2008-12-05

# A Population Genetic Study of Middle Eastern Populations Using DYS 458 Microvariants and Cohen Modal Haplotypes

Enass Nabeel Tinah
*Brigham Young University - Provo*

# A POPULATION STUDY OF MIDDLE EASTERN POPULATIONS USING DYS458 MICROVARIANTS AND COHEN MODAL SHORT TANDEM REPEATS

By: Enass Tinah

A thesis submitted to the faculty of

Brigham Young University

in partial fulfillment of the requirements for the degree of

Master of Science

Department of Microbiology and Molecular Biology

Brigham Young University

December 2008

GRADUATE COMMITTEE APPROVAL

of a thesis submitted by

Enass Tinah

This thesis has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

| | |
|---|---|
| Date | Brent L. Nielsen, Chair |
| Date | Paul E. Evans |
| Date | Arnold Green |

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the thesis of Enass Tinah in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

Date

Brent L. Nielsen
Chair, Graduate Committee

Accepted for the Department

Brent L. Nielsen
Chair, Department of Microbiology and Molecular Biology

Accepted for the College

Rodney J. Brown
Dean, College of Life Sciences

# ABSTRACT

A POPULATION STUDY OF MIDDLE EASTERN POPULATIONs USING DYS458 MICROVARIANTS AND COHEN MODAL SHORT TANDEM REPEATS

Enass Tinah

Department of Microbiology and Molecular Biology

Master of Science

A comprehensive population study in the Middle East was conducted using different genetic markers in order to establish a wider genetic profile of the Middle Eastern populations. The main goal of this study was to analyze DNA from samples collected from different locations, and produce genetic motifs and patterns that could be used to identify and distinguish the target populations. This information will allow us to analyze the ancestry of these populations, their interactions through time and space, and the effects these interactions have on the populations' structure.

We have collected around 1300 individual samples from different populations in the Middle East ranging from Oman in southern Arabia, to the West Bank and Gaza in Palestine. Our samples can be divided into two primary groups: 320 samples come from Oman; this region is important because of its geographical proximity to Yemen which is perceived as the historical area where the Arabs originated and 800 samples came from Palestine, a central region in the Middle East that connects Asia and Africa and was a passageway between the two continents through history. The samples collected from Oman have genealogy charts that were provided by the participants, while the samples from Gaza lack the genealogy charts.

DNA was extracted from the samples, amplified using PCR technology, sequenced and genotyped using non-recombining genetic markers. Short Tandem Repeats (STR) were screened in the samples. A specific STR marker, DYS458, exhibited an alternative allele at repeat number 18 which was a 2 base pair shift identified as microvariant 18.2. The samples showed an unusually high frequency of microvariant 18.2. When microvariant 18.2 haplotype was combined with Modal Cohen Haplotype (MCH) motifs, we were able to infer some genetic characteristics about our samples.

The Cohen Modal haplotype was used in combination with the results of the 18.2 DYS458 analyses to construct a snapshot of the Middle East, using NETWORK software to analyze the relationship within and between the samples from Oman, which is located in Southern Arabia, and the areas North of Arabia (Jordan, Syria, and Palestine), which is also known as the Levant.

# ACKNOWLDGEMNTS

I am so grateful for many people who helped bring this work and research to its culmination. I would like to express my gratitude and appreciation for Drs Brent Nielsen, Paul Evans and Arnold Green for the guidance, help and encouragement that made completing my degree possible. I am thankful for the love of my parents, Nabeel and Mariam, their unequivocal support helped me navigate some difficult times. For Laila, the best of friends, this could not have happened without you. To my family and friends, thank you.

TABLE OF CONTENTS

LIST OF FIGURES/TABLES

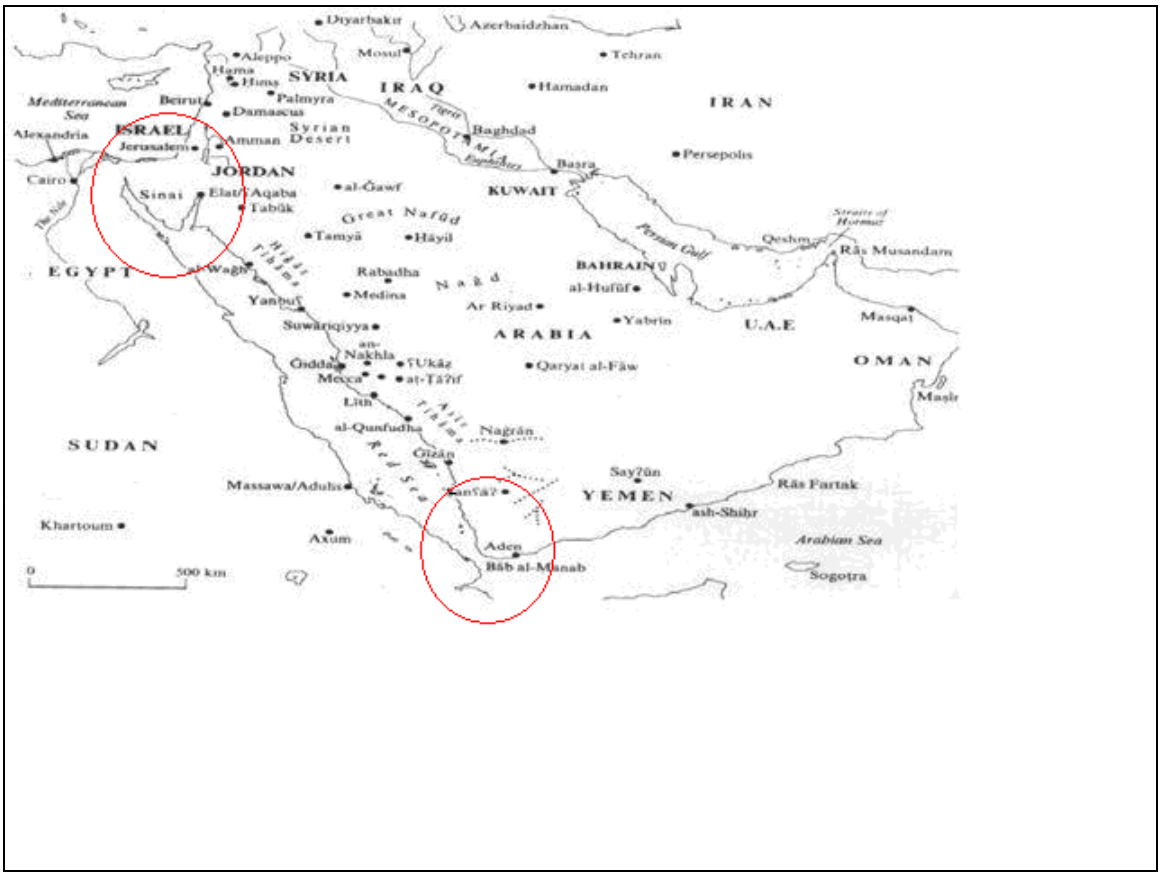
## Introduction

The Middle East has been the major crossroad between the major civilizations of Asia, Africa, and Europe. Although the current population of the Middle East is essentially a "snapshot in time", it is important to understand the current genetic structure of the Middle Eastern population to better understand the relationship between individuals in the Middle East and the surrounding areas. It may also be possible to reconstruct "echoes" of the past history of the region from the genetic record preserved in the people of the region. In some instances the genetic history of a people may predate the recorded historical information of a region. It is essential that appropriate references to known historical events be used to understand and interpret the genetic data.

## History of the Area

The Arabian Peninsula (Fig.1) is most likely the cradle of the Semitic race; it is the place where early populations migrated from, and later became the Babylonians, the Assyrians, the Phoenicians, Hebrews, and the Arabs (Hitti, 1970). The languages of western Asia and the horn of Africa, especially Arabic, Assyrian, Aramaic, Hebrew, and Ethiopic, resemble each other to a degree which points to the presence of a common linguistic root. The social institutions and the religious beliefs of the people who speak these languages reveal a similar connection, which indicates a common ancestry (Hitti, 1970,Wiess,1987).

**Figure 2: Map of the Arabian Peninsula. Yemen area is circled in the south, the Levant area is circled in the north (Resto 2003).**




Arab is a term which describes the indigenous inhabitants of the Arabian Peninsula (Fig.1) who roamed the desert, lived in tents, and used camels extensively. The history of the region is complex because the area has been a passageway between three continents: Asia, Africa, and Europe. Arabs today are not defined by any racial criterion that divides the population into sup-groups. Instead, the Arabs were divided geographically into the northern Arabs who were nomads, and the southern Arabs who inhabited the southern parts of the Arabian Peninsula in Yemen and Oman. (Gabrieli, 1963). It is from this division that the majority of the Arabs today claim an ancestry that originated in Yemen. It is useful to note that the genealogical division of Arabs is not uniform throughout the long history of the region,

but there are different divisions at different times, and that is what we will try to elaborate in this historical background.

The first division separates the Arabs into two groups: the extinct Arab (ba'idah), including tribes like Thamud, Ad, Tasm Jadis, and Jurhum which belonged mainly to two geographical locations, Hiğaz and al-Yamamah. Many of these populations were contemporary to the civilizations of that time. The Thamud were cited in the annals of the Assyrian King Sargon II (715 B.C). Those tribes no longer exist and the holy Quran tells the story of the destruction of the people of Ad and Thamud. The second group is the surviving Arabs (baqi'yah), and out of the surviving Arabs come the second division, which divides the Arabs into two groups: The Arabian Arabs, and the Arabicized Arabs. It is important to note that the division was more of a linguistic one in nature rather than racial. It reflected that a portion of the population were the original speakers of the developed Arabic, while the other portion had to learn it (Retso, 2003), The Arabian Arabs ('aribah) represented the original inhabitants of the peninsula, among the real Arabs the tribes of Qahtan (from Yemen), and Quda'a (from Northern Hiğaz and Syria). The Arabicised Arabs were the descendants of Ishmail, which came from the tribes of Adnan in central and eastern Arabia, in addition to the Najdis, and Nabateans.

In the beginning of the Abbasid period, the Arabs consisted of 5 major tribal groups: Qays which originated in the area south and east of Medina, the Quda'a tribes which were from Syria, Yaman tribes that originated from southern Arabia, Rabi'a tribes which inhabited the East and north of Nağd, and finally Mudar which represent a group of tribes that came from Mecca and the regions eastward (Retso 2003). Another division separated the Arab tribes into three major parts: Adnan encompassing Rabi'a and Mudar tribes, the Qahtan division encompassing the Yemenies, and the Quda'a tribes.

The major theme in the history of the populations of Arabia is the language; it is the acquisition of Arabic as a language that defined the hierarchy and genealogy of the people of the area. The term Arab could not have been a genealogical term (Retso, 2003).

The idea that Arabia could be the ancient ancestral land for all those populations and tribes may be due to its geographical nature. Arabia is mostly a desert with limited resources. When the population increased in the ancient times, the logical solution was migration. Waves of movement took place at different intervals.

One of the major events that occurred in Yemen and contributed to the displacement of the tribes was the destruction of the great dam of Ma'rib. This event happened between the years 542-570 A.D, which led to migration of several tribes to Ethiopia in the African horn, Al –Hijaz in the Arabian Peninsula, and to Syria in the north (Hitti, 1970). Some of the Arab tribes that claim a migration out of Yemen due to the great flood are: Banu Ghassan and Tanukh who inhabited Syria and Iraq, Banu Tayy and Kindah in central Arabia.

The other major population movement came after the rise of Islam, after 632-661 A.D, where the newly unified Arab tribes left the peninsula to spread the new religion to the Fertile Crescent, the Levant, Egypt, and North Africa. The Muslim Armies were composed mostly of Yemeni tribes who lived in Arabia at the time of the conquests. These tribes settled in the new lands, and they must have added to the genetic landscape of the area conquered.

**The study of Middle Eastern populations:**

**Classical markers**

As long as the human race existed, humans wondered about who they are and where they came from. Anthropology developed to answer those questions. Anthropology is divided into two main divisions: physical anthropology and cultural anthropology. The former deals with humans from a biological viewpoint and criteria like body size and form were the initial markers used to study populations. Later, with the development of genetics and biology, modern physical anthropology began to apply genetics and biochemistry to study the diversity of human populations. The latter can be divided into three branches: archeology, which studies history by examining the material artifacts left by ancient

cultures in order to understand the nature and the history of past populations, linguistics, which studies human languages, and deals with their construction, how were they developed, and their history, and ethnology which studies contemporary cultures by observing their different customs and habits.

In this review, we will concentrate on physical anthropology, since it is the history of the populations that we are interested in the most. We will concentrate on several physical and biological attributes, like blood groups, cranial features, and bone measurements.

**Blood Groups**

The main classification of human blood groups depends on the presence or lack of certain antigens on the surface of the red blood cell. The two antigens are A and B. If the cell carries the A antigen then the blood type is A, if it carries the B antigen then the blood type is B, if it carries both A and B, then the blood type is AB, if it carries neither, then the blood type is O. Other blood systems of interest are the Kell system, MN system and the Rh system.

The Arabs of Arabia resemble the mountainous and island populations of Europe and North Africa in having a high frequency of the O blood type (Mourant, 1976). They also exhibit a peak frequency of M type accompanied by high frequency of S type and Kell type.The High frequency of O, M, S, K may indicate an original characteristic of the populations in Arabia, which was retained in the populations because of geographic isolation of the tribes, and a social system of intermarriage, 75% of Saudi Arabs have O gene type and 70% are M type.

Going north, the levels of O blood group start to decrease and the levels of A blood group increase, so the Arab populations of Jordan and Palestine have lower levels of O and M and the A type is about 30% of the population. In Lebanon, the Arab populations show higher A type and a lower M type, and in Syria and Iraq, we start noticing an increase in the B type levels; the frequency of the B gene in Iraq ranges from 15-20%, while the A gene frequency ranges between 21-27%. Going west, the levels of A and B types increase in Egypt, with lower levels of M present; the A gene frequency is ~25%, and the

B gene frequency is ~19%. The M type is present in about 53.2% of the population, which is lower than the levels in Saudi Arabia and the Levant countries.

The Jewish communities reveal high variability in their blood grouping; the percentages of each blood group are similar to those of the host populations. The total Jewish populations in Eastern Europe have gene frequency of 28.5% for A, and 12.5% for B which is similar to the frequencies in Ashkenazi Jews in Israel and Eastern European. The Sephardic Jews, however, show a slightly different average frequency of ~25% of A, and ~13% of B, although different studies on several populations revealed a large variability (Mourant, 1976).

The Yemeni Jews have low frequencies of A and B, but very high M, and this profile resembles the Yemeni Arabs. The Jews in Iraq and Iran exhibit high frequency of the B gene, which is higher than any B frequency in the area for non Jewish populations (Mourant, 1976).

The study of blood group genes is more complex than the basic ABO group, but extensive studies covering all populations and all markers are rare, and it is important to remember that using ABO groups alone to study a population's makeup and history is faulty because blood groups frequencies may be affected by linkage to other hemoglobins or environmental factors like the presence of Malaria in an area.

**Anthropometry measures:**

Studying physical attributes in populations provides an excellent tool to investigate the variability among and between groups. Chistov (1994) studied human cranial remains in Yemen. He examined skulls belonging to a town that reached it peak during the Hadramawt civilization, which lasted for 100 years and began to decline in the first century A.D. Chistov (1994) compared the cranial measurements of the Yemeni remains with published measurements of several cranial series from the Near East, the Middle East, the South Eastern Mediterranean, and Asia. Modern Bedouins were also represented in the sample. The result of this study indicated that the ancient cranial series were closer to the morphology of the modern Bedouins living in the deserts of Yemen. These skulls did not have any

African admixture in them, which was explained by Chistov (1994) as a conservation of the Mediterranean type in Bedouins of south Yemen and Northern Arabia.

Chistov (1996) investigated 13 head and face measurements in 21 locations in Yemen, and the measurements were used to study the variation within the inhabitants of Yemen. Six markers were chosen to compare the variation of the Yemeni population with other populations in the Arabian Peninsula, the Near East, East Africa, North Africa, and central Africa. The markers that had the highest informational validity index were head length, head breadth, bizoygomatic diameter, facial height, nasal height, and nasal breadth.

The variation within the Yemeni populations revealed a significant difference between the samples from Hadramawt and the other samples from Yemen. It is believed that this difference is due to an African element that was introduced to the Hadramawt anthropological type via migration between Eastern Africa, a1nd southern Arabia.

Other differentiation traits(head breadth, height and facial height) were observed by Chistov (1996) between the recently settled nomadic groups of the desert and the settled agricultural group of the plateau areas. This may be explained by the different degree of admixture that occurred throughout history between the ancient autochthonic inhabitants of southern Yemen and the nomadic tribes that migrated from Arabia.

The study revealed that the south Yemen material is related more to the anthropological type described as southern Mediterranean more than the anthropological type of east Africa. The head and face measurement of the population of south Yemen have a large similarity with the agricultural groups inhabiting the central and northern parts of the Arabian Peninsula and Mesopotamia.

Physical traits provide much information about a population. Comparing modern traits with ancient ones can shed light on the history and the origin of a population. The problem is the availability of ancient remains, and whether their conditions can allow a thorough study of the physical traits.

**The genetic picture of the Middle East:**

The location of the Middle East played an important role in shaping the genetic structure of the people living there. The area was a crossroads between civilizations. Cultural and technological innovation spread from the area to Europe, Asia, and Africa (Cavalli-Sforza et al, 1994). It is important to increase the knowledge of genetic structures in the Middle Eastern populations in order to better understand the relationship between the Middle East, Asia Minor, Europe, and Africa.

The studies of the Middle Eastern populations reveal that there are distinguishing genetic characteristics in the people living today in the area. The Middle East was the probable homeland of numerous mtDNA and Y-chromosome haplogroups (Al-Zahery, 2003).

The genetic picture of the Middle Eastern populations is not complete. Some research was done on some populations, using selected or fewer markers (Krings, 1999, Hammer 2000, Nebel, 2001, Nebel, 2002, Rosenberg, 2002, Thomas, 2002, Bonne-Tamir, 2003, Luis, 2004). These studies reveal that there are unique genetic characteristics of the people in the Middle East, but they reveal also that the area has taken from and contributed to the genetic makeup of the neighboring populations.

**The Y-chromosome**

The Y-chromosome is a powerful tool to study genealogies. It is inherited paternally (Jobling, 1995), lacks recombination, and carries a wide range of polymorphisms (Underhill, 2000). The Y-chromosome is a large linear molecule (approx. 60 Mb), and it preserves a unique record of mutational events (Hammer, 1994), the Y-chromosome haplotypes combination have been used as tool to study human migrations (Hammer, 1998). It contains a simple record of the past that helps elaborate the evolutionary relationships of modern Y-chromosome (Jobling, 1995).

**STR markers**

Two types of markers are studied in the Y-chromosome, microsatellites and biallelic polymorphic sites. Microsatellites are blocks of 2-6 base pair tandem repeat units that are ubiquitous throughout the genome, abundant, and highly polymorphic. The variation in the number of repeats creates the different alleles (Carvalho-Silva.1999). They demonstrate high levels of heterozygosity due to the high mutation rate that allows for the inference of phylogenies among populations (Shriver.1997, White.1999). Microsatellites reflect a more recent genetic event because of their higher mutation rate (Nebel.2000). There are hundreds of STR markers that are used to characterize populations. They are very useful due to their high heterozygosity that comes from having 3 or more alleles of the same marker. Combining several Y-STR markers can yield a motif that can be tested in a population. One of the best known motifs is the Cohen Modal Haplotype (CMH) (Thomas et al 1998), which combines six Y-STR loci with certain repeat numbers. The CMH allele profile is: DYS19/14, DYS388/16, DYS390/23, DYS391/11, DYS392/12, and DYS393/10. This motif was found in high frequency in Jewish populations, and it is assumed that this represents the genetic type of the ancient Jewish priests. Variation in the repeat number produces different modals that fit different populations.

**Binary Markers**

Binary markers on the other hand, have a lower mutation rate, which allows the reconstruction of the ancestral state and can preserve the population specific haplotype information that spans human population history (Underhill, 1997). Binary markers represent unique event polymorphisms in human evolution; these events could be a single nucleotide polymorphism(Y-SNP) or insertion /deletions at specific sites on the Y-chromosome. They allow identification of deep splits in the Y-chromosome genealogy.

The nomenclature of Y-SNPs has been inconsistent depending on different research groups and their way of reporting their results. Sometimes the same haplogroups of the Y-SNPs have different

names, which caused confusion in reporting and understanding the genetics of the different populations around the world. The nomenclature was unified in 2002 by the Y Chromosome Consortium (YCC). The different haplogroups of Y-SNP were defined and named according to a system that allows easy incorporation of new discovered mutations.

Nebel et al (2001) examined the three Jewish populations and their genetic relationship to the Middle Eastern gene pool. They examined the Y chromosome using binary and STR markers and found that high numbers of the Middle Eastern and Jewish samples studied fell into Hg9, equivalent to Hg J in the current nomenclature. 55.2% of Palestinian samples, 65% of Bedouins, and 40% of Kurds belonged to HgJ.

This haplogroup was further divided into Eu9 and Eu10 sub branches and they suggested that the origin of Eu10 is the Arabian Peninsula, while Eu 9 came from the northern part of the Fertile Crescent. In addition to the CMH, Nebel et al (2002) found other modals that represented some samples in their study. The Muslim Kurds MH belonged to Eu 9, while CHM, Bedouins MH, Palestinians MH, and Galilee MH all clustered in Eu10. Table 5 identifies the motifs of each Modal Haplotype.

Nebel et al (2001) also identified that Eu9 and Eu10 were characterized by a very high frequency of allele DYS388 with ≥15 repeats, and those alleles with high numbers of repeats were found to be restricted to Eu9 and Eu10. When combining the Eu10 chromosome, with the modal haplotypes observed in Palestinians and Bedouins, Nebel (2001) found that 29% of the Palestinian sample and 37.5% of the Bedouin sample were found in this Arab specific chromosome haplotype.

The genetic makeup of the Y-chromosome in the Jewish populations was found to be closer to the Middle Eastern population (Palestinians, and Bedouins, Muslim Kurds, and Turks), and more than the host populations. It is worth noting that this paper was published in 2001, and the names of Eu9 and Eu10 have been changed to J-M172 and J-M267 (Semino, 2000, YCC 2002, Semino, 2004).

Semino(2004) investigated the Y chromosome haplogroup J and haplogroup E in the Mediterranean area. The most prevalent haplogroup in the Middle East is Hg J, and it is defined by the deletion mutation p12f2 (Casanova et al 1985). Subgroups J-M172 and J-M267 show the most variation, which indicates that these groups arose in the Middle East, but with different percentages in the populations. The majority of the J-M267 haplogroup have the STR marker motif YCAIIa 22-YCAIIb-22 in the Middle East (>70%) and North Africa (>90%). This is a motif that combines the Y-SNP data with the STR data.

According to Semino's interpretations (2004), the first migration in Neolithic times brought J-M267 to Europe and Ethiopia, and a recent migration brought the J-M267-YCAII-a22-YCAIIb22 motif. This clade also includes the STR modal haplotype DYS388-17/DYS390-23/DYS391-11/DYS392-11). Nebel (2001) and Bosch (2001) argue that this motif was diffused by the Arabs who expanded in the 7$^{th}$ century AD to Egypt and North Africa. Semino (2004) also studied Hg-E, which is presumed to have African origins. Some of the sub-branches such as E-M35 and E-M78 appear in different frequencies in the Middle East, but the highest ones are in North Africa.

Using binary markers, Luis et al (2004) suggest regional continuity between the Nile valley, the Middle East, and the Arabian Pininsula. The Afro-Asiatic populations of Egypt and Oman exibit the highest number of groups and haplogroups. There is a considerable decline in diversity from east to west along the central African corridore to Benin. J is the most common group of the Omani collection (47.9%), followed by E(23.1%) and R(10.7%). In Egypt the order is slightly different groups: E(39.5%), J(32%), G(8.8%). The J-12f2 is most prevalent in Egyptian (32%), and Omani (47.9%) which may indicate intra population variation of the Egyptian and Omani populations reflecting their geographical locations near strategic cross roads between Africa and Eurasia. The genetic makeup of these two populations reflects a Middle Eastern type more than a Sub-Saharan one.

Nearly all Sub-saharan samples that were used in Luis et al (2004) belong to groups A,B,and E, (92.2%). 9% of the Egyptian samples were of Sub-Saharan origin, while 10% of the Omani samples had that lineage. Eurasian haplogroups (C,D,F,Q) were found in 9% of the Egyptian samples and in 77% of the Omani samples.

The Y- haplotype diversity in Oman and Egypt are very distinctive from Sub-Saharan diversity and markers representing the sub-Saharan haplotypes are present in low frequency, which may indicate a recent gene flow, maybe through the slave trade. Markers of the neolithic expansion (12f2,M201,M35) are the predominant markers observed.

Quintana-Murci (2003) studied the variablity of six Y- chromosome markers in populations from the Mediterranean area. The samples came from Turkish, Lebanese, Jewish, Algerian, Tunisian, and other populations. The STR marker DYS392 with 11 repeats had its maximum frequency in the Middle Eastern samples; this allele was associated with J- Hg.This study used samples from the eastern and northern shores of the Mediterranean.Some genetic homogeneity regarding the Y markers was found which may suggest a recent common origin.

Manni et al (2002) investigated 10 binary markers in Egyptians, Moroccan Arabs and Moroccan Berbers in order to assess the diversity of these populations and to study the influence of neighboring groups on the genetic makeup of Egyptian and Moroccan populations. The study showed two major haplogroups expressed in the samples studied, Hg9, and Hg21. 34.9% of the Egyptian Arab sample belonged to Hg9, while 44.4% belonged to Hg21. Moroccan Arab sample had frequencies of 13.7% of Hg9 and 68.7% of Hg21.the Morrocan Berber samples had 4.8% of Hg9 and 76.0% of Hg21. Manni et al (2002) proposed that Hg21 may have originated in North Africa before spreading to Europe. In Egypt the reduced frequency of Hg21 could have happened by a population movement from the Levant that introduced Hg9. It should also be noted that Hg9 is now known as HgJ, and Hg21 is a branch of HgE.

Table 1 is a summary of the frequency of HgJ and HgE in different populations as reported in the literature:

**Table1. The frequency of HgJ and HgE in different populations of the world.**

| Population | N | % HgJ | % HgE | Reference |
|---|---|---|---|---|
| **Palestinians** | **73** | **51** | **-** | **Hammer 2000** |
| **Lebanese** | **24** | **46** | **-** | **Hammer 2000** |
| **Syrian** | **91** | **57** | **-** | **Hammer 2000** |
| **Sub Saharan** | **199** | **1** | **-** | **Hammer 2000** |
| **Mulim Kurds** | **95** | **40** | **-** | **Nebel 2001** |
| **Palestinians** | **134** | **55.2** | **-** | **Nebel 2001** |
| **Bedouins** | **32** | **65.5** | **-** | **Nebel 2001** |
| **Jewish Kurds** | **99** | **37.4** | **-** | **Nebel 2001** |
| **Sepharic Jews** | **78** | **28.2** | **-** | **Nebel 2001** |
| **Ashkenazi Jews** | **79** | **43** | **-** | **Nebel 2001** |
| **Turkey** | **167** | **33** | **-** | **Rosser 2000** |
| **Algeria** | **27** | **41** | **-** | **Rosser 2000** |
| **Egypt** | **63** | **34.9** | **4.8** | **Manni 2002** |
| **Moroccan Arabs** | **51** | **13.7** | **68.7** | **Manni 2002** |
| **Moroccan Berbers** | **50** | **4.8** | **76** | **Manni 2002** |
| **Iraqis** | **139** | **58.3** | **12.2** | **Al Zahery 2003** |
| **Omani** | **121** | **39.7** | **19** | **Luis 2004** |
| **Egypt** | **147** | **27.9** | **26.5** | **Luis 2004** |
| **Benin** | **100** | **0** | **95** | **Luis 2004** |

**Mitochondrial DNA**

Mitochondrial DNA is the maternal counterpart of the Y-chromosome. It is maternally inherited and lacks recombination which makes it an ideal tool to study populations through maternal lines. The haploid mitochondrial genome accumulates mutations faster than the nuclear genome; the most variable region is the 1,122 bp non-coding region between bp 16024 and bp 0057, which is called the control region (Richards 1996). The control region of the mitochondrial genome is highly variable, and it is used to determine the genetic structure and origin of populations (Parsons, 1997). The high rate of base substitution in the control region of the mitochondrial genome, and the fact that the effective population size of this region is ¼ of the nuclear genome which means increased genetic drift, allows maternal genealogy to be constructed with high specificity (Macaulay ,1999, Richards, 1996).

MtDNA can't be considered a unique identifier since unrelated individuals could share an unidentified maternal relative at some point in the past. One of the disadvantages of using mtDNA in population studies is that relationships of modern mtDNA sequences could be obscured by recurrent mutation, which affects the same nucleotide position and make the relationship more difficult to resolve (Macaulay, 1999).

The mt DNA control region is divided into 2 regions: HVR-I (hyper variable region-I) and HVR-II. It is HVR-I which is studied extensively. The numbering of the sequence data is done according to Anderson et al (1981). The haplogroup characteristics are determined according to the HVR-I sequence motif with additional testing with restriction enzymes. Restriction enzymes work on certain points in the HVR-I it and cut at specific sites, producing fragments that can be detected by gel electrophoresis. This method of using sequence fragments is called RFLP (Restriction fragment length polymorphism) (Macaulay, 1999, Torroni, 1999). Combining sequence data with RFLP analysis ensures the accurate reading of the mitochondrial haplogroups.

Macaulay et al. (1999) described the nomenclature of mtDNA clades or haplogroups, and major haplogroups are given roman letters (H, V, K, L, and B). These clades represent an early branch on the mtDNA evolutionary tree or a special geographic distribution (Richards 2000). Subclades branch from the major haplogroups, and are given additional small letters and numbers (L2, L1, V1a…).

Richards (2000) studied the mtDNA haplotypes in populations from Europe and the Near East. The paper's aim was to find founder lineages in the Near East that could have contributed to the gene pool of Europeans. This study included samples from Egypt, Yemen, Iraq, Syria, Jordan, Israeli Palestinians, Turks, Israeli Druze, and Jewish Israeli. This study revealed much about the maternal genetic makeup of the Near Eastern populations. 25% of the near Eastern population exhibited Cluster H, compared to 46% that is present in the European population, but this cluster was not present in Arabia. Other haplogroups had their highest frequency in Arabia also (Bedouins and Yemeni). Cluster J's frequency was 25%, and cluster HV's frequency was 22%. Other haplogroups that represent other populations appear in smaller frequencies in the Near Eastern populations. L1-L3A, which is of African origin appear in some populations, while haplogroup U5, which is present in Europe, is absent from Arabia, and appears sporadically in Kurds, Egyptians, and Armenians.

Al- Zahery (2003) studied the Iraqi population in particular, but used other populations from the Middle East to provide a comparison with the results. They used RFLP markers of the common Asian, African, and European haplogroups for their analysis. 78% of the Iraqi samples fell in this group of haplogroups (HV,H,V,J,T,K,U,I,X,W), which represent west Eurasian types. 90% of the European samples were encompassed by this group, while 60.4% of the Arabian samples included the mentioned haplogroups.

20% of the Iraqi samples have haplogroup H, which is lower than the European frequency (30-50%). Haplogroup V, which is a European one (Torroni et al, 2001), was in 0.5% of the Iraqi samples, while the African Haplogroup L1-L3 was in 4.2% of the samples. The one haplogroup that differentiated

the west Eurasian populations was Hg pre-HV. 15.2% of Arabian samples, 4.2% of Iraqi samples, 5.8% of Syrian samples and 2.6% of Palestinian samples carried this haplogroup, compared to 0.5% in Indian samples.

Richards (2003) investigated the female gene flow from sub Saharan Africa to Near Eastern populations, and the findings in this paper confirmed what other researchers found in their studies, that about 85% of near eastern populations fall in the N haplogroup which is a characteristic of West Eurasian mtDNA. This N group contains the following haplogroups (H, pre-HV, U1-U7, K, J, I, W, X,P).

Hg U6 is an African haplogroup. It originated in Eurasia, and then diverged in North Africa, which provides a descriptive characteristic of those samples from North Africa. This study confirmed also that pre-HV haplogroups and indeed a Near Eastern type, since its highest frequency was observed in samples from Arabia. Furthermore, the African haplogroup L1-L3 had a frequency of 10-15% in Arab samples, but had a larger frequency (35%) in samples from Yemen Hadramawt. This haplogroup was absent from the Jewish, Armenian, and Turkish samples.

In general, the research data indicates that there are specific haplogroups that define populations in the Middle East and North Africa, and the variation in the frequencies observed can be explained by the historical record, the population's movement and dynamics. Tables 2, 3, 4, summarize the frequencies of mtDNA haplogroups in different populations as reported in the literature.

**Table 2 The mtDNA haplogroup frequencies in Asian Middle Eastern populations.**

| Haplogroup | Estimated mtDNA haplogroups frequency in populations | | | | | |
|---|---|---|---|---|---|---|
| | Iraqi[a] | Arabia[b] | Palestinian[c] | Syrian[c] | Iranian[b] | Iranian[1,d] |
| **Sample size** | **216** | **389** | **117** | **69** | **451** | **42** |
| **A-G,M,N9*** | **-** | **7.7** | **-** | **-** | **6.2** | **4.8** |
| **Pre-HV** | **4.2** | **15.2** | **2.6** | **5.8** | **2.4** | **0** |
| **HV** | **10.6** | **3.6** | **1.7** | **4.3** | **5.5** | **16.7** |
| **H** | **19.9** | **12.9** | **30.8** | **24.6** | **17.1** | **14.3** |
| **V** | **0.5** | **0** | **0** | **2.9** | **0** | **0** |
| **J** | **9.3** | **20.8** | **9.4** | **10.1** | **13.5** | **16.7** |
| **T** | **8.8** | **4.6** | **12.8** | **10.1** | **8.4** | **7.1** |
| **K** | **3.2** | **3.6** | **6.8** | **4.3** | **7.5** | **7.1** |
| **U** | **19.0** | **10.5** | **7.6** | **15.9** | **21.5** | **0** |
| **I** | **1.9** | **0.8** | **0** | **0** | **2.0** | **2.4** |
| **X** | **2.8** | **1.8** | **3.4** | **0** | **2.9** | **0** |
| **W** | **1.9** | **1.8** | **2.6** | **2.9** | **2.0** | **2.4** |
| **B** | **0.9** | **nr** | **0** | **0** | **nr** | **0** |
| **M** | **1.4** | **nr** | **1.7** | **1.4** | **nr** | **4.8** |
| **L1** | **1.4** | **nr** | **0.9** | **2.9** | **nr** | **16.7** |
| **L2** | **2.8** | **nr** | **4.3** | **2.9** | **nr** | **0** |
| **L1-L3[#]** | | **10.5** | | | **2.2** | **0** |

**Table 3. The mtDNA haplogroup frequencies in several populations in Africa, Europe and Asia.**

| Haplogroup | Estimated mtDNA haplogroups frequency in population | | | | | |
|---|---|---|---|---|---|---|
| | Yemen[2,e] | Anatolia[b,f] | India[b] | Ethiopia[g] | German[h] | Mozambique[i,j] |
| **Sample size** | **56** | **388** | **1300** | **74** | **200** | **416** |
| **A-G,M,N9*** | | **0.8** | **66.3** | | | **0** |
| **Pre-HV** | **7** | **2.8** | **0.5** | | **0** | **0** |
| **HV** | | **3.6** | **0.6** | | **nr** | **0** |
| **H** | | **25.0** | **2.4** | | **50.0** | **0** |
| **V** | | **0** | **0** | | **2.5** | **0** |
| **J** | | **10.9** | **0.8** | | **7.5** | **0** |
| **T** | | **11.9** | **1.1** | | **8.5** | **0** |
| **K** | | **5.9** | **0.2** | | **6.5** | **0** |
| **U** | | **19.3** | **12.0** | | **13.5** | **0** |
| **I** | | **2.3** | **0.6** | | **2.5** | **0** |
| **X** | | **4.4** | **0.2** | | **0.5** | **0** |
| **W** | | **3.9** | **1.5** | | **1.0** | **0** |
| **B** | | **0** | **nr** | | **0** | **0** |
| **M** | | **4.4** | **nr** | | **0** | **0** |
| **L1** | | **nr** | | | **0** | **0** |
| **L2** | | **nr** | | | **0** | **0** |
| **L1-L3[#]** | **34** | **0.3** | | **55** | | **100** |

**Table 4. The mtDNA haplogroup frequencies in Jewish populations.**

| Haplogroup | Estimated mtDNA haplogroups frequency in population % | | | |
|---|---|---|---|---|
| | Iraqi [g] | Iranian[g] | Ashkenazi[g] | Ethiopian [g] |
| **Sample size** | **56** | **75** | **78** | **46** |
| **Pre-HV1** | **0** | **1** | **3** | **15** |
| **U6** | **0** | **0** | **3** | **0** |
| **L1-L3[#]** | **0** | **0** | **0** | **52** |

nr: Not Reported

*****Asian haplogroup, [#]African haplogroup, [1]southern and southern central Iran, [2]Hadramawt in Yemen. [a]Al-Zahery et al. (2003).[b]Kivisild et al.(2003).[c]Richards et al.(2000).[d]Quintana-Murci et al.(2004)[e]Di Rienzo and Wilson (1991)[f]Tambet et al. (2000).[g-]Thomas et al. (2002).[h]Kivisild et al. (1999).[i]Pereira et al. (2001). [j]Salas et al. (2002).

**Autosomal markers**

Because the Y-chromosome and mtDNA express different population and demographic structures, we may expect some trace of these differences in the autosomal DNA record (Hammer.1998). The use of autosomal markers is important because it gives a more general inference about the demographic history and the population's relationship than does the study of Y-chromosome and mt-DNA loci, as their history may be anomalous (Zhivotovsky.2003). The potential of using autosomal markers is great, they are very abundant, and they provide more loci for population genetic studies.

Most informative genetic studies have used a combination of the different genetic systems, since using the three systems provides a wide range of polymorphisms to study, and those same polymorphisms will be compared in the intended populations (Jorde.2000). Most autosomal markers used in population

studies are in non –recombining regions of the autosomal chromosomes, and they include short tandem repeats, restriction site polymorphisms, or autosomal Alu polymorphisms.

Studies investigating Middle Eastern populations using autosomal markers are rare because the use of these markers in population genetics studies is recent. Some studies like Rosenberg et al. (2001) examined Libyan Jews and compared autosomal markers with 6 other Jewish populations and two non-Jewish ones, but they did not come with any specific markers that identified Middle Eastern populations. They studied the genetic distance between the samples and correlated it with the history of these populations.

**The Jewish populations**

The Jewish populations trace their origins to tribes that lived in the Middle East 4,000 years ago in the Levant (Fig.1), where they built the ancient kingdoms of Israel. The Diaspora of the Jews led to the formation of communities in Europe, North Africa, and the Middle East. Modern Jews are divided into two main subdivisions: the Ashkenazi Jews from northeastern Europe, and the Sephardic Jews in North Africa and the Mediterranean Basin. Because of the complexity of the history of the Jews, it is expected to see similar complexity in the gene pool. One of the interesting questions to researchers was the genetic relationship between the Jewish communities and their host populations.

Thomas et al. (1998) found a motif of six STR markers that was called the Cohen Modal Haplotype, and was observed in high percentages in priests of Jewish groups that were separated geographically. The motif was found in high frequency in individuals with the last surname “Cohen”, which indicates membership in ancient Jewish priesthood groups. Those groups were highly regarded as priests and thus maintained a tradition of not marrying outside the Cohen priest class in order to preserve the purity of the bloodlines.

Hammer et al. (2000) studied Y biallelic markers in Jewish populations; they compared the variability of these markers in Jewish populations with non-Jewish ones in Africa, Europe, and the

Middle East. The data generated by Hammer et al. (2000) supported a common ancestor for Jewish samples that existed sometime in the Middle East. The results indicated a great affinity of the Jewish Y gene pool with the Middle Eastern one, especially Palestinian and Syrian Y markers.

Thomas et al. (2002) compared the maternal and paternal Jewish types with non -Jewish ones. The mtDNA analysis revealed that Jewish communities have lower diversity than non- Jewish mtDNA types, which indicated female-specific founding events in the Jewish populations. This means that there are fewer female ancestors of the Jewish communities accompanied by little genetic contribution of the host female gene pool. The Y- chromosome markers showed diversity similar to those of the Middle Eastern Samples, such as the Palestinians and Syrians.

Nebel et al. (2001) analyzed 13 binary markers and six STR polymorphisms in six Middle Eastern populations (Ashkenazi, Sephardic, and Kurdish Jews; Muslim Kurds, Muslim Arabs, and Bedouins). The study revealed that Sephardic and Kurdish Jews are closer genetically, and they differed slightly from Ashkenazi Jews. The analysis of Y haplogroups found that Jews were closely related to the populations of the North Fertile crescent.

Although the Jewish populations faced exile and Diaspora, they still maintained their genetic ties to the Middle East, which suggests that these populations did not mix genetically with the host populations.

**Summary and conclusions**:

Middle Eastern populations demonstrate characteristics that distinguish them from other populations, and using a combination of systems to identify the population gives us a more accurate and reliable picture.

Studies using the Y- chromosome have found that Middle Eastern populations belong to genetic haplogroups that are common in Eurasia. Certain haplogroups exhibit high frequencies in Middle Eastern

populations and could be used to identify those populations. HgJ was found to be the most common haplogroup in Arabia and the Levant and this haplogroup shows a geographical gradient that goes from East to West, which may indicate a Neolithic expansion from the area to North Africa and Europe. Two branches, J-M267 and J-M172, are especially interesting because they may point to migration events that introduced additional variability to the gene pool. HgE, which is believed to have an African origin, is expressed in the Middle Eastern populations in different frequencies. The geographical gradient of this haplogroup indicates its spread from West Africa to the Middle East and Europe,

Short tandem repeats are useful in estimating the times of appearance for specific traits. Several STR motifs are observed in association with the biallelic markers. The motif YCIIa/22-YC11B/22 was associated with Hg J-M267, while DYS388 was associated with repeat number $\geq 15$. DYS392 with 11 repeats was associated with the J-group in general, and had its highest frequencies in the Middle East.

Modal haplotypes provide another way of looking at the characteristics of the Middle Eastern populations, and certain combinations of STR markers with specific repeat numbers are found in high frequencies in Middle Eastern populations. The Cohen Modal Haplotype was found to be a mark of Jewish ancient priests (Thomas 1998).

Mitochondrial haplotypes reveal a greater variability in Middle Eastern populations than the Y-chromosome, which can be explained by the patrilocality of the populations. Female gene flow from Europe, East Africa, and Asia is demonstrated by the mitochondrial haplotype frequencies observed in the area. Still, the majority of mtDNA seen in the Middle East belongs to the West Eurasian type which includes clades H, preHV1, HV1,U1,U7,K,J,T,I,W,X,N. Haplogroup (preHV) 1 has its highest frequency in Arabia, while Hg U6 is common in North Africa. Hg L1-L3A, which is an African haplogroup, appears in high frequencies in Yemen and Arabia.

The geographical location of the Middle East contributed to the variation of its gene pool, with people migrating to or from and passing through the area, leaving their genetic marks. The Middle East

gene pool contributed also to populations in Africa and Europe. The theory of Neolithic diffusion that introduced farmers from the Middle East to Europe and Africa has been debated greatly, but the genetic record of the populations agrees with it.

The spread of Islam in the 7th century A.D by the tribes of the Arabian Peninsula added to the Y haplotype genetic pool of the North African populations. Haplotypes that are of Eurasian origins are now found in North Africa particularly because of the social customs of Muslim males of marrying into the new localities they settled in. The social practices of the populations left their mark also on the Y haplotypes, in communities like the Arabs and Bedouins, patrilocality and consanguineous marriages lead to a decrease in haplogroup numbers and diversity and added to more diverse female haplotypes in the gene pool.

The slave trade from East Africa to southern Arabia introduced maternal haplotypes to the population. Coupled with the practice of assimilating the offspring of slaves with the community as a result of miscegenation and manumissions, the high frequency of African mtDNA haplotypes becomes easier to understand. The relatively very low proportion of African Y haplotypes is explained by the fact that African men were brought to Arabia as workers, or castrated to be eunuchs, so they left no offspring that can add to the Y- haplotypes (Lewis, 1990).

The Middle East exhibits a linguistic diversity, and the languages spoken in the area differ in their origins and the linguistic family they belong to. Arabic and Hebrew belong to the Semitic family, while Kurdish and Armenian belong to the Indo-European family, and Turkish belongs to Altaic family. The Y gene pool of the Middle East is uniform, and the research suggests an ancient ancestry, all this may indicate that the genetic markers were defined in the area before the appearance of cultural and linguistic divides.

The Middle East has been an area of interest for genealogical studies due to its historical importance in terms of religious, linguistic, and social factors. Different approaches have been used to

study the Middle Eastern populations; the goal was always to learn more about an intriguing chapter of human history. The first record was oral history, and that of course lacks concrete verification of the events. Anthropological studies and blood types examinations contribute to our understanding of the population, but they are not specific enough, or prevalent enough.

Since the emergence of DNA markers as a way of examination the history of human populations, the Middle East has been a focal point to researchers. These markers represent unique evolutionary events, and they represent specific hereditary pathways that can give a clearer picture of the populations' dynamics.

The purpose of this study is to combine two specific DNA markers, a DYS458 microvariant and CMH motifs, and use them to investigate the relationship between two sets of samples. One set is from the south of Arabia, and the second is from the Northern parts (the Levant).

## Materials and methods

### Samples

The samples that were used in this study come from two primary sources, the first with 320 samples from Oman. Most of these samples were collected in southern Oman, in an area that was connected historically to Yemen. Social and trade connections contributed to the ties between southern Oman and Yemen. We collected 4-5 hairs from each participant, and they gave a 4-generation genealogy that contained information about the place of origin and the tribal affiliation of both the father and mother of the participant. The second group of samples comes from the West Bank and Gaza Strip, which represents the northern populations, and represents the other end of our investigation. The same collection procedures were conducted for both sets of samples. The Samples were collected on location in Oman and Palestine.

The samples collected from Oman were divided geographically into Northern samples (from the

capital Musqat) and Southern samples (from the city of Salalah in the south). The Southern samples were further divided by the researcher into a mountain people sample of those participants whose ancestors lived in the mountains, and the Arabi samples were from participants whose ancestors lived in the plains and considered themselves of pure Arabic origin. The Salalah division was done by simply using the tribe names provided by the participants and comparing them to Arabic genealogy references and social norms of the area.

The West Bank sample was collected by the researcher from college students and acquaintances. Most of the collection was done in the city of Ramallah, but we believe that the sample is very valid in the sense that Ramallah represents a melting pot for people from all around the West Bank, due to its economical, cultural, and social dynamics. The samples represent large part of the West Bank as illustrated by the genealogy charts provided by the participants.

**Methods**

**Transportation and storage:** The samples are hair samples, which makes storage and transportation more feasible. The hairs collected from the individuals were stored in plastic bags that were labeled in the field; we always labeled the sample in the presence of the participant, and checked the participant's genealogy charts before storage. The plastic bags containing the samples were kept and transported in dry containers.

**DNA extraction:** DNA was released from the cell using 500ul sterile 5% chelex®, the hair root was immersed in the chelex® and incubated at 56°C for 10 minutes, vortexed vigorously, incubated at 94°C for 10 minutes and centrifuged at 9000 rpm for 1 minute. The supernatant was then aliquoted into plates and dehydrated for storage purposes.

**DNA amplification:** In this study, we were looking at different locations on the genome; there were various PCR reactions were performed according to the marker in question. The Y-chromosome was amplified using multiplex PCR reactions in which combinations of fluorescent dyes attached to multiple

primers were used.

**Post amplification treatment**

**DNA sequencing**: The aim is to determine the base order in the DNA samples; we used fluorescence labeled dyes that were attached to ACGT extension products in DNA sequencing reactions. Dye labels were incorporated using either 5'-dye labeled primers, or 3'dye labeled dideoxynucleotide terminator (ABI PRISM® Genetic analyzer). This type of sequencing was used in analyzing the Y-chromosome markers.

**Genotyping:** 24 loci were chosen out of 34 loci to be genotyped and facts like the mutation rate, the number of alleles per loci, and the stability of the loci were factors in determining which loci to use in the genetic analysis. 8318 samples were analyzed for the DYS458 microvariant and the 458 was binned with single base wide bins. To investigate the nature of the 18.2 microvariant, sequencing of a selected subset of samples was performed; these samples were selected based on the geographical origin, the 458 allele sizes, and the level of concordance with Cohen model haplotypes (CHM).

**Network analysis:** The results were analyzed by constructing a phylogenetic tree using the software Network (www.fluxus-engineering.com); of the 24 loci typed, 19 were used to construct the network, DYS385 and DYS459 were eliminated because they exhibit duplicate loci with ≥2 alleles for an individual. DYS389, DYS389B, DYS390 were eliminated because they have a complex repeat structure and they both have 4 separately mutating repeat units inside the primer regions. Additional assays are required to separate these. We chose to eliminate these markers from the phylogenetic analysis in accordance with the literature, although when a test run was performed with DYS389/390, results were virtually indistinguishable.

Network calculations were performed with the MJ method (Bandelt et al, 1999), and the network was also post processed with MP calculations to reduce reticulations. Diversity indices like the average

gene diversity and Rst genetic differences were determined using Arlequin v.2.00 (Schneider et al, 2000), while haplotype diversity was calculated for geographies in which 458 microvariant was observed.

**Results and discussion:**

Cohen Modal Haplotypes (Thomas et al, 1998) combines six Y-STR loci with certain repeat numbers. The CMH allele profile is: DYS19/14, DYS388/16, DYS390/23, DYS391/11, DYS392/12, and DYS393/10. Table 4 shows the different haplotype motifs as reported in the literature. The following table shows five Modal Haplotypes that Nebel (2001) reported, and used to characterize Middle Eastern samples. The six microsatellites markers for each corresponding repeat appear in the Table, with the number of repeats shown.

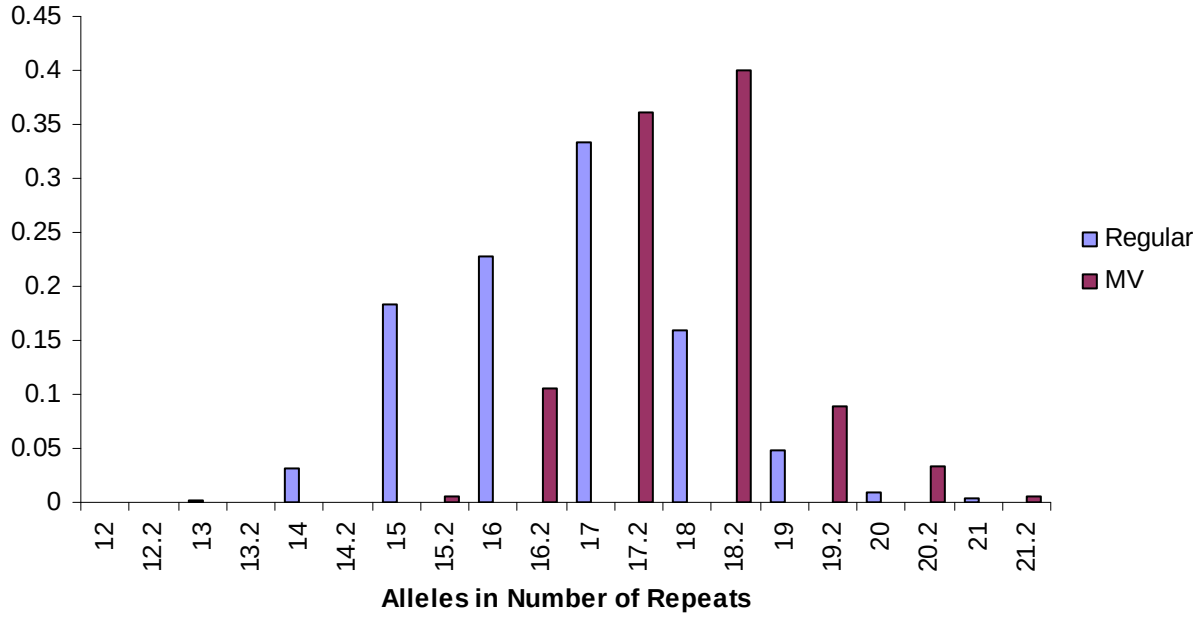
**Table 5. Middle Eastern Modal haplotypes using the six STR markers**

| | **DYS19** | **DYS388** | **DYS390** | **DYS391** | **DYS392** | **DYS393** |
|---|---|---|---|---|---|---|
| **CohenHM** | **14** | **16** | **23** | **10** | **11** | **12** |
| **GalileeMH** | **14** | **17** | **23** | **11** | **11** | **12** |
| **Bedouins MH** | **14** | **15** | **23** | **10** | **11** | **13** |
| **Muslim Kurds MH** | **14** | **15** | **23** | **10** | **11** | **12** |
| **Palestinian MH** | **14** | **17** | **22** | **11** | **11** | **12** |

**DYS458 microvariant:**

While examining the results of DYS458 haplotying for Middle Eastern samples, a two base pair shift in the marker produced a distinguishing set of results, and high numbers of the samples exhibited the alternative allele. The shift was in the 18 repeat bins, and after referring to the literature, the microvariant was identified as 18.2 DYS458 microvariant. Figure 2 shows the distribution of the microvariant throughout the bigger pool of samples (not just our Middle Eastern ones). Microvariant 18.2 is a significant alternative allele and was observed in high frequency in different populations.

**Fig 2: Frequency distribution of observed alleles in DYS458**



The following table (Table 6) has frequencies of the DYS458 microvariant in several populations, among them the Middle Eastern ones. We also combined the Cohen Modal Haplotype frequencies, and the last column represents the frequency of the samples that had both markers.

**Table 6: Frequency of the 18.2 DYS458 microvariants in populations around the world combined with the CMH haplotype frequencies**.

| Population | Total Sample number | Total number of DYS458 18.2MV | Frequency of DYS458 18.2 MV | Frequency of CMH in DYS458 MV Individuals |
|---|---|---|---|---|
| Middle East | 206 | 91 | 0.442 | 0.330 |
| *Oman* | 102 | 63 | 0.618 | 0.397 |
| *Gaza* | 41 | 11 | 0.268 | 0.182 |
| *West Bank* | 47 | 16 | 0.340 | 0.188 |
| South America | 985 | 28 | 0.028 | 0.679 |
| Europe | 2722 | 34 | 0.012 | 0.353 |
| North America | 4093 | 26 | 0.006 | 0.846 |
| Africa | 13 | 1 | *nr* | *Nr* |
| Asia | 114 | 0 | 0 | 0 |
| Oceania | 185 | 0 | 0 | 0 |

We selected all individuals with the DYS458 18.2MV and the CHM haplotypes. We performed a hierarchical scan for all major haplogroups on a subset of 12 samples, and found that 10 out of the 12 belong to haplogroup J. This haplogroup is of interest because of the high frequency in the tested group, and because the geographic coverage of the samples is similar to that observed in the J haplogroup. Semino et al (2004) elaborated that the J haplogroup not only shows its highest frequency in the Middle

East, but it originated there also. To confirm the membership of the samples to Haplogroup J, all samples were tested for the mutation 12f2 to confirm membership in the J group. Next, testing was done to confirm the M172 branch, which is a highly specific Middle Eastern branch, this binary marker testing helped us verify that this microvariant indeed originated in the Middle East.

We noticed that the 18.2 MV is most frequent in the Middle East, especially in Omani samples, while it is less frequent in Palestinian and Gaza samples. This MV is observed in Europe and the Americas at lower frequencies. The highest gene diversity of the MV was seen in Europe, which may indicate multiple distinct branches of ancient 458 mutations, or migration of distinct groups, rather than being an origin of the mutation. On the other hand the gene diversity in the Middle East is lower, which can be explained by population bottlenecks/compression, or it could be due to tribal homogeneity, which means that the mutation tapped into a few of the distinct branches of the ancient 458 mutation.

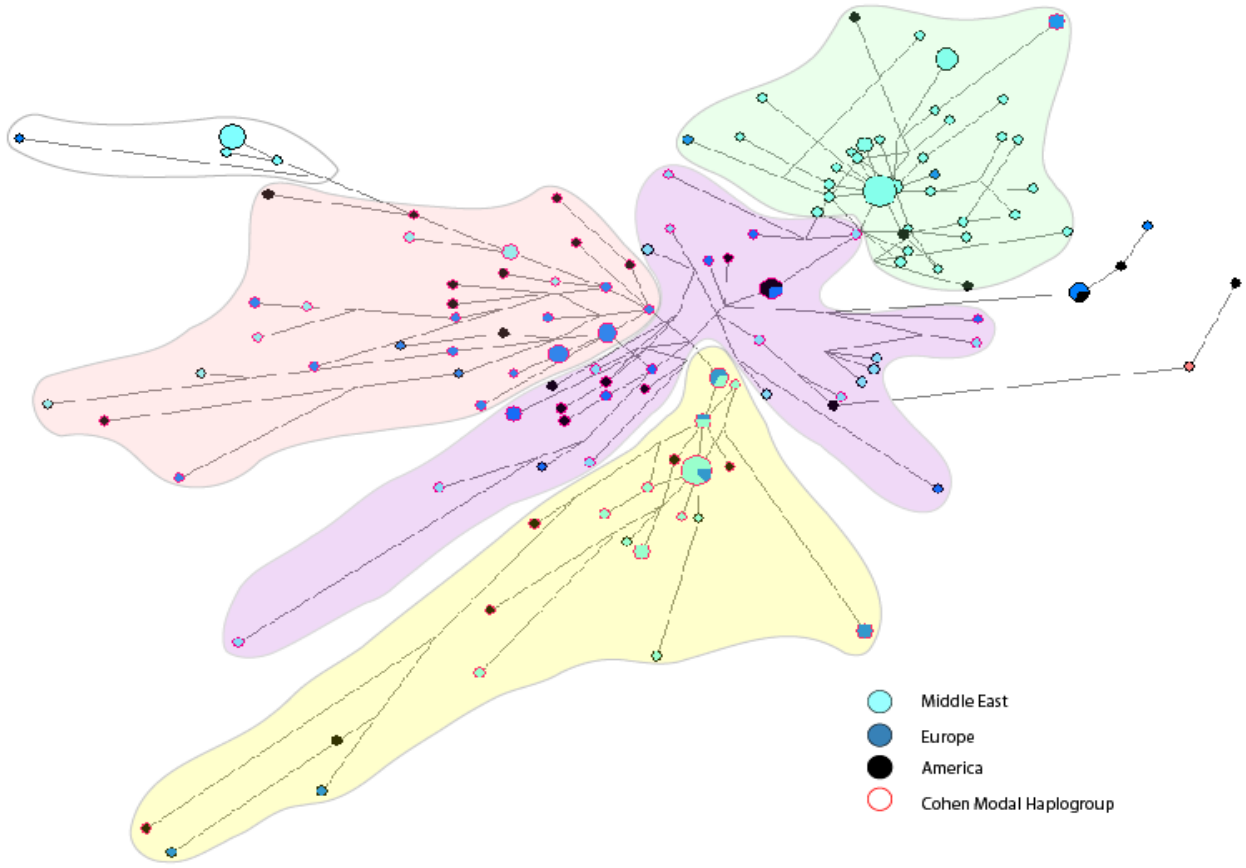
Three different mutations were observed with the DYS458 MV. The first one is an insertion in the 2$^{nd}$ repeat on the 3' side, and the second one is an insertion in the 5$^{th}$ repeat in the 3' side, while this mutation could be part of the same clade as the 3'-2. The third mutation is an insertion on the 3$^{rd}$ repeat on the 5' side (5'-3). The different mutations in the DYS458 MV appear in different populations. The 5'-3 mutation was observed in the Middle Eastern samples, and this mutation was also observed in the samples with the strongest concordance to CMH, so this mutational event is of great interest to our scope of research.

The data of the individuals used in this study was categorized according to phylogenetic groupings characterized by quantitative means, such as average gene diversity, Fst/Rst measures, AMOVA, exact test for population differentiation, and pairwise mismatch distribution. We used average gene diversity as a measure because haplotype diversity was approximately =1. Due to the highly discriminating nature of the 19-locus haplotype, virtually all the individuals were distinct even within groups. We chose not to reduce haplotypes to the standard 6 loci that will get us haplotype diversity, but

instead used average gene diversity to demonstrate the varying levels of haplotype diversity within the phylogenetic grouping. Fst/Rst measures and the AMOVA test will demonstrate the degree of substructure detected by phylogenetic groupings. In addition, the AMOVA test will enable us to compare phylogenetic groupings vs. groupings based only on geography. 22 loci were used in AMOVA test.

**Phylogenetic network analysis**

**Figure 3: Median joining tree of the DYS458 microvriant**




.

Figure 3 is a median joining tree of samples that had the 18.2 DYS458 microvariant. The Middle Eastern samples separated into two separate branches. The larger number of samples clustered in the branch with the light blue background, while a smaller group was in the branch with the yellow

background. We noticed that most of the samples that came from Omani participants who had tribal origins in the mountain area in Salalah were placed in the light yellow branch.

**Figure 4: Median joining tree of DYS458 microvariant and CMH traits combined.**

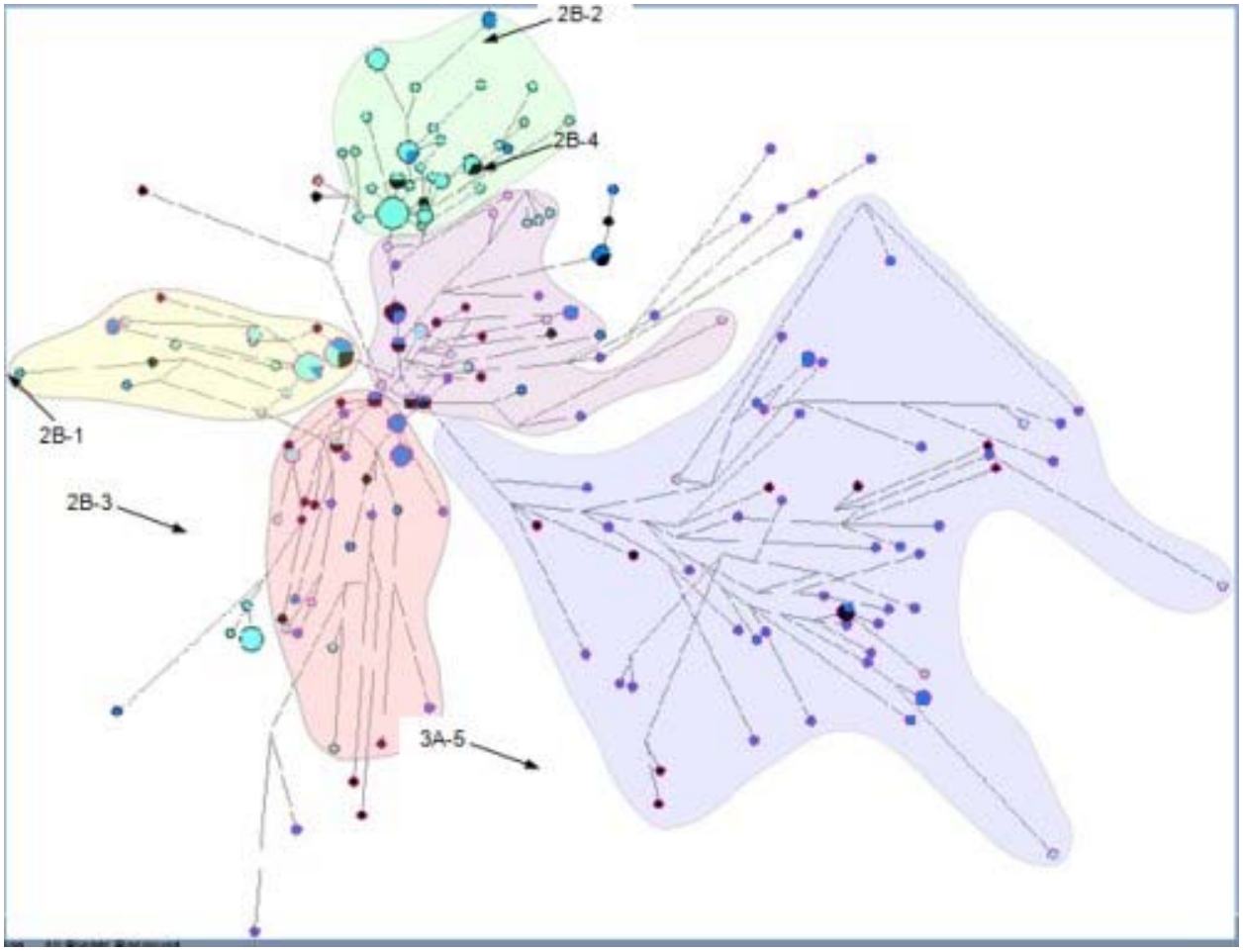


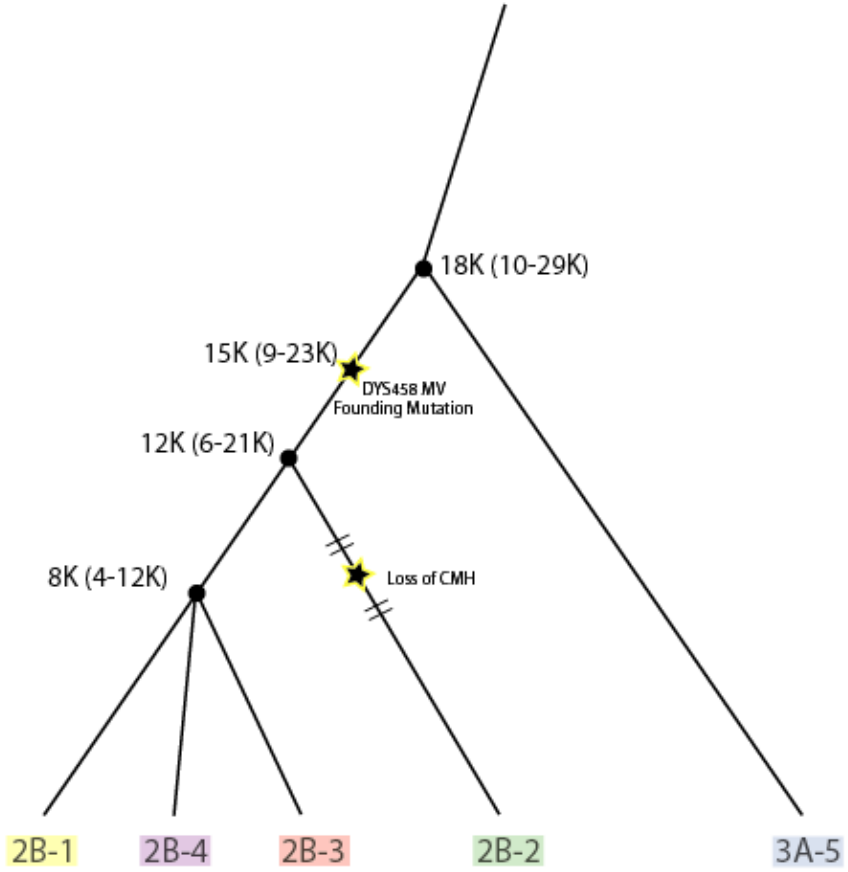

Figure .5 represents the median joining tree when both DYS458 microvariant and CMH were combined together. The samples which in the previous tree (Fig. 4) were divided into 2 branches, clustered together in this tree. The two main branches that had Middle Eastern Samples collapsed into one branch, when another criterion, in this case the CMH haplotype, was added to the tree building parameters. The added parameter helped distinguish the Middle Eastern samples more from the samples of other populations.

**Table 7: Quantitative characteristics of the branches from the median joining network trees**

| | | Proportion of Branch Membership | | | | | |
|---|---|---|---|---|---|---|---|
| Branch | Total Members | Middle East | Europe | America | Average No. Cohen Alleles | Average Gene Diversity | Average No. Pairwise Mismatches |
| 2B-1 | 28 | **0.643** | 0.214 | 0.143 | 5.179 | 0.241 (±0.132) | 5.788 (±2.854) |
| 2B-2 | 48 | **0.896** | 0.042 | 0.063 | 3.438 | 0.230 (±0.125) | 5.516 (±2.670) |
| 2B-3 | 31 | 0.194 | **0.548** | 0.258 | 5.226 | 0.266 (±0.144) | 6.396 (±3.114) |
| 2B-4 | 26 | 0.308 | 0.308 | 0.385 | 5.154 | 0.291 (±0.158) | 6.991 (±3.394) |
| 3A-5 | 59 | 0.085 | **0.695** | 0.220 | 5.203 | 0.445 (±0.228) | 10.67 (±4.928) |
| 3'-2 | 12 | 0.000 | **1.000** | 0.000 | 1.000 | 0.210 (±0.123) | 5.015 (±2.621) |

There is a strong geographic specifity to most of the branches, and there are certain branches that have clear concordance with CMH. 2B-2 has the DYS458 MV 5'-3 mutation, but does not belong to CMH. The diversity is much lower within branch definitions than with geography based classifications as demonstrated in Table 6, which demonstrates validity and usefulness of these groupings, and justifies using them for further analysis

**Figure 5: BATWING analysis to determine The Most Recent Common Ancestor (TMRCA) using Bayesian coalescent analysis (Wilson, 1998).**




Two methods were used to estimate the TMRCA (the most common ancient ancestor). The first method was the Bayesian coalescent analysis implanted in BATWING (Wilson, 1998). This method estimates the TMRCA for groups of individuals using a Marcov Chain Monte Carlo (MCMC) to generate approximate posterior distribution of the samples underlying the whole distribution of the genealogical tree. BATWING does not model the effects of recombination or selections, thus assuming these effects are negligible. Considering the nature of the Y chromosome short tandem repeats, we felt this method was appropriate to use in our analysis.

The DYS458 mutation appears to post-date development of CMH, while another branch of the lineage (2B-2) lost the CMH from its lineage, which may explain why some researchers thought the CMH was an exclusive Hebrew lineage, and a sampling error from only this branch may have led to such conclusions.

**Conclusion:**

The Middle Eastern samples used in this study showed unique characteristics such as a high frequency of a DYS458 microvariant 18.2. Compared with other samples from other populations, we concluded that this marker is useful as an identifying marker for certain Middle Eastern populations.

Furthermore, when our analysis was combined with the Cohen Modal Haplotype, it showed that the DYS458 microvariant post-dated the Cohen Modal Haplotype. Combining both sets of data helped construct a tree that revealed some descriptions of the samples. The high frequency of the Cohen Modal Haplotye in our samples indicates that it's not limited to Hebrew or Judaic populations; it represents more of a Semitic trait. Studying the trees produced, combined with ancestry analysis, we can detect bottle neck events in the population with the loss of CMH in some Middle Eastern branches on the tree.

The Presence of a DYS458 mutation in the Middle East gives us some insight into the history of the populations. There is evidence of Population divergence and constriction. The DYS458 microvariant appears to be a Middle Eastern trait through which gene flow appears in non Middle Eastern neighboring populations. One explanation could be the Jewish movement from the Middle East to Europe and the admixture of the populations during the crusades.

We also noticed a gradient distribution of the microvariant where the highest concentration appears in the southern samples (Oman). Although the current analysis is not sufficient to establish an immigration pattern, that could confirm an ancient event of population migration from the South to the North. The presence of such a gradient could be a first step for future studies that use such markers combined with other markers to investigate historic occurrences.

**References cited:**


Al-Zahery N, Semino O, Benuzzi G, Magri C, Passarino G, Torroni A and Santachiara-Benerecetti AS. 2003.Y-chromosome and mtDNA polymorphisms in Iraq, a crossroad of the early human dispersal and of post-Neolithic migrations. Mol Phylogenet Evol. 28:(3)458-472.

Arredi B, Poloni ES, Paracchini S, Zerjal T, Fathallah DM, Makrelouf M, Pascali VL, Novelletto A and Tyler-Smith C.2004.A predominantly neolithic origin for Y-chromosomal DNA variation in North Africa.Am J Hum Genet. 75: (2)338-345.

Behar DM, Thomas MG, Skorecki K, Hammer MF, Bulygina E, Rosengarten D, Jones AL, Held K, Moses V, Goldstein D, Bradman N and Weale ME.2003.Multiple origins of Ashkenazi Levites: Y chromosome evidence for both Near Eastern and European ancestries. Am J Hum Genet. 73: (4)768-779.

Bonné-Tamir B, Korostishevsky M, Redd AJ, Pel-Or Y, Kaplan ME and Hammer MF.2003.Maternal and paternal lineages of the Samaritan isolate: mutation rates and time to most recent common male ancestor. Ann Hum Genet. 67:(Pt 2)153-164.

Carvalho-Silva DR, Santos FR, Hutz MH, Salzano FM and Pena SD. 1999.Divergent human Y-chromosome microsatellite evolution rates. J Mol Evol 49: (2)204-214.

Casanova M, Leroy P, Boucekkine C, Weissenbach J, Bishop C, Fellous M, Purrello M, Fiori G and Siniscalco M.1985.A human Y-linked DNA polymorphism and its potential for estimating genetic and evolutionary distance.Science 230: (4732)1403-1406.

Cavalli-Sforza, LL ,Menozzi P, Piazza A.1994. The history and geography of human genes**.**Princeton University Press.Princeton, NJ.

Cavalli-Sforza LL and Minch E.1997.Paleolithic and Neolithic lineages in the European mitochondrial gene pool.Am J Hum Genet .61:(1)247-254.

Chistov YK.1996.Anthropometry of the South Yemen population: Between groups multivariate analysis. Homo. 47: (1-3)3-22.

Chistov YK.1994 .Human Cranial Remains from South Yemen. Homo. 45: (1)8-30.

Cinnioğlu C, King R, Kivisild T, Kalfoğlu E, Atasoy S, Cavalleri GL, Lillie AS, Roseman CC, Lin AA, Prince K, Oefner PJ, Shen P, Semino O, Cavalli-Sforza LL and Underhill PA .2004.Excavating Y-chromosome haplotype strata in Anatolia.Hum Genet 114:(2)127-148.

Cruciani F, Santolamazza P, Shen P, Macaulay V, Moral P, Olckers A, Modiano D, Holmes S, Destro-Bisol G, Coia V, Wallace DC, Oefner PJ, Torroni A, Cavalli-Sforza LL, Scozzari R and Underhill

PA.2002.A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes.Am J Hum Genet. 70:(5)1197-1214.

Di Rienzo A and Wilson AC.1991.Branching pattern in the evolutionary tree for human mitochondrial DNA.Proc Natl Acad Sci U S A .88:(5)1597-1601.

Gabrieli F, 1904-.1963.Arabs, a compact history. Translated by Salvator Attanasio. New York, Hawthorn Books

Hammer MF.1994.A recent insertion of an alu element on the Y chromosome is a useful marker for human population studies.Mol Biol Evol 11:(5)749-761.

Hammer MF, Redd AJ, Wood ET, Bonner MR, Jarjanazi H, Karafet T, Santachiara-Benerecetti S, Oppenheim A, Jobling MA, Jenkins T, Ostrer H and Bonne-Tamir B. 2000. Jewish and Middle Eastern non-Jewish populations share a common pool of Y-chromosome biallelic haplotypes.Proc Natl Acad Sci U S A. 97:(12)6769-6774.

Hitti PK. 1970.History of the Arabs from the earliest times to the present. Macmillan, St Martin's Press. London, New York.

Jobling MA and Tyler-Smith C.1995.Fathers and sons: the Y chromosome and human evolution.Trends Genet 11:(11)449-456.

Jorde LB, Watkins WS, Bamshad MJ, Dixon ME, Ricker CE, Seielstad MT and Batzer MA.2000.The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data.Am J Hum Genet 66:(3)979-988.

Kayser M, Kittler R, Erler A, Hedman M, Lee AC, Mohyuddin A, Mehdi SQ, Rosser Z, Stoneking M, Jobling MA, Sajantila A and Tyler-Smith C.2004.A comprehensive survey of human Y-chromosomal microsatellites.Am J Hum Genet 74:(6)1183-1197.

Kivisild,T., Kaldma,K., Metspalu, M., Parik, J., Papiha,S., Villms, R.,.1999.The Place of the Indian mtDNA Variants in the Global Network of Maternal Lineages and the Peopling of the Old World.

Kivisild,T., Rootsi,S., Metspalu,M., Metspalu,E., Parik,J., Kaldma,K., Usanga,E., Mastana,S., Papiha,S.S, Villems,R.2003.The Genetics of Language and Farming Spread in India.

Kivisild T, Rootsi S, Metspalu M, Mastana S, Kaldma K, Parik J, Metspalu E, Adojaan M, Tolk H-, Stepanov V, Golge M, Usanga E, Papiha SS, Cinnioglu C, King R, Cavalli-Sforza L, Underhill PA and Villems R.2003.The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations.Am J Hum. Genet 72:(2)313-332 ER.

Krings M, Salem AE, Bauer K, Geisert H, Malek AK, Chaix L, Simon C, Welsby D, Di Rienzo A, Utermann G, Sajantila A, Pääbo S and Stoneking M.1999.mtDNA analysis of Nile River Valley populations: A genetic corridor or a barrier to migration?Am J Hum Genet 64:(4)1166-1176.

Kunter M.1996.Human skeletal remains from the Oman Peninsula dating from the 4th to the 2nd millennium BC. A survey. Homo .47 :(1-3)43-60.

Lewis B.1990.Race and slavery in the Middle East: an historical
enquiry .Oxford University Press.New York

Luis JR, Rowold DJ, Regueiro M, Caeiro B, Cinnioğlu C, Roseman C, Underhill PA, Cavalli-Sforza LL and Herrera RJ.2004.The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations.Am J Hum Genet. 74:(3)532-544.

Lyovin A.1997.An introduction to the languages of the world. Oxford University Press. New York

Macaulay V, Richards M, Hickey E, Vega E, Cruciani F, Guida V, Scozzari R, Bonné-Tamir B, Sykes B and Torroni A.1999.The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs.Am J Hum Genet. 64:(1)232-249.

Manni F, Leonardi P, Barakat A, Rouba H, Heyer E, Klintschar M, McElreavey K and Quintana-Murci L.2002.Y-chromosome analysis in Egypt suggests a genetic regional continuity in Northeastern Africa.Hum Biol. 74:(5)645-658.

Metspalu E, Reidla M, Parik J, Kivisild T, Tambets K, Usanga E and Villems R.2002.Human mitochondrial DNA diversity in the near and Middle East and in northeastern Africa: A phylogeographic approach .European Journal of Human Genetics. 10:(Supplement 1)72-73 ER.

Mourant AE, Kopec AC, Domaniewska-Sobczak K.1976.The distribution of the human blood groups, and other polymorphisms.London : Oxford University Press ,Oxford,UK

Nebel A, Filon D, Brinkmann B, Majumder PP, Faerman M and Oppenheim A.2001.The Y chromosome pool of Jews as part of the genetic landscape of the Middle East.Am J Hum Genet. 69:(5)1095-1112.

Nebel A, Landau-Tasseron E, Filon D, Oppenheim A and Faerman M.2002.Genetic evidence for the expansion of Arabian tribes into the Southern Levant and North Africa.Am J Hum Genet. 70:(6)1594-1596.

Parsons TJ, Muniec DS, Sullivan K, Woodyatt N, Alliston-Greiner R, Wilson MR, Berry DL, Holland KA, Weedn VW, Gill P and Holland MM.1997.A high observed substitution rate in the human mitochondrial DNA control region. Nat Genet. 15:(4)363-368.

Pennarun E, Metspalu E, Kivisild T, El Amri H, El Kabbaj S, Chaventre A, Villems R and Moisan JP.2002.A comprehensive analysis of Arab and Berber maternal lineages in Morocco.European Journal of Human Genetics. 10:(Supplement 1)

Pereira L, Macaulay V, Torroni A, Scozzari R, Prata MJ and Amorim A.2001.Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade.Ann Hum Genet. 65:(Pt 5)439-458.

Qamar R, Ayub Q, Mohyuddin A, Helgason A, Mazhar K, Mansoor A, Zerjal T, Tyler-Smith C and Mehdi SQ.2002.Y-chromosomal DNA variation in Pakistan.Am J Hum Genet. 70:(5)1107-1124.

Quintana-Murci L, Krausz C, Zerjal T, Sayar SH, Hammer MF, Mehdi SQ, Ayub Q, Qamar R, Mohyuddin A, Radhakrishna U, Jobling MA, Tyler-Smith C and McElreavey K.2001.Y-chromosome lineages trace diffusion of people and languages in southwestern Asia.Am J Hum Genet. 68:(2)537-542.

Quintana-Murci L, Veitia R, Fellous M, Semino O and Poloni ES.2003.Genetic structure of Mediterranean populations revealed by Y-chromosome haplotype analysis .Am J Phys Anthropol .121:(2)157-171.

Quintana-Murci L, Chaix R, Wells RS, Behar DM, Sayar H, Scozzari R, Rengo C, Al-Zahery N, Semino O, Santachiara-Benerecetti AS, Coppa A, Ayub Q, Mohyuddin A, Tyler-Smith C, Qasim Mehdi S, Torroni A and McElreavey K.2004.Where west meets east: the complex mtDNA landscape of the southwest and Central Asian corridor.Am J Hum Genet .74:(5)827-845.

Retsö J.2003.the Arabs in antiquity: Their history from the Assyrians to the Umayyads .RoutledgeCurzon.New York

Richards M, Côrte-Real H, Forster P, Macaulay V, Wilkinson-Herbots H, Demaine A, Papiha S, Hedges R, Bandelt HJ and Sykes B.1996.Paleolithic and neolithic lineages in the European mitochondrial gene pool. Am J Hum Genet .59:(1)185-203

Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V and Rengo C, et al.2000.Tracing European founder lineages in the Near Eastern mtDNA pool. Am J Hum Genet. 67:(5)1251-1276..

Richards M, Rengo C, Cruciani F, Gratrix F, Wilson JF, Scozzari R, Macaulay V and Torroni A.2003.Extensive female-mediated gene flow from sub-Saharan Africa into near eastern Arab populations.Am J Hum Genet. 72:(4)1058-1064.

Rootsi S, Magri C, Kivisild T, Benuzzi G, Help H, Bermisheva M and Kutuev I, et al.2004.Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in europe. Am J Hum Genet .75:(1)128-137.

Rosenberg NA, Woolf E, Pritchard JK, Schaap T, Gefel D, Shpirer I, Lavi U, Bonne-Tamir B, Hillel J and Feldman MW.2001.Distinctive genetic signatures in the Libyan Jews. Proc Natl Acad Sci U S A .98:(3)858-863 ER.

Salas A, Richards M, De la Fe T, Lareu M, Sobrino B, Sánchez-Diz P, Macaulay V and Carracedo A.2002.The making of the African mtDNA landscape.Am J Hum Genet . 71:(5)1082-1111.

Santos FR, Pandya A, Tyler-Smith C, Pena SD, Schanfield M, Leonard WR, Osipova L, Crawford MH and Mitchell RJ.1999.The central Siberian origin for native American Y chromosomes. Am J Hum Genet. 64:(2)619-628.

Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, Marcikiae M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti AS, Cavalli-Sforza LL and Underhill PA. 2000. The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. Science. 290:(5494)1155-1159.

Semino O, Santachiara-Benerecetti AS, Falaschi F, Cavalli-Sforza LL and Underhill PA.2002.Ethiopians and Khoisan share the deepest clades of the human Y-chromosome phylogeny. Am J Hum Genet .70:(1)265-268.

Semino O, Magri C, Benuzzi G, Lin AA, Al-Zahery N, Battaglia V, Maccioni L, Triantaphyllidis C, Shen P, Oefner PJ, Zhivotovsky LA, King R, Torroni A, Cavalli-Sforza LL, Underhill PA and Santachiara-Benerecetti AS.2004.Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area.Am J Hum Genet. 74:(5)1023-1034.

Shriver MD, Smith MW, Jin L, Marcini A, Akey JM, Deka R and Ferrell RE.1997.Ethnic-affiliation estimation by use of population-specific DNA markers. Am J Hum Genet. 60:(4)957-964.

Thomas MG, Skorecki K, Ben-Ami H, Parfitt T, Bradman N and Goldstein DB.1998.Origins of Old Testament priests. Nature .394:(6689)138-140.

Thomas MG, Weale ME, Jones AL, Richards M, Smith A, Redhead N, Torroni A, Scozzari R, Gratrix F, Tarekegn A, Wilson JF, Capelli C, Bradman N and Goldstein DB.2002.Founding mothers of Jewish communities: geographically separated Jewish groups were independently founded by very few female ancestors.Am J Hum Genet. 70:(6)1411-1420.

Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T, Vollrath D, Davis RW, Cavalli-Sforza LL and Oefner PJ.1997.Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. Genome Res. 7:(10)996-1005.

Watkins WS, Rogers AR, Ostler CT, Wooding S, Bamshad MJ, Brassington AME, Carroll ML, Nguyen SV, Walker JA, Prasad BVR, Reddy PG, Das PK, Batzer MA and Jorde LB.2003.Genetic variation

among world populations: Inferences from 100 Alu insertion polymorphisms. Genome Res 13:(7)1607-1618.

Weiss BG and Green AH.1987. A survey of Arab history. American University in Cairo. Cairo, Egypt.

White PS, Tatum OL, Deaven LL and Longmire JL.1999.New, male-specific microsatellite markers from the human Y chromosome.Genomics 57:(3)433-437.

YCC. 2002. A nomenclature system for the tree of human Y-chromosomal binary haplogroups.Genome Res. 12 :( 2)339-348.