The Genetic Structure of the Kuwaiti and Failaka Island Populations: Y-chromosome & Mitochondrial DNA Variation

By

Jasem Bader Theyab
M.A., University of Kansas, 2010

Submitted to the graduate degree program in Anthropology and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

_____
Chairperson, Dr. Michael H. Crawford

_____
Dr. Majid Hannoum

_____
Dr. Deborah Smith

_____
Dr. Bartholomew C. Dean

_____
Dr. John Kelly

Date Defended: May 28, 2013

The Dissertation Committee for Jasem Bader Theyab
certifies that this is the approved version of the following dissertation:

The Genetic Structure of the Kuwaiti and Failaka Island Populations: Y-chromosome & Mitochondrial DNA Variation

_____
Chairperson, Dr. Michael H. Crawford

Date approved: May 31, 2013

Abstract

Recent studies applying multidisciplinary approaches suggest that the Anatomically Modern Homo sapiens (AMHS) passed through the Arabian Peninsula in their major diaspora out of Africa. The Arabian Peninsula is connected to three continents: Africa, Asia, and Europe. In addition to the major diaspora, the Arabian Peninsula has witnessed numerous migrations among the three continents. The populations of the Arabian Peninsula have been investigated to better understand their evolutionary history. This dissertation investigated the paternal genetic structure of the Kuwaiti and Failaka Island populations using 15 loci Y-STR data. In addition, the maternal genetic structure of Failaka Island has been investigated using mtDNA HVS-I sequence data. This is the first genetic study to characterize Failaka Island population.

The result showed that the Kuwaiti population has a high frequency of Y-haplogroup J1 (37%) similar to other Arabian populations. the highest Y-haplogroup frequency in Failaka Island is J2 (26%). According to the MDS plot, the Kuwaiti population is clustered with neighboring populations, including Iran and Saudi Arabia, and Failaka Island. However, maternal and paternal genetic structures of Failaka Island are closely related to Kuwait and Iran than Saudi Arabia. The paternal and maternal genetic structures of Kuwait are highly diverse and experienced more gene flow compared to Failaka Island which has experienced genetic drift with limited gene flow. Finally, Mantel tests show insignificant correlation between genetic and geographic distance matrices.

This dissertation is dedicated to my mother, father, wife, lovely daughter Mariam, and the rest of my family for their support, encouragement, and patience

Acknowledgment

I would like to thank my academic father and mentor Prof. Michael H. Crawford who gave me unlimited support, guidance, and encouragement during my M.A. and PhD studies. It is easy to be admitted to any graduate program, but it is difficult to have such a mentor. I would like thank Dr. Suzanne Al-Bustan from Kuwait University for her support during my research, and for giving me access to her lab to do my analyses. I would especially like to thank the Kuwaiti and Failaka Island participants whose participation made this study possible.

I would like to thank Dr. Yaqoub Al- Kandari for his continues support and encouragement since my acceptance of a Kuwait University scholarship. I would like to thank Dr. Mohammad Alenizi, forensic genetics specialist at the Ministry of Interior, and the forensic technician Mr. Ahmed M. H. Saleh, Senior Medical Laboratory Practitioner, for analyzing the electropherogram data using GeneMapper software.

I would like to thank my committee members for their guidance and support throughout my PhD program: Prof. Michael H. Crawford, Dr. Majid Hannoum, Prof. Deborah Smith, Dr. Bartholomew C. Dean, and Prof. John Kelly ; my fellow students including: Dr. Kristin Young, Dr. Phillip Melton, Kristine Beaty, Dr. Norberto Baldi, Orion Graf, Stephen Johnson, Jacob Boyd and my other friends and fellow students in the Department of Anthropology and the Laboratory of Biological Antrhopology; the Department of Anthropology office staff: Le-Thu Erazmus, Carol Archinal, and Kathleen Womack.

I would like to thank Mrs. Mariam Al-Khamis at the Department of Cultural Relations of Kuwait University for answering all my questions regarding my scholarship. Finally, I am in debt to my mother, father, wife, daughter, 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, located in the Middle East, is the world's largest peninsula covering 3,237,500 km². The significant location of the Arabian Peninsula comes from the fact that it connects three continents Africa, Asia, and Europe. As a result, several migrations took place over the Arabian Peninsula. One of the most important migrations which shaped the evolutionary history of humans is the migration of the Anatomically Modern *Homo sapiens* (AMHS) out of Africa circa 100,000 years BP (Crawford, 2007; Reich & Goldstein, 1998). Humans occupied three continents after their diaspora out of Africa. However, chronology of migrations and whether it has one episode of one way migration or series of mini-migration through different directions is currently debated (Al-Zahery *et al*., 2003; Derricourt, 2005; Luis *et al*., 2004).

Humans' origin and migrations is an active topic in the field of anthropology and molecular genetics (Cann *et al*., 1987; Weidenreich, 1943). After years of investigations, two models have been proposed to explain the origin of modern human: Multiregional model and Recent Out-of-Africa model. The Multiregional model is based mostly on the morphological traits of the fossil remains (Coon, 1963; Howells, 1980; Lahr & Foley, 1994; Thorne & Wolpoff, 1992; Weidenreich, 1943; Wolpoff *et al*., 1984; Wolpoff *et al*., 2000). According to the fossil remains the advocates of this model believed that *H. erectus* migrated from Africa to Asia and Europe circa 1.5 mya (Stringer & Andrews, 1988). Over a long period of time, Neandertals and anatomically modern *H. sapiens* emerged as a result of *H. erectus* evolution (Wolpoff *et al*., 2000). Although this evolution occurred in different geographical areas independently, there was possible gene flow between regions (Wolpoff *et al*., 2000). The other model, the Recent Out-of-Africa,

is based on the mitochondrial DNA (mtDNA) and Y-chromosome analyses and markers which are relatively new and advanced techniques in the field of molecular biology (Cann *et al*., 1987; Stringer & Andrews, 1988; Templeton, 1992; Wilson & Cann, 1992). The advocates of this model point out that the African populations have the highest genetic diversity among all other populations and consequently they are the oldest population among other continental groups (Gunz *et al.*, 2009; Nei, 1995). The reason behind the high genetic diversity is that fact that the recombination is absent in both the mtDNA and the Non-recombining Region of the Y-chromosome (NRY). As a result, the mutations accumulate overtime and the populations with the longest evolutionary history have the highest genetic diversity (Bowcock *et al.*, 1994; Watson *et al.*, 1996). Today, the debate continues but scientists are aware about the fact that to investigate humans' origin and migration they should apply a multidisciplinary approach including archaeological, paleoenvironmental, linguistic, and genetic data (Cavalli-Sforza *et al.*, 1988)

Studying population genetics is usually connected to the genetic structure of a population. Each population has a unique genetic structure which is determined by the distributions of the genes. Generally, genetic structure of a population refers to frequencies and distribution of all the alleles in a population. There are several evolutionary forces that can change the genetic structure of a population including: gene flow, gene drift, and natural selection (Hartl & Clark, 2007). Earlier, the genetic structure of a population was investigated through determining the frequency of protein polymorphisms, blood group, and enzyme electrophoretic polymorphisms which are known as classical markers (Al-Nassar *et al*., 1981; Sawhney *et al*., 1984). Major shift occurred in the field of physical anthropology and especially sub-field anthropological genetics during the 80s due to the development of new molecular genetics techniques

(Cavalli-Sforza & Feldman, 2003; Crawford, 2007). The following are examples of these molecular techniques and discoveries which provide greater information about population structure and history:

(1) Rapid and efficient DNA extraction methods

(2) Restriction Fragment Length Polymorphisms (RFLPs) as a method of mutations detection.

(3) DNA Polymerase Chain Reaction (PCR) amplification techniques

(4) Automated DNA sequencing

(5) New markers were discovered including mitochondrial DNA sequence variation, Short Tandem Repeat Polymorphisms (STRPs), and Single Nucleotide Polymorphisms (SNPs).

In the 1980s, studies of human evolution, migration, and origin using DNA technology were initiated by Brown (1980). Through analyzing mtDNA, Brown (1980) found that each population has a specific pattern of restriction enzyme cleavage. The next major step in understanding humans origin was made by Cann *et al.* (1987) who concluded that the existence of modern humans in Africa was no more than 200,000 years ago. During the 1990s, many studies using Y-chromosome and mtDNA markers have focused on the origin and expansion of *Homo sapiens* (Hammer, 1995; Hammer *et al.*, 1998; Hammer et al. 1997; Hedges *et al.*, 1992; Nei 1992; Templeton, 1992). Most of these studies have investigated European, Americans, African, and Asian populations. However, the Middle Eastern populations were not investigated during this period of time because of multiple wars during the 1980s and 1990s.

Recently, the critical role of the Arabian Peninsula in the expansion of early anatomically modern *Homo sapiens* has attracted scientists to begin their investigation, tracking the first foot steps of *Homo sapiens* out of Africa (Cabrera *et al.*, 2009; Jeffrey & Michael, 2009). A multidisciplinary approach including archaeology and molecular genetics has been used 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). Today, the Arabian Peninsula is occupied by seven major clusters of populations represented by the following countries: Yemen, Saudi Arabia, United Arab Emirates, Qatar, Bahrain, Kuwait, and Jordan (Rose & Petraglia, 2009). Some of the populations from these countries were characterized either through genetic data, archaeological data, or both (Abu-Amero *et al*., 2009; Alshamali *et al*., 2009; Al-Zahery *et al*., 2003; Jeffrey & Michael, 2009; Pérez-Miranda *et al*., 2006). The Kuwaiti population has been investigated previously through autosomal STRs (Alenizi *et al.*, 2008), and mtDNA (Theyab *et al.*, 2012).

This dissertation investigates the paternal genetic structure of Kuwait using 15 loci STRs markers. In addition to Kuwait, the maternal and paternal genetic structures of Failaka Island are investigated using mtDNA sequence data and 15 STRs loci markers. Failaka Island has not been studied genetically previously. This dissertation attempts to fill a gap in understanding the evolutionary history of the Arabian Peninsula through the characterization of the genetic structures of Kuwait and Failaka Island. The aims of this dissertation are to:

(1) Characterize the paternal genetic structure of the Kuwaiti population through Y-chromosome analyses.

(2) Investigate the paternal and the maternal genetic structures of Failaka Island population through Y-chromosome analyses and mtDNA sequence data.

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

(4) Examine the relationship between Failaka Island and neighboring populations, including Kuwait.

Chapter Two: Literature Review

Since this dissertation investigates the genetic structure of Kuwait and Failaka Island populations, located in the Arabian Peninsula, it is appropriate to begin by investigating the following questions: What and who is an Arab ? At the first glance the question seems to be simple questions and consequently should have simple answers. However, this question is difficult to answer because Arabs occupy an area of about 13 million square kilometers which stretches from the Arabian Peninsula and includes the Fertile Crescent in Western Asia, and extends across North Africa ending at the Atlantic Ocean. That includes twenty two countries with approximately estimated population of 315 million (Lewis 1993; Tadmouri *et al.* 2006).

Arab leaders had defined an Arab as "Whoever lives in our country, speaks our language, is brought up in our culture and takes pride in our glory is one of us" (Lewis, 1993; Rodinson, 1981). Another definition was used by Sir Hamilton Gibb (who was a well-qualified Western source) he defined Arabs as "All those are Arabs for whom the central fact of history is the mission of Muhammad and the memory of the Arab Empire and who in addition cherish the Arabic tongue and its cultural heritage as their common possession". Both definitions share a cultural prospective, and the latter definition is linked to Islam. However, both definitions neglect the deep historical roots that Arabs shared prior to the rise of Islam (Lewis, 1993; Rodinson, 1981).

Although "Arab" is an ethnic term used to describe millions of people worldwide, it should not be linked solely to language, culture, and history which are limited by time and source. The Arabs should be examined and defined from a multidisciplinary perspective including: archaeological, historical, cultural and biological aspects (Lewis,

1993; Rodinson, 1981). This chapter examines the Arabs by using this multidisciplinary approach. All available tools are utilized, to better understand origins of the Arabs using a chronology based on archaeological investigations of Arabia and ending with the genetical dimension of the contemporary inhabitants of the Arabian Peninsula.

Archaeological Background:

The history of the Arabian Peninsula transcends the history of the Arabs and has witnessed several migrations of early *Homo* out of Africa to occupy the rest of the world. Recent archaeological excavations belong to the Middle Paleolithic, are taking place over the Arabian Peninsula to understand its history. However, two archeological sites were discovered in Saudi Arabia: Wadi Fatimah and Dawādmi. These sites contain abundant lithic resources are located in elevated areas, near springs or stream channels. These lithic tools indicative of occupation of the sites by early *Homo* approximately 250,000 kya. (Scott-Jackson *et al.*, 2009). In Oman, more than 350 archaeological sites containing approximately 1 million artifacts were recently discovered in the Huqf region. 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). In the United Arab Emirates, assemblages of artifacts that belong to the Middle Paleolithic were discovered in three different locations: Abu Dhabi, Sharjah, Ras al Khaimah. One of the most known archaeological site is Jebel Faya which is a 10 km long limestone mountain located directly south of the Straits of Hormuz. The site contains archaeological artifacts which transect the historical periods from the Iron Ages, to the Paleolithic period. These southern sites support the theory of the occupation of the Arabian Peninsula through the Indian Ocean rim approximately 60,000 years ago. The

southern route was an alternative route to the Sinai Peninsula route (Armitage *et al.*, 2011; Scott-Jackson *et al.*, 2009; Wahida *et al.*, 2009)

The archaeological artifacts described 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 latter hypothesis states that the earliest settlers of 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).

In Qatar, 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. Potter (2009) suggests 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. The earliest evidence of settlement near the Persian shore dates to 6000-5000 B.C., when 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).

Archaeological excavations have confirmed the existence of a connection between the Gulf region and Mesopotamia. Pottery sherds, diagnostic of the Ubaid period, 6000 B.C.E. to 4000 B.C.E., 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).

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 anatomically modern *Homo sapiens* (AMHS). Recently, the archeologists' attention has been directed toward the Arabian Peninsula, but 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).

History of the Arabs:

The origin of the word Arab

According to philologists the origin of the word “Arab” is ambiguous. Rodinson (1981) and Lewis (1993) believed that the term “Arab” is derived from Semitic root meaning west. This meaning was first used by the people of Mesopotamia to describe the people who lived west of the Euphrates valley. However, this explanation is questionable since the term has been used by the Arabs themselves to describe another group of Arabs who lived in different location relative to another. Another explanation links the term “Arab” to the concept of nomadism. The term “Arab” is linked to many Hebrew words such as “Arābhā” which means dark land or steppe land, “Erebh” which means mixed and unorganized, or “Ăbhar” which means to move or pass (Lewis, 1993; Rodinson, 1981).

The term “Arab” first appeared in an Assyrian inscription in 853 B.C.E. during the reign of the King Shalmaneser III. From that time the Arab had frequent references in Assyrian and Babylonian inscriptions with many derivatives such as Aribi, Arabu, and Urbi. Some of the later inscriptions are accompanied by illustration of the Aribi and their camels. These inscriptions may refer to the nomadic people who lived in north but not southern Arabia (Halm, 2007; Rodinson 1981).

The earliest mention of Arabs and their land in a classical reference is in Aeschylus, who mentions Arabia as a land from which the warriors with sharp-pointed spears came. The terms “Arab” and “Arabia” was extended by Herodotus and other Greek and Latin writers to include the entire Peninsula and all its inhabitants, and even the eastern desert of Egypt. ‘Saracen’ is another term used in Greek literature to describe a single desert tribe in the Sinai Peninsula. However, the term ‘Saracen’ was extended in

Greek, Latin, and Talmudic literature to describe nomads generally, and applied to all Muslims during Byzantium and medieval West (Lewis, 1993; Rodinson, 1981).

The first mention of the term "Arab" by the Arabs themselves was in ancient remnants of southern civilization in Yemen by the southern branch of the Arabian people. In these inscriptions Arab means Bedouin, usually applied to distinguish nomadic (northern) from the sedentary population (southern) (Lewis, 1993; Rodinson, 1981).

Arabia before Islam

The original territory of Arab people was the Arabian Peninsula which is a rectangle of one and a quarter million square miles area. In the north, it is bordered by several countries called the Fertile Crescent which includes Syria, Palestine, Jordan, and Iraq. The Arabian Peninsula is surrounded by the Arabian Gulf in the east, Indian Ocean in the south, and the Red Sea in the west. The majority of the land consists of waterless steppes and desert except for Yemen which includes of mountains with rich of water resources and has led to the rise of agriculture and relatively advanced civilization. Most of the population was nomadic and pastoral living by its flocks (Hitti, 2002; Mansfield, 1992; Rodinson, 1981).

The early history of Arabia is obscure and not fully understood. However, there are a few hypotheses that have been proposed by historians to clarify this early history of Arabia. The first best-known theory is the Winckler-Caetani theory which is named after its two most distinguished proponents Hugo Winckler and Leone Caetani. According to Winckler-Caetani theory, Arabia was the first home of the Semitic peoples and originally a land of great fertility. Through the millennia Arabia experienced steady desiccation which led to drying waterways and subsequently the spread of the desert at the expense

of the cultivable land. The declining productivity, accompanied by the increase in the number of the inhabitants, led to a cycle of invasions from the neighboring countries by the Semitic peoples of the peninsula. Assyrians, Aramaeans, Canaanites (including the Phoenicians and Hebrews), and finally the Arabs entered the Fertile Crescent through this cycle of invasions (Hitti, 2002; Mansfield, 1992; Rodinson, 1981).

There is no systematic geological study in Arabia to evaluate the Winckler-Caetani theory. However, some evidence (such as the presence of the dried-up water ways and some indications of past fertility in the Arabian Peninsula) supports the theory. There is no evidence on the chronology of the desiccation and its effect on human affairs if it happened during the beginning of human life. Philological evidence supports the Winckler-Caetani theory in that the Arabic language, the most recent of the Semitic languages, has the oldest grammatical structure among other Semitic languages which led the philologists to presume that it is the original proto-Semitic tongue. Another explanation proposed that southern Mesopotamia was the homeland of the Smites. The last explanation suggested Africa or Armenia to be the homeland of Semites. However, all these theories are barely supported by geological or philological evidence (Hitti, 2002; Mansfield, 1992; Rodinson, 1981).

The chronology of the history of the Arabian Peninsula is highly complex. The Arabian people are known to be descendents of two branches-- the southern region and the northern-central regions. The northern and central tribes were descendents of one ancestor, who was Adnan, descendent of Ishmael, son of Abraham and were known as “Arabized Arabs”. The southern Arabs were supposed to be descendents of Qahtan who was identified with the biblical Yoqtan, direct descendent of Shem, son of Noah. They were known as “real Arabs”. The southern branch, well known for their flourishing

civilization, did not regard themselves as Arabs. They were mostly farmers and city-inhabitants. One explanation proposed that the southern Arabs applied the word Arab to the nomadic tribes of the central and northern portions of the Peninsula who spoke Arabic or proto-Arabic (Hitti, 2002; Mansfield, 1992; Rodinson, 1981).

Although southern Arabian history is obscure, there were many kingdoms established in southern Arabia and the best known in ancient records was the kingdom of Saba (Sheba). According to the Biblical sources, the queen of the kingdom of Saba entered into relations with King Solomon. The existence of the Saba can be traced back to the tenth century B.C.E. In 750 B.C.E., one of the Sabean kings built the Ma'rib dam which was used to regulate the agriculture life of the Saba Kingdom. The Sabeans had colonized the geographically nearest African coastlands and established the Kingdom of Abyssinia (Habasha). The Kingdom of Abyssinia was founded in the geographical location of current country Ethiopia (Lewis, 1993; Mansfield, 1992).

The history of northern and central Arabia was limited, when compared to that of southern Arabia. Most of the Assyrian, Biblical, and Persian resources provide little information regarding the nomadic peoples of Arabia. However, during the Hellenistic period, more detailed information has been recorded about the northern and central territories. Semi-sedentarized states were formed in the Syrian and northern desert during the Hellenistic period due to the western interest in the Arabian trade-route. These states were influenced by the Hellenized Aramaic culture and particularly their Aramaic language which appeared through their inscriptions. The Arabic characteristics of these states have vanished, except for their names which identify their Arabic origin. The best known of these people were the Nabataeans, who ruled an area surrounded by the Gulf of Aqaba (northern tip of the Red Sea) in the south, the Dead Sea in the north, and area of

the northern Hijāz (Saudi Arabia). Their capital city Petra was located east of the Dead Sea (Lewis, 1993; Mansfield, 1992).

During the third and the fourth centuries, the Arabian Peninsula witnessed a long series of wars between the Roman and Persian Empires which influenced the Arabian trade-routes and consequently the life of people especially in the southern part. In 384 C.E., a peace agreement between these two empires ended the wars. At that time the international trade routes returned to utilize direct routes through Egypt, the Red Sea, the Euphrates Valley, and the Arabian Gulf rather than the west Arabian trade-route which was abandoned due to its hazards. The situation in Arabia deteriorated when the southern civilization fell under foreign ruler and the Ma'rib dam was destroyed. That led the southern tribes to migrate toward Syria in the north, Ethiopia in the African horn, and AlHijāz in the Arabian Peninsula to search for new territories and to start a new life. The situation in the northern part was not much better than in the southern part. The established northern states reverted to nomadic society and nomadism spread across Arabia at the expense of cultivation and trade. Bedouin tribalism was the dominant characteristic of the northern and southern populations in which the group is the social unit rather than the individual. The relationship among the individuals is controlled by the blood-tie of the descent. The male line is considered the social bond through which the group provides the defense against the dangers of desert life (Hitti, 2002; Lewis, 1993; Rodinson 1981).

Arabia and the rise of Islam

During the pre-Islamic times the religion of Arabia was polydaemonism, which is belief and worship of evil spirit and power. However, Judaism and Christianity reached western Arabia from the north and the south. Yemen had been ruled by a Jewish dynasty followed by a Christian Abyssinian governor. In Yathrib (Medina) three of the five Arab tribes were Jewish. Christianity was introduced to Najran in the fifth century and formed a Christian community led by a bishop (Lewis, 1993).

Islam is an integrating element of the Arab world which was introduced into Arabia and the world through the prophet Muhammad and his followers. The prophet Muhammad was born around 570 in Mecca as a member of the Hashim clan in the Quraysh tribe. Around 610, the prophet Muhammad had received his revelation which introduced the monotheistic faith and rejected the ancient polytheistic religions of Arabia. Islam has five primary obligations, or pillars of faith, that each Muslim must fulfill in his or her lifetime and they are the framework of the Muslim life. These pillars are the testimony of faith, prayer, giving *zakat* (support of the needy), fasting during the month of Ramadan, and the pilgrimage to Makkah (Mecca) once in a lifetime for those who are able. In 622, Muhammad established the original Muslim community (umma) in Medina which was open to all tribes and clans. To reduce the social differences within the umma, the Islam had introduced the religion instead of the blood as a social bond among the umma. Within ten years, the Islamic umma expanded throughout the entire Arabian Peninsula with preliminary structures for a state. Many tribes joined the community and the umma became more powerful than before. In 632, the prophet Muhammad died and the entire Arabian Peninsula was associated with the umma and almost all Arabs were united in Islam. The death of the prophet Muhammad was followed by an Islamic

meeting by his original followers to select the new caliph (successor) of the umma. The fist caliph was Abu Bakr (632-634), followed by Umar (634-644), followed by Uthman (644-656), and finally Ali (656-661) who was Muhammad's cousin and son-in-law (Lewis, 1993; Mansfield, 1992; Rodinson, 1981).

In the 17th century, two great kingdoms existed and usually opposed each other: the Roman-Byzantine Empire in the west and the Persian Empire in the east. The Islamic umma arose as a new power on the political stage. The Islamic conquests started under the second caliph Umar, and quickly led to the defeat of the Byzantine Empire and the Persian-Sassanid Mesopotamia. Lewis (1993) and Rodinson (1981) believe that the rapid Islamic expansion had profound demographic affects. During the regime of the Umayyad Caliphate (661-750), the Islamic empire stretched from the borders of China and India in west Asia, Central Asia, the Middle East, North Africa, Sicily, and the Iberian Peninsula. Through the Islamic expansion, a large number of individuals from multiple ethnicities such as Persians, Chinese, and Africans converted to Islam. Islam became the religion of Arabs and non-Arabs alike, and the Arab elements diminished in importance as non-Arab cultures, particularly Persian, Indian, and Greek, contributed to the emergence of a new Islamic civilization (Hitti, 2002; Lewis, 1993; Rodinson, 1981).

Today, multiple sources have estimated that 1.2 to 1.57 billion Muslims of different ethnicities populate the world. However, most of the Arabs are inhabitants of the Arabian Peninsula-- a geographical area with 2.5 million km^2 landmass represented today by the following countries: Yemen, Oman, United Arab Emirates, Saudi Arabia, Qatar, Bahrain, Kuwait, and Jordan (Rose & Petraglia, 2009).

Occupation of Arabia: Genetic evidence

The Arabian Peninsula links three continents, Africa, Asia, and Europe. The location of the Arabian Peninsula played a major role in the dispersal of early *Homo sapiens* out of Africa and consequently in shaping the genetic structure of Arabia (Pérez-Miranda *et al.*, 2006; Petraglia, 2003; Rose, 2007). Studies of the Arabian Peninsula populations reveal that there are distinguishing genetic characteristics in the people living today in the area (Al-Zahery *et al.*, 2003).

There are two models to explain the early human migration out of Africa and consequently the occupation of Arabia, the closest land to Africa. These models are known as Out-of-Africa and Multiregional Origin models. Although the Recent Out-of-Africa model has been scientifically supported, questions regarding the timing, routes, and geographical destinations are still being debated. Recent studies have proposed four routes of human expansion out of Africa (Derricourt, 2005; Rose, 2007; Rose & Petraglia, 2009). The most widely accepted route is through 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 (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. The two routes that connect Africa to the Arabian Peninsula

Using either Sinai Peninsula or the Strait of Bab el-Mandab, anatomically modern *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.

The Sinai Peninsula is the only land bridge that connects Africa and Eurasia. However, there were various geographical and ecological obstacles in the Sinai Peninsula 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 (Derricourt, 2005). Recent genetic investigations are suggestive that both routes have been used by the early *Homo sapiens*

on their way out of Africa (Lawler, 2011; Petraglia *et al.*, 2010; Rose, 2007). The next section provides general information regarding the molecular markers, followed by investigation of the genetic structure of Arabia.

Mitochondrial DNA (mtDNA)

Analysis of mitochondrial DNA (mtDNA) 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. mtDNA is maternally inherited (uniparental) 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 mtDNA 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 separate from nuclear DNA. Each cell in the human body contains 10 to 100 mitochondrial compartments. Mitochondrial genomes vary within each mitochondrial compartment 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 16,589 base pairs (bp). The mtDNA 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 mitochondrial genome.

The mtDNA 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 mtDNA control region is divided into 3 regions, hypervariable region I (HVS-I), hypervariable region II (HVS-II), and hypervariable region III (HVS-III) (Figure 2). 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).

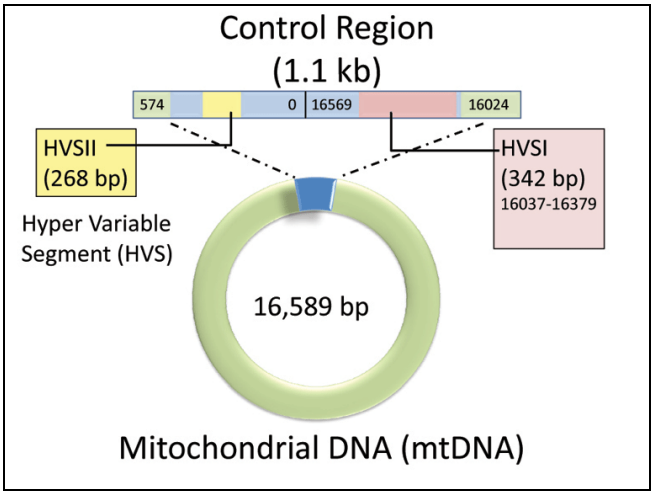

Figure 2. Mitochondrial DNA genome with hypervariable I and II

In 1987, Cann *et al.* presented data that indicated a recent African origin of modern humans between 140 and 280 thousand years ago. This study was based on restriction fragment length polymorphism (RFLPs) variation among the mtDNA of 147 humans from five geographic populations, including Africa, Asia, Australia, Caucasus, 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). In addition to the RFLP technique, sequencing the control region for use in population

genetics has been utilized because of the high degree of polymorphism that can be found in this relatively short sequence.

Y-chromosome

The Y-chromosome is a unique genetic characteristic of males with largest (95%) nonrecombining portion (NYR) of the human genome that can be used to study paternal human evolution, migration, and genetic processes (Hammer & Zegura, 2002). The chromatin of the Y-chromosome is found in three functionally different forms including: Pseudoautosomal regions (PARs), Euchromatin, and Heterochromatin. The PARs are located in the telomeric regions of the chromosome. During male meiosis, PARs pair and recombine with the X-chromosome. The euchromatin, consists of the functional genes, and is the sex-determining region of the Y-chromosome (SRY gene). Figure 3 shows different regions of the human Y-chromosome.

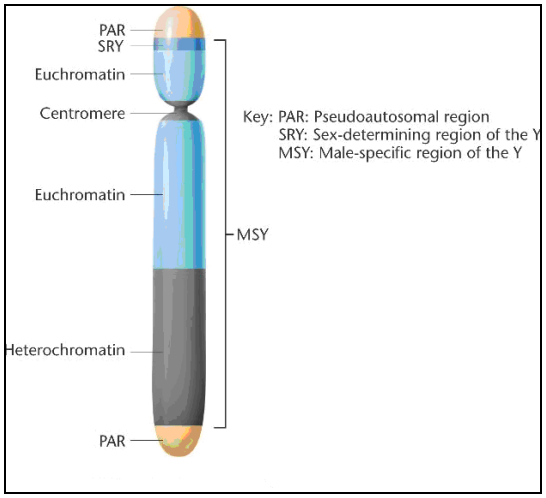

Figure 3. Different regions of the human Y-chromosome

The Y-chromosome preserves a unique record of mutational events and consists of about 60 million bases of DNA, most of which do not code for protein (Novelletto, 2007). The NYR region of the Y-chromosome comprises two types of Markers: slow mutating Single Nucleotide Polymorphisms marker (SNPs) and faster mutating short

tandem repeats (STRs). The mutation rate of the SNPs is 2 X 10^{-8} per base per generation which is slower than the STRs rate 1.2 X 10^{-3} per locus per generation (Crawford, 2007).

The STRs are blocks of 2-6 base pair tandem repeat units that are scattered throughout the genome, abundant, and highly polymorphic. They usually have multiple alleles and demonstrate high levels of heterozygosity of approximately 70% making them highly informative for genetic analysis and the inference of phylogenies among populations (Schlotterer, 2000; White *et al.*, 1999). The STRs account for approximately 3% of the human genome, have higher mutation rate than other parts of the Y-chromosome, and exhibit high variability in a population making them a useful tool to track the recent genetic events (Ellegren, 2004). There are thousands of genomic Y-STR markers that are used to characterize populations. They are very useful due to their high heterozygosity that comes from having 3 or more alleles of the same marker. Combining several Y-STR markers can yield a motif that can be tested in a population (Ellegren, 2004).

SNPs are the most common form of the binary markers which have a lower mutation rate that allow the reconstruction of the ancestral history and preserve the population specific haplotype information (Underhill *et al.*, 1997). In addition to SNPs, the insertion /deletions (indels) events at specific sites on the Y-chromosome are another form of binary markers. These markers are known as binary because the single base changes occur in two possibilities of alleles at defined position of the DNA sequence. In general, binary markers represent unique event polymorphisms in human evolution and they are useful in the identification of deep splits in the Y-chromosome genealogy (Crawford, 2007; Mills *et al.*, 2006).

In the 1990s, the application of Denaturing High Performance Liquid Chromatography (DHPLC) led to the increase in the number of the identified binary markers to more than 200. These markers contributed positively in understanding the genetic structure, variation, and history of multiple populations. However, the nomenclature of the binary markers has been inconsistent depending on the research groups and their method of reporting their results. Sometimes the same haplogroups of the Y-SNPs have different names, which caused confusion in reporting and understanding the genetics of the different populations around the world. In 2002, the nomenclature was unified by the Y-chromosome Consortium (YCC). The different haplogroups of Y-SNP were defined and named according to a system that allows easy incorporation of newly discovered mutations. In 2008, the Y-chromosome Consortium nomenclature was updated again and resulted in a total of 311 haplogroups with 600 binary markers (Karafet *et al.*, 2008).

The paternal haplogroups are identified by mutations at a given biallelic locus (SNPs or indels). Currently, there are 20 NRY haplogroups characterized by 20 capital English letters (A-T), and the primary SNP or indel that defines it (e.g., E-M81) (Y-chromosome Consortium 2002). While the M mutations indicate those identified by Peter Underhill's laboratory, the P mutations refer to mutations identified by Michael Hammer's laboratory. However, most of the P mutations from 123 to 297 were identified by different studies that published their mutations in public database (Karafet *et al.*, 2008).

Genetic Structure of Arabia

The presence of the anatomically modern *Homo sapiens* in the Arabian Peninsula has been traced genetically through both mtDNA and Y-chromosome analyses (Quintana-Murci *et al.*, 1999). The genetic analyses indicate that the early *Homo sapiens* had crossed the Arabian Peninsula through the Levant area and the Horn of African. Recent mtDNA studies indicate that the mtDNA haplogroup R0a (previously known as preHV) is presents at high frequency in the Arabian Peninsula with the highest frequency of R0a reported to date in western Yemen (25.6%). The high frequency of haplogroup R0a accompanied with the estimated coalescence time of about 19 ± 7.0 ka suggest it has a pre-Holocene occurrence in Yemen and followed by its appearance in Arabia and the Middle East. The presence of high frequency of R0a haplogroup and the African haplogroups are usually used to distinguish the Arabian Peninsula populations from European and Caucasus populations. Recent studies report the haplogroup frequencies of R0a and HV which occur at 18.8% in the Arabians (Kivisild *et al.*, 2003), 20.6% in the Arabian Bedouins (Di Rienzo & Wilson, 1991), 14.8% in Iraq (Al-Zahery *et al.*, 2003), and 23% in Kuwait (Theyab *et al.*, 2012). Recent phylogeographic analyses of R0a 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 era. According to Richards *et al.* (2000, 2003) the haplogroup R0a 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%).

The mtDNA analyses and more precisely the phylogenetic distribution of haplogroups M and N in southern Arabia and southwestern Asia support the early exodus out of Africa through Arabia. Both haplogroups M and N share similar founder ages and

they are derived from the African haplogroup L3. However, the greatest diversity among the M lineages is reported from India. It has been hypothesized that both lineages, M and N, either occurred in East Africa or were carried by early hominins during their expansion out of Africa. These genetic analyses suggest that Southern Arabia served as a connection area, connecting Africa and West Asia and subsequently South Asia in the first dispersal out of Africa. In addition to out-of-Africa dispersal, the Arabian Peninsula witnessed early migrations back to Africa. The mtDNA analyses in Arabia indicate that the M haplogroup is present in low frequencies, ranging from approximately 2% to 7% in Yemen and Kuwait (Cerny *et al.*, 2008; Theyab *et al.*, 2012). However, this haplogroup is represented mainly by recently derived haplotypes which are similar to those found in India, suggesting a recent Indian gene flow. This hypothesis supports the expansion along the Indian Ocean rim through the Strait of Bab el-Mandab during the pre-Holocene, and followed by a west Eurasian expansion derived from a radiation from India (Petraglia *et al.*, 2010). Figure 4 presents a map of the proposed route of migrations out of Africa.

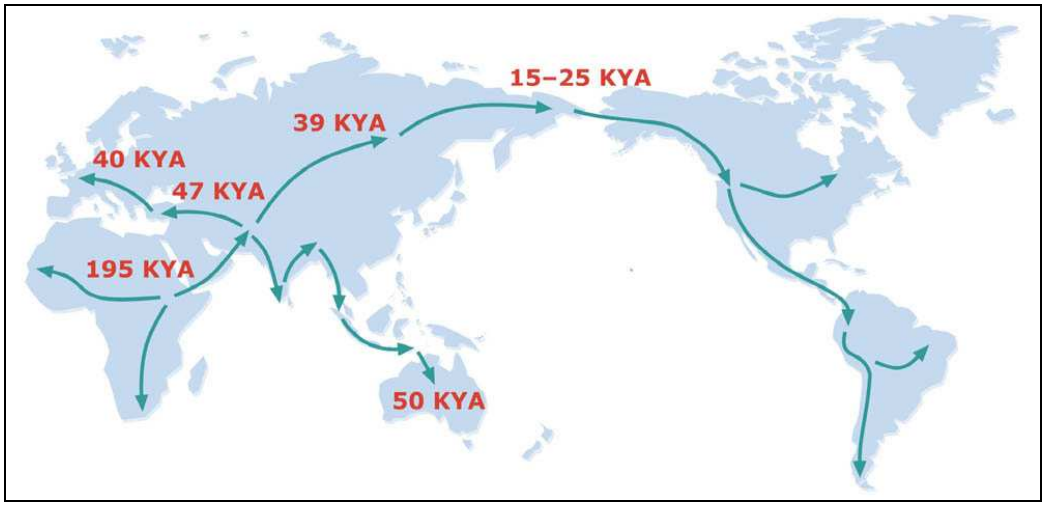

Figure 4. Map of one of the proposed out of Africa migration routs of anatomically modern *H. Sapiens* (http://www.sanger.ac.uk/research/projects/humanevolution/)

Although the previously discussed distribution is the standard mtDNA structure in the Arabian Peninsula, there are some unique mtDNA structures. Al-Zahery *et al.* (2003) have detected the presence of the European haplogroups U and V in Iraq with frequency of 20% and 0.5%. These results support the presence of the European gene flow in Iraq. In Jordan, a recent mtDNA study examined the population structure of two areas Amman (urban) and the Dead Sea (isolated). Although the genetic structure of the Jordanian subpopulation from Amman is similar to their surrounding populations, the inhabitants of the Dead Sea area showed a high frequency (39%) of Eurasian haplogroup U3, lacked the haplogroup R0a, and exhibited low frequencies of Neolithic mtDNA haplogroups J and T. The frequency of haplogroup U3 is highest in the Middle East. However, similar frequency has been reported in Northern Cameroon. The high frequency in both areas can be explained as a result of genetic isolation following an early human expansion.

The paternal genetic structure of the Arabian Peninsula is similar to the maternal structure in that both support the hypothesis that the Arabian Peninsula was inhabited by early *Homo sapiens*. However, the Y-chromosome STRs analyses show low genetic diversity compared to European populations. The finding can be explained by kin migration and the nonrandom mating by mean of consanguinity marriage. Consanguineous marriages, particularly among first cousins, are common in the Arabian Peninsula due to cultural beliefs. Multiple studies indicate the rate of consanguineous marriages in the Arabian Peninsula ranges from 35.9% to 57.7% (Cadenas *et al.*, 2008).

The Arabian Y-chromosome can be characterized by four major haplogroups: J, E, R, and T (Al-Zahery *et al.*, 2003). Haplogroup J with its subclades J1-M267 and J2-M172 are the most abundant in the Arabian Peninsula with frequencies 58% in Saudi Arabia (Abu-Amero *et al.*, 2009), 58.3% in Iraq (Al-Zahery *et al.*, 2003), 66.7% in

Qatar, 45.1% in United Arab Emirates (UAE), and 82.2% in Yemen (Cadenas *et al.*, 2008). Although J haplogroup is the most abundant in the Arabian Peninsula, its subclades show distinctive frequency distributions. While the frequency distribution of J1-M267 subclade increases in southern Arabia and decreases toward the North, the subclade J2-M172 increases in the Levant area and decreases toward the South. The recent phylogenetic analyses indicate that the J1-M267 lineage was brought to the Arabian Peninsula through the founder effect associated with small groups migrating from the Levant during the Neolithic period. Another ancestral population with J2a-M410 clade remained in the Levant and is associated with farming economy (Chiaroni *et al.*, 2010).

Haplogroup E is another Y-chromosome haplogroup identified in the Arabian Peninsula. Haplogroup E is widely distributed across the Middle East, Africa, and the Mediterranean basin. Previously, it had been hypothesized that the haplogroup E has Middle Eastern origin and was brought to Europe during the Neolithic expansion (Hammer *et al.*, 1998; Semino *et al.*, 2000). However, a recent study proposed that haplogroup E has African origin since its basal lineages are found among African pygmies and Bantu-speaking populations (Berniell-Lee *et al.*, 2009). It is difficult to interpret and draw conclusions from this geographically widespread haplogroup. Recent studies identified the presence of E1b1a, previously known as E3a-M2, which is defined by M2 mutation, lineage in Oman 7.4% (Luis *et al.*, 2004), Qatar 23.8% (Cadenas *et al.*, 2008), Yemen 3.2% (Cadenas *et al.*, 2008), UAE 5.5% (Cadenas *et al.*, 2008), Saudi Arabia 3.8% (Abu-Amero *et al.*, 2009), and Iraq 0.99% (Al-Zahery *et al.*, 2003). The presence of this lineage in the Arabian Peninsula is an indication of ancient gene flow from East Africa.

The distribution of the E1b1b1 (E-M35), previously known as E3b1-M35, and its derivatives, E1b1b1c (E-M123) and E1b1b1c1 (E-M34), in Yemen, Qatar and UAE, suggests that these haplotypes arrived in Arabia more likely through the Levantine corridor rather than through the Strait of Bab el-Mandab. This route follows the same pattern in which several mtDNA haplogroups have been introduced into the Arabian Peninsula during the Neolithic period (Cadenas *et al.*, 2008). The Near East is another possible place of origin of the E1b1b1c (E-M123) lineage due to it is high frequency in the area compared to it is Eastern African presence restricted to Ethiopia (Abu-Amero *et al.*, 2009; Semino *et al.*, 2004). In general, the frequency distribution of the E1b1b1c (E-M123) lineage in the Arabian Peninsula supports it is arrival along two routes: the Strait of Bab el-Mandab and the Levant and forming independent isolates. However, Flores *et al.* (2005) found that the highest frequency of the E1b1b1c1 (E-M34) to date is in a Dead Sea population sample from Jordan (31%). This result is similar to that found in Ethiopian sample (11%) and can be explained as a result of isolation and genetic drift effects (Abu-Amero *et al.*, 2009; Cruciani *et al.*, 2004). In Saudi Arabia, two individuals carrying the basal lineage E-M96 have been identified. This finding supports the proposed scenario in which early unipaternal human haplogroups such as DE-YAP evolved out side of Africa. The Levant and the Southern Arabia are the most likely regions in which the diversification of non-African haplogroup took place. One DE-YAP ancestor possibly spread into Asia and evolved into haplogroup D and another DE-YAP ancestor returned to Africa and evolved into haplogroup E (Abu-Amero *et al.*, 2009).

The ancient haplogroup R is defined by the M173 mutation and its highest frequency is in Eurasia (> 50%), especially subclades R1a1-M17 and R1b1b2-M269, are of relatively high frequencies when compared to other haplogroups in the Arabian

Peninsula. The origin of these widely distributed haplogroups in Europe and Asia are controversial. However, recent studies support the southwest Asian region rather than Central Asia as the origin due to the genetic diversity of Y-STR haplotype variation in R1 and R2 clades (Underhill *et al.*, 2010). The presence of these haplogroups in Arabia is probably due to back migration by an early ancestral population to Africa. The frequency distribution of the R-M17 in the Arabian Peninsula is 5.1% in Saudi Arabia (Abu-Amero *et al.*, 2009), 6.94% in Qatar (Cadenas *et al.*, 2008), 7.36% in UAE (Cadenas *et al.*, 2008), 9% in Oman (Luis *et al.*, 2004), 6.9% in Iraq (Al-Zahery *et al.*, 2003), and 1.37% in Jordan (Luis *et al.*, 2004). On the other hand, the highest reported subclade frequency R-M269 in the Arabian Peninsula is in Iraq (9.85%) and the rest of the region shows a frequency range from 1% to 4%.

The next haplogroup that has been identified in the Arabian Peninsula is haplogroup T, formerly known as K-M70 (K2). This haplogroup originated in Asia following the emergence of the K-M9 haplogroup (45–30 ky) (Underhill *et al.*, 2001). The haplogroup T present in the Arabian Peninsula resulted from the migration of haplogroup T individuals to Africa. The distribution frequency of the haplogroup T reported in the Arabian Peninsula which was 4.91% in UAE (Cadenas *et al.*, 2008), 5.1% in Saudi Arabia (Abu-Amero *et al.*, 2009), 5.91% in Iraq (Al-Zahery *et al.*, 2003), and 8% in Oman (Luis *et al.*, 2004). The Omani frequency is comparable to what has been found in Somalia (10.4%) (Sanchez *et al.*, 2005).

The Y chromosome variation in the Arabian Peninsula is not limited to those four haplogroups discussed previously. However, there are several haplogroups in the Arabian Peninsula but with low frequency ranging from 0.49% to 3%. These haplogroups have

not been discussed because such low frequencies would not contribute significantly to the Arabian genetic structure. Examples of these haplogroups are B, C, F, G, L, and Q.

Archaeology of Kuwait:

Although most of the archaeological sites in Kuwait found to date are located in the Failaka islands, recent excavations discovered the oldest boat model yet identified with related pottery remains (Figure 5). 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 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.

Figure 5. The boat model in the middle and related archaeological remains

Failaka (also known by Failakah or Faylakah, and locally known by the names Feileche, Feiliche or Feliche) is the only inhabited island off the coast of Kuwait, located 12.4 mile (20 km) from Kuwait. In antiquity, Failaka was named *Ikaros* by the Greek geographer Strabo and historian Arrian in their description about the Arabian Gulf

(Rajab, 1999). During 1970s, many names have been given to Failaka Island in early maps such as Ichara, Ilha de aquada, and I-Peleche (Rajab 1999). I-Peleche is the closest pronunciation to the local name of Failaka Island which is Feileche (Rajab, 1999).

Failaka Island has attracted the attention of archaeologists since 1958 when Danish archaeologists examined many archaeological artifacts from Failaka Island such as temple, stamp seals, and slabs (Figure 6) (Rajab, 1999). According to their findings, the history of Failaka Island is tied to the beginning of the second millennium B.C.E and extended over to the Bronze Age when the Dilmun civilization established a trade and/or a military station in Failaka Island (Casey, 2007; Rajab, 1999).

Figure 6. Temple of Artemis in Failaka Island

Most of the archaeological sites found in Failaka Island are associated with the Dilmun civilization. Tell Sa'ad wa Sa'id (F6), Al-Khidr, and Al-Awazim are the major archaeological sites in Failaka Island which are dated to the Bronze Age (Figure 7). The cuneiform and Proto-Aramaic inscriptions on vessel fragments, stamp seals and slabs from excavations indicated that the temple of the god Inzak, tutelary god of Dilmun, existed in Failaka Island during the Bronze Age (Rajab, 1999).

Figure 7. Archaeological sites in Failaka Island

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. At the south-western side of Failaka Island, the Seleucids built a fort plus multiple buildings for various purposes. 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 Late Pre-Islamic Period is another epoch in Failaka history. In the middle of Failaka Island, a large agriculture village was situated in an archaeological site known as Al-Qusur. This site was dated to the $5^{th}/6^{th}$ and $7^{th}/8^{th}$ centuries and was excavated by three international archaeological expeditions in the period from 1976 to 2006 (Rajab, 1999). During these excavations, a church was found in the middle of the village which belonged to a Nestorian Christian community which resided on Failaka Island. Approximately 140 farmsteads were found in that village and each farmstead has its own habitation and activity area. Al-Quraniya and Al-Zor are two archaeological sites that associated with the Middle and Late Islamic Period (Rajab, 1999). The archeological

excavations are still continuing in Failaka Island and the more archaeological sites found the better the reconstruction of Failaka Island history.

History of Kuwait:

The first European map of Kuwait territories was created by a Dutch sailor in 1645. The sailor was searching for his way to Basra (south Iraq), but by mistake he was in the Bay of Kuwait. In the early maps, Kuwait sometimes was known by Al-Qurayn (Grain) which means "the top of the hill," or 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.

The establishment of Kuwait as a state reflects a situation of traditional tribal societies that migrated from Najd in early 18th century to find new pastures and water sources that could sustain their livestock. The final stop for the migrating tribes was in Kuwait. The processes of state formation in Kuwait started by the arrival of the tribes, who were organized by means of the labor division (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 in Kuwait, while the second stage was initiated by the discovery of oil. Both stages had a vigorous impact on the evolution of Kuwait's political structure. Kuwait's political structure transformed from stateless to state society under British supervision and guidance. British colonialism also had its impact on the formation of Kuwait through the demarcation of Kuwait's border and oil discovery (Lienhardt & Al-Shahi, 2001).

Utub Families Migration

Kuwait was established in the 17th century by the arrival of nomads, known as Utub, who settled within the region of Kuwait. Utub is a group of families derived from the Arab tribe Aniza, which migrated from a plateau known as Najd. Najd has been the home of nomadic tribes of Arabian origin known as Bedouin. Utub families were highly organized and dependent on the presence of water and pasture locations for sustaining themselves and their livestock. In the absences of water, the Bedouin tribes migrated from one location to another searching for pasture and water (Casey, 2007).

The political patrilineal structure of the Bedouin is unique and consisted of three sub-structures: household, clan, and tribe. Household represent a group of individual families, and households then grouped to form a clan. The final sub-structure, tribe, is consisted of multiple clans. The name of most clans is derived from a common ancestor. Similar to most tribes in the Arabian Peninsula, Aniza tribal unity emerged from kinship sodalities. (Alghanim, 1998; Anscombe, 1997; Slot, 2003).

Prior to 1613, a small fishing village was established by Bani Khalid tribe in what is now Kuwait. This tribe controlled eastern Arabia and the coastline from Basrah to Qatar. Sheikh of Bani Khalid, Barrak bin Ghuraif, decided to construct a home in Qurain (Kuwait). This home was used as either a summer home or a storage place. Although Barrak bin Ghuraif was technically the first ruler of Kuwait, there is no evidence that indicates the period of his settlement in Kuwait (Casey, 2007).

In the 17th century, the drought was extreme and the lack of rainfall limited the availability of pasture land in Najd and there was no way in which the tribes could sustain their life. As a result, they started their journey of migration to find another place with superior availability of water and pasture. The migration routes started toward Qatar.

Historians believe that the Utub families settled in Qatar for approximately fifty years, and then they scattered all over the Arabian Gulf. The second migration route was from Qatar toward Kuwait. This route of migration was only part of different routes for the Aniza tribe (Anscombe, 1997; Slot, 2003). The availability of fresh water was the first obstruction for Kuwait's settlers. So they depended on two main sources of fresh water in Kuwait. The first source of fresh water was rainwater. During winter time they collected water and stored it in underground cisterns for drinking. The other source was the high reaches of Shatt Al-Arab (Lienhardt & Al-Shahi, 2001).

Overtime, Utub families were able to gain the control from Bani Khalid and minimize the role of Bani Khalid to control the port. The internal organization of Utub families gave them the power to control Kuwait. Utub families consisted of three principal divisions and they are called Al-Sabah (the ruling family of Kuwait), Al-Khalifa (the ruling family of Bahrain), and Al-Jalahima. The arrival of Utub families to Kuwait was the beginning of changing the tribal life. The first reason for settlement of Kuwait was the availability of water. Kuwait's location forced Utub families to adapt the new environment where the presence of water was accessible to them. Also, they were able take advantage of the critical geographic location of Kuwait. The adaptation to the new location started with the structural re-organization of Utub families. Also, they depended on the Arabian coast, its fish, its pearls, and its trading, as well as the advantages of Kuwait's harbor. At that time the commercial network in Kuwait was expanded and they started to exports the goods to Africa and India (Anscombe, 1997; Slot, 2003).

The Utub re-organization by means of dividing the work among the three main divisions was necessary in this critical period in order to sustain and control their lives on the one hand, and to share the total profits equally on the other hand. As a result, the Al-

Sabah family was responsible for administering the internal and military affairs of the community, and their rule was similar to the government rule in our day. While searching for pearls was the main job of Al-Khalifa family, the Al-Jalahima family was responsible for the maritime trades. The division of labor reflects the belief in the values of team work (Lienhardt & Al-Shahi, 2001).

In 1730, one of Oman's religious leader requested assistance from all Arabian settlements to defend the territory of Oman against the Portuguese military attack with intention to reduce the commercial expansion of Arabia. Kuwait responded by sending two ships loaded with munitions. That was the first documented foray into foreign affairs by the settlers of Kuwait (Casey, 2007).

By either 1752 or 1756, the election of Shaikh Sabah bin Jabir as a leader by the Utub families was the first sign of the establishing authority of Al-Sabah in Kuwait. In addition to local administration of justice, Shaikh Sabah bin Jabir was responsible for the collection of taxes and tariffs. In 1760, the creation of a mud wall was the first indication 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).

After the death of Sabah in 1764, the senior members of Al-Sabah family elected Sabah's youngest son, Abdullah. This election was based on Abdullah's sense of justice and intelligence in consulting the group in important matters. From then on, the tradition of election changed to be based on the appropriateness of the member to be the leader rather than on who was the oldest (Slot, 2003). Shaikh Abdullah has not changed the policies of his father which led to sharing the economic benefits among the community members (Slot, 2003).

Al-Sabah…The Ruling Family

In 1760s, dramatic changes occurred in Kuwait which attracted many merchants. At that time Kuwait has transferred from a small territory to a center of trade in the Arabian Peninsula. Kuwait was the main stop point for either camel caravans from southern Arabia or trading vessels. Wealthy Kuwaiti families were able to obtain their own ships and started a circular trade to either Africa or India (Casey, 2007).

Another dramatic change was the departure of the Al-Khalifa family and the Al-Jalahima family from Kuwait. The reason behind their departure was unclear. However, some evidence attribute the disagreement between the Al-Khalifa family and Shaikh Abdullah as the main reason. While others indicate that Al-Khalifa family and the Al-Jalahima family escaped an invasion threat from Bani Kaab tribe who settled in Arabistan region of Persia (Iran). These reasons forced both the Al-Khalifa family and the Al-Jalahima family to departure Kuwait toward Zubara (South Iraq) and then to Bahrain (Slot, 2003; Casey, 2007). The Al-Khalifa was able to conquer the island and they ruled Bahrain since that time. The Al-Sabah and the Al-Khalifa maintained good economic and political relationship.

The departure of the Al-Khalifa family removed the internal threat for Al-Sabah family, and the situation left the Al-Sabah to exercise their authority and control Kuwait without any internal threat that could replace the Al-Sabah family. In doing that, Al-Sabah gained economic and political advantages. As a result, the Al-Sabah family no longer shared their overall income with other families, and they increased their percentage of the income (Slot, 2003). The occupation of Basra (sea port) by Persians shifted the pearl fishing trade and the Indian trade routes to Kuwait. In addition to the trade shift, the successful relationship between Shaikh Abdullah and merchant families

raised the economic profit for Kuwait (Slot, 2003). The British East India Company had established a factory in Basrah. However, the bad situation in Basrah led to transferring the company to Kuwait for the period 1739-1795. Through the British East India Company, the Kuwaitis had their fist exposure to the British Empire. At that time, the concern of the British Empire was to enforce the stability through protecting their sea lines of communications from the Red Sea to India rather than gaining control over the Arabian Peninsula. Their first goal was preventing any foreign power from threatening their sea lines of communications through using the Arabian Peninsula as a base of attack. The second goal was controlling the local power in the region from gaining any military strength which eventually could threaten those sea lines of communications (Casey, 2007).

In 1812, Shaikh Jabir was elected by Utub families after his father's death. By this time, the Al-Sabah were firmly in control of Kuwait. The main task of Shaikh Jabir was keeping the interest of the wealthy family away from his controlling authority through creating a good economic environment for them to increase their profit. Kuwait depended on sailing for fishing and pearl fishing. In addition, making trade routes with other countries helped to increase profits. Utub families started to sail toward the Red Sea Coast, the Coast of Sind, and Bombay (Slot, 2003).

Economic Development

Fishing is the first source of livelihood for the tribe because of the simple technology that was used in fishing, such as nets. To start fishing the tribes needed to create fishing equipment to enable them to enter the sea and start fishing. Fish was the main protein source in early Kuwaiti life, and most people ate fish daily. On the other hand, consumption of red meat and chicken was very low and provided only a minor

source of protein in their diet. Fishing, as any another maritime occupation, is men's work. The women could help the men from their families by tying ropes onto the net (Lienhardt & Al-Shahi, 2001).

Men usually made their own fishing nets. Kuwait's tribes took advantage of their environment, so they used every natural source for their benefit. For example, the date palm was the only source of material used to construct fishing equipment. It is normal for fishers to make their own nets, but not in constructing boats. Consequently, carpenters specialized in constructing boats (Lienhardt & Al-Shahi, 2001). At first, the wealthy families were the owner of these ships and they gain the major profits with a small sharing to Al-Sabah. Ship captains were next in the line followed by crewmen and pearl divers.

In contrast to fishing, seafaring merchant was the next complex type of maritime activity in terms of organization of individual operation. Fish and pearls were used by fishermen for their own purpose, such as selling them in local markets. In case they had an extra amount of fish or pearls, they would ship them to external markets that were located in India or Africa. During the summer, hundreds of Kuwaiti ships were specialized in pearling and they were selling them at nearby market.

The travels by ships could take several months to reach foreign markets in either India of Africa. However, their need for imported goods was the main reason to motivate Kuwaiti tribes to undertake long and arduous trading voyages. The captain of the boat depended on astro-navigation to determine his way across the sea (Lienhardt & Al-Shahi, 2001). In their way to return to Kuwait, they imported many goods and raw materials. Their imports consisted of rice, wood, sugar, and spices from India; grain and dates from Basra. Although the slave trades were not the prime aspect of trade in Kuwait, their few

slaves brought to Kuwait from Africa. Although the details are unknown, the slaves were probably brought to Kuwait for transshipment by desert caravan to either western or northern markets (Slot, 2003; Casey, 2007). Shaikh Jabir collected one percent duty from the imports (Slot, 2003). In addition to the maritime trade with India and Africa, Kuwait also became an important stop on land for the trade routes that linked India, Arabia, and Persia to Europe (Slot, 2003).

Kuwait's Social Structure

Trade, as we have seen, divides Kuwait society 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 Shaikh depended on two sources to develop his final judgment. The first source was the '*Urf* (the customary rules), and the second was *Shari'a* law (the Islamic law). According to the nature of the ruler worked which it should be accessible to all Kuwaiti population, they introduced the *majlis*. *Majlis* was a place where the Kuwaiti population members could meet the ruler to introduce their disputes and discuss their needs (Alghanim, 1998; Lienhardt & Al-Shahi, 2001).

Kuwait's population consisted of a specific social structure. The merchants, similar to the ruling family, were small in number. The merchants were divided into two groups. The first group had almost all the ships, and they controlled commerce in Kuwait (Slot, 2003). The second group consisted of a large number of small merchant retailers. The first group of merchants was wealthier than the second group. Finally, Bedouin were a mobile group who searched for new pastures in the desert. The Bedouin were known to be good fighters and aided the Al-Sabah family in their wars (Alghanim, 1998; Lienhardt & Al-Shahi, 2001).

Bedouin tribes formed a political group in which the loyalty between its members united them. The political structure of the tribe consisted of the Shaikh of the tribe and the tribe members. The Shaikh of the tribe had two major roles. First, he was leader of the tribe in any attacks. Second, the Shaikh was involved in internal segmentary structure of his tribe, such as internal disputes among tribe members. The Shaikh of tribe recognized the necessity of the modification elementary segment of the tribe political structure in order to be part of the new society in the next stage of evolution. This modification happened due to the differences between the town state and Bedouin tribes. First, the town state is a commercial center attracts mixed population. Second, an increase of the authority in the town state occurred (Lienhardt & Al-Shahi, 2001).

The Bedouin of Najd consist of multiple tribes, most of them can trace their heritage back for centuries. That made a Bedouin man proud of his lineage and looked with disdain on non-Bedouin regardless of his social standing within his own tribe. In general, Arab can be classified in to two groups: Arab al Ariba and Arab al Musta'riba. Arab al Ariba or an "Arab of the Arabs" is a group of all Arabs (Bedouin) whom lived in the Arabian Peninsula before the Islam and they regarded as pure Arabs. To the Bedouin, all other Arabs belong to Arab al Musta'riba or "Arabs who became Arabs" group which is define those who became Arab after the Islamic conquests of northern Arabia and Africa. Arabs can also be classified according to their living location. Bedouin is a word derived from another word Badia (desert). Bedouin is used to describe people who are nomads, live in the desert, and constantly travel with their herd searching for new pastures. Al Hadar is a word used to describe the Arabs of the villages who live in permanent homes (Casey, 2007).

Kuwait in the 20th Century:

The twentieth century witnessed huge changes in Kuwait that impacted the social, political, and economic structures of Kuwait. At that time the estimation of the Kuwaiti population was approximately around fifty thousands. Most of them left their nomadic life and they started to settle in the city under the rule of Al-Sabah who was generally benign. In addition to the major merchant families, few Kuwaiti were growing rich. The economy in Kuwait was administrated and guided by Al-Sabah Family which led to better economic environment than any other place in the Arabian Peninsula (Casey, 2007; Crystal, 1992).

The life in Kuwait was simple and the common Kuwaitis still lived in mud homes. Since wood and stones were incredibly expensive, only the merchants were able to build their homes out of these materials. During the summer, the majority of Kuwaiti population preferred to sleep on the flat roof of their homes to avoid the nighttime heat (Casey, 2007).

In 1906, more than 400 boats available in Kuwait that employed more than nine thousands men in the pearl trade. Although the pearling season extends for four months (mid May to Mid September), the profit from pearling was enough to support the Kuwaitis until the next season. The pearl trade in Kuwait was in steady expansion over the following few decades. To cope with pearl trade expansion, more than fifteen thousands of Kuwaitis were employed as crew member and divers (Casey, 2007).

Two major external circumstances changed the economic situation of pearl trade for the worse: global economic crisis of the 1920s and the marketing of cultured pearls. Although the first trials for breeding cultured pearls were started in the 1700s by the Dutch, the marketing of cultured pearls was initiated in the 1920s by a Japanese

entrepreneur who created a fast and economically efficient technology for making cultured pearls. The effect of the cultured pearls on Kuwait economy was both devastating and predictable. Generally, the natural pearls markets collapsed. As a result, the Kuwait's pearls-based economy followed the general scheme and consequently Kuwait became poor country depending on the Great Britain more than ever (Casey, 2007).

Kuwait's economy was not in the collapse state for very long. After few years, a new discovery would change the economic structure of Kuwait. Moreover, the impact of the discovery would extend to the political and social structures. In 1932, oil was discovered by a British-American company in the oil field of Burgan (Casey, 2007).

Oil Discovery & its Impact

During the 1920s, the worldwide economy was in a depression and that has a great impact on Kuwait's economy. Kuwait was one of the leaders in pearling industry with approximately 800 boats involved in pearling to meet the world's need of pearls. However, Kuwait economy which was based on pearling reduced tremendously due to the world wide economic depression. In addition to the global economic depression, the Kuwait's economy was affected by the launching of a cultured pearl industry. As a result, the majority of Kuwaiti people were searching for new work for sustenance, except the rich families, who owned the boats, were able to survive the economic collapse (Casey, 2007; Crystal, 1992).

In 1932, oil was discovered in Kuwait, but commercial production did not start until June 1946 due to the outbreak of the Second World War. In 1950, Kuwait's production of oil had reached 17 million metric tons. During the next decade, oil production was over 87 million metric tons. The rise in oil production increased Kuwait's

income. Kuwait made huge amounts of money from oil production. The impact of the oil production was obvious on the society (Slot, 2003; Crystal, 1992).

The first impact of oil production was an increase in Kuwait's population. Oil production industries offered different kinds of jobs. Those jobs attracted approximately 100,000 skilled and unskilled workers. The advantages of work in Kuwait, beside the high salary, were free health, education, and other services of a welfare state. As a result, Kuwait's population doubled by the mid-1950s. The first census taken in 1957 indicated a total population of 206,473 and the second census in 1961gave a total population of 321,621 (Figure 8) (Slot, 2003; Crystal, 1992).

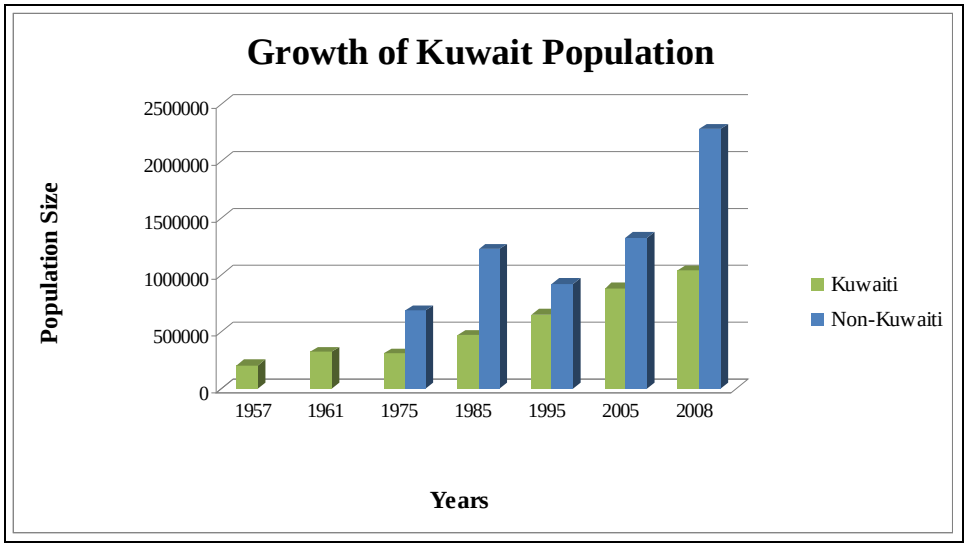

Figure 8. Censuses of the Kuwaiti population

Not only Kuwait population structure was affected by the Oil discovery, the population structure in Failaka Island was affected as well. Prior to 1957, there was no official census in Kuwait so that the total population size of Failaka was obscure. However, many historians and itinerants who visited Failaka Island gave estimate of the total population. In 1839, the traveler F. Jones claimed that there were 70 to 80 homes in Failaka Island which were occupied by approximately 150 individuals. In 1905, John Gordon Lorimer, who was a career administrator in the British overseas services, noticed

that the total men in Failaka Island were 200 approximately and the total population could reach 500 individuals. The Kuwaiti historian, Abdulaziz Al-Rushaid, stated that in Failaka Island there were 200 hundreds homes and approximately 1200 individuals. In 1950, another British colonial administrator known by H. Dickson estimated the total population to be approximately 1500. While most of them lived in Zor area which is located northwest coast, the reminder lived in Al-Quraniya which is located three miles northeast Zor area. According to these population estimates, the total inhabitants of Failaka Island increased from 150 to 1500 within 111 years.

The first official census of Failaka Island was in 1957 and followed by several censuses in the following years: 1961, 1965, 1970, 1980, and 1985 which was the last census before the Iraqi invasion of Kuwait (Figure 9).

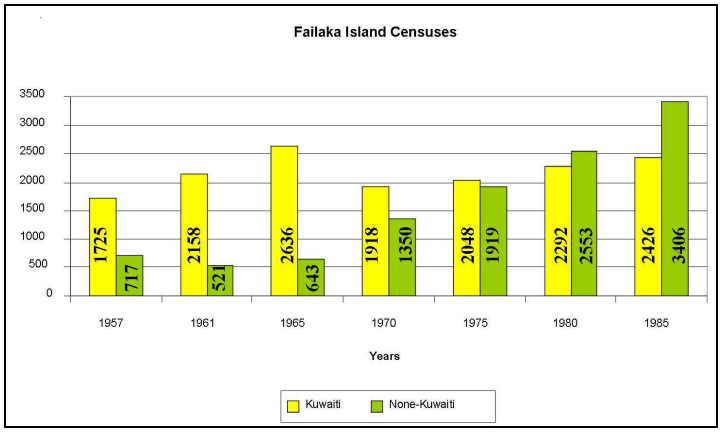

Figure 9. Censuses of Failaka Island

Figure 9 shows that the total Kuwaiti in Failaka Island increased from 1725 to 2636 within four years (1957-1965). According to 1970 census, there was a drop in the total Kuwaiti population in Failaka Island to reach 1918. The none-Kuwaiti population was in steady increase since 1961. Within five years (1965-1970) a major change occurred in the population structure of Failaka Island. While the total none-Kuwaiti was doubled from 643 to reach 1350, the total Kuwaiti population was declined from 2636 to

1918. In the last two censuses, the total number of none-Kuwaiti exceeded the total number of Kuwaiti in the Failaka Island.

The second impact of Oil discovery was the rapid expansion of government work. During this period, a great expansion and increased diversification in the machinery of government, and both legislative and executive powers of the government continued to be a feature of the Amir. In 1959, the Supreme Council was established in Kuwait. The Supreme Council is considered as the highest advisory and decision making body in the country. The Supreme consisted of ten members from society and the president of the departments. Decisions were made by majority vote (Slot, 2003

From 1950 to 1965, the government of Shaikh Abdullah Al-Salim Al-Sabah distributed the nation's wealth among Kuwait citizens. In addition, the government increased its income from different sources such as investments. The social order found in Kuwait prior to 1950 had changed after oil boom. Politically, the merchants had become the major beneficiaries of the new government policies of oil production and wealth distribution. As a result, merchants were known to be the strongest supporters of the ruling family. Moreover, the new constitution which adopted in 1962 had as much impact on government and society as did oil. The Constitution gives Kuwait a democratic rule in which "the citizens are provided with more political freedom, equality and social justice". We can say that Kuwait transformed from a tribal form to a complex civilization within only a short period of time (Slot, 2003).

Chapter Three: Material and Methods

Sample

Blood samples from 146 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. Out of the total number of the sample, 27 individuals who came from Failaka Island and represent Failaka Island population. The rest of the samples were categorized into one of three ethnicity based on their family's place of origin: Arab (n=47), Bedouin (n=26), or Iranian (n=46). 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 the discovery of oil 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 Y-STRs comparison with the Kuwaiti and Failaka Island population consisted of 7378 individuals representing 21 countries and three continents: Africa, Asia, and Europe (see Table 1). For the purpose of mtDNA comparison with Failaka Island population, 1854 individuals representing 19 countries were used (see Table 2). The mt-DNA sequence of HVS-I 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 tables represent the populations and the number of samples in each population.

Table 1. Y-chromosomes STR of the comparative populations included in this study

Continent	**Population**	**NO. of Samples**	**Y-chromosomes STR Source(s)**
Africa	Mozambique	112	Alves *et al.*, 2003
	Namibia (Ovambo)	54	Fujihara *et al.*, 2009
	Guinea	101	Arroyo-Pardo *et al.*, 2005
Asia	Yemen	50	Cadenas *et al.*, 2008
	United Arab Emirates	88	
	Qatar	47	
	Saudi Arabia	171	Abu-Amero *et al.*, 2009 Alshamali *et al.*, 2009
	Kuwait	**119**	**This Study**
	Failaka Island	**24**	
	Iran	325	Haber *et al.*, 2011
Europe	Austria	261	Erhart *et al.*, 2012
	Croatia	1035	Mrsic *et al.*, 2012
	Germany	996	Goedbloed *et al.*, 2009
	Poland	887	
	Ukraine	154	Mielnik-Sikorska *et al.*, 2013
Total	-	4449	-

Table 2. Mitochondrial DNA HVS-I of the comparative populations included in this study

Continent	Population	NO. of Samples	mtDNA sequence Source(s)
Africa	Nigeria	63	Watson *et al.* 1996
	Kenya	78	
	Somalia	5	Olivieri *et al.* 2006
	African San	17	Tishkoff *et al.* 2007
Asia	Yemen	90	Kivisild *et al.* 2004
	Saudi Arabia	15	Abu-Amero *et al.* 2007
	Kuwait	94	Theyab et al., 2012
	Failaka Island	**27**	**This study**
	Iraq	116	McEvoy *et al.* 2004
	Syria	69	
	Kurdistan	53	
	Iran	92	Nasidze *et al.*2008
Europe	Turkey	290	McEvoy *et al.* 2004
	Greece	179	
	England	242	
	Bulgaria	141	
	Romania	92	
Total	-	1663	-

Laboratory Methods

DNA Extraction

Two different methods were used to extract DNA from the blood samples collected from the Kuwaiti population: salting out procedure (Miller *et al.*, 1988) and Gentra Puregene Blood Kit (Qiagen, Valencia, CA). The salting out procedure 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_2EDTA, 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.

In the second extraction method, 3 ml of a volunteer's blood was transferred into a 15-ml polypropylene centrifuge tube. 30-ml of RBC lysis solution was added to the tube and mixed by inverting 10 times and then incubated at room temperature (15-25°C) for 5 minutes. Following the incubation, the tube was centrifuged for 15 minutes at 2000 rpm. The supernatant, containing the white blood cells pellets, was discarded and 200 µl of the supernatant left in the tube which was then vortexed to resuspend the pellet in the residual liquid and facilitate the cell lysis. 3ml of the cell lysis solution was added to the tube and mixed by pipeting up and down several times to lyse the cells. If cell clumps were visible after pipeting, the tube was incubated at 37°C until the solution was homogeneous.

Once the solution homogenized, 1ml of protein precipitation solution was added and vortexed for 20 seconds and then centrifuged for 5 minutes at 2000 rpm. Tight dark brown pellet should be formed as the protein precipitates. If the protein pellet was different, the tube was incubated on ice for 5 minutes and the centrifugation was repeated. The supernatant, containing the DNA, was transferred into a new 15-ml polypropylene tube containing 3-ml isopropanol and mixed by inverting 50 times until the DNA was visible as threads or a clump. The tube then was centrifuged for 3 minutes

at 2000 rpm and that the supernatant was discarded and the tube drained by inverting it on a clean piece of absorbent paper, making certain that the DNA pellet remained in the tube. 3-ml of 70% ethanol was added to the tube and inverted several times to wash the DNA pellet and then centrifuged for 1 minute at 2000 rpm. The supernatant was discarded and the tube was drained on a clean piece of absorbent paper, taking care that the pellet remained in the tube and air dried for 5-10 minutes. Once the tube was dry, 300 µl of DNA hydration solution was added and vortexed for 5 seconds at medium speed and incubated at 65°C for 1 hour to dissolve the DNA followed by overnight incubation. The next day, the tube was centrifuged and the sample transferred to a storage tube.

Y-chromosome STRs: Amplification and Sequencing

147 male individuals were characterized for 17 Y-chromosome STRs loci using the AmpFℓSTR® Yfiler® PCR Amplification Kit (Applied Biosystems, Foster City, CA). The amplified Y- chromosome STRs loci are: DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS438, DYS439, DYS437, DYS448, DYS456, DYS458, DYS635 and Y GATA H4. The samples were amplified using a Polymerase Chain Reaction (PCR) method and each amplification reaction tube contained: 5.25 µl AmpFℓSTR® Yfiler® PCR Reaction Mix, 2.75 µl of AmpFℓSTR® Yfiler® Primer Set, and 0.25 µl of AmpliTaq Gold® DNA Polymerase, 5 µl of sample DNA, and 2.25 double-distilled water. A 5 µl volume of deionized water was added to a tube, instead of DNA, and saved as a negative control. A 10 µl volume of AmpFℓSTR® Control DNA 007 was added as a positive control. Amplification reactions were run on an Applied Biosystems GeneAmp® PCR System 9700 (Applied Biosystems, Foster City, CA). The amplification reaction was adjusted to the following thermal profile: initial incubation at 94°C for 11

minutes (hold); 94°C for 1 minute (30 cycles); 61°C (annealing temperature) for 1 minute (30 cycles); 72°C for 1 minute (30 cycles); final extension at 60°C for 60 minutes; and then on hold at 4°C.

After the samples were amplified, they were quantified for DNA concentration using a NanoDrop ND-1000 spectrophotometer. Then, a master mix was created by mixing the following components and their appropriate volume per sample: 0.3 µl of GeneScan™ 500 LIZ® Size Standard and 10 µl of Hi-Di™ Formamide. A 10.3 µl volume of the master mix was pipetted into each well of the 96-well reaction plate. A 1 µl volume of the amplified DNA samples was pipetted into a corresponding well of the 96-well reaction plate. A 1 µl volume of the reagent blank, negative control, AmpFℓSTR® Control DNA 007 from the amplification, and allelic ladder were pipetted into corresponding wells of the 96-well reaction plate. 10 µl of Hi-Di™ Formamide was added to each blank well. The plate was heated at 95˚C for 4-5 minutes and then placed in ice-water for 5 minutes. Finally,the plate was placed into the Applied Biosystems 3130xl Genetic Analyzer. The data files were analyzed using GeneMapper® ID v3.2 software (Applied Biosystems, Foster City, CA).

Mitochondrial DNA: Amplification and Sequencing

Since the HVS-I region of the mtDNA of the Kuwaiti population was previously studied by the author of this dissertation (Theyab *at el.*, 2012), only the samples from Failaka Island population were amplified and sequenced for HVS-I region. The HVS-I (1600-16400 nt) region of the mtDNA control region was amplified and then 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 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). Table 3 provides information regarding the primers sequence used in the mtDNA analysis, mitochondrial location, and product size.

Table 3. List of mt-DNA primer, Annealing temperature, and PCR product fragment size.

mtDNA region	*Primers*	*Annealing *T (ºC)*	Product Size (bp)
HVS-I targeted nucleotides	F15971: 5'-TTAACTCCACCATTAGCACC-3' R16410: 5'-GAGGATGGTGGTCAAGGGAC-3'	56	443

The PCR reactions were run on an Applied Biosystems GeneAmp® PCR System 9700. The amplification reaction was adjusted to the following thermal profile: 94°C for 1 minute (1 cycle); 94°C for 50 seconds (40 cycles); 65°C (annealing temperature) for 1 minute (40 cycles); 72°C for 1 minute (40 cycles); 72°C for 5 minutes (1 cycle); and then on hold at 4°C. The 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 the gel was 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.

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).

Statistical Analyses:

For Y-chromosome analyses, DYS385a and DYS385b microsatellites were excluded since it is difficult to assign each allele to one locus. The repeats of DYS389I were subtracted to the DYS389II because the fragment size of DYS389II contains those of DYS389I.

Haplotype or gene diversity, nucleotide diversity, and neutrality test analyses were performed using a computer software package called Arlequin, version 3.11(Excoffier et al. 2005). Haplotype diversity (Nei, 1978) was calculated for mtDNA HVS-I and Y-chromosome STRs data. This measure is parallel to the expected heterozygosity for diploid data and it measures the probability that two haplotypes selected at random from the population are different. Haplotype diversity is calculated by using the following equation:

$$H = \left(\frac{n}{n-1}\right)\left(1-\sum_{i=1}^{k} P_i^2\right) \quad (1)$$

where n represents the sample size, k is the number of haplotypes, and p_i is the frequency of the ith haplotype. This measure is relatively stable and considered less responsive to genetic drift and recent demographic events (Helgason *et al*., 2003). Nucleotide diversity for the mtDNA sequence data (Nei and Li 1979) was calculated as:

$$\pi = \sum_{ij}^{q} x_i x_j d_{ij} \quad (2)$$

where q is the total number of alleles, x_i is the frequency of the i-th allele in the population, and d_{ij} is the number of nucleotide differences between alleles i and j. Haplotype and nucleotide diversity tests are two common measures of genetic diversity.

Measurements of selective neutrality, Fu's Fs and Tajima's D, were calculated using Arlequin version 3.11 (Fu, 1997; Tajima, 1993). These measures are used to determine whether the DNA sequences are statistically significant under the neutral equilibrium model. The neutral evolution model is based on the standard Wright-Fisher evolutionary model which has the following assumptions: 1) large constant diploid population size; 2) random mating; 3) discrete and non-overlapping generations: 4) no recombination; and 5) infinite sites with constant mutation rate (Simonsen *et al.*, 1995).

Fu's Fs and Tajima's *D* tests are used to differentiate between population growth versus constant population size. Tajima's *D* (Tajima, 1989) is based on the infinite-site model without recombination and is appropriate for use with short DNA sequences. It is defined as:

$$D = \frac{\theta_\pi - \theta_s}{\sqrt{Var(\theta_\pi - \theta_s)}} \quad (3)$$

where $\theta=2Ne\mu$ (for haploid data), where *Ne* is the effective population size and μ is the mutation rate; θ_π represents the mean number of pairwise differences between sequences (π); and θ_s is based on the number of observed polymorphic sites. If population growth accumulates an excess of low-frequency mutations, that would lead to larger value of θ_s compared to θ_π which would result in negative *D* value. The negative values are an indication of population expansion. On the contrary, positive *D* values are an indication of population undergone genetic bottlenecks (Aris-Brosou & Excoffier, 1996).

Fu's Fs is another neutrality test statistic that based on the infinite-site model without recombination. This test employs information from the haplotypes distribution and is defined by the following equation:

$$F_s = \ln\left(\frac{S'}{1-S'}\right) \quad (4)$$

where S' is the probability of observing a random neutral sample with *k* as the number of alleles equal to or smaller than the observed value given θ_π. Fu's F_s is more sensitive than Tajima's *D*, and may produce large negative values as an indication of large population expansion. Positive values may be indicative of genetic drift (Fu, 1997).

Mismatch distributions is another molecular diversity analysis that was used to investigate the demographic history of the populations. Population expansion can be

detected in genetic data by identifying specific signatures in the distribution of pairwise sequence differences. While a unimodal distribution represents a recent population expansion, a multimodal distribution is an indication of maintained constant population size over time (Rogers & Harpending, 1992). A raggedness index (*r*) is used to evaluate these distributions. While lower index values are usually associated with unimodal distribution, higher values are usually associated with constant population (Harpending *et al*., 1993). This index is defined by the following equation:

$$r = \sum_{i=1}^{d+1} (x_i - x_{i-1})^2 \qquad (5)$$

where *d* is the maximum differences between haplotypes, and *x* is the relative frequency of the mismatch. Mismatch analysis and *r* index were calculated for mtDNA sequences data based on pairwise distances using Arlequin 3.11.

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 both Y-chromosome STRs and 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 Neighbor-Joining tree (NJ) and Multi-Dimensional Scaling plot (MDS). 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 & Nei, 1987). The smallest sum of branch lengths was calculated by the following equation:

$$S_0 = \sum_{i=1}^{N} L_{iX} = \frac{1}{m-1} \sum_{i<j}^{m} d_{ij} = \frac{T}{m-1} \quad (6)$$

where m the number of neighbors, L_{iX} represents the branch length estimate between nodes i and X , and T represents the sum of distance estimates d_{ij} ((Nei & Kumar, 2000). 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. For the Y-chromosome STRs data, Slatkin's linearized *Fst* distances which is based on the stepwise mutation model for microsatellites were calculated using Arlequin (Slatkin, 1995). For mtDNA sequences, Tamura & Nei's (1993) distances were calculated since the measures take into account the differences in transversion and transition rates, in addition to transition rates between purines or pyrimidines. The distance measure also assumes equality of substitution rates among sites (Excoffier & Schneider, 2005). 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 and Failaka Island populations were compared to neighboring and worldwide populations to draw insight regarding the genetic relationship of Kuwaiti and Failaka Island populations 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).

A median-joining network (MJ) was computed for J1 Y-haplogroup for the Kuwaiti and Failaka Island populations. Using a statistical program NETWORK 4.611 (Fluxus Technology Ltd) following Bandelt *et al*., (1999). MJ network shows the

relationships between haplotypes by proposing likely ancestral haplotypes. Since each locus has different rate of evolution, each locus was weighted according to Kim *et al*., (2010). For MJ analysis, the DYS389I and II loci were omitted because these loci are interdependent and represent four independently mutating DNA segments. Since it is difficult to assign one allele to each locus, the bilocal DYS385a/b was excluded from this analysis.

To determine the relative amount of gene flow and genetic drift of Kuwait and Failaka Island populations, the heterozygosity of each subpopulation was plotted against the distance from the gene frequency centroid (r_{ii}). Where heterozygosity is the usual expected heterozygosity under Hardy-Weinberg, and the distance from the centroid (r_{ii}) for a population *i* is:

$$r_{ii} = \frac{(p_i - \bar{P})^2}{\bar{P}(1-\bar{P})} \qquad (7)$$

where r_{ii} is the distance from the centroid for a particular allele in the *i* population, p_i is the frequency of the allele in the *i* population, and $\bar{P}$ is the mean frequency of the allele for all populations. According to Harpending and Ward (1982), populations with more than average gene flow will fall above the regression line, while populations that are genetically isolated will fall below the regression line. Finally, Mantel test was used to test for the significance of the correlation between two matrices (Mantel, 1967). Mantel test, with 10,000 random permutations, was used to assess the correlations among mtDNA HVS-I, Y-STR, and geography distance matrices for Kuwait, Failaka Island, Saudi Arabia, and Iran. The Mantel tests were performed using the program Mantel ver. 3.1 (Relethford). While the genetic distance matrices were obtained from Arlequin 3.11, the geographic distance matrices computed in GEOG ver. 2.1 (Relethford, 2000).

Chapter Four: Results

Y-chromosome DNA

Y-chromosome STRs Haplotype and Population Diversities

The Y-chromosome haplotypes for Kuwait and Failaka Island are presented in Table 4 and 5, respectively. All the samples were typed for the following seventeen loci: DYS456, DYS389I, DYS390, DYS389II, DYS458, DYS19, DYS385a/b, DYS393, DYS391, DYS439, DYS635, DYS392, GATA_H4, DYS437, DYS438, and DYS448. A total of 119 haplotypes were identified in the Kuwaiti samples; 107 were unique (90%) and 12 (10%) were observed more than once. In Failaka Island's sample a total of 24 haplotypes were identified; 22 were unique (92%) and 2 (8%) were observed more than once. Shared haplotypes has not been found between Kuwait and Failaka Island samples. However, the sample of Kuwait and Failaka Island shared haplotypes with neighboring populations. Figure 10 shows the frequency of shared haplotypes of Kuwait and Failaka Island with neighboring populations.

Table 4. Y-chromosome haplotypes in Kuwait

Y-STR Haplotype[a]	Frequency
14,12,24,29,17,16,20/21,12,9,11,20,11,12,14,9,19	3
16,14,21,31,17,16,17/18,15,10,12,21,11,11,14,11,20	1
15,14,23,30,18.2,14,11/18,12,10,13,20,11,11,14,10,19	1
14,13,23,30,17.2,14,13/21,12,11,11,21,11,11,14,10,20	1
14,14,23,32,18.2,14,14/19,12,11,11,21,11,11,14,10,20	1
14,13,23,30,19.2,14,13/17,12,10,11,21,11,11,14,10,20	1
14,13,23,30,18.2,14,13/19,12,11,12,21,11,11,14,10,20	1
14,13,23,30,19.2,14,13/19,12,11,12,21,11,11,14,10,20	1
14,13,23,29,18.2,14,13/18,12,11,11,21,11,11,14,10,20	2
14,13,23,29,20.2,14,13/18,12,11,11,20,11,11,14,10,20	1
14,13,23,29,19.2,14,13/18,12,11,11,22,11,11,14,10,20	1
15,13,23,29,16,14,12/14,11,11,12,23,14,13,15,12,19	2
16,14,23,31,17,14,13/?,13,10,11,24,10,12,16,9,20	1
14,14,23,31,18.2,14,13.2/19,12,11,11,21.2,10.2,11,14,10.2,20	1
15,14,22,30,20.2,14,12/18,12,10,13,22,11,11,14,10,21	1
15,13,23,30,18.2,14,13/18,12,11,11,21,11,11,14,10,20	1
14,13,23,30,19.2,14,13/18,12,11,11,21,11,11,14,10,21	1

15,13,23,30,18.2,14,13/18,12,11,11,21,11,11,14,10,20	1
15,13,23,30,18.2,14,13/19,12,10,11,21,11,11,14,10,20	1
13,13,23,31,18,16,11,?/13,10,11,?,11,11,14,10,21	1
14,13,23,30,19.2,14,13/19,12,10,11,22,11,11,14,10,20	1
15,13,24,29,14,15,13/16,13,10,12,22,11,12,15,9,21	1
15,13,26,29,15,16,11/14,13,11,10,23,11,12,15,11,20	1
15,14,23,29,15,15,13/15,13,10,12,21,11,12,15,9,22	1
14,13,23,31,18.2,14,13/19,12,10,11,20,11,11,14,10,20	1
15,14,23,32,15,13,18/22,13,11,11,21,11,13,14,10,19	1
16,9,23,25,15,14,14/15,14,11,11,20,12,11,16,10,21	1
14,13,23,30,18.2,14,13/19,12,11,11,22,11,11,14,10,20	1
14,13,23,30,18.2,14,13/19,12,11,11,21,11,11,14,10,20	6
15,12,23,30,16,14,14/16,13,10,13,21,13,11,16,10,21	1
15,13,23,29,16,14,16/?,12,9,12,23,11,12,15,10,22	1
14,13,23,30,18.2,14,12/20,12,11,12,21,11,11,14,10,20	1
14,13,23,30,19.2,14,12/18,12,11,12,21,11,11,14,10,20	1
15,14,25,32,15,16,11/?,13,10,10,23,11,11,14,11,20	2
13,12,24,30,16,15,17/18,12,11,12,21,11,11,14,9,19	1
14,13,23,30,21.2,14,13/19,12,10,11,22,11,11,14,10,20	1
16,13,23,29,17,14,13/16,13,10,11,22,13,11,14,9,19	1
15,13,22,29,18,13,14/16,13,10,12,22,15,10,14,11,19	1
15,13,23,29,17.2,16,13/19,12,10,11,21,11,11,14,10,20	1
17,12,23,29,16,15,14/16,14,10,11,21,12,13,16,10,22	1
15,14,23,30,17,13,14/18,14,10,11,23,15,12,14,11,20	1
16,13,23,31,15,14,15/?,13,9,11,21,10,12,15,9,20	1
14,13,23,30,18.2,14,13/17,13,11,11,22,11,11,14,10,20	2
14,13,23,30,18.2,14,12/18,12,10,11,20,11,11,14,10,20	1
14,13,23,30,18.2,14,14/19,12,11,11,21,11,11,14,10,20	1
15,13,22,29,16,14,13/16,12,9,11,23,11,11,15,9,21	1
13,14,23,32,17,15,11/?,13,10,12,17,11,13,14,11,23	1
15,14,23,31,17,14,14/16,13,10,11,21,13,11,14,9,19	1
15,13,23,31,19.2,15,13/19,12,11,11,21,11,11,14,10,20	1
15,13,23,32,17.2,15,13/17,12,10,11,20,11,11,14,10,20	1
14,13,23,31,18.2,14,13/17,13,11,11,22,11,11,14,10,20	1
14,13,23,29,17.2,14,14/19,12,10,11,23,11,11,14,10,20	1
14,13,23,30,18.2,14,12/19,12,11,12,21,11,11,14,10,20	2
15,14,24,32,18,13,17/18,15,10,12,21,11,13,14,10,20	1
16,14,23,32,17,14,13/?,13,10,11,25,10,12,16,9,20	1
14,13,23,30,20.2,14,14/20,12,10,11,21,11,11,14,10,20	1
15,14,23,29,18,14,13/19,14,10,11,27,10,11,14,11,19	1
16,12,22,28,14,15,14/16,11,10,12,25,14,12,15,10,19	1
15,14,25,30,17,16,13/14,14,10,12,23,13,13,14,12,19	1
14,13,23,29,19.2,14,13/19,12,11,11,21,11,11,14,10,20	1
17,13,24,31,18,14,16/18,13,10,12,21,11,11,14,10,20	1
16,14,23,31,18,14,13/?,13,11,13,23,13,12,14,11,19	1
16,14,22,31,18,15,7/16,12,10,13,23,14,11,16,10,19	1
15,12,25,30,17,15,16/19,13,11,12,21,11,12,14,10,21	1
15,13,24,29,16,14,13/19,13,10,11,25,10,11,16,11,20	1
14,12,22,29,15,15,14/15,14,10,12,20,11,12,15,10,21	1
16,13,23,29,15,14,13/19,12,10,12,21,11,11,15,10,20	1
16,13,25,29,15,16,11/12,13,11,10,23,11,12,14,11,20	1
15,14,23,31,17,15,11/20,14,10,11,25,10,13,16,12,19	1
16,13,24,30,15,13,16/18,14,10,12,21,11,12,14,10,21	1
15,13,25,29,16,17,11/14,13,11,10,23,11,12,14,11,20	1

15,13,23,30,17,15,12/?,12,11,12,23,13,12,15,11,19	1
15,12,23,28,17,14,13/20,12,10,11,20,14,12,14,11,20	1
15,13,23,31,16,14,11/15,13,10,10,23,11,13,14,11,20	1
15,13,23,30,17,14,14/?,14,11,12,21,11,10,15,10,21	1
16,14,25,31,16,16,11/15,13,10,10,23,11,12,14,11,20	1
15,13,25,30,14,17,12/15,13,11,10,23,11,12,14,11,20	1
16,13,24,30,17,14,17/19,13,10,13,20,11,13,14,10,20	1
15,13,26,29,15,15,11/14,13,11,10,23,11,12,15,11,20	1
16,13,24,29,15,14,11/14,12,10,13,23,13,11,15,12,19	1
15,13,23,30,16,16,12/?,13,11,12,23,13,12,15,11,19	1
16,13,22,29,19,15,10/16,13,9,11,21,14,11,16,10,19	1
17,14,23,30,17,14,13/15,12,10,10,21,11,9,15,9,20	1
17,12,24,29,15,16,12/15,13,10,10,23,11,12,14,11,20	1
16,13,22,30,18,15,10/16,13,9,11,21,14,11,16,10,19	1
15,13,23,30,16,14,15/16,13,10,11,20,13,11,14,9,19	1
16,14,23,30,19,15,14/18,13,11,11,22,13,11,14,9,19	1
16,14,26,31,15,15,11/14,13,11,10,23,11,12,14,11,20	1
16,14,21,30,15,16,16/18,14,10,12,21,11,11,14,11,21	1
14,13,23,30,19.2,14,14/18,12,10,11,21,11,11,14,10,20	1
16,13,23,30,15,14,16/18,13,10,12,20,11,13,14,10,20	1
17,13,24,30,16,13,16/18,14,10,13,21,11,11,14,10,20	1
15,14,23,29,16,13,16/17,15,10,9,21,13,11,14,8,19	1
15,14,23,30,16,14,14/20,12,10,12,21,11,11,14,9,21	1
15,14,23,29,17,14,13/20,13,10,11,27,10,11,14,11,19	1
15,14,23,29,17,14,13/20,14,10,11,27,10,11,14,11,19	1
14,13,23,30,18.2,14,13/17,12,11,11,22,11,11,14,10,19	1
16,13,23,30,16,14,13/19,12,10,12,22,12,13,15,9,22	1
15,13,23,31,15,14,15/20,13,9,12,21,10,11,15,9,20	1
16,13,22,30,18,15,16/?,13,10,12,21,11,11,14,10,20	1
15,13,23,29,20,15,14/16,13,10,12,22,17,11,14,11,21	1
15,13,23,29,17.2,14,13/18,13,10,11,21,11,11,14,10,20	1
15,12,22,28,15,14,13/16,11,10,13,23,14,12,15,10,19	1
14,13,24,30,16,16,13/17,13,10,11,21,11,12,14,10,19.2	1
14,13,25,31,15,17,11/14,13,11,10,23,11,13,14,11,20	1
18,13,25,31,15,15,12/15,13,10,11,23,11,14,14,11,20	1
14,13,23,29,20.2,14,12/19,12,11,11,21,11,11,14,10,20	1
Total	**119**

[a]Haplotypes based on alleles for the following Y-STR loci: DYS456, DYS389I, DYS390, DYS389II, DYS458, DYS19, DYS385a/b, DYS393, DYS391, DYS439, DYS635, DYS392, GATA_H4, DYS437, DYS438, DYS448.

Table 5. Y-chromosome haplotypes in Failaka Island

Y-STR Haplotype[a]	Frequency
13-13-23-30-16-14-13/17-12-10-11-21-11-10-16-9-20	1
14-13-23-29-16-14-12/16-13-10-11-23-11-11-15-9-21	1
15-13-24-30-15-13-16/?-14-11-12-21-11-12-14-10-20	1
14-13-23-29-15-15-14/15-12-10-12-21-11-12-15-9-21	2
15-12-22-28-17-15-13/14-15-10-11-21-11-11-16-10-20	1
18-13-25-31-15-16-11/14-13-11-10-23-11-13-15-11-20	1
15-14-24-31-18-13-16/17-12-10-12-23-11-11-14-10-20	1
14-14-24-31-19.2-14-13/18-12-11-11-21-11-11-14-10-20	1
15-14-25-31-15-14-16/18-13-9-12-22-11-12-14-10-20	1
15-12-22-28-20-15-12/15-13-10-11-21-11-11-16-10-21	1
15-13-23-29-17.2-14-13/19-12-10-11-21-11-11-14-10-20	1
15-13-23-30-17.2-14-13/20-12-10-11-20-11-11-14-10-20	1
?-12-24-28-20-13-12/16 12-10-11-20-11-11-15-9-19	2
15-12-21-29-18-15-13/?-16-10-11-21-11-11-16-10-24	1
15-13-24-32-18-13-14/15-14-10-10-21-11-12-14-10-20	1
15-13-24-29-15-15-13/19-12-10-13-21-11-11-15-9-20	1
16-13-21-29-16-14-13/16-12-10-11-24-14-12-15-11-19	1
15-14-24-30-15-14-11/12-12-10-12-24-13-12-15-12-19	1
15-11-22-27-14-14-17/18-13-10-13-20-11-13-14-10-20	1
15-13-23-30-17.2-14-13/19-12-10-11-20-11-11-14-10-20	1
15-12-24-28-15-15-13/16-15-10-11-23-14-12-14-11-18	1
15-13-23-31-17.2-14-14/19-12-11-11-20-11-11-14-10-20	1
Total	**24**

[a]Haplotypes based on alleles for the following Y-STR loci: DYS456, DYS389I, DYS390, DYS389II, DYS458, DYS19, DYS385a/b, DYS393, DYS391, DYS439, DYS635, DYS392, GATA_H4, DYS437, DYS438, DYS448.

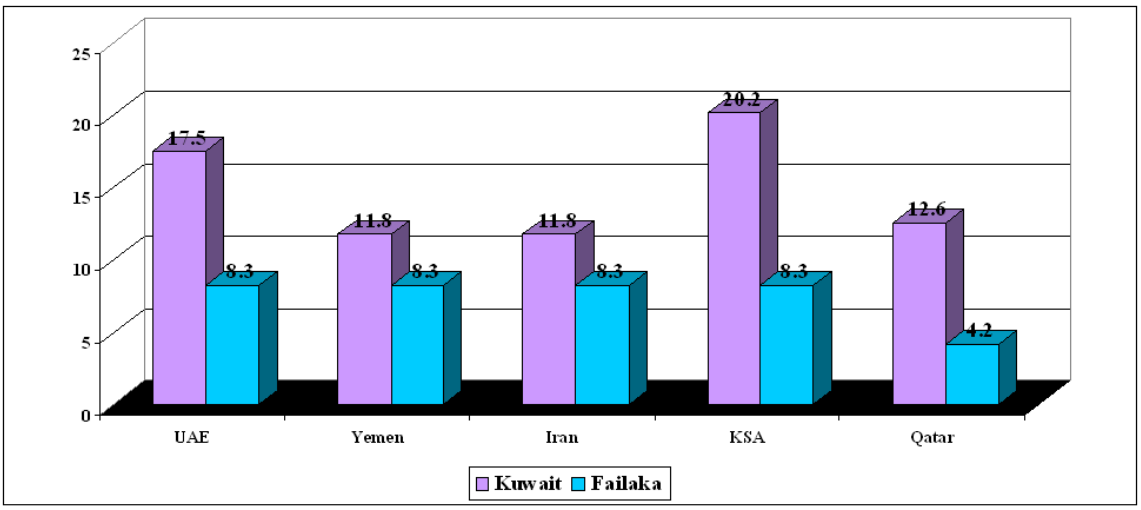

Figure 10. Percent of shared 15 loci Y-STR haplotypes between Kuwait and Failaka Island

Twenty percent of the Kuwaiti haplotypes are shared with Saudi Arabia, 17.5% are shared with United Arab Emirates, 12.6% are shared with Qatar, and 11.8% are shared with both Yemen and Iran. These shared haplotypes reflect the history of Kuwait. On the other hand, 8.3% percent of Failaka Island haplotypes are shared with Saudi Arabia, United Arab Emirates, Yemen, and Iran. Only 4.2% are shared with Qatar. According to the plot, there is a chance of bias in the shared haplotypes of Failaka due to the small sample size.

Table 6 shows the mean number of pairwise differences (MPD) and the mean number of gene diversity for Kuwait, Failaka Island, and neighboring populations. The MPD values are ranged from 5.521 for Qatar to 10.911 for Iran. The mean number of pairwise differences for Kuwait and Failaka Island is 8.951 and 8.775, respectively. While Iran has the highest mean number of gene diversity (0.69356), Qatar shows the lowest (0.37274). Both Kuwait and Failaka Island shared similar mean number of gene diversity (0.62578) and (0.64495), respectively.

Allele frequencies and estimated values of gene diversity for each locus of the17 Y-STR loci for Kuwait and Failaka Island populations are presented in table 7 and 8, respectively. Genetic diversity was greatest for the DYS385 locus in both populations, which is probably due to the duplicated nature of the DYS385 locus which includes two sub-loci DYS385a and DYS385b so that the genetic diversity values equivalent to two Y-STR loci. Figure 11 presents visual comparison of the frequency of genetic diversity for both Kuwait and Failaka Island populations.

Table 6. The mean number of pairwise differences and gene diversity

Population	MPD	Gene diversity
Kuwait	8.951859 (+/- 4.153213)	0.62578
Failaka Island	8.775362 (+/- 4.195306)	0.64495
Saudi Arabia	5.276717 (+/- 2.563487)	0.47404
Iran	10.911554 (+/- 4.974798)	0.69356
Yemen	5.942041 (+/- 2.883168)	0.39659
Qatar	5.521739 (+/- 2.702610)	0.37274
United Arab Emirates	7.121204 (+/- 3.373326)	0.46178

Table 7. Allele frequencies and estimated values of gene diversity (GD) of 17 Y-STR in the Kuwaiti population

	DYS456	DYS389I	DYS390	DYS389II	DYS458	DYS19	DYS393	DYS391	DYS439	DYS635	DYS392	GATA_H4	DYS437	DYS438	DYS448	Allelic Class	DYS385	Allelic Class	DYS385
8														0.008		7,16	0.008	17,18	0.025
9		0.008						0.076	0.008			0.008		0.151		10,16	0.017	17,19	0.008
10								0.513	0.118		0.076	0.017		0.555		11,12	0.008	18,22	0.008
10.2											0.008			0.008		11,14	0.059	20,21	0.008
11							0.034	0.403	0.521		0.706	0.613		0.235		11,15	0.017		
12		0.101					0.454	0.008	0.277		0.025	0.235		0.042		11,18	0.008		
13	0.025	0.647				0.059	0.387		0.076		0.092	0.118				11,20	0.008		
14	0.353	0.244			0.025	0.597	0.101				0.067	0.008	0.748			12,14	0.017		
15	0.378				0.176	0.185	0.025				0.017		0.168			12,15	0.025		
16	0.193				0.151	0.134							0.084			12,18	0.025		
17	0.042				0.168	0.025				0.008	0.008					12,19	0.025		
17.2					0.042											12,20	0.008		
18	0.008				0.084											13,14	0.008		
18.2					0.210											13,15	0.017		
19					0.017										0.244	13,16	0.034		
19.2					0.076										0.008	13,17	0.059		
20					0.008					0.118					0.588	13,18	0.059		
20.2					0.034											13,19	0.168		
21			0.017							0.445					0.118	13,20	0.025		
21.2					0.008					0.008						13,21	0.008		
22			0.084							0.118					0.034	13.2,19	0.008		
23			0.664							0.227					0.008	14,15	0.017		
24			0.126							0.008						14,16	0.050		
25			0.084	0.008						0.034						14,18	0.025		
26			0.025													14,19	0.025		
27										0.025						14,20	0.017		
28				0.025												15,16	0.008		
29				0.303												15,20	0.008		
30				0.429												16,17	0.008		
31				0.160												16,18	0.042		
32				0.076												16,19	0.008		
GD	0.698	0.516	0.533	0.699	0.783	0.593	0.638	0.574	0.637	0.717	0.477	0.559	0.409	0.611	0.577				0.957
NA	6	4	6	6	12	5	5	4	5	9	8	6	3	6	6			35	

GD: gene diversity; NA: allele number.

Table 8. Allele frequencies and estimated values of gene diversity (GD) of 17 Y-STR in Failaka Island

	DYS456	DYS389I	DYS390	DYS389II	DYS458	DYS19	DYS393	DYS391	DYS439	DYS635	DYS392	GATA_H4	DYS437	DYS438	DYS448	Allelic Class	DYS385
9								0.042						0.292		11,14	0.042
10								0.792	0.083			0.042		0.542		11,12	0.042
11		0.042						0.167	0.583		0.875	0.542		0.125		12,15	0.042
12		0.250					0.583		0.250			0.333		0.042		12,16	0.125
13	0.042	0.542				0.208	0.208		0.083		0.042	0.083				13,16	0.083
14	0.167	0.167			0.042	0.458	0.083				0.083		0.458			13,17	0.042
15	0.625				0.333	0.292	0.083						0.375			13,14	0.042
16	0.042				0.125	0.042	0.042						0.167			13,18	0.042
17					0.042											13,19	0.125
17.2					0.167											13,20	0.042
18	0.042				0.125										0.042	14,15	0.083
19															0.167	14,19	0.042
19.2					0.042											16,17	0.042
20					0.125					0.250					0.583	16,18	0.042
21			0.083							0.417					0.167	17,18	0.042
22			0.125							0.083							
23			0.333							0.167							
24			0.375							0.083					0.042		
25			0.083														
27				0.042													
28				0.208													
29				0.292													
30				0.208													
31				0.208													
32				0.042													
GD	0.519	0.641	0.750	0.815	0.784	0.688	0.627	0.359	0.609	0.754	0.236	0.612	0.649	0.630	0.627		0.976
NA	5	4	5	6	8	4	5	3	4	5	3	4	3	4	5	15	

GD: gene diversity; NA: allele number.

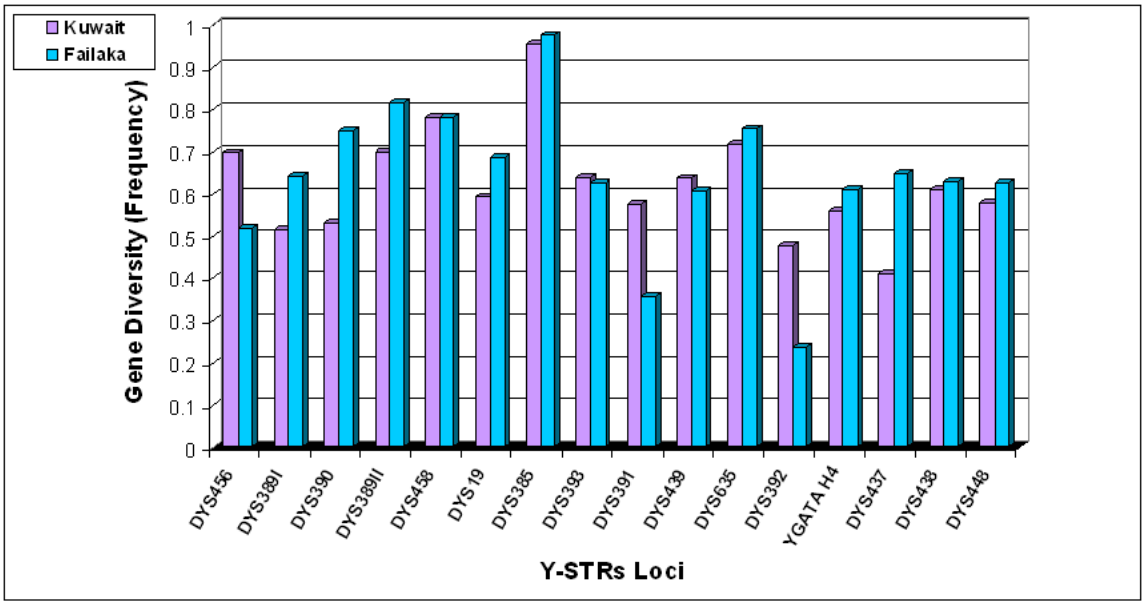

Figure 11. Gene diversity of the 16 loci for Kuwait and Failaka Island populations

For the Kuwaiti population, the lowest value of GD was observed for DYS437 (0.409), while the highest one (0.957) is presented by DYS385. The lowest genetic diversity value for Failaka Island population found in DYS392 (0.236), and the highest one (0.976) is presented by DYS385.

Y-chromosome Haplogroup Frequencies

A summary of the results of Y-STR haplotypes, Y haplogroup predictions, and their probabilities for the Kuwaiti and Failaka Island populations are shown in Table 9 and 10, respectively. Eight samples out of the total 119 Kuwaiti samples were not assigned to a specific haplogroup. The reminder were assigned to one of the following haplogroups with probability rate exceed 80%: J1, R1a, R1b, E1b1b, E1b1a, T, L, Q, H, G2a, J2b, J2a1b, J2a1h, and J2a1xJ2. The most common haplogroups in Kuwait are J1 (37%), R1a (11%), and E1b1b (7%). The Y-chromosome haplogroup frequencies of the Kuwaiti population are presented as pie charts in figure 12.

Twenty-two individuals were assigned to one of the following haplogroups with 80% probability rate: E1b1b, J2b, J1G2a, and J2a1xJ2, J2b, R1b, R1a, L, and Q. Only 2 samples were not assigned to a specific haplogroup. There is a good Y-haplogroups variation in Failaka for such a small sample number. The most common haplogroups in Failaka are J1 (20%), and E1b1b (17%). Each of the following haplogroups contributes by 13% of Failaka Island population gene pool: G2a, J2b, and J2a1xJ2. The Y-chromosome haplogroup frequencies of Failaka Island population are presented as a pie chart in Figure 13.

Table 9. Y-STR haplotypes, predicted haplogroups, and probability of haplogroup assignment (x 100) for the Kuwaiti population. A "batch" version of Whit Athey's haplogroup Predictor was used to obtain the haplogroups and their assigned probabilities (http://www.hprg.com/hapest5/index.html) (Athey, 2006). A (-) indicates missing data.

DYS456	*DYS389I*	*DYS390*	*DYS389II*	*DYS458*	*DYS19*	*DYS385a*	*DYS385b*	*DYS393*	*DYS391*	*DYS439*	*DYS635*	*DYS392*	*GATA_H4*	*DYS437*	*DYS438*	*DYS448*	**Haplogroup**	*Probability*
15	14	23	30	18.2	14	11	18	12	10	13	20	11	11	14	10	19	**J1**	99.3
14	13	23	30	17.2	14	13	21	12	11	11	21	11	11	14	10	20		100
14	14	23	32	18.2	14	14	19	12	11	11	21	11	11	14	10	20		99.9
14	13	23	30	19.2	14	13	17	12	10	11	21	11	11	14	10	20		99.8
14	13	23	30	18.2	14	13	19	12	11	12	21	11	11	14	10	20		100
14	13	23	29	18.2	14	13	18	12	11	11	21	11	11	14	10	20		99.9
14	14	23	31	18.2	14	13.2	19	12	11	11	21.2	10.2	11	14	10.2	20		100
15	14	22	30	20.2	14	12	18	12	10	13	22	11	11	14	10	21		99.6
15	13	23	30	18.2	14	13	18	12	11	11	21	11	11	14	10	20		99.3
15	13	23	30	18.2	14	13	19	12	10	11	21	11	11	14	10	20		99.9
14	13	23	31	18.2	14	13	19	12	10	11	20	11	11	14	10	20		100
14	13	23	30	18.2	14	13	19	12	11	11	22	11	11	14	10	20		99.9
14	13	23	30	18.2	14	13	19	12	11	11	21	11	11	14	10	20		100
14	13	23	30	18.2	14	12	20	12	11	12	21	11	11	14	10	20		100
14	13	23	30	19.2	14	12	18	12	11	12	21	11	11	14	10	20		99.9
14	13	23	30	18.2	14	13	19	12	11	11	21	11	11	14	10	20		100
14	13	23	30	19.2	14	13	18	12	11	11	21	11	11	14	10	21		99.9
14	13	23	30	21.2	14	13	19	12	10	11	22	11	11	14	10	20		100
15	13	23	29	17.2	16	13	19	12	10	11	21	11	11	14	10	20		99.8
14	13	23	30	18.2	14	13	17	13	11	11	21	11	11	14	10	20		98.2
14	13	23	30	18.2	14	12	18	12	10	11	20	11	11	14	10	20		100
14	13	23	29	20.2	14	13	18	12	11	11	20	11	11	14	10	20		100
14	13	23	30	18.2	14	14	19	12	11	11	21	11	11	14	10	20		99.9
14	13	23	30	18.2	14	13	19	12	11	11	21	11	11	14	10	20		100
15	13	23	31	19.2	15	13	19	12	11	11	21	11	11	14	10	20		99.9
14	13	23	30	18.2	14	13	17	13	11	11	22	11	11	14	10	20		89.9
15	13	23	32	17.2	15	13	17	12	10	11	20	11	11	14	10	20		99.9
14	13	23	31	18.2	14	13	17	13	11	11	22	11	11	14	10	20		85.5
14	13	23	30	18.2	14	13	19	12	11	11	21	11	11	14	10	20		100
14	13	23	29	17.2	14	14	19	12	10	11	23	11	11	14	10	20		93.1
14	13	23	30	18.2	14	13	19	12	11	11	21	11	11	14	10	20		100
14	13	23	30	18.2	14	12	19	12	11	12	21	11	11	14	10	20		100
14	13	23	30	20.2	14	14	20	12	10	11	21	11	11	14	10	20		100
14	13	23	29	18.2	14	13	18	12	11	11	21	11	11	14	10	20		99.9
14	13	23	29	19.2	14	13	19	12	11	11	21	11	11	14	10	20		100
14	13	23	29	19.2	14	13	18	12	11	11	22	11	11	14	10	20		99.8
14	13	23	30	18.2	14	12	19	12	11	12	21	11	11	14	10	20		100
14	13	23	30	18.2	14	13	19	12	11	11	21	11	11	14	10	20		100
14	13	23	30	19.2	14	14	18	12	10	11	21	11	11	14	10	20		99.8
14	13	23	30	18.2	14	13	17	12	11	11	22	11	11	14	10	19		98.3
14	13	23	30	19.2	14	13	19	12	11	12	21	11	11	14	10	20		100
15	13	23	29	17.2	14	13	18	13	10	11	21	11	11	14	10	20		98.9
14	13	23	29	20.2	14	12	19	12	11	11	21	11	11	14	10	20		100
15	12	23	28	17	14	13	20	12	10	11	20	14	12	14	11	20		83.4
14	13	23	30	19.2	14	13	19	12	10	11	22	11	11	14	10	20		100
15	13	24	32	18	15	11	14	13	12	10	23	11	13	14	11	20	**R1a**	100
15	13	26	29	15	16	11	14	13	11	10	23	11	12	15	11	20		100
15	14	25	32	15	16	11	-	13	10	10	23	11	11	14	11	20		100
16	13	25	29	15	16	11	12	13	11	10	23	11	12	14	11	20		100
15	13	25	29	16	17	11	14	13	11	10	23	11	12	14	11	20		100
15	14	25	32	15	16	11	-	13	10	10	23	11	11	14	11	20		100
15	13	23	31	16	14	11	15	13	10	10	23	11	13	14	11	20		100
16	14	25	31	16	16	11	15	13	10	10	23	11	12	14	11	20		100
15	13	25	30	14	17	12	15	13	11	10	23	11	12	14	11	20		100

15	13	26	29	15	15	11	14	13	11	10	23	11	12	15	11	20		100
17	12	24	29	15	16	12	15	13	10	10	23	11	12	14	11	20		100
16	14	26	31	15	15	11	14	13	11	10	23	11	12	14	11	20		100
14	13	25	31	15	17	11	14	13	11	10	23	11	13	14	11	20		100
18	13	25	31	15	15	12	15	13	10	11	23	11	14	14	11	20		100
15	14	23	32	15	13	18	22	13	11	11	21	11	13	14	10	19	**E1b1b**	100
15	14	24	32	18	13	17	18	15	10	12	21	11	13	14	10	20		100
17	13	24	31	18	14	16	18	13	10	12	21	11	11	14	10	20		100
15	12	25	30	17	15	16	19	13	11	12	21	11	12	14	10	21		98.6
16	13	24	30	15	13	16	18	14	10	12	21	11	12	14	10	21		100
16	13	24	30	17	14	17	19	13	10	13	20	11	13	14	10	20		100
16	13	23	30	15	14	16	18	13	10	12	20	11	13	14	10	20		100
17	13	24	30	16	13	16	18	14	10	13	21	11	11	14	10	20		100
15	13	23	29	16	14	12	14	11	11	12	23	14	13	15	12	19	**R1b**	99.9
15	14	25	30	17	16	13	14	14	10	12	23	13	13	14	12	19		93.7
15	13	23	30	17	15	12	-	12	11	12	23	13	12	15	11	19		99.7
15	14	23	31	17	15	11	20	14	10	11	25	10	13	16	12	19		81.1
16	13	24	29	15	14	11	14	12	10	13	23	13	11	15	12	19		100
15	13	23	30	16	16	12	-	13	11	12	23	13	12	15	11	19		97.8
15	13	23	29	16	14	12	14	11	11	12	23	14	13	15	12	19		100
16	9	23	25	15	14	14	15	14	11	11	20	12	11	16	10	21	**G2a**	99.9
15	12	23	30	16	14	14	16	13	10	13	21	13	11	16	10	21		97.5
17	12	23	29	16	15	14	16	14	10	11	21	12	13	16	10	22		100
14	12	22	29	15	15	14	15	14	10	12	20	11	12	15	10	21		100
15	13	22	29	18	13	14	16	13	10	12	22	15	10	14	11	19	**Q**	100
15	14	23	30	17	13	14	18	14	10	11	23	15	12	14	11	20		100
15	14	23	29	18	14	13	19	14	10	11	27	10	11	14	11	19		91.9
16	14	23	31	18	14	13	-	13	11	13	23	13	12	14	11	19		90.3
15	14	23	29	17	14	13	20	13	10	11	27	10	11	14	11	19		95.2
15	14	23	29	17	14	13	20	14	10	11	27	10	11	14	11	19		92
15	13	23	29	20	15	14	16	13	10	12	22	17	11	14	11	21		100
16	12	22	28	14	15	14	16	11	10	12	25	14	12	15	10	19	**L**	100
15	13	24	29	16	14	13	19	13	10	11	25	10	11	16	11	20		91.9
16	14	22	31	18	15	7	16	12	10	13	23	14	11	16	10	19		100
16	13	22	29	19	15	10	16	13	9	11	21	14	11	16	10	19		100
16	13	22	30	18	15	10	16	13	9	11	21	14	11	16	10	19		100
15	12	22	28	15	14	13	16	11	10	13	23	14	12	15	10	19		100
16	13	23	29	17	14	13	16	13	10	11	22	13	11	14	9	19	**T**	100
15	14	23	31	17	14	14	16	13	10	11	21	13	11	14	9	19		100
15	13	23	30	16	14	15	16	13	10	11	20	13	11	14	9	19		100
16	14	23	30	19	15	14	18	13	11	11	22	13	11	14	9	19		100
15	14	23	29	16	13	16	17	15	10	9	21	13	11	14	8	19		99.9
15	14	23	30	16	14	14	20	12	10	12	21	11	11	14	9	21	**J2a1xJ2**	98.2
16	13	23	30	16	14	13	19	12	10	12	22	12	13	15	9	22		85.4
15	13	23	31	15	14	15	20	13	9	12	21	10	11	15	9	20		91.4
15	14	23	29	15	15	13	15	13	10	12	21	11	12	15	9	22		83.2
16	14	21	31	17	16	17	18	15	10	12	21	11	11	14	11	20	**E1b1a**	100
13	14	23	32	17	15	11	-	13	10	12	17	11	13	14	11	23		100
16	14	21	30	15	16	16	18	14	10	12	21	11	11	14	11	21		100
17	14	23	30	17	14	13	15	12	10	10	21	11	9	15	9	20	**J2a1b**	86.1
15	13	22	29	16	14	13	16	12	9	11	23	11	11	15	9	21		84.4
16	14	23	32	17	14	13	-	13	10	11	25	10	12	16	9	20		83
13	12	24	30	16	15	17	18	12	11	12	21	11	11	14	9	19	**J2b**	87
15	13	24	29	14	15	13	16	13	10	12	22	11	12	15	9	21	**J2a1h**	99.2
14	12	24	29	17	16	20	-	12	9	11	20	11	12	14	9	19	**H**	88.2
14	12	24	29	17	16	20	-	12	9	11	20	11	12	14	9	19		88.2
14	12	24	29	17	16	20	21	12	9	11	20	11	12	14	9	19		97.1
16	14	23	31	17	14	13	-	13	10	11	24	10	12	16	9	20	J2a1b/L	42/41.6
15	13	23	29	16	14	16	-	12	9	12	23	11	12	15	10	22	**E1b1b/ J2a1xJ2**	66.4/17.7
13	13	23	31	18	16	11	-	13	10	11	-	11	11	14	10	21	**I2a1/I2a(x I2)**	60.5/38.5
16	13	23	31	15	14	15	-	13	9	11	21	10	12	15	9	20	**J2a1b/ J2a1xJ2**	64.8/26.5

16	13	23	29	15	14	13	19	12	10	12	21	11	11	15	10	20	**J2a1b/ J2a1xJ2**	51.7/39.1
15	13	23	30	17	14	14	-	14	11	12	21	11	10	15	10	21	**G2a/ I2a(xI2)**	38/23.8
16	13	22	30	18	15	16	-	13	10	12	21	11	11	14	10	20	**E1b1b/ E1b1a**	48.2/36.6
14	13	24	30	16	16	13	17	13	10	11	21	11	12	14	10	19.2	**E1b1b/ I2a(xI2)**	64/13.2

Table 10. Y-STR haplotypes, predicted haplogroups, and probability of haplogroup assignment (x 100) for Failaka Island populations. A "batch" version of Whit Athey's haplogroup Predictor was used to obtain the haplogroups and their assigned probabilities (http://www.hprg.com/hapest5/index.html) (Athey, 2006). A (-) indicates missing data.

DYS456	***DYS389I***	***DYS390***	***DYS389II***	***DYS458***	***DYS19***	***DYS385a***	***DYS385b***	***DYS393***	***DYS391***	***DYS439***	***DYS635***	***DYS392***	***GATA_H4***	***DYS437***	***DYS438***	***DYS448***	**Haplogroup**	***Probability***
15	13	24	30	15	13	16	-	14	11	12	21	11	12	14	10	20	**E1b1b**	100
15	14	24	31	18	13	16	17	12	10	12	23	11	11	14	10	20		100
15	14	25	31	15	14	16	18	13	9	12	22	11	12	14	10	20		100
15	13	24	32	18	13	14	15	14	10	10	21	11	12	14	10	20		99.6
13	13	23	30	16	14	13	17	12	10	11	21	11	10	16	9	20	**J2b**	99.2
-	12	24	28	20	13	12	16	12	10	11	20	11	11	15	9	19		91.6
-	12	24	28	20	13	12	16	12	10	11	20	11	11	15	9	19		91.6
14	14	24	31	19.2	14	13	18	12	11	11	21	11	11	14	10	20	**J1**	99.4
15	13	23	29	17.2	14	13	19	12	10	11	21	11	11	14	10	20		100
15	13	23	30	17.2	14	13	20	12	10	11	20	11	11	14	10	20		100
15	13	23	30	17.2	14	13	19	12	10	11	20	11	11	14	10	20		100
15	13	23	31	17.2	14	14	19	12	11	11	20	11	11	14	10	20		100
15	12	22	28	17	15	13	14	15	10	11	21	11	11	16	10	20	**G2a**	99.5
15	12	22	28	20	15	12	15	13	10	11	21	11	11	16	10	21		100
15	12	21	29	18	15	13	-	16	10	11	21	11	11	16	10	24		100
14	13	23	29	15	15	14	-	12	10	12	22	11	12	15	9	21	**J2a1xJ2**	99.7
15	13	24	29	15	15	13	19	12	10	13	21	11	11	15	9	20		96
14	13	23	29	15	15	14	15	12	10	12	21	11	12	15	9	21		98.6
15	12	24	28	15	15	13	16	15	10	11	23	14	12	14	11	18	**Q**	97.6
16	13	21	29	16	14	13	16	12	10	11	24	14	12	15	11	19	**L**	99.9
18	13	25	31	15	16	11	14	13	11	10	23	11	13	15	11	20	**R1a**	100
15	14	24	30	15	14	11	12	12	10	12	24	13	12	15	12	19	**R1b**	100
14	13	23	29	16	14	12	16	13	10	11	23	11	11	15	9	21	**J2a1xJ/J2a1b**	60.8/39
15	11	22	27	14	14	17	18	13	10	13	20	11	13	14	10	20	**E1b1b/L**	74.9/19.6

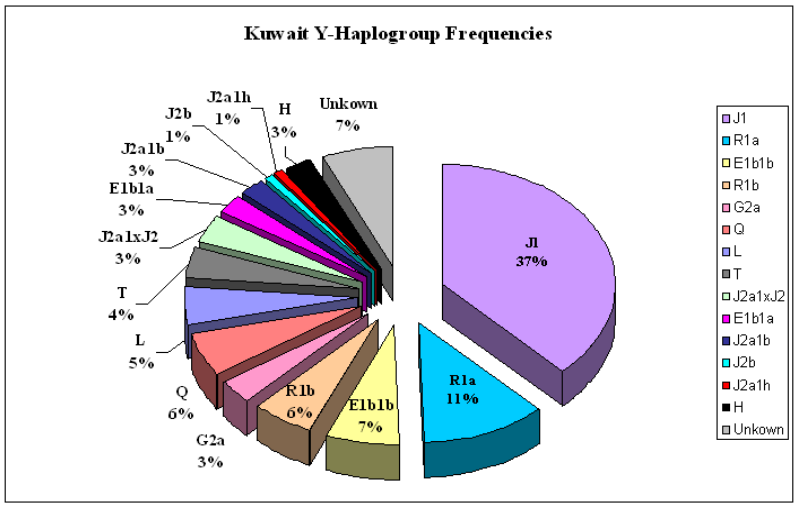

Figure 12. Y-chromosome haplogroup frequencies of the Kuwaiti Population

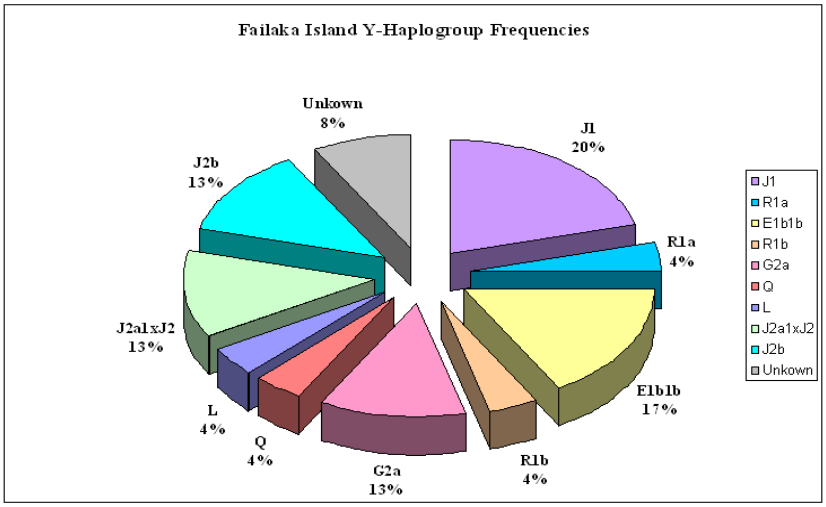

Figure 13. Y-chromosome haplogroup frequencies of the Failaka Island Population

Haplogroup J1 is the most abundant among Kuwait and Failaka Island population. While haplogroup E1b1b account for 17% of Failaka Island haplogroup frequencies, it accounts for only 7% percent of the Kuwaiti Y-haplogroup gene pool. Interestingly, Y-haplogroups J2b, J2a1xJ2, G2a are found in high frequencies in Failaka Island (13%) compared to Kuwait (1%), (3%), and (3%), respectively. Haplogroups H and T found in Kuwait in 3% and 4%, respectively, and not present from Failaka Island.

To examine how the Kuwaiti Y-haplogroup frequencies distributed among the Kuwaiti subpopulations, three pie charts were constructed and each one represents one of the following subpopulations: Arab, Bedouin, and Iranian ethnicities (Figure 14).

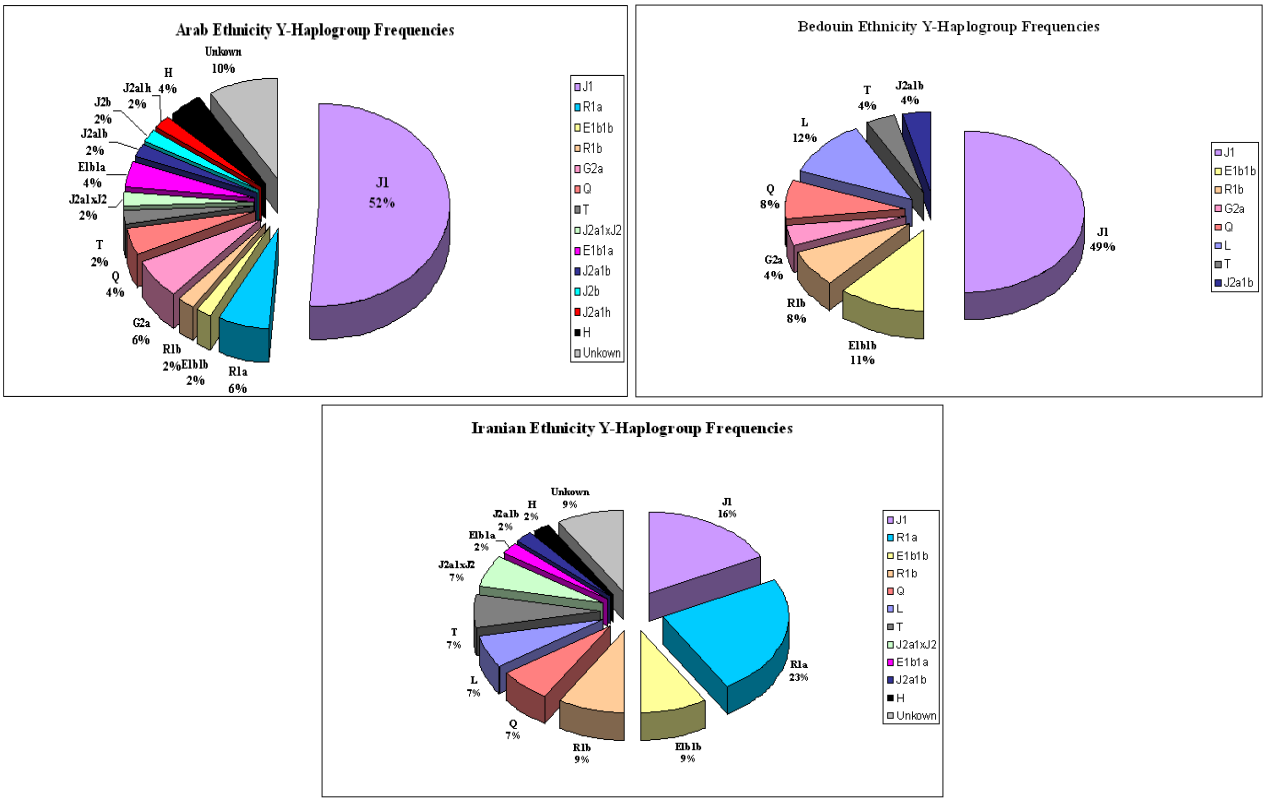

Figure 14. Y-haplogroup frequencies in each subpopulation

Haplogroup J1 is the most common haplogroup among Arabs and Bedouins 52%, 49% respectively, and it made 16% of the total Iranian subpopulation haplogroup frequencies. The Iranian ethnicity is characterized by the relatively high percentage of the haplogroup R1a (16%) which is found in Arab ethnicity at low frequency (6%), and is absent in the Bedouin Y-chromosome gene pool. While the haplogroup E1b1b and R1b shared in frequencies ranged from 8% to 11% among both Iranian and Bedouin ethnicities, each one of the haplogroups represent 2% of the Arabian Y-haplogroup gene pool. Haplogroup G2a is a unique haplogroup among Bedouin (4%) and Arab (6%) ethnicities, and not found among the Iranian ethnicity. Bedouin and Iranian ethnicities

shared the haplogroup L by 12% and 9%, respectively. E1b1a and J2a1xJ2 haplogroups found among both Arab and Iranian ethnicities but not in the Bedouin gene pool. J2b and J2a1h haplogroups are unique to the Arabian ethnicity and each one of them represents 2% of the total Y-haplogroup gene pool. The unassigned individuals represent 9% and 10% of the Arab and Iranian total sample, respectively.

Y-chromosome Median-Joining Tree

To visualize the diversity in Kuwait and Failaka Island, a median-joining tree was constructed based on Y-STR haplogroup variation in both populations. This tree is consisted of the most abundant haplogroup in both populations which is J1 Y-haplogroup figure 15. The network is based on 49 haplotypes, only 4 of them representing Failaka Island and they are represented in blue circles. The largest central node represents 6 Y-chromosome haplotypes.

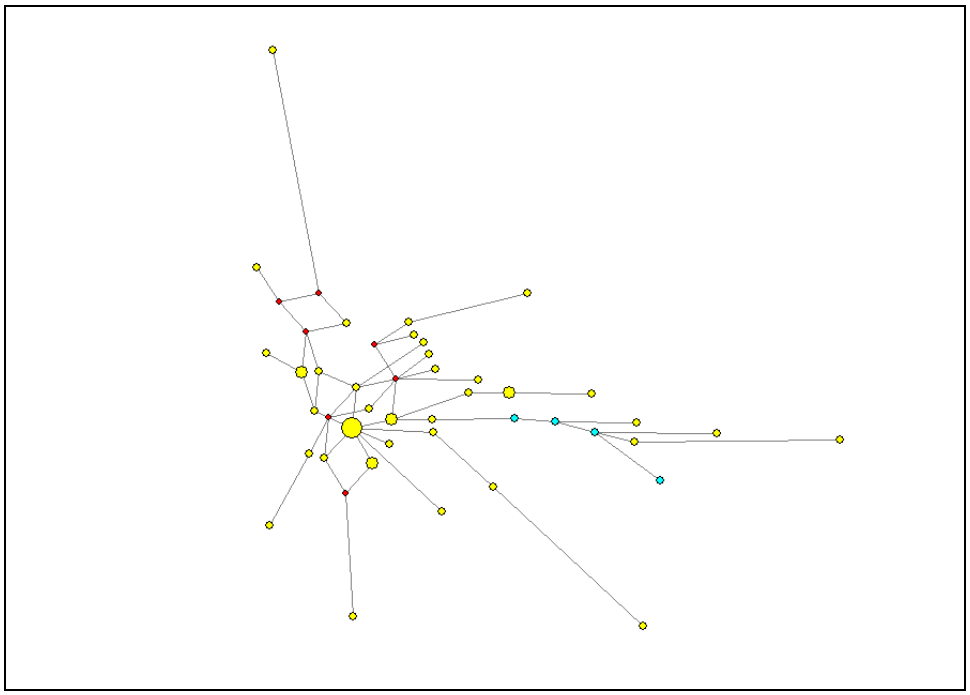

Figure 15. Median-Joining Network for Y-chromosome haplogroup J1 haplotypes in Kuwait (yellow circles) and Failaka Island (blue circles)

Neighbor-Joining Tree

A neighbor-joining tree based on a R_{ST} (sum of squared size differences) distance matrix for Kuwait, Failaka and comparative populations using Y-STR data present in figure 16. The African populations (Namibia (Ovambo), Guinea, and Mozambique) are all grouped together at the top of the tree. The European populations (Ukraine, Croatia, and Austria) grouped together. The rest of the populations shared a node and ended by grouping Qatar and Yemen together. Kuwait clustered between Saudi Arabia and Failaka Island which close to Iran. A two-way Mantel test was used to test the correlation between the original distance matrix and the cophenetic distance matrix which is represented in the tree. The correlation had an insignificant value of 0.432 which indicates that the tree is a poor fit to the data.

Multidimensional Scaling

A multidimensional scaling (MDS) method is a good alternative for neighbor-joining tree to represent the relationship of the Kuwaiti and Failaka Island populations to other comparative populations. Figure 17 shows a multidimensional scaling plot which is based on a R_{ST} (sum of squared size differences) distance matrix.

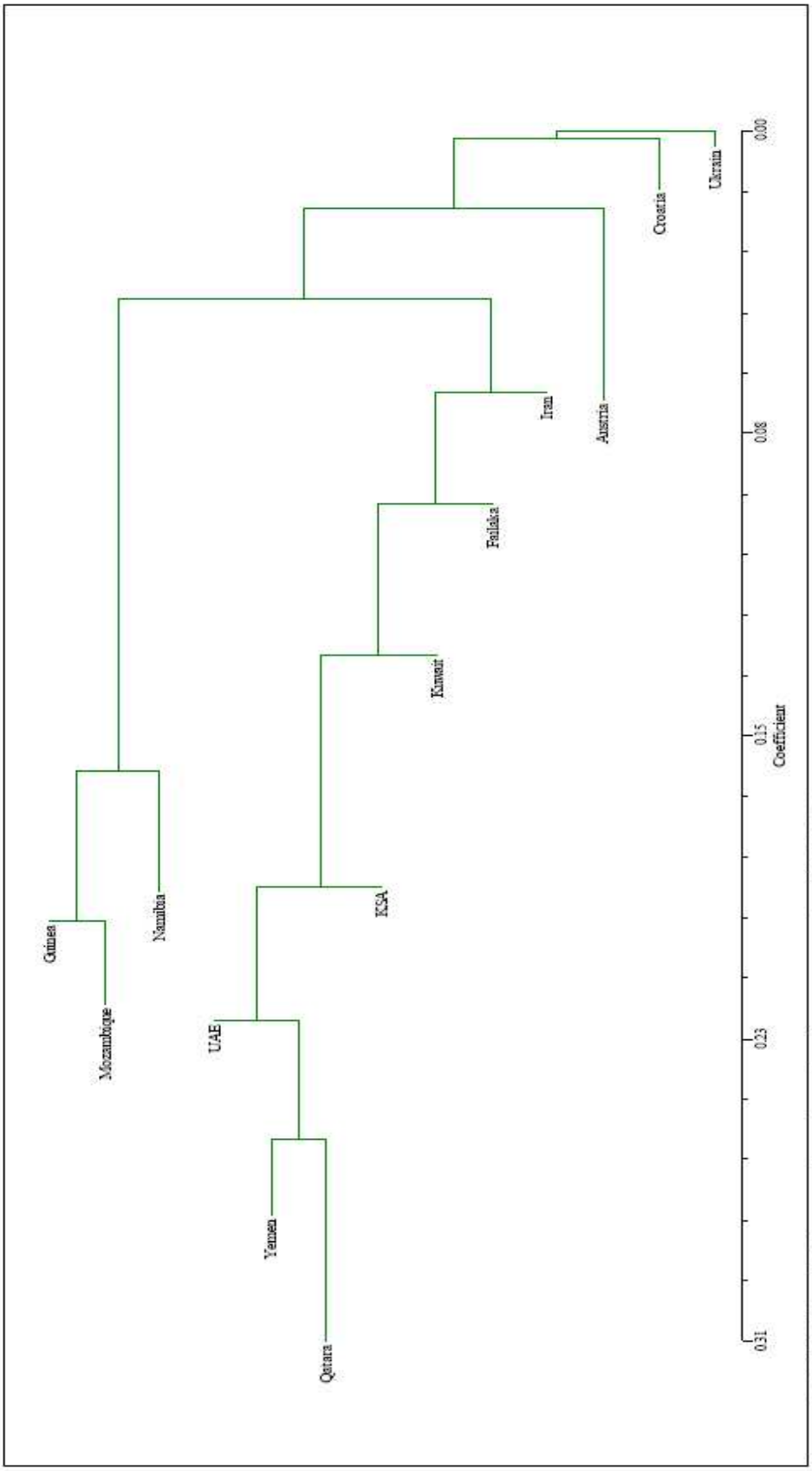

Figure 16. Neighbor-Joining tree based on R_{ST} distance matrix of Y-STR data

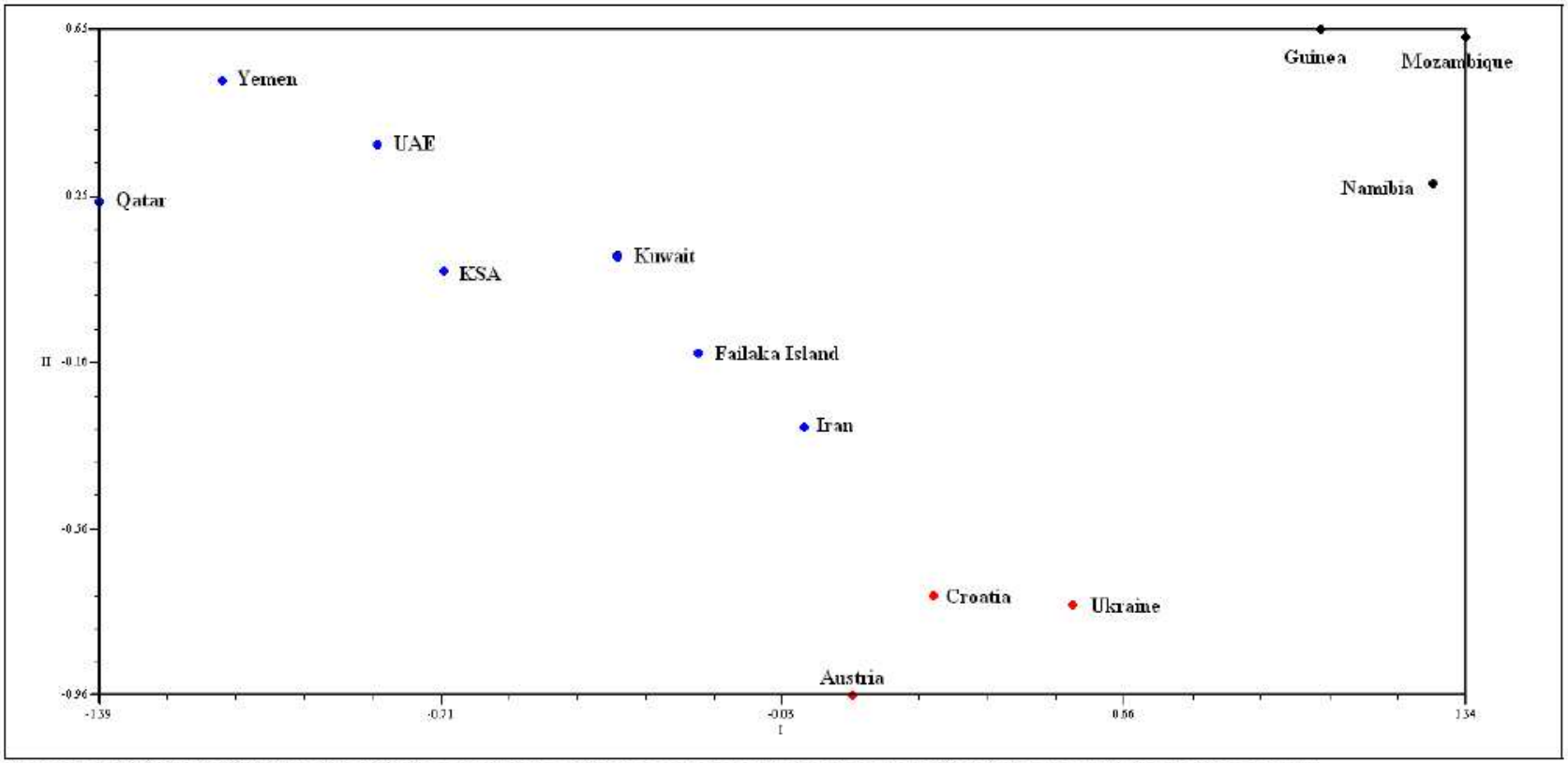

Figure 17. MDS plot of Y-STR data with 13 populations. African populations are black circles; West Asian populations are blue circles; European populations are red circles.

The MDS plot shows three major clusters. First cluster composed of the African population (Namibia (Ovambo), Guinea, and Mozambique) and they are represented in black circles. The second cluster includes the European populations (Ukraine, Croatia, and Austria) and they are represented in red circles. The final cluster contains the West Asian populations including: Iran, Kuwait, Failaka Island, Saudi Arabia (KSA), United Arab Emirates (UAE), Qatar, and Yemen. The West Asian populations are represented in blue circles. Failaka Island is located in the middle between Kuwait and Iran, and Kuwait close to Saudi Arabia. Qatar and Yemen are close to each other. The MDS plot reflects the geographical location of both Kuwait and Failaka Island, and the contribution of gene flow from the surrounding countries in the Kuwaiti and Failaka Island gene pools. The stress value of the MDS plot is good (0.15180). The original distance matrix was compared to a MDS matrix using a two-way Mantel test. The correlation was highly significant with value of 0.95224. This value indicates that the MDS plot is a good fit to the data.

To determine the relationship of the Kuwaiti subpopulations to Failaka Island and surrounding populations, another MDS plot was created (Figure 18). There are 5 clusters presented in the MDS plot. As the first plot, the first and second clusters contain the African populations (Namibia (Ovambo), Guinea, and Mozambique) and the European populations (Poland, Germany, Ukraine, Croatia, and Austria). The third cluster includes Iran, Failaka Island, and the Iranian subpopulation. Saudi Arabia with the Arab and Bedouin subpopulations formed the fourth cluster. The final cluster contains Qatar, Yemen, and United Arab Emirates. The stress value of the MDS plot is good (0.17711). The correlation was highly significant with value of 0.95518.

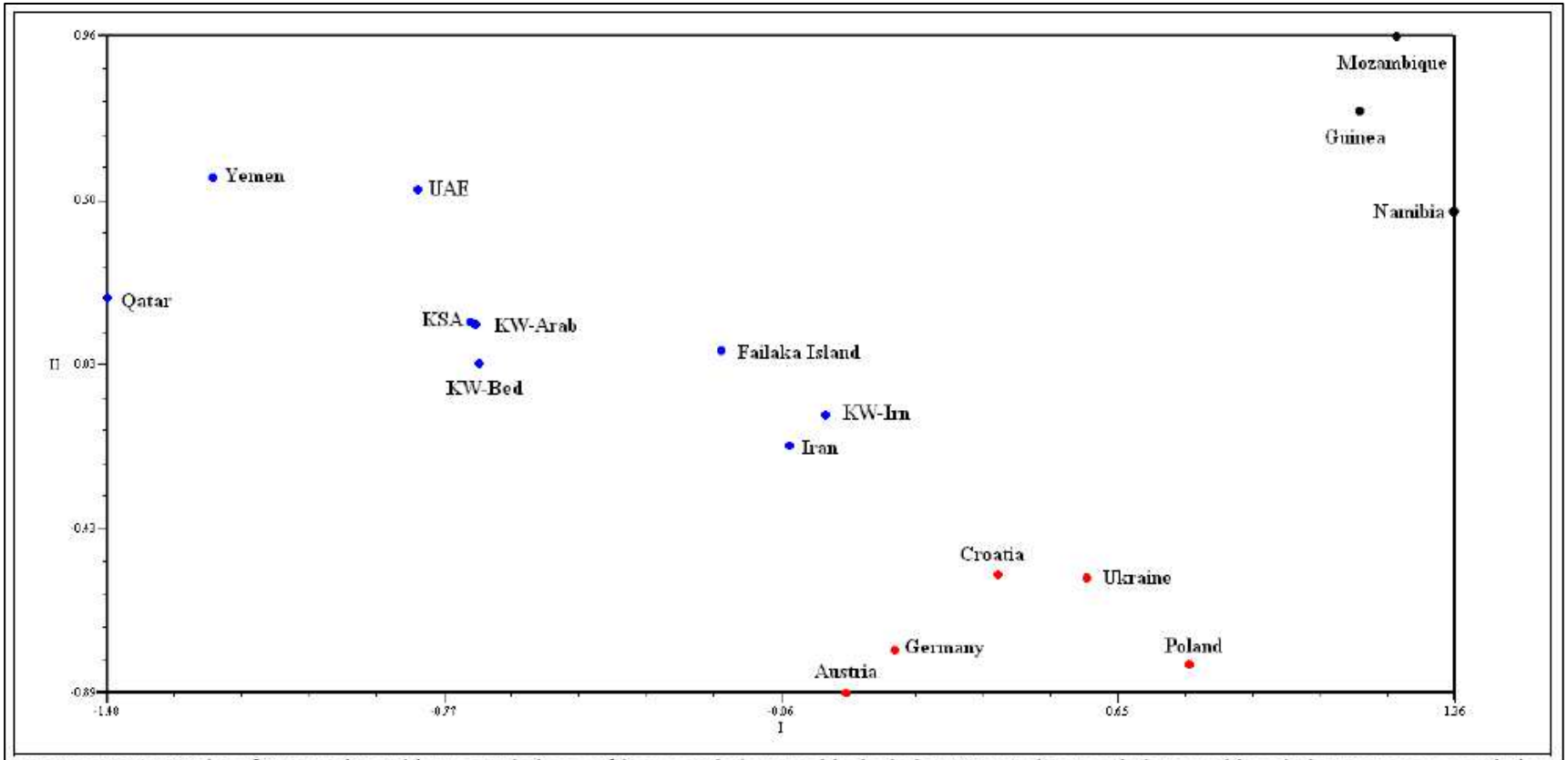

Figure 18. MDS plot of Y-STR data with 15 populations. African populations are black circles; West Asian populations are blue circles; European populations are red circles.

Gene Diversity vs. Distance from the Centroid (r_{ii})

The plot of heterozygosity vs. r_{ii} based on Y-chromosome STR data for 15 populations is presented in figure 19. Populations of Kuwait, Saudi Arabia (KSA), Poland, Austria, Germany, Iran, Qatar, and Guinea are located above the hypothetical regression line. This pattern indicates that these populations experienced more male gene flow and consequently more genetic diversity. The following populations are located below the theoretical regression line: Failaka Island, United Arab Emirates (UAE), Yemen, Croatia, Ukraine, Namibia, and Mozambique. This suggests that these populations experienced more genetic drift compared to gene flow which leads to low Y-chromosome diversity. Kuwait, Saudi Arabia (KSA), and Poland are located in the upper left corner, suggesting that they may have experienced more gene flow than the other populations.

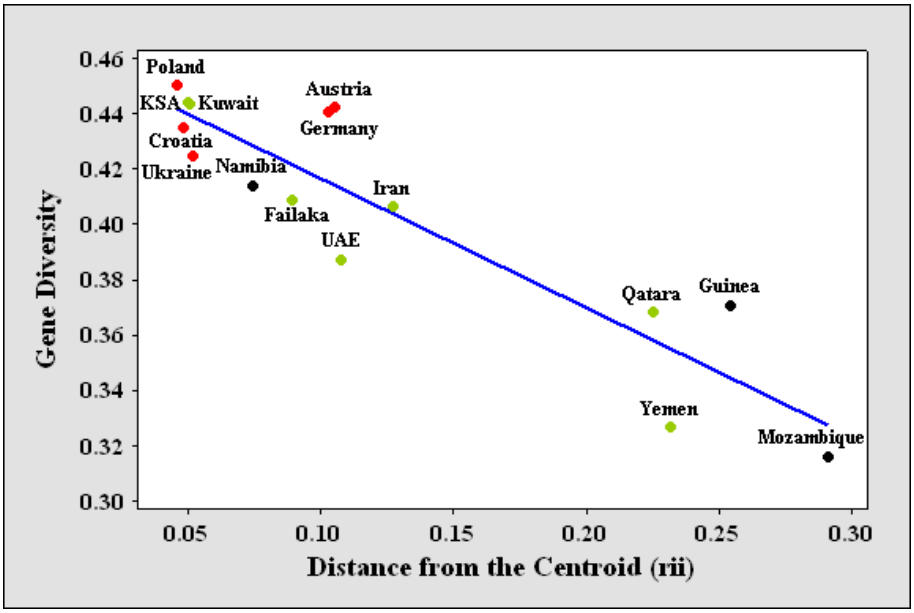

Figure 19. Y-STR Diversity vs. r_{ii} plot. African populations are black circles; West Asian populations are green circles; European populations are red circles.

Analysis of Molecular Variance (AMOVA)

Analysis of Molecular Variance was performed to determine the genetic structure of 15 populations based on the Y-chromosome STR haplotypes. Populations were classified according to their geographical regions. The first group represents the African populations: Namibia (Ovambo), Guinea, and Mozambique. The second group contains the European populations: Germany, Poland, Austria, Croatia, and Ukraine. The West Asian populations represent the last group: Iran, Kuwait, Failaka Island, United Arab Emirates, Qatar, Saudi Arabia, and Yemen. The result of AMOVA analysis is shown in table 11. The greatest amount of variation was observed among groups and accounted for 65.39% of the total variation, followed by within populations variation (22.98%), and finally within groups variation (11.63%). All variance components are significant.

Table 11. AMOVA of Y-STR data in 15 populations, grouped by geographic region (Africa, Asia, and Europe).

Source of Variation	d.f.	Sum of Squares	Variance Components	Percentage of Variation	F-Statistics
Among groups	2	41945.478	23.28185	65.39	Φ_{CT}=0.65390*
Within groups	12	13480.151	4.14072	11.63	Φ_{SC}=0.33602*
Within populations	4389	35911.149	8.18208	22.98	Φ_{ST}=0.77020*
Total	4403	91336.777	35.60465		

*P-value = 0.000

Mitochondrial DNA (mtDNA)

Sequence Diversities and Neutrality Test Statistics

Table 12 shows the diversity levels and neutrality test scores for the HVS-I data of Failaka Island and comparative populations. While England has the lowest genetic diversity (0.9651), the highest genetic diversity was observed in Kenya (0.9960). The measure of genetic diversity for Failaka Island (0.9886) is similar to Kuwait (0.9799) and the rest of neighboring populations, Iran (0.9895), Iraq (0.9918), and Saudi (0.9905). The gene diversities measures for the rest of populations fall between 0.9960 and 0.9651. For the neutrality test statistics, Failaka Island population shows a significant negative Tajima's D (-1.94476) accompanied with a significant Fu's Fs (-15.68033). These results indicate that Failaka Island population may be undergoing an expansion in relatively short period of time.

Table 12. Summary statistics for Failaka Island 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, mean number of pairwise differences between haplotypes (π).

Population	N	H	Nucleotide diversity	π	Tajima's D	Fu's FS
Nigeria	63	0.9949(0.0039)	0.0239(0.012661)	6.582(3.151327)	-1.40382	-25.12797*
Kenya	78	0.9960(0.0029)	0.032374(0.016713)	8.903(4.147357)	-1.36419	-24.68139*
Saudi Arabia	15	0.9905(0.0281)	0.022234(0.012566)	6.114(3.082149)	-1.32083	-7.63765*
Kuwait	94	0.9863(0.0051)	0.019106(0.010318)	5.273(2.570788)	-1.87839*	-25.40242*
Failaka Island	**27**	**0.9886(0.0131)**	**0.021481(0.011782)**	**5.929(2.919822)**	**-1.94476***	**-15.68033***
Iraq	116	0.9918(0.0036)	0.019862(0.010659)	5.482(2.656844)	-2.10200*	-25.27102*
Syria	69	0.9881(0.007)	0.019229(0.010419)	5.307(2.593951)	-2.14505*	-25.44560*
Kurdistan	53	0.9833(0.0095)	0.019239(0.010471)	5.310(2.604634)	-1.99134*	-25.45358*
Iran	92	0.9895(0.0049)	0.020453(0.010967)	5.645(2.732455)	-2.10033*	-25.30270*
Turkey	290	0.9851(0.0039)	0.017378(0.009419)	4.796(2.349858)	-2.22379*	-25.06957*
Greece	179	0.9760(0.0055)	0.014011(0.007825)	3.867(1.951536)	-2.17981*	-25.74642*
England	242	0.9651(0.0078)	0.014654(0.008122)	4.045(2.026110)	-2.24347*	-25.48480*
Bulgaria	141	0.9762(0.0065)	0.014925(0.008276)	4.119(2.063531)	-2.12006*	-25.72305*
Romania	92	0.9811(0.0051)	0.015303(0.008491)	4.224(2.115396)	-1.97053*	-25.81275*

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

Mismatch Distribution

Mismatch distributions of Failaka Island and comparative populations are provided in figure 20. Unimodal distributions were exhibited by Failaka Island, Kuwait, and Iran. These populations reach a peak in the number of pairwise differences between four and eight with frequency ranging between 15 and 20 percent. The values of Harpending's raggedness index were calculated for each population. All the values are insignificant and below 0.03. The smooth bell-shaped distributions accompanied by low raggedness index (below 0.03) are suggestive of a demographic expansion occurred for these populations. African Bushmen and Somalia show multimodal distributions. Their Harpending's raggedness index values were insignificant and below 0.03. According to the raggedness index and the mismatch distributions, these populations maintain a constant population size over a long period of time.

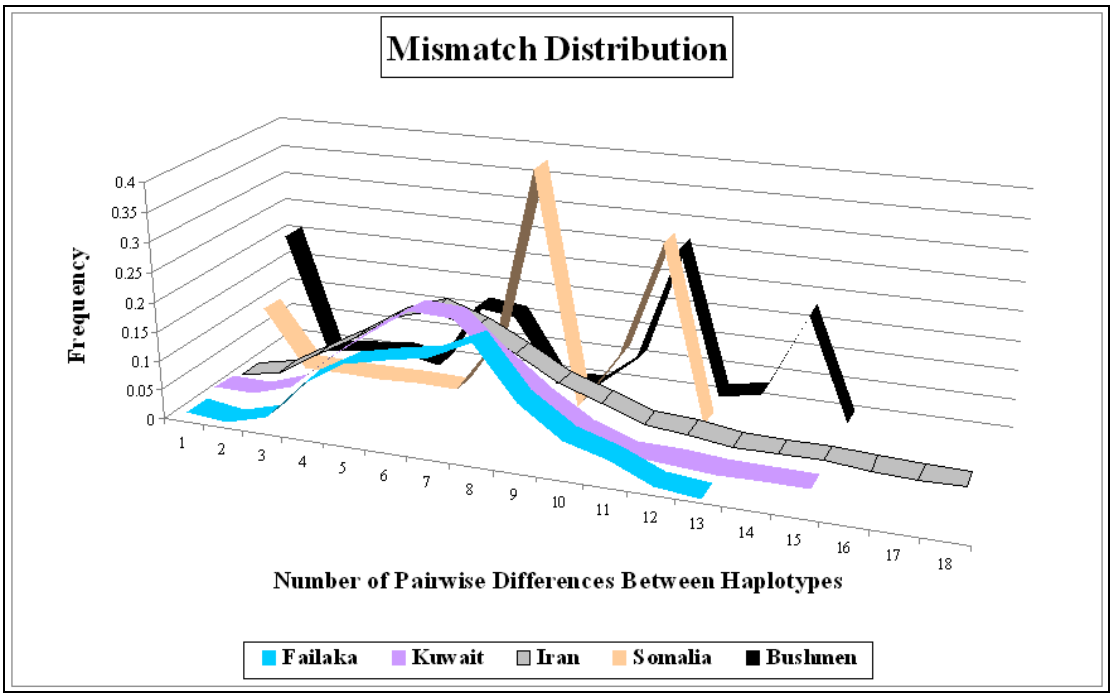

Figure 20. Mismatch distributions of Failaka Island and comparative populations

Neighbor-Joining Tree

A neighbor-joining tree based on mtDNA HVS-I sequence data of Failaka Island and comparative populations using is presented in figure 21. Nigeria and Kenya are grouped together at the bottom of the tree. While the following European populations (Romania, Bulgaria, Greece, and England) grouped together, Turkey is in the middle between Kurdistan and European populations. Each one of the remaining populations shared a separate node. Failaka Island clustered between Saudi Arabia and Kuwait. A two-way Mantel test was used to test the correlation between the original distance matrix and the cophenetic distance matrix which is represented in the tree. The correlation had a value of 0.86185 which indicates that the tree is a good fit to the data.

Multidimensional Scaling

A multidimensional scaling (MDS) method is another method to visualize the genetic relationship of Failaka Island population to other comparative populations. Figure 22 shows a multidimensional scaling plot which is based on mtDNA sequence data using Tamura and Nei's (1993) distances. The African populations, Nigeria and Kenya, are found in the left side of the plot. The European populations (Romania, Bulgaria, Greece, and England) are clustered in the lower right side of the plot. Turkey is located in the middle of European population and Kurdistan as represented in the Neighbor joining tree. Iraq, Syria, and Kurdistan are close to each other. Failaka Island located in the middle of Kuwait, Saudi Arabia, and Iran. The MDS plot has a good stress value of 0.04133. The original distance matrix was compared to a MDS matrix using a two-way Mantel test. The correlation was highly significant with value of 0.98349. This value indicates that the MDS plot is a very good fit to the data.

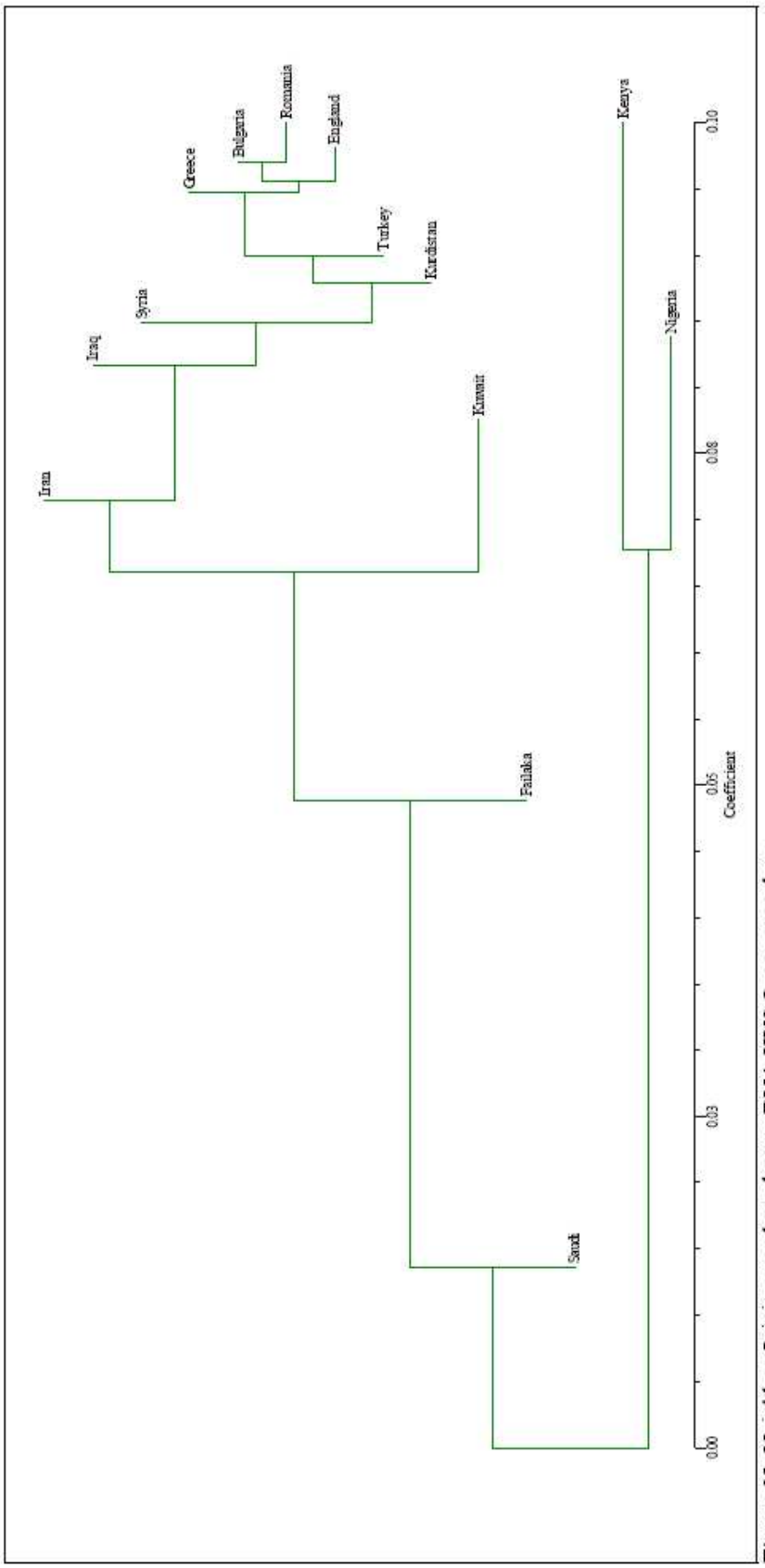

Figure 21. Neighbor-Joining tree based on mtDNA HVS-I sequence data

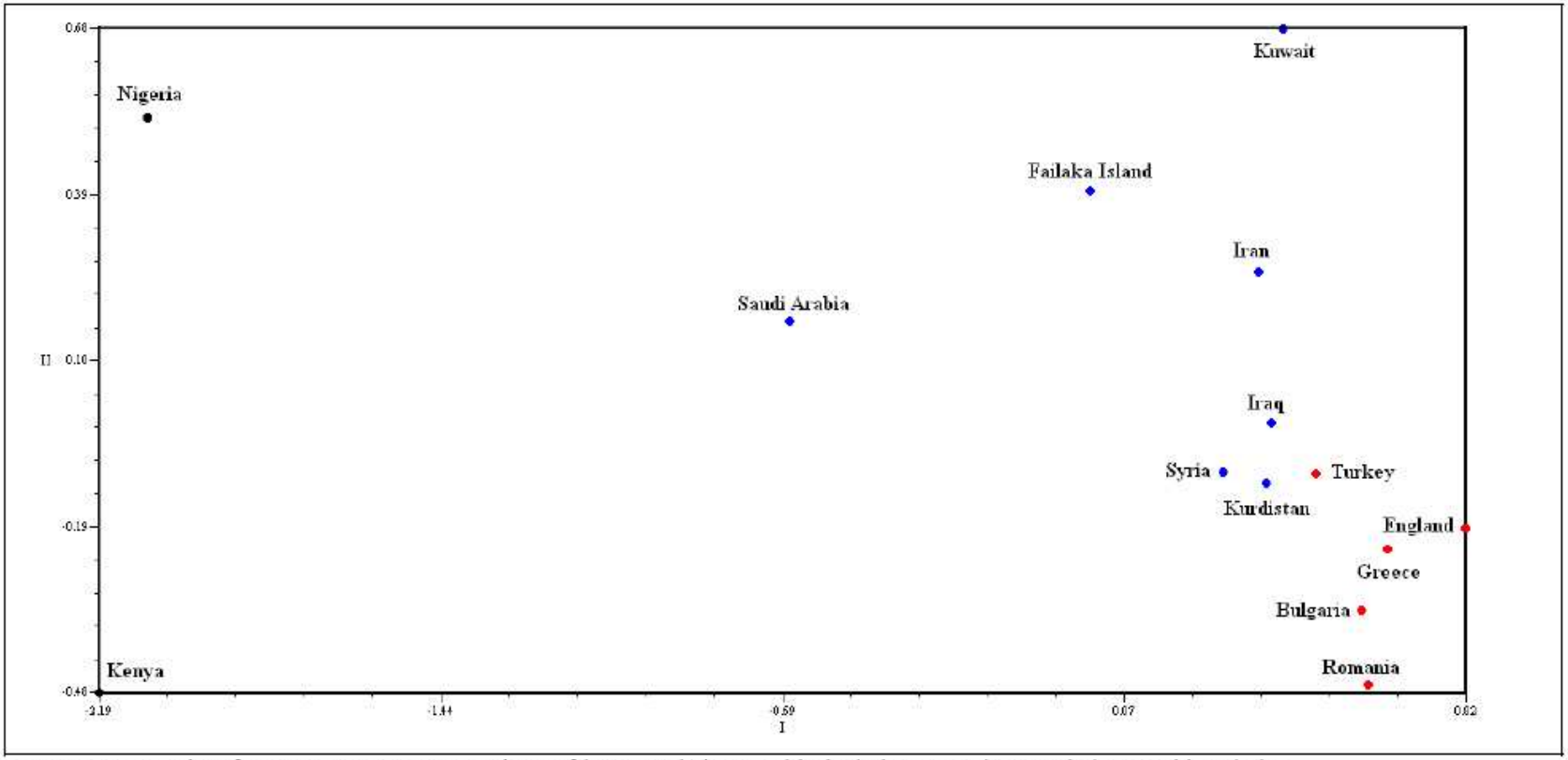

Figure 22. MDS plot of mtDNA HVS-I sequence data. African populations are black circles; W. Asian populations are blue circle; European populations are red circles.

Another MDS plot was constructed using three Kuwaiti subpopulations instead of Kuwait. As the Previous plot, the African populations, Kenya and Nigeria, are found in the left side of the plot. In the upper right side corner, the European populations (Romania, England, Greece, and Bulgaria) are grouped together. Turkey is located between the European populations group and Kurdistan. Iraq, Syria, and Kurdistan formed another group. Failaka Island is located in the middle of Saudi Arabia, Iran, and Bedouin ethnicity subpopulation. Iranian ethnicity subpopulation is located close to Iran. Arab ethnicity subpopulation is close to Bedouin ethnicity subpopulation than any other population. The MDS plot has a stress value of 0.06047. A comparison of the original distance to a MDS matrix using a two-way Mantel test indicated that they are highly correlated (0.98439). This value suggests that the MDS plot is a very good fit to the data.

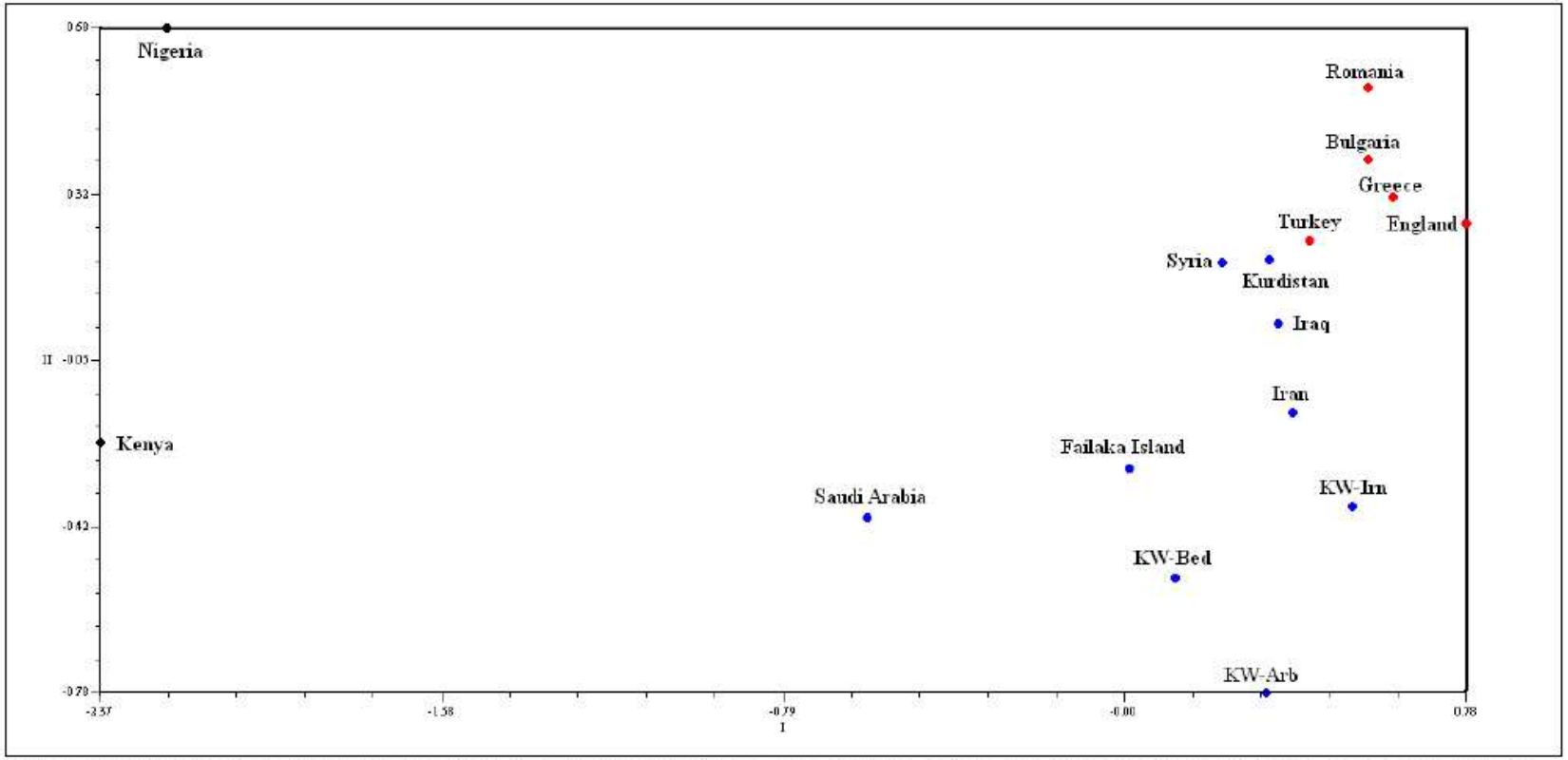

Figure 23. MDS plot of mtDNA HVS-I sequence data for Failaka and Kuwait subpopulations. African populations are black circles; W. Asian populations are blue circle; European populations are red circles.

Gene Diversity vs. Distance from the Centroid (r_{ii})

The plot of gene diversity vs. r_{ii} based on mtDNA HVS-I sequence data for 14 populations is presented in figure 24. Theoretically, an inverse linear relationship between diversity and r_{ii} is expected for human genetic variation which follows the isolation by distance model (Harpending and Ward 1982). Since this method was developed for small subpopulations, a positive linear relationship can be seen when applying this method for continent populations (Harpending and Ward 1982). The following populations are located above the hypothetical regression line: Kuwait, Iran, Iraq, Kurdistan, Syria, Turkey, Nigeria, and Kenya. Their location in the plot suggests that these populations experienced more female gene flow which is consequently increasing the genetic diversity. Failaka Island, Saudi Arabia, Romania, Greece, England, and Bulgaria are located below the hypothetical regression line. The plot suggests that these populations are less admixed and experienced more genetic drift than the rest of populations. Kenya is located in the upper right corner, suggesting that Kenya may have experienced more gene flow than the other populations.

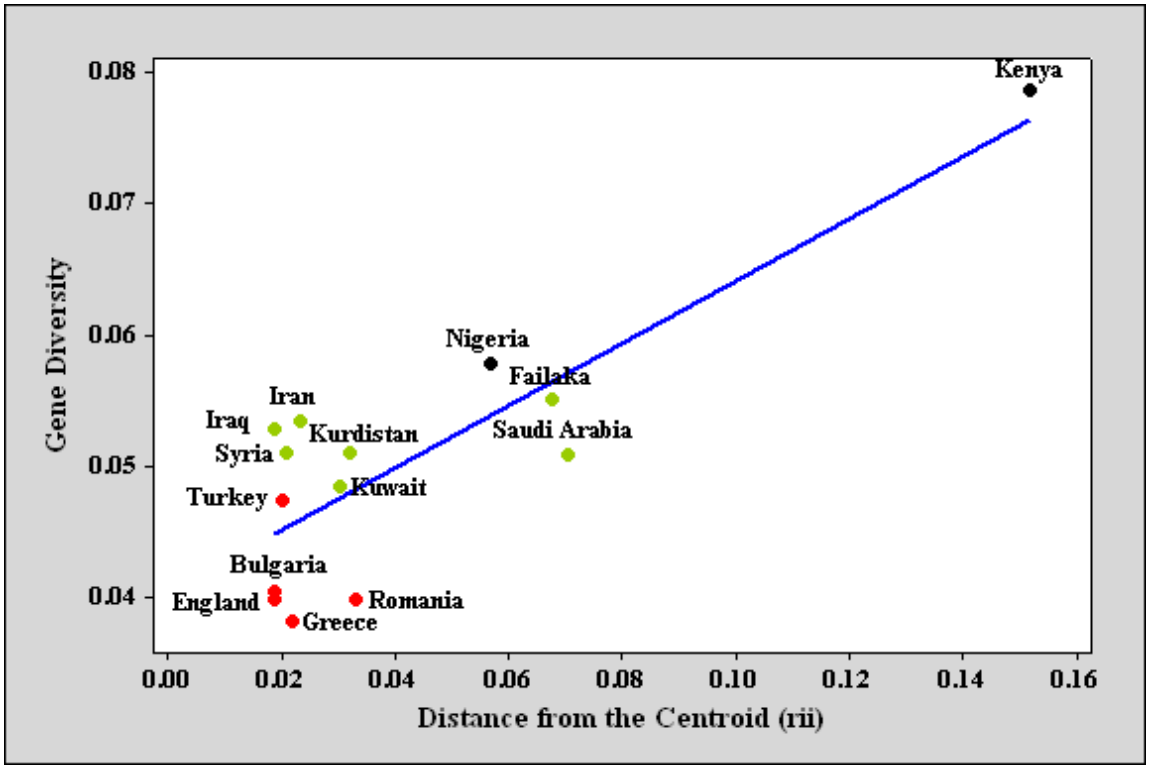

Figure 24. mtDNA HVS-I Diversity vs. r_{ii} plot. African populations are black circles; West Asian populations are green circles; European populations are red circles.

Analysis of Molecular Variance (AMOVA)

Analysis of Molecular Variance was performed to determine the genetic structure of 14 populations based on mtDNA HVS-I haplotype sequence data. Populations were classified according to their geographical regions. The first group represents the African populations include Nigeria and Kenya. The second group consisted of the following European populations: Turkey, Romania, Bulgaria, Greece, and England. The last group represents the West Asian populations and they are: Kuwait, Failaka Island, Saudi Arabia, Iraq, Iran, Syria, and Kurdistan. The result of AMOVA analysis is shown in table 13. The greatest amount of variation found within population and accounted for 91.5% of the total variation, followed by among groups variation (7.33%), and finally within groups variation (1.17%). All variance components are significant.

Table 13. AMOVA of Y-STR data in 14 populations, grouped by geographic region (Africa, Asia, and Europe).

Source of Variation	d.f.	Sum of Squares	Variance Components	Percentage of Variation	F-Statistics
Among groups	2	175.487	0.19905	7.33	Φ_{CT}=0.07332*
Within groups	11	64.677	0.03164	1.17	Φ_{SC}=0.01258*
Within populations	1537	3818.239	2.48422	91.5	Φ_{ST}=0.08497*
Total	1550	4058.403	2.71491		

*P-value = 0.000

Mantel Test:

Mantel tests were used to calculate correlation between genetic and geographic distances based on both mtDNA HVS-I and Y-STR data (Table 14). Mantel test indicates a strong and significant correlation between the mtDNA and Y-STR distances, weaker and insignificant correlation between genetics distance matrices (mtDNA/Y-STR) and geography distance matrice. The result suggests that there are no differences between patterns of mtDNA and Y-STR variation in the tested regions of Kuwait, Failaka Island, Saudi Arabia, and Iran. This result is consistent with the history of these countries where they experienced high gene flow among them.

Table 14. Mantel Correlation Test Results

Comparison	r	P
Mitochondrial DNA vs. Y-chromosome STR	0.8857	0.0435*
Y-chromosome STR vs. Geography	0.3714	0.2609
Mitochondrial DNA vs. Geography	0.0857	0.3913

*P-value > 0.05

Chapter Five: Discussion

Paternal Genetic Structure of Kuwait

The origins and the geographical expansion of humans have been studied using multiple approaches including molecular genetics, paleontology, and archaeology. Genetic findings indicate that Africa is the motherland of all humans and from which they migrated to rest of the world. The critical location of the Arabian Peninsula in understanding the first major migration of human history has triggered the investigation of the genetic structure of the Arabia populations in the last few years. The investigation of the genetic structures of Kuwait and Failaka Island assist in the understanding of the role of the Arabian Peninsula in human diaspora.

The major Y-haplogroup in Kuwait is J which is represented by its subclades J1 and J2 and these two subclades accounted for 45% of the total variation. The highest haplogroup frequency in Kuwait is J1 (37%). In Failaka Island, haplogroup J2 is found in higher frequency than J1, (26%) and (20%) respectively. Haplogroup J1 has been found in high frequencies in the Arabian Peninsula and more specifically in the southern regions. It is documented in Yemen (72.6%) (Cadenas *et al.*, 2008), Oman (38%) (Luis *et al.*, 2004), Qatar (58.3%) (Cadenas *et al.*, 2008), United Arab Emirates (34.8%) (Cadenas *et al.*, 2008), Saudi Arabia (42%) (Abu-Amero *et al.*, 2009), Iraq (33.1%) (Chiaroni *et al.*, 2008), Turkey (8.99%) (Cinnioglu *et al.*, 2004), Lebanon (12.5%) (Semino *et al.*, 2004), and Egypt (20%) (Luis *et al.*, 2004).

The split of haplogroup J into J1 and J2 most likely occurred 18 Kya (Semino *et al.*, 2004). Each clade has a specific geographical distribution pattern associated with a unique history. The geographical distribution of haplogroup J1 reaches it is highest

frequency in southern Arabia and decreased northwards. Haplogroup J2 has reverse pattern, highest frequency is found in the Levant and decreased southwards. According to Chiaroni *et al.* (2008), haplogroup J1 is found mostly inland area and associated with populations of herdsmen. Haplogroup J2 is found in the coastal area and linked to agricultural populations (Chiaroni *et al.*, 2008).

The geographical distribution of haplogroup J is consistent with my findings, especially when dividing the Kuwaiti population according to their ethnicity. The Bedouin subpopulation, whose members lived a nomadic life before oil discovery, have high frequency of haplogroup J1 (49%) and absence of haplogroup J2. The Arab subpopulation has high frequency of haplogroup (52%), much lower haplogroup J2 (8%). Compared to the other subpopulations, the Iranian subpopulation has low J1 haplogroup frequency (16%) and high J2 haplogroup frequency (11%).

The second most common haplogroup in Kuwait is R1a (R- M420) (11%), which is widely distributed among populations from South Asia to Central East Europe and Sothern region of Siberia (Underhill *et al.*, 2010). In Failaka Island the haplogroup R1a accounts for 4% of the total variation. The distribution pattern and the origin R1a Y-haplogroup are not well understood due to the absence of phylogenetic information of its subclades. In addition, it is difficult to distinguish between the Asian and the European origin of the R1a. Although, R1a1a (R-M17), also known as R-M198, is the most common subclade, its place and time of origin is still obscure (Underhill *et al.*, 2010). One hypothesis linked haplogroup R1a1a to the spread of Kurgan culture which originated in Southern Russia/Ukraine and between 3,000-1,000 B.C.E (Underhill *et al.*, 2010) and distributed over India, Central Asia, and Europe. Another explanation based on

R1a1a frequencies in Asia suggests an Asian Origin (Wells *et al.*, 2001). In central Asia, the frequencies of haplogroup R1a1a reach above 50% among Tajiks, Kyrgyz, and Altais and found below 10% among the Turkmenians and Kazakhs (Wells *et al.*, 2001; Zerjal *et al.*, 2002). In Iran, the frequency of haplogroup R1a reaches approximately 30% in the eastern regions and decreases in the western regions to a frequency below 10% (Quintana-Murci *et al.*, 2001). It is believed that both the Dasht-e Kavir and Dash-e Lut deserts of Iran act as gene barriers especially for the R1a1-M198 lineage (Quintana-Murci *et al.*, 2001; Regueiro *et al.*, 2006; Wells *et al.*, 2001).

The distribution of the haplogroup R1a among the Kuwaiti subpopulations brought more insight about the effect of the gene flow. The highest frequency found among the Iranian ethnicity subpopulation (22%), is followed by the Arab ethnicity (6%), and its absence among the Bedouin subpopulation. This distribution represents the presence of the Iranian gene flow to Kuwait. Iraq, as well, has experienced Iranian gene flow which was identified by Al-Zahery *et al.* (2003).

Haplogroup E1b1b (E-M215), known previously as E3b, is the second most common haplogroup in Failaka Island (17%) and found at low frequency in Kuwait (7%). Without further SNPs analyses it is difficult to assign a specific lineage or subclade for our sample. Most likely it belongs to one of the following most common lineages in the Arabian Peninsula: E1b1b1a1 (E-M78) and E1b1b1b2a (M-123). The haplogroups E1b1b1b2a (M-123) and E1b1b1a1 (E-M78) are present at low frequencies over a wide geographic areas including Eastern Africa, the Middle East and the Mediterranean basin (Cruciani *et al.*, 2004; Luis *et al.*, 2004; Semino *et al.*, 2004). However, isolation and genetic drift effects explain the increase in the haplogroup E1b1b1b2a (M-123) frequency

which has been detected in some Ethiopian samples (11%) (Cruciani *et al.*, 2004) and in the Dead Sea region of Jordan (31%) (Flores *et al.*, 2005). In the Arabian Peninsula, the southern region which includes Oman and Yemen has higher frequencies of the haplogroup E1b1b1b2a (M-123) compared to the northern region (Abu-Amero *et al.*, 2009). An East African origin of the haplogroups E1b1b1b2a (M-123) and E1b1b1a1 (E-M78) have been proposed which then spread to the Near East through the Levantine corridor (Cadenas *et al.*, 2008; Cruciani *et al.*, 2004). Horn of Africa and the Levant are probably the two gates through which the haplogroup E1b1b1b2a (M-123) was introduced to Arabia (Abu-Amero *et al.*, 2009). Failaka Island was part of the Dilmun civilization which has active maritime trading routes. The archaeological findings and the maritime activities of Failaka Island support the fact that the haplogroup E1b1b and it is subclades could be introduced and spread during the Neolithic period. While the Bedouin and the Iranian ethnicities of the Kuwaiti subpopulation exhibit similar frequencies of the haplogroup E1b1b 11% and 9% respectively, the Arab ethnicity shows low frequency (2%). These results cannot be further interpreted without additional information regarding the specific lineages they belong to.

The neighbor joining tree and the multidimensional scaling plot which are based on Y-STR data in figure 16 and 17, respectively, show that Kuwaiti populations are clustered in the middle between Iran and Saudi Arabia. Both plots suggest that the paternal genetic structure of Kuwait is similar to Saudi Arabia and Iran. Not only in the paternal genetic structure, but the maternal genetic structure of the Kuwaiti population is similar to Saudi Arabia and Iran (Theyab *et al.*, 2012). This pattern is expected given the fact that both countries shared geographical borders with Kuwait and gene flow is more

likely than in the distant countries. In the same plots, Failaka Island is located between Iran and Kuwait. Since Failaka is an Island located in the Arabian Peninsula between Iran and close to Kuwait, gene flow is expected from Iran. This pattern indicates that the paternal genetic structure of Failaka is similar to Kuwait and Iran. In addition, the gene flow from Saud Arabia is less likely according to the MDS plot. While Yemen and United Arab Emirates are genetically related, Qatar is located in the far left of the plot as an outlier.

The MDS plot shown in figure 18 reflects the relationship of the Kuwaiti subpopulations to neighboring populations. The Iranian subpopulation is clustered between Failaka Island and more proximal to Iran, implying that the paternal genetic structure of the Iranian subpopulation is closer to Iran and Failaka. Saudi Arabia, Arab subpopulation, and Bedouin subpopulation are clustered together which indicates that the paternal genetic structures of both subpopulations are similar to Saudi Arabia than any other populations.

The plot of gene diversity versus distance from the centroid (r_{ii}) indicates that the Kuwaiti paternal genetic structure is similar to Saudi Arabia paternal genetic structure in which both are highly diverse and have experienced more male gene flow than the other Arabian countries. The gene flow is expected in both countries especially after the discovery of oil when more jobs were filled by foreigner workers (Casey 2007). In addition, the presence of The Sacred Mosque in Saudi Arabia to which approximately 2 millions worshipers travel yearly to perform Hajj pilgrimage and some of them remain as illegal immigrants (Abu-Amero 2002). Moreover, the plot suggests that Failaka Island experienced genetic drift and less migration which reduced the genetic variation. The

genetic drift is a powerful evolutionary mechanism especially in island populations (Smith & Wayne, 1996).

Maternal Genetic Structure of Failaka Island

The maternal genetic structure of Failaka Island has been investigated and compared to Kuwait and neighboring populations using multivariate statistical analyses. In the multidimensional scaling plot in figure 22 Failaka clusters in the middle among Kuwait, Saudi Arabia, while Iran. Iraq, Syria, Kurdistan, and Turkey formed a second cluster separate from Failaka Island. This plot suggests that the maternal genetic structure of Failaka Island is a mixture from Kuwait, Saudi Arabia, and Iran. Interestingly, the plot suggests that gene flow from Iraq is less likely, although Iraq is geographically close to both Failaka Island and Kuwait. According to Theyab *et al.* (2012), the maternal genetic structure of Kuwait was not influenced by the Iraq gene flow. A neighbor joining tree was constructed to visualize the maternal genetic relationship of Failaka to other populations. According to the neighbor joining tree plot in figure 21, Failaka Island clustered between Kuwait and Saudi Arabia but was not distant from Iran. This finding is consistent with the MDS plot and both plots are good representations of the mtDNA sequence data. A second MDS plot was constructed including Failaka Island, the Kuwaiti subpopulations, and other populations (Figure 23). Failaka Island is clustered in the middle among the Bedouin ethnicity subpopulation, Iranian ethnicity subpopulation, and both Saudi Arabia and Iran, as previously described. The Arab ethnicity subpopulation is clustered next to the Bedouin ethnicity subpopulation. This pattern suggests that both the Bedouin and Iranian ethnicities subpopulations may have influenced the maternal genetic structure of Failaka Island.

The mismatch analysis of Failaka Island population indicates that the population of Failaka is a rapidly expanding population. This conclusion is supported by the history and the demography of Failaka Island combined with mismatch analysis the smooth bell-shape distribution with raggedness index value below 0.03. In addition, the significant Tajima's D statistic (-1.94476) and Fu's Fs (-15.68033) measurements of Failaka Island population support the demographic observation that Failaka Island underwent an expansion during a short period of time. Kuwait has experienced similar population expansion to Failaka Island with significant Tajima's D statistic (-1.36306) and Fu's Fs (-10.25005) measurements values (Theyab *et al.*, 2012). In addition, to Kuwait, Iran shows similar pattern of mismatch distribution. Compared to Failaka Island, the mismatch analyses of both Somalia and African Bushmen show ragged shaped distributions indicating that historically these populations were of constant size over a long period of time.

A plot of gene diversity versus distance from the centroid of distribution, based on mtDNA HVS-I of Failaka Island, Kuwait, and comparative populations are presented in figure 24. While Failaka Island and Saudi Arabia are located below the hypothetical regression line, Kuwait is located above the line. The location of Failaka Island suggests that the maternal genetic structure of Failaka Island has low female gene flow from neighboring populations and probably they experienced genetic drift at some time in history. The paternal and the maternal genetic structure of Failaka Island have similar pattern of gene flow and genetic drift. This pattern suggests that Island population is more likely to experience low gene flow and genetic drift (Smith and Wayne 1996). The diversity of the maternal genetic structure of Kuwait is low compared to the paternal

genetic structure. Similar to the pattern observed in Kuwait, the maternal genetic structure of Saudi Arabia has low genetic diversity. This pattern is supported by the fact that first cousin marriage is the most preferred in these populations (Teebi, 2010).

Chapter Six: Conclusion

The genetic structure of the Arabian Peninsula is complex and has a detailed evolutionary history. One of the most significant events in the Arabian Peninsula history was initiated by the major expansion early *Homo* out of Africa (Cabrera *et al.*, 2009; Petraglia *et al.*, 2009). Three major events occurred in the Arabian Peninsula that influenced the genetic structure of the Arabs: (1) the ancient civilizations that were established over areas of the Arabian Peninsula, such as Mesopotamia and Dilmun, were active trading routes with the Indus valley and Near East. These trading routes were documented by the archeological artifacts found in the Arabian Peninsula and related to the Ubaid period. (2) The long series of wars between the Roman and the Persian Empires occurred in the Arabian Peninsula during the fourth century. (3) The Islamic expansion was a major event in the Arabian Peninsula in which several ethnicities were introduced to Islam and to the Arabian Peninsula. These three events introduced new ethnicities to the Arabian Peninsula and altered the original genetic structure by introducing new haplogroups (Abu-Amero *et al.*, 2009; Cadenas *et al.*, 2008; Cerny *et al.*, 2008; Černý *et al.*, 2009).

Arabs lived in the Arabian Peninsula for an extended period of time and their early history is obscured and limited due to the absence of historians until the rise of Islam. However, recent archaeological artifacts found across the Arabian Peninsula indicate that the Arabian Peninsula is demographically an active region since it has been inhabited by several populations over a long period of time. One explanation for such activity is the location of the Arabian Peninsula which is the only land connected to Africa and acts as a bridge connecting three continents, Africa, Asia, and Europe.

(Armitage *et al.*, 2011; Lewis 1993; Petraglia *et al.*, 2010; Quintana-Murci *et al.*, 1999; Rodinson 1981).

In addition to the archaeological artifacts, recent analyses of genetic markers in the Arabian Peninsula support the Out of Africa theory. Multiple African, European, and Asian haplogroups have been identified in the Arabian Peninsula, including the mtDNA haplogroups H and L, and the Y-chromosome haplogroups E and T. However, the Arabs have their own maternal and paternal genetic markers which can be used to distinguish them from other ethnicities. The mtDNA structure of the Arabian Peninsula is distinguished by the presence of high frequency of haplogroup R0a which most likely originated in Yemen. On the other hand, the paternal structure of the Arabian Peninsula is distinguished by the high frequency of haplogroup J and especially subclades J1-M267 and J2-M172 (Abu-Amero *et al.*, 2009; Cadenas *et al.*, 2008; Černý *et al.*, 2009).

To better understand the evolutionary history of the Arabian Peninsula, this dissertation has presented an investigation of the paternal genetic structure of both Kuwait and Failaka Island based on Y-chromosome STR. In addition to the paternal genetic structure, the maternal genetic structure of Failaka Island has been examined using mtDNA sequence data.

The results of this dissertation demonstrate that the paternal genetic structure of Kuwait is similar to Failaka Island, Saudi Arabia, and Iran. This finding is supported by the multidimensional scaling plot and the neighbor-joining tree. These results are expected since they are a reflection of the recent history of Kuwait. About 300 years ago, tribal groups migrated from Najd in Saudi Arabia and established the current country of Kuwait (Alghanim 1998; Casey 2007). At that time, establishing a trading network with

neighboring populations such as Iran and Iraq was a way of sustaining the life for the tribal groups. Not only trading, but oil production in Kuwait attracted foreign laborers (Casey 2007). The modern genetic structure of the Kuwaiti population has been influenced significantly by two factors, trading and oil production. This finding is supported by the great genetic diversity of Kuwait compared to Failaka Island which shows low genetic diversity combined with possible genetic drift.

The most frequent Y-haplogroup in Kuwait is J1 and it accounted for 37% of the total variation and it is similar to other Arabian populations including: Oman (38%) (Luis *et al.*, 2004), Qatar (58.3%) (Cadenas *et al.*, 2008), United Arab Emirates (34.8%) (Cadenas *et al.*, 2008), and Saudi Arabia (42%) (Abu-Amero *et al.*, 2009). In Failaka Island, the most frequent haplogroup is J2 (26%) followed by J1 (20%).

The maternal genetic structure of Failaka is a mixture which is similar to Kuwait, Saudi Arabia, and Iran. The data suggest that the women of Failaka Island came originally from Iran and Saudi Arabia. The paternal genetic structure of Failaka Island is similar to its maternal genetic structure showing low genetic diversity and probable influence of genetic drift. In general, the maternal genetic structure of Failaka Island has experienced a demographic expansion over a short period of time. This finding is supported by the history of Failaka Island.

In conclusion, the history and the archaeological artifacts of Arabian Peninsula suggest that the Arabian Peninsula is an active region of the world and was occupied by multiple populations over a long period of time. Today, Kuwait population is one of several populations occupying the Arabian Peninsula. To understand the genetic structure of the Arabian Peninsula, this dissertation investigated the genetic structure of Kuwait

and Failaka Island. Both populations are similar to geographically neighboring populations especially Saudi Arabia and Iran.

Literature Cited

Abu-Amero, K. K. (2002). Review of the current status of drug-resistant tuberculosis in Saudi Arabia. *Annals of Saudi Medicine, 22*(3/4), 236-238.

Abu-Amero, K. K., Gonzalez, A. M., Larruga, J. M., Bosley, T. M., & Cabrera, V. M. (2007). Eurasian and African mitochondrial DNA influences in the Saudi Arabian population. *BMC Evol Biol, 7*, 32.

Abu-Amero, K. K., Hellani, A., Gonzalez, A. M., Larruga, J. M., Cabrera, V. M., & Underhill, P. A. (2009). Saudi Arabian Y-chromosome diversity and its relationship with nearby regions. *Bmc Genetics, 10*.

Abu-Amero, K. K., Larruga, J. M., Cabrera, V. M., & Gonzalez, A. M. (2008). Mitochondrial DNA structure in the Arabian Peninsula. *Bmc Evolutionary Biology, 8*.

Al-Nassar, K. E., Conneally, P. M., Palmer, C. G., & Yu, P. (1981). The Genetic Structure of the Kuwaiti Population. I. Distribution of 17 markers with genetic distance analysis. *Hum Genet, 57*(2), 192-198.

Al-Zahery, N., Semino, O., Benuzzi, G., Magri, C., Passarino, G., Torroni, A., et al. (2003). Y-chromosome and mtDNA polymorphisms in Iraq, a crossroad of the early human dispersal and of post-Neolithic migrations. [doi: DOI: 10.1016/S1055-7903(03)00039-3]. *Molecular Phylogenetics and Evolution, 28*(3), 458-472.

Alenizi, M., Goodwin, W., Ismael, S., & 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, NY In the U.S.A. and Canada distributed by St. Martin's Press,: I.B. Tauris ;.

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

Alves, C., Gusmao, L., Barbosa, J., & Amorim, A. (2003). Evaluating the informative power of Y-STRs: a comparative study using European and new African haplotype data. *Forensic Sci Int, 134*(2-3), 126-133.

Anderson, S., Bankier, A. T., Barrell, B. G., Debruijn, M. H. L., Coulson, A. R., Drouin, J., et al. (1981). Sequence and Organization of the Human Mitochondrial Genome. *Nature, 290*(5806), 457-465.

Anscombe, F. F. (1997). *The Ottoman Gulf : the creation of Kuwait, Saudi Arabia, and Qatar*. New York: Columbia University Press.

Aris-Brosou, S., & Excoffier, L. (1996). The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism. *Mol Biol Evol, 13*(3), 494-504.

Armitage, S. J., Jasim, S. A., Marks, A. E., Parker, A. G., Usik, V. I., & Uerpmann, H. P. (2011). The southern route "out of Africa": evidence for an early expansion of modern humans into Arabia. *Science, 331*(6016), 453-456.

Arroyo-Pardo, E., Gusmao, L., Lopez-Parra, A. M., Baeza, C., Mesa, M. S., & Amorim, A. (2005). Genetic variability of 16 Y-chromosome STRs in a sample from Equatorial Guinea (Central Africa). *Forensic Sci Int, 149*(1), 109-113.

Athey, T. W. (2006). Haplogroup prediction from Y-STR values using a Bayesian-allele-frequency approach. *J Genet Geneal, 2*, 34-39.

Bailey, G., Flemming, N., King, G., Lambeck, K., Momber, G., Moran, L., et al. (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., & Richards, M. (2006). *Human mitochondrial DNA and the evolution of homo sapiens*. Berlin ; New York: Springer.

Bandelt, H. J., Forster, P., & Rohl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol, 16*(1), 37-48.

Bermisheva, M. A., Viktorova, T. V., & Khusnutdinova, E. K. (2003). Polymorphism of human mitochondrial DNA. *Russian Journal of Genetics, 39*(8), 849-859.

Berniell-Lee, G., Calafell, F., Bosch, E., Heyer, E., Sica, L., Mouguiama-Daouda, P., et al. (2009). Genetic and demographic implications of the Bantu expansion: insights from human paternal lineages. *Mol Biol Evol, 26*(7), 1581-1589.

Bowcock, A. M., Ruiz-Linares, A., Tomfohrde, J., Minch, E., Kidd, J. R., & Cavalli-Sforza, L. L. (1994). High resolution of human evolutionary trees with polymorphic microsatellites. *Nature, 368*(6470), 455-457.

Brown, W. M. (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, M. R., DiZinno, J. A., Stauffer, C., Fasano, M. A., Holland, M. M., et al. (1999). Mitochondrial DNA regions HVI and HVII population data. *Forensic Science International, 103*(1), 23-35.

Cabrera, V., Abu-Amero, K., Larruga, J., & 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.

Cadenas, A. M., Zhivotovsky, L. A., Cavalli-Sforza, L. L., Underhill, P. A., & Herrera, R. J. (2008). Y-chromosome diversity characterizes the Gulf of Oman. *Eur J Hum Genet, 16*(3), 374-386.

Cann, R. L., Stoneking, M., & Wilson, A. C. (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., & Barrett, D. (1999). The Kuwait-British Archaeological Expedition to As-Sabiyah: report on the first season's work. *Iraq, 61*, 43-58.

Casey, M. S. (2007). *The history of Kuwait*. Westport, Conn.: Greenwood Press.

Cavalli-Sforza, L. L. (1998). The DNA revolution in population genetics. [doi: DOI: 10.1016/S0168-9525(97)01327-9]. *Trends in Genetics, 14*(2), 60-65.

Cavelier, L., Johannisson, A., & Gyllensten, U. (2000). Analysis of mtDNA Copy Number and Composition of Single Mitochondrial Particles Using Flow

Cytometry and PCR. [doi: DOI: 10.1006/excr.2000.4949]. *Experimental Cell Research, 259*(1), 79-85.
Cerny, V., Mulligan, C. J., Ridl, J., Zaloudkova, M., Edens, C. M., Hajek, M., et al. (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.
Černý, V., Pereira, L., Kujanová, M., Vašíková, A., Hájek, M., Morris, M., et al. (2009). Out of Arabia—The Settlement of Island Soqotra as Revealed by Mitochondrial and Y Chromosome Genetic Diversity. [Article]. *American Journal of Physical Anthropology, 138*(4), 439-447.
Chiaroni, J., King, R. J., Myres, N. M., Henn, B. M., Ducourneau, A., Mitchell, M. J., et al. (2010). The emergence of Y-chromosome haplogroup J1e among Arabic-speaking populations. *Eur J Hum Genet, 18*(3), 348-353.
Cinnioglu, C., King, R., Kivisild, T., Kalfoglu, E., Atasoy, S., Cavalleri, G. L., et al. (2004). Excavating Y-chromosome haplotype strata in Anatolia. *Hum Genet, 114*(2), 127-148.
Coon, C. S. (1963). *The origin of races*. London: Jonathan Cape.
Crassard, R. (2009). The Middle Paleolithic of Arabia: the view from the Hadramawt region, Yemen. In M. D. Petraglia & J. I. Rose (Eds.), *The evolution of human populations in Arabia: paleoenvironments, prehistory and genetics. The Netherlands: Springer* (pp. 151–168): Springer.
Crawford, M. H. (2007). *Anthropological genetics : theory, methods and applications*. Cambridge, UK ; New York: Cambridge University Press.
Cruciani, F., La Fratta, R., Santolamazza, P., Sellitto, D., Pascone, R., Moral, P., et al. (2004). Phylogeographic analysis of haplogroup E3b (E-M215) y chromosomes reveals multiple migratory events within and out of Africa. *Am J Hum Genet, 74*(5), 1014-1022.
Crystal, J. (1992). *Kuwait : the transformation of an oil state*. Boulder: Westview Press.
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., & Wilson, A. C. (1991). Branching pattern in the evolutionary tree for human mitochondrial DNA. *Proc Natl Acad Sci U S A, 88*(5), 1597-1601.
El-Sibai, M., Platt, D. E., Haber, M., Xue, Y., Youhanna, S. C., Wells, R. S., et al. (2009). Geographical structure of the Y-chromosomal genetic landscape of the Levant: a coastal-inland contrast. *Ann Hum Genet, 73*(Pt 6), 568-581.
Ellegren, H. (2004). Microsatellites: Simple sequences with complex evolution. *Nature Reviews Genetics, 5*(6), 435-445.
Ennafaa, H., Cabrera, V. M., Abu-Amero, K. K., Gonzalez, A. M., Amor, M. B., Bouhaha, R., et al. (2009). Mitochondrial DNA haplogroup H structure in North Africa. *BMC Genet, 10*, 8.
Erhart, D., Berger, B., Niederstatter, H., Gassner, C., Schennach, H., & Parson, W. (2012). Frequency data for 17 Y-chromosomal STRs and 19 Y-chromosomal SNPs in the Tyrolean district of Reutte, Austria. *Int J Legal Med, 126*(6), 977-978.

Excoffier, L., Laval, G., & Schneider, S. (2005). Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol Bioinform Online, 1*, 47-50.

Flores, C., Maca-Meyer, N., Larruga, J. M., Cabrera, V. M., Karadsheh, N., & Gonzalez, A. M. (2005). Isolates in a corridor of migrations: a high-resolution analysis of Y-chromosome variation in Jordan. *J Hum Genet, 50*(9), 435-441.

Fu, Y. X. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics, 147*(2), 915-925.

Fujihara, J., Yuasa, I., Muro, T., Iida, R., Tsubota, E., Nakamura, H., et al. (2009). Allele frequencies and haplotypes for 28 Y-STRs in Ovambo population. *Leg Med (Tokyo), 11*(4), 205-208.

Goedbloed, M., Vermeulen, M., Fang, R. N., Lembring, M., Wollstein, A., Ballantyne, K., et al. (2009). Comprehensive mutation analysis of 17 Y-chromosomal short tandem repeat polymorphisms included in the AmpFlSTR Yfiler PCR amplification kit. *Int J Legal Med, 123*(6), 471-482.

Gunz, P., Bookstein, F. L., Mitteroecker, P., Stadlmayr, A., Seidler, H., & Weber, G. W. (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.

Haber, M., Platt, D. E., Badro, D. A., Xue, Y., El-Sibai, M., Bonab, M. A., et al. (2011). Influences of history, geography, and religion on genetic structure: the Maronites in Lebanon. *Eur J Hum Genet, 19*(3), 334-340.

Halm, H. (2007). *The Arabs : a short history*. Princeton, N.J.: Markus Wiener Publishers.

Hammer, M. F. (1995). A recent common ancestry for human Y chromosomes. *Nature, 378*(6555), 376-378.

Hammer, M. F., Karafet, T., Rasanayagam, A., Wood, E. T., Altheide, T. K., Jenkins, T., et al. (1998). Out of Africa and back again: Nested cladistic analysis of human Y chromosome variation. *Molecular Biology and Evolution, 15*(4), 427-441.

Hammer, M. F., Spurdle, A. B., Karafet, T., Bonner, M. R., Wood, E. T., Novelletto, A., et al. (1997). The geographic distribution of human Y chromosome variation. *Genetics, 145*(3), 787-805.

Hammer, M. F., & Zegura, S. L. (2002). THE HUMAN Y CHROMOSOME HAPLOGROUP TREE: Nomenclature and Phylogeography of Its Major Divisions. *Annual Review of Anthropology, 31*(1), 303-321.

Harpending, H., & Ward, R. (1982). Chemical systematics and human populations. In M. H. Nitecki (Ed.), *Biochemical aspects of evolutionary biology* (pp. ix, 324 p). Chicago: University of Chicago Press.

Hartl, D. L., & Clark, A. G. (2007). *Principles of population genetics* (4th ed.). Sunderland, Mass.: Sinauer Associates.

Hedges, S. B., Kumar, S., Tamura, K., & Stoneking, M. (1992). Human origins and analysis of mitochondrial DNA sequences. *Science, 255*(5045), 737-739.

Helgason, A., Nicholson, G., Stefansson, K., & Donnelly, P. (2003). A reassessment of genetic diversity in icelanders: Strong evidence from multiple loci for relative homogeneity caused by genetic drift. *Annals of Human Genetics, 67*, 281-297.

Hitti, P. K. (2002). *History of the Arabs : from the earliest times to the present* (10th ed.). Basingstoke: Palgrave Macmillan.

Howells, W. W. (1980). Homo erectus—who, when and where: A survey. *American Journal of Physical Anthropology, 23*(S1), 1-23.
Jagher, R. (2009). The Central Oman Palaeolithic Survey: recent research in Southern Arabia and reflection on the prehistoric evidence. In P. MD & R. JI (Eds.), *The evolution of human populations in Arabia: paleoenvironments, prehistory and genetics. The Netherlands: Springer* (pp. 139–150): Springer.
Karafet, T. M., Mendez, F. L., Meilerman, M. B., Underhill, P. A., Zegura, S. L., & Hammer, M. F. (2008). New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res, 18*(5), 830-838.
Kim, S. H., Han, M. S., & Kim, W. (2010). Y chromosome homogeneity in the Korean population. *Int J Legal Med, 124*(6), 653-657.
Kivisild, T., Reidla, M., Metspalu, E., Rosa, A., Brehm, A., Pennarun, E., et al. (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., et al. (2003). The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *American Journal of Human Genetics, 72*(2), 313-332.
Lahr, M. M., & Foley, R. (1994). Multiple dispersals and modern human origins. *Evolutionary Anthropology: Issues, News, and Reviews, 3*(2), 48-60.
Lawler, A. (2011). Human evolution. Did modern humans travel out of Africa via Arabia? *Science, 331*(6016), 387.
Lewis, B. (1993). *The Arabs in history* (6th ed.). Oxford ; New York: Oxford University Press.
Lienhardt, P., & Al-Shahi, A. (2001). *Shaikhdoms of eastern Arabia*. Houndmills, Basingstoke, Hamshire ; New York: Palgrave.
Luis, J. R., Rowold, D. J., Regueiro, M., Caeiro, B., Cinnioglu, C., Roseman, C., et al. (2004). The Levant versus the Horn of Africa: Evidence for bidirectional corridors of human migrations. *American Journal of Human Genetics, 74*(3), 532-544.
Mansfield, P. (1992). *The Arabs* (3rd ed.). London ; New York: Penguin Books.
Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer research, 27*(2 Part 1), 209-220.
McEvoy, B., Richards, M., Forster, P., & Bradley, D. G. (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.
Mielnik-Sikorska, M., Daca, P., Wozniak, M., Malyarchuk, B. A., Bednarek, J., Dobosz, T., et al. (2013). Genetic data from Y chromosome STR and SNP loci in Ukrainian population. *Forensic Sci Int Genet, 7*(1), 200-203.
Mills, R. E., Luttig, C. T., Larkins, C. E., Beauchamp, A., Tsui, C., Pittard, W. S., et al. (2006). An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome Res, 16*(9), 1182-1190.
Mrsic, G., Grskovic, B., Vrdoljak, A., Popovic, M., Valpotic, I., Andelinovic, S., et al. (2012). Croatian national reference Y-STR haplotype database. *Mol Biol Rep, 39*(7), 7727-7741.

Nasidze, I., Quinque, D., Rahmani, M., Alemohamad, S. A., & 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. (1987). *Molecular evolutionary genetics*: Columbia Univ Pr.

Nei, M. (1992). Age of the common ancestor of human mitochondrial DNA. *Mol Biol Evol, 9*(6), 1176-1178.

Nei, M., & Kumar, S. (2000). *Molecular evolution and phylogenetics*: Oxford University Press, USA.

Nei, M., & Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci U S A, 76*(10), 5269-5273.

Novelletto, A. (2007). Y chromosome variation in Europe: continental and local processes in the formation of the extant gene pool. *Ann Hum Biol, 34*(2), 139-172.

Olivieri, A., Achilli, A., Pala, M., Battaglia, V., Fornarino, S., Al-Zahery, N., et al. (2006). The mtDNA legacy of the Levantine early Upper Palaeolithic in Africa. *Science, 314*(5806), 1767-1770.

Pérez-Miranda, A., Alfonso-Sánchez, M., Peña, J., & Herrera, R. (2006). Qatari DNA variation at a crossroad of human migrations. *Human Heredity, 61*(2), 67-79.

Petraglia, M. D. (2003). The Lower Paleolithic of the Arabian Peninsula: Occupations, adaptations, and dispersals. *Journal of World Prehistory, 17*(2), 141-179.

Petraglia, M. D., Haslam, M., Fuller, D. Q., Boivin, N., & Clarkson, C. (2010). Out of Africa: new hypotheses and evidence for the dispersal of Homo sapiens along the Indian Ocean rim. *Ann Hum Biol, 37*(3), 288-311.

Petraglia, M. D., Rose, J. I., & SpringerLink (Online service). (2009). The evolution of human populations in Arabia paleoenvironments, prehistory and genetics, Vertebrate paleobiology and paleoanthropology seriespp. 1 online resource.). Available from http://dx.doi.org/10.1007/978-90-481-2719-1

Pieta, K., Shehab, A. H. S., Tirpák, J., Bielich, M., Bartík, M., & ĎURIŠ, J. (2009). Archaeological and geophysical survey in deserted early islamic village Al-Qusur (Failaka, Kuwait). *Ve službách archeologie*, (1/09), 29–35.

Potter, L. G. (2009a). The Archaeology and Early History of the Persian Gulf. In L. G. Potter (Ed.), *The Persian Gulf in History* (pp. 27-56). New York: Palgrave Macmillan.

Potter, L. G. (2009b). *The Persian Gulf in history* (1st ed.). New York: Palgrave Macmillan.

Quintana-Murci, L., Chaix, R., Wells, R. S., Behar, D. M., Sayar, H., Scozzari, R., et al. (2004). Where west meets east: The complex mtDNA landscape of the southwest and Central Asian corridor. *American Journal of Human Genetics, 74*(5), 827-845.

Quintana-Murci, L., Krausz, C., Zerjal, T., Sayar, S. H., Hammer, M. F., Mehdi, S. Q., et al. (2001). Y-chromosome lineages trace diffusion of people and languages in southwestern Asia. *Am J Hum Genet, 68*(2), 537-542.

Quintana-Murci, L., Semino, O., Bandelt, H. J., Passarino, G., McElreavey, K., & Santachiara-Benerecetti, A. S. (1999). Genetic evidence of an early exit of Homo sapiens sapiens from Africa through eastern Africa. *Nat Genet, 23*(4), 437-441.

Regueiro, M., Cadenas, A. M., Gayden, T., Underhill, P. A., & Herrera, R. J. (2006). Iran: tricontinental nexus for Y-chromosome driven migration. *Hum Hered, 61*(3), 132-143.

Reich, D. E., & Goldstein, D. B. (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, J. H. MANTEL (Version 3.1). New York.: John Relethford's Software Page, Departmen of Anthropology, SUNY College at Oneota, New York.

Relethford, J. H. (2000). GEOG (Version 2.1). New York: Distributed by the author, Departmen of Anthropology, SUNY College at Oneota, New York.

Richards, M., CorteReal, H., Forster, P., Macaulay, V., WilkinsonHerbots, H., Demaine, A., et al. (1996). Paleolithic and neolithic lineages in the European mitochondrial gene pool. *American Journal of Human Genetics, 59*(1), 185-203.

Richards, M., Macaulay, V., Hickey, E., Vega, E., Sykes, B., Guida, V., 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, J. F., Scozzari, R., et al. (2003). Extensive female-mediated gene flow from sub-Saharan Africa into near eastern Arab populations. *American Journal of Human Genetics, 72*(4), 1058-1064.

Rodinson, M. (1981). *The Arabs*. Chicago: University of Chicago Press.

Rogers, A. R., Fraley, A. E., Bamshad, M. J., Watkins, W. S., & Jorde, L. B. (1996). Mitochondrial mismatch analysis is insensitive to the mutational process. *Mol Biol Evol, 13*(7), 895-902.

Rogers, A. R., & 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*.

Rose, J. I., & Petraglia, M. D. (2009). Tracking the Origin and Evolution of Human Populations in Arabia. In P. MD & R. JI (Eds.), *The evolution of human populations in Arabia: paleoenvironments, prehistory and genetics. The Netherlands: Springer* (pp. 1–12): Springer.

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol, 4*(4), 406-425.

Sanchez, J. J., Hallenberg, C., Borsting, C., Hernandez, A., & Morling, N. (2005). High frequencies of Y chromosome lineages characterized by E3b1, DYS19-11, DYS392-12 in Somali males. *Eur J Hum Genet, 13*(7), 856-866.

Sawhney, K. S., Sunderland, E., & Woolley, V. (1984). Genetic Polymorphisms in the Kuwaiti Arabs. *Human Heredity, 34*(5), 303-307.

Schlotterer, C. (2000). Evolutionary dynamics of microsatellite DNA. *Chromosoma, 109*(6), 365-371.

Scott-Jackson, J., Scott-Jackson, W., & Rose, J. (2009). Paleolithic stone tool assemblages from Sharjah and Ras al Khaimah in the United Arab Emirates. In M. D. Petraglia & J. I. Rose (Eds.), *The evolution of human populations in*

Arabia: paleoenvironments, prehistory and genetics. The Netherlands: Springer (pp. 125–138): Springer.
Semino, O., Magri, C., Benuzzi, G., Lin, A. A., Al-Zahery, N., Battaglia, V., et al. (2004). Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. *Am J Hum Genet, 74*(5), 1023-1034.
Semino, O., Passarino, G., Oefner, P. J., Lin, A. A., Arbuzova, S., Beckman, L. E., et al. (2000). The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. *Science, 290*(5494), 1155-1159.
Sharma, S., Saha, A., Rai, E., Bhat, A., & 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.
Simonsen, K. L., Churchill, G. A., & Aquadro, C. F. (1995). Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics, 141*(1), 413-429.
Slatkin, M. (1995). A Measure of Population Subdivision Based on Microsatellite Allele Frequencies. *Genetics, 139*(1), 457-462.
Slot, B. (2003). *Kuwait : the growth of a historic identity*. London: Arabian.
Smith, T. B., & Wayne, R. K. (1996). *Molecular genetic approaches in conservation.* New York, NY: Oxford University Press.
Stringer, C. B., & Andrews, P. (1988). Genetic and Fossil Evidence for the Origin of Modern Humans. *Science, 239*(4845), 1263-1268.
Tadmouri, G. O., Al Ali, M. T., Al-Haj Ali, S., & Al Khaja, N. (2006). CTGA: the database for genetic disorders in Arab populations. *Nucleic Acids Res, 34*(Database issue), D602-606.
Tajima, F. (1993). Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics, 135*(2), 599-607.
Teebi, A. S. (2010). *Genetic disorders among Arab populations* (2nd ed.). Berlin ; New York, NY: Springer-Verlag.
Templeton, A. R. (1992). Human origins and analysis of mitochondrial DNA sequences. *Science, 255*(5045), 737-737.
Thangaraj, K., Sridhar, V., Kivisild, T., Reddy, A. G., Chaubey, G., Singh, V. K., 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.
Theyab, J. B., Al-Bustan, S., & Crawford, M. H. (2012). The genetic structure of the Kuwaiti population: mtDNA Inter- and intra-population variation. *Hum Biol, 84*(4), 379-403.
Thorne, A. G., & Wolpoff, M. H. (1992). The multiregional evolution of humans. *Scientific American, 266*(4), 76-&.
Tishkoff, S. A., Gonder, M. K., Henn, B. M., Mortensen, H., Knight, A., Gignoux, C., 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.
Uerpmann, H., Potts, D., & Uerpmann, M. (2009). Holocene (re-) occupation of Eastern Arabia. In M. D. Petraglia & J. I. Rose (Eds.), *The evolution of human*

populations in Arabia: paleoenvironments, prehistory and genetics. The Netherlands: Springer (pp. 205–214): Springer.
Underhill, P. A., Jin, L., Lin, A. A., Mehdi, S. Q., Jenkins, T., Vollrath, D., et al. (1997). Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res, 7*(10), 996-1005.
Underhill, P. A., Myres, N. M., Rootsi, S., Metspalu, M., Zhivotovsky, L. A., King, R. J., et al. (2010). Separating the post-Glacial coancestry of European and Asian Y chromosomes within haplogroup R1a. *Eur J Hum Genet, 18*(4), 479-484.
Underhill, P. A., Passarino, G., Lin, A. A., Marzuki, S., Oefner, P. J., Cavalli-Sforza, L. L., et al. (2001). Maori origins, Y-chromosome haplotypes and implications for human history in the Pacific. *Hum Mutat, 17*(4), 271-280.
Wahida, G., Al-Tikriti, Y., Beech, M., & Al Meqbali, A. (2009). A Middle Paleolithic assemblage from Jebel Barakah, Coastal Abu Dhabi Emirate. In M. D. Petraglia & J. I. Rose (Eds.), *The evolution of human populations in Arabia: paleoenvironments, prehistory and genetics. The Netherlands: Springer* (pp. 117–124): Springer.
Watson, E., Bauer, K., Aman, R., Weiss, G., von Haeseler, A., & 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.
Wells, R. S., Yuldasheva, N., Ruzibakiev, R., Underhill, P. A., Evseeva, I., Blue-Smith, J., et al. (2001). The Eurasian Heartland: A continental perspective on Y-chromosome diversity. *Proceedings of the National Academy of Sciences, 98*(18), 10244-10249.
White, P. S., Tatum, O. L., Deaven, L. L., & Longmire, J. L. (1999). New, male-specific microsatellite markers from the human Y chromosome. *Genomics, 57*(3), 433-437.
Wilson, A. C., & Cann, R. L. (1992). The Recent African Genesis of Humans. *Scientific American, 266*(4), 68-73.
Wolpoff, M. H., Hawks, J., & Caspari, R. (2000). Multiregional, not multiple origins. *American Journal of Physical Anthropology, 112*(1), 129-136.
Wolpoff, M. H., Wu, X., & Thorne, A. G. (1984). Modern Homo sapiens origins: a general theory of hominid evolution involving the fossil evidence from East Asia. *The origins of modern humans: a world survey of the fossil evidence, 6*, 411-483.
Zerjal, T., Wells, R. S., Yuldasheva, N., Ruzibakiev, R., & Tyler-Smith, C. (2002). A Genetic Landscape Reshaped by Recent Events: Y-Chromosomal Insights into Central Asia. *American Journal of Human Genetics, 71*(3), 466-482.