4-2019

Study of Autosomal STR Markers in United Arab Emirates Population

Mohammed Naji Mohammed

United Arab Emirates University

College of Science

Department of Biology

STUDY OF AUTOSOMAL STR MARKERS IN UNITED ARAB EMIRATES POPULATION

Mohammed Naji Naji Mohammed

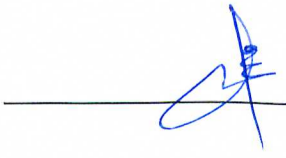
This thesis is submitted in partial fulfilment of the for the degree of Master of Science in Molecular Biology and Biotechnology

Under the Supervision of Dr. Synan Abu-Qamar

April 2019

Declaration of Original Work

I, Mohammed Naji Naji Mohammed, the undersigned, a graduate student at the United Arab Emirates University (UAEU), and the author of this thesis entitled "*Study of Autosomal STR Markers in United Arab Emirates Population*", hereby, solemnly declare that this thesis is my own original research work that has been done and prepared by me under the supervision of Dr. Synan Abu-Qamar in the College of Science at UAEU. This work has not previously been presented or published or formed the basis for the award of any academic degree, diploma or a similar title at this or any other university. Any materials borrowed from other sources (whether published or unpublished) and relied upon or included in my thesis have been properly cited and acknowledged in accordance with appropriate academic conventions. I further declare that there is no potential conflict of interest with respect to the research, data collection, authorship, presentation and/or publication of this thesis.

Student's Signature:  Date: 26/05/2019

Approval of the Master Thesis

This Master Thesis is approved by the following Examining Committee Members:

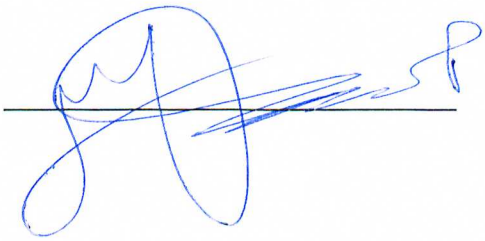
1) Advisor (Committee Chair): Dr. Synan AbuQamar
   Title: Associate Professor
   Department of Biology
   College of Science
   Signature _____ Date 21/04/2019

2) Member: Prof. Rabah Iratni
   Title: Professor
   Department of Biology
   College of Science
   Signature _____ Date 21/04/2019

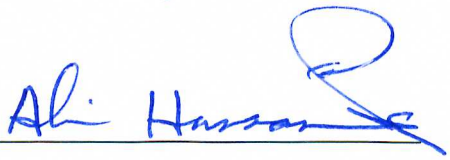
3) Member (External Examiner): Dr. Mohammed Nazier
   Title: Assistant Professor
   College of Biotechnology
   Institution: University of Modern Science, UAE
   Signature _____ Date 21/4/2019

This Master Thesis is accepted by:

Dean of the College of Science: Professor Ahmed Murad

Signature  Date 26/5/2019

Acting Dean of the College of Graduate Studies: Professor Ali H. Al-Marzouqi

Signature  Date 26/5/2019

Copy 1 of 10

Advisory Committee

1) Advisor: Dr. Synan Abu-Qamar

Title: Associate Professor

Department of Biology

College of Science

2) Co-advisor: Dr. Rashed Alghafri

Title: Adjunct Faculty

Department of Biology

College of Science

Abstract

A study of the United Arab Emirates (UAE) population is important due to high consanguineous marriage, which may affect the power of discrimination of some loci. The genetic polymorphisms of 23 autosomal short tandem repeat (STR) loci including D3S1358, vWA, D16S539, CSF1PO, TPOX, D8S1179, D21S11, D18S51, D2S441, D19S433, TH01, FGA, D22S1045, D5S818, D13S317, D7S820, D10S1248, D1S1656, D12S391, D2S1338, D6S1043, Penta D and Penta E were evaluated in 571 random unrelated UAE Arabs population. Blood samples were collected on FTA cards. Targeted loci were amplified using Verifiler Express PCR Amplification Kit. PCR products were run on the ABI 3500 Genetic analyzer. Arlequin and Forstat softwares were used to determine the forensic parameters and population structure analysis for 23 autosomal STRs. A total of 305 alleles were observed with the corresponding allelic frequencies ranging between 0.000876 and 0.49387. Data of forensic statistical parameters such as locus diversity ranged from 0.67406 (TPOX) to 0.9149 (Penta E). The most variable autosomal STR loci observed was Penta E (observed heterozygosity: 0.90368, match probability: 0.0147). Results suggest that the 23 STR loci had a relatively high genetic variation, which was suitable for forensic personal identification and paternity testing in the UAE population. The significance of this work is to build an allelic frequency database for the latest and most powerful amplification kit using current forensic workflow aiding statistical evaluation of generated STR profiles in the corresponding populations.

Keywords: short tandem repeat, allele frequency, genetic polymorphisms, power of discrimination, forensic personal identification.

Title and Abstract (in Arabic)

دراسة ترددات المواقع الجينية المستخدمة في تطبيقات الحمض النووي للأدلة الجنائية في شعب دولة الامارات العربية المتحدة.

الملخص

تعد دراسة المواقع الجينية لسكان دولة الإمارات العربية المتحدة مهمة نظرًا لارتفاع نسبة زواج الأقارب الذي قد يؤثر على قوة التمييز في بعض المواقع الجينية. التنوع الجيني لـ 23 موقع قصير التكرار (D3S1358, vWA, D16S539, CSF1PO, TPOX, D8S1179, D21S11, D18S51, D2S441, D19S433, TH01, FGA, D22S1045, D5S818, D13S317, D7S820, D10S1248, D1S1656, D12S391, D2S1338, D6S1043, Penta D, وPenta E) تم دراسة عينة عشوائية مكونة من 571 شخص غير مرتبطين ببعضهم بصلة قرابة في دولة الإمارات العربية المتحدة. تم جمع عينات الدم على بطاقات (FTA Cards). تم مضاعفة كمية المواقع المستهدفة باستخدام (VeriFiler™ Express PCR Amplification Kit)، ثم تم فصلها وتحليلها باستخدام المحلل الجيني (Genetic analyzer ABI 3500). تم استخدم برامج Arlequin وForstat لتحديد المعلمات الجنائية وتحليل التركيبة السكانية ل 23 موقع. وبعد تحليل نتائج الترددات الاليلية لوحظ ما مجموعه 305 أليلات مع ترددات أليلية تتراوح بين 0.000876 و0.49387. وتراوحت بيانات المعلمات الإحصائية الجنائية مثل التنوع الجيني للمواقع من 0.67406 (TPOX) إلى 0.9149 (Penta E). لوحظ أيضا إن Penta E يعتبر أكثر المواقع تنوعا (Observed heterozygosity: 0.90368، احتمالية المطابقة: 0.0147). تشير النتائج إلى أن المواقع ال 23 للترددات المتكررة (STR) لديها تنوعا جينيا مرتفعا نسبيا وهي مناسبة لتحديد الهوية الشخصية في المجال الجنائي واثبات الأبوة في سكان دولة الإمارات العربية المتحدة. تكمن أهمية هذا العمل في إنشاء قاعدة بيانات تردد أليلية لأحدث وأقوى Amplification Kitباستخدام سير عمل الاجراءات الجنائية الحالية بمساعدة التقييم الإحصائي للملفات الجينية (STR Profiles) التي تم إنشاؤها في المجموعات السكانية المقابلة.

مفاهيم البحث الرئيسية: المواقع الجينية، الترددات الاليلية، التنوع الجيني، قوة التمييز، المجال الجنائي.

Acknowledgements

I would like to express my special appreciation and thanks to my advisors Dr. Synan AbuQamar and Dr. Rashed Alghafri for supporting and encouraging me through my thesis. I have been extremely lucky to have supervisors who cared so much about my work and who responded to my questions and queries. They consistently allowed this thesis to be my own work, but steered me in the right direction whenever they thought I needed it.

A special thanks to my family. Words cannot express how grateful I am to my mother, father, brother and sisters for all the sacrifices that you have made on my behalf and for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. In addition, special thanks are extended to Ahmed Adan and Tariq Zeyad for their assistance and friendship.

Finally, I would like to express my deepest appreciation and thanks to Dubai Police General Head Quarter for providing me with the facilities during my work in my thesis research. In addition, I would like to thank Dubai Health Authority for providing me with the blood samples.

Dedication

To my beloved parents and family

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AMOVA	Analysis of Molecular Variance
bp	Base Pair
CDP	Combined Discrimination Power
CEP	Combined Exclusion Probability
CE	Capillary Electrophoresis
CMP	Combined Matching Probability
DNA	Deoxyribonucleic Acid
HWE	Hardy Weinberg Equilibrium
MB	Matching Probability
MDS	Multidimensional Scaling
PCR	Polymerase Chain Reaction
PD	Power of Discrimination
PE	Power of Exclusion
RFLP	Restriction Fragment Length Polymorphism
STR	Short Tandem Repeats
VNTR	Variable Number of Tandem Repeats

Chapter 1: Introduction

1.1 DNA Typing History

In the 1985, deoxyribonucleic acid (DNA) fingerprinting or DNA typing was described by an English geneticist Alec Jeffreys (Carracedo, 2015). He discovered that certain regions of DNA contained sequences that were repeated continuously. He additionally found that the number of repeated sequences present in a specific sample may differ from person to another (Butler & Hill, 2012). By developing a technique to analyze the length variation of these DNA sequence repeats, Alec created the ability to perform human identity tests. These DNA repeats regions became known as variable number of tandem repeats (VNTRs). Samples obtained from crime scene investigations or used in other forms of human identity testing were analyzed using both biological and genetics methods. DNA analysis depends on the comparison conducted between two samples: crime scene sample and a known reference sample or crime scene sample and another crime scene sample. The result of any forensic examination is a laboratory technical report, which represents a summary of work conducted by the forensic expert (Butler, 2012).

Blood is a core stain in forensic science for identifying individuals suspected of committing crimes. In early twentieth century, Karl Landsteiner described the ABO blood grouping system and found that there are differences in blood type between individuals (Harbison, 2016). Karl showed that different blood types could be divided into different groups those are A, B, AB, and O groups. This was the first step in the development of forensic haemogenetics. In 1915, Leone Lattes published a book explaining how to use ABO typing to resolve a paternity case; and by 1931, the ABO

typing technique became standard method in forensic laboratories (Goodwin, Linacre, & Hadi, 2011). In forensic science, scientists are using blood typing to exclude suspects to be the sources of blood samples deposited at crime scene. Though, this test does not clearly identify the suspect involved because there are many people sharing the same blood type.

In 1980, the first discovery of Restriction Fragment Length Polymorphisms (RFLP), Alec Jeffreys used RFLP to visualize length polymorphisms by treating genomic DNA samples with one or more restriction enzymes which cleave the DNA (each restriction enzyme will cut DNA in a unique restriction site); to produce number of fragments of the DNA with different lengths (Thompson, Zoppis, & McCord, 2012). The RFLP technique became the first scientifically accepted forensic DNA analysis method in the United States with few limitations. One of these limitations is it need large amounts of DNA, which is difficult when dealing with crime scene samples.

In 1986, polymerase chain reaction (PCR) was developed by Kary Mullis. PCR was the most important invention in molecular biology and widely supported technique for analyzing DNA (Watson, 2012). PCR facilitate the duplication of a specific DNA segment. Kary Mullis was awarded the Nobel Prize in Chemistry in 1993 for his work in developing this method (Heinze, 2013). The advantages of using PCR-based technology compared to other technologies are the analysis of small amount of samples and the analysis of degraded DNA samples.

1.2 Human Genome

The DNA sequence of the entire human genome was revealed by 2003 (Lander, 2011). Approximately three billion base pairs of deoxyribonucleic acid (DNA) within the 23 chromosome pairs in diploid cell nuclei and also there is DNA available in the mitochondria organelle (mtDNA). About 99.5 % of human genome is shared across all human individuals and that is what makes us human beings. The other 0.5% varies between people and that make people different from one to another such as in individual's traits like blood types, hair color and eye color (Consortium, 2015). This small portion is of interest to forensic scientists and it makes each individual's DNA unique with the exception of identical twins.

Human cells contain 23 pairs of chromosomes, including 22 pairs of autosomal chromosomes and a pair of sex chromosomes (Chu & Giles, 1959) (Figure 1). Females contain two X chromosomes while males contain a single copy of the X chromosome and a single copy of the Y chromosome. All body (somatic) cells contain two copies of each chromosome: one copy of the genome is inherited from the father and the other copy is inherited from the mother. These types of cells are called diploid cells. On the other hand, gametes cells (sperm and egg) contain only one copy of each chromosome. These types of cells are called haploid cells, so normal human cells contain 23 pairs of chromosomes which means 46 different chromosomes (Chu & Giles, 1959).

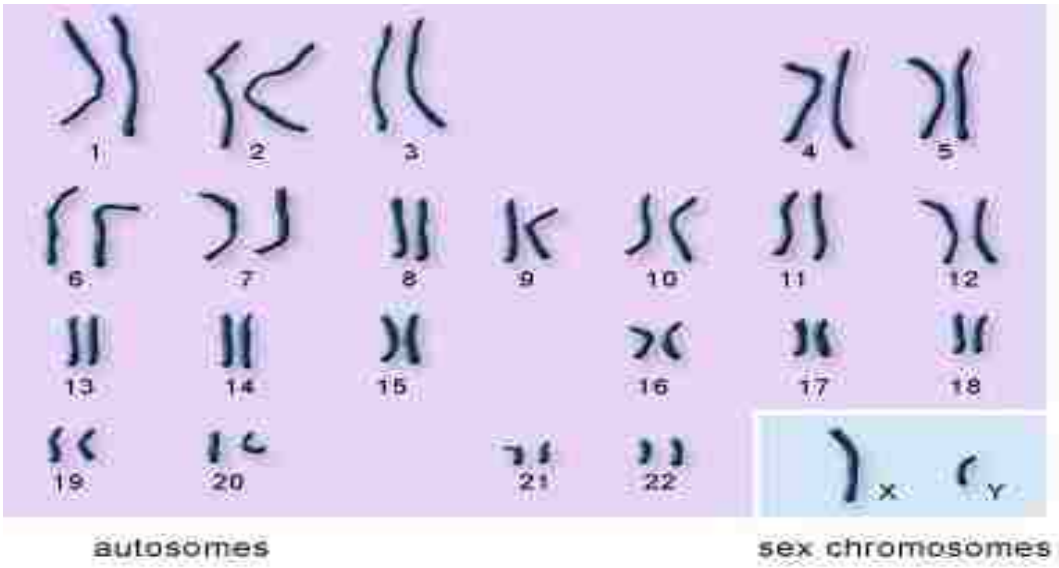


Figure 1: Human genome chromosomes in a diploid cell (J. M. Butler, 2005)

Approximately 95% of human genome is non-coding DNA which is not encoding protein sequences. The other 5% of human genomic DNA is called genes coding DNA that is responsible for coding protein. A human has approximately 25,000 to 30,000 genes. The average gene is sized from thousands to ten thousand of base pairs (Premalatha, Ramesh, Babu, & Balamurali, 2014). Genes consist of exons and introns. Introns are parts of genes that do not directly code for proteins whereas exons are the parts of the gene that are translated into protein (De Conti, Baralle, & Buratti, 2013). DNA is used as a template to create mRNA in a specific process called transcription. This mRNA then undergoes a further process called translation where the mRNA is used to synthesize proteins (Figure 2). Loci which are used in forensic human identification are found in the non-coding region either within genes or between genes.

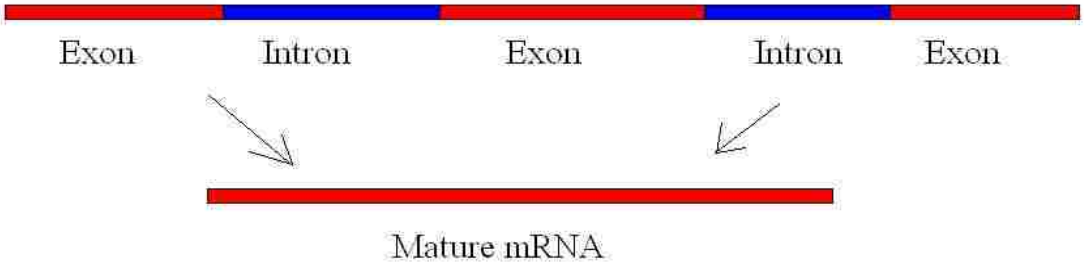


Figure 2: Gene composition and transcription step (J. M. Butler, 2011)

1.3 Short Tandem Repeats (STRs)

Various factors contribute in making DNA an excellent source of information that aids individualization in forensic science. There is almost no difference in DNA between cell types. An individual's DNA does not generally change throughout human lifetime, and DNA is less prone to degradation as compared to proteins and lastly high variations are exhibited among individuals and between species (J. M. Butler, 2011).

Repeated DNA sequences are spread in the human genome and show a good variability among individuals in a population that they have become important in several fields including genetic mapping, linkage analysis, and human identity testing (Alaeddini, Walsh, & Abbas, 2010). These tandemly repeated regions of DNA are typically classified into several groups depending on the size of the repeat region.

Minisatellites (VNTRs) have core repeat unit for a medium length repeat with

the range of approximately 10–100 bases in length (Pakzad, Mozdarani, Izadi-Mood, & Niromanesh, 2014). The DNA marker D1S80 is a minisatellite with a 16 bp repeat unit and contains alleles as shown in (Figure 3). While DNA regions with repeat units that are 2–7 bp in length are called microsatellites, Simple Sequence Repeats (SSRs), or STRs (Guichoux et al., 2011). Table (1) illustrates the types of human repeated DNA sequence. STRs loci are the most 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 with between 5 and 20 length variants (alleles) on average per locus, which makes these STRs effective for human identification purposes.

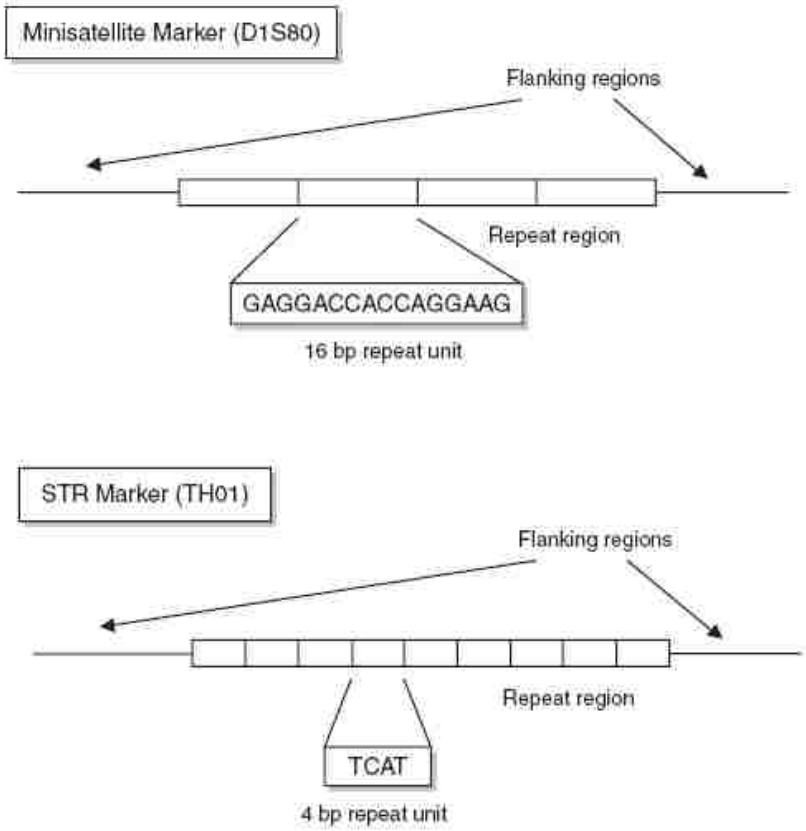


Figure 3: Minisatellite and Microsatellite (STR) DNA markers
(Halima, Bernal, & Sharif, 2009)

Table 1:Types of human repeated DNA sequences

Type of DNA	Description	Length
Satellite DNA	Regions of long stretches of repeated DNA	< 1000
Minisatellite DNA	Regions of medium stretches of repeated DNA	10-100
Microsatellite DNA	Regions of small stretches of repeated DNA	2-7

There are specific criteria that used for the selection of STR loci in forensic purposes and human identification applications: firstly, high discriminating power; secondly, robustness and reproducibility of results when multiplexed with other loci; thirdly, separate chromosomal locations to ensure that closely linked loci are not chosen; fourthly, STR alleles that have lower mutation rates; and finally small size of STR alleles make STR loci better candidates in human identifications for use in forensic applications, because DNA in such samples are either degraded or mixed (M. Ahmed & Zeyad, 2017).

1.3.1 Types of STR loci

STR repeat sequences are named by the length of the repeat unit. STR loci vary in the length of the repeat unit and the number of repeats and in the repeat pattern. Some DNA sequences have two nucleotides repeated next to each other over and over again that type of sequences called Dinucleotide repeats. Other sequences have three nucleotides in the repeat unit that type of sequences called trinucleotides repeats, tetranucleotides have four repeats, pentanucleotides have five repeats, and hexanucleotides have six repeat units in the core repeat (Lance et al., 2013) (Table 2).

Table 2: Types and examples of STR repeat units

Length of Repeat Unit	Example
Dinucleotide	(CA) (CA) (CA) (CA)
Trinucleotide	(GCC) (GCC) (GCC)
Tetranucleotide	(ACTG) (ACTG) (ACTG)
Pentanucleotide	(ATATG) (ATATG) (ATATG)
Hexanucleotide	(ATATGA) (ATATGA) (ATATGA)

1.3.2 STR markers Categories

STRs that are usually used have either 4 bp or 5 bp core repeat motif and often divided into several categories based on the repeat pattern, those are simple repeats or simple repeats with non-consensus alleles, compound repeats, and complex repeats (John Marshall Butler, 2001) (Table 3). Not all alleles for an STR locus contain complete repeat units but there are alleles that contain incomplete repeat units called microvariants.

Some DNA sequences contain one unit of identical length (e.g. AGATAGATAGATAGAT). These sequences are called Simple repeats STRs and the examples of these include are CSF1PO, TH01 ,TPOX, D5S818, D7S820, D13S317, D16S539, D18S51, D10S1248, and D22S1045. There is another type of simple repeats called simple repeats with non-consensus alleles which is alleles with incomplete repeat units (e.g. AGATAGATAGA). An example of this repeat is HUMTH01 .

Some DNA sequences comprise two or more adjacent simple repeats (e.g . TCTATCTG TCTATCTG). These sequences called Compound repeats STRs and the

examples of these types are FGA, VWA ,D3S1358, D8S1179, D2S1338, D19S433, D12S391, and D6S1043.

Complex repeats may contain several repeat blocks of variable unit length as well as variable intervening sequences. Example of repeats is D21S11.

Table 3: Categories for STR markers

Category	Example of Repeat Structure	Example of Loci
Simple repeats	(GCCA) (GCCA) (GCCA)	TPOX, CSF1PO, D5S818, D13S317, D16S539
Simple repeats with non-consensus alleles	(GCCA) (GCC-) (GCCA)	TH01, D18S51, D7S820
Compound repeats	(GCCA) (GCCA) (GCTA)	VWA, FGA, D3S1358, D8S1179
Complex repeats	(GCCA) (GCTA) (CA)(CACA)	D21S11

1.4 Multiplex STR typing

1.4.1 First generation multiplex STR typing

Quadruplex amplification kit considered the first multiplex STR typing system. It was created by the Forensic Science Service (FSS) in the United Kingdom and based on 4 STR loci. These loci include TH01, FES, vWA, and F13A1. Quadruplex had a matching probability of approximately 1 in 10,000 (Budowle, Moretti, Keys, Koons, & Smerick, 1997).

1.4.2 Second generation multiplex STR typing

In 1996, Forensic Science Service has developed also six-locus STR system combined with the Amelogenin sex test (Gill, 2002). This system, known as the “second-generation multiplex” (SGM) and consisted of two of the Quadruplex

minisatellite loci (TH01 and vWA) combined with the STR microsatellite loci D8S1179, D18S51, D21S11 and FGA which are highly polymorphic. Results from these loci provided a matching probability of approximately 1 in 50 million.

1.4.3 Expansion of the Second-generation multiplex system

In 1999, Applied Biosystems was expanded to 10 loci (Cotton et al., 2000). 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 hundreds of millions. The SGM plus system was fully accredited for use in forensic cases.

Nowadays Applied Biosystems, Promega Corporation and other companies have many different STR kits that help 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 in new STR Kits.

Today, the results can be obtained in few hours by using STR Kits compared to RFLP methods which takes weeks. The STR technology has developed quickly in the late 1990s for more sensitive, rapid, and accurate measurements of STR alleles (Gymrek, Golan, Rosset, & Erlich, 2012). At the same time, the number of STRs that can be simultaneously amplified has increased from three or four STR to over 15 STRs using multiple-color fluorescent tags. A list of commercially available STR multiplexes and companies are shown in Table 4.

Table 4: Some of Commerciale available STR multiplexes

Commerciale available STR multiplexes	Company
AmpFISTR COfiler	Applied Biosystems
AmpFISTR Identifiler	Applied Biosystems
AmpFISTR GlobalFiler	Applied Biosystems
Investigator IDplex Plus Kit	QIAGEN
Investigator HDplex Kit	QIAGEN
PowerPlex 16	Promega Corporation
PowerPlex Fusion	Promega Corporation

1.4.4 Multiplex STR typing studies in different populations

Many Studies have employed STR multiplex PCR using various types of STRs. For example, allele 18 in SE33 locus represented the most polymorphic allele described in Saudi population (Khubrani, Wetton, & Jobling, 2019) , whereas Penta E showed the most discriminating locus between Iraqi population (Farhan, Hadi, Iyengar, & Goodwin, 2016).

1.5 STR Genotyping

1.5.1 FTA Cards

Most of DNA extraction techniques for blood or any biological material are time consuming and require advanced laboratory equipment. Another approach to DNA extraction involves the use of FTA cards. In 1980, FTA card was discovered by Lee Burgoyne at Flinders University in Australia (Sulistyawati, 2010). He developed a special method for storage of DNA. FTA cards as in Figure (4) contain chemical

components which lyze cells on contact, denature proteins, and protect DNA from degradation. One of the advantages of FTA cards is ability to inactivate organisms including pathogens and prevent the growth of bacteria and other microorganisms so the nucleic acids collected on FTA Cards remain stable for years at room temperature.

Figure 4: FTA Card (H. A. Ahmed, MacLeod, Hide, Welburn, & Picozzi, 2011)

1.5.2 Electrophoresis

Electrophoresis is the process of DNA fragment separation based on the differential migration features of charged molecules in an electric field. In the electrophoresis, two electrodes are immersed in two buffer chambers (anode which is positively charged, and cathode which is negatively charged) (Dunbar, 2012). By using a power supply, electric potential difference is generated between the two electrodes. As a result, electrons flow from one of the electrodes, the anode, towards the other electrode, the cathode. The electrophoretic mobility of the sample can be affected by the net charge, molecular weight of the molecules, shape, size, the buffer type, and the voltage power.

1.5.3 Capillary Electrophoresis

Capillary electrophoresis (CE) is a powerful separation method. This type of separation becomes increasingly prominent in bioanalytical research, e.g., in genetic, biotechnology, diagnostic, and forensic applications. In forensic field, Capillary electrophoresis considered as a principal method to detect genetic variation. Separation and detection of short tandem repeat (STR) alleles can be achieved rapidly and reliably, providing sufficient resolution for genotyping the large range of markers currently utilized for forensic casework and research (Pascali, Bortolotti, & Tagliaro, 2012).

The main components of a basic capillary electrophoresis instrument include a narrow glass capillary, two buffer vials, and two electrodes connected to a high-voltage power supply. CE systems also contain a laser excitation source, a fluorescence detector, an autosampler to hold the sample tubes or tray, and a computer to control the sample injection and detection.

Capillary electrophoresis (CE) was used to achieve reliable STR typing resolve the

DNA fragments. During capillary electrophoresis the sample or PCR products enter to capillary by applied voltage to induce the sample to migrate into the capillary which forced the negatively charged fragments into the capillaries. DNA fragments or PCR products can be detected using fluorescent intercalating dyes. These dyes bind to the DNA molecule by binding to the DNA helix, affecting the configuration of the aromatic rings of the dye and producing a large enhancement of the fluorescence signal. The fluorescently labeled DNA fragments are separated based on the size before reaching the positive electrode and move across the path of a laser beam. When

the separated labeled DNA fragments reach to the laser beam, the dyes attached to the fragments to fluoresce. The diffraction system separates the dye signals and a Charge Coupled Device (CCD) camera detects the fluorescence. Each dye has different wavelength of the emitted light (Figure 5). The data collection software collects the fluorescence signal and convert it into an electropherogram (St. Claire, 1996).

Capillary electrophoresis (CE) provides a number of advantages for analysts: Firstly, the all three steps which include injection, separation, and detection are fully automated. Secondly, injection process consumed only very small amount of DNA samples so, the researcher can retest same samples if needed which will be more useful in forensic samples that often cannot be easily replaced. Thirdly, Time consuming in separation step in capillaries is much less than that in slab gel due to higher voltages that are permitted with improved heat dissipation from capillaries. Fourthly, data can be obtained and storage following the completion of the run from data collection software. Unlike gel no further steps are needed such as scanning the gel or taking a picture and no need for lane tracking. (John M Butler, 1998).

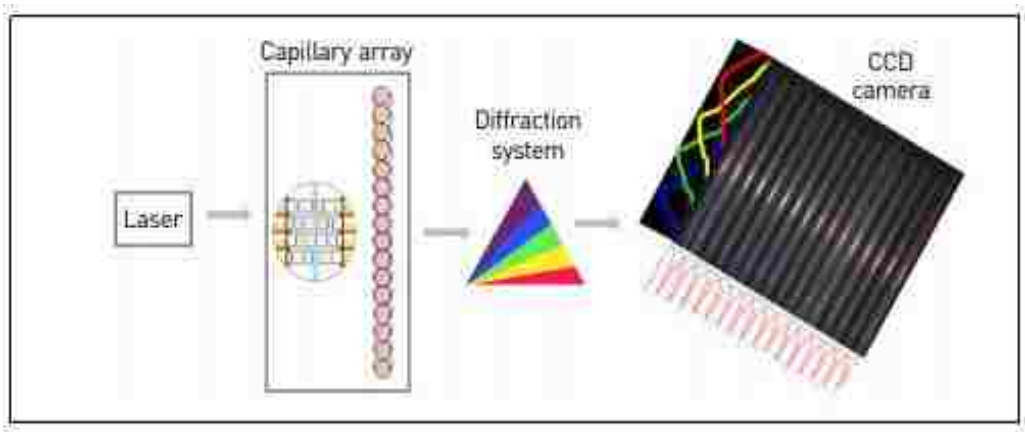


Figure 5: Capillary electrophoresis process (John M Butler, 1998)

The 3500 Genetic Analyzers (Applied Biosystems) is the first genetic analyzer designed with a specific feature set and workflow for human identification (HID) applications. It is a fluorescence-based DNA analysis instrument using capillary electrophoresis technology with 8 capillaries each 36 cm in length. It is automated systems capable of sequencing DNA or analyzing fragments for a variety of applications.

1.6 Importance of the study

The significance of this study is to build an allelic frequency database for the latest and most powerful amplification kit using current forensic workflow aiding statistical evaluation of generated STR profiles in the corresponding populations. This study is very important, and it contributes to other studies in the Gulf region to study the genetic diversity of populations

1.7 Hypothesis and Objectives

I hypothesize that 23 autosomal STR markers are highly discriminatory in the UAE population.

Objectives

1- To study the autosomal chromosomes loci polymorphisms and estimate alleles frequencies of 23 STR loci in UAE Arab population.

2- To assess the importance of increasing the number of STR loci utilized in forensic DNA analysis and to compare the results with neighborhood population's studies.

Chapter 2: Materials and Methods

2.1 Sample collection

In this study, blood samples were collected from 571 males and females unrelated individuals from UAE Arab population. Samples were collected and provided via Dubai Health Authority (DHA). Samples were treated as anonymous for the purpose of this research. No information existed that could have linked a specific sample to a specific individual. This helping to ensure the credibility, integrity and objectivity of the results obtained.

Samples were protected from contamination and deterioration before their arrival in the laboratory by using only recommended sample containers. Samples were processed and analyzed according to specific standard operating procedures. All possible measures and procedures were taken to prevent contamination.

2.2 Ethical considerations

Ethical approval has been obtained from Dubai scientific research ethics committee in (DHA) under the reference number (DSREC-SR-08/2018-03).

2.3 Sample preparation

One drop from each sample tube was applied to separate Whatman® FTA cards (GE Healthcare). Each card was assigned a unique barcode number and left to dry.

2.4 Storage conditions

FTA cards that contains the blood samples were stored at room temperature while whole blood samples tubes received from DHA were stored at -20°C.

VeriFiler™ Express PCR Amplification Kit (Thermofisher Scientific) was used to analyze 23 autosomal STR loci, 1 insertion/deletion polymorphic marker on the Y chromosome (Y indel) and Amelogenin (sex-determining marker) in all DNA samples. The kit contents were stored as shown in Table 5. Primer set, amplified DNA, allelic ladder, and size standard was protected from light when not in use. Gene Scan "600 LIZ" Size Standard was stored in 2°C to 8°C. Hi−Di Formamide was stored in –15°C to -25°C.

Table 5: Kit content and storage conditions (According to manufacturer protocol)

Kit contents	Storage conditions
VeriFiler™ Express Master Mix Contains $MgCl_2$, dATP, dGTP, dCTP, and dTTP, bovine serum albumin, enzyme, and 0.05% sodium azide in buffer and salt.	–25°C to –15°C on receipt. 2°C to 8°C after first use.
VeriFiler™ Express Primer Set Contains locus-specific 6-FAM™, VIC™, NED™, TAZ™, and SID™ dye-labeled and unlabeled primers in buffer. The primers amplify the STR loci D3S1358, vWA, D16S539, CSF1PO,TPOX, D8S1179, D21S11, D18S51, Penta E, D2S441, D19S433, TH01, FGA, D22S1045, D5S818, D13S317, D7S820, D6S1043, D10S1248, D1S1656, D12S391, D2S1338, Penta D, and the sex-determining markers, Y indel and Amelogenin.	–25°C to –15°C on receipt. 2°C to 8°C after first use. Store protected from light.
VeriFiler™ Express Allelic Ladder Contains the following amplified alleles: • 6-FAM™ dye (blue): D3S1358 9–20; vWA 11–24; D16S539 5, 8–15; CSF1PO 6–15; TPOX 5–15. • VIC™ dye (green): Y indel 1, 2; Amelogenin X, Y; D8S1179 5-19; D21S11 24, 24.2, 25–28, 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; D18S51 7, 9, 10, 10.2, 11–13, 13.2, 14, 14.2, 15–27; DYS391 7–13; Penta E 5–26. • NED™ dye (yellow): D2S441 8–11, 11.3, 12–17; D19S433 6–12, 12.2, 13, 13.2, 14, 14.2, 15, 15.2, 16, 16.2, 17, 17.2, 18.2, 19.2; TH01 4–9, 9.3, 10, 11, 13.3; FGA 13–26, 26.2, 27–30, 30.2, 31.2, 32.2, 33.2, 42.2, 43.2, 44.2, 45.2, 46.2, 47.2, 48.2, 50.2, 51.2. • TAZ™ dye (red): D22S1045 8–19; D5S818 7–18; D13S317 5– 16; D7S820 6–15; D6S1043 9–25. • SID™ dye (purple): D10S1248 8–19; D1S1656 9–14, 14.3, 15, 15.3, 16, 16.3, 17, 17.3, 18.3, 19.3, 20.3; D12S391 14–19, 19.3, 20–27; D2S1338 11–28; PentaD 2.2, 3.2,5–17.	–25°C to –15°C on receipt. 2°C to 8°C after first use. Store protected from light. IMPORTANT! The allelic ladder contains PCR products. Do not amplify. To avoid contamination, store the allelic ladder separate from the other kit components and unamplified DNA.
DNA Control 007 Contains 2 ng/µL of human male genomic DNA in 0.05% sodium azide and buffer.	–25°C to –15°C on receipt. 2°C to 8°C after first use.

2.5 Sample preparation for PCR amplification

An FTA card disc sized (0.5 mm diameter) was punched using micro-puncher (Harries®) and placed in 0.2ml PCR tube. The disc then was ready to proceed to PCR amplification step without washing process because VeriFiler™ Express PCR Amplification Kit (Thermofisher Scientific) has been designed to resist some molecules such as hematein or other cell components which are considered PCR inhibitors. The quantification step was skipped in this process because the amount of DNA in the disc is sufficient for PCR and it is optimized as per the constant size of the disc, so this consider one of the major advantages of using FTA card and this step help forensic laboratory to complete DNA typing very quickly (Figure 6).

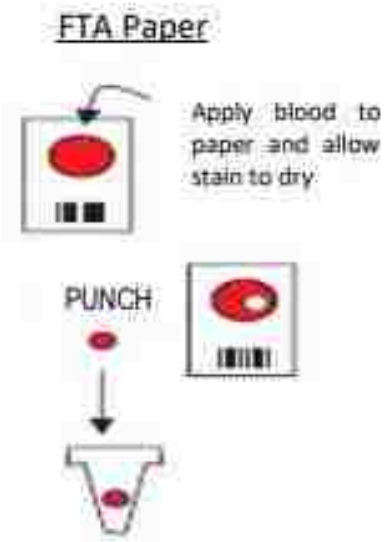


Figure 6: DNA extraction process (M. Ahmed & Zeyad, 2017)

2.6 PCR Amplification Step

2.6.1 Kit overview

The VeriFiler™ Express PCR Amplification Kit also includes: 23 autosomal STR loci and VeriFiler™ Express Allelic Ladder as shown in Table (6).

Table 6: Some of VeriFiler™ Express PCR Amplification Kit loci and alleles (The table is adopted from the manufacture protocol

Locus	Alleles range included in Allelic Ladder	Dye Label	DNA Control 007
D3S1358	9 to 20	6-FAM	15,16
vWA	11 to 24		14,16
D16S539	5 to 15		9,10
CSF1PO	6 to 15		11,12
D8S1179	5 to 19	VIC	12,13
D21S11	24 to 38		28,31
D18S51	7 to 27		12,15
Penta E	5 to 26		7,12
D2S441	8 to 17	NED	14,15
D19S433	6 to 19.2		14,15
TH01	4 to 13.3		7,9.3
D22S1045	8 to 19	TAZ	11,16
D5S818	7 to 18		11,11
D13S317	5 to 16		11,11
D7S820	6 to 15		7,12
D10S1248	8 to 19	SID	12,15
D1S1656	9 to 20.3		13,16
D2S1338	11 to 28		20,23
Penta D	2.2 to 17		11,12

2.6.2 Multiplex PCR using VeriFiler™ Express PCR Amplification Kit

Amplification of the 23 loci was performed according to manufacturer's protocol with some modification. Quarter volume has been validated and used due to high number of samples needed to be tested and high cost of the kit). The total number of samples in this study were 571 samples. The reaction mix was prepared by mixing the master mix, primer set, and distilled water. The total amount of the reaction mix solution for each sample was 6.25 µL (Table 7).

Table 7: kit component quantity for each sample

DNA sample	Volume per reaction (µL) (full volume)	Quarter of the quantity (µL)
VeriFiler Express Master Mix	10.0	2.5
VeriFiler Express Primer Set	10.0	2.5
Distilled Water	5.0	1.25
Total Volume	25.0	6.25

First step in reaction preparation mix was to thaw the VeriFiler™ Express master mix and primer set, then vortexed each tube for 3 seconds followed by quick spin down step. Thawing was required only during first use of the kit. After first use, reagents were stored at 2 to 8°C and, therefore, they do not require subsequent thawing.

Second step is to calculate and prepare the required volumes of components according to the number of samples in addition to positive, negative controls and some extra samples for pipetting errors, then the master mix was prepared accordingly in a

sterilised 2.0 mL microcentrifuge Eppendorf tube. The reaction mix was vortexed for 3 seconds, then a spinning-down step was conducted.

2.6.3 FTA Punch procedure

6.25 µL reaction mix was dispensed into the PCR tube before adding the punches to prevent discs jumping out of the tube due to static effect. The final reaction volume in the positive control tube was 6.75 µL because 0.5 µL of the DNA control 007 was added to the positive control tube.

0.5 mm disc of the FTA paper of each sample was punched by using micro-puncher (Harries®) and placed into a labeled PCR tubes which were preloaded by 6.25µL of the reaction mix. 0.5 mm punch of blank FTA paper was added to the negative control (Table 8). The reaction mix was vortexed for 3 seconds and spined down to remove bubbles. The samples were amplified in a GeneAmp PCR System 9700 (Applied Biosystems).

Table 8: Size of the disc for test samples and negative control and the amount of positive control

DNA sample	To prepare
Negative control	0.5 mm blank disc
Test Sample	0.5 mm sample disc
Positive control	Add 0.5 µL of Control DNA 007

2.6.4 Thermal Cycling Parameters for VeriFiler™ Express PCR Amplification Kit

Amplification was performed on GeneAmp PCR System 9700 (Applied Biosystems), using the following protocol: initial 1 minute incubation at 95°C for the activation of DNA polymerase, followed by 26 cycles of denaturation at 94°C for 3 seconds, annealing and extension at 59°C for 3 seconds per cycle. The last cycle was followed by a final extension step at 60°C for 5 minutes to promote complete non-template 3'- nucleotide addition as shown in (Table 9).

Table 9: Condition of Thermal Cycling Parameters VeriFiler™ Express PCR

Initial incubation step	Optimum cycle number		Final extension	Final hold
	Denature	Anneal/Extend		
HOLD	26 CYCLE		HOLD	HOLD
95°C, 1 minute	94°C, 3 seconds	59°C, 16 seconds	60°C, 5 minutes	4°C, up to 24 hours

2.7 Separation and detection methods of VeriFiler™ Express PCR Amplification Kit Product

Amplifying STR alleles using PCR produce a mixture of DNA fragments called amplicons. This mixture need to be separated and resolved from one another by electrophoresis which was used for fragments separation and either performed in a slab-gel or capillary environment.

2.7.1 Capillary electrophoresis

Amplified STR alleles were run on ABI 3500 Genetic Analyzer according to VeriFiler™ Express PCR Amplification Kit protocol. There are three reagents required for this step including Hi-Di Formamide and LIZ600 size standard which were mixed with the samples and the allelic Ladder. Internal Lane Size Standard 600 that was provided in the kit was included with every sample to allow automatic sizing of alleles.

1. Hi-Di formamide

Hi-Di Formamide is required in CE. It is an injection solvent in DNA analysis methods on the Genetic Analyzers and its main function is stabilizing denatured DNA samples for fluorescence detection on Genetic Analyzer. It helps to stop reannealing of the DNA single strands in DNA sample strands after denaturation at a high temperature to eliminate the formation of secondary structure of nucleic acid (M. Ahmed & Zeyad, 2017).

2. GeneScan 600 LIZ Size Standard

The GeneScan™ 600 LIZ™ Size Standard is an internal lane size standard developed for use with fluorescence-based DNA electrophoresis systems. GeneScan size standards are fluorescently labeled DNA fragments of known sizes and it has different dye color from the PCR products as shown in (Table 10). Fragments of known sizes can be used by the data analysis software to generate a calibration curve to determine the size of unknown fragments in each sample (Shewale et al., 2013). The GeneScan 600 LIZ Size standard comprising of 36 single stranded labeled fragments which are: 20, 40, 60, 80, 100, 114, 120, 140, 160, 180, 200, 214, 220, 240, 250, 260, 280, 300, 314, 320, 340, 360, 380, 400, 414, 420, 440, 460, 480, 500, 514, 520, 540,

560, 580, and 600 nucleotides. Each kit contains two-200 µL tubes of the size standard. This is sufficient for 800 analyses when using the recommended loading amount of 0.5 µL. The kit was stored at 2°C to 8°C.

Table 10: Dyes used in the VeriFiler™ Express kit

Dye	Color	Label
6-FAM	Blue	Samples, allelic ladders and controls
VIC	Green	
NED	Yellow	
TAZ	Red	
SID	Purple	
LIZ	Orange	GeneScan™ 600 LIZ™ Size Standard

3. Veifiler™ Express Allelic Ladder

The ladder consists of DNA fragments that represent common alleles for all loci. So, it represents all the possible alleles for each locus. Allelic Ladder in conjunction with LIZ600 size standard allows verification of unknown STR lengths.

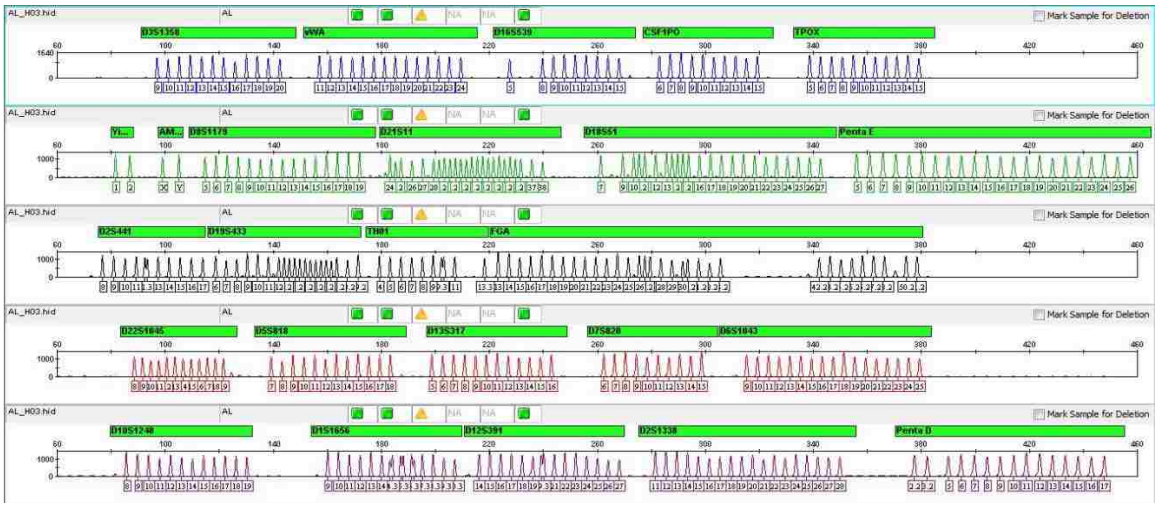


Figure 7: GeneMapper ID X Software plot of the VeriFiler™ Express Allelic Ladder Allelic Ladder (The figure is adopted from manufacture protocol)

Prior to loading the samples into the ABI 3500 Genetic Analyzer, formamide/size standard mixture was prepared by mixing the size standard and Hi-Di Formamide as follows for each sample:

- 0.5µL of LIZ-600 Size Standard

- 9.5µL of HiDi-Formamide

Then the Formamide/size standard mixture was mixed for 10–15 seconds with a vortex mixer and 10µl of the Formamide/size standard mixture was then loaded into each well in MicroAmp optical 96-well plate. 0.5 µL of PCR product was added to each well. 0.5 µL of allelic ladder was added to ladder well. A minimum of 1 ladder sample was loaded per plate. 0.5 µL of positive control sample (007) was added to to the plate where a minimum of 1 positive control sample was presented per plate. 0.5 µL of negative control was added into the negative control well where a minimum of 1 negative control sample was presented per plate.

Plate was covered with a septa, then briefly vortexed and spin down to remove air bubbles. After that plate was loaded into the 9700 GeneAmp PCR System for denaturation for 5 minutes at 95°C and cooled 2 minutes at 4°C and loaded the plate immediately into the instrument.

2.8 Statistical analysis

GeneMapper ID-X version 1.4 analysis software is a powerful data analysis tool. It is an automated genotyping software application for forensic casework, databasing, and paternity data analysis. It was used to analyze the data generated from the capillary electrophoresis platforms such as 3500 Genetic analyzer (John M Butler, 2014).

In this study, this software was required to analyze data that was generated using the VeriFiler™ Express kit. The raw data file extracted from the GeneMapper ID-X was analyzed using Microsoft Excel and an integrated software package for population genetics data analyses (Arlequin version 3.5) and forensic statistics analysis toolbox (FORSTAT). FORSTAT software program was used to calculate the heterozygosity, homozygosity, power of discrimination (PD), probability of exclusion (PE). Arlequin software program was used to calculate the Genetic Diversity, Hardy Weinberg Equilibrium (HWE) as well as F-statistics of population differentiation .(Schneider, Roessli, & Excoffier, 2000).

2.8.1 Match Probability

Match Probability defined as the probability of a match between two unrelated individuals and it was calculated the following formula (Fisher, 1951), where xi is the frequency of the genotype i at a given locus in the population.

$$MP = \sum_{i=1}^{n} xi^2$$

2.8.2 Power of discrimination

Power of discrimination (PD) indicates the probability that two randomly unrelated selected individuals in a population will have different genotypes for a given locus and it was calculated using the following formula (Brenner, 1990).

$$PD = 1 - MP$$

2.8.3 Power of Exclusion

Power of exclusion (PE) is the probability of a given locus to exclude one individual and it was calculated using the following formula (Brenner, 1990).

$$PE = h^2(1 - 2hH^2)$$

where *h* is the proportion of heterozygous individuals and *H* the proportion of homozygous individuals in the population sample.

2.8.4 Genetic Diversity

Genetic Diversity is the probability that two alleles drawn at random from the population will be different and it was calculated using the following formula (Nei, 1987).

$$GD = \frac{n}{n-1}(1 - \sum_i p^2)$$

where *n* is the number of gene copies sampled and *pi* is the frequency of the i allele in the population.

2.8.5 Hardy-Weinberg equilibrium

The Hardy-Weinberg equilibrium (HWE) predicts how gene frequencies will be inherited from generation to generation and it was calculated using the following formula (John M Butler, 2005).

$$p^2 + 2pq + q^2$$

where p2 represents the frequency of the homozygous genotype AA, q2 represents the frequency of the homozygous genotype aa, and 2pq represents the frequency of the heterozygous genotype Aa.

Chapter 3: Results

3.1 profiling samples

571 samples that obtained from unrelated individuals of UAE population were profiled with VeriFiler™ Express PCR Amplification Kit for 23 autosomal STR markers. An example of the Autosomal STR profile result is shows in the Figure 8.

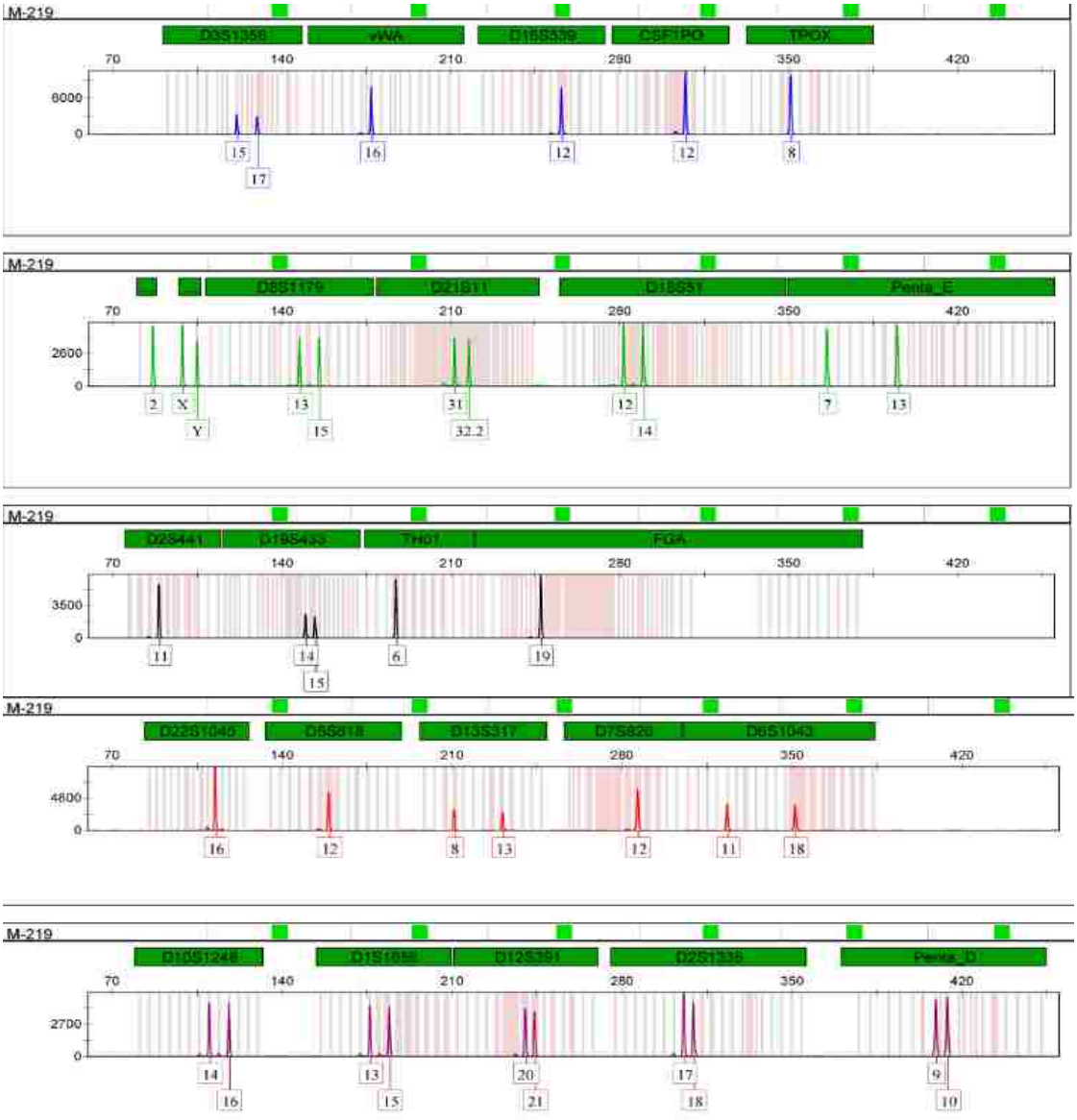


Figure 8: Autosomal STR profile generated using VeriFiler™ Express kit

This profile shows the different dyes that correspond to the different loci. Every profile shown that there are 25 loci in each sample, because there is one insertion/deletion polymorphic marker on the Y chromosome (Y indel) and Amelogenin (sex-determining marker) in all DNA samples, number of total alleles detected using this kit was brought up to 46 alleles in each sample for 23 autosomal loci. Numbers which below each peak indicate the number of repeats for each allele.

3.2 Alleles Frequencies

Table 11 and Figure 9 (A-W): Alleles frequencies of different loci with bar charts for UAE Arabs population.

A. Locus D3S1358

Table 11-A: Alleles for the locus D3S1358

Alleles No.	Allele repeats	Freq.	Percentage
1	13	2	0.2
2	14	61	5.3
3	15	319	27.9
4	16	339	29.7
5	17	258	22.6
6	18	152	13.3
7	19	10	0.9
8	20	1	0.1
Total		571	100

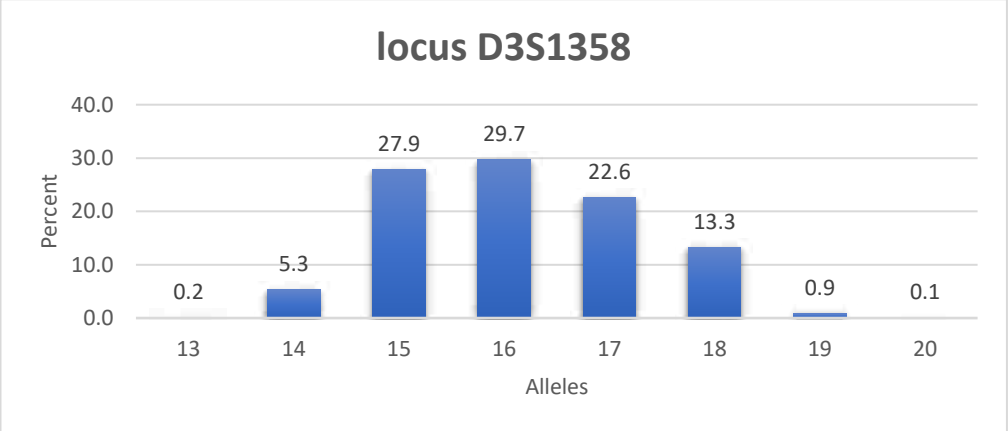


Figure 9-A: Alleles distrubtion for the locus D3S1358

Locus D3S1358 have 8 different numbers of alleles in 571 UAE ppulation. Allele 16 showed the highest frequency with 29.7% wheras allele 20 represented the lowest frequency with 0.1%.

B. Locus vWA

Table 11-B: Alleles for the locus vWA

Alleles No.	Allele repeats	Freq.	Percentage
1	13	3.0	0.3
2	14	80.0	7.0
3	15	153.0	13.4
4	16	266.0	23.3
5	17	332.0	29.1
6	18	230.0	20.1
7	19	67.0	5.9
8	20	10.0	0.9
9	21	1.0	0.1
Total		571	100

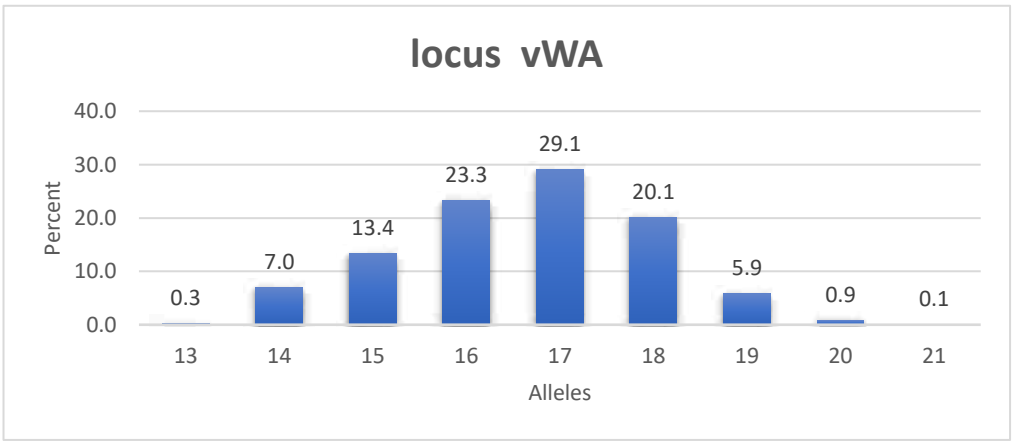


Figure 9-B: Alleles distrubtion for the locus VWA

C. Locus D16S539

Table 11-C: Alleles for the locus D16S539

Alleles No.	Allele repeats	Freq.	Percentage
1	8	41.0	3.6
2	9	175.0	15.3
3	10	98.0	8.6
4	11	402.0	35.2
5	12	277.0	24.3
6	13	136.0	11.9
7	14	13.0	1.1
Total		571	100

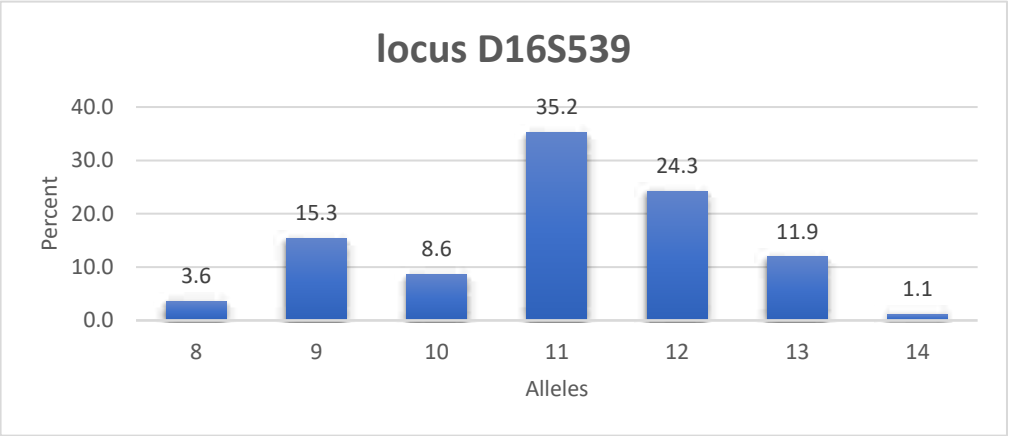


Figure 9-C: Alleles distrubtion for the locus D16S539

D. Locus CSF1PO

Table 11-D: Alleles for the locus CSF1PO

Alleles No.	Allele repeats	Freq.	Percentage
1	7	2.0	0.2
2	8	7.0	0.6
3	9	21.0	1.8
4	10	296.0	25.9
5	11	330.0	28.9
6	12	416.0	36.4
7	13	64.0	5.6
8	14	4.0	0.4
9	15	2.0	0.2
Total		571	100

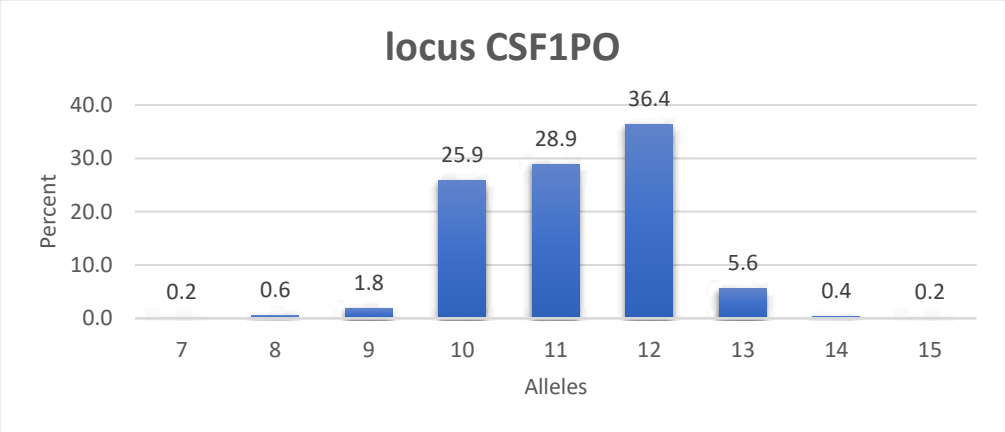


Figure 9-D: Alleles distrubtion for the locus CSF1PO

E. Locus TPOX

Table 11-E: Alleles for the locus TPOX

Alleles No.	Allele repeats	Freq.	Percentage
1	6	8.0	0.7
2	7	2.0	0.2
3	8	564.0	49.4
4	9	124.0	10.9
5	10	124.0	10.9
6	11	274.0	24.0
7	12	43.0	3.8
8	13	2.0	0.2
9	15	1.0	0.1
Total		571	100

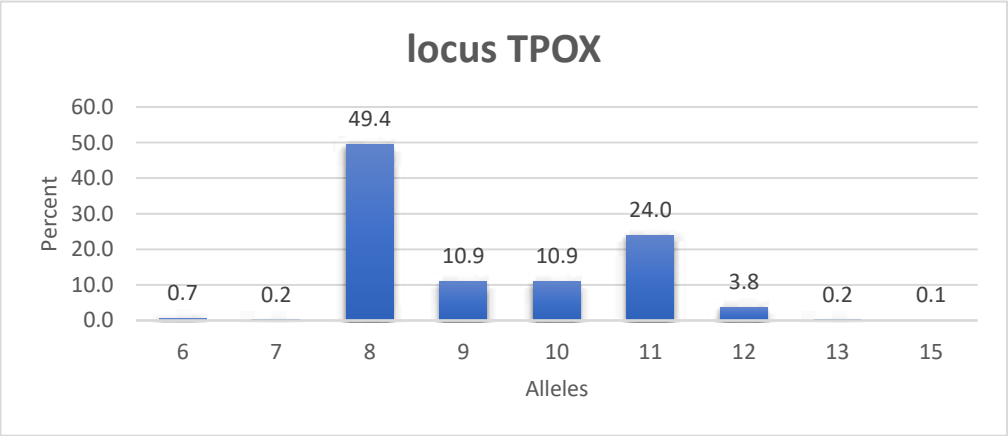


Figure 9-E: Alleles distrubtion for the locus TPOX

F. Locus D8S1179

Table 11-F: Alleles for the locus D8S1179

Alleles No.	Allele repeats	Freq.	Percentage
1	8	13.0	1.1
2	9	6.0	0.5
3	10	98.0	8.6
4	11	82.0	7.2
5	12	163.0	14.3
6	13	275.0	24.1
7	14	197.0	17.3
8	15	213.0	18.7
9	16	80.0	7.0
10	17	14.0	1.2
11	18	1.0	0.1
Total		571	100

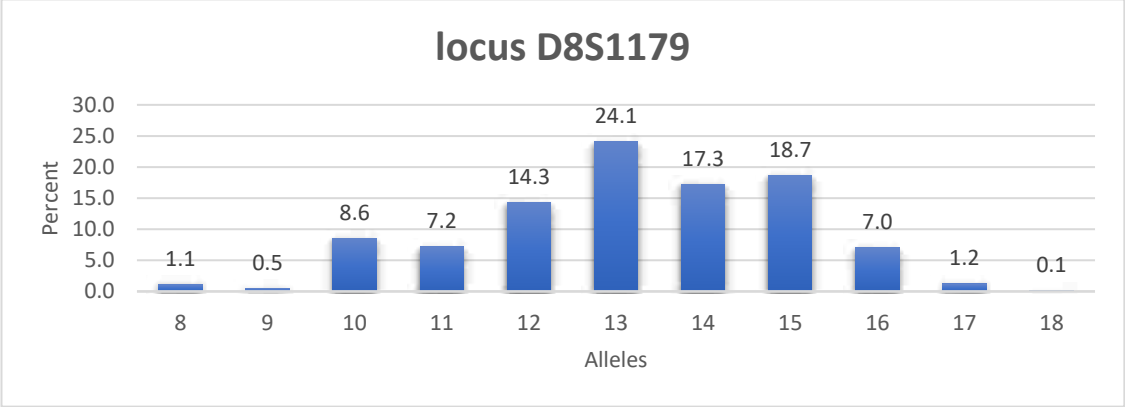


Figure 9-F: Alleles distrubtion for the locus D8S1179

G. Locus D21S11

Table 11-G: Alleles for the locus D21S11

Alleles No.	Allele repeats	Freq.	Percentage
1	27	20.0	1.8
2	28	180.0	15.8
3	29	265.0	23.2
4	29.3	3.0	0.3
5	30	253.0	22.2
6	30.2	28.0	2.5
7	31	51.0	4.5
8	31.2	125.0	10.9
9	32	8.0	0.7
10	32.2	134.0	11.7
11	33.2	51.0	4.5
12	34	2.0	0.2
13	34.2	12.0	1.1
14	35	7.0	0.6
15	36	2.0	0.2
16	37	1.0	0.1
Total		571	100

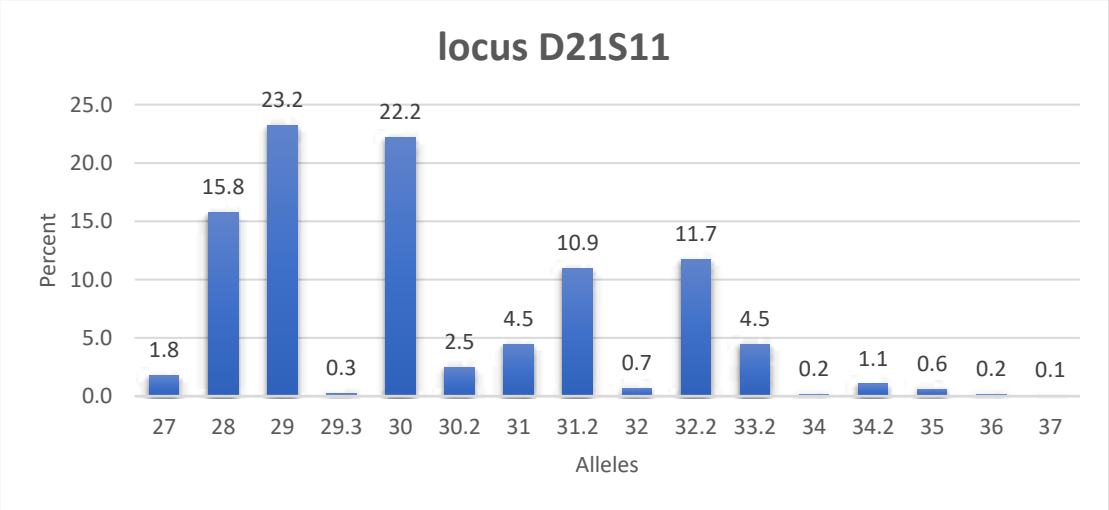


Figure 9-G: Alleles distrubtion for the locus D21S11

H. Locus D18S51

Table 11-H: Alleles for the locus D18S51

Alleles No.	Allele repeats	Freq.	Percentage
1	12	124.0	10.9
2	14	160.0	14.0
3	16	150.0	13.1
4	13	176.0	15.4
5	17	138.0	12.1
6	15	169.0	14.8
7	11	25.0	2.2
8	18	77.0	6.7
9	10	6.0	0.5
10	19	74.0	6.5
11	20	22.0	1.9
12	21	8.0	0.7
13	17.2	2.0	0.2
14	22	4.0	0.4
15	10.2	3.0	0.3
16	16.2	2.0	0.2
17	8	1.0	0.1
18	12.2	1.0	0.1
Total		571	100

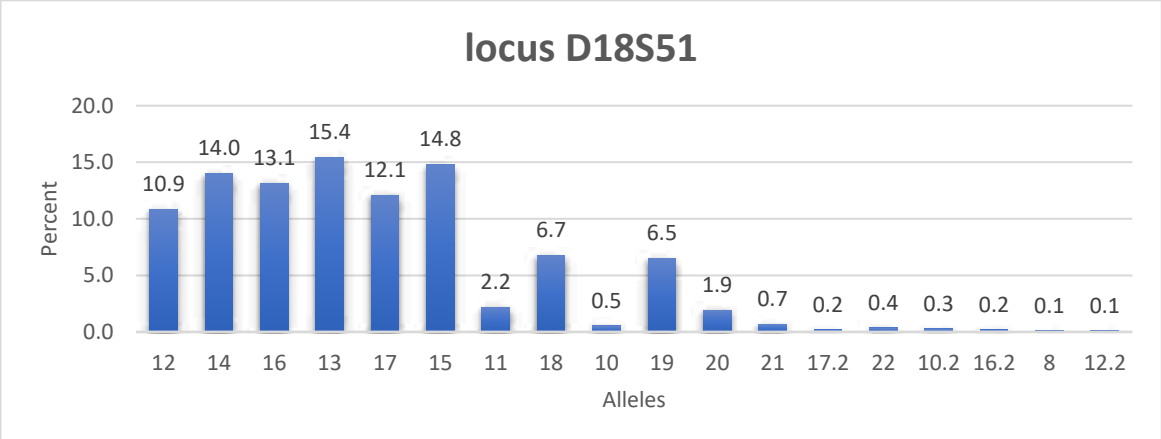


Figure 9-H: Alleles distrubtion for the locus D18S51

I. Locus Penta E

Table 11-I: Alleles for the locus Penta E

Alleles No.	Allele repeats	Freq.	Percentage
1	5	63.0	5.5
2	6	2.0	0.2
3	7	113.0	9.9
4	8	58.0	5.1
5	9	19.0	1.7
6	10	79.0	6.9
7	11	152.0	13.3
8	12	178.0	15.6
9	12.4	1.0	0.1
10	13	107.0	9.4
11	14	56.0	4.9
12	15	66.0	5.8
13	15.3	1.0	0.1
14	15.4	1.0	0.1
15	16	65.0	5.7
16	16.4	5.0	0.4
17	17	58.0	5.1
18	18	47.0	4.1
19	19	32.0	2.8
20	19.4	1.0	0.1
21	20	26.0	2.3
22	21	8.0	0.7
23	22	2.0	0.2
24	24	1.0	0.1
25	25	1.0	0.1
Total		571	100

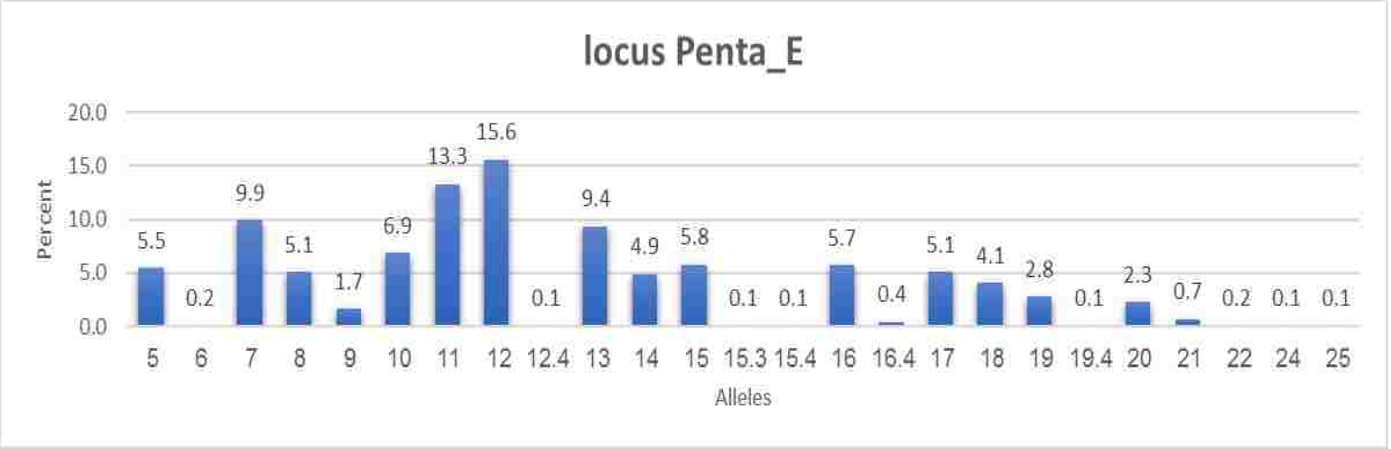


Figure 9-I: Alleles distrubtion for the locus Penta E

Penta E locus represented the largest number of different alleles (25 alleles). Allele 12 showed the highest frequency among other alleles with 15.6% and alleles 24 and 25 showed the lowest frequency with 0.1%.

J. Locus D2S441

Table 11-J: Alleles for the locus D2S441

Alleles No.	Allele repeats	Freq.	Percentage
1	9	4.0	0.4
2	9.1	1.0	0.1
3	10	195.0	17.1
4	11	440.0	38.5
5	11.1	1.0	0.1
6	11.3	55.0	4.8
7	12	96.0	8.4
8	12.3	2.0	0.2
9	13	37.0	3.2
10	14	255.0	22.3
11	15	54.0	4.7
12	16	2.0	0.2
Total		571	100

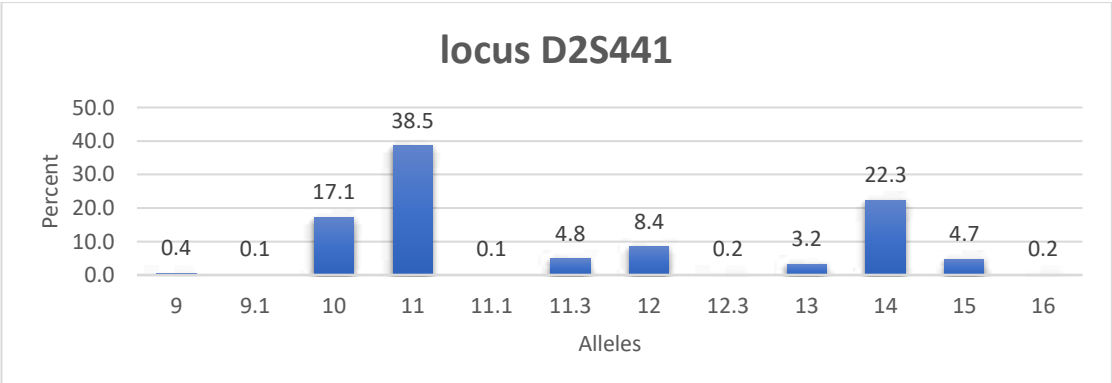


Figure 9-J: Alleles distrubtion for the locus D2S441

K. Locus D19S433

Table 11-K: Alleles for the locus D19S433

Alleles No.	Allele repeats	Freq.	Percentage
1	6.2	2.0	0.2
2	9	1.0	0.1
3	10	1.0	0.1
4	11	19.0	1.7
5	12	96.0	8.4
6	12.2	12.0	1.1
7	13	295.0	25.8
8	13.2	36.0	3.2
9	14	264.0	23.1
10	14.2	76.0	6.7
11	15	129.0	11.3
12	15.2	105.0	9.2
13	16	55.0	4.8
14	16.2	43.0	3.8
15	17	3.0	0.3
16	17.2	4.0	0.4
17	18.2	1.0	0.1
Total		571	100

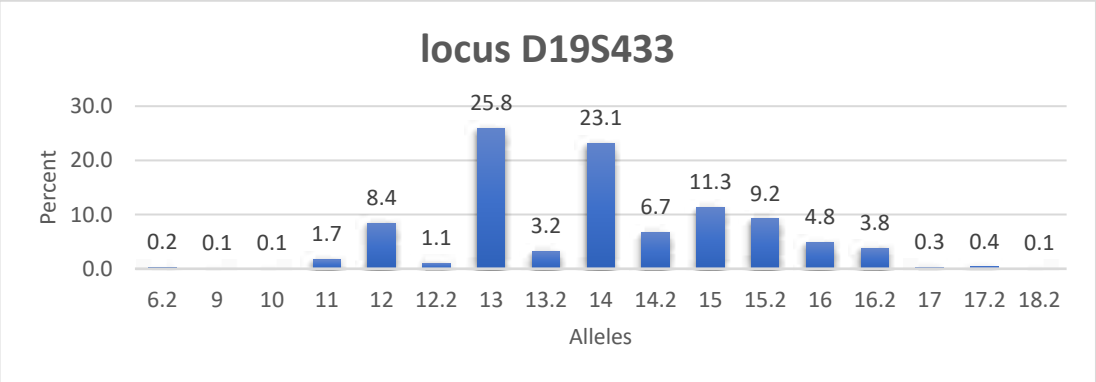


Figure 9-K: Alleles distrubtion for the locus D19S433

L. Locus TH01

Table 11-L: Alleles for the locus TH01

Alleles No.	Allele repeats	Freq.	Percentage
1	6	313.0	27.4
2	7	261.0	22.9
3	8	146.0	12.8
4	9	289.0	25.3
5	9.3	118.0	10.3
6	10	15.0	1.3
Total		571	100

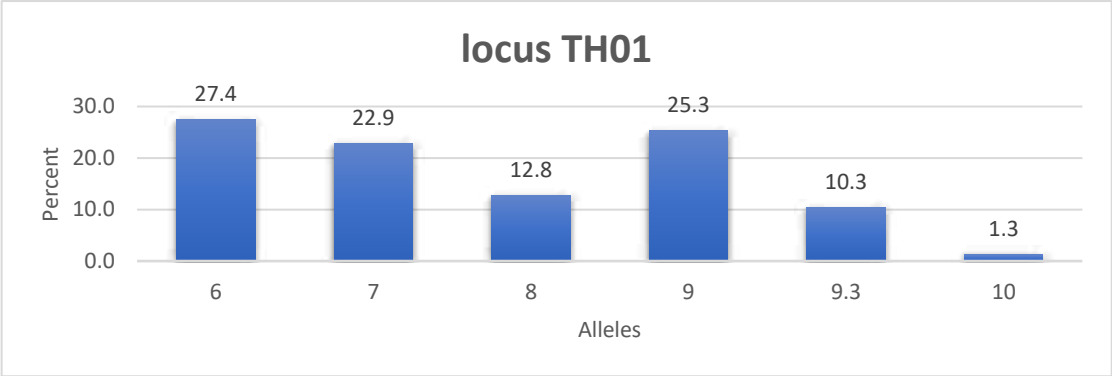


Figure 9-L: Alleles distrubtion for the locus TH01

TH01 locus represented the smallest number of different alleles (6 alleles) Allele 6 showed the highest frequency among other alleles with 27.4% whereas allele 10 showed the lowest frequency with 1.3%

M. Locus FGA

Table 11-M: Alleles for the locus FGA

Alleles No.	Allele repeats	Freq.	Percentage
1	16.1	3.0	0.3
2	17	5.0	0.4
3	18	10.0	0.9
4	18.2	2.0	0.2
5	19	61.0	5.3
6	19.2	1.0	0.1
7	20	109.0	9.5
8	20.2	2.0	0.2
9	21	142.0	12.4
10	21.2	7.0	0.6
11	22	175.0	15.3
12	22.2	11.0	1.0
13	23	216.0	18.9
14	23.2	7.0	0.6
15	23.3	1.0	0.1
16	24	213.0	18.7
17	24.2	4.0	0.4
18	25	96.0	8.4
19	26	48.0	4.2
20	27	16.0	1.4
21	28	7.0	0.6
22	29	3.0	0.3
23	30.2	2.0	0.2
24	31.2	1.0	0.1
Total		571	100

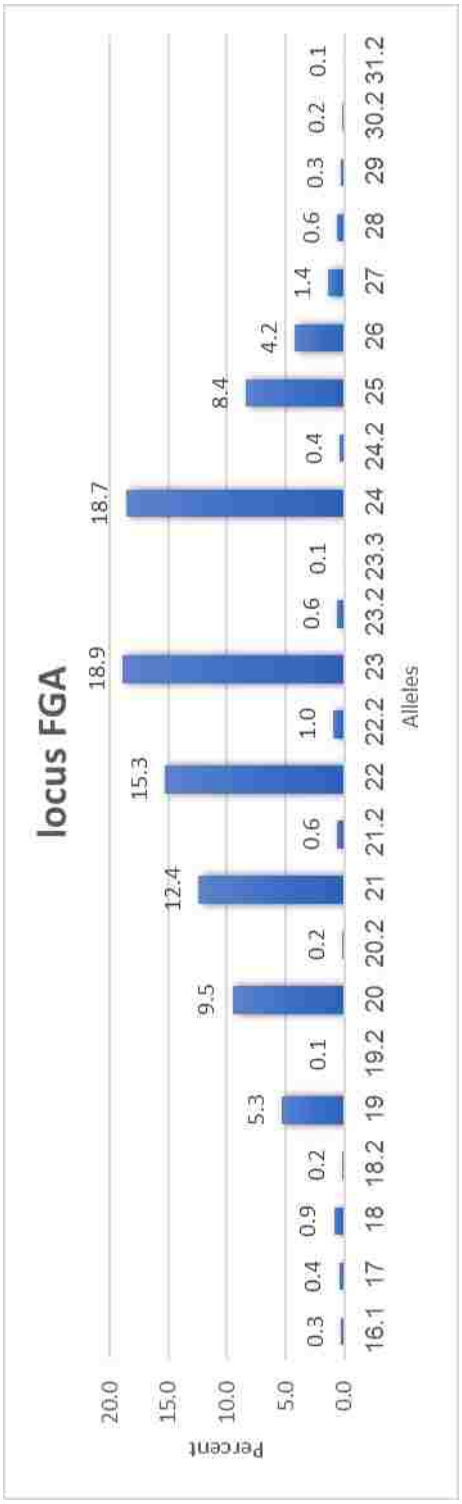


Figure 9-M: Alleles distrubtion for the locus FGA

N. Locus D22S1045

Table 11-N: Alleles for the locus D22S1045

Alleles No.	Allele repeats	Freq.	Percentage
1	10	11.0	1.0
2	11	166.0	14.5
3	11.2	2.0	0.2
4	12	16.0	1.4
5	13	6.0	0.5
6	14	77.0	6.7
7	15	501.0	43.9
8	16	283.0	24.8
9	17	73.0	6.4
10	18	7.0	0.6
Total		571	100

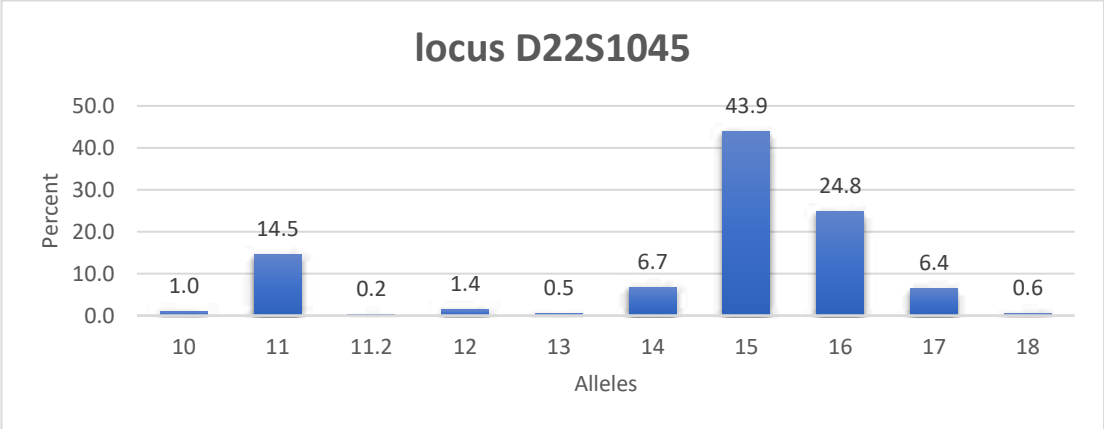


Figure 9-N: Alleles distrubtion for the locus D22S1045

O. Locus D5S818

Table 11-O: Alleles for the locus D5S818

Alleles No.	Allele repeats	Freq.	Percentage
1	8	17.0	1.5
2	9	57.0	5.0
3	10	120.0	10.5
4	11	311.0	27.2
5	12	433.0	37.9
6	13	195.0	17.1
7	14	7.0	0.6
8	15	1.0	0.1
9	16	1.0	0.1
Total		571	100

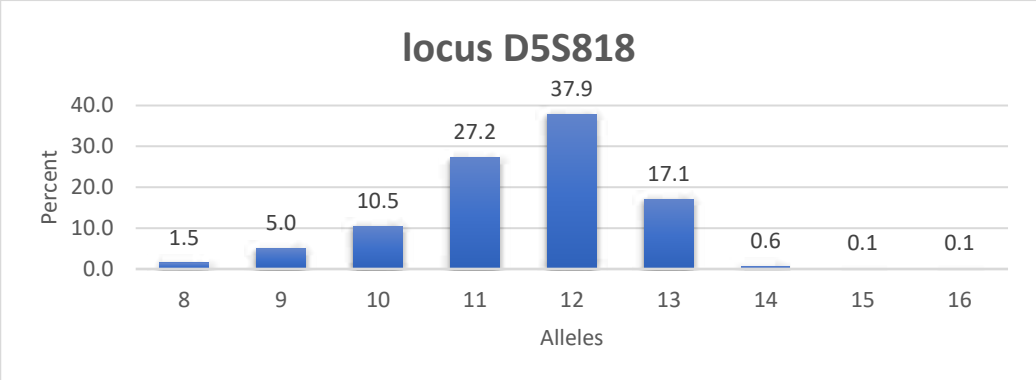


Figure 9-O: Alleles distrubtion for the locus D5S818

P. Locus D13S317

Table 11-P: Alleles for the locusD13S317

Alleles No.	Allele repeats	Freq.	Percentage
1	8	154.0	13.5
2	9	57.0	5.0
3	10	86.0	7.5
4	11	318.0	27.8
5	12	360.0	31.5
6	13	127.0	11.1
7	14	39.0	3.4
8	15	1.0	0.1
Total		571	100

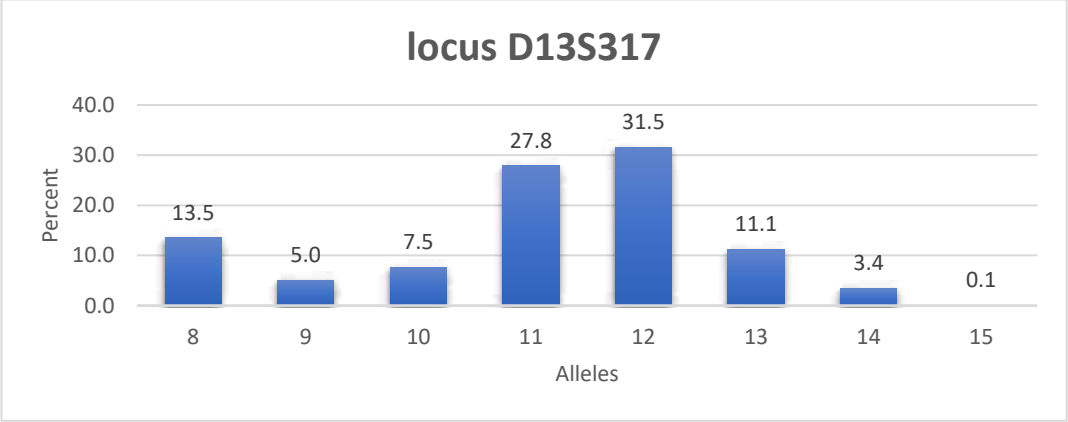


Figure 9-P: Alleles distrubtion for the locus D13S317

Q. Locus D7S820

Table 11-Q: Alleles for the locus D7S820

Alleles No.	Allele repeats	Freq.	Percentage
1	6	1.0	0.1
2	7	19.0	1.7
3	8	211.0	18.5
4	8.3	1.0	0.1
5	9	113.0	9.9
6	9.1	1.0	0.1
7	10	313.0	27.4
8	11	279.0	24.4
9	12	169.0	14.8
10	13	32.0	2.8
11	14	3.0	0.3
Total		571	100

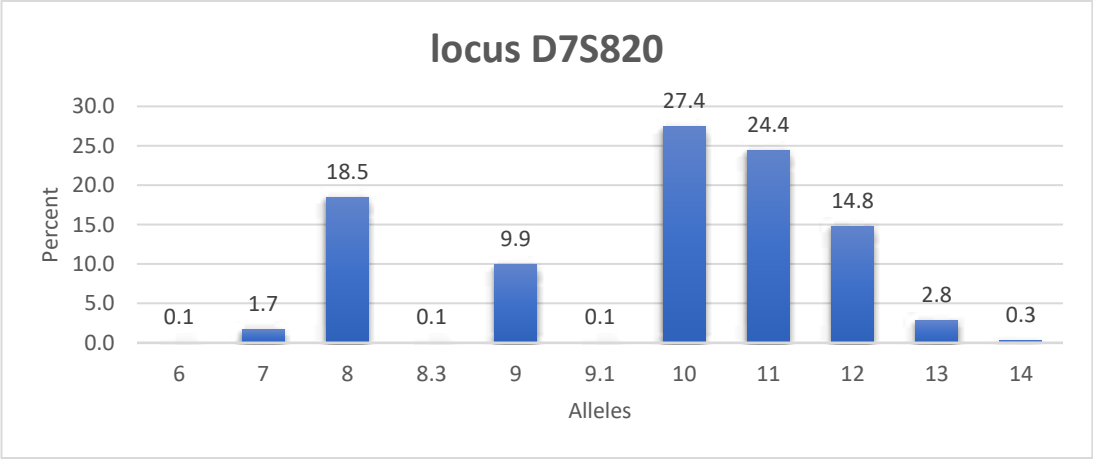


Figure 9-Q: Alleles distrubtion for the locus D7S820

R. Locus D6S1043

Table 11-R: Alleles for the locus D6S1043

Alleles No.	Allele repeats	Freq.	Percentage
1	7	1.0	0.1
2	8	1.0	0.1
3	9	2.0	0.2
4	10	17.0	1.5
5	11	327.0	28.6
6	12	282.0	24.7
7	13	93.0	8.1
8	14	55.0	4.8
9	15	19.0	1.7
10	16	2.0	0.2
11	17	36.0	3.2
12	18	119.0	10.4
13	19	130.0	11.4
14	19.3	1.0	0.1
15	20	48.0	4.2
16	21	8.0	0.7
17	22	1.0	0.1
Total		571	100

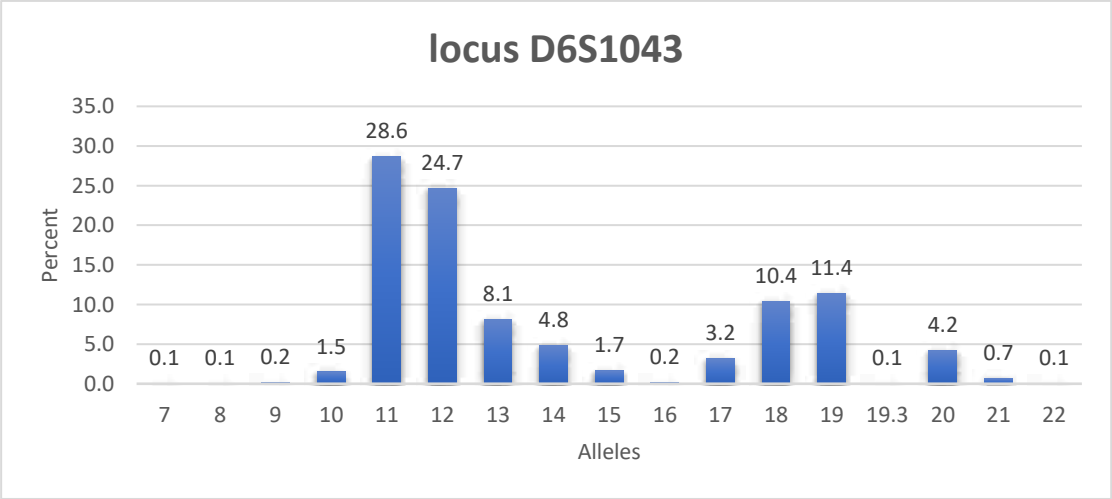


Figure 9-R: Alleles distrubtion for the locus D6S1043

S. Locus D10S1248

Table 11-S: Alleles for the locus D10S1248

Alleles No.	Allele repeats	Freq.	Percentage
1	8	2.0	0.2
2	9	11.0	1.0
3	10	4.0	0.4
4	11	12.0	1.1
5	12	35.0	3.1
6	13	234.0	20.5
7	14	347.0	30.4
8	15	318.0	27.8
9	16	137.0	12.0
10	17	35.0	3.1
11	18	7.0	0.6
Total		571	100

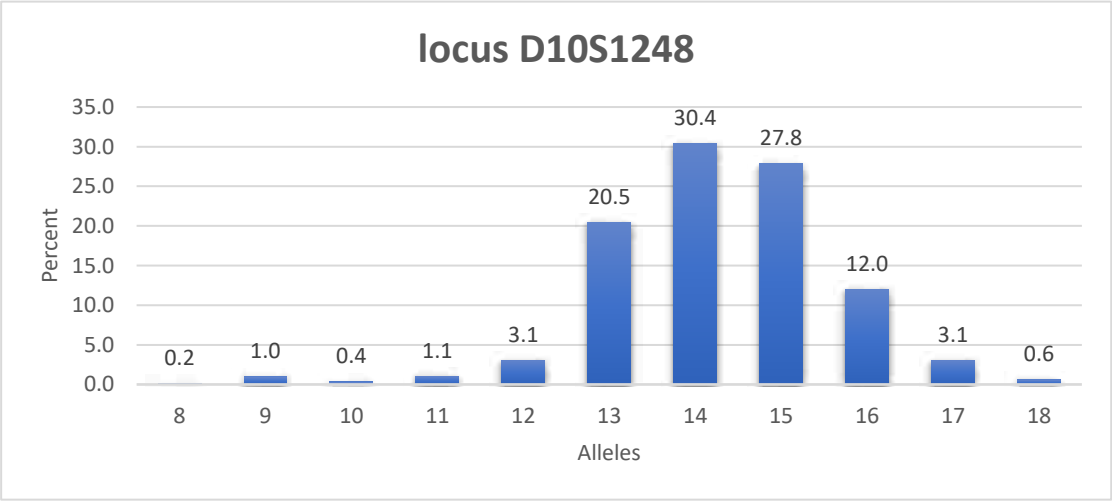


Figure 9-S: Alleles distrubtion for the locus D10S1248

T. Locus D1S1656

Table 11-T: Alleles for the locus D1S1656

Alleles No.	Allele repeats	Freq.	Percentage
1	8	12.0	1.1
2	10	3.0	0.3
3	11	89.0	7.8
4	12	137.0	12.0
5	13	119.0	10.4
6	14	122.0	10.7
7	14.3	7.0	0.6
8	15	179.0	15.7
9	15.3	38.0	3.3
10	16	237.0	20.8
11	16.3	68.0	6.0
12	17	59.0	5.2
13	17.3	43.0	3.8
14	18	7.0	0.6
15	18.3	15.0	1.3
16	19	1.0	0.1
17	19.3	5.0	0.4
18	20.3	1.0	0.1
Total		571	100

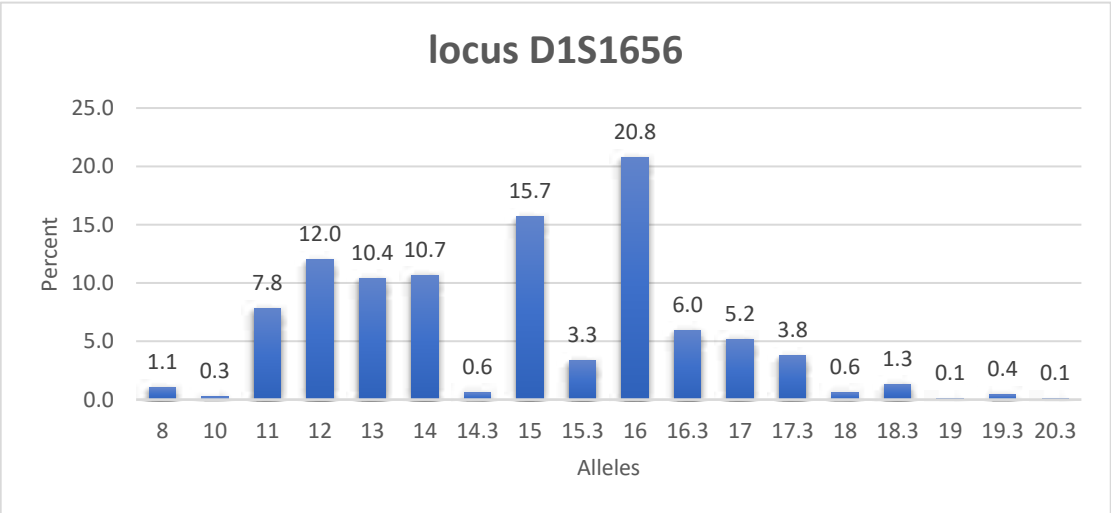


Figure 9-T: Alleles distrubtion for the locus D1S1656

U. Locus D12S391

Alleles for the locus D12S391

Table 11-U:

Alleles No.	Allele repeats	Freq.	Percentage
1	15	24.0	2.1
2	16	25.0	2.2
3	16.3	1.0	0.1
4	17	141.0	12.3
5	17.3	2.0	0.2
6	18	212.0	18.6
7	18.3	17.0	1.5
8	19	143.0	12.5
9	19.2	1.0	0.1
10	19.3	2.0	0.2
11	20	128.0	11.2
12	20.3	1.0	0.1
13	21	130.0	11.4
14	22	134.0	11.7
15	23	98.0	8.6
16	23.3	1.0	0.1
17	24	46.0	4.0
18	24.3	1.0	0.1
19	25	29.0	2.5
20	26	5.0	0.4
21	27	1.0	0.1
Total		571	100

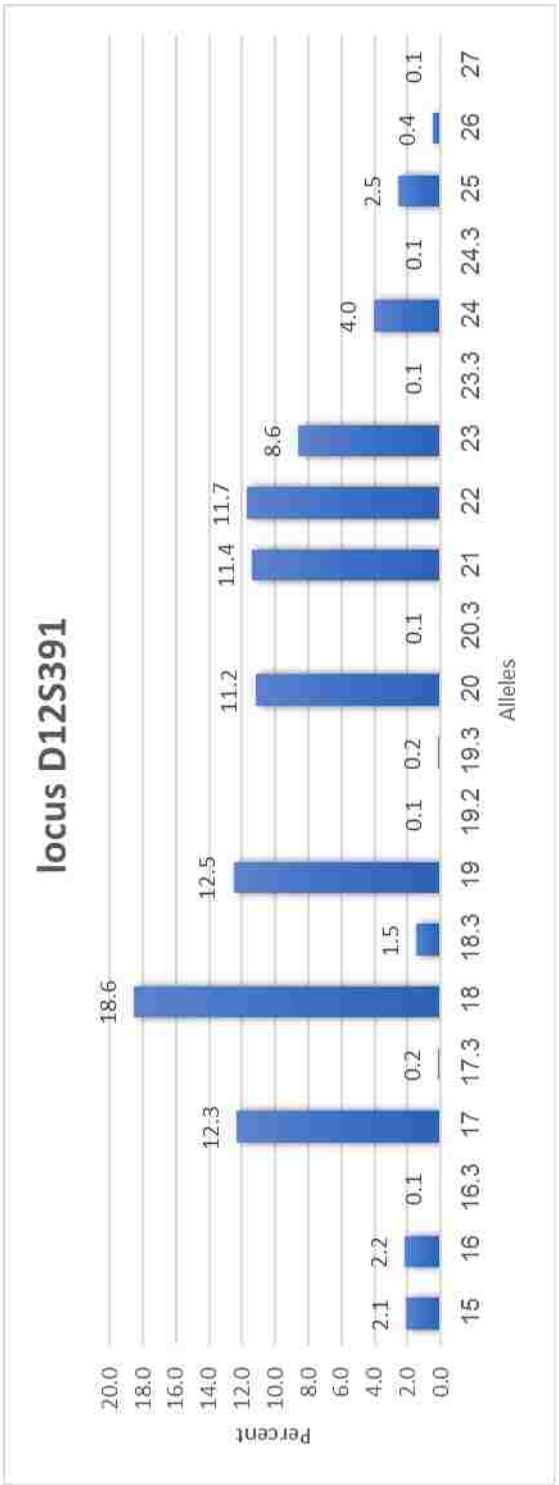


Figure 9-U: Alleles distrubtion for the locus D12S391

V. Locus D2S1338

Alleles for the

Table 11-V: locus D2S1338

Alleles No.	Allele repeats	Freq.	Percentage
1	14	1.0	0.1
2	15	1.0	0.1
3	16	39.0	3.4
4	17	215.0	18.8
5	18	105.0	9.2
6	19	162.0	14.2
7	20	170.0	14.9
8	21	74.0	6.5
9	22	60.0	5.3
10	23	139.0	12.2
11	24	79.0	6.9
12	25	72.0	6.3
13	26	21.0	1.8
14	27	4.0	0.4
Total		571	100

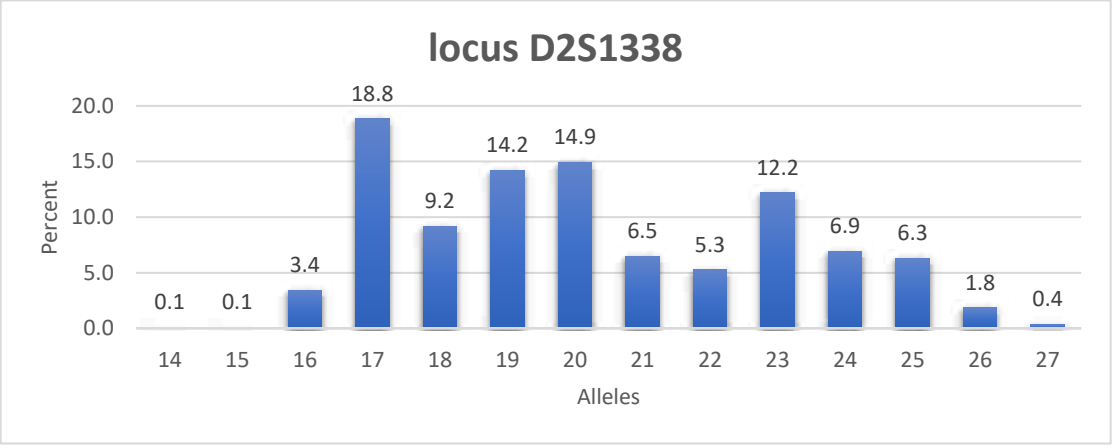


Figure 9-V: Alleles distrubtion for the locus D2S1338

W. Locus Penta D

Alleles for the locus Penta D

Alleles No.	Allele repeats	Freq.	Percentage
1	2.2	19.0	1.7
2	3.2	1.0	0.1
3	5	2.0	0.2
4	6	3.0	0.3
5	7	7.0	0.6
6	8	39.0	3.4
7	9	232.0	20.3
8	10	205.0	18.0
9	11	216.0	18.9
10	12	153.0	13.4
11	13	141.0	12.3
12	14	90.0	7.9
13	15	26.0	2.3
14	16	7.0	0.6
15	18	1.0	0.1
Total		571	100

Table 11-W:

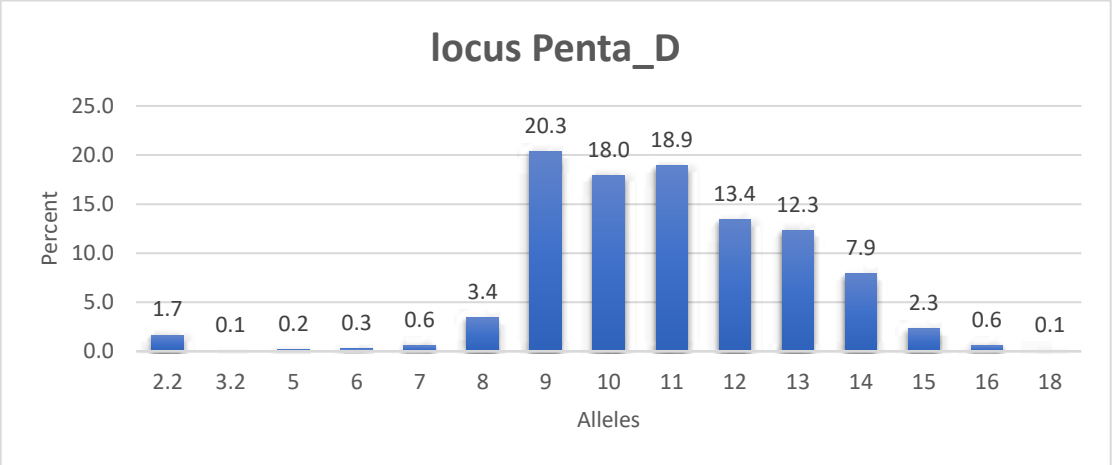


Figure 9-W: Alleles distrubtion for the locus Penta D

Tables 11 and Figures 9 (A-W) present a detailed description of allelic distribution for each loci and the corresponding bar charts. Each locus has different number of alleles among the UAE population with different degree of prevalence. Locus Penta E has highest number of alleles which has 25 alleles among the UAE Arabs population while TH01 Locus has 6 alleles among the UAE Arabs population which is the lowest number of alleles.

Table 12: Predominant alleles for the 23 loci detected in UAE Arabs population using VeriFiler™ Express PCR Amplification Kit

Locus	Predominant Alleles	Freq.	Locus	Predominant Alleles	Freq.
D3S1358	16	0.2968	FGA	23	0.1891
vWA	17	0.2907	D22S1045	15	0.4387
D16S539	11	0.3520	D5S818	12	0.3792
CSF1PO	12	0.3643	D13S317	12	0.3152
TPOX	8	0.4939	D7S820	10	0.2741
D8S1179	13	0.2408	D6S1043	11	0.2863
D21S11	29	0.2320	D10S1248	14	0.3039
D18S51	13	0.1541	D1S1656	16	0.2075
Penta_E	12	0.1559	D12S391	18	0.1856
D2S441	11	0.3853	D2S1338	17	0.1883
D19S433	13	0.2583	Penta_D	9	0.2032
TH01	6	0.2741			

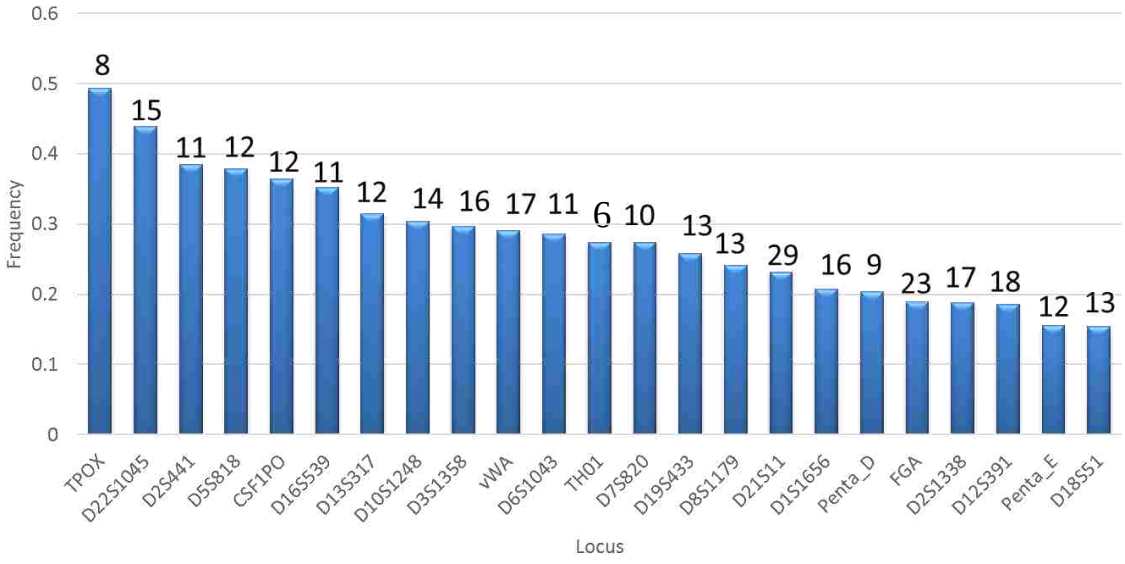


Figure 10: Allele frequency for the predominant alleles in the UAE Arabs population

Analysis of the allele frequency from Tables 11 (A-W) in the UAE Arabs population clearly showed that each locus has predominant alleles as shows in Table 12 and Figure 10.

3.3 Gene Diversity for 23 loci in UAE Arabs population

Table 13: Genetic diversity for VeriFiler™ Express loci in UAE Arabs population

Locus	Diversity	Locus	Diversity
D3S1358	0.76283	FGA	0.86992
vWA	0.79498	D22S1045	0.71664
D16S539	0.77148	D5S818	0.73977
CSF1PO	0.71371	D13S317	0.78389
TPOX	0.67406	D7S820	0.79899
D8S1179	0.84009	D6S1043	0.82167
D21S11	0.84211	D10S1248	0.7723
D18S51	0.88215	D1S1656	0.88127
Penta_E	0.9149	D12S391	0.8853
D2S441	0.76051	D2S1338	0.88254
D19S433	0.84272	Penta_D	0.85002
TH01	0.7821		

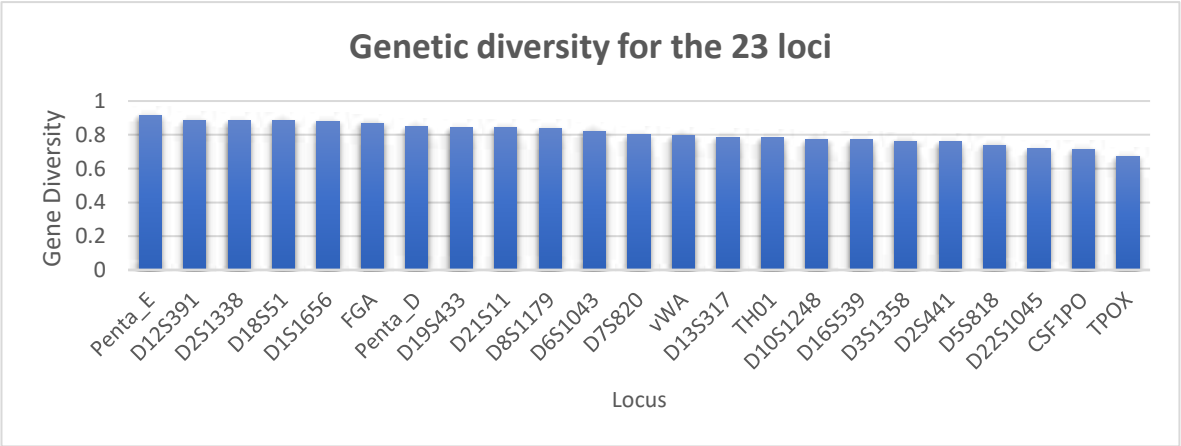


Figure 11: Genetic diversity for VeriFiler™ Express loci in UAE Arabs population

Genetic diversity was calculated for all 23 loci using Arlequin 3.5. Table 13 shows the genetic diversity for all loci and it is clear that the Penta E locus has the highest genetic diversity while TPOX has the lowest genetic diversity. Genetic diversity has been represented in bar chart to make it more readable and clear (Figure11).

3.4 Heterozygosity and Homozygosity

Table 14: Heterozygosity and Homozygosity in each Locus

Locus	Observed Heterozygosity	Observed Homozygosity
D3S1358	78.5 %	21.5 %
vWA	76.9 %	23.1 %
D16S539	76.9 %	23.1 %
CSF1PO	72.0 %	28.0 %
TPOX	68.8 %	31.2 %
D8S1179	83.5 %	16.5 %
D21S11	81.6 %	18.4 %
D18S51	87.6 %	12.4 %
Penta E	90.4 %	9.6 %
D2S441	77.6 %	22.4 %
D19S433	84.1 %	15.9 %
TH01	75.8 %	24.2 %
FGA	85.3 %	14.7 %
D22S1045	69.4 %	30.6 %
D5S818	68.3 %	31.7 %
D13S317	78.3 %	21.7 %
D7S820	77.8 %	22.2 %
D6S1043	80.4 %	19.6 %
D10S1248	72.7 %	27.3 %
D1S1656	86.2 %	13.8 %
D12S391	87.6 %	12.4 %
D2S1338	88.1 %	11.9 %
Penta_D	82.8 %	17.2 %

Study of heterozygosity is one of the first parameters that presents in a data set. It is important to study the genetic variation in populations to get an idea about the structure of a population. Heterozygosity is the condition of having two different alleles at a locus. The heterozygosity and homozygosity calculated for 23 loci. 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. Table 14 shows the heterozygosity and homozygosity for all loci and it is clear that the Penta E locus has the highest heterozygosity while D5S818 has the lowest heterozygosity.

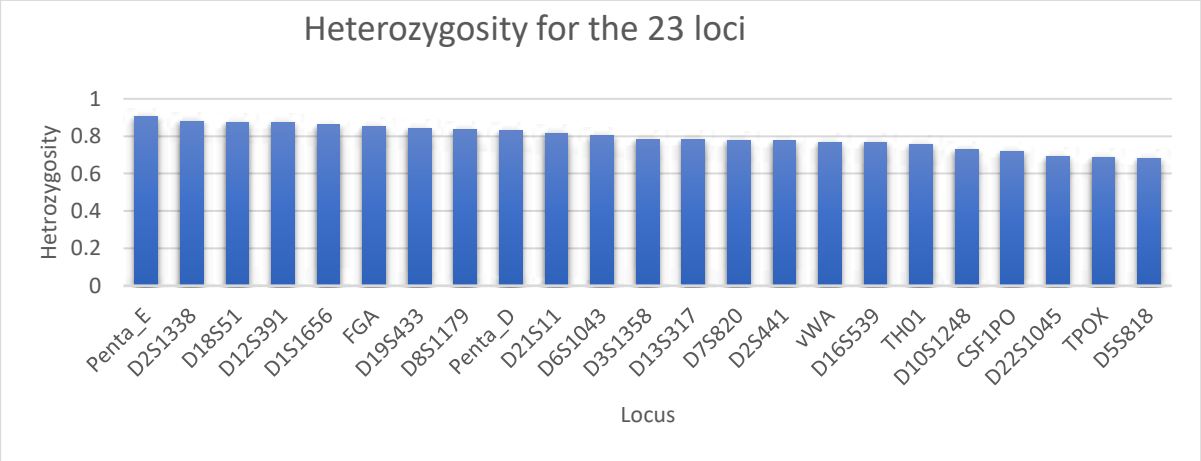


Figure 12: Heterozygosity for VeriFiler™ Express loci in UAE Arabs population

3.5 Test for Hardy- Weinberg Equilibrium

Table 15: Hardy- Weinberg Equilibrium in each Locus

Locus	HWE
D3S1358	0.46523
vWA	0.31257
D16S539	0.46757
CSF1PO	0.50447
TPOX	0.0772
D8S1179	0.14863
D21S11	0.09887
D18S51	0.06147
Penta_E	0.44002
D2S441	0.01042
D19S433	0.26643
TH01	0.04252
FGA	0.90589
D22S1045	0.34794
D5S818	0.19068
D13S317	0.55129
D7S820	0.80331
D6S1043	0.48434
D10S1248	0.07346
D1S1656	0.65247
D12S391	0.07082
D2S1338	0.07252
Penta_D	0.70776

Hardy Weinberg Equilibrium indicates that allele and genotype frequencies in a population will remain constant from generation to generation in the absence of other evolutionary influences.

3.6 Forensic and population genetic parameters

Statistical analysis is used to interpret DNA results for genetic identity. 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 for the 23 STR loci of the UAE population were determined and are indicated in Table 16.

Table 16: Forensic efficiency parameters of the 23 STR loci in 571 unrelated UAE individuals

Locus	Matching Probability	Power of Discrimination	Power of Exclusion
D3S1358	0.100	0.900	0.571
vWA	0.072	0.928	0.543
D16S539	0.090	0.910	0.543
CSF1PO	0.138	0.862	0.460
TPOX	0.150	0.850	0.410
D8S1179	0.047	0.953	0.666
D21S11	0.046	0.954	0.629
D18S51	0.028	0.972	0.746
Penta_E	0.015	0.985	0.803
D2S441	0.096	0.904	0.555
D19S433	0.042	0.958	0.676
TH01	0.082	0.918	0.524
FGA	0.031	0.969	0.701
D22S1045	0.117	0.883	0.418
D5S818	0.106	0.894	0.402
D13S317	0.078	0.922	0.568
D7S820	0.070	0.930	0.558
D6S1043	0.053	0.947	0.606
D10S1248	0.086	0.914	0.471
D1S1656	0.027	0.973	0.718
D12S391	0.026	0.974	0.746
D2S1338	0.027	0.973	0.757
Penta_D	0.040	0.960	0.653
Combined	2.9×10^{-29}	0.99999	0.99999

3.7AMOVA results in UAE Arabs population

Table 17: Fst AMOVA results using 15 loci within different populations

Population	UAE	Turkey	Portugal	Morocco	Pakistan	India
UAE	-					
Turkey	0.00159	-				
Portugal	0.00464	0.00337	-			
Morocco	0.00510	0.00436	0.00398	-		
Pakistan	0.00465	0.00496	0.00930	0.00899	-	
India	0.00657	0.00670	0.001214	0.01082	0.00002	-

Analysis of Molecular Variance (AMOVA) is a method of estimating population differentiation directly from molecular data and testing hypotheses about such differentiation.

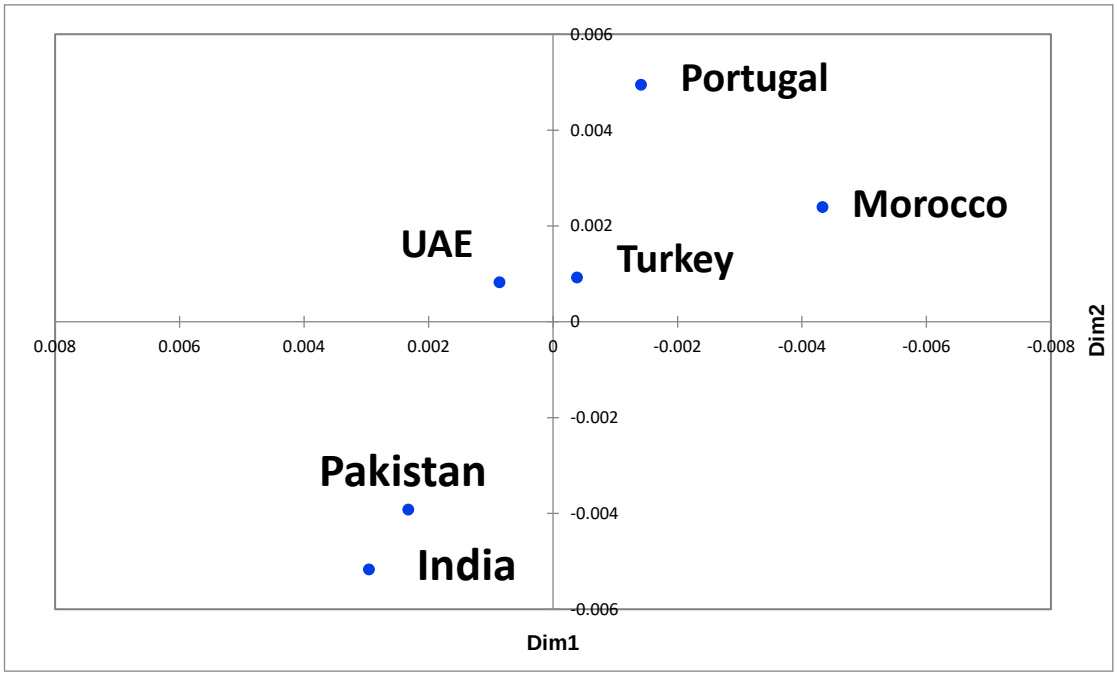


Figure 13: Multidimensional scaling (MDS) plot based on pairwise Fst genetic distance among UAE Arabs population and other populations using the 23 Autosomal STR Verifieler ® kit.

Multidimensional scaling (MDS) plot based on pairwise Fst genetic distance using the 15 Autosomal STR Veifiler® kit (Figure13) among UAE Arabs population and other populations.

Chapter 4: Discussion

DNA analysis has become one of the most definitive methods for human identification in forensic science, paternity testing, missing person's identifications and mass disaster investigations. In this study, results were obtained from 23 autosomal STR loci (D3S1358, vWA, D16S539, CSF1PO, TPOX, D8S1179, D21S11, D18S51, D2S441, D19S433, TH01, FGA, D22S1045, D5S818, D13S317, D7S820, D10S1248, D1S1656, D12S391, D2S1338, D6S1043, Penta D and Penta E) in United Arab Emirates (UAE) Arab population in order to investigate its usefulness in such population.

Statistical evaluation of population structure and forensic parameters including observed homozygosity and heterozygosity, genetic diversity, the exact testing for Hardy-Weinberg equilibrium (HWE), the power of discrimination (DP) and the power of exclusion (PE) were estimated based on allelic frequencies.

4.1 Alleles frequency distribution

Allelic frequencies of 571 UAE Arabs individuals at the 23 autosomal STR loci are shown in (Table 11A-W). The total number of alleles observed across the population was found to be 305 alleles indicating high level of polymorphism of the selected microsatellites. The highest allele frequencies were detected in D22S1045 and TPOX loci (allele 15 for D22S1045 and allele 8 for TPOX). Penta E locus represented the largest number of different alleles (25 alleles) whereas FGA locus (24 alleles) coming second largest number of alleles. The TH01 locus showed the smallest number of different alleles (6 alleles). Therefore, most polymorphic loci is Penta E and the least polymorphic is TH01. Penta E exhibited the highest genetic diversity 0.9149

(Table 11); whereas TH01 showed a genetic diversity of 0.7821. TPOX locus showed the lowest genetic diversity with 0.67406.

According to the resulted genetic diversity, the three most diverse loci are Penta E = 0.9149, D12S391 = 0.8853 and D2S1338 = 0.88254. Therefore, these loci should be taken into consideration in forensic analysis, especially when evaluating the validity of profiles obtained from unknown biological stains.

4.2 Predominant Alleles

In the present study, the most predominant alleles at the 23 loci are shown in (Table 12). At the CSF1P0 locus, a total of 9 alleles were observed, with repeat numbers ranging from 7 –15. The repeat numbers 10, 11 and 12 are the three most predominant alleles in the Emirati as well as in the Saudi, Kuwaiti, Iraqi and Egyptian populations, respectively (Table 18) (Alshamali, Alkhayat, Budowle, & Watson, 2005) (Alsafiah, Goodwin, Hadi, Alshaikhi, & Wepeba, 2017) (Hadi, Abdullah, Jaber, & Yoke, 2014). However, allele 12 occurred with a higher frequency (0.364, 0.342, 0.360 and 0.324) in the Emirati, Saudi, Kuwaiti and Egyptian populations, respectively, while allele 11 with a higher frequency (0.315) in the Iraqi population. The number of alleles is dependent on the sample size because the number of observed alleles tends to increase with increases in population size and also because the presence of unique alleles in populations, which occur in low frequencies. The number of alleles for each marker is an important indicator of the usefulness of the marker for genetic screening.

Table 18: Predominant Alleles in different populations

Predominant Alleles						
Locus / Pop	Emirati	Saudi	Kuwaiti	Iraqi	Egyptian	Lebanese
D8S1179	13	14	15	13	13	13
D21S11	29	29	29	30	29	29
D7S820	10	10	10	10	10	10
CSF1PO	12	12	12	11	12	11
D3S1358	16	16	15	15	15	17
TH01	6	6	6	9	9	9
D13S317	12	12	12	12	12	11
D16S539	11	11	11	11	11	11
D2S1338	17	20	17	17	17	17
D19S433	13	14	14	14	14	14
vWA	17	17	17	16	17	17
TPOX	8	8	8	8	8	8
D18S51	13	13	13	13	12	14
D5S818	12	12	12	12	12	12
FGA	23	23	23	22	23	24

4.3 Heterozygosity and Homozygosity

Observed heterozygosity and homozygosity all over the 23 loci are presented in Table 13 and the observed heterozygosity between UAE population and other studied populations as illustrated in Table 19

The observed heterozygosity in any population depends on the number and the frequency of alleles of each locus. Out of 23 autosomal loci in UAE Population, Penta E showed the highest heterozygote loci with more than 90% from 571 samples, whereas D5S818 showed the lowest heterozygote loci with less than 69% from 571 samples (Table 14)

Table 19: Observed heterozygosity in different populations

Heterozygosity Frequency						
Locus / Pop	Emirati	Saudi	Kuwaiti	Iraqi	Egyptian	Lebanese
D8S1179	0.8354	0.7960	0.7908	0.8151	0.8052	0.8200
D21S11	0.8161	0.8040	0.8247	0.8219	0.8290	0.8050
D7S820	0.7776	0.7920	0.7470	0.8138	0.7506	0.7390
CSF1PO	0.7198	0.6560	0.6753	0.7260	0.6714	0.6820
D3S1358	0.7846	0.7480	0.7629	0.7808	0.7743	0.7220
TH01	0.7583	0.7258	0.7649	0.7466	0.7838	0.7260
D13S317	0.7828	0.7600	0.7669	0.7123	0.7720	0.7550
D16S539	0.7688	0.8080	0.7390	0.7466	0.7862	0.7450
D2S1338	0.8809	0.8440	0.8347	0.8356	0.8575	0.8180
D19S433	0.8406	0.8440	0.8207	0.8356	0.8290	0.7960
vWA	0.7688	0.7640	0.7590	0.8082	0.7862	0.8030
TPOX	0.6883	0.6280	0.6235	0.6370	0.6295	0.6020
D18S51	0.8757	0.8600	0.8307	0.9041	0.8622	0.8500
D5S818	0.6830	0.7320	0.7171	0.7123	0.7363	0.7470
FGA	0.8529	0.8760	0.8466	0.9041	0.8452	0.8460

Observed heterozygosity from different populations was observed from 15 common autosomal STR loci (Table 19) (Alsafiah et al., 2017; Alshamali et al., 2005) (Hadi et al., 2014). On average, there are populations with low heterozygosity, approximately lower than 65% in most tested loci (Halima et al., 2009). On the contrary, Emirati and other populations such as Saudi, Kuwaiti, Iraqi, Egyptian and Lebanese populations showed highest percentage of heterozygotes loci in most of the 15 loci. By comparing the results from different populations, D2S1338 showed the highest heterozygote loci with more than 88% in the UAE Arab population. Saudi, Kuwaiti and Iraqi populations shared the same highest heterozygote loci, which are FGA with more than 87% in Saudi population, more than 84% in Kuwaiti population and more than 90% in Iraqi population. Egyptian and Lebanese populations shared the

same highest heterozygote loci which is D18S51 with more than 86% in Egyptian population and 85% in Lebanese population.

4.4 Hardy Weinberg Equilibrium

Testing for deviations from Hardy-Weinberg Equilibrium (HWE) is a widely recommended test for population-based genetic association studies. P-values test was conducted to detect significant deviation from HW equilibrium. There were 23 within-locus tests conducted on the UAE population (Table 14). No significant deviation from HWE expectations was observed (a 5% significance level was taken) in the UAE population for 21 of the 23 STR loci analyzed ($P>0.05$). The exceptions were D2S441 (P-value = 0.01042) and TH01 (P-value = 0.04252) 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(John M Butler, 2009). Thus, a set of 23 comparisons would lower the alpha value from 0.05 to 0.0022 (0.05/23) so only p-values below 0.0022 would be considered statistically significant rather than the conventional $P < 0.05$. The D2S441 and TH01 loci P-value did not fail at this newly generated P-value of 0.0022. 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 (Halima et al., 2009).

4.5 Power of Discrimination

Power of discrimination (PD) for all the 23 loci is presented in Table 16. These values range between 0.985 and 0.850 being the highest and the lowest values given, respectively, by Penta E and TPOX markers. Power of discrimination for all tested loci was above 85% for the CSF1PO and TPOX loci, above 88% for the D5S818, D22S1045 and D3S1358 loci and ranged from 90% to 98.5 % for the rest of the loci. The PD is the probability that two randomly chosen persons would not have matching DNA profiles. Otherwise stated, this is the probability that an innocent person will be excluded as the donor of a biological evidence sample.

The combination of 23 STRs proved to be extremely discriminating; The Combined Discrimination Power (CDP) for the United Arab Emirates population for the corresponding 23 STR loci used, has been calculated to be greater than 0.999999999 (Table 16), meaning that in the population, two randomly selected unrelated individuals will have 0.999999999 probability having different genotypes. These figures mean that those loci can be confidently used to establish a DNA-based database for United Arab Emirates population.

Table 20: Power of Discrimination (PD) in different populations

Power of Discrimination (PD) Frequency						
Locus / Pop	Emirati	Saudi	Kuwaiti	Iraqi	Egyptian	Lebanese
D8S1179	0.9529	0.9507	0.9485	0.9527	0.9490	0.9460
D21S11	0.9541	0.9485	0.9495	0.9559	0.9535	0.9530
D7S820	0.9297	0.9296	0.9163	0.9256	0.9156	0.9280
CSF1PO	0.8621	0.8837	0.8604	0.8785	0.8752	0.8640
D3S1358	0.8998	0.8977	0.9076	0.9083	0.9104	0.9040
TH01	0.9181	0.9118	0.9173	0.9199	0.9185	0.9200
D13S317	0.9221	0.9244	0.9202	0.9109	0.9228	0.9240
D16S539	0.9102	0.9101	0.9121	0.9248	0.9181	0.9180
D2S1338	0.9733	0.9661	0.9688	0.9685	0.9715	0.9670
D19S433	0.9582	0.9680	0.9651	0.9437	0.9597	0.9470
vWA	0.9284	0.9310	0.9258	0.9440	0.9367	0.9350
TPOX	0.8501	0.8387	0.8213	0.8387	0.8412	0.8290
D18S51	0.9719	0.9737	0.9704	0.9735	0.9720	0.9690
D5S818	0.8936	0.9034	0.8957	0.8966	0.9114	0.8840
FGA	0.9688	0.9632	0.9639	0.9680	0.9689	0.9670

Power of Discrimination from different populations was obtained from 15 common autosomal STR loci (Table 20) (Alsafiah et al., 2017; Alshamali et al., 2005) (Hadi et al., 2014). The Penta E and Penta D loci included in the Verifiler Express Amplification kit were not typed in the Saudi, Kuwaiti, Iraqi, Egyptian and Lebanese populations because they used different kits in their genotyping studies. Emirati and other populations such as Saudi, Kuwaiti, Iraqi, Egyptian and Lebanese populations showed highest percentage of Power of Discrimination loci in most of the 15 loci. By comparing the results from different populations, D2S1338 showed the highest power of discrimination loci with more than 97% in UAE Population. Saudi, Kuwaiti, Iraqi, Egyptian and Lebanese populations shared the same highest power of discrimination loci which is D18S51 with greater than 97.3% in Saudi Population, greater than 97%

in Kuwaiti Population, greater than 97.3 % in Iraqi Population, 97.2 % in Egyptian and greater than 96% in Lebanese population.

4.6 Random matching probability

Matching Probability (MP) is the probability to find two individuals with identical genotype (the same DNA profile) if you select them randomly in a population. The MP between unrelated individuals for the 23 STR loci used in this study was estimated for UAE population (Table 16).

In the present study, the calculated MP ranged from 0.149 to 0.015. The highest value for this parameter is given by TPOX marker and the lowest by Penta E. The CMP value (for all possible genotypes) for the United Arab Emirates population for the corresponding 23 STR loci used in this study was one person in 2.9×10^{-29}. Therefore, the probability that two randomly selected samples will match at all 23 of the loci is $2.9 \times 10^{-29,}$ 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 contributor to the biological stain found in crime scene.

4.7 Exclusion probability

Exclusion probability (EP) is the probability of a given locus to exclude one individual, as the proportion of individuals that have a DNA profile different from that a randomly selected individual, The EP between unrelated individuals for the 23 STR loci used in this study was estimated for UAE population (Table 16).

In the present study and from the genotyping data, PE for every locus was calculated and presented in Table 16. As expected, the power of exclusion was high for all the microsatellites analyzed rangeing from 0.402 (D5S818) to 0.803 (Penta E), with an average of 0.597. The combined power of exclusion, which is the exclusion probability considering all 23 loci, was greater than 99.99%, indicating that these loci are appropriate to determine parentage in UAE population beyond any reasonable doubt.

4.8 Analysis of Molecular variance (AMOVA)

Analysis of Molecular Variance (AMOVA) was calculated using Arlequin 3.5 software. The resulted frequencies of UAE population gentoypes were compared to other population's genotypes from different regions of world (Turkey, Morocco, Portugal, Pakistan and India) using only 15 common loci (Table 17).

The MDS plot structured from Fst distance matrix for the 15 common STR loci (Figure 12) showed that UAE population clustered along with Turkey populations which mean that they have a closer genetic distance. These results also correlated with the geographical distribution of these populations.

Chapter 5: Conclusion

The main objective of this study was to determine the genetic structure of UAE population using 23 autosomal STR loci and to evaluate the usefulness of these loci for forensic genetic purposes.

The number of different alleles observed across the population was found to be 305 alleles. The highest allele frequencies occurred in the allele 12 for D13S317 locus (43.9 %) and allele 8 for TPOX (49.4 %), while the Penta E locus showed the largest number of different alleles (25 alleles) and TH01 locus represented the smallest number of different alleles (6 alleles). The heterozygosity of the 23 STR loci ranged from 68.3 % to 90.4 % (mean value 79.6%), the locus with the highest heterozygosity was Penta E, while locus D5S818 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, D22S1045 and D3S1358 loci and ranged from 90% to 97% for the rest of loci except Penta E which is consider the highest power of discrimination loci with 98.5%. The combined probability of exclusion, power of discrimination, probability of matching value for all the 15 STR loci were 0.999999999695, 0.999999999 and 2.9 x 10^ -29, respectively. The genetic distance between UAE Arabs population and Turkey population was closest to each other, whereas it appeared to be far from Pakistan, Indian and Portugal populations.

This study is very important, and it contributes to other studies in the Gulf region to study the genetic diversity of populations. In conclusion, it is noticeable that the 23 Autosomal STRs in VeriFiler™ Express PCR Amplification Kit show high power of forensic discrimination and high genetic diversity.

References

Ahmed, H. A., MacLeod, E. T., Hide, G., Welburn, S. C., & Picozzi, K. (2011). The best practice for preparation of samples from FTA® cards for diagnosis of blood borne infections using African trypanosomes as a model system. *Parasites & vectors, 4*(1), 4-68.

Ahmed, M., & Zeyad, T. (2017). Study of Y-Chromosome STR Markers in United Arab Emirates Population.

Alaeddini, R., Walsh, S. J., & Abbas, A. (2010). Forensic implications of genetic analyses from degraded DNA—a review. *Forensic Science International: Genetics, 4*(3), 148-157.

Alsafiah, H. M., Goodwin, W. H., Hadi, S., Alshaikhi, M. A., & Wepeba, P.-P. (2017). Population genetic data for 21 autosomal STR loci for the Saudi Arabian population using the GlobalFiler® PCR amplification kit. *Forensic Science International: Genetics, 31*, e59-e61.

Alshamali, F., Alkhayat, A. Q., Budowle, B., & Watson, N. (2005). STR population diversity in nine ethnic populations living in Dubai. *Forensic science international, 152*(2-3), 267-280.

Brenner, C. (1990). *Paternity index calculations in single locus hypervariable DNA probes: validation and other studies.* Paper presented at the Proceedings for the international symposium on human identification 1989. Madison, Promega.

Budowle, B., Moretti, T. R., Keys, K. M., Koons, B. W., & Smerick, J. (1997). Validation studies of the CTT STR multiplex system. *Journal of Forensic Science, 42*(4), 701-707.

Butler, J. (2012). Sample collection, storage and characterization. *Butler JM. Advanced Topics in Forensic DNA Typing: Methodology. Waltham: Elsevier*, 1-28.

Butler, J., & Hill, C. (2012). Biology and genetics of new autosomal STR loci useful for forensic DNA analysis. *analysis, 24*(15), 16-26.

Butler, J. M. (1998). The use of capillary electrophoresis in genotyping STR loci. In *Forensic DNA Profiling Protocols* (pp. 279-289): Springer.

Butler, J. M. (2001). *Forensic DNA typing: biology & technology behind STR markers*: Academic Press.

Butler, J. M. (2005). *Forensic DNA typing: biology, technology, and genetics of STR markers*: Elsevier.

Butler, J. M. (2009). *Fundamentals of forensic DNA typing*: Academic Press.

Butler, J. M. (2011). *Advanced topics in forensic DNA typing: methodology*: Academic Press.

Butler, J. M. (2014). *Advanced topics in forensic DNA typing: interpretation*: Academic Press.

Carracedo, A. (2015). Forensic genetics: history. *Forensic Biology, 19*.

Chu, E. H., & Giles, N. H. (1959). Human chromosome complements in normal somatic cells in culture. *American journal of human genetics, 11*(1), 63-79.

Consortium, G. P. (2015). A global reference for human genetic variation. *Nature, 526*(7571), 68-74.

Cotton, E., Allsop, R., Guest, J., Frazier, R., Koumi, P., Callow, I., . . . Sparkes, R. (2000). Validation of the AMPFlSTR® SGM Plus™ system for use in forensic casework. *Forensic science international, 112*(2-3), 151-161.

De Conti, L., Baralle, M., & Buratti, E. (2013). Exon and intron definition in pre-mRNA splicing. *Wiley Interdisciplinary Reviews: RNA, 4*(1), 49-60.

Dunbar, B. S. (2012). *Two-dimensional electrophoresis and immunological techniques*: Springer Science & Business Media.

Farhan, M. M., Hadi, S., Iyengar, A., & Goodwin, W. H. (2016). Population genetic data for 20 autosomal STR loci in an Iraqi Arab population: Application to the identification of human remains. *Forensic Science International: Genetics, 25*, e10-e11.

Fisher, R. (1951). Standard calculations for evaluating a blood-group system. *Heredity, 5*(1), 95-102.

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.

Goodwin, W., Linacre, A., & Hadi, S. (2011). *An introduction to forensic genetics* (Vol. 2): John Wiley & Sons.

Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Léger, P., Lepais, O., . . . Salin, F. (2011). Current trends in microsatellite genotyping. *Molecular ecology resources, 11*(4), 591-611.

Gymrek, M., Golan, D., Rosset, S., & Erlich, Y. (2012). lobSTR: a short tandem repeat profiler for personal genomes. *Genome research*.

Hadi, I., Abdullah, M., Jaber, A., & Yoke, C. (2014). Genetic variation of twenty autosomal STR loci and evaluate the importance of these loci for forensic genetic purposes. *African Journal of Biotechnology, 13*(11), 1-9.

Halima, M. S. A., Bernal, L. P., & Sharif, F. A. (2009). Genetic variation of 15 autosomal short tandem repeat (STR) loci in the Palestinian population of Gaza Strip. *Legal medicine, 11*(4), 203-204.

Harbison, C. (2016). ABO Blood Type Identification and Forensic Science (1900-1960). *Embryo Project Encyclopedia*.

Heinze, T. (2013). Creative accomplishments in science: definition, theoretical considerations, examples from science history, and bibliometric findings. *Scientometrics, 95*(3), 927-940.

Khubrani, Y. M., Wetton, J. H., & Jobling, M. A. (2019). Analysis of 21 autosomal STRs in Saudi Arabia reveals population structure and the influence of consanguinity. *Forensic Science International: Genetics, 39*, 97-102.

Lance, S. L., Love, C. N., Nunziata, S. O., O'Bryhim, J. R., Scott, D. E., Flynn, R. W., & Jones, K. L. (2013). 32 species validation of a new Illumina paired-end approach for the development of microsatellites. *PloS one, 8*(11), e81853.

Lander, E. S. (2011). Initial impact of the sequencing of the human genome. *Nature, 470*(7333), 187-197.

Nei, M. (1987). *Molecular evolutionary genetics*: Columbia university press.

Pakzad, Z., Mozdarani, H., Izadi-Mood, N., & Niromanesh, S. (2014). Variable number tandem repeat (VNTR) genotyping of hydatidiform mole in Iranian patients. *Avicenna journal of medical biotechnology, 6*(4), 246-253.

Pascali, J. P., Bortolotti, F., & Tagliaro, F. (2012). Recent advances in the application of CE to forensic sciences, an update over years 2009–2011. *Electrophoresis, 33*(1), 117-126.

Premalatha, B., Ramesh, V., Babu, S., & Balamurali, P. (2014). Procedures to view aberrations-A travel from protein to gene: Literature review. *Indian Journal of Dental Research, 25*(1), 91-91.

Schneider, S., Roessli, D., & Excoffier, L. (2000). Arlequin ver. 2.000. *A software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland*.

Shewale, J., Wang, D., Zhong, C., Gopinath, S., Mulero, J., Razdan, A., . . . Dallett, C. (2013). From sample collection to report generation–A new integrated workflow solution for paternity and relationship testing. *Forensic Science International: Genetics Supplement Series, 4*(1), e232-e233.

St. Claire, R. L. (1996). Capillary electrophoresis. *Analytical Chemistry, 68*(12), 569-586.

Sulistyawati, P. (2010). *Forest diseases: FTA cards and DNA sampling.* University of Tasmania.

Thompson, R., Zoppis, S., & McCord, B. (2012). An overview of DNA typing methods for human identification: past, present, and future. In *DNA Electrophoresis Protocols for Forensic Genetics* (pp. 3-16): Springer.

Watson, J. D. (2012). *The polymerase chain reaction*: Springer Science & Business Media.

Appendix

Appendix I: DNA Profiles Data for 50 random unrelated UAE Arabs population

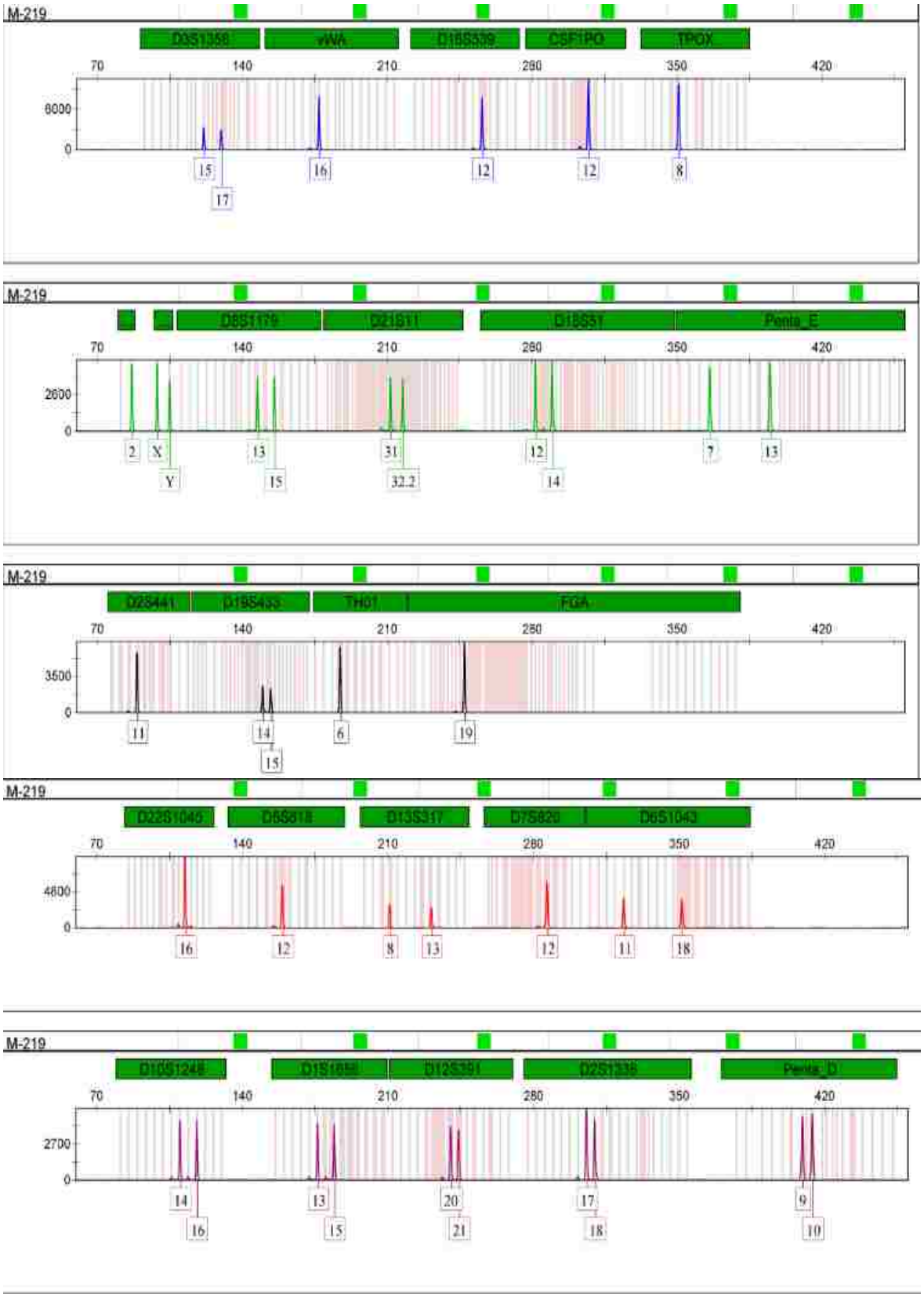
ID	D3S1358	vWA	D16S539	CSF1PO	TPOX	D8S1179	D21S11	D18S51	Penta_E	D2S441	D19S433	TH01	FGA	D22S1045	D5S818	D13S317	D7S820	D6S1043	D10S1248	D1S1656	D12S391	D2S1338	Penta_D
S1	16	17	9	10	8	10	28	12	10	11	14	6	18	15	11	11	9	19	14	11	17	17	9
	17	17	11	13	8	11	29	12	16.4	14	15.2	6	23	16	12	11	11	20	14	13	19	23	12
S2	16	16	11	12	8	10	28	14	11	11	12	7	19	15	13	12	7	11	14	11	16	20	11
	17	16	12	12	9	13	31	16	16	14	13	9	21	16	13	13	11	17	15	15	18	24	11
S3	15	16	9	11	8	12	30	13	11	14	14	7	21.2	15	8	11	10	11	15	12	19	20	11
	17	18	11	12	8	15	32.2	17	13	14	14	8	23	15	11	11	10	20	15	13	23	24	11
S4	15	16	9	10	8	13	28	15	11	14	12	7	20	14	12	8	11	11	14	15	18	18	6
	15	17	11	11	10	14	30	15	17	14	13	9	24	16	12	9	11	19	14	16	21	22	9
S5	16	17	10	11	8	11	30	13	12	10	14	8	23	11	11	11	9	11	14	13	22	18	10
	17	19	11	11	8	13	31.2	17	12	11	15.2	8	24	16	12	12	13	12	15	16	24	25	10
S6	16	14	11	11	8	13	28	13	12	11	13	6	25	11	9	8	8	11	14	12	17	17	9
	17	15	11	12	8	13	31.2	13	12	11	15	7	25	14	11	10	10	19	15	16.3	21	25	9
S7	15	18	13	12	8	11	30	11	5	11	14	7	19	15	10	10	10	17	14	15	20	18	12
	17	19	14	13	11	13	30	15	12	14	15.2	9.3	22	17	12	11	11	18	15	16	22	23	13
S8	18	16	11	10	8	13	31	13	10	11	13	6	20	15	9	8	9	11	14	14	17	18	9
	18	17	12	11	9	14	33.2	18	13	12	14	9	26	17	12	12	10	19.3	15	16	19	25	13
S9	16	15	9	11	9	11	29	10	12	10	13	7	19	11	11	12	11	13	15	14	19	19	12
	17	18	13	12	11	16	31.2	14	13	11	14	8	26	16	12	12	12	13	16	16	20	24	12
S10	14	14	11	11	8	14	27	17	5	11.3	13	7	23	15	8	10	11	12	14	14	19	24	9
	15	15	13	12	9	14	35	19	8	14	14.2	7	24	16	12	11	12	12	14	16.3	21	26	13
S11	17	16	11	10	8	10	32.2	12	10	10	13	7	22	14	12	11	8	12	13	12	17	18	10
	18	16	12	10	10	13	32.2	17	13	15	15.2	9	23	16	12	13	11	13	13	16.3	24	22	14
S12	16	16	11	10	8	14	28	14	9	10	14	9	22	15	11	11	10	11	14	11	18	19	11
	17	16	14	12	10	15	32.2	15	11	11	15.2	9	23	15	13	12	10	19	16	13	18.3	19	14

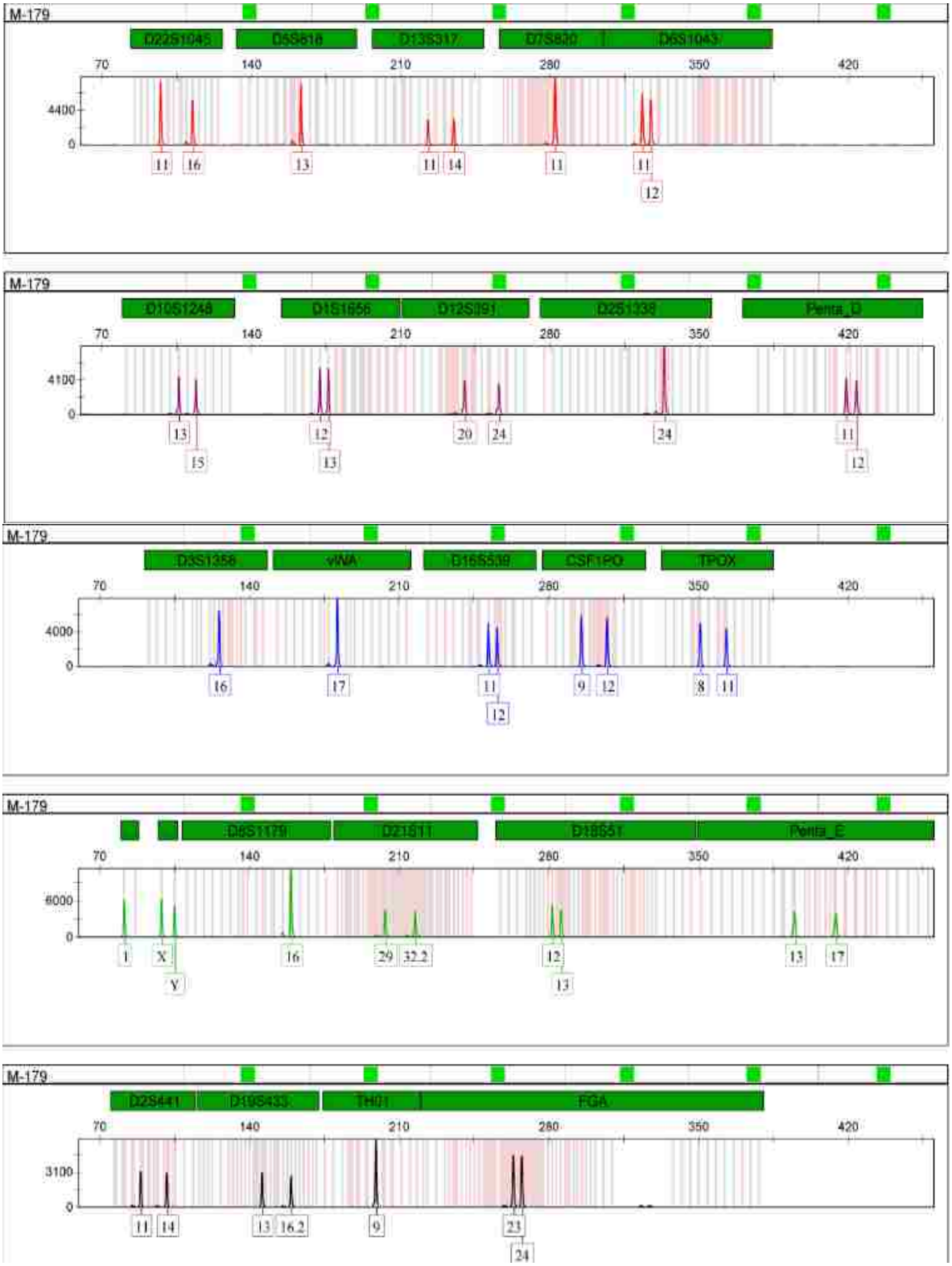
ID	D3S1358	vWA	D16S539	CSF1PO	TPOX	D8S1179	D21S11	D18S51	Penta_E	D2S441	D19S433	TH01	FGA	D22S1045	D5S818	D13S317	D7S820	D6S1043	D10S1248	D1S1656	D12S391	D2S1338	Penta_D
S13	16	15	10	12	8	13	30	12	11	12	12.2	6	19	11	10	8	10	11	14	13	22	21	9
	17	18	11	12	9	14	31	19	13	13	14	9	25	15	11	11	10	20	15	15	24	22	12
S14	16	17	9	12	8	11	30	13	10	10	14	8	24	11	11	12	8	11	13	14	20	18	10
	17	18	11	12	8	15	32.2	15	13	11	15.2	8	26	15	11	14	13	11	16	15	23	25	10
S15	16	15	9	12	8	13	28	17	10	11	13	8	20	11	8	11	11	19	12	15	19	18	10
	19	16	11	12	8	15	31.2	17	13	11	15.2	9	21	15	13	12	12	19	15	16	23	23	12
S16	16	14	9	12	8	12	30	13	7	11	12	6	19	15	12	12	10	12	13	12	17	22	9
	16	16	11	12	10	13	30	18	11	14	14	8	20	16	12	12	10	12	15	15	19	23	14
S17	15	15	12	12	8	11	29	17	12	11	14	6	19	16	11	10	8	12	13	16	19	20	12
	15	17	12	12	11	11	30	20	15	11	15	6	23	16	12	11	11	19	14	16	21	23	15
S18	16	15	9	12	10	14	27	13	12	11.3	13	7	22	11.2	11	11	8	18	14	12	18	17	11
	17	18	9	12	11	15	30	14	18	14	13	9.3	24	16	11	12	9	18	15	16	18.3	20	12
S19	15	18	11	12	8	13	30	14	12	14	12	7	23	16	10	12	12	10	13	15	20	18	6
	16	18	14	12	8	16	30	16	13	14	14	9	23	16	10	13	12	11	15	15	23	20	13
S20	17	15	10	10	8	10	30	14	7	11	13	6	24	15	10	8	9	12	13	11	20	17	12
	18	18	13	11	11	15	31.2	17	15	11	16	9	25	16	11	12	9	12	15	13	20	24	12
S21	15	16	11	11	11	9	28	13	11	10	12.2	7	22	14	10	12	10	19	13	14	17	18	9
	16	17	11	12	12	14	31.2	14	12	11	13	8	25	17	11	12	12	19	13	16	23	20	14
S22	16	16	11	10	8	12	28	13	12	11	14	7	21	11	12	9	9	18	14	15	17	16	12
	17	18	11	12	8	16	30	17	15	12	16	9.3	23	15	12	11	9	20	14	15.3	19	23	13
S23	16	15	10	10	8	13	27	13	11	12	15	6	20	11	12	11	8	12	15	14	18	17	9
	18	16	12	11	10	15	29	15	13	14	15.2	6	23	14	12	12	10	12	15	16	22	21	13
S24	15	16	12	12	8	14	29	13	11	14	12	6	19	15	11	12	8	11	13	12	19	17	8
	18	17	13	13	8	14	30	14	14	14	14	10	24	16	12	13	11	14	14	14	23	20	10
S25	17	14	10	10	8	14	29	12	9	11	14.2	8	24	15	11	10	10	11	15	11	19	17	9
	19	19	12	10	8	14	32.2	13	13	12	15.2	9.3	25	16	12	11	11	11	15	15	21	17	11

ID	D3S1358	vWA	D16S539	CSF1PO	TPOX	D8S1179	D21S11	D18S51	Penta E	D2S441	D19S433	TH01	FGA	D22S1045	D5S818	D13S317	D7S820	D6S1043	D10S124	D1S1656	D12S391	D2S1338	Penta D
S26	17	16	10	11	8	10	30	14	11	10	13	8	25	17	11	10	10	11	13	13	17	17	13
	18	17	11	12	11	16	30	15	12	11	13	9	25	17	11	10	11	13	14	15	21	17	13
S27	15	15	11	12	12	12	31.2	12	11	11	14	6	22	14	11	8	10	12	14	11	18	18	11
	15	18	12	12	12	13	32	16	12	11	16.2	7	23	15	12	12	11	18	14	12	21	19	12
S28	15	15	9	10	8	13	28	18	8	11	13	7	23	15	9	12	10	11	14	12	17	20	11
	18	15	11	10	12	17	31	21	12	11	15.2	8	23	15	12	14	11	18	14	14	18	23	11
S29	15	17	10	10	8	11	29	16	11	10	14.2	6	20	11	10	8	9	12	14	14	21	23	11
	15	19	13	12	9	14	29	19	14	11	15	9.3	20	17	12	11	12	21	15	15	22	27	13
S30	15	17	9	10	8	10	29	17	11	10	12	9	20	15	12	11	10	12	13	12	16	20	9
	16	20	10	12	11	13	30	19	11	11	14	9	20	16	12	14	10	13	17	15	18.3	26	9
S31	15	14	8	10	8	11	29	10	5	11	13	7	21	15	12	12	10	11	14	12	18	19	11
	16	19	11	11	8	13	31.2	17	10	14	14	9	25	15	12	14	13	12	15	17	21	25	14
S32	15	17	10	12	8	10	30	17	12	10	13	8	24	11	12	10	8	14	13	15	22	17	13
	15	18	12	12	10	15	32.2	17	18	11	14	9	26	16	13	10	9	20	14	16	24	21	14
S33	17	15	11	10	8	13	29	14	5	11	13	6	20	15	11	9	11	13	14	15	21	18	10
	17	17	12	11	11	15	32.2	15	12	11	15.2	9.3	21	16	12	12	11	13	16	20.3	21	22	13
S34	15	16	12	10	9	15	29	17	16	11	13	7	21	15	11	10	10	13	13	14	21	22	10
	16	17	13	10	12	15	29	19	18	14	14	9	22	16	13	12	11	16	15	18	22	26	15
S35	15	17	11	11	8	10	29	11	15	11	12	6	22	14	11	11	10	12	14	15	16	18	11
	17	19	11	11	11	15	30	17	18	14	14	8	25	15	13	12	11	14	15	16	20	19	12
S36	15	17	11	10	8	14	29	18	7	10	13	6	21	15	11	11	8	11	13	12	18	17	9
	18	17	12	12	11	14	31.2	18	12	11	13	6	24	16	11	12	9	14	16	15	20	26	12
S37	15	15	11	11	11	13	28	17	7	10	12	6	21	14	11	11	10	11	15	11	18	21	12
	16	17	12	12	12	14	29	19	7	12.3	14	7	23.2	15	12	11	12	14	15	12	22	23	14
S38	14	16	10	11	8	14	29	12	15	12	14	7	24	11	12	8	9	13	14	15	15	19	9
	16	18	12	12	11	15	29	14	18	13	15	9.3	25	13	12	12	11	17	17	17.3	23	19	11

ID	D3S1358	vWA	D16S539	CSF1PO	TPOX	D8S1179	D21S11	D18S51	Penta E	D2S441	D19S433	TH01	FGA	D22S1045	D5S818	D13S317	D7S820	D6S1043	D10S124	D1S1656	D12S391	D2S1338	Penta D
S39	15	18	13	10	8	13	29	11	12	14	12	6	25	14	10	11	9	17	14	14	18	17	11
	18	18	13	12	8	16	30	14	18	14	13	6	25	16	12	12	11	18	16	15.3	20	24	14
S40	15	16	8	10	8	12	28	16	7	11	14	8	23	11	8	12	10	11	12	12	17	18	10
	17	16	11	11	10	13	29	17.2	11	11	15.2	7	25	11	12	13	10	15	12	16	18	20	11
S41	16	15	9	12	10	13	30.2	13	12	10	15	6	21	11	11	12	11	11	14	17	19	17	10
	17	16	13	12	11	13	31.2	14	17	10	16	7	23	15	11	14	12	13	16	17	21	21	13
S42	15	16	9	10	8	13	29	12	11	11	12	8	22	15	10	11	8	11	13	15	18	20	14
	18	17	11	12	8	16	31.2	17	14	11	14.2	9	24	16	12	12	8	14	14	17	22	24	14
S43	16	15	10	11	8	12	30	13	20	10	14	8	21	14	11	9	8	12	15	12	17	20	10
	16	16	12	11	11	13	32.2	15	20	12	16	9	23	15	11	13	12	18	15	14	24	20	13
S44	16	16	11	10	9	11	29	13	16	14	13	8	22	12	10	11	10	12	15	12	18	17	9
	18	18	11	12	11	12	29	18	17	15	14	10	24	16	12	12	11	12	15	15	19	18	10
S45	15	16	11	10	8	12	29	13	12	10	13	8	23	15	11	11	10	11	15	14	20	17	10
	18	17	12	12	11	13	32.2	19	20	12	15.2	9	27	15	11	12	11	11	15	16	22	20	13
S46	15	16	11	10	8	13	30	13	11	11	13	8	20	15	11	11	8	10	13	16	17	16	9
	16	18	13	12	8	13	32.2	13	18	15	15.2	9	24	15	12	13	10	18	15	17	26	20	9
S47	16	15	9	10	8	11	30	15	11	10	14.2	6	20	16	11	11	8	12	14	16	19.3	23	10
	18	18	13	11	11	13	30	19	12	10	15	9.3	20	16	13	13	8	12	15	17	20	25	13
S48	15	17	11	11	10	14	29	17	7	11	13	9	21	15	10	8	8	12	16	16	19	20	11
	16	18	13	11	11	15	29	19	15	11	14	9.3	23	15	11	12	8	12	17	17	19	20	13
S49	16	14	11	11	8	12	29	16	11	10	14	6	28	15	12	11	11	12	14	13	19	17	9
	18	17	12	12	9	16	29	19	21	11	16	9	28	15	13	11	13	13	14	15	21	24	10
S50	15	17	11	11	8	13	28	13	13	11	14	6	23	11	11	8	11	12	13	12	18	18	10
	16	17	12	12	11	14	31	13	16	12	14	6	24	16	12	11	11	20	15	13	19	19	11

Appendix II: DNA Profiles for 3 random unrelated UAE Arabs population




M-179
D22S1045
D5S818
D13S317
D7S820
D6S1043
70
140
210
280
350
420
4400
0
11
16
13
11
14
11
11
12
M-179
D10S1248
D1S1656
D12S391
D2S1338
Penta_D
70
140
210
280
350
420
4100
0
13
15
12
13
20
24
24
11
12
M-179
D3S1358
vWA
D16S539
CSF1PO
TPOX
70
140
210
280
350
420
4000
0
16
17
11
12
9
12
8
11
M-179
D8S1179
D21S11
D18S51
Penta_E
70
140
210
280
350
420
6000
0
1
X
Y
16
29
32.2
12
13
13
17
M-179
D2S441
D19S433
TH01
FGA
70
140
210
280
350
420
3100
0
11
14
13
16.2
9
23
24