6-2020

Short Tandem Repeats Diversity in Indian and Pakistani Population Living In United Arab Emirates

Ruksar Salim Damji

United Arab Emirates University

College of Science

Department of Biology

SHORT TANDEM REPEATS DIVERSITY IN INDIAN AND PAKISTANI POPULATION LIVING IN UNITED ARAB EMIRATES

Ruksar Salim Damji

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 Professor Synan AbuQamar

June 2020

Declaration of Original Work

I, Ruksar Salim Damji, the undersigned, a graduate student at the United Arab Emirates University (UAEU), and the author of this thesis entitled "*Short Tandem Repeats Diversity in Indian and Pakistani Population Living in United Arab Emirates*", 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 Synan AbuQamar, 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 Any materials borrowed from other sources (whether published or unpublished) and relied upon or included in my thesis have been properly cited and acknowledged in accordance with appropriate academic conventions.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: 29/7/20

Advisory Committee

1) Advisor: Professor Synan AbuQamar

Title: Professor

Department of Biology

College of Science

2) Co-advisor: Dr. Rashed Alghafri

Title: Adjunct Associate Professor

Department of Biology

College of Science

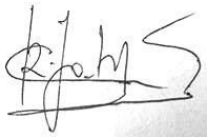
Approval of the Master Thesis

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

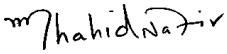
1) Advisor (Committee Chair): Professor Synan AbuQamar
   Title: Professor
   Department of Biology
   College of Science

   Signature  Date 14/06/20

2) Member: Professor Rabah Iratni
   Title: Professor
   Department of Biology
   College of Science

   Signature  Date 14/06/20

3) Member (External Examiner): Dr. Muhammad Shahid Nazir
   Title: Associate Professor
   Department of Biotechnology
   Institution: University of Modern Sciences, Dubai, UAE

   Signature  Date 14/06/20

This Master Thesis is accepted by:

Acting Dean of the College of Science: Professor Maamar Benkraouda

Signature Maamar Benkraouda Date August 6, 2020

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

Signature Ali Hassan Date August 6, 2020

Copy ____ of ____

Abstract

Short Tandem Repeats (STRs), have become increasingly popular markers of choice due to its wide array of advantages in the sector of forensic investigation. It is necessary to expand genetic research into residing populations i.e., Indian and Pakistani, in the same geographic region such as in the United Arab Emirates (UAE). The objectives of this study were to: (1) assess the forensic efficiency parameters and population structure analysis of the autosomal STR loci for the most potent amplification kit (2) estimate the allele frequencies; and (3) determine the significance of increasing the number of STR loci used in forensic DNA analysis. This study focused on 23 autosomal STR loci, namely D3S1358, vWA, D16S539, CSF1PO, TPOX, D8S1179, D21S11, D18S51, D2S441, D19S433, TH01, FGA, D22S1045, D5S818, D13S317, D7S820, D10S1248, D1S1656, D12S391, D2S1338, D6S1043, Penta D and Penta E which were evaluated in a total of 701 Indian and Pakistani population, living in the UAE. Blood samples were collected and the studied loci were amplified using VeriFiler™Express PCR Amplification Kit and electrophoresed using ABI 3500 Genetic Analyzer. Arlequin and Powerstat software was utilized to analyze the forensic parameters and population structure analysis. A total of 248 and 298 alleles was observed in the Indian and Pakistani population, respectively. Among all studied loci, prominent results were attained from Penta E. Genetic diversity between the two populations, ranged from 70% (TPOX) to 92% (Penta E). The combined probability of power of exclusion (CPE), power of discrimination (CPD) and random match probability (CMP) in the Indian population was 0.999999991519, 0.999999999999999999999999989897 and $4.8x10^{-26}$, respectively. Whereas, the Pakistani population showed combined probability of 0.999999990719, 0.999999999999999999999999999901 and $4.4x10^{-26}$, following the same order of parameters previously mentioned. This confirms that these 23 loci are suitable for individual identification, paternity testing, kinship analysis and population diversity studies in forensic practice.

Keywords: Indian Population, Pakistani Population, Penta E, Population Diversity, Power of Discrimination, Short Tandem Repeats, United Arab Emirates.

Title and Abstract (in Arabic)

تنوع تكرارات الترددات القصيرة في عينات من الشعب الهندي والشعب الباكستاني المقيمين في دولة الامارات العربية المتحدة

الملخص

أصبح استخدام البصمة الوراثية واسع الانتشار في مجال الأدلة الجنائية وبالأخص تقنية الترددات القصية المتكررة (STR) المنتشرة في الحمض النووي البشري في مواقع متفرقة نظراً لإيجابياتها المتعددة. فهذه التقنية لها استخدامات في الأبحاث الجينية وتشمل التعرف على هوية الأشخاص وإثبات النسب للأشخاص ودراسة التركيبات السكانية كالشعب الهندي والباكستاني على نفس المنطقة الجغرافية وهي دولة الإمارات العربية المتحدة في هذه الدراسة. أهداف هذا البحث شملت: 1) تقييم كفاءة العوامل الجنائية، وتحليل التركيبة السكانية. 2) تقدير تكرر الترددات الأليلية لأقوى طريقة مستخدمة لتكثيف المواقع. 3) تحديد أهمية زيادة عدد العينات و والترددات القصيرة المنتشرة في مواقع متفرقة من الحمض النووي في التحليل الجنائي. في هذه الدراسة تم التركيز على تحديد نسب الترددات لثلاثة وعشرون موقع؛ D3S1358, vWA, D16S539, CSF1PO, TPOX, D8S1179, D21S11, D18S51, D2S441, D19S433, TH01, FGA, D22S1045, D5S818, D13S317, D7S820, D10S1248, D1S1656, D12S391, D2S1338, D6S1043, Penta D and Penta E. وذلك من خلال تحليل 701 عينة من الشعبين الباكستاني والهندي المقيمين في دولة الامارات العربية المتحدة. لهذه الدراسة تم جمع عينات الدم وتحليلها باستخدام البصمة الوراثية (Verifiler™ Express Amplification Kit). بعد إتمام عملية تكثيف تلك المواقع من خلال استخدام تقنية التفاعل البلمري (PCR) تمت قراءة تلك المواقع باستخدام تقنية الفصل الكهربائي عن طريق الشعيرات (Capillary Electrophoresis) باستخدام جهاز (Genetic Analyzer 3500). بالإضافة للقيام بالحسابات الإحصائية باستخدام برنامج (PowerStat) و (Arlequin) من أجل دراسة المتغيرات والمؤمشرات الجنائية. عن طريق دراسة 248 و 298 أليل من الشعب الهندي والباكستاني على التوالي. ضمن كل المواقع الجينية التي تمت دراستها معظم النتائج كانت من التكرار الخماسي E. فقد كان التنوع الجيني بين الشعب الهندي والباكستاني بنسبة تتراوح ما بين 70% للتكرار خماسي E و 92% لمواقع TPOX. والتي من خلالها تم تحديد قوة التمييز للبصمة الوراثية المستخدمة وقوة استبعاد الأشخاص واحتمالية التطابق العشوائي مع شخص مجهول كالآتي: 0.999999991519

و0.99999999999999999999999989897 و$4.8x10^{-26}$ على التوالي في الشعب الهندي، بينما في الشعب الباكستاني كانت القياسات كالتالي 0.999999990719 و0.99999999999999999999999999990 و $4.4x10^{-26}$ على التوالي. والذي بدوره أثبت أن هذه ال 23 مواقع الجينية التي تم تحليلها يمكن الاعتماد عليها للتعرف على هوية الأشخاص، اختبار الأبوة، وإثبات النسب للأشخاص، ودراسة التركيبة السكانية والتي عادة تستخدم بالتحاليل الجنائية.

مفاهيم البحث الرئيسية: الشعب الهندي، الشعب الباكستاني، التكرار الخماسي E، التنوع السكاني، قوة التمييز، تكرارات الترددات القصيرة، الإمارات العربية المتحدة.

Acknowledgements

I would like to extend my deepest gratitude and respect to Prof. Synan AbuQamar and Dr. Rashed Alghafri for their sincere guidance and unceasing help. This accomplishment would have not been possible without them.

I humbly extend my thanks to Mr. Mohammed Naji Mohammed, Ms. Reem Almheiri and Mr. Ahmed Adan for their contribution. Their timely suggestions with kindness and enthusiasm kept me going.

I also profusely thank Dubai Police HQ for providing me with all the necessary facilities; and all the staff members for their cooperation.

Last but not least, I am eternally grateful to my parents and siblings for showing me unfailing support and encouragement throughout my years of study. Thank you!

Dedication

To my beloved parents

Table of Contents

List of Tables

List of Figures

List of Abbreviations

A, T, C, G	Adenine, Thymine, Cytosine, Guanine
AMOVA	Analysis of Molecular Variance
bp	Base Pairs
CMP	Combined Match Probability
CPD	Combined Power of Discrimination
CPE	Combined Power of Exclusion
DNA	Deoxyribonucleic acid
HWE	Hardy Weinberg Equilibrium
HE	Heterozygosity
HO	Homozygosity
MP	Match Probability
MDS	Multidimensional Scaling Plot
PI	Paternity Index
PCR	Polymerase Chain Reaction
PIC	Polymorphic Information Content
PD	Power of Discrimination
PE	Power of Exclusion
RFLP	Restriction Fragment Length Polymorphism
STR	Short Tandem Repeat
VNTR	Variable Number Tandem Repeat

Chapter 1: Introduction

1.1 Overview on the Discovery of DNA

In early 1950s, James Watson and Francis Crick demonstrated that DNA had a complex structure; yet was simple enough to be the master molecule of life. Using X-ray crystallography, Rosalind Franklin contributed to determine the helical, two-stranded structure of DNA (Watson & Crick, 1953). Moreover, Erwin Chargaff identified that Adenine (A) stringently complements with Thymine (T); whereas Guanine (G) complements Cytosine (C) by two and three hydrogen bonds, respectively between the two strands of DNA. The purines A and G have a double ring structure; whereas T and C are considered pyrimidines with a single ring structure (Griffiths et al., 1999).

1.2 Chromosomal DNA

In eukaryotes, human nuclear DNA is divided into chromosomes. Human genome consists of 22 pairs of autosomal chromosomes and a pair of sex-determining chromosomes (a total of 46 or 23 pairs of chromosomes). Chromosomal pairs 1-22 are categorized according to decreasing size. Males have a single copy of the X-chromosome and a single copy of the Y-chromosome, designated as XY (Butler, 2001). However, females have two copies of the X-chromosomes, designated as XX. DNA in chromosomes is made of coding and noncoding regions. Merely 1 % of DNA is made up of protein-coding genes; whereas the remaining 99 % is considered as noncoding regions that does not serve instructions for making proteins. Throughout the noncoding regions of the human genome, polymorphic markers that vary among individuals can be found. Identity tests are performed using markers on the autosomal

chromosomes and gender determination is carried out using markers on the sex chromosomes (Butler, 2009).

A DNA marker, or a specific location of a gene in the chromosome, is generally specified as a locus. Thousands of loci have been identified and mapped to specific regions of human chromosomes through the achievement of the global effort of the human genome project. The alternate possibilities for a gene or a genetic locus are referred to as alleles. In a genetic locus of a homologous chromosome, different alleles are heterozygous and identical alleles are homozygous. The perceptible differences in alleles at equivalent loci are important for human identity testing (Butler, 2009). Genetic testing can provide information about the individual's genes and/or chromosomes (Saad, 2005). Despite the fact that 99 % of the DNA sequences are identical amongst individuals, 1 % of the DNA sequences dissimilarities can be used to differentiate between individuals, and are, therefore, targeted for human identity testing.

1.3 Discovery of DNA fingerprinting

Historically, identity testing in the field of forensics began with the analysis of the ABO blood group system. In human, blood groups with multiple alleles display a range of dominance patterns (Adams, 2008). Subsequently, new markers for identification were based on differences in red blood cell enzymes and serum proteins (Abu Halimah, 2008). In 1984, Professor Alec Jeffreys, identified that variations could be detected within the human DNA rather than from the protein products. This lead to the discovery of the technique DNA fingerprinting. The term DNA fingerprint represents a pattern that is unique to an individual. By examining an adequate number of DNA regions showing variability among each individual, one can lessen the

probability of an inclusion of two individuals to an immensely low level. The probability is so low that DNA fingerprinting is a method not only for inclusion or exclusion, but also for absolute identification (Jeffreys et al., 1985).

DNA fingerprinting has undergone three primary phases of technological advancement. These included multi-locus restriction fragment length polymorphism (multi-locus RFLP), single-locus restriction fragment length polymorphism (single-locus RFLP) and lastly the short tandem repeats (STRs) stage (Abu Halimah, 2008). Multi-locus RFLP probes was the first method to be developed (Jeffreys et al., 1985). Tandemly repeated DNA sequences that varied among individuals was digested into fragments of distinct lengths using restriction enzymes and separated on agarose gels. The radioactive multi-locus probes collectively hybridized various minisatellite loci. These probes are highly variable among individuals but require a great deal of labor, time and expertise to produce a DNA profile (Butler, 2009). Deciphering sample mixture is common in forensic cases, but is a challenge with the multi-locus RFLP method (Butler, 2005). The following stage in the development of DNA fingerprinting used the same RFLP technology. Nonetheless, the probes utilized to visualize the product was altered to aim at one distinct locus at a time. This system was known as single-locus RFLP probes in New Zealand & United Kingdom (Buckleton et al., 2016). In the United States it was, however, referred to as variable number tandem repeats (VNTRs) (Houck & Siegel, 2015). The VNTRs contain sets of tandemly repeated base pair sequences that differ in their lengths (10-100 bp). The forensic DNA marker (D1S80) is an example of VNTR that consists of a 16-bp repeat unit (Figure 1).

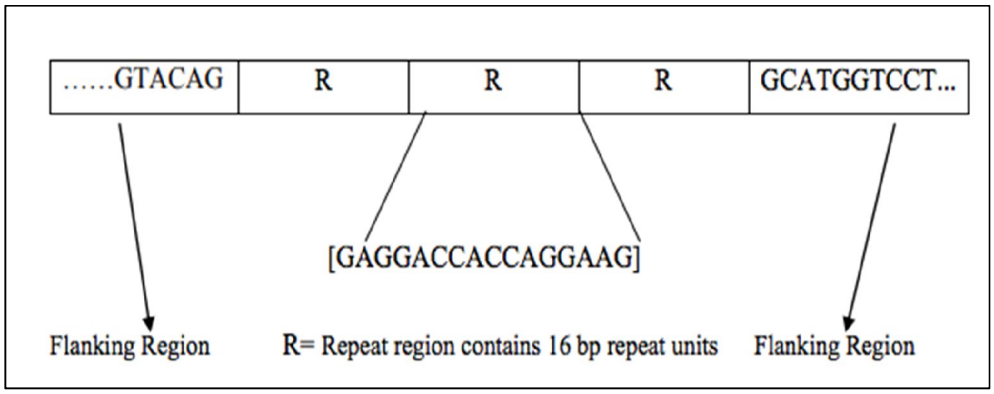


Figure 1: VNTR marker D1S80 (Aydin, 2015)

VNTR analysis can be utilized thoroughly for paternity disputes and for applications where large quantities of intact DNA can be attainable to accumulate (Kashyap et al., 2004). There are plenty of samples that can be recovered from crime scenes, yielding barely picogram or nanogram of DNA, which is intermittently degraded and thereupon not applicable for VNTR analysis (Herrera et al., 2016). In order to overcome this barrier, scientists relied upon the polymerase chain reaction (PCR) to intensify the shorter hypervariable regions which are known as short tandem repeats (STRs). In mid1990s, technology was amended to include the usage of PCR for STR loci (Buckleton et al., 2016).

1.4 Short Tandem Repeats

Short tandem repeats (STRs), also referred to as simple sequence repeats (SSRs) or microsatellites, are DNA sequences consisting of repetitive units (2-6 bp in length) (El-Alfy & Abd El-Hafez, 2012). STRs are known to be highly beneficial in applications such as human identification, population genetic studies, paternity testing, kinship analysis and much recently disease diagnosis. The majority of the human identity testing is executed utilizing STR markers on the autosomal chromosomes

since they are well-established and highly effective for elucidating genetic diversity (Butler, 2009).

The density of the STRs differ among the chromosomes. In humans, the highest density of STRs can be found in chromosome 19. Roughly, one STR appears per 2,000 bp in the human genome. A-rich units including: A, AC, AAAN, AAN, and AG are the most common STRs (Fan & Chu, 2007).

STR analysis centralizes on increasing discrimination power, sensitivity and genetic polymorphism (Preet et al.,2016). High degree of polymorphism in STR as a forensic genetic marker is extremely essential to increase the discrimination power and lower the match probability (Jebor, 2013). Studies have shown that STRs are diverse at both individual and population levels. Therefore, population studies have been conducted to determine the individualizing power of STR markers (Grubb, 2015).

STR markers are the preferred and the ideal genetic markers for a number of reasons (Abu Halimah, 2008). Firstly, they are distributed randomly and found in abundance in the human genome. Secondly, they are hypervariable within populations and portray higher mutation rates in comparison to other nuclear regions. Most importantly, STR markers genotyping necessitates only minuscule amounts of template DNA, considering that it is PCR-based (Norrgard, 2008). Sufficient amount of DNA can be extracted from a single shed of hair, a piece of tissue, minute amount of blood or epithelial cells sloughed off in saliva, feces or urine (Abu Halimah, 2008). In STR typing, PCR is utilized to recover information from small quantities of biological material that is available. The short PCR product sizes (100-500 bp) achieved with STR testing are compatible with the DNA that is degraded, and are likely to be present on the biological matter due to environmental insults found at the

crime scene. PCR amplification of multiple STR loci simultaneously (also known as multiplexing) is attainable with various colored fluorescent dyes and PCR product sizes. Utilizing multiple loci allows a high power of discrimination in just one single test without consuming high amounts of DNA (Butler, 2009).

1.4.1 Types of STR Markers

STR markers are distinguished by the length of repeats the locus accommodates as well as the type of repeat unit present (Michelle & Graham, 2015). Dinucleotide repeats have two nucleotides repeated next to each other several times. Trinucleotides have three nucleotides repeated, tetranucleotides have four, pentanucleotides have five and hexanucleotides have six repeat units. Among these various types of STRs, tetranucleotides are more favorable and popular in comparison to di- or tri-nucleotides. Tetranucleotides have a narrow allele size range that allows multiplexing and reduces allelic dropouts from amplifying smaller alleles. However, penta- and hexa-nulceotides are not very common in the human genome. STR sequences not only differ in the length of the repeat unit but also in the accuracy with which they conform to an accumulative repeat pattern (Abu Halimah, 2008).

STRs are often divided into various categories based on the repeat pattern. Simple repeats consist of one unit of similar length and sequence, such as Penta E, TPOX, TH01 D5S818, CSF1PO and D16S539 as shown in Figure 2 (Woerner et al., 2017). Simple repeats can also contain non-consensus alleles that lay between alleles with full repeat units. Compound repeats, on the other hand, constitute of two or more adjacent simple repeats, such as FGA, vWA, D3S1358 and D8S1179 (Butler, 2012; Abu Halimah, 2008). The complex repeat STR *i.e* D21S11 may include many repeat blocks of variable length and intervening sequences. Complex hypervariable also

comes with non-consensus alleles that varies in the sequence as well as the size; and therefore, greater efforts are required to reproduce the genotype (Abu Halimah, 2008; Venables, 2012).

Not all alleles for a STR locus consist of complete repeat units. Micro-alleles (or micro-variants) are an STR with a fractional value rather than a whole integer, that complicate assigning alleles (Huel et al., 2007). They are named for the number of full repeat unit and the portion of the partial repeat unit. For example: The locus TH01; 9.3 micro-variant has nine complete tetranucleotide repeats and a partial incomplete repeat unit of three nucleotides. TH01, D18S51 and D7S820 loci are examples of simple repeat STRs that have common micro-variant alleles. These fractional values usually arise from a deletion, a mutational loss of one or more nucleotides (Abu Halimah, 2008).

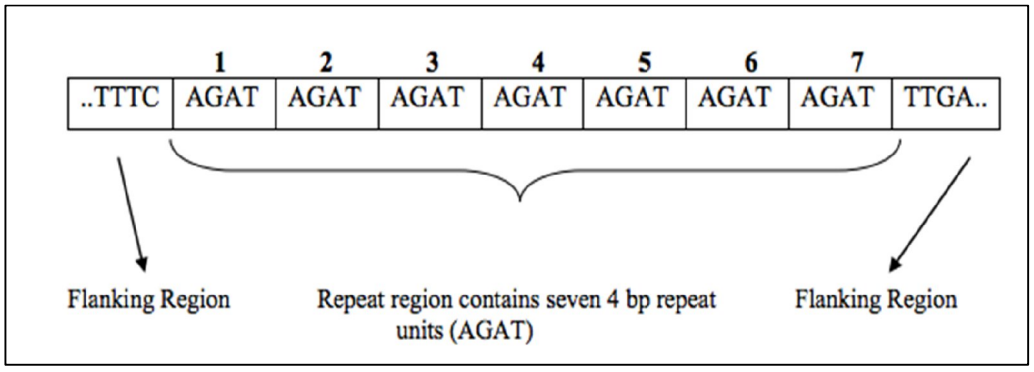


Figure 2: STR marker – D5S818 simple repeats (Aydin, 2015)

A biological occurrence known as stutter products may result when STR alleles are PCR amplified. Stutter artefacts are amplicons that vary in length in comparison to the true allele by one or more whole repeat units (Brookes et al., 2012). These artefacts arise during PCR when the DNA polymerase detaches from the DNA and is misaligned when it re-associates (replication slippage), leading to deletions or

additions of repeat units. Stutter artefacts associated with tetranucleotide repeats have approximately a peak height of less than 15 % in comparison to the true alleles peak height (Abu Halimah, 2008). Dinucleotide and trinucleotide repeats can, however, show artefacts where the peak height is 30 % of the true allele. Using tetranucleotide results in a decline of stutter product formation when compared to dinucleotide or trinucleotide repeats (Butler, 2005).

Furthermore, by the process of replication slippage, these repetitive sequences have the ability to expand or contract. When the template and nascent strands disassociate and misalign, hairpin loops are produced. These loops can lead to repeat expansion on the nascent strand or repeat contraction on the template strand (Ryan, 2019). Expansion of STRs have been recently identified as the causal DNA mutation in over 40 Mendelian disorders (Gymrek et al., 2016). Several developmental and neurological disorders including: Huntington's disease, fragile X syndrome, polyalanine disorders and spinocerebellar ataxias have been associated with the unstable expansion of STRs (Halman & Oshlack, 2020). The likelihood that an individual is affected with the disease increases with the repeat length. Additionally, the severity of the disease is dependent upon the gender of the parent who transmitted the expansion (Dashnow et al., 2018). Further understanding of the molecular mechanisms causing the instability of the STRs will ease the efforts in developing treatment and novel preventative approaches for repeat expansion disorders (Sun et al., 2018).

In the human genome, STRs are known to be rapidly mutating markers, with mutation rates of 10-100,000 times greater than the average rate, and therefore contribute to a wide fraction of genetic variation (Gymrek, 2017). The autosomal STR

loci which have the lowest observed mutation rates include CSFIPO, TH01, TPOX, D8S1179 and D5S818. As expected, the locus with the highest mutation rates (D18S51, D7S820, D16S539, D21S11 and FGA) are amongst the most polymorphic loci with the highest number of alleles observed (Omran, 2008). Having knowledge of the mutation rates of STR loci is essential, particularly in population genetics and paternity testing, where conclusions are being made from the genetic data from one or several generations.

When a microsatellite locus is identified in the genome, oligonucleotide primers are designed from the DNA sequences upstream and downstream of the microsatellite loci in order to amplify that specific fragment by PCR. By utilizing electrophoresis, the microsatellite marker variation can be directly assayed. PCR products can be visualized in polyacrylamide gels. Two PCR product bands will be shown for heterozygous individuals, whereas a single band will be displayed for homozygotes (Abu Halimah, 2008).

1.5 STR Genotyping

1.5.1 Capillary Electrophoresis

In forensic DNA laboratories, capillary electrophoresis (CE) is the primary method utilized for separating and detecting STR alleles (Butler, 2012). Capillaries were brought into electrophoresis as a heat controlling and anti-convective innovation (Petersen & Mohammad, 2001). It is now widely used due to its simple operation and high peak resolution. The apparatus for CE instrument consists of a narrow capillary tube, high voltage direct current power supply, two buffer vials which are placed at the same level and accommodating cathodic and anodic solutions. Two electrodes

assemblies (cathode and anode) connected to the power supply. The CE instrument also contains an auto-sampler to hold the tray containing samples, a laser excitation source, a fluorescence detector and a computer in order to control the sample injection and detection (Butler, 2005).

The CE differs from the other forms of electrophoresis as it is performed within a confined narrow capillary tube (Whatley, 2001). This tube is filled with a polymer solution rather than a gel in order to carry out DNA size separation. The ratio of higher surface area to volume allows more effective dissipation of heat obtained by the process of electrophoresis and consequently allows greater separation voltage to be applied. The fluorescently labelled DNA fragments are separated according to size before it reaches the positive electrode and proceeds across the path of a laser beam. The dyes attached to the fragment fluoresce due to the laser beam. The dye signals are separated by the diffraction system and the charge coupled device camera is used to detect the fluorescence. Every dye has a distinct wavelength of the light emitted. The data collection software gathers the fluorescence signals and converts it into an electrophorogram. Using an internal size standard with a fluorophore permits resolving the sizes of the DNA fragments present in the sample (Butler, 2005).

1.5.2 Genetic Analyzer

In this project, the 3500 Genetic Analyzer (Applied Biosystems) was utilized. It is a fluorescence-based DNA analysis instrument utilizing the technology of capillary electrophoresis with 8 capillaries, each 36 cm in span. This analyzer has advanced multiplexing capabilities for DNA analysis with up to six dyes. Since it is fully automated, the possibilities of human error during the post amplification handling

as well as samples analysis can be reduced. This is a major advantage when large amounts of samples are processed.

1.6 Significance of the Study Undertaken

The present study determined the significance of ongoing research into specific populations, even those residing in the same geographic region. It is necessary to expand genetic research to all the populations globally since the allele frequencies of each marker is specific to each population. The data achieved can consequently be used to build an allelic frequency database for the most potent amplification kit in the Indian and Pakistani population, residing in UAE.

Moreover, this study revealed the importance of increasing the number of STR loci examined in DNA analysis. The VeriFiler™Express kit may eventually replace the amplification kits used today in forensic laboratories since this is at present the only kit that amplifies the highest number of autosomal STR loci simultaneously, leading to obtaining the maximum power of discrimination with just miniscule amounts of DNA.

1.6.1 Hypothesis

1) The 23 autosomal STR loci was highly polymorphic and discriminatory in the Indian and Pakistani population, living in the UAE.

2) The higher the number of STR loci examined, the higher the power of discrimination and the lower the match probability.

1.6.2 Objectives

1) To assess the forensic efficiency parameters and population structure analysis of 23 autosomal STR loci using the VeriFiler™Express PCR Amplification Kit, in the Indian and Pakistani population, residing in the UAE.

2) To estimate the allele frequencies of the 23 autosomal STR loci; and

3) To determine the significance of increasing the number of STR loci used in forensic DNA analysis.

Chapter 2: Materials and Methods

2.1 Sample Collection

Consented blood samples were collected from a total of 701 unrelated individuals. This consisted of two populations residing in the UAE; Indian (n = 352), Pakistani (n = 349). Ethical approval was granted from Dubai scientific research ethics committee in Dubai Health Authority (DSREC-SR-08/2018-03).

2.2 Kit Overview

The VeriFiler™ Express PCR Amplification Kit is a 6-dye, STR multiplex assay for the amplification of single-source human genomic DNA. The kit amplifies the following loci within the read region of 75–465 nt:

- 23 autosomal STR loci (D3S1358, vWA, D16S539, CSF1PO, TPOX, D8S1179, D21S11, D18S51, D2S441, D19S433, TH01, FGA, D22S1045, D5S818, D13S317, D7S820, D10S1248, D1S1656, D12S391, D2S1338, D6S1043, Penta D, Penta E).

The VeriFiler™ Express PCR Amplification Kit includes:

- Master mix which contains $MgCl_2$, dATP, dGTP, dCTP, and dTTP, bovine serum albumin, enzyme and 0.05 % sodium azide in buffer and salt
- Primer Set which contains locus-specific 6-FAM™, VIC™, NED™, TAZ™, and SID™ dye-labeled and unlabeled primers in buffer
- DNA control 007 contains 2.0 ng/µL human male genomic DNA in 0.05 % sodium azide and buffer
- Allelic ladder which consists of the most commonly found alleles per locus obtained from large population base study (Table 1).

Table 1: Common alleles per locus obtained from a large population base study

Locus designation	Chromosome location	Alleles included in the allelic ladder	Dye label	DNA control 007
D3S1358	3p21.31	9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20	6-FAM™	15, 16
vWA	12p13.31	11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24		14, 16
D16S539	16q24.1	5, 8, 9, 10, 11, 12, 13, 14, 15		9, 10
CSF1PO	5q33.3-34	6, 7, 8, 9, 10, 11, 12, 13, 14, 15		11, 12
TPOX	2p23-2per	5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15		8, 8
D8S1179	8q24.13	5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19	VIC™	12, 13
D21S11	21q11.2-q21	24, 24.2, 25, 26, 27, 28, 28.2, 29, 29.2, 30, 30.2, 31, 31.2, 32, 32.2, 33, 33.2, 34, 34.2, 35, 35.2, 36, 37, 38		28, 31
D18S51	18q21.33	7, 9, 10, 10.2, 11, 12, 13, 13.2, 14, 14.2, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27		12, 15
Penta E	15q26.2	5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26		7, 12
D2S441	2p14	8, 9, 10, 11, 11.3, 12, 13, 14, 15, 16, 17	NED™	14, 15
D19S433	19q12	6, 7, 8, 9, 10, 11, 12, 12.2, 13, 13.2, 14, 14.2, 15, 15.2, 16, 16.2, 17, 17.2, 18.2, 19.2		14, 15
TH01	11p15.5	4, 5, 6, 7, 8, 9, 9.3, 10, 11, 13.3		7, 9.3
FGA	4q28	13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 26.2, 27, 28, 29, 30, 30.2, 31.2, 32.2, 33.2, 42.2, 43.2, 44.2, 45.2, 46.2, 47.2, 48.2, 50.2, 51.2		24, 26

Table 1: Common alleles per locus obtained from a large population base study (continued)

Locus designation	Chromosome location	Alleles included in the allelic ladder	Dye label	DNA control 007
D22S1045	22q12.3	8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19	TAZ™	11, 16
D5S818	5q21-31	7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18,		11, 11
D13S317	13q22-31	5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16,		11, 11
D7S820	7q11.21-22	6, 7, 8, 9, 10, 11, 12, 13, 14, 15		7, 12
D6S1043	6q16.1	9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25		12, 14
D10S1248	10q26.3	8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19	SID™	12, 15
D1S1656	1q42.2	9, 10, 11, 12, 13, 14, 14.3, 15, 15.3, 16, 16.3, 17, 17.3, 18.3, 19.3, 20.3		13, 16
D12S391	12p13.2	14, 15, 16, 17, 18, 19, 19.3, 20, 21, 22, 23, 24, 25, 26, 27		18, 19
D2S1338	2q35	11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28		20, 23
Penta D	21q22.3	2.2, 3.2, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17		11, 12

2.3 DNA Extraction

Whatman FTA Card, is a filter paper product which was manufactured by GE Health Care (GE Healthcare, 2011). It is a paper matrix permeated with a proprietary formula containing chemicals that promote cell lysis and denaturation of protein with subsequent release of nucleic acids that are entrapped within the matrix of the card. It is stabilized within room temperature, therefore allowing long term storage and is now being utilized as an alternative method for DNA extraction. A major advantage of FTA cards includes low cost, transportation made easy, minimal storage space and simple

extraction protocols. Moreover, quantification is not needed since the amount of DNA in the punched disc is adequate for PCR. The manufacturer's protocol was followed to perform the amplification of the 23 STR loci; however, some alterations were done due to the large numbers of samples that were tested in this study. Only quarter volume was used per reaction (2.5 µL master mix, 2.5 µL primer set, 1.25 µL distilled water = total volume of 6.25 µL). Positive and negative controls were being run concurrently. For the positive control, DNA control 007, provided in the kit, was used (2.5 µL master mix, 2.5 µL primer set, 0.75 µL distilled water, 0.5 µL DNA Control 007). Negative control (2.5 µL master mix, 2.5 µL primer set, 1.25 µL distilled water, blank disc) was used to inspect for any DNA contamination. A stock was prepared of the master mix, primer set and distilled water in an eppendorf tube. A total volume of 6.25 µL was dispensed into each labeled (sample number) PCR tubes. FTA Cards that contained the blood samples were then punched using the 0.5-mm Harris Micro Puncher and the disc was added into each tube.

2.4 PCR Amplification

The PCR tubes containing the reaction mix and the punched disc were then vortexed followed by a quick spin. Amplification was performed on Veriti™PCR Thermocycler, with the following parameters shown in Table 2.

Table 2: Thermal cycling parameters for VeriFiler™Express PCR Amplification Kit

Initial Denaturation	Denaturation	Annealing	Extension	Final Extension	Overnight Hold
	26 Cycles				
95°C 1 minute	94°C 3 seconds	59°C 16 seconds	65°C 29 seconds	60°C 5 minutes	4°C up to 24 hours

After the run was completed, the samples were removed from the thermocycler and proceeded for capillary electrophoresis.

2.5 Reagents for STR

- GeneScan™600 LIZ™size standard v2.0: An internal size standard that yields precise sizing results for PCR products.
- Highly deionized formamide (Hi-Di formamide): Formulated with a stabilizer for ready use as an injection solvent in DNA analysis methods in the genetic analyzer. Formamide in DNA samples helps to halt the reannealing of the single strands of DNA after denaturation.
- VeriFiler™Express allelic ladder: Developed for accurate characterization of the alleles amplified by the kit. This allows automatic genotyping of most of the reported alleles for the loci in the kit as shown in (Figure 3).

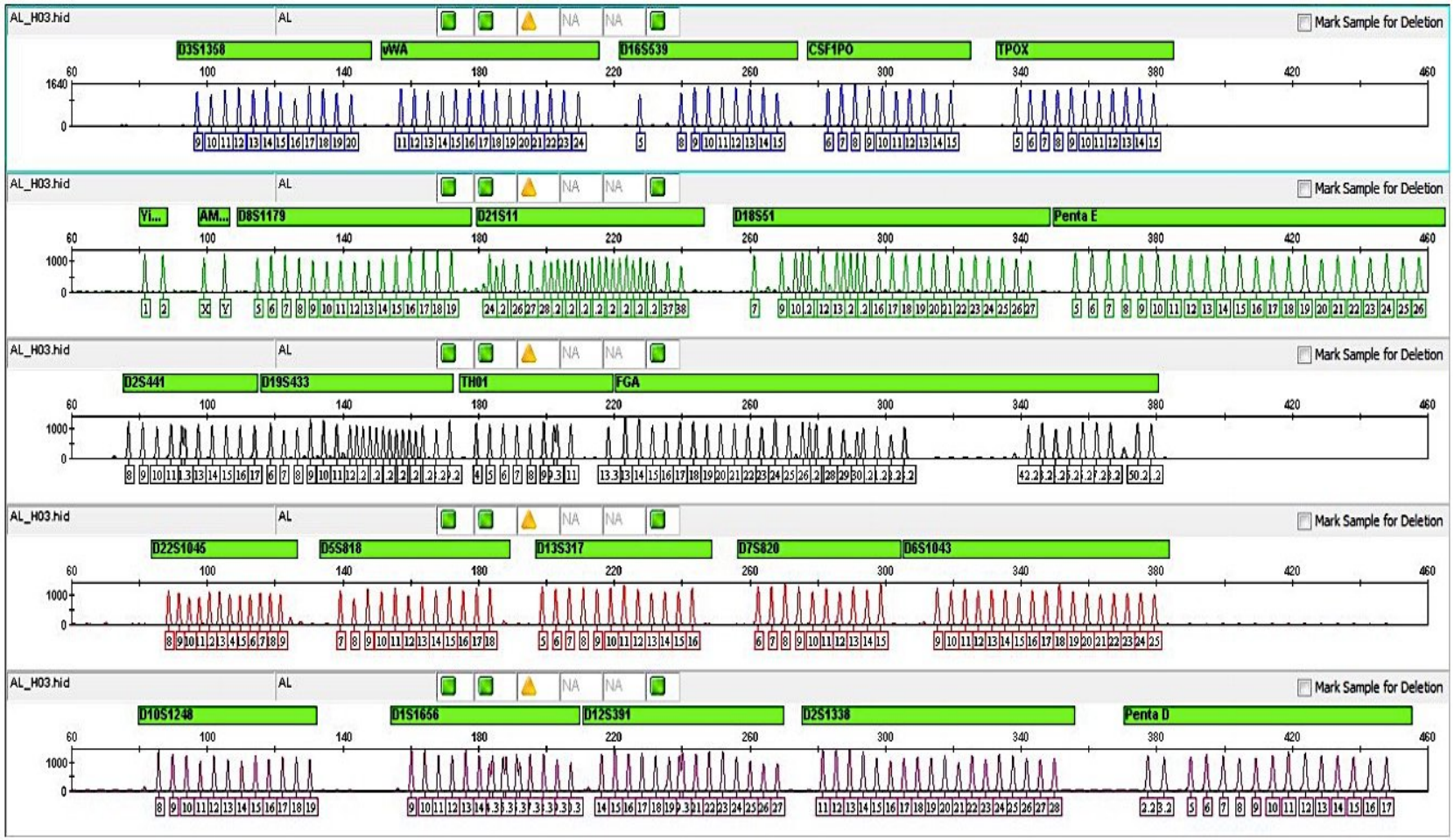


Figure 3: VeriFiler™Express allelic ladder profile. (manufacturer's protocol; Thermo Fisher Scientific)

2.6 Sample Preparation for Capillary Electrophoresis

A mix of Hi-Di formamide (10 μL per sample) and LIZ (0.5 μL per sample) was prepared. In the 96-well plate, 10 μL of the mix and 1 μL of the sample were dispensed into each well. A 10 μL of Hi-Di formamide was added to the empty wells to avoid injecting into dry wells. A 0.5 μL of VeriFiler™ Express Allelic Ladder was added to the well assigned for the ladder. A 0.5 μL of each positive and negative control was added to the respective wells. The 96-well plate was then covered with a septa and centrifuged for a few seconds. The well plate was placed into the 3500 Genetic Analyzer and the data were recorded into the software accordingly. Injection time was set to 8 seconds.

2.7 Statistical Analysis

The data for the samples were exhibited in the form of peaks that correspond to different STR alleles for each locus. The peaks were shown at different locations in the electropherogram of a sample and were often plotted as the intensity of the fluorescent signal versus the time passing the detector. The colors of the dye were separated and the DNA fragments, represented in the form of peaks, were classified and associated with a suitable color. The fragments of the DNA were compared and sized according to the internal sizing standard. The STR alleles were then compared to the allelic ladder to attain the genotype of the samples.

Data achieved from the 3500 Genetic Analyzer was then evaluated using the Gene Mapper ID-X Version 1.4 analysis software (Thermo Fisher Scientific, 2012). The raw data (genotypes) obtained from the Gene Mapper were exported and analyzed using Arlequin version 3.5 (Excoffier & Lischer, 2010) and Powerstats software.

Arlequin was utilized to attain the allele frequencies, predominant alleles, homozygosity, heterozygosity, analysis of molecular variance (AMOVA) and the deviation probability from the Hardy Weinberg Equilibrium (HWE) test. Powerstats was utilized to calculate the polymorphic information content (PIC), power of discrimination (PD), match probability (MP), power of exclusion (PE) and paternity index (PI). Multidimensional (MDS) plot was designed using XLSTAT software (XLSTAT, 2019).

Chapter 3: Results

3.1 DNA Profile

DNA profiles for 701 (352 Indian and 349 Pakistani) unrelated individual blood samples were compiled. Figure 4 shows an example of an autosomal STR profile displayed in Gene Mapper.

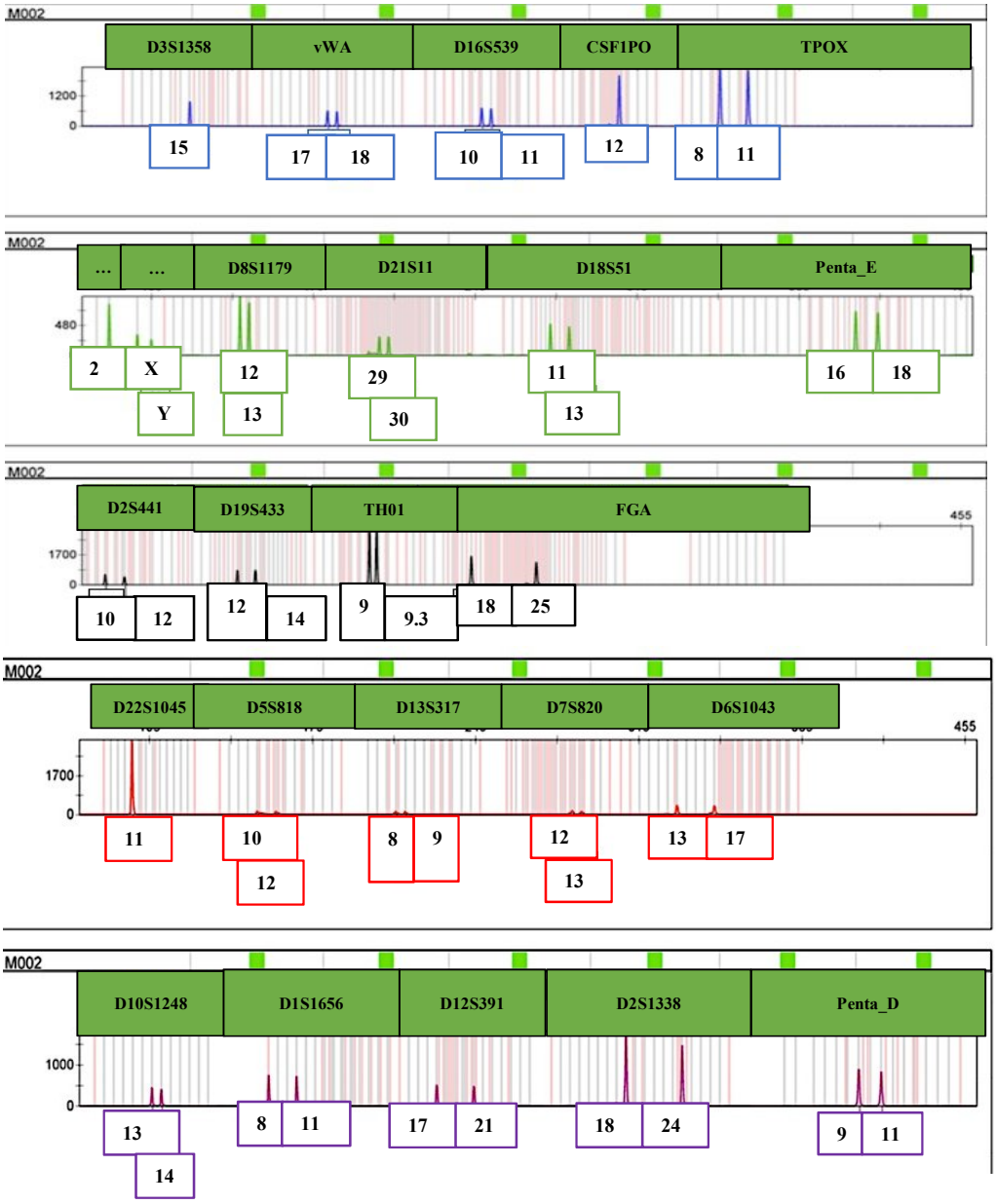


Figure 4: Autosomal STR profile generated utilizing VeriFiler™Express Kit

The allele frequency in a population indicates the degree of population genetic diversity as well as plays a vital role in interpreting DNA profiles for individual identification. Tables 3-25 and Figures 5-27 below show the allele frequencies of 23 autosomal STR loci examined among the Indian population. Tables 26-48 and Figures 28-50 show the allele frequencies observed in the Pakistani population.

Locus D3S1358 (Indian population)

Table 3: Allele frequency for the locus D3S1358 in the Indian population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	13	4.9984	0.71
2	14	27.0336	3.84
3	15	209.0176	29.69
4	16	213.0304	30.26
5	16.2	0.9856	0.14
6	17	181.984	25.85
7	17.1	1.9712	0.28
8	18	57.024	8.1
9	19	6.9696	0.99
10	20	0.9856	0.14

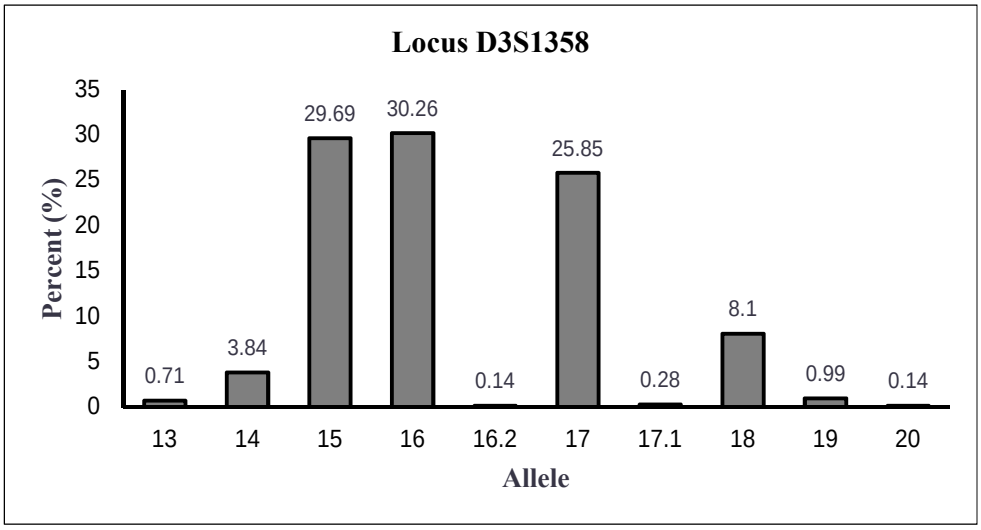


Figure 5: Percentage of allele frequency for the locus D3S1358 in the Indian population

Locus vWA (Indian population)

Table 4: Allele frequency for the locus vWA in the Indian population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	13	3.0272	0.43
2	14	92.0128	13.07
3	15	53.0112	7.53
4	16	161.9904	23.01
5	17	195.008	27.7
6	18	136.9984	19.46
7	19	53.9968	7.67
8	20	8.0256	1.14

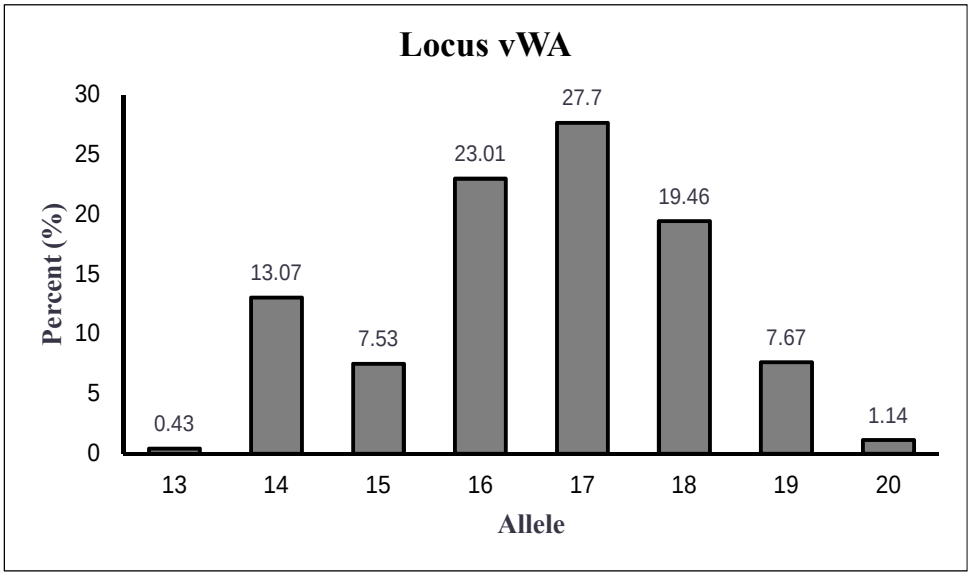


Figure 6: Percentage of allele frequency for the locus vWA in the Indian population

Locus D16S539 (Indian population)

Table 5: Allele frequency for the locus D16S539 in the Indian population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	8	43.0144	6.11
2	9	118.976	16.9
3	10	83.0016	11.79
4	11	229.0112	32.53
5	12	145.024	20.6
6	13	74.976	10.65
7	14	9.9968	1.42

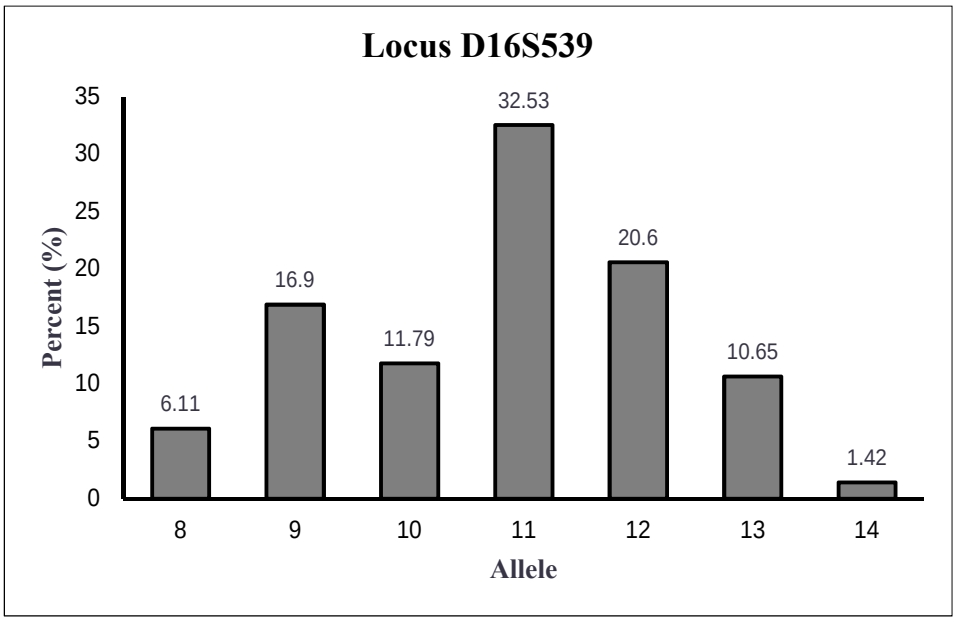


Figure 7: Percentage of allele frequency for the locus D16S539 in the Indian population

Locus CSF1PO (Indian population)

Table 6: Allele frequency for the locus CSF1PO in the Indian population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	7	1.9712	0.28
2	9	11.968	1.7
3	10	142.9824	20.31
4	11	224.0128	31.82
5	12	265.9712	37.78
6	13	47.0272	6.68
7	14	9.9968	1.42

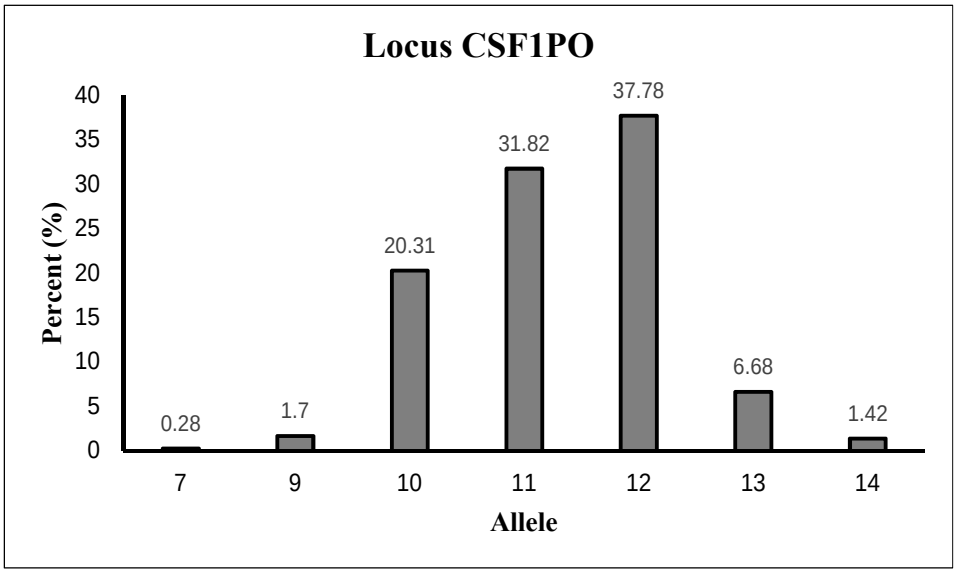


Figure 8: Percentage of allele frequency for the locus CSF1PO in the Indian population

Locus TPOX (Indian population)

Table 7: Allele frequency for the locus TPOX in the Indian population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	8	245.9776	34.94
2	9	106.0224	15.06
3	10	62.0224	8.81
4	11	270.9696	38.49
5	12	19.008	2.7

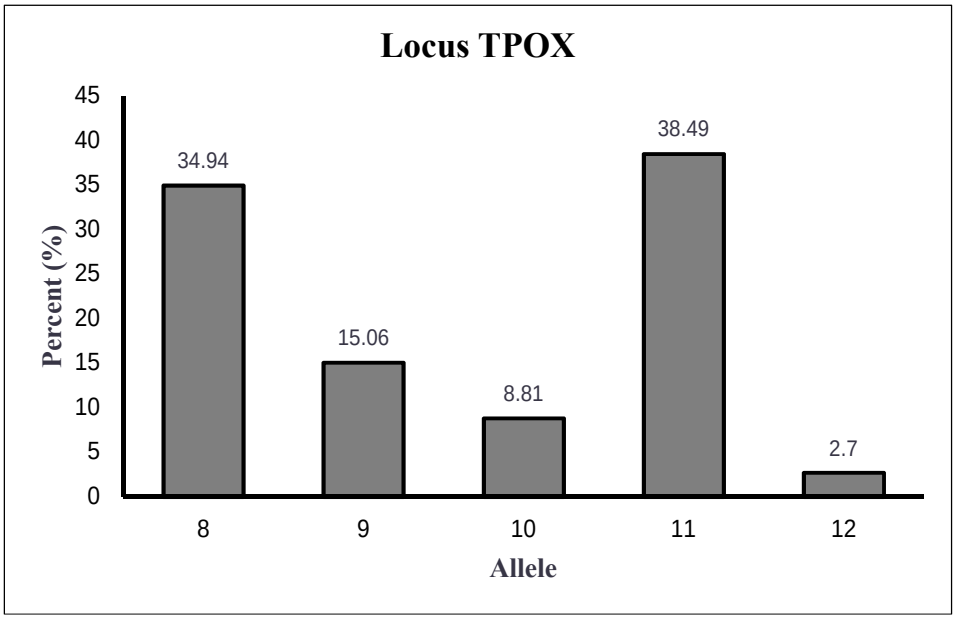


Figure 9: Percentage of allele frequency for the locus TPOX in the Indian population

Locus D8S1179 (Indian population)

Table 8: Allele frequency for the locus D8S1179 in the Indian population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	5	0.9856	0.14
2	8	6.9696	0.99
3	9	3.0272	0.43
4	10	113.9776	16.19
5	11	60.9664	8.66
6	12	68.992	9.8
7	13	121.0176	17.19
8	14	142.9824	20.31
9	15	122.9888	17.47
10	16	39.9872	5.68
11	17	20.9792	2.98
12	18	0.9856	0.14

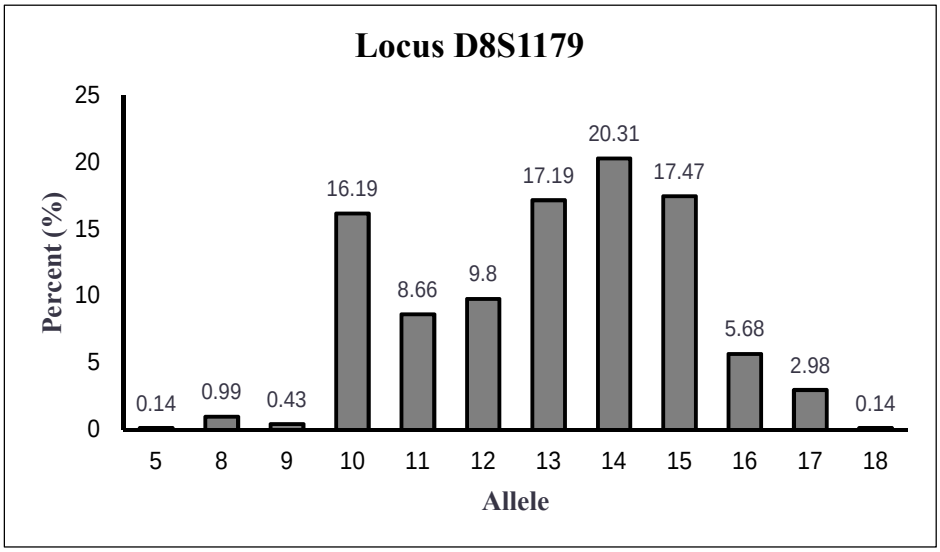


Figure 10: Percentage of allele frequency for the locus D8S1179 in the Indian population

Locus D21S11 (Indian population)

Table 9: Allele frequency for the locus D21S11 in the Indian population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	27	9.9968	1.42
2	28	88	12.5
3	28.3	0.9856	0.14
4	29	148.9664	21.16
5	29.2	4.0128	0.57
6	29.3	1.9712	0.28
7	30	120.032	17.05
8	30.2	16.9664	2.41
9	31	37.0304	5.26
10	31.2	94.9696	13.49
11	31.3	0.9856	0.14
12	32	6.9696	0.99
13	32.2	115.0336	16.34
14	33	1.9712	0.28
15	33.2	49.984	7.1
16	34.1	0.9856	0.14
17	34.2	3.0272	0.43
18	35.2	1.9712	0.28

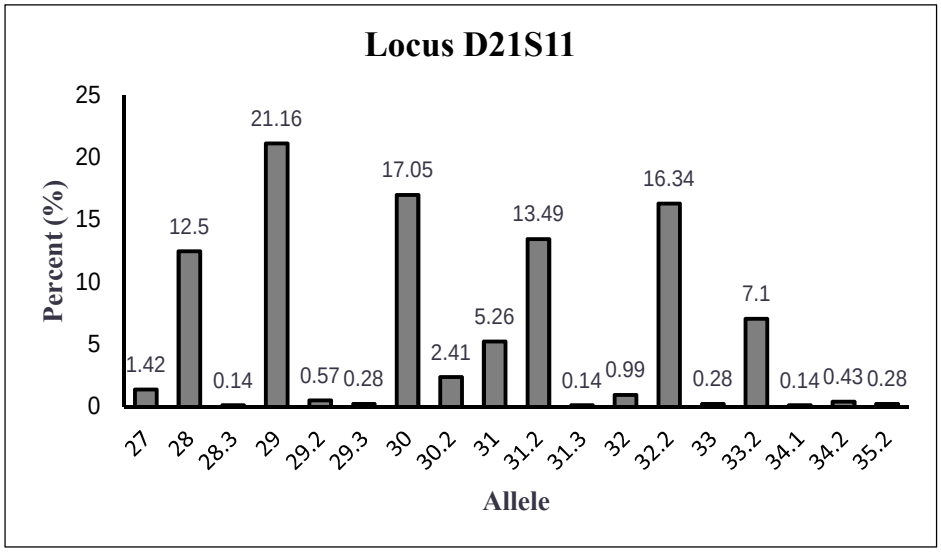


Figure 11: Percentage of allele frequency for the locus D21S11 in the Indian population

Locus D18S51 (Indian population)

Table 10: Allele frequency for the locus D18S51 in the Indian population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	10	9.9968	1.42
2	11	8.0256	1.14
3	12	55.968	7.95
4	13	87.0144	12.36
5	14	177.9712	25.28
6	15	117.9904	16.76
7	15.2	0.9856	0.14
8	16	97.0112	13.78
9	16.2	1.9712	0.28
10	17	55.968	7.95
11	18	32.032	4.55
12	19	23.0208	3.27
13	20	18.0224	2.56
14	21	9.0112	1.28
15	22	4.9984	0.71
16	23	3.0272	0.43
17	24	0.9856	0.14

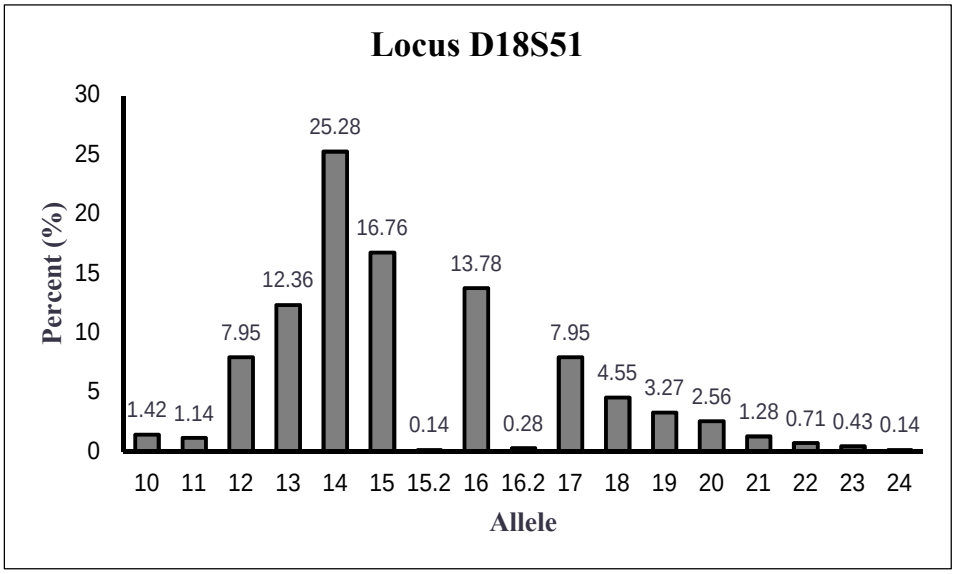


Figure 12: Percentage of allele frequency for the locus D18S51 in the Indian population

Locus Penta E (Indian population)

Table 11: Allele frequency for the locus Penta E in the Indian population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	5	40.9728	5.82
2	7	53.9968	7.67
3	8	9.9968	1.42
4	9	6.9696	0.99
5	10	48.9984	6.96
6	11	96.0256	13.64
7	12	68.992	9.8
8	12.4	0.9856	0.14
9	13	58.9952	8.38
10	14	37.0304	5.26
11	15	58.9952	8.38
12	16	68.0064	9.66
13	17	52.0256	7.39
14	17.4	0.9856	0.14
15	18	38.016	5.4
16	18.5	0.9856	0.14
17	19	32.032	4.55
18	20	15.9808	2.27
19	21	9.0112	1.28
20	22	3.0272	0.43
21	23	1.9712	0.28

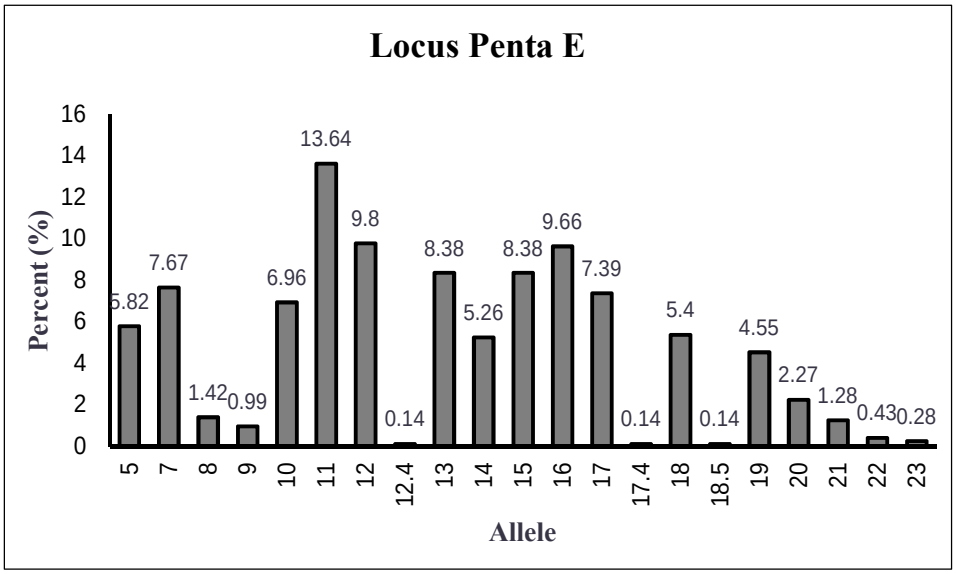


Figure 13: Percentage of allele frequency for the locus Penta E in the Indian population

Locus D2S441 (Indian population)

Table 12: Allele frequency for the locus D2S441 in the Indian population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	9	0.9856	0.14
2	10	223.0272	31.68
3	11	285.9648	40.62
4	11.3	30.976	4.4
5	12	45.9712	6.53
6	12.3	3.0272	0.43
7	13	11.968	1.7
8	14	91.0272	12.93
9	15	9.0112	1.28
10	16	1.9712	0.28

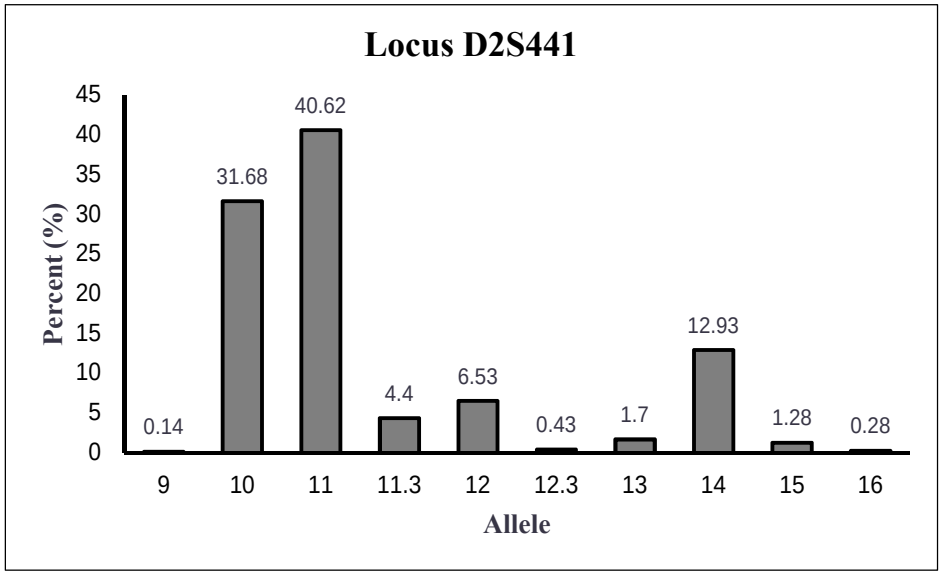


Figure 14: Percentage of allele frequency for the locus D2S441 in the Indian population

Locus D19S433 (Indian population)

Table 13: Allele frequency for the locus D19S433 in the Indian population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	11	5.984	0.85
2	11.2	0.9856	0.14
3	12	53.0112	7.53
4	12.2	5.984	0.85
5	13	218.0288	30.97
6	13.2	24.0064	3.41
7	14	147.9808	21.02
8	14.2	54.9824	7.81
9	15	76.032	10.8
10	15.2	60.9664	8.66
11	16	33.0176	4.69
12	16.2	14.9952	2.13
13	17	4.9984	0.71
14	17.2	1.9712	0.28
15	18	0.9856	0.14

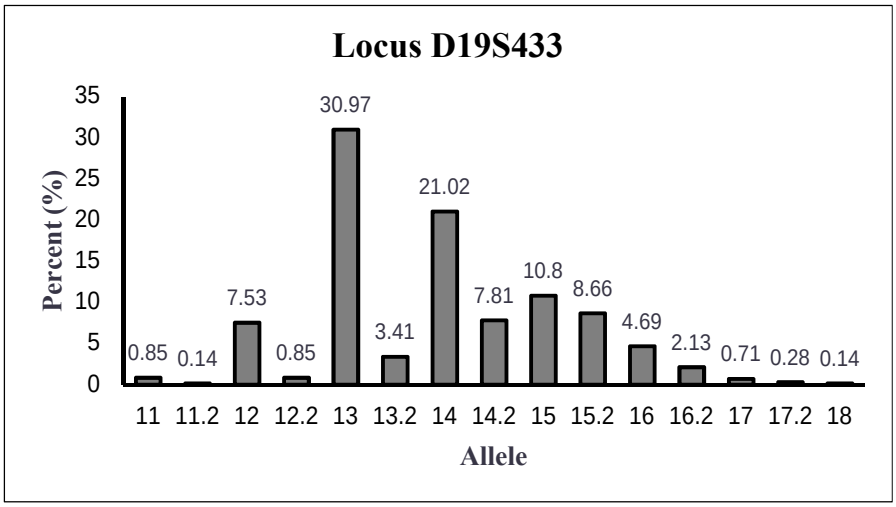


Figure 15: Percentage of allele frequency for the locus D19S433 in the Indian population

Locus TH01 (Indian population)

Table 14: Allele frequency for the locus TH01 in the Indian population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	5.3	0.9856	0.14
2	6	151.008	21.45
3	7	132	18.75
4	8	94.9696	13.49
5	8.3	1.9712	0.28
6	9	208.032	29.55
7	9.3	108.9792	15.48
8	10	4.9984	0.71
9	10.3	0.9856	0.14

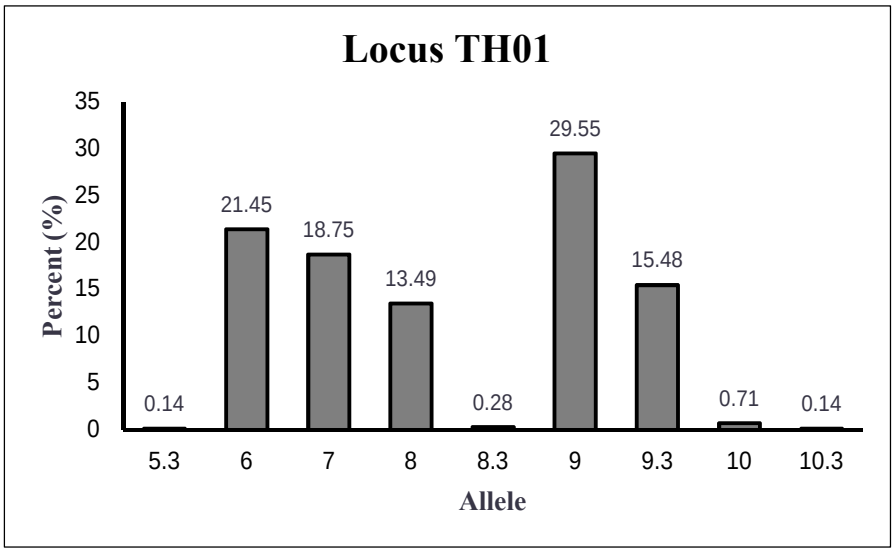


Figure 16: Percentage of allele frequency for the locus TH01 in the Indian population

Locus FGA (Indian population)

Table 15: Allele frequency for the locus FGA in the Indian population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	18	4.0128	0.57
2	19	34.9888	4.97
3	20	64.9792	9.23
4	20.2	1.9712	0.28
5	21	86.0288	12.22
6	21.2	4.0128	0.57
7	22	107.9936	15.34
8	22.2	6.9696	0.99
9	22.3	0.9856	0.14
10	23	137.984	19.6
11	23.2	4.9984	0.71
12	24	118.976	16.9
13	24.2	4.0128	0.57
14	25	74.976	10.65
15	25.2	1.9712	0.28
16	26	35.9744	5.11
17	26.2	0.9856	0.14
18	27	4.9984	0.71
19	28	4.0128	0.57
20	29	3.0272	0.43

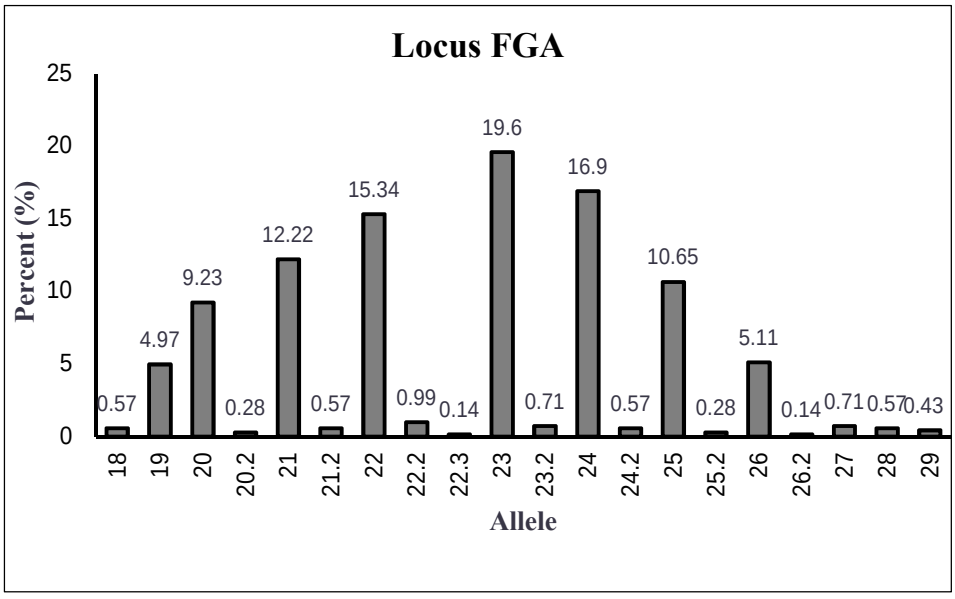


Figure 17: Percentage of allele frequency for the locus FGA in the Indian population

Locus D22S1045 (Indian population)

Table 16: Allele frequency for the locus D22S1045 in the Indian population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	9	0.9856	0.14
2	10	1.9712	0.28
3	11	199.0208	28.27
4	12	3.0272	0.43
5	13	1.9712	0.28
6	14	43.0144	6.11
7	15	270.9696	38.49
8	16	126.016	17.9
9	17	50.9696	7.24
10	18	4.9984	0.71
11	19	0.9856	0.14

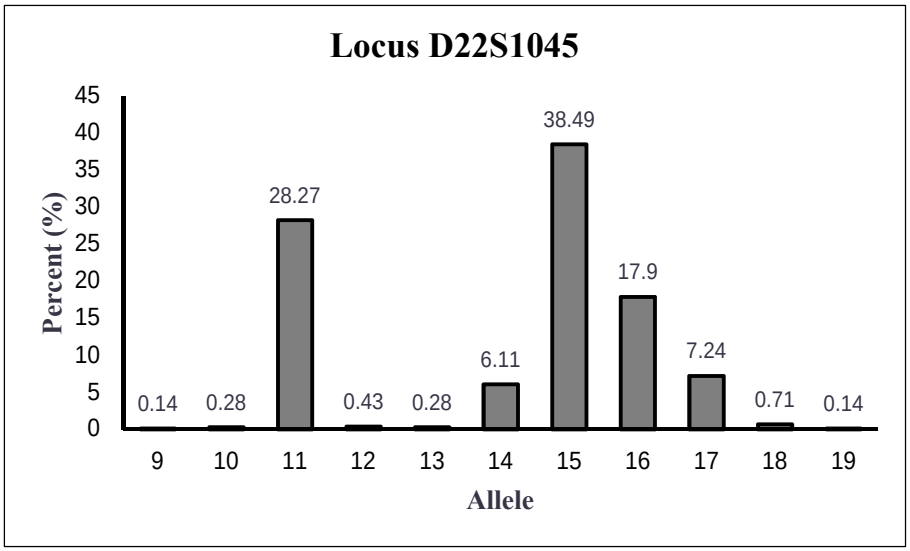


Figure 18: Percentage of allele frequency for the locus D22S1045 in the Indian population

Locus D5S818 (Indian population)

Table 17: Allele frequency for the locus D5S818 in the Indian population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	8	3.0272	0.43
2	9	29.9904	4.26
3	10	74.976	10.65
4	11	240.9792	34.23
5	12	214.016	30.4
6	13	130.0288	18.47
7	14	10.9824	1.56

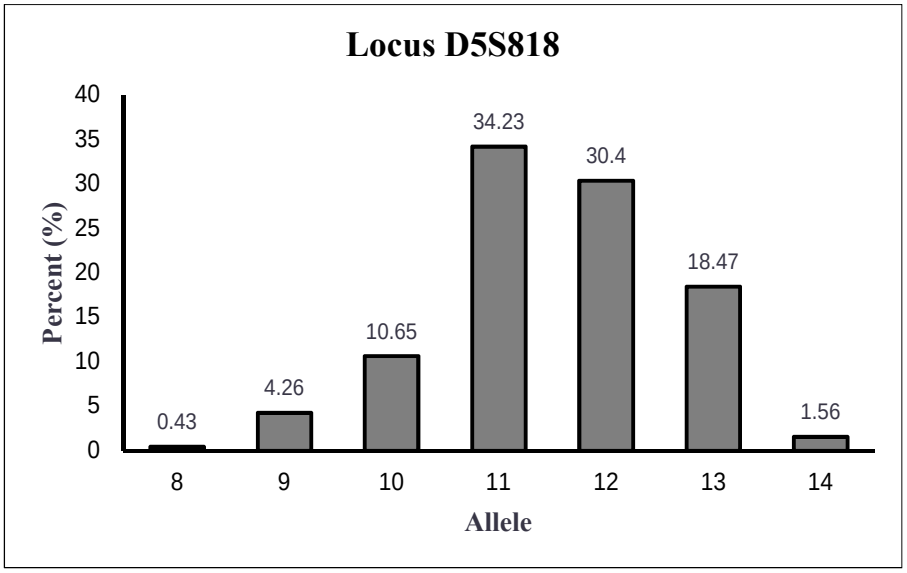


Figure 19: Percentage of allele frequency for the locus D5S818 in the Indian population

Locus D13S317 (Indian population)

Table 18: Allele frequency for the locus D13S317 in the Indian population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	6	0.9856	0.14
2	7	8.0256	1.14
3	8	132.9856	18.89
4	9	69.9776	9.94
5	10	60.9664	8.66
6	11	164.032	23.3
7	12	190.0096	26.99
8	13	58.9952	8.38
9	14	16.9664	2.41
10	15	0.9856	0.14

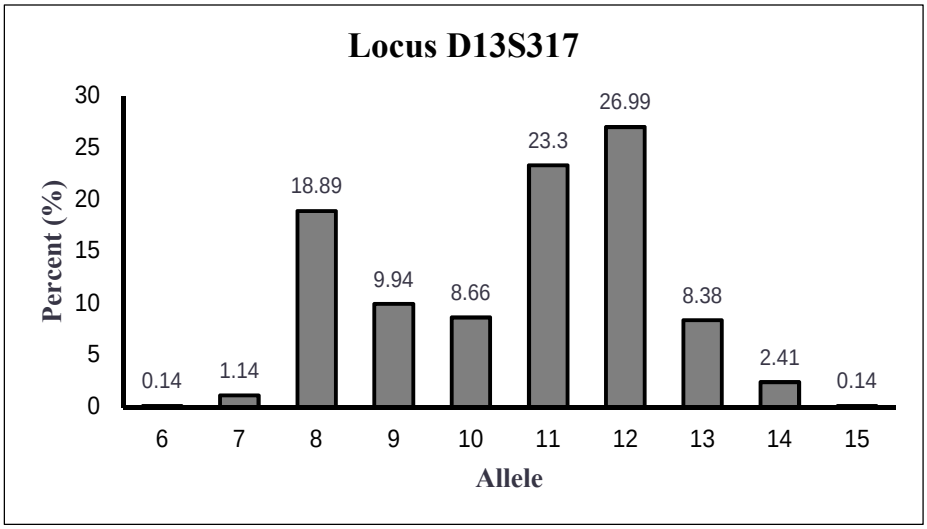


Figure 20: Percentage of allele frequency for the locus D13S317 in the Indian population

Locus D7S820 (Indian population)

Table 19: Allele frequency for the locus D7S820 in the Indian population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	7	21.9648	3.12
2	7.1	0.9856	0.14
3	8	155.0208	22.02
4	9	62.0224	8.81
5	10	150.0224	21.31
6	11	172.9728	24.57
7	11.3	0.9856	0.14
8	12	122.0032	17.33
9	13	15.9808	2.27
10	14	1.9712	0.28

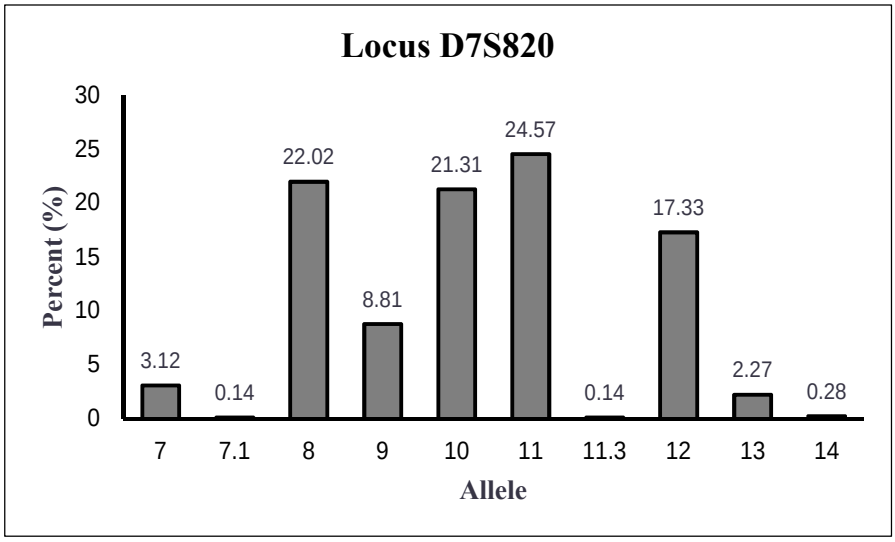


Figure 21: Percentage of allele frequency for the locus D7S820 in the Indian population

Locus D6S1043 (Indian population)

Table 20: Allele frequency for the locus D6S1043 in the Indian population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	10	15.9808	2.27
2	11	196.9792	27.98
3	12	151.9936	21.59
4	13	73.0048	10.37
5	14	45.9712	6.53
6	15	5.984	0.85
7	16	4.0128	0.57
8	17	34.0032	4.83
9	18	77.0176	10.94
10	19	68.992	9.8
11	20	24.0064	3.41
12	21	4.9984	0.71
13	23	0.9856	0.14

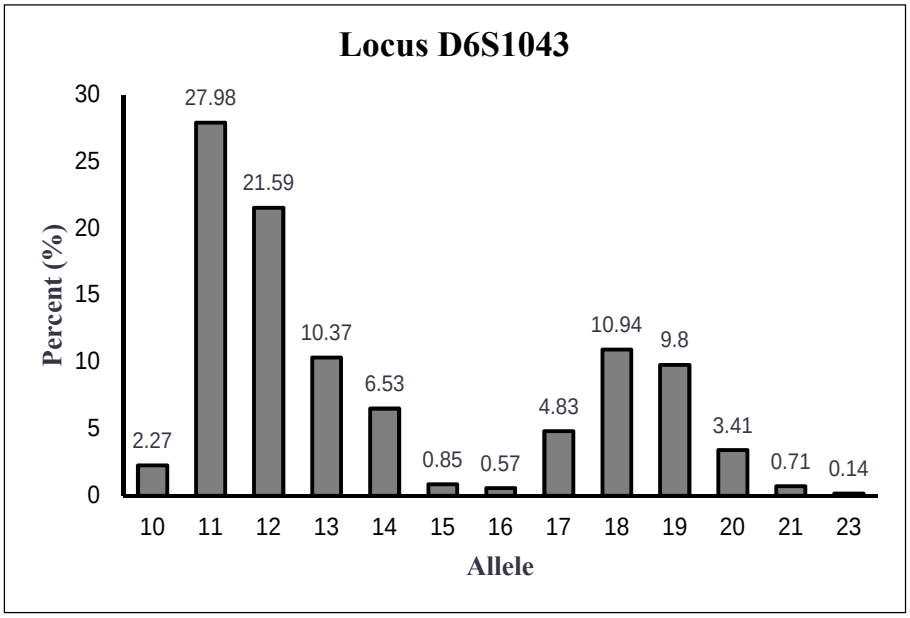


Figure 22: Percentage of allele frequency for the locus D6S1043 in the Indian population

Locus D10S1248 (Indian population)

Table 21: Allele frequency for the locus D10S1248 in the Indian population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	11	6.9696	0.99
2	12	14.0096	1.99
3	13	117.0048	16.62
4	14	182.9696	25.99
5	15	204.0192	28.98
6	16	131.0144	18.61
7	17	39.9872	5.68
8	18	6.9696	0.99
9	19	0.9856	0.14

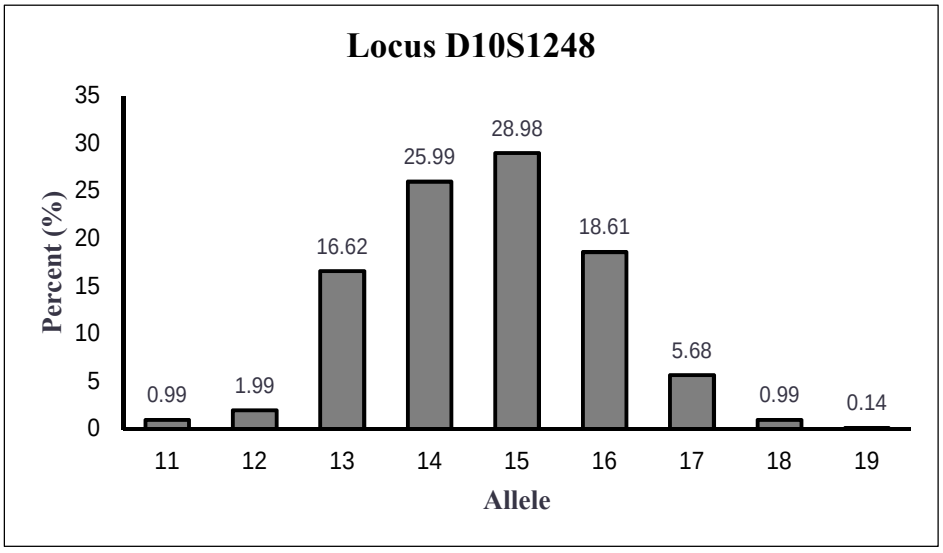


Figure 23: Percentage of allele frequency for the locus D10S1248 in the Indian population

Locus D1S1656 (Indian population)

Table 22: Allele frequency for the locus D1S1656 in the Indian population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	6	0.9856	0.14
2	8	25.9776	3.69
3	9	4.0128	0.57
4	10	9.9968	1.42
5	11	99.968	14.2
6	12	63.008	8.95
7	13	93.984	13.35
8	14	69.9776	9.94
9	14.3	1.9712	0.28
10	15	123.9744	17.61
11	15.3	10.9824	1.56
12	16	101.024	14.35
13	16.3	25.9776	3.69
14	17	30.976	4.4
15	17.3	10.9824	1.56
16	18	9.0112	1.28
17	18.3	14.0096	1.99
18	19.3	6.9696	0.99

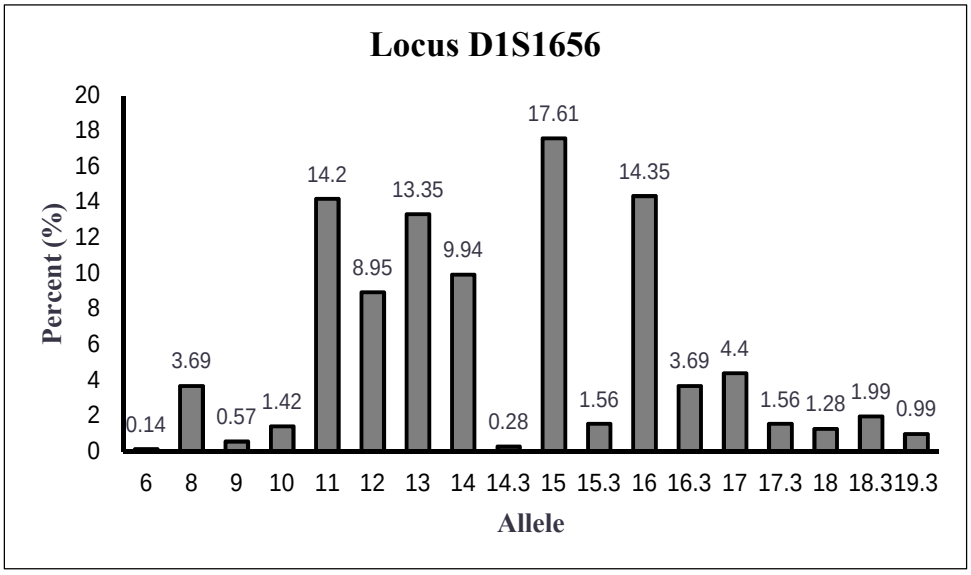


Figure 24: Percentage of allele frequency for the locus D1S1656 in the Indian population

Locus D12S391 (Indian population)

Table 23: Allele frequency for the locus D12S391 in the Indian population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	14	0.9856	0.14
2	15	1.9712	0.28
3	16	19.008	2.7
4	17	96.0256	13.64
5	17.3	11.968	1.7
6	18	164.032	23.3
7	18.3	8.0256	1.14
8	19	115.0336	16.34
9	19.3	4.0128	0.57
10	20	82.016	11.65
11	20.3	0.9856	0.14
12	21	67.0208	9.52
13	22	64.9792	9.23
14	23	39.9872	5.68
15	24	18.0224	2.56
16	25	8.0256	1.14
17	26	0.9856	0.14
18	27	0.9856	0.14

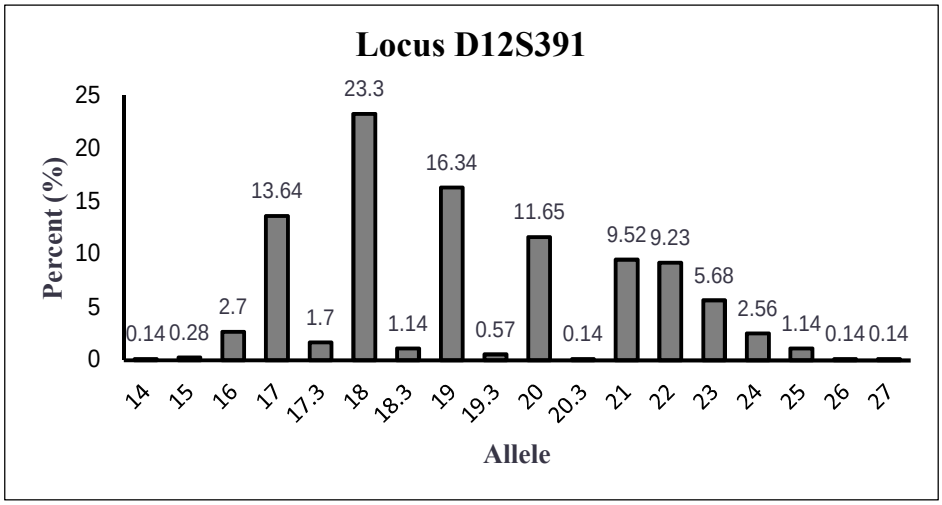


Figure 25: Percentage of allele frequency for the locus D12S391 in the Indian population

Locus D2S1338 (Indian population)

Table 24: Allele frequency for the locus D2S1338 in the Indian population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	13	0.9856	0.14
2	15	0.9856	0.14
3	16	5.984	0.85
4	17	49.984	7.1
5	18	104.9664	14.91
6	19	112.992	16.05
7	20	81.0304	11.51
8	21	30.976	4.4
9	22	45.9712	6.53
10	23	121.0176	17.19
11	24	71.0336	10.09
12	25	64.9792	9.23
13	26	8.0256	1.14
14	27	4.9984	0.71

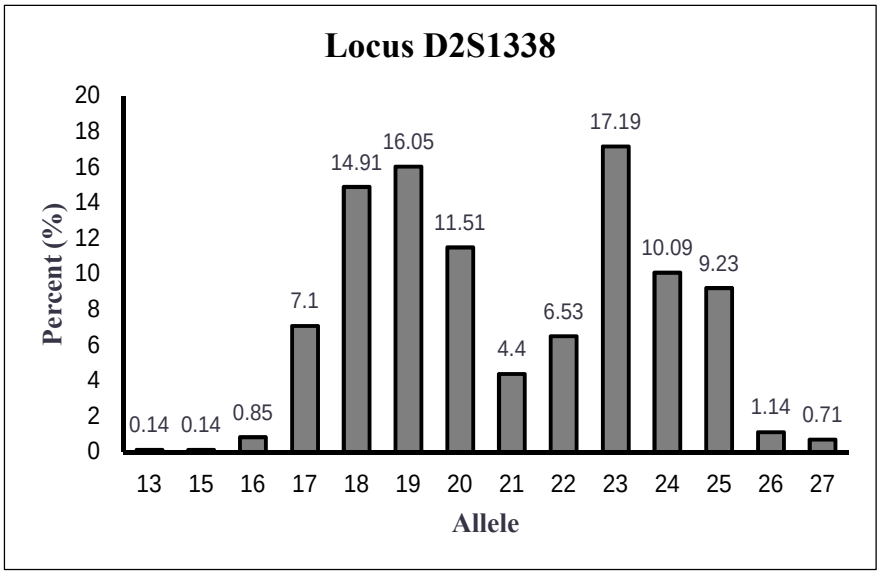


Figure 26: Percentage of allele frequency for the locus D2S1338 in the Indian population

Locus Penta D (Indian population)

Table 25: Allele frequency for the locus Penta D in the Indian population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	2.2	4.0128	0.57
2	7	5.984	0.85
3	8	14.0096	1.99
4	9	141.0112	20.03
5	9.4	0.9856	0.14
6	10	141.0112	20.03
7	11	180.9984	25.71
8	12	88	12.5
9	13	79.9744	11.36
10	14	23.0208	3.27
11	15	20.9792	2.98
12	16	3.0272	0.43
13	17	0.9856	0.14

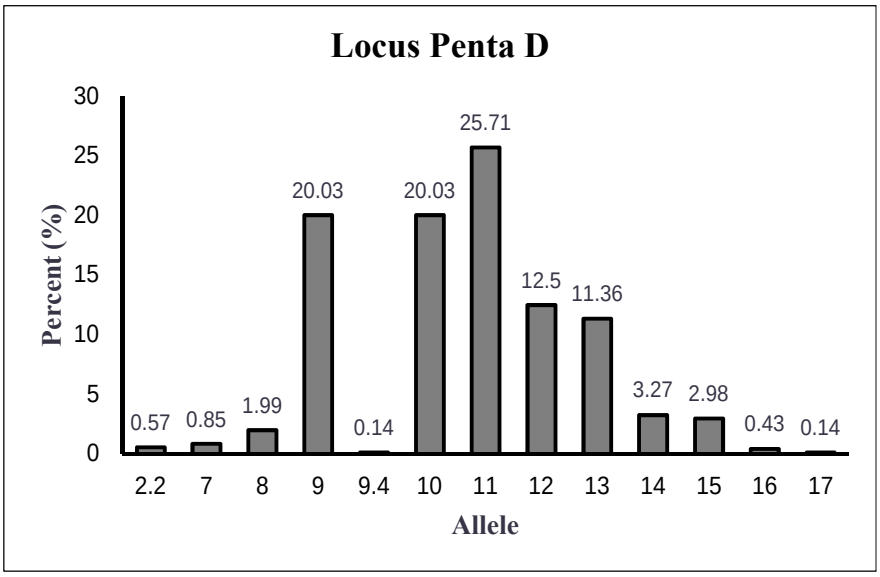


Figure 27: Percentage of allele frequency for the locus Penta D in the Indian population

Locus D3S1358 (Pakistani population)

Table 26: Allele frequency for the locus D3S1358 in the Pakistani population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	13	2.792	0.4
2	14	32.108	4.6
3	15	189.86	27.2
4	16	210.1	30.1
5	17	170.31	24.4
6	17.1	0.698	0.1
7	18	80.968	11.6
8	19	11.168	1.6

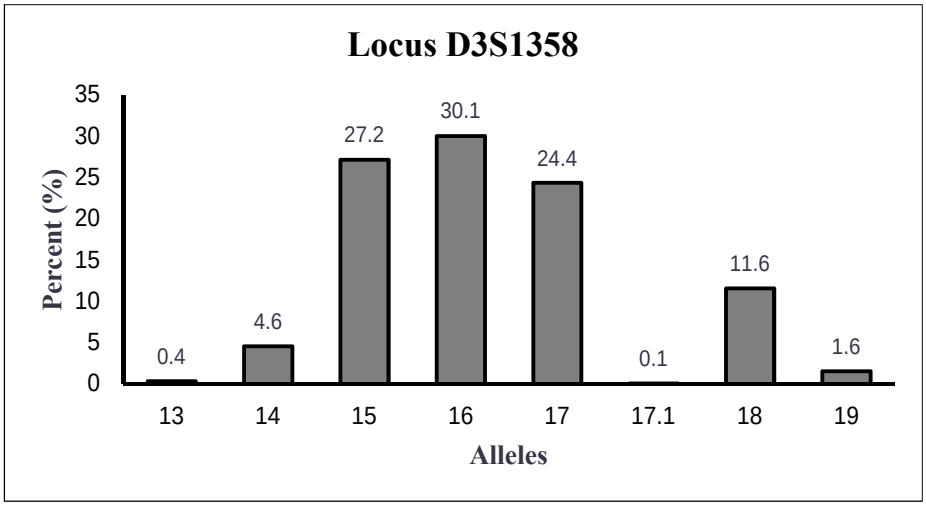


Figure 28: Percentage of allele frequency for the locus D3S1358 in the Pakistani population

Locus vWA (Pakistani population)

Table 27: Allele frequency for the locus vWA in the Pakistani population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	13	2.792	0.4
2	14	83.76	12
3	15	62.82	9
4	15.2	0.698	0.1
5	16	164.73	23.6
6	17	168.92	24.2
7	18	152.86	21.9
8	19	53.048	7.6
9	20	6.282	0.9
10	21	0.698	0.1

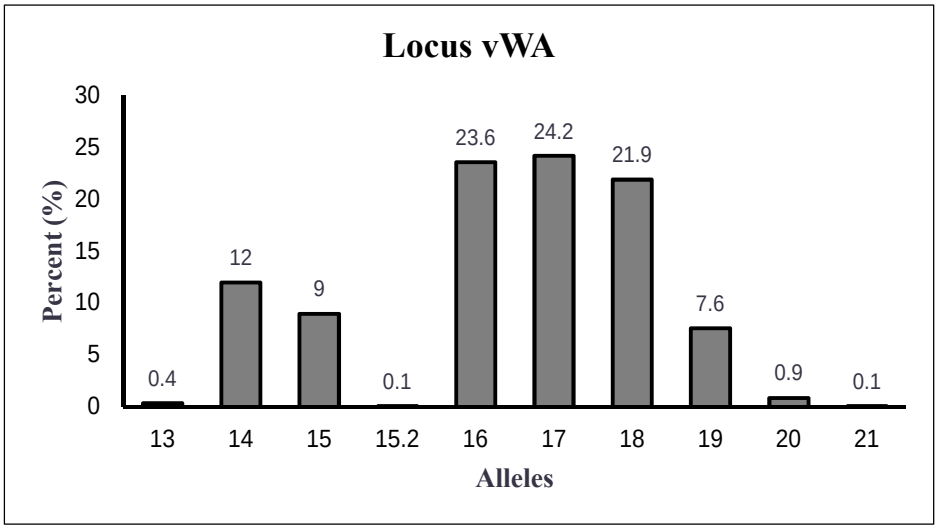


Figure 29: Percentage of allele frequency for the locus vWA in the Pakistani population

Locus D16S539 (Pakistani population)

Table 28: Allele frequency for the locus D16S539 in the Pakistani population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	8	43.276	6.2
2	9	115.17	16.5
3	10	85.854	12.3
4	11	228.94	32.8
5	12	145.88	20.9
6	13	71.196	10.2
7	14	7.678	1.1

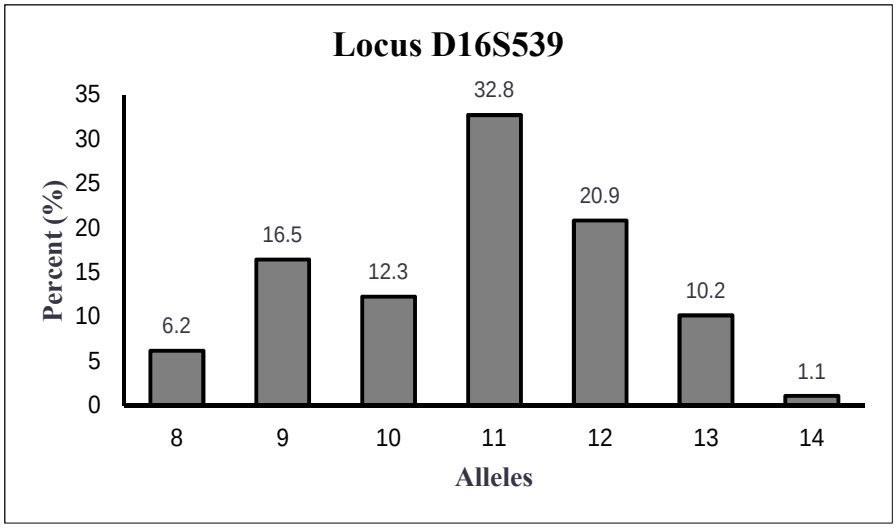


Figure 30: Percentage of allele frequency for the locus D16S539 in the Pakistani population

Locus CSF1PO (Pakistani population)

Table 29: Allele frequency for the locus CSF1PO in the Pakistani population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	7	2.094	0.3
2	8	0.698	0.1
3	9	22.336	3.2
4	10	155.65	22.3
5	11	196.84	28.2
6	12	258.96	37.1
7	13	53.746	7.7
8	14	6.98	1

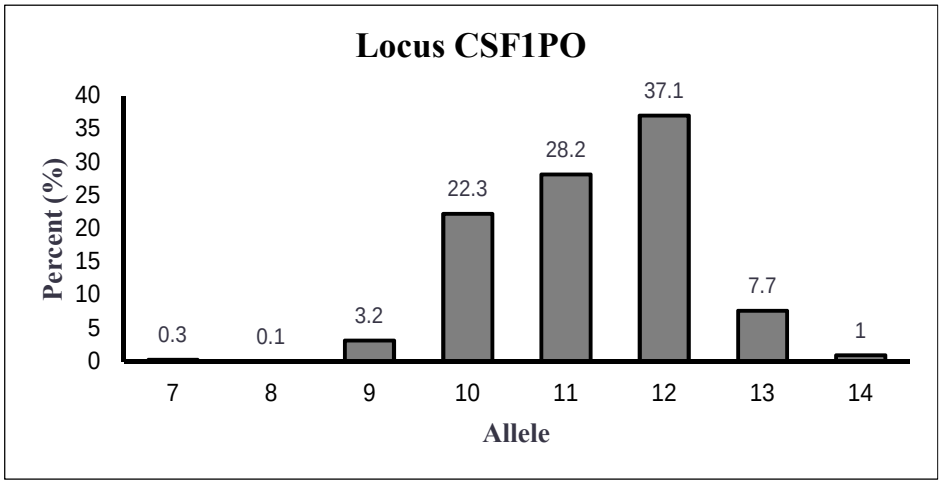


Figure 31: Percentage of allele frequency for the locus CSF1PO in the Pakistani population

Locus TPOX (Pakistani population)

Table 30: Allele frequency for the locus TPOX in the Pakistani population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	7	0.698	0.1
2	8	261.75	37.5
3	9	108.19	15.5
4	10	50.954	7.3
5	11	252.68	36.2
6	12	20.94	3
7	13	2.094	0.3

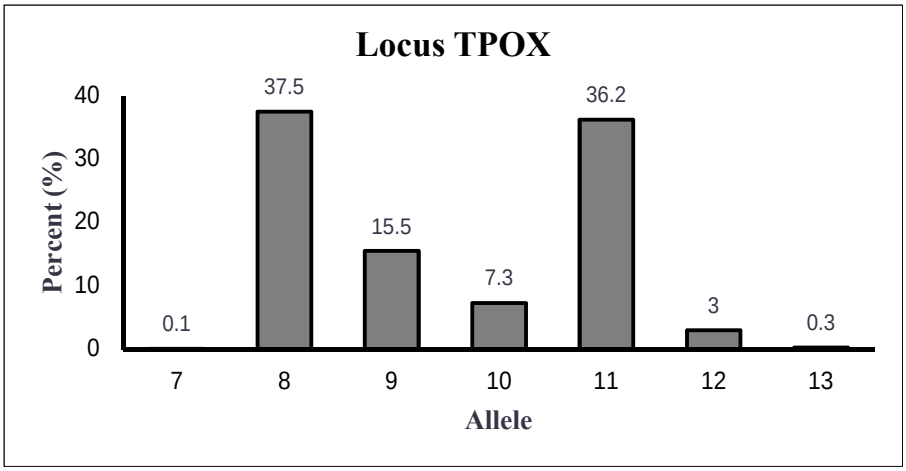


Figure 32: Percentage of allele frequency for the locus TPOX in the Pakistani population

Locus D8S1179 (Pakistani population)

Table 31: Allele frequency for the locus D8S1179 in the Pakistani population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	8	6.98	1
2	9	4.188	0.6
3	10	103.3	14.8
4	11	51.652	7.4
5	12	67.706	9.7
6	13	120.75	17.3
7	14	148.67	21.3
8	15	118.66	17
9	16	60.726	8.7
10	17	11.866	1.7
11	18	2.094	0.3

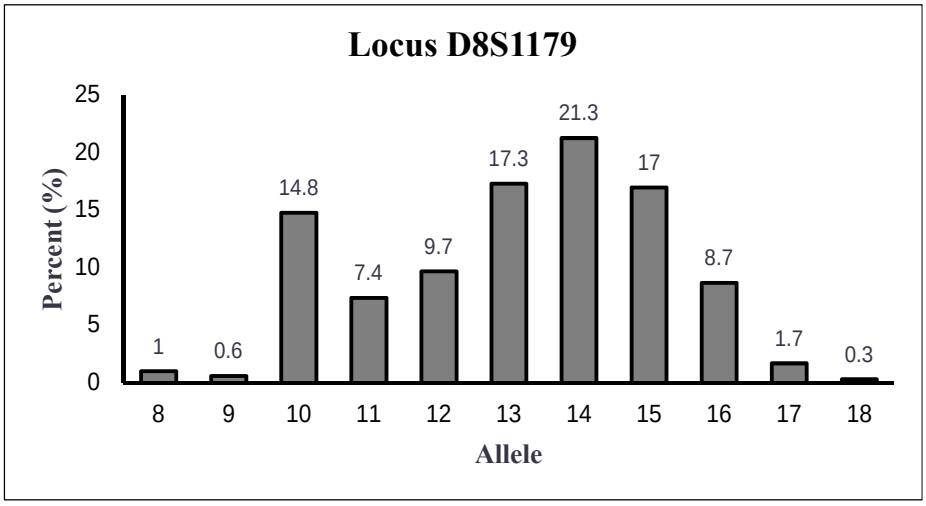


Figure 33: Percentage of allele frequency for the locus D8S1179 in the Pakistani population

Locus D21S11 (Pakistani population)

Table 32: Allele frequency for the locus D21S11 in the Pakistani population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	26	0.698	0.1
2	27	11.866	1.7
3	28	83.062	11.9
4	28.2	0.698	0.1
5	29	148.67	21.3
6	29.2	2.792	0.4
7	29.3	0.698	0.1
8	30	138.9	19.9
9	30.2	18.846	2.7
10	30.3	0.698	0.1
11	31	32.806	4.7
12	31.2	99.116	14.2
13	32	4.188	0.6
14	32.2	94.23	13.5
15	33	2.094	0.3
16	33.2	48.86	7
17	34	0.698	0.1
18	34.2	4.886	0.7
19	35	0.698	0.1
20	36	0.698	0.1

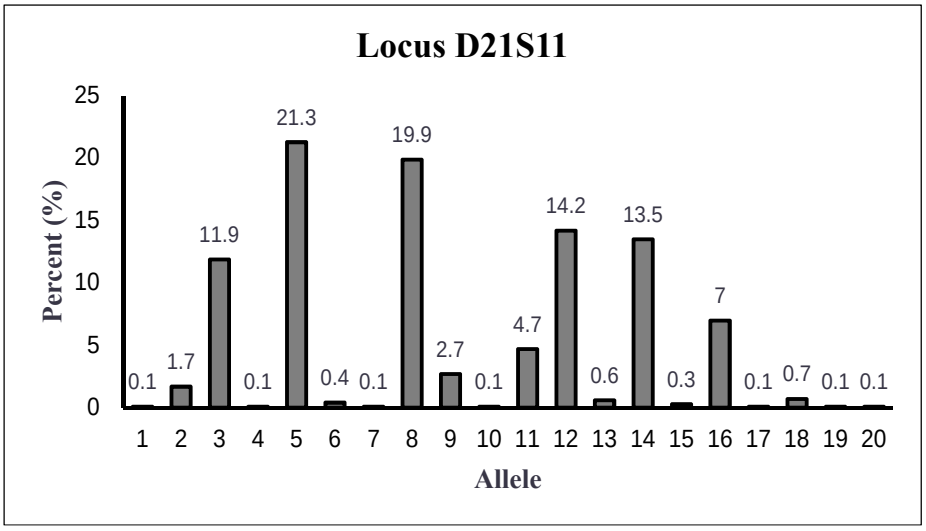


Figure 34: Percentage of allele frequency for the locus D21S11 in the Pakistani population

Locus D18S51 (Pakistani population)

Table 33: Allele frequency for the locus D18S51 in the Pakistani population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	10	2.792	0.4
2	11	13.262	1.9
3	11.2	0.698	0.1
4	12	59.33	8.5
5	13	73.988	10.6
6	13.2	0.698	0.1
7	14	182.18	26.1
8	15	117.26	16.8
9	16	97.022	13.9
10	16.2	0.698	0.1
11	17	60.028	8.6
12	18	32.806	4.7
13	19	25.826	3.7
14	20	18.846	2.7
15	21	4.886	0.7
16	22	4.886	0.7
17	23	0.698	0.1
18	25	0.698	0.1

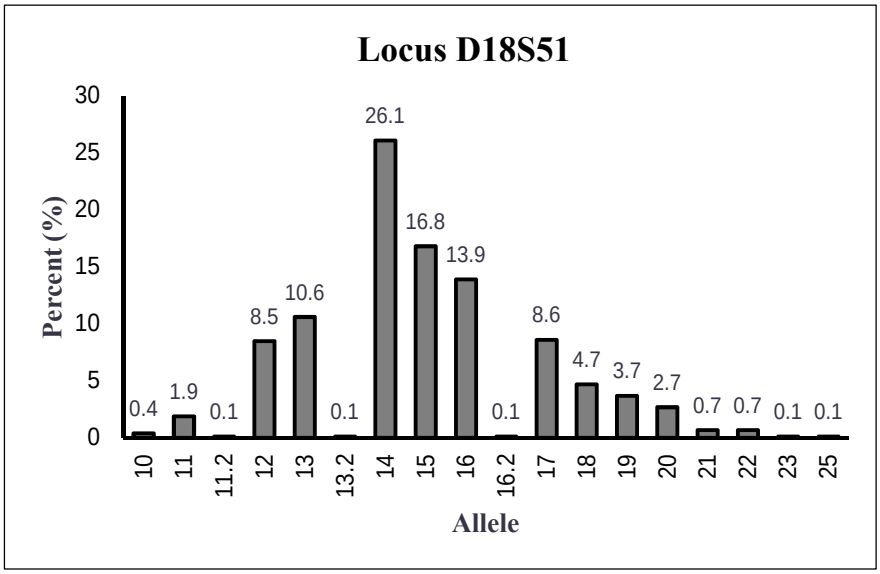


Figure 35: Percentage of allele frequency for the locus D18S51 in the Pakistani population

Locus Penta E (Pakistani population)

Table 34: Allele frequency for the locus Penta E in the Pakistani population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	5	48.86	7
2	7	50.954	7.3
3	8	13.262	1.9
4	9	7.678	1.1
5	10	44.672	6.4
6	11	90.74	13
7	11.4	0.698	0.1
8	12	78.176	11.2
9	12.4	2.094	0.3
10	13	50.256	7.2
11	14	51.652	7.4
12	15	53.048	7.6
13	16	64.914	9.3
14	16.4	0.698	0.1
15	17	53.746	7.7
16	18	37.692	5.4
17	19	27.92	4
18	20	7.678	1.1
19	21	6.98	1
20	22	2.094	0.3
21	23	0.698	0.1
22	24	0.698	0.1

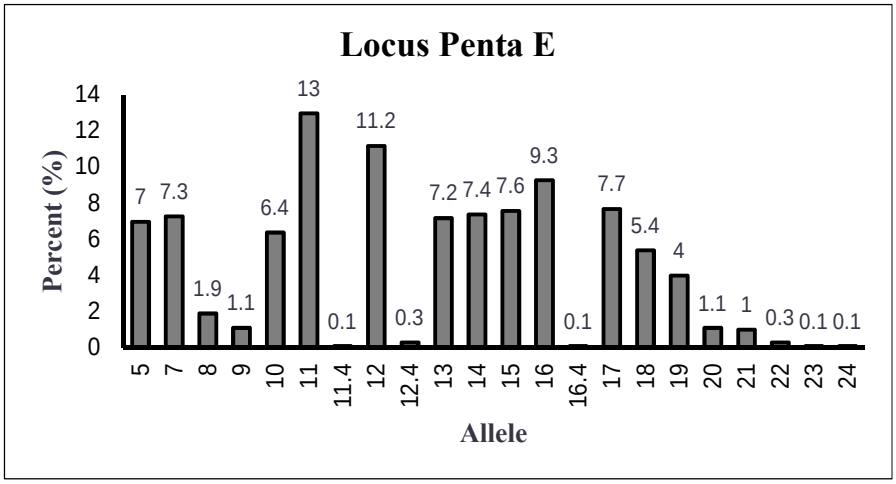


Figure 36: Percentage of allele frequency for the locus Penta E in the Pakistani population

Locus D2S441 (Pakistani population)

Table 35: Allele frequency for the locus D2S441 in the Pakistani population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	8	0.698	0.1
2	9	0.698	0.1
3	9.1	0.698	0.1
4	10	194.04	27.8
5	11	251.98	36.1
6	11.2	0.698	0.1
7	11.3	34.202	4.9
8	12	55.142	7.9
9	12.3	2.094	0.3
10	13	18.148	2.6
11	14	113.08	16.2
12	14.1	0.698	0.1
13	15	23.034	3.3
14	16	2.094	0.3

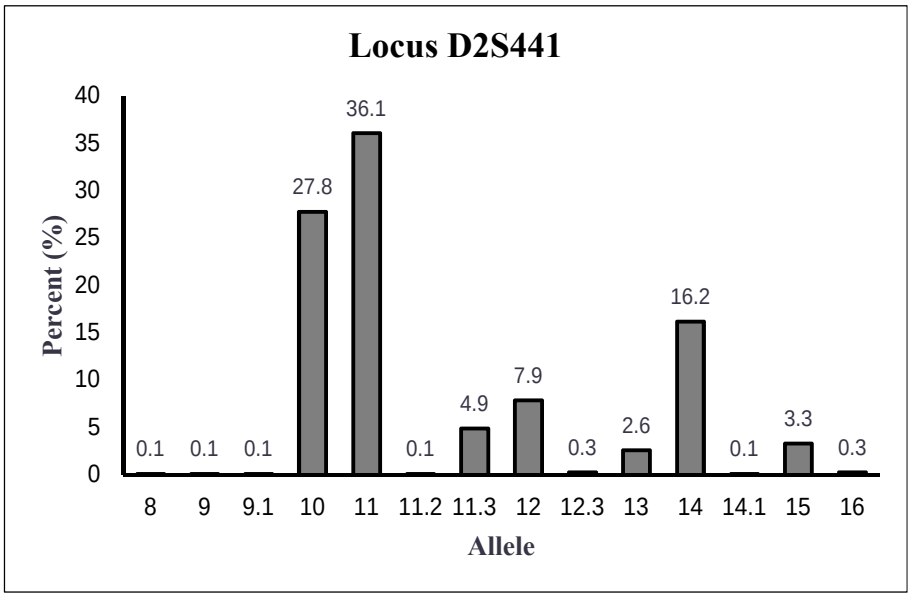


Figure 37: Percentage of allele frequency for the locus D2S441 in the Pakistani population

Locus D19S433 (Pakistani population)

Table 36: Allele frequency for the locus D19S433 in the Pakistani population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	6	0.698	0.1
2	8	2.094	0.3
3	9	2.094	0.3
4	11	4.188	0.6
5	11.2	0.698	0.1
6	12	41.182	5.9
7	12.2	4.886	0.7
8	13	224.06	32.1
9	13.2	18.148	2.6
10	14	171.01	24.5
11	14.2	46.766	6.7
12	15	66.310	9.5
13	15.2	64.914	9.3
14	16	32.108	4.6
15	16.2	13.96	2
16	17	4.188	0.6
17	17.2	0.698	0.1

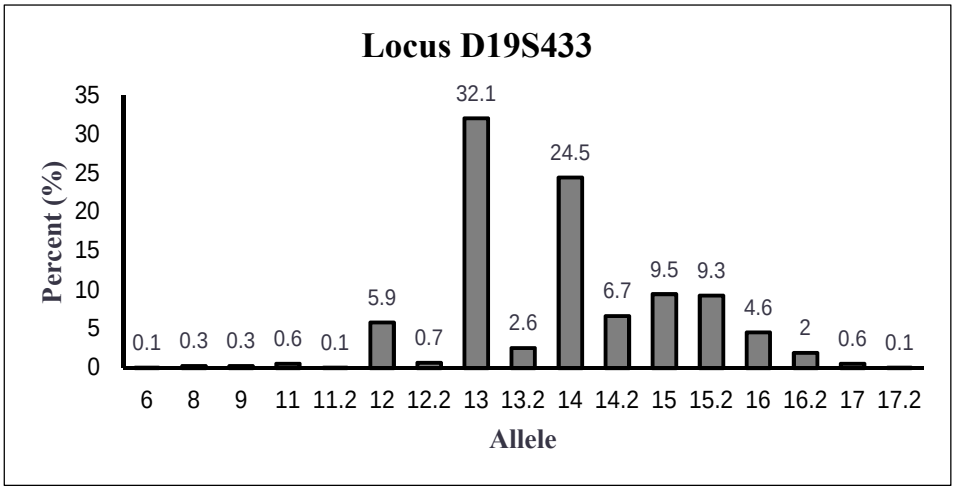


Figure 38: Percentage of allele frequency for the locus D19S433 in the Pakistani population

Locus TH01 (Pakistani population)

Table 37: Allele frequency for the locus TH01 in the Pakistani population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	6	171.01	24.5
2	7	124.24	17.8
3	8	111.68	16
4	9	171.01	24.5
5	9.3	94.928	13.6
6	10	14.658	2.1
7	10.3	2.094	0.3
8	20	0.698	0.1
9	24	0.698	0.1

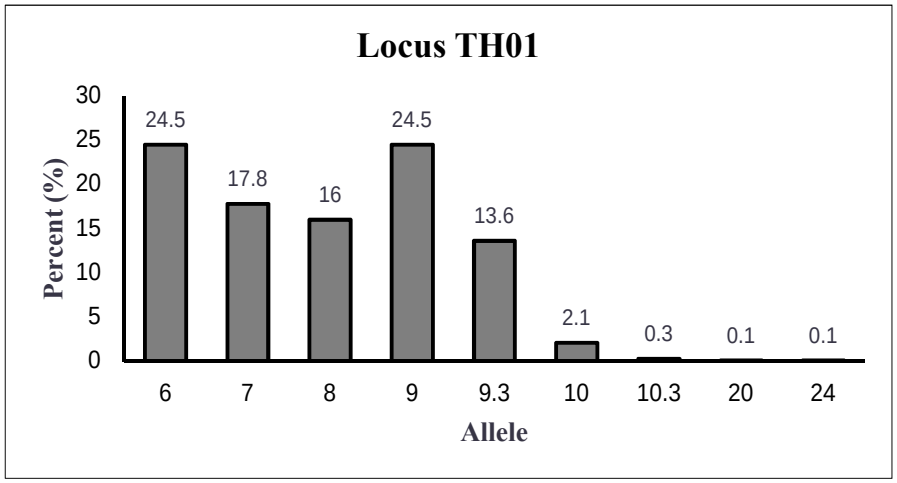


Figure 39: Percentage of allele frequency for the locus TH01 in the Pakistani population

Locus FGA (Pakistani Population)

Table 38: Allele frequency for the locus FGA in the Pakistani population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	11	0.698	0.1
2	17	2.094	0.3
3	18	9.772	1.4
4	18.2	0.698	0.1
5	19	43.276	6.2
6	20	73.988	10.6
7	20.2	0.698	0.1
8	21	94.23	13.5
9	21.2	7.678	1.1
10	22	101.21	14.5
11	22.2	2.792	0.4
12	23	127.73	18.3
13	23.2	2.094	0.3
14	24	113.77	16.3
15	24.2	2.792	0.4
16	24.3	0.698	0.1
17	25	71.196	10.2
18	25.2	2.094	0.3
19	25.3	0.698	0.1
20	26	25.128	3.6
21	27	9.772	1.4
22	28	0.698	0.1
23	29	0.698	0.1
24	30.2	0.698	0.1

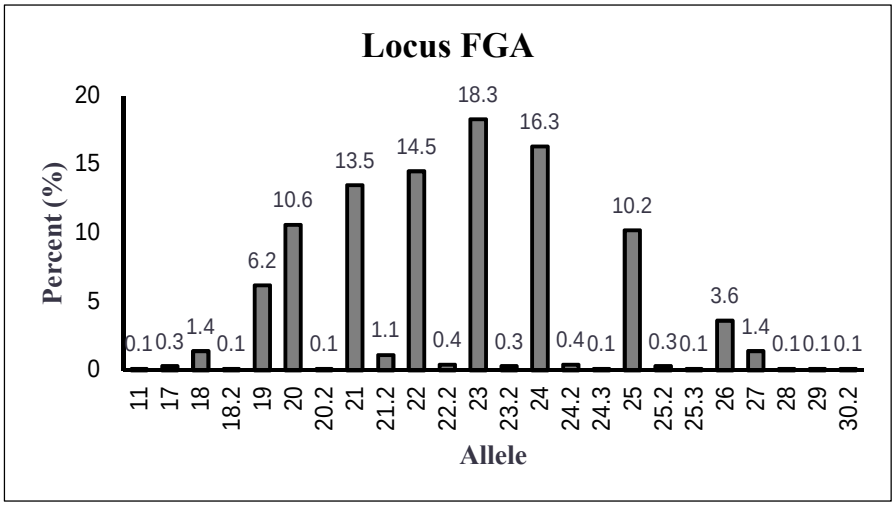


Figure 40: Percentage of allele frequency for the locus FGA in the Pakistani population

Locus D22S1045 (Pakistani population)

Table 39: Allele frequency for the locus D22S1045 in the Pakistani population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	10	0.698	0.1
2	11	182.18	26.1
3	12	9.074	1.3
4	13	2.792	0.4
5	14	39.088	5.6
6	15	252.68	36.2
7	16	143.09	20.5
8	17	60.726	8.7
9	18	6.98	1

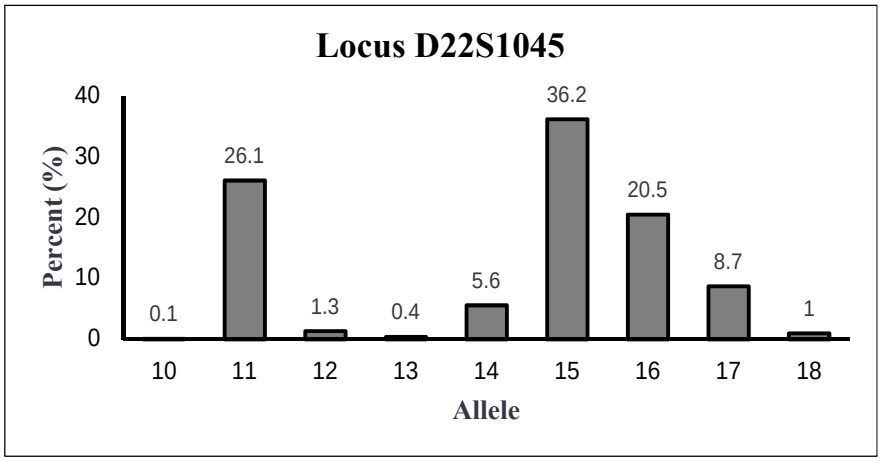


Figure 41: Percentage of allele frequency for the locus D22S1045 in the Pakistani population

Locus D5S818 (Pakistani population)

Table 40: Allele frequency for the locus D5S818 in the Pakistani population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	7	0.698	0.1
2	8	7.678	1.1
3	9	30.014	4.3
4	10	101.91	14.6
5	11	233.83	33.5
6	12	203.82	29.2
7	13	106.1	15.2
8	14	13.262	1.9

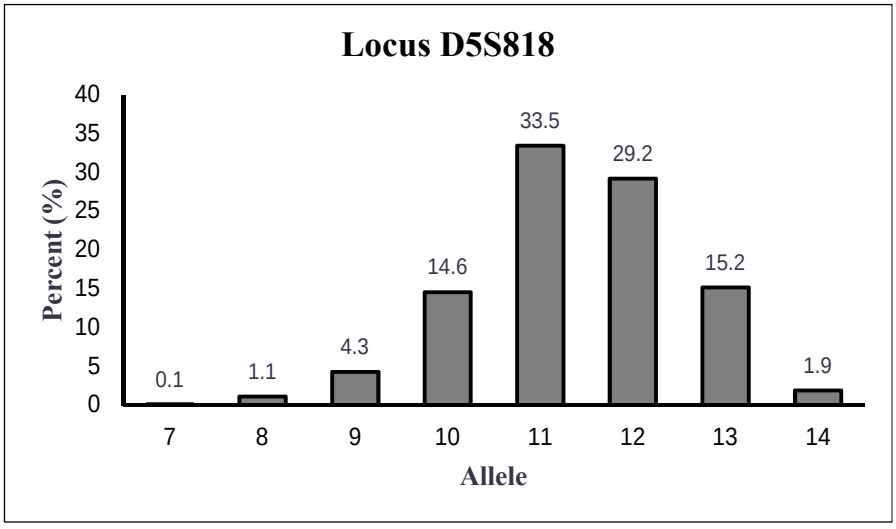


Figure 42: Percentage of allele frequency for the locus D5S818 in the Pakistani population

Locus D13S317 (Pakistani population)

Table 41: Allele frequency for the locus D13S317 in the Pakistani population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	4	0.698	0.1
2	7	4.188	0.6
3	8	129.83	18.6
4	9	62.82	9
5	10	51.652	7.4
6	11	173.8	24.9
7	12	189.16	27.1
8	13	64.216	9.2
9	14	20.94	3

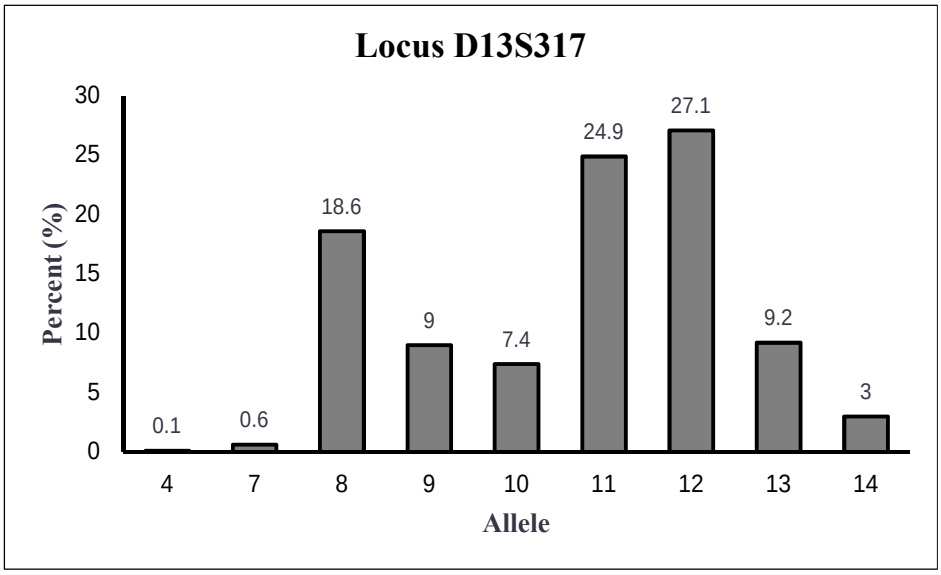


Figure 43: Percentage of allele frequency for the locus D13S317 in the Pakistani population

Locus D7S820 (Pakistani population)

Table 42: Allele frequency for the locus D7S820 in the Pakistani population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	7	20.242	2.9
2	8	138.9	19.9
3	9	46.068	6.6
4	10	171.01	24.5
5	11	170.31	24.4
6	12	124.24	17.8
7	13	23.034	3.3
8	14	2.792	0.4
9	18	0.698	0.1
10	26.2	0.698	0.1

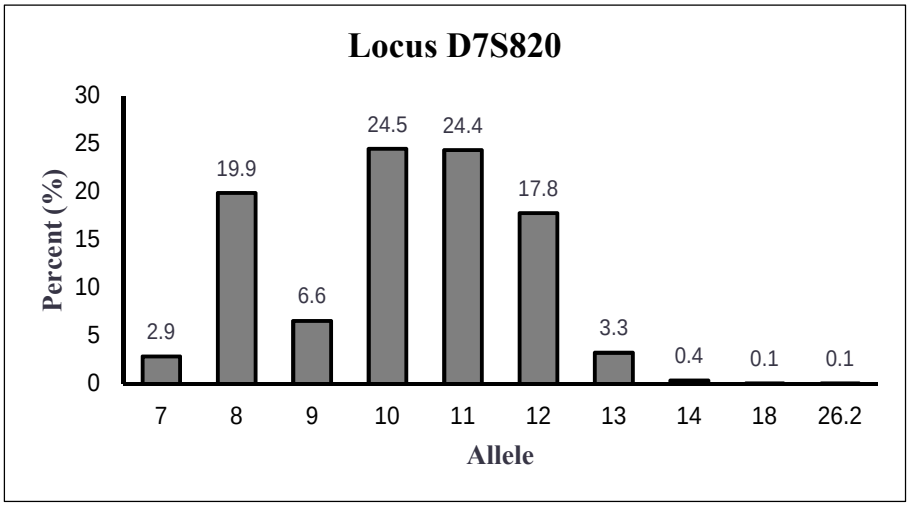


Figure 44: Percentage of allele frequency for the locus D7S820 in the Pakistani population

Locus D6S1043 (Pakistani population)

Table 43: Allele frequency for the locus D6S1043 in the Pakistani population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	9	2.094	0.3
2	10	20.94	3
3	11	212.19	30.4
4	12	166.82	23.9
5	13	53.746	7.7
6	14	51.652	7.4
7	15	6.98	1
8	16	4.188	0.6
9	17	25.826	3.7
10	18	69.102	9.9
11	19	44.672	6.4
12	20	32.806	4.7
13	21	6.282	0.9

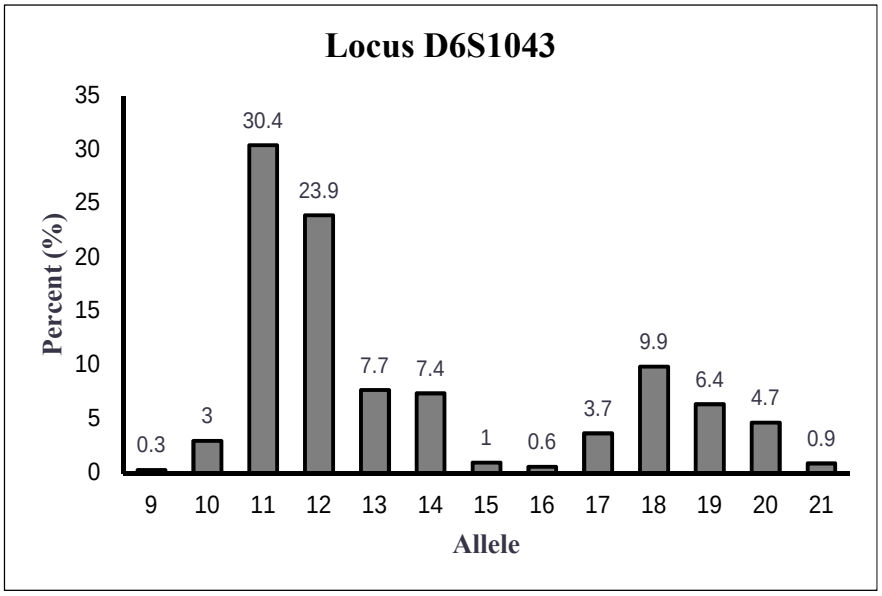


Figure 45: Percentage of allele frequency for the locus D6S1043 in the Pakistani population

Locus D10S1248 (Pakistani population)

Table 44: Allele frequency for the locus D10S1248 in the Pakistani population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	10	0.698	0.1
2	11	9.074	1.3
3	12	18.846	2.7
4	13	131.22	18.8
5	14	210.1	30.1
6	15	166.82	23.9
7	16	120.75	17.3
8	17	34.202	4.9
9	18	4.886	0.7
10	19	0.698	0.1

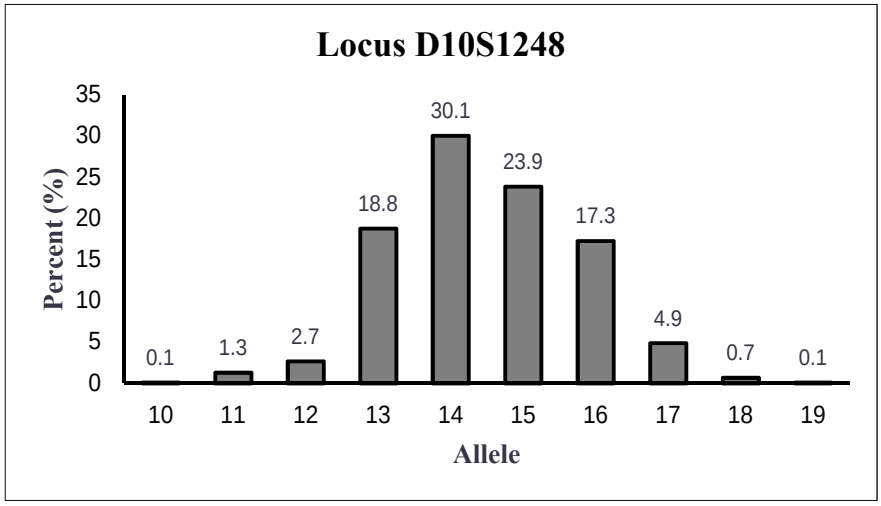


Figure 46: Percentage of allele frequency for the locus D10S1248 in the Pakistani population

Locus D1S1656 (Pakistani population)

Table 45: Allele frequency for the locus D1S1656 in the Pakistani population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	8	27.222	3.9
2	9	2.094	0.3
3	10	6.98	1
4	11	85.156	12.2
5	12	53.746	7.7
6	13	83.062	11.9
7	14	80.27	11.5
8	14.3	2.094	0.3
9	15	129.13	18.5
10	15.3	11.168	1.6
11	16	101.91	14.6
12	16.3	209.4	30
13	17	53.746	7.7
14	17.3	23.034	3.3
15	18	4.188	0.6
16	18.3	9.074	1.3
17	19.3	4.188	0.6
18	20.3	0.698	0.1

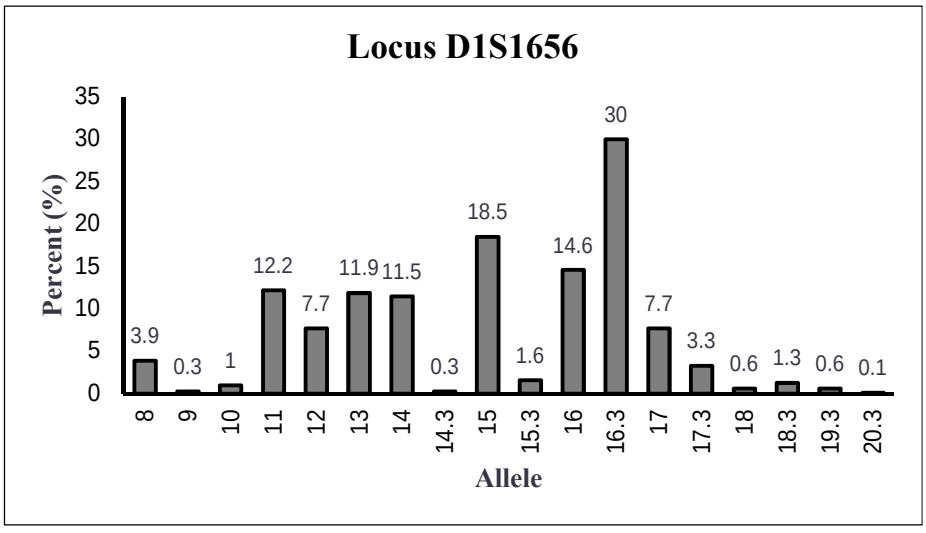


Figure 47: Percentage of allele frequency for the locus D1S1656 in the Pakistani population

Locus D12S391 (Pakistani population)

Table 46: Allele frequency for the locus D12S391 in the Pakistani population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	14	0.698	0.1
2	15	9.074	1.3
3	16	13.262	1.9
4	17	71.894	10.3
5	17.3	6.282	0.9
6	18	175.2	25.1
7	18.1	0.698	0.1
8	18.3	9.772	1.4
9	19	99.116	14.2
10	19.3	2.094	0.3
11	20	85.156	12.2
12	21	85.156	12.2
13	22	74.686	10.7
14	23	41.182	5.9
15	24	14.658	2.1
16	25	6.282	0.9
17	26	2.792	0.4

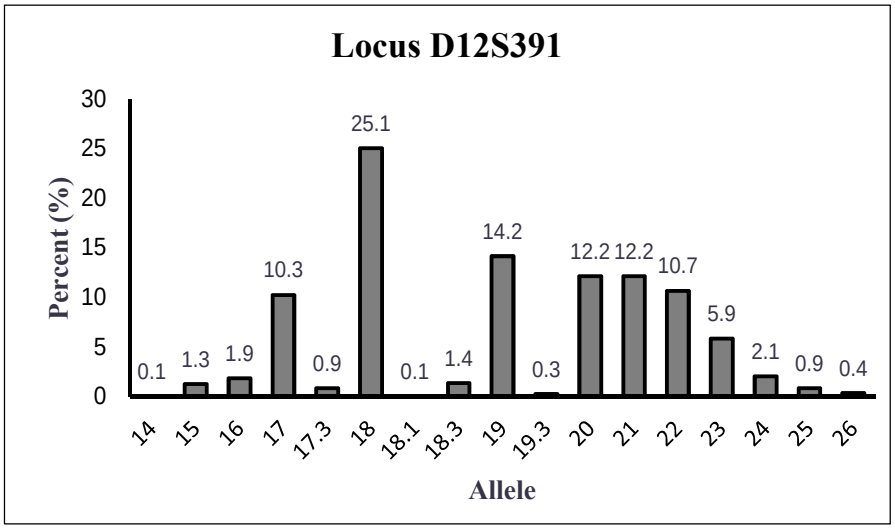


Figure 48: Percentage of allele frequency for the locus D12S391 in the Pakistani population

Locus D2S1338 (Pakistani population)

Table 47: Allele frequency for the locus D2S1338 in the Pakistani population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	12	2.094	0.3
2	16	2.792	0.4
3	17	53.746	7.7
4	18	104	14.9
5	19	103.3	14.8
6	20	99.116	14.2
7	21	30.014	4.3
8	22	46.766	6.7
9	23	120.06	17.2
10	24	66.31	9.5
11	25	55.84	8
12	26	13.262	1.9
13	27	0.698	0.1

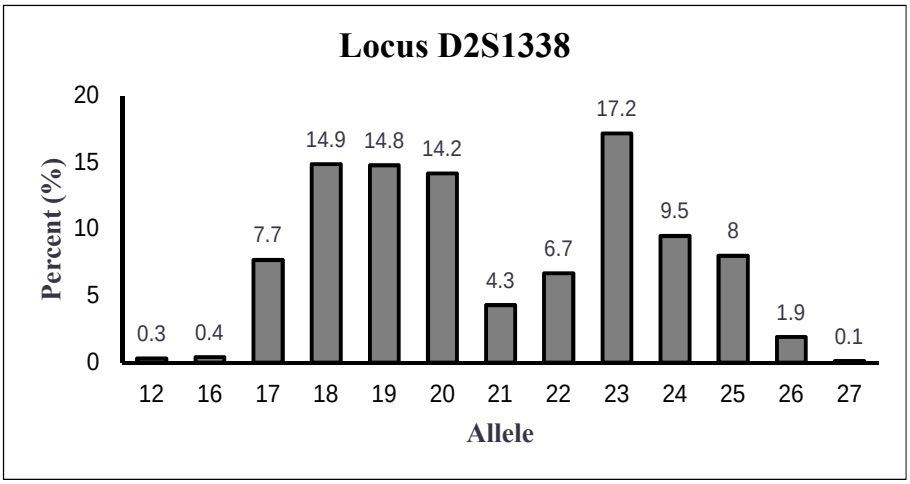


Figure 49: Percentage of allele frequency for the locus D2S1338 in the Pakistani population

Locus Penta D (Pakistani population)

Table 48: Allele frequency for the locus Penta D in the Pakistani population

Allele No.	Allele Repeats	Frequency	Percent (%)
1	2.2	4.886	0.7
2	3.2	0.698	0.1
3	5	2.094	0.3
4	6	4.188	0.6
5	7	4.188	0.6
6	8	20.242	2.9
7	9	147.28	21.1
8	9.4	4.188	0.6
9	10	113.77	16.3
10	11	161.94	23.2
11	12	97.72	14
12	13	87.948	12.6
13	14	29.316	4.2
14	14.2	0.698	0.1
15	15	18.148	2.6
16	16	0.698	0.1

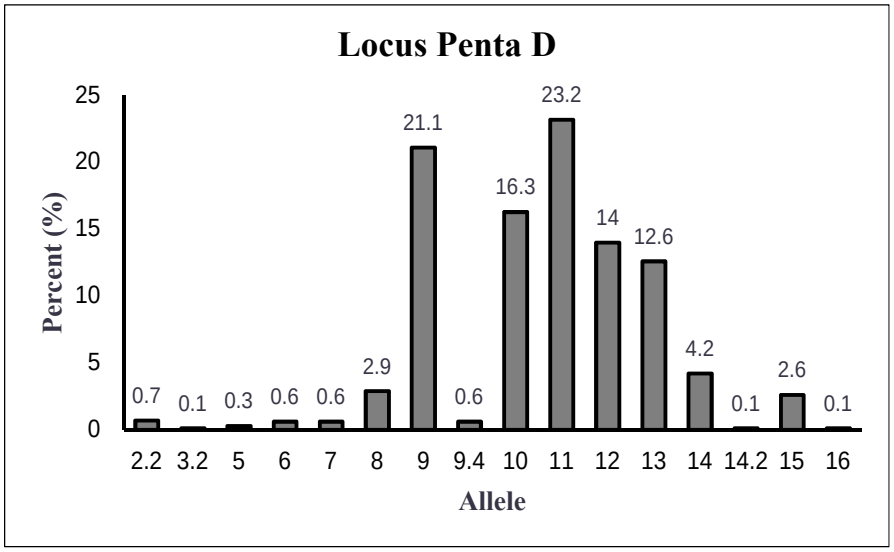


Figure 50: Percentage of allele frequency for the locus Penta D in the Pakistani population

Examining the allele frequencies in the Indian population, the locus Penta E (21 alleles) and FGA (20 alleles) showed the highest number of total observed alleles. Whereas, locus TPOX (5 alleles) showed the lowest number of total alleles. Considering the Pakistani population, the locus FGA (24 alleles) and Penta E (22 alleles) displayed the highest number of total alleles, whilst the least number of total alleles was observed in locus D16S539 (7 alleles) and TPOX (7 alleles). Thereupon, the following results reveals that Penta E and FGA are the most polymorphic loci, while the locus TPOX is the least polymorphic loci among the 23 autosomal STR loci, in both the studied populations.

3.2 Predominant Alleles

Predominant alleles of 23 autosomal STR loci in the Indian and Pakistani population residing in UAE, is presented in Tables (49-50) and Figures (51-52).

Table 49: Predominant alleles of 23 STR loci in the Indian population

Locus	Predominant Alleles	Frequency
D3S1358	16	0.303
vWA	17	0.277
D16S539	11	0.324
CSF1PO	12	0.378
TPOX	11	0.385
D8S1179	14	0.203
D21S11	29	0.212
D18S51	14	0.253
Penta E	11	0.136
D2S441	11	0.406
D19S433	13	0.310
TH01	9	0.296
FGA	23	0.200
D22S1045	15	0.385
D5S818	11	0.342
D13S317	12	0.270
D7S820	11	0.246

Table 49: Predominant alleles of 23 STR loci in the Indian population (continued)

Locus	Predominant Alleles	Frequency
D6S1043	12	0.216
D10S1248	15	0.290
D1S1656	15	0.176
D12S391	19	0.163
D2S1338	18	0.149
Penta D	9,10	0.200

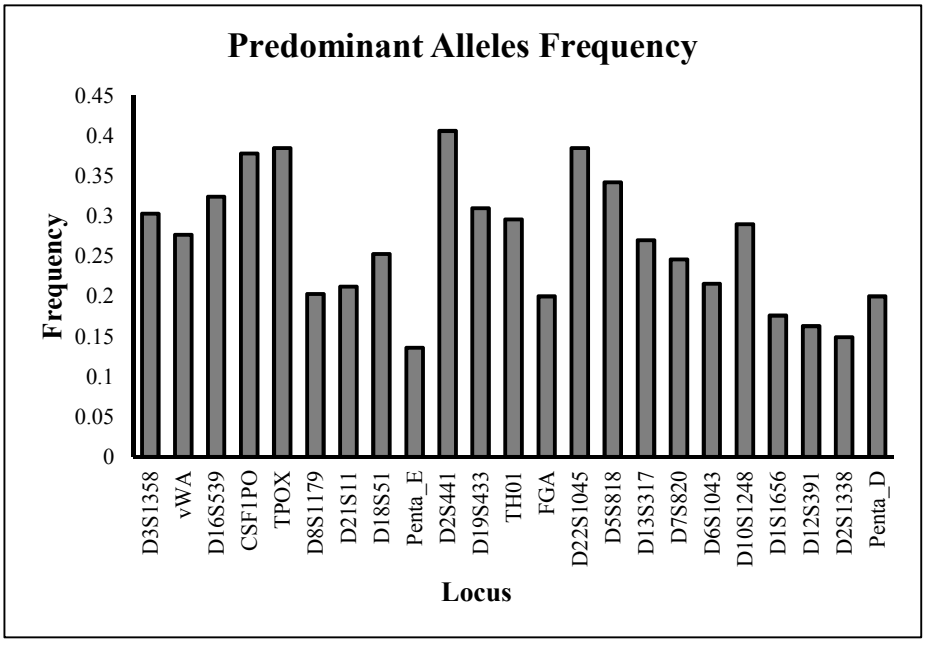


Figure 51: Frequency of the predominant alleles for 23 STR loci in the Indian population

Table 50: Predominant alleles of 23 STR loci in the Pakistani population

Locus	Predominant alleles	Frequency
D3S1358	15	0.272
vWA	17	0.242
D16S539	11	0.328
CSF1PO	12	0.371
TPOX	11	0.362
D8S1179	14	0.213
D21S11	29	0.213
D18S51	14	0.261
Penta E	11	0.130
D2S441	11	0.361
D19S433	13	0.321
TH01	6,9	0.245
FGA	23	0.183
D22S1045	15	0.362
D5S818	11	0.335
D13S317	12	0.271
D7S820	10	0.245
D6S1043	11	0.304
D10S1248	14	0.301
D1S1656	15	0.185
D12S391	19	0.142
D2S1338	23	0.172
Penta D	11	0.232

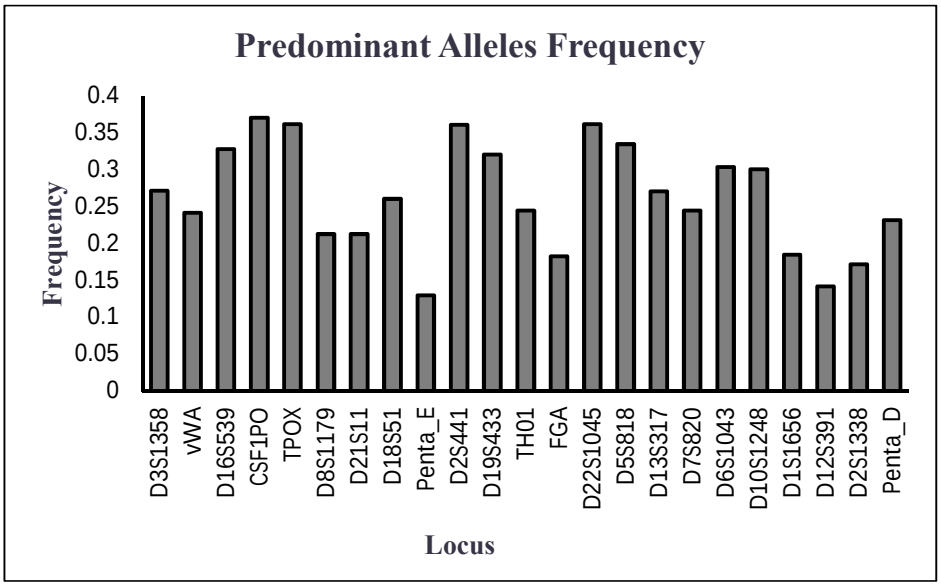


Figure 52: Frequency of the predominant alleles for 23 STR loci in the Pakistani population

Results demonstrated that allele 11, 12 and 15 were the three most predominant alleles in the Indian and Pakistani population. In the Indian population, the highest predominant allele frequency; 0.406, was observed for allele "11" in the locus D2S441. This was followed by the locus TPOX and D2SS1045, having an allele frequency of 0.385 for allele "11" and "15", respectively. Whereas, in the Pakistani population, the maximum predominant allele frequency; 0.371, was shown for the allele "12" in the locus CSF1PO, followed by the locus TPOX and D22S1045 having an allele frequency of 0.362 for the allele "11" and "15", respectively.

3.3 Genetic Diversity

The polymorphic nature of the STR nominates them to be the preferred markers for genetic diversity studies. The genetic diversity of 23 autosomal STR loci in the Indian and Pakistani population, residing in UAE is represented in Tables (51-52) and Figures (53-54).

Table 51: Genetic diversity of 23 STR loci in the Indian population

Locus	Genetic Diversity
D3S1358	0.746
vWA	0.805
D16S539	0.795
CSF1PO	0.711
TPOX	0.700
D8S1179	0.852
D21S11	0.858
D18S51	0.858
Penta E	0.921
D2S441	0.712
D19S433	0.826
TH01	0.790
FGA	0.870
D22S1045	0.732
D5S818	0.744
D13S317	0.813

Table 51: Genetic diversity of 23 STR loci in the Indian population (continued)

Locus	Genetic Diversity
D13S317	0.813
D7S820	0.808
D6S1043	0.836
D10S1248	0.784
D1S1656	0.888
D12S391	0.865
D2S1338	0.880
Penta D	0.824

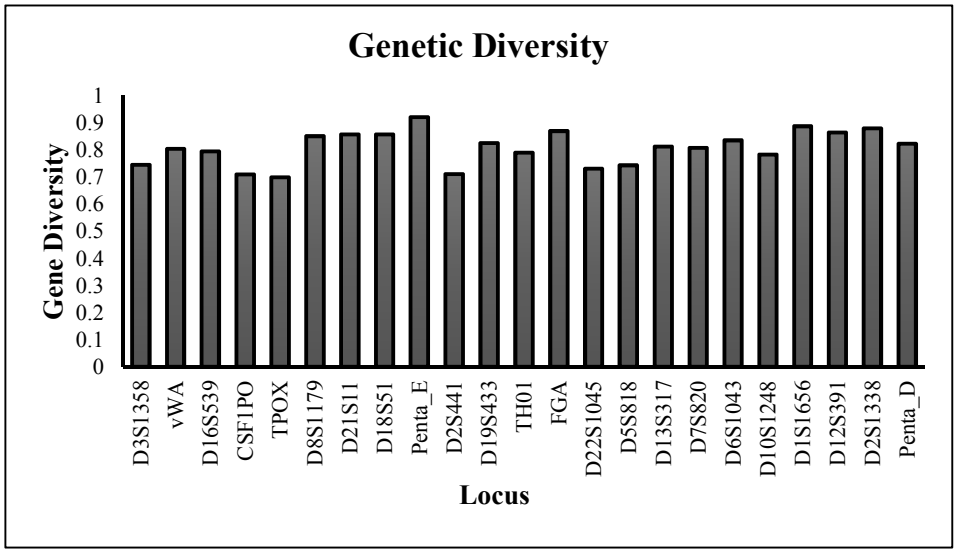


Figure 53: Genetic diversity of 23 STR loci in the Indian population

Table 52: Genetic diversity of 23 STR loci in the Pakistani population

Locus	Genetic Diversity
D3S1358	0.761
vWA	0.810
D16S539	0.793
CSF1PO	0.727
TPOX	0.699
D8S1179	0.852
D21S11	0.855
D18S51	0.855
Penta E	0.921
D2S441	0.757
D19S433	0.809
TH01	0.801

Table 52: Genetic diversity of 23 STR loci in the Pakistani population (continued)

Locus	Genetic Diversity
FGA	0.875
D22S1045	0.749
D5S818	0.757
D13S317	0.808
D7S820	0.804
D6S1043	0.821
D10S1248	0.785
D1S1656	0.888
D12S391	0.862
D2S1338	0.879
Penta D	0.837

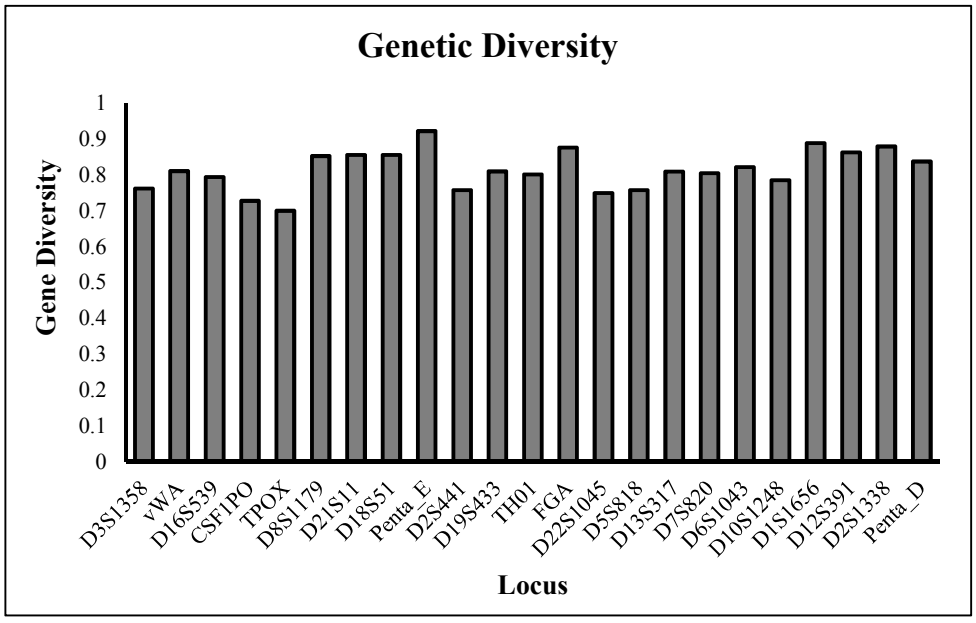


Figure 54: Genetic diversity of 23 STR loci in the Pakistani population

The genetic diversity between the populations ranged from 70 % (TPOX) – 92 % (Penta E). In the Indian population, the highest diversity was observed at locus Penta E = 0.921 (92 %), followed by locus D1S1656 = 0.888 (89 %). Whereas, the lowest diversity was observed in locus TPOX = 0.700 (70 %) and locus CSF1PO = 0.711 (71 %). In the Pakistani population, the highest diversity was perceived at the same locus as observed in the Indian population; locus Penta E = 0.921 (92 %) followed by locus

D1S1656 = 0.888 (89 %). Similarly, the least diversity was shown in locus TPOX = 0.699 (70 %) and locus CSF1PO = 0.727 (73 %).

3.4 Evaluation of Forensic Parameters

The forensic efficiency parameters analysis aids in facilitating easier resolution of forensic and paternity cases. Match Probability (MP), Power of Discrimination (PD), Power of Exclusion (PE), Polymorphism Information Content (PIC) and Paternity Index (PI) in both the populations was evaluated and presented in Tables 53 & 54, respectively.

Table 53: Forensic parameters of 23 autosomal STR loci in the Indian population

Locus	Match Probability	Power of Discrimination	Power of Exclusion	Polymorphic Information Content	Paternity Index
D3S1358	0.113	0.887	0.529	0.70	2.10
vWA	0.070	0.930	0.580	0.78	2.38
D16S539	0.070	0.930	0.500	0.77	1.96
CSF1PO	0.141	0.859	0.444	0.66	1.73
TPOX	0.144	0.856	0.422	0.65	1.64
D8S1179	0.041	0.959	0.677	0.83	3.14
D21S11	0.039	0.961	0.666	0.84	3.03
D18S51	0.036	0.964	0.722	0.84	3.67
Penta E	0.014	0.986	0.820	0.91	5.68
D2S441	0.137	0.863	0.458	0.67	1.78
D19S433	0.052	0.948	0.633	0.81	2.75
TH01	0.076	0.924	0.534	0.76	2.12
FGA	0.033	0.967	0.716	0.86	3.59
D22S1045	0.112	0.888	0.401	0.69	1.57
D5S818	0.111	0.889	0.495	0.70	1.93
D13S317	0.063	0.937	0.580	0.79	2.38
D7S820	0.066	0.934	0.544	0.78	2.17
D6S1043	0.048	0.952	0.677	0.82	3.14
D10S1248	0.083	0.917	0.565	0.75	2.29
D1S1656	0.026	0.974	0.733	0.88	3.83
D12S391	0.037	0.963	0.797	0.85	5.03
D2S1338	0.029	0.971	0.722	0.87	3.67
Penta D	0.058	0.942	0.606	0.80	2.55

Table 54: Forensic parameters of 23 autosomal STR loci in the Pakistani population

Locus	Match Probability	Power of Discrimination	Power of Exclusion	Polymorphic Information Content	Paternity Index
D3S1358	0.101	0.899	0.561	0.72	2.27
vWA	0.065	0.935	0.587	0.78	2.42
D16S539	0.074	0.926	0.561	0.76	2.27
CSF1PO	0.120	0.880	0.392	0.68	1.54
TPOX	0.144	0.856	0.401	0.64	1.57
D8S1179	0.041	0.959	0.669	0.83	3.06
D21S11	0.039	0.961	0.652	0.84	2.91
D18S51	0.041	0.959	0.680	0.84	3.17
Penta E	0.014	0.986	0.783	0.91	4.72
D2S441	0.097	0.903	0.477	0.72	1.86
D19S433	0.065	0.935	0.603	0.79	2.53
TH01	0.072	0.928	0.593	0.77	2.46
FGA	0.03	0.970	0.658	0.86	2.96
D22S1045	0.100	0.900	0.427	0.71	1.66
D5S818	0.101	0.899	0.501	0.72	1.96
D13S317	0.061	0.939	0.536	0.78	2.13
D7S820	0.071	0.929	0.598	0.77	2.49
D6S1043	0.054	0.946	0.630	0.8	2.73
D10S1248	0.083	0.917	0.587	0.75	2.42
D1S1656	0.026	0.974	0.674	0.88	3.12
D12S391	0.036	0.964	0.743	0.85	3.97
D2S1338	0.029	0.971	0.697	0.87	3.36
Penta D	0.052	0.948	0.714	0.82	3.56

The matching probability (MP) is the probability of finding a random match between two unrelated individuals selected from the same population. Based on the results obtained, in both the populations, the lowest MP was observed at locus Penta E = 0.014 followed by locus D1S1656 = 0.026, whereas the highest MP was for locus TPOX = 0.144. The combined match probability (CMP) of the 23 STR loci was $4.8x10^{-26}$ and $4.4x10^{-26}$ in the Indian and Pakistani population, accordingly.

The power of discrimination (PD) is the probability that two randomly selected individuals would not carry matching DNA profiles. As proposed from the MP, the PD value for both the populations, was the highest at the locus Penta E = 0.986, followed by locus D1S1656 = 0.974. Whereas, the least PD value was observed at locus TPOX = 0.856. The combined power of discrimination (CPD) of the 23 autosomal STR loci has been calculated as 0.99999999999999999999999989897 and 0.99999999999999999999999999901 in the Indian and Pakistani population, respectively.

The power of exclusion (PE) is utilized to signify how it's very unlikely to find a random man to be excluded as the child's biological father. In the Indian population, the PE was shown to be the highest for locus Penta E = 0.820, followed by locus D12S391 = 0.797. In the Pakistani population, the PE was the highest at the same loci, however with slightly lower PE values; Penta E = 0.783 followed by D12S391 = 0.743. The combined power of exclusion (CPE), considering all the 23 autosomal STR loci, was 0.999999991519 in the Indian population and 0.999999990719 in the Pakistani population.

Polymorphic information content (PIC) quantitively measures the informativeness of markers. The range of PIC value is commonly 0 to 1. When the marker has only one allele present, the PIC value is 0. The marker would comprise of limitless number of alleles if the PIC value is 1. A PIC value above 0.7 is acknowledged to be considerably informative. According to the results acquired, for both the populations, the PIC range was 0.64 (TPOX) – 0.91 (Penta E).

The paternity index (PI) is the genetic likelihood in favor of an alleged father being the child's biological father. The optimal PI value is 1. In both the populations, the PI value was greater than 1 in all the 23 autosomal STR loci.

3.5 Heterozygosity and Homozygosity

The higher the percentage of heterozygosity obtained, the higher the existence of allele diversity and thereupon, the minor the possibility of random sample matching. The results achieved from the tests for Heterozygosity (observed and expected) and Homozygosity in the Indian and Pakistani population living in UAE, are demonstrated in Tables 55 and 56, respectively.

Table 55: Heterozygosity and Homozygosity of 23 STR loci in the Indian population

Locus	No. of alleles	Observed Heterozygosity (%)	Expected Heterozygosity (%)	Observed Homozygosity (%)
D3S1358	10	76.10	74.64	23.90
vWA	8	79.0	80.48	21.0
D16S539	7	74.40	79.51	25.60
CSF1PO	7	71.00	71.08	29.00
TPOX	5	69.60	69.96	30.40
D8S1179	12	84.10	85.23	15.90
D21S11	18	83.50	85.81	16.50
D18S51	17	86.40	85.79	13.60
Penta E	21	91.20	92.14	8.80
D2S441	10	71.90	71.22	28.10
D19S433	15	81.80	82.61	18.20
TH01	9	76.40	79.04	23.60
FGA	20	86.10	87.05	13.90
D22S1045	11	68.20	73.19	31.80
D5S818	8	74.10	74.39	25.90
D13S317	11	79.00	81.32	21.00
D7S820	10	77.00	80.76	23.00
D6S1043	13	84.10	83.55	15.90
D10S1248	9	78.10	78.35	21.90
D1S1656	18	86.90	88.77	13.10
D12S391	18	90.10	86.53	9.90
D2S1338	14	86.40	88.03	13.60
Penta D	13	80.40	82.38	19.60

The observed heterozygosity in the Indian population ranged from 69.60 % (TPOX) to 91.20 % (Penta E). Likewise, the loci TPOX and Penta E showed the least and the highest expected heterozygosity, correspondingly.

Table 56: Heterozygosity and Homozygosity of 23 STR loci in the Pakistani population

Locus	No. of alleles	Observed heterozygosity (%)	Expected heterozygosity (%)	Observed homozygosity (%)
D3S1358	8	77.90	76.13	22.10
vWA	10	79.40	81.01	20.60
D16S539	7	77.90	79.32	22.10
CSF1PO	8	67.60	72.67	32.40
TPOX	7	68.20	69.85	31.80
D8S1179	11	83.70	85.16	16.30
D21S11	20	82.80	85.53	17.20
D18S51	18	84.20	85.52	15.80
Penta E	22	89.40	92.10	10.60
D2S441	14	73.10	75.69	26.90
D19S433	17	80.20	80.93	19.80
TH01	9	79.70	80.05	20.30
FGA	24	83.10	87.46	16.90
D22S1045	9	69.90	74.87	30.10
D5S818	8	74.50	75.65	25.50
D13S317	9	76.50	80.80	23.50
D7S820	10	79.90	80.43	20.00
D6S1043	13	81.70	82.15	18.30
D10S1248	10	79.40	78.48	20.60
D1S1656	18	84.00	88.76	16.00
D12S391	17	87.40	86.16	12.60
D2S1338	13	85.10	87.95	14.90
Penta D	16	86.00	83.73	14.00

In the Pakistani population, the observed heterozygosity ranged from 67.60 % (CSF1PO) to 89.40 % (Penta E). However, the expected heterozygosity was exhibited to be the maximum at Penta E and minimum at TPOX. According to the results attained, the locus Penta E shows the highest level of diversity amongst the remaining autosomal STR loci, in both the populations.

3.6 Hardy Weinberg Equilibrium Test

The Hardy Weinberg Equilibrium (HWE) test is executed by taking the observed allele frequencies and calculating the expected genotype frequencies depending on the allele frequencies (Lauretto et al., 2009). Conceding that the observed genotype frequencies are adjacent to the expected genotype frequencies, calculated from the observed allele frequencies, then the population is said to be in Hardy-Weinberg equilibrium and the combination of the alleles are thereupon presumably independent of each other (Guo & Thompson, 1992). The results attained from the HWE test for the Indian and Pakistani population living in UAE, are shown in Tables (57-58) and Figures (55-56).

Table 57: HWE value of 23 autosomal STR loci in the Indian population

Locus	HWE
D3S1358	0.1946
vWA	0.2442
D16S539	0.1597
CSF1PO	0.4493
TPOX	0.5487
D8S1179	0.6031
D21S11	0.3250
D18S51	0.4869
Penta E	0.4273
D2S441	0.2968
D19S433	0.6510
TH01	0.9614
FGA	0.2022
D22S1045	0.0272
D5S818	0.2725
D13S317	0.2704
D7S820	0.0199
D6S1043	0.6129
D10S1248	0.3000
D1S1656	0.1357
D12S391	0.0569
D2S1338	0.1529
Penta D	0.1861

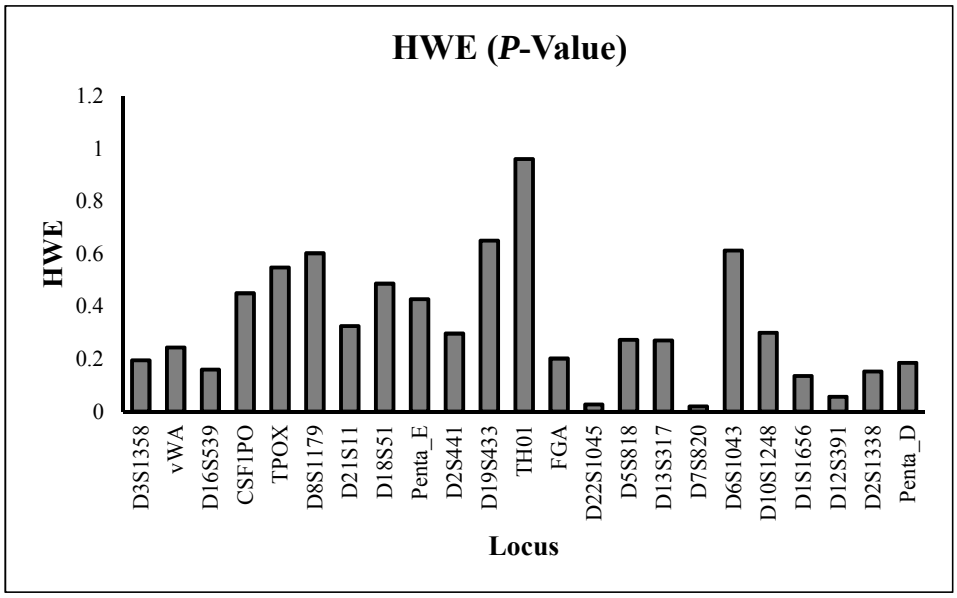


Figure 55: HWE value of 23 autosomal STR loci in the Indian population

Table 58: HWE value of 23 autosomal STR loci in the Pakistani population

Locus	HWE
D3S1358	0.4279
vWA	0.4072
D16S539	0.0282
CSF1PO	0.3636
TPOX	0.3573
D8S1179	0.7117
D21S11	0.4243
D18S51	0.3023
Penta E	0.5197
D2S441	0.1882
D19S433	0.3033
TH01	0.3220
FGA	0.3724
D22S1045	0.7704
D5S818	0.1243
D13S317	0.5257
D7S820	0.3680
D6S1043	0.8001
D10S1248	0.7355
D1S1656	0.0315
D12S391	0.5467
D2S1338	0.0765
Penta D	0.0738

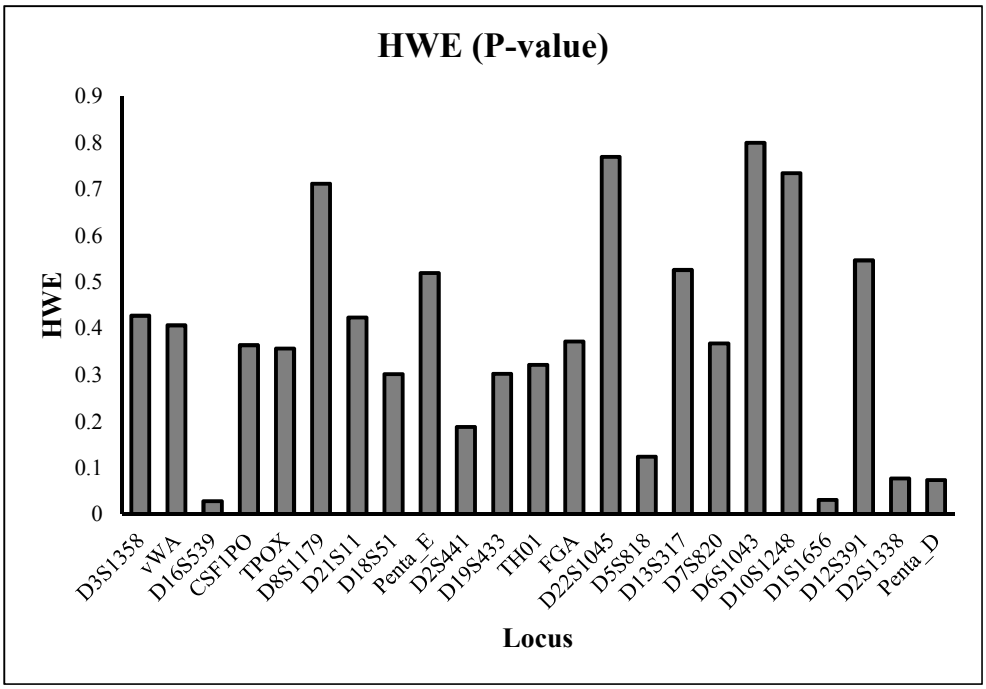


Figure 56: HWE value of 23 autosomal STR loci in the Pakistani population

The departures from HWE and *P*-values (P > 0.05) was generated by utilizing the Markov-chain method. In the Indian population, significant departure from HWE was only observed for the following loci: D22S1045 (P-value = 0.0272) and D7S820 (P-value = 0.0199) Whereas, in the Pakistani population, locus D16S539 (P-value = 0.0282) and D1S1656 (P-value = 0.0315) showed significant departure from HWE.

3.7 Analysis of Molecular Variance

The genotypes of the Indian and Pakistani population were compared to genotypes of various other populations as shown below in Table 59.

Table 59: Fst AMOVA results using 15 loci in various populations

Population	Indian	Pakistani	Mongolian	American	Egyptian	Libyan
Indian	0					
Pakistani	-0.00034	0				
Mongolian	0.00800	0.00848	0			
American	0.03476	0.03368	0.02579	0		
Egyptian	0.00938	0.00832	0.01508	0.04102	0	
Libyan	0.05278	0.05147	0.06472	0.07576	0.04699	0

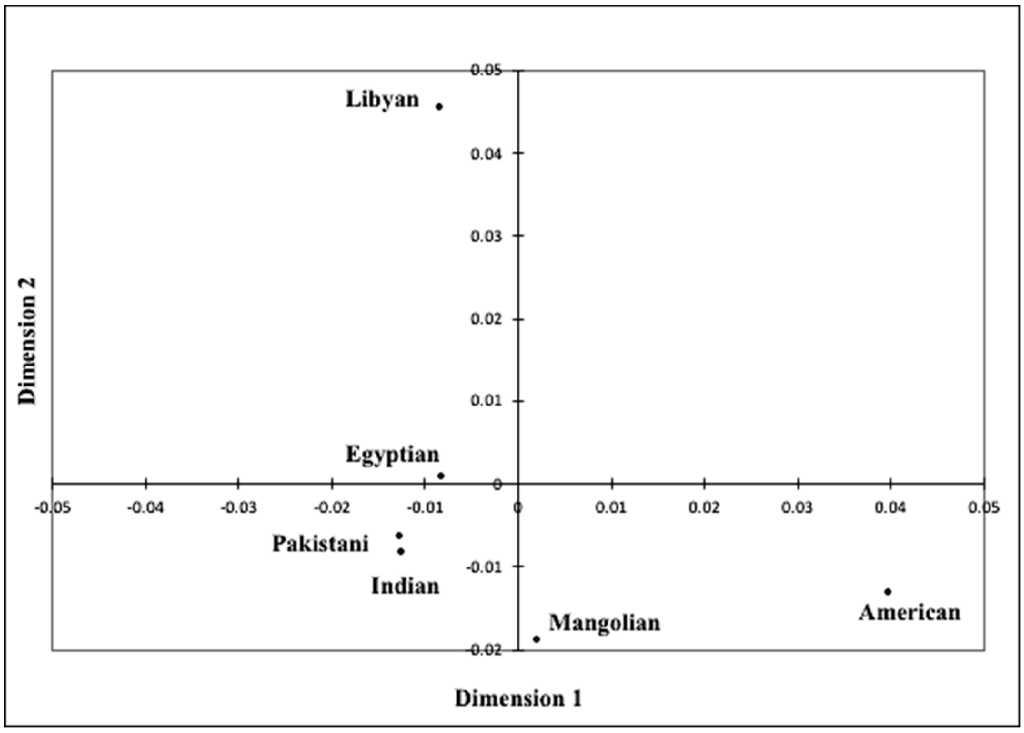


Figure 57: Multidimensional scaling plot based on pairwise Fst genetic distance

The MDS result showed that the Indian and Pakistani population share a closer genetic relation to the Egyptian and Mongolian population in comparison to the American and Libyan population.

3.8 Novel Alleles

Amplified STR alleles are compared to commercially produced allelic ladders to accurately assign the allele designations, as previously mentioned. Majority of the time, alleles acquired correlate and are represented within the STR kit allelic ladder. However, as more samples are analyzed globally, novel alleles are being discovered. Nevertheless, this signifies that it should be mandatory for the companies to specifically test all the populations before the kits are released in the market. In the present study, several novel alleles absent in the allelic ladder were identified in the Indian and Pakistani population, as shown in Tables 60 and 61, respectively.

Table 60: Novel alleles observed in the Indian population using the VeriFiler™Express Kit

Locus	Allele
D3S1358	16.2*, 17.2*
D21S11	28.3, 29.3, 31.3, 34.1
D18S51	15.2, 16.2
Penta E	12.4, 17.4, 18.5*
D2S441	12.3
D19S433	11.2, 18*
TH01	5.3*, 8.3*, 10.3
FGA	20.2, 21.2, 22.2, 22.3, 23.2, 24.2, 25.2
D7S820	7.1*, 11.3
D1S1656	18
D12S391	17.3, 18.3, 20.3
Penta D	9.4*

*not previously reported in literature

Table 61: Novel alleles observed in the Pakistani population using the VeriFiler™Express Kit

Locus	Allele
D3S1358	17.1*
vWA	15.2*
D21S11	29.3, 30.3
D18S51	11.2*, 16.2
Penta E	11.4, 12.4, 16.4
D2S441	9.1*, 11.2, 12.3, 14.1
D19S433	11.2
TH01	10.3, 20*, 24*
FGA	11*, 18.2, 20.2, 21.2, 22.2, 23.2, 24.2, 24.3, 25.2, 25.3
D13S317	4*
D7S820	18*, 26.2*
D1S1656	8, 18
D12S391	17.3, 18.1, 18.3
Penta D	9.4*, 14.2*

*not previously reported in literature

Chapter 4: Discussion

STRs have revolutionized the field of forensic genetics and is currently the standard technique utilized for individual identification, paternity testing and population diversity studies. In the present study, results were attained from 23 autosomal STR loci (D3S1358, vWA, D16S539, CSF1PO, TPOX, D8S1179, D21S11, D18S51, D2S441, D19S433, TH01, FGA, D22S1045, D5S818, D13S317, D7S820, D10S1248, D1S1656, D12S391, D2S1338, D6S1043, Penta D and Penta E) from the Indian (n = 352) and Pakistani (n = 349) population, residing in the UAE. Forensic efficiency parameters and population structure analysis including the PD, PE, MP, PIC, PI, observed HO and HE, HWE and AMOVA was evaluated.

Identifying the best markers for applications in the forensic arena is complex since the allele frequencies for each STR marker is specific to each population. Therefore, this necessitates the analysis of known STR markers in all the populations worldwide (Lahmi & Vallian, 2009). Good indicators of genetic polymorphism and the usefulness of these STR markers is particularly verified by the number of alleles observed for each marker (Hameed et al., 2015).

In the current study, the total number of alleles observed for the 23 autosomal STR loci was 248 alleles in the Indian population and 298 alleles in the Pakistani population. As depicted from the results, Penta E and FGA displayed the highest number of alleles in the Indian and Pakistani population and thus are considered to be the two most polymorphic loci. Similar results were achieved by (Okamoto et al., 2003) in the Japanese population, by using the PowerPlex®16 kit which includes 15 common STR loci. Locus Penta E (19 alleles) and FGA (15 alleles) showed the highest number of alleles and similarly TPOX showed the least number of alleles (5 alleles).

In addition, allele 12.4 in the Penta E loci was observed in the Indian and Pakistani population but was not detected in the Japanese population. The appearance of same alleles in the loci in different populations may indicate common ancestries and may be utilized in subsequent time to determine the identity of the population (Alshamali et al., 2003).

Allele 11, 12 and 15 were shown to be the three most predominant alleles in the Indian and Pakistani population. Considering the common STR loci among the different kits used in previously conducted genotyping studies, equivalent results were achieved in the Bangladeshi, Bahraini, Yemenites, Saudi Arabian and Omani population, as presented in Table 62 (Alshamali et al., 2005; Al-Snan et al., 2019).

Table 62: Predominant alleles in different populations

Predominant Alleles							
Locus	**Indian**	**Pakistani**	**Bangladeshi**	**Bahraini**	**Yemenites**	**Saudi Arabian**	**Omani**
D3S1358	16	15	15	16	16	17	15
vWA	17	17	17	16	18	16	18
D8S1179	14	14	14	13	14	15	13, 14, 15
D18S51	14	14	14	14	12	13	14
D5S818	11	11	11	12	12	12	12
D13S317	12	12	11	12	12	11, 12	12
D7S820	11	10	11	10	10	10	10
D16S539	11	11	11	11	11	11	11
TH01	9	6,9	9	6	6, 9	6	7
FGA	23	23	23	23	24	23	24
TPOX	11	11	8	8	8	8	11
CSF1PO	12	12	12	11	10	13	10

The genetic diversity in the Indian and Pakistani population, ranged from 70 % (TPOX) – 92 % (Penta E). A study conducted by (Vieira et al., 2013) in the population of Goaias State, Brazil, using the PowerPlex®16 kit showed similar results. The highest percentage of genetic diversity amongst the 15 common STR loci was for Penta E = 91 % and the lowest percentage was observed for locus TPOX = 73 %. These results

confirm that locus Penta E should be considered in forensic analysis, particularly when determining the validity of profiles from unknown biological evidence stains.

The forensic genetic parameters evaluated in the present study, determined the practicality of analyzing these 23 autosomal STR loci in forensic DNA identification among individuals from the Indian (n = 352) and Pakistani population (n = 349). The greater the number of subjects studied, the more accurate is the distribution of forensic parameters. The lowest MP value attained was for locus Penta E = 0.014 followed by locus D1S1656 = 0.026. The combined match probability (CMP) of the 23 STR loci was $4.8x10^{-26}$ and $4.4x10^{-26}$ in the Indian and Pakistani population, accordingly. The CMP value achieved affirms that the likelihood of a biological stain to have a genotype matching to that of the suspect is extremely low and therefore, it is beyond the bounds of possibility to allege that the suspect is innocent. The data achieved is comparable to (Batham et al., 2019) study for the Kahar population of Uttar Pradesh, India, using the PowerPlex®21 kit. The lowest MP was shown for the two same loci however with marginally higher MP values: Penta E = 0.024 and D1S1656 = 0.028. Nevertheless, this verifies the applicability of locus Penta E in forensics and population genetic studies.

As proposed from the MP, the PD value for both the populations, was maximum at locus Penta E = 0.986. Whereas, the least PD value was observed at locus TPOX = 0.856. The combined power of discrimination (CPD) of the 23 autosomal STR loci has been calculated as 0.99999999999999999999999989897 and 0.99999999999999999999999999901 in the Indian and Pakistani population, respectively. A study conducted by (Hadi et al., 2014) for the Iraq population showed PD values ranging from 76% to 97%. The least PD value was shown for locus TPOX

= 0.756 and the highest PD was shown for locus D21S11 = 0.969. The CPD was 0.999999972. The PD values and the CPD obtained in the present study using the VeriFiler™ Express kit (23 autosomal STR loci) was greater in contrast to the PowerPlex®21 kit (20 autosomal STR loci). This shows that the greater the array of autosomal STR loci examined, the greater the discrimination value, since the likelihood that two different individuals having an identical STR profile and possessing the same number of repeat units for all the STR loci becomes exceedingly small (Panneerchelvam & Norazmi, 2003). This affirms that the combination of these 23 autosomal STR loci are nonetheless remarkably discriminating and can be used to create a DNA-based database for the Indian and Pakistani population, residing in UAE.

In the Indian population, the PE was shown to be the highest for the locus Penta E = 0.820, followed by locus D12S391 = 0.797. In the Pakistani population, the PE was the highest at the same loci, however with slightly lower PE values; Penta E = 0.783 followed by D12S391 = 0.743. The combined power of exclusion (CPE), considering all the 23 autosomal STR loci was 0.999999991519 in the Indian population and 0.999999990719 in the Pakistani population. The CPE assuredly indicates that these 23 autosomal STR loci are applicable to determine parentage in the observed populations. Furthermore, a study conducted by (Lucassen et al., 2013) using the AmpF*l*STR®Identifiler Plus™ kit (15 STR loci) for the South African Indian population, achieved the highest PE for locus D21S11 = 0.724 and locus D18S51 = 0.724. The CPE was 0.9997874. Despite acquiring high PE values in their study, the PE values were greater in locus Penta E and D12S391 in the present study. This shows that the additional loci included in the VeriFiler™Express kit has greater power in excluding a non-related individual chosen by chance, in comparison to the

loci typed in AmpF*l*STR®Identifiler Plus™ kit. Moreover, a study accomplished by (Bhinder et al., 2018) for the Punjabi population in Pakistan, showed that locus SE33 which is currently known to be the most discriminating STR marker due to its high mutation rate and high power of individualization showed substantially lower PE value (SE33 = 0.745) in comparison to locus Penta E from the present study. Therefore, locus Penta E should be designated as an imperative marker to help in forensic investigation systems.

For the Indian and Pakistani population, the PIC range was 0.64 (TPOX) – 0.91 (Penta E). This result was concordant with the findings of (Abu Halima et al., 2009) for the Palestinian population of Gaza. The PIC ranged from 0.65 (TPOX) – 0.90 (Penta E). Similar degree of polymorphism was also found in the Eastern Chinese Han population (Sheng et al., 2018) where the highest PIC value was 0.91 (Penta E) and the lowest was 0.56 (TPOX). The high PIC values achieved in this study validated the usefulness of using these markers in genetic polymorphism studies.

In the studied populations, the PI ranged from 1.54 (CSF1PO) to 4.72 (Penta E). A PI value greater than 1 in all the loci signifies the advantages of using these markers in paternity testing applications. A study accomplished by (Pilav et al., 2017) for the Bosnian and Herzegovinian population using the PowerPlex®16 kit (15 STR loci) showed the highest PI 4.35 (D18S51) followed by 4.07 (Penta E). Whereas, the lowest PI was 1.29 (TPOX). Obtaining a high PI range implies that the combination of the 23 autosomal STR loci in the VeriFiler™Express kit will be highly beneficial and efficient for paternity testing applications.

The observed heterozygosity (HE) in a population relies on the number and the frequency of alleles in each locus. The observed HE in the Indian population ranged

from 69.60 % (TPOX) to 91.20 % (Penta E). Whereas, in the Pakistani population, it ranged from 67.60 % (CSF1PO) to 89.40 % (Penta E). The result achieved was similar to the study conducted by (Mishra et al., 2019) and (Kumawat et al., 2020) for the population of Gujarat and Rajasthan in India, respectively, by using the PowerPlex®21 kit. The highest observed HE in the Gujarat population was 91.1 % shown in locus Penta E. Similarly, in the Rajasthan population, locus Penta E showed the topmost HE of 90.0 %. Penta E exhibiting the topmost heterozygosity affirms that more allege diversity exists and thence, lesser chance of random sample matching.

P-value test ($P > 0.05$) was performed to detect any significant deviation from HWE. In the Indian population, significant departures were only observed in the following loci: D22S1045 (P-value = 0.0272) and D7S820 (P-value = 0.0199). Whereas, in the Pakistani population, locus D16S539 (P-value = 0.0282) and D1S1656 (P-value = 0.0315) showed significant departure from HWE. However, after applying the Bonferroni's correction ($P > 0.002$), all of the loci in both the populations were in HWE.

Determining the genetic structure of populations is progressively becoming an essential aspect of genetic studies. The most commonly utilized method is analysis of molecular variance (AMOVA). The genotypes of four other populations including Mongolian (Zhan et al., 2018), Libyan (Elmrghni et al., 2012), Egyptian (Omran et al., 2009) and American (Ng et al., 2016) was compared to the Indian and Pakistani population, considering the 15 common STR loci. The Multidimensional Scaling Plot (MDS) designed using XLSTAT software displayed that the Indian and Pakistani population share a closer genetic relation to the Egyptian and Mongolian population

in comparison to the American and Libyan population. The results obtained also equate with the geographical distribution of these populations.

Chapter 5: Conclusion

On the basis of low matching probability, high power of discrimination and polymorphic characteristics, the combination of the studied 23 autosomal STR loci has been proven to be highly valuable and significant for population genetics and individual identification. Particularly, prominent results were attained from locus Penta E and thence; it should be allocated as an imperative marker. The results obtained affirmed that the greater the number of STR loci examined, the greater the discrimination value. The data achieved can be utilized to build an allelic frequency database as well aid in contriving genetic studies in the future. For further study, to understand the Indian and Pakistani population extensively, next-generation sequencing technology can be used to provide greater depth of data on the STR alleles. Additionally, more populations should now be analyzed since promising results were acquired by using the VeriFiler™Express PCR amplification kit.

References

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

Abu Halimah, M. (December 2008). *Genetic variation of 15 autosomal short tandem repeat (STR) loci in the Palestinian population of Gaza Strip*. MSc, The Islamic University, Gaza.

Adams, J. (2008). Paternity testing: blood types and DNA. *Nature Education,* 1(1), 146-147.

Alshamali, F., Alkhayat, A., Budowle, B. & Watson, N. (2003). Allele frequency distributions and other population genetic parameters for 13 STR loci in a UAE local population from Dubai. *International Congress Series 1239*, 249–258.

Alshamali, F., Alkhayat, A., Budowle, B. & Watson, N. (2005). STR population diversity in nine ethnic populations living in Dubai. *Forensic Science International*, 152, 267-279.

Al-Snan, N., Messaoudi, S., R. Babu, S. & Bakhiet, M. (2019). Population genetic data of the 21 autosomal STRs included in GlobalFiler kit of a population sample from the Kingdom of Bahrain. *Plos One*, 14(8), 1-9.

Aydin, H. (January 2015). *Development of a fast, simple profiling method for sample screening using HRM of Y-STRs*. MSc, Fatih University, Turkey.

Batham, M.S., Kushwaha, K.P.S., Chauhan, T., Kumawat, R.K. & Shrivastava, P. (2019). Autosomal STR allele frequencies in Kahars of Uttar Pradesh, India, drawn with PowerPlex® 21 multiplex system. *International Journal of Legal Medicine*, 134(2), 1–3.

Bhinder, M., Zahoor, M., Sadia, H., Qasim, M., Parveen, R., Anjum, G., Iqbal, M., Ullah, N., Shehzad, W., Tariq, M. & Waryah, A. (2018). SE33 locus as a reliable genetic marker for forensic DNA analysis systems. *Turkish Journal of Medical Sciences*, 48(3), 611-614.

Brookes, C., Bright, J., Harbison, S. & Buckleton, J. (2012). Characterising stutter in forensic STR multiplexes. *Forensic Science International: Genetics*, 6(1), 58-63.

Buckleton, J., Bright, J. & Taylor, D. (2016). Forensic DNA Evidence Interpretation. 2nd ed. CRC Press, 12-15.

Butler, J. (2001). Forensic DNA Typing: Biology & Technology Behind STR Markers. Academic Press, 16-17.

Butler, J. (2005a). Forensic DNA Typing: Biology, Technology and Genetics of STR Markers. 2nd ed. Elsevier, 319-320.

Butler, J. (2005b). Forensic DNA typing: Biology, Technology and Genetics of STR Markers. 2nd ed. Academic Press, 5-6.

Butler, J. (2009). Fundamentals of Forensic DNA Typing. Academic Press, 24-149.

Butler, J. (2012). Advanced topics in Forensic DNA Typing. Academic Press, 99-165.

Dashnow, H., Lek, M., Phipson, B., Halman, A., Sadedin, S., Lonsdale, A., Davis, M., Lamont, P., Clayton, J., Laing, N., MacArthur, D. & Oshlack, A. (2018). STRetch: detecting and discovering pathogenic short tandem repeat expansions. *Genome Biology*, 19(1), 121. https://doi.org/10.1186/s13059-018-1505-2

El-Alfy, S. & Abd El-Hafez, A. (2012). Paternity testing and forensic DNA typing by multiplex STR analysis using ABI prism 310 genetic analyzer. *Journal of Genetic Engineering and Biotechnology*, 10(1), 101-112.

Elmrghni, S., Dixon, R., Coulson-Thomas, Y. & Williams, D. (2012). Genetic data provided by 15 autosomal STR loci in the Libyan population living in Benghazi. *Forensic Science International: Genetics*, 6(3), 93-94.

Excoffier, L., & Lischer H.E.L. (2010). Arlequin Suite Ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources. Available from: cmpg.unibe.ch/software/arlequin35/. Accessed 5th October, 2019.

Fan, H. & Chu, J. (2007). A brief review of short tandem repeat mutation. *Genomics, Proteomics & Bioinformatics*, 5(1), 7-14.

GE Healthcare. (2011). FTA Cards. Available from https://www.gelifesciences.co.jp/tech_support/manual/pdf/28984354%20AA_FTA%20CardsDataSheet.pdf. Accessed 28th June, 2020

Griffiths, A., Gelbart, W., Miller, J. & Lewontin, R. (1999). *Modern Genetic Analysis*. New York: W. H. Freeman. Available from: https://www.ncbi.nlm.nih.gov/books/NBK21248/. Accessed 28th June, 2020.

Grubb, N. (May 2015). *Genetic diversity in Sioux Indians of South Dakota with 21 autosomal STR loci and their forensic utility.* MSc, University of North Texas, Fort Worth, USA.

Guo, S. & Thompson, E. (1992). Performing the exact rest of Hardy-Weinberg proportion for multiple alleles. *Biometrics*, 48(2), 361-362.

Gymrek, M. (2017). A genomic view of short tandem repeats. *Current Opinion in Genetics & Development*, 44, 9-16.

Hadi, I., Abdullah, M., Jaber, A. & Yoke, C. (2014). Genetic variation of twenty autosomal STR loci and evaluate the importance of these loci for forensic genetic purposes. *African Journal of Biotechnology*, 13(11), 1210-1218.

Halman, A. & Oshlack, A., 2020. Accuracy of short tandem repeats genotyping tools in whole exome sequencing data. *F1000 Research*, 9(200), 3-4

Hameed, I., Ommer, A., Murad, A. & Mohammed, G. (2015). Allele frequency data of 21 autosomal short tandem repeat loci in Mesan and Basra provinces in South Iraq. *Egyptian Journal of Forensic Sciences*, 5(4), 150-156.

Herrera, R., Garcia-Bertrand, R. & Salzano, F. (2016). Genomes, evolution, and culture. 3rd ed. *John Wiley & Sons*, 46-47.

Houck, M. & Siegel, J. (2015). Fundamentals of Forensic Science. Academic Press, 270-271.

Huel, R., Basic, L., Todorovic, K., Smajlovic, L., Eminovic, I., Berbic, I., Milos, A. & Parsons, T. (2007). Variant Alleles, Triallelic Patterns, and Point Mutations Observed in Nuclear Short Tandem Repeat Typing of Populations in Bosnia and Serbia. *Crotian Medical Journal*, 48(4), 494-502.

Jebor, M. (2013). Autosomal STR: From locus information to next generation sequencing technology. *Research Journal of Biotechnology*, 8(10), 92-105.

Jeffreys, A., Wilson, V. & Thein, S. (1985). Hypervariable 'minisatellite' regions in human DNA. *Nature*, 314, 67-73.

Kashyap, V., Sitalaximi, T., Chattopadhyay, P. & Trivedi, R. (2004). DNA profiling technologies in Forensic analysis. *International Journal of Human Genetics*, 4(1), 11-12.

Kumawat, R., Shrivastava, P., Shrivastava, D., Mathur, G. & Dixit, S. (2020). Genomic blueprint of population of Rajasthan based on autosomal STR markers. *Annals of Human Biology*, 47(1), 1-6.

Lahmi, R. & Vallian, S. (2009). Genetic variation of informative short tandem repeat (STR) loci in an Iranian population. *Iranian Journal of Biotechnology*, 7(3), 137-141.

Lauretto, M., Nakano, F., Faria Jr, S., Pereira, C. & Stern, J. (2009). A straightforward multiallelic significance test for the hardy-weinberg equilibrium law. *Genetics and Molecular Biology*, 32(3), 619-625.

Lucassen, A., Ehlers, K., Grobler, P.J. & Shezi, A.L. (2013). Allele frequency data of 15 autosomal STR loci in four major population groups of South Africa. *International Journal of Legal Medicine*, 128(2), 275–276.

Michelle, L. & Graham, B. (April 2015). *Discovery and characterization of tetranucleotide short tandem repeats in North American bears*. MSc, University of North Texas Health Science Center, Forth Worth, USA.

Mishra, A., Dixit, S., Choudhary, S., Sharma, H. & Shrivastava, P. (2019). Forensic genetic analysis of the population of Gujarat with PowerPlex 21 multiplex system. *Forensic Science International: Genetics Supplement Series*, 7(1), 167-168.

Ng, J., Oldt, R., McCulloh, K., Weise, J., Viray, J., Budowle, B., Smith, D. & Kanthaswamy, S. (2016). Native American population data based on the GlobalFiler ® autosomal STR loci. *Forensic Science International: Genetics*, 24, 12-13.

Norrgard, K. (2008). Forensics, DNA Fingerprinting, and CODIS. *Nature Education,* 1(1), 35-36.

Okamoto, O., Yamamoto, Y., Inagaki, S., Yoshitome, K., Ishikawa, T., Imabayashi, K., Miyaishi, S. & Ishizu, H. (2003). Analysis of short tandem repeat (STR) polymorphisms by the PowerPlex 16 System and capillary electrophoresis: Application to Forensic practice. *ActaMedica Okayama*, 57(2), 59-71.

Omran, G.A.F. (April 2008). *Forensic DNA profiling of a population sample from upper (south) Egypt.* PhD, University of Leicester, England.

Omran, G., Rutty, G. & Jobling, M. (2009). Genetic variation of 15 autosomal STR loci in upper (southern) Egyptians. *Forensic Science International: Genetics*, 3(2), 39-44.

Panneerchelvam, S. & Norazmi, M.N. (2003). Forensic DNA profiling and database. *The Medical Journal of Medical Sciences*, 10(2), 20–26.

Petersen, J. & Mohammad, A. (2001). Clinical and Forensic applications of capillary electrophoresis. *Springer Science & Business Media*, 23.

Pilav, A., Pojskić, N., Ahatović, A., Džehverović, M., Čakar, J. & Marjanović, D. (2017). Allele frequencies of 15 STR loci in Bosnian and Herzegovinian population. *Croatian Medical Journal*, 58(3), 250-256.

Preet, K., Malhotra, S., Shrivastava, P., Jain, T., Rawat, S., Varte, L., Singh, S., Singh, I. & Sarkar, S. (2016). Genetic diversity in Gorkhas: an autosomal STR study. *Scientific Reports*, 6, 1-9.

Ryan, C. (2019). Tandem Repeat Disorders. *Evolution, Medicine & Public Health*, (1), 17. doi:10.1093/emph/eoz005

Saad, R. (2005). Discovery, development, and current applications of DNA identity testing. *Baylor University Medical Center Proceedings*, 18(2), 130-133.

Sheng, X., Wang, Y., Zhang, J., Chen, L., Lin, Y., Zhao, Z., Li, C. & Zhang, S. (2018). Forensic investigation of 23 autosomal STRs and application in Han and Mongolia ethnic groups. *Forensic Sciences Research*, 3(2), 138–144.

Sun, J., Zhou, L., Emerson, D., Phyo, S., Titus, K., Gong, W., Gilgenast, T., Beagan, J., Davidson, B., Tassone, F. & Phillips-Cremins, J. (2018). Disease-Associated Short Tandem Repeats Co-localize with Chromatin Domain Boundaries. *Cell*, 175(1), 224-238.

Thermo Fisher Scientific. (2017). VeriFiler™Express PCR amplification kit, user guide.assets.thermofisher.com/TFSAssets/LSG/manuals/100043588_VFE_UG.pdf. Accessed 20 September, 2019.

Thermo Fisher Scientific. (2012) *GeneMapper* ID-X *Software v1.4,* Available from: www.thermofisher.com/order/catalog/product/4479707#/4479707 Accessed 25th October, 2019.

Venables, S. (December 2012). *A population genetic analysis of forensic identity and ancestry markers in Indonesia.* PhD, National Centre for Forensic Studies, University of Canberra, Australia.

Vieira, T., Silva, D., Gigonzac, M.A., Ferreira, V.L., Gonçalves, M. & Da Cruz, A. (2013). Allelic frequencies and statistical data obtained from 15 STR loci in a population of the Goiás State. *Genetics and Molecular Research*, 12(1), 23–27.

Watson, J. & Crick, F. (1953). Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, 171, 737-738.

Whatley, H. (2001). Clinical and Forensic applications of capillary electrophoresis. *Humana*, 21-58.

Woerner, A., King, J. and Budowle, B. (2017). Flanking variation influences rates of stutter in simple repeats. *Genes,* 8(11), 329-330.

XLSTAT. (2019). Statistical software for Excel. Available from: www.xlstat.com/en/. Accessed 25th October, 2019.

Zhan, X., Adnan, A., Zhou, Y., Khan, A., Kasim, K. & McNevin, D. (2018). Forensic characterization of 15 autosomal STRs in four populations from Xinjiang, China, and genetic relationships with neighboring populations. *Scientific Reports*, 8, 1-7.

List of Publications

1) Jarrar, M., El Shafey, A., Essa, R., Damji, R., Rauf, A. & Hindi, I. (2017). Melanin relieves oxidative stress in adipocytes culture in vitro. *Pharmaceutical and Biological Evaluations*, 4(4), 212-215.

2) Shrivastava, P., Alghafri, R., Almheiri, R., Mishra, A., Alfalasi, H., Khoory, S., Damji, R., Kishan Kumawat, R., Kumar Choudhary, S., Dixit, S., Kaitholia, K., Ranjan Dash, H., Ben Trivedi, V. & Shrivastava, D. (2019). Investigation on rapidly mutating Y-STRs multiplex in Indian population: A pilot study. *Forensic Science International: Genetics Supplement Series*, 7(1), 805-806.

3) Naji, M., Damji, R., Adan, A., Abu Qamar, S. & Alghafri, R. (2019). Population genetics data of 23 autosomal STR loci for three Populations in United Arab Emirates. *Forensic Science International: Genetics Supplement Series*. 7(1), 187-188.