ilaria S., Laura G., Silvia G., Silvia C., et al. (2002). Analysis of 13 tetrameric short tandem repeat loci in a population of Tuscany (Central Italy) performed by means of an automated infrared sequencer. Forensic Science International; 125: 83–85.

Ulises T., Gabriela B. and Eduardo R. (2003). STR data for PowerPlex116 System from Neuquen population, SW Argentina. Forensic Science International; 134: 219–22.

Urquhart A., Kimpton C.P., and Gill P. (1993).Sequence variability of the tetranucleotide repeat of the human beta-actin related pseudogene H-beta-Ac-psi-2 (ACTBP2) locus. Human Genetic; 92: 637-638.

Urquhart A., Kimpton C.P., Downes T.J., and Gill P. (1994). Variation in short tandem repeats sequences survey of twelve microsatellite loci for use as forensic identification markers. International Journal of Legal Medicine; 107: 13-20.

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 Research; 24: 2807–2812.

Weber J.L. and Wong C. (1993). Mutation of human short tandem repeats. Human Molecular Genetic; 2: 1123-1128.

Weir B.S. (1996). Genetic data analysis 2: methods for discrete population genetic data**.** Second Edition. Sunderland**,** MA: Sinauer Associates, Inc, 92–5.

Wilgland P., Budowle B., Rand S. and Brinkmann B. (1993). Forensic validation of the STR systems SE33 and TC11. International Journal of Legal Medicine; 105:315-320.

Wright S. (1951). The genetical structure of populations. Annals of Eugenics; 15: 323-354.

Yanmei S., Jian W., Chunying Z., Bingtao H., Yi L., et al. (2002). Allele frequencies for nine PCR-typed STR loci in a population from middle China. Forensic Science International; 127:145–146.

Zhou H.G., Sato K., Nishimaki Y., Fang L., and Hasekura H. (1997) The HumD21S11 system of short tandem repeat DNA polymorphisms in Japanese and Chinese. Forensic Science International; 86(1-2):109-188.

Appendix 1: Allele frequencies for the fifteen STR loci in the Palestinian population of Gaza Strip [n = 125 or 250 chromosomes measured].

Allele Frequency

Allele	D3S1358	TH01	D21S11	D18S51	Penta E	D5S818	D13S317	D7S820	D16S539	CSF1PO	Penta D	vWA	D8S1179	TPOX	FGA
2.2											0.032				
5					0.056					0.004					
6		0.256									0.004			0.004	
7		0.248			0.100			0.012			0.012				
8		0.108			0.020	0.016	0.160	0.172	0.028	0.016	0.028		0.020	0.436	
9		0.280			0.028	0.096	0.084	0.120	0.164	0.024	0.204		0.016	0.184	0.004
9.3		0.072													
10		0.036		0.004	0.084	0.096	0.044	0.300	0.148	0.304	0.140		0.096	0.052	
11				0.028	0.100	0.276	0.256	0.224	0.308	0.312	0.160		0.104	0.276	
12				0.188	0.184	0.348	0.364	0.132	0.224	0.304	0.132		0.104	0.040	
13	0.004			0.144	0.140	0.156	0.052	0.040	0.104	0.032	0.144		0.220	0.008	
14	0.036			0.140	0.048	0.012	0.040		0.024	0.004	0.104	0.068	0.220		
15	0.196			0.096	0.056						0.032	0.144	0.152		
16	0.304			0.148	0.036						0.008	0.212	0.052		
17	0.308			0.116	0.032							0.292	0.016		
17.2				0.004											
18	0.148			0.092	0.036							0.176			0.004
19	0.004			0.020	0.040							0.096			0.068
20				0.012	0.024							0.012			0.116
21				0.008	0.016										0.132
21.2															0.012
22															0.164
23															0.160
23.2															0.008
24															0.132
24.2															0.004

Continue …

Allele Frequency															
Allele	D3S1358	TH01	D21S11	D18S51	Penta E	D5S818	D13S317	D7S820	D16S539	CSF1PO	Penta D	vWA	D8S1179	TPOX	FGA
25															0.144
26															0.048
27			0.012												
28			0.120												0.004
29			0.232												
30			0.248												
30.2			0.044												
31			0.036												
31.2			0.128												
32			0.016												
32.2			0.104												
33.2			0.040												
34			0.012												
35			0.004												
35.2			0.004												

Appendix 2: Alleles genotype for the fifteen STR loci in a Palestinian population of Gaza Strip [n = 125 or 250 chromosomes measured].

Sample #	D3S1358	TH01	D21S11	D18S51	Penta E	D5S818	D13S317	D7S820	D16S539	CSF1PO	Penta D	vWA	D8S1179	TPOX	FGA
1	16	9	29	13	14	12	12	11	11	10	11	16	13	8	19
	16	9.3	32.2	18	18	13	12	13	11	12	13	17	14	8	20
2	15	6	28	13	7	11	8	9	11	11	10	17	14	8	20
	18	6	32	14	13	12	11	12	12	12	10	17	16	11	23
3	16	6	32.2	13	12	12	12	8	11	10	12	17	10	11	21
	16	6	33.2	14	19	13	14	9	12	12	14	18	12	11	24
4	17	6	32.2	11	12	10	8	8	9	10	12	17	10	8	19
	17	7	32.2	21	12	12	11	10	11	12	13	18	12	9	26
5	16	6	28	12	5	9	8	10	10	11	9	19	11	8	21.2
	18	8	29	14	7	9	12	10	11	12	9	20	13	11	23
6	17	8	29	14	7	9	12	10	11	11	9	19	11	11	19
	18	9	29	18	18	9	12	11	12	11	12	20	15	11	23
7	17	9.3	30	14	20	11	9	8	10	10	11	15	11	10	20
	17	10	32.2	14	20	12	11	10	10	11	14	17	14	11	25
8	16	6	28	12	13	11	8	10	11	10	9	14	14	10	20
	17	7	29	15	13	11	12	11	12	12	9	15	14	12	20
9	15	6	31.2	12	12	11	11	12	9	10	10	14	12	8	23
	16	7	31.2	15	15	11	11	12	13	11	13	14	17	11	24
10	15	6	29	11	11	9	9	10	9	9	2.2	16	11	8	25
	16	8	30	18	13	10	11	10	9	11	2.2	19	16	9	25
11	16	6	30	12	5	12	8	8	12	11	11	14	10	8	23
	18	7	32.2	16	14	13	13	11	13	12	14	17	11	8	24
12	15	6	29	11	12	11	8	8	9	11	9	15	14	9	19
	15	9	30	12	12	13	9	11	14	12	13	15	14	11	23
13	15	9	29	12	11	8	8	10	11	10	9	15	14	8	22
	17	9	29	12	12	12	12	10	12	11	13	17	16	11	23
14	15	6	28	12	12	9	12	8	9	12	11	16	12	9	20
	17	7	30	18	13	13	12	10	11	12	12	16	14	9	25
15	16	6	29	12	12	11	8	8	11	9	13	15	10	8	22
	19	8	29	15	13	12	12	10	14	10	13	18	13	8	24
16	16	6	31	12	13	9	9	10	9	10	9	17	14	8	22
	17	8	31.2	17	14	10	12	12	12	10	9	17	16	8	22
17	15	6	29	14	12	11	9	8	10	10	2.2	15	8	8	21
	17	9	30	16	12	12	11	11	12	12	13	17	13	9	22
18	15	6	30	12	11	11	12	11	9	5	12	15	12	11	20
	16	6	35	14	16	12	12	12	11	13	14	16	15	11	21
19	15	9	27	13	10	11	8	11	9	9	8	16	15	8	23
	18	9	31.2	15	12	12	12	11	12	11	9	17	15	8	25
20	15	7	30.2	12	11	10	11	7	9	10	13	14	10	8	19
	17	8	32.2	12	13	11	12	10	10	10	14	17	11	8	21
21	15	6	30	11	11	8	9	10	12	11	9	15	14	8	20
	17	9	32.2	18	14	11	14	11	14	12	12	18	15	11	22
22	14	8	30	17	11	12	12	9	8	10	10	14	13	8	21
	15	9	30	18	12	12	12	11	13	10	13	17	14	11	23
23	15	9	29	12	13	12	8	10	11	11	11	16	10	8	21
	17	9.3	30	13	13	13	12	11	11	12	12	19	15	11	25
24	16	6	27	12	7	11	11	10	9	10	9	17	10	8	21
	17	7	29	13	19	13	13	11	12	12	11	18	15	8	24
25	16	9	28	12	13	11	12	9	8	11	12	16	11	8	22
	17	9	31	16	14	13	13	12	12	12	12	19	17	8	24
26	15	6	30	15	8	9	11	10	9	10	10	14	12	8	20
	17	9	30	15	10	13	12	11	11	11	11	15	15	12	21

27	16	8	30	13	15	9	8	8	11	8	9	19	9	8	23
	17	9	31.2	16	16	13	9	13	13	11	9	19	16	11	26
28	16	6	28	10	5	11	12	8	11	10	9	15	11	8	19
	17	8	30	16	5	13	12	10	13	12	10	18	15	11	25
29	16	6	29	16	12	10	8	9	11	12	11	16	10	9	20
	18	9	30	17	13	12	12	12	12	13	11	16	12	11	23
30	15	7	31.2	12	10	11	11	10	9	11	12	17	13	8	22
	15	9	31.2	13	11	13	12	13	12	12	13	17	14	11	24
31	17	6	29	12	10	10	12	8	11	11	10	17	14	11	20
	18	8	31.2	16	13	12	12	12	11	12	10	18	15	12	20
32	15	6	31.2	16	14	11	11	10	11	11	11	17	13	11	21
	17	9	33.2	18	16	12	12	11	14	11	11	18	14	11	25
33	16	6	29	12	8	12	12	9	10	11	9	15	13	9	19
	17	9	31.2	16	13	13	12	11	10	12	10	18	14	10	23.2
34	16	6	30	14	7	11	11	9	9	11	12	16	10	8	18
	18	9.3	30.2	18	19	12	12	11	12	11	13	17	15	11	23
35	14	7	29	12	11	11	12	9	10	10	9	17	13	11	20
	17	7	33.2	14	11	12	12	11	10	10	14	17	14	11	22
36	15	7	28	12	9	9	12	11	9	10	11	18	12	8	21
	17	9	32.2	13	21	11	12	12	9	12	12	18	14	10	22
37	16	9	29	12	15	9	11	10	9	10	10	16	11	8	20
	17	10	29	14	18	9	12	13	13	12	10	17	13	8	20
38	14	7	31.2	16	7	10	9	9	11	12	9	16	8	8	24.2
	16	7	33.2	17	9	12	12	11	13	13	10	17	13	11	25
39	15	9.3	30	12	7	13	13	8	12	8	13	16	14	8	22
	17	9.3	32.2	15	10	14	13	11	13	13	16	18	15	8	25
40	17	6	30.2	14	13	11	9	7	9	10	10	16	14	10	19
	17	7	30.2	16	20	11	12	8	12	11	10	17	14	11	21
41	16	9	29	17	14	12	11	9	9	10	9	16	13	8	23
	16	9.3	33.2	20	14	12	12	12	12	10	16	19	14	11	25
42	17	6	30	13	12	11	11	8	12	10	13	16	10	8	20
	18	9.3	31	19	17	12	11	8	12	11	14	17	16	11	23
43	16	7	30	17	7	10	8	10	9	11	11	15	13	8	24
	17	9	31.2	17	12	11	8	11	11	12	11	18	13	11	24
44	15	6	30	12	7	11	8	11	9	12	9	15	10	8	24
	16	9	30	17	11	11	9	12	10	14	13	18	15	11	24
45	15	6	29	14	10	10	8	11	9	10	2.2	14	14	8	25
	18	9	29	17	13	11	14	12	11	11	12	17	15	9	25
46	15	7	29	15	10	12	10	8	12	11	11	15	9	8	19
	16	7	32.2	16	15	13	12	12	13	12	11	17	12	11	25
47	15	8	31	13	10	12	11	8	9	11	9	14	10	9	20
	16	9	31.2	15	16	13	11	10	11	12	13	17	17	11	25
48	17	7	30	12	7	11	8	10	12	10	11	17	10	9	23
	18	8	32.2	18	17	12	11	12	13	12	13	18	13	9	26
49	17	6	29	14	10	10	8	8	9	10	9	16	13	8	24
	18	9	29	17	12	12	14	11	11	12	14	19	14	9	25
50	14	7	31.2	13	9	12	10	9	11	12	9	15	9	8	19
	14	9	32	21	10	13	12	12	11	12	13	17	14	9	23
51	17	7	30	14	10	10	11	10	10	11	2.2	17	10	8	22
	17	7	30	18	18	13	12	11	11	11	12	19	14	11	25
52	17	7	30	16	10	10	11	8	13	10	2.2	18	11	11	24
	17	8	31.2	18	11	11	12	10	13	12	9	19	15	11	25
53	15	7	29	12	5	10	11	10	11	10	10	17	13	9	20
	15	9	29	17	13	10	11	10	12	11	11	18	15	11	21
54	13	7	28	13	7	11	8	8	11	11	8	17	10	8	23
	16	7	31.2	14	13	11	10	11	11	12	9	18	14	8	23
55	17	6	28	12	11	10	10	10	11	11	9	17	11	8	24
	17	9.3	28	13	15	11	12	10	12	12	10	17	15	9	24
56	15	9	29	14	10	12	10	9	11	9	9	14	13	11	21

	17	9.3	31.2	15	16	12	12	10	12	12	14	16	15	12	24
57	16	9.3	29	17	9	12	12	11	9	11	9	17	14	8	22
	18	10	32.2	17	17	12	14	12	12	12	12	19	15	11	28
58	15	6	30	13	10	12	12	11	9	10	10	16	13	8	26
	18	9	30.2	17	13	13	12	11	13	12	11	16	15	9	26
59	16	6	29	13	5	10	12	10	11	10	14	17	12	8	20
	18	10	30.2	15	20	12	12	11	13	11	14	18	14	9	24
60	16	6	29	14	15	12	11	8	12	11	9	17	11	8	22
	16	9	29	15	15	13	12	9	13	12	12	17	13	9	25
61	15	6	28	13	13	10	11	10	10	11	11	16	13	8	25
	16	9	30	17	13	12	12	10	11	12	14	17	14	11	25
62	15	7	29	12	11	11	9	9	9	10	9	14	11	8	22
	17	9	31.2	13	13	12	14	10	11	12	14	16	14	8	23
63	16	7	28	12	10	11	9	8	11	12	9	17	10	8	19
	17	8	32.2	16	13	12	12	10	12	12	11	17	12	8	24
64	16	6	31	14	12	9	10	9	10	10	12	17	13	8	24
	16	9	33.2	16	16	13	13	10	11	11	15	18	15	8	24
65	16	6	28	15	12	10	8	10	9	11	10	16	13	10	21
	16	6	30	17	13	13	12	12	12	12	15	19	13	10	23
66	16	9	29	14	5	12	11	8	11	11	8	16	11	8	22
	18	9	30	15	10	12	11	11	12	11	11	16	14	10	25
67	16	8	28	12	12	11	8	10	10	10	8	16	13	11	21
	18	9	30	12	18	13	11	11	11	10	9	17	15	11	22
68	16	8	28	13	11	11	10	12	8	10	10	16	14	8	21.2
	16	9	28	19	15	13	13	12	12	10	12	18	16	11	22
69	16	6	28	16	7	13	11	11	12	10	11	16	8	11	21
	17	9.3	28	16	13	14	11	12	13	13	12	16	14	11	25
70	16	6	30.2	13	17	10	14	9	9	10	9	18	11	10	26
	16	6	32.2	17	17	11	14	12	11	12	11	19	11	11	26
71	16	6	29	13	7	12	8	9	12	12	15	15	12	8	21
	18	9	30	16	15	12	14	11	13	12	15	15	14	9	23
72	16	7	28	15	12	12	11	11	8	12	9	14	10	9	21
	17	9	31	16	12	12	12	11	9	13	14	14	12	12	21
73	15	6	29	13	11	11	11	11	11	10	9	17	15	10	23
	16	9	34	17	12	12	12	11	11	12	13	18	15	11	23
74	15	9	29	14	11	9	11	9	11	11	12	16	13	11	22
	16	9	33.2	18	12	9	12	10	11	12	13	19	14	12	24
75	17	9	29	16	7	12	10	8	12	11	9	18	13	8	21
	18	9.3	30	20	10	13	11	9	14	12	12	18	15	9	23
76	15	7	27	17	5	11	8	10	9	10	11	15	13	8	21
	16	9	33.2	18	13	11	11	10	12	12	11	17	14	11	26
77	17	6	29	12	13	11	8	11	10	10	10	18	10	9	22
	17	7	31.2	16	19	11	11	12	10	12	11	19	13	9	24
78	14	6	29	13	12	11	13	8	8	12	10	17	13	8	24
	18	6	30	18	12	12	13	8	9	12	14	17	16	8	26
79	15	7	30	11	5	12	11	8	11	10	10	14	13	8	23
	17	7	32.2	14	15	12	11	10	11	11	12	16	14	9	24
80	15	6	29	12	14	10	12	9	8	10	10	16	8	8	23
	16	6	33.2	14	16	12	14	11	9	11	11	16	11	8	23
81	15	7	30	14	11	11	10	10	11	11	13	17	13	8	23
	16	9.3	31	16	12	13	13	13	12	12	13	18	15	12	25
82	15	7	29	14	5	12	11	8	10	8	7	14	11	8	20
	15	7	32.2	17	15	13	11	10	11	10	12	15	13	11	20
83	15	7	28	16	9	12	11	10	9	10	13	17	12	8	22
	17	9.3	31.2	18	12	13	12	10	11	13	13	20	13	10	23
84	15	7	30	11	5	12	11	8	11	8	7	15	10	8	23
	17	7	32.2	17	15	12	11	10	11	11	10	15	14	11	25
85	16	7	30	13	13	12	8	9	12	10	10	15	12	11	19
	18	9	31.2	15	13	12	11	13	12	12	14	16	16	11	22

86	15	6	29	13	5	11	11	10	11	12	9	15	13	8	22
	17	7	32.2	17	13	12	12	12	11	12	11	16	14	9	23
87	16	6	30	14	13	9	8	10	12	11	10	16	12	9	25
	17	9	31.2	16	14	12	8	12	12	12	11	17	15	13	25
88	16	8	30	12	18	12	8	11	9	10	13	15	10	8	21
	18	8	30	13	21	13	12	13	12	10	15	15	15	11	25
89	16	7	29	13	10	12	11	10	10	10	13	16	13	8	22
	17	9	32.2	17	12	12	12	11	11	12	14	18	15	8	23
90	17	6	31.2	13	10	11	12	10	10	10	8	16	12	8	22
	17	9.3	34	18	16	12	12	13	12	11	9	17	14	10	23
91	17	6	31.2	16	12	8	11	10	10	10	9	16	14	9	20
	18	8	32.2	16	12	10	11	10	12	11	14	16	15	9	22
92	16	6	30	14	13	9	8	10	12	11	10	16	12	9	25
	17	9	31.2	16	14	12	8	12	12	12	11	17	15	13	25
93	16	6	28	11	7	9	8	8	10	10	10	15	13	8	22
	18	7	30.2	19	7	12	12	13	10	11	11	19	16	8	22
94	15	6	31.2	12	5	11	9	9	11	12	9	17	12	8	23
	16	9	31.2	15	17	13	12	11	11	12	13	17	16	11	24
95	16	9.3	28	14	8	9	9	10	10	11	8	19	13	8	22
	18	10	30	18	13	12	11	12	11	12	10	19	14	8	26
96	16	7	30	12	11	12	8	10	11	11	11	17	11	8	21
	18	9	30.2	14	12	13	9	11	12	11	15	18	14	8	22
97	16	7	32	17	12	11	9	8	10	10	13	17	15	11	21
	17	9	32.2	20	17	12	12	12	13	12	14	17	15	11	25
98	15	6	30	13	8	12	11	9	10	10	6	15	8	8	24
	17	9	31.2	16	12	13	12	11	11	11	8	17	14	8	24
99	15	7	29	12	7	12	10	9	10	10	11	15	13	8	19
	17	8	29	12	7	12	12	9	13	12	11	15	13	9	26
100	17	7	29	13	11	11	8	8	11	9	10	16	12	8	24
	17	9	30	16	11	11	12	8	12	10	14	17	13	9	25
101	16	7	30	13	12	11	9	8	12	10	9	16	14	9	19
	17	7	31	15	12	12	12	8	13	11	15	18	14	10	22
102	16	7	28	12	12	11	12	8	9	11	12	17	10	6	19
	17	7	32.2	17	17	12	13	11	11	11	14	17	10	11	24
103	14	9	28	12	8	11	12	12	11	10	12	17	14	9	20
	17	9.3	34	17	12	12	12	12	13	12	13	17	15	11	23.2
104	16	7	30	15	11	11	11	11	10	10	9	15	9	8	20
	18	9	30	17	18	12	12	11	12	11	9	18	13	8	23
105	17	8	28	12	13	12	9	7	10	12	10	15	14	9	21
	17	9	30	15	20	13	9	12	10	12	11	19	14	11	21
106	18	6	30	12	21	12	11	10	11	11	9	15	11	8	21
	18	9	31.2	16	21	13	12	11	12	11	13	18	16	9	23
107	18	7	30	13	12	11	12	11	13	10	9	17	13	8	21
	18	8	35.2	17	19	12	12	11	13	13	12	18	15	12	23
108	16	6	30	12	7	12	12	8	8	10	7	18	10	11	19
	18	9	32.2	18	10	14	12	10	9	11	12	18	13	11	24
109	16	8	32.2	15	7	11	11	10	11	11	11	16	12	9	22
	18	9	33.2	19	12	12	12	11	14	11	12	16	16	11	26
110	15	6	28	12	5	11	8	10	10	11	9	16	11	11	21
	15	7	31.2	14	7	12	8	11	11	11	14	18	14	11	24
111	16	7	30	14	7	12	11	9	11	10	9	15	13	9	21
	18	10	31.2	15	11	13	11	13	11	11	11	19	13	11	21
112	17	7	29	13	19	11	12	8	10	11	9	14	10	8	22
	18	9	30	16	19	12	13	9	11	12	11	17	14	8	22
113	15	7	29	16	12	11	8	8	11	10	10	14	11	8	21
	17	9	31.2	17.2	12	12	8	10	12	10	13	15	12	9	23
114	15	7	28	14	12	9	12	10	10	10	2.2	16	13	8	22
	16	10	30	18	15	11	12	10	11	10	14	17	13	12	22
115	17	9	29	15	12	11	8	9	9	10	10	18	13	8	22

	17	9	29	18	19	12	11	11	10	11	15	18	13	8	24
116	16	6	29	16	10	11	9	10	10	10	10	18	14	8	20
	17	9	30	18	12	12	12	12	10	10	13	18	14	9	20
117	15	6	28	14	9	11	11	8	9	10	12	17	13	8	25
	16	9	32.2	14	12	12	12	10	12	10	13	17	14	9	25
118	14	7	29	13	14	10	9	10	9	10	13	16	11	8	19
	17	10	32.2	15	19	11	10	10	9	11	14	18	15	8	20
119	17	6	29	16	7	8	12	9	10	11	11	15	11	8	21
	17	9	30.2	17	19	11	13	10	13	11	14	16	17	12	25
120	16	7	28	13	16	11	11	12	11	11	12	17	12	8	20
	18	8	31	19	20	12	11	12	11	12	13	17	13	9	22
121	15	7	30	16	7	10	11	8	13	11	2.2	18	11	8	22
	17	9	30	18	18	13	11	10	13	12	12	19	15	11	23
122	14	8	30	12	11	13	11	9	11	12	9	16	12	9	22
	16	10	32	16	13	13	11	10	12	12	11	18	12	9	22
123	17	7	29	13	11	11	8	8	11	9	10	16	12	8	24
	17	9	30	16	11	11	12	8	12	10	14	17	13	9	25
124	16	7	29	12	15	9	12	10	12	11	12	16	13	9	19
	17	9	30	18	18	9	12	11	12	12	13	19	15	11	20
125	16	6	28	12	5	9	8	10	10	11	9	19	11	8	21.2
	18	8	29	14	7	9	12	10	11	12	9	20	13	11	23