2-1-2016

Analysis of the D1S80 VNTR locus polymorphism in the Nubian population

Soha Rabie Ahmed

The American University in Cairo
Biotechnology Program

Analysis of the D1S80 VNTR locus polymorphism in the Nubian Population

A Thesis Submitted to
The Biotechnology Program

In partial fulfillment of the requirements for
The degree of Master of Science

By Soha Rabie Ahmed
Bachelor of Science

Under supervision of Dr. Hamza El Dorry

Table of Content:

Contents

List of Figures:

List of Tables:

Abstract

D1S80 (1p35-p36), variable Number of tandem repeat (VNTR,) is a minisatellite molecular genetic marker. It is known by its high polymorphism that makes it a powerful marker in the forensic applications, paternity testing, and evolutionary studies. Hypervariable D1S80 locus has been applied to study the genetic structure of the Nubian population in Southern Egypt. PCR has been applied for repeats amplification and analyzed on 2% agarose gel. 75 genotypes and 36 alleles with their frequencies were detected for 105 unrelated individuals. Allele 27 has the highest frequency 0.119% followed by allele 26 with frequency 0.09%. Heterozygosity was calculated as 0.69%. P values of Chi^2 and exact tests indicate departure from Hardy-Weinberg equilibrium which points to the inbreeding within the population. Non-metric dimensional scaling and hierarchical cluster dedrograms were applied to determine the phylogenetic relationship between the Nubian population and other worldwide populations. Both plots indicated the proximity of the Nubian population to Modern Malay population as a result of dominance of 27 alleles in both populations.

1. Introduction:

1.1. Human Races:

Dividing people into different races has been a controversial issue for a long time scientifically and socially. It was accepted by some and rejected by others (Jorde & Wooding, 2004). However, it is used until now in different applications such as the forensic data of USA which are classified as ethnic categories such as Hispanic, European-American, or African-American. Moreover, some of the biomedical scientists found that races do not have any scientific bases behind their nomenclature although others found that it is important to apply races in some experimental designs and research studies (Jorde & Wooding, 2004). Luckily, human genetics provided a huge array of data that enabled researchers to study the human genetic polymorphism. Thus, this polymorphism yielded information about the population history and variants which might contribute in specific diseases (Jorde & Wooding, 2004).

1.2. Human genetic variation:

Generally, human species have variability from one individual to another since any two unrelated individual differ approximately by 3 million variant (0.1 % of DNA). About 85% of that variation is amongst linguistic or local populations. This variation is based on immigration events into them from other groups and also based on the population size (Jorde & Wooding, 2004; R.C. Lewontin, 2006). The remaining percentage of variation is racial variation (different countries) and geographic variation (different continents). These percentages are not sharp because scientists do not have a cut-off for races as different geneticists and anthropologists claimed the presence of 3 to 30 different human races(R.C. Lewontin, 2006).The evolutionary biologist and geneticist Dobzhansky introduced the definition of races geographically. He emphasized that populations differ genetically from each other and this was proved by time after evolution of genetic markers. So, every population is considered as a distinct geographic race (R.C. Lewontin, 2006).

Accordingly, in order to make clusters for differentiations, multivariate analysis must be implemented for different genetic markers because using one marker might cause overlapping among populations and there would not be a systematic classification (Cavalli-Sforza, Menozzi, & Piazza, 1994; R.C. Lewontin, 2006).

2. Literature Review:

2.1. Genetic markers importance:

Human genetic variation studies are profound as it has an evolutionary importance and medical applications. It allows the researchers to differentiate human groups who might be related to each other genetically in addition to population migrations incidence. Moreover, in medicine studying genetic variation could be significant as some alleles cause diseases that might occur commonly in specific geographic regions more than others(Imad et al., 2014). Discovery of genetic markers was an epic since they are inherited and resistant to the effect of rapid environmental change. Instead, they undergo evolution by the means of natural forces as natural selection, mutation, migration or genetic drift (Cavalli-Sforza et al., 1994). The first breakthrough in genetic markers was the ABO blood group in the beginning of the last century. As from the data that have been collected and gathered, it was obvious that different populations have different proportions of the blood groups. Thus, ABO gene has been widely used for human population differentiation (Cavalli-Sforza et al., 1994). This differentiation was a result of the polymorphism, presence of allele in ≥1% within a population; however some genes or loci are much more polymorphic than others. Accordingly, they could be used as genetic markers.

2.2. Genetic markers evolution:

The cephalic index was the favored anthropometric marker for a long time; however, it has a serious disadvantage because this trait might not be under the biological control as it might be affected by the environmental factors in short terms. Consequently, blood groups were the first attempt for human differentiation including the ABO blood group, RH factor (it has a large number of alleles), and the MN blood groups (a system of two genes glycophorin A and glycophorin B on chromosome 4) (Daniels, 2002). They became favored rather than the cephalic index (Cavalli-Sforza et al., 1994).After the establishment of the electrophoresis system, proteins have been used as a genetic marker. The first protein discovered was hemoglobin, because of the presence of its mutant form in sickle cell anemia, beside other serum protein as haptoglobin which is also found to be polymorphic. In the early 60s Human leukocyte antigen (HLA) proteins, which known as major histocompatibility complex (MHC) in

vertebrates, have been discovered to be polymorphic and very informative system (Cavalli-Sforza et al., 1994). Immunoglobulins also have been reported to be a rich resource of genetic variability. All of the above markers exhibit variation in the protein level. Some of them yielded a lot of information however others have some difficulties in their practical work (Cavalli-Sforza et al., 1994).

2.2.1. Restriction Fragment Length Polymorphism (RFLP) marker:

The previous polymorphic markers were in progress until the introduction of restriction fragment length polymorphism technique (RFLP) that works on DNA polymorhism by using restriction enzymes that cut the DNA at their specific restriction site. Then, after resolving the DNA segments by gel electrophoresis according to their sizes the radioactive probes bind to the DNA segments on membranes (Southern blot). This technique revealed a huge number of genetic polymorhisms. Nevertheless, this technique is very laborious and time consuming. It also needs a high amount of intact DNA to reach the technique sensitivity(B Budowle, Chakraborty, Giusti, Eisenberg, & Allen, 1991; Cavalli-Sforza et al., 1994).

2.2.2. Polymerase Chain Reaction (PCR) breakthrough:

In the late 80s, the powerful technique polymerase chain reaction (PCR) has been evolved. It was a revolution in the molecular techniques field and it is the most prevailed and currently used tool. It overcomes the drawbacks of RFLP because it is a rapid reaction and does not consume a large amount of DNA but just few nanograms(B Budowle et al., 1991).Table 1 shows the differences between RFLP and PCR typing methods(Butler, 2010).

2.2.3. PCR based markers:

There are different polymorphic markers based on PCR techniques that are divided into two types; biallelic polymorphic markers and multiallelic polymorphic markers.

Examples of bialleleic markers are single nucleotide polymorphisms (SNPs) and inserted and deleted sequence (INDELs). Examples of multiallelic polymorphisms are microsatellites or short tandem repeats (STRs) and minisatellites or variable number tandem repeats (VNTRs)(Imad et al., 2014).

2.2.4 Single Nucleotide Polymorphism (SNPs):

SNPs are considered as the most abundant genetic variant in human genome. It occurs at a frequency about 1 in 500-1000 base pairs in different regions as coding and non-coding sequences(Shastry, 2002). The variation amount differed from one population to another. Subsequently, they could be used in evolutionary and medical studies (Imad et al., 2014). SNPs can indicate the past mutations that were distinctive events. If there are two individuals who share the same variant, they share the same evolution process since they are attributed to a single particular mutational event that could happen once during the species history (Imad et al., 2014; Stoneking, 2001).

2.2.5 Insertion/Deletion (INDELs) markers:

The inserted/deleted sequences vary from only few nucleotides to hundreds of nucleotides long (Batzer & Deininger, 2002). Alu elements are one of the most important INDELs. They are about 300 bp long. They are stable genetic markers and can reflect the phylogenetic and evolutionary relationships within populations (Guo et al., 2015). In addition, using Alu elements facilitate the date interpretation because it is always known that Alu insertion is not the ancestral state but always the evolved state. Other genetic markers cannot inform which is the ancestral status and which is the derived one (Guo et al., 2015; Imad et al., 2014).

Table1: The differences between RFLP typing method and PCR typing method(Butler, 2010)

Characteristics	RFLP Typing Method	PCR Typing Method
Time required to obtain results	6-8 weeks with radioactive probes; ~1 week with chemiluminescent probes	1-2 days
Required DNA amount	50-500 ng	0.1-1 ng
DNA condition required	Intact DNA' High molecular weight	Degraded DNA might be used
Capability of usage of sample mixtures	Yes for single locus probes	Yes
Allele identification	Binning required as distribution of sizes observed	Discrete alleles obtained and identified
DNA form used	Must be double stranded	Could be single or double stranded
Power of discrimination	~1 in 1 billion with 6 loci	~1 in 1 billion with 8-13 loci
Automation and high volume sample processing	No	Yes

2.2.6 VNTRs and STRs

In 1980, Wayman and White reported for the first time that human genome includes a huge number of polymorphic segments which afterward were called Variable Number of Tandem Repeats (VNTRs) or minisatellites. Alleles on different loci within the genome can be appropriately detected and identified by their number of repeated units with a definite core sequence (Silva et al., 1999). Subsequently, in 1984, Tautz and Renz identified a small sequence repeats which then was called Short Tandem Repeats (STR) or microsatellites. The length variation between them was classified since the core sequence is 6-100 base pairs for VNTRs while for STRs is 2-6 base pairs. Those tandem repeats constitute approximately 3% of the human genome

(Tamaki & Jeffreys, 2005). A reasonable number of these loci has been known as markers to understand the evolutionary history of human populations (Silva et al., 1999).

Microsatellites are scattered through the genome and often appear in the transcription units. On the other hand, minisatellites are dispersed but most frequently in the telomeric regions. VNTRs show higher mutation rates than any other DNA markers so they could be used in tracing back the recent historic events but cannot be used to determine the ancestral allele (Imad et al., 2014; Wan, Wu, Fujihara, & Fang, 2004). Microsatellites have multiple loci with different alleles by which the level of heterzygosity and consequently the genetic variation increases since heterozygosity might reach >75% (Imad et al., 2014). By introduction of PCR, microsatellites have become a highly functional genetic marker. As variation of microsatellites is very high, in a population of large number of individuals and a small number of loci, all individuals will have a distinct genotype. Thus, many issues such as population structure, relationships, and classifications could be addressed(Wan et al., 2004). STRs have also become a powerful tool in paternity testing and forensics (Imad et al., 2014).

2.2.6.1 Minisatellites (VNTRs) as polymorphic markers:

Minisatellites will be focused on in the thesis topic. Minisatellite or Variable Number Tandem Repeats (VNTRs) were the first highly polymorphic genetic marker utilized for genetic analyses (Näslund et al., 2005). VNTRs are located in the non-coding regions of the genome and they do not affect the phenotypic traits (Alford, Hammond, Coto, & Caskey, 1994). They have a core sequence that has a great variability in their copy number that makes them exhibit a higher degree of polymorphism than other classical biochemical and serological genetic markers(Chakraborty, Fornage, Gueguen, & Boerwinkle, 1991). They are inherited in a codominant Mendelian manner and demonstrate high heterozygosity which might reach 95-99%. Accordingly, they are powerful in paternity testing, forensic applications, anthropological and phylogenetic studies (Chakraborty et al., 1991; Mahdieh, Tafsiri, Karimipour, & Akbari, 2005).

There are many different VNTR loci utilized for studying the heterogeneity of different populations of different ethnicity and origins such as D1S80, APOB (Mahdieh et al., 2005), D4S43 (Maria et al., 2003), D12S67, DYS19 (Falcone, Spadafora, Luca, & Ruffolo, 1995), D17S5, D19S20 (Birajalaxmi Das, Anu Ghosh, P. S. Chauhan, 2002). We will focus on the D1S80 locus in our study.

2.2.6.2 D1S80 VNTR locus as a genetic marker:

The VNTR locus D1S80 is located on chromosome 1 (location: 1p35-p36). Its GenBank sequence accession number #D28507 with a core sequence 16 nucleotides base pairs (Duncan, Balamurugan, Budowle, Smerick, & Tracey, 1997). The core consensus sequence is G-N-Pu-G-A-C-C-A-C-(A or C)-G-G-N-A-A-G since N represents any kind of nucleotide bases while Pu represent a purine base A or G (Kasai et al., 1990). Twenty variations based on this sequence were detected (Kuppareddi Balamurugan, Tracey, Heine, Maha, & Duncan, 2012). There are about 29 different known alleles identified for D1S80 locus in different populations which have sizes from 370 to 802 base pairs. Accordingly, this number of alleles will allow the presence of a huge number of different genotypes in each population; Thus, it could be used in genetic variation studies (Holewinsky, 2002). The most common alleles reported in different human populations are 18 and 24 and it is suggested that they are the ancestral alleles of this locus (Kuppareddi Balamurugan et al., 2012).

Most of the D1S80 alleles are heterozygous and they are inherited in a Mendelian mode. They have a flanking sequence that includes primer sites that are used in the PCR reaction to amplify this DNA segment (Figure 1, 2).The highest observed heterozygosity has been reported as 90.5% while the lowest was 24% with more than 27 alleles (Kuppareddi Balamurugan et al., 2012).

D1S80 locus is composed of VNTRs that have discrete alleles of small size which facilitate PCR amplification process; and it is considered as an informative and highly polymorphic DNA marker. It is used in forensic applications, paternity testing, and population genetics applications(Kuntze, Stancu, & Bonte, 1996).

Figure 1: Diagram of two different alleles in heterozygous individual with primer sites

(Holewinsky, 2002)

2.3 Nubia as a case study for genetic variation:

Nubia has been an ancient region and one of the most important lands in Africa. Although it was an arid land, due to its geography, it witnessed a number of events that influenced the development of its civilization. Nubia was a battlefield of the ancient world where black and white people conflict occurred for North Africa predominance (Emery, 1965).

2.3.1 Geography of Nubia:

Ancient Nubia location was restricted to the farmed area tied with desert that borders the river Nile. The northern part was usually rainless. On the other hand, it extended south of 17º 50´ beyond the river bank where rains were annually expected and remarks of ancient human existed at a short distance from the river (Shinnie, 1996). The river Nile is the main characteristic of the Nubian topography. Nubia starts at Sennar on the Blue Nile (130º 34´ north 33º 35´ east) 250 kilometers upstream its juncture with the White Nile at Khartoum where the united single river flows for 2000 kilometers through Sudan and Egypt till its estuary at the Mediterranean (Shinnie, 1996). The river flow is hindered at six locations where blazing rocks of the basement complex caused formation of six cataracts extended from upstream at Aswan to 75 kilometers downstream from Khartoum (Shinnie, 1996). The geographic location of ancient Nubia extended from the first cataract at southern Egypt to the sixth cataract central Sudan (figure3) (Carlson & van Gerven, 1979; Kirwan & Kirwan, 2015).

Ancient Nubia has been divided into two parts. The lower Nubia was the region between the first and the second cataract and it was called Wawat. The upper Nubia is

the region extended from the second to the sixth cataract and it was called Kush (figure 2) (Emery, 1965). In contemporary Nubia the geography has been changed due o the successive construction of dams starting from Aswan dam construction in 1898-1902 then its heightening in 1908-1912 and then in 1929-1932 (Kirwan & Kirwan, 2015). In 1963, because of the high dam construction, people of lower Nubia had to be relocated to Komombo in Aswan governorate which is 50 kilometers north away of Aswan city. On the other hand, people of upper Nubia were relocated in a region called Kashm al Gerba, now known as new Halfa project, which is 800 kilometers far from their land. Now this region extends between the first and fourth cataracts (figure 4) (Ammar, 1996).

2.3.2 Brief History of Nubia:

Nubia history has been divided by archaeologists and historians into different groups each one represents an era with distinct characteristics.

A-Group: prolonged from 3400 B.C. to 2400 B.C (Adams, 1977). This period of time exhibited the conversion from the Mesolithic to the Neolithic era. It was claimed that people of that group represented Egyptians of the predynastic era in Nubia (Carlson & van Gerven, 1979).

B-Group: the archeologist Reisner claimed that B group followed the A group however it was socially and economically less than A group. He suggested that there was an invasion of another group (Negroid) (Carlson & van Gerven, 1979). Afterwards, in 1966 H.S. Smith proved that the archeological evidences on which B group was claimed are assigned to group A. Accordingly, there is no mean to represent this period until the evolution of the C group (Török, 2009).

C-Group: Extended from 2300 B.C. to 1200 B.C (Carlson & van Gerven, 1979). This period was in lower Nubia. In the earlier times of this period, it was occupied by Egyptians through six dynasties. Consequently, it was the first time to be documented in Egyptian old kingdom account for trading as Nubia has become a corridor between north side and the tropical Africa (Carlson & van Gerven, 1979; Török, 2009). It was suggested again that people of that period were Negroid. Another suggestion is that they were people of A-group who migrated to Upper Nubia as a result of Egyptian enforcement and climatic factors but they came back with improvement of these

factors. During the eleventh and twelfth dynasties C-group was exposed to political intervention of Egyptian forces but it was for a short time then they returned back to Egypt (Carlson & van Gerven, 1979; Török, 2009).

Nubian-Pharaonic interaction during New Kingdom: prolonged from 1650 B.C. to 1000 B.C. but some historians suggested later date while others suggested earlier dates (Carlson & van Gerven, 1979).During the new Egyptian kingdom, from eighteenth to twenty-fourth dynasties, the Egyptian military forces had returned to occupy the lower Nubia and a part of the upper Nubia (till the third cataract) (Buzon, 2006). Here, a query is raised. Has this occupation affected the Nubian race? Some archeologists anticipated extensive fusion of the two races according to the skeletal remains found there (Buzon, 2006). Others claimed the prevalence of the Egyptian race but the rest assure the prevalence of the Negroid race. However, in lower Nubia by seventieth dynasty there were graves of C-group indigenous Nubians beside the Egyptians but the graves' number started to decrease by time till the end of the twentieth dynasty as there were not any graves either for Nubians or Egyptians. This was attributed to depopulation of the lower Nubia totally to the upper Nubia between 1000B.C. to 100 A.D (Carlson & van Gerven, 1979).

Napatan and Meroitic kingdoms: during the migration of C-group people to upper Nubia they established the Napatan kingdom that controlled Upper Nubia and Egypt around 750 B.C. and established the twenty-fifth dynasty which lasted till 600 B.C. In 350B.C. the Meroitic kingdom developed at upper Nubia and by 100 A.D. the lower Nubia was reoccupied (Carlson & van Gerven, 1979).

X-group: around 350 A.D. the Meroitic kingdom has fallen and replaced by the X-group. It was suggested that the origin of this group's people are Negroid who came from the South either Blemmy (nomadic Nubian lived in Eastern desert near Meroe) or Nobatae (Emery, 1965; Kirwan & Kirwan, 2015).

Christianity and Islam: by 550 A.D the X-group ended and Christianity was accepted (Carlson & van Gerven, 1979). In the 7th century Abdullah ibn Saad, the Arab ruler of Egypt, invaded the Nubian lands and made a treaty (El Baqt treaty) which lasted for six centuries. Nevertheless, since the eleventh century an Arabian tribe called Rabia was compulsorily moved and settled in lower Nubia. It brought their religion and its political constitution although they gained the Nubian language

and intermarried within the Nubian population. Those people are called BeniKanz or Kenuz who reside at north Nubia until now (Fernea, Fernea, Rouchdy, & American University in Cairo, 1991).

Mamelukes and Ottomans Periods: the Mameluke rule extended from the thirteenth century till the beginning of the sixtieth century. When the Ottoman rule prevailed, it ended the Mameluke rule which completely ended by Mohamed Ali at the nineteenth century. The Mameluke survivors escaped to the South and destroyed many Nubian communities. Mohamed Ali sent garrisons to the present border at south of Egypt. Some of them, who were Bosnian, Hungarians, Kurdish and Turkish, intermarried with the Nubian girls and they were entitled Kashifs (Fernea et al., 1991).

In 1848 Mohamed Ali declared that Egypt and Sudan are independent from the Ottoman governance and during this period Nubia has become a corridor between Africa and the Mediterranean for gold, ivory and slaves trade. Subsequently, the English colonialism has started in 1882 till liberation from it in 1952 and separation of Egypt and Sudan at 1956. Accordingly, Nubia has been separated as lower Nubia in Egypt and Upper Nubia in Sudan (Ammar, 1996; Fernea et al., 1991).

As we can notice along the history that Nubian people are of mixed origins. Therefore, our research aims to study the genetic variation among different Nubian individuals.

2.3.3 People of Nubia and the linguistic affiliations:

Nubian languages are one of the Eastern Sudanic languages which descends from the Nilo-Saharan group of languages (Morris F. Goodman, n.d.). Egyptian Nubians have three different linguistic groups. Kenuz group, who is located at the north part of Aswan extending to 150 kilometers to the South before the relocation during the high dam construction, speaks Kunuz Nubian or Metokkii (Fahim, 1983). The group includes 17 villages (Sokarno, 2007). The Arab group resided in the next 40 kilometers to south and their mother tongue is Arabic (Fahim, 1983). This is the smallest group and includes 8 villages (Sokarno, 2007). The last group is the Fadjikkii resided in 130 kilometers in the southern part of the Egyptian Nubia, speaks Mahas (Fahim, 1983). This group includes 16 villages (Sokarno, 2007). The Sudanese Nubians have two linguistic groups. First, Halfans, located at Wadi Halfa extended

for 170 kilometers from the Egyptian border to the Dal cataract in Sudan, who speak Mahas or Sukkot language (Fahim, 1983). Second, the Dongolawi language spoken by Dongola people located at a region from south Dal cataract to Dongola (Shinnie, 1996).

2.4 Previous work on genetic variation among contemporary Nubians:

Genetic information of the Nubian region is limited. In earliest studies of genetic variation among the forty-nine African populations by using around 47 genes in average, Nubian population was located within the cluster of North African population as Moroccan, Bedouin, Tunisian, Libyan, Egyptian and Berber in the phylogenetic tree (figure 5) (Cavalli-Sforza et al., 1994). This was confirmed with principal component analysis (figure 6)in addition to the existence of north Nubians (might be North Nubians of Sudan) in proximity with Ethiopian Somali ethnic group which suggested presence a genetic relationship between Nubians and some of Sub-Sharan African groups (Cavalli-Sforza et al., 1994).

In 1997, another study has been conducted through analysis of *Hpa*I African marker that represents the point mutation position 3592 on the mitochondrial genome of ancient Meroitic mummies. It is prevalent in the sub-Saharan Africa. The study revealed that *Hpa*I marker was present in a high percentage (39%) which suggested migration and gene flow from south to the Meroitic kingdom location in ancient times (Fox, 1997).

Later, another study complemented the previous one in 1999. The *Hpa*I mitochondrial marker was analyzed among 224 contemporary individuals from different locations along the river Nile, passing by Egypt, Nubia, and southern Sudan, to find the genetic differences between northern and southern parts (Krings et al., 1999). The difference in sequences was designated as the Northern mitochondrial-DNA and Southern-mitochondrial DNA. The study stated that the percentages of the northern and southern types are different significantly among them. The diversity of the northern mitochondrial DNA was highest in Egypt and lowest in northern Sudan while the southern mitochondrial DNA was highest in northern Sudan and lowest in Egypt. In addition, both Nubia and Egypt display comparable amounts of variance in the both

types of mitochondrial DNA. This proves the historical evidence of the bidirectional movements and the interactions between Nubia and Egypt (Krings et al., 1999).

In 2003 another study has been implemented on assessing Y chromosome haplotypes distribution along the river Nile through Delta in Egypt, Upper Egypt, and lower Nubia (Lucotte & Mercier, 2003).There was a northern Y-chromosome haplotype and a southern Y-chromosome haplotype. Haplotype V was the northern one which was dominant in Delta and the least in Lower Nubia. On the other hand, haplotype IV was the southern one which was almost deficient in the North Egypt and predominant in lower Nubia (Lucotte & Mercier, 2003). Obviously, the results elucidate that river Nile was a corridor for migration and gene flow in a bidirectional mode which is consistent with both the pre-dynastic and post-dynastic events (Lucotte & Mercier, 2003).

2.5 Objective:

In this thesis, we will apply using the VNTR D1S80 locus in studying the genetic polymorphism among the non-Arab Nubian population in Egypt.

5’ ACCGGCCCT CACGGTGCCA AGGAAACAGC CCCACCATGA GGCGCTGAGA
GAAACTGGCC TCCAAACACT GCCCGCCGTC CACGGCCGGC CGGTCCTGCG
TGTGAATGAC TCAGGAGCGT ATTCCCCACG CGCCAGCACT GCATTCAGAT
AAGCGCTGGC TCAGT

GTCAGCCCAA GGAAGA

CAGACCACAG GCAAGG

AGGACCACCG GAAAGG

AAGACCACCG GAAAGG

AAGACCACAG GCAAGG

AAGACCACAG GCAAGG

AGGACCACCG GAAAGG

AAGACCACCG GCAAGG

AGGACCACCG GCAAGG

AGGACCACCA GGAAGG

AGGACCACCA GCAAGG

AGGACCACCA GCAAGG

AGGACCACCA GGAAGG

AGGACCACCA GGAAGG

AGGACCACCG GCAAGG

AGGACCACCA GGAAGG

AGGACCACCA GGAAGG

AGGACCACCG GCAAGG

AGGACCACCA GGAAGG

AGAACCACCA GGAAGG

AGGACCACCA GGAAGG

AGGACCACCA GGAAGG

AGGACCACTG GCAAGG

AAGACCACCG GCAAGC

CTGCAAGGGG CACGTGCATC TCCAACAAGA CAAAATAAAC AAGCCAGAGA

GGGCTTGTGA CCAGTGTGGC ATTTGTCAC 3’

Figure 2: D1S80 nucleotide sequence with flanking regions including the underlined primer sites. The red sequence represents one repeat(Kasai et al., 1990)

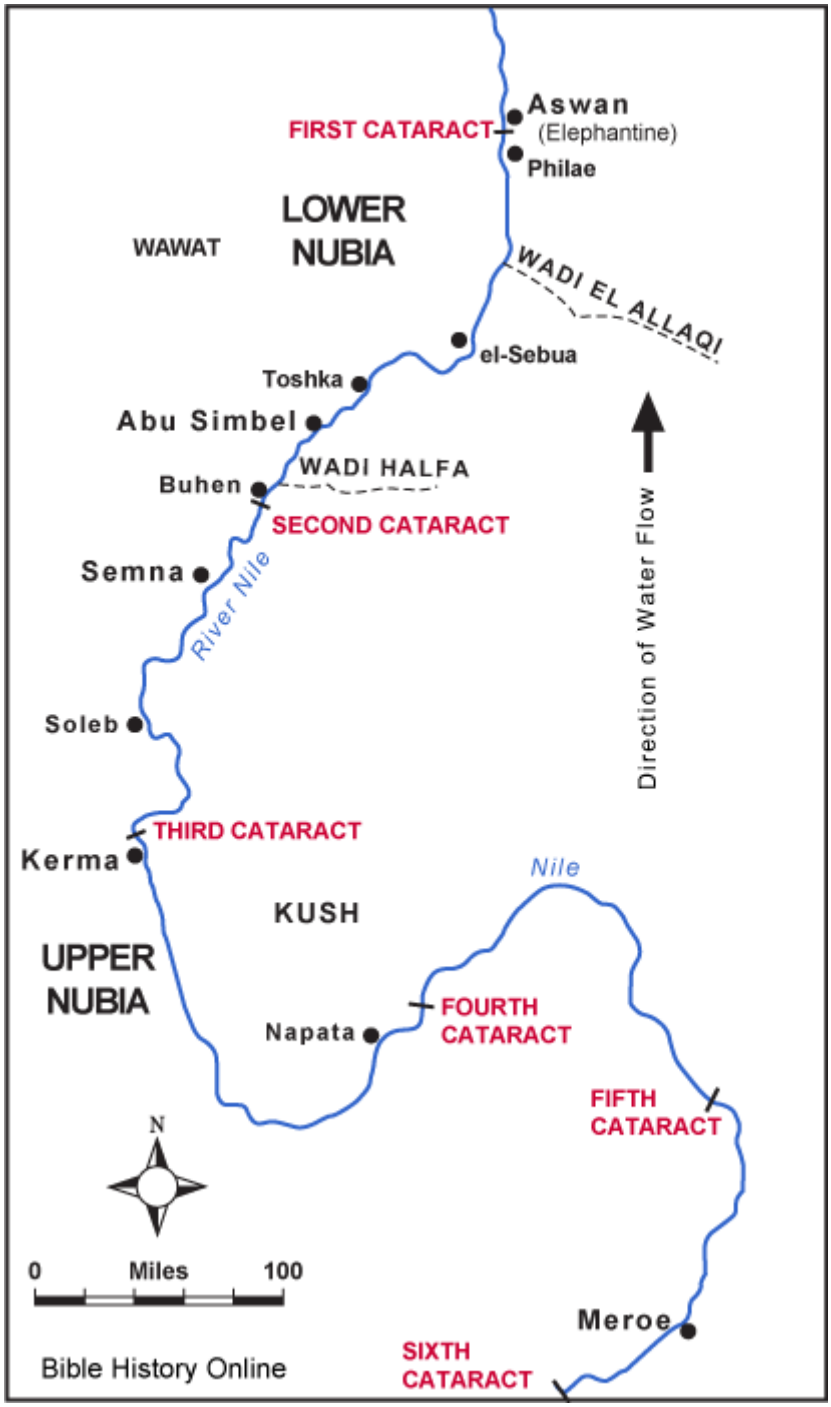

Figure 3: Map of Ancient Nubia
http://www.bible-history.com/geography/maps/map_nubia.html

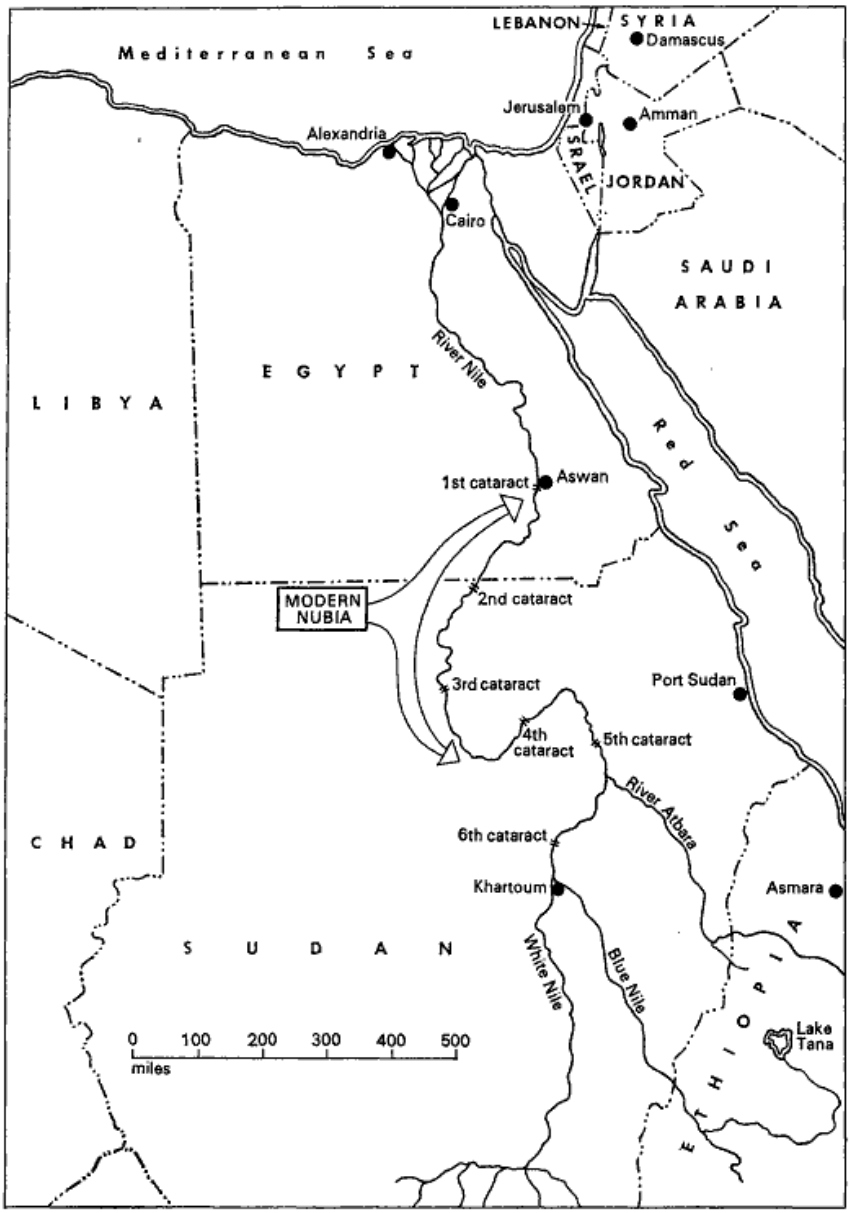

Figure 4: Location of Modern Nubia

http://quod.lib.umich.edu.libproxy.aucegypt.edu:2048/cache/h/e/b/heb09247.0001.001/00000038.tif100.gif

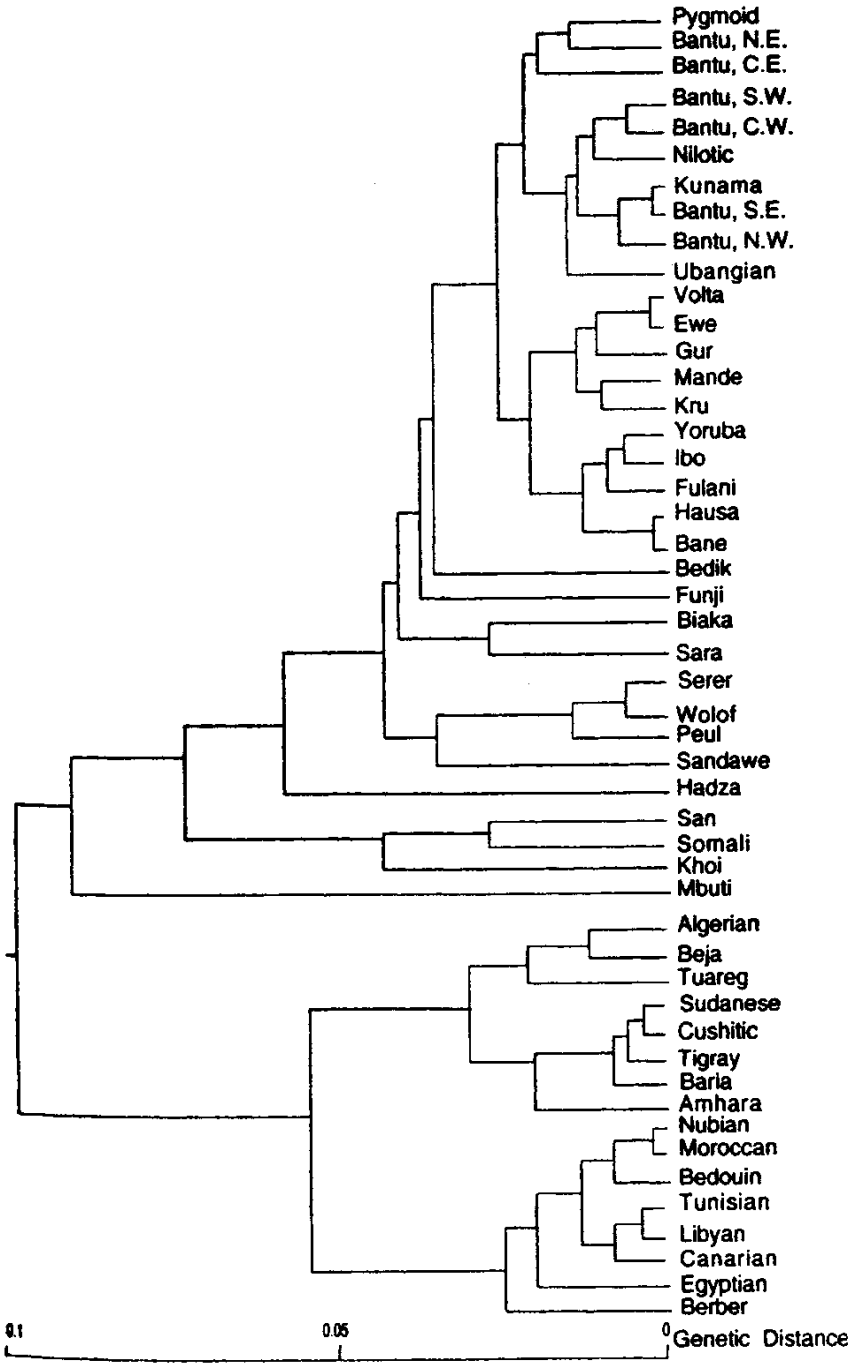

Figure 5: phylogenetic tree of 49 African populations(Cavalli-Sforza et al., 1994)

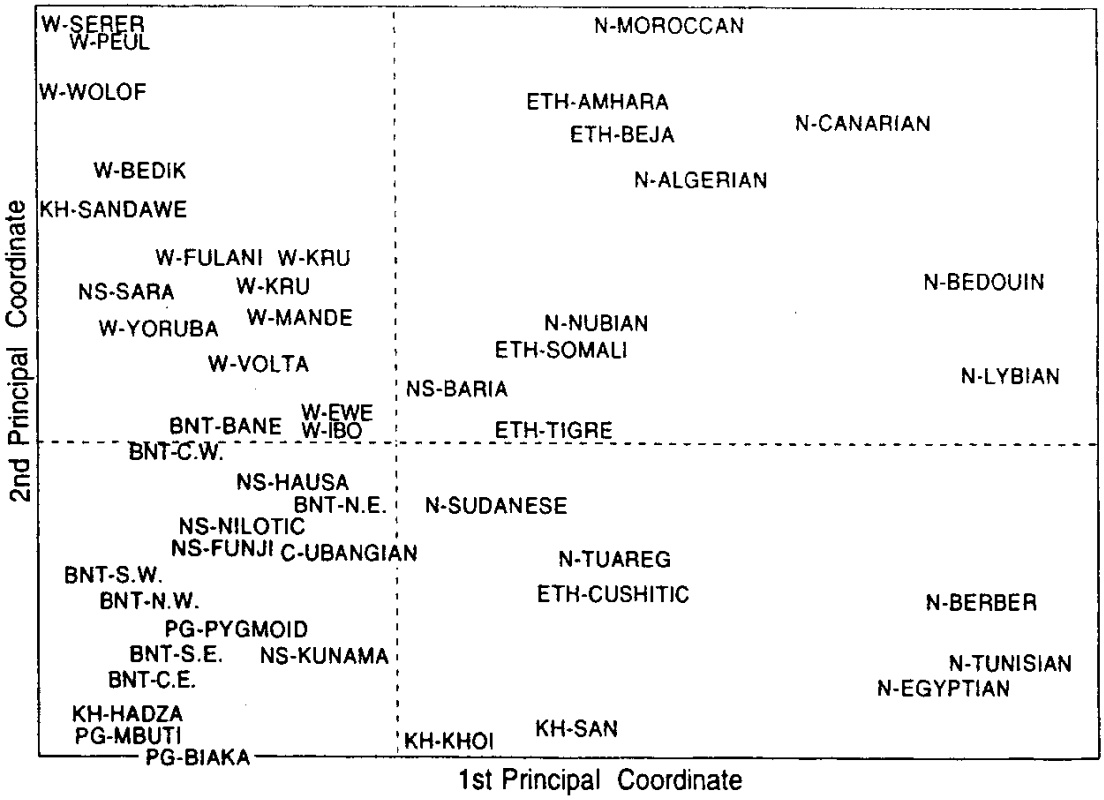

Figure 6: Principal component analysis of 49 African populations(Cavalli-Sforza et al., 1994)

3. Material, methods and statistical analysis:

3.1 Study location and sample size:

The study included 105 randomly selected unrelated Nubian males and females who signed consent forms. 20 are residing in Cairo and 85 residing in different Nubian villages in Aswan (Table 2). It was declared that their ancestors had lived in their villages at least for three generations. Each donor has filled a pedigree chart including their ethnicity, their spoken language and birth place in addition to similar information about their parents, paternal and maternal grandfathers and grandmothers. Samples from Fadjikki and Matokkii groups were included in the study while the Arab group was excluded.

3.2 Ethnographic information about samples' donors:

Each donor filled a pedigree chart to ensure the Nubian origins at least for three previous generations. The information was recorded as follows:

-The donor's full name

-Birth date

-The ethnicity (Fadjikior Matokii)

-Place of birth and current residence of the donor

-The Identity of donor's father, mother, paternal, and maternal grandfathers and grandmothers

- The birth date, birth place and residence of the previous subjects at least for three generations

3.3 Blood samples collection:

2 ml of whole blood sample were collected from each individual by hypodermic needle syringes and reserved in Ethylene diamine tetra-acetic acid (EDTA) vacutainer tubes. The tubes were labeled by sample donor and its village. Then, the tubes were transported to the science and technology research center (STRC) Biology Department, School of Science and Engineering (SSE) at American university in Cairo (AUC). The blood samples were stored in -20^{o} C for further analysis.

Table2: Samples number and their ethnic affiliations

Ethnic group	Males	Females
Matokkii	42	24
Fadjikkii	33	6
Total	75	30

3.4 Genomic DNA isolation:

Genomic DNA was isolated by QIAamp DNA Extraction kit (QIAGENE, Germany). The extracted DNA quality was determined by resolving on 0.5% agarose gel with 1X TBE buffer (0.5 M Tris, 0.05 M boric acid and 1mM EDTA, pH 8.0). Staining was by 2.0 µl ethidium bromide. The loading buffer was also 1X TBE buffer. 5 µl DNA of each sample was loaded. The gel was run at 90V using a power supply system.

3.5 VNTR Analysis:

Materials:

-Primers pMCT118 as described by (Kasai et al, 1990)

Forward upstream primer (28 mer).

5'-GAAACTGGCCTCCAAACACTGCCCGCCG-3'

Reverse downstream primer (29mer).

5'-GTCTTGTTGGAGATGCACGTGCCCCTTGC-3'

-10X Dream Taq Green Buffer (Thermo Scientific) include 20 mM MgCl2.

-DNTPs Mix 10 mM (Promega)

-Dream Taq DNA Polymerase 5u/µl (Thermo Scientific)

-H2O (Gibco® Distilled Water)

-GeneRuler 100bp plus ladder 0.1µg/µl (Thermo Scientific)

Each PCR tube contained 2.5 µl 10X Buffer with MgCl2, 1µl dNTPs, 2µl forward primer, 2µl reverse primer, 0.5 µl Taq Polymerase, 14 µl sterilized H2O, and 3µl of . DNA amplification implemented by Veriti® 96 well thermal cycler (Applied Biosystem). DNA concentration was 3-60 ng.

The amplification program was as follows:
Step 1: The initial denaturing step is for 5 min at 94°C;
Step 2: 30 cycles of cyclic denaturaion for 30 sec at 94°C, 30 sec for primer annealing at 65°C, and 30 sec for primer extension at 72°C;
Steps 3: 7 minutes are added for further extension at 72°C. Samples will remain at 4°C for 15 minutes at thermal cycler before opening the thermal block.
The PCR products were resolved on 2% agarose gel (SeaKem® LE Agarose - Lonza) in 1X TBE buffer. Staining was done by 10µl ethidium bromide and 5µl of each PCR product were loaded into wells of the gels. 5µl of Generuler 100 bp DNA plus ladder was loaded in each gel to estimate the bands' size. The gel was run at 120V, 100mA using a power system.

.Table 3: Populations included in the study

Population name	Country	Reference
Congo	Congo	(Silva et al., 1999)
Cameroon	Congo	(Silva et al., 1999)
Cametá	Brazil	(Silva et al., 1999)
Trombetas	Brazil	(Silva et al., 1999)
Cajueiro	Brazil	(Silva et al., 1999)
Paredão	Brazil	(Silva et al., 1999)
Curiepo	Venzuela	(Silva et al., 1999)
Sotillo	Venzuela	(Silva et al., 1999)
Panaquire	Venzuela	(Silva et al., 1999)
Birongo	Venzuela	(Silva et al., 1999)
Benin	Benin	(Herrera et al., 2004)
Kenya	Kenya	(Herrera et al., 2004)
Rwanda	Rwanda	(Herrera et al., 2004)
Tanta	Egypt	(Herrera et al., 2004)
Cameroon2	Cameroon	(Herrera et al., 2004)
Arabs	Dubai	(Alkhayat, 1996)
Turkish	Turkey	(Çakir et al., 2001)
Düsseldorf	Germany	(Kuntze et al., 1996)
Southern Italian	Italy	(Falcone et al., 1995)
Greek	Greece	(Falcone et al., 1995)
Qatari	Qatar	(Ismail M. Sebetan, 1998)
Pacoval	Amazon	(Maria et al., 2003)
Curiau	Amazon	(Maria et al., 2003)
Metro Manila	Filipino	(Miranda & Benecke, 1998)
Hasidic and non-Hasidic Jews	New York, USA	(Medintz et al., 1998)
Jordanian	Jordan	(K. Balamurugan et al., 1998)
Black African	Zimbabwe	(Wolfarth et al., 2000)
African and African derived	Venezuela	(Bernal & Villasmil, 2000)
Maracaibo	Venezuela	(Bernal & Villasmil, 2000)
Iranian	Iran	(Mahdieh et al., 2005)
North Indian	India	(Mukherjee et al., 2005)

Table 3 cont.: Populations included in the study

Population name	Country	Reference
Caucasians	USA	(Bruce Budowle et al., 1995)
African American	USA	(Bruce Budowle et al., 1995)
Southeastern Hispanic	USA	(Bruce Budowle et al., 1995)
South Western Hispanic	USA	(Bruce Budowle et al., 1995)
Orientals	USA	(Bruce Budowle et al., 1995)
Hungarians	Hungary	(B Budowle et al., 1996)
Bahrainian	Bahrain	(Tahir et al., 1999)
Saudi Arabian	Saudi Arabia	(Muhammad A. Taher et al., 2000)
Poland	Poland	(Raczek, 2001)
Arab	Gaza, Judaea, Samaria, Israel	(Jeanne M. Hayes, Bruce Budowle, 1995)
Antioquia	Colombia	(Builes et al., 2002)
Kholmogory	Russia	(D. A. Verbenko et al., 2003)
Lviv	Ukraine	(D. A. Verbenko et al., 2003)
Klav	Ukraine	(D. A. Verbenko et al., 2003)
Grodno	Byelorussia	(D. A. Verbenko et al., 2003)
Khoiniki	Byelorussia	(D. A. Verbenko et al., 2003)
Mjadel	Byelorussia	(D. A. Verbenko et al., 2003)
Staro-Shaiginsky	Mordovia	(D. A. Verbenko et al., 2003)
Romodanovsky	Mordovia	(D. A. Verbenko et al., 2003)
Sysolsky	Eastern Europe	(D. A. Verbenko et al., 2003)
Almetievsk	Eastern Europe	(D. A. Verbenko et al., 2003)
Ilishevsky	Eastern Europe	(D. A. Verbenko et al., 2003)
Abzelilovsky	Eastern Europe	(D. A. Verbenko et al., 2003)
Sterlibashevsky	Eastern Europe	(D. A. Verbenko et al., 2003)
Arkhangelsky	Eastern Europe	(D. A. Verbenko et al., 2003)
Beloretzky	Eastern Europe	(D. A. Verbenko et al., 2003)
Kuban Cossaks	Russia	(D. A. Verbenko et al.,2004)
Circassians	North Caucasians	(D. A. Verbenko et al.,2004)
Abkhazinas	North Caucasians	(D. A. Verbenko et al.,2004)
Shapsugs	North Caucasians	(D. A. Verbenko et al.,2004)
West Adygeis	North Caucasians	(D. A. Verbenko et al.,2004)
East Adygeis	North Caucasians	(D. A. Verbenko et al.,2004)

Table 3 cont.: Populations included in the study

*AfA: African American

Population	Country	Reference
Valencia	Spain	(Peterson et al., 2000)
Madrid	Spain	(Peterson et al., 2000)
Andalusia	Spain	(Peterson et al., 2000)
Barcelona	Spain	(Peterson et al., 2000)
North England	United Kingdom	(Peterson et al., 2000)
Baranya	Hungary	(Peterson et al., 2000)
Caucasian	Hungary	(Peterson et al., 2000)
Belgian	Brussels	(Peterson et al., 2000)
Moroccan	Brussels	(Peterson et al., 2000)
Turks	Brussels	(Peterson et al., 2000)
Malays	Malaysians	(Peterson et al., 2000)
Orientals	Malaysians	(Peterson et al., 2000)
Mapuche	Argentina	(Peterson et al., 2000)
Mataco	Argentina	(Peterson et al., 2000)
Caucasians	Israel	(Peterson et al., 2000)
Riyadh	Saudi Arabia	(Peterson et al., 2000)
Nigerians	Nigeria	(Peterson et al., 2000)
Dogrib Indians	Canada	(Peterson et al., 2000)
Caucasians Alabama	USA	(Peterson et al., 2000)
CaucasiansMinessota	USA	(Peterson et al., 2000)
Caucasians Nevada	USA	(Peterson et al., 2000)
Caucasians Texas	USA	(Peterson et al., 2000)
Caucasians Virginia	USA	(Peterson et al., 2000)
AfAAlabama	USA	(Peterson et al., 2000)
AfAMinessota	USA	(Peterson et al., 2000)
AfANevada	USA	(Peterson et al., 2000)
AfATexas	USA	(Peterson et al., 2000)
AfAVirginia	USA	(Peterson et al., 2000)
AfA California	USA	(Peterson et al., 2000)
AfAMissouri	USA	(Peterson et al., 2000)

Table 3 cont.: Populations included in the study

*NA: Native Americans

Population	Country	Reference
Hispanic California	USA	(Peterson et al., 2000)
Hispanic Florida	USA	(Peterson et al., 2000)
Hispanic Nevada	USA	(Peterson et al., 2000)
Hispanic Southeast	USA	(Peterson et al., 2000)
Hispanic Southwest	USA	(Peterson et al., 2000)
Hispanic Texas	USA	(Peterson et al., 2000)
Hispanic Virginia	USA	(Peterson et al., 2000)
NA Alaska N. Slope	USA	(Peterson et al., 2000)
NA Alaska B. Wade	USA	(Peterson et al., 2000)
NA Alaska	USA	(Peterson et al., 2000)
NA Minessota	USA	(Peterson et al., 2000)
NA Navajo	USA	(Peterson et al., 2000)
NA New Mexico	USA	(Peterson et al., 2000)
NA Pueblo	USA	(Peterson et al., 2000)
NA Sioux	USA	(Peterson et al., 2000)
Mexico city	Mexico	(Peterson et al., 2000)
Konkanasthas	India	(Birajalaxmi et al., 2002)
Marathas	India	(Birajalaxmi et al., 2002)
Nairs	India	(Birajalaxmi et al., 2002)
Ezhavas	India	(Birajalaxmi et al., 2002)
Muslims	India	(Birajalaxmi et al., 2002)
Punjabi	India	(Trivedi et al., 2002)
Kuki	India	(Trivedi et al., 2002)
Hmar	India	(Trivedi et al., 2002)
Ekere	India	(Trivedi et al., 2002)
Reddy	India	(Trivedi et al., 2002)
Baniya	India	(Trivedi et al., 2002)
Deshast Brahmin	India	(Trivedi et al., 2002)
Bengali Kayastha	India	(Trivedi et al., 2002)

Table 3 cont.: Populations included in the study

Population	Country	Reference
Tamil	Madras, India	(K Balamurugan et al., 2001)
Tamil	South India	(Vanaja, M. et al 2001)
Bahari	Patalkot, India	(Rajesh Biswas, 2002)
Thoti	Andhra Pradesh, India	(Sachdeva et al., 2004)
Kolam	Andhra Pradesh, India	(Sachdeva et al., 2004)
Modern Malay	Sarawak, Malaysia	(Roslan, Azizan, & Saat, 2009)
Romanian	Romania	(Ceacareanu et al., 2004)

3.6 Statistical analysis:

The PCR products' size was determined by the regression analysis. The size was calculated by the linear regression through using the migration distance of each band relative to the 100 bp marker which is inversely proportional to repeat number units at the loci (Roslan et al., 2009).

For other statistical analyses R program with different packages was used. For allelic representation Population and Evolutionary Genetics Analysis System (PEGAS) package was utilized for writing, reading, analyzing, and plotting the allelic haplotypic data (Paradis, 2010). It is incorporated with two other packages Analysis of Phylogenetic and Evolution (APE) and Adegenet for implementing multivariate analysis for genetic markers including the observed and expected heterozygosity (Jombart, 2008; Paradis, Claude, & Strimmer, 2004). Additionally, Hardy Weinberg equilibrium agreement was determined by Chi square test and the exact test for the whole population and for the two subpopulations. Pegas package was also used to measure F statistics (Fis and Fst).Vegan package was used to provide functions for diversity analyses and plotting of results by non-metric multidimensional scaling and hierarchical clustering dendrograms for different populations included in the study (Oksanen et al., 2016). Clustering of the Nubian population with other different population groups as African, American, Arabian, Asian, Caucasians, and European has been also performed.

4.0 Results:

The alleles of the D1S80 VNTR locus were detected according to the amplified fragment polymorphism technique (Amp-FLP) followed by agarose gel electrophoresis. The 2% agarose gel was efficient to resolve the convergent alleles (Figure 7, 8, 9, 10, and 11).

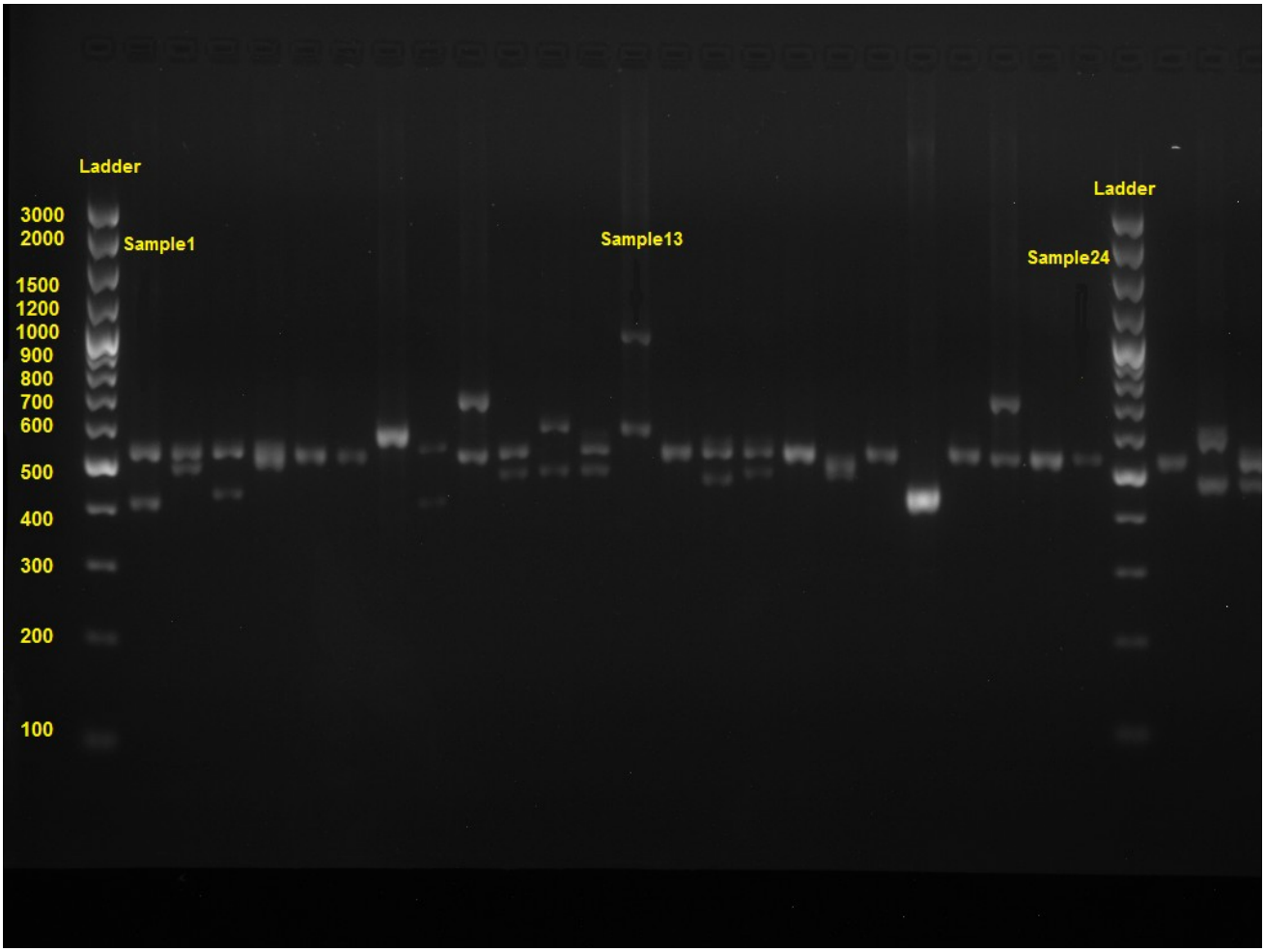

Figure 7: 24 PCR products detected on 2% (w/v) agarose gel and 1X TBE buffer (Gel 1, Part 1).

75 genotypes were detected in our study among 105 Nubian individual (Table 4). The most prevailed genotype was 27/27 with 0.066 % followed by 26/26 with 0.057 %. The study resulted in detection of 36 alleles (Table 5). The highest frequency was for allele 27 represents 0.119 % followed by allele 26 represents 0.09 % (figure 12). The observed heterozygosity is 0.69 % while the expected heterozygosity is 0.94%. The corresponding observed homozygosity is 0.31% while the expected homozygosity is 0.06% respectively.

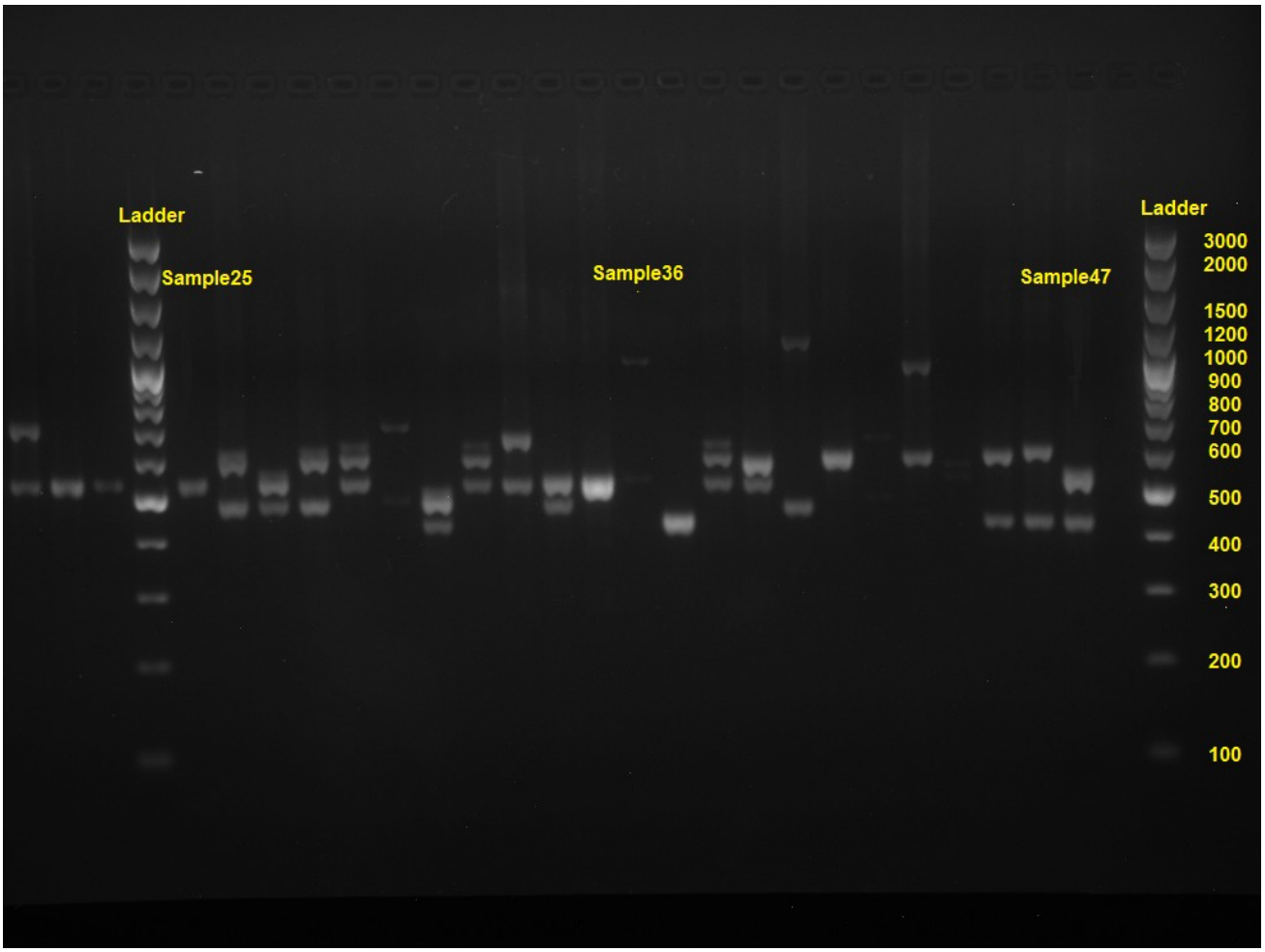

Figure 8: 23 PCR products detected on 2% (w/v) agarose gel and 1X TBE buffer (Gel 1, Part 2).

For each of the subpopulations, Matokkiis has 29 alleles with observed heterozygosity 0.62% and expected heterozygosity 0.93%. On the other hand, the Fadjikkiis has 27 alleles and the observed heterozygozity is 0.79% while the expected is 0.93%. The departure from the HWE was confirmed by Chi square test for the whole population and for Matokkii subpopulation. In addition, the probability exact test for the total population and for the two subpopulationsconfirmed the same results (table 6). The F statistics for the total population and for subpopulation indicated that there is no differentiation between them (table 6).

Figures 13, 15, 17, 19, 21 and 23 show non-metric dimensional scaling of Egyptian Nubian population in association with different countries and ethnic groups divided

into different geographical regions as African, American, Arabian, Asian, Caucasian and European respectively. In all figures, except in the Asian group, the Nubian population is placed in a position while other populations are aggregated in another position.

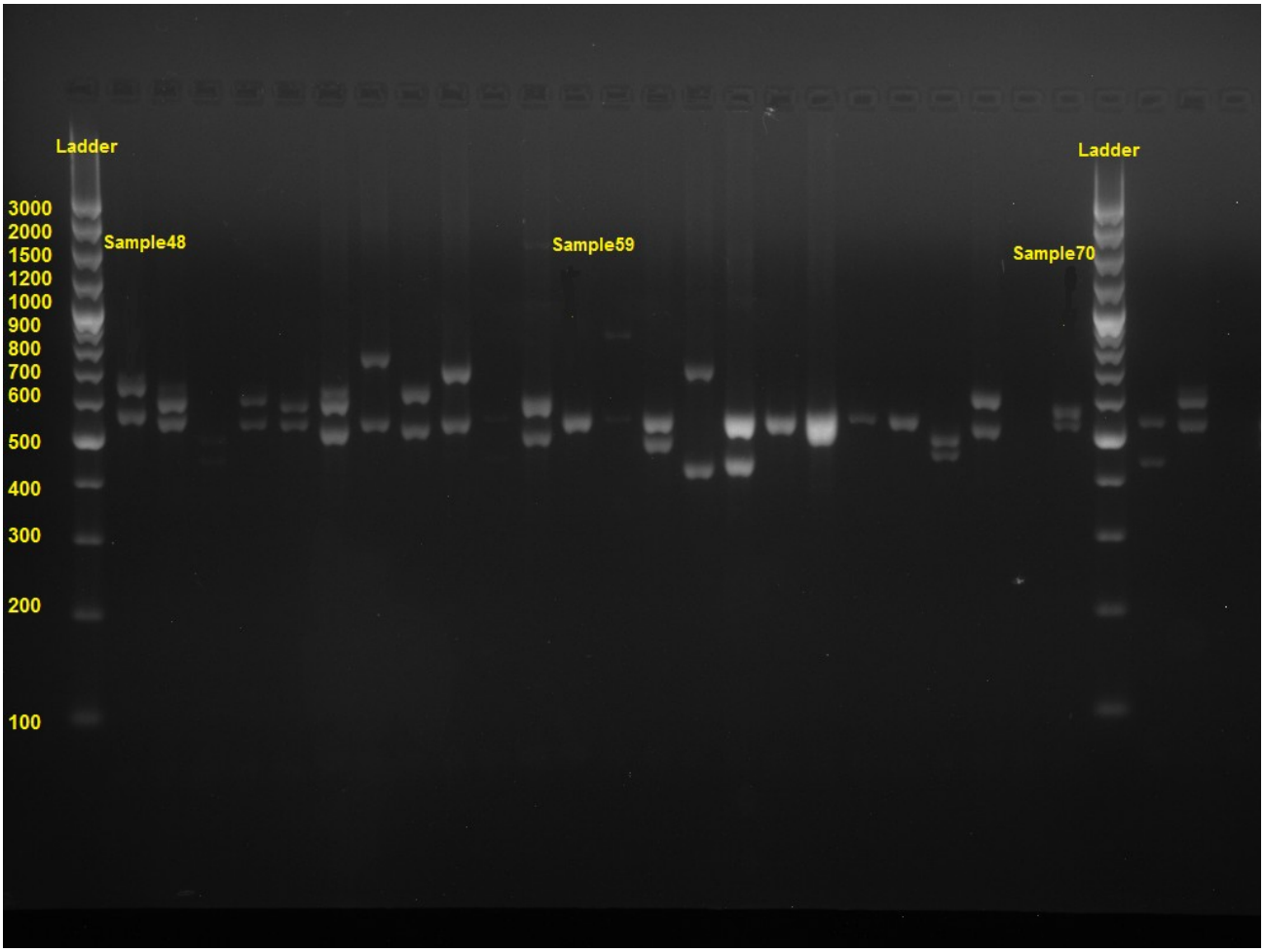

Figure 9: 23 PCR products detected on 2% (w/v) agarose gel and 1X TBE buffer (Gel 2, Part 1).

In the Asian group the North Indian and Modern Malay populations are in proximity of the Nubian population.

Figure 14, 16, 18, 20, 22 and 24 display dendrogram clustering of the same previous populations with the Nubian populations. As the NMDS graphs, the dendrograms place the Nubian population in on clade while other populations are clustered in other clade that are branched in different clades. In the Asian group, the Nubian population is clustered with Modern Malay population in the same clade which is differentiated in another clade containing North Indian population.

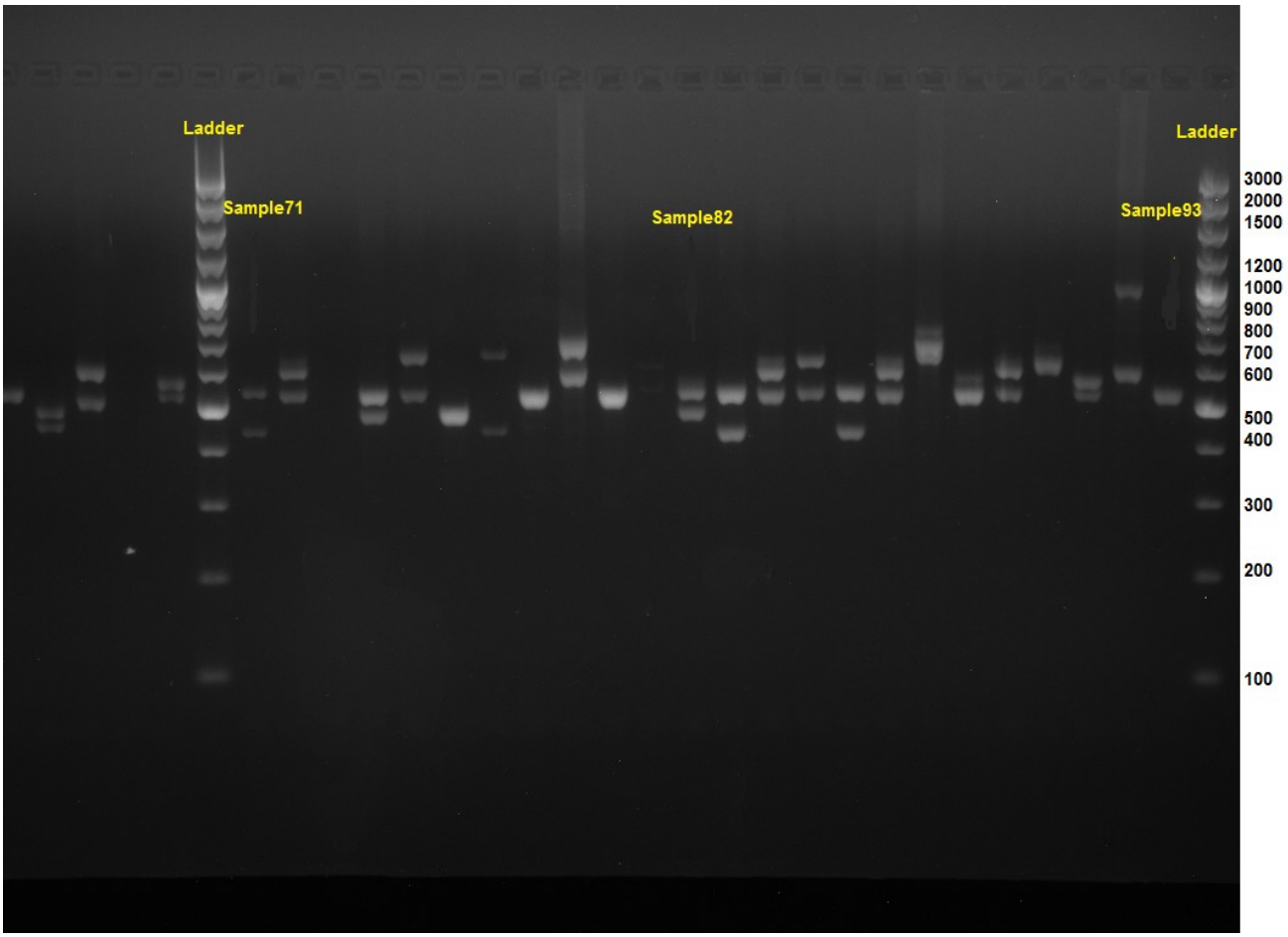

Figure 10: 23 PCR products detected on 2% (w/v) agarose gel and 1X TBE buffer (Gel 2, Part 2).

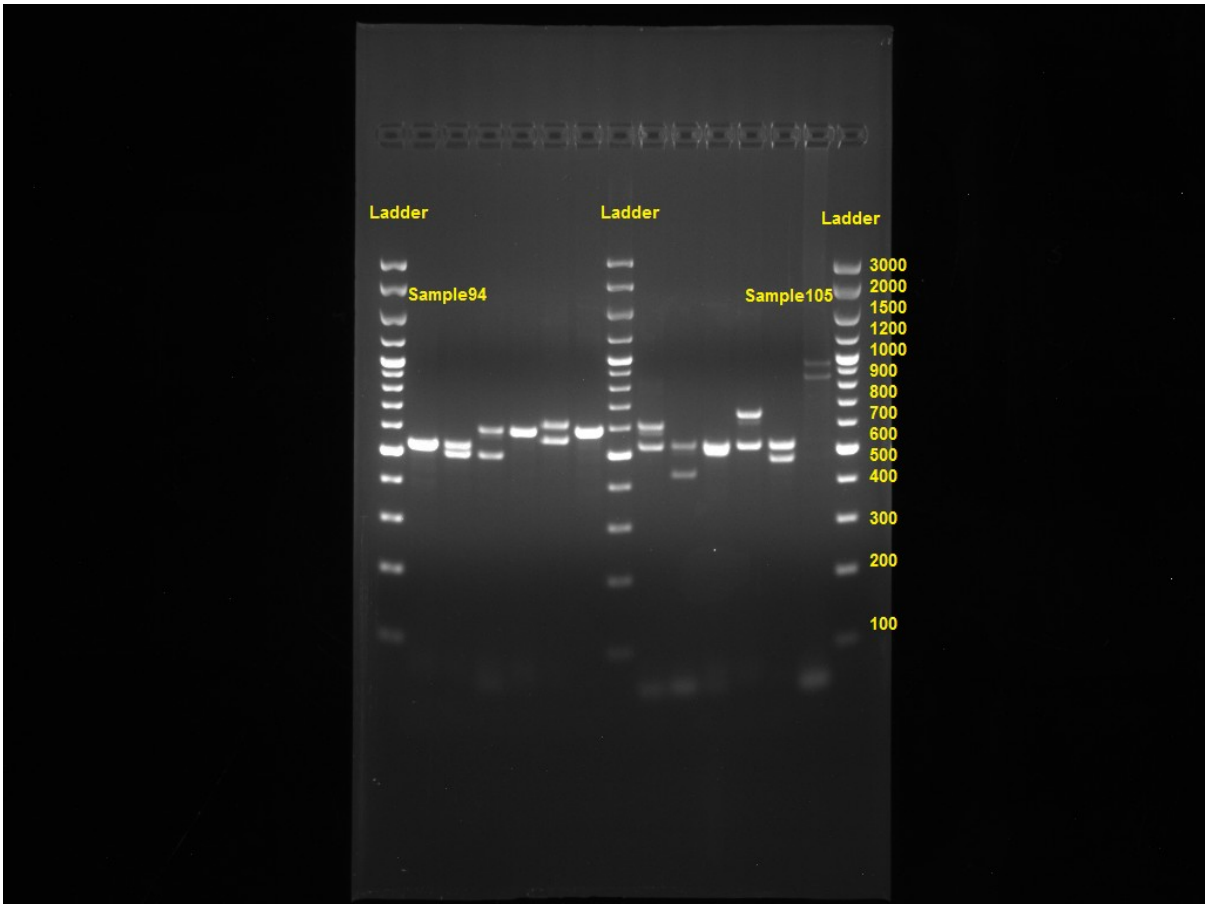

Figure 11: 12 PCR products detected on 2% (w/v) agarose gel and 1X TBE buffer (Gel 3).

Table 4: Shows genotype frequencies of the D1S80 locus among 105 Nubian individual. 75 genotypes were detected

Genotype	Frequency %	Genotype	Frequency %	Genotype	Frequency %
13/26	0.0095	22/24	0.0095	27/31	0.019
16/24	0.0095	22/28	0.0095	27/32	0.019
17/17	0.0095	23/23	0.0285	27/40	0.0095
17/21	0.0095	23/27	0.0285	27/42	0.0095
17/25	0.0095	23/29	0.0095	27/45	0.0095
17/26	0.0095	23/31	0.0095	28/28	0.0285
17/30	0.0095	23/33	0.0095	28/33	0.0095
17/31	0.0095	24/24	0.0095	28/35	0.0095
18/26	0.0095	24/28	0.019	28/38	0.0095
18/28	0.019	24/29	0.0095	29/29	0.019
18/41	0.0095	24/30	0.0095	29/30	0.0095
19/19	0.0095	24/34	0.0095	29/31	0.0095
19/23	0.0095	25/25	0.0285	29/35	0.0095
19/27	0.0095	25/28	0.0095	29/36	0.0095
20/23	0.0095	25/29	0.0095	29/37	0.0095
20/29	0.019	25/30	0.0095	30/30	0.0285
20/71	0.0095	25/33	0.0095	30/37	0.0095
21/21	0.0095	25/31	0.019	30/55	0.0095
21/23	0.0095	25/35	0.0095	30/60	0.0095
21/25	0.019	26/26	0.057	31/40	0.0095
21/28	0.0095	26/29	0.0095	33/63	0.0095
21/30	0.019	26/33	0.0095	34/65	0.0095
21/36	0.0095	26/41	0.0095	36/36	0.0095
21/39	0.0095	26/63	0.0095	39/39	0.0095
21/40	0.0095	27/27	0.066	50/59	0.0095

Table 5: Allele frequencies of the D1S80 locus among 105 Nubian individual. 36 allele were observed

Allele	Observation	Frequency
13	1	0.48%
16	1	0.48%
17	7	3.33%
18	4	1.90%
19	4	1.90%
20	4	1.90%
21	12	5.71%
22	2	0.95%
23	15	7.14%
24	9	4.29%
25	16	7.62%
26	19	9.05%
27	25	11.90%
28	16	7.62%
29	15	7.14%
30	15	7.14%
31	8	3.81%
32	2	0.95%
33	5	2.38%
34	2	0.95%
35	3	1.43%
36	4	1.90%
37	2	0.95%
38	1	0.48%
39	3	1.43%
40	3	1.43%
41	2	0.95%
42	1	0.48%
45	1	0.48%
50	1	0.48%
55	1	0.48%
59	1	0.48%
60	1	0.48%
63	2	0.95%
65	1	0.48%
66	1	0.48%
71		

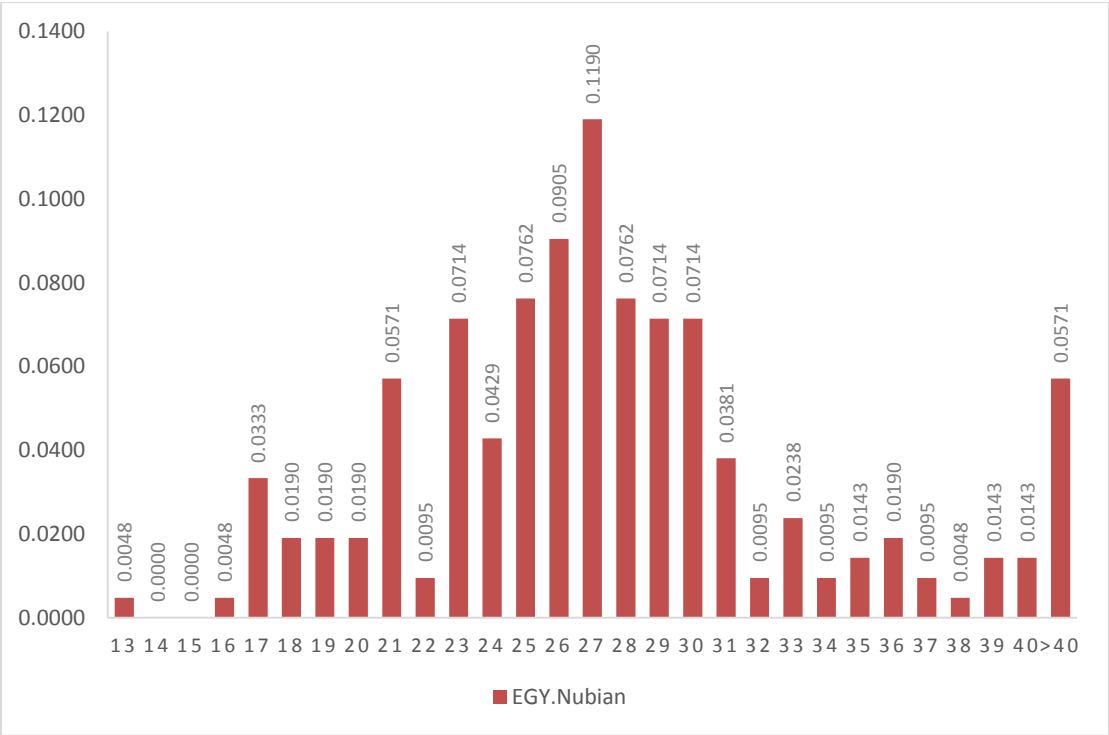

Figure 12: Allelic distribution chart among the Nubian population

Table 6: Statistics of the total Nubian population and the two subpopulations

Population	Total Nubian Population	Fadjikkii	Matokkii
Group size	105	39	66
Alleles number	36	27	29
Expected heterozgosity	0.94	0.93	0.93
Observed heterozygosity	0.69	0.79	0.62
Chi^2 p value	$6.92E^{-11}$	0.2	$1.14E^{-09}$
Exact p value	0	0.003	0
Fst	0.005	0.01	0.01
Fis	0.27	---	---

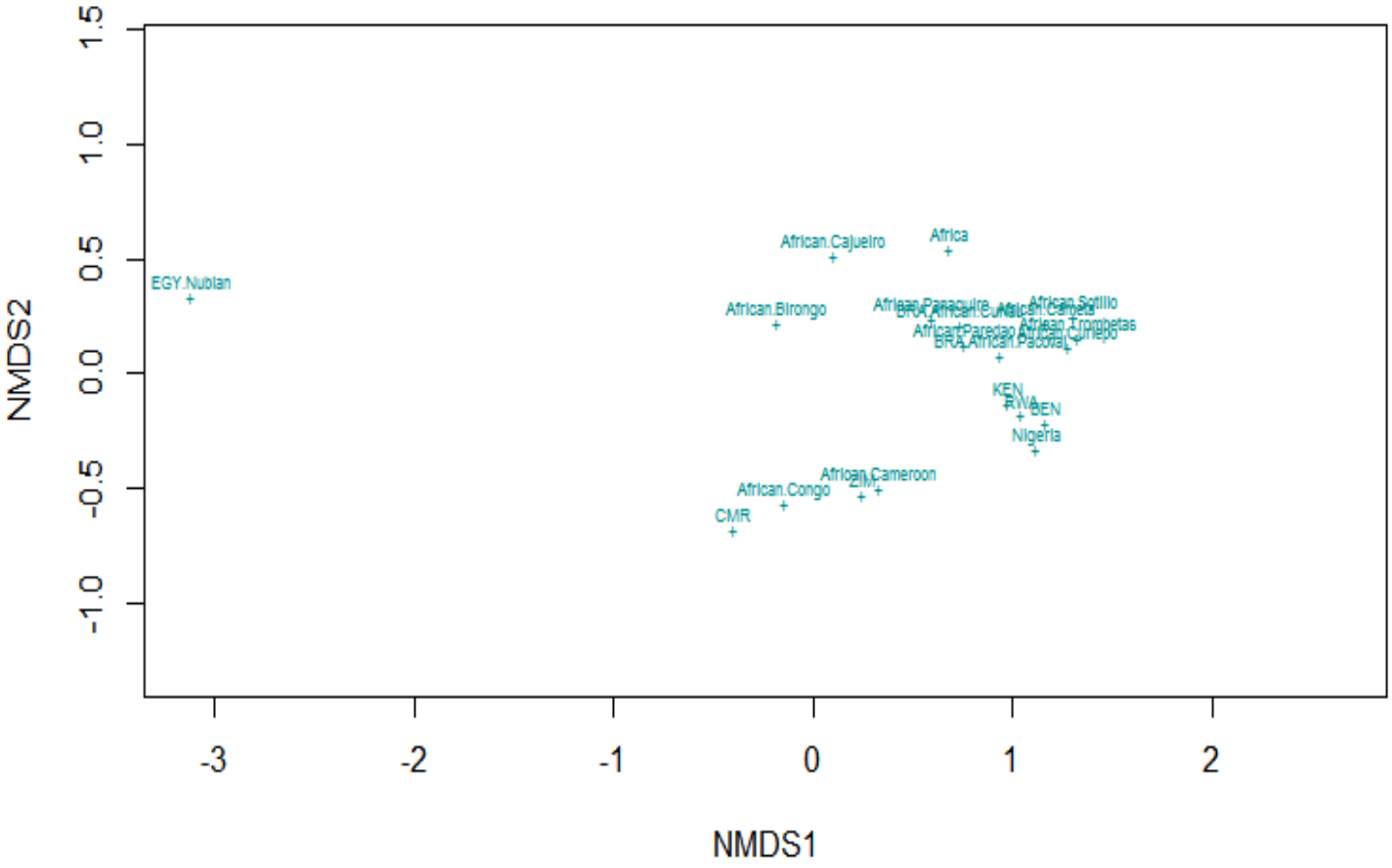

Figure 13: Non-metric multidimensional scaling of Nubian population with African populations

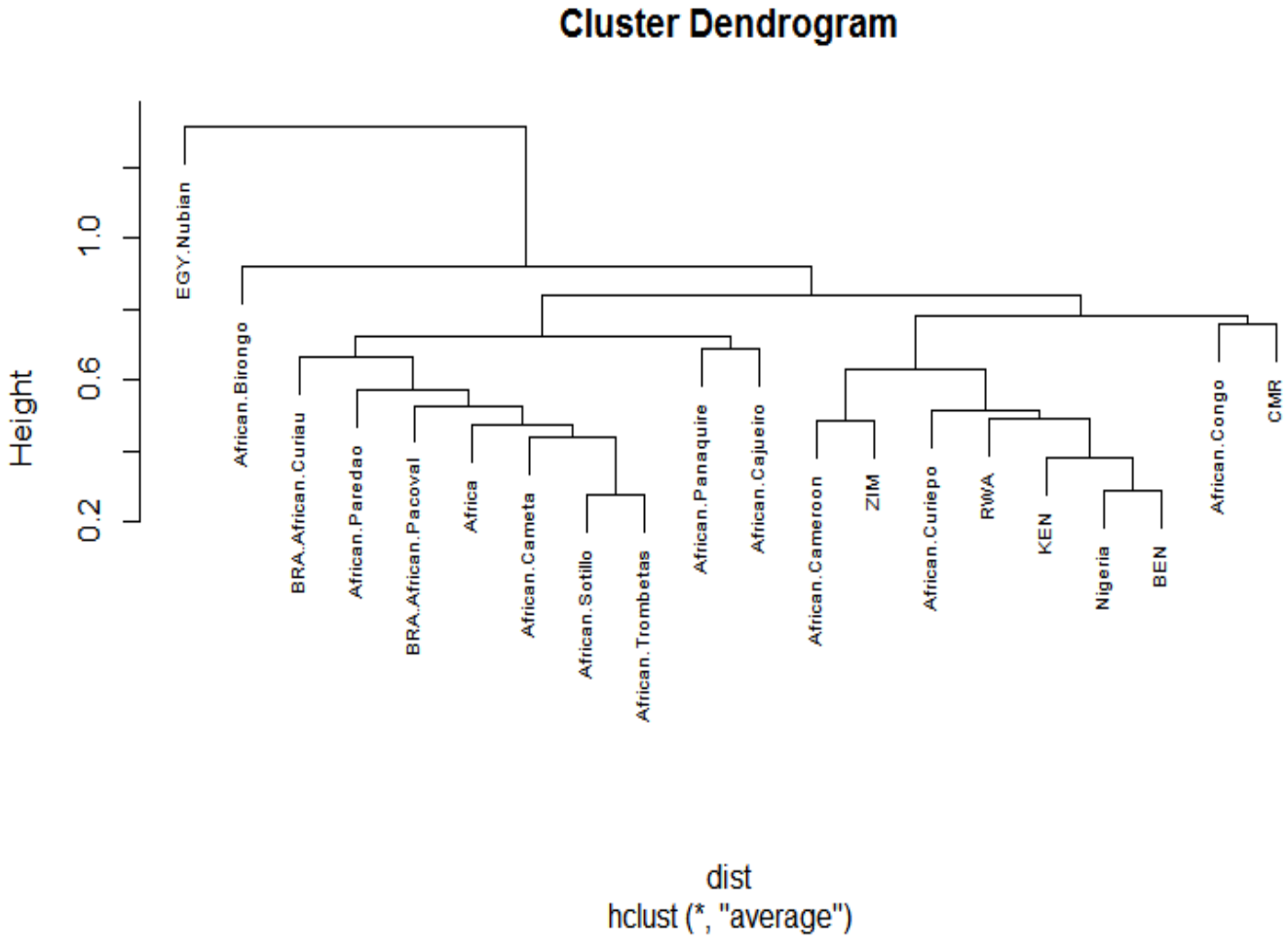

Figure 14: Cluster dendrogram of Nubian population with African populations

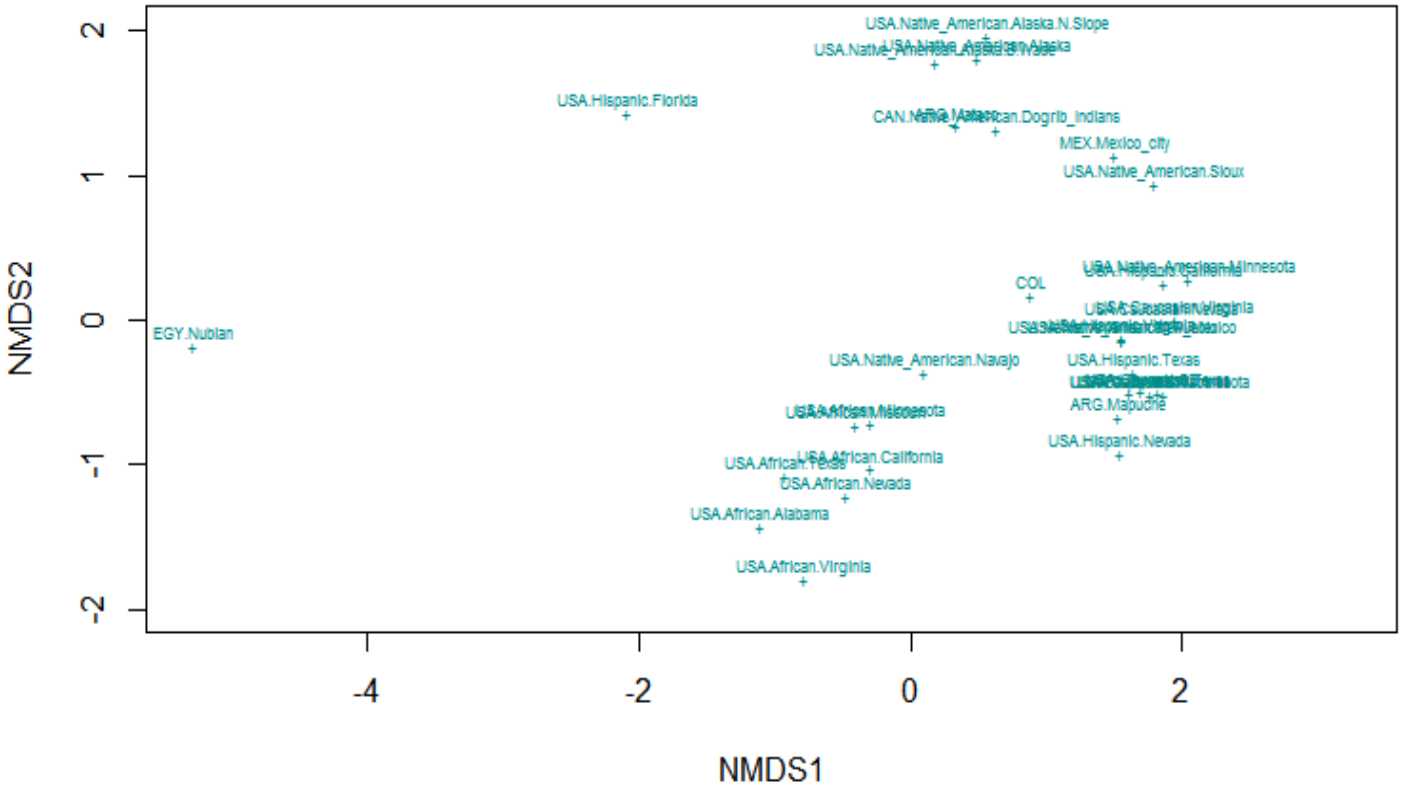

Figure 15: Non-metric multidimensional scaling of Nubian population with American populations

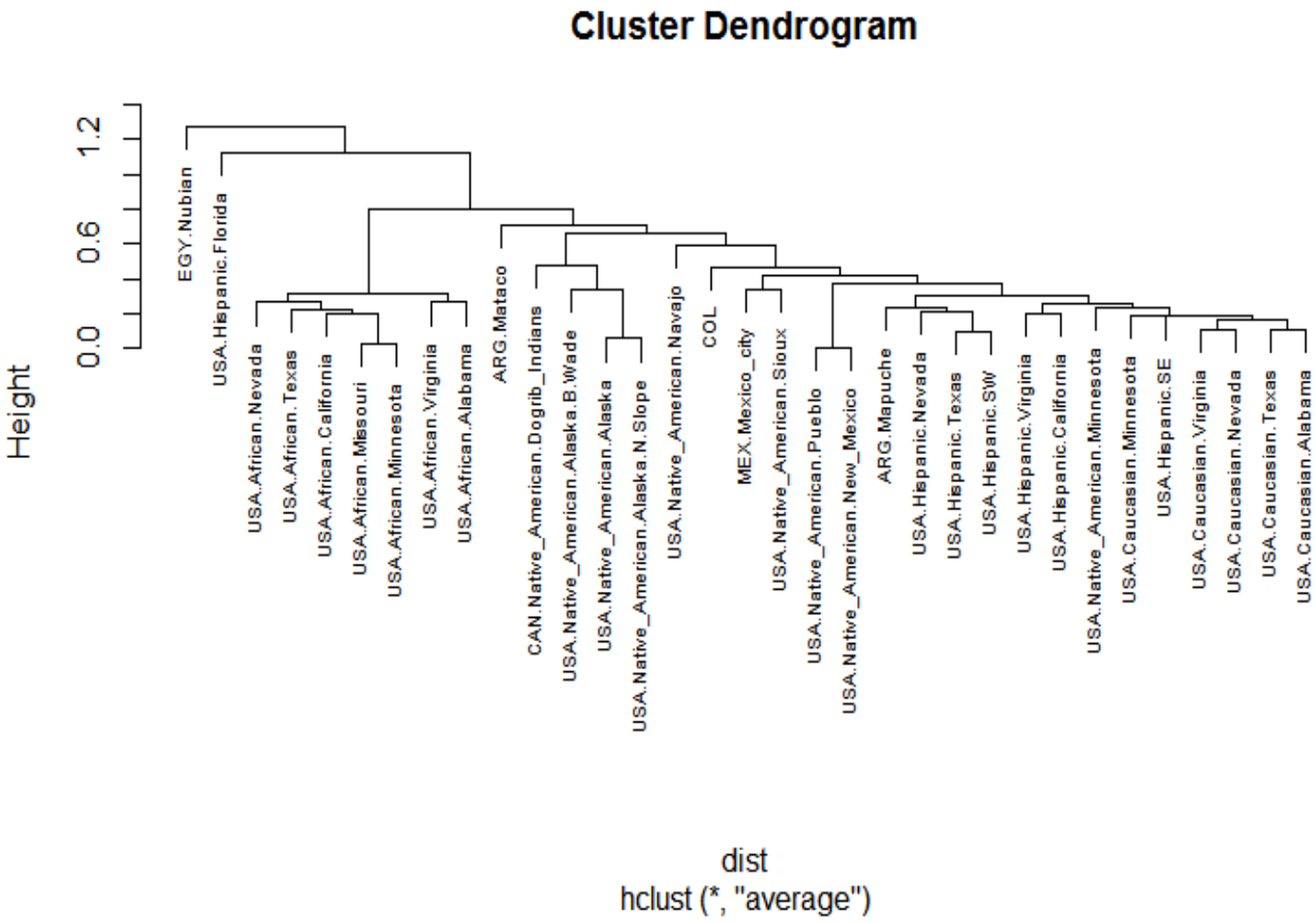

Figure 16: Cluster dendrogram of Nubian population with American populations

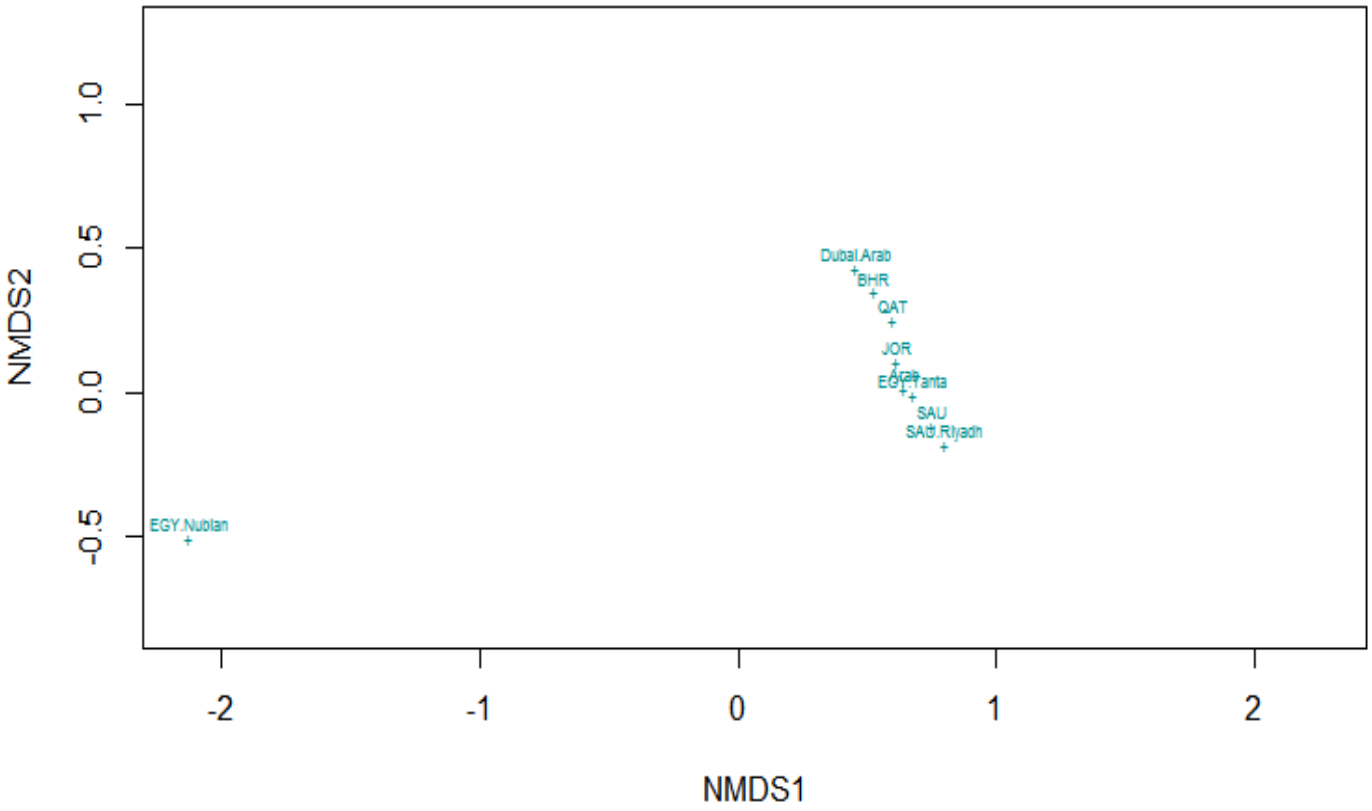

Figure 17: Non-metric multidimensional scaling of Nubian population with Arabian populations

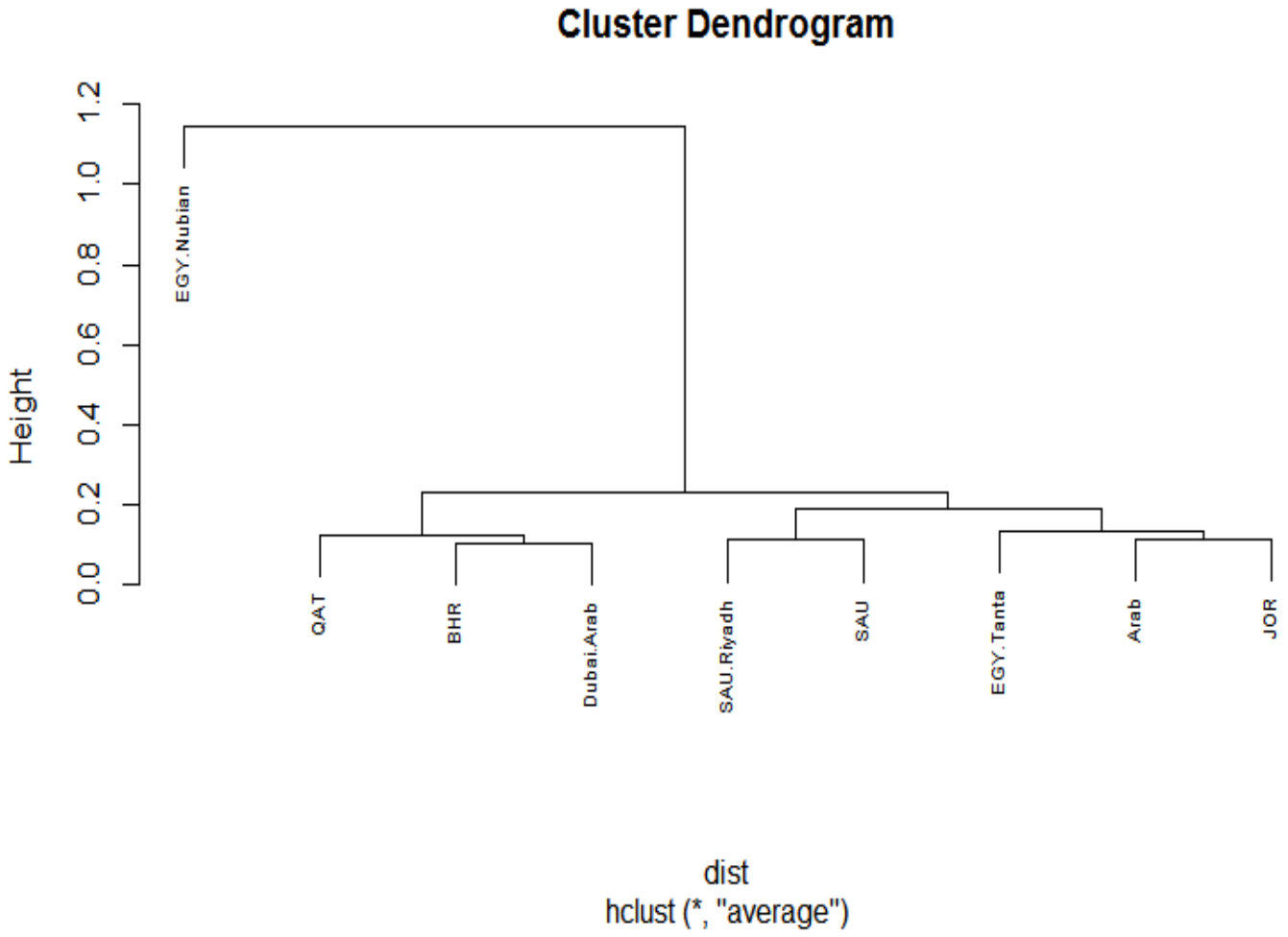

Figure 18: Cluster dendrogram of Nubian population with Arabian populations

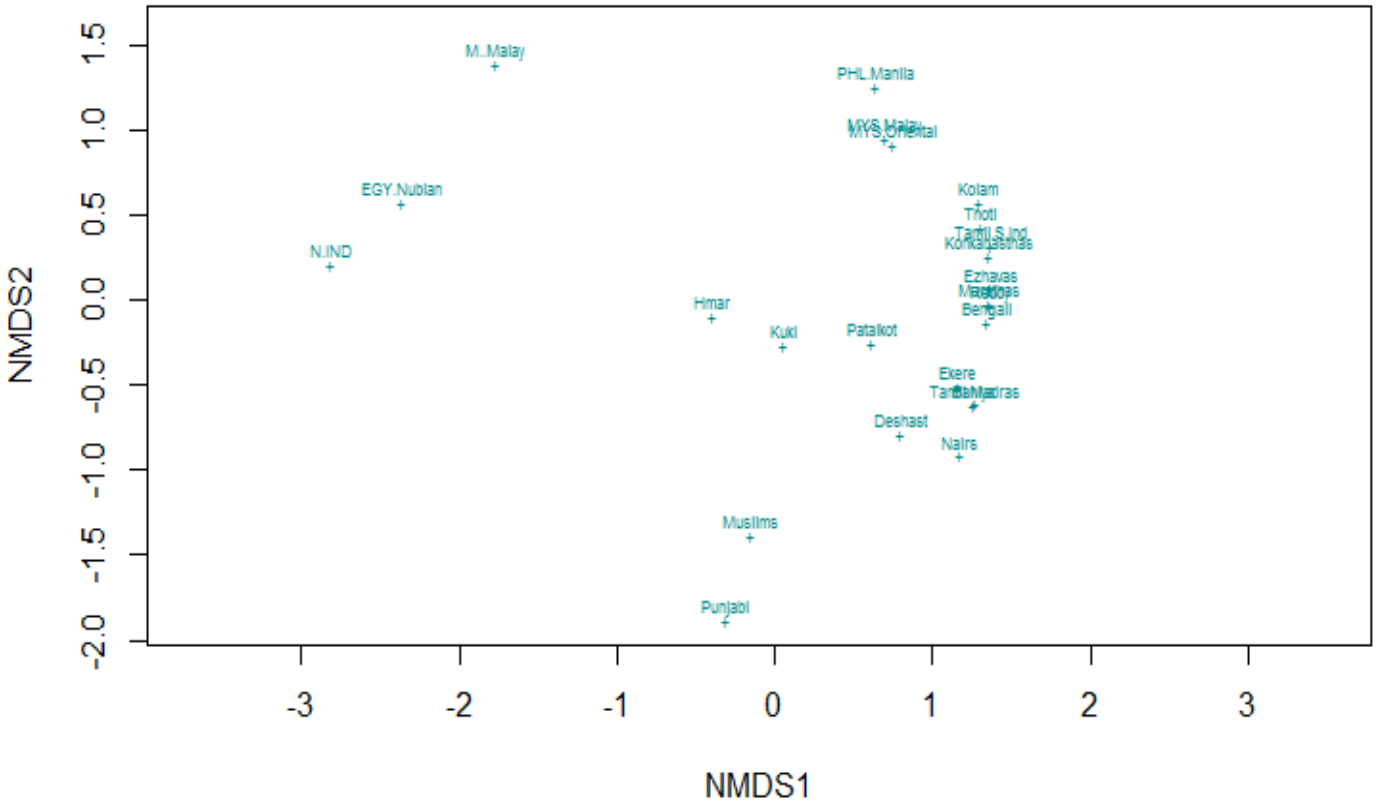

Figure 19: Non-metric multidimensional scaling of Nubian population with Asian populations

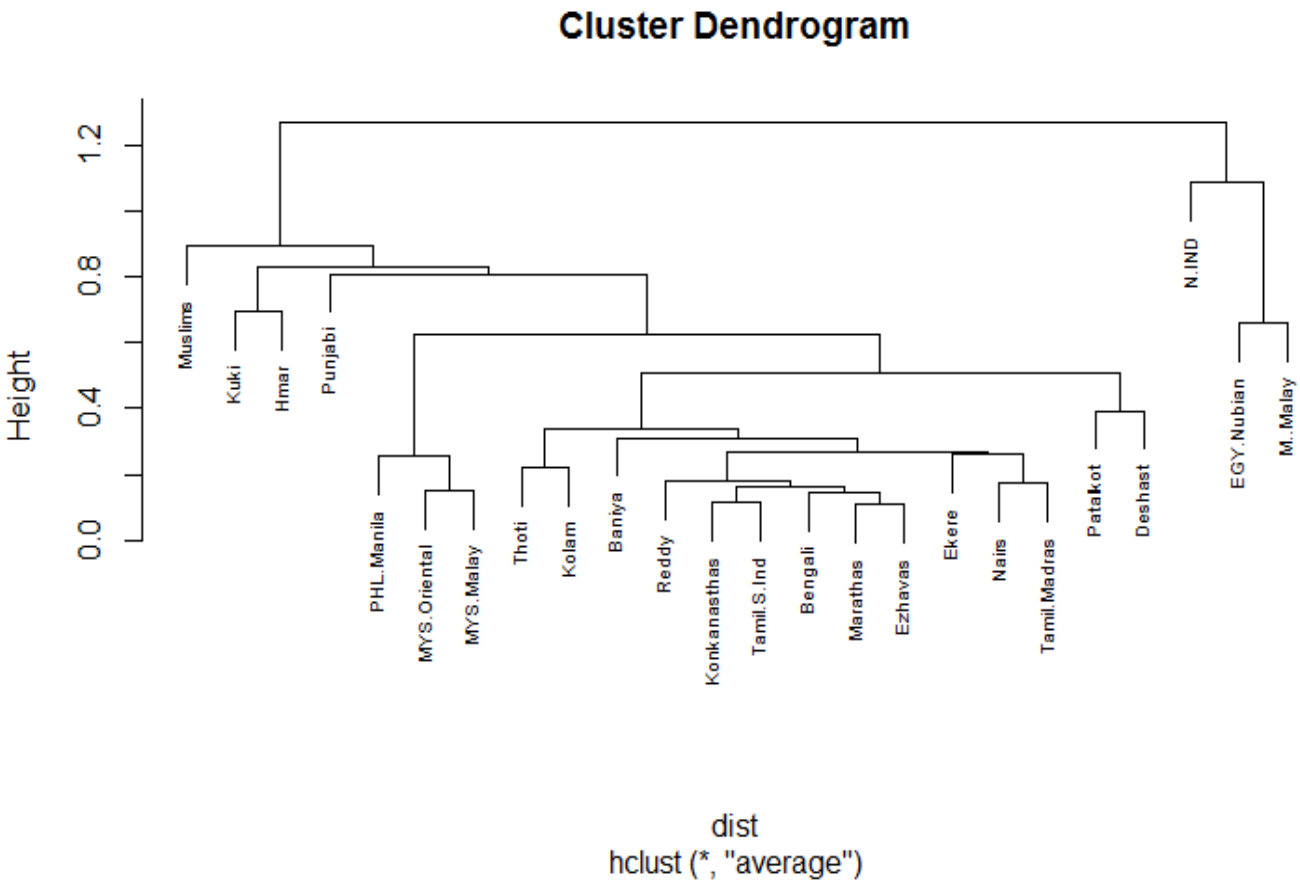

Figure 20: Cluster dendrogram of Nubian population with Asian populations

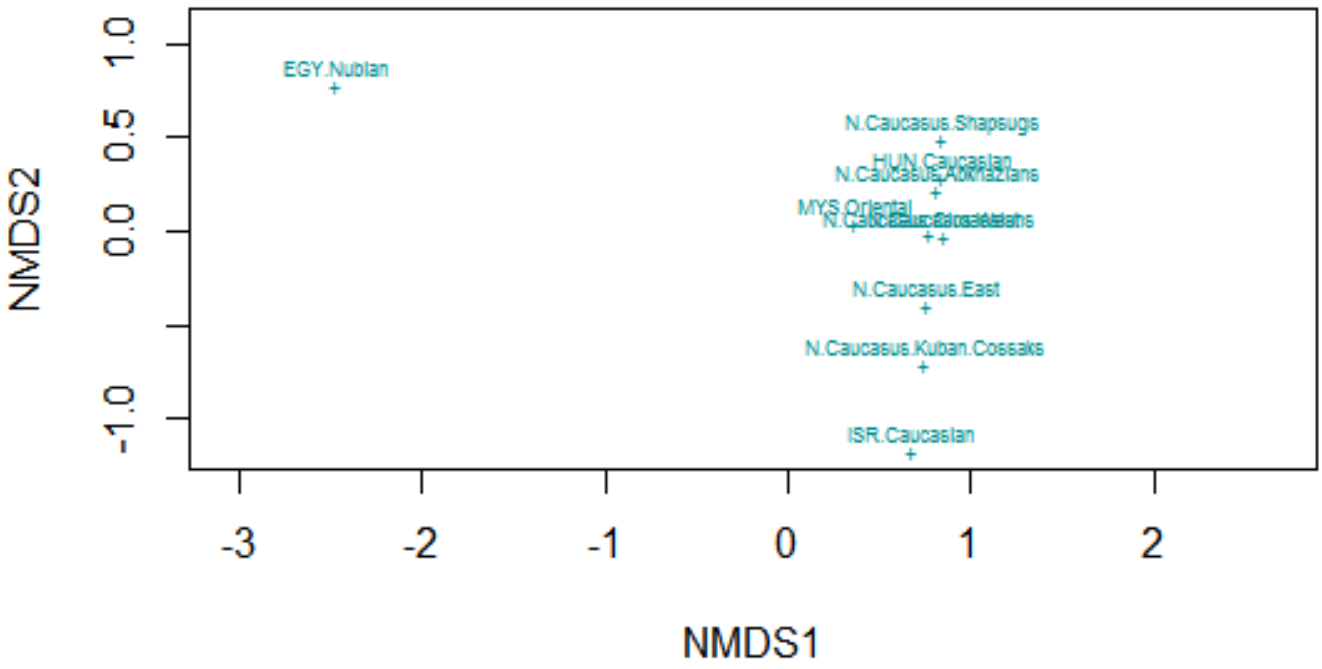

Figure 21: Non-metric multidimensional scaling of Nubian population with Caucasian populations

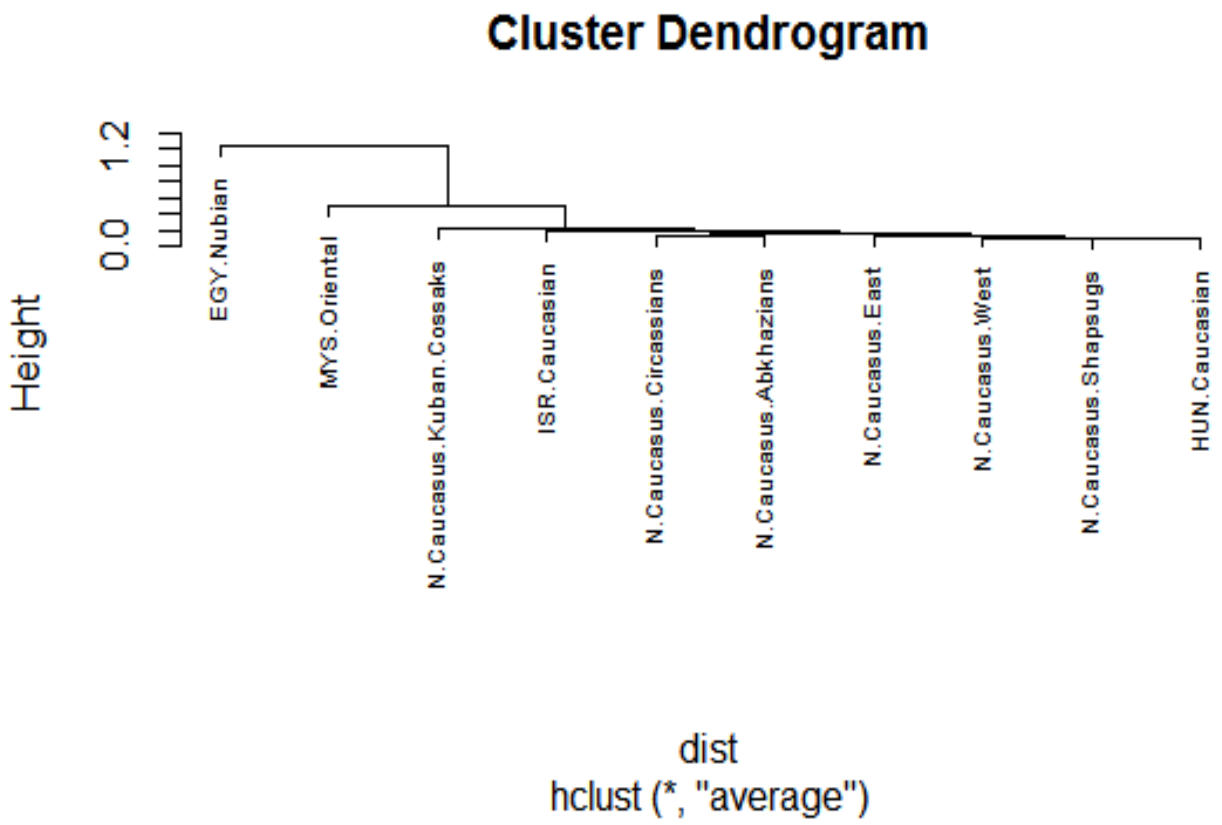

Figure 22: Cluster dendrogram of Nubian population with Caucasian populations

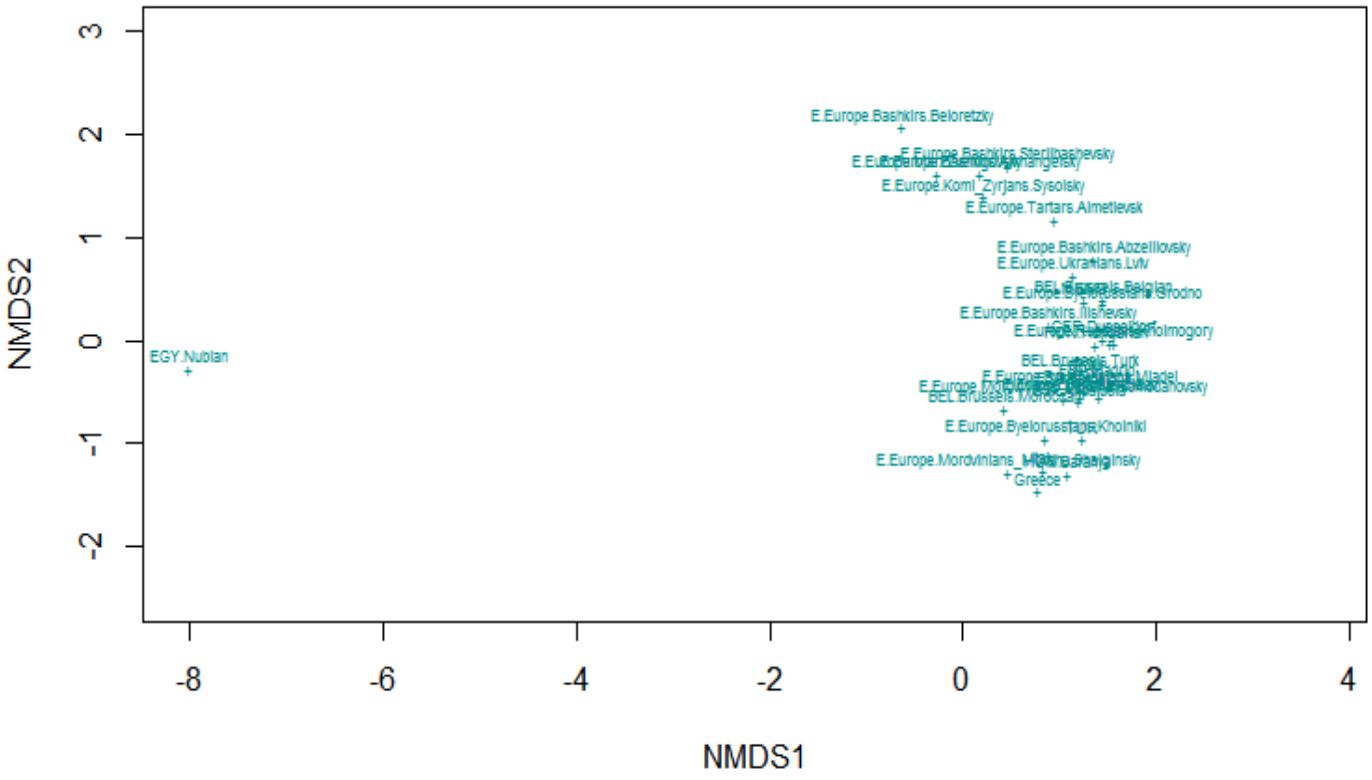

Figure 23: Non-metric multidimensional scaling of Nubian population with European populations

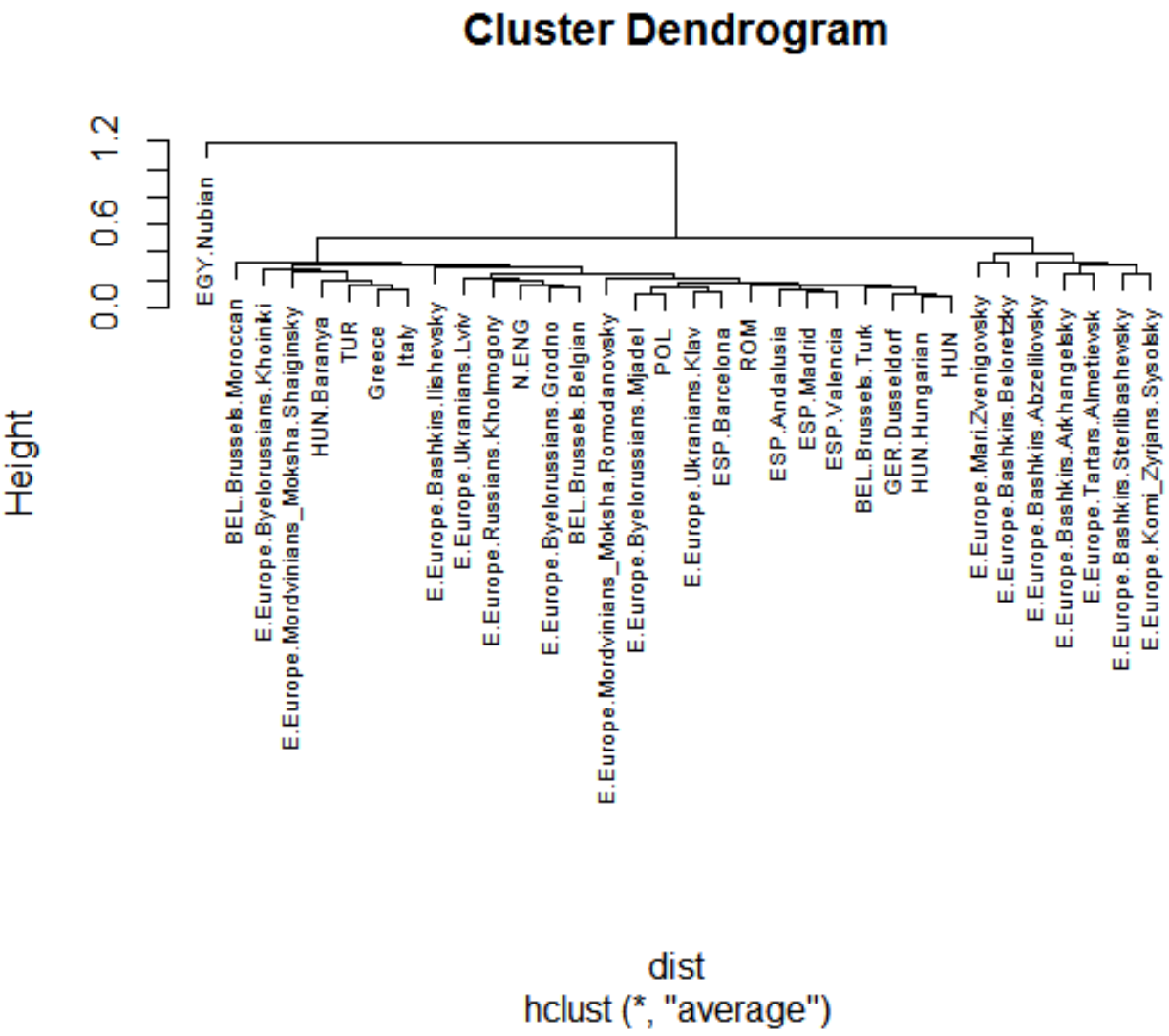

Figure 24: Cluster dendrogram of Nubian population with European populations

Chapter Five:

5.0 Discussion:

D1S80 polymorphism is attributed mainly to the different number of repeats of the core sequence which results in different alleles in addition to some microvariations due to change in the core sequence of the repeats (Duncan et al., 1997). Detection of number of repeats of D1S80 PCR product was facilitated by using the allelic ladder by which it can distinguish the difference by one repeat in the PCR product(Kloosterman, Budowle, & Daselaar, 1993). Although the most known allele of D1S80 locus ranges from 14-42 alleles, there are smaller and lager alleles which have been detected in our study such as 13, 45, 50, 55, 59, 60, 63, 65, and 71. The large alleles were previously reported in different studies (Ingrid Eisenbarth, Günter Assum & Krone, 1999; Watanabe & Shimizu, 2002). It is obvious that the most common allele within the Nubian population is 27 followed by allele 26 unlike most of worldwide populations who have common alleles28 and 14(Gross, Carmody, & Guerrieri, 1997).

Since the p value of the total population is highly significant for both Chi^2 test and the exact test, this indicates rejection of the null hypothesis and the deviation from the Hardy-Weinberg equilibrium. Additionally, because of the excess of the observed homozygosity rather than the expected, it might be ascribed to the non-random mating by inbreeding. This is compatible with the prevailed consanguinity marriage among the Nubian population. Within the subpopulations, the p values of the exact test are significant while the p value of the Chi^2 test was not significant in Fadjikkii subpopulation which could be due to the small sample size(Wang, 2005). But generally, the p values confirm the departure from HWE.

In order to know if there is a genetic differentiation (substructuring of the Nubian population) between the two subpopulations Fst is calculated. Because the result is 0.005 ($P < 0.05$), this indicates no differentiation between them which is confirmed by the similar results of Fst that are obtained for each of them. Consequently, they should be considered as a one population.

Fis is the inbreeding indicator and non-random mating within a population. The calculated Fis is 0.27 points to prevalence of inbreeding among the Nubian population

since the range of inbreeding is 0<Fis≤1 (Silva et al., 1999). This is ensures the fact that inbreeding is attributed to the dominance of consanguinity marriage among the Nubian individuals.

To facilitate the comparison between of the Nubian population and other worldwide countries and ethnic groups, cluster dendrograms were performed to infer the genetic relatedness. In figure 13, African populations are aggregated together as they are characterized by their high genetic variation and heterozygosity(Herrera et al., 2004). Although the Nubian population is a part of the African continent, it is separated from other African population which might suggest that it is considered as an isolated area. Figure 14 demonstrates the cluster dendrogram as the Nubian population is clustered in one clade that is separated from the other populations.

Figures 15 and 16 of the American population group reveal approximate closeness between the Nubian population and the American Hispanic population in Florida. However, the allele frequency data do not reveal any similarity. Accordingly, this needs confirmation by other genetic markers and application of multivariate analyses.

Figure 17 and 18 comparing the Nubian population with Arabian countries. According to the historical events and the Arab conquer and intermarriage with the Nubians, it was expected to find a strong relationship between them. Nevertheless, both figures reveal the distant relationship between Arab and non-Arab Egyptian Nubians.

Figure 19 illustrates the proximity between the Nubian, Modern Malay and North Indian populations in the Asian population group. Figure 20 depicts presence of Nubian and Modern Malay populations (lives in Sarawak province in Malaysia) in one clade and the North Indians in the same cluster. This is attributed to the dominance of allele 27 followed by allele 26 among Malay population as the Nubian population. Allele 27 was reported to have high frequency in Oriental populations (Harashima, Liu, Katsuyama, Ota, & Fukushima, 1997). This might suggest occurrence of gene flow into the Nubian population since the ancient times.

In figure 21 and 22 illustration of the Nubian populations with Caucasian race, populations from different countries are shown. There is also no obvious relatedness between them. They display delineation of the Caucasian populations distant from the

Nubian population and presence of the Nubian population in separated clade in the dendrogram. Therefore, it indicates a genetic distinction between Nubian and Caucasian populations.

Figure 23 and 24 display the relationship between the European populations and the Nubian one. Similar to the Caucasion populations, the NMDS display a large genetic distance between the aggregated populations and the Nubian population. In the cluster dendrogram the Nubian population is separated in one clade while all other populations are aggregated in another clade clustering all the European populations. Nonetheless, the first proximate population in the European cluster is the Moroccan population which lives in Brussels. The relationship between Nubian and Moroccan population was mentioned in "History and Geography of the Human Genes" (figure5). This was depicted according to analysis of different genes among the African continent (Cavalli-Sforza et al., 1994).

Chapter six: Conclusion and Future perspective:

6.1 Conclusion:

In this thesis, D1S80 database, as an autosomal marker, was reported for the Nubian population for the first time with elucidation of its phylogenetic relationship with worldwide populations. Despite the richness of the ancient and modern Nubian history and the different migration events into it, the study reveals that the contemporary Nubians are isolated to a great extent which is attributed to prevalence of the consanguinity marriages. This was revealed by the commonness of allele 27 which is not dominant except in one Asian population. This requires comparative studies of different markers to ensure if there is a distinction of the Nubian population that might lead to discover a specific Nubian genetic fingerprint. In addition, the study proves that D1S80 is powerful in population discrimination.

6.2 Future perspective:

This study is considered as a preliminary study to recognize the genetic structure of the Egyptian Nubian population. Next studies should include the Arab group since it might contribute to the genetic profile of the Nubian population. We aim to work on short tandem repeats (STRs) in order to generate haplotypes that are specific to the Nubian population which might be useful for human identification and tracing back through the HapMap database.

Acknowledgement:

I would like to thank Professor Hamza El Dorry for helping me to choose this point, his guidance and patience to end up with this study. Special thanks to Mr. Amged Auf for his help in the lab work and providing me with the required materials. I would also like to thank Mo Ibrahim Foundation for sponsoring my scholarship. Finally, I am grateful and I appreciate my family to endure me for a long time till finishing my thesis.

References:

Adams, W. Y. (1977). *Nubia: corridor to Africa*. London: Allen Lane. http://doi.org/AUC Library Catalog

Alford, R. L., Hammond, H. A., Coto, I., & Caskey, C. T. (1994). Rapid and efficient resolution of parentage by amplification of short tandem repeats. *American Journal of Human Genetics*, *55*(1), 190–5.

Alkhayat, A. Alshamali, F., & Budowle, B. (1996). Population data on the PCR-based loci LDLR, GYPA, HBGG, D7S8, Gc, HLA-DQA1, and DlS80 from Arabs from Dubai. *Forensic Science International 81*, 29–34.

Ammar, N. (1996). Nubians. Retrieved May 2, 2016, from http://www.encyclopedia.com/topic/Nubia.aspx

Balamurugan, K., Abdel-Rehman, H., Duncan, G. T., Budowle, B., Anderson, S., Macechko, J., & Tahir, M. (1998). Distribution of D1S80 alleles in the Jordanian population. *International Journal of Legal Medicine*, *111*(5), 276–277.

Balamurugan, K., Prabakaran, N., Duncan, G., Budowle, B., Tahir, M., & Tracey, M. (2001). Allele frequencies of 13 STR loci and the D1S80 locus in a Tamil population from Madras, India.

Balamurugan, K., Tracey, M. L., Heine, U., Maha, G. C., & Duncan, G. T. (2012). Mutation at the Human D1S80 Minisatellite Locus. *The Scientific World Journal*, *2012*(2), 1–8.

Batzer, M. A., & Deininger, P. L. (2002). Alu Repeats and Human Genomic Diversity. *Nature Reviews Genetics*, *3*(5), 370–379.

Bernal, L. P., & Villasmil, M. G. (2000). VNTR Locus D1S80 : Application to the Study of a Mixed Venezuelan Sample. *American Journal of Human Biology,622*(July 1999), 616–622.

Birajalaxmi Das, Anu Ghosh, P. S. Chauhan, and M. S. (2002). Genetic Polymorphism Study at Four Minisatellite Loci (D1S80, D17S5, D19S20, and APOB) among Five Indian Population Groups. *Human Biology*, *74*(3), 345–361.

Budowle, B., Chakraborty, R., Giusti, A. M., Eisenberg, A. J., & Allen, R. C. (1991). Analysis of the VNTR locus D1S80 by the PCR followed by high-resolution PAGE. *American Journal of Human Genetics*, *48*(1), 137–44.

Budowle, B., Baechtel, E. S., Smerick, J. B., Presley, Budowle, R. (1995). D1S80 Population Data in African Americans , Caucasians , Southeastern Hispanics , Southwestern Hispanics , and Orientals. *Journal of Forensic Science*, 38–44.

Budowle, B., Woller, J., Koons, B. W., Furedi, S., Errera, J. D., & Padar, Z. (1996). Hungarian population data on seven PCR-based loci. *J Forensic Sci*, *41*(4), 667–670.

Builes, J. J., Sc, M., Moreno, M. A., Gaviria, A. M., & Espinal, C. (2002). FOR THE RECORD Allele Frequency Data in Antioquia (Colombia) Population. *Journal of Forensic Science,47*(4), 1–2.

Butler, J. M. (2010). *Fundamentals of forensic DNA typing*. Amsterdam;Boston;

Buzon, M. R. (2006). Biological and Ethnic Identity in New Kingdom Nubia: A Case Study from Tombos. *Current Anthropology*, *47*(4), 683–695.

Çakir, A. H., Ph, D., Açik, L., Ph, D., Kesici, T., & Ph, D. (2001). FOR THE RECORD Distribution of Amplified Fragment Length Polymorphism D1S80 Alleles in Turkish Population. *Journal of Forensic Science46*(4), 2001.

Carlson, D. S., & van Gerven, D. P. (1979). Diffusion, Biological Determinism, and Biocultural Adaptation in the Nubian Corridor. *American Anthropologist*, *81*(3), 561–580.

Cavalli-Sforza, L. L., Menozzi, P., & Piazza, A. (1994). *The history and geography of human genes*. Princeton, N.J: Princeton University Press.

Chakraborty, R., Fornage, M., Gueguen, R., & Boerwinkle, E. (1991). Population genetics of hypervariable loci: analysis of PCR based VNTR polymorphism within a population. In *DNA fingerprinting: approaches and applications* (pp. 127–143).

Daniels, G. (2002). *Human blood groups*. *Human Blood Groups*.

Duncan, G. T., Balamurugan, K., Budowle, B., Smerick, J., & Tracey, M. L. (1997). Microvariation at the human D1S80 locus. *International Journal of Legal Medicine*, *110*(3), 150–154.

Emery, W. B. (1965). *Egypt in Nubia*.

Fahim, H. M. (1983). *Egyptian Nubians: resettlement and years of coping*. Salt Lake City: University of Utah Press.

Falcone, E., Spadafora, P., Luca, M. D. E., & Ruffolo, R. (1995). *Human Biology*, *67*(5), 689–701.

Fernea, E. W., Fernea, R. A., Rouchdy, A., & American University in Cairo, L. (1991). *Nubian ethnographies*. Prospect Heights, Ill: Waveland Press.

Fox, C. L. (1997). mtDNA analysis in ancient Nubians supports the existence of gene flow between sub-Sahara and North Africa in the Nile Valley. *Annals of Human Biology*, *24*(3), 217–227.

Gross, A., Carmody, G., & Guerrieri, R. (1997). Validation Studies for the Genetic Typing of the D1S80 Locus for Implementation into Forensic Casework. Journal of Forensic Science, 42(6), 1140-1146.

Guo, H., Jiang, J., Cui, Y., Yi, Y., Jia, X., Wang, H., … Li, J. (2015). Identification and characterization of polymorphic Alu insertions in the Tibetan macaque (Macaca thibetana). *European Journal of Wildlife Research*, *61*(1), 143–149.

Harashima, N., Liu, C., Katsuyama, Y., Ota, M., & Fukushima, H. (1997). Sequence variation of allele 27 at the D1S80 locus. *Int J Legal Med*, *110*(1), 22–26.

Herrera, R. J., Adrien, L. R., Ruiz, L. M., Sanabria, N. Y., Biology, S. H., February, N., … Ruiz, L. M. (2004). D1S80 Single-Locus Discrimination Among African Populations. Human Biology, 76(1), 87-108.

Holewinsky, M., Pott, U., (2002). Chapter 11, DNA Fingerprinting, 1–7.

Imad, H. H., Mohammed, A. J., Aamera, J. O., Cheah, Y., Haider, K. Z., Ali, H. A.-S., & Muataz, a. A. (2014). Genetic variation and DNA markers in forensic analysis. *African Journal of Biotechnology*, *13*(31), 3122–3136.

Ingrid Eisenbarth, Günter Assum, & Krone, W. (1999). Description of Three New Alleles at the D1S80 (MCT118) Locus. *Biosystems*, *80*(May 1998), 841–842.

Ismail M. Sebetan, H. A. H. and E. I. (1998). Frequency Distribution of D1S80 (MCT118) Locus Polymorphism in a Qatari Population. *Human Biology*, *70*(1), 129–135.

Jeanne M. Hayes, Bruce Budowle, A. M. F. (1995). Arab Population Data on the PCR-Based Loci LDLR , GYPA, HBGG, D7S8, Gc, and D1S80. *Journal of Forensic Science,* 645–648.

Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405.

Jorde, L., & Wooding, S. (2004). Genetic variation, classification and'race'. *Nature Genetics*, *36*(11), 28–33.

Kasai, K., Ph, D., Nakamura, Y., Ph, D., White, R., Ph, D., & Kasai, R. (1990). Amplification of a Variable Number of Tandem Repeats (VNTR) Locus (pMCT118) by the Polymerase Chain Reaction (PCR) and Its Application to Forensic Science. *Journal of Forensic Science*, 1196–1200.

Kirwan, A. L. P., & Kirwan, L. P. (2015). Nubia and Nubian Origins Published by : The Royal Geographical Society (with the Institute of British Geographers)

Kloosterman, A. D., Budowle, B., & Daselaar, P. (1993). PCR-amplification and detection of the human D1S80 VNTR locus. *International Journal of Legal Medicine*, 257–264.

Krings, M., Salem, a E., Bauer, K., Geisert, H., Malek, a K., Chaix, L., … Stoneking, M. (1999). mtDNA analysis of Nile River Valley populations: A genetic corridor or a barrier to migration? *American Journal of Human Genetics*, *64*(4).

Kuntze, K., Stancu, V., & Bonte, W. (1996). VNTR Locus D1S80: Allele frequencies and genotype distributionin the region of Düsseldorf. *Anthropologischer Anzeiger,1*(März 1996), 7–17.

Lucotte, G., & Mercier, G. (2003). Brief communication: Y-chromosome haplotypes

in Egypt. *American Journal of Physical Anthropology*, *121*(1), 63–66.

Mahdieh, N., Tafsiri, E., Karimipour, M., & Akbari, M. T. (2005). Heterozygosity and allele frequencies of the two VNTRs (ApoB and D1S80) in Iranian popuiation. *Indian Journal of Human Genetics11*(1), 31-34.

Maria, I., Cayres, V., Carlos, A., Vallinoto, R., Maria, C., Valente, D., & Guerreiro, J. F. (2003). Allele frequency distributions of six hypervariable loci (D1S80 , APOB , D4S43 , vW1 , F13A and DYS19) in two African-Brazilian communities from the Amazon region. *Genetics and Molecular Biology*, 26(3) 235–240.

Medintz, I., Kingston, C., Levine, L., Fogarty, P., Mar, E., McCurdy, L., & Kobilinsky, L. (1998). D1S80 allele frequencies in Hasidic and non-Hasidic New York City Jewish populations. *International Journal of Legal Medicine*, *111*(5), 273–275.

Miranda, J. Y. C. J., & Benecke, S. H. M. (1998). Allele frequency distributions of the polymorphic STR loci HUMVWA , HUMFES , HUMF13A01 and the VNTR D1S80 in a Filipino population from Metro Manila. International Journal of Legal Medicine, 224–226.

Morris F. Goodman. (n.d.). Nilo-Saharan languages. Retrieved from http://www.britannica.com/topic/Nilo-Saharan-languages

Muhammad A. Taher, Sudhir K. Sinha, Nahedh A-AL-Kubaidan, Usman A. Tahir, Bruce Budowle, and M. A. (2000). For the Record.Distribution of Amplified Fragment Length Polymorphism D1S80 Alleles in a Saudi Arabian Population*Journal of Forensic Science*, *45*(5), 46236.

Mukherjee, M., Srivastava, A., Kesari, A., & Mittal, B. (2005). Analysis of VNTR loci , ApoB 3 ′ HVR and D1S80 in North Indians. *Indian Journal of Biotechnology*, *(4*), 358–362.

Näslund, K., Saetre, P., Von Salomé, J., Bergström, T. F., Jareborg, N., & Jazin, E. (2005). Genome-wide prediction of human VNTRs. *Genomics*, *85*(1), 24–35.

Oksanen, A. J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., …

Szoecs, E. (2016). Package “ vegan ”(Version 2.4-0).

Paradis, E. (2010). Pegas: An R package for population genetics with an integrated-modular approach. *Bioinformatics*, *26*(3), 419–420.

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, *20*(2), 289–290.

Peterson, B. L., Su, B., Ph, D., Chakraborty, R., Ph, D., Budowle, B., … Budowle, B. (2000). World Population Data for the HLA-DQA1 , PM ® and D1S80 Loci with Least and Most Common Profile Frequencies for Combinations of Loci Estimated Following NRC II Guidelines *. *Journal of Forensic Science*,45(1),118–146.

R.C. Lewontin. (2006). Confusions About Human Races. Retrieved from http://raceandgenomics.ssrc.org/Lewontin/

Raczek, E. (2001). Population data on the D1S80 locus in the Upper Silesia (Poland). *Journal of Forensic Sciences*, *46*(6), 1521.

Rajesh Biswas, R. G. (2002). Allele Frequencies of D1S80 Locus in the Bharias , a Primitive Indian Tribal Population from Patalkot , India. *J Forensic Sci*, *47*(6).

Roslan, H. A., Azizan, N. H., & Saat, R. (2009). DNA Polymorphism of D1S80 locus in modern Malay sample population of Sarawak. *Sains Malaysiana*, *38*(2), 143–147.

Sachdeva, M. P., Mastana, S. S., Saraswathy, K. N., Elizabeth, A. M., Chaudhary, R., & Kalla, A. K. (2004). Genetic variation at three VNTR loci (D1S80, APOB and D17S5) in two tribal populations of Andhra Pradesh, India. *Annals of Human Biology*, *31*(1), 95–102.

Shastry, B. S. (2002). SNP alleles in human disease and evolution. *Journal of Human Genetics*, *47*(11), 561–566.

Shinnie, P. L. (1996). *Ancient Nubia*. New York;London; Kegan Paul International.

Silva, J. O. D. A., Ca, M., Elion, J., Krishnamoorthy, R., Paula, M., & Schneider, C. (1999). Genetic Diversity of Two African and Sixteen South American

Populations Determined on the Basis of Six Hypervariable Loci. American Journal of Physical Anthropology, 425–437.

Sokarno, A. (2007). The attitudes of Egyptian Nubians towards their vernaculars and Arabic. *Occasional Papers in the Development of English Language Education*, *43*, 101–126.

Stoneking, M. (2001). From the evolutionary past. *Nature*, *409*(February), 821–822. h

Tahir, M. A., Ph, D., Rogers, C., Alkhayyat, M., Ph, D., El-gohary, M., … Alkhayyat, M. (1999). Distribution of D1S80 Alleles in the Bahrainian Population. *International Journal of Legal Medicine*, 1314–1315.

Tamaki, K., & Jeffreys, A. J. (2005). Human tandem repeat sequences in forensic DNA typing. *Legal Medicine*, *7*(4), 244–250.

Török, L. (2009). *Between two worlds: the frontier region between ancient Nubia and Egypt, 3700 BC-AD 500* (Vol. 29. Bd.). Boston;Leiden; Brill.

Trivedi, R., Chattopadhyay, P., Dutta, R., Sarkar, N., & Sharma, A. K. (2002). FOR THE RECORD D1S80 Population Data in Eight Predominant Populations of India * References : *J Forensic Sci*, 676–677.

Vanaja, M., Panneerchelvam, S., Sankaranarayanan, K. R., Baskar, D., Rajmohan, S. Nalina, K. & Damodaran, C. (2001). For the Record,Allele Frequency for D1S80 Locus in Tamil Population(South India). *Journal of Forensic Science*,*46*(2), 177.

Verbenko, D. A., Kekeeva, T. V, Pogoda, T. V, Khusnutdinova, E. K., Mikulich, A. I., Kravchenko, S. A., … Limborska, S. A. (2003). Allele frequencies for D1S80 (pMCT118) locus in some East European populations. *J Forensic Sci*, *48*(1), 207–208.

Verbenko, D. a., Pocheshkhova, E. a., Marshanija, E. Z., Kvitzinija, P. K., & Limborska, S. a. (2004). Polymorphisms of D1S80 and 3′ApoB Minisatellite Loci in Northern Caucasus Populations. *Journal of Forensic Sciences*, *49*(1), 1–3.

Wan, Q. H., Wu, H., Fujihara, T., & Fang, S. G. (2004). Which genetic marker for which conservation genetics issue? *Electrophoresis*, *25*(14), 2165–2176.

Wang, J. (2005). Estimation of effective population sizes from data on genetic markers. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *360*(1459), 1395–1409.

Watanabe, G., & Shimizu, K. (2002). DNA sequence analysis of long PCR amplified products at the D1S80 locus. *Legal Medicine*, *4*(1), 37–39.

Wolfarth, R., Nhari, L. T., Budowle, B., Kanoyangwa, S. B., & Masuka, E. (2000). Polymarker, HLA-DQA1 and D1S80 allele data in a Zimbabwean Black sample population. *International Journal of Legal Medicine*, *113*(5), 300–301.