Arab American University

Faculty of Graduate Studies

Analysis of the genetic landscape of the Palestinian population by using Autosomal and Y-STRs

By

Wafa'a Mohammad Ahmad Al-Beiruti

Supervisor

Dr. Zaidoun Mahmoud Hasan Salah

This thesis was submitted in partial fulfillment of the requirements for the Master`s degree in

Molecular Genetics and Genetic Toxicology

February/ 2023

Thesis Approval

Analysis of the genetic landscape of the Palestinian population by using Autosomal and Y-STRs

By

Wafa'a Beiruti

This thesis was defended successfully on 9/3/2023 and approved by:

Committee members	Signature
1. Supervisor Name: Dr. Zaidoun Salah	.....
2. Internal Examiner Name: Dr. Reham Khaled Marzad	.....
3. External Examiner Name: Dr. Abdallah Abu Taha	.....

Thesis Declaration

I, Wafa'a Beiruti (student ID: 202012162), declare that I am the sole author of this thesis. To the best of my knowledge this thesis contains no material previously published by any other person except where due acknowledgement has been made.

Signature

Date: March 14th, 2023

Acknowledgment

To the martyrs of my family

To my brother Muntaser, who left this world early to save others

To my father Mohammad, who dedicated his life fighting for Palestine

To my mother, who taught me to be strong

To Dr. Zaidoun Salah, for all of the dedication and everything he taught me

Abstract

Genetic diversity fosters species ability to evolve and their resilience to adapt. Thus, it has been a research interest for decades. It does not only allow scientists to study the evolutionary process that occurred in the past, but it has many important implications in population genetics and forensic science. Worldwide, forensic testing and legal systems both depend on forensic DNA analysis. According to international forensic regulations and guidelines, each community should create its allele frequency databases to put this branch of science into practice. Thus, we present here one of the first studies that aimed to study the Palestinian population genetic structure, substructure, and the first allele frequency databases that are essential for the Palestinian justice system. Genetic diversity of the Palestinian population was investigated by screening 23 autosomal STRs and 23 Y-STRs. 617 samples were extracted using SwabSolution®. Samples were directly amplified using PowerPlex® Fusion system. Autosomal allelic frequencies ranged from 53.97% to 0.081%, with 336 alleles genotyped along the loci. No significant departure from Hardy-Weinberg equilibrium in the total population was detected at any of the tested loci after applying Bonferroni-type corrections (p-value>0.00217). Moreover, linkage disequilibrium was not detected at any of the loci co-located on the same chromosome. Observed heterozygosity ranged from 58% (TPOX) to 92.5% (SE33). Combined random match probability (CMP), combined power of discrimination (CPD), and combined power of exclusion (CPE) were 4.20677E-30, 1, and 0.999999999, respectively. Deviations from HWE were detected at the subpopulation level with the Samaritan subpopulation (p-value=0.0006) indicating a higher degree of inbreeding. For Y-STRs genotyping, 303 samples were successfully amplified using PowerPlex® Y23, 288 different haplotypes were detected. 275 were unique haplotypes (90.8% of samples). Y-STRs

Allelic frequencies ranged from 79.2% to 0.33% with 156 different alleles among all loci. The average gene diversity for the 23 Y-STRs was 65%. Haplotype match probability, haplotype diversity, and discrimination capacity were 0.39%., 99.9%, and 95%, respectively. Our analysis revealed relatedness to the surrounding countries and our In-silico analysis predicted haplogroup Q to be widespread in the Palestinian population. We reveal that the Palestinian population is quite diverse, moreover, our results indicate that both autosomal and Y STRs systems presented in this study are useful forensically and can be implemented in the West Bank-Palestine. In case of encountering a Samaritan Autosomal DNA profile or a sample of unknown sources, a theta (Θ) correction factor of 0.03 needs to be applied to correct for the deviation.

List of contents

List of tables

Palestinian population and 29 worldwide populations.

- **Table .17: Haplogroup prediction based on White Athey Algorithm.**

List of Figures

List of abbreviations

Abbreviation	Definition
ABI	Applied Biosystems, Inc.
AF	Alleged Father
al	Allele
bp	Base pair
CDP	Combined Discrimination Power
CEP	Combined Exclusion Probability
CMP	Combined Matching Probability
CODIS	Combined DNA Index System
CPI	Combined Paternity Index
DNA	Deoxyribonucleic Acid
He	Expected Heterozygosity
Ho	Observed Heterozygosity
HWE	Hardy Weinberg Equilibrium
ILS	Internal Lane Standard
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
PD	Power of Discrimination
PE	Probability of Exclusion
PIC	Polymorphism Information Content
POP	Probability of Paternity

POP-4	Performance Optimized Polymer 4
SNP	Single Nucleotide Polymorphism
SSR	Simple Sequence Repeats
STR	Short Tandem Repeats
TPI	Typical Paternity Index
DNMs	De Novo Mutations
BCE	Before the Common Era
WGS	Whole Genome Sequencing
NGS	Next Generation Sequencing
OL	Off-Ladder
OB	Out-of-bin
AT	Analytical threshold
AMOVA	Analysis of Molecular Variance
HD	Haplotype diversity
DC	Discrimination capacity
HMP	Haplotype random match probability
HF	Haplotype frequency
PCO	Poly component analysis
MDS	Multidimensional scaling

Chapter 1

Introduction

Human diversity and evolution have been a research interest for decades and the pattern of genetic variation was studied over time to understand the adaptation and evolution behavior of species and populations (Lowe et al., 2004). DNA analysis provides the privilege of investigating the genetic relationship at the individual, population, or species level. Moreover, it helps to investigate the origin, migration, and past events that resulted in the fixation or loss of genetic variations within and between populations and subpopulations. Such analysis is applied to understand population genetics issues studied nowadays like genetic drift, population size, population structure, natural selection, or migration (Ryman & Utter, 1987, Lowe et al., 2004, Dudgeon et al., 2012). Genetic diversity has important implications in many fields. For example, harvest regulation is performed by analyzing the genetic population structure and the distribution of subpopulations (Ryman & Utter, 1987). Genetic variation is a key determinant of disease susceptibility, personalized medicine or response to therapy, and clinical outcome (Marian, 2020,Baker & Antonovics, 2012, Eichler, 2019). Furthermore, forensic investigation is one of the greatest implications of population genetics parameters (Jobling, 2022, Weir, 1992), which is one of the main aims of this thesis.

The human genome which consists of approximately 3 billion base pairs includes important information that enables scientists and experts to identify each individual through typing of specific non-coding regions in the chromosomes known as repetitive sequences that are widely distributed throughout the human genome. These non-coding regions are highly variable and polymorphic. However, because variations in these regions are usually tolerated and do not affect the survival of humans, these repeated sequences become very diverse with generations. Thus, they can be analyzed

by different techniques to identify individuals (Panneerchelvam & Norazmi, 2003, AJ, 1987). Generally, DNA fingerprinting aided in the development of forensic science enormously through the implementation of specific guidelines to analyze DNA profiles and therefore, achieving justice in court by convicting criminals and acquitting innocent people (Roewer, 2013, Stanley et al., 2020, Panneerchelvam & Norazmi, 2003).

In the introduction we will talk about the process and factors that promote genetic diversity, we will make demographic inferences from genetic diversity, then reveal modern humans' origin and history, and finally, we will touch on short tandem repeats (STRs) typing.

1.1 Processes and factors promoting genetic diversity

Population genetics investigate alteration in population genetic structure, allele frequency, and genotype frequency, which are affected by many factors through time and space. Thus, understanding the processes that shape this genetic diversity is essential to interpret the current population genetic structure. This understanding can aid in the deduction of past human genetic diversity and past process or events that occurred and resulted in the current genetic state such as divergence, coalescence, admixture, and growth rate (Krawczak, 2009). Continuous variation events are the main driver for genetic variation, which introduces new variants into a large pool of existing variants in a population at a specific time (Hartl & Borot, 1994). However, critical biological processes and mechanisms play a role in increasing or decreasing this genetic variation (Etxeberria & Umerez, 2008). Forces such as natural selection, migration, gene flow, and genetic drift drive this evolution in genetic variability (Taboun et al., 2021). These factors and processes will be discussed below.

1.1.1 Genetic variations

Variations are the leading cause of genetic diversity as they are the only way to that introduce new variants in species (Hartl & Borot, 1994). Variations can be either somatic or germline. Even though somatic variations have a great impact on health and disease state (García-Nieto et al., 2019), they are not inherited from a parent to a child and thus, they are not considered a source of interest for evolution or genetic diversity research. Evolutionary significant variations are variations inherited through a parental line and transmitted through generations (Berdan et al., 2021). Examples on variations are insertion-deletion (indel), frameshift, inversion, or duplication (Vogelstein et al., 2013). DNA generally exhibits a low variation rate, especially for unique sequences with approximately 10^{-9} nucleotides per generation. However, STRs have higher variation rates compared to unique sequences, and different STRs have different variation rates ranging from 10^{-6} to 10^{-2} (Chen et al., 2020, Fan & Chu, 2007). In a study that was conducted on a father-offspring trio, they found that paternal age at conception of offspring is strongly correlated with the rate of de novo mutations (DNMs), with an average increase of about two DNMs per year (Kong et al., 2012). Maternal age is also strongly correlated with the increased incidence of DNMs. In another study that was conducted using whole genome sequencing (WGS) on 61 parents–offspring trios to investigate the origin of parental DNMs rate in offspring, they documented an increase of paternal DNMs rate by 0.31 per year. However, the maternal DNMs rate was 0.12 per year (Wong et al., 2016).

1.1.2 Recombination

Recombination or crossing over is an important cause of genetic diversity. It does so by generating new allele combinations between neighboring loci (Stapley et al., 2017). The mechanism involves the physical breakage and subsequent joining of a segment of DNA in sister chromatids in homologous chromosomes. These combinations of alleles are known as haplotypes and such continuous shuffling of genetic material can produce infinite haplotypes number (Strachan & Read, 2018). The degree of association between loci is measured by linkage disequilibrium; high linkage disequilibrium represents a strong interaction between loci, meaning that these loci act dependently. Usually, recombination behaves in a way to reduce linkage disequilibrium between physically linked loci over successive generations (Hamilton Matthew B, 2009). Factors affecting linkage disequilibrium other than recombination fraction and distance between loci are natural selection, genetic drift, and admixture. Subsequently, different populations can show marked differences in the extent of linkage disequilibrium depending on their history(D. E. Reich et al., 2001).

1.1.3 Migration and gene flow

A population in panmixia is a population that acts as a single entity with a uniform movement of individuals. However, this assumption cannot be true in reality. Individuals' mating potential depends largely on their location and physical distance, which leads to population structure. The migration of individuals from one area to another and the successful reproduction of the migrant leads to gene flow (Hamilton, 2014). Several models describe paths and rates of migration and gene flow. The continent-island model considers the gene flow to be unidirectional from a large to a smaller population. In other words, allele frequency in the continent is unchanged by immigration, however, allele frequency in the small population in the island is strongly affected (Haldane, 1930). The infinite island model depends on the assumption that a population is divided into an infinite number of identical

subpopulations, each subpopulation has an equal potential for migration and exchanging genetic material with equal rates regardless of their location (Wright, 1931). The isolation-by-distance model proposes a continuous population where distance limits mating potential (Wright, 1943). Finally, the stepping-stone model approximates the isolation-by-distance model by the gene flow between only adjacent subpopulations, which produces a cumulative effect on allele frequency (Motoo Kimura & Weiss, 1964).

1.1.4 Genetic drift

Genetic drift refers to the random change of allele frequency in successive generations. This occurs due to the stochastic process of sampling and sampling error from the gene pool. When we take a sample from a population to generate the next population, allele frequency can change by chance alone (Hamilton Matthew B, 2009). Genetic drift extent is a function of population growth and sample size. Small populations can exhibit large genetic drift and reduction in genetic diversity, which can result in the founder effect and genetic bottleneck (Star & Spencer, 2013). Moreover, such genetic drift can result in the differentiation of allele frequency between two populations. Migration and natural selection can alter the extent of genetic drift. One of the most famous models to describe genetic drift is the Wright-Fisher model (Fisher, 1930, Wright, 1931), which holds the assumption of random mating, non-overlapping generations, and finite population. This introduces the idea of effective population size (N_e), as we expect that not all the population contribute to the next generation. Effective population size is the size adopted by an ideal Wright-Fisher model and it is the size that experiences the same genetic drift as the census size of the population (Deffner D, Kandler A, 2022)(Fig.1).

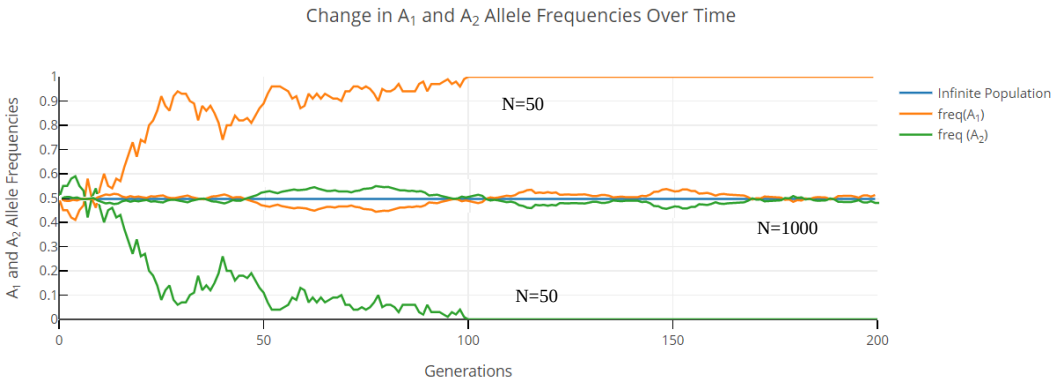


Figure.1: Genetic drift simulation in populations of different sizes. Simulations show the allele frequency from generation to generation of populations of size either 50 or 1000 for a total of 200 generations. Each starts from the same initial allele frequency (0.5). The allele rapidly becomes either fixed or lost in small populations, whereas more subtle variations are seen in larger populations. The figure was simulated using online Web PopGen II (https://sites.radford.edu/~rsheehy/Gen_flash/popgen/).

1.1.5 Natural selection

Natural selection can be defined as the maximization of fitness which results in adaptive evolution (Teotónio et al., 2009, Birch, 2016). Generally, organisms show differential survival, which can result in natural selection that depend on their survival, and reproduction success. Gamete viability is measured by the potential success of fertilization. This depends on the potential fertilization compatibility genes, bias in frequency, and combination of specific alleles in gametes (Springate & Frasier, 2017). Natural selection can influence genetic diversity by either maintaining, increasing, or decreasing the amount of variability. If natural selection results in weeping out deleterious alleles or the fixation of fitness alleles, it reduces genetic diversity, which is called negative or purifying selection. Genetic variability can be maintained or increased in the case of the heterozygote allele combinations being more compatible and fit compared to either homozygote combination, a process known as balancing or positive selection (Lohmueller et al., 2011).

1.1.6 Neutral theory

In contrast to the natural selection theory, the neutral theory claims that the majority of evolutionary variations are selectively neutral and caused by random fixation due to random genetic drift in finite populations (Motoo Kimura, 1991). It is based mainly on the observation made by Kimura where some substitutions found in proteins and RNA are tolerated because they do not affect viability or function, or risk for extinction. They are also coupled with hemostasis in organisms (M. Kimura, 1986). It is worth mentioning that the neutral theory provides an essential null model to test and detect natural selection.

1.1.7 Inbreeding and consanguinity in humans

Many forces influence random mating between individuals. Geographic distribution, ethnicity, religion, and social and economic status play an important role in individuals choosing their mates. This can also influence genetic diversity by causing a drift in allele frequency between homogenous subpopulations (Bittles & Black, 2010). Positive assortative mating is a pattern of nonrandom mating where, usually, individuals with similar genotypes or phenotypes tend to mate with each other. This results in a decrease in genetic variation with an increased homozygous genotype portion in such subgroups. A less common phenomenon is known as negative assortative mating in which individuals who are different from each other tend to mate. As opposed to positive assortative mating. This mating pattern results in increased genetic variation with increased heterozygous genotype frequency (Rios Moura et al., 2021, Border et al., 2022).

1.2 Drawing demographic conclusions from genetic diversity

Apart from natural selection, demographic factors such as divergence, migration, and admixture can change effective population size. Such factors can dramatically shape the genetic diversity along the entire genome. (Lawton-Rauh, 2008). In this section, we will summarize the tools and methods used to measure genetic variations and make some conclusions on the past genetic events and processes that shaped this diversity as well as populations' relationships.

1.2.1 Molecular markers used for measuring genetic variation

Genetic diversity can be measured by different methods. For example, genetic variation between humans was initially studied by detecting protein differences and blood groups. Using these classical markers, variation was studied on the allele frequency level because the molecular basis of DNA sequence was not discovered (Bodmer, 2014). However, with the development of advanced molecular techniques such as polymerase chain reaction (PCR), genetic variation can be investigated at the DNA sequence level by using unique molecular markers that can be analyzed based on the allele frequencies and distances between alleles at a locus in terms of evolution. (Erlich & Arnheim, 1992, Bodmer, 2014).

The non-recombining region of the Y chromosome (NYR) and mitochondrial DNA (mtDNA) are of special interest in lineage studies as they keep a sequential record of gene diversity throughout history. mtDNA is inherited mainly from the maternal line meanwhile the Y chromosome is inherited as a haplotype from the paternal line. Such features allow for the analysis of lineage and unique phylogenies construction (Kundu & Ghosh, 2015). The main types of genetic markers used in genetic diversity studies nowadays are microsatellites and single nucleotide polymorphisms (SNPs) (Liu et al., 2005). microsatellites are repetitive sequences known as short tandem repeats (STRs) or simple sequence repeats (SSRs) that can range in their length between 2-6 base pairs (bp) long (Mossallam et

al., 2005, Stanley et al., 2020). In the current study, we analyzed the Palestinian population diversity using autosomal and Y-chromosome STRs (see section 1.4 for more detailed information about STRs).

1.2.2 Measures of molecular diversity

Genetic diversity can be simply estimated by either calculating the mean number of alleles present over a range of molecular markers or loci or by counting the number of a specific allele at a specific locus. This method can be used to compare populations or subpopulations and thus, deduce some past demographic processes such as admixture or bottleneck by changing the mean number of alleles (Caballero & García-Dorado, 2013). Statistically, gene diversity can be measured by calculating the polymorphic loci present across the genome, a method that is known as Nei's formula (Nei, 1973). A similar method to Nei's utilizes heterozygosity measurement at a molecular marker position (Pagnotta, 2018). Haplotype count can also infer on genetic diversity of a population. Initial studies that compared populations based on genetic diversity analysis aided the development of theories on the origin of humans and migration.

1.2.3 Measures of diversity apportionment

As mentioned earlier, subpopulation isolation and divergence can result in differential fixation and loss of alleles in these subpopulations, and loss of heterozygosity in the total population. Consequently, population structure can develop and the extent of differentiation and diversity apportionment can be measured (Wright, 1943). F_{st} is a common measure of population structure, it is derived from the fixation indices calculations. Fixation indices measure the deviation of observed heterozygosity compared to expected heterozygosity under Hardy-Weinberg equilibrium (HWE). The F_{st} formula measures the differentiation between the total population and the subpopulation by comparing the heterozygosity levels between them (the formula is described in methodology section; chapter 2)

(Kitada et al., 2021). Analysis of Molecular Variance (AMOVA) is widely used to measure the extent of population subdivision. It depends on estimating variance components and calculating F-statistics, thus, it measures genetic distances between different alleles (L. Excoffier et al., 1992).

Genetic distance between populations can be measured by many means. For example, Nei's standard genetic distance holds the assumption that differences are caused by mutation and genetic drift (Nei, 1973). Another model that measures the genetic distance is the Cavalli-Sforza and Edwards model. In this model, populations are considered points in a multidimensional space and assumes the presence of genetic drift only. It is a very good model for generating phylogenetic trees and many studies applied it to STRs (Cavalli-Sforza & Edwards, 1967).

Another method used to measure population diversity is the clustering method. It can be used to detect population structure by the resemblance between the populations analyzed. It can also deduce shared ancestry or admixture events that occurred in the past. The clustering method is divided into two types: the distance-based method and the model-based method. The distance-based method can be used by either Multidimensional Scaling (MDS) or Component Analysis (PCA) (fig.2B). PCA depends on the allele frequency rather than genetic distances considered by MDS(Price et al., 2006). Constructing a tree using Neighbor-Joining (NJ) method also depends on the distance-based method, which combines populations that have the least genetic distance (fig.2A). The model-based method estimates the population allele frequency and attempts to assign individuals to ancestral populations (individual grouping) based on their genotypes (fig.2C) (Pritchard et al., 2000).

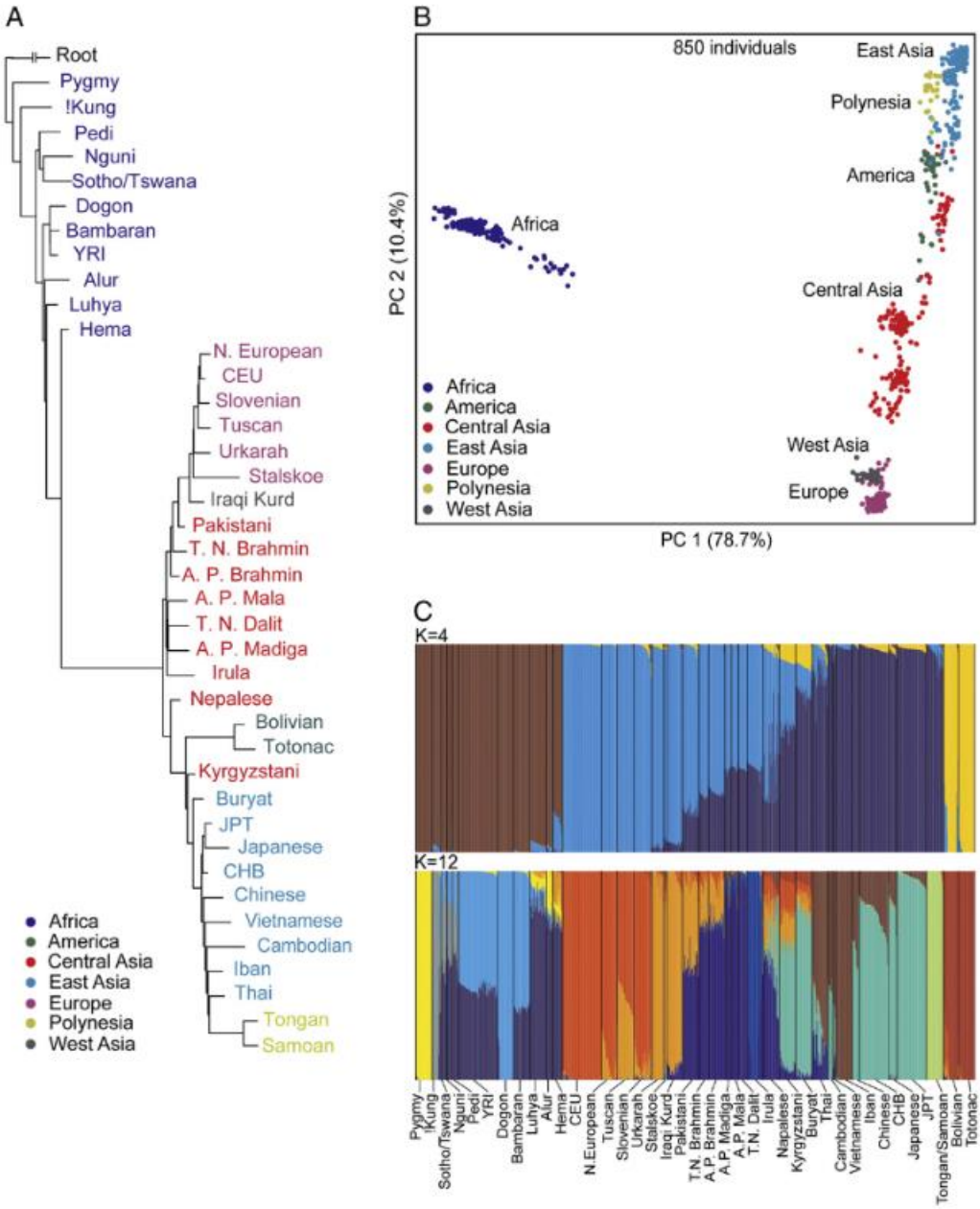


Figure.2: Examples of clustering methods used to reveal population relationships between the 40 populations. A) Neighbor-joining tree. Populations are color-coded based on their continental origins. The hypothetical ancestral population is shown. Bootstrap support values for most branches are larger than 95%. B) Principal component analysis. The first two principal components (PCs) are shown. Each individual is represented by one dot and the color label corresponding to their regional origin. The percentage of variance explained by each PC is shown on the axis. C) Individual grouping inferred by ADMIXTURE. Results from K = 4 and K = 12 are shown. Each individual's genome is represented by a vertical bar composed of colored sections, where each section represents the proportion of an individual's ancestry derived from one of the K ancestral populations. Individuals are arrayed horizontally and grouped by population as indicated (Xing et al., 2010).

1.3 Modern humans' history and origin

Understanding the mechanisms shaping genetic diversity and the necessary measures used to analyze genetic variation which we summarized up in the preceding sections, allow analysts to use these genetic data to infer the populations' demographic history. The following section will review the knowledge gained about the origin and migration of modern humans. In addition, we will do an overview of the Middle East and the Palestinian population demographic history. Finally, we will do some literature review on genetic diversity research conducted using STRs.

1.3.1 The origin and disperse of modern humans

The origin of modern humans was extensively studied through the last decades. Based on these studies, hypotheses were proposed and two main theories (models) were generated based on them. These models are the multiregional evolution and the recent replacement models. The multiregional model assumes that our ancestors surrounded the entire old world. They suppose that the archaic human lived and evolved through the past 1.5-2 million years. They also assume that this archaic human lived in different areas around the world without speciation by having enough gene flow between these areas (Thorne & Wolpoff, 1992). On the other hand, the replacement model assumes that modern humans have recent African origin, with later distribution and migration which replaced the archaic human (Disotell, 1999). Until now, both archeological and genetic data supported the second model of modern human origin and dispersion. Fossils found in Africa present the gradual development of anatomically modern skeletal features. Omo I & II and Herto skulls that belong to 195,000 to 160,000 years ago were found in Ethiopia and they were suggested as the immediate modern ancestors of the modern human(McDougall et al., 2005). Fossil analysis predicted that the first migration occurred out of Africa approximately 100,000 years ago mainly to the Levant (Schwarcz & Grun, 1992). A second migration was suggested by archaeological and genetic data approximately 50,000 years ago, which is

considered the source of all non-African populations (Quintana-Murci et al., 1999)(Stringer, 2012). Until now, the route of migration of modern humans is still unknown and controversial. However, some data suggest a favorable route from Bab el Mandeb that separates Djibouti and the Arabian Peninsula (Fernandes et al., 2012). Furthermore, DNA analysis from ancient humans that belong to the Neanderthals and Denisovans suggested two episodes of limited gene flow to modern humans. The first episode was the initial dispersion from Africa, and the second episode was directed to Oceania, which resulted in the modern Melanesian(Green et al., 2010)(D. Reich et al., 2010). Understanding the migration pattern facilitates genetic diversity analysis because migration and thus, expansion has a great effect on the molecular level. For example, migration from Africa resulted in the founder effect model where new subpopulations were formed from the main population with a subset of genetic and variant combinations, which results in a decrease in genetic diversity. All of these findings encourage scientists to study population differences to deduce the origin of the modern human. In addition, they will be able to conclude the events that occurred in the past and shaped the current genetic variation found worldwide.

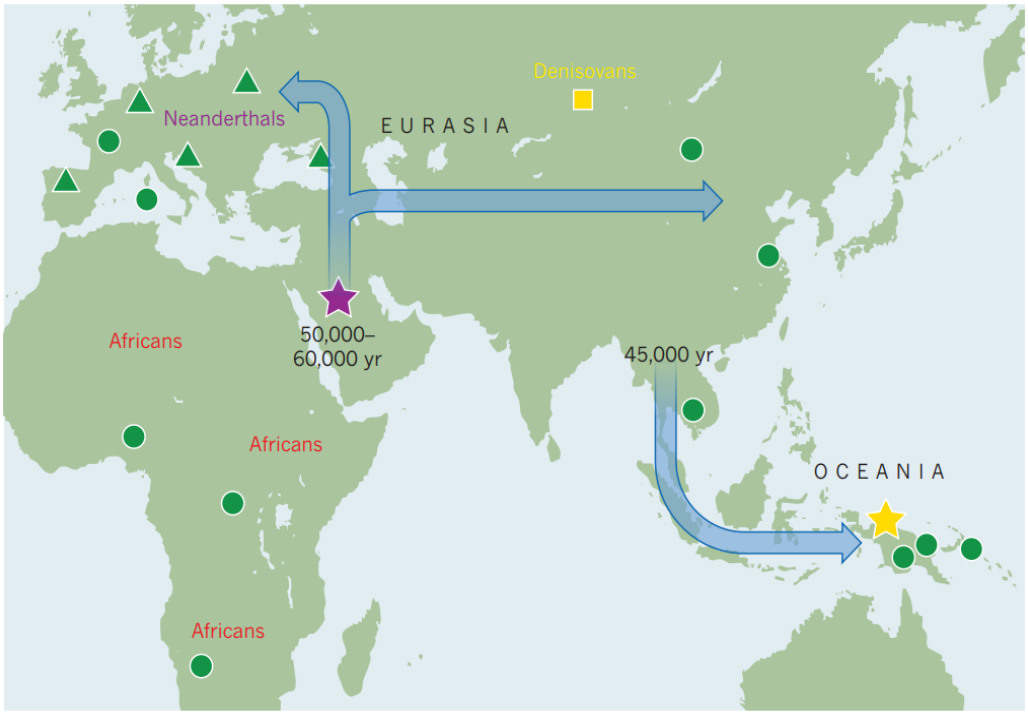


Figure.3: Modern human migration hypotheses. Triangles and circles respectively represent sampling locations of Neanderthal remains and present-day human genomes. The blue arrows indicate generally accepted major migrations of anatomically modern humans, following their departure from Africa 50,000–60,000 years ago. At this time, there were two primary archaic species in Eurasia, Neanderthals, and Homo erectus; Reich, Pääbo, and co-workers suggest that a third group was also present, represented by the ancient Denisovan genome. From ancient DNA, they identify additional putative events involving two episodes of limited gene flow: first, genetic admixture from Neanderthals to modern humans, shortly after the exit from Africa; second, subsequent admixture with the archaic population exemplified by the nuclear DNA extracted from the Denisova finger bone. This second event seems to affect only the ancestors of present-day Melanesians, who are thought to have colonized Papua New Guinea some 45,000 years ago. African populations, both past, and present are genetically highly diverse, as indicated by the multiple labels. Figure taken from (Bustamante & Henn, 2010).

1.3.2 Bidirectional migration and trade through the Middle East

Analyzing the genetic diversity of the Middle East is interesting because of its geographical location, cultural complexity, religion, political, social, and economic status. Many historical events and factors played a role in the interaction and migration between areas in the Middle East, North Africa, and South Asia. Examples of these players include the early migration of modern humans, trade, climate conditions, food, war, colonization, and Islamic expansion in this region and bidirectional migration. All of these factors aided in the complexity of genetic variation in populations inhabiting this region (Barbujani et al., 2013, Maca-Meyer et al., 2003).

mtDNA haplogroups analysis performed to infer the evolution of modern humans and migration revealed important results. For example, they discovered the L1 haplotype which was restricted to African populations, afterward, they discovered the L2 and L3 haplotypes in Africa which were derived from the L1 haplotype. It is hypothesized that migration events that occurred from Africa resulted in the development of the M and N haplotypes detected in the Middle East (Forster, 2004). Other mtDNA analyses revealed bidirectional dispersion between the Middle East and Africa. U6 mtDNA lineage that is originating from the Middle East was detected in Africa suggesting back migration from the Middle East to Africa 40,000 to 50,000 years ago. This is expected, as the Middle East is a multidirectional region that connects three continents, Asia, Africa, and Europe (Maca-Meyer et al., 2003).

Archaeological data and genetic analysis revealed plant and animals trade routes between Africa and the Arabian Peninsula back to 6000 BCE (Before the Common Era) (Hodgson et al., 2014). During the 6th century, cotton trade and successive population dispersion were documented across the Red Sea coasts of the Arabian Peninsula and Africa(Abu-Amero et al., 2007). More recent marine trade directions were documented in the 20th century across the United Arab Emirates (UAE), Iran, and Qatar(Potter, 2009).

It is important to note that even though the Middle East shows high heterogeneity, however, some homogeneity can be detected in some populations in this region. The Arabian Gulf is a good example presenting genetic homogeneity probably due to isolated subpopulations (Petraglia et al., 2010) and consanguinity which affects heterozygosity rates such as what happens in the Bedouins (Reilly, 2013).

1.3.3 Demographic history of Palestine

Palestine is a small country located in the Middle East. According to the Palestinian Central Bureau of Statistics (PCBS), the total Palestinian population in Palestine is approximately 5,360,000, with 3,190,000 living in the West Bank and 2,170,000 living in Gaza Strip(*PCBS*). Moreover, there are Palestinians in the occupied regions inside the green line (1948 Palestinian regions) and Palestinians in diaspora. Due to its strategic location and religious importance for Muslims, Christians, and Jewish people, Palestine was occupied and colonized many times throughout history including by Romans, Byzantines (Broshi, 2018), British mandate (Tsimhoni, 1984), Ottoman empire, and many others. Moreover, according to archeological studies, Jericho is considered one of the oldest cities worldwide which goes back to more than 10,000 BP (Sparks, Rachel Thyrza;Wagemakers, 2020). Currently, Palestine is divided into the West Bank, Gaza Strip, and a few locations in Jerusalem; these are the main places where Palestinians are allowed to live. Furthermore, the West Bank is completely separated from Gaza Strip, which means that Palestinians from Gaza Strip are not allowed to move to the West Bank and vice versa. Furthermore, Palestinian population is very diverse and inclusive for some minor groups (Christians, Samaritans, and Palestinian Africans), who have been living in the region for long time. All of these variables drive curiosity about the Palestinian genetic population structure (Fig.4).

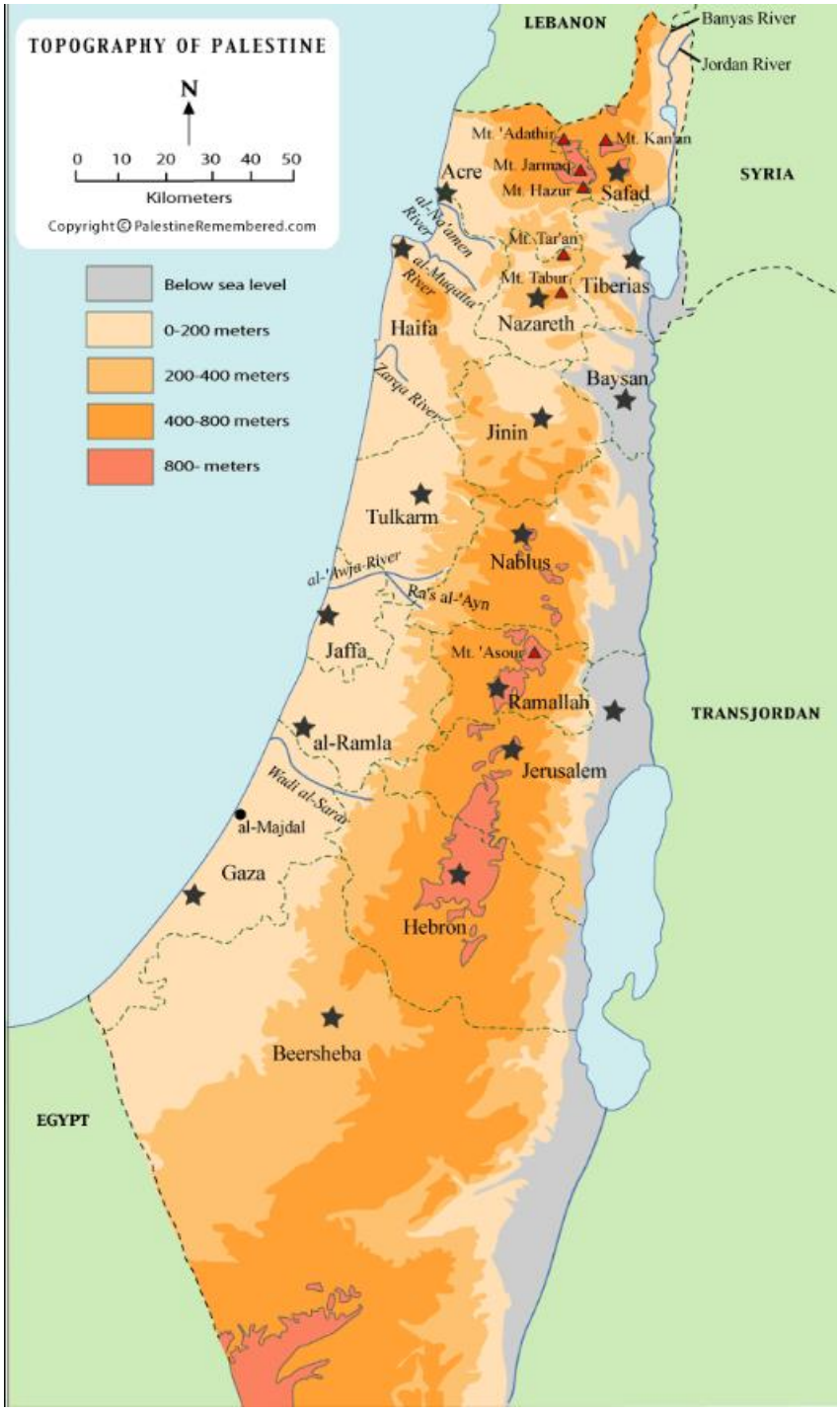


Figure.4: The historical map of Palestine.
(https://www.palestineremembered.com/Maps/index.html)

1.3.4 Literature review on genetic diversity research using STRs

Genetic variation degrees between population was extensively studied in some global projects. The HapMap Project was established to study the genetic variation, distance, allele frequency, and genotype frequency among populations (Garte, 2010). Unluckily, it lacked significant data about the Middle East. Moreover, the 1000 Genome Project measured allele frequency from North America, Europe, East, and South Asia, and West Africa (Kumar et al., 2017, Kanavakis & Xaidara, 1995). Such projects by excluding the Middle East, resulted in a gap in understanding the evolution, genetic variation, and migration of modern humans because the Middle East is the crossroad of migration and trade. Even though there are some studies performed on some of the populations in the Middle East, however, not every country has been analyzed and compared to each other. In this section, we will go through some of the knowledge gained from the Middle East and worldwide conducted using mainly STRs.

A population genetic structure study was performed in Gaza Strip-Palestine in 2008 using 15 autosomal STR loci. This study established some of the first essential information about allele frequencies in the Palestinian population. In this study, 125 unrelated individuals from Gaza strip were included. PowerPlex® 16 PCR was used to generate these data. They found no departure from Hardy-Weinberg (HWE) or linkage equilibrium (LD). In addition, they were able to prove that Palestinians in Gaza strip are randomly mating with very high heterozygosity percentages (Abu Halima et al., 2009a). In another study, analysis of 21 STRs in Saudi Arabia population revealed that the population structure is influenced heavily by consanguinity. Using GlobalFiler® kit, twenty-one autosomal STRs were analyzed in 523 males recruited from five different locations in Saudi Arabia. In this study, significant differences were found among regions. Deviation from Hardy-Weinberg equilibrium was detected, with a deficiency in heterozygosity approximately in all loci, and D2S1338 showed the

largest deviation (Khubrani et al., 2019). Allelic frequencies of 12 X-STRs were also studied in 501 males from the United Arab Emirates by using the Investigator Argus© X-12 QS kit. The study was used to report allele frequency, haplotype diversity, and forensic statistical parameters of the tested STRs. They also analyzed haplotypes of the four previously defined linkage groups. No shared profile was observed, and significant linkage disequilibrium was detected only within the four linkage groups. Moreover, they identified 20 distinct off-ladder alleles in seven different loci. In addition, their analysis revealed that the 12 loci are informative and discriminatory in the Emirati population (Almarri & Lootah, 2018).

Research conducted in Central Sahel was concerned about the effect of ethnic fragmentation and degree of urbanization on the discrimination ability of Y-STRs. Samples from 431 males were analyzed by using Yfiler© Plus multiplex and 82 Y-SNPs. In brief, their findings showed that reduced intra-population diversity and high inter-population differentiation in the central Sahel have some consequences on forensic applications. The level of differentiation between villages is so high to make nontrivial (if not impossible) to define an adequate reference population database to be used for statistical evaluation of Y-haplotype matches for crimes committed in rural areas (Della Rocca et al., 2020). In another study that was conducted in Jordan, they aimed to determine the efficiency of the PowerPlex® Fusion system. Blood samples from 500 individuals were collected and genotyped for 22 autosomal STRs amplified by PowerPlex® Fusion assay. The study findings confirmed that these 22 loci are useful for forensic and paternity testing, and that the Jordanian population is in Hardy-Weinberg equilibrium (HWE) after Bonferroni correction. Penta E and TPOX loci were the highest and lowest informative loci with power of discrimination of 0.99 and 0.834 respectively (Al-Eitan & Tubaishat, 2018). Also in Iraq, the Iraqi population was analyzed for 15 autosomal STRs variations in six cities. 1061 samples collected from the cities of Anbar, Baghdad, Basra, Diyala, Najaf, and

Wasit were genotyped. It was found that Iraq is similar to neighboring countries in the Middle East especially Iran and Turkey with a closer resemblance to Europe compared to Asia or Africa. They concluded that the analyzed loci are useful for forensic and paternity testing (Al-Zubaidi, 2022).

In addition to the studies conducted on Middle Eastern populations, other studies were conducted on different South American populations. In a study that targeted the eastern Paraguayan population, 537 blood samples were analyzed for 23 Y-STRs. Some of the analyzed loci showed unusual PowerPlex©Y23 (PPY23) profiles. These results were confirmed with the Y filer plus kit. For example, DYS385 locus showed a different allelic pattern (Biallelic and triallelic patterns). Also, a null allele pattern was observed at locus DYS448. Moreover, a new allele (10.3) at DYS358 was detected. This new allele was not detected at the Y-STR Haplotype Reference Database (YHRD). Furthermore, one sample showed null alleles at locus DYS456 and additional three alleles at locus DYS385. Haplotype diversity and differentiation analysis showed that there is homogeneity through parental ancestry. They used pairwise genetic analysis to measure the distance between populations through measuring (FST). Their results proved that Paraguay is closer to the European population than Africans or Native Americans. Moreover, differences were documented between Paraguay and other populations such as Panama, Ecuador, and Bolivia due to high paternal ancestry who are native Americans(Ribeiro et al., 2018). Also in Southern America, allele variations were reported for 23 Y-STRs in three Brazilian populations. 407 father-son samples were analyzed. In addition to the genotypes generated, mutations were analyzed for the Y-STRs using PPY23. In this population, haplotype diversity was found to be high (Ambrosio et al., 2020).

In Africa, the allele frequencies for 23 Y-chromosomal STRs for the South African population were recorded. This study aimed to generate Y-STR frequency data from South African individuals, in four population groups, for use in human identification by the local forensic community. Gene diversity

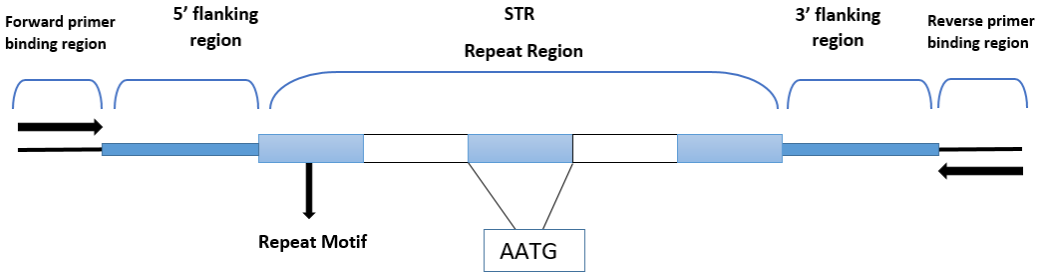
(GD) varied between the four population groups, with the DYS385a/b combined marker having the greatest diversity across all population groups. This combined marker had the lowest GD within the South African European population compared to the other three groups (Reid & Heathfield, 2020).

In a study that aimed to analyze the Y-chromosome allele variation of three major ethno-linguistic subpopulations in the Republic of North Macedonia. 314 individuals including Macedonians, Albanians, and Turks were studied using Y-SNPs and Y-STRs through mini-sequencing and fragment analysis. In their population study, they were able to show that only the Macedonian speakers could be completely resolved (identified), whereas the Albanian and Turkish speakers have still identical 27-loci haplotypes (Jankova et al., 2019). A study was conducted in the Han population from Jining Shandong province, eastern China to study the genetic variation and structure of 27 Y-STR loci, they observed 798 distinct haplotypes. Among these, 739 were unique, 45 were observed twice, nine were shared among three individuals, four were detected four times, and only one was found in five individuals. As expected, the Yfiler© Plus system indicated higher genetic diversity (Wang et al., 2019).

1.4 Short Tandem Repeats (STRs)

STRs have been widely used in forensic testing or human identification testing (HID), paternity testing, population genetic studies, and disease diagnosis. STRs density differs throughout the human genome. For example, chromosome 19 has the highest density among all chromosomes with a rough estimation of one STR locus per 2,000 base pairs (bp) (Fan & Chu, 2007). STRs are diverse between individuals and populations, and such diversity and polymorphism content is essential in forensic testing per the needs of powerful discrimination capacity and lower match probability (Mohammad & Imad, 2013, Preet et al., 2016). STR sequences contain 5’ and 3’ flanking sequences and the STR repeat region including the repeat motif, and sometimes, this repeat region could include intervening

sequences (Fig.5). In this section, we will go over characteristics, types, advantages, and factors that complicate STRs interpenetration, including repeat slippage and stutter formation, non-templet addition, microvariant alleles, null alleles, and copy number variants (CNVs).



Figrue.5: Schematic illustration of STR locus structure with AATG tetranucleotide as an example.

1.4.1 Characteristics and types of STRs

STRs are identified by the type and length of the repeat (Graham et al., 2015) (Table.1). Also, STRs can be classified according to their characteristics and repeat pattern into simple repeats, non-consensus simple repeats, compound repeats, and complex repeats (Butler, 2014) (Table.2).

Table.1: STRs types based on their repeat length.

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

Table.2: STRs classification based on their repeat construction and pattern.

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

1.4.2 Advantages of STRs

The evolution of STRs genotyping was accompanied by powerful advantages, the ability to amplify and genotype very minute quantities of DNA (0.1-1 ng) from intact, old, or degraded biological samples was of substantial importance in forensic science (Jordan & Mills, 2021, Peter Gill et al., 1985). Moreover, STR genotyping it is much faster compared to the traditional RFLP method. This is due to the small size of the repeats (100 bp to 500 bp) and the use of fluorescence techniques, which can produce results within a day compared to the 6-8 weeks needed with RFLP (Jordan & Mills, 2021). In addition, STRs with the current databases hold a high degree of discrimination with global kits used worldwide, and such discrimination power can reach one in one billion. Moreover, the development of multiplex amplification which is used to amplify many loci in a single reaction is heavily employed in the forensic field because it also decreases the time, labor, and costs required to generate a single profile (Lins et al., 2018).

1.4.3 Factors that complicate STRs interpretation

There are many factors that complicate STRs interpretation including stutter formation, non-templet addition, microvariants, and null alleles. These factors are discussed below.

1.4.3.1 Repeat slippage and stutter formation

Stutters are artefacts produced during PCR amplification. They result from replication slippage which results in the deletion or addition of repeat units, and therefore, it can cause expansion or contraction of the produced product due to loop formation. If the loop is formed on the nascent strand then it will cause expansion, but if the loop is formed on the templet strand then it will cause contraction (Fig.6). Stutter are identified as amplicons having a different length in comparison to the true alleles present in a biological sample. They usually differ by one or more repeat unit (Brookes et al., 2012). Tetranucleotide repeats are associated with artefacts that are roughly 15% of the true allele present in a locus. However, Dinucleotide and trinucleotide repeats can result in stutters that are 30% of the allele present in a locus (Butler, 2014). It is worth noting that more than 40 disorders are associated with repeat expansion disorders of which mostly are neurological (La Spada et al., 1994, Paulson, 2018).

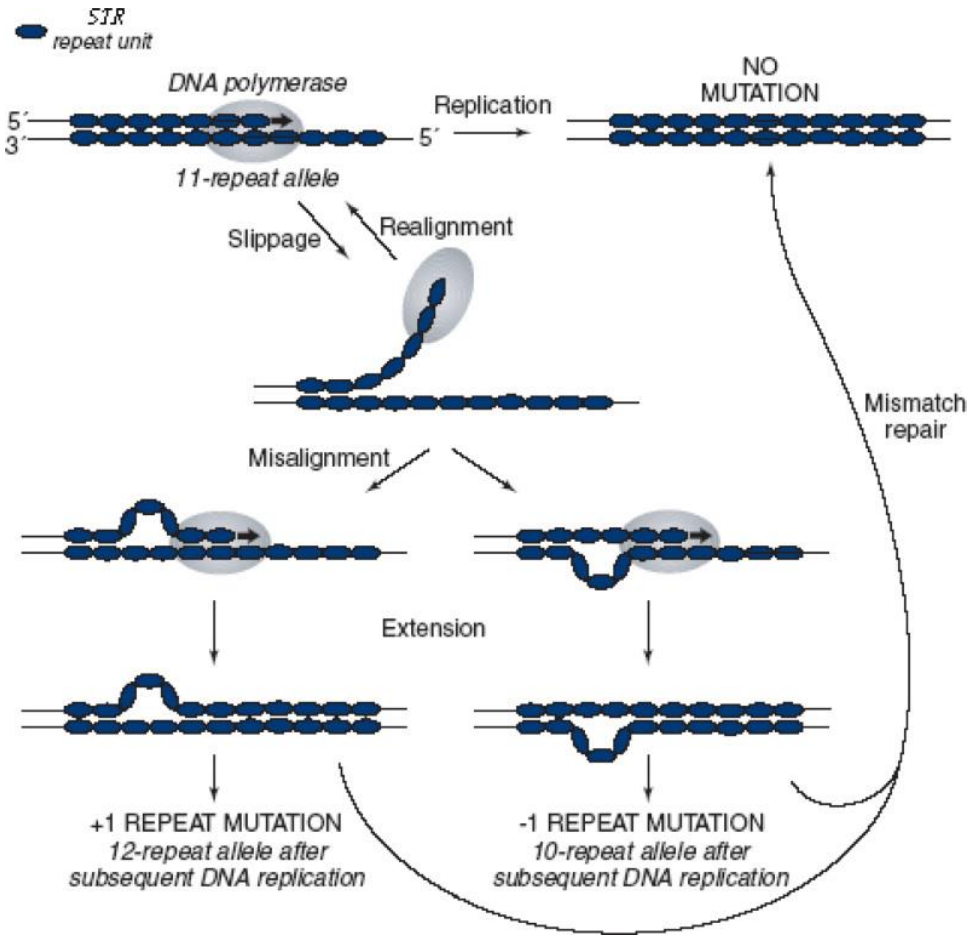


Figure.6: Schematic illustration of the strand-slippage replication at STR locus(Fan & Chu, 2007)**.**

1.4.3.2 Non-template addition

Non-template addition occurs when Taq DNA polymerase adds a nucleotide, generally adenine, to the 3´ ends of amplified DNA fragments in a template-independent manner which results in the appearance of a false peak that is one nucleotide different than the true allele. Some primer sequences are more prone to this condition compared to others (J. R. Smith et al., 1995). This phenomenon can be avoided by adding a final extension step, in the amplification program, at 60°C or 72°C for 30 or 45 seconds (Kimpton et al., 1993).

1.4.3.3 Microvariant alleles

Alleles that differ from a usual or common full-length allele by one or more base pairs are usually rare in human populations. The mechanism behind this phenomenon is either insertion or deletion mutations (Fig.7). Microvariants complicate the assigning process of alleles by software because usually, they do not size the same as the commonly known alleles present in the allelic ladder. Thus, they are assigned by the software as "off-ladder" alleles(Lang et al., 2019).

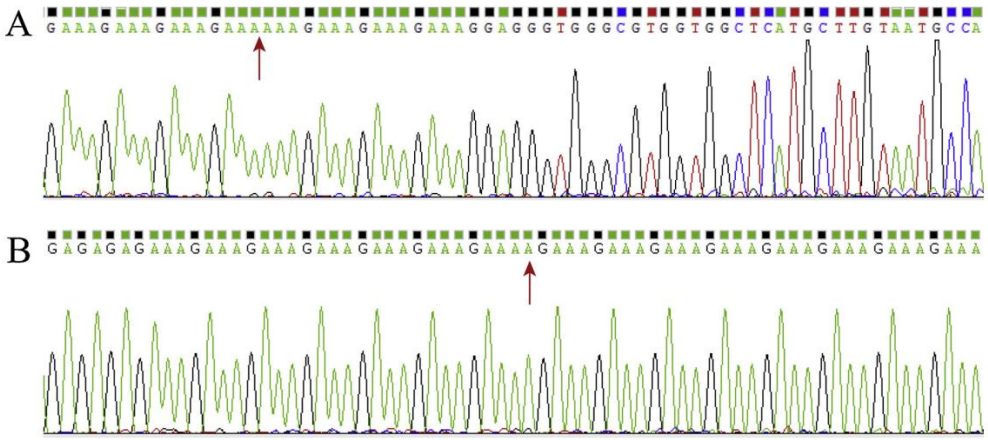


Figure.7: Sequencing results of intermediate or microvariant allele at DYS458 and DYS627. A) The sequence of DYS458 showing a lost G (the red arrow). B) DYS627 sequence showing A insertion (blue arrow) (Lang et al., 2019).

1.4.3.4 Null Alleles due to allele dropout and chromosomal region deletions

Polymorphisms with neutral biological and physiological effects are widely distributed and detected throughout the human genome and STR loci are not exceptions. In STR genotyping, missing alleles can occur due to deletions or polymorphisms encompassing the binding sites of primers used to genotype a marker of interest. These genetic variations results in amplification failure and thus allele dropout (Crooks et al., 2013) (Fig.8). Such silent alleles can be revealed with concordance studies (Dion et al., 2011).

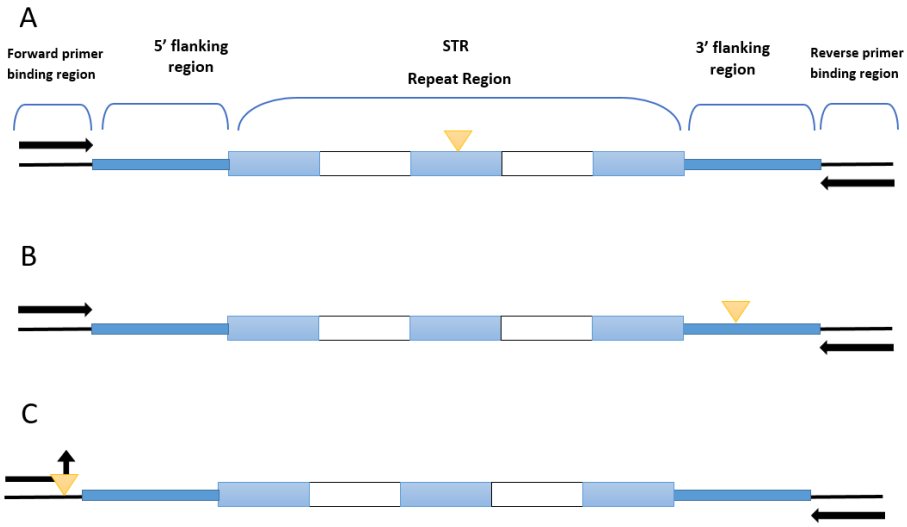


Figure.8: The effect of STR locus sequence mutation on STR genotyping. A) A mutation occurring within the repeat region does not affect the amplification, however, it results in the production of a different genotype. B) A mutation occurring in the flanking region does not affect amplification however it produces a different genotype. C) A mutation occurring at the 3'end of the forward primer (or the reverse primer) results in amplification failure and thus, silent allele.

1.4.3.5 Copy number variations (CNVs)

Occasionally, duplication events are observed in the genome. These duplication events occur due to either a duplication of DNA segment or a total chromosome duplication. A common example of the later condition is Down syndrome or trisomy 21. D21S11 locus and other STR loci have been used extensively in the diagnosis of trisomy 21(Yoon et al., 2002). Duplication can result in tri-allelic pattern observed at a specific autosomal locus or more in an electropherogram of a single source sample. In addition, they can result in duplication or triplication patterns on the Y and X chromosomes. These tri-allelic patterns usually fall into two types. Type 1, which is the most common type, presents with trialleles where in electropherogram analysis, the sum of two alleles peak heights equals to that of the third allele. Type 2 presents three alleles with relatively equal peak heights. Until March 2017, STRBase documented 401 different tri-allelic patterns observed at different loci (https://strbase.nist.gov/tri_tab.htm).

1.5 Thesis significance, statement, and aims

In the present study we tried to elucidate the genomic landscape variation of the Palestinian population in respect to both the autosomal and Y-chromosome STRs. This study elaborated on the significance of conducting population genetic research specific to each population. It shed the light on the Palestinian population genetic diversity, population relationship, and genetic distance from neighboring populations in the Middle East and more distant populations worldwide. Moreover, it analyzed the population genetic structure in the total population, within and among subpopulations that exist in the West Bank. The aims of this study were focused on analyzing the genetic diversity of the Palestinian population and generating allele frequency databases of 23 autosomal and 23 Y-chromosome STRs using PowerPlex© Fusion 6C and PowerPlex© Y23, respectively. We also assessed the forensic efficiency parameters of these systems. Moreover, we analyzed the Palestinian population's genetic structure including isolated subpopulations such as the Samaritans. In addition, this study highlighted the effect of including more loci to increase the discrimination capacity, which is of great importance in forensic testing performance.

Chapter 2

Methodology

2.1 Ethical approval and consent to participate

Ethical approval was obtained from the graduate studies at the university. Moreover, informed consent was obtained from all subjects before sampling. All methods were performed in accordance with the relevant guidelines and regulations of the AAUP ethical committee.

2.2 Sample collection

Buccal swabs were collected from 647 randomly chosen unrelated volunteers. Subjects declared their nationality, ethnicity, religion, and language. Inclusion criteria included age older than eighteen years, holding Palestinian nationality, and their ancestors had lived in Palestine for at least three successive generations. Samples were collected from all eleven provinces in the West Bank.

2.3 DNA Extraction

DNA was extracted from 647 Buccal swabs using SwabSolution® kit (Promega, USA) according to the manufacturer recommendations. In brief, 2 ml of SwabSolution was added to each swab and incubated at 70 °C for 30 minutes using a heat block (Labnet, USA).

2.4 STRs DNA Amplification

Twenty-three autosomal STRs and twenty-three Y-STRs were directly amplified using PowerPlex® Fusion 6C and PowerPlex® Y23 (Promega, USA) according to the manufacturer recommendations. For PowerPlex® fusion 6C, we amplified the DNA in 12.5µl reaction that included (2.5µl AmpSolution, 2.5µl Master Mix, 2.5µl Primers, 2µl swab extract, and 3µl molecular grade water). Afterward, samples were amplified by using Biometra Tone thermal cycler. The amplification program was as follows; 96°C for 1 minute, then 26 cycles of 96°C for 5 seconds, and 60°C for 1 minute followed by 10 minutes of incubation at 60°C. For PowerPlex® Y23 (Promega, USA), we amplified the DNA in 12.5µl reaction mix that included (2.5µl AmpSolution, 2.5µl Master Mix, 1.25µl Primers, 2µl swab extract, and 4.25µl molecular grade water) samples were amplified by using Biometra Tone thermal cycler. The amplification program was as follows: 96°C for 2 minutes, then 26 cycles of 96°C for 10 seconds, 61°C for 1 minute, and 72°C for 30 seconds, followed by 10 minutes of incubation at 60°C.

2.5 Genotyping using capillary electrophoresis (CE)

2.5.1 Spectral calibration

In order to be able to detect fragments on ABI 3500 genetic analyzer (Applied Biosystem, Foster City, CA), we performed spectral calibration using 5C matrix for PowerPlex® Y23 and 6C matrix for PowerPlex® Fusion 6C to correct for spectral overlap. Either 10µl of the 6C or 5C Matrix Mixes was added to a tube of Matrix Dilution Buffer. Then, 10µl of the diluted Matrix Mix was added to 500µl of Hi-Di™ formamide. 15µl of this Matrix Mix mixed with formamide was then added to the first eight wells in the 96-plate and placed in the ABI 3500 genetic analyzer. All injections passed the spectral calibration.

2.5.2 Fragment analysis

Amplified samples were prepared for fragment analysis for both kits per the manufacturer recommendation. A cocktail was prepared by mixing (0.5 µl WEN ILS, 9.5µl Hi-Di^{TM} formamide). Then, 1 to 1.5 µl of amplified samples were added to each well containing the cocktail mix. Afterwards, plates were denatured using a heat block for three minutes and then placed immediately on ice for three minutes per manufacturer recommendations. Fragment analysis was performed using ABI 3500 genetic analyzer, using 50 cm capillaries and Pop-7. The injection voltage was set at 1.6 kVolts and the injection time was set for 8 seconds.

2.6 Electropherograms analysis and interpretation

For quality check reasons fragment analysis results were analyzed using both GeneMapper version 5 (Thermo Fisher Scientific) and GeneMarker®HID Software, Version 3.0.0 (Promega, USA) using the local southern analysis method. All alleles were designated according to the guidelines and nomenclature of the International Society for Forensic Genetics (ISFG) for STR interpretation. Electropherograms were interpreted and analyzed by identifying peaks at the loci that were amplified using PowerPlex® Y23 and PowerPlex® Fusion 6C. Sizes of fragments were determined through comparison to 500 WEN internal lane standard (WEN 500 ILS) (Promega, USA). In order to identify alleles, all peaks were compared to an allelic ladder. A universal minimum analytical threshold (AT) of 50 Relative Fluorescence Units (RFU) was applied to reduce the back noise and other possible artifacts. Pull-up correction, spike removal, and saturation detection icons were selected in the run wizard.

Non-allelic peaks such as background noise or stutter were removed according to the AT and stutter filters per Y23 and PowerPlex® Fusion 6C technical manuals recommendations for stutter filter

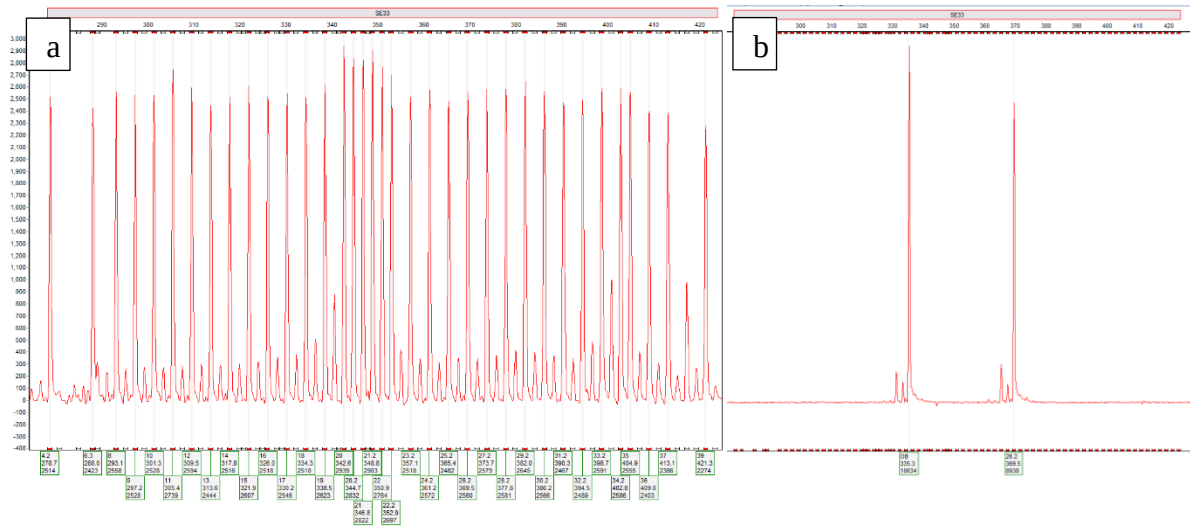
application. The following stutter behavior was considered during the interpretation of electropherograms. Peaks that are one repeat smaller or larger than the main allele identify Back and forward stutter. Two repeat units smaller than the main allele identify double back stutter. We took into consideration that some tetranucleotide loci might have halfback stutter, which is two base pairs (bp) smaller. Finally, a 20% global max filter was applied.

Pull-up peaks if occurred due to oversaturation were treated by diluting the PCR product and re-injecting the sample in case the sizing failed.

Spikes that occur due to power surges, crystals, or air bubbles were characterized by a narrow peak width and are present usually in all dye colors at the same location. If this occurred, they were identified by looking at all dye colors including the ILS, and by looking at other samples in the same run.

New variants (microvariant alleles) can occur and the software labels them as “Off-ladder” alleles. These alleles are not present in the allelic ladder as they are usually rare variants. However, it is worth noting that these could occur due to a migration shift. We examined the off-ladder alleles in comparison to the allelic ladder in the same run and other alleles in the same locus, if it was less than 0.6 bp (not at least a full nucleotide); we considered it a migration shift. Off-ladder allele designation was calculated according to Butler’s textbook (Advanced Topics in Forensic DNA Typing: Interpretation) (Butler, 2014, P. Gill et al., 1996). For example, a microvariant is resolved as 18.1 for the SE33 locus, if the off-ladder allele size is 335.3 bp and the allelic ladder 18.2 allele size is 336.3

bp and the other allele size in the same locus, which is allele 26.2, is 369.5 bp in size, which is also in the ladder with the size of 369.5(Fig.9).



$\delta_1 = S_{26.2} - L_{26.2} = 369.5 - 369.5 = 0$

$\delta_2 = S_{OB} - L_{18.2} = 335.3 - 336.3 = -1$

$C = |\delta_1 - \delta_2| = |0 - -1| = 1$ nucleotide

Figrue.9: Microvariant detection and analysis using PowerPlex Fusion® 6C, capillary electrophoresis, and GeneMarker® HID software. a) Allelic ladder of SE33 locus. b) Anonymous sample showing the diallelic pattern for the SE33 locus with Out-of-Bin (OB) or off-ladder microvariant. Peak labels indicate allele designation, RFU intensity, and calculated fragments sizes using the local southern method and WEN 500 ILS.

2.7 Statistical Analysis

2.7.1 Autosomal STRs

STRAF version 2.1.5 (Gouy, A., & Zieger, 2017) (https://straf-p7bdrhm3xq-ew.a.run.app/) was used to calculate population genetics parameters. These include polymorphism information content (PIC), observed heterozygosity (Ho), expected heterozygosity, probability of exclusion (PE), and power of discrimination (PD). Arlequin version 3.5(L. and H. E. L. L. Excoffier, 2010) was used to calculate allele frequencies. F-Statistics was conducted and included measuring inbreeding-like effect or coefficient among subpopulations (F_{ST} or Θ), within populations (F_{IS} or f), and within the entire

population (F_{IT} or F) using Genepop software version 4.7(Raymond M. & Rousset F, 1995) (https://genepop.curtin.edu.au/).

The following formulas were used to calculate these various population genetics parameters:

2.7.1.1 Hardy-Weinberg equation

This method calculates the expected allele frequency from the observed allele frequency.

$$p^2+2pq+q^2=1$$

Where p is the frequency of the dominant allele in a population, 2pq is the heterozygous frequency, and q is the recessive allele in a population.

2.7.1.2 HWE Exact Test

This test calculates any deviations from Hardy-Weinberg Equilibrium (HWE) which states that allele frequencies are constant through generations when random mating is occurring. The exact test is always conservative and better to implement when having a smaller cohort to avoid a type I error rate (Wigginton et al., 2005). The test computes the p-value based on 1,000 Monte Carlo permutations of alleles. By consensus, a p-value lower than 0.05 is considered significant and indicates deviation for HWE. In case any locus result was less than 0.05, a Bonferroni correction should be applied which states that the p-value (0.05) should be divided by the number of comparisons (α/n). Where α is the critical p-value and n is the number of comparisons performed in the study. The HWE exact test depends on the distribution of heterozygotes (NAB) depending on the minor allele N^A frequency (Graffelman & Weir, 2016).

$$P(N_{AB} = n_{AB}|N, n_A) = \frac{2^{n_{AB}} N!}{n_{AA}! n_{AB}! n_{BB}!} \times \frac{n_A! n_B!}{(2N)!}$$

2.7.1.3 F-statistics

A) F_{ST} : Fixation Index

This test measures genetic differentiation among subpopulations and it indicates a reduction in heterozygosity due to divergence in allele frequency. It is always positive. If the F_{ST} result is zero then there are no genetic differentiations among subpopulations. If F_{ST} equals one, then there is strong genetic differentiation among subpopulations.

$$F_{ST} = H_T - H_S / H_T$$

Where H_T is the total population heterozygosity and H_S is heterozygosity among subpopulations.

B) F_{IT}: Overall Fixation Index

This test measures deviation from HWE in the total population, it indicates the correlation between two alleles in a genotype sampled at random from a single subpopulation given the possibility of non-random mating within populations and allele frequency divergence among populations. If the F_{IT} result is zero then there is no deviation from HWE in the total population. If positive: Deviation exists due to deficiency of heterozygotes in the total population. If negative: Deviation exists due to an excess of heterozygotes the in total population.

$$F_{IT} = H_T - H_I / H_T$$

Where H_T is the total population heterozygosity and H_I is the individual heterozygosity.

C) F_{IS}: Inbreeding Co-Efficient

This test measures deviation from HWE within each subpopulation due to non-random mating. If it is zero then there is no deviation from HWE within subpopulations. If it is positive then there is deviation due to a deficiency of heterozygotes within the subpopulation. If it is negative then there is deviation due to an excess of heterozygotes within the subpopulation.

$$F_{IS}= H_S\text{-}H_I/H_S$$

Where H_I is the individual heterozygosity and H_S is heterozygosity within subpopulations.

2.7.1.4 Expected Heterozygosity (H_e) or Gene Diversity (GD)

This type of analysis is used to calculate an estimation of expected heterozygosity. Higher heterozygosity is an indication of high gene diversity(Edwards et al., 1992). Expected Heterozygosity (H_e) is calculated according to the following equation.

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

Where n is the allele count for specific alleles at a locus in a sample taken from a certain population, while pj is the allele frequency.

2.7.1.5 Power of Exclusion (PE)

Also called the probability of exclusion (Fisher, 1951). It measures the efficiency of the loci tested in excluding an unrelated subject during paternity investigations (Plodthong et al., 2014). It can be calculated using the following formula

$$PE= Ho^2(1\text{-}2Ho(1\text{-}Ho)^2)$$

Where Ho is the observed heterozygosity.

2.7.1.6 Power of Discrimination (PD)

Also known as the probability of discrimination (Fisher, 1951). It can be calculated according to the following formula.

$$PD = 1 - P_I$$

Where P_I is the probability of identity.

2.7.1.7 Probability of Identity (P_I) or Matching Probability (pM)

It measures the probability of having an identical genotype at specifically tested loci when two subjects are selected randomly(Sensabaugh, 1982, Jones, 1972). It was calculated as follows:

$$pM = \sum_{k=1}^{m} p_k^2$$

Where P_k is the frequency of every distinct genotype (k), while m is the number of distinct genotypes.

2.7.1.8 Polymorphism information content (PIC)

Also called the power of information content. Usually, it is used for linkage studies and it reflects the ability to infer the parental genotype if the child has a rare allele at a specific locus (Guo & Elston, 1999). It is calculated according to the following equation.

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

Where *pi* is the population frequency of the i allele.

2.7.1.9 Paternity Index (PI)

Is the probability of a child's genotype supporting the assumption of the tested parent is actually the biological parent and not any unrelated subject chosen by random. It is calculated according to the following formula.

$$TPI = \frac{1}{2H}$$

Where Pi is the frequency of the I allele in the number of samples (n) in the population tested;

H =homozygosity.

2.7.1.10 Combined Paternity Index (CPI)

The combined Paternity index is calculated by multiplying the individual PIs of the tested loci.

2.7.1.11 Probability of Paternity (POP)

POP tests the probability of the alleged father passing the genotype of a specific STR locus compared to the possibility of a random person passing this genotype (Chakraborty & Stivers, 1996).

The probability of paternity can be calculated from the following formula

Probability of Paternity (POP): CPI/CPI+1

2.7.2 Y-STRs analysis

2.7.2.1 Allelic frequencies

They were calculated using Excel and they were estimated by the direct counting method.

2.7.2.2 Gene Diversity

Nei's formula for calculating gene diversity for each locus was implemented.

$$GD= 1-\sum p_i^2$$

Where p_i is the frequency of the i^{th} allele.

2.7.2.3 Haplotype Diversity

Haplotype diversity was calculated using the following formula:

$$HD=n\ (1-\sum p_i^2)/\ (n-1)$$

Where n is the sample size and p_i is the frequency of the i^{th} haplotype.

2.7.2.4 Haplotype Match probability (HMP)

Haplotype match probability was calculated using the following formula:

$$HMP= \sum p_i^2$$

Where p_i is the frequency of the i^{th} haplotype.

2.7.2.5 Discrimination Capacity

Discrimination capacity was calculated by dividing the total number of observed haplotypes by the total number of individuals in the study population, according to the following formula:

$$DC= H/N$$

H: observed number of haplotypes. N equals number of individuals.

2.8 Quality control and performance check

In order to validate the usage of 50cm capillary and 7-pop with the used kits on the 3500 genetic analyzer, two samples were genotyped in a forensically accredited laboratory and results were compared to our genotypes at the university's laboratory. Genotypes showed a complete match. Moreover, the 2800 positive control that was included in both kits was amplified multiple times and with each run to check for the performance of the instrument. Genotypes were as expected and reproducible each time. More than one allelic ladder was included in each run; all ladders passed the quality check of the analysis software. Furthermore, allele calls were revised for each ladder to make sure they were correctly assigned as expected. Allelic ladders are an effective method of assessing the adequacy of the system's analytical performance because they contain abundant DNA fragments, which can adequately assess the spectral and spatial resolution of the system. For instance, our system was able to show single-base resolution between ladder alleles, which are different from each other in a single nucleotide sequence such as allele 9.3 and allele10 in TH01. Thus, the system was considered reliable for high-resolution allele designation(Butler, 2014). WEN Internal lane standard (ILS) 500 was included in each sample, whether it is a real sample, negative control, positive control, or allelic ladder were revised to make sure that fragment sizing was done correctly. For all samples, the sizing quality score ranged from 96-99% match. Molecular grade water was included in each run as negative control and we never encountered contamination of the negative control. This indicates that contamination measures were followed at best. Loci that were amplified in PowerPlex® Fusion 6C and PowerPlex® Y23 for the same sample were compared when applicable to check for the performance of the systems used.

2.9 Anomalous samples processing

2.9.1 Primer design

First, primers were picked for PCR and Sanger sequencing based on previous primer sequences published on STRBase or other literature reviews. UCSC *In Silico* PCR tool (https://genome.ucsc.edu/cgi-bin/hgPcr) was used to confirm the specificity of the primers. Primers used are listed below.

Table.3: Forward and reverse primer sequences for autosomal and Y-chromosome STRs

#	Locus	Primer sequence	Product size range (bp)	Ref
1	Penta D	FW 5'- TGGAAGGTCGAAGCTGAAGT-3' REV 5'-ATTAGAATTCTTTAATCTGGACACAAG-3'	370-454	STRBase
2	Vwa	FW 5'-GCCCTAGTGGATGATAAGAATAATCAGTATGTG-3' REV 5'-GGACAGATGATAAATACATAGGATGGATGG-3'	123-183	STRBase
3	TH01	FW 5'-ATTCAAAGGGTATCTGGGCTCTGG-3' REV 5'-GTGGGCTGAAAAGCTCCCGATTAT-3'	171-215	STRBase
4	D21S11	FW 5'-ATATGTGAGTCAATTCCCCAAG-3' REV 5'-TGTATTAGTCAATGTTCTCCAG-3'	155-273	STRBase
5	SE33	FW 5'-AATCTGGGCGACAAGAGTGA-3' REV 5'-ACATCTCCCCTACCGCTATA-3'	197-343	STRBase
6	D16S539	FW 5'-GGGGGTCTAAGAGCTTGTAAAAAG-3' REV 5'-TTAGCGTTTGTGTGTGCATCT-3'	260-308	STRBase
7	Penta E	FW 5'-ATTACCAACATGAAAGGGTACCAATA-3' REV 5'-TGGGTTATTAATTGAGAAAACTCCTTACAATTT-3'	379-484	STRBase
8	DYS393	FW 5'-GTGGTCTTCTACTTGTGTCAATAC-3' REV 5'-GAACTCAAGTCCAAAAAATGAGG-3	108-140	STRBase
9	DYS385 a/b	FW 5'-AGCATGGGTGACAGAGCTA-3' REV 5'-GCCAATTACATAGTCCTCCTTTC -3'	242-326	STRBase
10	DYS392	FW 5'-TAGAGGCAGTCATCGCAGTG-3' REV 5'-GACCTACCAATCCCATTCCTT-3'	290-323	STRBase
11	DYS437	FW 5'-GACTATGGGCGTGAGTGCAT -3' REV 5'-GAGACCCTGTCATTCACAGATGA -3'	182-200	STRBase
12	DYS389 | &||	FW 5'-CCAACTCTCATCTGTATTATCTATG -3' REV 5'-GTTATCCCTGAGTAGTAGAAGAATG -3'	143-171/259-295	STRBase
13	DYS439	FW 5'-TCGAGTTGTTATGGTTTTAGGTCT-3' REV 5'-GTGGCTTGGAATTCTTTTACCC-3'	213-233	STRBase
14	DYS448	FW 5'-TGGGAGAGGCAAGGATCCAA-3' REV 5'-GTCATATTTCTGGCCGGTCTGG-3'	299-341	STRBase
15	Y-GATA H4	FW 5'-ATGCTGAGGAGAATTTCCAA-3' REV 5'-GCTATTCATCCATCTAATCTATCCATT-3'	122-142	STRBase
16	DYS 549	FW 5'-GTGTAAGCCAAACCCAAATATAGC-3' REV 5'-TGCTTAATATTAGTAGGTTGTTTGCAT-3'	372–412 bp	(Turrina et al., 2013)

17	DYS 456	FW 5'-CTGTTGTGGGACCTTGTGATA-3' REV 5'- ACTCAGCCCAAAACTTCTTAAA-3'	90–138	(Moon et al., 2022)
18	DYS576	FW 5'-GCGTATTTGTCTTGGCTTTTT-3' REV 5'-CATAGCAAGACCTCATCTCTGAA-3'	90–138	(Moon et al., 2022)

2.9.2 PCR

PCR was performed in 20µL reaction volumes consisting of 10 µL of GoTaq PCR master mix 2x (Promega, Madison, WI, USA), 1µL of DNA (100 ng), 1 µL of forward, and reverse primers (10 mmol each), and nuclease-free free water to complete the volume to 20 µL. Thermal cycling conditions were as follows: Initial denaturation for 5 min at 95°C, and 31 cycles of 30 sec at 95°C for denaturation, 30 sec at 58°C for annealing, then 30 sec at 72°C with a final extension of 5 min at 72°C.

2.9.3 Agarose gel electrophoresis.

To confirm PCR amplification success and specificity, agarose gel electrophoresis was carried out using 2.5% (w/v) agarose gel in TAE 1X. 50 µLs of (0.5 µg/mL) ethidium bromide (Hy labs, Jerusalem) were added to aid DNA visualization. Seven µl of PCR product were loaded into each well of the gel. 50 or 100 bp DNA ladder (GeneDirex, Hy.labs) was run alongside the samples at 120 mV for 30-50 minutes. PCR products were visualized by using a UV trans-illuminator.

2.9.4 Polyacrylamide gel electrophoresis (PAGE)

Polyacrylamide gel electrophoresis was used to confirm the presence of CNVs. PCR products were separated in 12% polyacrylamide gel that was made by mixing 3.6 ml acrylamide, 6 ml H_2O, 2.4 ml TAE 5X, 200µL of 10% APS, 10µL TEMED. The gel was run at 30V for 3hours. Then the gel was stained ethidium bromide and visualized and photographed using a UV trans-illuminator (Bio Rad, USA).

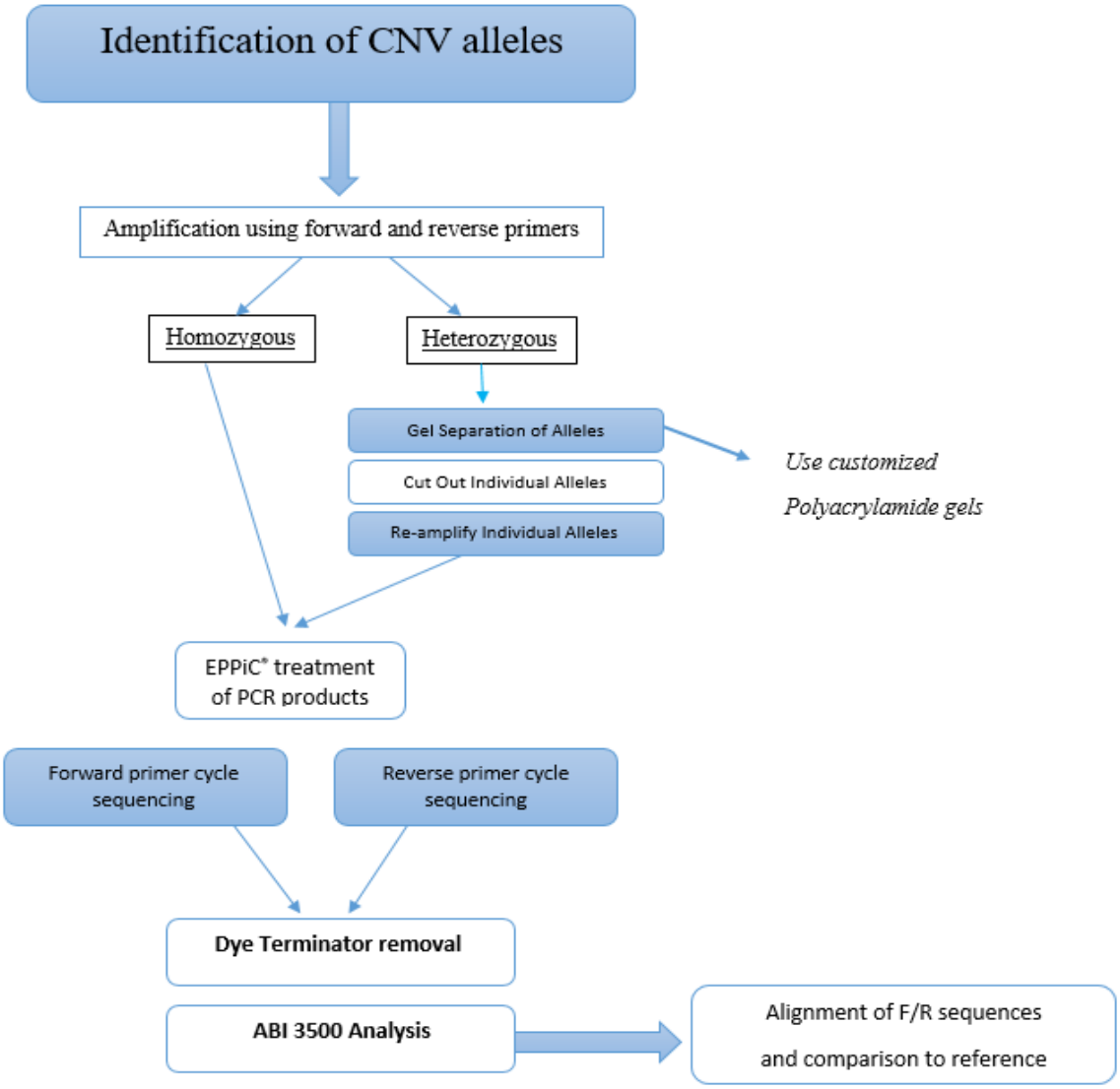
2.9.5 Sanger sequencing

Leftover primers and dNTPs were cleaned from PCR products using the EPPiC Fast kit (A&A Biotechnology). 5µl of cleaned PCR product was added to 1µl of the EPPiC Fast enzymatic solution. Then, the mixture was incubated at 37°C for 10 minutes followed by 1 minute at 80°C in a thermal cycler. Afterwards, the clean PCR product was amplified and labelled using the BigDye Terminator v3 kit (Applied Biosystems) according to the manufacturer's instructions. Afterwards, labelled DNA was cleaned using ethanol precipitation technique. Briefly, DNA product was mixed with 5 µl of 0.5 M EDTA and 60 µl of absolute Ethanol to each sample. Then the mixture was mixed by vortexing and then incubated at 4 °C for 12 min, and centrifuged at 2200 x g for 12 min. Afterwards, 80 µl of absolute ethanol was added and centrifuged for 15 min at 1600 x g. Finally, the samples were air dried for 15 minutes and 10 µl of formamide (Thermo Fisher Scientific, USA) were added to each samples. The samples were then boiled at 95 °C for 3 min in a dry heat bath and immediately transferred , to ice for 3 min. Finally, the sanger sequencing machine (Hitachi3500 Genetic Analyzer, Thermo Fisher Scientific, USA) was used to read the plate.

2.9.6 Sequence Analysis

The sequences of the PCR products for variant alleles, null alleles, and CNVs (Fig.10) were viewed and analyzed against reference sequences obtained from STRBase, reference sequences for most of the STR loci used are available at http://www.cstl.nist.gov/biotech/strbase/seq_ref.htm) and STRs sequencing project (Gettings et al., 2017), these sequences can be found at

(https://www.ncbi.nlm.nih.gov/bioproject/380127). Heterozygous genotypes for autosomal loci were analyzed using Poly Peak Parser (http://yosttools.genetics.utah.edu/PolyPeakParser/).



.

Fig. 10. Schematic illustration of the steps used in sequencing STR CNV allele.

Chapter 3

Results

3.1 Sample description and demographics

Out of the 647 samples genotyped, 28 samples failed PCR and/or had severe partial profiles. Thus, these samples were excluded from the study (fig.11A).

3.1.1 Gender, race, and ethnicity

39.3% of the samples were females (243 samples) and 60.7% of the samples were males (376 samples) (Fig.11A). 99.03% of samples were Caucasian (613 samples) and 0.97% of samples were Palestinians of African origin (6 samples) (Fig.11B). Finally, 95.6% of samples were from subjects born to Palestinian parents and grandparents (592 samples). However, 4.4% of samples were for admixture subjects (27 samples) with one of their parents or grandparents from a different nationality, ethnicity, and/or religion (Fig.11C).

3.1.2 Religion

540 samples (87.23%) were from Muslims, 62 samples (10.01%) were from Christians, 16 samples (2.6%) were from Samaritans, and one sample (0.16%) was from an atheist (Fig.12A). Out of the 62 Christian samples, 4 samples (6.5%) were Armenian, 3 samples (4.8%) were Syriac, 1 sample (1.6%) was Coptic, and 54 samples (87.1%) declared themselves as other (Meaning Christian Arabs) (Fig.12B). The atheist sample was combined with the Muslim samples for allele frequency analysis as he declared he was born for Muslim parents and Muslim ancestry.

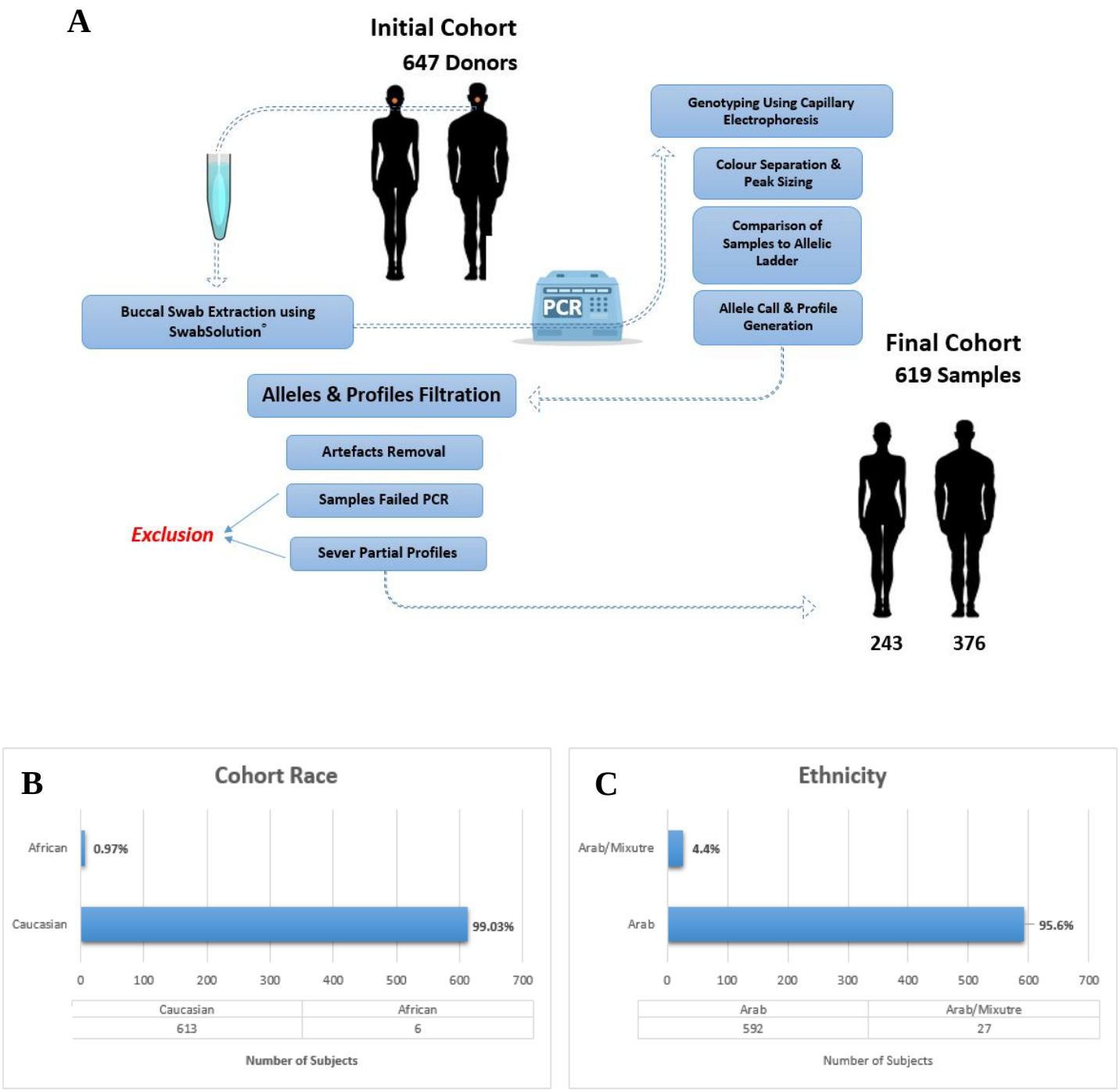


Figure.11: Study cohort, samples gender, race, and ethnicity. A) Initial and final cohort included in the study including the exclusion criteria. B) Race as self-declared by subject. C) Ethnicity as self-declared by subjects.

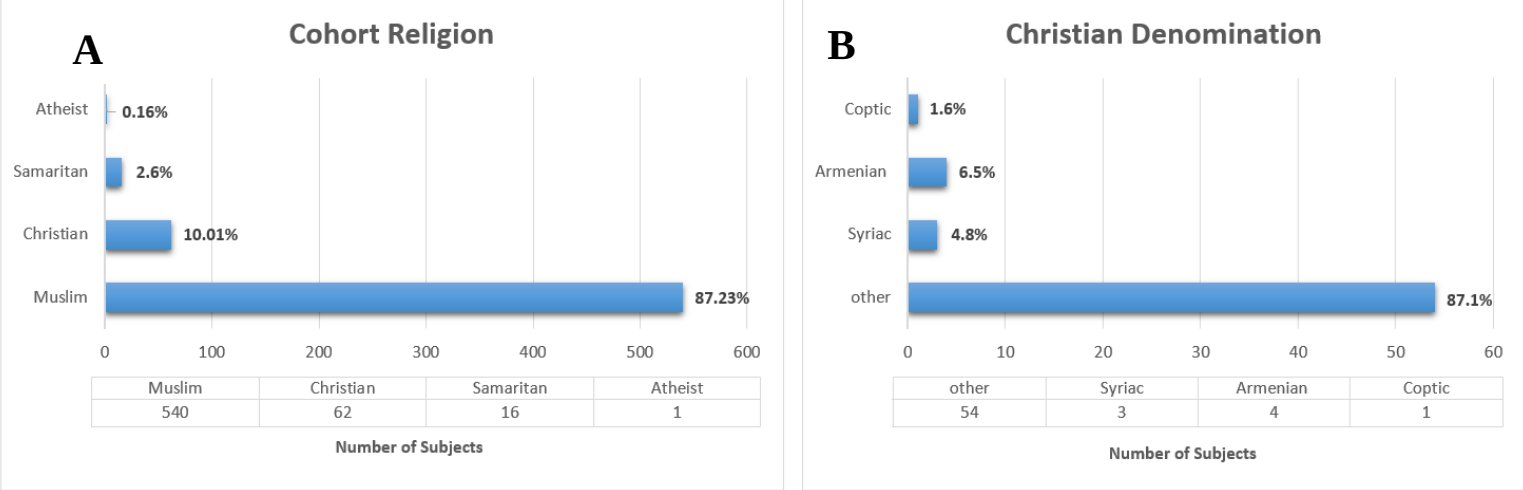


Figure.12: Subject religion and Christian Denomination. A) Subjects religion as self-declared by the subject. B) Christian Denomination as self-declared by subjects. Others in figure B refers to Christian Arabs.

3.1.3 Geographic distribution

Our 619 samples represented the 11 West Bank provinces and Gaza Strip. Samples distribution was as follows; 135 samples (21%) from Ramallah & Al-Bireh, 60 samples (10%) from Bethlehem, 114 samples (18%) from Hebron, 69 samples (11%) from Nablus, 6 samples (1%) from Tubas, 18 samples (3%) from Salfit, 11 samples (2%) from Jericho, 46 samples (7%) from Jenin, 64 samples (10%) from Jerusalem, 55 samples (9%) from Tulkarem, and 34 samples (6%) from Qalqylia. Moreover, we collected seven samples (1%) for subjects from Gaza Strip (Fig.13). The West Bank is divided into three main regions: middle, north, and south with the 11 provinces distributed throughout these regions. 210 samples (34.3%) were from the middle of the West Bank, 228 samples (37.3%) were from the north of the West Bank, and 174 samples (28.4%) were from the south of the West Bank (Fig.14).

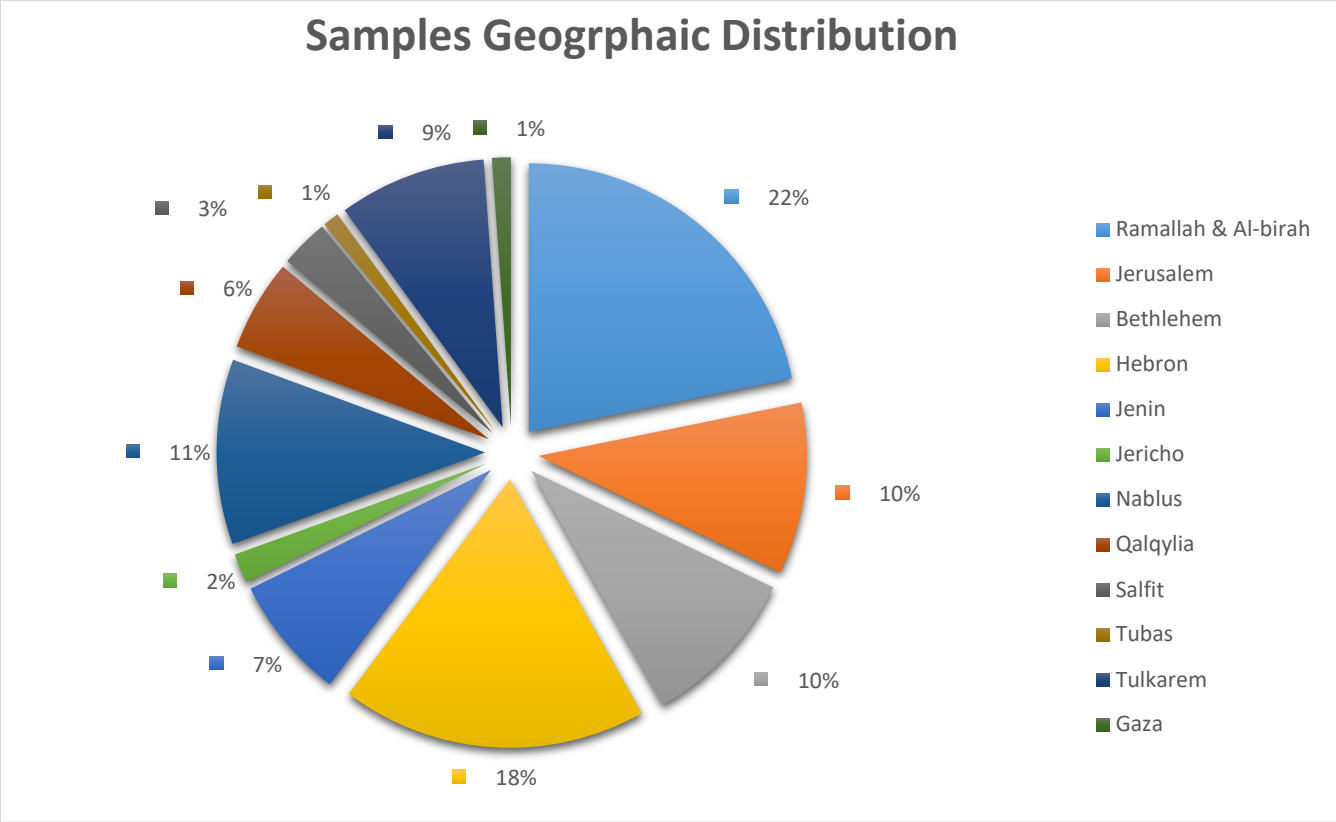


Figure.13: Sample geographic distribution throughout 11 provinces in the West Bank and Gaza Strip.

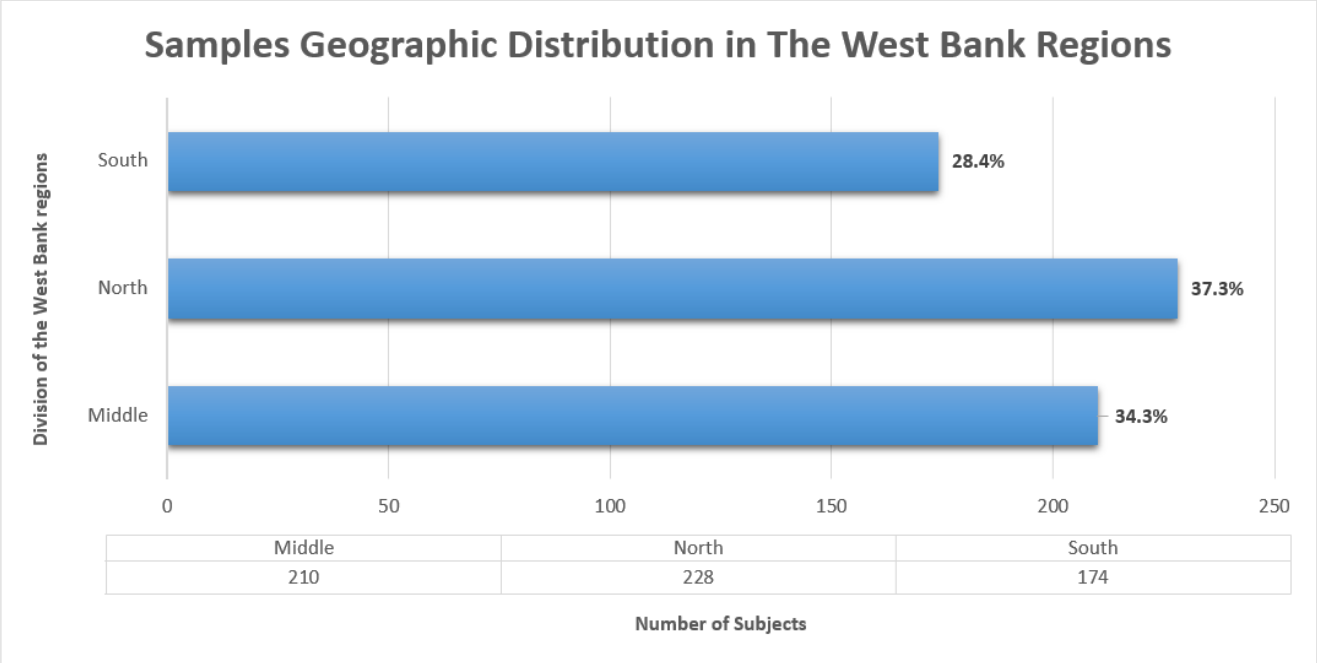
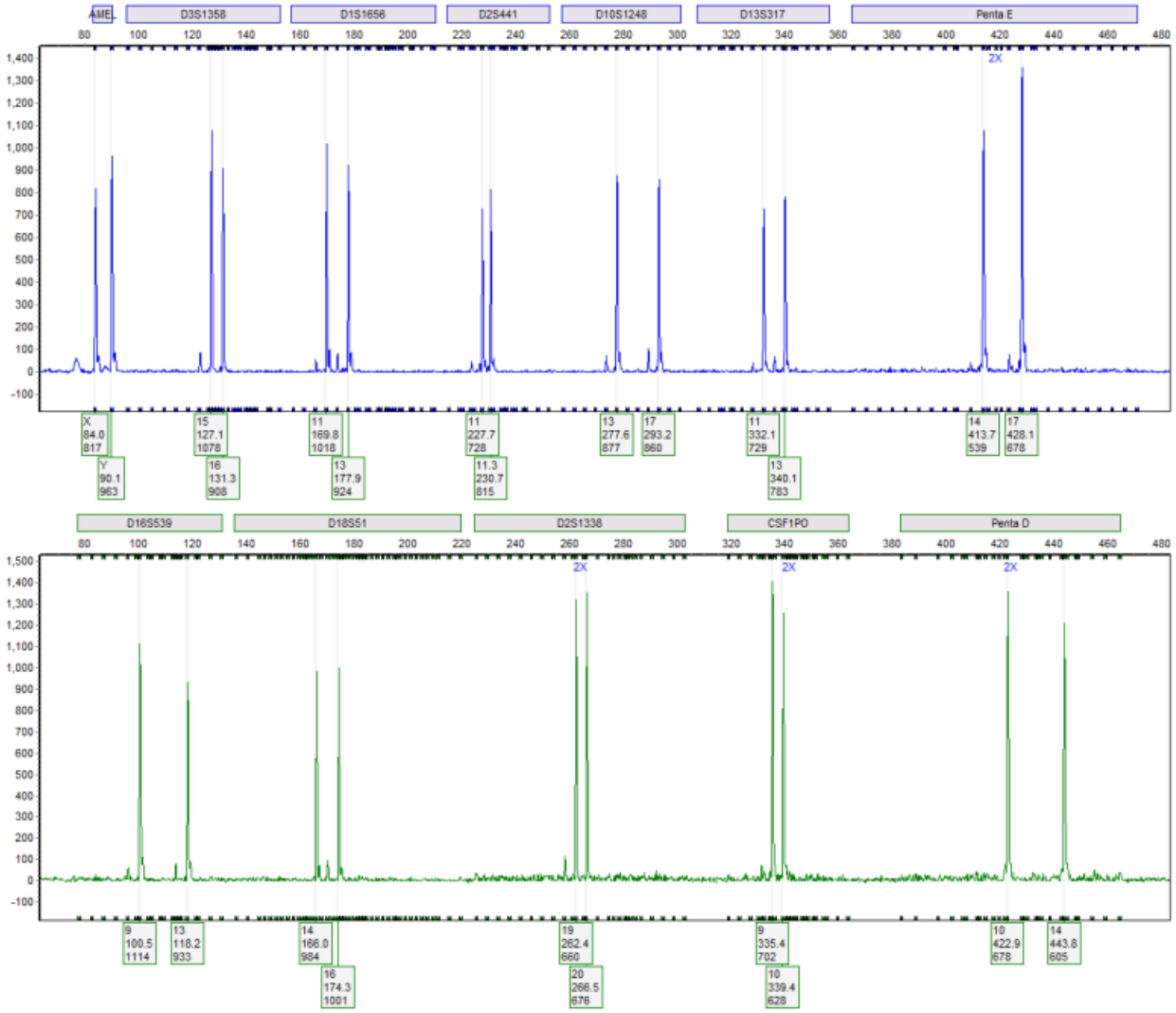


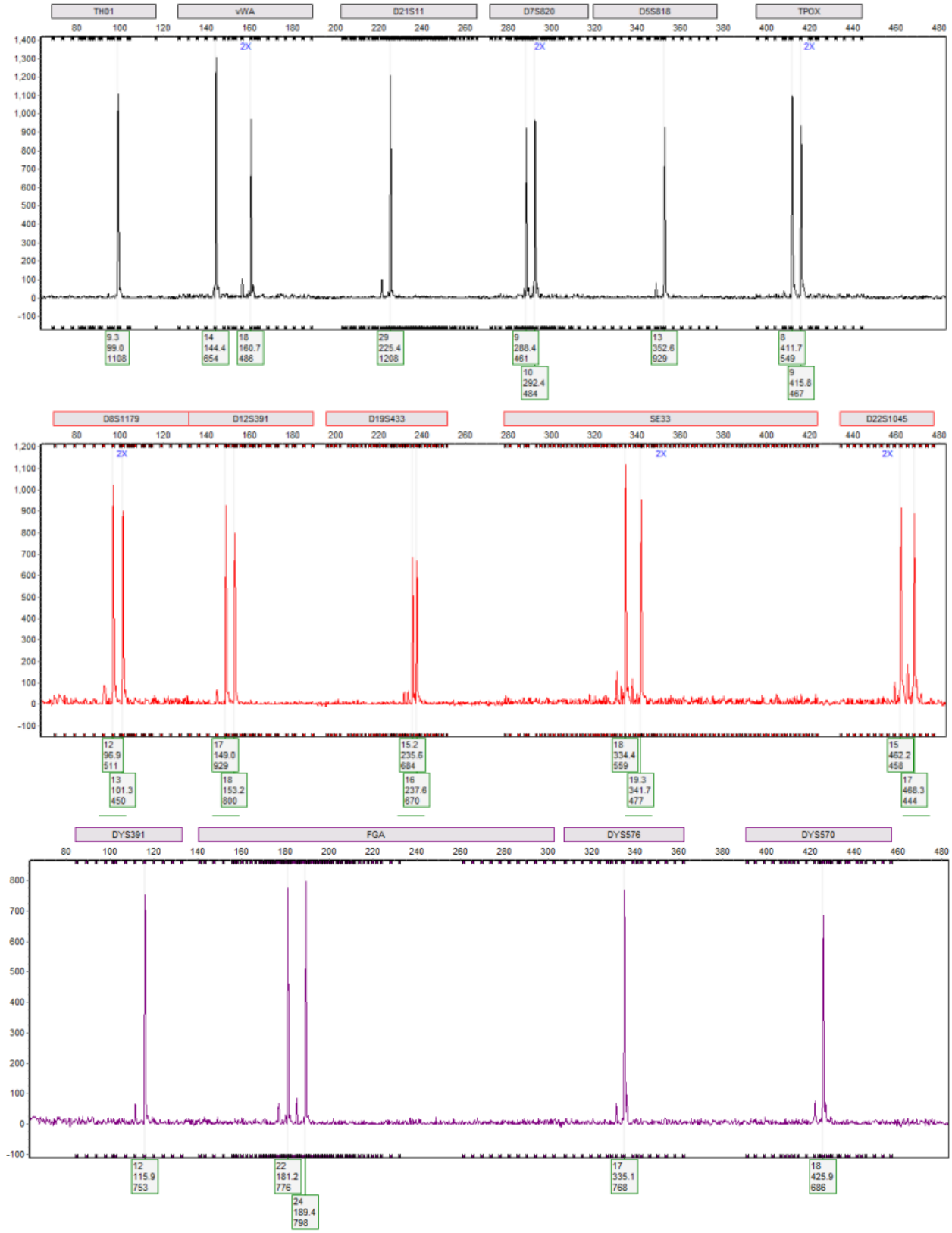
Figure.14: Samples geographic distribution throughout the middle, north, and south of the West Bank.

3.2 Autosomal STRs analysis

3.2.1 DNA genotyping

In order to study the genetic landscape of the Palestinian population, 619 samples obtained from unrelated individuals of the Palestinian population were profiled with the PowerPlex Fusion 6C® system for 23 autosomal STR markers, Amelogenin, and three male specific markers. Figure 15 shows an example of the STR profile obtained using PowerPlex Fusion 6C® system and capillary electrophoresis.




TH01
vWA
D21S11
D7S820
D5S818
TPOX
2X
9.3 99.0 1108
14 144.4 654
18 160.7 486
29 225.4 1208
9 288.4 461
10 292.4 484
13 352.6 929
8 411.7 549
9 415.8 467
D8S1179
D12S391
D19S433
SE33
D22S1045
12 96.9 511
13 101.3 450
17 149.0 929
18 153.2 800
15.2 235.6 684
16 237.6 670
18 334.4 559
19.3 341.7 477
15 462.2 458
17 468.3 444
DYS391
FGA
DYS576
DYS570
12 115.9 753
22 181.2 776
24 189.4 798
17 335.1 768
18 425.9 686

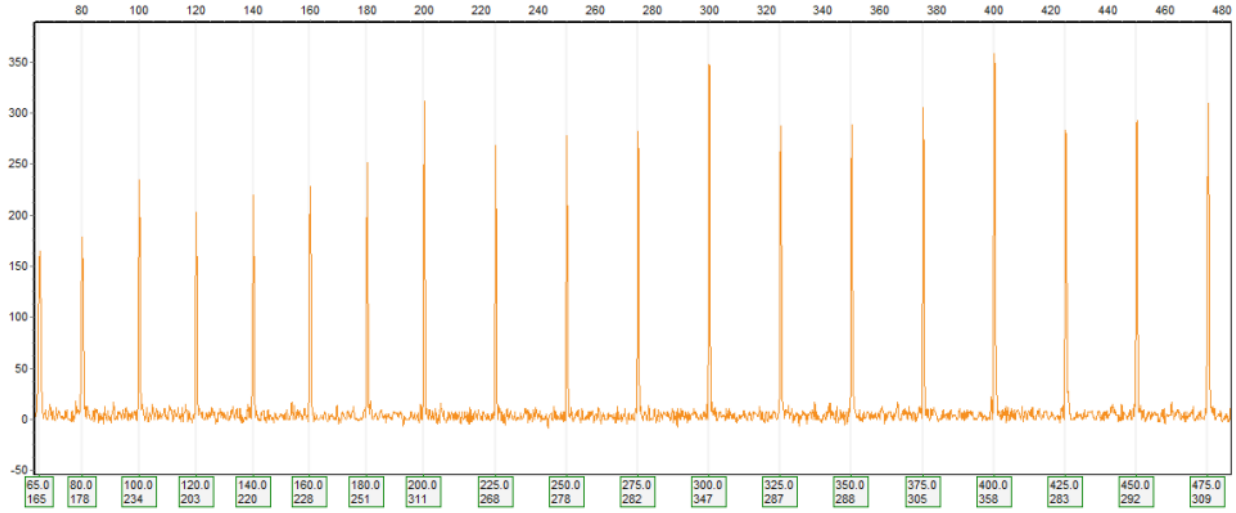
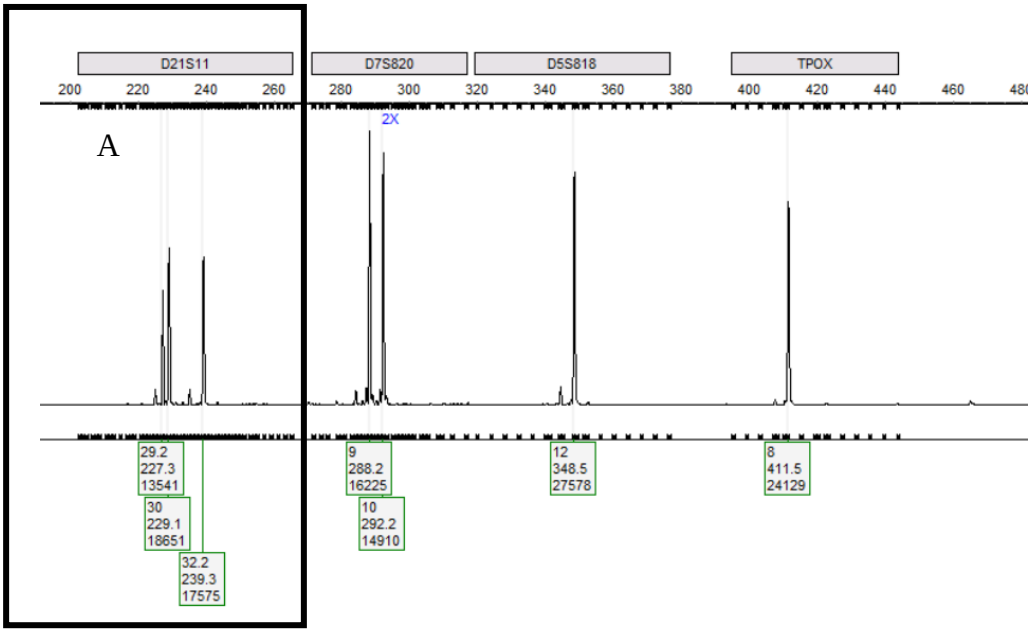


Figure.15: A random male STR profile as obtained from the PowerPlex Fusion 6C® system. This profile shows the different dyes that correspond to the different loci. The first number below each peak indicates the number of repeats for each allele (allele call), below the allele call is the allele size in bp, and finally, the last number indicates the height in relative fluorescence units (RFU). 500 WEN Internal lane standard is shown in orange color.

3.2.2 Allele frequency of 23 autosomal STRs in the total population

Throughout the analysis of the genotypes of the 619 samples studied, two samples demonstrated a tri-allelic pattern at the locus D21S11 with the genotype 29.2, 30, 32.2 for the first sample and the genotype 29, 29.2, 32 for the second sample. These samples were confirmed with re-amplification and re-injection. Of note, none of these tri-allelic genotypes were reported in STRBase (fig.16A & B). These two samples were excluded from allelic frequency analysis but were considered in further anomalous sample analysis. Thus, the total genotypes used for allele frequency calculations and further analysis was 617. In total, 336 alleles were genotyped at the different loci. TPOX*8 had the highest allele frequency (53.97%). Allele 15 of D22S1045 exhibited the second highest allele frequency (43.11%). Other alleles at other loci showed allele frequency ranging from 34.27% to 0.081%. 53 alleles were rare and detected only once in the 617 samples with a frequency of 0.081% (Table.4). These rare alleles were D3S1358*12*20, D10S1248*10*19, D2S1338*27, CSF1PO*7*15, Penta D*5*18, TH01*4*11, D21S11*29.2*33, D16S539*10.3, TPOX*13, D8S1179*7*23,

D12S391*13*14.2*18.1*19.1*20.1*27, D18S51*14.2, Penta E*13.3,*14.2*15.4*22*23, D7S820*14, D19S433*10*21, vWA*21, SE33*8*10*12.1*13.2*13.3*18.2*19.1*23*26*34.2*35*36*37, FGA*22.3*24.2*25.2*30, D22S1045*13*20*22. TPOX, THO1, D13S317, and D5S818 exhibited the lowest number of alleles with only 8 alleles genotyped in our cohort. SE33 showed the largest number of alleles with 55 different alleles genotyped along the locus (Table.5 & Fig.17).



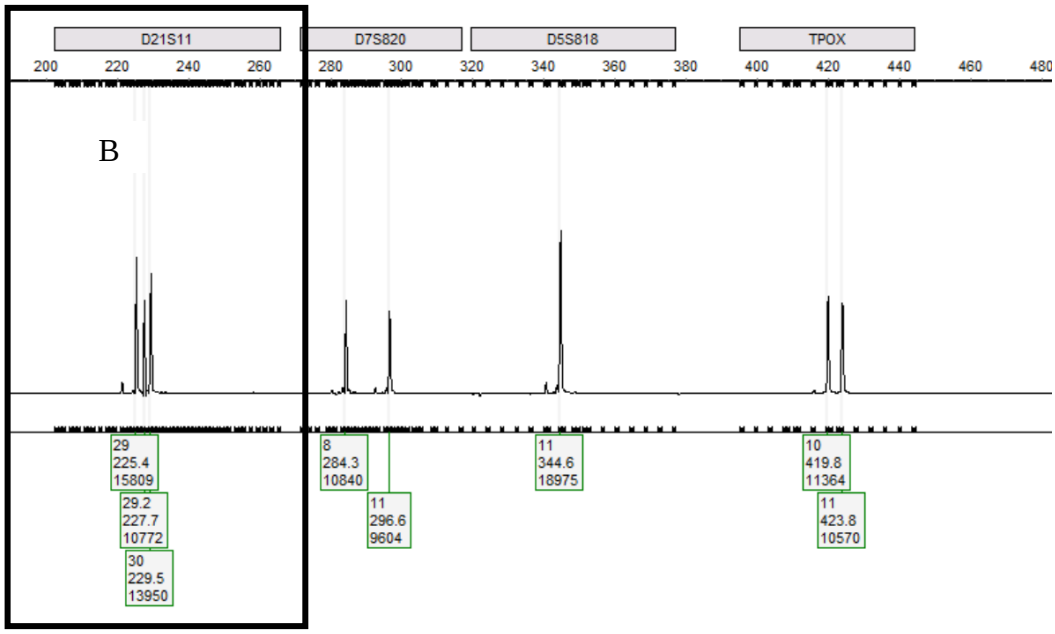


Figure.16: Tri-allelic pattern detected in two samples in at D21S11 locus. A) Tri-allelic pattern with 29.2, 30, 32.2 genotype. B) Tri-allelic pattern with 29, 29.2, 32 genotype.

Table.4: Autosomal STRs allele Frequencies Results for the Total Population: Combined (Sample Size = 617)*

Locus	D3S-1358	D1S-1656	D2S-441	D10S-1248	D13-S317	Penta E	D16-S539	D18-S51	D2S-1338	CSF1PO	Penta D	TH01	vWA	D21-S11	D7S-820	D5S-818	TPOX	D8S-1179	D12-S391	D19-S433	SE33	D22S-1045	FGA
N	1234	1234	1234	1234	1234	1234	1234	1234	1234	1234	1234	1234	1234	1234	1234	1234	1234	1234	1234	1234	1234	1234	1234
Allele																							
2.2	-	-	-	-	-	-	-	-	-	-	0.0170	-	-	-	-	-	-	-	-	-	-	-	-
4	-	-	-	-	-	-	-	-	-	-	-	0.0008	-	-	-	-	-	-	-	-	-	-	-
5	-	-	-	-	-	0.0567	-	-	-	-	0.0008	-	-	-	-	-	-	-	-	-	-	-	-
6	-	-	-	-	-	-	-	-	-	-	0.0073	0.2350	-	-	-	-	0.00405	-	-	-	-	-	-
6.3	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	0.0032	-	-
7	-	-	-	-	-	0.1029	-	-	-	0.00081	0.0049	0.1815	-	-	0.0113	-	0.0024	0.0008	-	-	-	-	-
7.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0016	-	-
8	-	0.0024	-	-	0.1442	0.0373	0.0300	-	-	0.0113	0.0267	0.1175	-	-	0.1572	0.0130	0.5397	0.0122	-	-	0.0008	-	-
9	-	-	0.0032	0.0049	0.0673	0.0154	0.1540	0.0024	-	0.0154	0.1686	0.2958	-	-	0.1135	0.0729	0.1297	0.0097	-	-	0.0122	-	-
9.3	-	-	-	-	-	-	-	-	-	-	-	0.1394	-	-	-	-	-	-	-	-	-	-	-
10	-	0.0016	0.1199	0.0008	0.0438	0.0616	0.0867	0.0049	-	0.2909	0.1669	0.0292	-	-	0.3015	0.0964	0.0810	0.0656	-	0.0008	0.0008	0.0041	-
10.1	-	-	-	-	-	-	-	-	-	-	0.0065	-	-	-	-	-	-	-	-	-	-	-	-
10.2	-	-	-	-	-	-	-	0.0024	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10.3	-	-	-	-	-	-	0.0008	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-
11	-	0.0729	0.3363	0.0065	0.2788	0.1240	0.3071	0.0243	-	0.2950	0.1896	0.0008	0.0016	-	0.2350	0.2747	0.2196	0.0940	-	0.0146	-	0.1175	-
11.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0024	-	-	-	-	-	-	-	-
11.3	-	-	0.0956	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12	0.0008	0.1037	0.0616	0.0332	0.3363	0.1840	0.2569	0.1378	-	0.3209	0.1896	-	-	-	0.1556	0.3290	0.0227	0.1224	-	0.0908	0.0049	0.0219	-
12.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0008	-	
12.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0016	-	-	-
12.3	-	-	0.0016	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
13	0.0064	0.0997	0.0316	0.2099	0.0997	0.1070	0.1410	0.1686	-	0.0559	0.1613	-	0.0032	-	0.0227	0.2010	0.0008	0.2528	0.0008	0.2091	0.0267	0.0008	-
13.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0381	0.0008	-	-
13.3	-	-	0.0024	-	-	0.0008	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0008	-	-
14	0.0535	0.1070	0.3128	0.3428	0.0284	0.0511	0.0219	0.1807	-	0.0089	0.0802	-	0.0875	-	0.0008	0.0097	-	0.1969	-	0.2731	0.0324	0.0640	-
14.2	-	-	-	-	-	0.0008	-	0.0008	-	-	-	-	-	-	-	-	-	-	0.0008	0.0502		-	-
14.3	-	0.0016	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0016	-	-
14.4	-	-	-	-	-	0.0016	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
15	0.2520	0.1783	0.0292	0.2091	0.0016	0.0616	0.0016	0.1021	-	0.0008	0.0429	-	0.1013	-	-	0.0032	-	0.1775	0.0211	0.1264	0.0446	0.4311	-
15.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0948	-	-	-
15.3	-	0.0429	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0024	-	-
15.4	-	-	-	-	-	0.0008	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16	0.2536	0.1750	0.0057	0.1499	-	0.0535	-	0.1136	0.0397	-	0.0097	-	0.2326	-	-	-	-	0.0527	0.0186	0.0575	0.0462	0.2731	-
16.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0276	-	-	-
16.3	-	0.0421	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0032	-	-
16.4	-	-	-	-	-	0.0016	-	-		-	-	-		-	-	-	-	-	-	-	-	-	-

17	0.2650	0.060	-	0.0373	-	0.0446	-	0.1094	0.2358	-	0.0024	-	0.2812	-	-	-	-	0.0113	0.1045	0.0065	0.0891	0.0746	0.0032
17.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0032	0.0024	-	-
17.3	-	0.0746	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0073	-	0.0016	-	-
18	0.1556	0.0057	-	0.0049	-	0.0462	-	0.0875	0.1353	-	0.0008	-	0.1994	-	-	-	-	0.0032	0.1856	0.0049	0.0932	0.0113	0.0049
18.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0008	-	0.0032	-	-
18.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0008	-	-
18.3	-	0.0267	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0227	-	0.0016	-	-
19	0.0122	-	-	0.0008	-	0.0276	-	0.0381	0.1183	-	-	-	0.0754	-	-	-	-	-	0.1248	-	0.0656	-	0.0551
19.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0008	-	0.0008	-	-
19.3	-	0.0057	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0057	-	0.0024	-	-
20	0.0008	-	-	-	-	0.0146	-	0.0186	0.1394	-	-	-	0.0170	-	-	-	-	-	0.1232	-	0.0446	0.0008	0.0972
20.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0008	-	-	-	
20.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0024	-	-
21	-	-	-	-	-	0.0049	-	0.0057	0.0608	-	-	-	0.0008	-	-	-	-	-	0.0948	0.0008	0.0235	-	0.1548
21.2	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0105	-	0.0057
22	-	-	-	-	-	0.0008	-	-	0.0397	-	-	-	-	-	-	-	-	-	0.1224	-	0.0041	0.0008	0.1783
22.2		-	-	-	-		-		-	-	-	-		-	-	-	-	-	-	-	0.0162	-	0.0024
22.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0008
23	-	-	-	-	-	0.0008	-	-	0.0875	-	-	-	-	-	-	-	-	0.0008	0.1005	-	0.0008	-	0.1718
23.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0365	-	0.0049
24	-	-	-	-	-		-		0.0746	-	-	-		-	-	-	-	-	0.0438	-	0.0016	-	0.1677
24.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0284	-	0.0008
25	-	-	-	-	-	-	-	-	0.0616	-	-	-	-	-	-	-	-	-	0.0162	-	0.0016	-	0.0924
25.2	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-	0.0332	-	0.0008
26	-	-	-	-	-	-	-	-	0.0065	-	-	-	-	0.0008	-	-	-	-	0.0041	-	0.0008	-	0.0462
26.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0421	-	-
27	-	-	-	-	-		-	-	0.0008	-	-	-	-	0.0219	-	-	-	-	0.0008	-	-	-	0.0073
27.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0526	-	-
28	-	-	-	-	-	-	-	-	-	-	-	-	-	0.1548	-	-	-	-	-	-	-	-	0.0032
28.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0519	-	-
29	-	-	-	-	-	-	-	-	-	-	-	-	-	0.2545	-	-	-	-	-	-	-	-	0.0016
29.2	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0008	-	-	-	-	-	-	0.0900	-	-
30	-	-	-	-	-	-	-	-	-	-	-	-	-	0.2504	-	-	-	-	-	-	-	-	0.0008
30.2	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0219	-	-	-	-	-	-	0.0519	-	-
31	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0421	-	-	-	-	-	-	-	-	-
31.2	-	-	-	-	-	-	-	-	-	-	-	-	-	0.1070	-	-	-	-	-	-	0.0308	-	-
32	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0065	-	-	-	-	-	-	0.0016	-	-
32.2	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0932	-	-	-	-	-	-	0.0097	-	-
33	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0008	-	-	-	-	-	-	0.0016	-	-
33.2	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0332	-	-	-	-	-	-	0.0065	-	-
34	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0032	-	-	-	-	-	-	0.0081	-	-
34.2	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0049	-	-	-	-	-	-	0.0008	-	-
35	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0016	-	-	-	-	-	-	0.0008	-	-

35.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0016	-	-
36	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0016	-	-	-	-	-	-	0.0008	-	-
37	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0008	-	-

***Allele frequency measured among 617 Palestinian from the West Bank genotyped at 23 autosomal loci governed by PowerPlex fusion® 6C kit.** Allele frequencies were calculated using Arlequin 3.5. N is the total number of individuals included in the study multiplied by 2 (2 N to measure the frequency of ploidy organisms.

Table.5: Forensic parameters measured to check for the quality and adequacy of this system to be used in forensic applications.*

Locus	D3S-1358	D1S-1656	D2S-441	D10S-1248	D13-S317	Penta E	D16-S539	D18-S51	D2S-1338	CSF1PO	Penta D	TH01	vWA	D21-S11	D7S-820	D5S-818	TPOX	D8S-1179	D12-S391	D19-S433	SE33	D22S-1045	FGA
Allele (n)[a]	9	16	11	11	8	23	9	16	12	9	16	8	10	17	9	8	8	13	21	16	55	11	20
Allele (v)[b]	0	6	3	0	0	5	1	2	0	0	2	1	0	6	1	0	0	0	7	7	36	0	6
Allele (C)[c]	17	15	11	14	12	12	11	14	17	12	11	9	17	29	10	12	8	13	18	14	18	15	22
Allele (R)[d]	12	10, 14.3	12.3	9, 10	15	13.3 14.2 15.4 22 23	10.3	14.2	27	7, 15	5, 18	4, 11	21	29.2, 33	14	15	13	7, 23	13 14.2 18.1 19.1 20.1 27	10, 21	8, 10, 12.1, 13.2, 13.3, 18.2, 19.1, 23 26, 34.2 35, 36, 37	13 20 22	22.3 24.2 25.2 30
H_o [e]	0.7439	0.8412	0.7536	0.7682	0.7618	0.8947	0.7812	0.8428	0.8266	0.6872	0.8314	0.7374	0.7666	0.8055	0.7974	0.7293	0.5802	0.7909	0.8703	0.7942	0.9254	0.6953	0.82001
He [f]	0.7753	0.8873	0.7604	0.7703	0.7718	0.9065	0.7878	0.8753	0.8695	0.7223	0.8600	0.7909	0.8038	0.8252	0.7921	0.7616	0.6371	0.8352	0.8870	0.8408	0.9483	0.7160	0.8640
pHW [g]	0.2093	0.1014	0.0125	0.1015	0.1202	0.1668	0.3800	0.1330	0.0460	0.0871	0.1426	0.1369	0.6851	0.0148	0.9625	0.2123	0.0109	0.2975	0.1057	0.0153	0.0195	0.1013	0.0568
RMP [h]	0.0853	0.0236	0.0978	0.0890	0.0866	0.0182	0.0761	0.0289	0.0311	0.1257	0.0364	0.0736	0.0666	0.0533	0.0739	0.0952	0.1788	0.0480	0.0247	0.0442	0.0069	0.1211	0.0338
TPI [i]	1.9525	3.1480	2.0296	2.1573	2.0986	4.7462	2.2852	3.1804	2.8832	1.5984	2.9663	1.9043	2.1424	2.5708	2.468	1.8473	1.1911	2.3915	3.8563	2.4291	6.7065	1.6410	2.7793
PD [j]	0.9147	0.9764	0.9022	0.9110	0.9134	0.9818	0.9239	0.9711	0.9689	0.8743	0.9636	0.9264	0.9334	0.9467	0.9261	0.9048	0.8212	0.9520	0.9753	0.9558	0.9931	0.8789	0.9662
PE [k]	0.4994	0.6775	0.5160	0.5415	0.5301	0.7845	0.5646	0.6807	0.6493	0.40873	0.6586	0.4885	0.5386	0.6093	0.5942	0.4751	0.2678	0.5823	0.7353	0.5883	0.8476	0.4210	0.6369
PIC [l]	0.7376	0.8761	0.7248	0.7348	0.7382	0.8985	0.7562	0.8615	0.8556	0.6698	0.8431	0.7586	0.7754	0.8027	0.7606	0.7230	0.5920	0.8141	0.8755	0.8225	0.9451	0.6749	0.8480
Ht[m]	0.7387	0.8833	0.7334	0.7332	0.7158	0.8781	0.8006	0.8667	0.825	0.6912	0.873	0.814	0.7551	0.7798	0.7958	0.7835	0.5034	0.8311	0.8789	0.8448	0.942	0.735	0.8153

*[a] Number of alleles genotyped at each loci, [b] Number of different variants genotyped at each locus, [c] most common allele for each locus, [d] rarest allele for each locus, [e] observed heterozygosity, [f] Expected heterozygosity, [g] Hardy-Weinberg equilibrium, [h] random match probability, [i] typical paternity index, [j] power of discrimination, [k] power of exclusion, [l] polymorphic information content. [m] heterozygosity expectation if all subpopulations were at HWE. Red color indicates deviations from HWE before applying Bonferroni's correction.

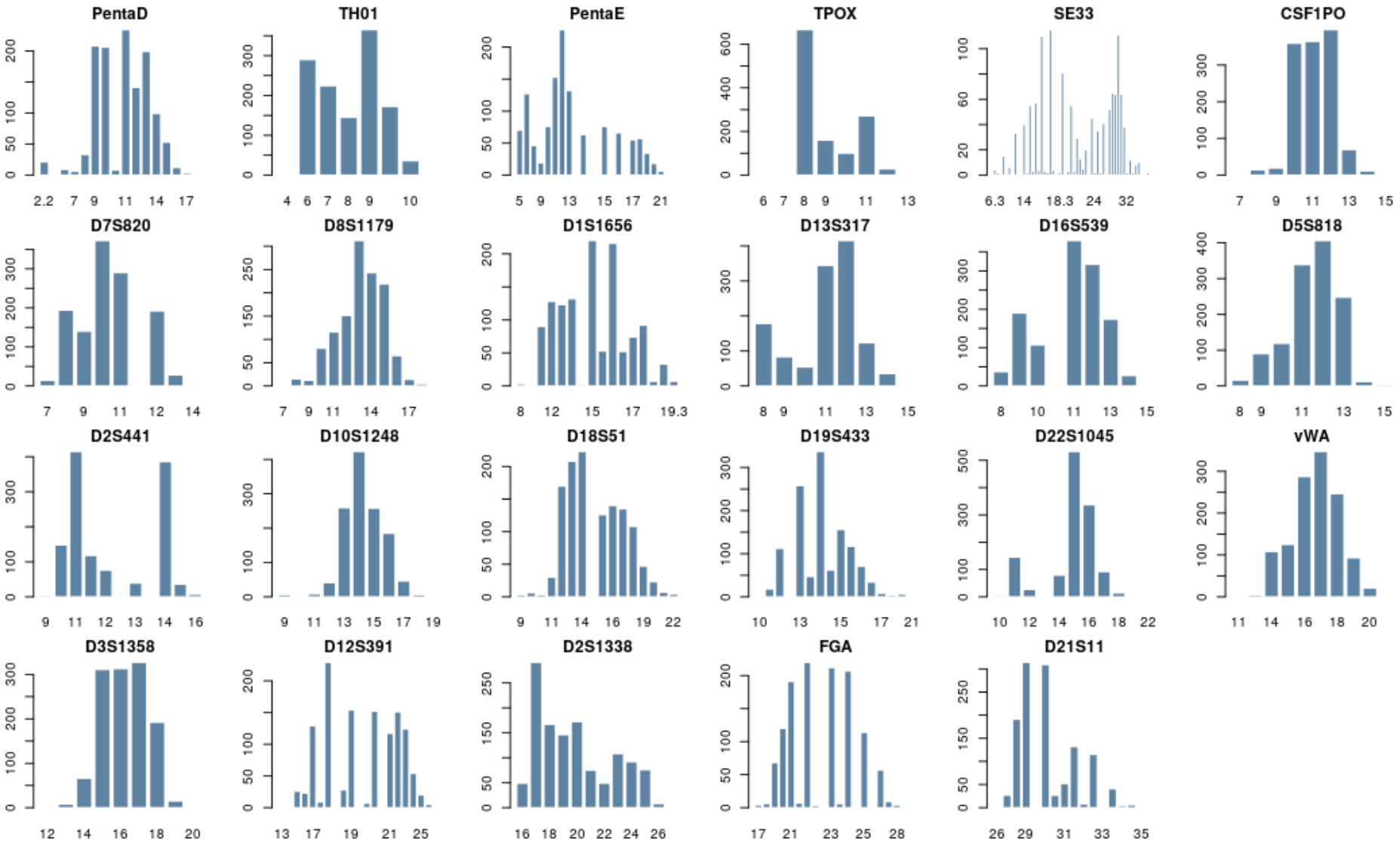


Figure.17: Allele frequency diagram for each locus. The x-axis is the allele designation (alleles genotyped in our cohort) and the y-axis is the frequency of each allele for each locus. Figure was created using STRAF software version 2.1.5.

3.2.3 Microvariant “off-ladder” and novel allele

While analyzing our data, we were able to detect 11 microvariant “off-ladder” in three loci (SE33, Penta D, and Penta E) (Table.6). Interestingly, variant 10.1 of Penta D was detected only in Samaritans, SE33 alleles 13.3 and 19.1, and Penta E allele 13.3 were detected in the Christian cohort only. Figure 18 shows sequencing results for some of these samples (the rest will be confirmed later as some of the samples failed PCR or sequencing). These microvariants were reported on STRBase except for Penta E*13.3 which was never reported to our knowledge (Fig.18).

Table.6: Microvariant “off-ladder” and novel alleles detected in our cohort.

Locus	Allele	Size in electropherogram (bp)	Number of individuals	Frequency
SE33	7.3	292.3	2	0.003241491
SE33	12.1	310.7	1	0.001620746
SE33	13.3	316.8	1	0.001620746
SE33	17.3	333.5	2	0.003241491
SE33	18.1	335.3	4	0.006482982
SE33	18.3	337.5	2	0.003241491
SE33	19.1	339.5	1	0.001620746
Penta E	13.3	399.5	1	0.001620746
Penta E	14.4	417.7	2	0.003241491
Penta E	15.4	422.1	1	0.001620746
Penta D	10.1	423.9	7	0.011345219

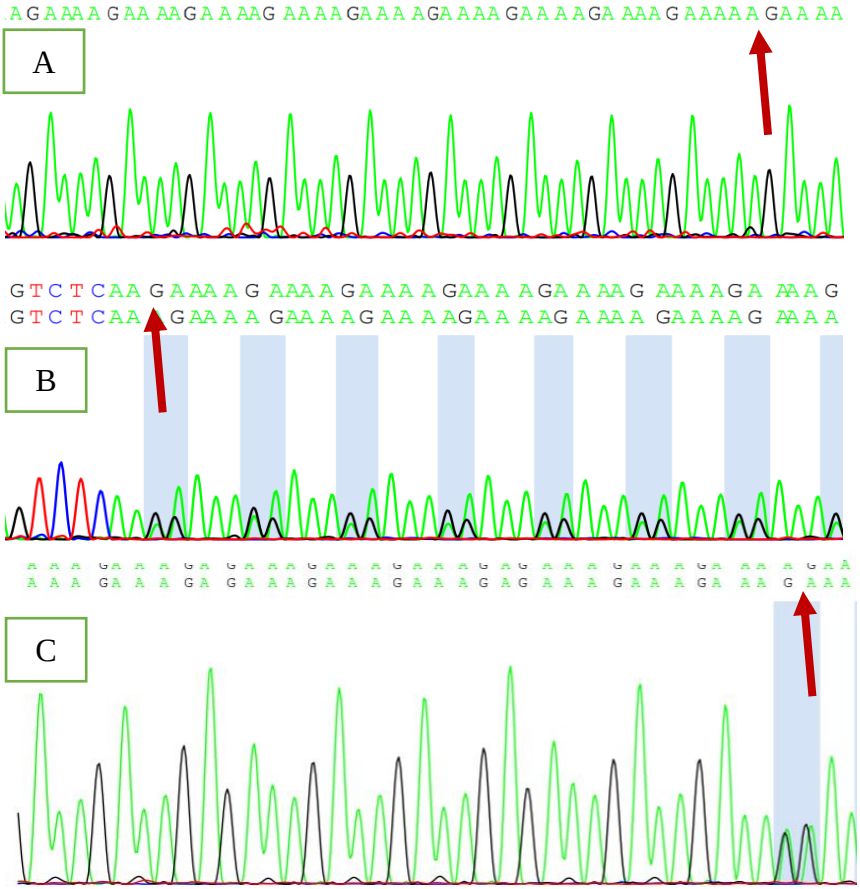


Figure.18: Sequencing results of intermediate or microvariant allele at Penta D, Penta E, and SE33. A) The sequence of Penta D*10.1, the red arrow, and the line indicates the 10th repeat unit where one base "A" is inserted. B) The sequence of Penta E*14.4, the red arrow, and the line indicates the first repeat unit where one base "A" is lost in a heterozygous sample. C) The sequence of SE33*12.1, the red arrow, and the line indicates the 10th repeat unit where one base "A" is inserted in a heterozygous sample. Sequences were analyzed using Poly Peak Parser online tool

3.2.4 HWE Exact Test and Linkage Disequilibrium

In order to use a database in forensic testing, it needs to be confirmed that the sample population included in the database is representative of the actual whole population. This issue can be confirmed by Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium analyses. To test for the null hypothesis of HWE in our study cohort for each locus, Fisher exact test was used to calculate the p-value using the Markov-chain Monte Carlo method. The p-values for most of the loci were larger than 0.05, indicating no significant departure from HWE. Five loci - D21S11, TPOX, D2S441, SE33, and D19S433 showed a slight departure from HWE with p-values of

0.0148, 0.0109, 0.0125, 0.0195, and 0.0153, respectively. However, after applying the Bonferroni's correction (0.05/23= 0.00217) (Fujii et al., 2014), there were no significant discrepancies, suggesting that our study cohort represents the Palestinian population and that the observed allele frequencies are quite close to the expected frequencies (Table.5). For linkage disequilibrium, loci that are located closely on the same chromosome can exhibit high linkage disequilibrium which means that they are inherited dependently. In this case, knowing a specific allele allows the prediction of the second allele in a diploid sample. Some factors can affect linkage equilibrium and result in population structure. These factors include natural selection, genetic recombination, and genetic drift. To prove that genotypes at one locus are independent of genotypes at another locus, we performed linkage disequilibrium testing on loci that are located on the same chromosomes by using the log-likelihood ratio statistic (G-test) using Genepop software. Loci D2S441, D2S1338, and TPOX are located on chromosome 2. Loci CSF1PO and D5S818 are located on chromosome 5. Loci vWA and D12S391 are located on chromosome 12. Loci Penta D and D21S511 are located on chromosome 21. The p-values obtained from G-test for the loci tested above were larger than 0.05, indicating that loci located on the same chromosome have no linkage disequilibrium (Table.7). These results suggest that the analyzed loci in PowerPlex® fusion system are independent of each other. Thus, the product rule can be used to calculate the combined match probability (CMP), the combined power of discrimination (CPD), and the combined power of exclusion (CPE).

Table.7: linkage disequilibrium. P-value of Likelihood ratio statistic (G-test) between loci located on the same chromosome.

Locus	D2S441	TOPX	CSF1PO	Penta D	vWa
D2S1338	0.919120	0.388660	-	-	
TPOX	0.559220	-	-	-	
D5S818	-	-	0.331020	-	
D21S511	-	-	-	0.213620	
D12S391					1.000000

3.2.5 Forensic efficiency parameters

One of our aims from studying the genetic structure of the Palestinian population was to test of the validity of using the data generated for forensic applications. Random match probability (RMP) is important for measuring the probability of having an identical genotype when two subjects are selected randomly. In our study, RMP ranged from 0.69% for the SE33 locus to 17.88% for the TPOX locus. Polymorphism information content ranged from 59.2% with TPOX to 94.5% with SE33. Measuring how informative each locus tested in a forensic STR system is necessary in order to build the basis for any calculations made in real forensic casework to make sure that loci picked are actually highly polymorphic and can distinguish individuals with high confidence. This informative ability can also be measured by the power of discrimination and the power of exclusion. In general, a locus can be informative when the power of discrimination is larger than 80% and/or the power of exclusion is larger than 50% (Shriver et al., 1995). For all tested loci, the power of discrimination ranged from the lowest value of 82.11 % for TPOX, to the highest value of 99.3% for SE33, and the power of exclusion ranged from the lowest value of 26.78% for TPOX to the highest value of 84.76% for SE33 (Fig.19). This indicates that all loci in the PowerPlex®

Fusion system are informative and can be used with high confidence in real casework (Shriver et al., 1995). In this case, CMP, CPD, and CPE equal 4.20677E-30, 1, and 0.999999999, respectively.

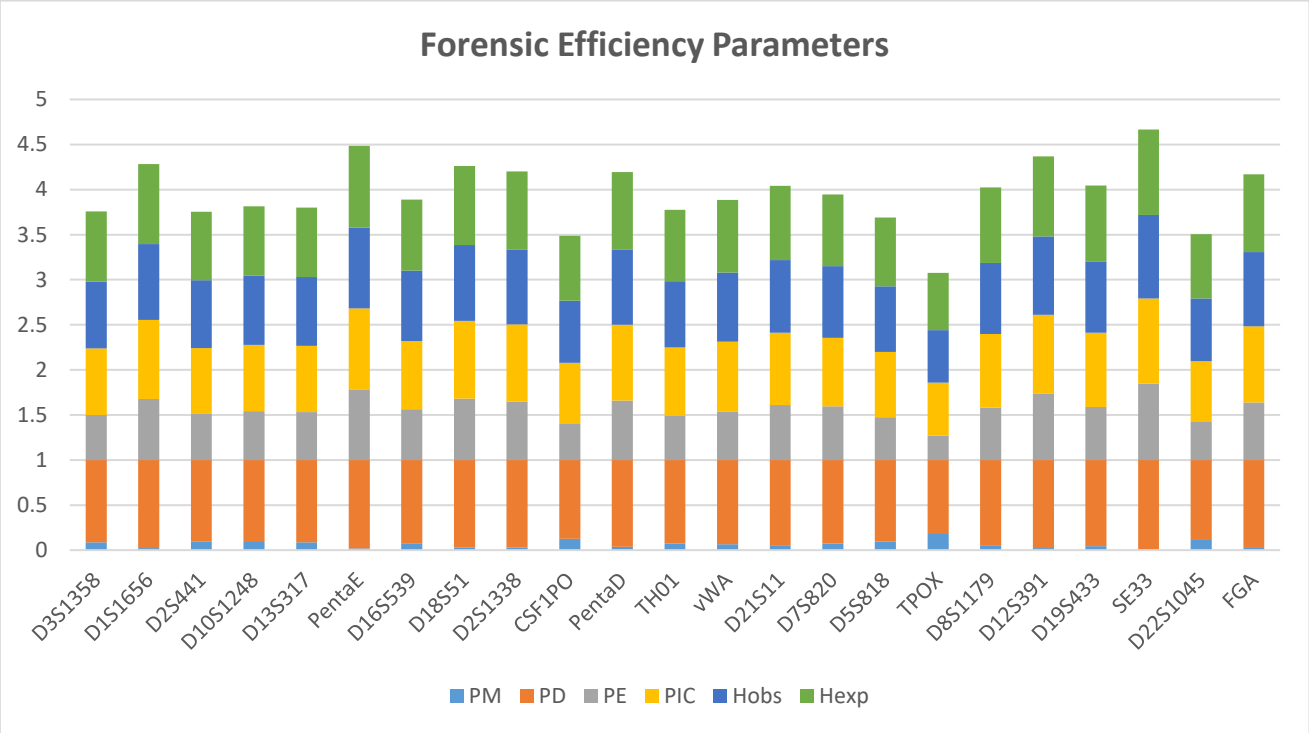


Figure 19: Stacked column analysis for forensic parameters analysis results. PM: probability matching. PD: power of discrimination. PE: power of exclusion. PIC: polymorphism information content. H_{obs}: observed heterozygosity. H_{exp}: expected heterozygosity.

3.2.6 Paternity testing

Measuring the usefulness of this database in paternity testing is important before implementing this database in real forensic casework. This allows users of this database to be confident in interpreting STR genotypes and include or exclude a specific individual from being the alleged father. Typical paternity index is useful in comparing the questioned DNA profile to an untested random person from the public population, considering the subject racial background. In other words, it tests the probability of the alleged father passing the genotype of a specific STR locus compared to the possibility of a random person passing this genotype. In our cohort, SE33 locus had the highest typical paternity index (6.706), indicating that an inherited genotype from the alleged father to son is very informative. On the other hand, TPOX had the lowest informative

index (1.1911). Combined paternity index (CPI) is important in order to calculate POP; CPI is the product of multiplying all the paternal indices together. A CPI greater than 1000 means that the POP is greater than 99%. In our case, CPI equals 1,039,358,320. POP is the chance of paternity based on the DNA profile. In other words, it is the chance of the alleged father being the actual alleged father and not any other individual. Table 8 shows an example of the genotype for a theoretical alleged father and son with the alleged allele indicated. In this case, the probability of paternity equals 99.9999999%.

Table.8: Power of Paternity calculations.

Locus	AF	Mother	SON	Alleged allele	Freq. of the obligate allele	TPI
D3S1358	**17,17**	**14,18**	**17,18**	**17**	**0.2650**	**1.9525**
D1S1656	**14,15**	**12,14**	**12,15**	**15**	**0.1783**	**3.148**
D2S441	**11,14**	**10, 11**	**10, 14**	**14**	**0.3128**	**2.0296**
D10S1248	**13,15**	**13,16**	**13,16**	**13**	**0.2099**	**2.1573**
D13S317	**13,13**	**11,12**	**12,13**	**13**	**0.0997**	**2.0986**
Penta E	**8,18**	**14, 14**	**14,18**	**18**	**0.0462**	**4.7462**
D16S539	**11,11**	**11,12**	**11,12**	**11**	**0.3071**	**2.2852**
D18-S51	**14,14**	**12,17**	**14,17**	**14**	**0.1807**	**3.1804**
D2S1338	**20,25**	**18,21**	**20,21**	**20**	**0.1394**	**2.8832**
CSF1PO	**10,11**	**10,12**	**11,12**	**11**	**0.2950**	**1.5984**
Penta D	**11,12**	**9, 14**	**11, 14**	**11**	**0.1896**	**2.9663**
TH01	**6,9**	**9,10**	**9,10**	**9**	**0.2958**	**1.9043**
vWA	**14,18**	**16, 17**	**14,17**	**14**	**0.0875**	**2.1424**
D21S11	**29,29**	**30,31**	**29,30**	**29**	**0.2545**	**2.5708**
D7S820	**8,9**	**9,10**	**8,10**	**8**	**0.1572**	**2.468**
D5S818	**9,12**	**11,13**	**9,13**	**9**	**0.0729**	**1.8473**
TPOX	**8,10**	**11,11**	**8,11**	**8**	**0.5397**	**1.1911**
D8S1179	**13,13**	**10,16**	**13,16**	**13**	**0.2528**	**2.3915**
D12S391	**18.3,22**	**19,23**	**18.3,23**	**18.3**	**0.0227**	**3.8563**
D19S433	**12,13.2**	**14,14**	**13.2,14**	**13.2**	**0.0381**	**2.4291**
SE33	**18,19**	**17,19**	**17,18**	**18**	**0.0932**	**6.7065**
D22S1045	**11, 15**	**11,14**	**11,14**	**11**	**0.1175**	**1.641**
FGA	**19,20**	**22,25**	**19,25**	**19**	**0.0551**	**2.7793**
CPI	**1039358320**					
POP	**99.9999999%**					

*Theoretical father-son genotypes indicated with their corresponding obligate allele frequencies used to calculate typical paternity index, combined paternity index, and power of paternity, AF: alleged father.

3.2.7 Most common genotypes frequencies

It is important to have a glimpse on the frequency of the most common possible STR profile to know how strong is our database in differentiating between two different unrelated individuals (Edwards et al., 1992). To do so, we calculated the frequency of a theoretical most common profile by assuming a person has all of his 23 loci in heterozygous genotype. With the two heterozygous alleles being the most common alleles according to the allele frequency we detected above. By multiplying the frequency of these two most common alleles, we were able to obtain the frequency of the most common genotype for each locus. Thus, the total frequency of this genotype is 2.98139E-23 (Table.9). This indicates that even if the individual has all of the most common genotypes by chance, this system is highly informative and allows distinguishing individuals with high confidence (Foreman & Evett, 2001).

Table.9: Most common profile allele frequency. *

Locus	Allele 1	Allele 2	Allele 1 frequency	Allele 2 frequency	Formula	Most common genotype frequency
D3S1358	16	17	0.2537	0.2650	2pq	0.134461
D1S1656	15	16	0.1783	0.1750	2pq	0.062405
D2S441	11	14	0.3363	0.3128	2pq	0.21038928
D10S1248	13	14	0.3428	0.2099	2pq	0.14390744
D13S317	11	12	0.2788	0.3363	2pq	0.18752088
Penta E	11	12	0.1240	0.1840	2pq	0.045632
D16S539	11	12	0.3071	0.2569	2pq	0.15778798
D18S51	13	14	0.1686	0.1807	2pq	0.06093204
D2S1338	17	20	0.2358	0.1394	2pq	0.06574104
CSF1PO	11	12	0.2950	0.3209	2pq	0.189331
Penta D	9	11	0.1686	0.1896	2pq	0.06393312
TH01	6	9	0.2350	0.2958	2pq	0.139026
vWA	16	17	0.2326	0.2812	2pq	0.13081424
D21S11	29	30	0.2545	0.2504	2pq	0.1274536
D7S820	10	11	0.3015	0.2350	2pq	0.141705
D5S818	11	12	0.2747	0.3290	2pq	0.1807526
TPOX	8	11	0.5397	0.2196	2pq	0.23703624
D8S1179	13	14	0.2528	0.1969	2pq	0.09955264
D12S391	18	19	0.1856	0.1248	2pq	0.04632576
D19S433	13	14	0.2091	0.2730	2pq	0.1141686
SE33	18	29.2	0.0932	0.0900	2pq	0.016776
D22S1045	15	16	0.4311	0.2730	2pq	0.2353806
FGA	22	23	0.1783	0.1718	2pq	0.06126388
Profile frequency: 2.98139E-23						

*Allele frequency for each designated allele is calculated using Arlequin version 3.5. 2pq formula was used in order to calculate the most common heterozygous allele combination at a given locus. Excel was used in order to calculate the frequency of 2pq genotypes and the total profile frequency.

3.2.8 Homozygosity and Heterozygosity

Heterozygosity is the condition of having two different alleles at a locus. The heterozygosity and homozygosity were calculated for the 23 loci tested. The number of homozygotes (h) plus the number of heterozygotes (H) equals 100% of the samples tested. Thus, since h +H= 1, then H = 1 − h and h = 1 − H. A higher heterozygosity means that more allele diversity exists in the population and therefore there is less chance of a random sample matching between two random people in a population. Our cohort showed approximately almost high heterozygosity among all loci with SE33 having the highest heterozygosity level with 92.54% and TOPX having the lowest heterozygosity level with 58.02%. Meaning that the chance of having an individual picked at random from the Palestinian population with a heterozygous genotype at the SE33 locus is approximately 94%, in contrast, the chance of this individual having a homozygous genotype at TPOX is approximately 42% (Fig.20).

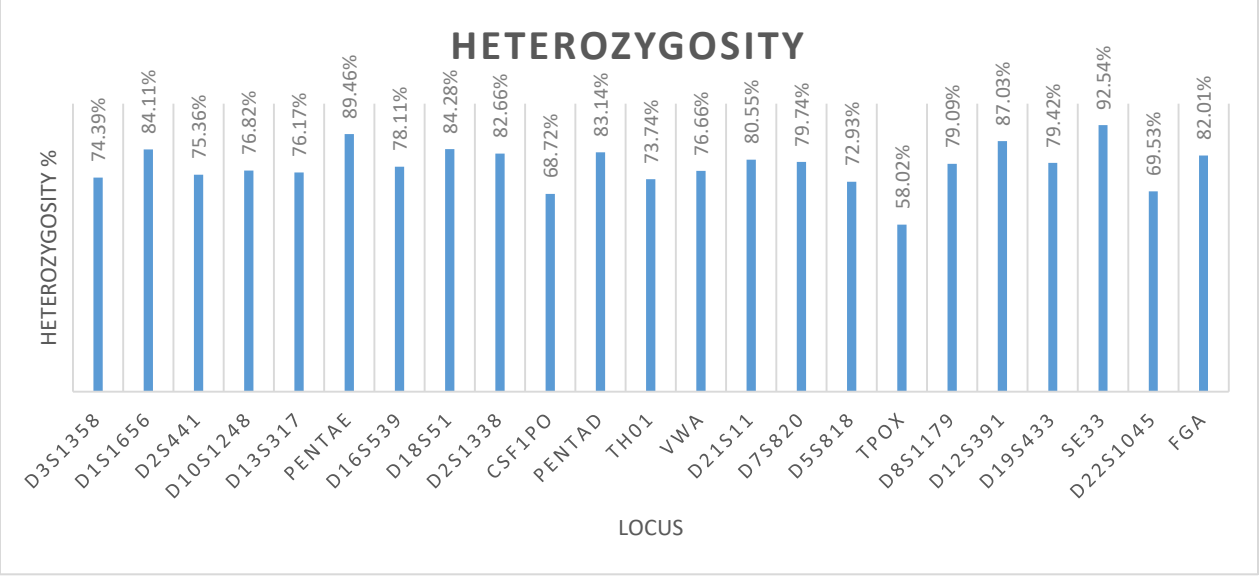


Figure.20: Heterozygosity percentage as detected for all loci in the Palestinian population for the 23 autosomal loci genotyped.

3.2.9 Subpopulations: allele frequency, HWE, and F-statistics

Random mating can be affected by many factors such as religion, language, culture, and geographic distances. In the West Bank, the majority of the Palestinian population is Muslim, with a minority of Christians that does not exceed 2.5% of the total population(*PCBS*). In our study cohort, we were able to genotype 62 Christians, 541 Muslims, and 16 Samaritans. We calculated the allele frequency of each subpopulation in order to measure the p-value for HWE exact test and F-statistics parameters, which include measuring deviation from HWE in the total population (F_{IT}), deviation from HWE within subpopulation (F_{IS}), and differentiation among subpopulation (F_{ST}). Allele frequencies for the three subpopulations for each locus are indicated in Table 10A-W. We tested each subpopulation for HWE using an exact probability test using Genepop software. Muslims exhibited insignificant p-values (>0.05) for most of the loci except for CSF1PO, D21S11, TPOX D19S433, D22S1045, and FGA with p-values of 0.0389, 0.0085, 0.0083, 0.0328, 0.0229, and 0.0298, respectively. This slight discrepancy between the expected and observed allele frequency was corrected after applying Bonferroni's correction with a p-value greater than 0.00217 (Table.11). Christians subpopulations also confirmed to be in HWE for all the tested loci with a p-value >0.05 except for two loci, D1S1656 and D2S441 with a p-value of 0.0050 and 0.0273. However, after applying Bonferroni's correction none of these p-values were significant indicating the Christian subpopulation is in HWE. On the other hand, Samaritans showed a departure from HW for the loci D2S1338 and D1S1656 with a p-value of 0.0433 and 0.0006, respectively. This departure for D2S1338 was corrected after applying Bonferroni's correction. However, D1S1656 departure was still significant even after applying Bonferroni's correction. Taking into account the small sample size, this indicates that this population has some sort of non-random mating that results in deviation from HWE in this subpopulation. Also, each subpopulation was tested for linkage disequilibrium for loci located on the same chromosome, none of the physically linked

loci showed a significant linkage disequilibrium p-value that was larger than 0.05. This indicates that alleles at these genotypes are inherited independently.

F_{IS} for the total loci equals 0.0304. This low positive value means that there is a slight deficiency in individual heterozygosity within subpopulations due to genetic drift. Moreover, this positive value can be attributed mainly to the Samaritans subpopulation, as it is the only subpopulation that showed significant deviation for HW expectations under random mating. F_{ST} is a reflection of subpopulations having different allele frequencies. In other words, it is an indicator of population differentiation. Since F_{ST} equals 0.0180 (lower than 0.05), then there is a very small differentiation among the three subpopulations. F_{IT} is an indicator of deviation from HWE in the total population due to a deficiency in heterozygosity among subpopulations compared to the total population. In our cohort, F_{IT} equals 0.0479, since it is positive then this indicates there is a slight deficiency in heterozygosity in the total population (Table.11).

Some alleles were detected only in specific subpopulations so they were considered specific to these subpopulations. For example, Allele 10.3 of locus D16S539, and allele 10.1 of Penta D were specific only to Samaritans. Alleles 8, 13.3, 17.3 of SE33 and allele 13.3 of Penta E were detected only in Christians. Allele 26 of D12S391 was detected 5 times mainly in Christians, however, when we analyzed the demographic data of these samples, it appeared that two Muslim subjects had mixed ancestry with Christian grandparents and received this 26 allele. Many alleles were detected only in the Muslim cohort and they were never genotyped in any of the Christian or Samaritan subjects. Nevertheless, we can not give an absolute confirmation that these alleles are Muslim specific because the Christian and Samaritan cohorts are smaller compared to the Muslim cohort, thus, we would assume that some of these alleles will be detected in the other cohorts if a larger sample size were analyzed. Some alleles were detected specifically in a combination of

subpopulations. For example, Allele 8 of D5S818, allele 19.3 of D1S1656, Allele 13 of D3S1358, allele 30.2 of D21S11, allele 8 of Penta D, allele 18 of D22S1045 were detected only in the Samaritan and Muslim cohort but was never detected in the Christian cohort (Table.10). Poly component analysis (PCA) was conducted to dimensionally compare the three subpopulations together and calculate the variation percentage. The dimensional representation shows that the three subpopulations are closely clustered and the Christian subpopulation overlaps with the Muslim subpopulation. Our analysis revealed that the variation along the x-axis (PC1) is higher than the variation along the y-axis (2.23% compared to 1.06% variance). This indicates that variation at the individual level attributes more than the variation among subpopulations in the Palestinian population (Fig.21).

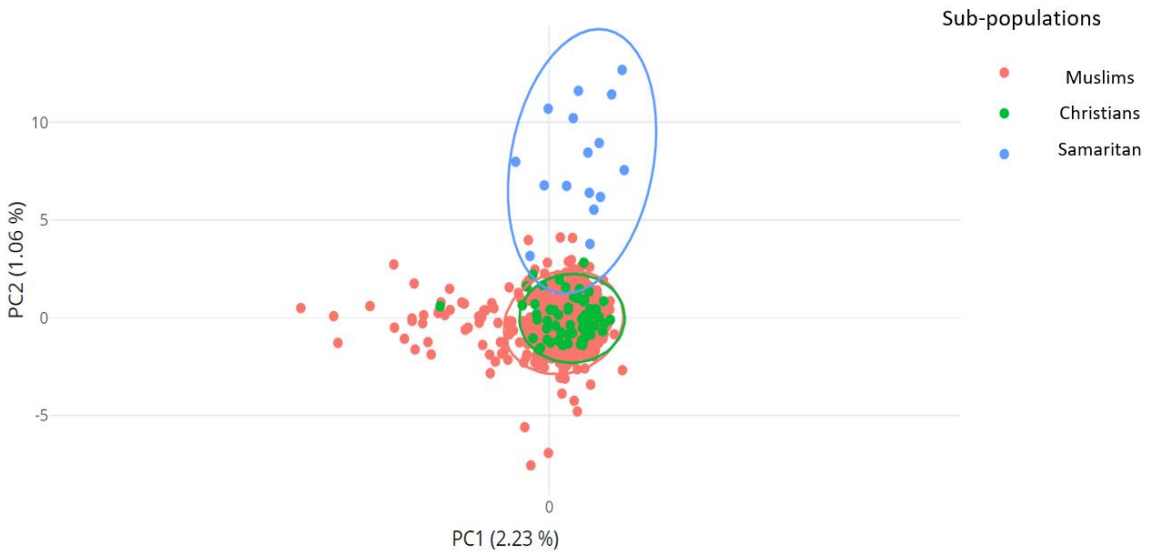


Figure.21: Poly Component analysis (PCA) of the three subpopulations showing the variation percentage along the x-axis (PC1) and the y-axis (PC2). Figure was created using STRAF software version 2.1.5.

Table.10: Allele frequency of the three subpopulations (Muslims, Christians, and Samaritans) in the West Bank for each locus from A-W.*

A) CSF1PO	Muslim	Christian	Samaritan
	1078	124	32
7	0.000927644	0	0
8	0.012987013	0	0
9	0.015769944	0.016129032	0
10	0.296846011	0.25	0.25
11	0.298701299	0.290322581	0.1875
12	0.307050093	0.379032258	0.5625
13	0.058441558	0.048387097	0
14	0.008348794	0.016129032	0
15	0.000927644	0	0

B) D16S539	Muslim	Christian	Samaritan
	1078	124	32
8	0.027829314	0.056451613	0
9	0.141001855	0.258064516	0.1875
10	0.086270872	0.10483871	0.03125
10.3	0	0	0.03125
11	0.313543599	0.258064516	0.28125
12	0.269016698	0.161290323	0.21875
13	0.138218924	0.137096774	0.25
14	0.022263451	0.024193548	0
15	0.001855288	0	0

C) D10S1248	Muslim	Christian	Samaritan
	1078	124	32
9	0.005565863	0	0
10	0.000927644	0	0
11	0.006493506	0.008064516	0
12	0.037105751	0.008064516	0
13	0.209647495	0.217741935	0.1875
14	0.334879406	0.362903226	0.53125
15	0.212430427	0.169354839	0.25
16	0.149350649	0.185483871	0.03125
17	0.037105751	0.048387097	0
18	0.005565863	0	0
19	0.000927644	0	0

D) D22S1045	Muslim	Christian	Samaritan
	1078	124	32
10	0.003710575	0.008064516	0
11	0.121521336	0.096774194	0.0625
12	0.023191095	0.016129032	0
13	0.000927644	0	0
14	0.062152134	0.096774194	0
15	0.435992579	0.427419355	0.28125
16	0.265306122	0.314516129	0.375
17	0.078849722	0.040322581	0.0625
18	0.006493506	0	0.21875
20	0.000927644	0	0
22	0.000927644	0	0

E) TH01	Muslim	Christian	Samaritan
	1078	124	32
4	0.000927644	0	0
6	0.239332096	0.233870968	0.09375
7	0.184601113	0.177419355	0.09375
8	0.115955473	0.10483871	0.21875
9	0.292207792	0.370967742	0.125
9.3	0.141001855	0.064516129	0.375
10	0.025046382	0.048387097	0.09375
11	0.000927644	0	0

F) TPOX	Muslim	Christian	Samaritan
	1078	124	32
6	0.003710575	0.008064516	0
7	0.002782931	0	0
8	0.523191095	0.580645161	0.9375
9	0.136363636	0.10483871	0
10	0.083487941	0.080645161	0
11	0.227272727	0.193548387	0.0625
12	0.022263451	0.032258065	0
13	0.000927644	0	0

G) D3S1358	Muslim 1078	Christian 124	Samaritan 32
12	0.000927644	0	0
13	0.004638219	0	0.09375
14	0.059369202	0.016129032	0
15	0.250463822	0.314516129	0.0625
16	0.24025974	0.274193548	0.625
17	0.262523191	0.298387097	0.21875
18	0.167903525	0.088709677	0
19	0.012987013	0.008064516	0
20	0.000927644	0	0

H) D7S820	Muslim 1078	Christian 124	Samaritan 32
7	0.012987013	0	0
8	0.152133581	0.233870968	0.03125
9	0.112244898	0.10483871	0.1875
10	0.305194805	0.266129032	0.3125
11	0.233766234	0.225806452	0.3125
11.1	0.000927644	0.016129032	0
12	0.161410019	0.120967742	0.09375
13	0.020408163	0.032258065	0.0625
14	0.000927644	0	0

I) D5S818	Muslim 1078	Christian 124	Samaritan 32
8	0.012059369	0	0.09375
9	0.06864564	0.088709677	0.15625
10	0.092764378	0.153225806	0
11	0.279220779	0.25	0.21875
12	0.340445269	0.298387097	0.0625
13	0.192022263	0.209677419	0.46875
14	0.011131725	0	0
15	0.003710575	0	0

J) D13S317	Muslim 1078	Christian 124	Samaritan 32
8	0.152133581	0.10483871	0.03125
9	0.069573284	0.056451613	0.03125
10	0.040816327	0.080645161	0
11	0.275510204	0.282258065	0.375
12	0.327458256	0.35483871	0.5625
13	0.10296846	0.096774194	0
14	0.029684601	0.024193548	0
15	0.001855288	0	0

K) vWA	Muslim 1078	Christian 124	Samaritan 32
11	0.001855288	0	0
13	0.003710575	0	0
14	0.093692022	0.056451613	0
15	0.107606679	0.064516129	0.03125
16	0.229128015	0.233870968	0.34375
17	0.269016698	0.322580645	0.53125
18	0.206864564	0.161290323	0.09375
19	0.070500928	0.137096774	0
20	0.016697588	0.024193548	0
21	0.000927644	0	0

L) D2S441	Muslim 1078	Christian 124	Samaritan 32
9	0.003710575	0	0
10	0.125231911	0.056451613	0.1875
11	0.324675325	0.387096774	0.53125
11.3	0.096474954	0.112903226	0
12	0.064007421	0.032258065	0.09375
12.3	0.001855288	0	0
13	0.031539889	0.032258065	0.03125
13.3	0.002782931	0	0
14	0.314471243	0.338709677	0.15625
15	0.028756957	0.040322581	0
16	0.006493506	0	0

M) D8S1179	Muslim 1078	Christian 124	Samaritan 32
7	0.000927644	0	0
8	0.012987013	0.008064516	0
9	0.009276438	0.016129032	0
10	0.067717996	0.048387097	0.0625
11	0.09554731	0.080645161	0.09375
12	0.127087199	0.088709677	0.09375
13	0.255102041	0.282258065	0.0625
14	0.192022263	0.161290323	0.5
15	0.17903525	0.201612903	0.03125
16	0.043599258	0.10483871	0.15625
17	0.012059369	0.008064516	0
18	0.003710575	0	0
23	0.000927644	0	0

N) D2S1338	Muslim 1078	Christian 124	Samaritan 32
16	0.042671614	0.024193548	0
17	0.22541744	0.258064516	0.5
18	0.135435993	0.169354839	0
19	0.119666048	0.129032258	0.03125
20	0.142857143	0.14516129	0
21	0.048237477	0.080645161	0.40625
22	0.040816327	0.040322581	0
23	0.086270872	0.10483871	0.0625
24	0.081632653	0.032258065	0
25	0.06864564	0.016129032	0
26	0.00742115	0	0
27	0.000927644	0	0

O) D1S1656	Muslim 1078	Christian 124	Samaritan 32
8	0.002782931	0	0
10	0.001855288	0	0
11	0.079777365	0.032258065	0
12	0.093692022	0.14516129	0.28125
13	0.091836735	0.137096774	0.21875
14	0.111317254	0.096774194	0
14.3	0.001855288	0	0
15	0.184601113	0.120967742	0.1875
15.3	0.046382189	0.024193548	0
16	0.179962894	0.177419355	0
16.3	0.038961039	0.080645161	0
17	0.05380334	0.129032258	0
17.3	0.073283859	0.040322581	0.25
18	0.006493506	0	0
18.3	0.028756957	0.016129032	0
19.3	0.004638219	0	0.0625

P) Penta D	Muslim 1078	Christian 124	Samaritan 32
2.2	0.019480519	0	0
5	0.000927644	0	0
6	0.008348794	0	0
7	0.005565863	0	0
8	0.02690167	0	0.125
9	0.177179963	0.137096774	0
10	0.167903525	0.169354839	0.125
10.1	0	0	0.25
11	0.186456401	0.225806452	0.15625
12	0.119666048	0.064516129	0.125
13	0.154916512	0.209677419	0.1875
14	0.078849722	0.112903226	0
15	0.04174397	0.056451613	0.03125
16	0.008348794	0.024193548	0
17	0.002782931	0	0
18	0.000927644	0	0

Q) D12S391	Muslim 1078	Christian 124	Samaritan 32
13	0.000927644	0	0
14.2	0.000927644	0	0
15	0.020408163	0.024193548	0.03125
16	0.019480519	0.016129032	0
17	0.112244898	0.064516129	0
17.3	0.006493506	0.016129032	0
18	0.183673469	0.241935484	0.03125
18.1	0.000927644	0	0
18.3	0.018552876	0.064516129	0
19	0.127087199	0.072580645	0.25
19.1	0.000927644	0	0
19.3	0.006493506	0	0
20	0.119666048	0.129032258	0.21875
20.1	0.000927644	0	0
21	0.084415584	0.120967742	0.34375

R) D19S433	Muslim 1078	Christian 124	Samaritan 32
10	0.000927644	0	0
11	0.013914657	0.024193548	0
12	0.098330241	0.048387097	0
12.2	0.001855288	0	0
13	0.204081633	0.282258065	0.09375
13.2	0.040816327	0.024193548	0
14	0.273654917	0.338709677	0
14.2	0.052875696	0.040322581	0
15	0.123376623	0.120967742	0.25
15.2	0.089053803	0.048387097	0.46875
16	0.060296846	0.048387097	0
16.2	0.023191095	0.024193548	0.1875
17	0.00742115	0	0
17.2	0.003710575	0	0
18.2	0.005565863	0	0
21	0.000927644	0	0

22	0.131725417	0.040322581	0.125
23	0.102040816	0.112903226	0
24	0.044526902	0.048387097	0
25	0.015769944	0.024193548	0
26	0.001855288	0.024193548	0
27	0.000927644	0	0

S) **FGA**	**Muslim** 1078	**Christian** 124	**Samaritan** 32
17	0.003710575	0	0
18	0.005565863	0	0
19	0.060296846	0.024193548	0
20	0.097402597	0.120967742	0
21	0.158627087	0.153225806	0.03125
21.2	0.005565863	0.008064516	0
22	0.163265306	0.274193548	0.3125
22.2	0.002782931	0	0
22.3	0.000927644	0	0
23	0.178107607	0.153225806	0.03125
23.2	0.002782931	0.024193548	0
24	0.159554731	0.129032258	0.59375
24.2	0.000927644	0	0
25	0.093692022	0.10483871	0
25.2	0.000927644	0	0
26	0.051020408	0.008064516	0.03125
27	0.008348794	0	0
28	0.003710575	0	0
29	0.001855288	0	0
30	0.000927644	0	0

T) **D18S51**	**Muslim** 1078	**Christian** 124	**Samaritan** 32
9	0.001855288	0.008064516	0
10	0.005565863	0	0
10.2	0.002782931	0	0
11	0.020408163	0.016129032	0.1875
12	0.136363636	0.177419355	0.03125
13	0.178107607	0.120967742	0.03125
14	0.177179963	0.177419355	0.3125
14.2	0.000927644	0	0
15	0.104823748	0.10483871	0
16	0.114100186	0.10483871	0.125
17	0.096474954	0.177419355	0.28125
18	0.090909091	0.080645161	0
19	0.038961039	0.032258065	0.03125
20	0.021335807	0	0
21	0.006493506	0	0
22	0.003710575	0	0

U) **D21S11**	**Muslim** 1078	**Christian** 124	**Samaritan** 32
26	0.001855288	0	0
27	0.025046382	0	0
28	0.161410019	0.120967742	0.0625
29	0.255102041	0.298387097	0.0625
29.2	0.000927644	0	0
30	0.230983302	0.330645161	0.59375
30.2	0.020408163	0	0.15625
31	0.045454545	0.024193548	0
31.2	0.108534323	0.120967742	0
32	0.00742115	0	0
32.2	0.094619666	0.072580645	0.125
33	0.000927644	0	0
33.2	0.035250464	0.024193548	0
34	0.003710575	0	0
34.2	0.004638219	0.008064516	0
35	0.001855288	0	0
36	0.001855288	0	0

V) SE33	Muslim 1078	Christian 124	Samaritan 32
6.3	0.003710575	0	0
7.3	0.001855288	0	0
8	0	0.008064516	0
9	0.012059369	0.016129032	0
10	0.000927644	0	0
12	0.001855288	0.032258065	0
12.1	0.000927644	0	0
13	0.027829314	0.024193548	0
13.2	0.000927644	0	0
13.3	0	0.008064516	0
14	0.031539889	0.048387097	0
14.3	0.001855288	0	0
15	0.043599258	0.064516129	0
15.3	0.001855288	0.008064516	0
16	0.048237477	0.040322581	0
16.3	0.003710575	0	0
17	0.092764378	0.080645161	0
17.2	0.002782931	0	0
17.3	0	0.016129032	0
18	0.096474954	0.088709677	0
18.1	0.003710575	0	0
18.2	0.000927644	0	0
18.3	0.000927644	0.008064516	0
19	0.065862709	0.056451613	0.09375
19.1	0.000927644	0	0
19.3	0.002782931	0	0
20	0.045454545	0.024193548	0.09375
20.2	0.002782931	0	0
21	0.025974026	0.008064516	0
21.2	0.010204082	0.016129032	0
22	0.004638219	0	0
22.2	0.017625232	0.008064516	0
23	0.000927644	0	0
23.2	0.038033395	0.032258065	0
24	0.001855288	0	0
24.2	0.02690167	0.048387097	0
25	0.001855288	0	0
25.2	0.037105751	0.008064516	0
26	0.000927644	0	0
26.2	0.040816327	0.008064516	0.21875
27.2	0.042671614	0.096774194	0.21875
28.2	0.049165121	0.072580645	0.0625
29.2	0.093692022	0.040322581	0.15625
30.2	0.051948052	0.048387097	0.0625
31.2	0.02690167	0.048387097	0.09375
32	0.000927644	0.008064516	0
32.2	0.009276438	0.016129032	0
33	0.001855288	0	0
33.2	0.00742115	0	0

W) Penta E	Muslim 1078	Christian 124	Samaritan 32
5	0.057513915	0.048387097	0.0625
7	0.094619666	0.169354839	0.125
8	0.038033395	0.008064516	0.125
9	0.017625232	0	0
10	0.06122449	0.080645161	0
11	0.114100186	0.14516129	0.375
12	0.180890538	0.193548387	0.25
13	0.107606679	0.129032258	0
13.3	0	0.008064516	0
14	0.055658627	0.024193548	0
14.2	0.000927644	0	0
14.4	0.001855288	0	0
15	0.067717996	0.024193548	0
15.4	0.000927644	0	0
16	0.057513915	0.032258065	0
16.4	0.000927644	0.008064516	0
17	0.046382189	0.040322581	0
18	0.049165121	0.016129032	0.0625
19	0.027829314	0.032258065	0
20	0.012059369	0.040322581	0
21	0.005565863	0	0
22	0.000927644	0	0
23	0.000927644	0	0

*red color indicates alleles differentiated between subpopulations.

Table.11: HWE exact test p-values for each subpopulation and F-statistics.*

Locus	pHW Muslim	pHW Christian	pHW Samaritan	F_{IS}	F_{ST}	F_{IT}
D3S1358	0.3812	0.8394	0.6588	0.0334	0.0314	0.0638
D1S1656	0.2256	0.0050	0.0006	0.0478	0.0193	0.0662
D2S441	0.0700	0.0273	0.3035	0.0066	0.0104	0.0169
D10S1248	0.0735	0.9940	0.3906	0.0018	0.0037	0.0056
D13S317	0.2901	0.7094	0.3652	0.0111	0.0085	0.0196
Penta E	0.2707	0.3678	0.9798	0.0102	0.0125	0.0226
D16S539	0.3538	0.5947	0.9155	0.006	0.0107	0.0167
D18S51	0.0599	0.8174	0.9420	0.0339	0.0145	0.048
D2S1338	0.0601	0.5472	0.0433	0.0428	0.0297	0.0712
CSF1PO	0.0389	0.5415	0.1214	0.0468	0.009	0.0554
Penta D	0.4977	0.0816	0.8910	0.0303	0.0129	0.0427
TH01	0.4138	0.3698	0.6465	0.0639	0.0175	0.0802
vWA	0.5712	0.6612	1.0000	0.0433	0.014	0.0567
D21S11	0.0085	0.1334	0.2695	0.0175	0.0279	0.045
D7S820	0.9721	0.8846	0.7786	-0.0074	0.0031	-0.0043
D5S818	0.3939	0.2892	0.6085	0.0375	0.0221	0.0588
TPOX	0.0083	0.4025	1.0000	0.0822	0.0331	0.1127
D8S1179	0.5155	0.0976	0.5219	0.0489	0.0189	0.0669
D12S391	0.1481	0.2220	0.9296	0.0142	0.02	0.0339
D19S433	0.0328	0.1665	0.4779	0.047	0.0385	0.0836
SE33	0.0855	0.9713	0.0588	0.0219	0.0102	0.0319
D22S1045	0.0229	0.3347	0.7780	0.0269	0.0096	0.0362
FGA	0.0298	0.2439	0.0845	0.043	0.0349	0.0765
			All	**0.0304**	**0.0180**	**0.0479**

*red color indicates loci with significant deviations that became insignificant after Bonferroni correction (p-value>0.00217). Yellow color indicted significant deviation even after Bonferroni correction (p-value<0.00217)

3.2.10 The West Bank regions: middle, north, and south

In order to reveal if there is any differentiation between the north, middle, and south of the West Bank that can result due to isolation-by-distance or the presence of subpopulations within each region Pairwise Fst was conducted. In other words, we measured the amount of genetic variance that can be explained by population structure if present based on Wright's F-statistics (Wright, 1965) using STRAF and Genepop software. Our results indicates insignificant differentiation between the three regions of the West Bank with very low F_{ST} values (Table.12). Moreover, when we conducted an overall locus-by-locus Analysis of MOlecular VAriance (AMOVA) test using Arlequin v3.5.2.1, the resulting overall fixation index (F_{ST}) of 0.00090

showed no significant variation between the three regions. AMOVA analysis also revealed that the source of variation of the Palestinian population in these three regions is mainly because of variation within populations (meaning at the individual level) rather than among populations residing in these three regions with a percentage variation of 99.90% (within populations) and 0.09% (among populations) with a p-value of 0.0000, respectively.

Table.12: Pairwise F_{ST} matrix analysis for the three divisions of the West Bank.

Subpopulation/ F_{ST}	North	Middle
South	0.001013715	4.26606E-06
Middle	0.001270048	

3.2.11 Differentiation between the West Bank and Gaza Strip

A locus-by-locus analysis for population differentiation between the West Bank and Gaza Strip was performed using Fisher's exact probability test to determine whether there are significant variations between them for the shared 15 loci (D3S1358, TH01, D21S11, D18S51, Penta E, D5S818, D13S317, D7S820, D16S539, CSF1PO, Penta D, vWA, D8S1179, TPOX, FGA)between the two subpopulations. Moreover, seven individuals who self-declared as being born and belonged to Gaza Strip in our cohort were combined with the previously published STRs data for125 samples for Gaza Strip(Abu Halima et al., 2009b), thus a total of 610 samples from the West Bank were compared 133 samples from Gaza Strip. Two loci (THO1 and TPOX) showed significant differentiation with a P-value <0.05. However, after applying Bonferroni's correction (0.00217) only TPOX exhibited significant differentiation. Furthermore, when an overall pairwise F_{ST} was computed for all loci using STRAF software, the resulting fixation index (F_{ST}) of 0.0011 showed no significant variation between the overall two datasets from Gaza Strip and the West Bank (Table.13).

Table.13: Population differentiation locus-by-locus exact test between the two Palestinian datasets from the West Bank and Gaza Strip for each locus.

Locus	P-Value +-S.E	Locus	P-Value +-S.E
D3S1358	0.34203 +- 0.017869	CSF1PO	0.40919 +- 0.01603
TH01	0.03657 +- 0.00695855	D16S539	0.25628 +- 0.0170016
D21S11	0.30126 +- 0.0171738	Penta D	0.52395 +- 0.019194
D18S51	0.26388 +- 0.0170876	vWA	0.34643+- 0.0170247
Penta E	0.71113 +- 0.0181002	D8S1179	0.61172 +- 0.0180113
D5S818	0.66944 +- 0.0142745	TPOX	5e-05 +- 5e-05
D7S820	0.8343 +- 0.0129656	FGA	0.67725 +- 0.0186396
D13S317	0.18612 +- 0.0140938		

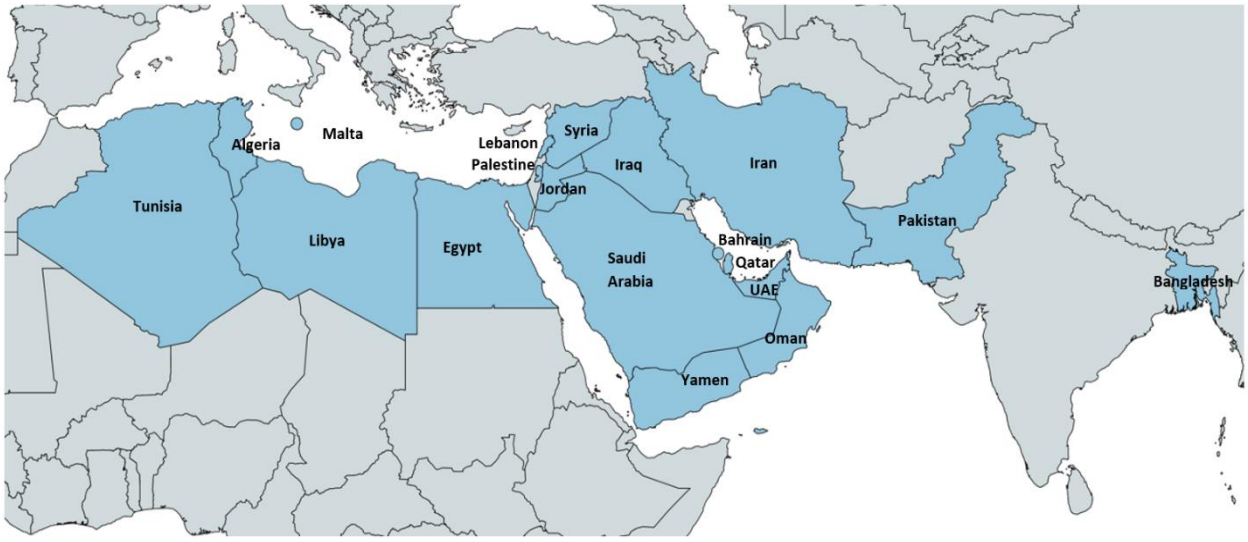
*red color indicates significantly differentiated loci p<0.05 and p-value<0.00217 before and after applying Bonferroni's correction.

3.3 A comparative analysis of autosomal STRs allele frequency of the Palestinian population and surrounding populations

In order to study the relationship from a genetic perspective between the Palestinian population and other populations. The allele frequency generated in this study was compared to another 18 countries in the Middle East, South Asia, and North Africa using 13 shared autosomal STRs (D3S1358, vWA, D8S1179, FGA, CSF1PO, TPOX, TH01, D21S11, D18S51, D5S818, D13S317, D7S820, D16S539). To complete this meta-analysis, previously published population data from these countries were collected. The Middle East populations included UAE(Garcia-Bertrand et al., 2014), Saudi Arabia(Khubrani et al., 2019), Yemen(Alshamali et al., 2005), Oman(Alshamali et al., 2005), Qatar(Pérez-Miranda et al., 2006), Iraq(Barni et al., 2007), Jordan(Al-Eitan & Tubaishat, 2018), Syria(Abdin, Shimada, et al., 2003), Bahrain(Al-Snan et al., 2019), Lebanon(El Andari et al., 2013), and Palestine (current study). North African countries included Egypt(Coudray et al., 2007), Libya(Khodjet-El-Khil et al., 2012), Algeria(Bosch et al., 2001), Tunisia(L. Cherni et al., 2005), and Malta(Cassar et al., 2008)

(Morocco had missing data for some loci so it was excluded). South Asian countries included Iran(Shepard & Herrera, 2006), Pakistan(Rakha et al., 2009), and Bangladesh(Ferdous et al., 2009) (Fig.22).

Figure.22: Geographical locations of the countries with published population data used for the meta-analysis in this study. Map generated using MapChart website (https://www.mapchart.net/world.html).



The locus-by-locus exact test was carried out to between each of the 19 populations and compared for the shared 13 loci. Figure 23 summarizes the number of loci with significant differences in the F_{ST} values (p-value< 0.05) between every two populations. Initially, we compared the Middle Eastern countries to each other, we found that the highest different loci number were found between Saudi Arabia and Jordan (13/13 loci), Saudi Arabia and Lebanon (11/13 loci), Saudi Arabia and Palestine (10/13 loci), Saudi Arabia and Bahrain (10/13 loci). The lowest number of significantly differentiated loci with Palestine were Lebanon (3/ 13 loci), Oman (2/13), and Jordan (4/13 loci). Among all of the Middle Eastern countries, Yemen and Saudi Arabia, Palestine and Oman showed the lowest number of significantly differentiated loci (2/13). When comparing North African populations together, Libya and Algeria showed the lowest significant number of loci (3/13), followed by Malta and Algeria (4/13 loci), and Libya and Malta (4/13 loci). Egypt

showed more significant differentiation compared to other North African countries with 8 loci with Libya, 9 loci with Tunisia, and 10 loci with Algeria and Malta. Tunisia showed significant differentiation even with the neighboring North African populations with Libya (11/13 loci), Malta (12/13 loci), and Algeria (10/13). South Asian countries showed different degree of differentiation with Bangladesh and Iran exhibiting the highest number of loci with (7/13 loci) and the lowest degree of differentiation was between Pakistan and Iran (5/13 loci). When we compared the number of significantly differentiated loci between Palestine and North Africa we found the highest differentiation was between Palestine and Tunisia (13/13 loci), Egypt (8/13 loci), and Algeria (7/13 loci). Palestine and Libya showed a low number of differentiated loci (4/13). Finally, when we compared Palestine with South Asian countries, we found the highest degree of differentiation was for Pakistan and Bangladesh (9/13 loci). Iran showed a low number of differentiated loci compared to Palestine (4/13 loci) (Fig.23). Thus, we conclude that the Palestinian population share common history and genetic makeup with the neighboring Middle Eastern countries.

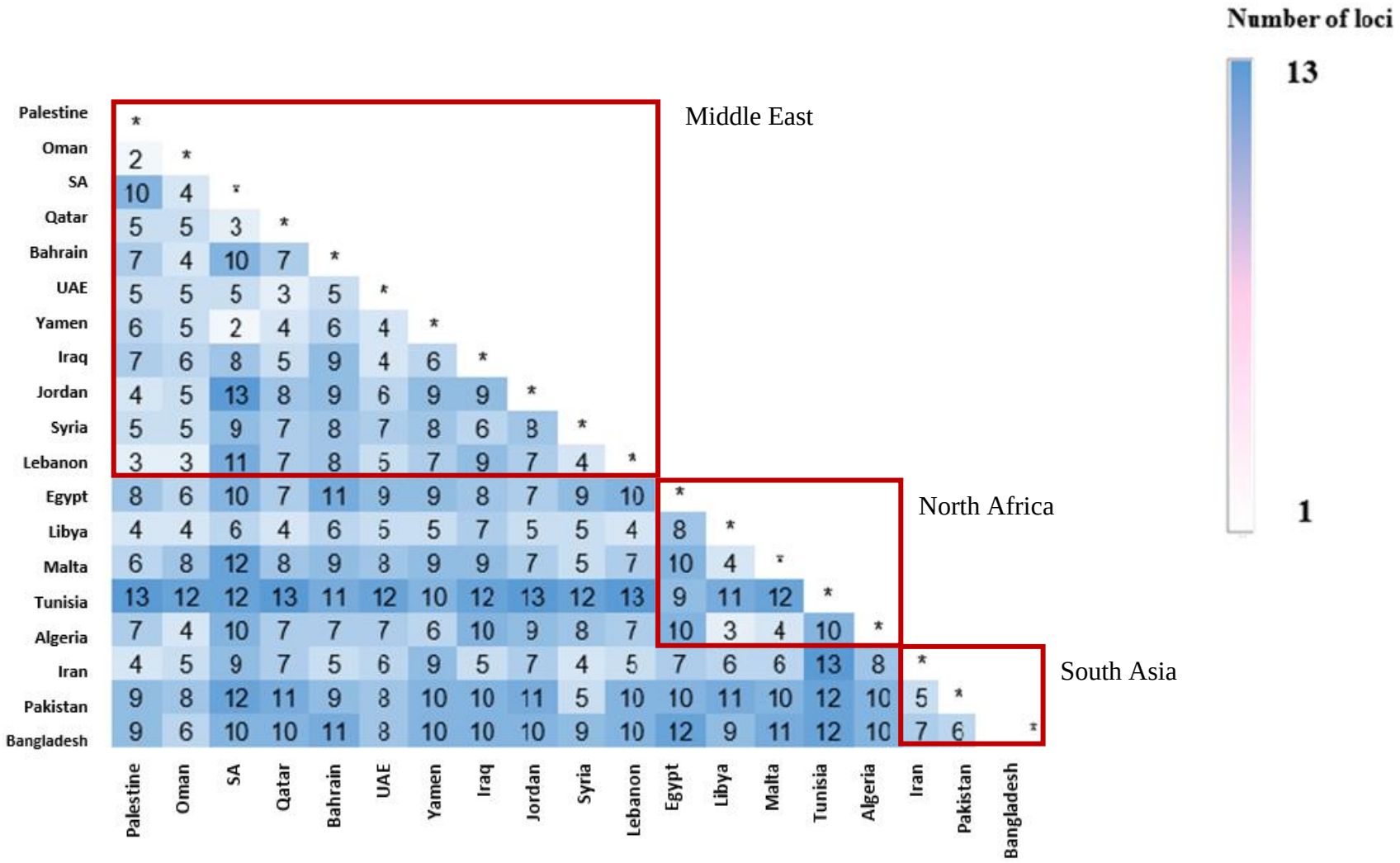
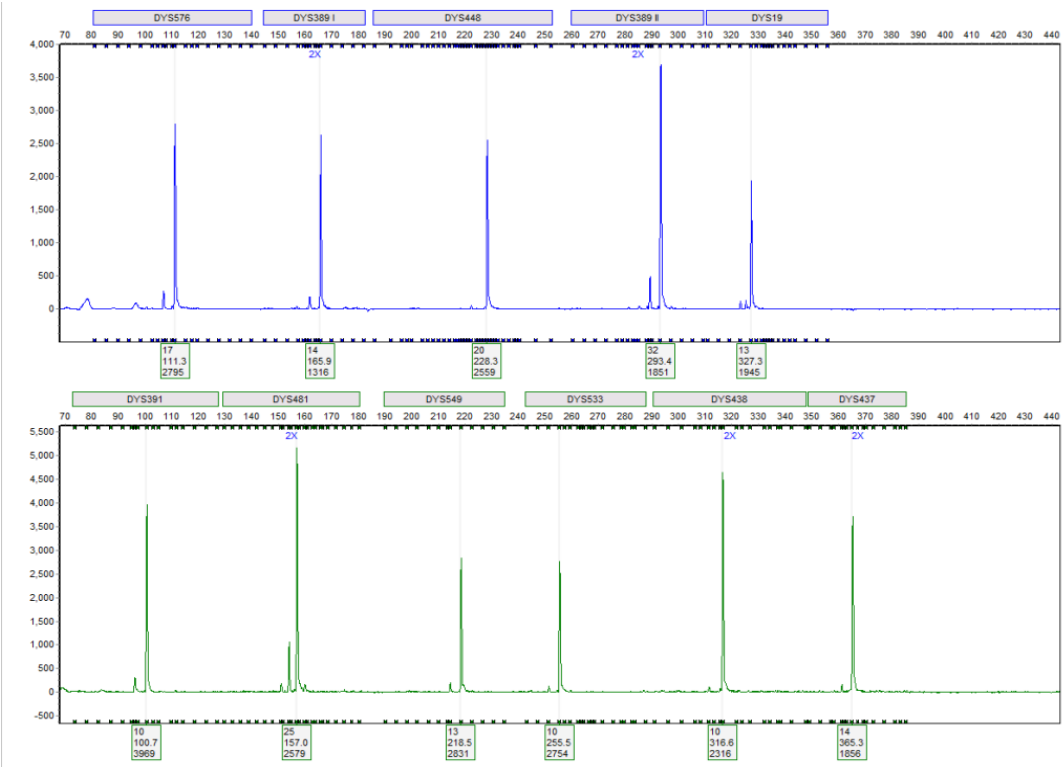


Figure.23: Heat map showing the number of significantly different loci (p-value<0.05) between two populations for the shared 13 loci.

3.4 Y-STRs Haplotypes can be used to differentiate lineages in the Palestinian population

3.4.1 Y-STRs electropherograms and fingerprints

In order to test the utility of using Y-STRs to differentiate lineages in the Palestinian population, 303 individual male were genotyped using PowerPlex® Y23 system. Then, Then capillary electrophoresis was used to generate their DNA profile for 23 Y-chromosome STR markers (DYS576, DYS389I/II, DYS448, DYS19, DYS391, DYS481, DYS549, DYS533, DYS438, DYS437, DYS570, DYS635, DYS390, DYS439, DYS392, DYS643, DYS393, DYS458, DYS385a/b, DYS456, and Y-GATA-H4). Figure 24 represents an example of the Y-STR profile generated throughout our study.



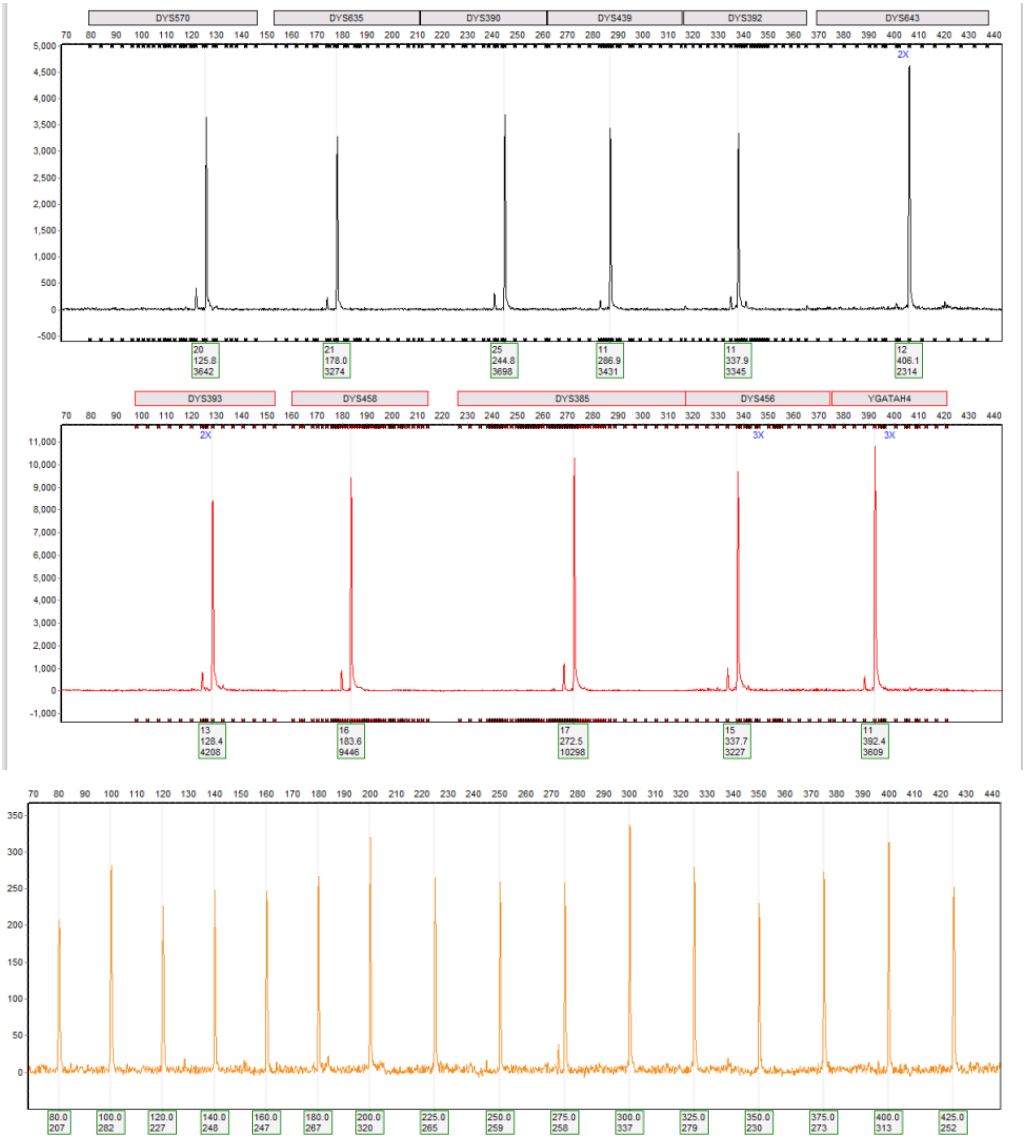


Figure.24: A random male STR profile as obtained from the PowerPlex® Y23 system. This profile shows the different dyes that correspond to the different loci. The first number below each peak indicates the number of repeats for each allele (allele call), below the allele call is the allele size in bp, and finally, the last number indicates the height in relative fluorescence units (RFU). 500 WEN Internal lane standard is shown in orange color.

3.4.2 Haplotype diversity

After genotyping all the loci for all study participants, we wanted to test the Y-STR haplotype diversity in the Palestinian population. Out of 303 Y-STR profiles generated, 288 different haplotypes were detected among individuals genotyped. 275 (90.8% of samples) haplotypes were unique and detected only once thorough our cohort with a frequency of 0.35%. In contrast, 13-shared haplotypes were observed in different frequencies. Each of the haplotype 7 (H7), H14, H63,

H87, H91, H99, H100, H199, H206, H216, and H288 were shared between two participants (7.2% op haplotypes) with a frequency of 0.7%. However, each of H106 and 287 was shared between three participants (2% of haplotypes) with a frequency of 1% (Table.14); these haplotypes are considered more common in the Palestinian population and have less discrimination power compared to the unique haplotypes.

3.4.3 Y-STRs Allelic Frequencies

Allele frequencies ranged from 79.2% (DYS392*11) to 0.33% (DYS19*17, DYS389I*15*16, DYS391*8, DYS392*9*15, DYS438*8, DYS448*16, DYS458*16.2*20, DYS481*20*30, DYS533*14*19, DYS570*14, DYS635*17*25, DYS643*14). A total of 156 different alleles were genotyped along all loci. The Number of Alleles ranged from 3 alleles at the DYS437 locus to 15 alleles genotyped at the DYS458 locus. Few intermediate alleles were reported in our study, 19.2 was reported in locus DYS448. Additionally, 16.2, 17.2, 18.2, 19.2, 20.2, 21.2, and 22.2 were seen at locus DYS458, and intermediate allele 15.2 was reported at locus DYS385a/b. The average gene diversity for 23 Y-STRs was 65%, ranging from 35.98% (DYS392) to 88.6% (DYS458). Gene diversity of loci DYS576, DYS389 II, DYS481, DYS570, DYS390, DYS643, DYS456, DYS635, and DYS385b exceeded 70% (Table.15 & Fig.25). Interestingly, no novel variants reported in our cohort. The most predominant and common allele was allele 11 of DYS392 with 79.21%.

3.4.4 Y-STRs population genetics and forensic parameters

The overall haplotype diversity for the Palestinian population at 23 Y-STRs was 99.9312812%, which indicates that the Palestinian population is quite diverse. Moreover, the discrimination capacity of the 23 loci was 95.049505%. The overall haplotype random match probability was 0.398525%. The low haplotype match probability, high haplotype diversity, and high discrimination capacity presented in our cohort are indications of the uniqueness of our population, which results in a highly useful system for forensic application (Table.15).

Table.14: Y-STRs Haplotypes detected in the West Bank (Sample size= 303)*.

HN	HF	576	389 I	448	389 II	19	391	481	549	533	438	437	570	635	390	439	392	643	393	458	385 a	385 b	456	Y H4	n
H1	0.0035	18	14	20	30	14	11	22	12	11	9	16	18	20	24	11	11	9	12	16	13	15	15	10	1
H2	0.0035	16	13	21	30	14	10	23	13	12	9	15	18	23	23	13	11	11	12	15	15	17	15	12	1
H3	0.0035	19	13	20	30	15	10	22	13	12	10	14	20	21	24	11	11	11	13	19	17		16	13	1
H4	0.0035	18	13	20	29	14	11	25	12	11	10	14	18	21	22	11	11	9	12	18.2	13	19	13	11	1
H5	0.0035	16	12	20	28	14	10	26	13	11	10	16	20	22	22	11	11	11	13	14	14		14	11	1
H6	0.0035	17	13	20	28	14	11	24	12	11	11	14	18	21	22	11	11	9	12	18.2	13	18	13	11	1
H7	0.007	15	13	21	32	13	10	26	12	11	10	14	18	21	24	12	11	11	12	19	16	18	15	11	2
H8	0.0035	17	13	19	31	13	11	25	12	11	9	14	17	21	23	12	14	11	13	15	14	15	15	11	1
H9	0.0035	18	14	22	31	16	10	26	11	10	10	16	15	19	21	13	11	9	12	19	11	12	14	12	1
H10	0.0035	17	14	20	31	14	10	25	12	11	10	14	18	22	23	11	11	9	12	17.2	13	18	15	11	1
H11	0.0035	18	13	20	30	15	11	26	13	11	10	14	18	21	23	11	11	9	12	18.2	13	20	17	11	1
H12	0.0035	18	13	20	30	13	10	23	12	11	10	14	17	24	24	12	11	12	14	15	16	17	15	12	1
H13	0.0035	18	13	20	30	13	10	24	12	12	11	14	18	21	23	11	11	13	13	16	16		15	11	1
H14	0.007	18	13	19.2	30	16	10	25	11	8	10	14	19	21	24	12	11	12	13	15	13	17	14	12	2
H15	0.0035	17	13	20	30	14	11	26	12	11	11	14	18	21	24	11	11	9	12	18.2	13	18	15	11	1
H16	0.0035	16	13	19	28	13	10	23	12	11	11	14	16	23	22	12	16	12	13	18	14	16	16	10	1
H17	0.0035	18	14	20	29	14	11	25	12	12	10	14	20	21	23	10	11	9	12	18.2	13	18	14	12	1
H18	0.0035	18	13	20	29	14	10	26	12	11	10	14	17	22	22	11	11	9	12	18.2	14	17	13	11	1
H19	0.0035	17	14	18	31	14	11	23	14	13	9	14	17	22	24	12	13	10	13	17	15		15	11	1
H20	0.0035	18	14	20	30	14	11	22	12	11	9	16	17	21	24	11	11	9	12	17	13	15	15	10	1
H21	0.0035	17	14	20	31	14	11	22	12	11	9	16	17	20	24	12	11	9	12	16	13	15	15	10	1
H22	0.0035	20	13	20	28	14	10	25	12	11	10	14	18	21	23	11	11	9	13	17.2	13	20	15	11	1
H23	0.0035	16	13	20	30	14	10	25	12	11	10	14	19	20	23	11	11	9	12	17.2	13	17	14	11	1
H24	0.0035	19	13	20	29	14	10	25	13	11	10	14	17	21	23	12	11	9	13	17.2	13	20	15	11	1
H25	0.0035	18	14	16	30	15	10	23	12	12	11	15	15	23	22	11	15	11	11	17	12	17	15	11	1
H26	0.0035	18	13	20	30	14	11	27	13	11	10	14	18	21	23	11	11	9	12	18.2	13	18	14	11	1
H27	0.0035	18	13	20	29	14	11	27	12	11	10	14	19	21	22	11	11	9	12	19.2	13	18	13	11	1
H28	0.0035	18	14	21	31	13	10	27	12	8	11	15	19	22	26	13	11	12	14	16	14	18	16	12	1
H29	0.0035	16	13	19	30	14	10	21	13	12	12	15	17	23	24	12	13	10	12	17	11	15	16	14	1
H30	0.0035	17	13	20	29	13	11	26	12	10	10	14	19	21	22	11	11	9	12	20.2	14	18	13	11	1
H31	0.0035	16	14	20	31	13	10	25	14	10	10	14	19	21	25	11	11	12	13	16	17		15	11	1
H32	0.0035	19	13	20	29	14	11	26	12	11	10	14	20	21	22	11	11	9	12	19.2	13	18	13	11	1
H33	0.0035	16	12	19	28	14	10	26	14	11	10	16	20	21	22	11	11	11	13	14	14		14	11	1
H34	0.0035	16	13	18	30	14	11	22	14	13	9	14	17	21	24	12	13	10	13	17	15		15	11	1
H35	0.0035	18	12	20	30	13	10	26	11	10	10	14	20	21	25	12	11	12	13	18	16	18	15	9	1
H36	0.0035	17	13	20	29	14	11	25	12	11	10	14	20	21	22	11	11	9	12	19.2	12	18	12	11	1
H37	0.0035	18	14	20	30	14	10	22	13	11	9	16	17	20	24	11	11	9	12	16	13	15	15	10	1
H38	0.0035	19	12	23	30	11	10	24	12	12	11	14	16	22	24	12	12	13	13	14	17	18	15	11	1

H39	0.0035	17	14	20	31	13	10	24	13	10	10	14	20	21	26	11	11	12	13	16	17		15	11	1
H40	0.0035	16	13	20	31	14	10	27	13	12	10	14	18	20	23	12	11	9	12	17.2	13	17	15	11	1
H41	0.0035	19	13	20	31	14	12	26	13	11	10	14	17	22	23	11	11	9	12	18.2	13	20	14	11	1
H42	0.0035	18	13	20	29	13	9	28	11	11	10	14	23	24	23	12	11	11	13	17	13	14	15	12	1
H43	0.0035	16	12	22	29	15	11	21	13	9	10	16	18	20	22	12	11	11	14	16	13	14	15	11	1
H44	0.0035	16	13	20	31	13	10	25	12	10	10	14	18	23	24	12	11	12	13	21	16	17	15	11	1
H45	0.0035	18	13	20	30	14	10	25	12	12	10	14	18	20	23	11	11	9	12	20.2	13	18	14	12	1
H46	0.0035	17	13	20	31	14	10	25	12	11	10	14	18	21	23	11	11	9	13	17.2	13	18	15	11	1
H47	0.0035	18	13	20	29	14	11	24	12	11	10	14	19	21	23	11	11	9	12	18.2	13	18	14	11	1
H48	0.0035	18	13	20	29	14	10	27	12	11	10	14	17	22	22	11	11	9	12	18.2	13	17	13	11	1
H49	0.0035	16	13	19	30	15	11	22	14	12	12	15	17	23	24	12	13	10	12	16	11	14	15	12	1
H50	0.0035	15	13	19	30	15	11	22	14	12	12	15	17	23	24	12	13	10	12	16	11	14	15	12	1
H51	0.0035	18	14	20	30	14	11	22	12	11	9	16	17	20	24	11	11	9	12	16	13	15	15	10	1
H52	0.0035	17	13	22	30	15	11	21	14	9	10	16	18	20	22	13	11	10	14	16	13	14	13	11	1
H53	0.0035	19	14	20	30	14	11	22	12	11	9	16	18	20	24	11	11	9	12	16	13	15	15	10	1
H54	0.0035	17	13	20	29	14	11	25	12	11	10	14	19	21	22	11	11	9	12	20.2	12	17	13	11	1
H55	0.0035	17	12	19	27	14	10	22	11	12	10	15	16	21	22	12	14	10	11	16	14	19	16	13	1
H56	0.0035	15	13	22	31	15	11	27	11	11	11	14	20	21	21	13	11	13	14	16	15	17	15	11	1
H57	0.0035	15	13	20	30	13	10	25	14	11	11	14	18	22	24	12	11	12	14	16	15	19	15	11	1
H58	0.0035	16	12	19	27	14	11	23	12	12	10	15	15	25	22	12	14	11	11	15	13	18	16	12	1
H59	0.0035	17	14	20	32	15	10	28	12	12	10	14	22	22	23	11	12	12	14	16	14	15	13	11	1
H60	0.0035	18	13	19	30	14	10	24	12	11	9	15	16	21	23	11	11	11	12	17	14	16	15	11	1
H61	0.0035	17	13	20	29	15	12	26	12	11	10	14	19	21	22	11	11	9	12	20.2	13	17	13	11	1
H62	0.0035	17	12	19	28	15	10	21	15	13	9	16	18	22	24	12	11	9	12	18	14	17	13	11	1
H63	0.007	17	14	20	31	13	10	25	13	10	10	14	20	21	25	11	11	12	13	16	17		15	11	2
H64	0.0035	18	13	20	31	13	11	24	11	12	10	14	18	23	25	12	11	12	13	19	15	17	15	11	1
H65	0.0035	17	14	20	31	13	10	25	14	10	10	14	19	21	25	11	11	12	13	16	17		15	11	1
H66	0.0035	16	13	20	29	14	11	26	12	11	10	14	19	21	22	11	11	9	12	22.2	13	17	12	11	1
H67	0.0035	16	12	20	28	14	10	26	13	11	10	16	20	21	22	11	11	11	13	15	14		14	11	1
H68	0.0035	19	13	20	29	14	11	26	12	11	10	14	19	21	22	11	11	9	12	18.2	13	18	13	11	1
H69	0.0035	16	13	18	30	14	12	22	14	13	9	14	17	22	24	12	13	10	13	17	15		15	11	1
H70	0.0035	16	13	18	30	14	11	22	14	13	9	14	17	22	24	13	13	10	13	17	15		15	11	1
H71	0.0035	18	13	20	29	14	10	23	12	11	9	14	17	24	23	10	11	10	13	15	13	18	17	11	1
H72	0.0035	19	13	20	29	14	11	26	12	11	10	14	19	21	22	11	11	9	12	19.2	13	18	13	11	1
H73	0.0035	15	13	20	29	15	10	22	12	11	9	15	16	23	23	11	11	10	12	16	13	16	16	11	1
H74	0.0035	18	13	20	30	13	10	22	12	12	10	14	17	21	24	13	11	13	12	15	18		15	11	1
H75	0.0035	19	13	19	30	14	10	21	13	12	12	15	17	23	25	12	12	10	12	17	11	14	15	14	1
H76	0.0035	17	13	21	30	14	10	22	12	11	9	15	19	24	23	11	12	10	13	15	13	15	15	11	1
H77	0.0035	17	13	20	30	14	10	20	12	12	10	14	20	20	25	12	11	12	13	16	18	19	15	13	1
H78	0.0035	18	14	20	30	13	9	27	12	10	10	14	23	21	25	10	11	13	13	18	13	14	14	11	1
H79	0.0035	18	13	20	30	13	10	26	11	11	10	14	19	24	24	11	12	11	15	16	17	19	15	11	1
H80	0.0035	18	13	20	29	14	11	26	12	10	10	14	19	22	22	11	11	9	12	19.2	13	18	13	11	1
H81	0.0035	17	13	20	31	14	10	25	12	11	10	14	18	21	24	11	11	9	12	18.2	13	17	14	11	1

H82	0.0035	17	14	21	31	14	10	21	12	12	10	14	19	20	23	11	11	9	12	18.2	13	18	14	11	1
H83	0.0035	15	12	19	28	15	10	24	12	11	9	14	19	20	25	11	11	9	12	16	16	19	14	11	1
H84	0.0035	17	13	21	29	16	10	24	13	11	9	15	18	21	23	12	11	10	12	15	13	17	15	12	1
H85	0.0035	19	12	21	28	15	10	21	12	9	11	16	16	21	23	12	11	11	14	19	14	15	15	11	1
H86	0.0035	19	13	20	29	14	10	25	13	11	10	14	19	21	22	11	11	9	13	20.2	13	13	13	11	1
H87	0.007	21	13	20	29	14	11	26	12	11	10	14	19	21	22	11	11	9	12	19.2	13	18	13	11	2
H88	0.0035	17	14	20	30	14	11	26	12	11	10	14	20	21	22	11	11	8	12	19.2	13	18	13	11	1
H89	0.0035	18	13	20	30	15	10	22	12	12	9	14	15	22	22	11	11	8	12	16	15		15	12	1
H90	0.0035	17	13	20	29	14	11	27	12	11	10	14	19	21	22	12	11	9	12	20.2	13	17	13	11	1
H91	0.007	18	13	19	29	14	11	22	12	12	12	15	17	23	25	12	13	10	13	17	10	16	17	13	2
H92	0.0035	18	14	20	31	14	11	25	11	11	10	14	17	21	25	11	11	9	12	18.2	15	19	15	11	1
H93	0.0035	19	13	20	29	14	10	24	12	11	10	14	17	20	23	12	9	9	13	17.2	13	18	15	11	1
H94	0.0035	17	14	20	32	13	10	25	13	10	10	14	20	21	25	11	11	12	13	16	17		15	11	1
H95	0.0035	20	12	19	28	14	11	22	13	12	12	15	18	23	24	12	13	10	12	16	11	15	15	12	1
H96	0.0035	17	13	20	30	13	10	25	13	10	10	14	20	21	26	11	11	12	13	17	17	18	15	11	1
H97	0.0035	16	13	21	29	15	10	25	12	11	10	14	17	20	24	11	11	9	12	17.2	14	17	15	11	1
H98	0.0035	16	14	19	31	15	11	22	13	13	12	14	18	23	24	12	13	12	13	17	13	16	15	13	1
H99	0.007	19	13	20	30	14	10	24	14	11	11	14	18	20	23	11	11	9	12	17.2	13	17	16	11	2
H100	0.007	16	14	22	31	15	11	22	12	10	10	16	18	21	21	11	11	12	14	17	14		16	11	2
H101	0.0035	18	13	20	29	14	10	23	12	11	9	16	17	20	24	11	11	9	12	17	13	15	15	10	1
H102	0.0035	16	12	19	28	15	10	25	12	11	9	14	18	20	26	11	11	9	12	16	17	19	14	11	1
H103	0.0035	17	13	20	30	14	10	25	12	11	10	14	18	21	24	11	11	9	12	19.2	13	17	14	11	1
H104	0.0035	16	12	23	28	15	10	21	12	10	10	15	18	20	21	12	11	12	15	16	12	15	18	11	1
H105	0.0035	19	13	20	30	15	10	23	12	12	10	14	21	20	24	13	11	11	13	18	16	17	16	14	1
H106	0.01	16	13	21	30	14	10	23	13	11	9	15	17	23	23	13	11	11	12	15	15	17	15	12	3
H107	0.0035	16	13	21	30	14	10	23	13	11	9	15	17	23	23	12	11	11	12	15	15	17	15	12	1
H108	0.0035	18	13	21	30	15	9	25	12	11	10	14	15	21	21	13	11	11	14	18	12		15	11	1
H109	0.0035	19	12	21	28	14	11	22	12	11	9	14	16	21	24	12	11	9	12	19	13	17	15	11	1
H110	0.0035	18	12	20	31	13	10	26	13	10	10	14	19	21	24	12	11	12	13	17	17	18	16	11	1
H111	0.0035	16	13	21	30	14	10	23	13	11	9	14	17	23	23	13	11	11	12	15	15	17	15	12	1
H112	0.0035	17	12	20	31	13	10	25	12	10	10	14	17	21	24	11	11	12	12	16	17	18	15	11	1
H113	0.0035	18	14	20	30	13	9	27	12	10	10	14	23	22	25	10	11	13	13	18	13	14	14	11	1
H114	0.0035	16	14	20	30	14	10	25	11	11	10	14	19	20	23	12	11	10	12	19.2	12	17	13	11	1
H115	0.0035	18	13	19	30	15	11	24	13	12	12	14	18	23	24	12	13	11	14	18	13	16	15	12	1
H116	0.0035	16	14	22	31	15	12	22	12	10	10	16	18	21	21	11	11	12	14	18	14		16	11	1
H117	0.0035	17	13	20	29	14	12	25	12	11	10	14	18	21	22	12	11	9	12	20.2	13	18	13	11	1
H118	0.0035	17	14	20	32	14	9	25	13	11	10	14	17	22	24	11	11	13	14	16	16	17	15	12	1
H119	0.0035	18	13	19	29	14	11	22	12	12	12	14	17	23	25	12	13	10	13	18	10	16	17	12	1
H120	0.0035	15	13	22	29	14	10	25	12	10	10	16	21	23	24	12	12	11	13	17	16		17	11	1
H121	0.0035	16	13	20	29	14	10	22	13	12	9	15	20	21	23	11	11	9	12	17	13	15.2	16	12	1
H122	0.0035	18	14	19	30	15	10	23	12	12	12	14	19	23	24	13	13	10	13	17	13	15	16	11	1
H123	0.0035	17	13	20	31	14	10	25	12	11	10	14	18	21	24	11	11	9	12	18.2	13	18	14	11	1
H124	0.0035	18	13	20	30	14	10	25	11	12	10	14	19	20	23	11	11	9	12	21.2	13	18	14	11	1

H125	0.0035	19	13	20	28	14	10	25	13	11	10	14	18	20	23	11	11	9	12	16.2	14	17	15	11	1
H126	0.0035	17	13	20	31	13	9	23	14	11	10	14	19	23	24	13	11	12	14	16	17		15	12	1
H127	0.0035	17	13	21	29	13	10	22	13	11	9	14	20	21	22	11	11	11	13	15	13	17	15	12	1
H128	0.0035	17	13	20	30	13	9	25	12	11	11	14	19	23	25	11	11	12	15	16	16	17	16	12	1
H129	0.0035	18	14	20	30	14	10	25	12	12	10	14	16	20	23	11	11	9	12	17.2	13	18	16	10	1
H130	0.0035	17	13	20	31	13	9	22	12	13	10	14	18	21	23	12	11	12	13	16	15	17	18	11	1
H131	0.0035	17	14	20	32	16	10	28	12	12	10	14	21	22	23	11	12	12	14	16	14	15	13	11	1
H132	0.0035	17	12	20	30	11	10	24	13	12	12	14	18	21	24	11	12	13	13	14	18	19	15	12	1
H133	0.0035	17	13	20	28	14	10	23	12	11	10	14	17	21	22	11	11	9	11	18.2	13	18	13	11	1
H134	0.0035	17	13	20	30	14	10	24	12	12	10	14	18	20	23	11	11	9	12	19.2	13	18	14	11	1
H135	0.0035	17	13	20	30	14	10	24	12	12	10	14	19	20	23	11	11	9	12	19.2	13	20	14	11	1
H136	0.0035	18	12	20	30	13	8	26	12	10	10	14	22	21	23	12	11	12	13	17	17	18	15	9	1
H137	0.0035	16	13	20	30	14	10	22	12	11	10	14	19	21	24	11	11	12	14	16	18	20	16	13	1
H138	0.0035	18	12	20	30	13	10	25	11	10	10	14	20	21	25	13	11	12	13	18	16	18	15	9	1
H139	0.0035	18	13	20	29	14	11	25	12	12	10	14	18	21	22	11	11	9	12	18.2	13	18	13	12	1
H140	0.0035	18	13	21	29	15	10	22	12	10	11	16	17	21	22	12	11	12	13	18	14	16	15	11	1
H141	0.0035	19	13	20	30	13	10	24	12	12	11	14	18	21	23	11	11	13	13	16	16		15	11	1
H142	0.0035	19	13	21	30	14	10	24	13	11	10	14	18	21	23	11	11	9	12	18.2	13	16	15	11	1
H143	0.0035	18	13	20	29	14	10	25	12	11	10	14	18	21	23	11	11	9	12	18.2	13	16	13	11	1
H144	0.0035	18	13	20	30	14	10	24	13	11	11	14	18	20	23	11	11	9	12	17.2	13	17	16	11	1
H145	0.0035	21	14	19	30	15	10	22	14	13	12	14	19	23	24	12	13	11	13	16	13	16	15	12	1
H146	0.0035	16	14	19	31	14	10	21	12	13	9	14	16	21	22	11	13	10	13	17	13	16	15	11	1
H147	0.0035	18	13	20	30	14	12	27	14	11	10	14	17	21	23	11	11	9	12	18.2	13	19	16	11	1
H148	0.0035	16	14	22	31	16	10	26	11	10	10	16	15	19	21	13	11	9	12	18	11	12	14	12	1
H149	0.0035	17	13	20	30	14	10	22	12	11	10	14	19	20	24	13	11	11	13	15	16	17	16	12	1
H150	0.0035	15	13	20	29	16	10	22	12	11	9	15	16	22	23	11	11	10	12	16	13	16	16	11	1
H151	0.0035	18	12	20	28	14	11	24	12	11	10	14	18	22	23	11	11	8	12	18.2	13	19	14	11	1
H152	0.0035	19	13	21	30	16	10	22	12	9	10	15	17	21	22	11	11	12	14	16	15		15	11	1
H153	0.0035	16	13	20	30	15	10	21	12	9	10	16	17	21	23	11	11	11	14	16	13	15	17	12	1
H154	0.0035	19	13	20	30	14	11	26	13	12	10	14	17	21	23	11	11	9	12	18.2	13	19	14	11	1
H155	0.0035	16	13	19	28	14	11	24	12	11	10	14	18	21	22	11	11	9	12	18.2	14	18	13	11	1
H156	0.0035	17	13	19	29	13	11	25	12	11	9	14	17	21	23	12	14	11	13	14	14	15	15	11	1
H157	0.0035	18	13	20	28	14	11	25	12	11	10	14	18	20	22	11	11	9	12	20.2	13	18	13	11	1
H158	0.0035	18	13	20	29	14	11	24	12	11	10	14	18	23	22	12	11	9	12	20.2	13	18	13	11	1
H159	0.0035	19	13	20	29	14	11	25	12	11	10	14	19	21	22	11	11	9	12	18.2	13	17	13	11	1
H160	0.0035	20	13	20	30	14	11	24	13	11	11	14	18	20	23	11	11	9	12	17.2	13	17	16	11	1
H161	0.0035	17	13	20	29	14	11	25	12	11	10	14	19	22	22	12	11	9	12	22.2	13	18	14	11	1
H162	0.0035	19	12	23	30	11	10	24	12	12	11	14	16	22	24	13	12	13	13	14	17	18	15	11	1
H163	0.0035	17	13	19	30	13	10	23	12	13	11	14	19	23	24	12	11	13	14	15	16	18	17	12	1
H164	0.0035	18	12	19	29	15	10	24	12	12	10	16	16	21	23	12	13	11	11	15	13	17	16	10	1
H165	0.0035	15	13	20	30	15	10	22	11	11	9	14	16	20	23	11	11	10	12	15	13	15	15	11	1
H166	0.0035	17	13	18	30	14	11	23	14	13	9	14	17	22	24	12	13	10	13	17	15		15	11	1
H167	0.0035	18	13	20	29	13	10	26	11	11	10	14	23	22	24	11	11	12	13	16	13	14	16	12	1

H168	0.0035	18	13	20	30	13	10	24	12	11	11	14	18	21	23	11	11	13	14	16	16		16	11	1
H169	0.0035	18	12	20	30	13	10	25	11	10	10	14	20	21	25	12	11	12	13	18	16	19	15	9	1
H170	0.0035	18	14	23	32	15	10	25	11	11	10	14	17	17	24	12	11	13	13	17	11	12	13	13	1
H171	0.0035	20	13	20	29	14	10	25	12	11	10	14	17	21	23	11	11	9	13	18.2	13	20	15	11	1
H172	0.0035	18	14	21	31	14	10	22	12	13	10	14	17	20	24	11	11	12	13	17	17	18	15	13	1
H173	0.0035	17	13	20	30	15	10	25	12	12	10	14	18	20	23	11	11	9	12	19.2	14	18	15	11	1
H174	0.0035	16	12	21	29	16	11	25	12	10	10	14	17	22	23	12	12	11	12	17	14	15	15	10	1
H175	0.0035	18	13	18	31	14	10	28	13	12	11	14	23	23	23	11	12	12	14	16	18		15	12	1
H176	0.0035	18	12	19	30	13	10	23	12	12	10	14	19	21	24	10	10	12	13	14	18		15	11	1
H177	0.0035	18	13	20	30	14	11	26	12	11	10	14	19	21	22	11	11	9	12	19.2	13	17	13	11	1
H178	0.0035	18	12	20	29	13	10	25	11	10	10	14	20	21	25	12	11	12	13	18	16	18	15	9	1
H179	0.0035	16	14	22	31	15	11	22	12	10	10	16	19	21	21	11	11	12	14	17	14		15	11	1
H180	0.0035	19	14	20	30	14	10	25	12	11	10	14	17	21	23	11	11	9	13	17.2	13	19	15	11	1
H181	0.0035	16	13	21	32	13	10	25	12	11	10	14	18	21	24	12	11	11	13	18	16	17	15	11	1
H182	0.0035	17	13	19	29	14	10	23	13	14	12	16	18	23	24	12	13	10	12	18	11	14	16	12	1
H183	0.0035	17	13	20	28	14	11	24	12	11	11	14	18	21	22	11	11	9	12	21.2	13	18	13	11	1
H184	0.0035	17	12	20	29	13	10	23	13	12	10	14	19	21	25	10	10	12	13	14	18		15	11	1
H185	0.0035	19	13	20	30	14	10	24	12	11	11	14	18	20	23	11	11	9	12	18.2	13	17	16	11	1
H186	0.0035	18	14	21	30	13	10	25	12	8	10	14	20	23	25	13	11	12	13	17	14	19	15	12	1
H187	0.0035	16	14	20	30	14	10	22	11	12	9	15	17	21	23	10	11	10	12	17	13	18	15	12	1
H188	0.0035	18	13	19	30	15	10	23	12	12	9	15	17	20	24	11	11	9	12	17	13	15	15	11	1
H189	0.0035	19	13	20	30	14	10	25	14	11	11	14	18	20	23	11	11	9	12	17.2	13	17	16	11	1
H190	0.0035	19	13	20	30	14	10	24	13	11	11	14	19	20	23	11	11	9	12	17.2	13	17	15	11	1
H191	0.0035	21	12	23	29	11	10	24	12	12	11	14	16	22	24	13	12	13	13	14	17	19	15	11	1
H192	0.0035	18	14	20	30	14	11	24	12	11	10	14	21	22	22	12	11	9	12	20.2	13	19	13	11	1
H193	0.0035	16	12	19	28	14	10	25	13	12	10	15	16	24	21	12	14	13	11	15	13	18	17	11	1
H194	0.0035	18	14	20	31	13	10	24	12	11	10	14	18	21	24	13	11	12	13	15	15	17	15	10	1
H195	0.0035	16	12	19	28	14	10	24	13	12	10	15	16	24	21	12	14	13	11	15	13	18	17	11	1
H196	0.0035	18	14	21	30	14	10	22	11	12	9	14	19	23	23	13	12	11	12	17	15	19	14	11	1
H197	0.0035	18	13	20	30	13	11	23	12	11	10	14	19	23	25	11	11	12	13	16	16	18	17	12	1
H198	0.0035	19	13	21	29	14	10	22	13	12	9	15	14	23	23	12	11	10	12	14	13	18	15	11	1
H199	0.007	15	13	20	30	15	10	22	12	11	9	15	17	22	23	11	11	10	12	15	14	16	17	12	2
H200	0.0035	17	13	21	30	14	10	25	12	12	10	14	19	20	23	11	11	9	12	18.2	13	18	14	11	1
H201	0.0035	17	14	20	30	14	11	26	12	11	10	14	19	21	22	10	11	9	12	19.2	13	18	13	11	1
H202	0.0035	19	13	20	29	14	11	25	12	11	10	14	19	21	23	11	11	9	12	18.2	13	18	13	11	1
H203	0.0035	18	13	20	29	14	11	26	12	11	10	14	18	21	22	10	11	9	12	19.2	13	18	13	11	1
H204	0.0035	19	13	20	29	14	11	25	13	11	10	14	19	21	22	12	11	9	12	20.2	13	19	13	11	1
H205	0.0035	17	13	19	30	14	10	21	13	12	12	15	16	23	24	13	12	10	12	17	11	14	15	13	1
H206	0.007	20	13	20	29	14	11	26	12	11	10	14	19	21	22	11	11	9	12	19.2	13	18	13	11	2
H207	0.0035	18	13	20	29	13	10	22	12	11	10	14	20	22	22	11	11	12	13	16	17	18	15	11	1
H208	0.0035	18	13	20	28	14	11	26	12	11	10	14	19	21	22	11	11	8	12	19.2	13	19	13	11	1
H209	0.0035	20	13	20	29	15	10	25	12	11	10	14	18	21	23	11	11	9	12	18.2	13	17	15	11	1
H210	0.0035	20	13	20	29	14	10	26	12	11	10	14	19	20	22	11	11	9	12	19.2	13	18	13	11	1

H211	0.0035	18	13	20	29	14	10	26	12	11	10	14	17	22	22	11	11	9	12	19.2	13	18	13	11	1
H212	0.0035	17	14	20	31	13	10	25	13	10	10	14	20	21	25	11	11	12	13	17	17		15	11	1
H213	0.0035	17	13	20	28	14	11	24	12	11	11	14	18	21	22	11	11	9	12	20.2	13	18	13	11	1
H214	0.0035	17	14	20	31	13	10	25	13	10	10	14	20	21	25	12	11	12	13	16	17		15	11	1
H215	0.0035	17	13	20	30	15	10	22	12	12	9	14	15	22	22	11	11	8	12	16	15		15	12	1
H216	0.007	19	13	20	28	14	10	25	13	11	10	14	16	22	22	11	11	9	12	18.2	12	18	13	11	2
H217	0.0035	18	14	20	31	13	10	25	13	10	10	14	20	21	25	11	11	12	14	17	17		15	11	1
H218	0.0035	18	13	20	31	15	10	23	14	12	10	14	22	20	24	12	11	11	13	17	17		16	14	1
H219	0.0035	18	13	20	28	14	10	26	12	11	10	14	18	21	22	11	11	9	12	20.2	13	18	13	11	1
H220	0.0035	16	12	19	28	15	10	25	12	11	9	14	18	20	25	11	11	9	12	16	15	19	14	11	1
H221	0.0035	18	13	20	29	14	10	25	12	11	10	14	16	22	23	11	11	9	13	17.2	13	19	15	11	1
H222	0.0035	16	13	19	31	14	10	24	14	12	9	14	16	21	23	12	13	11	13	18	14	16	14	11	1
H223	0.0035	17	13	20	29	14	11	25	11	11	9	14	17	22	23	10	11	10	12	15	13	16	16	11	1
H224	0.0035	17	13	19	28	13	10	23	12	11	11	14	16	23	22	12	16	12	13	18	14	16	16	10	1
H225	0.0035	17	13	19	29	14	11	22	14	12	12	15	17	23	24	12	13	10	12	17	11	14	16	12	1
H226	0.0035	19	13	20	28	14	11	26	12	11	10	14	18	20	22	11	11	9	12	19.2	13	18	13	11	1
H227	0.0035	17	13	20	28	14	11	24	12	11	11	14	19	21	22	11	11	9	12	19.2	13	18	13	11	1
H228	0.0035	19	12	19	28	15	11	23	13	12	12	15	18	23	24	12	13	8	12	16	11	15	15	13	1
H229	0.0035	16	13	21	30	17	10	24	11	11	11	14	16	20	21	13	11	14	12	16	16	18	15	11	1
H230	0.0035	17	13	20	30	15	12	26	13	11	10	14	18	22	23	11	11	9	12	18.2	13	19	15	11	1
H231	0.0035	20	12	23	30	11	10	24	12	12	11	14	16	22	24	13	12	13	13	14	17	18	15	10	1
H232	0.0035	17	12	20	29	14	10	26	12	11	10	14	17	21	24	12	11	9	12	18.2	13	17	15	11	1
H233	0.0035	18	13	20	31	15	10	23	12	12	10	14	19	21	24	12	11	11	14	16	17	18	17	13	1
H234	0.0035	18	13	20	29	14	11	24	12	11	10	14	19	21	22	11	11	10	12	21.2	13	17	13	11	1
H235	0.0035	16	13	20	31	14	10	26	12	11	10	14	17	20	23	11	11	9	12	18.2	14	17	15	11	1
H236	0.0035	17	14	19	30	14	11	23	13	12	12	15	16	23	24	12	13	9	12	15	11	14	16	12	1
H237	0.0035	20	13	20	29	14	10	25	12	11	10	14	18	21	23	11	11	9	13	18.2	13	19	15	11	1
H238	0.0035	18	13	20	29	14	11	25	12	11	10	14	19	20	23	11	11	9	12	20.2	13	18	13	11	1
H239	0.0035	16	14	20	31	15	10	26	13	9	10	14	20	23	24	12	11	12	13	14	13	14	15	12	1
H240	0.0035	17	13	20	29	14	11	26	12	11	10	14	19	21	22	11	11	8	12	19.2	13	16	13	11	1
H241	0.0035	19	13	20	29	14	10	25	12	11	11	14	17	21	23	11	11	9	13	17.2	13	18	15	11	1
H242	0.0035	17	13	19	29	14	11	22	14	12	12	15	17	24	25	11	13	10	12	17	11	14	15	12	1
H243	0.0035	17	13	19	30	14	9	25	12	11	11	14	18	21	25	11	12	12	13	15	17	18	16	11	1
H244	0.0035	17	13	20	30	16	11	23	12	13	11	14	17	24	24	11	11	10	13	15	11	14	17	12	1
H245	0.0035	17	13	20	28	14	11	26	12	11	10	14	18	21	22	11	11	9	12	19.2	13	18	13	11	1
H246	0.0035	17	14	20	30	13	9	27	12	10	10	14	22	21	25	10	11	12	13	18	13	14	14	11	1
H247	0.0035	18	13	19	29	14	11	22	12	12	12	15	17	24	25	12	13	10	13	18	10	16	17	13	1
H248	0.0035	17	13	19	29	15	10	22	13	12	12	14	20	24	24	12	13	12	13	15	13	15	15	13	1
H249	0.0035	16	12	19	28	15	10	26	12	11	9	14	18	20	25	11	11	9	12	16	16	20	14	11	1
H250	0.0035	18	12	20	29	13	10	23	12	12	10	14	22	21	24	10	10	12	13	14	18		15	11	1
H251	0.0035	21	12	23	30	11	10	24	12	12	11	14	16	22	24	13	12	13	13	14	17	18	15	11	1
H252	0.0035	15	14	21	32	14	10	22	13	11	9	15	17	23	23	12	11	11	12	15	15	17	15	12	1
H253	0.0035	15	13	20	29	15	10	22	12	11	9	15	16	22	23	11	11	10	12	16	13	16	16	11	1

H254	0.0035	17	14	18	31	14	10	21	13	12	12	15	17	23	24	12	12	10	12	17	11	14	15	14	1
H255	0.0035	16	12	19	28	15	10	25	12	11	9	14	17	20	25	11	11	9	12	16	17	19	14	11	1
H256	0.0035	16	13	21	29	15	10	22	12	11	9	14	18	22	23	12	11	11	12	14	13	16	15	12	1
H257	0.0035	17	14	21	31	13	10	27	12	8	11	15	19	22	26	13	11	12	14	16	14	18	16	12	1
H258	0.0035	19	13	20	29	14	10	26	12	11	10	14	17	21	22	11	11	9	12	19.2	13	18	13	11	1
H259	0.0035	15	13	21	30	14	10	30	11	11	11	14	18	22	21	11	11	12	13	17	16		14	11	1
H260	0.0035	19	13	21	30	16	10	22	12	9	10	15	17	21	22	11	11	12	14	16	14	15	15	11	1
H261	0.0035	16	16	19	32	14	10	22	15	12	9	14	17	21	23	10	13	10	13	18	14	16	16	11	1
H262	0.0035	18	12	20	30	13	10	25	11	10	10	14	20	21	25	12	11	12	13	17	16	18	15	9	1
H263	0.0035	17	12	19	29	13	12	25	13	11	9	14	17	21	23	13	14	11	13	15	14	15	15	11	1
H264	0.0035	17	13	21	29	15	10	25	13	12	10	14	18	20	23	11	11	9	12	19.2	14	18	14	11	1
H265	0.0035	19	13	19	30	14	10	22	13	12	12	15	17	23	24	12	12	10	12	17	11	14	15	14	1
H266	0.0035	18	13	20	29	14	11	25	12	11	10	14	18	21	22	11	11	9	12	19.2	13	18	13	11	1
H267	0.0035	17	13	20	30	14	10	25	11	11	11	14	17	21	23	11	11	9	12	17.2	13	16	15	12	1
H268	0.0035	18	12	20	30	11	10	24	13	12	12	14	18	21	24	11	12	13	13	14	18	19	15	12	1
H269	0.0035	18	14	21	32	16	10	28	12	12	10	14	22	22	23	11	12	12	14	16	14	15	13	11	1
H270	0.0035	17	13	20	32	13	11	26	12	10	10	14	20	22	25	12	11	12	13	18	15	16	15	11	1
H271	0.0035	16	14	19	30	14	10	22	13	12	9	14	17	21	24	11	11	10	13	19	14	16	16	11	1
H272	0.0035	18	13	20	30	14	11	25	11	11	10	14	16	21	25	11	11	9	12	18.2	15	19	15	11	1
H273	0.0035	18	13	21	31	15	10	25	11	12	11	14	19	22	21	12	11	13	14	17	15		16	11	1
H274	0.0035	18	15	20	31	15	10	22	12	19	9	15	15	24	24	11	11	10	12	21	14	16	16	11	1
H275	0.0035	17	13	20	29	14	10	22	12	11	8	15	16	21	23	13	11	10	12	15	14	17	16	11	1
H276	0.0035	18	14	20	30	14	10	22	12	9	9	15	15	24	24	11	11	10	12	21	14	16	16	11	1
H277	0.0035	18	13	21	31	14	11	22	11	12	10	14	20	20	24	12	11	11	13	16	17	19	16	13	1
H278	0.0035	18	13	21	31	14	11	22	11	12	10	14	20	20	24	11	11	11	13	16	17	19	16	13	1
H279	0.0035	18	14	20	30	14	10	22	12	9	9	15	15	24	25	11	11	10	12	20	14	16	16	11	1
H280	0.0035	17	14	20	30	14	10	22	12	9	9	15	15	24	24	11	11	10	12	21	14	16	16	11	1
H281	0.0035	17	13	20	29	15	10	25	13	11	10	14	17	21	23	11	11	9	13	17.2	13	19	15	11	1
H282	0.0035	16	14	19	30	15	12	22	13	12	12	14	18	23	24	12	13	11	13	19	14	17	15	13	1
H283	0.0035	17	14	20	31	13	10	24	13	10	10	14	19	21	26	11	11	12	13	16	17		15	11	1
H284	0.0035	17	12	21	29	16	10	22	13	10	10	15	17	20	22	12	11	12	15	16	15		15	13	1
H285	0.0035	17	13	20	28	14	12	26	12	11	10	14	18	21	22	11	11	9	12	19.2	13	18	13	11	1
H286	0.0035	17	13	20	30	14	10	22	14	11	10	14	18	21	23	12	11	12	13	14	16	19	16	14	1
H287	0.01	19	13	20	30	14	10	24	13	11	11	14	18	20	23	11	11	9	12	17.2	13	17	16	11	3
H288	0.007	19	13	20	30	14	10	24	12	11	11	14	18	20	23	11	11	9	12	17.2	13	17	16	11	2

*****HN is the haplotype number or type detected. HF is the haplotype frequency. n is the number of times each haplotype was observed. Shared haplotypes are color coded with light blue.

Locus	576	389 I	448	389 II	19	391	481	549	533	438	437	570	635	390	439	392	643	393	458	385a/b	456	YH4
Allele	N=303	N=303	N=303	N=303	N=303	N=303	N=303	N=303	N=303	N=303	N=303	N=303	N=303	N=303	N=303	N=303	N=303	N=303	N=303	N=303	N=303	303
8	-	-	-	-	-	0.0033	-	-	0.0165	0.0033	-	-	-	-	-	-	0.0231	-	-	-	-	-
9	-	-	-	-	-	0.0330	-	-	0.0330	0.1881	-	-	-	-	-	0.0033	0.4059	-	-	-	-	0.0198
10	-	-	-	-	-	0.6370	-	-	0.1221	0.5809	-	-	-	-	0.0429	0.0099	0.1518	-	-	0.0076	-	0.0462
11	-	-	-	-	0.0231	0.2937	-	0.0990	0.5280	0.1452	-	-	-	-	0.5644	0.7921	0.1386	0.0231	-	0.0324	-	0.6766
12	-	0.1419	-	-	-	0.0330	-	0.5842	0.2508	0.0825	-	-	-	-	0.2937	0.0693	0.2112	0.534	-	0.0190	0.0066	0.1749
13	-	0.6469	-	-	0.1881	-	-	0.2343	0.0429	-	-	-	-	-	0.0990	0.0924	0.0660	0.3333	-	0.2652	0.1881	0.0594
14	-	0.2046	-	-	0.5611	-	-	0.0759	0.0033	-	0.7558	0.0033	-	-	-	0.0231	0.0033	0.0957	0.0561	0.1068	0.1188	0.0231
15	0.0495	0.0033	-	-	0.1815	-	-	0.0066	-	-	0.1617	0.0363	-	-	-	0.0033	-	0.0132	0.1122	0.0820	0.4488	-
15.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0019	-	-
16	0.1749	0.0033	0.0033	-	0.0429	-	-	-	-	-	0.0825	0.0957	-	-	-	0.0066	-	-	0.2046	0.0973	0.1782	-
16.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0033	-	-	-
17	0.2805	-	-	-	0.0033	-	-	-	-	-	-	0.2376	0.0033	-	-	-	-	-	0.1419	0.1469	0.0528	-
17.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0825	-	-	-
18	0.3036	-	0.0231	-	-	-	-	-	-	-	-	0.2772	-	-	-	-	-	-	0.0759	0.1660	0.0066	-
18.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.1188	-	-	-
19	0.1419	-	0.1617	-	-	-	-	-	0.0033	-	-	0.1947	0.0066	-	-	-	-	-	0.0297	0.0591	-	-
19.2	-	-	0.0066	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0957	-	-	-
20	0.0330	-	0.6205	-	-	-	0.0033	-	-	-	-	0.1056	0.1881	-	-	-	-	-	0.0033	0.0152	-	-
20.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0462	-	-	-
21	0.01650	-	0.1287	-	-	-	0.0396	-	-	-	-	0.0132	0.4521	0.0462	-	-	-	-	0.0132	-	-	-
21.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0099	-	-	-
22	-	-	0.0330	-	-	-	0.2178	-	-	-	-	0.0198	0.1518	0.2277	-	-	-	-	-	-	-	-
22.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0066	-	-	-
23	-	-	0.0231	-	-	-	0.1023	-	-	-	-	0.0165	0.1452	0.3168	-	-	-	-	-	-	-	-
24	-	-	-	-	-	-	0.1551	-	-	-	-	-	0.0495	0.2673	-	-	-	-	-	-	-	-
25	-	-	-	-	-	-	0.2673	-	-	-	-	-	0.0033	0.1221	-	-	-	-	-	-	-	-
26	-	-	-	-	-	-	0.1551	-	-	-	-	-	-	0.01980	-	-	-	-	-	-	-	-
27	-	-	-	0.0066	-	-	0.0396	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
28	-	-	-	0.1155	-	-	0.0165	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
29	-	-	-	0.2739	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
30	-	-	-	0.3993	-	-	0.0033	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
31	-	-	-	0.1650	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
32	-	-	-	0.0396	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Locus	576	389 I	448	389 II	19	391	481	549	533	438	437	570	635	390	439	392	643	393	458	385a/b	456	YH4
A (n)[a]	7	5	8	6	6	5	10	5	8	5	3	10	8	6	4	8	7	5	15	12	7	6
A(v)[b]	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	1	0	0
A(C)[c]	18	13	20	30	14	10	25	12	11	10	14	18	21	23	11	11	9	12	16	13	15	11
A(R)[d]	21	15, 16	16	27	17	8	20, 30	15	14, 19	8	16	14	17, 25	26	10	9, 15	14	15	16.2, 20	15.2	12, 18	9
(GD)[e]	0.7771	0.5213	0.5720	0.7257	0.6165	0.5075	0.8218	0.5902	0.6422	0.6013	0.3971	0.8090	0.7159	0.7614	0.5855	0.3598	0.7459	0.5951	0.8860	0.8493	0.7168	0.5067
(DC)[f]		0.95049505																				
(HMP)[g]		0.00398525																				
(HD)[h]		0.999312812																				

Table.15: Y STRs forensic parameters. a Number of alleles genotyped at each locus, b Number of different variants genotyped at each locus, c most common allele for each locus, d rarest allele for each locus. Gene diversity and haplotype discrimination capacity were calculated using Excel. 385a/b loci were treated as independent single loci.

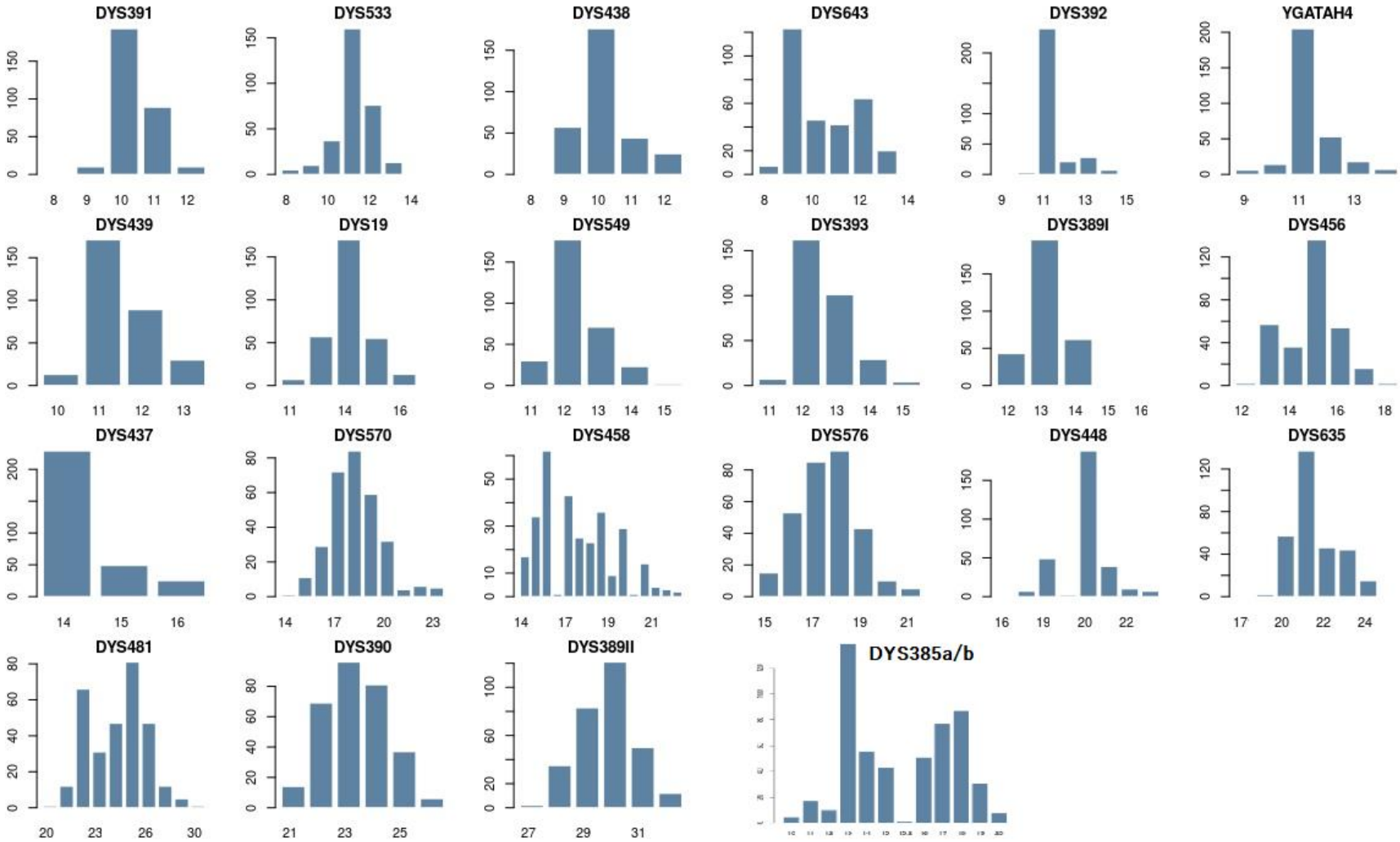


Figure.25: Allele frequency diagram for each locus. The x-axis is the allele designation (alleles genotyped in our cohort) and the y-axis is the frequency of each allele for each locus. Figure was created using STRAF software version 2.1.5.

3.4.5 Y-STRs allele dropout (null alleles) and deletions

10 samples exhibited Y-STR allele dropout and possible deletions of some of the Y-STRs loci genotyped, even though these DNA profiles had relatively balanced amplification and peak pattern. Locus DYS549 exhibited the highest frequency of this phenomenon with 8 samples, one sample presented with allele dropout at locus DYS576, and one sample presented with the same phenomenon at locus DYS448 (Fig. 26).

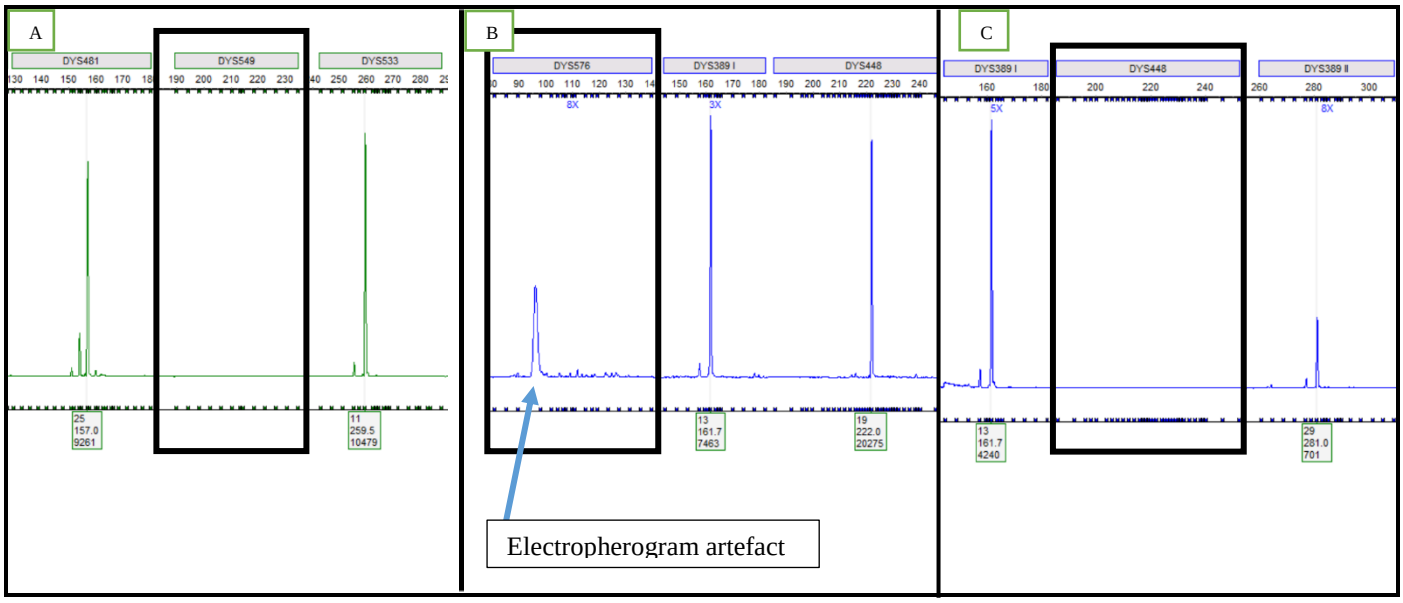


Figure.26: Three loci (DYS549, DYS576, and DYS448) presented with absence of alleles and amplification.

In order to confirm whether these samples as true anomalous cases, we amplified the suspected null loci presented in these samples with a second set of primers. This allowed us to confirm if these allelic dropout are due to primer binding site mutation or a chromosomal deletion occurring at that locus. When we amplified these samples with our primers, five samples with allele dropout at locus DYS549 were amplified successfully, indicating, may be, the presence of a point mutation in the binding site of the original primer set (fig.27 A). The rest of the samples are probably

microdeletion occurring at the above-mentioned loci (Fig.27 A&B&C). To further prove whether the drop out alleles resulted from primer binding site variations, we sequenced the PCR products obtained from drop out locus amplification with the new primer sets. Of note, the new primer binding sites flank the binding sites of the primers provided in the kit and gave us a larger PCR amplicon size. the sequences obtained from these samples were aligned, using nucleotide BLAST against a reference sequence (GenBank Accession number AC010133.4) and against the positive control (2800 positive control) sequence that is included in PowerPlex©Y23 kit. Our sequence comparison revealed a single nucleotide polymorphism (SNP) (A>T), at position 107502 which seems to fall in the kit primer binding site (Fig. 28). Moreover, this mutation was detected in 4/5 samples analyzed. Therefore, the molecular basis of this null allele can probably be ascertained due to imperfect PowerPlex©Y23 binding site of some primers.

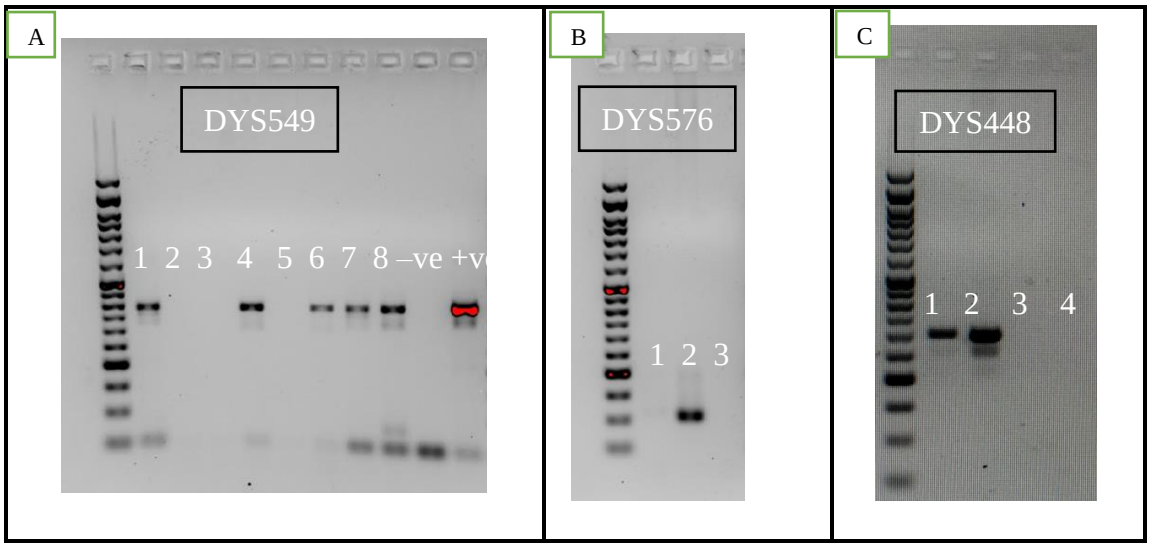


Figure.27: Amplification of DYS549, DYS576, and DYS448 in null allele samples using a new primer sets. A) From right to left, well 1, 4, 6, 7, and 8 are samples with null alleles at locus DYS549, well 2, 3, and 5 are samples that failed to amplify, well 9 and 10 are negative and positive control respectively. B) From right to left, well 1 is a sample presented with a null allele pattern at locus DYS576 that failed to amplify with the new set of primers. well 2 and 3 are positive and negative control respectively. C) From right to left, well 3 is a sample presented with null allele pattern at locus DYS448, well 2 and 4 are positive and negative control respectively (well one is a randomly chosen sample).

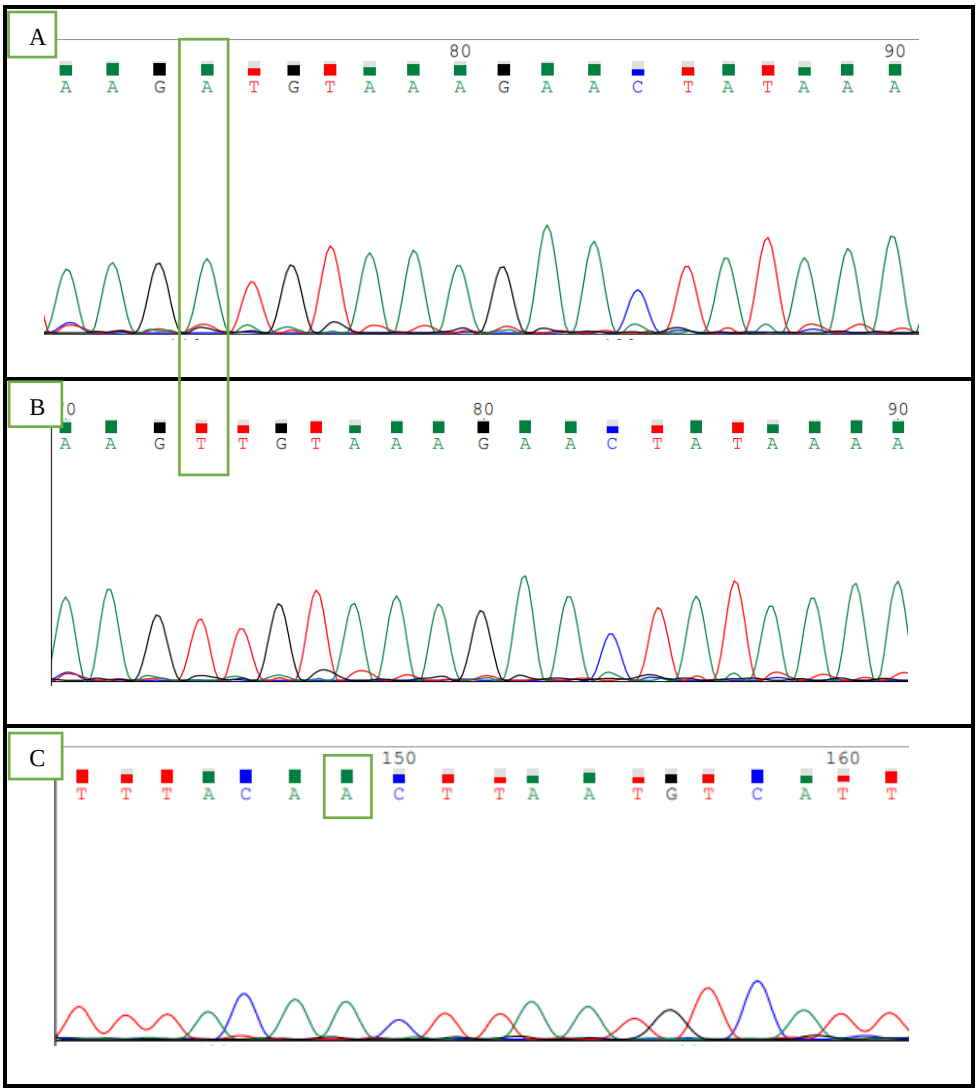


Figure.28: Sanger sequencing result of representative null allele at locus DYS549 showing the transversion mutation (A>T) at position 107502. A) 2800 positive control sample. B) Forward sequence & C) reverse sequences of one of the samples presented with this variation.

3.4.6 Tri-allelic pattern observed at the locus DYS385a/b

Locus DYS385 a/b is a multi-copy locus that presents in most of the populations with two copies. However, we were able to detect one sample with a three (17, 18, and 19) genotype as seen in the Electropherogram (Fig.29A). We tried to confirm this rare condition with 12% acrylamide gel (Fig.29B). However, we obtained an unexpected result of bands pattern. Even though when we run the same sample on Agarose gel, we obtained the expected result of bands pattern (Fig.29C).

We cannot explain the result we obtained with the Acrylamide gel at the moment. Nevertheless, we can still see a difference in band numbers pattern between the positive control (expected two bands with the genotype 13, 16 (band sizes 266 and 278 bp respectively)) and the sample (expected three bands with the genotype 17, 18, 19 (band sizes 282, 286, 292bp).

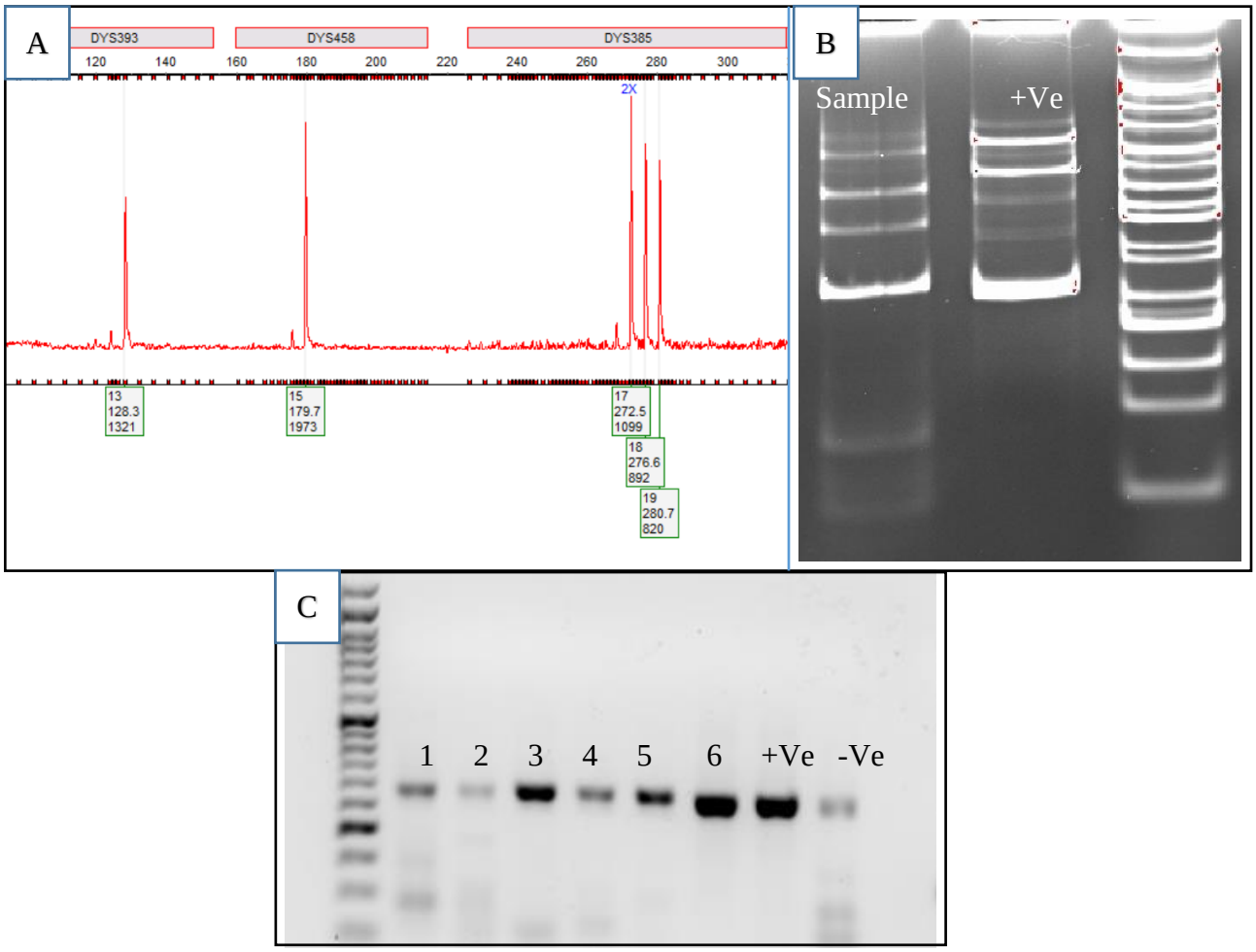


Figure.29: A sample presented with a triplication pattern at the DYS385 a/b locus. A) Electropherogram of the sample genotype at locus DYS385a/b. B) 12% acrylamide gel showing the ladder, 2800 positive control, and the triplicated sample. C) Wells 1-6 are replicate of the same triplicated sample, well 7 and 8 are positive and negative control, respectively.

3.4.7 The power of 23 Y-STRs loci system

Initially, Y-STRs commercial kits included only 9 core Y-STRs loci. In which all of them are single copy loci (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393) expect for they highly polymorphic multi-copy locus DYS385 a/b. Later, the number of loci was increased to 17 (by adding DYS438, DYS439, DYS437, DYS448, DYS458, DYS456, DYS635, and Y-GATA-H4 to the minimal haplotype system) to increase the discrimination capacity. Nowadays, the most common systems used include 23 or even 27 loci. To evaluate the added discrimination power of the system we used over other systems, we calculated the number of detected haplotypes, HMP, DC, and HD to see how increasing the number of loci genotyped in the Palestinian population will increase the efficiency of these systems forensically while fixing the genotypes and number of individuals. Our analysis revealed that the number of different haplotypes detected increased from 207 haplotypes for the minimal system, 268 for the 17-STR system, to 288 for the 23-STRs system. The discrimination capacity increased from 68.32% to 86.79% to 95.05%, HMP decreased from 0.95% to 0.54%, to 0.40%, and HD increased from 99.37% to 99.79%, to 99.93% for the minimal haplotype, 17-STRs system to 23-STRs system, respectively (Fig.30). Thus, implementing the 23 system in the Palestinian population is more efficient and such enhancement in forensic parameters has great power in real casework.

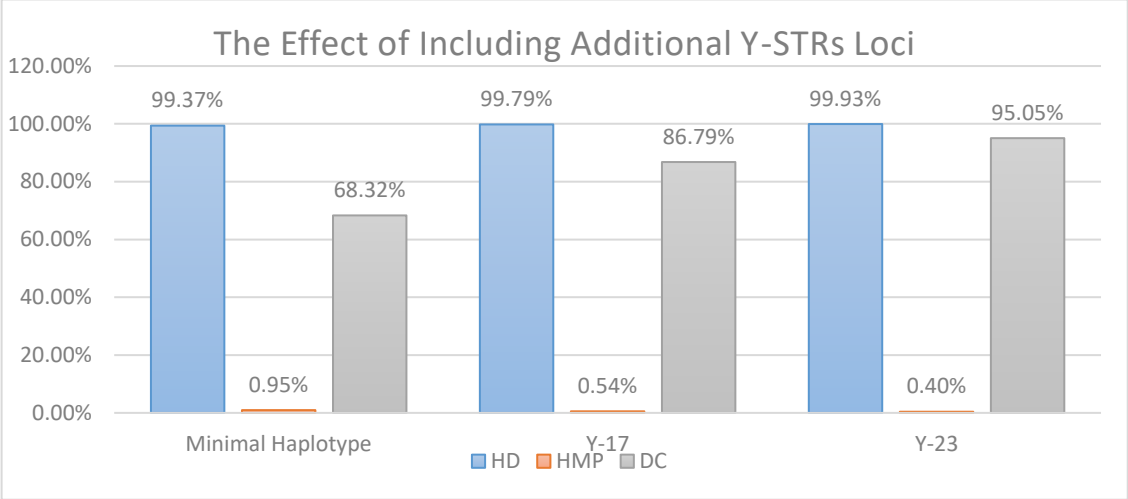


Figure.30: Effect of increasing tested loci on the discrimination power of used systems. The HD, DC, and HMP efficiency of three different systems containing different number of loci on the Palestinian population. HD: Haplotype diversity. HMP: haplotype match probability. DC: discrimination capacity.

3.4.8 Christians and Samaritans haplotype uniqueness

To test the uniqueness of some Palestinian populations, we evaluated the haplotype uniqueness of the Samaritan and Christian populations. Interestingly, among the six Samaritan males genotyped there was no shared haplotype detected even though they are a very small cohort and they are known for their high inbreeding and consanguinity. In contrast, in the 34 Christians, four haplotypes were shared between these samples, H100 was shared between 2 subjects, H99 was also shared between 2 subject, and H287 was shared among 3 subjects, H288 was shared between another 2 Christians. Interestingly, one haplotype was shared between a Muslim and a Christian, which can infer on the gene flow between these two subpopulations. We conducted a PCA test to see how the haplotypes present in these three subpopulations are close to each other dimensionally and we found that these groups are dispersed along the PC1 and PC2 axis with overlapping regions. The variance on the PC1 axis was 6.84% and on the PC2 axis was 4.61%, which can infer that the variation among individuals in each subpopulation is higher than the variation among these subpopulations (Fig.31).

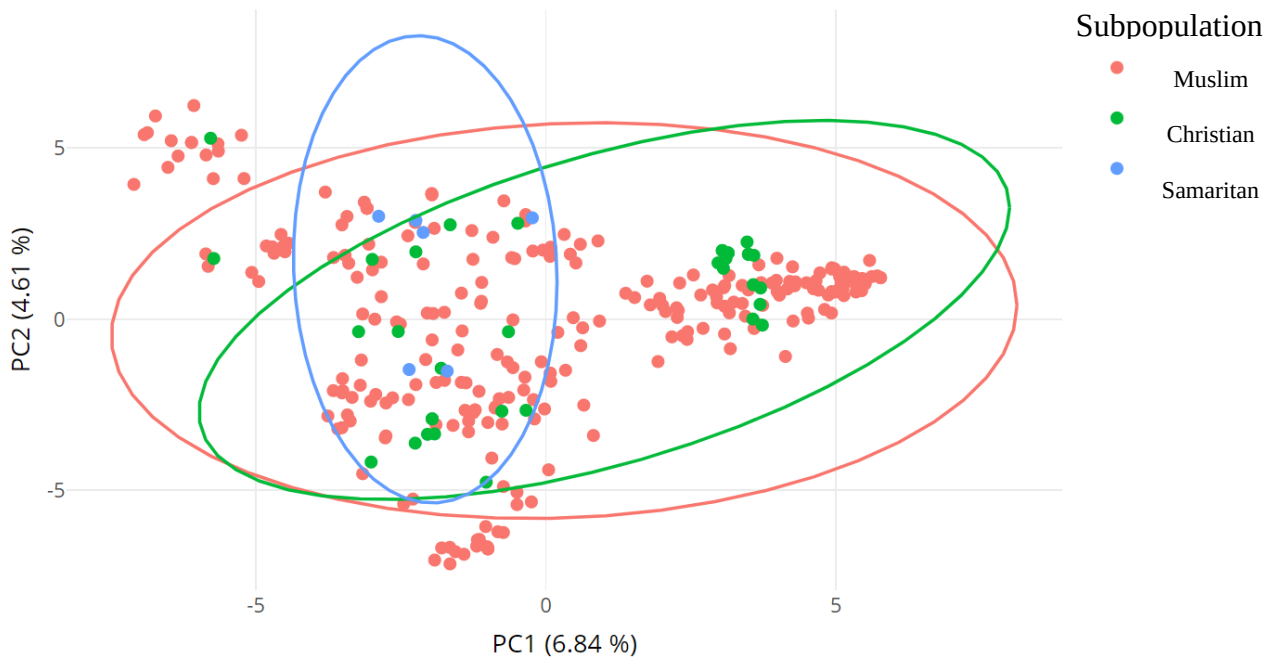


Figure.31: Poly component analysis (PCA) of haplotypes detected in each subpopulation.

3.5 Population analysis

We conducted a meta-analysis comparison for the Y-STRs haplotypes detected in the Palestinian population with other 29 neighbouring and distant countries using AMOVA, based on Slatkin's R_{st} values to accommodate for allele repeat number variation. Subsequently, we performed multidimensional scaling (MDS) analysis using STRAF software version 2.1.5. Finally, we construed a neighbour-joining (N-J) phylogenetic tree based on Nei's discriminant analysis (D_A) genetic distance metric using POPTREE2 software, and bootstraps values were calculated based on 1000 replications. The following populations with equivalent population dataset were chosen for this analysis (n refers to the sample size, YA refers to accession number of population data on

the YHRD (https://yhrd.org/) database for Y chromosome haplotypes) : Belgium (n=206, YA003651), (Spain= 534,YA003046, YA003887, YA003047, YA003890), Hungary (n= 384, YA002979, YA002978, YA003187, YA003853, YA003188), Switzerland (n=862, YA003886, YA003040, YA003269), Germany (n=810, YA002904, YA002904, YA003038, YA00288), Poland (n=803, YA003253, YA003034, YA003048, YA003463, YA003801, YA002967), Sweden (n=302, YA003071, YA003072, YA003073, YA003460, YA003075, YA003074, YA003076), Bosnia and Herzegovina [Bosnian] (n= 100, YA003787), Italy (n=704, YA003295, YA003726, YA003724, YA003069, YA003095, YA003327), England (n=195, YA003880, YA003882), Portugal (n=242, YA003016, YA003015, YA003017), Greece (n=214, YA003862, YA003127), Argentina (n= 364, YA002987, YA003002, YA002997, YA003855, YA003854), Costa Rica (n= 166, YA003353), Jamaica (n=66, YA003885), USA (n=550, YA003683, YA003865, YA003310, YA003309, YA003311), China (n=577, YA003470, YA003861, YA003307, YA003308, YA003866, YA003876, YA003877), Japan (n=323, YA003386, YA003789, YA003394, YA003868), South Korea (n=300, YA003406), Singapore (n= 104, YA003270), Philippine (n=798, YA003892, YA003206), Vietnam (n=45, YA003869), Iraq (n=124, YA003858), Lebanon (n=505, YA003859), Palestine (n=303, current study), Bangladesh (n= 137, (Hasan et al., 2016)), Nigeria (n=81, YA003874), Kenya (n=144, YA003871, YA003870), Zimbabwe (n=55, YA003884), South Africa (n=114, YA003258).

Upon the comparison of the haplotype data of this study with the other above-mentioned populations, the Palestinian population was found to be distinct from other populations. The genetic distance from closest to the furthest t for our dataset were as follows: Jamaica (0.00021), Italy (R_{ST}=0.00022), Germany (R_{ST}=0.00023), Poland (R_{ST}=0.00024), China (R_{ST}=0.00024), Switzerland, Greece (R_{ST}=0.00025), Spain, Japan and Philippines (R_{ST}=0.00026), Lebanon

(R_{ST}=0.00027), Iraq (R_{ST}=0.00028), England (R_{ST}=0.00029), Belgium, Sweden, USA, South Korea, and (R_{ST}=0.0003), Bosnia and Herzegovina (R_{ST}=0.00031), Singapore (R_{ST}=0.00039), Portugal and Argentina (R_{ST}=0.0004), Hungary (R_{ST}=0.00066), Costa Rico (R_{ST}=0.00072), and Vietnam (0.003) (Table.16). Even though Jamaica showed a low R_{st} value with an insignificant p-value with Palestine, however, it showed the same pattern with all of the other populations with insignificant p-value, this probably due to the dataset that was taken from YHRD for this population was not enough to represent the actual population efficiently. Multidimensional scaling (MDS) analysis based on Nei's genetic distance matric revealed a relatively clustered pattern between the individuals analyzed and their geographic locations, the proximity of populations that are close in geographical locations was assuring. Palestine is clustered close to Lebanon, Iraq, and Bangladesh. The European continent countries (Spain, Belgium, Hungary, Switzerland, Germany, Poland, Sweden, Italy, England, Portugal, and Greece, Bosnia and Herzegovina) clustered together. The Latin American countries, Argentina and Costa Rico also clustered with European countries, when checked the samples, they had mixed ancestry with the European countries. The USA also clustered close to the European countries; this is because the USA population analyzed had many mixed samples with ancestry history from nations worldwide, in which most of them are European ancestry. The African countries (Nigeria, Zimbabwe, South Africa, and Kenya) clustered together. Even though Jamaica is a Latin American country, however, it is clustered close to Kenya, which is probably ascertained because of immigration back in history from Africa to Jamaica. Lastly, Japan, South Korea, China, Singapore, Vietnam, and the Philippines clustered together (Fig.32). We constructed a phylogenetic tree to have a deeper insight into the relationships among populations analyzed, our result revealed that the Palestinian population is closer to Lebanon and Iraq, and these three countries are more closely related to the European countries

compared to African or Asian countries. In addition, it shows that the African countries clustered together and the Asian countries clustered together (Fig.33). AMOVA analysis was conducted using Arlequin 3.5 to measure the genetic variance between the analyzed populations. It revealed the variation between these populations is caused mainly by variation within the population at the individual level (92.7%) compared to variation among populations (7.3%) with an overall fixation index of 0.07256 and p-value of 0.0000.

Table.16: Pairwise genetic distance matrix based on the R_{ST} and p-values between the Palestinian population and 29 worldwide populations*.

	Pop	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	Belgium	-	0	0	0	0	0	0	**0.09009**	0	0	0	**0.00901**	0	0	**0.31532**	0	0	0	0	**0.01802**	0	0	**0.04505**	0	0	**0.02703**	0	0	0	0
2	Spain	0.00014	-	0	0	0	0	0	**0.04505**	0	0	0	**0.01802**	0	0	**0.31532**	0	0	0	0	**0.00901**	0	0	**0.04505**	0	0	**0.00901**	0	0	0	0
3	Hungary	0.00054	0.0005	-	0	0	0	0	0.01802	0	0	0	0	0	0	**0.10811**	0	0	0	0	**0.01802**	0	0	0	0	0	0	**0.00901**	0	**0.01802**	0
4	Switzerland	0.00013	0.00009	0.00048	-	0	0	0	**0.04505**	0	0	0	**0.00901**	0	0	**0.13514**	0	0	0	0	0	0	0	**0.03604**	0	0	0	0	0	0	0
5	Germany	0.00011	0.00007	0.00047	0.00006	-	0	0	0.00901	0	0	0	**0.00901**	0	0	**0.21622**	0	0	0	0	0	0	0	**0.04505**	0	0	0	0	0	0	0
6	Poland	0.00013	0.00008	0.00047	0.00007	0.00005	-	0	**0.04505**	0	**0.00901**	0	**0.01802**	0	0	**0.04505**	0	0	0	0	**0.00901**	0	0	**0.03604**	0	0	0	0	0	0	0
7	Sweden	0.00018	0.00014	0.00054	0.00013	0.0001	0.00012	-	**0.06306**	0	0	0	**0.00901**	0	0	**0.17117**	0	0	0	0	0	0	0	**0.08108**	0	0	**0.00901**	0	0	0	0
8	Bosina	0.0002	0.00015	0.00055	0.00014	0.00012	0.00013	0.00019	-	**0.01802**	**0.08108**	**0.009**	**0.18018**	**0.027**	**0.00901**	**0.56757**	**0.01802**	**0.04505**	**0.03604**	**0.036**	**0.10811**	**0.01802**	0	**0.16216**	**0.02703**	**0.05405**	**0.0991**	**0.00901**	0	**0.00901**	0
9	Italy	0.00011	0.00006	0.00046	0.00005	0.00003	0.00005	0.0001	0.00011	-	0	0	**0.00901**	0	0	**0.40541**	0	0	0	0	0	0	0	**0.04505**	0	0	0	0	0	0	0
10	England	0.00017	0.00013	0.00053	0.00012	0.0001	0.00011	0.00017	0.00018	0.00009	-	0	**0.00901**	0	0	**0.36036**	0	0	**0.00901**	**0.009**	**0.00901**	0	0	**0.03604**	0	0	**0.03604**	**0.00901**	0	0	0
11	Portugal	0.00032	0.00028	0.00068	0.00026	0.00025	0.00026	0.00032	0.00033	0.00024	0.00031	-	0	0	0	**0.13514**	0	0	0	0	0	0	0	**0.01802**	0	0	**0.00901**	0	0	0	0
12	Greece	0.00014	0.00009	0.00048	0.00009	0.00006	0.00008	0.00013	0.0001	0.00006	0.00012	0.0003	-	0	0	**0.37838**	0	**0.00901**	**0.00901**	0	**0.01802**	**0.00901**	0	**0.08108**	**0.01802**	0	**0.02703**	0	0	0	0
13	Argentina	0.00026	0.00021	0.00061	0.0002	0.00018	0.0002	0.00024	0.00026	0.00018	0.00024	0.0004	0.00021	-	0	**0.15315**	0	0	0	0	**0.01802**	0	0	**0.00901**	0	0	0	0	0	0	0
14	Costa Rico	0.00061	0.00057	0.00096	0.00056	0.00053	0.00055	0.0006	0.00062	0.00053	0.0006	0.0008	0.00056	0.0007	-	**0.10811**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	Jamaica	0.0001	0.00002	0.00045	0.00004	0.00002	0.00003	0.00009	0.0001	0.00001	0.00008	0.0002	0.00004	0.0002	0.00052	-	**0.1982**	**0.27027**	**0.33333**	**0.2793**	**0.27027**	**0.15315**	0	**0.55856**	**0.20721**	**0.09009**	**0.41441**	**0.10811**	0	**0.00901**	0
16	USA	0.00019	0.00015	0.00054	0.00014	0.00012	0.00013	0.00018	0.0002	0.00011	0.00018	0.0003	0.00014	0.0003	0.00061	0.0001	-	0	0	0	**0.02703**	0	0	**0.03604**	0	0	**0.00901**	0	0	0	0
17	China	0.00013	0.00008	0.00048	0.00007	0.00005	0.00006	0.00012	0.00013	0.00005	0.00011	0.0003	0.00007	0.0002	0.00055	0.00003	0.00012	-	**0.00901**	0	**0.01802**	0	0	**0.03604**	0	0	**0.01802**	0	0	0	
18	Japan	0.00014	0.0001	0.0005	0.00009	0.00007	0.00008	0.00014	0.00015	0.00006	0.00013	0.0003	0.00009	0.0002	0.00056	0.00005	0.00015	0.00008	-	**0.018**	**0.02703**	0	0	**0.05405**	0	0	**0.00901**	0	0	**0.00901**	0
19	South Korea	0.00014	0.00009	0.00049	0.00009	0.00006	0.00008	0.00013	0.00015	0.00006	0.00012	0.0003	0.00009	0.0002	0.00056	0.00005	0.00014	0.00008	0.00008	-	0	0	0	**0.04505**	0	0	**0.01802**	0	0	0	0
20	Singapore	0.00028	0.00024	0.00064	0.00023	0.0002	0.00022	0.00027	0.00029	0.0002	0.00027	0.0004	0.00023	0.0004	0.00071	0.00019	0.00025	0.00018	0.00023	0.0002	-	**0.01802**	0	**0.04505**	**0.00901**	0	**0.04505**	**0.01802**	0	0	0
21	Philippines	0.00015	0.00011	0.0005	0.0001	0.00007	0.00009	0.00014	0.00016	0.00007	0.00014	0.0003	0.0001	0.0002	0.00057	0.00006	0.00015	0.00008	0.0001	0.0001	0.00024	-	0	**0.00901**	0	0	**0.01802**	0	0	0	0
22	Vietnam	0.0032	0.00315	0.00356	0.00314	0.00311	0.00313	0.00319	0.00323	0.00311	0.00319	0.0033	0.00315	0.0033	0.00364	0.00315	0.0032	0.00313	0.00315	0.0032	0.00331	0.00315	-	0	0	0	0	0	0	0	0
23	Iraq	0.00016	0.00012	0.00052	0.00011	0.00009	0.0001	0.00013	0.00017	0.00008	0.00015	0.0003	0.00011	0.0002	0.00059	0.00007	0.00016	0.0001	0.00012	0.0001	0.00025	0.00012	0.003	-	**0.02703**	0	**0.15315**	0	0	0	0
24	Lebanon	0.00016	0.00011	0.00051	0.0001	0.00008	0.0001	0.00015	0.00016	0.00008	0.00014	0.0003	0.00011	0.0002	0.00058	0.00006	0.00016	0.00009	0.00011	0.0001	0.00025	0.00012	0.003	0.00013	-	0	**0.00901**	0	0	0	0
25	Palestine	0.0003	0.00026	0.00066	0.00025	0.00023	0.00024	0.0003	0.00031	0.00022	0.00029	0.0004	0.00025	0.0004	0.00072	0.00021	0.0003	0.00024	0.00026	0.0003	0.00039	0.00026	0.003	0.00028	0.00027	-	**0.00901**	0	0	0	0
26	Bangladesh	0.00021	0.00016	0.00056	0.00015	0.00013	0.00014	0.0002	0.00021	0.00012	0.00019	0.0003	0.00015	0.0003	0.00063	0.00011	0.00021	0.00014	0.00016	0.0002	0.0003	0.00017	0.003	0.00018	0.00017	0.00032	-	0	0	0	0
27	Nigeria	0.00063	0.00058	0.00099	0.00058	0.00055	0.00057	0.00062	0.00064	0.00055	0.00062	0.0008	0.00058	0.0007	0.00106	0.00054	0.00063	0.00057	0.00058	0.0006	0.00073	0.00059	0.004	0.00061	0.0006	0.00074	0.00065	-	0	**0**	**0**
28	Kenya	0.00482	0.00476	0.00516	0.00474	0.00472	0.00473	0.00481	0.00486	0.00472	0.00481	0.005	0.00477	0.0049	0.00525	0.00479	0.0048	0.00474	0.00476	0.0048	0.00495	0.00476	0.008	0.00482	0.00477	0.00493	0.00486	0.00532	-	**0.018**	**0.01802**
29	Zimbabwe	0.00113	0.00108	0.00149	0.00107	0.00105	0.00106	0.00112	0.00114	0.00104	0.00111	0.0013	0.00108	0.0012	0.00156	0.00104	0.00113	0.00106	0.00108	0.0011	0.00123	0.00109	0.004	0.00111	0.00109	0.00124	0.00115	0.00159	0.00584	-	0
30	South Africa	0.00419	0.00413	0.00453	0.00411	0.00409	0.0041	0.00417	0.00422	0.00409	0.00418	0.0043	0.00414	0.0042	0.00462	0.00415	0.00417	0.00411	0.00413	0.0041	0.00431	0.00413	0.007	0.00418	0.00414	0.00429	0.00422	0.00467	0.00886	0.0052	-

*P-values are above and the R_{ST} values are below the diagonal; the significance level is $p<0.05$ and $p<0.0004$ (0.05/110; 110 is the number of pairwise comparisons made) before and after Bonferroni correction, respectively.

Figure.32: A two-dimensional scaling (MDS) plot showing the genetic relationship between the Palestinian population and 29 representative populations worldwide. Numbers are related to countries when coded through analysis.

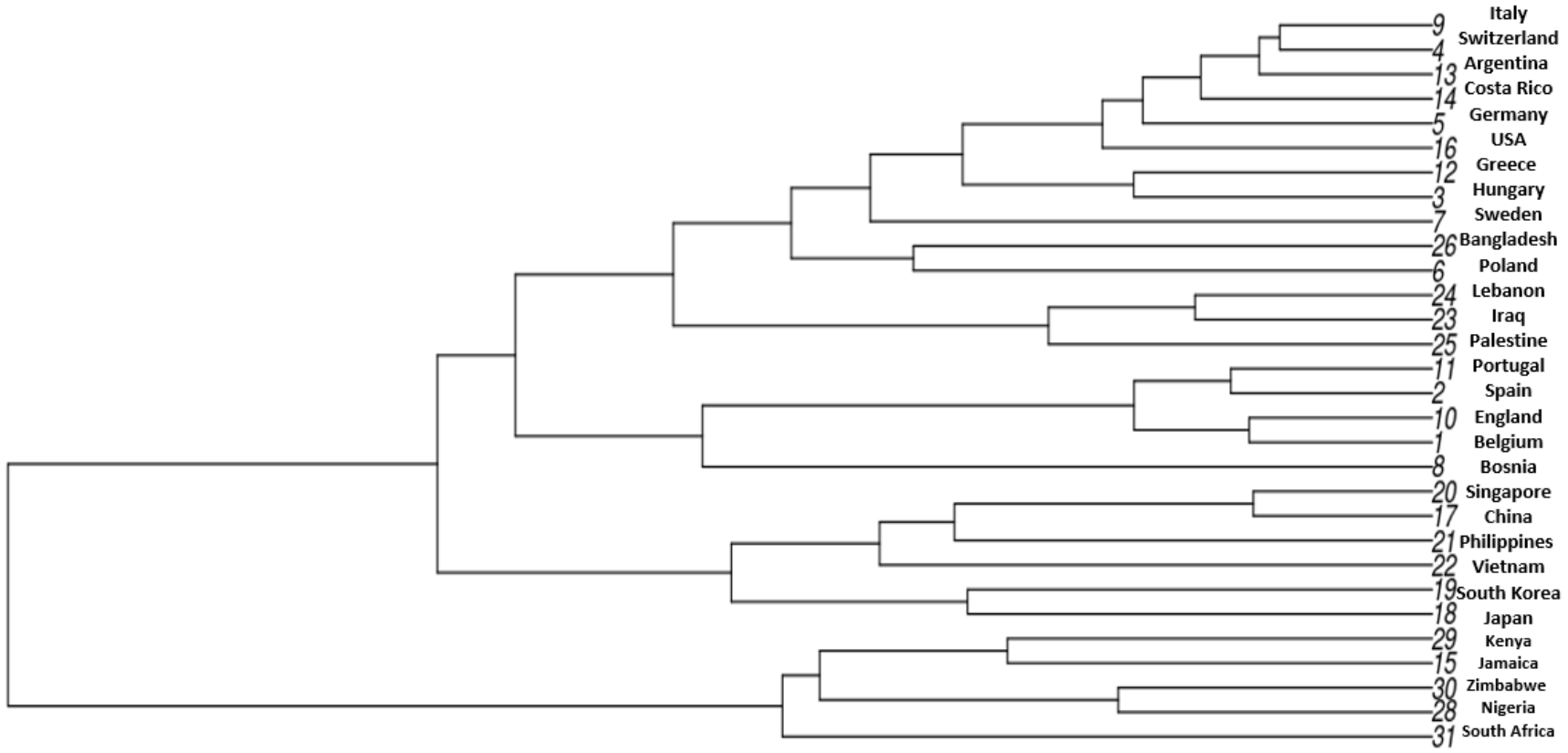


Figure.33: A neighbouring-joining (N-J) phylogenetic tree constructed based on Nei's discriminant analysis (D_A) genetic distance metric showing the clustering and relationships of analyzed populations. Numbers are related to countries when coded through analysis.

3.6 Insilico haplogroup prediction

Insilico haplogroup prediction based on Y-STR data was performed using White Athey Algorithm to infer the haplogroup and possibly the ancestral background of the population tested. Samples that exceeded that minimum fitness score of 25% and Bayesian probabilities (50%) were only considered as a possibly reliable estimate of the haplogroup. The most common haplogroup in the Palestinian males tested is Haplogroup Q (47.7%), followed by L haplogroup (25.13%), followed by I1 haplogroup (10.3%), followed by R1b haplogroup (5.1%), followed by E1b1b and J1 haplogroup (2.56%), followed by T haplogroup (3.1%), followed by G2A haplogroup (2.05%), and followed by R1a haplogroup (0.5%). The fact that haplogroup Q is the most common haplogroup detected was expected because it is widely spread in the Middle East and this haplogroup harbours mainly the Middle East Y-chromosome (Table.17).

Table .17: Haplogroup prediction based on White Athey Algorithm.

Haplogroup	Number of individuals	Frequency
E1b1b	5	2.5%
G2A	4	2.05%
G2C	0	0
H	0	0
I1	20	10.3%
(xI2a1)	0	0
I2a1	0	0
I2b (xI2b1)	0	0
I2b1	0	0
J1	5	2.56%
J2a1b	0	0
J2a1h	0	0
J2a1 x J2a1-bh	0	0
J2b	0	0
L	49	25.13%
N	0	0
Q	93	47.7%
R1a	1	0.5%
R1b	10	5.1%
T	6	3.1%

Chapter 4

Discussion

The Middle East is one of the most important locations to consider conducting population genetics research given its geographical location at the crossroad of three continents. In addition, the migration of modern humans out of Africa, which is hypothesized to occur through the Middle East, was followed by dispersion throughout the world, and subsequent migration back to Africa from this region. All of these factors drive curiosity about the genetic pool of this region. Thus, this thesis was designated with the objective to increase the knowledge about the Palestinian population genetic structure and diversity, which lacks significant and deep analysis compared to other populations worldwide. A deep analysis based on short tandem repeats (STRs) of autosomal and Y-chromosomes genetic pool was conducted in this project to infer the population genetic structure of the Palestinian population. Along with comparison to other neighbouring Middle Eastern, Arab, and Muslim countries to see if they share any common genetic pool or if they differentiated over history, time, and place. Moreover, a worldwide survey was conducted with reference populations with available genetic data on Y-STRs to analyse the genetic distance and population relationship between these countries. Y-STR genotyping provided us with the privilege to perform an Insilico haplogroup prediction algorithm to predict the common haplogroup present in Palestine. In addition, this thesis greatest aim was to establish the necessary allele frequency databases for autosomal and Y-STRs to put the basic ground of forensic DNA analysis in Palestine, which was lacking in the Palestinian justice system for all of these years. These databases will allow us to solve crimes that have been waiting for years with high reliability, and to be more specific, sexual assault crimes have been one of the biggest weaknesses in the Palestinian justice system, because solving these crimes depends largely on Y-STR genotyping.

Albeit next-generation sequencing (NGS) techniques are overtaking the world with the ability to sequence whole exomes and whole genomes, most of the population genetic studies are based on specific markers such as autosomal, X-STRs, Y-STRs, and mitochondrial STRs. Moreover, in recent decades, population geneticists started to focus on SNPs to infer population relationships, distance, ancestry, and past genetic events such as migration, bottleneck, genetic drift, etc. In addition, the fact that STRs hold a great discrimination power between individuals made them powerful tools in forensic science. These markers are the most extensively studied, established, and used markers for forensic application. Moreover, these STRs are produced widely and available worldwide with relatively cost-effective prices and forensic accreditation, which allows users to have reliable results in many different lab settings. Thus, these tools will remain one of the most important tools in both forensic application and population genetics. Thus, we characterized the allele frequencies and forensic parameters for the Palestinian population for core autosomal and Y-STR loci, using PowerPlex© Fusion 6C and PowerPlex© Y23 systems, respectively. Which are accredited and used in forensic DNA analysis worldwide.

For the autosomal STRs, we genotyped 23 different autosomal STRs that are dispersed on the different chromosomes. Despite the rarity of tri-allelic pattern events at STR loci (different loci have different tri-allelic frequencies, which can reach 1 in 10,000 samples and higher according to STRBase), We were able to detect two tri-allelic pattern samples in apparently healthy individuals that happen to occur at the same locus (D21S11) which is located on chromosome 21. These samples were initially confirmed by re-amplification and re-injection; however, they need to be confirmed by subsequent sequencing. This chromosome is well known for its anomalies, which can cause trisomy 21 or Down syndrome (Yoon et al., 2002). However, to our knowledge, the genotype combination of these samples we reported here was not reported previously, and thus, it

is novel. A total of 336 alleles were detected along the 23 loci in the total population with allele 8 of TPOX having the highest allele frequency (53.97%) which is the lowest informative locus as it presented with low PD, PE, and PIC of 82.11 %, 26.78%, and 59.2%, respectively. TPOX showed the lowest number of alleles (8 alleles) and the least heterozygosity (58.02%). Our results are in agreement with results reported around the world. For example, in a study conducted in a Southern Chinese Han population that analyzed 23 autosomal STRs, they reported TPOX to have the lowest PD (77.91%), PIC (53.80%), PE (34.30%) for trio paternity cases and 19.20% for duo paternity cases (H. xia Li et al., 2018). However, in their population, TOPX was found to be non-informative marker (H. xia Li et al., 2018), unlike our population, which was still considered informative. This also agrees with a study performed on the Qatari population using 15 autosomal STRs (Pérez-Miranda et al., 2006). In this study, they reported TPOX to have the highest allelic frequency with allele frequency of 48.11% with PIC and PD of 63.86% and 85.53% respectively. The highest informative marker in our population was SE33 with PIC of 94.5%, PD of 99.3%, and PE of 84.76%. SE33 locus also showed the highest allele number with 55 alleles and 92.54% heterozygosity. Our results are also in concordance with other surrounding populations such as a Circassian subpopulation in Jordan(AL-Eitan et al., 2020), where they found that locus SE33 is the most informative locus with a PIC of 93%. In our study, the most common genotype frequency analysis showed a very low probability to have a random genotype match in the Palestinian population with a frequency of 2.98139E-23. In addition, paternity analysis revealed that the PowerPlex© Fusion 6C is quite effective in paternity testing in the Palestinian population with a POP of 99.9999999%. This high value indicates that the system can identify the actual biological father with high confidence. In order to implement this highly informative and discriminatory system in the Palestinian population we had to confirm that the Palestinian population complies with the rules and guidelines of implementing a database into a population, which states that a

population needs to be in HWE and linkage equilibrium to be able to use the product rule in forensic testing safely. Our results proved that the whole population is in HWE and LD (p-value>0.05& p>.00217; before and after applying Bonferroni's correction, respectively). However, when we looked deeper at the subpopulations level (Muslims, Christians, and Samaritans) that compose the Palestinian population, we found that Samaritans exhibit a different pattern of inbreeding which affected allele frequencies and caused deviations from Hardy-Weinberg expected allele frequencies (p-value= 0.0006 at D1S1656). This inbreeding can be seen with low positive values obtained from F_{IS} and F_{IT} (0.0304 and 0.0479, respectively) which indicates a slight loss of heterozygosity. Moreover, Samaritans showed more genetic differentiation compared to Muslims and Christians. This genetic differentiation was visually demonstrated in the PCA plot in figure 21 where Samaritans clustered together along the PC2 axis, while Muslims and Christians showed more homogenous distribution between the genotypes of the two subpopulations. Genetic differentiation between the Palestinian subpopulation was seen not only at the allelic frequency level but it was seen at the individual allele genotype level, where some alleles were specific to specific populations (e.g. Penta D*10.1 for Samaritans and Penta E*13.3 for Christians). Such alleles are a great example of population divergence, genetic drift, and admixture events, which lead to the introduction or the fixation of these alleles in a subpopulation and the loss or absence of the same allele in another subpopulation. The fixation and loss of alleles can have forensic implications, for example, when we encounter an unknown sample from a crime scene with Penta D*10.1 allele, we can reliably assume that this profile belongs to a Samaritan. Moreover, this pattern of deviated genetic behavior is not strange for small isolated subpopulations where consanguinity records are high, because genetic drift in such cases can cause enormous effects where the effective population size can lead to genetic fixation or loss as can be seen in the simulation conducted in figure 1. In addition, such inbreeding in the Samaritan

population is the reason for heterozygosity loss detected in our results. In this regard, our results were similar to previously published data from other isolated populations. For example, in a study that was conducted on the Christian minority in Pakistan, they found that there are significant deviations from HWE at two loci (D13S317, D18S51) after Bonferroni's correction ($p \leq 0.003$). Moreover, they revealed that the Christians have unique genetic structure regarding specific alleles and their frequencies compared to the general Pakistani population (Rubab et al., 2020). Such genetic structure caused by non-random mating in isolated subpopulations can affect the implementation of these databases to populations during forensic casework. Thus, a correction factor called Theta (θ) of 0.03 needs to be implemented whenever a homozygous genotype is encountered at a locus in small isolated subpopulations. In our case, this correction factor needs to be implemented whenever we encounter a DNA fingerprint from a Samaritan or a sample of an unknown source per the recommendation of The National Institute of Standards and Technology (NIST). In population genetics, Gaza strip and the West Bank currently represent a two-island model where they are separated by a barrier with minimal gene flow and potential to exchange genetic material. Thus, we compared these two subpopulations to see if there was any effect on this forced divergence. When we compared the data available from Gaza Strip to the data we have generated for the West Bank, we found significant differentiation at one locus TPOX (p-value=5e-05) between the two datasets. This indicates that there are actual biological forces that are playing to differentiate between the two subpopulations. Over a longer period of time, the separation of the two populations may lead to even greater deviations if the genetic flow between the two subpopulations is not restored. While there is no similar situation present worldwide that resembles the condition of Gaza Strip and the West Bank taking into consideration that Gaza Strip has the most condensed population worldwide, we tried to look for very similar situation in other countries. We found that Gaza Strip-West bank situation might be similar to the situation between

South and North Korea. South and North Korea used to be a single country; however, they became culturally, geographically, and politically divided from each other after Korea's liberation from Japan in 1945. In a study conducted on butt cigarettes believed to belong to Northern Korean men using PowerPlex© Y23, they reported no significant variation among the haplotypes detected between the two populations and they were able to confirm that they remained a single population regarding Y chromosome genetic pool (Jeong et al., 2018). Nevertheless, we would assume that genetic differentiation caused by 70 years of separation could be detected at the autosomal level if were tested.

The comparison conducted in this project to reveal genetic relationships between populations from the Middle East, South Asia, and North Africa allowed us to gain more understanding regarding the genetic makeup, ethnic admixture, social, religious, and human migration that caused a high diversity within these regions. Unfortunately, the available population data does not include all of the loci we tested in our study; therefore, this limited the number of loci used in this comparison. Nevertheless, the common 13 loci that were analyzed in different populations still produced great exclusionary, polymorphic, and discriminatory results from all populations in south Asia, North Africa, and the Middle East. This also highlights the power of polymorphic STRs in forensic applications and even possibly disease susceptibility. The locus-by-locus comparison performed here revealed that Saudi Arabia exhibits strong genetic differentiation from Palestine (10/13 loci) and most of the other Middle Eastern countries except for Yemen (2/13 loci), Qatar (3/13 loci), and Oman (4/13 loci). Saudi Arabia is a large country with many isolated subpopulations such as Bedouins and tribes that inhabited their land for thousands of years without significant migration, colonization, or wars(Osman & Alsafar, 2015). Such history and lifestyle lead to significant differentiation compared to the other Middle Eastern countries as seen from our analysis; this is also supported by the differentiation degree between Yamen, Qatar, and Oman. Our results support

the migration history of individuals from these populations to the surrounding Peninsula countries before the 1970s which caused more homogenous genetic makeup between these neighboring populations(Anthony jd, 1980, Osman & Alsafar, 2015). Minimal significant differences were detected between the Palestinian population and the neighboring countries, Lebanon (3/13 loci) and Jordan (4/13 loci). This is indicative of these neighboring counties sharing common historical events and gene flow. Moreover, the lacking of significant differences found between Oman, Yamen, and UAE is indicative of the genetic relationship that resulted from initial modern human migration from the Horn of Africa to the Arabian Peninsula. The difference in differentiation significance between the Levant countries and the Arabian Peninsula reflects how countries such as Lebanon and Palestine had a different initial migration route from Africa which was mainly through Egypt with subsequent dispersion to Europe(Kundu & Ghosh, 2015). Moreover, Palestine, Lebanon, Jordan, and Syria exhibited a large expansion of ethnicities compared to the Arabian Peninsula, which resulted in different cultural behavior, mating, and genetic makeup. This was also supported by exhibiting variable significantly differentiated loci with South Asian and North African countries (Fig. 23). Tunisia appeared to be the most genetically distant country because it exhibited a large number of significantly different loci compared to any other country in this analysis, which can be explained by historical migration and invasion history in this area(Lotfi Cherni et al., 2016). Iran showed a relatively lower significant number of loci compared to most of the Middle Eastern countries, this can be explained by the migration or dispersion between South Asia and Europe, maritime trade through the Persian Gulf, and other shared historical events between Iran and the Levant(Shepard & Herrera, 2006, Grugni et al., 2012). Other South Asian countries (Pakistan and Bangladesh) showed more significant differentiation compared to the Middle East, which can be explained by the geographic distance.

Y-STRs are powerful and superior to autosomal STR in cases of sexual assaults for two reasons: First, the reduction of labor work by replacing differential extraction with Y chromosome amplification selection. Second, it facilitated the identification and recognition of male perpetrator profile(s) because the female profile is not amplified in the first place. Nevertheless, before applying Y-STRs, a national database should be built and investigated forensically to have a solid basis that this database is valid and safe to use for forensic statistics in this particular population. Moreover, it is necessary to make sure that the high-throughput technique used such as multiplex PCR and electrophoresis is applied with the correct STRs combination. In other words, to show that the Y-STRs system being applied is efficient. Many efficient Y-STR systems have been invented and applied in the forensic field worldwide and PowerPlex® Y23 is one of them. This is the system we applied in this study to build the necessary database and to investigate the Palestinian lineage, haplogroup prediction, and ancestry history.

In order to characterize the haplotypes and their frequencies, and their discrimination capacity in the Palestinian population, we examined 303 samples for 23 Y-STRs which were co-amplified in PowerPlex® Y23 system. Interestingly, our study cohort contained 275 unique haplotypes out of 303 samples (90.8%), eleven distinct haplotypes were repeated twice (7.2%), and another two distinct haplotypes were repeated three times (2%). This leads to a high haplotype diversity (99.9312812%) and discrimination capacity (95.049505%). Many loci showed high gene diversity above 70% such as DYS576, DYS389 II, and DYS48, etc. This indicates that this system can have high discrimination between different unrelated individuals in the Palestinian population. Moreover, our results revealed that gene diversity ranged from the lowest locus DYS392 with 35.98% gene diversity to the highest locus DYS458 with 88.6% gene diversity. These results are similar to results obtained from a previous study conducted on the Iraqi Arab population. In this

study, they found that the highest gene diversity in their cohort was for DYS385a/b and DYS458 loci (84.8% and 82.8%), respectively, and that the lowest genetic diversity was 40.6% in the DYS392 locus (Albarzinji et al., 2022). In a study that was performed in Syria and analyzed 13 Y-STRs, they also found allele 11 of locus DYS392 to be the most frequent and predominant (lowest diversity) among all alleles in all loci tested, which is also similar to our results (Abdin, Dewa, et al., 2003). Throughout samples processing and analysis, we were able to detect 10 samples with allele absence due to allele dropout or loci deletions. These loci were DYS549 (8 samples), DYS448 (1 sample), and DYS576 (1 sample). After processing these samples with primers that amplify larger regions that flank the locus of interest, the DYS549 locus was amplified in five samples, In the successfully amplified samples at the DYS549 locus, we identified a possible rare polymorphism (A>T) at position 107502 that reduced the binding stability of PowerPlex© Fusion primers. Which caused allele dropout in 4/5 samples. Such events reduce the efficiency of the system used in forensic applications because a different genotype could be generated from two different labs if different kits and primers are used for the same person, which can lead to the exclusion of the perpetrator by error. Moreover, such events can alter the allele frequency of the national database. This is the reason why we excluded these samples from allele frequency analysis in our project. Moreover, the same situation is encountered worldwide, for example, a published study described four bp deletions at the primer-binding site of locus D19S433 of AmpFlSTR® Identifiler kit in two cases. Generally, the frequency of such events is considered to be two events in 1026 allelic transfers (Dauber et al., 2008). In addition to the five samples that we were able to amplify with different sets of primers, we failed to amplify the dropout loci in five other samples, and thus we suspect possible deletions in these regions. DYS549 and DYS448 are found at the azoospermia factor (AZF) region of the Y-chromosome. This region plays a major role in infertility in males, and such deletions usually occur due to chromosomal rearrangements (Turrina et al.,

2015, Wang et al., 2014). We tested these samples for Y chromosome microdeletions but did not find any. For these samples we need to do further investigation to elucidate the reason(s) behind amplification failure. In addition, in our study we report a single case of the tri-allelic pattern at the DYS385 a/b locus with the genotype 17, 18, 19. This locus is originally a multi-copy locus present in the Y chromosome with two copies in the majority of populations. However, some studies described rare situations with some individuals having more than two copies. In a very rare occasion; they were able to identify a very rare situation with two individuals having 5 copies of this marker (rate= 2 out of 7760), and 1 individual with 6 copies (1 out of 4920) (F. Li et al., 2019).

Forensic industrial companies have been putting huge efforts to increase the discrimination and efficiency capacity of their systems. This was seen when we analyzed the same Palestinian dataset with the initial core 9 Y-STR, 17 Y-STR loci, and the 23 Y-STR loci where the discrimination capacity increased from 68.32%, 86.79%, to 95.05%. The importance of implementing an STR genotyping system with high discrimination power was seen when we analyzed the specific haplotypes generated from Muslims, Christians, and Samaritans. We observed that there are shared haplotypes, especially with the Christian subpopulation where we detected four-shared haplotypes among 34 samples, which is relatively quite high. In these cases, it is recommended to increase the number of Y-STRs implemented in forensic casework to 27 loci, with the addition of rapidly mutating Y-STRs. In a study that was conducted to evaluate the value of increasing the number of loci from 12 to 27 in common haplotypes and related people in the USA, they were able to prove that the 27 loci system used on samples with the common genotypes in the USA, were completely resolved (Decker et al., 2007). Thus, our results suggest that the database we generated from our study is adequate and can be implemented in forensic casework. However, additional loci

inclusion is recommended in special cases, or, in some cases, we recommend conducting both autosomal and Y-STR analysis to increase the resolution when needed.

Y-STRs are known for their heterogeneous distribution in relation to the geographical location of populations. This is mainly exhibited by the high variation detected between populations, unlike autosomal STRs where they exhibit more variation within a population. This is mainly caused by the smaller effective size, which causes stronger genetic drift. Moreover, clustering in population is evident in the Y-chromosome STRs due to patrilocality (Seielstad et al., 1998). Thus, studying population genetics and ancestry becomes more relevant with Y chromosome. Therefore, we conducted genetic distance and relationship analysis between Palestine and another 29 neighboring and distant populations worldwide to see the pattern of clustering using R_{ST} values. The clustering pattern observed in our analysis is reassuring. The African countries clustered together and showed more distance compared to non-African countries with R_{ST} values larger than 0.004 and p-value almost equal to zero for most comparisons, proving that these populations are more distinct. Of course, this result supports the idea of initial migration from Africa. Jamaica clustered together with other African countries, which goes along the migration history between Africa and Jamaica. Asian countries were also clustered together proving that they share a common ancestry, and they share relatively common historical events and gene flow. European countries clustered together along with USA and Latin America. When we looked at the samples, we revealed that many of these individuals had immigrant and admixed ancestry from Europe. The Arab countries analyzed in the study, Palestine, Iraq, and Lebanon clustered together proving that populations in these regions are more related. Moreover, the close clustering of these countries with the European countries reflects the relationships suggested previously between Arabs and Europeans(Garcia-Bertrand et al., 2014). This was also evident in the phylogenetic tree we constructed, where these

three Arab populations branched together, however, they diverged more closely from European countries compared to other Asian or African countries. Our results were further proved by the Insilico haplogroup prediction analysis we performed, our analysis revealed that the most common haplogroup in Palestine is the Q haplogroup. It was reported that this haplogroup with subclade Q-M378 harbors mainly the Middle East and it is present in the West, Central, and parts of South Asia as well as European samples. It is hypothesized that haplogroup Q originated in Eurasia around 30,000 years ago (Medina et al., 2020). This also confirms that the Palestinian population shares a common Y chromosome genetic pool with the surrounding Middle Eastern populations.

In summary, we present for the first time the first genetic data for 617 unrelated individuals from Palestine by genotyping forensic genetic markers using PowerPlex® fusion and PowerPlex® Y23 systems. Allele frequency and other population and forensic parameters were assessed for the whole population in the West Bank. Our results demonstrated that all STRs loci tested here are polymorphic, informative, and useful for forensic application. Moreover, we were able to shed the light on the relationships between the Palestinians and other populations worldwide from a genetic perspective.

Funding

This thesis was co-funded by the Arab American University-Palestine and the United Nations on Drugs and Crimes (UNODC), which supports the Palestinian forensic science laboratory through Haya project: Eliminating Violence Against Women in the West Bank and Gaza Strip.

Bibliography

1. Abdin, L., Dewa, K., Rand, S., Hohoff, C., & Brinkmann, B. (2003). Analysis of 13 Y-chromosomal STRs in an Arab population sample from Syria. *International Congress Series*. https://doi.org/10.1016/S0020-1693(02)00893-9
2. Abdin, L., Shimada, I., Brinkmann, B., & Hohoff, C. (2003). Analysis of 15 short tandem repeats reveals significant differences between the Arabian populations from Morocco and Syria. *Legal Medicine*. https://doi.org/10.1016/S1344-6223(02)00129-3
3. Abu-Amero, K. K., Gonźlez, A. M., Larruga, J. M., Bosley, T. M., & Cabrera, V. M. (2007). Eurasian and African mitochondrial DNA influences in the Saudi Arabian population. *BMC Evolutionary Biology*, *7*(1), 1–15. https://doi.org/10.1186/1471-2148-7-32/FIGURES/4
4. Abu Halima, M. S. H., Bernal, L. P., & Sharif, F. A. (2009a). Genetic variation of 15 autosomal short tandem repeat (STR) loci in the Palestinian population of Gaza Strip. *Legal Medicine*. https://doi.org/10.1016/j.legalmed.2009.02.072
5. Abu Halima, M. S. H., Bernal, L. P., & Sharif, F. A. (2009b). Genetic variation of 15 autosomal short tandem repeat (STR) loci in the Palestinian population of Gaza Strip. *Legal Medicine*, *11*(4), 203–204. https://doi.org/10.1016/j.legalmed.2009.02.072
6. AJ, J. (1987). Highly variable minisatellites and DNA fingerprints. *Biochemical Society Transactions*, *15*(3), 309–317. https://doi.org/10.1042/BST0150309
7. AL-Eitan, L. N., Darwish, N. N., Hakooz, N. M., & Dajani, R. B. (2020). Investigation of the forensic GlobalFiler loci in the genetically isolated Circassian subpopulation in Jordan. *Gene*. https://doi.org/10.1016/j.gene.2019.144269
8. Al-Eitan, L. N., & Tubaishat, R. R. (2018). Evaluation of forensic genetic efficiency parameters of 22 autosomal STR markers (PowerPlex® Fusion system) in a population sample of Arab descent from Jordan. *Australian Journal of Forensic Sciences*.

https://doi.org/10.1080/00450618.2016.1212401

9. Al-Snan, N. R., Messaoudi, S., Babu, S. R., & 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*. https://doi.org/10.1371/journal.pone.0220620
10. Albarzinji, B. M., Abdulkarim, F. M., Hussein, S. A., Rashid, D., & Lazim, H. (2022). Population genetic study of 17 Y-STR Loci of the Sorani Kurds in the Province of Sulaymaniyah, Iraq. *BMC Genomics*, *23*(1), 763. https://doi.org/10.1186/s12864-022-09005-6
11. Almarri, M. A., & Lootah, R. A. (2018). Allelic and haplotype diversity of 12 X-STRs in the United Arab Emirates. *Forensic Science International: Genetics*, *33*, e4–e6. https://doi.org/10.1016/j.fsigen.2017.12.013
12. Alshamali, F., Alkhayat, A. Q., Budowle, B., & Watson, N. D. (2005). STR population diversity in nine ethnic populations living in Dubai. *Forensic Science International*. https://doi.org/10.1016/j.forsciint.2004.09.133
13. Ambrosio, I. B., Braganholi, D. F., Orlando, L. B. M., Andrekenas, N. C., da Mota Pontes, I., da Silva, D. A., Astolfi-Filho, S., de Carvalho, E. F., Cicarelli, R. M. B., & Gusmão, L. (2020). Mutational data and population profiling of 23 Y-STRs in three Brazilian populations. *Forensic Science International: Genetics*, *48*, 102348. https://doi.org/10.1016/j.fsigen.2020.102348
14. Anthony jd, H. J. (1980). Eastern Arabian States: kuwait, Bahrain, Qatar United Arab Emirates, Oman. *The Goverment and Politics of the Middle East and North Africa. Colorado, Westview p*.
15. Baker, C., & Antonovics, J. (2012). Evolutionary determinants of genetic variation in susceptibility to infectious diseases in humans. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0029089
16. Barbujani, G., Ghirotto, S., & Tassi, F. (2013). Nine things to remember about human genome diversity. *Tissue Antigens*, *82*(3), 155–164. https://doi.org/10.1111/TAN.12165

17. Barni, F., Berti, A., Pianese, A., Boccellino, A., Miller, M. P., Caperna, A., & Lago, G. (2007). Allele frequencies of 15 autosomal STR loci in the Iraq population with comparisons to other populations from the middle-eastern region. *Forensic Science International*. https://doi.org/10.1016/j.forsciint.2006.03.005

18. Berdan, E. L., Blanckaert, A., Slotte, T., Suh, A., Westram, A. M., & Fragata, I. (2021). Unboxing mutations: Connecting mutation types with evolutionary consequences. *Molecular Ecology*, *30*(12), 2710–2723. https://doi.org/10.1111/MEC.15936

19. Birch, J. (2016). Natural selection and the maximization of fitness. *Biological Reviews*, *91*(3), 712–727. https://doi.org/10.1111/BRV.12190

20. Bittles, A. H., & Black, M. L. (2010). Consanguinity, human evolution, and complex diseases. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(SUPPL. 1), 1779–1786. https://doi.org/10.1073/PNAS.0906079106/SUPPL_FILE/0906079106SI.PDF

21. Bodmer, W. (2014). Genetic characterization of human populations: From abo to a genetic map of the british people. *Genetics*. https://doi.org/10.1534/genetics.114.173062

22. Border, R., O'Rourke, S., de Candia, T., Goddard, M. E., Visscher, P. M., Yengo, L., Jones, M., & Keller, M. C. (2022). Assortative mating biases marker-based heritability estimators. *Nature Communications*, *13*(1). https://doi.org/10.1038/S41467-022-28294-9

23. Bosch, E., Clarimón, J., Pérez-Lezaun, A., & Calafell, F. (2001). STR data for 21 loci in northwestern Africa. *Forensic Science International*. https://doi.org/10.1016/S0379-0738(00)00344-3

24. Brookes, C., Bright, J. A., Harbison, S., & Buckleton, J. (2012). Characterising stutter in forensic STR multiplexes. *Forensic Science International: Genetics*, *6*(1), 58–63. https://doi.org/10.1016/j.fsigen.2011.02.001

25. Broshi, M. (2018). The Population of Western Palestine in the Roman-Byzantine Period.

Https://Doi.Org/10.2307/1356664, *236*, 1–10. https://doi.org/10.2307/1356664

26. Bustamante, C. D., & Henn, B. M. (2010). Shadows of early migrations. *Nature 2010 468:7327*, *468*(7327), 1044–1045. https://doi.org/10.1038/4681044a
27. Butler, J. M. (2014). *Advanced topics in forensic DNA typing: interpretation*.
28. Caballero, A., & García-Dorado, A. (2013). Allelic diversity and its implications for the rate of adaptation. *Genetics*. https://doi.org/10.1534/genetics.113.158410
29. Cassar, M., Farrugia, C., & Vidal, C. (2008). Allele frequencies of 14 STR loci in the population of Malta. *Legal Medicine*. https://doi.org/10.1016/j.legalmed.2007.09.004
30. Cavalli-Sforza, L. L., & Edwards, A. W. (1967). Phylogenetic analysis. Models and estimation procedures. *American Journal of Human Genetics*. https://doi.org/10.2307/2406616
31. Chakraborty, R., & Stivers, D. N. (1996). Paternity Exclusion by DNA Markers: Effects of Paternal Mutations. *Journal of Forensic Sciences*. https://doi.org/10.1520/jfs13976j
32. Chen, Z., Xuan, Y., Liang, G., Yang, X., Yu, Z., Barker, S. C., Kelava, S., Bu, W., Liu, J., & Gao, S. (2020). Precise annotation of tick mitochondrial genomes reveals multiple copy number variation of short tandem repeats and one transposon-like element. *BMC Genomics*. https://doi.org/10.1186/s12864-020-06906-2
33. Cherni, L., Loueslati Yaâcoubi, B., Pereira, L., Alves, C., Khodjet El Kill, H., Ben Ammar El Gaaied, A., & Amorim, A. (2005). Data for 15 autosomal STR markers (Powerplex 16 System) from two Tunisian populations: Kesra (Berber) and Zriba (Arab). *Forensic Science International*. https://doi.org/10.1016/j.forsciint.2004.04.009
34. Cherni, Lotfi, Pakstis, A. J., Boussetta, S., Elkamel, S., Frigi, S., Khodjet-El-Khil, H., Barton, A., Haigh, E., Speed, W. C., Ben Ammar Elgaaied, A., Kidd, J. R., & Kidd, K. K. (2016). Genetic variation in Tunisia in the context of human diversity worldwide. *American Journal of Physical Anthropology*. https://doi.org/10.1002/ajpa.23008

35. Coudray, C., Guitard, E., El-Chennawi, F., Larrouy, G., & Dugoujon, J. M. (2007). Allele frequencies of 15 short tandem repeats (STRs) in three Egyptian populations of different ethnic groups. *Forensic Science International*. https://doi.org/10.1016/j.forsciint.2006.03.017

36. Crooks, L., Carlborg, Ö., Marklund, S., & Johansson, A. M. (2013). Identification of null alleles and deletions from snp genotypes for an intercross between domestic and wild chickens. *G3: Genes, Genomes, Genetics*. https://doi.org/10.1534/g3.113.006643

37. Decker, A. E., Kline, M. C., Vallone, P. M., & Butler, J. M. (2007). The impact of additional Y-STR loci on resolving common haplotypes and closely related individuals. *Forensic Science International: Genetics*. https://doi.org/10.1016/j.fsigen.2007.01.012

38. Deffner D, Kandler A, F. L. (2022). Effective population size for culturally evolving traits. *PLoS Comput Biol*, *18(4): e10*.

39. Della Rocca, C., Cannone, F., D'Atanasio, E., Bonito, M., Anagnostou, P., Russo, G., Barni, F., Alladio, E., Destro-Bisol, G., Trombetta, B., Berti, A., & Cruciani, F. (2020). Ethnic fragmentation and degree of urbanization strongly affect the discrimination power of Y-STR haplotypes in central Sahel. *Forensic Science International: Genetics*, *49*, 102374. https://doi.org/10.1016/j.fsigen.2020.102374

40. Dion, D., Sulzer, A., & Kratzer, A. (2011). Concordance testing with the new STR kits and the influence on the Swiss National DNA database. *Forensic Science International: Genetics Supplement Series*. https://doi.org/10.1016/j.fsigss.2011.08.042

41. Disotell, T. R. (1999). Human evolution: Origins of modern humans still look recent. *Current Biology*, *9*(17), R647–R650. https://doi.org/10.1016/S0960-9822(99)80414-X

42. Dudgeon, C. L., Blower, D. C., Broderick, D., Giles, J. L., Holmes, B. J., Kashiwagi, T., Krück, N. C., Morgan, J. A. T., Tillett, B. J., & Ovenden, J. R. (2012). A review of the application of molecular genetics for fisheries management and conservation of sharks and rays. In *Journal of*

Fish Biology. https://doi.org/10.1111/j.1095-8649.2012.03265.x

43. Edwards, A., Hammond, H. A., Jin, L., Caskey, C. T., & Chakraborty, R. (1992). Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics*. https://doi.org/10.1016/0888-7543(92)90371-X
44. Eichler, E. E. (2019). Genetic Variation, Comparative Genomics, and the Diagnosis of Disease. *New England Journal of Medicine*. https://doi.org/10.1056/nejmra1809315
45. El Andari, A., Othman, H., Taroni, F., & Mansour, I. (2013). Population genetic data for 23 STR markers from Lebanon. In *Forensic Science International: Genetics*. https://doi.org/10.1016/j.fsigen.2013.04.010
46. Erlich, H. A., & Arnheim, N. (1992). Genetic analysis using the polymerase chain reaction. In *Annual Review of Genetics*. https://doi.org/10.1146/annurev.ge.26.120192.002403
47. Etxeberria, A., & Umerez, J. (2008). Conceptual Issues in Evolutionary Biology (3rd ed.). Elliott Sober (Ed.). (2006, Cambridge MA: MIT Press.). *Artificial Life*, *14*, 235–238. https://doi.org/10.1162/artl.2008.14.2.235
48. Excoffier, L. and H. E. L. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*, 564–567.
49. Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of Molecular Variance Inferred from Metric Distances among DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data. *Genetics*, *131*(2), 479. https://doi.org/10.1093/GENETICS/131.2.479
50. Fan, H., & Chu, J. Y. (2007). A brief review of short tandem repeat mutation. *Genomics, Proteomics & Bioinformatics*, *5*(1), 7–14. https://doi.org/10.1016/S1672-0229(07)60009-6
51. Ferdous, A., Ali, M. E., Alam, S., Hasan, M., Hossain, T., & Akhteruzzaman, S. (2009). Forensic evaluation of STR data for the PowerPlex^{TM} 16 System loci in a Bangladeshi population. *Legal*

Medicine. https://doi.org/10.1016/j.legalmed.2009.02.027

52. Fernandes, V., Alshamali, F., Alves, M., Costa, M. D., Pereira, J. B., Silva, N. M., Cherni, L., Harich, N., Cerny, V., Soares, P., Richards, M. B., & Pereira, L. (2012). The Arabian cradle: Mitochondrial relicts of the first steps along the Southern route out of Africa. *American Journal of Human Genetics*. https://doi.org/10.1016/j.ajhg.2011.12.010
53. Fisher, R. A. (1951). Standard calculations for evaluating a blood-group system. *Heredity*. https://doi.org/10.1038/hdy.1951.5
54. Foreman, L. A., & Evett, I. W. (2001). Statistical analyses to support forensic interpretation for a new ten-locus STR profiling system. *International Journal of Legal Medicine*. https://doi.org/10.1007/s004140000138
55. Forster, P. (2004). Ice Ages and the mitochondrial DNA chronology of human dispersals: a review. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *359*(1442), 255. https://doi.org/10.1098/RSTB.2003.1394
56. Fujii, K., Iwashima, Y., Kitayama, T., Nakahara, H., Mizuno, N., & Sekiguchi, K. (2014). Allele frequencies for 22 autosomal short tandem repeat loci obtained by PowerPlex Fusion in a sample of 1501 individuals from the Japanese population. *Legal Medicine*. https://doi.org/10.1016/j.legalmed.2014.03.007
57. Garcia-Bertrand, R., Simms, T. M., Cadenas, A. M., & Herrera, R. J. (2014). United Arab Emirates: Phylogenetic relationships and ancestral populations. *Gene*. https://doi.org/10.1016/j.gene.2013.09.092
58. García-Nieto, P. E., Morrison, A. J., & Fraser, H. B. (2019). The somatic mutation landscape of the human body. *Genome Biology*, *20*(1). https://doi.org/10.1186/S13059-019-1919-5
59. Garte, S. (2010). Human population genetic diversity as a function of SNP type from HapMap data. *American Journal of Human Biology*, *22*(3), 297–300. https://doi.org/10.1002/AJHB.20984

60. Gettings, K. B., Borsuk, L. A., Ballard, D., Bodner, M., Budowle, B., Devesse, L., King, J., Parson, W., Phillips, C., & Vallone, P. M. (2017). STRSeq: A catalog of sequence diversity at human identification Short Tandem Repeat loci. *Forensic Science International: Genetics*. https://doi.org/10.1016/j.fsigen.2017.08.017

61. Gill, P., Urquhart, A., Millican, E., Oldroyd, N., Watson, S., Sparkes, R., & Kimpton, C. P. (1996). A new method of STR interpretation using inferential logic - Development of a criminal intelligence database. *International Journal of Legal Medicine*. https://doi.org/10.1007/BF01369596

62. Gill, Peter, Jeffreys, A. J., & Werrett, D. J. (1985). Forensic application of DNA "fingerprints." *Nature*. https://doi.org/10.1038/318577a0

63. Gouy, A., & Zieger, M. (2017). STRAF - A convenient online tool for STR data evaluation in forensic genetics. *Forensic Science International: Genetics*, *30*, 148–151.

64. Graffelman, J., & Weir, B. S. (2016). Testing for Hardy-Weinberg equilibrium at biallelic genetic markers on the X chromosome. *Heredity*. https://doi.org/10.1038/hdy.2016.20

65. Graham, M., Planz, J., Allen, M., & Shewale, S. J. (2015). *Discovery and Characterization of Tetranucleotide Short Tandem Repeats in North American Bears (Ursids)*. https://unthsc-ir.tdl.org/handle/20.500.12503/26593

66. Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M. H. Y., Hansen, N. F., Durand, E. Y., Malaspinas, A. S., Jensen, J. D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H. A., … Pääbo, S. (2010). A draft sequence of the neandertal genome. *Science*, *328*(5979), 710–722. https://doi.org/10.1126/SCIENCE.1188021/SUPPL_FILE/GREEN_SOM.PDF

67. Grugni, V., Battaglia, V., Hooshiar Kashani, B., Parolo, S., Al-Zahery, N., Achilli, A., Olivieri, A., Gandini, F., Houshmand, M., Sanati, M. H., Torroni, A., & Semino, O. (2012). Ancient

migratory events in the Middle East: New Clues from the Y-chromosome variation of modern Iranians. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0041252

68. Guo, X., & Elston, R. C. (1999). Linkage information content of polymorphic genetic markers. *Human Heredity*. https://doi.org/10.1159/000022855
69. Haldane, J. B. S. (1930). A mathematical theory of natural and artificial selection. (Part VI, Isolation.). *Mathematical Proceedings of the Cambridge Philosophical Society*. https://doi.org/10.1017/S0305004100015450
70. Hamilton Matthew B. (2009). *Population Genetics*. Wiley-Blackwell.
71. Hartl, D. L., & Borot, N. (1994). *Génétique des populations*. 305.
72. Hasan, M. M., Momtaz, P., Hossain, T., Mazumder, A. K., Khondaker, M. J. A., Sufian, A., Makhdum, M. N., & Akhteruzzaman, S. (2016). Phylogenetic and forensic studies of the Bangladeshi population using next-generation PowerPlex® Y23 STR marker system. In *International Journal of Legal Medicine*. https://doi.org/10.1007/s00414-016-1367-7
73. Hodgson, J. A., Mulligan, C. J., Al-Meeri, A., & Raaum, R. L. (2014). Early Back-to-Africa Migration into the Horn of Africa. *PLoS Genetics*. https://doi.org/10.1371/journal.pgen.1004393
74. Jankova, R., Seidel, M., Videtič Paska, A., Willuweit, S., & Roewer, L. (2019). Y-chromosome diversity of the three major ethno-linguistic groups in the Republic of North Macedonia. *Forensic Science International: Genetics*, *42*, 165–170. https://doi.org/10.1016/j.fsigen.2019.07.007
75. Jeong, K. S., Shin, H., Lee, S. J., Kim, H. S., Kim, J. Y., Han, M. S., Lee, Y. H., Park, K. W., & Chun, B. W. (2018). Genetic characteristics of Y-chromosome short tandem repeat haplotypes from cigarette butt samples presumed to be smoked by North Korean men. *Genes and Genomics*. https://doi.org/10.1007/s13258-018-0701-5
76. Jobling, M. A. (2022). Forensic genetics through the lens of Lewontin: population structure, ancestry and race. *Philosophical Transactions of the Royal Society B: Biological Sciences*,

377(1852). https://doi.org/10.1098/RSTB.2020.0422

77. Jones, D. A. (1972). Blood Samples : Probability of Discrimination. *Journal of the Forensic Science Society*. https://doi.org/10.1016/S0015-7368(72)70695-7
78. Jordan, D., & Mills, D. E. (2021). Past, Present, and Future of DNA Typing for Analyzing Human and Non-Human Forensic Samples. In *Frontiers in Ecology and Evolution*. https://doi.org/10.3389/fevo.2021.646130
79. Kanavakis, E., & Xaidara, A. (1995). The Human Genome Project. *Alcohol Health and Research World*, *19*(3), 190. https://doi.org/10.5005/jp/books/10279_22
80. Khodjet-El-Khil, H., Fadhlaoui-Zid, K., Gusmão, L., Alves, C., Benammar-Elgaaied, A., & Amorim, A. (2012). Allele frequencies for 15 autosomal STR markers in the Libyan population. In *Annals of Human Biology*. https://doi.org/10.3109/03014460.2011.630678
81. Khubrani, Y. M., Wetton, J. H., & Jobling, M. A. (2019). Analysis of 21 autosomal STRs in Saudi Arabia reveals population structure and the influence of consanguinity. *Forensic Science International: Genetics*, *39*, 97–102. https://doi.org/10.1016/j.fsigen.2018.12.006
82. Kimpton, C. P., Gill, P., Walton, A., Urquhart, A., Millican, E. S., & Adams, M. (1993). Automated DNA profiling employing multiplex amplification of short tandem repeat loci. *PCR Methods and Applications*, *3*(1), 13–22. https://doi.org/10.1101/GR.3.1.13
83. Kimura, M. (1986). DNA and the neutral theory. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *312*(1154), 343–354. https://doi.org/10.1098/RSTB.1986.0012
84. Kimura, Motoo. (1991). The neutral theory of molecular evolution: A review of recent evidence. *The Japanese Journal of Genetics*, *66*(4), 367–386. https://doi.org/10.1266/JJG.66.367
85. Kimura, Motoo, & Weiss, G. H. (1964). The Stepping Stone Model of Population Structure and the Decrease of Genetic Correlation with Distance. *Genetics*, *49*(4), 561.

https://doi.org/10.1093/GENETICS/49.4.561

86. Kitada, S., Nakamichi, R., & Kishino, H. (2021). Understanding population structure in an evolutionary context: population-specific FST and pairwise FST. *G3: Genes|Genomes|Genetics*, *11*(11). https://doi.org/10.1093/G3JOURNAL/JKAB316
87. Kong, A., Frigge, M. L., Masson, G., Besenbacher, S., Sulem, P., Magnusson, G., Gudjonsson, S. A., Sigurdsson, A., Jonasdottir, A., Jonasdottir, A., Wong, W. S. W., Sigurdsson, G., Walters, G. B., Steinberg, S., Helgason, H., Thorleifsson, G., Gudbjartsson, D. F., Helgason, A., Magnusson, O. T., … Stefansson, K. (2012). Rate of de novo mutations and the importance of father's age to disease risk. *Nature 2012 488:7412*, *488*(7412), 471–475. https://doi.org/10.1038/nature11396
88. Krawczak, M. (2009). Mathew B. Hamilton: Population genetics. *Human Genetics*. https://doi.org/10.1007/s00439-009-0755-1
89. Kumar, S., Kingsley, C., & DiStefano, J. (2017). *The Human Genome Project : Where AreWe Now and Where Are We Going ?*
90. Kundu, S., & Ghosh, S. K. (2015). Trend of different molecular markers in the last decades for studying human migrations. *Gene*, *556*(2), 81–90. https://doi.org/10.1016/J.GENE.2014.12.023
91. La Spada, A. R., Paulson, H. L., & Fischbeck, K. H. (1994). Trinucleotide repeat expansion in neurological disease. *Annals of Neurology*, *36*(6), 814–822. https://doi.org/10.1002/ANA.410360604
92. Lang, M., Song, F., Ye, Y., Xie, M., Zhu, H., Wang, Z., & Hou, Y. (2019). Sequence characterization of microvariant alleles at DYS627 and DYS458. *Forensic Science International: Genetics Supplement Series*. https://doi.org/10.1016/j.fsigss.2019.09.043
93. Lawton-Rauh, A. (2008). Demographic processes shaping genetic variation. In *Current Opinion in Plant Biology*. https://doi.org/10.1016/j.pbi.2008.02.009

94. Li, F., Zhao, P., Xiao, C., Feng, C., Chen, L., & Du, W. (2019). Identification of extra alleles in DYS385a/b multi-allelic patterns. *Legal Medicine*. https://doi.org/10.1016/j.legalmed.2019.01.003

95. Li, H. xia, Peng, D., Wang, Y., Wu, R. ga, Zhang, Y. ming, Li, R., & Sun, H. yu. (2018). Evaluation of genetic parameters of 23 autosomal STR loci in a Southern Chinese Han population. *Annals of Human Biology*. https://doi.org/10.1080/03014460.2018.1480731

96. Lins, A. M., Sprecher, C. J., Puers, C., & Schumm, J. W. (2018). Multiplex Sets for the Amplification of Polymorphic Short Tandem Repeat Loci — Silver Stain and Fluorescence Detection. *Https://Doi.Org/10.2144/96205rr01*, *20*(5), 882–889. https://doi.org/10.2144/96205RR01

97. Liu, N., Chen, L., Wang, S., Oh, C., & Zhao, H. (2005). Comparison of single-nucleotide polymorphisms and microsatellites in inference of population structure. *BMC Genetics*, *6*(Suppl 1), S26. https://doi.org/10.1186/1471-2156-6-S1-S26

98. Lohmueller, K. E., Albrechtsen, A., Li, Y., Kim, S. Y., Korneliussen, T., Vinckenbosch, N., Tian, G., Huerta-Sanchez, E., Feder, A. F., Grarup, N., Jørgensen, T., Jiang, T., Witte, D. R., Sandbæk, A., Hellmann, I., Lauritzen, T., Hansen, T., Pedersen, O., Wang, J., & Nielsen, R. (2011). Natural Selection Affects Multiple Aspects of Genetic Variation at Putatively Neutral Sites across the Human Genome. *PLOS Genetics*, *7*(10), e1002326. https://doi.org/10.1371/JOURNAL.PGEN.1002326

99. Lowe, A. (Plant ecologist), Harris, S., & Ashton, P. (2004). *Ecological genetics : design, analysis, and application*. 326. https://www.wiley.com/en-us/Ecological+Genetics%3A+Design%2C+Analysis%2C+and+Application-p-9781405100335

100. Maca-Meyer, N., González, A. M., Pestano, J., Flores, C., Larruga, J. M., & Cabrera, V. M. (2003). Mitochondrial DNA transit between West Asia and North Africa inferred from U6

phylogeography. *BMC Genetics*. https://doi.org/10.1186/1471-2156-4-15

101. Marian, A. J. (2020). Clinical Interpretation and Management of Genetic Variants. In *JACC: Basic to Translational Science*. https://doi.org/10.1016/j.jacbts.2020.05.013
102. McDougall, I., Brown, F. H., & Fleagle, J. G. (2005). Stratigraphic placement and age of modern humans from Kibish, Ethiopia. *Nature 2004 433:7027*, *433*(7027), 733–736. https://doi.org/10.1038/nature03258
103. Medina, L. S. J., Sepúlveda, P. B. P., Ramallo, V., Sala, C., Beltramo, J., Schwab, M., Motti, J. M. B., Santos, M. R., Cuello, M. V., Salceda, S., Dipierri, J. E., Gómez, E. L. A., Muzzio, M., Bravi, C. M., & Bailliet, G. (2020). Continental origin for Q haplogroup patrilineages in Argentina and Paraguay. *Human Biology*. https://doi.org/10.13110/humanbiology.92.2.01
104. Mohammad, A., & Imad, H. (2013). Autosomal STR: From locus information to next generation sequencing technology. In *Research Journal of Biotechnology*.
105. Mohammed.M Al-Zubaidi, Majeed .A Sabbah, T. Y. S. (2022). Short Tandem Repeat (STR) Variation from 6 Cities in Iraq Based on 15 Loci. *Research Square, REPRINT (Version 1) available at Research Square [https://doi.org/10.21203/rs.3.rs-2338240/v1]*.
106. Moon, M. H., Hong, S. R., & Shin, K. J. (2022). Sequence Variations of 31 Y-Chromosomal Short Tandem Repeats Analyzed by Massively Parallel Sequencing in Three U.S. Population Groups and Korean Population. *Journal of Korean Medical Science*, *37*(6). https://doi.org/10.3346/JKMS.2022.37.E40
107. Mossallam, G. I., Smith, A. G., & McFarland, C. (2005). Comparison of variable number tandem repeat and short tandem repeat genetic markers for qualitative and quantitative chimerism analysis post allogeneic stem cell transplantation. *Journal of the Egyptian National Cancer Institute*.
108. Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.70.12.3321

109. Osman, A. E., & Alsafar, H. (2015). Autosomal Short Tandem Repeat (STR) Variation Based on 15 Loci in a Population from the Central Region (Riyadh Province) of Saudi Arabia. *Journal of Forensic Research*. https://doi.org/10.4172/2157-7145.1000267

110. Pagnotta, M. (2018). Comparison among Methods and Statistical Software Packages to Analyze Germplasm Genetic Diversity by Means of Codominant Markers. *J*. https://doi.org/10.3390/j1010018

111. Panneerchelvam, S., & Norazmi, M. N. (2003). Forensic DNA profiling and database. In *Malaysian Journal of Medical Sciences*.

112. Paulson, H. (2018). Repeat expansion diseases. *Handbook of Clinical Neurology*, *147*, 105–123. https://doi.org/10.1016/B978-0-444-63233-3.00009-9

113. *PCBS*. (n.d.). Retrieved October 29, 2021, from https://www.pcbs.gov.ps/default.aspx

114. Pérez-Miranda, A. M., Alfonso-Sánchez, M. A., Peña, J. A., & Herrera, R. J. (2006). Qatari DNA variation at a crossroad of human migrations. *Human Heredity*. https://doi.org/10.1159/000092648

115. Petraglia, M. D., Haslam, M., Fuller, D. Q., Boivin, N., & Clarkson, C. (2010). Out of Africa: new hypotheses and evidence for the dispersal of Homo sapiens along the Indian Ocean rim. *Http://Dx.Doi.Org/10.3109/03014461003639249*, *37*(3), 288–311. https://doi.org/10.3109/03014461003639249

116. Plodthong, N., Chareonsirisuthigul, T., Lumjiaktase, P., & Rerkamnuaychoke, B. (2014). Analysis of paternity testing results by Identifiler system in Thailand. *Genomics and Genetics*.

117. Potter, L. G. (2009). The Persian Gulf in History. In *The Persian Gulf in History*. https://doi.org/10.1057/9780230618459

118. Preet, K., Malhotra, S., Shrivastava, P., Jain, T., Rawat, S., Varte, L. R., Singh, S., Singh, I., & Sarkar, S. (2016). Genetic Diversity in Gorkhas: an Autosomal STR Study. *Scientific Reports*

2016 6:1, *6*(1), 1–9. https://doi.org/10.1038/srep32494

119. Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics 2006 38:8*, *38*(8), 904–909. https://doi.org/10.1038/ng1847

120. Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945. https://doi.org/10.1093/GENETICS/155.2.945

121. Quintana-Murci, L., Semino, O., Bandelt, H. J., Passarino, G., McElreavey, K., & Santachiara-Benerecetti, A. S. (1999). Genetic evidence of an early exit of Homo sapiens sapiens from Africa through eastern Africa. *Nature Genetics 1999 23:4*, *23*(4), 437–441. https://doi.org/10.1038/70550

122. Rakha, A., Yu, B., Hadi, S., & Sheng-bin, L. (2009). Population genetic data on 15 autosomal STRs in a Pakistani population sample. *Legal Medicine*. https://doi.org/10.1016/j.legalmed.2009.08.001

123. Raymond M. & Rousset F. (1995). *GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J. Heredity*. *86*, 248–249.

124. Reich, D. E., Cargili, M., Boik, S., Ireland, J., Sabeti, P. C., Richter, D. J., Lavery, T., Kouyoumjian, R., Farhadian, S. F., Ward, R., & Lander, E. S. (2001). Linkage disequilibrium in the human genome. *Nature 2001 411:6834*, *411*(6834), 199–204. https://doi.org/10.1038/35075590

125. Reich, D., Green, R. E., Kircher, M., Krause, J., Patterson, N., Durand, E. Y., Viola, B., Briggs, A. W., Stenzel, U., Johnson, P. L. F., Maricic, T., Good, J. M., Marques-Bonet, T., Alkan, C., Fu, Q., Mallick, S., Li, H., Meyer, M., Eichler, E. E., … Pääbo, S. (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature 2010 468:7327*, *468*(7327), 1053–1060. https://doi.org/10.1038/nature09710

126. Reid, K. M., & Heathfield, L. J. (2020). Allele frequency data for 23 Y-chromosome short tandem repeats (STRs) for the South African population. In *Forensic Science International: Genetics* (Vol. 46). Elsevier Ireland Ltd. https://doi.org/10.1016/j.fsigen.2020.102270

127. Reilly, B. (2013). Revisiting consanguineous marriage in the greater Middle East: Milk, blood, and bedouins. *American Anthropologist*. https://doi.org/10.1111/aman.12023

128. Ribeiro, J., Romero, M., Simão, F., Ferreira Almeida, A. P., Quiroz, A., Machado, P., Velázquez, V., de Carvalho, E. F., Vullo, C., & Gusmão, L. (2018). Analysis of 23 Y-STRs in a population sample from eastern Paraguay. In *Forensic Science International: Genetics* (Vol. 37, pp. e20–e22). Elsevier Ireland Ltd. https://doi.org/10.1016/j.fsigen.2018.08.007

129. Rios Moura, R., Oliveira Gonzaga, M., Silva Pinto, N., Vasconcellos-Neto, J., & Requena, G. S. (2021). Assortative mating in space and time: patterns and biases. *Ecology Letters*, *24*(5), 1089–1102. https://doi.org/10.1111/ELE.13690

130. Roewer, L. (2013). DNA fingerprinting in forensics: past, present, future. *Investigative Genetics*, *4*(1), 22. https://doi.org/10.1186/2041-2223-4-22

131. Rubab, A., Shafique, M., Javed, F., Saleem, S., Zahra, F. T., McNevin, D., & Shahid, A. A. (2020). Population genetic portrait of Pakistani Lahore-Christians based on 32 STR loci. *Scientific Reports*. https://doi.org/10.1038/s41598-020-76016-2

132. Ryman, N., & Utter, F. (1987). *Population genetics & fishery management*. Washington Sea Grant Program.

133. Schwarcz, H. P., & Grun, R. (1992). Electron spin resonance (ESR) dating of the origin of modern man. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *337*(1280), 145–148. https://doi.org/10.1098/RSTB.1992.0091

134. Seielstad, M. T., Minch, E., & Cavalli-Sforza, L. L. (1998). Genetic evidence for a higher female migration rate in humans. *Nature Genetics*. https://doi.org/10.1038/3088

135. Sensabaugh, G. F. (1982). Biochemical markers of individuality. *Forensic Science Handbook*.

136. Shepard, E. M., & Herrera, R. J. (2006). Iranian STR variation at the fringes of biogeographical demarcation. *Forensic Science International*. https://doi.org/10.1016/j.forsciint.2005.05.012

137. Shriver, M. D., Jin, L., Boerwinkle, E., Deka, R., Ferrell, R. E., & Chakraborty, R. (1995). A novel measure of genetic distance for highly polymorphic tandem repeat loci. *Molecular Biology and Evolution*. https://doi.org/10.1093/oxfordjournals.molbev.a040268

138. Smith, J. R., Carpten, J. D., Brownstein, M. J., Ghosh, S., Magnuson, V. L., Gilbert, D. A., Trent, J. M., & Collins, F. S. (1995). Approach to genotyping errors caused by nontemplated nucleotide addition by Taq DNA polymerase. *Genome Research*, *5*(3), 312–317. https://doi.org/10.1101/GR.5.3.312

139. Smith, R. L., & Smith, T. M. (2001). Plant adaptations to the environment II: thermal, moisture, and nutrient environment. *Ecology and Field Biology, Sixth Edition. Benjamin Cummings, San Francisco, California*, 98–119.

140. Sparks, Rachel Thyrza;Wagemakers, B. (n.d.). *Digging Up Jericho : Past, Present and Future / , Bill - Oxford : Archaeopress, - ISBN: 9781789693522 - Permalink: http://digital.casalini.it/9781789693522 - Casalini id: 4732488. 320 p.*

141. Springate, L., & Frasier, T. R. (2017). Gamete compatibility genes in mammals: Candidates, applications and a potential path forward. In *Royal Society Open Science*. https://doi.org/10.1098/rsos.170577

142. Stanley, U. N., Khadija, A. M., Bukola, A. T., Precious, I. O., & Davidson, E. A. (2020). Forensic DNA Profiling: Autosomal Short Tandem Repeat as a Prominent Marker in Crime Investigation. *The Malaysian Journal of Medical Sciences : MJMS*, *27*(4), 22. https://doi.org/10.21315/MJMS2020.27.4.3

143. Stapley, J., Feulner, P. G. D., Johnston, S. E., Santure, A. W., & Smadja, C. M. (2017).

Recombination: The good, the bad and the variable. *Philosophical Transactions of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rstb.2017.0279

144. Star, B., & Spencer, H. G. (2013). Effects of genetic drift and gene flow on the selective maintenance of genetic variation. *Genetics*, *194*(1), 235–244. https://doi.org/10.1534/GENETICS.113.149781/-/DC1/GENETICS.113.149781-2.PDF
145. Strachan, T., & Read, A. (2018). Human Molecular Genetics. In *Human Molecular Genetics*. https://doi.org/10.4324/9780203833544
146. Stringer, C. (2012). What makes a modern human. *Nature 2012 485:7396*, *485*(7396), 33–35. https://doi.org/10.1038/485033a
147. Taboun, Z. S., Walter, R. P., Ovenden, J. R., & Heath, D. D. (2021). Spatial and temporal genetic variation in an exploited reef fish: The effects of exploitation on cohort genetic structure. *Evolutionary Applications*. https://doi.org/10.1111/eva.13198
148. Teotónio, H., Chelo, I. M., Bradić, M., Rose, M. R., & Long, A. D. (2009). Experimental evolution reveals natural selection on standing genetic variation. *Nature Genetics 2009 41:2*, *41*(2), 251–257. https://doi.org/10.1038/ng.289
149. Thorne, A. G., & Wolpoff, M. H. (1992). The multiregional evolution of humans. *Scientific American*. https://doi.org/10.1038/scientificamerican0492-76
150. Tsimhoni, D. (1984). The Status of the Arab Christians under the British Mandate in Palestine. *Middle Eastern Studies*. https://doi.org/10.1080/00263208408700605
151. Turrina, S., Caratti, S., & De Leo, D. (2013). Powerplex® Y 23 System: Molecular characterization of a null allele at locus DYS549. *Forensic Science International: Genetics Supplement Series*, *4*(1), e117–e118. https://doi.org/10.1016/J.FSIGSS.2013.10.060
152. Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. A., & Kinzler, K. W. (2013). Cancer genome landscapes. *Science (New York, N.Y.)*, *339*(6127), 1546–1558.

https://doi.org/10.1126/SCIENCE.1235122

153. Wang, Y., Dang, Z., Zhang, G., Li, S., Liu, Q., Li, C., Hou, X., Li, H., Chen, S., Cui, W., Wang, D., Kong, X., & Man, D. (2019). Genetic diversity and haplotype structure of 27 Y-STR loci in a Han population from Jining, Shandong province, eastern China. In *Forensic Science International: Genetics* (Vol. 42, pp. e25–e26). Elsevier Ireland Ltd. https://doi.org/10.1016/j.fsigen.2019.06.008

154. Weir, B. S. (1992). Population genetics in the forensic DNA debate. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(24), 11654–11659. https://doi.org/10.1073/pnas.89.24.11654

155. Wigginton, J. E., Cutler, D. J., & Abecasis, G. R. (2005). A note on exact tests of Hardy-Weinberg equilibrium. *American Journal of Human Genetics*, *76*(5), 887–893. https://doi.org/10.1086/429864

156. Wong, W. S. W., Solomon, B. D., Bodian, D. L., Kothiyal, P., Eley, G., Huddleston, K. C., Baker, R., Thach, D. C., Iyer, R. K., Vockley, J. G., & Niederhuber, J. E. (2016). New observations on maternal age effect on germline de novo mutations. *Nature Communications*. https://doi.org/10.1038/ncomms10486

157. Wright, S. (1931). Evolution in Mendelian Populations. *Genetics*, *16*(2), 97–159. https://doi.org/10.1093/GENETICS/16.2.97

158. Wright, S. (1943). Isolation by Distance. *Genetics*, *28*(2), 114–138. https://doi.org/10.1093/GENETICS/28.2.114

159. Xing, J., Watkins, W. S., Shlien, A., Walker, E., Huff, C. D., Witherspoon, D. J., Zhang, Y., Simonson, T. S., Weiss, R. B., Schiffman, J. D., Malkin, D., Woodward, S. R., & Jorde, L. B. (2010). Toward a more uniform sampling of human genetic diversity: A survey of worldwide populations by high-density genotyping. *Genomics*. https://doi.org/10.1016/j.ygeno.2010.07.004

160. Yoon, H. R., Park, Y. S., & Kim, Y. K. (2002). Rapid Prenatal Detection of Down and Edwards Syndromes by Fluorescent Polymerase Chain Reaction with Short Tandem Repeat Markers. *Yonsei Medical Journal*, *43*(5), 557–566. https://doi.org/10.3349/YMJ.2002.43.5.557

ملخص الرسالة

الهدف من هذه الرسالة هو دراسة وتحديد التنوع الجيني ونسبة البوليمورفية في 23 موقع جيني تكراري ترادفي متنوع في الكروموسومات الجسدية و23 موقع جيني تكراري ترادفي متنوع في كروموسوم Y للشعب الفلسطيني في الضفة الغربية. هذه المواقع الجينية تدعى Short tandem repeats (STRs)، دراسة التنوع الجيني لهذه المواقع الجينية يعزز فهم اختلاف الجينات السكانية للشعوب المختلفة، وكذلك يساعد في بناء قواعد البينات الجينية الضرورية لقطاع العدالة. تم دراسة التنوع الجيني للشعب الفلسطيني لهذه المواقع الجينية عن طريق تحليل 617 عينة مأخوذة من جميع محافظات الضفة الغربية باستخدام أنظمة PowerPlex Fusion® 6C and PowerPlex® Y23، بعد ذلك تم تحليل العينات على جهاز ABI genetic analyzer 3500. فيما يخص Autosomal STRs، نسبة تكرار الأليلات تفاوتت من 53.97% إلى 0.081%، بالإضافة إلى أنه تم إيجاد 336 اليل في جميع هذه المواقع الجينية. وبعد تطبيق نظام Bonferroni، لم يتم إيجاد إي انحراف لهذه المواقع الجينية عن قاعدة Hardy Weinberg Equilibrium (HWE). وكذلك، لم يتم إيجاد انحراف عن قاعدة Linkage equilibrium للمواقع الجينية المتواجدة على نفس الكروموسوم. نسبة heterozygosity تراوحت من 58% في TPOX إلى 92.5% في SE33. وتم قياس نسبة الاستثناء، نسبة التفرقة، ونسبة التوافق لجميع المواقع وكانت 0.999999999، 1، 4.20677E-30، على التوالي. بالرغم من ذلك، عند دراسة المجتمعات الجزئية في المجتمع الفلسطيني، تبين وجود انحراف عن قاعدة HWE عند السامريين والتي تؤشر إلى وجود نسبة أعلى من تزاوج الأقارب لديهم. أما فيما يخص ال 23 موقع جيني المحللة على كروموسوم Y، تم تحليل 303 عينة من جميع أنحاء الضفة الغربية. وتم إيجاد 275 نمط فرداني Haplotype في المجتمع الفلسطيني. نسبة تكرار الأليلات تراوحت بين 79.2% إلى 0.33% مع مجموع 156 اليل مختلف على طول كروموسوم Y، ومعدل التنوع الجيني لجميع المواقع الجينية يساوي 65%. وتم قياس نسبة تنوع النمط الفرداني (Haplotype diversity)، نسبة التفرقة، ونسبة التوافق لجميع المواقع وكانت 99.9%، 95%، 0.39%، على التوالي. وعلى ضوء هذه النتائج نستطيع أن نستنتج بأنه من الممكن استخدام هذه المواقع الجينية وقواعد البيانات الجينية المستخلصة من هذه الدراسة لفحوصات الطب الشرعي وفحوصات اثبات الأبوة في الضفة الغربية.