11-2017

Study of Y-Chromosome STR Markers in United Arab Emirates Population

Tareq Zeyad Mohammed Ahmed

United Arab Emirates University

College of Science

Department of Biology

STUDY OF Y-CHROMOSOME STR MARKERS IN UNITED ARAB EMIRATES POPULATION

Tareq Zeyad Mohammed Ahmed

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

Under the Supervision of Prof. Rabah Iratni

November 2017

Declaration of Original Work

I, Tareq Zeyad Mohammed Ahmed, the undersigned, a graduate student at the United Arab Emirates University (UAEU), and the author of this thesis entitled "*Study of Y-Chromosome 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 Professor Rabah Iratni, 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 Any materials borrowed from other sources (whether published or unpublished) and relied upon or included in my thesis have been properly cited and acknowledged in accordance with appropriate academic conventions.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: 27/ 11/2017

Approval of the Master Thesis

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

1) Advisor (Committee Chair): prof. Rabah Iratni
   Title: Professor
   Department of Biology
   College of Science
   Signature _____ Date 16/11/17

2) Member: Dr. Rashed Alghafri
   Title: Adjunct Faculty / Head of Biology and DNA Section
   Department of Biology in UAEU / General Department of Forensic Sciences and Criminology in Dubai Police G.H.Q.
   College of Science
   Signature _____ Date 16. Nov. 2017

3) Member: Dr. Ranjit Vijayan
   Title: Associated Professor
   Department of Biology
   College of Science
   Signature _____ Date 16/11/17

4) Member (External Examiner): Dr. Mohammed Nazier
   Title: Assistant Professor
   Department of Biotechnology
   Institution: University of Modern Science
   Signature _____ Date 16/11/17

This Master Thesis is accepted by:

Dean of the College of Science: Professor Ahmed Murad

Signature _____ Date 29/11/2017

for Dean of the College of Graduate Studies: Professor Nagi T. Wakim

Signature Ali Hassan Date 29/11/2017

Copy 4 of 9

Advisory Committee

1) Advisor: Prof. Rabah Iratni

Title: Professor

Department of Biology

College of Science

2) Co-advisor: Dr. Rashed Alghafri

Title: Adjunct Faculty

Department of Biology

College of Science

Abstract

The recently introduced 6-dye Yfiler Plus multiplex which includes 27 Y-STR loci (DYS576, DYS389I, DYS635, DYS389II, DYS627, DYS460, DYS458, DYS19, YGATAH4, DYS448, DYS391, DYS456, DYS390, DYS438, DYS392, DYS518, DYS570, DYS437, DYS385 a/b, DYS449, DYS393, DYS439, DYS481, DYF387S1a/b and DYS533) has been used to study 343 UAE Arab male individuals using Yfiler Plus® amplification kit. This set includes seven rapidly mutating loci (RM Y-STRs). These RM Y-STRs are useful for discriminating between closely related and unrelated males..

According to measures of genetic diversity the highest diversity were observed at loci DYS385=(0.94984), DYF387S1=(0.930523) and DYS449=(0.895402). Therefore, these loci should be considered the most diverse and polymorphic for forensic testing which can be used to distinguish between male relatives. 313 haplotypes were observed in UAE Arab male population and 15 haplotypes were shared between two individuals. Discrimination capacity for 27 loci among the UAE Arab male population was determined to be 95.43% whereas haplotype diversity was found to be 0.99973. AMOVA results showed that UAE Arab male population was placed at far genetic distance from European populations such as Denmark, Italy, Spain and United States. While it shows closer genetic distance to the regional populations from Iran, Iraq, Egypt, Yemen and Kuwait.

Keywords: Y- chromosome, STR, haplotype, genetic diversity, discrimination capacity, haplotype diversity.

Title and Abstract (in Arabic)

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

الملخص

باستخدام التقنية التي أدخلت مؤخراً (6-dye Yfiler Plus multiplex) والتي تحتوي على 27 موقعاً في الكروموسوم الذكري (DYS576, DYS389I, DYS635, DYS389II, DYS627, DYS460, DYS458, DYS19, YGATAH4, DYS448, DYS391, DYS456, DYS390, DYS438, DYS392, DYS518, DYS570, DYS437, DYS385 a/b, DYS449, DYS393, DYS439, DYS481, DYF387S1a/b and DYS533) تم دراسة 343 عينة من أفراد شعب دولة الامارات العربية المتحدة العرب الذكور باستخدام (Yfiler Plus® amplification kit). هذه المجموعة من المواقع تحتوي ايضاً على سبعة مواقع ذات تحور موضعي سريع (RM Y-STRs). هذه المواقع ذات التحور الموضعي السريع مفيده للتمييز بين الاقارب الذكور وايضا بين الذكور غير الاقارب.

وبعد تحليل نتائج التنوع الجيني لوحظ أن أعلى تنوع جيني كان في موقع = DYS385 0.94984، DYF387S1 = 0.930523 و DYS449 = 0.895402. ولذلك، ينبغي اعتبار هذه المواقع الاكثر تنوعاً وتعدداً لفحوص الادلة الجنائية والتي يمكن استخدامها للتمييز بين الأقارب الذكور. وقد لوحظ 313 نمط فردي في شعب الامارات العرب الذكور ، كما تم ملاحظة 15 نمط مشترك بين شخصين. تم تحديد القدرة على التمييز ل 27 موقعاً بين شعب الامارات العرب من الذكور لتكون 95.43٪ في حين أن التنوع النوعي كان 0.99973. وأظهرت نتائج (AMOVA) أن شعب الامارات العرب الذكور متباعدة وراثياً عن الشعوب الاوروبية مثل الدنمارك وإيطاليا وإسبانيا والولايات المتحدة. في حين أظهرت مسافة وراثية أقرب إلى شعوب الدول المحيطة مثل إيران والعراق ومصر واليمن والكويت.

مفاهيم البحث الرئيسية: الكروموسوم الذكري، نمط، التنوع الجيني، القدرة على التمييز، التنوع النوعي.

Acknowledgements

I would first like to thank my thesis advisors Prof. Rabah Iratni and Dr. Rashed Alghafri for supporting and encouraging me through my thesis. The door to Prof. Rabah Iratni and Dr. Rashed Alghafri offices were always open whenever I ran into a trouble spot or had a question about my research or writing. They consistently allowed this thesis to be my own work, but steered me in the right direction whenever they thought I needed it.

Special thanks go to my parents, brothers, sisters and my wife who helped me along the way 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 for his 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. And I would like to thank Dubai Health Authority for providing me with the blood samples.

Dedication

To my beloved mother, father, wife and to my sons

Table of Contents

Title……i

Declaration of Original Work ii

Copyright iii

Advisory Committee iv

Approval of the Master Thesis v

Abstract vii

Title and Abstract (in Arabic) viii

Acknowledgements ix

Dedication x

Table of Contents xi

List of Tables xiii

List of Figures xiv

List of Abbreviations xv

Chapter 1: Introduction 1

1.1 A Brief History of DNA Typing 1

1.2 Human Genome 2

1.3 Structure of Human Genome 5

1.4 Human Y-Chromosome 6

1.5 Short Tandem Repeats (STRs) 8

1.5.1 Types of STR Loci 10

1.5.2 Y- STR Loci 11

1.6 STR Genotyping 13

1.6.1 Principles of FTA Cards 13

1.6.2 Principles of Electrophoresis 13

1.6.3 Capillary Electrophoresis Reagents 16

1.7 Objectives 16

Chapter 2: Materials and Methods 17

2.1 Sample Collection 17

2.2 Preparation and Sample Batching 17

2.3 Storage Condition 17

2.4 DNA Extraction and Quantification Process 19

2.5 PCR Amplification Process 20

2.5.1 Kit Overview 20

2.5.2 Multiplex PCR Using Yfiler Plus® PCR Amplification Kit 21

2.5.3 FTA Punch Procedure 23

2.5.4 Thermal Cycling Parameters for Yfiler Plus® PCR Kit 23

2.5.5 Multiplex PCR Using RM-YPlex Kit: 24
2.5.6 Thermal Cycling Parameters for RM-YPlex PCR Kit 25
2.6 Separation and Detection Methods of Yfiler Plus® PCR Kit Product 26
2.6.1 Sample Preparation Before Capillary Electrophoresis 26
2.7 Statistical Analysis 29

Chapter 3: Results 31
3.1 Profiling Samples 31
3.2 Y-STR Loci Frequencies 33
3.3 Gene Diversity for 27 Loci in UAE Arab Male Population 37
3.4 Haplotype Analysis 38
3.5 Forensic and Population Genetic Parameters 40
3.6 AMOVA Results in UAE Arab Male Population 41
3.7 Abnormal Alleles 44
3.8 RM-Yplex 47

Chapter 4: Discussion 48

Chapter 5: Conclusion 58

References 60

Appendix 66

List of Tables

Table 1: Kit content and storage conditions 18
Table 2: Yfiler Plus® PCR Amplification Kit loci and alleles 21
Table 3: Kit component quantity for each sample 22
Table 4: Size of the disc for test samples and negative control and the amount of positive control 23
Table 5: Thermal Cycling Parameters Yfiler Plus® PCR Amplification Kit 24
Table 6: RM-Yplex kit component quantity for each sample 25
Table 7: Thermal Cycling Parameters RM-YPlex PCR Amplification Kit 25
Table 8: Dyes used in the Yfiler Plus® kit 27
Table 9: Alleles frequency of locus DYS385a/b for UAE Arab male population 33
Table 10: Predominant alleles for the 27 loci detected in UAE Arab male population using Yfiler Plus® amplification kit 36
Table 11: Genetic diversity for Yfiler plus® loci in UAE Arab male population. Loci with an asterisk are added to the Yfiler Plus kit compared to the Yfiler kit 37
Table 12: Haplotype frequency for shared haplotype. 39
Table 13: Percentage of discrimination capacity and haplotype diversity 40
Table 14: Fst AMOVA results using 27 loci within different populations 41
Table 15: Fst AMOVA results using 17 loci within different populations 41
Table 16: UAE Arab male population intermediate alleles 46
Table 17: Predominant alleles compared with Different populations. Allele shared between UAE Arab male populations are highlighted 53
Table 18: Comparison of HD and DC between UAE population and other populations 54
Table 19: Intermediate allele in different loci for different populations 57
Table 20 (A-X): Alleles frequency of different loci with bar charts for UAE Arab male population 66

List of Figures

Figure 1: Human genome chromosomes in diploid cell 3
Figure 2: Gene composition 4
Figure 3: DNA double strands structure 6
Figure 4: Schematic of X and Y chromosome 8
Figure 5: The structure of three commonly used STR loci 11
Figure 6: Capillary electrophoresis process 14
Figure 7: Schematic of capillary electrophoresis instruments used for DNA analysis 15
Figure 8: FTA extraction process 19
Figure 9: GeneMapper ID-X Software plot of the Yfiler Plus® Allelic Ladder 28
Figure 10: Y STR profile generated using Yfiler plus® kit 31
Figure 11: Alleles frequency of locus DYS385a/b for UAE Arab male population 34
Figure 12: Allele frequency for the predominant alleles in the UAE Arab male population 36
Figure 13: Genetic diversity for Yfiler plus® loci in UAE Arab male population 37
Figure 14: Multidimensional scaling (MDS) plot based on pairwise Fst genetic distance among UAE Arab male population and other European populations using the 27 Y-STR Yfiler Plus® kit 42
Figure 15: Multidimensional scaling (MDS) plot based on pairwise Fst genetic distance among UAE Arab male population and other Arab populations as well as Iran population using the 17 Y-STR Yfiler® kit 43
Figure 16: Different profiles shows the null alleles in different loci (null alleles indicated by circles) 44
Figure 17: Different profiles shows the extra alleles in different loci (extra alleles indicated by circles) 45
Figure 18: RM-YPlex results of abnormal alleles in DYF387S1 and DYS570 loci 47

List of Abbreviations

AMOVA	Analysis of Molecular Variance
CE	Capillary Electrophoresis
CODIS	Combined DNA Index System
DC	Discrimination Capacity
HD	Haplotype Diversity
MDS	Multidimensional Scaling
RM	Rapidly Mutating
STR	Short Tandem Repeats

Chapter 1: Introduction

1.1 A Brief History of DNA Typing

In 1944 DNA (deoxyribonucleic acid) was defined by Oswald Avery as the chemical basis for specific and apparently heritable transformation [1]. Frances Harry Compton Crick (1916-2004) was a British molecular biologist, biophysicist, and neuroscientist while James Dewey Watson (1928) was an American molecular biologist, geneticist, and zoologist. Both were working together on the DNA project in 1951, Crick and Watson were putting together the evidences that came from the work of other scientists such as Chargaff's realization that Adenine (A) equal to Thymine (T) and Cytocine (C) equal to Guanine (G), Rosalind X-ray diffraction photo and her explanation. In February 1953, they discovered the structure of DNA as double helix, and the paper was published on 25 April 1953 in Nature. Cardboard cutouts was used for DNA double helix model to represent the individual chemical components of the four bases and other nucleotide subunits. They showed in their model that each strand of DNA was a template for the other. DNA's structure discovery was the most important biological work of the last century [2-4].

Fifty years ago, differentiation between human individuals was obtained by using blood groups such as ABO system or serum proteins and red blood cell enzymes. In some forensic applications such as paternity testing Human Leukocyte Antigen (HLA) was used due to its association with tissue types. In 1970s sufficient number of regions of DNA were analyzed resulting in higher variability between individuals which shifted the study of human variation from the protein products of DNA to DNA itself [5].

In 1980, the first discovery of Restriction Fragment Length Polymorphisms (RFLP), which is small variation found between people at the genetic level was achieved by David Botstein and co-workers and it was considered as a landmark in the construction of a human gene map [6]. In 1985, Jeffreys and his colleagues discovered a unique application of RFLP for individual identification which he termed "DNA fingerprint" and it became the concept of human DNA identification. In the same year the first application of DNA typing in forensic casework was achieved in the United Kingdom. It was also initiated in the United States in late 1986 by commercial laboratories and in 1988 by the Federal Bureau of Investigation (FBI) [5].

In 1986 Polymerase Chain Reaction (PCR) was invented by Kary Mullis. PCR was the most important invention in molecular biology which elucidate the structure of DNA [7]. In 1990, the human genome project began under the leadership of James Watson. In June 2000, a draft sequence of the human genome project was announced and it was published in February 2001. Finally, the announcement of the final reference sequence for the human genome was done in April 2003 [8].

1.2 Human Genome

All human individuals share approximately 99.5% of human genome and that is what makes us human beings. It is only 0.5% which varies between individuals and it is reflected in some individual's traits such as eye color, hair color, and blood types. This small portion is of interest to forensic scientists and it makes each individual's DNA unique with the exception of identical twins [9].

Human cells contain 23 pairs of chromosomes. (22) pairs of autosomal chromosome and a pair of sex chromosomes (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 [1, 5, 11]. All body (somatic) cells contain two copies of each chromosome: one inherited from mother and other one inherited from father. This types of cells are called diploid cells. On the other hand gametes cells (sperm and egg) contain only one copy of each chromosome. This types of cells are called haploid cells. Thus, normal human cells contain 23 pairs of chromosomes which means 46 different chromosomes. The haploid human genome contain approximately three billions DNA base pairs, which means six billion base pairs per diploid cell [12].

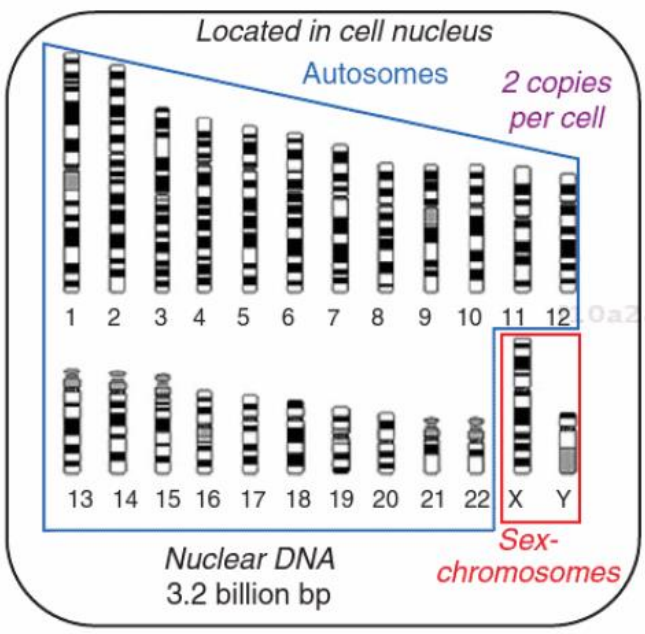


Figure 1: Human genome chromosomes in diploid cell [10]

DNA is identical in all human body cells and it is produced in somatic cells through mitosis. In mitosis chromosomes are replicated and then separated into two new nuclei. The daughter cells are identical to the parent cell as well as to each other. Mitosis daughter cells containing 23 pair chromosomes in each. While in sex cells (gametes) the process of cell division is called meiosis. In meiosis homologous chromosomes separate leading to four daughter cells rather than two. Meiosis daughter

cells are not identical to each other and they contain only one set of each chromosome [13].

Approximately 95% of human DNA is non-genes coding DNA which is not relevant to any proteins production. The other 5% of human genomic DNA is called genes which is responsible for protein structure [5, 9]. A human has approximately 30,000 genes. The average gene is sized from thousands to ten thousand of base pairs. Genes consist of exons (translated to form protein) and introns (the intervening sequences) (Figure 2). Loci that used in forensic human identification are found in the non-coding region either within genes or between genes [15].

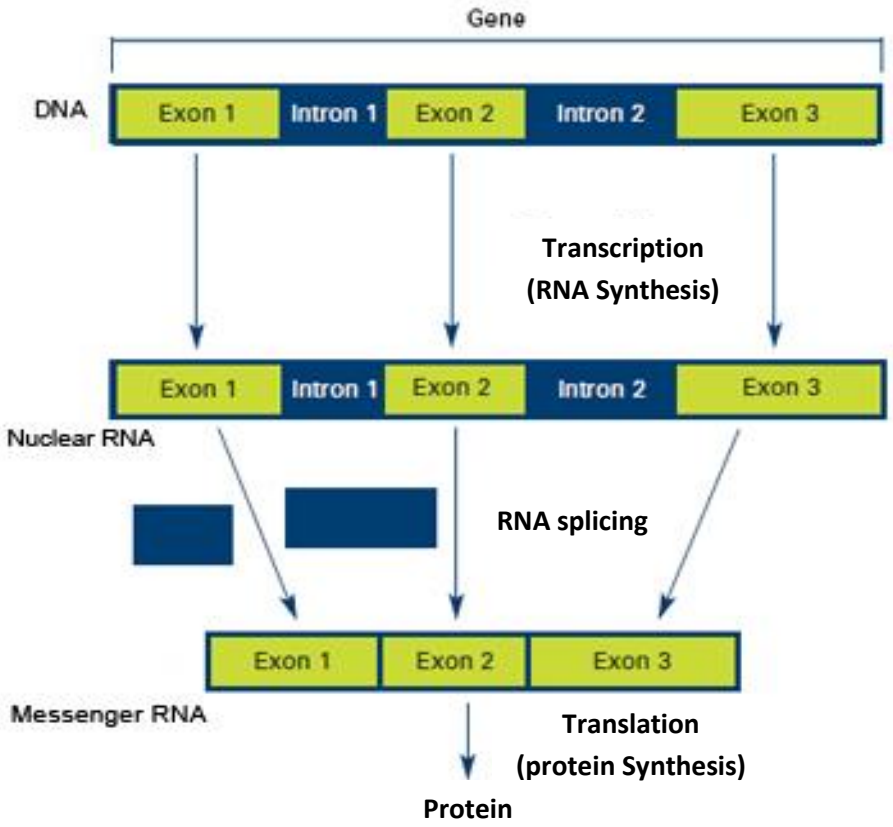


Figure 2: Gene composition [14]

1.3 Structure of Human Genome

DNA carries the hereditary traits which are transmitted from one generation to another. These genetic materials are used in the development, function and reproduction of all known living organisms. DNA consists of a double helical strand, where each strand is a polymer of nucleotides linked by covalent phosphodiester bonds and the two strands are held together by hydrogen bonds. Each nucleotide contains a sugar group, a phosphate group, and a nitrogen base. The four types of nitrogen bases are cytosine (C), guanine (G), adenine (A), and thymine (T). These four bases are divided into two groups; purine bases (adenine and guanine) which have a double ring structure, and pyrimidine bases (thymine and cytosine) which have only a single ring [11] (Figure 3). A purine always pairs with a pyrimidine. In this way the width of the DNA molecule stays the same. The direction of DNA strands is from 5' to 3'. This numbering scheme refers to the position of carbon atoms in the sugar ring. Also the DNA polymerases adding the nucleotides from 5' to 3'. The order of these bases is what determines the genetic code (genes). DNA is too long to fit inside the cell so it coiled to form structure called chromosome [12].

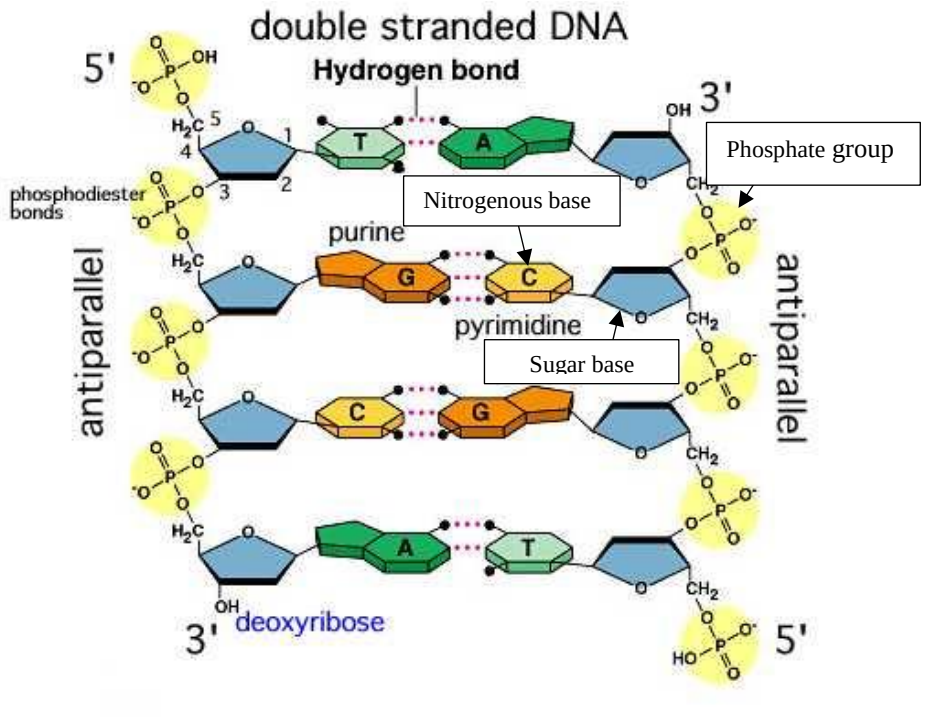


Figure 3: DNA double strands structure [16]

1.4 Human Y-Chromosome

In 1905, Nettie Stevens identified Y chromosome as a sex-determining chromosome. Y chromosome is one of two sex chromosomes which called allosomes. Traits that are inherited via the Y chromosome are called Y linked traits "holandric traits" [17]. Y chromosome contain the SRY gene (Sex-determining Region Y) which encode a protein that triggers testes development [18].

There are everal uses of Y chromosomes testing such as: forensic casework on sexual assault evidence, paternity testing, missing persons investigations, human migration and evolutionary studies, and historical and genealogical research. One of the advantages of using Y chromosome tests in forensic cases where autosomal tests are limited by the evidence such as high levels of female DNA in the presence of minor amount of male DNA [19]. Patrilineal male relatives may be used in paternity testing

or in missing person investigation where the father is not available. An extreme example of using Y chromosome in paternity analysis was in 1998 which helped linking the third US president, Thomas Jefferson to the child of one of his slaves, Sally Hemmings [18].

Y human chromosome represents about 2% of the total human genome and contain 78 genes which codes for proteins [20]. Y human chromosome spans approximately 60 Mb (million base pairs) while X chromosome spans approximately 154 Mb [18].

There are two small portions PAR1 and PAR2 (pseudoautosomal region 1 and pseudoautosomal region 2) located on Y chromosome. PAR1 located at the tip of the short arm (Yp) while PAR2 located at the tip of the long arm (Yq). PAR1 and PAR2 recombine with the X chromosome. These two small region consist only 5% (3 million base pairs) of the Y chromosome. The remaining 95% (57 million base pairs) is non-recombinant (NRY) and is inherited from father to son unchanged [1, 21]. NRY renamed by Skaletsky in 2003 to male specific region (MSY) [21] (Figure 4). Unlike autosomal chromosome most of the Y chromosome can only accumulate variation through mutation [18].

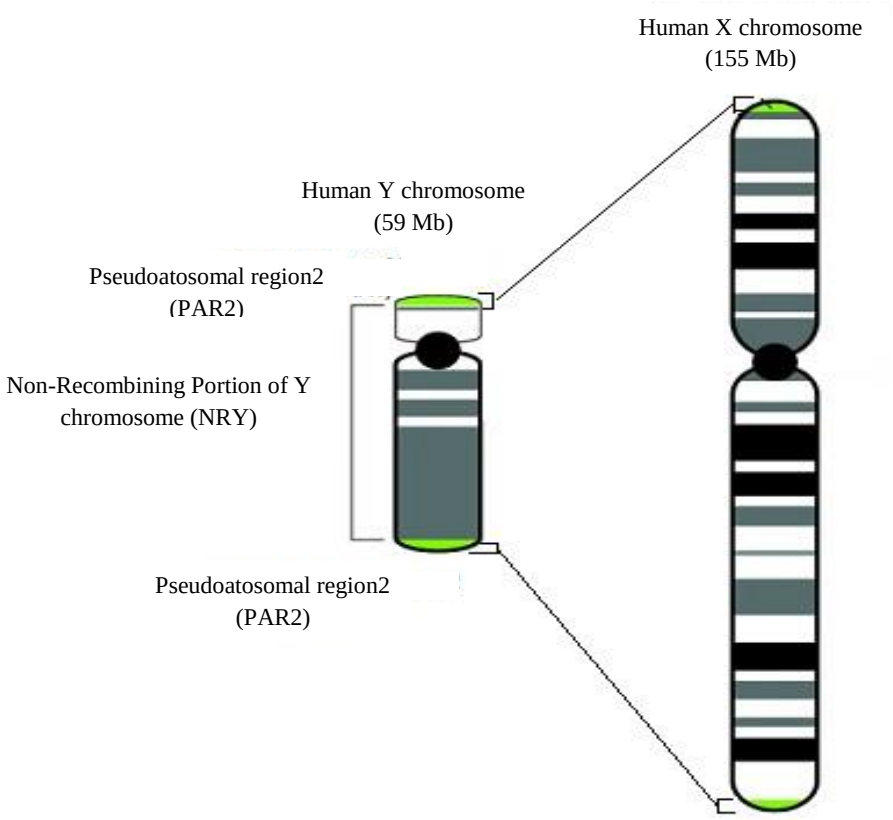


Figure 4: Schematic of X and Y chromosome [22]

1.5 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, individual's DNA do not generally change throughout human lifetime, DNA is less prone to degradation as compared to proteins and lastly high variations are exhibited among individuals and between species [23]. DNA regions with repeat units that are 2 to 7 bp in length are mainly used in forensic applications. These repeats are called microsatellites, simple sequence repeats (SSRs), or most usually short tandem repeats (STRs) and they are mostly found surrounding the chromosomal centromere in all chromosomes. The STRs loci selected for forensic

purposes have been shown to have minimal linkages with diseases and are mainly in non-coding regions of DNA [10]. In 1990s were the first used of STRs in forensic casework. Nowadays STRs usage has become the gold standard application in forensic laboratories around the world [18].

In an effort to establish core STR loci for inclusion within the national DNA database known as CODIS (Combined DNA Index System), the Federal Bureau of Central Investigation (FBI) laboratory sponsored a community-wide forensic science in the beginning of 1996. The 13 CODIS core loci are CSF1PO, FGA, TH01, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, and D21S11. These 13 core STR loci were chosen to be the basis of the future CODIS national DNA database [11]. In April 2011 the FBI announced plans to add more loci to the core 13 loci to become 20 loci which will increase the global DNA data sharing [24].

The average random match probability is rarer than one in a trillion among unrelated individuals for these 13 CODIS loci. However, these 13 loci are by no means the only STRs that have been evaluated or used by forensic labs around the world. Many other loci have been used which are unlinked to the core STR loci (non-CODIS loci) [25]. Analysis of such polymorphic non-CODIS loci in addition to the core CODIS loci may increase the power of exclusion in complex kinship testing involving deficient paternity cases as well as increase the possibility to have more loci in the case of weak or degraded samples [26].

The criteria that used for the selection of STR loci in human identification applications include: firstly, high discriminating power; secondly, separate chromosomal locations to ensure that closely linked loci are not chosen; thirdly, robustness and reproducibility of results when multiplexed with other loci; fourthly,

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; and finally STR alleles that have lower mutation rates [11].

1.5.1 Types of STR Loci

STRs are easily amplified by polymerase chain reaction (PCR) and the size of STRs loci are small and variable among individuals which make these STRs effective for human identification purposes. STR loci vary in the length of the repeat unit and the number of repeats and in the repeat pattern. 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 [24], those are simple repeats or simple repeats with non-consensus alleles, compound repeats, and complex repeats [18] (Figure 5). Not all alleles for an STR locus contain complete repeat units but there are alleles that contain incomplete repeat units called microvariants.

Simple repeats STRs are those which contain one unit of identical length and sequence (e.g. AGATAGATAGATAGAT). Example of this type 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 [24].

Compound repeats comprise two or more adjacent simple repeats (e.g. TCTATCTG TCTATCTG). Example of this type of repeats are FGA, VWA, D3S1358, D8S1179, D2S1338, D19S433, D12S391, and D6S1043 [24].

Complex repeats may contain several repeat blocks of variable unit length as well as variable intervening sequences. Example of this type of repeats is D21S11 [24].

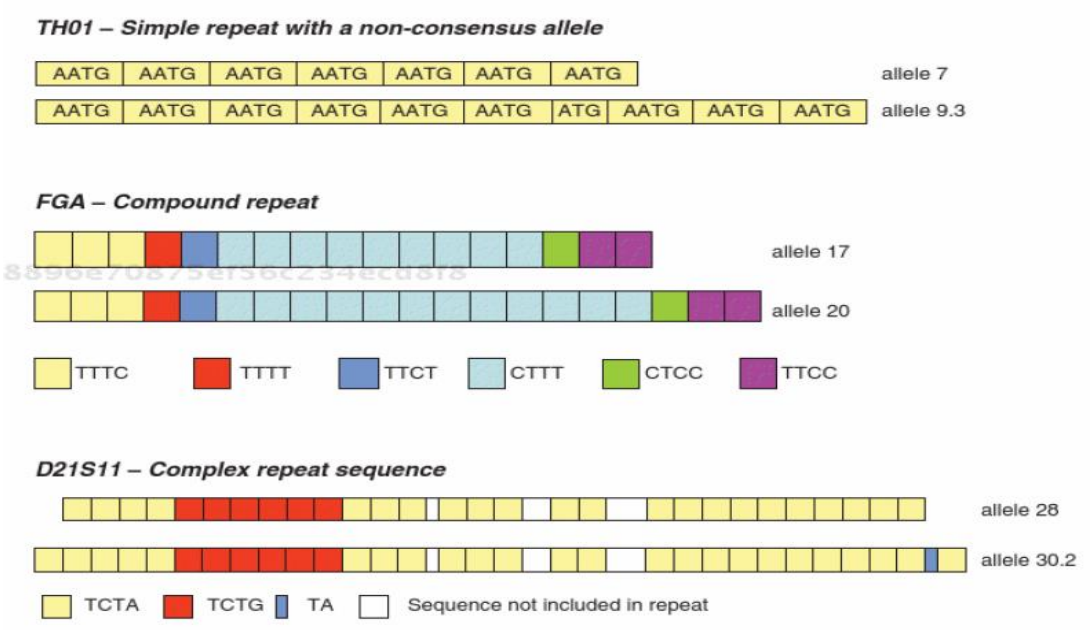


Figure 5: The structure of three commonly used STR loci [18]

1.5.2 Y- STR Loci

There are two ways to analyze Y-chromosome DNA Y-STRs and Y-SNPs. The mutation rate of Y-STR is 1 in 10^3 while the mutation rate of Y-SNPs is 1 in 10^9. That means Y-STRs change more rapidly compared to Y-SNPs. Y-STRs are described as defining haplotypes while Y-SNPs define haplogroups [19]. Y-STR loci were characterized and available for use in 1990s [11]. DYS19 was the first Y-STR locus reported by Roewer in 1992 [20]. Number of Y-STR loci were increased to 30 loci in the beginning of 2002. Nowadays there are hundreds of new Y-STR loci [11].

Using Y-STR typing is helpful in the casework that contain mixtures of DNA from two or more males and when there are female-male mixture even with low level

of male DNA in comparison to high level of female DNA with female:male ratios of up to 2000:1 [18, 20].

In 1997 a core set of Y-STR loci were selected which serves as “minimal haplotype”. These loci are single copy Y-STR loci: DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, and the highly polymorphic multi-copy locus DYS385 a/b. In 2003 the minimal haplotype loci plus two additional single copy Y-STRs, DYS438 and DYS439 were recommended to be used in forensic application by U.S. Scientific Working Group on DNA Analysis Methods (SWGDAM) [27]. In 2001 first kit was offered by ReliaGene Technologies company called Y-Plex6® which amplifies six loci (DYS393, DYS19, DYS389II, DYS390, DYS391 and DYS385 a/b) using only two dye colors (blue and yellow). In 2002 Y-Plex5® were offered by the same company which amplifies five loci (DYS389I, DYS389II, DYS439, DYS438 and DYS392) using three dye colors (blue, green and yellow). In 2003 same company developed kit called Y-Plex12® which amplifies the SWGDAM recommended loci plus the amelogenin locus using three dye colors (blue, green and yellow). In the same year Promega Corporation offered a new kit which amplified the same loci of Y-Plex12® but removed amelogenin and DYS439 loci and add new locus DYS437. In 2004 Applied Biosystems company offered new kit Yfiler® which amplifies 17 loci the SWGDAM recommended loci plus (DYS456, DYS458, DYS437, DYS448, YGATAH4 and DYS635) [11]. In present research, new kit developed by Applied Biosystems company called Yfiler Plus® which amplifies 27 Y-STR loci including the 17 loci in Yfiler plus and ten new loci (DYF387S1(a/b), DYS449, DYS481, DYS518, DYS458, DYS627, DYS570, DYS460 and DYS533) was studied among UAE Arab male population.

1.6 STR Genotyping

1.6.1 Principles of FTA Cards

FTA card was developed as a method for storage of DNA in the late 1980s by Lee Burgoyne at Flinders University in Australia. Nowadays FTA cards are used as an alternative way for DNA extraction. FTA cards contain chemicals that lyse cells, denature proteins, and protect DNA molecules from nuclease degradation and preserve the paper from bacterial growth. DNA is entrapped in the fibers of the matrix of the paper and remain immobilized within the matrix. Therefore FTA card provide a simple solution for storage the DNA at the room temperature for several years, and it is an easy and fast way to isolate nucleic acid samples for analysis [28].

1.6.2 Principles of Electrophoresis

Electrophoresis is a process used to separate DNA fragments from one another based on differences in their migration speed. In the electrophoresis, two electrodes are immersed in two separate buffer chambers (anode which is positively charged, and cathode which is negatively charged). Charged particles can migrate from one chamber to the other based on their net charge by using a voltage generated between the two electrodes. The electrophoretic mobility of the sample can be affected by the net charge, size, shape, molecular weight of the molecules, the buffer type, and the voltage power [29]. In this experiment, Capillary Electrophoresis (CE) was used to resolve the DNA fragments. During capillary electrophoresis the PCR products enter the capillary due to the high voltage charge applied to the sample which forced the negatively charged fragments into the capillary. 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. The laser beam causes the dyes attached to the fragments to fluoresce. The diffraction system separate the dye signals and a Charge Coupled Device (CCD) camera detects the fluorescence. Each dye has different wavelength of the emitted light (figure 6). The data collection software collect the fluorescence signal and convert it into an electropherogram [11].

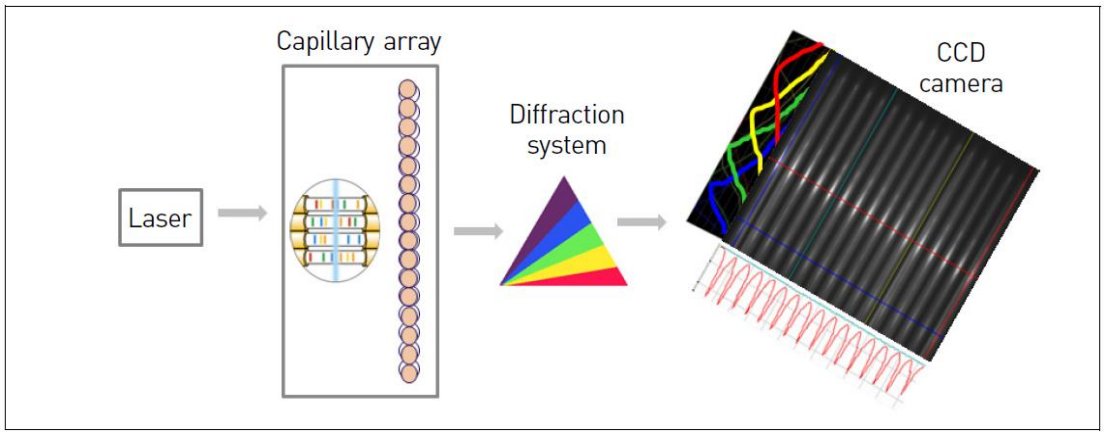


Figure 6: Capillary electrophoresis process
(This figure is adopted from Applied Biosystem 3500 Genetic Analyzer manual)

The advantages of using CE are: first of all it is fully automated in all steps (injection, separation, and detection). Second, only very small amount of DNA samples are consumed in the injection process so sample can be retested if needed. This advantage is important for forensic samples that often cannot be easily replaced. Third, separation time in capillaries is much less than that in slab gel due to higher voltages that are permitted with improved heat dissipation from capillaries. Another advantage is that data can be obtained 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. Lastly, there is no issue of cross-contamination from samples leaking over from adjacent wells with CE.

The primary elements of a CE instrument include a narrow 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, and a computer to control the sample injection and detection (Figure 7).

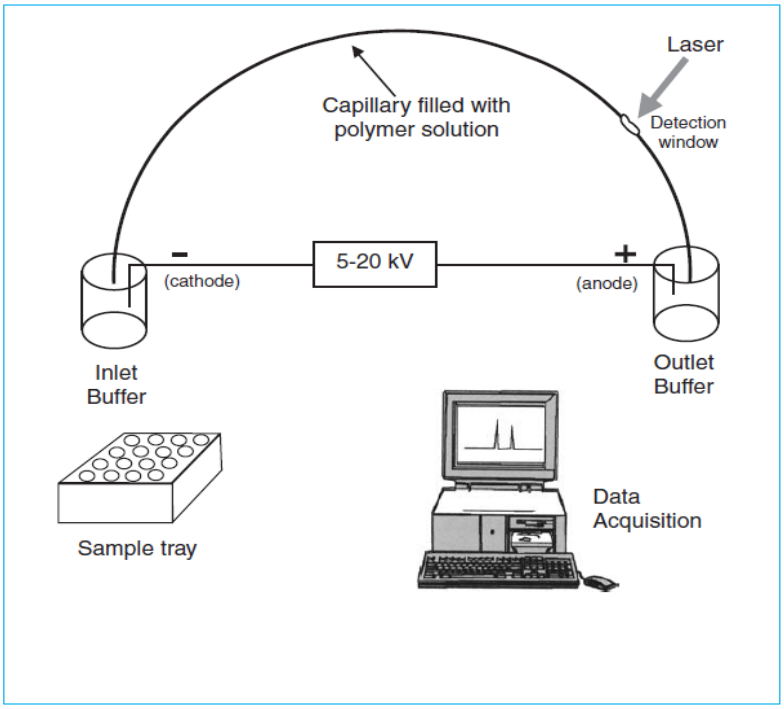


Figure 7: Schematic of capillary electrophoresis instruments used for DNA analysis [10]

In this research 3500 Genetic Analyzers (Applied Biosystems) which is a fluorescence based DNA analysis instrument using capillary electrophoresis technology with 8 capillaries each 36 cm in length. The 3500 Genetic Analyzer is fully

automated, from sample loading to primary data analysis, for sequencing, fragment analysis, and HID (human identification) analysis.

1.6.3 Capillary Electrophoresis Reagents

Performance Optimized Polymer (POP-4):

Performance Optimized polymer (POP-4) dynamically coat the capillary wall to control electro-osmotic flow. It acts as a matrix for the samples to migrate in the capillaries of the capillary electrophoresis. POP-4 polymers are specifically prepared and formulated to separate DNA fragments of a known size range at a desired resolution and run time. Pop4 polymers are designed specifically for Human Identification (HID) forensic applications. Other reagents that required for capillary electrophoresis are Anode (positive charge) and Cathode (negative charge) buffers.

In this study, we hypothesize that the extra nine Y-STR markers that added to Yfiler Plus® PCR Amplification Kit are highly discriminatory and polymorphic in the UAE Arab male population.

1.7 Objectives

There are three objectives for this study:

1. To study the Y chromosome polymorphism among UAE Arab male population.
2. To study the forensic parameters of 27 Y-STR loci among UAE Arab male population.
3. To assess the importance of increasing the number of Y- chromosome STR loci utilized in forensic DNA analysis.

Chapter 2: Materials and Methods

2.1 Sample Collection

Blood samples were collected from 343 males from the United Arab Emirates Arab male population. The samples were collected randomly from unrelated individuals. Consented samples were collected and provided via Dubai Health Authority (DHA). Ethical approval has been obtained from Dubai scientific research ethics committee in (DHA) under the reference number (DSREC-SR-10/2017-02).

2.2 Preparation and Sample Batching

Once the blood sample tubes were received from DHA, one drop from each sample were applied to separate whatman FTA cards (GE Healthcare), Each was assigned a unique barcode number and left to dry.

2.3 Storage Condition

Yfiler Plus® PCR Amplification Kit (Thermofisher Scientific) was used to analyze 27 Y-STR loci in all DNA samples. The kit contents were stored as shown in (Table 1). Kit contents should not be refrozen after thawing. The fluorescent dyes that are attached to the primers are light sensitive. The primer set, amplified DNA, allelic ladder, and size standard should be 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. FTA cards that contains the blood samples were stored at room temperature while the blood samples tubes that received from DHA were stored at -20°C.

Table 1: Kit content and storage conditions
(The table is adopted from manufacture protocol)

Kit contents	Storage conditions
Yfiler Plus® Master Mix Contains MgCl2, 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.
Yfiler Plus® Primer Set Contains locus-specific 6-FAM™, VIC™, NED™, TAZ™, and SID™ dye-labeled primers in buffer. The primers amplify the Y-STR loci DYS19, DYS385 a/b, DYF387S1 a/b, DYS389 I/II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS449, DYS456, DYS458, DYS460, DYS481, DYS518, DYS533, DYS570, DYS576, DYS627, DYS635 (Y GATA C4), and Y GATA H4.	–25°C to –15°C on receipt. 2°C to 8°C after first use. Store protected from light.
Yfiler Plus® Allelic Ladder Contains the following amplified alleles: • 6-FAM™ dye (blue): DYS389I 9-17; DYS389II 24-35; DYS576 10-25; DYS627 11-27; DYS635 15-30. • VIC™ dye (green): DYS19 10-19; DYS391 5-16; DYS448 14-24; DYS458 11-24; DYS460 7-14; Y GATA H4 8-15. • NED™ dye (yellow): DYS390 17-29; DYS392 4-20; DYS438 6-16; DYS456 10-24; DYS518 32-49. • TAZ™ dye (red): DYS385 a/b 32-49; DYS437 10-18; DYS449 22-40; DYS570 10-26. • SID™ dye (purple): DYF387S1 a/b 30-44; DYS393 7-18; DYS439 6-17; DYS481 17-32; DYS533 7-17.	–25°C to –15°C on receipt. 2°C to 8°C after first use. Store protected from light.
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.

2.4 DNA Extraction and Quantification Process

FTA papers that contained the blood samples has been punched (0.5 mm) using micro-puncher (Harries®) and placed into a PCR tubes. The small punched discs were directly proceeded to the PCR reaction without washing step because the Yfiler Plus® PCR Amplification Kit has been designed to resist the hematein and other cell components inhibition effect of the PCR reaction. A major advantage of using FTA card is that the extracted DNA does not required quantification because the amount of DNA in the disc is sufficient for PCR and it is optimized as per the constant size of the disc (Figure 8).

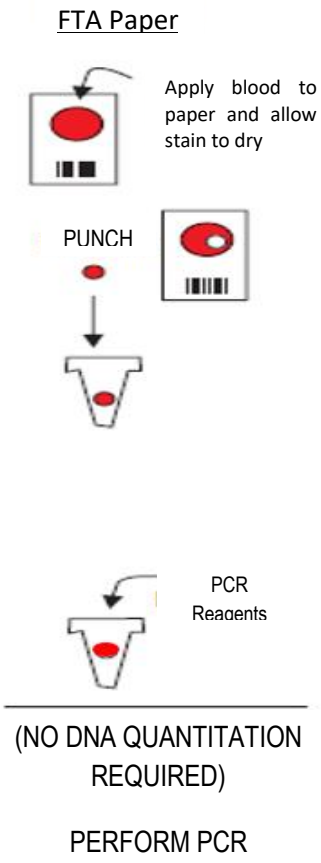


Figure 8: FTA extraction process [10]

2.5 PCR Amplification Process

2.5.1 Kit Overview

Yfiler Plus® PCR Amplification Kit was used to amplify 27 Y-chromosome STR loci. These loci are: " DYS576, DYS389I, DYS635, DYS389II, DYS627, DYS460, DYS458, DYS19, YGATAH4, DYS448, DYS391, DYS456, DYS390, DYS438, DYS392, DYS518, DYS570, DYS437, DYS385 a/b, DYS449, DYS393, DYS439, DYS481, DYF387S1a/b and DYS533". The Yfiler Plus® PCR Amplification Kit includes:

1- Y filer Plus master mix which contains $MgCl_2$ as a co-factor for DNA polymerase, (dNTPs), polymerase enzyme, bovine serum albumin to stabilize the polymerase, and 0.05% sodium azide in buffer and salt as a preservative.
2- Y filer Plus primer set which contains locus-specific (6-FAM dye (blue), VIC dye (green), NED dye (yellow), TAZ dye (red), and SID dye (purple)) dye-labeled and unlabeled primers in buffer.
3- DNA control 007 which contains 2 ng/µL of human male genomic DNA in 0.05% sodium azide and buffer.
4- Y filer Plus Allelic Ladder which contains the most commonly found alleles per locus that obtained from large population base study as shown in (Table 2).

Table 2: Yfiler Plus® PCR Amplification Kit loci and alleles
(The table is adopted from manufacture protocol)

Locus designation	Allele Range included in Allelic Ladder	Dye label	DNA Control 007
DYS576	10 to 25	6-FAM™	19
DYS389I	9 to 17		13
DYS635	15 to 30		24
DYS389II	24 to 35		29
DYS627	11 to 27		21
DYS460	7 to 14	VIC™	11
DYS458	11 to 24		17
DYS19	9 to 19		15
YGATAH4	8 to 15		13
DYS448	14 to 24		19
DYS391	5 to 16		11
DYS456	10 to 24	NED™	15
DYS390	17 to 29		24
DYS438	6 to 16		12
DYS392	4 to 20		13
DYS518	32 to 49		37
DYS570	10 to 26	TAZ™	17
DYS437	10 to 18		15
DYS385a/b	6 to 28		11,14
DYS449	22 to 40		30
DYS393	7 to 18	SID™	13
DYS439	6 to 17		12
DYS481	17 to 32		22
DYS387S1	30 to 44		35,37
DYS533	7 to 17		13

2.5.2 Multiplex PCR Using Yfiler Plus® PCR Amplification Kit

Amplification of the 27 loci were performed according to manufacturer's protocol with some modification. Due to the high cost of the kit and the limited budget for the project (Enough only for 100 full reaction volume samples) and high number of samples that were to be tested, quarter volume has been used. The total samples

which have been included in this study were 343 samples. Some of the kit reagents were consumed to validate the quarter reaction. 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 samples was 6.25µL (Table 3).

Table 3: Kit component quantity for each sample

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

To prepare reaction mix first the Y-filer Plus® master mix and primer set were thawed, then vortexed for 3 seconds followed by quick spin down. Thawing is required only during first use of the kit. After first use, reagents are stored at 2 to 8 °C and, therefore, they do not require subsequent thawing.

The required volumes of components were prepared according to the number of samples plus positive and negative controls then it was pipetted into a 2.0 mL microcentrifuge Eppendorf tube. Additional reaction volume was included in the calculations to provide excess volume for the loss that occurs during reagent transfers and to cover pipetting errors. The reaction mix was vortexed for 3 seconds, then centrifuged briefly.

2.5.3 FTA Punch Procedure

To prevent static issues with the paper discs, the 6.25 µL reaction mix was dispensed into the PCR tube before adding the punches (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 tube). Then 0.5 mm disc of the FTA paper of each sample was punched and placed into a separate labeled PCR tubes which is preloaded with 6.25 µL of the reaction mix (for negative control 0.5 mm punch of blank FTA paper was added) (Table 4). 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 4: 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

2.5.4 Thermal Cycling Parameters for Yfiler Plus® PCR Kit

All PCR reactions consisted of initial incubation cycle at 95°C for 1 minute, followed by 27 cycles of denaturation at 94°C for 4 seconds, and annealing and extension at 61.5°C for 1 minute per cycle. The PCR reactions were completed by post-extension for 22 minutes at 60°C as shown in (Table 5).

Table 5: Thermal Cycling Parameters Yfiler Plus® PCR Amplification Kit

Initial incubation step	Optimum cycle number		Final extension	Final hold
	Denature	Anneal/Extend		
HOLD	27 CYCLE		HOLD	HOLD
95°C, 1 minute	94°C, 4 seconds	61.5°C, 1 minute	60°C, 22 minutes	4°C, up to 24 hours

2.5.5 Multiplex PCR Using RM-YPlex Kit:

RM-Yplex kit was used to confirm the results of abnormal loci following previously published method [31]. 0.5 mm disc of the FTA paper of each sample was punched and placed into a separate labeled PCR tubes. Then FTA disc was washed using 200µL FTA buffer for 30 to 45 minutes. After that FTA buffer was carefully removed from the tube. Then disc was washed again with 200 µL TE buffer for 5 minutes. RM-Yplex master mix and primer set were vortexed for 3 seconds followed by quick spin down step. The required volumes of components were prepared according to the number of samples plus positive and negative controls then it was pipetted into an appropriately 2.0 mL microcentrifuge Eppendorf tube. 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 samples was 15µL. Then a total amount of 15 ml of the reaction mix has been transferred into each PCR tubes (Table 6).

Table 6: RM-Yplex kit component quantity for each sample

DNA sample	Volume per reaction (µL)
RM-Yplex Master Mix	7
RM-Yplex Primer Set	1.8
Distilled Water	6.2
Total Volume	15

2.5.6 Thermal Cycling Parameters for RM-YPlex PCR Kit

All PCR reactions consisted of initial incubation cycle at 95°C for 10 minutes, followed by (stage 2) 12 cycles of denaturation at 94°C for 30 seconds, annealing at 58°C for 45 seconds and extension at 72°C for 1 minute per cycle. Then (stage 3) 20 cycles of denaturation at 94°C for 30 seconds, annealing at 55°C for 45 seconds and extension at 72°C for 1 minute per cycle. PCR reactions were completed by post-extension for 60 minutes at 72°C as shown in (Table 7).

Table 7: Thermal Cycling Parameters RM-YPlex PCR Amplification Kit

Stage 1	Stage 2			Stage 3			stage 3	Final hold
	Denature	Anneal	Extend	Denature	Anneal	Extend		
HOLD	12 CYCLE			20 CYCLE			HOLD	HOLD
95°C, 10 minutes	94°C, 30 seconds	58°C, 45 seconds	72°C, 1 minute	94°C, 30 seconds	55°C, 45 seconds	72°C, 1 minute	72°C, 60 minutes	4°C, up to 24 hours

2.6 Separation and Detection Methods of Yfiler Plus® PCR 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. Electrophoresis is a process that used for fragments separation and is either performed in a slab-gel or capillary environment.

2.6.1 Sample Preparation Before Capillary Electrophoresis

After the end of PCR, samples were removed from the thermocycler and proceed to CE setup using 3500 Genetic Analyzer. 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.

1. **Hi-Di Formamide**

Hi-Di Formamide is a highly deionized formamide to stabilize denatured DNA samples for fluorescence detection using the Genetic Analyzer. Formamide is a reagent that is ionizing solvent in aqueous buffer. It's widely used in biochemistry and molecular biology, especially in nucleic acids research. Formamide in DNA sample helps to stop reannealing of the DNA single strands after denaturation at a high temperature (95 °C) or eliminate formation of secondary structure of nucleic acid.

2. **GeneScan 600 LIZ Size Standard**

Fragment lengths generated from DNA samples were calculated by direct comparison with constant set of fragments of known size. GeneScan size standards were fluorescently labeled DNA fragments of known size and it has different dye color from the PCR products as shown in (Table 8). Fragments of known sizes were used by the data analysis software to generate a calibration curve, which is then used to

determine the size of unknown fragments in a sample. The GeneScan 600 LIZ Size standard is designed to size DNA fragments in the 20–600 bp range and provides 36 single stranded labeled fragments of: 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 bp.

Table 8: Dyes used in the Yfiler Plus® 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. Yfiler Plus® Allelic Ladder

The ladder consist 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 9).

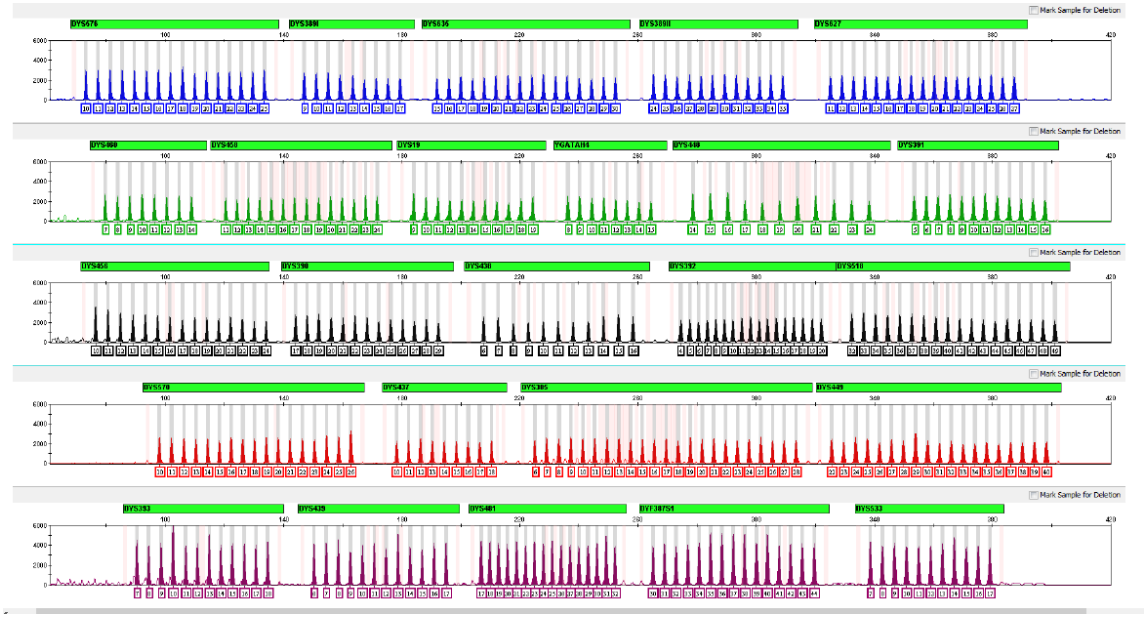


Figure 9: GeneMapper ID-X Software plot of the Yfiler Plus® Allelic Ladder (The figure is adopted from the manufacture's protocol)

Master mix was created with the following components ratio for each sample:

- 0.5µL of LIZ-600 Size Standard

- 9.5µL of HiDi-Formamide

10µL of master mix was dispensed into each well used. To prevent injection into a dry well 10µL of HiDi-Formamide was added into extra available wells of the injection plate. 0.5 µL of PCR products was added to each well. In case of samples

need to be re-injected, 0.5 µL of PCR products was added again. 0.5 µL of Y filer Plus® allelic ladder was added to each ladder sample. A minimum of 1 ladder sample was presented per plate. 0.5 µL of positive control sample (007) was added to each positive control well. A minimum of one positive control sample was presented per plate. 0.5 µL of negative control was added into the negative control well. A minimum of one negative control sample was presented per plate. Then plate was covered with a septa. Plate was briefly vortexed and briefly centrifuged. After that plate was placed into the 9700 GeneAmp PCR System for denaturation for 5 minutes at 95°C and 2 minutes at 4°C.

2.7 Statistical Analysis

Data analysis was carried out using GeneMapper ID-X version 1.4 analysis software. It analyzes data generated from the capillary electrophoresis platforms such as 3500 Genetic analyzer. In this project, this software was used to analyze the results and to generate raw data for all samples. The raw data file generated from the GeneMapper ID-X was analyzed using Microsoft Excel and an integrated software package for population genetics data analyses Arlequin version 3.5 to estimate allele and haplotype frequencies [32]. By using Arlequin software, Discrimination Capacity (DC) was calculated using the following formula:

$$Discrimination\ Capacity = \frac{Number\ of\ unique\ hap.}{Number\ of\ hap.}$$

Using the frequency results from Arlequin, Haplotype Diversity (HD) was calculated using the following formula:

$$HD = \frac{N}{N-1}(1 - \sum x^2)$$

Where x is the (relative) haplotype frequency of each haplotype in the sample and N is the sample size. Haplotype diversity is a measure of the uniqueness of a particular haplotype in a given population.

HapYDive program was used to estimate the Genetic diversity for the 27 loci in UAE Arab male population. The Genetic distance was identified using the analysis of molecular variance (AMOVA and MDS) tool available at the YHRD website (http://www.yhrd.org). Several methods have been used to score the variations and diversity of Y STR in UAE Arab male population. F statistic (F_{st}) or fixation index is one of the most used biometric method. F_{st} simply measures the level of heterozygosity or differentiation between populations. F_{st} ranges from 0 to 1 where 0 indicating no differences in allele frequencies between two populations and 1 indicating that the two populations are fixed for alternate alleles. In this study the UAE population was compared with other populations in different part of the world. Using the same software Nonmetric Multidimensional Scaling MDS was used to represent the distance between populations using bidimensional scaling graph.

Chapter 3: Results

3.1 Profiling Samples

343 samples that obtained from unrelated males in the UAE were profiled with Yfiler Plus® PCR Amplification Kit for 27 Y-chromosome loci. An example of the Y STR profile result is shows in the Figure 10.

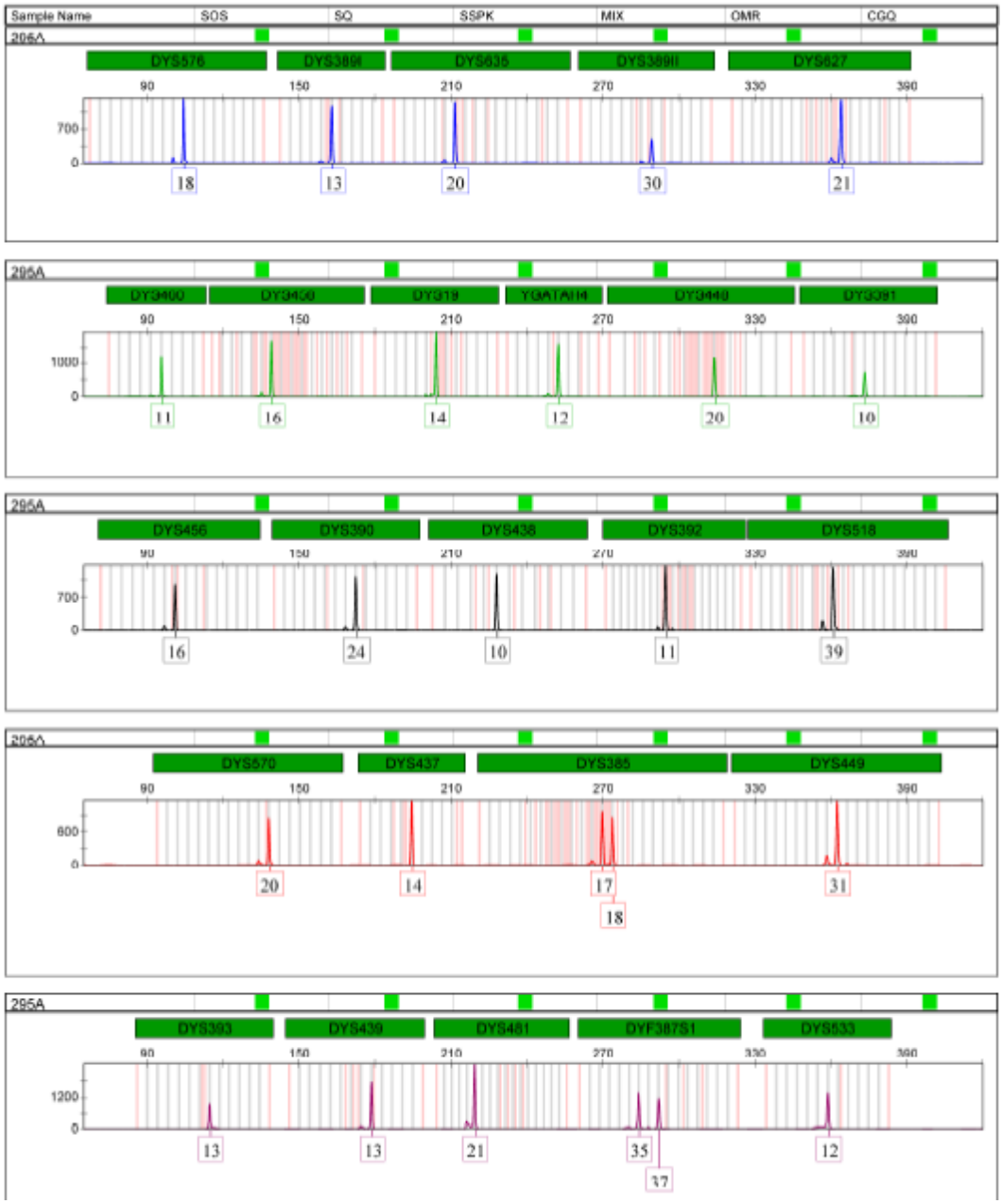


Figure 10: Y STR profile generated using Yfiler plus® kit

This profile shows the different dyes that correspond to different loci. Moreover, it is obvious there are 25 loci shown in this profile, but because there are two multi-copy loci available in this profile (DYS385(a/b) and DYF387S1(a/b), the number of total alleles detected using this kit is brought up to 27 alleles. Numbers which are shown below each peak indicate the number of repeats for each allele following the recommended nomenclature by ISFG [31, 33].

3.2 Y-STR Loci Frequencies

Table 9: Alleles frequency of locus DYS385a/b for UAE Arab male population

Alleles for the locus DYS385(a/b)							
Haplotype No.	Haplotype	Freq.	Percentage	Haplotype No.	Haplotype	Freq.	Percentage
1	10,14	1	0.3	24	14,15	9	2.6
2	11,12	2	0.6	25	14,16	25	7.3
3	11,13	1	0.3	26	14,17	6	1.7
4	11,14	31	9	27	14,18	9	2.6
5	11,15	8	2.3	28	14,19	6	1.7
6	11,16	2	0.6	29	14,20	2	0.6
7	12,13	4	1.17	30	14,21	3	0.9
8	12,14	9	2.6	31	14	4	1.17
9	12,15	1	0.3	32	15,16	5	1.5
10	12,17	4	1.17	33	15,17	2	0.6
11	12,18	1	0.3	34	15,18	4	1.17
12	12,19	3	0.9	35	15,19	2	0.6
13	12,20	4	1.17	36	15,20	1	0.3
14	12	4	1.17	37	15	4	1.17
15	12.2,18	1	0.3	38	16,17	15	4.4
16	13,14	4	1.17	39	16,18	8	2.3
17	13,15	7	2	40	16,19	2	0.6
18	13,16	14	4.1	41	16	3	0.9
19	13,17	19	5.5	42	17,18	21	6.1
20	13,18	20	5.8	43	17,19	3	0.9
21	13,19	46	13.4	44	17	8	2.3
22	13,20	11	3.2	45	18	2	0.6
23	13,21	1	0.3	46	19	1	0.3
Total						343	100

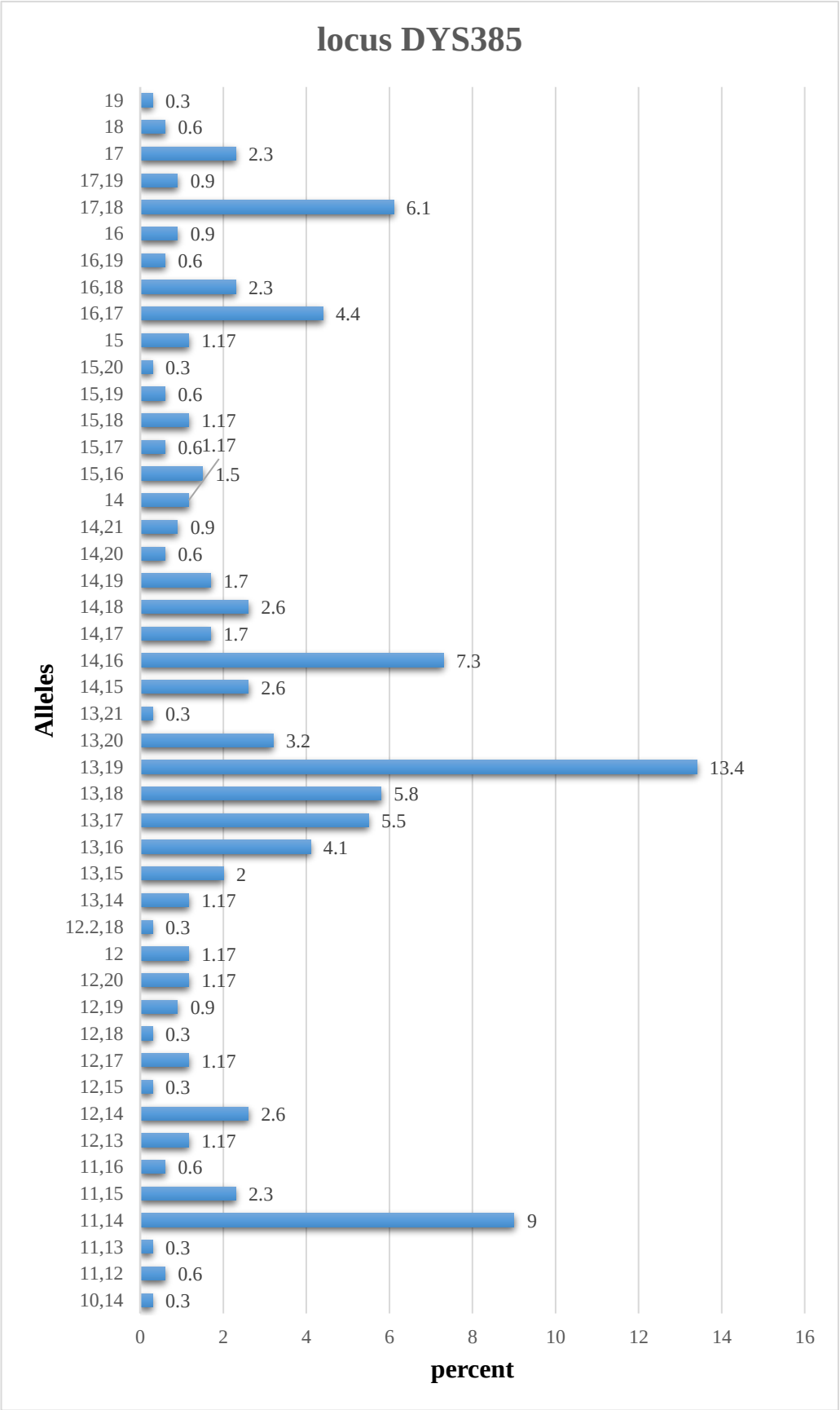


Figure 11: Alleles frequency of locus DYS385a/b for UAE Arab male population

Table 9 present a detailed description of allelic distribution for locus DYS385a/b and the corresponding bar chart (Figure 11). This locus has different number of alleles among the UAE Arab male population with different degree of prevalence. As we can notice that locus DYS385(a/b) has 46 alleles among the UAE Arab male population. Alleles Frequency for the other loci has been added in (Appendix).

Table 10: Predominant alleles for the 27 loci detected in UAE Arab male population using Yfiler Plus® amplification kit

Locus	Predominant Alleles	Freq.	Locus	Predominant Alleles	Freq.
DYS392	11	0.764	DYS456	15	0.461
DYS437	14	0.717	DYS390	23	0.458
DYS391	10	0.685	DYS393	13	0.391
DYS389I	13	0.644	DYS576	17	0.297
YGATAH4	11	0.557	DYS627	20	0.292
DYS19	14	0.534	DYS570	18	0.265
DYS448	20	0.531	DYS518	39	0.248
DYS389II	30	0.501	DYS458	16	0.23
DYS438	10	0.501	DYS481	23	0.222
DYS635	21	0.478	DYS449	26	0.19
DYS460	10	0.475	DYS385(a/b)	13,19	0.134
DYS439	11	0.466	DYF387S1(a/b)	37	0.134
DYS533	11	0.464			

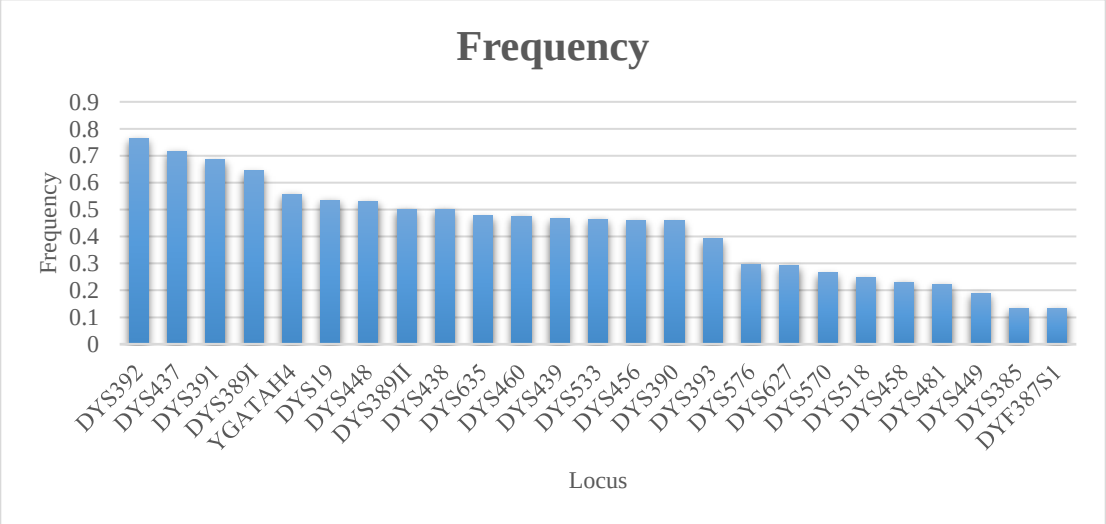


Figure 12: Allele frequency for the predominant alleles in the UAE Arab male population

Analysis of the allele frequency from (Table 9) and (Appendix) in the UAE Arab male population clearly showed that each locus has predominant alleles as shown in Table 10 and Figure 12.

3.3 Gene Diversity for 27 Loci in UAE Arab Male Population

Table 11: Genetic diversity for Yfiler plus® loci in UAE Arab male population. Loci with an asterisk are added to the Yfiler Plus kit compared to the Yfiler kit

Locus	Diversity	Locus	Diversity
DYS385(a/b)	0.94984	DYS456	0.676845
DYF387S1(a/b)*	0.930523	DYS19	0.643377
DYS449*	0.895402	DYS438	0.643326
DYS481*	0.854262	DYS393	0.641826
DYS518*	0.848294	DYS533*	0.628817
DYS458	0.837178	DYS448	0.625595
DYS627*	0.819907	DYS460*	0.601112
DYS570*	0.802107	YGATAH4	0.590183
DYS576*	0.784836	DYS389I	0.525174
DYS390	0.707705	DYS391	0.455561
DYS635	0.707023	DYS437	0.437932
DYS389II	0.684074	DYS392	0.402315
DYS439	0.678039		

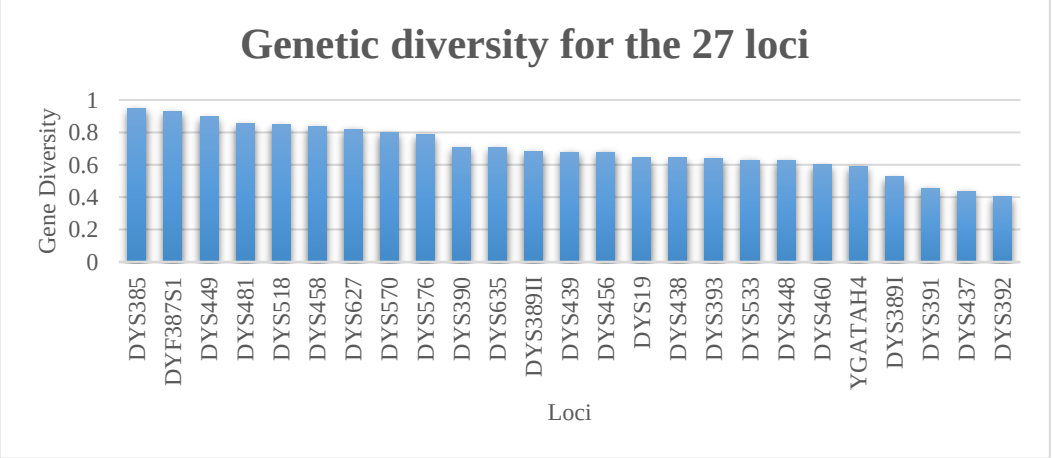


Figure 13: Genetic diversity for Yfiler plus® loci in UAE Arab male population

Genetic diversity was calculated for all 25 loci using HapYDive Microsoft Excel Macro [34]. Table 11 shows the genetic diversity for all loci and it is clear that the multi-copy loci DYS385(a/b) and DYF387S1(a/b) have the highest genetic diversity while DYS392 has the lowest genetic diversity. Genetic diversity has been represented in bar chart to make it more readable and clear (Figure 13).

3.4 Haplotype Analysis

By using the 27 loci there were 15 haplotypes shared between two individuals. Whereas 313 individuals have unique haplotype (Table 12 A). While using only 17 loci there were 27 haplotypes shared between two individuals. Moreover, there were 8 haplotypes shared between three individuals. Also, there was one haplotype shared between 18 individuals. Whereas the remaining 247 individuals have unique haplotype (Table 12 B).

Table 12: Haplotype frequency for shared haplotype
(MP represent match probability which is sum of the squared haplotype frequencies where N is the number of sample)

A. Yfiler plus® (27loci)

Shared Haplotype	27 Y-STR Yfiler Plus®				
	Freq.	(Freq.)²	No. of Hap.	MP	N
1	0.00292	$8.526*10^{-6}$	313	0.00267	313
2	0.00583	$3.399*10^{-5}$	15	0.00051	30
Total	-	-	328	0.00318	343

B. Yfiler® (17loci)

Shared Haplotype	17 Y-STR Yfiler®				
	Freq.	(Freq.)²	No. of Hap.	MP	N
1	0.00292	$8.526*10^{-6}$	247	0.00211	247
2	0.00583	$3.399*10^{-5}$	27	0.000918	54
3	0.00875	$7.656*10^{-5}$	8	0.000613	24
18	0.0525	$2.756*10^{-3}$	1	0.00276	18
Total	-	-	283	0.00640	343

3.5 Forensic and Population Genetic Parameters

Table 13: Percentage of discrimination capacity and haplotype diversity

Forensic Parameter of Y chromosome in UAE	Yfiler Plus®	Yfiler®
samples in population	343	343
Number of haplotypes	328	283
Number of unique haplotypes	313	247
Discrimination Capacity (%)	95.43	87.28
Haplotype Diversity	99.973	99.65

Discrimination capacity (DC) means that the number of haplotypes observed only once in the population. The DC calculated for both 27 loci and 17 loci by dividing the number of unique haplotype over the total number of the haplotypes. DC for 27 loci among the UAE Arab male population was determined to be 95.43% and for the 17 loci was determined to be 87.28% (Table 13). Whereas haplotype diversity (HD) is measure of the uniqueness of a particular haplotype in a given population. The (HD) was calculated using the following equation:

$$HD = \frac{N}{N-1}(1 - \sum x^2)$$

Where x is the relative haplotype frequency of each haplotype in the sample and N is the number of sample [35]. Haplotype diversity in the UAE Arab male population when using 27 loci was 99.973 while when using 17 loci it was 99.65 (Table 11).

3.6 AMOVA Results in UAE Arab Male Population

Table 14: F_{st} AMOVA results using 27 loci within different populations

Population	UAE	Denmark	Germany	Hungary	Italy	Serbia	Spain	United States
UAE	-							
Denmark	0.0001	-						
Germany	0.0002	0.0001	-					
Hungary	0.001	0.0009	0.001	-				
Italy	0.0001	0	0.0001	0.0009	-			
Serbia	0.0002	0.0001	0.0001	0.001	0.0001	-		
Spain	0.0001	0	0.0001	0.0009	0	0.0001	-	
United States	0.0001	0	0.0001	0.0009	0	0.0001	0	-

Table 15: F_{st} AMOVA results using 17 loci within different populations

Population	UAE	Egypt	Iran	Iraq	Jordan	Kuwait	Sudan	Tunisia	Yemen
UAE	-								
Egypt	0.0005	-							
Iran	0.0004	0.0003	-						
Iraq	0.0005	0.0003	0.0002	-					
Jordan	0.0074	0.0072	0.0071	0.0072	-				
Kuwait	0.0021	0.0019	0.0019	0.0019	0.0088	-			
Sudan	0.087	0.0878	0.0849	0.0893	0.0948	0.088	-		
Tunisia	0.0035	0.0033	0.0034	0.0035	0.0104	0.0051	0.0912	-	
Yemen	0.0011	0.0008	0.0008	0.0008	0.0078	0.0025	0.0917	0.0041	-

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

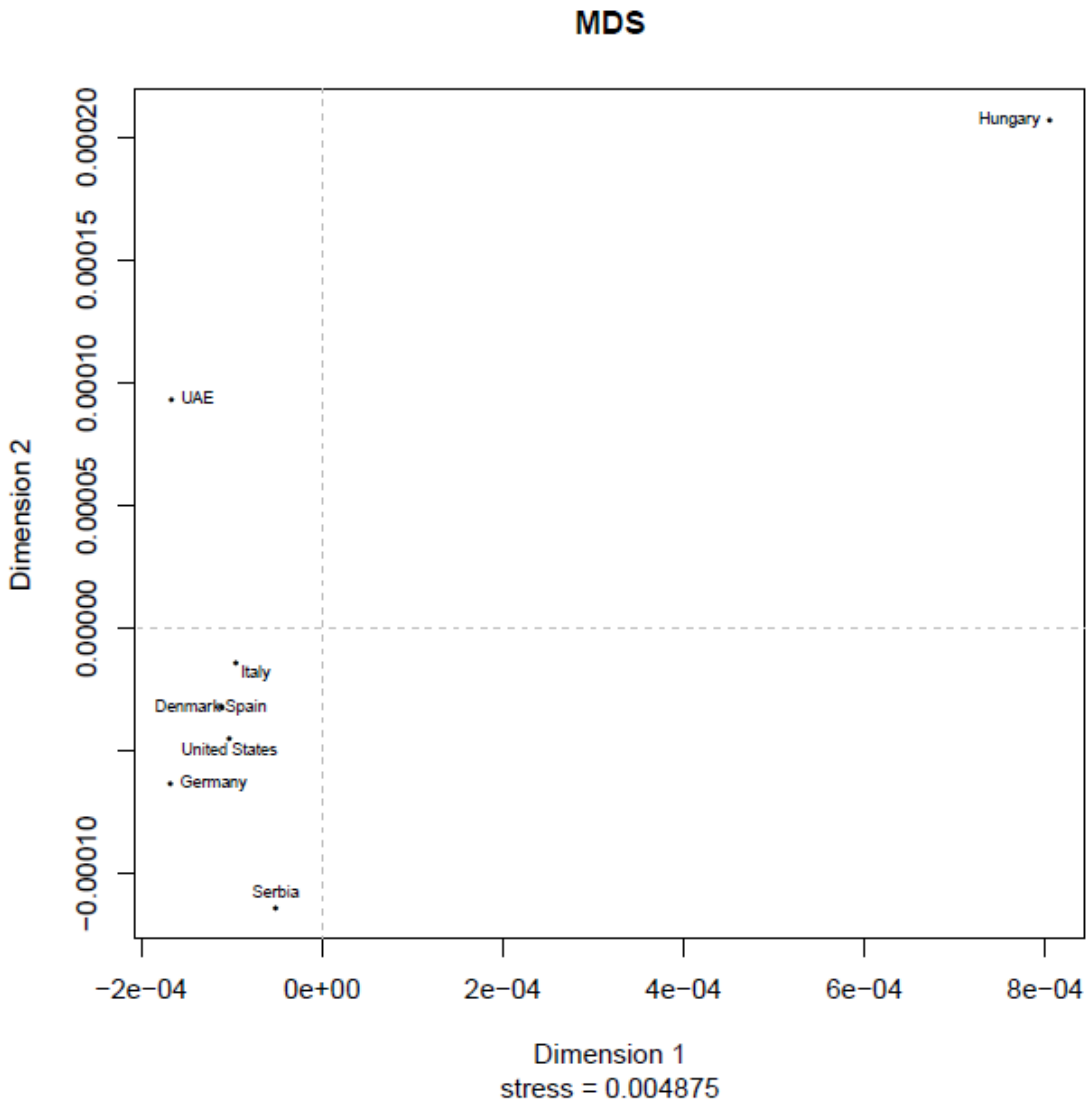


Figure 14: Multidimensional scaling (MDS) plot based on pairwise F_{st} genetic distance among UAE Arab male population and other European populations using the 27 Y-STR Yfiler Plus® kit

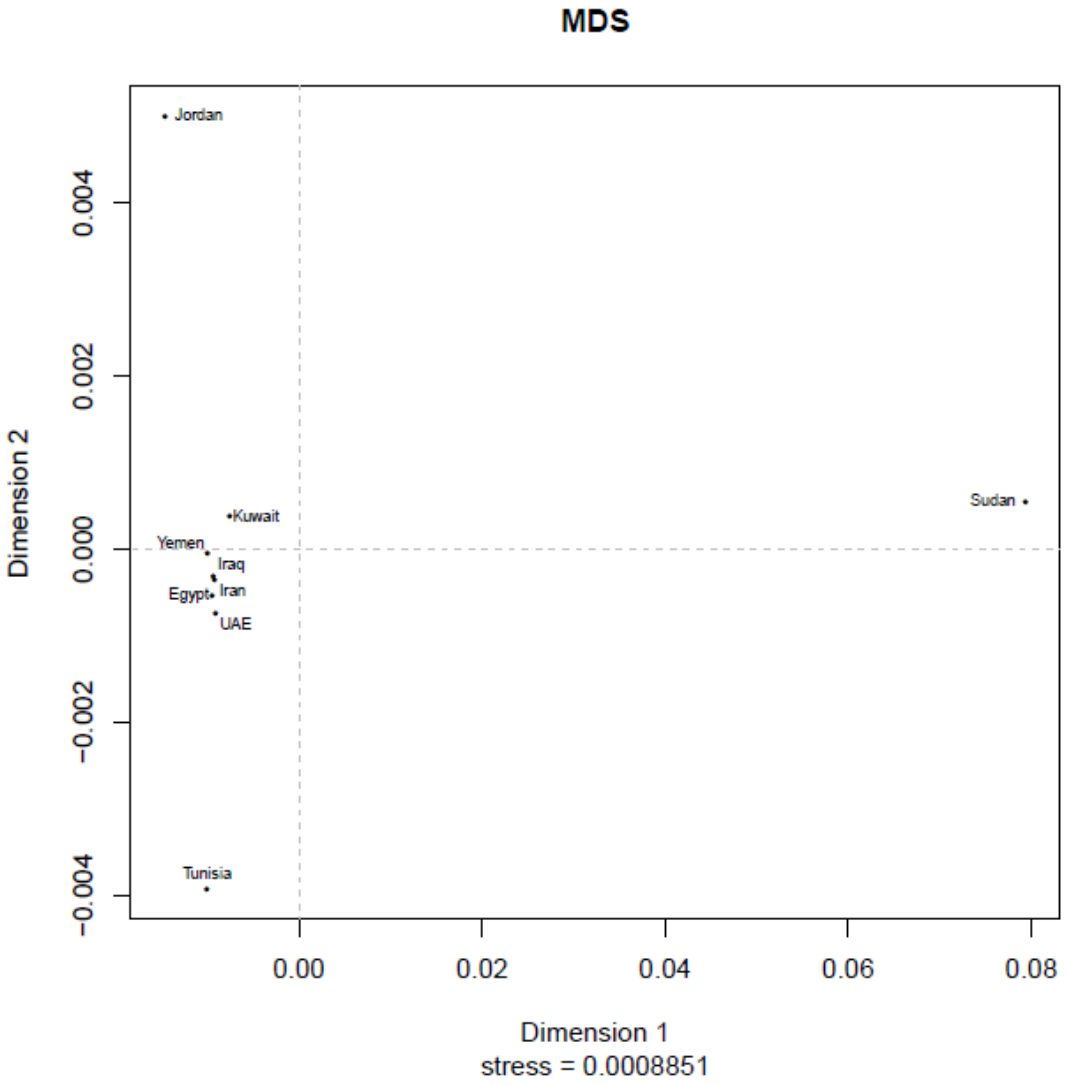


Figure 15: Multidimensional scaling (MDS) plot based on pairwise F_{st} genetic distance among UAE Arab male population and other Arab populations as well as Iran population using the 17 Y-STR Yfiler® kit

Multidimensional scaling (MDS) plot based on pairwise F_{st} genetic distance using the 27 Y-STR Yfiler® kit (Figure14) among UAE Arab male population and other European populations. While (Figure15) shows (MDS) plot based on pairwise F_{st} genetic distance among UAE Arab male population and other Arab populations as well as Iran population using the 17 Y-STR Yfiler® kit.

3.7 Abnormal Alleles

1- Null Alleles

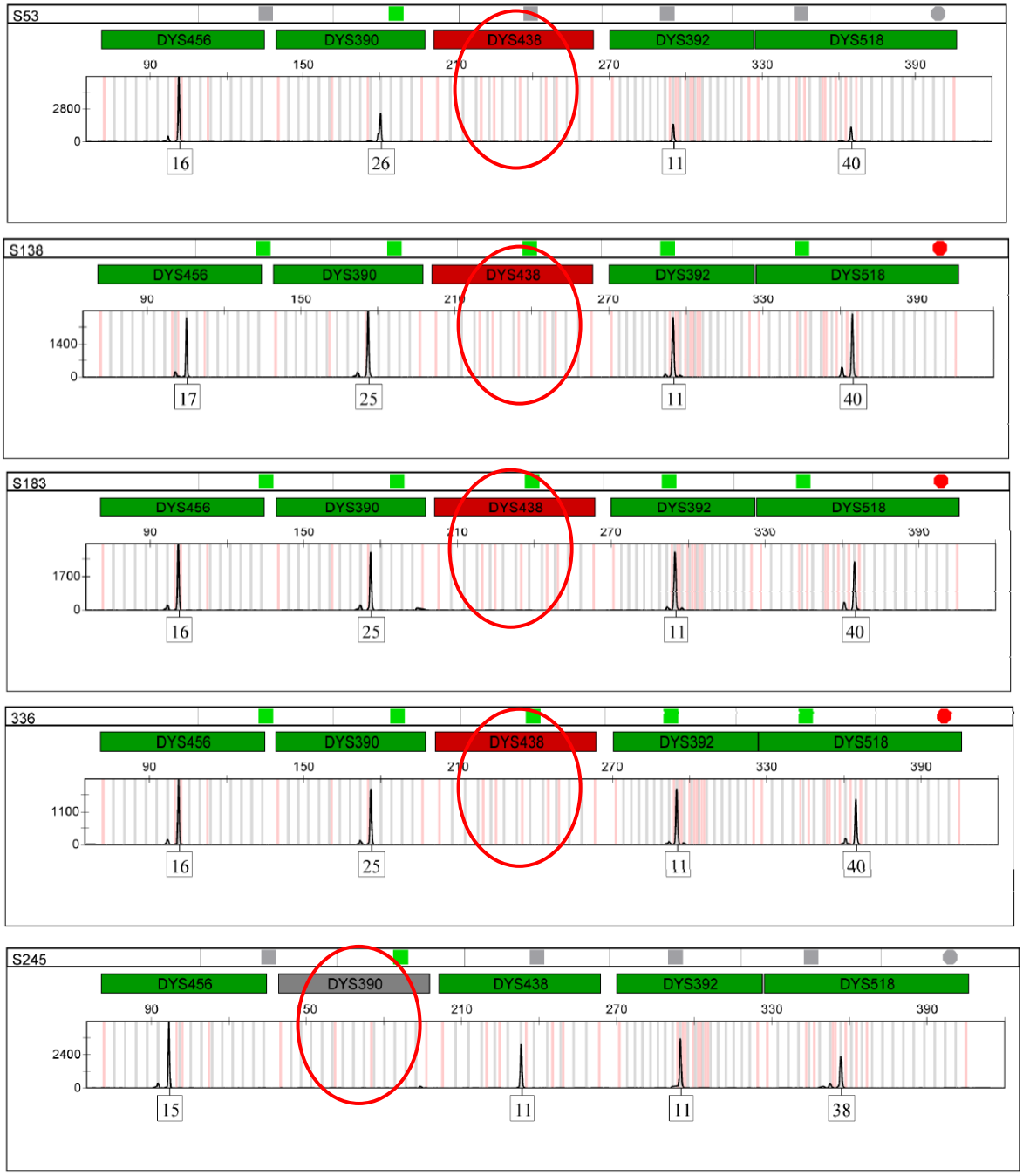


Figure 16: Different profiles shows the null alleles in different loci (null alleles indicated by circles)

2- Multiple alleles

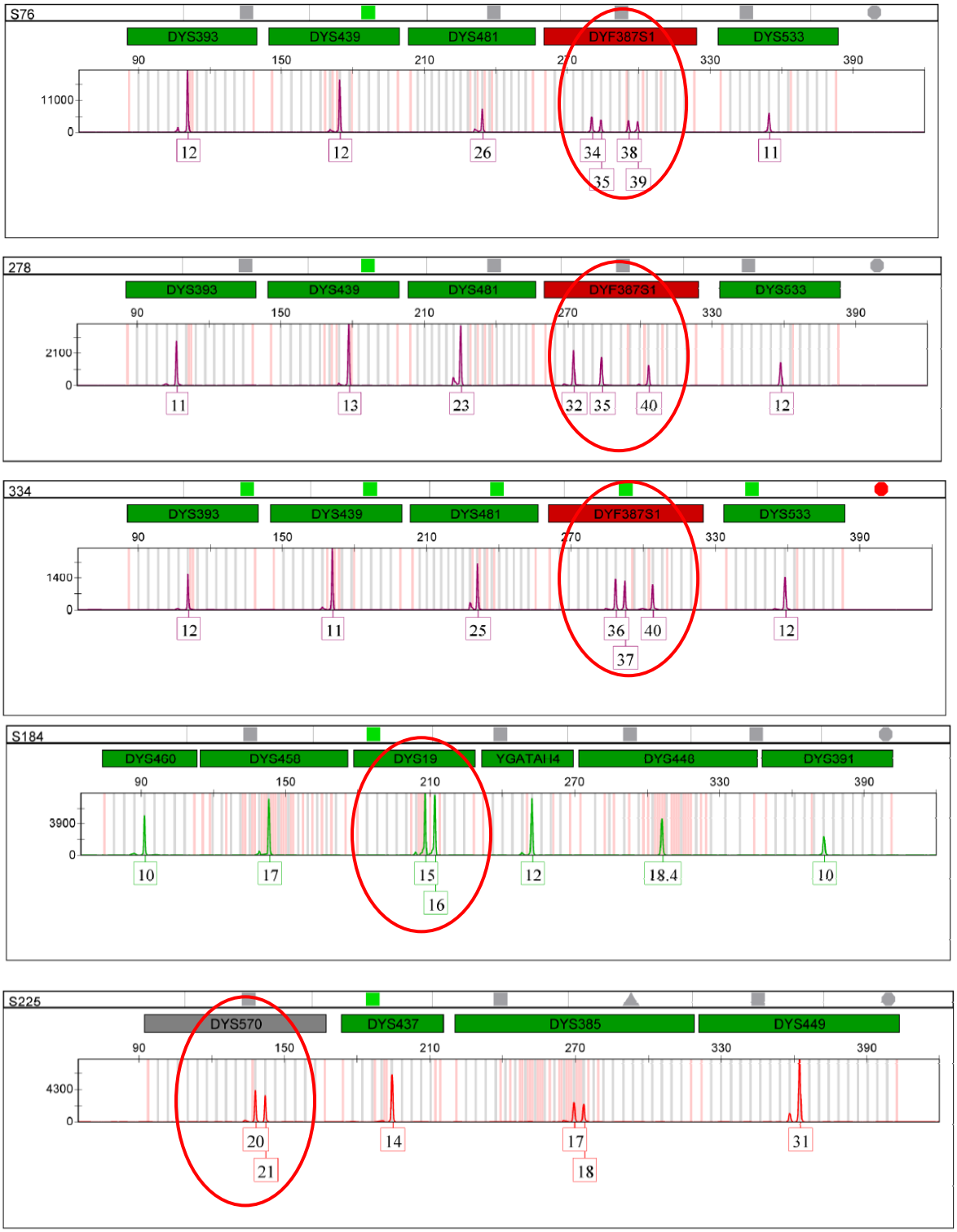


Figure 17: Different profiles shows the extra alleles in different loci (extra alleles indicated by circles). DYS385 is a multi-copy locus

3- Intermediate Alleles

Table 16: UAE Arab male population intermediate alleles

Loci	No. of samples	Intermediate alleles
DYS458	90	16.2, 17.2, 18.2, 19.2, 20.2, 21.2
DYF387S1(a/b)	3	9.2, 41.2, 42.2
DYS627	2	17.2, 21.2
DYS576	1	17.2
YGATAH4	1	8.2
DYS448	1	18.2
DYS518	1	38.1
DYS385(a/b)	1	12.2
DYS481	1	28.1

Null alleles were observed in four individuals at DYS438 and in one individual at DYS390 (Figure 16). Four alleles were found in one individual in multi-copy locus DYF387S1a/b and three alleles in two individuals in the same locus. While there were two alleles found in single allelic loci DYS19 and DYS570 (Figure 17). In all the samples, 101 different intermediate alleles were found in 9 loci whereas locus DYS458 showed the highest number (six different intermediate alleles with 16.2, 17.2, 18.2, 19.2, 20.2 and 21.2 where observed in 90 individuals) (Table 16).

3.8 RM-Yplex

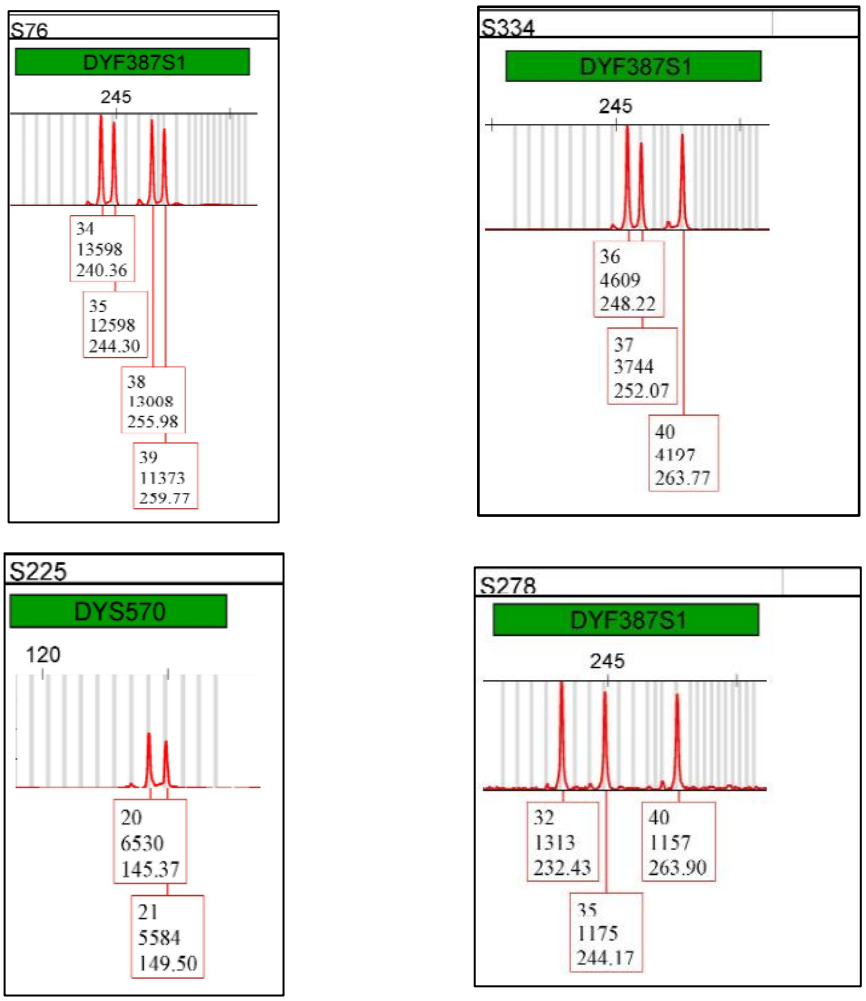


Figure 18: RM-YPlex results of abnormal alleles in DYF387S1 and DYS570 loci

Abnormal loci which have multiple alleles (DYF387S1 and DYS570) has been repeated using RM-YPlex. Figure 18 shows the RM-YPlex which confirm the results that has been produced by Yfiler Plus® kit.

Chapter 4: Discussion

Over the past decades forensic genetics has been continually improved specially in the quality of STR analysis. In this study forensic parameters have been determined and alleles distribution was investigated in the UAE Arab male population in comparison to other male populations using the new 27 Y-STR Yfiler Plus® amplification kit. It was found that, compared to the 17 Y-STR Yfiler® kit, the new Yfiler Plus® kit shows greater power of discrimination, greater haplotype diversity and the additional loci shows more genetic diversity. The additional loci are (DYF387S1(a/b), DYS449, DYS481, DYS518, DYS458, DYS627, DYS570, DYS460 and DYS533). Seven of these additional loci are rapidly mutating (RM) Y-STRs.

Allele frequency of 343 UAE Arab male individuals that scored for the 27 Y-chromosome STRs are shown in (Table 9) and (Appendix). The most polymorphic loci were DYS385a/b with 46 total number of alleles (Table 9) and DYF387S1a/b with 36 total number of alleles (Appendix). Corresponding to that DYS385a/b and DYF387S1a/b exhibit the highest genetic diversity 0.94984 and 0.930523 respectively (Table 11). These loci have such high diversity because it is a multi-copy loci. DYS458, DYS481, DYS518 and DYS449 have 13, 13, 12 and 12 total number of alleles respectively (Appendix) with genetic diversity of 0.837178, 0.854262, 0.84894, and 0.895402 respectively. Loci which exhibited lower polymorphisms were DYS460, DYS439 and DYS533 which have a total number of 6 alleles (Appendix) with genetic diversity 0.601112, 0.678039 and 0.628817 respectively. DYS393 have 5 total number of alleles with 0.641826 genetic diversity. DYS389I, DYS391 and DYS437 have 4

total number of alleles with genetic diversity of 0.525174, 0.455561, and 0.437932 respectively.

According to measures of genetic diversity the highest diversity were observed at locus DYS385=0.94984, DYF387S1=0.930523 and DYS449=0.895402. Therefore, these loci should be considered the most diverse and polymorphic for forensic testing which are useful to differentiate between male relatives. While loci with the lowest diversity such as loci DYS391=0.455561, DYS437=0.437932 and DYS392=0.402315 are the least polymorphic loci and are more useful for paternity lineage testing. Among the nine markers showed the highest gene diversity, seven are new in the Yfiler Plus® kit compared to the Yfiler® kit (Table 11). These results confirm what has been reported in previous studies for different populations such as Italian, Austrian, Nigeria, Serbia and Eastern Turkey populations [27, 36–39]. As there are no data or researche published yet for Arab populations in Arabian Peninsula using the 27 Y-STR Yfiler Plus® the results of 17 Y-STR Yfiler® loci of -Iraq, Syria and Tunisia male populations [40–43] were used to compare with the present study results and it was found that there is no significant difference between all the results taking into consideration forensic parameters and loci diversity.

Although the number of alleles at each locus affects the genetic diversity, it is not necessarily that the diversity should be higher for loci with high number of alleles detected. Because genetic diversity affected by the way of the prevalence of the locus alleles among the population. For example DYS392 has 7 total number of alleles while DYS437 has 4 total number of alleles but the genetic diversity for these loci are 0.402315 and 0.437932 respectively. A total number of 266 distinct alleles were observed across the 27 loci in the UAE Arab male population.

Alleles are distributed in any population at different rates. The analysis of the allele frequency in the UAE Arab male population clearly showed that each locus has predominant allele (Table 10 and Figure 12). There was difficulty in determining a predominant allele in some loci such as DYS458, DYS481, DYS449 and DYS570 (Appendix). DYS458 has 4 alleles (16, 15, 17 and 18.2) which distributed among UAE Arab male population in close ratio. While DYS481 has 3 alleles (23, 25 and 26) which almost has the same probability of prevalence in the UAE Arab male population. Moreover, DYS449 almost has equally distributed alleles which means this locus is good for discrimination because the more the allele is equally distributed the high genetic diversity. So, it is clearly noticed that DYS449 has the highest genetic diversity among the single-copy loci (Table 11). In addition, for locus DYS570 it was not easy to determine the predominant allele because it has 3 alleles which are very close in the prevalence among the UAE Arab male population. There is a high possibility the predominant allele can be changed in these loci in case of increasing the number of the samples. (Table 10) shows the most predominant alleles in UAE Arab male population were allele 11 of DYS392 locus which was shared among 76.4% of the population; almost similarly allele 14 of DYS437 locus was shared among 71.7% of the population, allele 10 of DYS391 locus was shared in 68.5% of the population and allele 13 of DYS389I locus is shared among 64.4% of the population which mean these alleles in such loci should be treated carefully in interpretation process for paternity testing because there is high possibility to have the same allele in these loci shared among non-related individuals. Even with minimal rate of allele prevalence still there are predominant alleles such as allele 13, 19 of the multi-copy locus DYS385(a/b) is shared between 13.4% of the population; similarly, allele 37 of the multi-copy locus DYF387S1(a/b) is shared among 13.4% of the population. In fact, this study shows

approximately half of the population share at least 9 alleles across the 27 STR loci. Comparison of allele frequencies in UAE Arab male population with other populations [40, 42, 44–52] showed that UAE Arab male population share most of its predominant alleles with Turkey (shares 12 loci), Tunisia (shares 10 loci), India (shares 9 loci) and Syria and Iraq (shares 8 loci) populations (Table 17). It can be observed that there were shared alleles across most of the populations. DYS389I shared allele 13 among all populations except Iraq and Finland populations which shared allele 14. And allele 14 in DYS19 was shared among all population except Somalia, Tunisia and India populations which have different predominant alleles 11, 13 and 15 respectively. Likewise, there were alleles shared with neighboring populations such as Iraq, Turkey and Syria populations. These alleles are allele 10 in locus DYS391 and allele 11 in locus DYS92.

A total of 313 unique haplotypes were obtained from 343 individuals using the 27 Y-STR Yfiler plus® kit. And there were 15 haplotypes which shared among two individuals. On the other hand a total of 247 unique haplotypes were obtained from the same individuals using the 17 Y-STR Yfiler® kit. And there were 36 haplotypes which shared among at least two individuals (Table 12A, B). Compared to the 17 Y-STR Yfiler® kit, the 27 Y-STR Yfiler Plus® kit discrimination capacity and haplotype diversity is greater, as expected, because of the number of analyzed loci, as shown in (Table 13). DC of the 17 Y-STR Yfiler® and 27 Y-STR Yfiler Plus® was 87.28 and 95.43 respectively while HD was 99.65 and 99.973 respectively. The increase in the power of discrimination and haplotype diversity with the addition of the 10 extra Y-STRs of the Yfiler Plus® was pronounced in the UAE Arab male population. DC and HD results of UAE Arab male population were compared with the results of other populations [27, 36, 37, 53–56] (Table 18). Results shows that HD and DC were

increased in all populations. Greenlanders showed the lowest HD and DC (99.71, 79) respectively

Table 17: Predominant alleles compared with Different populations. Allele shared between UAE Arab male populations are highlighted

Population	UAE N=343	Guinea Bissau N=215	Iraq N=400	Turkey N=76	Syria N=113	Somalia N=201	Spain N=347	Tunisia N=218	Brazil N=412	Finland N=400	Colombia N=173	Italy N=155	India N=106
DYS456	15	_	15	15	_	_	14	15	_	_	_	_	15
DYS389I	13	13	14	13	13	13	13	13	13	14	13	13	13
DYS390	23	21	23	23	23	25	24	25	24	24	24	24	22
DYS389II	30	30	29	20	30	31	16	31	29	30	29	29	30
DYS458	16	_	15	16	_	_	16	17	_	_	_	_	16
DYS19	14	14	14	14	14	11	14	13	14	14	14	14	15
DYS385-A	13	_	13	10	16	27	15	15	_	11	_	12	11
DYS385-B	18	_	14	11	18	28	18	18	_	13	_	_	14
DYS393	13	13	14	12	12	13	13	13	13	14	13	13	13
DYS391	10	10	10	10	10	10	11	9	10	11	11	10	10
DYS439	11	13	10	11	12	11	12	10	_	10	14	12	11
DYS635	21	_	24	21	_	_	22	21	_	_	24	_	_
DYS392	11	11	11	11	11	13	14	11	13	14	13	11	11
GATAH4	11	_	12	11	_	_	12	11	_	11	29	_	12
DYS437	14	14	14	13	14	14	15	14	_	14	15	15	14
DYS438	10	11	10	10	9	11	12	10	_	10	13	_	11
DYS448	20	_	19	20	_	_	19	20	_		_	_	19
number of shared loci	17	7	8	12	8	4	5	10	4	5	3	5	9

Table 18: Comparison of HD and DC between UAE population and other populations

populations	Yfiler®		Yfiler Plus®	
	HD	DC	HD	DC
UAE	99.65	87.28	99.973	95.43
Danes	99.99	98	100	100
Greenlanders	98.72	57	99.71	79
Somalia	98.03	61	99.93	99
Serbia	-	98.92	99.985	99.985
Nigeria	99.92	-	99.98	-
Peru	-	-	100	100
Han (China)	99.987	96.5	99.999	99.6
autochthonous (Spain)	-	95.55	99.99996	99.9996
Austrian	99.99	97.88	99.9989	99.76
Italian	99.97	97.04	99.99	98.52

In relationship testing, the inclusion of seven RM Y-STRs in Yfiler Plus® may cause some concern, because these loci have high mutation rates and, thus, genetic inconsistencies will rather frequently be observed between paternally closely related individuals [53]. Discrimination capacity can be increased between unrelated individuals and males of the same patrilineage by using RM Y-STRs even between first degree relatives [36].

Analysis of Molecular Variance (AMOVA) was calculated using online software YHRD. Using the YHRD database, the resulted haplotypes of UAE Arab male population were compared to European populations haplotypes generated using the Yfiler plus® kit since there is no database for Arab male population available for this kit (Table 14). Results shows that Denmark, Italy, Spain and United States have a short genetic distance between each other, while UAE Arab male population showed higher value of genetic distance with these populations. For parallel comparison purposes with nearby populations the new loci that added to the 27 Y-STR Yfiler plus®

were eliminated and the analysis was done using only 17 Y-STR Yfiler® loci (Table 15). AMOVA showed that the genetic distance between UAE Arab male population and Iran, Iraq, Egypt, Yemen and Kuwait populations were closest to each other respectively. Whereas it appeared to be far from Sudan, Jordan and Tunisia populations.

The MDS plot structured from F_{st} distance matrix for the 27 Y-STR Yfiler plus® (Figure 14) showed that UAE Arab male population is far from other European populations. Corresponding to the AMOVA results Denmark, Italy, Spain and United States are clustered together while Hungary was observed far from them. This reflects that the UAE Arab male population has statistically significant different genetic structure from European population. In contrast, The MDS plot for the 17 Y-STR Yfiler® (Figure 15) showed that UAE Arab male population clustered along with Iran, Iraq, Egypt, Yemen and Kuwait populations which mean that they have a closer genetic relation. These results also correlated with the geographical distribution of these populations.

Abnormal alleles were noticed in UAE Arab male population. Null alleles were observed in two loci DYS438 and DYS390 (Figure 16). Comparing to other studies null alleles were observed in two Iraqi individuals and one Italian male at DYS448 locus [57, 58]. In other study they noticed null alleles in DYS448, DYS392, DYS549 and DYS385a/b [59]. That means the null alleles in UAE Arab male population is unique in DYS438 and DYS390 loci. These null alleles could be because of two reason. Firstly primer binding site mutation which lead to un-amplified the locus. Secondly it could be a true null alleles due to deletion mutation of the whole locus [57] which happens due to the rearrangement on the long arm of the human Y-chromosome

“azoospermia factor regions” (AZFa, AZFb, AZFc) [58]. Other type of abnormal alleles were found in UAE Arab male population; there were multiple alleles in three different loci DYF387S1a/b, DYS19 and DYS570 (Figure 16). Four alleles were observed in one individual in DYF387S1a/b locus and three alleles were observed in two individuals in the same locus where double alleles were observed in DYS19 and DYS570. Three alleles in DYF387S1a/b and DYS385a/b were also previously observed in Serbia population [39] and double alleles were also observed in Iraqi population in two loci DYS439 and DYS635 [57]. Moreover, double alleles were observed in Bahia (Brazil) in three loci DYS389 II, DYS437 and DYS439 [60]. There are other previous studies that showed double alleles in DYS19 [61, 62]. That means the four alleles that were observed in DYF387S1a/b locus and double alleles in DYS570 were unique mutation in UAE Arab male population. The reason behind that could be because if insertion of chromosomal region including the STR locus during the transmission of Y chromosome among paternal lineage [63]. On the other hand, intermediate alleles were found in 9 loci (DYS458, DYF387S1(a/b), DYS627, DYS576, YGATAH4, DYS448, DYS518, DYS385(a/b) and DYS481) (Table 16). Intermediate alleles for other populations are shows in (Table 19) [39, 64]. The reason for that could be one, two or three base deletions within or far from the repeat region [19].

Table 19: Intermediate allele in different loci for different populations

Loci	Intermediate alleles	Populations
DYS392	10.2, 11.2	Adygeis, Russia
DYS458	17.2	Han, Taiwan, China
DYS481	25.1, 26.1	India
DYS385	12.2	
DYS549	11.3	
DYS390	24.3	Serbia
DYS458	17.2, 19.2, 20.2	

Abnormal loci (DYF387S1a/b and DYS570) which have multiple alleles was repeated using RM-Yplex kit [31]. RM-Yplex results confirmed the results that have been obtained from Yfiler Plus® kit (Figure 18).

Chapter 5: Conclusion

It can be concluded that the new 27 Y-STR Yfiler Plus® shows more power of forensic discrimination, more haplotype diversity and the additional loci shows more genetic diversity. The dramatic increase in the power of discrimination between male individuals from previous version, Yfiler, and new version, Yfiler Plus, will clearly results in no match when inquiry is done using one of the largest Y-STR haplotypes databases such as YHRD. Therefore, Y-STR haplotypes should be continuously produced and studied in order to enrich the database for realistic estimation of the haplotype frequency for forensic purposes.

The genetic structure of UAE Arab male population showed close distance with the surrounding populations especially to Arab populations. Observing a null locus should not represent a problem for profile interpretation. More than one peak at one or more loci could lead up to deep misinterpreting of a profile and therefore from which kind of sample the profile come out. This study is very important and it contributes to other studies in the Gulf region to study the genetic diversity of populations. Further studies are needed such as sequencing the abnormal alleles to know what exactly the reason behind that. However, this study calls for more studies such as disease-associated haplotypes, archeological, and historical studies.

According to measures of genetic diversity the highest diversity should be considered the most diverse and polymorphic for forensic testing which are useful to differentiate between male relatives. While loci with the lowest diversity are the least polymorphic loci and are more useful for paternity lineage testing.

Among the nine markers showed the highest gene diversity, seven are new in the Yfiler Plus® kit compared to the Yfiler® kit which confirm our hypothesis.

References

[1] N. Rudin and K. Inman, An introduction to forensic DNA analysis, *CRC press*, vol. 3, 2001.

[2] D. A. Parry and J. Squire, Fibrous Proteins: Structures and Mechanisms. *Springer*, 2017.

[3] F. A. Khan, Biotechnology in medical sciences. *CRC Press*, 2014.

[4] L. Pray, Discovery of DNA structure and function: Watson and Crick. *Nature Education*, vol. 1, no. 1, p. 100, 2008.

[5] N. R. Council, DNA technology in forensic science. *National Academies Press*, 1992.

[6] D. Botstein, R. L. White, M. Skolnick, and R. W. Davis, Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American journal of human genetics*, vol. 32, no. 3, p. 314, 1980.

[7] J. M. Bartlett and D. Stirling, A short history of the polymerase chain reaction. *PCR Protocols*, pp. 3–6, 2003.

[8] I. H. G. S. Consortium, Finishing the euchromatic sequence of the human genome. *Nature*, vol. 431, no. 7011, pp. 931–945, 2004.

[9] G. Palsson, Anthropology, Genomics, and Human Variation: National Roots, in International Encyclopedia of the Social & Behavioral Sciences (Second Edition). *Elsevier*, pp. 758–764, 2015.

[10] J. M. Butler, Fundamentals of forensic DNA typing. *Academic Press*, 2009.

[11] J. M. Butler, Advance topics in forensic DNA typing: interpretation. *Elsevier Academic Press*, 2015.

[12] L. Buckingham, Molecular diagnostics: fundamentals, methods and clinical applications. *FA Davis*, 2011.

[13] M. D. Griswold and P. A. Hunt, “Meiosis,” in Brenner’s Encyclopedia of Genetics (Second Edition). *Academic Press*, pp. 338–341, 2013.

[14] “Multiple codes in DNA,” [Online]. Available: https://pandasthumb.org/archives/2007/10/konstantin-b-ze.html. [Accessed: 08-Oct-2007].

[15] J. M. Butler, “Genetics and genomics of core short tandem repeat loci used in human identity testing,” *J. Forensic Sci.*, vol. 51, no. 2, pp. 253–265, 2006.

[16] "DNA structure." [Online]. Available: http://academic.brooklyn.cuny.edu /biology/bio4fv/page/molecular%20biology/dna-structure.html. [Accessed: 14-Oct-2017].

[17] M. E. Turner, C. Jenkins, A. Milsted, and D. L. Ely, Sex chromosomes. *Advances in Molecular and Cell Biology*, vol. 34, pp. 1–13, 2004.

[18] W. Goodwin, A. Linacre, and S. Hadi, An introduction to forensic genetics. *John Wiley & Sons*, vol. 2, 2011.

[19] J. M. Butler, Advanced topics in forensic DNA typing: methodology. *Academic Press*, 2011.

[20] J. S. Buckleton, J.-A. Bright, and D. Taylor, Forensic DNA evidence interpretation. *CRC press*, 2016.

[21] H. Skaletsky, T. Kuroda-Kawaguchi, P. J. Minx, H. S. Cordum, L. Hillier, L.G. Brown, S. Repping, T. Pyntikova, J. Ali, Ti. Bieri and A. Chinwalla, The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*, vol. 423, no. 6942, p. 825, 2003.

[22] P. J. Wijchers and R. J. Festenstein, Epigenetic regulation of autosomal gene expression by sex chromosomes. *Trends Genet.*, vol. 27, no. 4, pp. 132–140, 2011.

[23] P. Thanakiatkrai, T. Kitpipit, R. Gupta, S. Sharma, I. Babahan, E.P. Coban, H. Biyik, K. Alizadeh, J.A.T. da Silva, A. Babnik and R. Petkovsek, Current STR-based techniques in forensic science. *Maejo Int J Sci Technol*, vol. 7, pp. 1–15, 2013.

[24] J. M. Butler and C. R. HILL, Biology and genetics of new autosomal STR loci useful for forensic DNA analysis. *CRC press*, 2012.

[25] L. Zha, Y. Liu, Y. Guo, J. Li, K. Wang, K. Geng, Q. Liao, J. Liu, H. Chen and J. Cai, Genetic polymorphism of 21 non-CODIS STR loci in the Chinese Mongolian ethnic minority. *Forensic Sci. Int. Genet.*, vol. 9, pp. e32–e33, 2014.

[26] H. Asamura, M. Ota, and H. Fukushima, Population data on 10 non-CODIS STR loci in Japanese population using a newly developed multiplex PCR system. *J. Forensic Leg. Med.*, vol. 15, no. 8, pp. 519–523, 2008.

[27] I. Pickrahn, E. Müller, W. Zahrer, B. Dunkelmann, J. Cemper-Kiesslich, G. Kreindl and F. Neuhuber, Yfiler® Plus amplification kit validation and calculation of forensic parameters for two Austrian populations. *Forensic Sci. Int. Genet.*, vol. 21, pp. 90–94, 2016.

[28] D. C. Linhares, A. Rovira, and M. Torremorell, Evaluation of Flinders Technology Associates cards for collection and transport of samples for detection of Porcine reproductive and respiratory syndrome virus by reverse transcription polymerase chain reaction. *J. Vet. Diagn. Invest.*, vol. 24, no. 2, pp. 328–332, 2012.

[29] R. Kuhn and S. Hoffstetter-Kuhn, Capillary electrophoresis: principles and practice. *Springer Science & Business Media*, 2013.

[30] J. M. Butler, Forensic DNA typing: biology, technology, and genetics of STR markers. *Academic Press*, 2005.

[31] L. Excoffier, G. Laval, and S. Schneider, Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary bioinformatics online*, vol. 1, p. 47, 2005.

[32] L. Gusmão, J. M. Butler, A. Carracedo, P. Gill, M. Kayser, W. R.Mayr, N. Morling, M. Prinz, L. Roewer, C. Tyler-Smith and P. M. Schneider, DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis. *Int. J. Legal Med.*, vol. 120, no. 4, pp. 191–200, 2006.

[33] R. Alghafri, W. Goodwin, A. Ralf, M. Kayser, and S. Hadi, A novel multiplex assay for simultaneously analysing 13 rapidly mutating Y-STRs. Forensic Sci. *Int. Genet.*, vol. 17, pp. 91–98, 2015.

[34] C. Alves, L. Gusmao, J. Meirinhos, and A. Amorim, Making the most of Y-STR haplotypes: The HapYDive. *International Congress Series*, vol. 1288, pp. 201–203, 2006.

[35] M. Nei, Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, vol. 89, no. 3, pp. 583–590, 1978.

[36] C. Rapone, E. D'Atanasio, A. Agostino, M. Mariano, M. T. Papaluca, F. Cruciani and A. Berti, Forensic genetic value of a 27 Y-STR loci multiplex (Yfiler® Plus kit) in an Italian population sample. *Forensic Sci. Int. Genet.*, vol. 21, pp. e1–e5, 2016.

[37] B. Martinez, L. Catelli, M. Romero, M. V. O. Okolie, S. O. Keshinro, E. F. Carvalho, C. Vullo and L. Gusmão, Forensic evaluation of 27 y-str haplotypes in a population sample from nigeria. *Forensic Sci. Int. Genet. Suppl. Ser.*, 2017.

[38] M. Tokdemir and F. T. Tunçez, Genetic polymorphisms of 17 Y-STR loci in Eastern Turkey population. *Gene Rep.*, vol. 6, pp. 15–18, 2017.

[39] D. Zgonjanin, R. Alghafri, M. Antov, G. Stojiljković, S. Petković, R. Vuković and D. Drašković, Genetic characterization of 27 Y-STR loci with the Yfiler (®) Plus kit in the population of Serbia. *Forensic Sci. Int. Genet.*, 2017.

[40] H. Imad, Q. Cheah, J. Mohammad, and O. Aamera, Genetic variation of 17 Y-chromosomal short tandem repeats (STRs) loci from unrelated individuals in Iraq. *Int. J. Biotechnol. Mol. Biol. Res.*, vol. 4, no. 8, pp. 119–129, 2013.

[41] M. A. Kareem, M. A. Jebor, and I. H. Hameed, Allele frequency present within the DYS635, DYS437, DYS448, DYS456, DYS458, YGATA H4, DYS389I, DYS389II, DYS19, DYS391, DYS438, DYS390, DYS439, DYS392, DYS393, DYS385a and DYS385b of unrelated individuals in Iraq. *Afr. J. Biotechnol.*, vol. 14, no. 10, pp. 851–858, 2015.

[42] L. Abdin, K. Dewa, S. Rand, C. Hohoff, and B. Brinkmann, Analysis of 13 Y-chromosomal STRs in an Arab population sample from Syria. *International Congress Series*, vol. 1239, pp. 319–321, 2003.

[43] S. Frigi, F. Pereira, L. Pereira, B. Yacoubi, L. Gusmão, C. Alves, H. K. el Khil, L. Cherni, A. Amorim and A. El Gaaied, Data for Y-chromosome haplotypes defined by 17 STRs (AmpFLSTR® YfilerTM) in two Tunisian Berber communities. *Forensic Sci. Int.*, vol. 160, no. 1, pp. 80–83, 2006.

[44] B. Ambrosio, A. Novelletto, C. Hernandez, J. M. Dugoujon, C. Fortes-Lima, J. N. Rodriguez and R. Calderon, Y-STR genetic diversity in autochthonous Andalusians from Huelva and Granada provinces (Spain). *Forensic Sci. Int. Genet.*, vol. 6, no. 2, pp. e66–e71, 2012.

[45] Y. M. Chang, R. Perumal, P. Y. Keat, R. Y. Yong, D. L. Kuehn, and L. Burgoyne, A distinct Y-STR haplotype for Amelogenin negative males characterized by a large Y p 11.2 (DYS458-MSY1-AMEL-Y) deletion. *Forensic Sci. Int.*, vol. 166, no. 2, pp. 115–120, 2007.

[46] L. Dönbak, T. Bajanowski, B. Brinkmann, and C. Hohoff, Y-STR haplotypes in populations from the Eastern Mediterranean region of Turkey. *Int. J. Legal Med.*, vol. 120, no. 6, pp. 395–396, 2006.

[47] K. Fadhlaoui-Zid, S. Chennakrishnaiah, R. Zemni, S. Grinberg, R. J. Herrera, and A. Benammar-Elgaaied, Sousse, Tunisia: Tumultuous history and high Y-STR diversity. *Electrophoresis*, vol. 33, no. 23, pp. 3555–3563, 2012.

[48] C. Hallenberg, B. Simonsen, J. Sanchez, and N. Morling, Y-chromosome STR haplotypes in Somalis. *Forensic Sci. Int.*, vol. 151, no. 2, pp. 317–321, 2005.

[49] M. Hedman, V. Pimenoff, M. Lukka, P. Sistonen, and A. Sajantila, Analysis of 16 Y STR loci in the Finnish population reveals a local reduction in the diversity of male lineages. *Forensic Sci. Int.*, vol. 142, no. 1, pp. 37–43, 2004.

[50] Z. Rahman and T. Husnain, Y-Chromosomal STR Haplotype Profiling in Yousafzai's living in Swat Valley Pakistan. *National Centre of Excellence in Molecular Biology*, 2012.

[51] J. A. Soares-Vieira, A. E. Billerbeck, E. S. Iwamura, B. B. Mendonca, L. Gusmão, and P. A. Otto, Population and mutation analysis of Y-STR loci in a sample from the city of São Paulo (Brazil). *Genet. Mol. Biol.*, vol. 31, no. 3, pp. 651–656, 2008.

[52] A. Rosa, C. Ornelas, A. Brehm, and R. Villems, Population data on 11 Y-chromosome STRs from Guine-Bissau. *Forensic Sci. Int.*, vol. 157, no. 2, pp. 210–217, 2006.

[53] J. K. Olofsson, H. S. Mogensen, A. Buchard, C. Børsting, and N. Morling, Forensic and population genetic analyses of Danes, Greenlanders and Somalis typed with the Yfiler® Plus PCR amplification kit. *Forensic Sci. Int. Genet.*, vol. 16, pp. 232–236, 2015.

[54] D. H. Tineo, S. Loiola, F.V. Paredes, L. R. Noli, Y. C. Amaya, F. Simão, E.F. Carvalho and L. Gusmão, Genetic characterization of 27 Y-STR loci in the native population of Ashaninka from Peru. *Forensic Sci. Int. Genet. Suppl. Ser.*, vol. 5, pp. e220–e222, 2015.

[55] Y. Wang, Y. J. Zhang, C. C. Zhang, R. Li, Y. Yang, X. L. Ou, D. Y. Tong and H. Y. Sun, Genetic polymorphisms and mutation rates of 27 Y-chromosomal STRs in a Han population from Guangdong Province, Southern China. *Forensic Sci. Int. Genet.*, vol. 21, pp. 5–9, 2016.

[56] O. García, I. Yurrebaso, I. D. Mancisidor, S. López, S. Alonso, and L. Gusmão, Data for 27 Y-chromosome STR loci in the Basque Country autochthonous population. *Forensic Sci. Int. Genet.*, vol. 20, pp. e10–e12, 2016.

[57] I. Carboni and U. Ricci, Unexpected patterns in Y-STR analyses and implications for profile identification. *Forensic Sci. Int. Genet. Suppl. Ser.*, vol. 2, no. 1, pp. 55–56, 2009.

[58] S. Turrina, S. Caratti, M. Ferrian, and D. De Leo, Deletion and duplication at DYS448 and DYS626 loci: unexpected patterns within the AZFc region of the Y-chromosome. *Int. J. Legal Med.*, vol. 129, no. 3, pp. 449–455, 2015.

[59] J. Beltramo, M. A. Pena, and M. M. Lojo, The finding of Y-STR microdeletion involving DYS448, DYS392, DYS549 and DYS385a/b markers in a paternity case with deceased alleged father. Forensic Sci. *Int. Genet. Suppl. Ser.*, vol. 5, pp. e141–e143, 2015.

[60] M. Diederiche, P. Martin, A. Amorim, F. Corte-Real, and L. Gusmao, A case of double alleles at three Y-STR loci: forensic implications. *Int. J. Legal Med.*, vol. 119, no. 4, pp. 223–225, 2005.

[61] C. L. Johnson, J. H. Warren, R. C. Giles, and R. W. Staub, Validation and uses of a Y-chromosome STR 10-plex for forensic and paternity laboratories. *J. Forensic Sci.*, vol. 48, no. 6, pp. 1260–1268, 2003.

[62] M. Kayser and A. Sajantila, Mutations at Y-STR loci: implications for paternity testing and forensic analysis. *Forensic Sci. Int.*, vol. 118, no. 2, pp. 116–121, 2001.

[63] H. Skaletsky, T. Kuroda-Kawaguchi, P. J. Minx, H. S. Cordum, L. Hillier, L. G. Brown, S. Repping, T. Pyntikova, J. Ali, T. Bieri and A. Chinwalla, The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*, vol. 423, no. 6942, pp. 825–838, 2003.

[64] T. Gao, L. Yun, Y. Gu, W. He, Z. Wang, and Y. Hou, Phylogenetic analysis and forensic characteristics of 12 populations using 23 Y-STR loci. *Forensic Sci. Int. Genet.*, vol. 19, pp. 130–133, 2015.

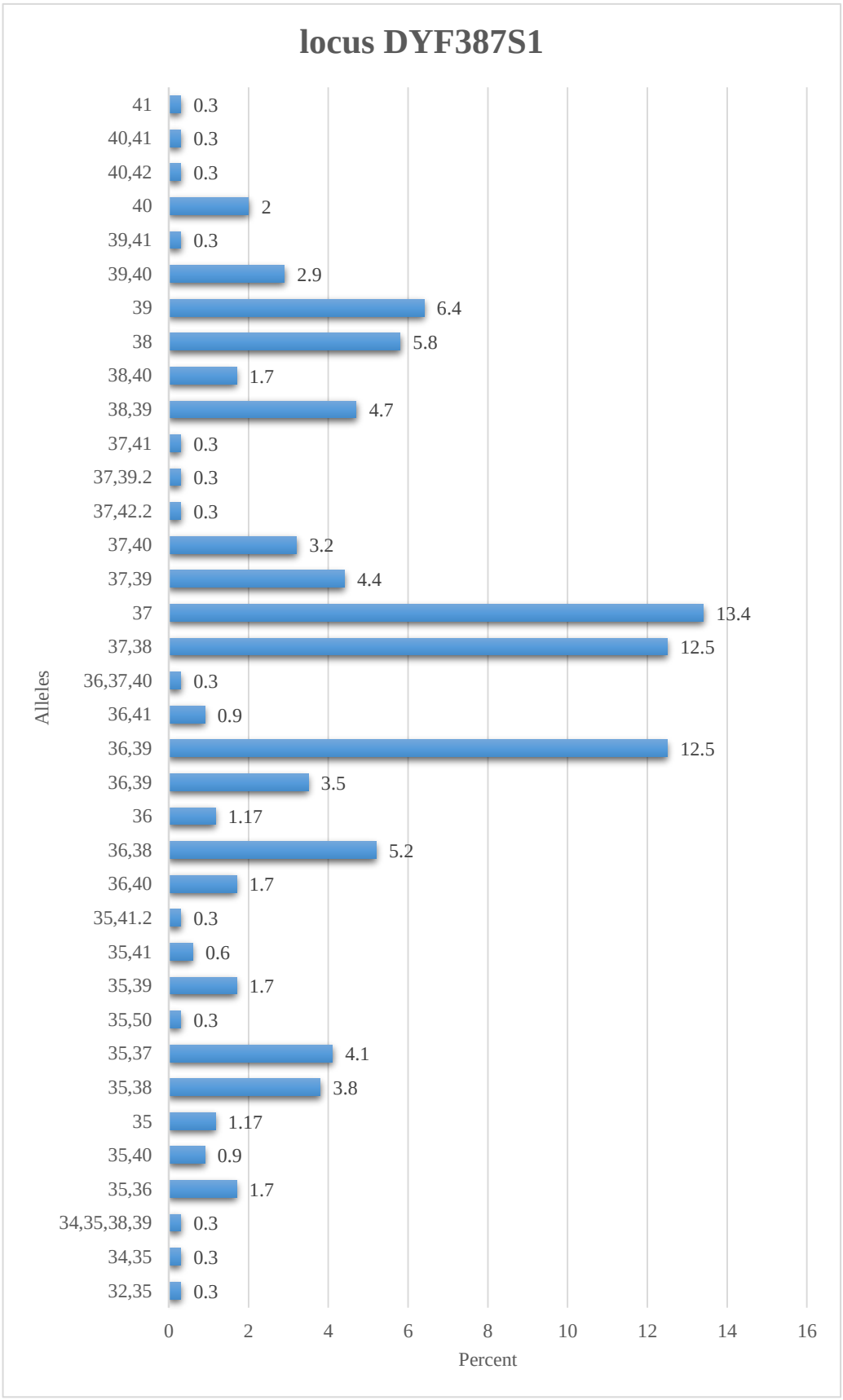
Appendix

Y-STR loci frequencies

Table 20 (A-X): Alleles frequency of different loci with bar charts for UAE Arab male population

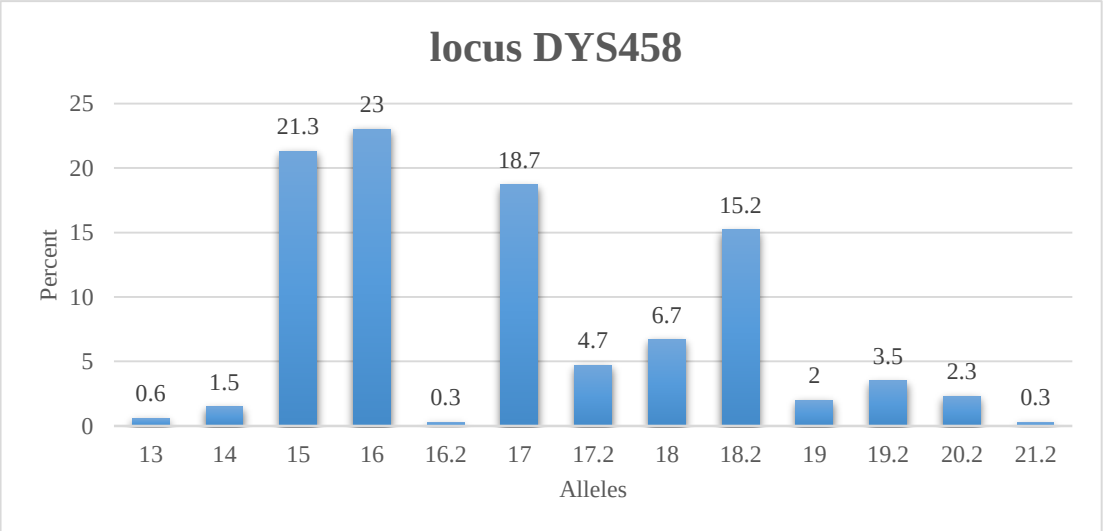
A. Locus DYF387S1(a/b)

Alleles for the locus DYF387S1(a/b)								
Haplotype No.	Haplotype	Freq.	Percentage		Haplotype No.	Haplotype	Freq.	Percentage
1	32,35	1	0.3		19	36,37,40	1	0.3
2	34,35	1	0.3		20	37,38	43	12.5
3	34,35,38,39	1	0.3		21	37	46	13.4
4	35,36	6	1.7		22	37,39	15	4.4
5	35,40	3	0.9		23	37,40	11	3.2
6	35	4	1.17		24	37,42.2	1	0.3
7	35,38	13	3.8		25	37,39.2	1	0.3
8	35,37	14	4.1		26	37,41	1	0.3
9	35,50	1	0.3		27	38,39	16	4.7
10	35,39	6	1.7		28	38,40	6	1.7
11	35,41	2	0.6		29	38	20	5.8
12	35,41.2	1	0.3		30	39	22	6.4
13	36,40	6	1.7		31	39,40	10	2.9
14	36,38	18	5.2		32	39,41	1	0.3
15	36	4	1.17		33	40	7	2
16	36,39	12	3.5		34	40,42	1	0.3
17	36,39	43	12.5		35	40,41	1	0.3
18	36,41	3	0.9		36	41	1	0.3
Total							343	100


locus DYF387S1
Alleles
41 0.3
40,41 0.3
40,42 0.3
40 2
39,41 0.3
39,40 2.9
39 6.4
38 5.8
38,40 1.7
38,39 4.7
37,41 0.3
37,39.2 0.3
37,42.2 0.3
37,40 3.2
37,39 4.4
37 13.4
37,38 12.5
36,37,40 0.3
36,41 0.9
36,39 12.5
36,39 3.5
36 1.17
36,38 5.2
36,40 1.7
35,41.2 0.3
35,41 0.6
35,39 1.7
35,50 0.3
35,37 4.1
35,38 3.8
35 1.17
35,40 0.9
35,36 1.7
34,35,38,39 0.3
34,35 0.3
32,35 0.3
0 2 4 6 8 10 12 14 16
Percent

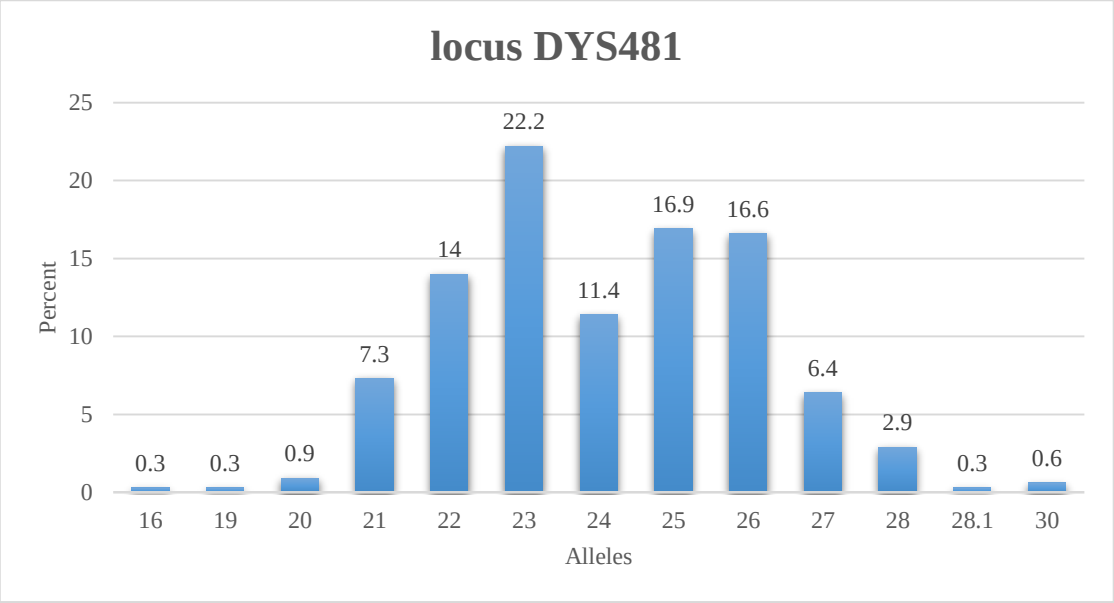
B. Locus DYS458

Alleles for the locus DYS458			
Alleles No.	Allele repeats	Freq.	Percentage
1	13	2	0.6
2	14	5	1.5
3	15	73	21.3
4	16	79	23
5	16.2	1	0.3
6	17	64	18.7
7	17.2	16	4.7
8	18	23	6.7
9	18.2	52	15.2
10	19	7	2
11	19.2	12	3.5
12	20.2	8	2.3
13	21.2	1	0.3
Total		343	100



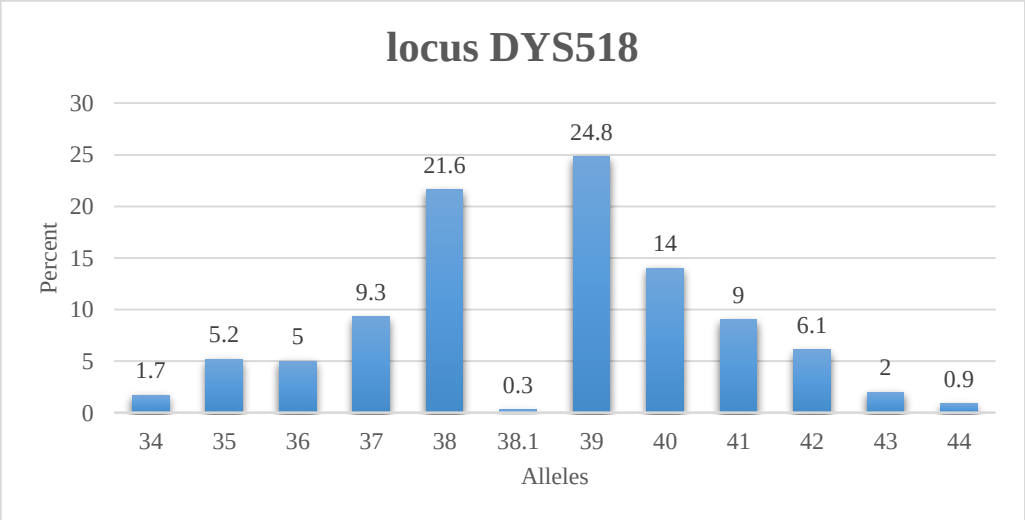
C. Locus DYS481

Alleles for the locus DYS481			
Alleles No.	Allele repeats	Freq.	Percentage
1	16	1	0.3
2	19	1	0.3
3	20	3	0.9
4	21	25	7.3
5	22	48	14
6	23	76	22.2
7	24	39	11.4
8	25	58	16.9
9	26	57	16.6
10	27	22	6.4
11	28	10	2.9
12	28.1	1	0.3
13	30	2	0.6
Total		343	100



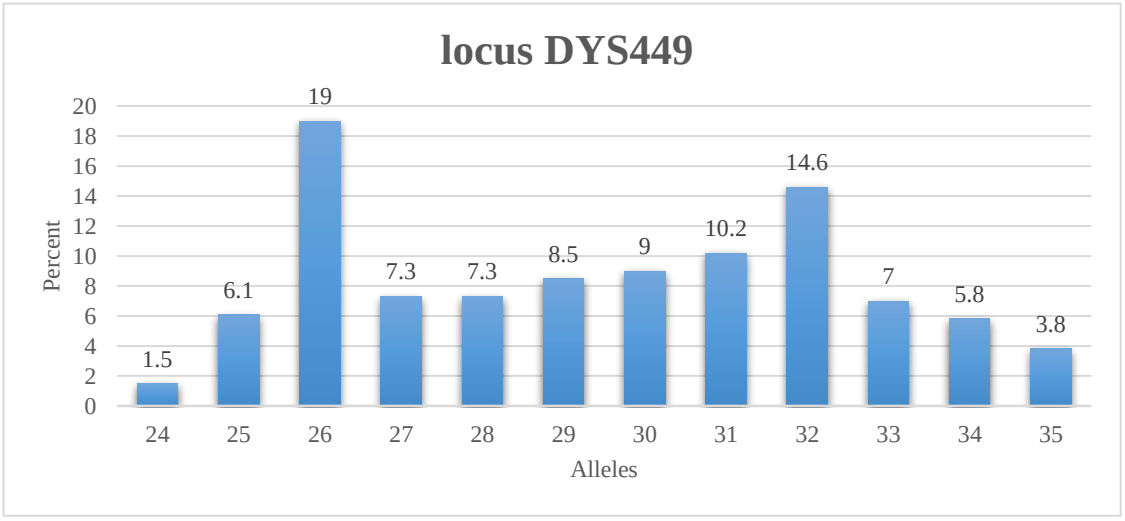
D. Locus DYS518

Alleles for the locus DYS518			
Alleles No.	Allele repeats	Freq.	Percentage
1	34	6	1.7
2	35	18	5.2
3	36	17	5
4	37	32	9.3
5	38	74	21.6
6	38.1	1	0.3
7	39	85	24.8
8	40	48	14
9	41	31	9
10	42	21	6.1
11	43	7	2
12	44	3	0.9
Total		343	100



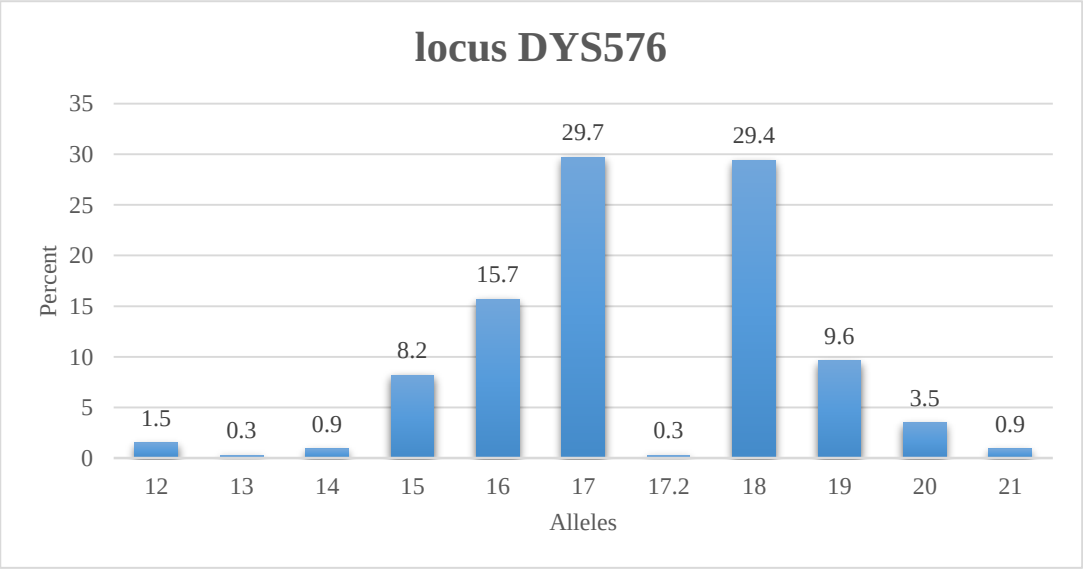
E. Locus DYS449

Alleles for the locus DYS449			
Alleles No.	Allele repeats	Freq.	Percentage
1	24	5	1.5
2	25	21	6.1
3	26	65	19
4	27	25	7.3
5	28	25	7.3
6	29	29	8.5
7	30	31	9
8	31	35	10.2
9	32	50	14.6
10	33	24	7
11	34	20	5.8
12	35	13	3.8
Total		343	100



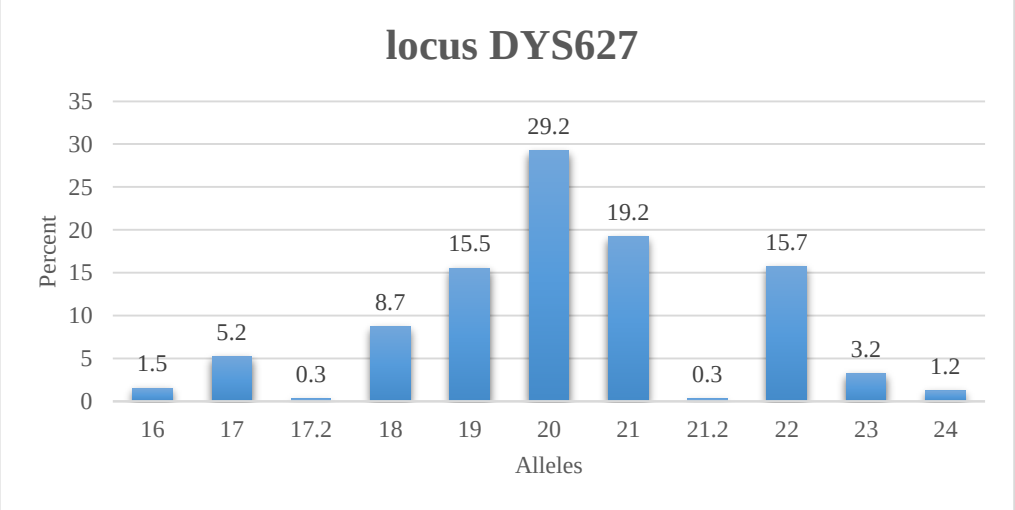
F. Locus DYS576

Alleles for the locus DYS576			
Alleles No.	Allele repeats	Freq.	Percentage
1	12	5	1.5
2	13	1	0.3
3	14	3	0.9
4	15	28	8.2
5	16	54	15.7
6	17	102	29.7
7	17.2	1	0.3
8	18	101	29.4
9	19	33	9.6
10	20	12	3.5
11	21	3	0.9
Total		343	100



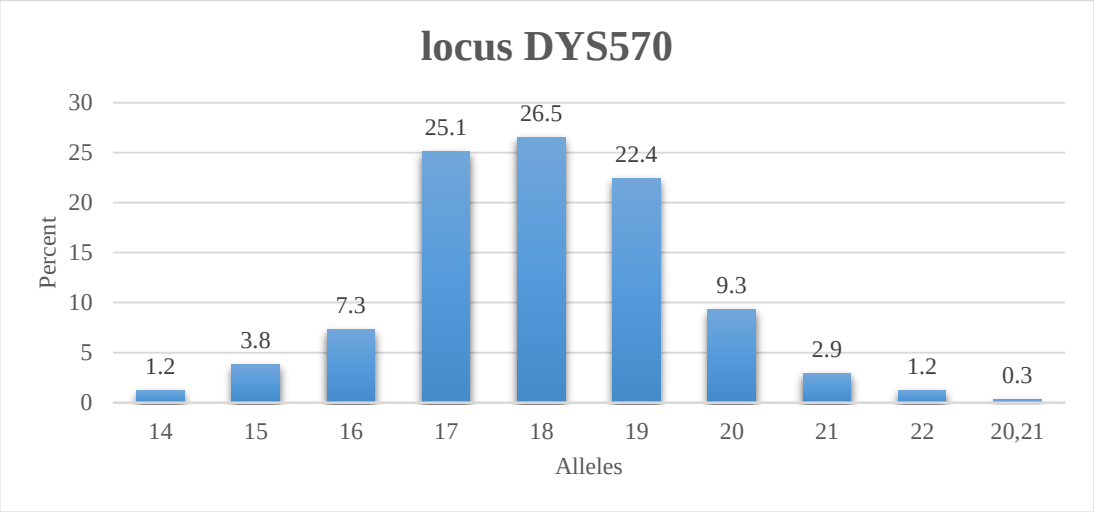
G. Locus DYS627

Alleles for the locus DYS627			
Alleles No.	Allele repeats	Freq.	Percentage
1	16	5	1.5
2	17	18	5.2
3	17.2	1	0.3
4	18	30	8.7
5	19	53	15.5
6	20	100	29.2
7	21	66	19.2
8	21.2	1	0.3
9	22	54	15.7
10	23	11	3.2
11	24	4	1.2
Total		343	100



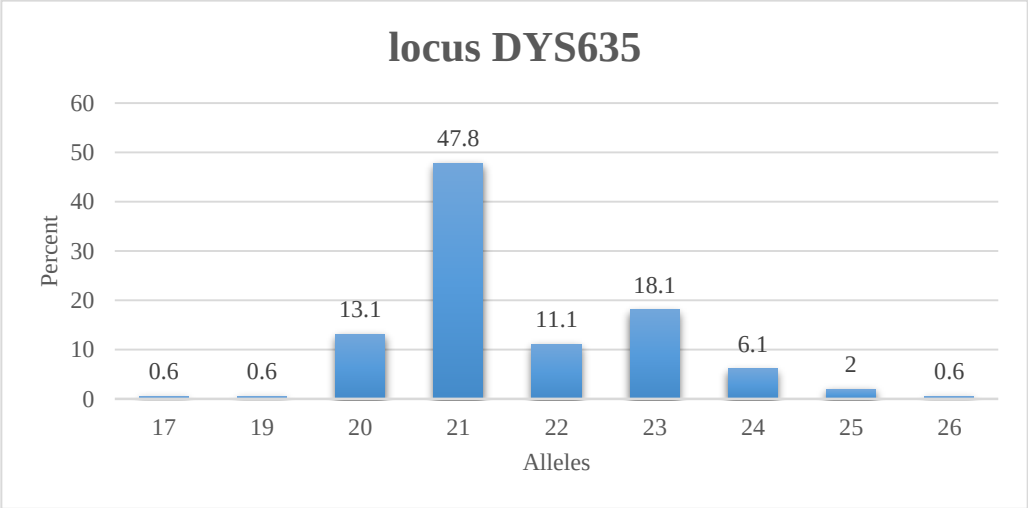
H. Locus DYS570

Alleles for the locus DYS570			
Alleles No.	Allele repeats	Freq.	Percentage
1	14	4	1.2
2	15	13	3.8
3	16	25	7.3
4	17	86	25.1
5	18	91	26.5
6	19	77	22.4
7	20	32	9.3
8	21	10	2.9
9	22	4	1.2
10	20,21	1	0.3
Total		343	100



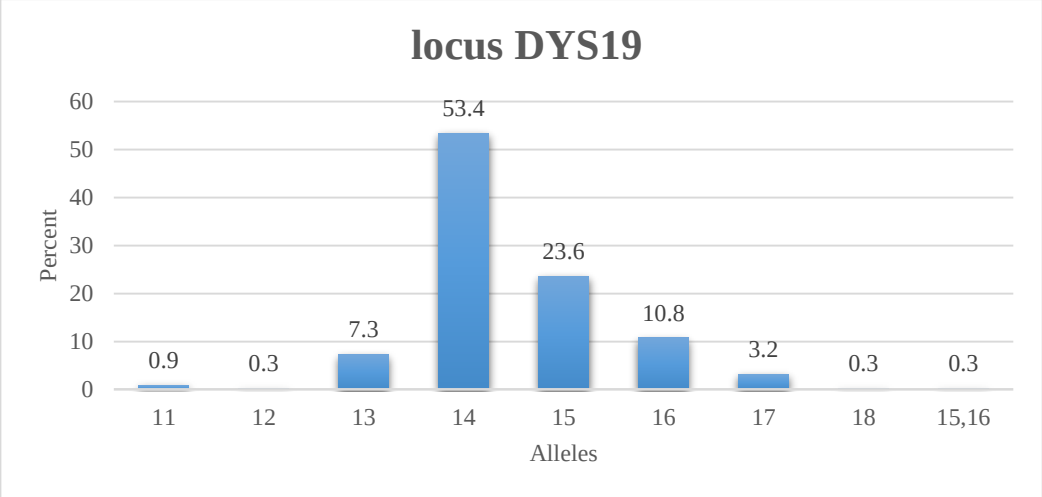
I. Locus DYS635

Alleles for the locus DYS635			
Alleles No.	Allele repeats	Freq.	Percentage
1	17	2	0.6
2	19	2	0.6
3	20	45	13.1
4	21	164	47.8
5	22	38	11.1
6	23	62	18.1
7	24	21	6.1
8	25	7	2
9	26	2	0.6
Total		343	100



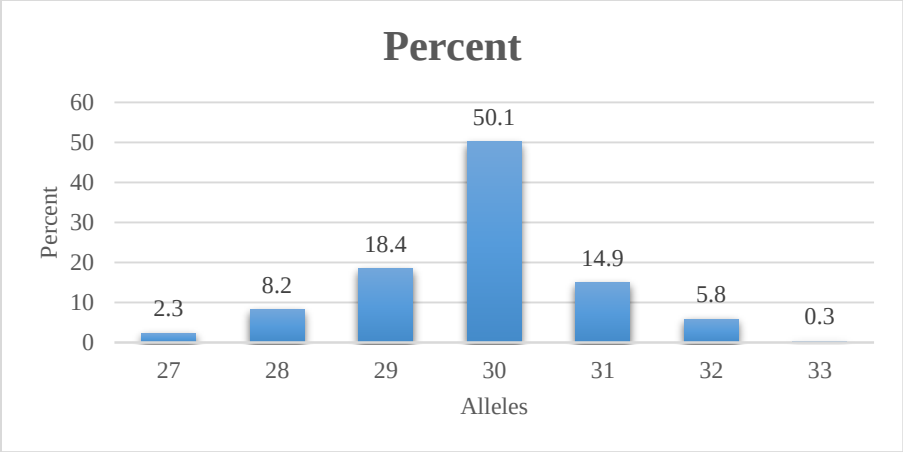
J. Locus DYS19

Alleles for the locus DYS19			
Alleles No.	Allele repeats	Freq.	Percentage
1	11	3	0.9
2	12	1	0.3
3	13	25	7.3
4	14	183	53.4
5	15	81	23.6
6	16	37	10.8
7	17	11	3.2
8	18	1	0.3
9	15,16	1	0.3
Total		343	100



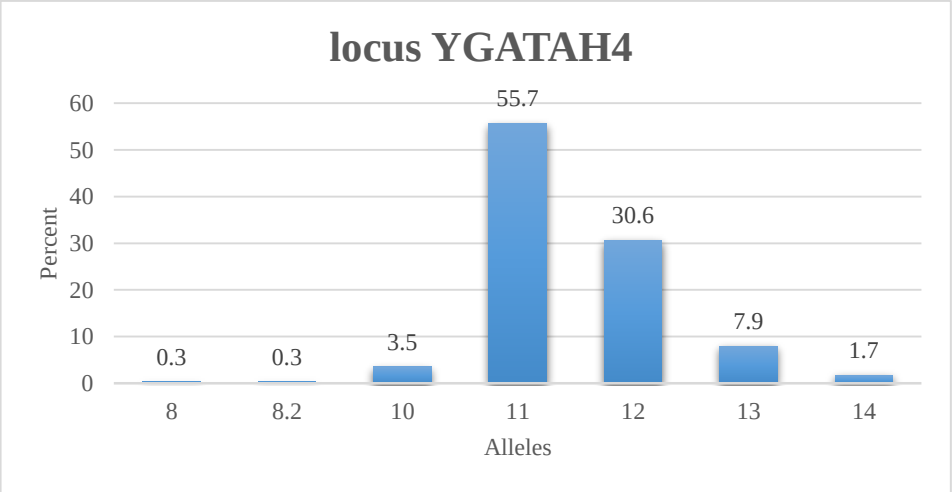
K. Locus DYS389II

Alleles for the locus DYS389II			
Alleles No.	Allele repeats	Freq.	Percentage
1	27	8	2.3
2	28	28	8.2
3	29	63	18.4
4	30	172	50.1
5	31	51	14.9
6	32	20	5.8
7	33	1	0.3
Total		343	100



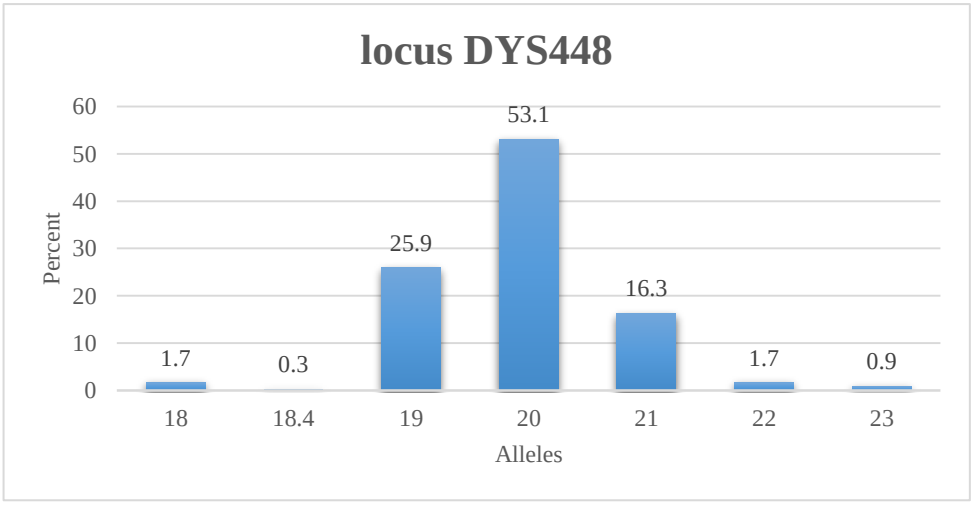
L. Locus YGATAH4

Alleles for the locus YGATAH4			
Alleles No.	Allele repeats	Freq.	Percentage
1	8	1	0.3
2	8.2	1	0.3
3	10	12	3.5
4	11	191	55.7
5	12	105	30.6
6	13	27	7.9
7	14	6	1.7
Total		343	100



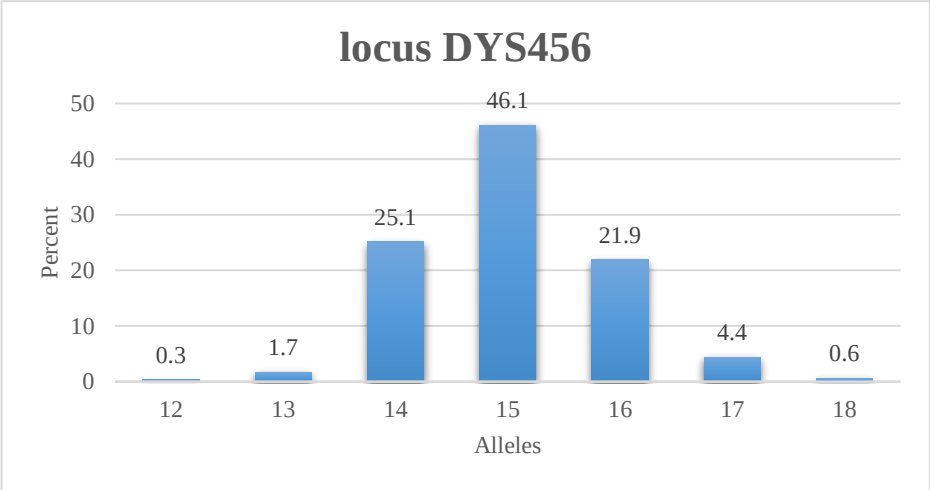
M. Locus DYS448

Alleles for the locus DYS448			
Alleles No.	Allele repeats	Freq.	Percentage
1	18	6	1.7
2	18.4	1	0.3
3	19	89	25.9
4	20	182	53.1
5	21	56	16.3
6	22	6	1.7
7	23	3	0.9
Total		343	100



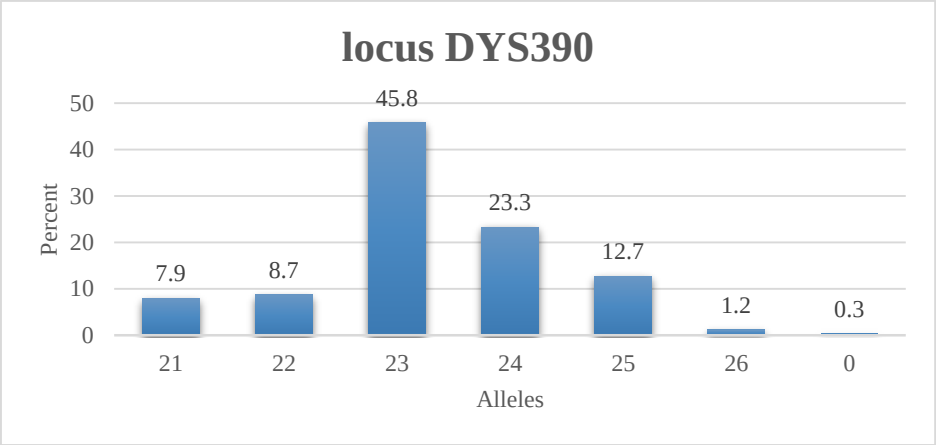
N. Locus DYS456

Alleles for the locus DYS456			
Alleles No.	Allele repeats	Freq.	Percentage
1	12	1	0.3
2	13	6	1.7
3	14	86	25.1
4	15	158	46.1
5	16	75	21.9
6	17	15	4.4
7	18	2	0.6
Total		343	100



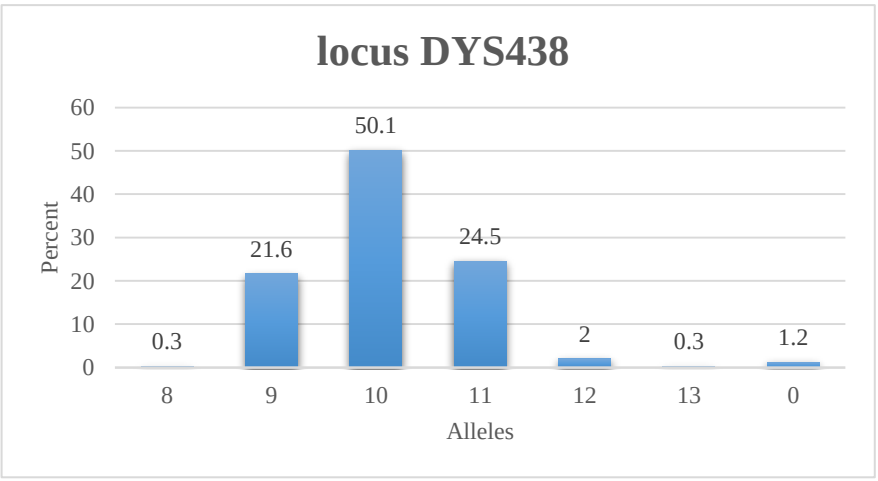
O. Locus DYS390

Alleles for the locus DYS390			
Alleles No.	Allele repeats	Freq.	Percentage
1	21	27	7.9
2	22	30	8.7
3	23	157	45.8
4	24	80	23.3
5	25	44	12.7
6	26	4	1.2
7	Null Allele	1	0.3
Total		343	100



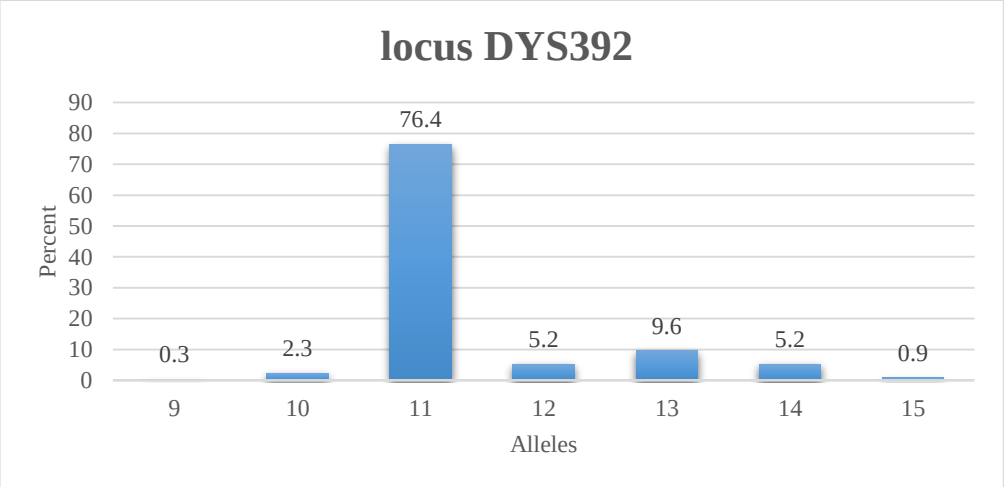
P. Locus DYS438

Alleles for the locus DYS438			
Alleles No.	Allele repeats	Freq.	Percentage
1	8	1	0.3
2	9	74	21.6
3	10	172	50.1
4	11	84	24.5
5	12	7	2
6	13	1	0.3
7	Null Allele	4	1.2
Total		343	100



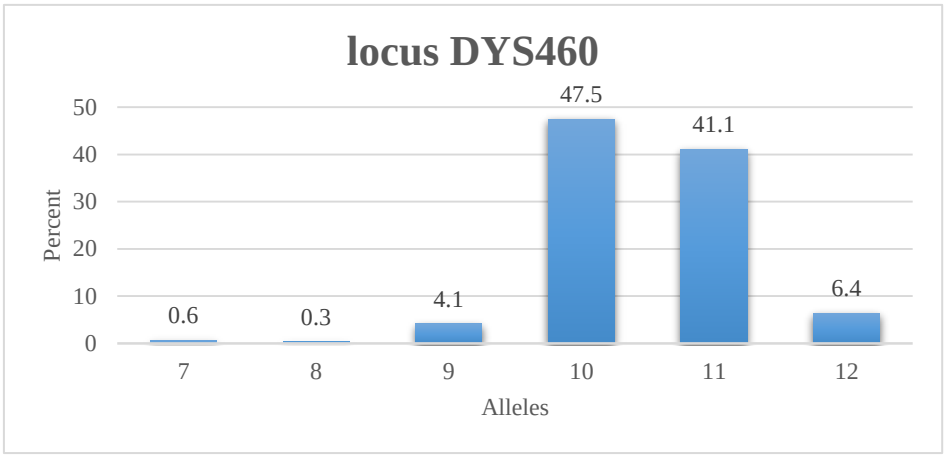
Q. Locus DYS392

Alleles for the locus DYS392			
Alleles No.	Allele repeats	Freq.	Percentage
1	9	1	0.3
2	10	8	2.3
3	11	262	76.4
4	12	18	5.2
5	13	33	9.6
6	14	18	5.2
7	15	3	0.9
Total		343	100



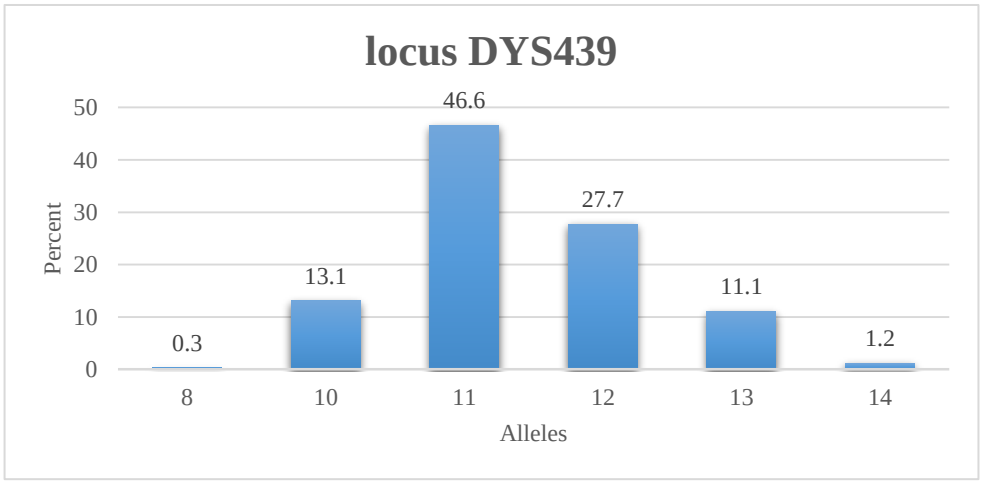
R. Locus DYS460

Alleles for the locus DYS460			
Alleles No.	Allele repeats	Freq.	Percentage
1	7	2	0.6
2	8	1	0.3
3	9	14	4.1
4	10	163	47.5
5	11	141	41.1
6	12	22	6.4
Total		343	100



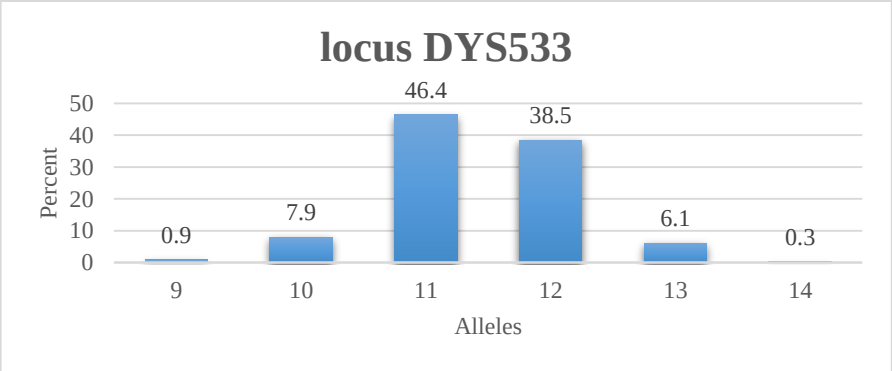
S. Locus DYS439

Alleles for the locus DYS439			
Alleles No.	Allele repeats	Freq.	Percentage
1	8	1	0.3
2	10	45	13.1
3	11	160	46.6
4	12	95	27.7
5	13	38	11.1
6	14	4	1.2
Total		343	100



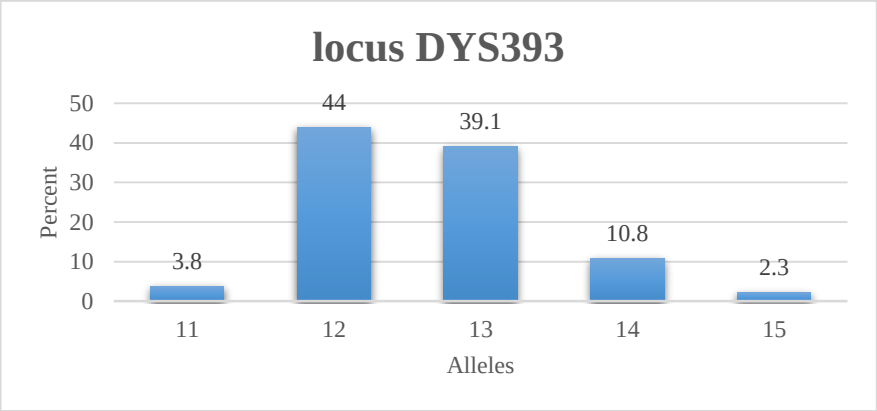
T. Locus DYS533

Alleles for the locus DYS533			
Alleles No.	Allele repeats	Freq.	Percentage
1	9	3	0.9
2	10	27	7.9
3	11	159	46.4
4	12	132	38.5
5	13	21	6.1
6	14	1	0.3
Total		343	100



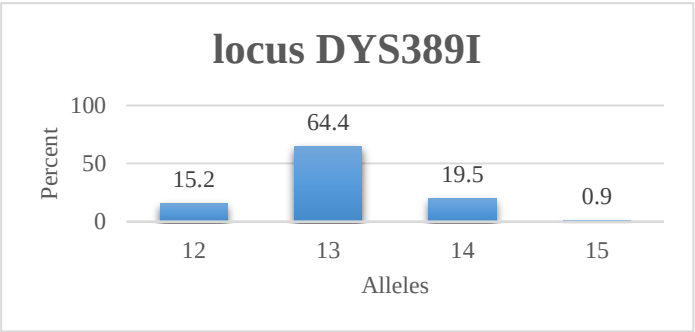
U. Locus DYS393

Alleles for the locus DYS393			
Alleles No.	Allele repeats	Freq.	Percentage
1	11	13	3.8
2	12	151	44
3	13	134	39.1
4	14	37	10.8
5	15	8	2.3
Total		343	100



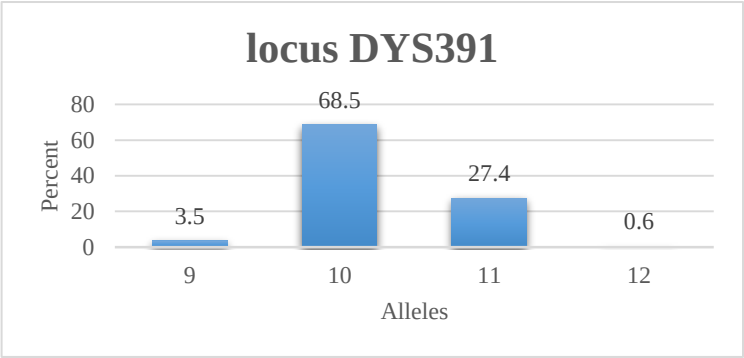
V. Locus DYS389I

Alleles for the locus DYS389I			
Alleles No.	Allele repeats	Freq.	Percentage
1	12	52	15.2
2	13	221	64.4
3	14	67	19.5
4	15	3	0.9
Total		343	100



W. Locus DYS391

Alleles for the locus DYS391			
Alleles No.	Allele repeats	Freq.	Percentage
1	9	12	3.5
2	10	235	68.5
3	11	94	27.4
4	12	2	0.6
Total		343	100



X. Locus DYS437

Alleles for the locus DYS437			
Alleles No.	Allele repeats	Freq.	Percentage
1	14	246	71.7
2	15	72	21
3	16	24	7
4	17	1	0.3
Total		343	100