# The Genetic Structure of the Kuwaiti Population:

# Mitochondrial DNA Markers

By


Jasem Bader Theyab

Submitted to the Graduate Degree Program in Biological Anthropology and the Graduate Faculty of the University of Kansas
in partial fulfillment of the requirements for the degree of
Master's of Arts

Dr. Michael Crawford, Chairperson

Committee members

Dr. James Mielke

Dr. Deborah Smith

Date Defended: June 10th, 2010

The Thesis Committee for Jasem B. Theyab certifies
That this is the approved Version of the following thesis:

# The Genetic Structure of the Kuwaiti Population:

# Mitochondrial DNA Markers

Committee:

Dr. Michael Crawford, Chairperson

Dr. James Mielke

Dr. Deborah Smith

Date Accepted: June 14th, 2010

# Abstract

In the past few decades, researchers using human mitochondrial DNA (mt-DNA) have significantly contributed to our knowledge of human evolution and migration. The Arabian Peninsula is assumed to be one of the first inhabited regions following the expansion of early *Homo sapiens* out of Africa. Kuwait is located in the Northeast portion of the Arabian Peninsula.

This thesis investigated the mitochondrial DNA (mtDNA) genetic variation in 117 unrelated individuals to determine the genetic structure of the Kuwaiti population and compared the Kuwaiti population to their neighboring populations. Restriction fragment length polymorphism (RFLP) and mt-DNA sequencing analyses were used to answer the investigated questions. The result showed that the Kuwaiti population has a high frequency of haplogroup pre-HV similar to other Arabian populations. Furthermore, the MDS plot showed that the Kuwaiti population is clustered with neighboring populations, including Iran and Saudi Arabia, but not Iraq.

*This thesis is dedicated to my mother, father, wife, and the rest of my family for their support, encouragement, and patience*

## Acknowledgment

I would like to thank Dr. Suzanne Al-Bustan for providing me the DNA sample used in this study, for her support during my research, and for giving me access to her lab to do my analyses. Without her help and encouragement this work would not completed.

I would like to thank my committee member for their guidance and support throughout my master program: Dr. Michael Crawford, Dr. James Mielke, and Dr. Deborah Smith; my fellow students including: Anne Justice, Kristine Beaty, Norberto Baldi, Orion Graf, and my other friends and fellow students in the Department of Anthropology and the Laboratory of Biological Antrhopology; the Department of Anthropology office staff: Judy Ross and Carol Archinal.

I would like to thank my mother, father, wife, and the rest of my family and friends for their support and encouragement during my studies.

## Table of Contents

**List of Figures**

**List of Tables**

# Chapter One: Introduction

The Arabian Peninsula has a critical geographical location linking Africa, Asia, and Europe. Since ancient times, this location witnessed numerous migrations between Africa, Asia and Europe (Cabrera et al. 2009). From the evolutionary standpoint, one of the most important migrations is the diaspora of *Homo sapiens* out of Africa circa 60,00-100,000 years BP (Crawford 2007; Reich and Goldstein 1998). This migration enabled humans to migrate to Asia, Europe, and the reminder of the world. This past migration has attracted the attention of many scientists from different fields to investigate the chronology of this major diaspora and to determine whether it was one major expansion or bidirectional migration (Al-Zahery et al. 2003; Derricourt 2005; Luis et al. 2004). Besides this major migration, scientists have investigated the origins of modern humans, and have tried to determine whether modern humans are descendants from a single ancestor or several ancestral populations (Lahr and Foley 1994; Ramachandran et al. 2005; Relethford 2008; Relethford and Harpending 1994; Stringer and Andrews 1988).

Genetic data from *H. erectus* and archaic *H. sapiens* fossil remains have been used by scientists to investigate the origin of modern humans (Lahr and Foley 1994; Relethford 1995; Stringer and Andrews 1988; Wilson and Cann 1992). The fossil evidence indicates that *H. erectus* migrated from Africa to Asia and Europe circa 1.5 mya (Stringer and Andrews 1988). Over that long expanse of time, *H. erectus* evolved into Neandertals and anatomically modern *H. sapiens* in different geographical areas independently with possible gene flow between regions (Wolpoff et al. 2000). This model is known as Multiregional model and is supported by morphological traits in the fossil remains (Lahr 1994; Weidenreich 1943). The molecular genetic evidence, on the other hand, supports a

recent African origin of all modern humans. This model is known as Recent Out-of-Africa (Hedges et al. 1992; Vigilant et al. 1991). The Recent Out-of-Africa model is supported by mt-DNA and Y-chromosome analyses which indicate that the African populations have the highest genetic diversity among all other populations (Gunz et al. 2009; Nei 1995). In general, worldwide genetic diversity has a specific pattern which reaches its maximum in Africa and decreases as a function of geographical distance (Serre and Paabo 2004). Because of the absence of recombination in mt-DNA and NRY, genetic diversity increased as a result of mutation accumulation overtime. In other words, populations with the longest evolutionary history have the highest genetic diversity (Bowcock et al. 1994; Watson et al. 1996). The origin of modern humans is one of the most contentious topics in the field of evolutionary anthropology. Today, these questions are addressed through multidisciplinary approaches including archaeological, paleo-environmental, and genetic data (Cavalli-Sforza et al. 1988). Genetic data are a key to understanding the source of the genetic variation in current populations and reconstructing their evolutionary history.

A population is defined by geneticists as "a group of organisms of the same species living within a sufficiently restricted geographical area that any member can potentially mate with any other member" (Hartl and Clark 2007). Each population has a unique genetic structure which is determined by the genes of its members. Generally, genetic structure of a population refers to frequencies and distribution of all the alleles in a population. The genetic structure can be altered due to the effect of gene flow, gene drift, and natural selection (Hartl and Clark 2007). Genetic structure of a population was initially studied through the use of frequency variation of the classical markers, such as

protein polymorphisms, blood group, and enzyme electrophoretic polymorphisms (Al-Nassar et al. 1981a; Sawhney et al. 1984). During the1980s, a major transition in the field of anthropological genetics took place with the development of molecular genetic techniques (Cavalli-Sforza and Feldman 2003; Crawford 2007). These included:

(1) Rapid and efficient DNA extraction methods.

(2) The detections of mutations through Restriction Fragment Length Polymorphisms (RFLPs).

(3) DNA Polymerase Chain Reaction (PCR) amplification techniques which permitted the use of small quantities of DNA.

(4) Automated and high through-put DNA sequencing which enabled scientists to sequence either region of the DNA or the complete human genome.

(5) New markers were discovered including mitochondrial DNA sequence variation, Single Nucleotide Polymorphisms (SNPs), and Short Tandem Repeat Polymorphisms (STRPs). These techniques and discoveries provided greater information about population structure and history.

The first attempt at reconstructing the evolutionary history of populations through the use of DNA technology was initiated by Brown (1980). Brown (1980) studied mitochondrial DNA (mt-DNA) and proposed the existence of population specific patterns of restriction enzyme cleavage. In 1987, a groundbreaking study by Cann et al (1987) presented genetic evidence that the last shared ancestor of all contemporary humans existed in Africa no more than 200,000 years ago. Since then, many studies have focused

on the origin and expansion of *Homo sapiens* (Hedges et al. 1992; Nei 1992; Templeton 1992). These studies investigated African, Asian, and European populations, but paid little attention to Middle Eastern populations especially those from the Arabian Peninsula.

Recently, scientists reconsidered the critical role of the Arabian Peninsula in the expansion of early *Homo sapiens* (Cabrera et al. 2009; Jeffrey and Michael 2009). As a result, archaeological and molecular investigations in the Arabian Peninsula began to reconstruct the evolutionary history of the contemporary populations of the Arabian Peninsula (Abu-Amero et al. 2008; Alshamali et al. 2009; Beyin 2006; Carter 2006; Rose 2007). The Arabian Peninsula consists of seven countries: Yemen, Saudi Arabia, United Arab Emirates, Qatar, Bahrain, Kuwait, and Jordan (Rose and Petraglia 2009). Almost all of these countries were investigated either through genetic data, archaeological data, or both (Abu-Amero et al. 2009; Alshamali et al. 2009; Jeffrey and Michael 2009; Pérez-Miranda et al. 2006). The Kuwaiti population was sampled recently using molecular data such as STRs and Y-chromosome (Alenizi et al. 2008; Mohammad et al. 2009). These molecular data permit the separation of the paternal genetic or recombined structure of the Kuwaiti population. The results of this analysis allow the reconstruction of paternal portion of the evolutionary history of Kuwaiti population. The other half of the Kuwaiti population history is the maternal structure--which is investigated in this thesis.

In sum, this thesis investigates the maternal genetic structure of Kuwait and consequently fills a gap in the understanding of the evolutionary history of the Arabian Peninsula. The aims of this thesis are to:

(1) Investigate the maternal structure of the Kuwaiti population through mt-DNA sequencing and RFLPs analyses.

(2) Examine the relationship between Kuwait and neighboring populations.

# Chapter Two: Literature Review

## Origin of Modern Humans:

The origin and dispersal of *Homo sapiens* is one of the most debated topics in the field of evolutionary biology. Over the past decade, this topic has been discussed and debated extensively by molecular geneticists and paleontologists (Lahr and Foley 1994; Relethford 1998; Relethford 2008). However, there is no single agreement to the question of modern human origins. Scientists have proposed two models to explain the origin of modern *Homo sapiens*: the Multiregional Origin model and the Recent Out-of-Africa model (Lahr and Foley 1994; Wilson and Cann 1992).

The Multiregional Origin model was originally proposed by Franz Weidenreich (1943) who hypothesized that the regional differences in morphology between modern human populations are caused by independent evolution from *Homo erectus*. Weidenreich's theory has since been used to explain the origins of modern *Homo sapiens*. This model proposes that some local populations in Europe, Asia, and Africa continued their indigenous evolutionary development from *Homo erectus* to *Homo sapiens*. In this model, regional continuity and similarity between populations was explained by genetic drift, gene flow, and selection (Relethford 2008; Templeton 2007; Wolpoff et al. 2000). Thorne & Wolpoff, 1992, have interpreted the fossil record and certain morphological traits of *H. erectus* found in Asia, Europe, and Australia as support for the model.

In contrast, the Recent Out-of-Africa model proposed a single recent common ancestral population for *Homo sapiens*. According to the proponents of this model, the common ancestor evolved in Africa approximately 200 thousand years ago (Kya). This single population went through a demographic expansion throughout the world, replacing the archaic *Homo* populations (Cann et al. 1987; Stringer and Andrews 1988). This model is supported by the archaeological and molecular evidence of the early anatomically modern *Homo sapiens* in Africa and the Middle East (Bar-Yosef and Belfer-Cohen 2001; Goldstein and Chikhi 2002).

The Multiregional Origin and Recent Out-of-Africa models differ in their assumptions about the sizes of ancestral populations, dates of the bottleneck before the expansion, and the cause of the diversity within our ancestral population. However, the Recent Out-of-Africa model is overwhelmingly supported by molecular genetic data (Carter 2007; Omkar et al. 2009).

Although the Recent Out-of-Africa model has been scientifically supported, questions regarding the timing, routes, and geographical destination are still being debated. Recent studies have proposed four routes of human expansion out of Africa (Derricourt 2005; Jeffrey and Michael 2009; Rose 2007; Rose and Petraglia 2009). The most widely supported route is the Sinai Peninsula (a land route) while the other routes are water crossings, including the Sicilian Channel, the Strait of Gibraltar and the Strait of Bab el-Mandab in the southern Red Sea (Derricourt 2005).

The Sinai Peninsula and the Strait of Bab el-Mandab are the two routes leading toward the Arabian Peninsula, which is a junction connecting Africa, Asia, and Europe

(see Figure 1). The Arabian Peninsula is assumed to have played a critical role in early *Homo sapiens* migration out of Africa. Hence this discussion will focus on the Sinai Peninsula and the Strait of Bab el-Mandab routs (Derricourt 2005).


**Figure 1. Map of the Arabian Peninsula. Strait of Bab el-Mandab is circled in the south, and the Sinai Peninsula is circled in the north. Modified from connect.in.com.**

Using the Sinai Peninsula and the Strait of Bab el-Mandab, early *Homo sapiens* were able to reach the Arabian Peninsula and Eurasia. The Bab el-Mandab is the Strait that separates the southwestern portion of the Arabian Peninsula and Africa. According to Petraglia (2003), the early Bab el-Mandab Strait crossing model is less likely because of the absence of a land bridge and a paucity of archaeological evidence. However, Bailey *et al.* (2007) provide paleo-climatic and paleo-geographical evidence that support the early

*Homo sapiens* expansion toward the Arabian Peninsula and Eurasia via the Bab el-Mandab Strait during the Pleistocene.

Although the Sinai Peninsula is the only land bridge that connects Africa and Eurasia, the Sinai Peninsula various geographical and ecological obstacles, ranging from limited water resources to the presence of marshy areas which may have hindered human expansion. Archaeologically, there is no indication of the early *Homo sapiens* settlements in the Sinai Peninsula. Although neither route is supported by archaeological evidence, the assumption of both routes being used remains active (Derricourt 2005).

Questions concerning the origins and expansions of modern humans are still being debated. Statistical methods, molecular techniques, and computer simulations have been applied by scientists to test the Multiregional Origin model *versus* the Recent Out-of-Africa model. These tools are used by scientists to search for genetic patterns that may reflect human migration out of Africa (Cavalli-Sforza and Feldman 2003; DeGiorgio et al. 2009; Mithen and Reed 2002).

Using migration, admixture, and fluctuation in population size, molecular anthropologists and paleoanthropologists have utilized DNA markers and the demographic history of modern humans to identify the genetic patterns of diversification. Markers include mitochondrial DNA (mt-DNA), non-recombining Y chromosomes (NRY), X chromosomes, and autosomal STRs (Cavalli-Sforza and Feldman 2003; Crawford 2007). This study will focus on the maternal side of human migration, using mt-DNA sequences and restriction fragment length polymorphisms (RFLPs) rather than any other markers.

**Mitochondrial DNA (mt-DNA):**

The analysis of mitochondrial DNA (mt-DNA) has been applied to the studies of human origins and evolution by tracking dispersal patterns and estimating times of divergence and coalescence within and between species. Mitochondrial DNA has distinct features which distinguish it from nuclear DNA i.e. mt-DNA is maternally inherited and lacks recombination making it an ideal tool for the investigation of genetic structure of any population through maternal lines. The high substitution rates in the control region of the mt-DNA allows for estimates of coalescence and phylogenetic studies of closely related populations (Bermisheva et al. 2003; Richards et al. 1996)

Mitochondria are double membrane organelles that exist in the cytoplasm of cells and have their own distinct DNA which is distinct from nuclear DNA. Each cell in the human body contains 10 to 100 mitochondrial compartments. The copy of the mitochondrial genome within each mitochondrial compartment varies from 2 to 11 (Bandelt et al. 2006). In the cell, mitochondria function as the centers of energy production which is involved in cellular homeostasis. This energy is produced through the oxidative phosphorylation pathway involving adenosine triphosphate (ATP) (Bandelt et al. 2006; Cavelier et al. 2000).

The mitochondrial genome is a circular, double-stranded DNA, maternally inherited, and consists of approximately 16,600 base pairs (bp). The mt-DNA encodes 13 polypeptides as well as the 22 transfer RNA (tRNA) genes and two ribosomal RNA (r-RNA) genes necessary for the transcription and translation of the genome. The mt-DNA non-coding control region (D-loop) is about 1,100 bp long (i.e. about 7% of the genome)

and is involved in transcription and replication processes. The mt-DNA control region is divided into 3 regions, hypervariable region I (HVS-I), hypervariable region II (HVS-II), and hypervariable region III (HVS-III). The hypervariable region I is the one which is most commonly studied in evolutionary genetics research (Anderson et al. 1981; Bandelt et al. 2006; Bermisheva et al. 2003).

In 1987, a group of researchers presented data that indicated a recent African origin of modern humans occurred between 140 and 280 thousand years ago (Cann et al. 1987). This study was based on restriction fragment length polymorphism (RFLPs) variation among the mt-DNA of 147 humans from five geographic populations, including Africa, Asia, Australia, Caucus, and New Guinea. The RFLP technique was the most commonly used molecular technique during the 1990s. Today, some researchers still apply the RFLP technique but with finer resolution. However, improvements in rapid sequencing technology have moved researchers toward using rapid sequencing rather than the RFLP technique (Abu-Amero et al. 2008; Bandelt et al. 2006; Bermisheva et al. 2003; Budowle et al. 1999).

Sequencing the control region for use in population genetics has been utilized because of the high degree of polymorphism that can be examined in this relatively short sequence. Recent studies using mt-DNA sequencing of HVS-I and HVS-II support the Recent Out-of-Africa model (Gunz et al. 2009; Ingman et al. 2000).

The investigation of the origin of modern humans using advanced molecular techniques continues. Today, the Arabian Peninsula attracts many molecular biologists and archeologists who think that early *Homo sapiens* originated in Africa and passed

through the Arabian Peninsula on their way out of Africa (Jeffrey and Michael 2009; Pérez-Miranda et al. 2006; Rose and Petraglia 2009). The next section will provide recent archaeological and genetics findings specifically for the Arabian Peninsula.

**Arabian Peninsula:**

The Arabian Peninsula links three continents, Africa, Asia, and Europe. Scientists believe that the location of the Arabian Peninsula played a major role in the dispersal of early *Homo sapiens* out of Africa (Jeffrey and Michael 2009; Pérez-Miranda et al. 2006; Petraglia 2003; Rose 2007). Archaeological and evolutionary studies of the Arabian Peninsula have recently started to uncover its role in human migration from Africa.

The Arabian Peninsula refers to a geographical area with 2.5 million $km^2$ landmass. It is surrounded by three seas, the Red Sea in the west, the Arabian Sea in the south, and the Persian Gulf in the east. Yemen, Oman, United Arab Emirates, Saudi Arabia, Qatar, Bahrain, Kuwait, and Jordan are the countries located on the Arabian Peninsula (Rose and Petraglia 2009). The next section provides background information about the archaeology and genetic structure of contemporary populations in the Arabian Peninsula

**Archaeological Background:**

In 1879, the first archaeological excavations of the Arabian Peninsula started in Bahrain. During the last few decades, many archaeological sites have been discovered from all prehistoric periods. The earliest evidence of settlement near the Persian shore dates to 6000-5000 B.C., where the climatic conditions supported human occupation of

the Arabian Peninsula. By this time, evidence of agriculture, animal domestication, and village life were already established in the Arabian Peninsula (Potter 2009).

In Qatar, the earliest stone tools were excavated and dated back to the sixth and fifth millennium. However, these stone tools showed similarities to those excavated from the Levantine area. Scientists suggest that the tools found in Qatar had originated in southern Levantine area, and they belonged to those people who migrated from the southern Levantine area into eastern Arabia (Potter 2009).

Archeological excavations have confirmed the existence of a relationship between the Gulf region and Mesopotamia. Diagnostic pottery sherds associated with to the Ubaid period, 6000 B.C. to 4000 B.C. were found in Kuwait, eastern Saudi Arabia, Bahrain, Qatar, and United Arab Emirates. These sherds belonged to the sedentary people of southern Iraq who contacted southern Arabia through their seasonal fishing voyages or trading expeditions (Potter 2009). In addition, archeological artifacts, such as the Dilmun seal stamp found in Bahrain, eastern Saudi Arabia, United Arab Emirates, and Kuwait from the third millennium confirmed that the Dilmun society also existed in the Arabian Peninsula (Potter 2009).

Recent archaeological excavations are taking place over the Arabian Peninsula. In Saudi Arabia, two archeological sites were discovered and they are known as Wadi Fatimah and Dawādmi. These sites are located in elevated areas and near springs or stream channels that contained abundant lithic resources. These tools suggest that these sites were occupied by early *Homo* approximately 250,000 kya. (Scott-Jackson et al. 2009). In the United Arab Emirates, assemblages of artifacts that belong to the Middle

Paleolithic were discovered in three different locations: Abu Dhabi, Sharjah, and Ras al Khaimah (Scott-Jackson et al. 2009; Wahida et al. 2009). In Oman, more than 350 archaeological sites were discovered recently in the Huqf region containing approximately 1 million artifacts. Archaeologists attributed these findings to the extended occupation of the sites over a long period of time (Jagher 2009). In the Wadi Hadramaut region of Yemen archaeological sites were discovered with an abundance of lithic tools (Crassard 2009).

The archaeological artifacts described above have been attributed to the early and Middle Paleolithic period. However, those artifacts of the Upper Paleolithic period are limited in number and provide little insight into the Arabian Peninsula occupation during that period. The scarcity of the artifacts raised a question regarding the population continuity or discontinuity between the Pleistocene and Holocene. However, the dynamics of the Neolithic expansion into Arabia is explained by three schemes introduced by Uerpmann *et al.* (2009). The first hypothesis considers the occupation of eastern Arabia to be a result of the climatic deterioration in the northern Arabian Peninsula around 6,200 BC. The second hypothesis considers the occupation of eastern Arabia to be a result of extensive population expansions during the early Holocene. The last hypothesis states that the earliest inhabitants in southeastern Arabia came as a result of repopulation by the indigenous population from South Arabia and/or the northeastern African population. The second hypothesis is most strongly supported by the chronology of *Homo sapiens* expansion and the environmental conditions associated with that period (Uerpmann et al. 2009).

The abundant archeological sites in the Arabian Peninsula reflect the critical location of the Arabian Peninsula which served as a corridor for the dispersal of early *Homo sapiens*. Recently, the archeologists' attention is directed toward the Arabian Peninsula, and more research is needed to understand the role of the Arabian Peninsula in the dispersal of early *Homo sapiens* (Petraglia 2003; Uerpmann et al. 2009). As mentioned earlier, the expansion of *Homo sapiens* cannot be understood without a multidisciplinary approach. As a result, it is essential to present the results of genetic findings in the Arabian Peninsula.

**Genetics Background:**

Pérez-Miranda *et al.* (2006) have investigated the Qatari genetic structure by screening 15 autosomal short tandem repeats (STRs) markers. The results showed that the Qatari population is genetically similar to that of the United Arab Emirates (UAE) and slightly different from Syria, Egypt and Turkey. In addition, a genetic affinity with sub-Saharan populations was detected. These authors conclude that the Qatar geographical area served as a bi-directional corridor for recent and ancient migration routes.

Gonzalez *et al.* (2008) studied the genetic structure of the contemporary Jordanian population using mt-DNA sequencing analysis. The sample for this study consisted of 145 individuals representing two populations from different Jordanian regions: Amman (urban) and the Dead Sea (isolated). The aim of this study was to reconstruct the genetic structure of the Jordanian population, and to compare the results with neighboring populations. The results show that the Jordanian population has a high frequency (39%)

of Eurasian haplogroup U3, but lacks the haplogroup pre-HV1. The genetic structure of the Jordanian sub-population from Amman is similar to their surrounding populations. However, the Jordanian sub-population (Druze) that settled near the Dead Sea is genetically isolated and differs from neighboring populations.

Abu-Amero *et al.* (2008) studied the mt-DNA structure of Saudi Arabia to determine if the Strait of Bab el-Mandab (Southern route) was preferred by early *Homo sapiens* in their expansion out of Africa or not. Five hundred fifty-three mt-DNA samples were sequenced and the results demonstrate that Saudi Arabia has been influenced by African, Asian, and European gene flow. The African gene flow was represented by L, M1 and U6 lineages (20%). Of the 553 sequenced mt-DNA, 18% have an eastern origin which has been detected by the presence of U lineages and Indian M lineages. The remainder (62%) belongs to R and N lineages. Abu-Amero *et al.* (2008) have claimed that many population expansions were detected in Saudi Arabia starting during the Neolithic period. During that period, Saudi Arabia acted as a receptor for the gene flow that originated from Africa, Indonesia, India, and Australia. In 2009, Abu-Amero *et al.* (2009) studied the paternal genetic structure of the Saudi Arabian population by using Y-chromosome STRs and single nucleotide polymorphism (SNPs) analyses to determine the role of Saudi Arabia in modern *Homo sapiens* expansion. The total number of the Saudi Arabian males who participated in this study was 157. The results of this study showed that 14% of Saudi Arabian Y-chromosome originated in Africa, 17% originated in the eastern part of Arabia and migrated to Saudi Arabia through Iran, and the majority (69%) had Levantine origin. This study confirmed that Saudi Arabia acted as a major gene flow

recipient since the last Glacial maximum. According to this study, modern *Homo sapiens* expansion was through the Sinai Peninsula but not the Strait of Bab el-Mandab.

Cerny *et al.* (2008) analyzed the genetic structure of the Yemeni population by sequencing 185 mt-DNA of unrelated Yemeni individuals and excluding those who recently immigrated to Yemen. These results were compared to different populations and showed that the western Yemeni sample is genetically related to Middle Eastern and North African populations. The eastern Yemeni sample is genetically similar to East African populations. Different haplogroups were detected in Yemen, and they included: R0a, M1, and L. Cerny *et al.* (2008) explain the finding as a result of extensive gene flow from West Eurasia, Northeast Africa, and South Asia.

Alshamali *at el.* (2009) studied the population dynamics in the Arabian Peninsula. Using Y-STRs; she compared the genetic structure of Oman, Saudi Arabia, Yemen, and United Arab Emirates (UAE) with African, Asian, Middle Eastern, and other neighboring populations. The results suggested that the geographical isolation of the Arabian Peninsula played a major role in the clustering of Oman and UAE in one group proximal to Iran, Egypt, Syria, and Ethiopia on one hand, and Saudi Arabia and Yemen on the other. Alshamali *at el.* (2009) claimed that the Y-STR analysis indicates an expansion from Iran through the Ormuz Strait.

The investigation of the genetic structure of the Arabian Peninsula is still continuing with more countries being investigated in order to reconstruct the evolutionary history of the Arabian Peninsula. More analyses are required to study all the countries in this region. This study will try to fill a gap by identifying the genetic structure of Kuwaiti

population using mt-DNA sequencing and RFLP analyses. The next section focuses on the archeology, history, and recent genetic structure of Kuwait.

**Kuwait:**

Kuwait is a small country located in the Middle East on the coast of the Arabian Gulf (see Figure 2). Kuwait is bordered by Iraq on the north and Saudi Arabia on the south. Its area is estimated at about 17,820 sq km (6,880 sq mi) (Casey 2007). There are nine islands off the coast of Kuwait: Failaka, Bubiyan, Miskan, Warba, Auhha, Umm Al-Maradim, Umm Al-Naml, Kubbar and Qaruh (Casey 2007).


**Figure 2. Map of Kuwait was created at http://www.planiglobe.com/omc_set.html.**

**Archaeology and History of Kuwait:**

Although most of the archaeological sites in Kuwait found to date are located in the Failaka islands, recent excavations in Kuwait discovered the oldest boats yet identified with related pottery remains. These remains were found in a coastal area known as As-Sabiyah (H3) (Carter et al. 1999). According to Robert Carter (2006) and Potter (2009), these remains are evidence of maritime exchange and trade between the

Arabian Neolithic communities of Eastern Arabia and Ubaid communities of southern Mesopotamia during the sixth and fifth millennia BC (Carter 2006; Potter 2009). As-Sabiyah (H3) is currently the only archaeological site in Kuwait, while the rest are found on Failaka Island.

Failaka is the only inhabited island of Kuwait, located 12.4 mile (20 km) off the coast of Kuwait. The artifacts found in Failaka date back to the beginning of the second millennium B.C.E. to the Bronze Age when the ancient Dilmun civilization established a trade and/or a military station in Failaka. Archaeological artifacts found in Failaka include a temple, a stamp seal, and slabs. (Casey 2007).

During the Hellenistic period, the army of Alexander the Great and Greek sailors reached Failaka and established a trading colony. Failaka was named Icaros or Ikaros during that time. The name was engraved on a stone tablet which helped archaeologists determine the history of Failaka. After the death of Alexander the Great in 323 B.C.E., his empire was divided by his generals; one of them was his friend Seleucus who became a king of the Seleucid Empire which occupied and controlled Failaka. The possible events following the arrival of the Seleucids to Failaka are not well documented and not supported by any archaeological evidence (Casey 2007; Pieta et al. 2009).

The first European map showing Kuwait territories was created by a Dutch sailor in 1645. The sailor was searching for his way to Basra (southern Iraq), but by mistake found himself in the Bay of Kuwait. On the early maps, Kuwait sometimes was labeled as Al-Qurayn (Grain) which means “the top of the hill,” and it was also known as Kut, which means “the small castle or fort” (Alghanim 1998; Slot 2003). The archaeological

sites uncovered in Kuwait reflect a vital location which has been used by ancient civilizations as a military or trade station.

**Kuwait Formation:**

The evolution of Kuwait as a state reflects the migration of traditional tribal societies from Najd in early 18th century to find new pastures and water sources that could sustain their livestock. These groups found the necessary pastures and water in Kuwait. The processes of state formation in Kuwait were started by the arrival of the many tribal groups. The social and political organization of these tribal groups facilitated the formation of the State of Kuwait (Khoury et al. 1990).

Kuwait experienced two important stages of state formation, each composed of a major period of change followed by a period of alteration and adaptation to these changes. The first period started with the arrival of Utub families to Kuwait, while the second stage started with the discovery of oil. Both stages had a vigorous impact on the evolution of Kuwait's political structure (Lienhardt and Al-Shahi 2001).

**Utub Families Migration:**

The formation of Kuwait started in the 18th century with the arrival of Utub families into the region as the first settlers. Utub is a group of families derived from the Arab tribe Aniza, which migrated from Najd, currently located in Saudi Arabia. The Aniza tribe had a decentralized political system in which the authority was distributed among small groups. Utub families were dependent on the presence of water and pasture locations for sustaining themselves and their cattle, which were important as a

domesticated food source. Because of the absence of water, they had to relocate to other locations. Similar to most tribes in the Arab Gulf, Aniza tribal unity emerged from kinship sodalities, which trace a common descent through male line. As a result, the political structure of the Aniza tribe depended on kinship sodalities. A sodality is a formal kinship institution that united geographically scattered populations based on common age, gender, activity, or kinship as in Aniza tribe (Alghanim 1998; Anscombe 1997; Slot 2003).

In the 18th century, drought was severe in Najd and there was no way in which the tribes could sustain their lifestyle. As a result, they migrated in search of water and pasture. The migration continued to Kuwait after a short stay in Qatar. Utub families consisted of three principal families called Al-Sabah (the ruling family of Kuwait), Al-Khalifa (the ruling family of Bahrain), and Al-Jalahima (Anscombe 1997; Slot 2003).

By 1785, the election of Shaikh Sabah bin Jabir as a leader of the Utub families was the first sign of authority being established in Kuwait. In addition, the creation of a mud wall was the first indication of the consolidation of Kuwait's territories which were located under Al-Sabah power. Also, the mud wall provided a kind of protection for the Al-Sabah authority and territories against other tribes (Slot 2003).

In the 1760s, dramatic changes occurred in Kuwait. These changes attracted many merchants and Kuwait became a center of trade. These changes started with the disagreement between the Al-Khalifa family and Shaikh Abdullah Al-Sabah and ended with the departure of the former to Zubara (southern Iraq) and then to Qatar. In 1812, Shaikh Jabir Al-Sabah was elected by Utub families after his father's death (Slot 2003).

To overcome the difficulties of living under harsh environmental conditions, the tribes organized themselves and divided the work to sustain their lives. In terms of organization of individual operations, fishing and merchant seafaring were among the maritime activities established in Kuwait. The extra fish or pearls were shipped to external markets that were located in India or Africa. The need for imported goods was the main reason motivating the Kuwaiti tribes to undertake long and arduous trading voyages (Lienhardt and Al-Shahi 2001). In addition to maritime trade with India and Africa, Kuwait also became an important terminus for the trade routes that linked India, Arabia, and Persia to Europe (Slot 2003).

Trade divided the Kuwait population into three social groups: the ruling family, the merchants, and the Bedouin. The main function of the ruler was as arbitrator in all major disputes. The few merchants controlled the commerce in Kuwait. Finally, the Bedouins, remained a mobile group who lived in black tents in the desert while their cattle searched for new pastures in the desert (Casey 2007; Crystal 1992).

**Oil Discovery & its Impact:**

In 1932, oil was discovered in Kuwait, but commercial production did not start until June 1946 because of the outbreak of the Second World War. In 1950, Kuwait's production of oil reached 17 million metric tons. The rise in oil production increased Kuwait's income. Oil production had a marked impact on the social and political systems of Kuwait. The most critical impact of oil production was an increase in Kuwait's population. In addition, oil production industries offered many job opportunities which

attracted approximately 100,000 skilled and unskilled workers from Asian and African countries (Slot 2003).

Because of the attraction of high salary, free health care, education, and other services of a welfare state, Kuwait's population doubled by the mid-1950s. The first census taken in 1957 indicated a total population of 206,473, the second census in 1961 gave a total population of 321,621 (Slot 2003). The 2008 edition of the Kuwait census indicates a total population of 3.3 million people. However, more than half, 2.3 million, are immigrants. Kuwaiti population makes up 45% of the total population, followed by Asians 9%, Iranians 4%, and 32% minorities from other populations (Casey 2007) (see Figure 3). Kuwaiti population can be classified into two major groups, Persian and Arab either from Saudi Arabia (Najd) or Iraq (Alenizi et al. 2008). However, history indicates that Kuwait acts as a gene flow recipient from neighboring and other populations which may alter the gene pool of Kuwaiti population and result in heterogeneous genetic structure.

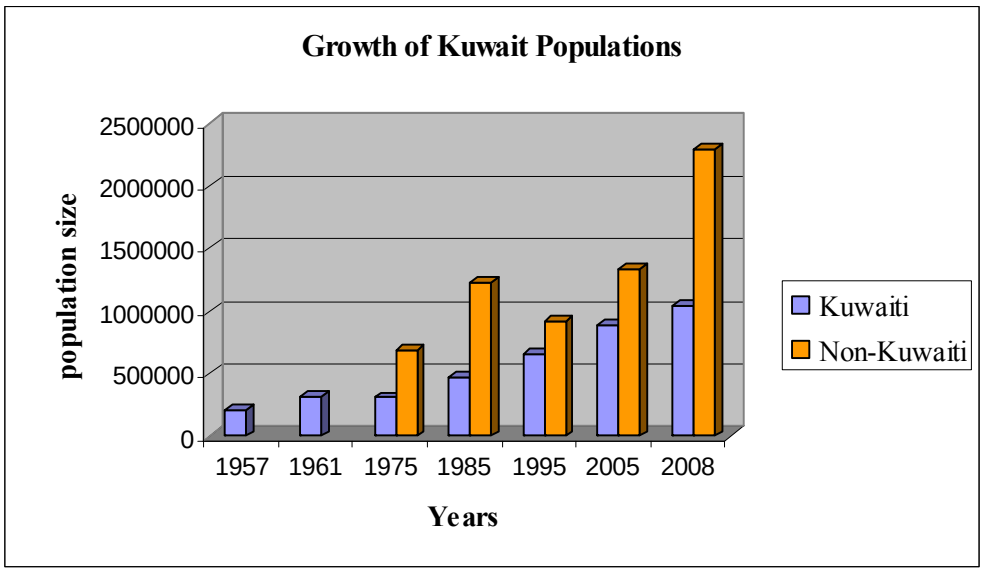


**Figure 3. The growth of Kuwaiti and non-Kuwaiti populations from 1957 to 2008.**

**Genetic Structure of the Kuwaiti Population:**

The genetic structure of the Kuwaiti population was investigated previously using both classical and molecular markers. Classical markers such as blood groups, serum proteins, and red-cell enzymes were used widely in beginning of the 20th century when the ABO blood groups were discovered and used to study population similarities. The ABO blood groups system of Kuwaiti population is characterized by high frequencies of O alleles (44.6%). According to Al-Nassar *et al.* (1981a) and Sawhney *et al.* (1984), the Kuwaiti population is similar to neighboring populations of the Arabian Peninsula. In contradiction to the findings of Al-Nassar *et al.* (1981a) and Sawhney *et al.* (1984), Al-Bustan *et al.* (2002) and El-Zawahri and Luqmani (2008) studied the frequency of ABO alleles in Kuwaiti population and they concluded that the distribution of ABO alleles of the Kuwait are significantly different from neighboring populations in Saudi Arabia, Iraq, and Iran. According to Al-Bustan *et al.* (2002), the distribution of ABO allele frequencies of 18,558 unrelated Kuwaiti individuals are 44.6% for O, 26.7% for A, 24.1% for B, and 4.6% for AB. These differences are attributed to the fact that Kuwait is a heterogeneous population which originally started with small founding populations from Saudi Arabia, Iraq, and Iran. Over short period of time, Kuwait acts as a gene flow recipient from other countries such as Syria, Palestine, Lebanon, and India which doubled the total population size (Al-Bustan et al. 2002).

For the MNS system, the Kuwaiti population has high frequencies of Ms and Ns alleles (54% and 46%, respectively), but considerably lower than neighboring populations (Al-Nassar et al. 1981a). In addition, the Kuwaiti population exhibits low frequency of MS allele whereas neighboring populations have high frequencies. The

Kuwaiti population is similar to neighboring populations in that they share the same frequency of NS allele. According to the Rh system, the presence of an African component in the Kuwaiti gene pool is demonstrated by the high frequency of Dce (Ro) haplotype (Sawhney et al. 1984).

Recently, the genetic structure of Kuwaiti population was investigated through screening Y-STRs, autosomal STRs, and SNPs markers. Alenizi *et al.* (2008) investigated the genetic structure of the Kuwaiti population through screening 15 Y-STRs markers in 502 unrelated individuals. The results were compared to neighboring populations, and showed that the genetic structure of the Kuwaiti population was similar to neighboring groups. In another study, Mohammad *et al.* (2009) studied the genetic structure of six Bedouin tribes from Kuwait. According to the analyses of autosomal STRs and Y-chromosome STRs and SNPs, the Bedouin tribes are clustered with their geographical neighbors in Iran and Saudi Arabia.

**Summary:**

The origin of the modern humans has been extensively debated by scientists over the past decades. To address the question of human origin, scientists have proposed two models, the Recent Out-of-Africa and the Multiregional Origin (Lahr and Foley 1994; Wilson and Cann 1992). In comparison to Multiregional Origin model, the Recent Out-of-Africa model has been supported by recent molecular studies and is more widely accepted in the scientific community (Goldstein and Chikhi 2002; Takahata and Satta 1997).

The dispersal of early *Homo sapiens* out of Africa toward the Arabian Peninsula and Eurasia was through the Sinai Peninsula, and three water crossing, the Sicilian Channel, the Strait of Gibraltar and the Strait of Bab el-Mandab. How and when the early *Homo sapiens* migrated out of Africa is still debated (Derricourt 2005).

Recent advances in molecular genetic technology have enabled scientists to trace the origins of humans by using mt-DNA, Y-chromosome, and autosomal DNA analyses. RFLP and mt-DNA sequencing techniques are used widely to study the genetic structure and maternal pattern of migration in different population. Moreover, these techniques enable scientists to reconstruct the dispersal routes of early *Homo sapiens* out of Africa (Cavalli-Sforza and Feldman 2003; Crawford 2007).

The critical location of the Arabian Peninsula with its complicated archaeological history attracts molecular geneticists to study its genetic structure and reconstruct the dispersal routes of early *Homo sapiens* Out of Africa (Potter 2009). Today, extensive genetic studies are taking place in the Arabian Peninsula; however, not all the countries have been studied (Abu-Amero et al. 2009; Abu-Amero et al. 2008; Alshamali et al. 2009; Černý et al. 2009; Crassard 2009; Pérez-Miranda et al. 2006). The aim of this study is to:

(1) Investigate the genetic structure of Kuwaiti population using mt-DNA sequencing and RFLP techniques

(2) Compare the Kuwaiti population to neighboring countries

# Chapter Three: Material and Methods

## Samples:

Blood samples from 117 unrelated Kuwaiti volunteers who participated in this study were drawn by the Ministry of Health's certified nurses into EDTA vacutainer tubes. Surnames of volunteers were checked to eliminate any potential relatives. The samples were categorized into one of three ethnicity based on their family's place of origin: Arab (n=48), Bedouin (n=32), or Iranian (n=36). The Arab group represents individuals whose maternal and paternal ancestors originated from the Arabian Peninsula. The second group, the Bedouins, represents individuals whose maternal and paternal ancestors originated from the Arabian Peninsula and who are still members of the Bedouin tribes. The Iranian group represents individuals whose families migrated from Iran to Kuwait before oil discovery and have lived in Kuwait for several generations. Permission from the Human Subject Committee of Kuwait University was acquired to conduct the research. In addition, written informed consent was obtained from each volunteer prior to blood drawing.

## Comparative Populations:

The populations used for comparison with the Kuwaiti population consisted of 1735 individuals representing 18 countries and three continents: Africa, Asia, and Europe (see Table 1). The mt-DNA sequence of HVSI for each individual was obtained from GenBank (www.ncbi.nih.gov/Genbank/) except the following populations: Iraq, Syria, Kurdistan, and European. The mtDNA sequences of these populations were obtained

from (www.gen.tcd.ie/molpopgen/resources.php) (Benson et al. 2006). The following table represents the populations and the number of samples in each population.

**Table 1. Mitochondrial DNA HVSI of the populations included in this study.**

| Continent | Population | NO. of Samples | Mt-DNA sequence Source(s) |
|---|---|---|---|
| Africa | Nigeria | 63 | Watson *et al.* 1996 |
| | Kenya | 78 | Watson *et al.* 1996 |
| | Somalia | 5 | Olivieri *et al.* 2006 |
| | Ethiopia | 75 | Watson *et al.* 1996 |
| | African San | 17 | Tishkoff *et al.* 2007 |
| Asia | Yemen | 90 | Kivisild *et al.* 2004 |
| | Saudi Arabia | 15 | Abu-Amero *et al.* 2007 |
| | Kuwait | 94 | This study |
| | Iraq | 116 | McEvoy *et al.* 2004 |
| | Syria | 69 | McEvoy *et al.* 2004 |
| | Jordan | 9 | Ennafaa *et al.* 2009 |
| | Kurdistan | 53 | McEvoy *et al.* 2004 |
| | Iran | 92 | Nasidze *et al.*2008 |
| | India | 109 | Sharma *et al.* 2005 & Thangaraj *et al.* 2005 |
| Europe | Turkey | 290 | McEvoy *et al.* 2004 |
| | Greece | 179 | McEvoy *et al.* 2004 |
| | England | 242 | McEvoy *et al.* 2004 |
| | Bulgaria | 141 | McEvoy *et al.* 2004 |
| | Romania | 92 | McEvoy *et al.* 2004 |
| *Total* | - | 1829 | - |

**DNA extraction:**

All DNA samples from Kuwait used in this study were extracted by a laboratory technician at Kuwait University, Faculty of Science. DNA was extracted from 5 ml of blood according to the method described by Miller *et al.* (1988) using proteinase K and salting-out procedures. This DNA extraction method takes 2 days to obtain an excellent yield of DNA (50-200 ug).

During the first day, 5 ml of a volunteer's blood was transferred into a 15-ml polypropylene centrifuge tube. 3ml of ice-cold nuclear lysis buffer (10mM Tris-HCL, 400mM NaCl, 2mM $Na_2$EDTA, pH 8.2) was added to the tube and incubated on ice for 5-10 minutes and re-suspended. Then, the cell lysates were digested with 0.2 ml of 10% SDS and 0.5 ml of proteinase K solution (1 mg proteinase K in 1% SDS and 2mM $Na_2$ EDTA), overnight at 37 °C. During the second day, 1ml of saturated NaCl (6M) was added to each tube and mixed vigorously for 15 seconds and then centrifuged for 15 minutes at 2500 rpm. The supernatant, containing the DNA, was transferred into a new 15-ml polypropylene tube. Two volumes of ice-cold ethanol were added to the tube and gently inverted several times until the DNA precipitated. The precipitated DNA was removed with a pipette and transferred to a 1.5-ml microcentrifuge tube containing a 100-200 microliter TE buffer (10mM Tris-HCL, 0.2mM $Na_2$ EDTA, pH 7.5) and was left to dissolve for 2 hours at 37°C.

**DNA amplification:**

The mitochondrial DNA hypervariable region I (HVS-I) was amplified by the author of this thesis and sequenced to identify the haplogroup for each sample. Furthermore, different regions of the mt-DNA were amplified by the author of this thesis for each sample to conduct the Restriction Fragment Length Polymorphism (RFLP) analysis. The samples were amplified using a Polymerase Chain Reaction (PCR) method which amplifies a short segment of DNA in a relatively short period of time. Each amplification reaction tube contained: 2.5 µL of 10X PCR Buffer (Promega), 4.0 µL of $MgCl_2$ (25mM), 0.5 µL of dNTP nucleotide mix, 0.2 µL of Taq polymerase (Promega), 7.8 µL of $ddH_2O$, 2.5 µL of forward primer (10 pmole/µL), 2.5 µL of reverse primer (10

pmole/μL), and 1 to 3 μL of sample DNA (if greater quantities of DNA was used, the $ddH_2O$ was adjusted accordingly).

The PCR reactions were run on an Applied Biosystems GeneAmp® PCR System 9700 by the author of this thesis. Each amplification reaction required specific annealing temperatures and primer pairs to yield a successful amplification of the designated region of the mt-DNA. This information is listed in Table 2.

Each amplification reaction was adjusted to the following thermal profile: 94°C for 1 minute (1 cycle); 94°C for 50 seconds (40 cycles); primer annealing temperature for 1 minute (40 cycles); 72°C for 1 minute (40 cycles); 72°C for 5 minutes (1 cycle); and a hold at 4°C. Amplification reactions were checked for PCR products on a 1.5% agarose amplification gel using electrophoresis at 95 volts for approximately one hour. The gels were prepared by mixing 150 mL of 1X TBE (Tris-borate-EDTA) and 2.25 g of Nusieve agarose (FMC BioProducts, USA) in a flask, and heating to boiling temperature. Once the Nusieve agarose powder dissolved completely, the flask was allowed to cool down to 45°C and stained with ethidium bromide.

A total of 5 μL of the PCR product and 2 μL of loading dye (Promega, Madison, USA) were mixed and added to each well. The first well contained a 50 bp ladder (Invitrogen, UK) to size the PCR products against, while the second well contained a negative control (containing PCR mix but with no DNA). The remaining wells contained the amplified DNA. After an hour, the PCR products were visualized under ultraviolet light.

Each PCR amplification product was subjected to site-specific restriction endonucleases to identify base substitutions or insertion/deletion events. This research investigated the most abundant eight mt-DNA haplogroups in the Arabian Peninsula to determine the genetic variation of Kuwaiti population. These haplogroups are identified through a specific cut at the amplified sequence by a specific restriction enzyme. List of each haplogroup and its associated restriction enzyme is provided in Table 2.

The restriction enzyme digestion was prepared by mixing 5 µl of the amplified DNA, 4.5 µl of distilled water, 2.0 µL of 10X RFLP buffer (New England Biolabs, UK), and 1 µl of restriction enzyme (New England Biolabs, UK). Each digestion reaction was incubated in a water bath at a specific temperature overnight. Table 2 provides the digestion temperature for each restriction enzyme. Then, 5 µl of the digested DNA fragments was mixed with 2 µL of loading dye and electrophoretically separated on a 2% NuSieve gel (FMC BioProducts, USA), at 97 volts for two hours. The digested products were visualized under ultraviolet light.

**Mitochondrial DNA Sequencing:**

The HVS-I (1600-16400 nt) region of the mtDNA control region was sequenced on an ABI 3730 automated capillary sequencer (Applied Biosystems, Foster City, CA) using Applied Biosystems v3.1 big dye kit and following the manufacturer's protocol. The DNA template was prepared by using 25 µl of the PCR reaction (same protocol as RFLP PCR). The annealing temperatures and the primers that were used are listed in Table 2. The PCR products were separated on a 1.5% agarose amplification gel. Once the presence of the DNA fragment was detected, the DNA purification was completed using

AMPure magnetic bead purification (Agencourt Bioscience Corp., USA) on a Biomek NX robot (Beckman Coulter, USA). Amplification products were then sequenced in both directions at the DNA Laboratory of Arizona State University by Dr. Scott Bingham. Each sequencing reaction contained 3 µl of a PCR product, 1 µl of primer, 6 µl of water, and 4 µl of dye (v3.1 )/Taq enzyme mixture.

This PCR reaction was run on the ABI 3730 automated capillary sequencer according to the following thermal profile: 94°C for 1 minute (1 cycle) followed by 40 cycles of 94°C for 10 seconds, 50°C for 15 seconds, and 60°C for 2.5 minutes. Sequences were purified using CleanSeq magnetic bead purification (Agencourt Bioscience Corporation) on a Biomek NX robot (Beckman Coulter) to remove unused primers and dyes. At the end of the sequencing cycles, chromatogram data were recorded on a computer and used for mt-DNA analysis. Chromatogram data which contain mt-DNA sequence were edited and aligned using the following freeware; BioEdit and MEGA 4. All the mt-DNA sequences were aligned against the revised human Cambridge Reference Sequence (rCRS) (Anderson et al. 1981; Andrews et al. 1999).

**Table 2. List of mt-DNA primers, Annealing temperature, PCR product fragment size, and RFLPs reaction condition for each polymorphic site.**

| mtDNA Haplogroup | Primer | Annealing *T (ºC) | PCR Product Size (bp) | Digestion *T (ºC) | RFLP Product Size (bp) |
|---|---|---|---|---|---|
| H (7025 AluI) | L6949: 5'-CCGTAGGTGGCCTGACTGGC- 3'<br>H7052: 5'-TGATGGCAAATACAGCTCCT- 3' | 56 | 123 | 37 | 78; 30; 15 |
| pre-HV (14766 MseI) | L14603: 5' -CTAAACCCCCATAAATAGGAG- 3'<br>H14791: 5' -AGGTCGATGAATGAGTGG- 3' | 50 | 226 | 65 | 184; 21;17; 4 |
| HV (11719 SmaI) | L11852: 5' GGGGGTAAGGCGAGGTTAGC 3'<br>H11718: 5' CGCAGTCATTCTCATAATCGCCCCCGG 3' | 58.8 | 180 | 25 | 155; 25 |
| U (12308 HinfI) | L12216: 5' -CACAAGAACTGCTAACTCATGC- 3'<br>H12338: 5'- ATTACTTTTATTTGGAGTTGCACCAAGATT- 3' | 53 | 123 | 37 | 93; 30 |
| L (3592 HpaI) | L3388: 5'-CTAGGCTATATACAACTACGC-3'<br>H3717: 5'-GGCTACTGCTCGCAGTG-3' | 50.9 | 330 | 37 | 207; 123 |
| M (10397 AluI) | L10252: 5' -TTGATCTAGAAATTGCCCTC- 3'<br>H10527: 5' -GTATTCCTAGAAGTGAGATG- 3' | 48.2 | 276 | 37 | 147; 129 |
| N (10873 MnlI) | L10727: 5' -CTCAATCTCCAACACATATGGC- 3'<br>H10920: 5' -GGTCGGAGGAAAAGGTTG- 3' | 51 | 232 | 37 | 176; 56 |
| *HVS-I targeted nucleotides* | F15971: 5'-TTAACTCCACCATTAGCACC-3'<br>R16410: 5'-GAGGATGGTGGTCAAGGGAC-3' | 56 | 443 | Sequencing | |

**Statistical Analysis:**

Haplotype diversity and nucleotide diversity analyses were performed using a computer software package called Arlequin, version 3.11(Excoffier et al. 2005). Haplotype diversity and nucleotide diversity are two common measures of genetic diversity. These diversity tests are used to differentiate between population growth versus constant size groups. In addition, Arlequin version 3.11 was used to calculate Fu's Fs and Tajima's D which are measurements of selective neutrality (Fu 1997; Tajima 1993). These measures are used to determine whether the mt-DNA HVS-I sequences are statistically significant under the neutral equilibrium model. Mismatch distributions is another molecular diversity analysis that was used to investigate the demographic history of the populations (Rogers et al. 1996; Rogers and Harpending 1992).

Analysis of molecular variance (AMOVA) was conducted using Arlequin 3.11 to examine the genetic structure of populations. Based on molecular data, AMOVA estimates population subdivision through analyzing variation between mt-DNA sequences. In addition, AMOVA estimates the percentage of genetic variation found within and among populations (Excoffier et al. 1992).

The biological relationships among different populations were represented in a Multi-Dimensional Scaling plot (MDS) and Neighbor-Joining tree (NJ). The Neighbor-Joining tree method searches for the best tree that describes the evolutionary relationship among populations. This tree has the minimal total branch lengths (Saitou and Nei 1987). The NJ tree was constructed in NTSYS 2.1 computer program (Rohlf 2008).

The MDS plot is based on $F_{ST}$ pairwise differences between populations. The $F_{ST}$ distance matrix is used as a measure of genetic distance among populations and its values can range from 0 to 1. Values from 0 to 0.05 represent little genetic differentiation among populations, values from 0.05 to 0.15 indicate moderate genetic differentiation, values from 0.15 to 0.25 reflect high differentiation, and values above 0.25 indicate great differentiation. The Kuwaiti population was compared to neighboring and worldwide populations to draw insight regarding the genetic relationship of Kuwaiti population to other populations. The goodness of fit is presented in the stress value; the higher stress value, the lower accuracy of relationship among populations. The MDS plot was constructed in NTSYS 2.1 computer program (Rohlf 2008).

# Chapter Four: Results

In this study, a total of 117 individuals were analyzed using RFLPs analysis. Out of these 117, only 94 individuals were successfully sequenced for HVS-I region of the mitochondrial DNA. To examine the contribution of the Kuwaiti sub-populations to the total Kuwaiti gene pool the Kuwaiti samples were sub-divided into three populations according to their origins: 50 Arabian Peninsula, 36 Iranian, and 31 Bedouins who represent more than one tribal group. According to Figure 3, all Kuwaiti sub-populations have similar haplogroup frequencies. However, the Iranian sub-population shows higher frequency of haplogroup H than the other sub-populations. In addition, the Bedouins sub-population demonstrates high frequency of haplogroup pre-HV which expected since they migrated from Arabia. Furthermore, a chi-square test was used to determine if the 3 sub-populations were homogenous. The chi-square test indicates significant similarities among the three sub-populations ($X^2 = 17.35$; $p < 0.001$, $df = 12$). Since all sub-populations contributed equally in the Kuwaiti gene pool, there is no need to divide the Kuwaiti populations according to their origin in further analyses.

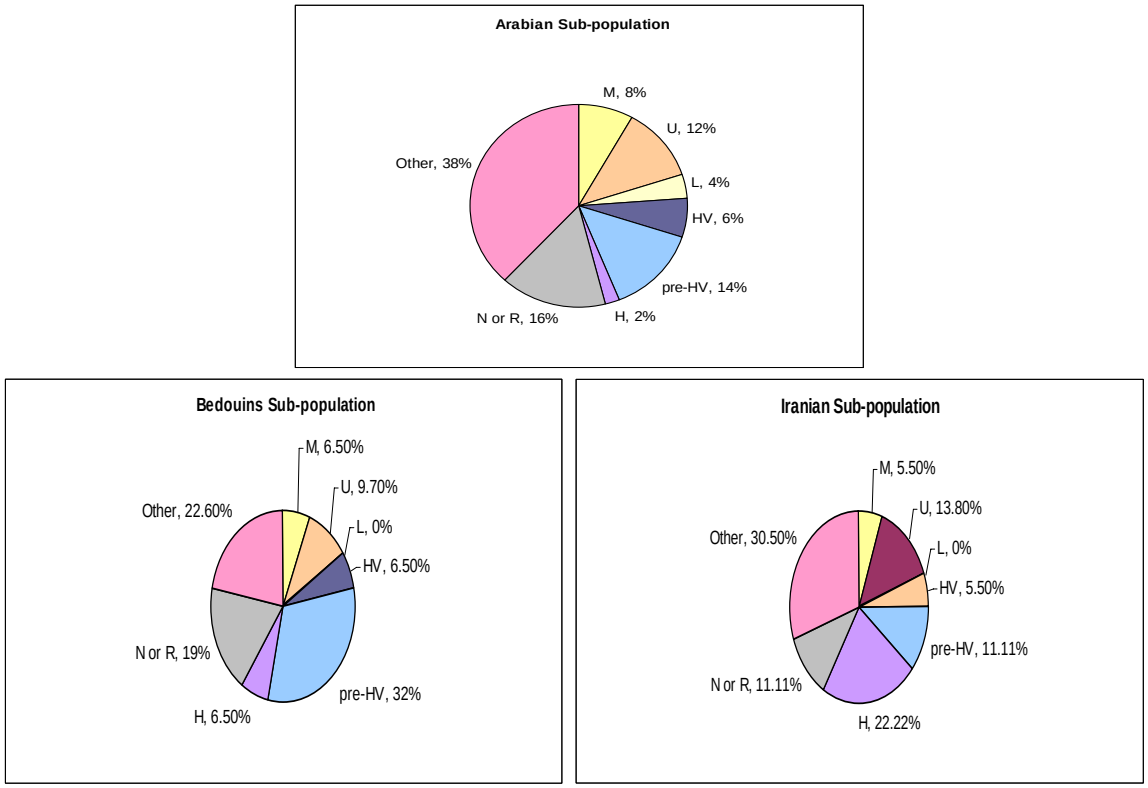


**Figure 4. Mitochondrial haplogroup frequencies in Arabian, Bedouins, and Iranian sub-populations based on RFLPs analyses.**

According to the RFLPs analysis, the Kuwaiti population is a mix of many haplogroups that originated in Africa, Asia, and Eurasia. The most common haplogroup among Kuwaiti individuals is pre-HV (18%), followed by U haplogroup (12%). Haplogroup L is the least common in Kuwaiti population. However, more than one third of Kuwaiti population is not assigned to a specific haplogroup. Since Kuwait is similar to Arabia in its mitochondrial haplogroup frequencies, it is expected that majority of the unassigned Kuwaiti individuals would be assigned to haplogroup J which is most abundant in Arabia. Figure 4 shows the haplogroup frequencies in Kuwaiti population.

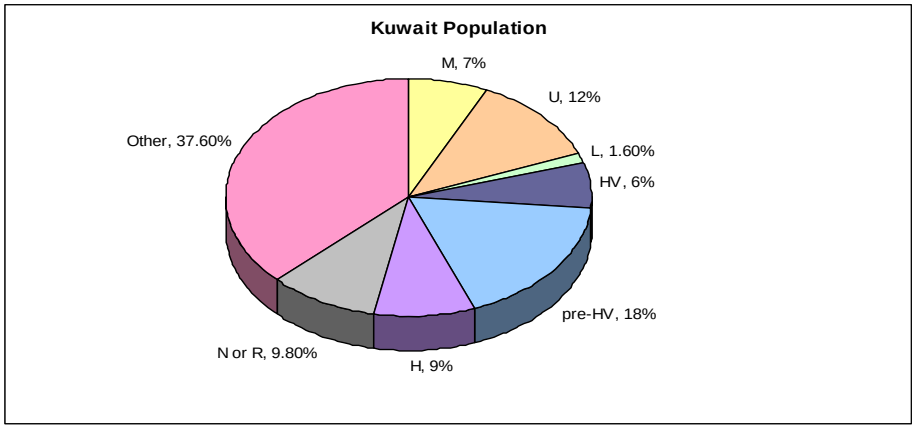


**Figure 5. Mitochondrial haplogroup frequencies in the Kuwaiti population based on**

Table 3 shows mt-DNA haplogroup frequencies in Kuwaiti and compares them to other populations of the Middle East. The haplogroup frequency of the Kuwaiti population is, as expected, similar to Iranian and Arabian populations. However, the Kuwaiti haplogroup frequencies differ from those observed in either Syria or Palestine. Haplogroups pre-HV, HV, H, M, N or R, L and U encompass about 62% of the Kuwaiti subjects. Compared to other haplogroups, pre-HV is the most common one observed in Kuwait and Arabia, 18% and 15.2%, respectively. Furthermore, the influence of African L haplogroup is low in Kuwait (1.6) and Iran (2.2%), but much higher in Arabia (10.5%). According to RFLPs analysis, Kuwait served as a gene recipient from both the Asian and African continents. The presence of some particular haplogroups is an indication of gene flow to Kuwait and possible founder effect of the founding families. The presence of haplogroup L in Kuwait (1.6%) reveals the African gene flow to Kuwait. In addition, haplogroup M and N or R are an indication of Asian gene flow.

**Table 3. Estimated haplogroups frequencies in Kuwait and comparative populations.**

| Haplogroup | Estimated mt-DNA haplogroups frequency (%) in populations | | | | | | |
|---|---|---|---|---|---|---|---|
| | Iraqi[a] | Arabia[b] | **Kuwait** | Palestinian[c] | Syrian[c] | Iranian[b] | Anatolia[b,f] |
| **Sample size** | 216 | 389 | **117** | 117 | 69 | 451 | 388 |
| **Pre-HV** | 4.2 | 15.2 | **18** | 2.6 | 5.8 | 2.4 | 2.8 |
| **HV** | 10.6 | 3.6 | **5** | 1.7 | 4.3 | 5.5 | 3.6 |
| **H** | 19.9 | 12.9 | **9** | 30.8 | 24.6 | 17.1 | 25.0 |
| **V** | 0.5 | 0 | **nr** | 0 | 2.9 | 0 | 0 |
| **J** | 9.3 | 20.8 | **nr** | 9.4 | 10.1 | 13.5 | 10.9 |
| **T** | 8.8 | 4.6 | **nr** | 12.8 | 10.1 | 8.4 | 11.9 |
| **K** | 3.2 | 3.6 | **nr** | 6.8 | 4.3 | 7.5 | 5.9 |
| **U** | 19.0 | 10.5 | **12** | 7.6 | 15.9 | 21.5 | 19.3 |
| **I** | 1.9 | 0.8 | **nr** | 0 | 0 | 2.0 | 2.3 |
| **X** | 2.8 | 1.8 | **nr** | 3.4 | 0 | 2.9 | 4.4 |
| **W** | 1.9 | 1.8 | **nr** | 2.6 | 2.9 | 2.0 | 3.9 |
| **B** | 0.9 | nr | **nr** | 0 | 0 | Nr | 0 |
| **M** | 1.4 | nr | **7** | 1.7 | 1.4 | Nr | 4.4 |
| **N or R** | nr | 12.6 | **9.8** | nr | nr | 6.1 | 1.8 |
| **L1-L3#** | 4.2 | 10.5 | **1.6** | 5.2 | 5.8 | 2.2 | 0.3 |
| **Other** | 11.5 | 6.2 | **37.6** | 15.4 | 11.6 | 8.6 | 4.5 |

**nr: Not reported**

*Asian haplogroup, #African haplogroup. [a]Al-Zahery *et al.* (2003).[b]Kivisild *et al.*(2003).[c]Richards *et al.*(2000).[d]Quintana-Murci *et al.*(2004).[f]Tambet *et al.* (2000).

The results of the mitochondrial DNA diversity and neutrality measures for Kuwaiti and comparative populations are presented in Table 4. The measure of genetic diversity indicates that African Bushmen have the lowest genetic diversity (0.7941). However, recent study of complete mt-DNA sequence indicates that African Bushmen have the greatest genetic diversity (Gonder et al. 2007). The highest genetic diversity was observed in Kenya (0.9960).

The genetic diversity of Kuwait (0.9799) is similar to neighboring populations, Iran (0.9895), Iraq (0.9918), and Saudi (0.9905). The remaining populations all have gene diversities that fall between 0.9960 and 0.7941. Tajima's D (-1.87839) and Fu's Fs (-

25.40242) for Kuwaiti population are significant at $p<0.05$ and $p<0.005$, respectively. These results suggest that the Kuwaiti population may be undergoing an expansion in relatively short period of time. Populations with negative Fs values indicate a population expansion. Kuwait, Somali, Saudi Arabia, and African Bushmen have the lowest Fs values, -25.40242, -0.96395, -10.25, and -11.52, respectively.

**Table 4. Summary statistics for Kuwaiti and comparative populations based on mt-DNA HVS-I sequence data (np 16080-16380), including the number of samples (n), gene diversity (H), nucleotide diversity, the number of pairwise differences between haplotypes ($\pi$).**

| Population | N | H | Nucleotide diversity | $\pi$ | Tajima's D | Fu's FS |
|---|---|---|---|---|---|---|
| Nigeria | 63 | 0.9949(0.0039) | 0.024693(0.013067) | 6.815(3.60656) | -1.40382 | -25.12808* |
| Kenya | 78 | 0.9960(0.0029) | 0.034584(0.017767) | 9.545(4.90382) | -1.34071 | -24.63881* |
| Somalia | 5 | 0.9000(0.1610) | 0.023980(0.015946) | 6.618(4.40119) | 0.18621 | -0.96395 |
| Ethiopia | 75 | 0.9910(0.0044) | 0.034005(0.017499) | 9.385(4.82985) | -1.19564 | -24.68836* |
| African San | 17 | 0.7941(0.0783) | 0.027642(0.015181) | 7.629(4.18986) | -0.05484 | -11.62028* |
| Yemen | 90 | 0.9793(0.007) | 0.028242(0.014703) | 7.795(4.05790) | -1.42630* | -24.87694* |
| Saudi Arabia | 15 | 0.9905(0.0281) | 0.023429(0.013172) | 6.466(3.63536) | -1.36306 | -10.25005* |
| Kuwait | 94 | 0.9799(0.0074) | 0.019810(0.010660) | 5.448(2.93155) | -1.87839* | -25.40242* |
| Iraq | 116 | 0.9918(0.0036) | 0.020573(0.011) | 5.678(3.03595) | -2.10200* | -25.27107* |
| Syria | 69 | 0.9881(0.007) | 0.019950(0.010766) | 5.506(2.97134) | -2.14505* | -25.44552* |
| Kurdistan | 53 | 0.9833(0.0095) | 0.019922(0.010801) | 5.498(2.98117) | -1.99134* | -25.45367* |
| Iran | 92 | 0.9895(0.0049) | 0.021020(0.011240) | 5.802(3.10213) | -2.10033* | -25.30259* |
| India | 109 | 0.9674(0.0088) | 0.019047(0.010274) | 5.257(2.83567) | -1.95946* | -25.39018* |
| Turkey | 290 | 0.9858(0.0037) | 0.017938(0.009686) | 4.951(2.67337) | -2.22379* | -25.06563* |
| Greece | 179 | 0.9760(0.0055) | 0.014409(0.008016) | 3.977(2.21243) | -2.17981* | -25.74642* |
| England | 242 | 0.9667(0.0076) | 0.014809(0.008196) | 4.087(2.26215) | -2.24347* | -25.49169* |
| Bulgaria | 141 | 0.9762(0.0065) | 0.015306(0.008459) | 4.224(2.33476) | -2.12006* | -25.72313* |
| Romania | 92 | 0.9811(0.0051) | 0.015691(0.008678) | 4.331(2.39506) | -1.97053* | -25.81281* |
| | | | 0.022064 | 6.0885(3.33172) | -1.59729 | -21.38537 |

***Tajima's D are significant at p<0.05 & Fu's Fs are significant at p<0.005**

Mismatch distributions of Kuwaiti and comparative populations are provided in Figure 5. The distributions of Kuwait, Iran, Iraq, and Yemen show unimodal distributions. These populations reach a peak in the number of pairwise differences between six and eight with frequency ranging between 15 and 20 percent. The Harpending's raggedness index was calculated for each population and it was below 0.03, and not significant at p= 0.05. The smooth bell-shaped distributions accompanied by low

raggedness index (below 0.03) indicate that these populations are in demographic expansion.

However, the mismatch distributions of African Bushmen, Saudi Arabia, Somalia, and Ethiopia show multimodal distributions. In addition, the Harpending's raggedness index for Saudi Arabia and Jordan were above 0.03. None of the values were significant at p= 0.05. According to the raggedness index and the mismatch distributions, these populations had a constant population size over a long period of time.

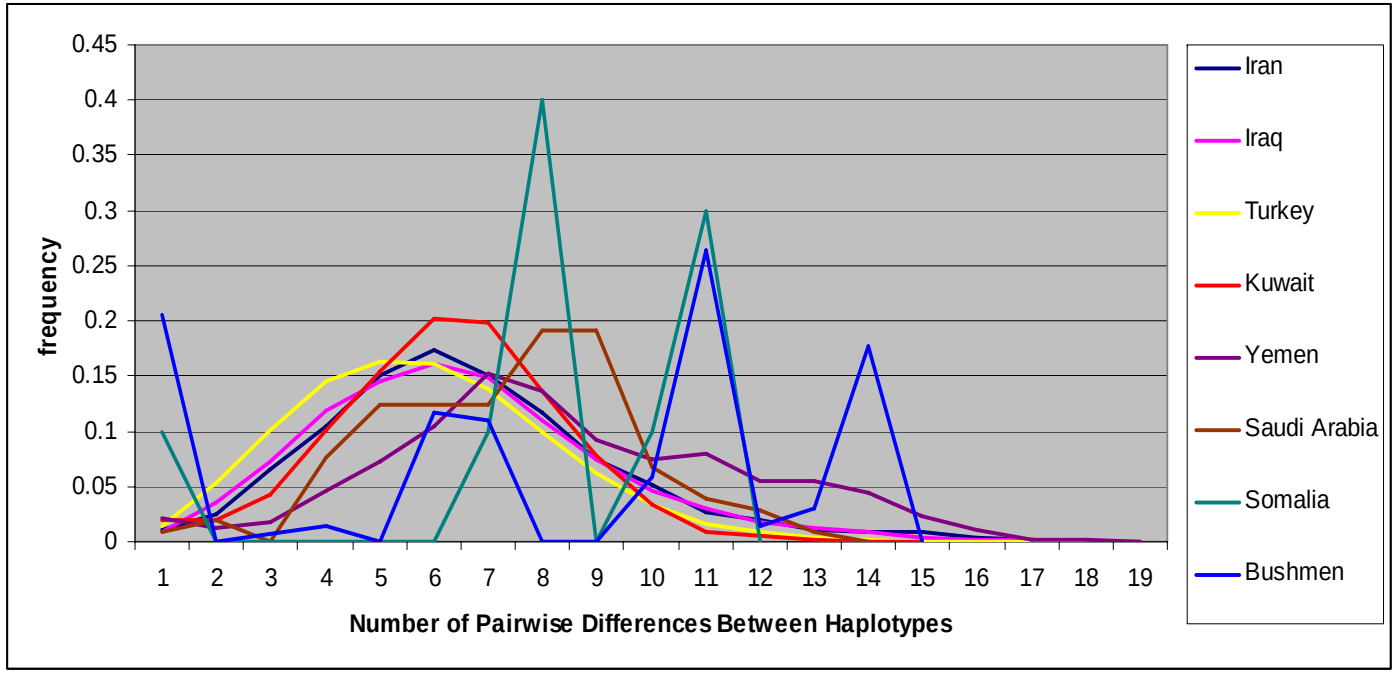


**Figure 6. Mismatch distributions of the Kuwaiti and geographically neighboring populations.**

A neighbor-joining tree (NJ) based on mt-DNA HVS-I sequence data (np 16080-16380) for Kuwaiti and comparative populations are presented in Figure 6. In this plot, the Kuwaiti population clusters with geographically neighboring populations including Iran and Iraq, but not Saudi Arabia. As expected, the European populations cluster together.

Bushmen being the oldest population among the other populations exhibit the longest length of branch in the NJ tree. The European populations are genetically closer to each other. Kuwait and Iran share a common ancestor and they cluster with Iraq which is geographical neighbors of Kuwait. Furthermore, Syria is closely related to Turkey and Kurdistan. The original distance matrix was compared to a cophenetic distance matrix using a two-way Mantel test. The correlation had an insignificant value of 0.73721. This value indicates that the NJ tree is a poor fit to the data and does not reflect the genetic distance among these populations.

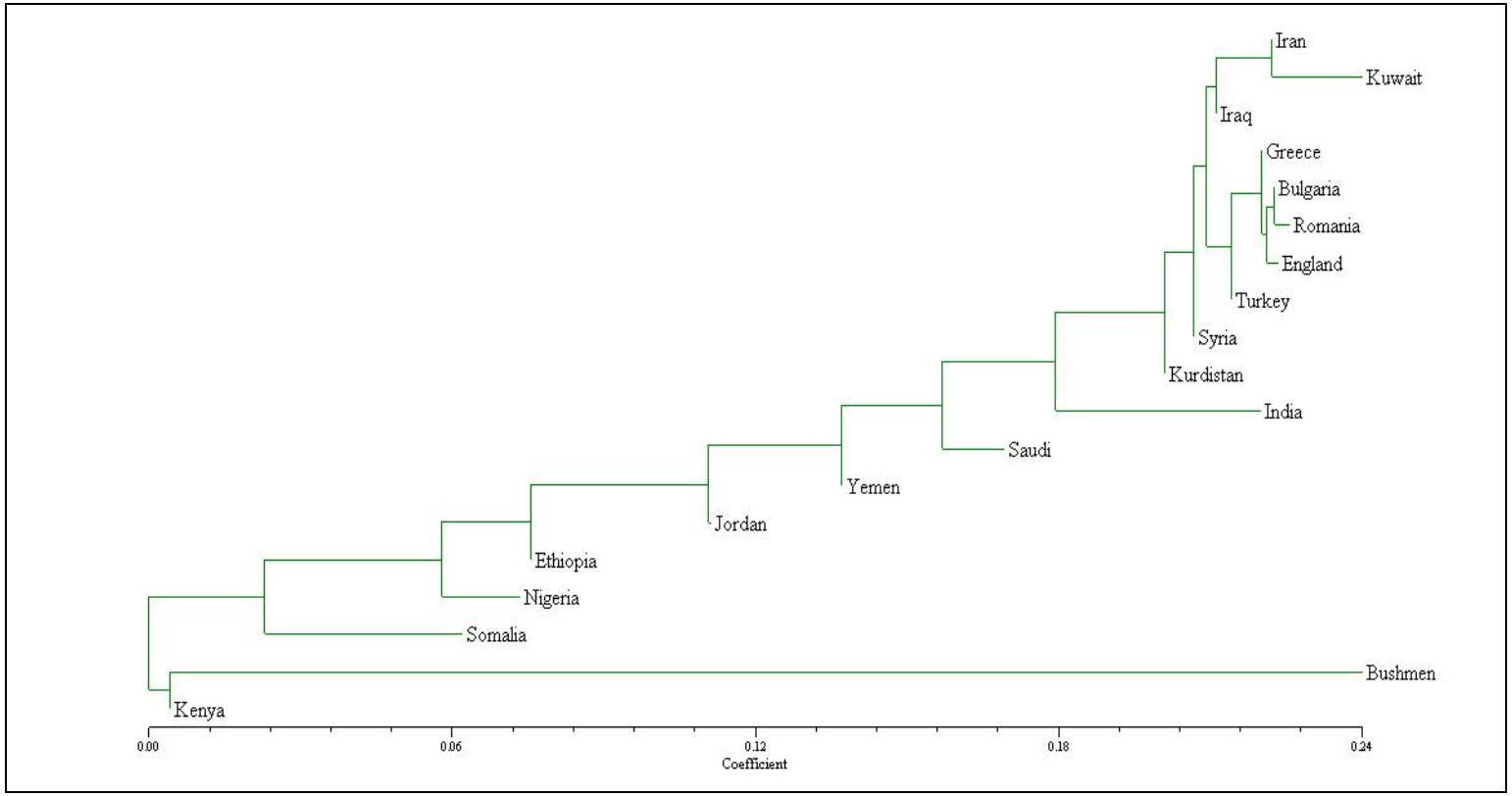


**Figure 7. Neighbor-joining tree based on mt-DNA HVS-I sequence data.**

Since the NJ is a poor fit to the data, a multidimensional scaling (MDS) method is a good alternative to representing the relationship of the Kuwaiti population to comparative populations. Figure 7 shows a multidimensional scaling plot which is based on Tamura and Nei's (1993) distances. The plot indicates that there are three distinct clusters of populations. The first cluster, which is represented in black circles, is composed of African populations except African Bushmen. The African Bushmen group is located in the upper left corner of the graph. This is probably due to the high genetic diversity of African Bushmen group. Furthermore, Ethiopia is the closest African population to the Asian populations. Another MDS plot was constructed excluding African Bushmen group. However, the stress value is reduced and the plot became less informative and does not reflect the genetic distance among populations.

The second cluster represents the Asian populations which are intermediate between African and European populations. However, India is located in the lower right side of the graph away from other Asian (Middle Eastern) populations. In addition, as expected Iraq, Kurdistan, and Syria are close to Turkey. The final cluster contains the European populations and they are represented in red circle. Kuwait is located in the upper right side of the graph near both Iranian and Saudi populations. That reflects the geographical location of Kuwait and the contribution of gene flow from these countries in the Kuwaiti gene pool. The stress value of the MDS plot is good (0.09584). The original distance matrix was compared to a MDS matrix using a two-way Mantel test. The correlation was highly significant with value of 0.98228. This value indicates that the MDS plot is a very good fit to the data.

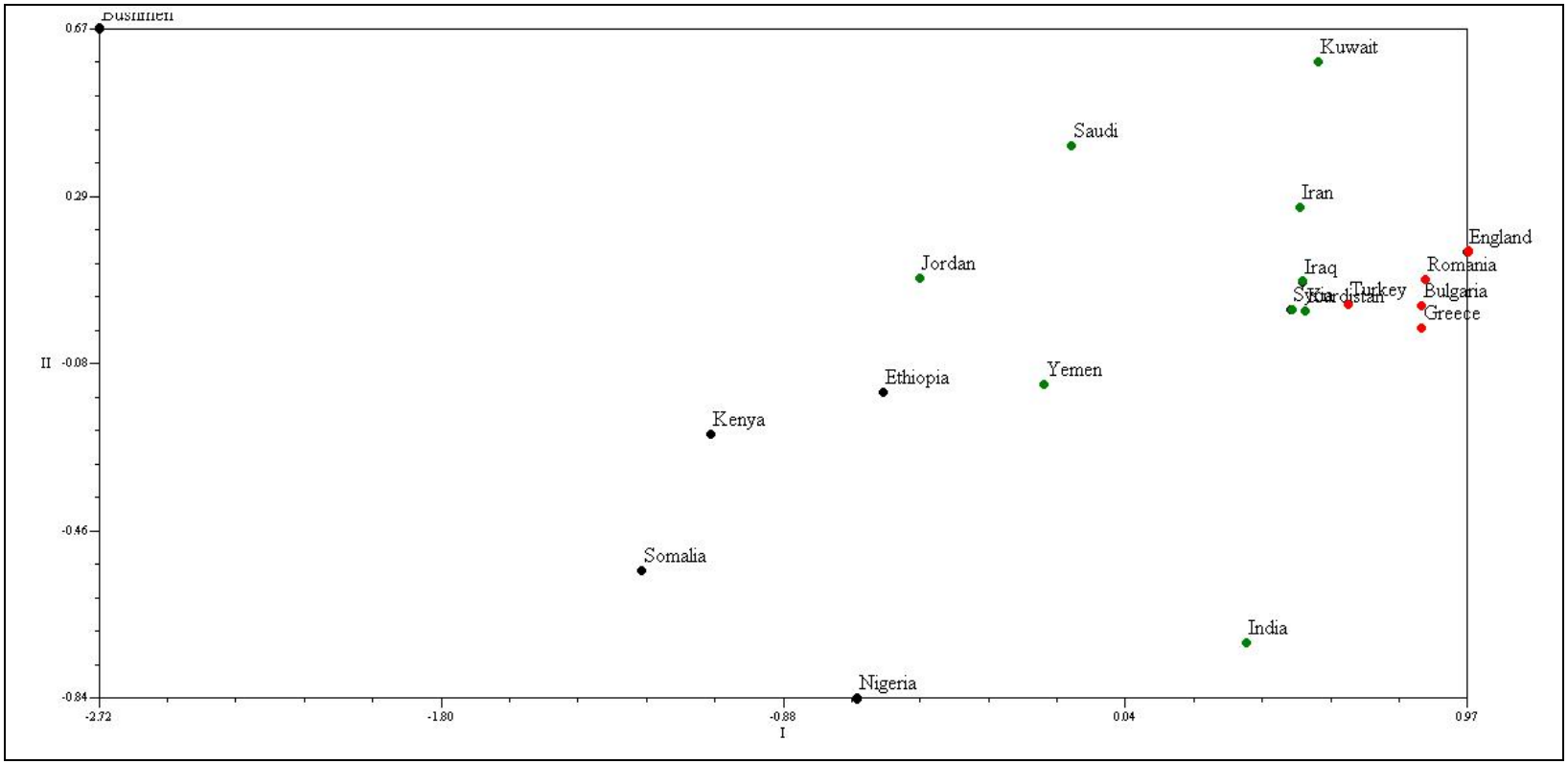


**Figure 8. MDS plot of mtDNA HVS-I sequence data with 19 populations. African populations are black circles; Asian populations are green circle; European populations are red circles.**

To determine the genetic structure of 19 populations based on the mt-DNA HVS-I sequence data, an analysis of molecular variance (AMOVA) was conducted. All populations were classified according to their geographical locations. The first group contains the African populations: Nigeria, Somalia, Kenya, Ethiopia, and African Bushmen. The second group is composed of Asian populations: India, Yemen, Jordan, Saudi Arabia, Iran, Iraq, Kurdistan, Syria, and Kuwait. The last group contains European populations: Turkey, England, Romania, Bulgaria, and Greece. The result of AMOVA analysis is shown in Table 5. The greatest amount of variation was observed within populations and accounted for 89.78% of the total variation, followed by among groups variation (6.68%), and finally within groups variation (3.54%). All variance components are significant.

**Table 5. AMOVA results for comparison of three groups (Africa, Asia, and Europe).**

| Source of variation | d.f. | Sum of Squares | Variance Components | Percentage of Variation | F-Statistics |
|---|---|---|---|---|---|
| Among groups | 2 | 240.779 | 0.19637 | 6.68 | $\Phi_{CT}$ = 0.06676* |
| Within groups | 16 | 194.017 | 0.10423 | 3.54 | $\Phi_{SC}$= 0.03797* |
| Within populations | 1810 | 4779.490 | 2.64060 | 89.78 | $\Phi_{ST}$ = 0.10220* |
| *Total* | 1828 | 5214.286 | 2.94120 | | |

*P-value = 0.00000

# Chapter Five: Discussion

## Genetic Structure of the Kuwaiti Population:

Studies of mt-DNA variation have proven to be useful for examining evolutionary processes in humans and reconstructing population histories (Horai et al. 1995; Torroni et al. 1994; Vigilant et al. 1991). This research investigates the genetic structure of the Kuwaiti population through mt-DNA analyses to understand the relationship of the Kuwaiti population to other Middle Eastern groups. The genetic structure of the Kuwaiti population was previously investigated using classical genetic markers such blood groups (Al-Bustan et al. 2002; Al-Nassar et al. 1981a; El-Zawahri and Luqmani 2008; Sawhney et al. 1984). These studies indicated that the Kuwaiti population differs significantly from the genetic structure of Iraq, Iran, and Saudi Arabia (Al-Bustan et al. 2002; El-Zawahri and Luqmani 2008). In contrast to Al-Bustan *et al.* (2002) and El-Zawahri and Luqmani (2008), and Sawheny *et al.* (1984), demonstrated that the Kuwait genetic structure is similar to geographically neighboring populations. However, this study indicated that the maternal genetic structure of the Kuwaiti population resembles the maternal genetic structure of both Iran and Saudi Arabia and that is expected since Kuwaiti population is a mixture mostly from Iran and Saudi Arabia. Since Iraq is geographically close to Kuwait as is Saudi Arabia and Iran, it was expected that the Kuwaiti gene pool would be influenced by Iraqi gene flow. However, this study did not detect Iraqi gene flow influence on the Kuwaiti gene pool and that may due to the small sample size in this used in this study.

The geographic origins of some mt-DNA haplogroups have been documented, thus these haplogroups can be used to reconstruct the genetic history of populations (Abu-Amero et al. 2007; Abu-Amero et al. 2008; Ennafaa et al. 2009; Kivisild et al. 2003a; Santos et al. 2004; Torroni et al. 1996; Torroni et al. 1994; Underhill and Kivisild 2007). According to the RFLPs analysis, Kuwaiti maternal genetic structure is similar to other populations of Arabia in that they share a high incidence of haplogroup pre-HV (18%) and low incidence African haplogroup L (1.6%). These two haplogroups are unique to the maternal gene pool of Arabia (Abu-Amero et al. 2007; Al-Zahery et al. 2003). The European and Caucasian populations exhibit a high frequency of haplogroup H (45%) and very low frequency of haplogroups HV or pre-HV (when present) (Abu-Amero et al. 2007). However, the haplogroup frequency of pre-HV and HV occurs at 18.8% in the Arabians (Kivisild et al. 2003b), 20.6% in the Arabian Bedouins (Di Rienzo and Wilson 1991), 14.8% in Iraq (Al-Zahery et al. 2003), and in this study of a Kuwaiti sample the rate occurs at 23%, which is high compared to other haplogroups frequencies in the Arabian populations (Abu-Amero et al. 2007). Recent phylogeographic analyses of pre-HV haplogroup detected several radiations of this clade and several of its subclades in areas centered in the Middle East and Arabia and dated from the Neolithic and beyond. According to Richards *et al.* (2000, 2003) the haplogroup pre-HV has a Middle Eastern origin which eventually evolved into haplogroup H which is the most abundant haplogroup in Europe (45%) and the Near East (25%).

This study demonstrates the presence of the African haplogroup L (1.6%) in the Kuwaiti gene pool similar to the patterns observed in Saudi Arabian and Yemeni gene pools (Abu-Amero et al. 2007; Cerny et al. 2008). Approximately 2,500 years ago, the

slave trade from Africa to the Arabian Peninsula introduced maternal haplotypes to the Kuwaiti gene pool. The high frequency of African mt-DNA haplotypes in the Kuwaiti gene pool can be explained by the practice of assimilating the offspring of slaves into the community (Abu-Amero et al. 2007; Lewis 1990).

According to the mt-DNA HVS-I MDS plot in Figure 7, the Kuwaiti population clusters with Iranian and Saudi Arabian populations. This finding indicates that the maternal genetic structure of the Kuwaiti population resembles both the Iranian and Saudi Arabian maternal genetic structure. Furthermore, gene flow from Saudi Arabia and Iran to Kuwait is also likely. This finding is consistent with the findings of Alenizi *et al.* (2008) and Mohammad *et al.* (2009) who indicate that the paternal genetic structure of the Kuwaiti population is similar to the paternal genetic structure of Saudi Arabia, Iraq, and Iran. However, this analysis of maternal markers failed to support the presence of maternal gene flow from Iraq. This can be explained by the limited genetic contribution of Iraq in the Kuwaiti gene pool compared to Iran and Saudi Arabia and/or the small sample size used in this study which may not have included any individuals from Iraq.

The mismatch analysis indicates that the Kuwaiti population is a rapidly expanding population. This conclusion is supported by the demographic data (see Figure 3) and the smooth bell-shape distribution with raggedness index value below 0.03. Both Tajima's D statistic (-1.36306) and Fu's Fs (-10.25005) measurements have significant values. In contrast to Kuwait, the mismatch analyses of Saudi Arabia, Somalia, and African Bushmen show ragged shaped distributions indicating that historically these populations were of constant size over a long period of time.

The genetic affinities between Kuwait and the neighboring populations are supported by the archaeological artifacts sites which have been found in Kuwait, Saudi Arabia, Iran, Iraq, Bahrain, Qatar, and Bahrain. These archaeological artifacts are sherds of well-fired buff-ware with black geometric decorations dated to 5000 and 2500 B.C. These finding suggest a cultural exchange among these populations during the Ubaid period. Therefore, it is likely that migrations took place in the past between these regions (Potter 2009). These archaeological sites are located mostly in coastal areas (e.g. Kuwait H3) (Carter 2006; Potter 2009). The locations of these sites suggest that sailing was one form of migration and cultural contact at that time. In addition to sailing, certain traditional routes across the desert had been followed by the Kuwaiti to travel through the Arabian Peninsula using camels.

# Chapter Six: Conclusion

Kuwait is one of seven countries located in the Arabian Peninsula. The critical location of the Arabian Peninsula is believed to have played a major role in the expansion of early *Homo* out of Africa (Cabrera et al. 2009; Jeffrey and Michael 2009). To better understand its role in the early expansion, recent studies investigated the genetic structure and the archaeology of the Arabian Peninsula. The genetic structure of the population of the Arabian Peninsula was investigated through classical markers or genetic markers such as mt-DNA, Y-chromosome, or both (Abu-Amero et al. 2007; Al-Bustan et al. 2002; Rose 2007). However, to date, not all of the countries in the Arabian Peninsula have been investigated and this thesis tries to fill a gap in the understanding of the genetic structure of the Arabian Peninsula. This thesis investigates the genetic structure of Kuwait through mt-DNA analyses, using RFLPs and HVSI sequencing.

The results demonstrate that the maternal genetic structure of Kuwait resembles both Saudi Arabia and Iran. This finding is not supported by the neighbor-joining tree in which Kuwait clusters together with Iran and Iraq. However, the stress value indicates that the tree is not a good reflection of the original distance matrix. Based on a good fit stress value, the multidimensional scaling plot provides a more accurate relationship between Kuwait and neighboring populations. These results reflect the recent history of population formation in Kuwait. About 300 years ago, tribal groups migrated from Najd in Saudi Arabia and established the contemporary country of Kuwait (Alghanim 1998; Casey 2007). At that time, these tribal groups began trading networks with populations in Iraq, Iran, and other countries. In addition to trading, oil production in Kuwait attracted foreign laborers (Casey 2007). These two factors, trading and oil production, influenced

the contemporary genetic structure of the Kuwaiti population as demonstrated by this thesis.

Kuwait has high frequencies of mt-DNA haplogroup pre-HV and L. This research indicates that the haplogroup frequency in Kuwait is similar to other Arabian populations in that they share high frequencies of haplogroups pre-HV and L (Abu-Amero et al. 2007; Di Rienzo and Wilson 1991; Kivisild et al. 2003b). The presence of haplogroup L indicates past gene flow from Africa during the slave trading period and the Islamic expansion (Abu-Amero et al. 2007; Lewis 1990).

In sum, the Arabian Peninsula is a significant region in the history of modern humans in that it served and continues to serve as a cross-road between Africa, Asia, and Europe. Kuwait is one of the Arabian Peninsula countries in which the same haplogroup frequencies are shared with Arabia. In addition, Kuwait is influenced by gene flow from neighboring populations and especially from both Saudi Arabia and Iran. Further studies are needed to investigate the other haplogroup frequencies in Kuwait. In addition, the genetic study of Y-chromosome markers is needed to better understand the genetic structure and past history of Kuwait.

# Literature Cited


Abu-Amero KK, Gonzalez AM, Larruga JM, Bosley TM, and Cabrera VM. 2007. Eurasian and African mitochondrial DNA influences in the Saudi Arabian population. BMC Evol Biol 7:32.

Abu-Amero KK, Hellani A, Gonzalez AM, Larruga JM, Cabrera VM, and Underhill PA. 2009. Saudi Arabian Y-Chromosome diversity and its relationship with nearby regions. Bmc Genetics 10.

Abu-Amero KK, Larruga JM, Cabrera VM, and Gonzalez AM. 2008. Mitochondrial DNA structure in the Arabian Peninsula. Bmc Evolutionary Biology 8.

Al-Bustan S, El-Zawahri M, Al-Azmi D, and Al-Bashir AA. 2002. Allele frequencies and molecular genotyping of the ABO blood group system in a Kuwaiti population. International Journal of Hematology 75(2):147-153.

Al-Nassar KE, Conneally PM, Palmer CG, and Yu P. 1981a. The Genetic Structure of the Kuwaiti Population. I. Distribution of 17 markers with genetic distance analysis. Hum Genet 57(2):192-198.

Al-Nassar KE, Palmer CG, Conneally PM, and Yu PL. 1981b. The Genetic Structure of the Kuwaiti Population II: The distribution of Q-band chromosomal heteromorphisms. Hum Genet 57(4):423-427.

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. Molecular Phylogenetics and Evolution 28(3):458-472.

Alenizi M, Goodwin W, Ismael S, and Hadi S. 2008. STR data for the AmpFlSTR Identifiler loci in Kuwaiti population. Leg Med (Tokyo) 10(6):321-325.

Alghanim S. 1998. The reign of Mubarak al-Sabah, Shaikh of Kuwait, 1896-1915. London ; New York

New York, NY In the U.S.A. and Canada distributed by St. Martin's Press,: I.B. Tauris ;. xiii, 242 p. p.

Alshamali F, Pereira L, Budowle B, Poloni ES, and Currat M. 2009. Local Population Structure in Arabian Peninsula Revealed by Y-STR Diversity. Human Heredity 68(1):45-54.

Anderson S, Bankier AT, Barrell BG, Debruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F et al. . 1981. Sequence and Organization of the Human Mitochondrial Genome. Nature 290(5806):457-465.

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, and Howell N. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 23(2):147.

Anscombe FF. 1997. The Ottoman Gulf : the creation of Kuwait, Saudi Arabia, and Qatar. New York: Columbia University Press. xiv, 270 p. p.

Bailey G, Flemming N, King G, Lambeck K, Momber G, Moran L, Al-Sharekh A, and Vita-Finzi C. 2007. Coastlines, submerged landscapes, and human evolution: the Red Sea Basin and the Farasan Islands. The Journal of Island and Coastal Archaeology 2(2):127-160.

Bandelt H-J, Macaulay V, and Richards M. 2006. Human mitochondrial DNA and the evolution of homo sapiens. Berlin ; New York: Springer. xiv, 271 p. p.
Bar-Yosef O, and Belfer-Cohen A. 2001. From Africa to Eurasia - early dispersals. Quaternary International 75:19-28.
Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, and Wheeler DL. 2006. GenBank. Nucleic Acids Res 34(Database issue):D16-20.
Bermisheva MA, Viktorova TV, and Khusnutdinova EK. 2003. Polymorphism of human mitochondrial DNA. Russian Journal of Genetics 39(8):849-859.
Beyin A. 2006. The Bab al Mandab vs the Nile-Levant: an appraisal of the two dispersal routes for early modern humans out of Africa. African Archaeological Review 23(1):5-30.
Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, and Cavalli-Sforza LL. 1994. High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368(6470):455-457.
Brown WM. 1980. Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. Proc Natl Acad Sci U S A 77(6):3605-3609.
Budowle B, Wilson MR, DiZinno JA, Stauffer C, Fasano MA, Holland MM, and Monson KL. 1999. Mitochondrial DNA regions HVI and HVII population data. Forensic Science International 103(1):23-35.
Cabrera V, Abu-Amero K, Larruga J, and González A. 2009. The Arabian peninsula: gate for human migrations out of Africa or cul-de-sac? A mitochondrial DNA phylogeographic perspective. The evolution of human populations in Arabia: paleoenvironments, prehistory and genetics The Netherlands: Springer:79–87.
Cann RL, Stoneking M, and Wilson AC. 1987. Mitochondrial DNA and Human Evolution. Nature 325(6099):31-36.
Carter R. 2006. Boat remains and maritime trade in the Persian Gulf during the sixth and fifth millennia BC. ANTIQUITY-OXFORD- 80(307):52.
Carter R, Crawford H, Mellalieu S, and Barrett D. 1999. The Kuwait-British Archaeological Expedition to As-Sabiyah: report on the first season's work. Iraq 61:43-58.
Carter RW. 2007. Mitochondrial diversity within modern human populations. Nucleic Acids Research 35(9):3039-3045.
Casey MS. 2007. The history of Kuwait. Westport, Conn.: Greenwood Press. xv, 160 p. p.
Cavalli-Sforza L, and Feldman M. 2003. The application of molecular genetic approaches to the study of human evolution. Nature Genetics 33(3 s):266-275.
Cavalli-Sforza LL, Piazza A, Menozzi P, and Mountain J. 1988. Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data. Proc Natl Acad Sci U S A 85(16):6002-6006.
Cerny V, Mulligan CJ, Ridl J, Zaloudkova M, Edens CM, Hajek M, and Pereira L. 2008. Regional differences in the distribution of the sub-Saharan, West Eurasian, and South Asian mtDNA lineages in Yemen. Am J Phys Anthropol 136(2):128-137.
Crassard R. 2009. The Middle Paleolithic of Arabia: the view from the Hadramawt region, Yemen. In: Petraglia MD, and Rose JI, editors. The evolution of human populations in Arabia: paleoenvironments, prehistory and genetics The Netherlands: Springer: Springer. p 151–168.

Crawford MH. 2007. Anthropological genetics : theory, methods and applications. Cambridge, UK ; New York: Cambridge University Press. x, 476 p. p.
Crystal J. 1992. Kuwait : the transformation of an oil state. Boulder: Westview Press. xii, 194 p. p.
DeGiorgio M, Jakobsson M, and Rosenberg NA. 2009. Explaining worldwide patterns of human genetic variation using a coalescent-based serial founder model of migration outward from Africa. Proceedings of the National Academy of Sciences of the United States of America 106(38):16057-16062.
Derricourt R. 2005. Getting "Out of Africa": sea crossings, land crossings and culture in the Hominin migrations. Journal of World Prehistory 19(2):119-132.
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.
El-Zawahri MM, and Luqmani YA. 2008. Molecular genotyping and frequencies of A(1), A(2), B, O-1 and O-2 alleles of the ABO blood group system in a Kuwaiti population. International Journal of Hematology 87(3):303-309.
Ennafaa H, Cabrera VM, Abu-Amero KK, Gonzalez AM, Amor MB, Bouhaha R, Dzimiri N, Elgaaied AB, and Larruga JM. 2009. Mitochondrial DNA haplogroup H structure in North Africa. BMC Genet 10:8.
Excoffier L, Laval G, and Schneider S. 2005. Arlequin (version 3.0): An integrated software package for population genetics data analysis. Evol Bioinform Online 1:47-50.
Excoffier L, Smouse P, and Quattro J. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131(2):479.
Fu YX. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147(2):915-925.
Goldstein DB, and Chikhi L. 2002. Human migrations and population structure: What we know and why it matters. Annual Review of Genomics and Human Genetics 3:129-152.
Gonder MK, Mortensen HM, Reed FA, de Sousa A, and Tishkoff SA. 2007. Whole-mtDNA genome sequence analysis of ancient African lineages. Mol Biol Evol 24(3):757-768.
Gonzalez AM, Karadsheh N, Maca-Meyer N, Flores C, Cabrera VM, and Larruga JM. 2008. Mitochondrial DNA variation in Jordanians and their genetic relationship to other Middle East populations. Annals of Human Biology 35(2):212-231.
Gunz P, Bookstein FL, Mitteroecker P, Stadlmayr A, Seidler H, and Weber GW. 2009. Early modern human diversity suggests subdivided population structure and a complex out-of-Africa scenario. Proceedings of the National Academy of Sciences of the United States of America 106(15):6094-6098.
Hartl DL, and Clark AG. 2007. Principles of population genetics. Sunderland, Mass.: Sinauer Associates. xv, 652 p. p.
Hawksworth D. 1996. Biodiversity: measurement and estimation: Springer.
Hedges SB, Kumar S, Tamura K, and Stoneking M. 1992. Human origins and analysis of mitochondrial DNA sequences. Science 255(5045):737-739.
Horai S, Hayasaka K, Kondo R, Tsugane K, and Takahata N. 1995. Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial

DNAs. Proceedings of the National Academy of Sciences of the United States of America 92(2):532-536.
Ingman M, Kaessmann H, Paabo S, and Gyllensten U. 2000. Mitochondrial genome variation and the origin of modern humans. Nature 408(6813):708-713.
Jagher R. 2009. The Central Oman Palaeolithic Survey: recent research in Southern Arabia and reflection on the prehistoric evidence. In: MD P, and JI R, editors. The evolution of human populations in Arabia: paleoenvironments, prehistory and genetics The Netherlands: Springer: Springer. p 139–150.
Jeffrey IR, and Michael DP. 2009. Tracking the Origin and Evolution of Human Populations in Arabia.
Khoury PS, Kostiner J, Massachusetts Institute of Technology., and Harvard University. 1990. Tribes and state formation in the Middle East. Berkeley: University of California Press. xv, 351 p. p.
Kivisild T, Reidla M, Metspalu E, Rosa A, Brehm A, Pennarun E, Parik J, Geberhiwot T, Usanga E, and Villems R. 2004. Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. Am J Hum Genet 75(5):752-770.
Kivisild T, Rootsi S, Metspalu M, Mastana S, Kaldma K, Parik J, Metspalu E, Adojaan M, Tolk HV, Stepanov V et al. . 2003a. The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. American Journal of Human Genetics 72(2):313-332.
Kivisild T, Siiri Rootsi, Mait Metspalu, Ene Metspalu, Juri Parik, Katrin Kaldma, Esien Usanga, Sarabjit Mastana, Surinder S. Papiha, and Villems R. 2003b. The Genetics of Language and Farming Spread in India. In: Renfrew C, Boyle, K., editor. Examining the Farming/Language Dispersal Hypothesis. Cambridge: McDonald Institute for Archaeological Research. p 215-222.
Lahr MM. 1994. The multiregional model of modern human origins: a reassessment of its morphological basis. Journal of Human Evolution 26(1):23-56.
Lahr MM, and Foley R. 1994. Multiple dispersals and modern human origins. Evolutionary Anthropology: Issues, News, and Reviews 3(2):48-60.
Lewis B. 1990. Race and slavery in the Middle East: Oxford Univ. Pr.
Lienhardt P, and Al-Shahi A. 2001. Shaikhdoms of eastern Arabia. Houndmills, Basingstoke, Hamshire ; New York: Palgrave. xx, 257 p. p.
Luis JR, Rowold DJ, Regueiro M, Caeiro B, Cinnioglu 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. American Journal of Human Genetics 74(3):532-544.
McEvoy B, Richards M, Forster P, and Bradley DG. 2004. The Longue Duree of genetic ancestry: multiple genetic marker systems and Celtic origins on the Atlantic facade of Europe. Am J Hum Genet 75(4):693-702.
Miller SA, Dykes DD, and Polesky HF. 1988. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 16(3):1215.
Mithen S, and Reed M. 2002. Stepping out: a computer simulation of hominid dispersal from Africa. Journal of Human Evolution 43(4):433-462.
Mohammad T, Xue Y, Evison M, and Tyler-Smith C. 2009. Genetic structure of nomadic Bedouin from Kuwait. Heredity 103(5):425-433.

Nasidze I, Quinque D, Rahmani M, Alemohamad SA, and Stoneking M. 2008. Close genetic relationship between Semitic-speaking and Indo-European-speaking groups in Iran. Ann Hum Genet 72(Pt 2):241-252.
Nei M. 1992. Age of the common ancestor of human mitochondrial DNA. Mol Biol Evol 9(6):1176-1178.
Nei M. 1995. Genetic support for the out-of-Africa theory of human evolution. Proceedings of the National Academy of Sciences of the United States of America 92(15):6720-6722.
Olivieri A, Achilli A, Pala M, Battaglia V, Fornarino S, Al-Zahery N, Scozzari R, Cruciani F, Behar DM, Dugoujon JM et al. . 2006. The mtDNA legacy of the Levantine early Upper Palaeolithic in Africa. Science 314(5806):1767-1770.
Omkar D, Serafim B, Marcus WF, and Cavalli-Sforza LL. 2009. A serial founder effect model for human settlement out of Africa. Proceedings: Biological Sciences 276(1655):291-300.
Pérez-Miranda A, Alfonso-Sánchez M, Peña J, and Herrera R. 2006. Qatari DNA variation at a crossroad of human migrations. Human Heredity 61(2):67-79.
Petraglia MD. 2003. The Lower Paleolithic of the Arabian Peninsula: Occupations, adaptations, and dispersals. Journal of World Prehistory 17(2):141-179.
Potter LG. 2009. The Archaeology and Early History of the Persian Gulf. In: Potter LG, editor. The Persian Gulf in History. New York: Palgrave Macmillan. p 27-56.
Ramachandran S, Deshpande O, Roseman CC, Rosenberg NA, Feldman MW, and Cavalli-Sforza LL. 2005. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. Proceedings of the National Academy of Sciences of the United States of America 102(44):15942-15947.
Reich DE, and Goldstein DB. 1998. Genetic evidence for a Paleolithic human population expansion in Africa. Proceedings of the National Academy of Sciences of the United States of America 95(14):8119-8123.
Relethford JH. 1995. Genetics and modern human origins. Evolutionary Anthropology: Issues, News, and Reviews 4(2):53-63.
Relethford JH. 1998. Genetics of modern human origins and diversity. Annual Review of Anthropology 27:1-23.
Relethford JH. 2008. Genetic evidence and the modern human origins debate. Heredity 100(6):555-563.
Relethford JH, and Harpending HC. 1994. Craniometric variation, genetic theory, and modern human origins. American Journal of Physical Anthropology 95(3):249-270.
Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C, Sellitto D, Cruciani F, Kivisild T et al. . 2000. Tracing European founder lineages in the near eastern mtDNA pool. American Journal of Human Genetics 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. American Journal of Human Genetics 72(4):1058-1064.

Rogers AR, Fraley AE, Bamshad MJ, Watkins WS, and Jorde LB. 1996. Mitochondrial mismatch analysis is insensitive to the mutational process. Mol Biol Evol 13(7):895-902.
Rogers AR, and Harpending H. 1992. Population growth makes waves in the distribution of pairwise genetic differences. Mol Biol Evol 9(3):552-569.
Rohlf F. 2008. NTSYS-pc: Numerical Taxonomy System, version 2.1. Setauket, NY: Exeter Publishing Limited.
Rose J. 2007. The Arabian Corridor Migration Model: archaeological evidence for hominin dispersals into Oman during the Middle and Upper Pleistocene. p 1-19.
Rose JI, and Petraglia MD. 2009. Tracking the Origin and Evolution of Human Populations in Arabia. In: MD P, and JI R, editors. The evolution of human populations in Arabia: paleoenvironments, prehistory and genetics The Netherlands: Springer: Springer. p 1–12.
Saitou N, and Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4(4):406-425.
Santos C, Montiel R, Angles N, Lima M, Francalacci P, Malgosa A, Abade A, and Aluja MP. 2004. Determination of human Caucasian mitochondrial DNA haplogroups by means of a hierarchical approach. Human Biology 76(3):431-453.
Sawhney KS, Sunderland E, and Woolley V. 1984. Genetic Polymorphisms in the Kuwaiti Arabs. Human Heredity 34(5):303-307.
Scott-Jackson J, Scott-Jackson W, and Rose J. 2009. Paleolithic stone tool assemblages from Sharjah and Ras al Khaimah in the United Arab Emirates. In: Petraglia MD, and Rose JI, editors. The evolution of human populations in Arabia: paleoenvironments, prehistory and genetics The Netherlands: Springer: Springer. p 125–138.
Serre D, and Paabo S. 2004. Evidence for gradients of human genetic diversity within and among continents. Genome Res 14(9):1679-1685.
Sharma S, Saha A, Rai E, Bhat A, and Bamezai R. 2005. Human mtDNA hypervariable regions, HVR I and II, hint at deep common maternal founder and subsequent maternal gene flow in Indian population groups. J Hum Genet 50(10):497-506.
Slot B. 2003. Kuwait : the growth of a historic identity. London: Arabian. xiii, 129 p. p.
Stringer CB, and Andrews P. 1988. Genetic and Fossil Evidence for the Origin of Modern Humans. Science 239(4845):1263-1268.
Tajima F. 1993. Simple methods for testing the molecular evolutionary clock hypothesis. Genetics 135(2):599-607.
Takahata N, and Satta Y. 1997. Evolution of the primate lineage leading to modern humans: Phylogenetic and demographic inferences from DNA sequences. Proceedings of the National Academy of Sciences of the United States of America 94(9):4811-4815.
Templeton AR. 1992. Human origins and analysis of mitochondrial DNA sequences. Science 255(5045):737-737.
Thangaraj K, Sridhar V, Kivisild T, Reddy AG, Chaubey G, Singh VK, Kaur S, Agarawal P, Rai A, Gupta J et al. . 2005. Different population histories of the Mundari- and Mon-Khmer-speaking Austro-Asiatic tribes inferred from the mtDNA 9-bp deletion/insertion polymorphism in Indian populations. Hum Genet 116(6):507-517.

Thorne AG, and Wolpoff MH. 1992. The multiregional evolution of humans. Scientific American 266(4):76-&.

Tishkoff SA, Gonder MK, Henn BM, Mortensen H, Knight A, Gignoux C, Fernandopulle N, Lema G, Nyambo TB, Ramakrishnan U et al. . 2007. History of click-speaking populations of Africa inferred from mtDNA and Y chromosome genetic variation. Mol Biol Evol 24(10):2180-2195.

Torroni A, Huoponen K, Francalacci P, Petrozzi M, Morelli L, Scozzari R, Obinu D, Savontaus ML, and Wallace DC. 1996. Classification of European mtDNAs from an analysis three European populations. Genetics 144(4):1835-1850.

Torroni A, Lott MT, Cabell MF, Chen YS, Lavergne L, and Wallace DC. 1994. mtDNA and the origin of Caucasians: Identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-loop region. American Journal of Human Genetics 55(4):760-776.

Uerpmann H, Potts D, and Uerpmann M. 2009. Holocene (re-) occupation of Eastern Arabia. In: Petraglia MD, and Rose JI, editors. The evolution of human populations in Arabia: paleoenvironments, prehistory and genetics The Netherlands: Springer: Springer. p 205–214.

Underhill PA, and Kivisild T. 2007. Use of Y chromosome and mitochondrial DNA population structure in tracing human migrations. Annual Review of Genetics 41:539-564.

Vigilant L, Stoneking M, Harpending H, Hawkes K, and Wilson AC. 1991. African Populations and the Evolution of Human Mitochondrial DNA. Science 253(5027):1503-1507.

Watson E, Bauer K, Aman R, Weiss G, von Haeseler A, and Paabo S. 1996. mtDNA sequence diversity in Africa. American Journal of Human Genetics 59(2):437-444.

Weidenreich F. 1943. The skull of Sinanthropus pekinensis: a comparative study on a primitive hominid skull: Palaeontologia Sinica.

Wilson AC, and Cann RL. 1992. The Recent African Genesis of Humans. Scientific American 266(4):68-73.

Wolpoff MH, Hawks J, and Caspari R. 2000. Multiregional, not multiple origins. American Journal of Physical Anthropology 112(1):129-136.