# ANALYSIS OF MITOCHONDRIAL DNA VARIATION IN THE EGYPTIAN POPULATION AND ITS IMPLICATIONS FOR FORENSIC DNA ANALYSIS

**RANIA GOMAA**

**A thesis submitted in partial fulfilment of the requirements of Nottingham Trent University for the degree of Doctor of Philosophy**

***July 2010***

# ABSTRACT

The genetic sequence of human mitochondrial DNA (mtDNA) is of particular interest to forensic investigations involving human identification, as well as population genetics. The current mtDNA database is lacking sufficient representatives of African mitochondrial DNA sequences compared to European and Asian sequences. The present study was concerned with the analysis of mtDNA in the Egyptian population. FTA cards were used for blood sample collection, storage, and shipment and DNA extraction. An optimised laboratory protocol for rapid PCR amplification of the mitochondrial hypervariable regions was developed. A database of 261 mitochondrial hypervariable region I (HVI) sequences and 78 hypervariable region II (HVII) sequences was established from 261 adult Egyptians. A total of 113 polymorphic sites were reported in the HVI region (nt16024-16365) which identified a total of 187 different haplotypes, of which 151 were unique to single individuals. The most commonly observed HVI haplotype was identical to the Cambridge Reference Sequence (CRS). Analysis of 78 HVII sequences (nt73-340) revealed a total of 42 polymorphic sites that identified 62 different haplotypes, of which 51 were unique to single individuals. Sites that showed the highest variability in the HVI and HVII regions agreed with the previously reported mutational hotspots. Combination of the HVI and HVII data resulted in identification of 207 different mitochondrial haplotypes, of which 183 (~88%) were unique to single individuals. Such a large number of unique mitochondrial haplotypes indicates a high diversity of mtDNA in the Egyptian population, which has a direct impact on forensic applications, since the significance of a match between an evidence sample and a reference sample depends on the population frequency of a profile. The random match probabilities in the HVI and HVII datasets were 2.34% and 3.22%, respectively. Combination of the two datasets reduced the random match probability to 1.28%. The genetic diversities in the HVI and HVII sequences were estimated to be 0.9804 and 0.9803, respectively, whereas, the genetic diversity in the combined dataset was 0.9911. A new strategy was developed to facilitate Restriction Fragment Length Polymorphism (RFLP) analysis of the whole mitochondrial genome for the purpose of haplogroup assignment. All the Egyptian individuals were assigned to well known mitochondrial haplogroups and the frequency distribution of different lineages was estimated. A mixed maternal ancestry of the Egyptian population was reported via detection of a mixture of mitochondrial DNA lineages with varying racial backgrounds. The most frequently reported mitochondrial

lineages were haplogroup T (13.8%) followed by haplogroups L3 (12.6%), H (12.3%), U (~10%) and M (8.4%). The genetic relationship between Egypt and its neighbours was revealed, and it was found that present day Egyptians were the closest African population to the Middle East and Europe, with an overall 62.5% European, 25% African, and 12.5% Asian mitochondrial lineages. The data presented here will enrich the limited Egyptian mitochondrial DNA reference data currently available for forensic applications.

*To the memory of my late Mother-in-law, Mrs Fawkiya*

*Who was looking after my family while I was doing my PhD.*

*Without her kind help, this thesis could never have been completed.*

# ACKNOWLEGEMENTS

I would like to thank my director of study Dr David C Hughes for guiding me into this project and giving me the opportunity to benefit from his valuable expertise in Molecular Biology and Bioinformatics. He was always supportive and sympathetic. I would also like to acknowledge my second supervisor Prof. Ellen Billet for her interest and support.

I wish to thank my past supervisor Prof. Mark Darlison who is really missed by leaving NTU. I would also like to acknowledge Mr David Butler for his continual support and encouragement, for informing me about Forensic Science Events at NTU and elsewhere and for nominating me to represent NTU at the International Networking Event in Forensic Science, in Thailand.

I would like to thank Dr Christian Thode for being my internal assessor for three years and for providing me with valuable comments and advice. I would also like to thank all my friends at NTU, past and present, especially Dr Sara Croft, Amy Pole, Tina Schwabe, Noha Juma, Fathiya Asteal, Vinaya Phatak, and Muralitharan Ponniah. They have always been supportive and helpful. I have greatly enjoyed my time at NTU.

I wish to thank the Egyptian government for the generous fund for this research as a part of a PhD scholarship. I would also like to thank my Professors at Alexandria University who nominated me for this scholarship.

I would like to thank my family, especially my Mum and my Dad for giving me all kinds of support and encouragement. I would definitely like to acknowledge my son Seif for being a sensible young man. I wish to thank my lovely daughter Hana for being good.

Last but certainly not least, my biggest heartfelt gratitude must go to my husband Mohamed. I could not have completed this project without his indefinite support, love, sympathy and encouragement.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# ABBREVIATIONS

| | |
|---|---|
| A | adenine |
| AFDIL | Armed Forces DNA identification Laboratory |
| AIDS | acquired immune deficiency syndrome |
| AMELX | amelogenin Y gene |
| AMELY | amelogenin X gene |
| APS | ammonium per sulphate |
| ATP | adenine triphosphate |
| BH | Bosnia and Herzegovina |
| bp | base pairs |
| BSA | bovine serum albumin |
| C | cytosine |
| CR | control region |
| CRS | Cambridge Reference Sequence |
| $ddH_2O$ | double distilled water |
| deoxyNTP | deoxynucleotide triphosphate |
| D-loop | displacement loop |
| DEPC | Diethylpyrocarbonate |
| DNA | deoxyribonucleic acid |
| EDTA | ethylenediaminetetraacetic acid |
| FTA | Flinders Technology Associates |
| G | guanine |
| H-strand | heavy strand |
| HVI | hypervariable region I |
| HVII | hypervariable region II |
| L-strand | light strand |
| LOHN | Leber Hereditary Optic Neuropathy |
| mtDNA | mitochondrial DNA |
| $Mg^{2+}$ | magnesium |
| $MgCl_2$ | magnesium chloride |
| MITOMAP | a human mitochondrial genome database |
| MRCA | most recent common ancestor |
| MS | multiple sclerosis |
| $Na_2CO_3$ | sodium carbonate |
| NAHR | non-allelic homologous recombination |
| NCBI | National Centre for Biotechnology Information |
| PCR | polymerase chain reaction |
| rRNA | ribosomal ribonucleic acid |
| RFLP | restriction fragment length polymorphism |
| RMP | random match probability |
| RNA | ribonucleic acid |
| ROS | Reactive Oxygen Species |
| SDS | sodium dodecyl sulfate |
| SNPs | single nucleotide polymorphisms |
| STRs | short tandem repeats |
| STS | sequence-tagged site mapping |
| SWGDAM | Scientific Working Group on DNA Analysis Methods |
| tRNA | transfer ribonucleic acid |
| T | Thymine |
| TAE | tris-acetate-EDTA buffer |

| | |
|---|---|
| TB | tuberculosis |
| TBS | tris-buffered saline |
| TSPY | testis-specific protein Y encoded gene |
| UV | ultra violet |
| v/v | volume per volume |
| VNTRs | variable number tandem repeats |
| w/v | weight per volume |
| WGA | whole genome amplification |
| Y-STRs | short tandem repeats on Y- chromosome |

# 1. GENERAL INTRODUCTION

## Biological Issues

## 1.1 DNA and human identification

The use of DNA in human identification and forensic investigation has increased dramatically since DNA fingerprinting was first demonstrated by Sir Alec Jeffreys in 1984. The many uses of DNA based identification include paternity testing. The human genome incorporates 3 billion base pairs in its structure that encode only for 20,000-25,000 genes corresponding to roughly 5% of the genome. Thus, the majority of the human genome is non-coding. These regions are of particular interest to DNA identity testing as they are more liable to mutate and change without causing any physical or physiological consequences. Therefore, they provide useful markers for human identification.

These markers have contributed enormously to the fields of medical and forensic science including the forensic investigative work and identity testing. The human genome is full of repeated DNA sequences and these repeated loci are classified according to the length of the core repeat unit and the number of repeats into: Variable Number Tandem Repeats (VNTRs) which are also known as minisatellites, and Short Tandem Repeats (STRs) or microsatellites. Unlike VNTRs, which number a few thousand in our genome, STRs are present at > 100 000 regions that cover most of our genome. The lengths of the core unit repeat range from 2-6 bp in microsatellites to 10-100 bp in minisatellites (Habano et al. 1999; Butler 2005a).

STRs are more convenient to forensic applications compared to VNTRs due to the fact that they are scattered all through the human genome with allele size ranging between 100-400 bp compared to 400-1000 bp of VNTRs. Hence, this relatively small size of STRs is more convenient to the forensic applications where degraded DNA is common. Moreover, the number of the repeat units of STR markers can be highly variable among individuals and thus makes them more effective for human identification purposes (Edwards et al. 1991).

Currently, there are many commercial kits available for multiplex PCR amplification of STR loci that are routinely and extensively used in DNA profiling. However, all these kits are of limited value when dealing with highly degraded forensic samples. That in turn explains the increasing interest in Single Nucleotide Polymorphism (SNP) typing in the forensic arena (Sobrino et al. 2005; Dixon et al. 2006).

Several potential applications of SNPs exist for human identity testing that could further aid criminal investigations. These are mainly targeting three areas;

- Estimation of the ethnic origin of samples retrieved at the crime scene

Frudakis et al. have screened 211 SNPs in the human pigmentation and xenobiotic metabolism genes and they have found remarkable differences in the frequencies of 56 SNPs (mostly in pigmentation genes) between groups of unrelated individuals descending from different ancestral background (Frudakis et al. 2003a).

- Prediction of the physical traits of the perpetrators

In the last 10 years, the attention of researchers has been drawn to the possibility of anticipation of the human phenotypic characteristics through identification of certain genetic variants. For example, an SNP typing assay has been developed by the Forensic Science Service, that involved a set of mutations in the human melanocortin 1 receptor gene which is associated with the red hair phenotype (Grimes et al. 2001; Frudakis et al. 2003b).

- Recovery of much more information from degraded DNA samples

Being easily recoverable from degraded samples compared to STR markers, SNPs have been very successful in victims' identification in many circumstances particularly in mass disasters in which body parts have been highly degraded. As experienced in Tsunami earthquake in Thailand, 2004, where DNA samples were badly damaged and it was not possible to obtain fully informative STR profiling, meanwhile the use of mitochondrial SNPs gave the opportunity to carry out further analysis that helped in the identification of the human remains (Deng et al. 2005). Furthermore, identification of victims of the World Trade Centre attacks in New York, 2001, was more successful after supplementation of STR profiling results with mitochondrial SNPs data (Holland et al. 2003).

## 1.2 Mitochondrial genome

### 1.2.1 Mitochondria

Mitochondria are double-membraned subcellular organelles that are present in high copy number (several hundreds to several thousands) in the mammalian cell. Mitochondria are the powerhouse of the cell in which many crucial metabolic processes occur such as oxidative phosphorylation (Holland and Parsons 1999).

They have their own unique genome (mtDNA), which is distinct from the nuclear DNA. Each mitochondrion has more than 2 copies of the mtDNA; thus there are on average thousands of copies of mtDNA in each somatic cell compared to the two copies of the nuclear genome (Kobilinsky et al. 2004).

### 1.2.2 Structure and gene content of the mitochondrial DNA

The mitochondrial genome is a double-stranded, antiparallel, double helical circular DNA molecule, approximately 16,569 bp in length encoding 37 densely packed genes. Twenty-two of these genes encode tRNAs, two encode 12S and 16S rRNAs, and 13 encode enzymes involved in the processes of oxidative phosphorylation and ATP production (Holland and Parsons 1999).

The two strands of the mtDNA are the light (L-strand) and the heavy (H-strand) strands due to a selective distribution of the nucleotides with more guanine and thymine (the highest molecular weights) on the latter. There is asymmetric distribution of genes within the mitochondrial genome since the heavy strand carries most of the genetic information. Twelve of the polypeptides, the two ribosomal RNAs, in addition to 14 tRNAs are encoded from the heavy strand while L strand transcribes the remaining eight tRNAs and only one enzyme corresponding to ND6 subunit of complex I (NADH-dehydrogenase-ubiquinone reductase) (**Figure 1.1)** (Fernandez-Silva et al. 2003).

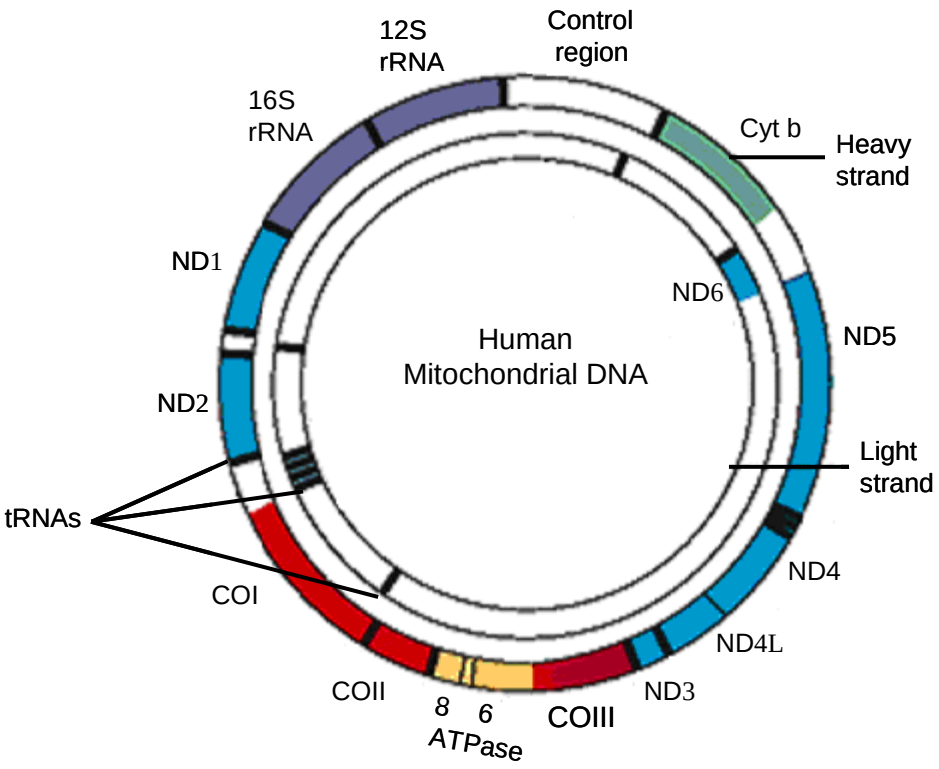



**Figure 1.1: A map of the human mitochondrial genome.** The subunits of NADH dehydrogenase (*ND1–ND6* and *ND4L*) are shown in blue; subunits of cytochrome *c* oxidase (*COI–COIII*) are shown in red; cytochrome *b* is shown in green; and the subunits of the ATP synthase (*ATPase 6* and *8*) are shown in yellow. The two ribosomal RNAs (rRNAs; 12*S* and 16*S*) are shown in purple. The black lines represent the 22 tRNAs. Reproduced from Taylor and Turnbull (2005).

There is a specific region consisting approximately of 1122 bp which does not code for any gene products and hence named 'non-coding' region. Whereas, all the remaining part of the DNA molecule is known as the coding region in which the genes are tightly arranged with no introns in between. The non-coding region, despite not transcribing any gene products, it contains the origin of replication of the heavy strand as well as promoters for the transcription initiation and therefore, it is mostly referred to as the control region or CR (Shadel and Clayton 1997; Fish et al. 2004).

The CR is also known as the displacement loop or D-loop based on the mode of replication. In the displacement model, replication of mtDNA starts within the CR at the origin of replication of the heavy strand (**Figure 1.2)** (Brown et al. 2005).

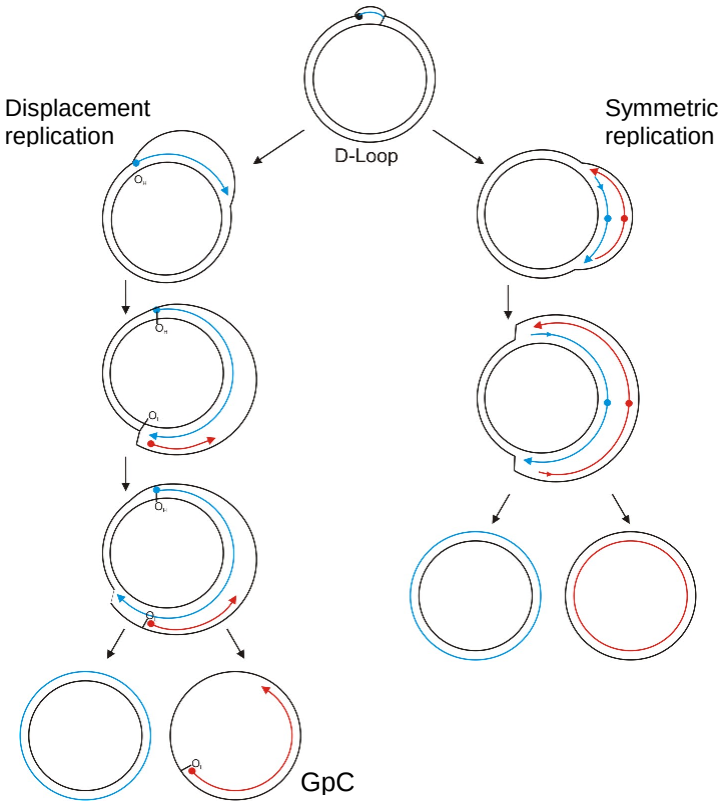



**Figure 1.2: Displacement (asymmetric) replication versus strand-coupled (symmetric) model of mtDNA replication.** The model of asymmetric replication (shown on the left) illustrates how the circular genome starts replication of one strand (H-strand) with expansion and displacement of the D-loop. This expansion proceeds till the origin of replication of the other strand (L-strand) gets exposed with subsequent replication. This asymmetry of strand synthesis results in the production of one segregated daughter molecule with an incompletely synthesized L-strand, hence called a gapped circle (GpC). On the contrary, the strand-coupled replication model (shown on the right) illustrates a synchronous replication of the double strands bidirectionally leading to completely synthesized double stranded molecules (Brown et al. 2005).

### 1.2.3 Cambridge Reference Sequence (CRS)

The primary sequence of the mitochondrial genome was first determined in 1981 and was known as the Anderson Reference Sequence (Anderson et al. 1981). This sequence was primarily derived from an individual of European descent. However, in 1999 following the improvement in DNA sequencing technology, several minor errors have been identified and corrected to derive the Cambridge Reference Sequence (CRS) (Andrews et al. 1999).

The base numbering system of the Cambridge Reference Sequence (CRS) starts in the middle of the control region (CR). The CR starts at the nucleotide position 16024 to 16569 and then continues from position 1 to 576 (**Figure 1.3)**. Despite of the corrections that have been done for the minor errors in the Anderson Reference

Sequence, the historical nucleotide numbering system was maintained to avoid any possible confusion.

This control region involves two distinct segments that exhibit a very high level of variability among individuals and so named hypervariable regions. The hypervariable region one (HVI) extends from base position 16024 to 16365, whereas the second hypervariable region (HVII) extends from position 73 to 340 (**Figure 1.3)** (Butler 2005b).

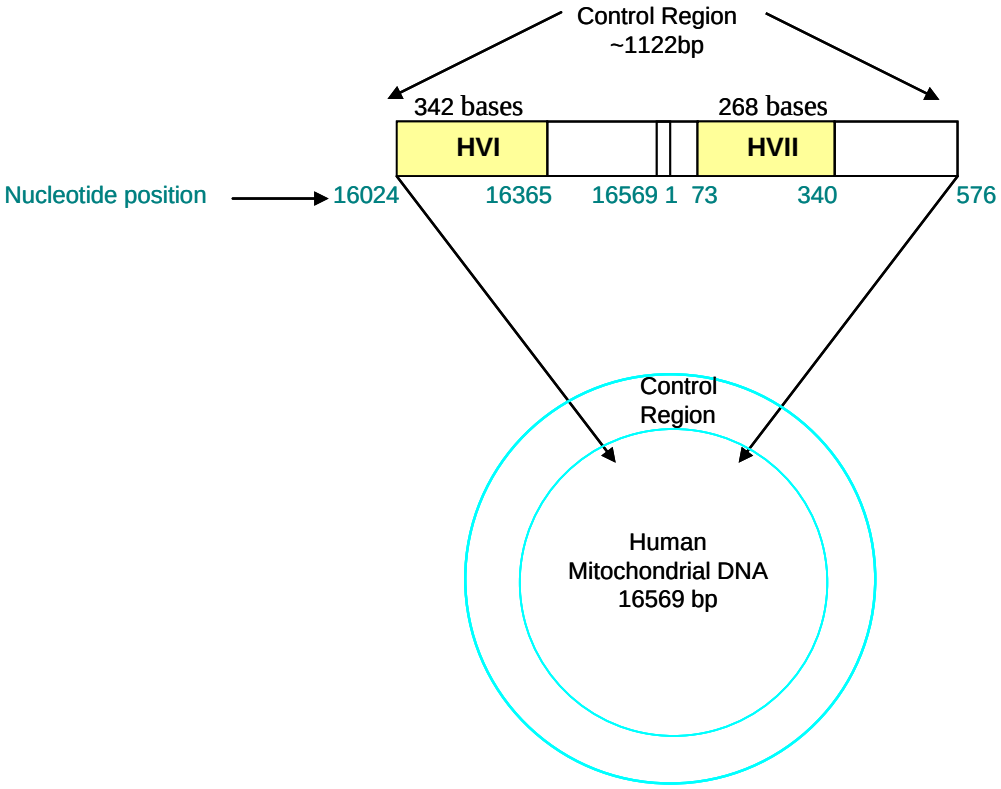



**Figure 1.3: A schematic representation of the mitochondrial human genome.** The control region (CR) is opened to show the extension of the hypervariable regions (HVI and HVII) with lengths of 342 and 268 bases respectively.

### 1.2.4 Mitochondrial inheritance and recombination

The most widely accepted model for mitochondrial inheritance is that it is transmitted strictly through the maternal line. Conception occurs by introducing the sperm's nucleus into the ovum to constitute the nuclear genetic material of the zygote. Therefore, the cytoplasm and all other cellular components including the mitochondria will be derived from the maternal side solely (Giles et al. 1980; Gyllensten et al. 1985; Primorac and Schanfield 2000).

Despite the widespread agreement to the simple clonal mode of mitochondrial inheritance, many studies have been conducted over the last 25 years that support the concept of paternal participation as well as recombination probabilities.

#### *1.2.4.1 Paternal inheritance of mtDNA*

A current debate exists about a different pattern of mitochondrial inheritance which involves paternal transmission through a phenomenon termed "paternal leakage" that is thought to occur during fertilization. Although, until recently, paternal leakage was thought to be impossible, there are some studies that support this phenomenon (Ankel-Simons and Cummins 1996; Kvist et al. 2003). Through analysis of mitochondrial DNA in sheep, Zhao et al. (2004) have provided direct evidence of paternal inheritance. They suggested presence of both parents' mtDNA at an early stage of development after which only one type of mtDNA survived to later development. The most striking evidence of paternal leakage in human was reported by Schwartz and Vissing (2002), while investigating a case with mitochondrial myopathy caused by a 2-bp mitochondrial deletion. During this investigation which involved analysis of the mitochondrial DNA of the patient's parents and his sister, a paternal mitochondrial haplotype was detected in the patient's quadriceps muscle in a ratio of 10: 1 paternal to maternal type. Nonetheless, this was the only case that showed evidence of paternal inheritance of the mtDNA in human since 2002. In addition, the paternal mtDNA was only found in a specific muscle tissue and not in blood which is commonly used for mitochondrial DNA analysis. A finding which might explain the occurrence of paternal mtDNA transmission in that case on the basis of certain selection forces, and thus the phenomenon of paternal leakage can not be generally accepted as a crucial route of mitochondrial inheritance. Therefore, it was not expected for this phenomenon to have an actual impact on the current study which involved analysis of mtDNA in human blood samples.

In contrast, there are many other researchers who disagree with the phenomenon of paternal leakage and have documented various evidences that support their opinions. First, it was found that even if paternal leakage occurred, the parental mitochondria would be diluted by the enormous number of the maternal ones especially after passing through the mitochondrial genetic bottleneck (Chen et al. 1995a; Wolff and Gemmell 2008). Second, there is a proposed mechanism of ubiquitination of sperm mitochondria outer membrane which makes them recognizable, therefore, they will be targeted and

destroyed selectively through a process of proteolytic digestion by the newly formed embryo cellular machinery (Thompson et al. 2003; Sutovsky et al. 2004; Butler 2005b). Moreover, the sequestration of mtDNA molecules within the inner mitochondrial membrane hinders against heterologous recombination (Elson et al. 2001).

#### *1.2.4.2 Recombination characteristics of mtDNA*

The other feature of the mitochondrial inheritance that is still under discussion is the mitochondrial recombination. There is growing evidence which supports the likelihood of mtDNA recombination at both intermolecular as well as intramolecular levels.

MtDNA recombination has been well documented in yeast and plants (Gray 1989) as well as in invertebrates (Ladoukakis and Zouros 2001). There are studies that have demonstrated evidences of mtDNA recombination in fish (Ciborowski et al. 2007) and in lizard (Ujvari et al. 2007). Extracts of mitochondria from human cells were shown to harbour enzymes that catalyze the homologous recombination of DNA plasmids (Thyagarajan et al. 1996). Different recombinant molecules were reported within human cytoplasmic hybrid clones suggesting that recombination events involve different mtDNA regions (D'Aurelio et al. 2004).

Despite the evidence that supports the possibility of mtDNA recombination, there are still barriers to recombination to occur. First, there was an early evidence against the free exchange of mtDNA molecules within a cell (Enriquez et al. 2000; Jacobs et al. 2000). However, more recently, mitochondrial fusion and fission have been identified and analysed in model organisms, such as yeasts, worms and flies (Hoppins et al. 2007; Westermann 2008, 2009). Second, it is not very well known how long the paternal mitochondria will need to survive to enable fusion/recombination (Howell 1997). Lastly, the limited number of the recombinant DNA molecules (if recombination does occur) within the zygote, will then have to undergo clonal expansion to allow passage through the mitochondrial genetic bottleneck and transmission to subsequent generations (Elson et al. 2001), *see also heteroplasmy*.

Following all these discussions, the answer to the question whether mitochondrial DNA recombines sufficiently to influence human evolution and disease etiology is still equivocal. Overall, the currently available population genetic data suggests that recombination is not occurring at detectable levels (White et al. 2008).

### 1.2.5 Heteroplasmy within the mitochondrial genome

The general rule for an individual is to have a single mtDNA genotype (homoplasmy). However, occasionally, an individual may carry two sequences that differ in one or two nucleotide positions. This is referred to as heteroplasmy.

Heteroplasmy arises because of the high mutation rate of mitochondrial DNA compared to the nuclear genome (Ingman 2001; Kivisild et al. 2006) which is attributed to the mutagenic properties of Reactive Oxygen Species (ROS) (the harmful by-products of oxidative phosphorylation), an error-prone polymerase and deficient DNA repair mechanisms (Clayton 1991; Bogenhagen 1999). This high mutation rate is more prone to produce a state of heteroplasmy when it is coupled with deficiency in the protective mechanisms that are known to effectively clear any genetic diversity from the mitochondrial DNA population, e.g. mitochondrial germline bottleneck, gene transfer, selection against deleterious mutations and vegetative segregation (Martin and Herrmann 1998; Rispe and Moran 2000; Rand 2001). Moreover, heteroplasmy may be a result of the recently presumed paternal leakage phenomenon as revealed in the reported human case with mitochondrial myopathy in which two different mitochondrial haplotypes were identified in the patient's muscle with a ratio of 10:1 paternal to maternal, whereas only the maternal mtDNA haplotype was detected in the patient's blood (Schwartz and Vissing 2002).

Heteroplasmy is now accepted as a widespread phenomenon but still much less common than homoplasmy. Although mitochondrial heteroplasmy can complicate the interpretation of phylogenetic networks of population data, heteroplasmy may provide a useful tool for defining haplotypes (if heteroplasmy remains stable) (White et al. 2008). Heteroplasmy may further aid in resolving the relationships among individuals at the population level as was experienced in the identification of the remains of Russia's last Tsar, Nicholas II. By analysis of mtDNA from the presumed skeleton of the Tsar, heteroplasmy was detected at the nucleotide position 16169 (C/T – 1:1.6) and his identification was confirmed by analysis of his brother's mitochondrial DNA that was heteroplasmic at the same nucleotide position, thus providing a powerful evidence supporting the identification of Tsar Nicholas II (Gill et al. 1994; Ivanov et al. 1996). A more recent analysis has demonstrated rapid segregation of the two mitochondrial haplotypes (16169C/ 16169T) in two maternal lines derived from Tsar's mother (Rogaev et al. 2009).

Heteroplasmy can be one of two types;

- Site heteroplasmy (which is also known as sequence heteroplasmy or point heteroplasmy) occurs when two different nucleotides are detected at a single nucleotide position (**Figure 1.4)** (Carracedo et al. 1998; Carracedo et al. 2000; Butler 2001).

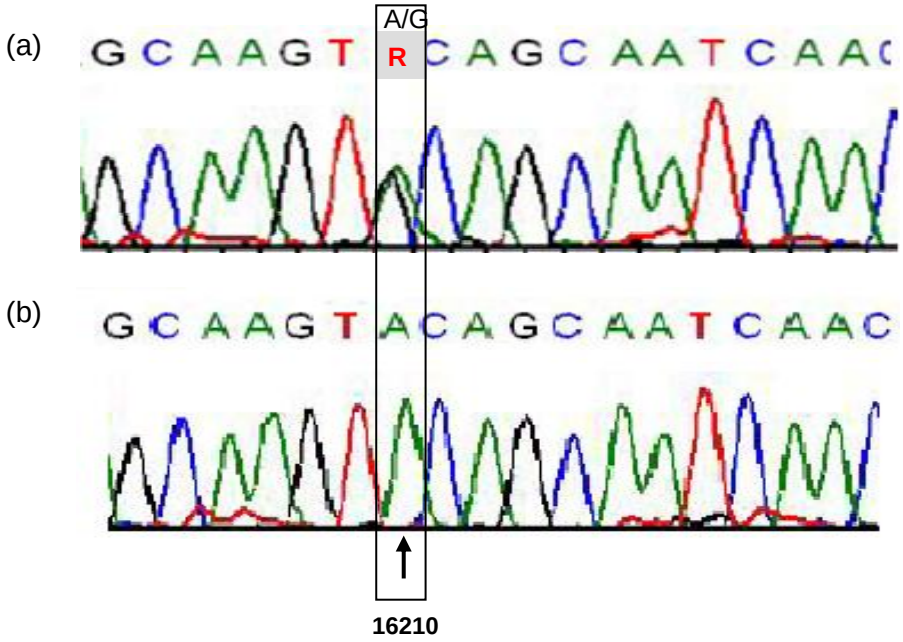



**Figure 1.4: Mitochondrial HVI sequence between positions 16204 and 16220.** The electropherogram at the top (a) shows sequence heteroplasmy at position 16210 possessing both G and A nucleotides with clearly defined peaks well above the background noise compared to (b) the same region on a different sample containing only an A peak at position 16210.

- Length heteroplasmy results from the slippage of the DNA polymerase during replication of a repetitive sequence resulting in either deleted or inserted bases. If this error is not repaired, the individual will posses two distinct populations of mitochondrial DNA with one genome slightly larger than the other. The poly C stretch in HVII found between nucleotide positions 303 to 315 is an example of this phenomenon (Kobilinsky et al. 2004). It was reported that length heteroplasmy is much more prevalent than point heteroplasmy (Carracedo et al. 2000).

Heteroplasmy can be detected in an individual in several forms (Stewart et al. 2001);

- Individuals may have more than one mtDNA haplotype in a single tissue.
- Individuals may exhibit one mtDNA haplotype in one tissue and other haplotype in another tissue.
- Individuals may be heteroplasmic in one tissue and homoplasmic in another tissue.

The rate of heteroplasmy differs between different tissues of the body, among which the brain was suggested to be the most common site of heteroplasmy (Lightowlers et al. 1997). Whereas, a higher frequency of heteroplasmy was observed in muscle compared to blood, heart and brain (Calloway et al. 2000).

Comparison of HVII length heteroplasmy between mothers and children identified variation in length heteroplasmic level between generations. Variability between individuals was greater in hair samples compared to blood, buccal cells and nails (Asari et al. 2007; Asari et al. 2008).

In contrast, HVI length heteroplasmy seems to be more stable in various tissues within individuals (Malik et al. 2002a; Lutz-Bonengel et al. 2004) and among different individuals within the same lineage (Cavelier et al. 2000; Malik et al. 2002b) despite showing wide variation between unrelated individuals (Lee et al. 2004).

Several studies have identified a higher frequency of heteroplasmy in hair shafts compared to other tissues (Tully et al. 2004; Paneto et al. 2007). Moreover, it has been shown that not only does the rate of heteroplasmy differ between different tissue types within an individual, but it also varies between extracts from different segments of a single hair (Salas et al. 2001; Tully et al. 2004). This is a consequence of the stochastic segregation of mitochondrial DNA types following the bottleneck in the female germ line (**Figure 1.5**) (Kirches et al. 2001).

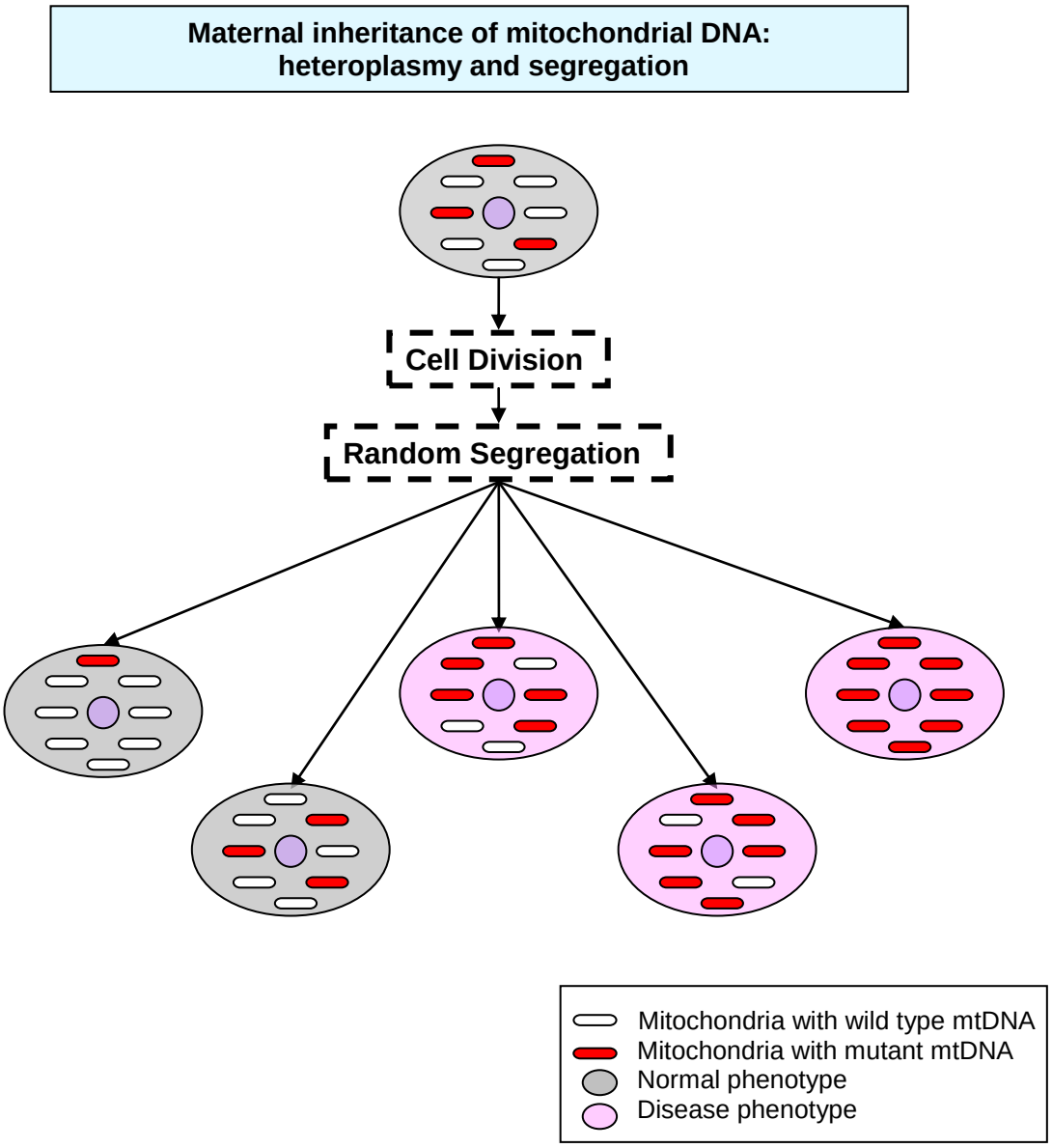



**Figure 1.5: Heteroplasmy and mitochondrial segregation.** During cell division, mitochondria are randomly distributed into daughter cells. Their mitochondrial genotype will then evolve into a normal or a mutant form depending on the percentages of wild type and mutant mtDNA involvement. Reproduced from (www1).

## 1.3 Modern human populations

Evidence of mitochondrial recombination and paternal leakage is not strong enough to play a fundamental role in mitochondrial inheritance. Therefore, mitochondrial inheritance down the maternal line should still be considered as the primary route of mitochondrial inheritance, and mtDNA is expected to pass unchanged from generation to generation along the maternal line apart from spontaneous mutations. Thus, all modern human populations could be tracked back along the maternal lineages to a common ancestor.

### 1.3.1 Population genetic variation

Differences between mtDNA sequences are attributed to mutations. which are known to pass on under certain circumstances: firstly, the mutation has to occur in a female; specifically in the germ line cells that give rise to oocytes, then the mature egg carrying the mutated mtDNA must be fertilized by an X-bearing sperm to give rise to a female embryo that will develop later into a female adult, who can then pass it on to her children (Ingman 2001).

### 1.3.2 Mitochondrial Eve

Mitochondrial Eve represents the woman whose mtDNA exists in all the humans living now on Earth and she is estimated to have lived 100,000 to 200,000 years ago. However, that does not mean that she was the only woman on earth at that time, but she was the only woman who contributed her mtDNA to modern human (**Figure 1.6**) (Cann et al. 1987).

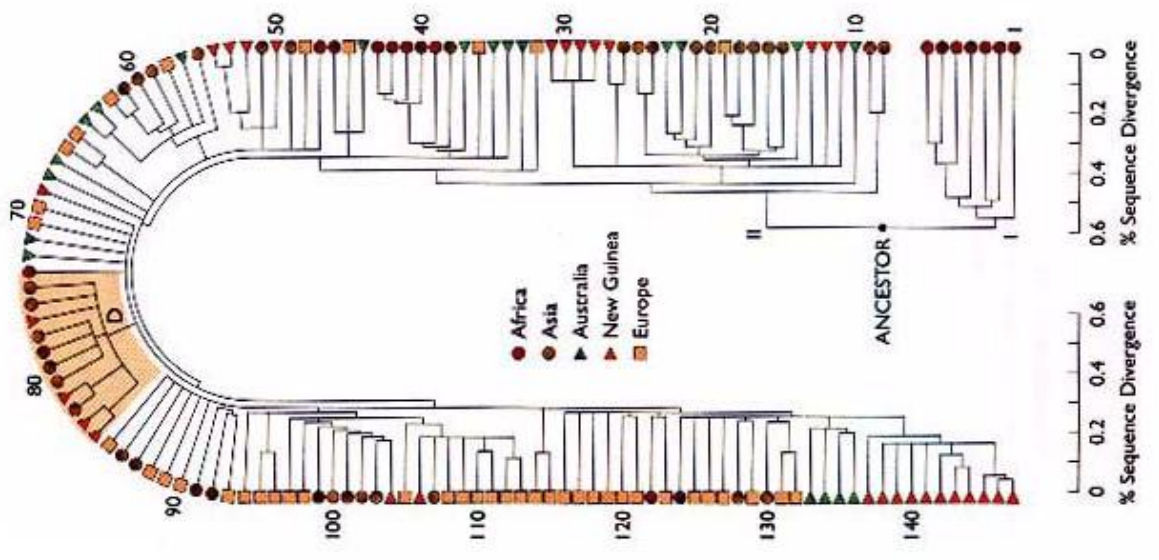



**Figure 1.6: Phylogenetic tree representing mitochondrial DNA sequence relationships among modern humans.** The evolutionary tree has its roots in Africa and its branches are short enough to hypothesize that all humans have a fairly recent common ancestor " mitochondrial eve" (Cann et al. 1987).

### 1.3.3 Pattern of modern human migration

According to Cann et al. (1987), humans who lived on earth before the emergence of mitochondrial eve were completely replaced by her descendants by migrating out of Africa somewhere between 100,000 and 200,000 years ago.

Several phylogenetic studies have been conducted on mitochondrial DNA sequences that support the "Out-of-Africa" scenario for the evolution of modern human populations. These studies have shown that the deepest-rooting branches of the mitochondrial DNA tree are located in sub-Saharan Africa. Moreover, they have demonstrated that the highest mtDNA diversity in the world is found within Africa (**Figure 1.7)** (Torroni et al. 2006; Salas et al. 2007).

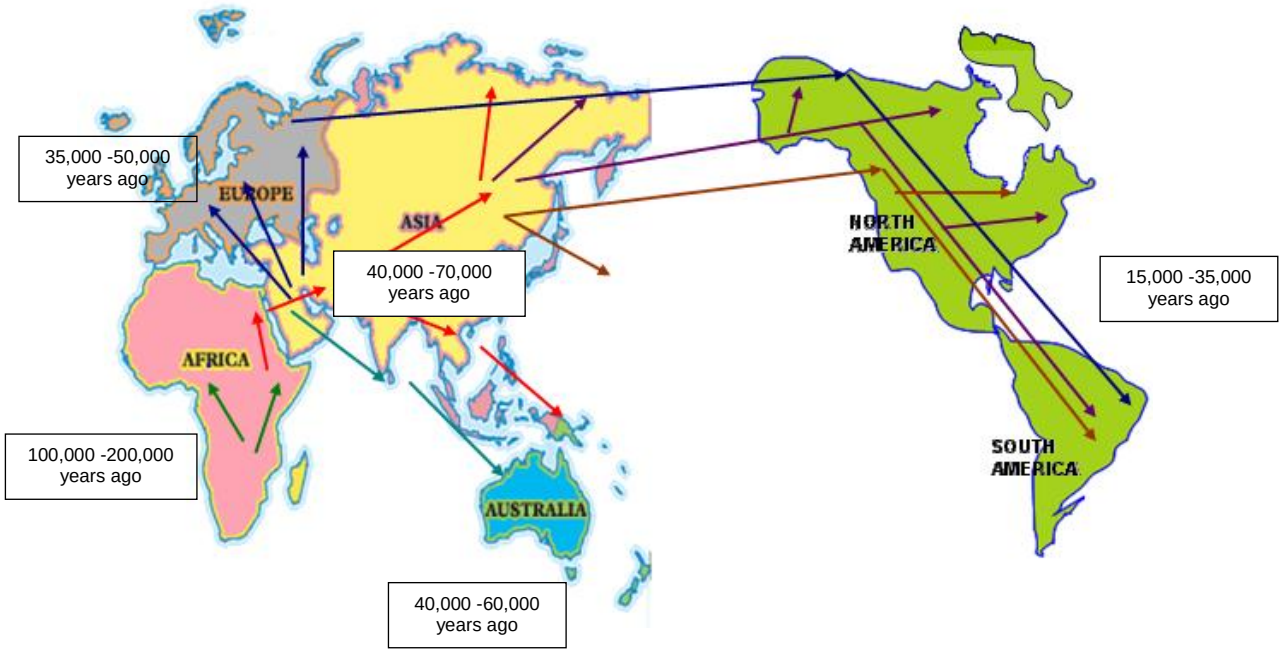



**Figure 1.7: Tracing modern human evolution through mitochondrial DNA mutations.** Mitochondrial DNA map representing different routes of modern human migrations out of Africa based on previous phylogenetic studies. Times of early human expansion across the world were shown in boxes. Data were obtained from the mitochondrial DNA migrations map-1 produced by Family Tree DNA at (www2).

However there is a persistent debate about the "Out-of-Africa" replacement theory and the other opposing theory is the multidirectional hypothesis. Wolpoff et al. (2000), have suggested that human populations in both Africa and Eurasia have contributed to the evolution of modern human (**Figure 1.8**).

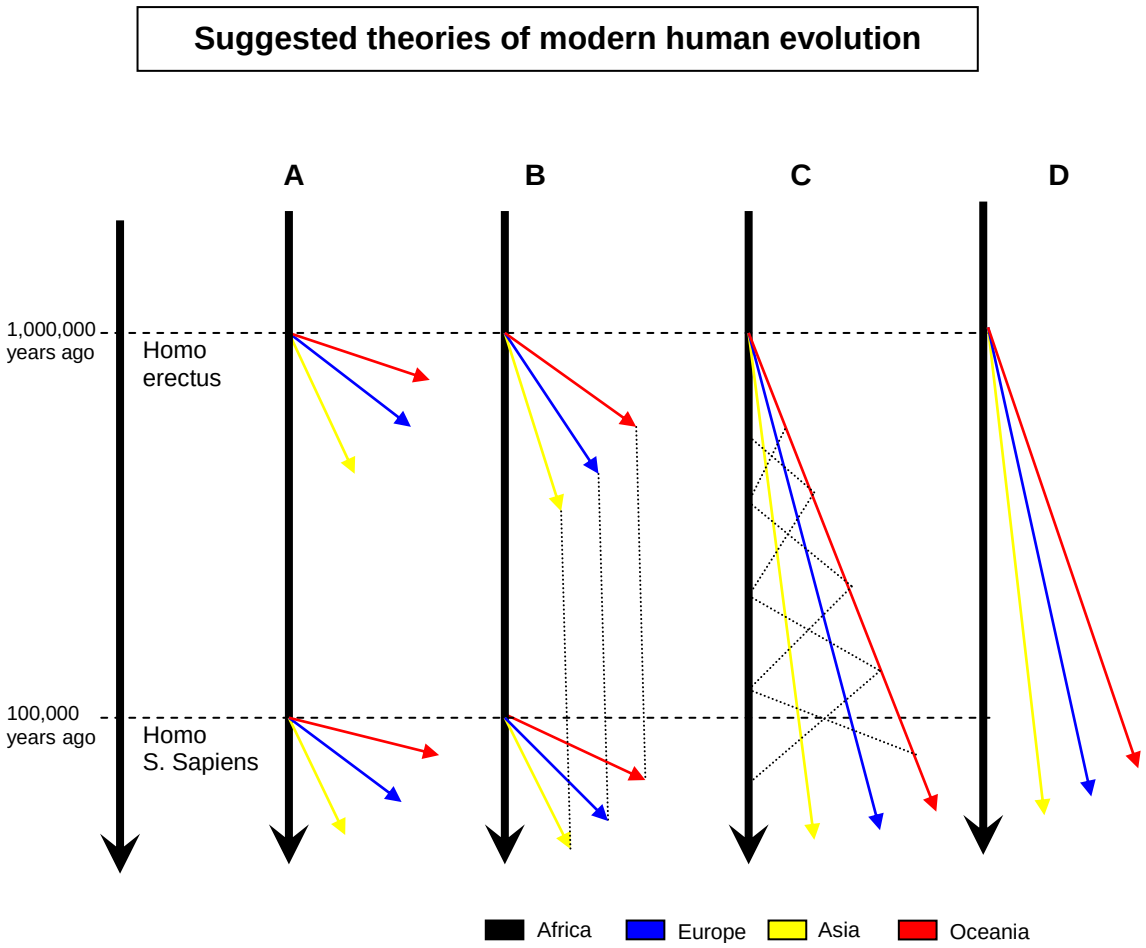



**Figure 1.8: Out-of-Africa versus the multidirectional hypothesis.** The diagram represents the current theories of modern human evolution. The first hypothesis (model **A**) suggests that there was a second wave of migration out of Africa that has occurred 100,000 years ago, in which the modern humans of African origin have invaded the world and have completely replaced archaic human populations. Model **B** is a compromised version of hypothesis **A** that emphasizes the African origin of the modern human populations with possibility of minor local contributions. Models **A** and **B** support the out of Africa theory. Model **C** represents the multidirectional theory in a way that allows continuous mixing between continental populations. Lastly, model **D** represents the multidirectional theory with completely independent multiple origins. Reproduced from Jin and Su (2000).

## 1.4 Mitochondrial DNA haplogroups

### 1.4.1 Definition of mitochondrial DNA haplogroups

Mitochondrial mutations accumulate sequentially leading to constitution of independent lineages. Relationships among different lineages can be estimated by phylogenetic networks (Bandelt et al. 1995; Maca-Meyer et al. 2001) in which mutations are classified in hierarchal levels.. The haplotype is defined as a specific DNA sequence observed in a population sample (Budowle et al. 2003) and thus describes the entire combination of substitutions retrieved from the complete sequence in any given sample and therefore indicates the tips of the phylogeny (Behar et al. 2008b).

Closely related haplotypes are clustered as clades, which represent closely related maternal lineages. A cluster of clades that share common mutations is referred to as a haplogroup. A haplogroup is defined as a set of haplotypes that share basal mutations, indicating a common ancestry (Budowle et al. 2003). Thus, mitochondrial haplogroups are identified by certain clusters of mtDNA mutations, both in coding and non-coding regions, relative to the CRS. Figure 1.9 illustrates this hierarchy.

Currently, there are over six thousand complete mitochondrial DNA sequences available in the mitochondrial database (www3). These sequences can be integrated into the construction of a phylogenetic tree of global human mtDNA. Continuous updating of the global phylogenetic tree is always required in order to include newly identified mitochondrial lineages (Van Oven and Kayser 2009). Fortunately, this continuous refinement of the mitochondrial phylogeny does not change the high level structure of the earlier phylogenies (Logan 2009).

The major haplogroups are specific to geographical locations and/or ethnic groupings (Richards et al. 2000; Maca-Meyer et al. 2001). Therefore, mapping of mitochondrial DNA haplogroups clarifies modern human migration and colonisation around the world.

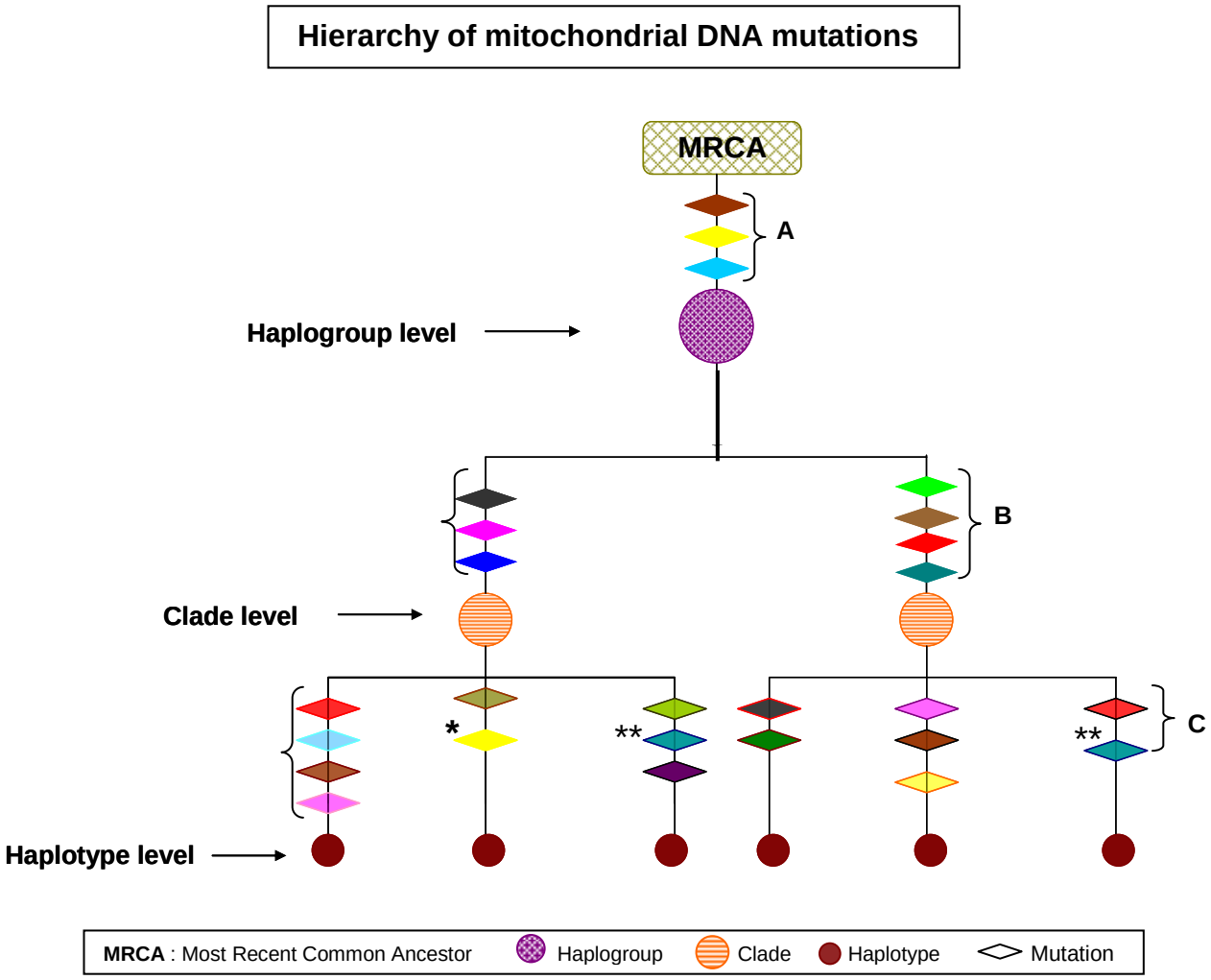



**Figure 1.9: Simplified diagram of the mitochondrial phylogeny showing the hierarchal levels of mitochondrial mutations with subsequent classification of mitochondrial DNA into haplogroups, clades and haplotypes.** Diamond shapes with different colours represent mutations in the mitochondrial DNA relative to the CRS. A haplogroup (represented by a violet patterned sphere) is present at the top of the phylogeny, it is derived from the most recent common ancestor or alternatively from one of the other major haplogroups by having a certain set of basal mutations **(A)** and it gives rise to various clades (represented by stripped brown spheres). Each clade is differentiated from one another by having a different set of additional mutations **(B)** and it gives rise to various haplotypes (represented by brown solid small spheres at the tips of the phylogeny) which represent closely related maternal lineages. Each haplotype has a specific DNA sequence with one more set of mutations **(C)**. The haplotype is identified by the entire combination of mutations **(A+B+C)** within the whole mitochondrial genome relative to the CRS. (*****) indicates a back mutation at a certain nucleotide position and (******) indicates homoplasy (parallel mutation events).

### 1.4.2 Nomenclature of the mitochondrial DNA haplogroups

A universal, agreed nomenclature has been defined for mitochondrial DNA haplogroups (Torroni et al. 1994; Torroni et al. 1996; Wallace et al. 1999; Finnila et al. 2001b; Herrnstadt et al. 2002). This nomenclature system is based on the use of both letters and numbers (**Figure 1.10**). However, the given letters and numbers do not reflect the chronological order of emergence of these haplogroups within the human evolution instead, they reflect the order of their discovery.

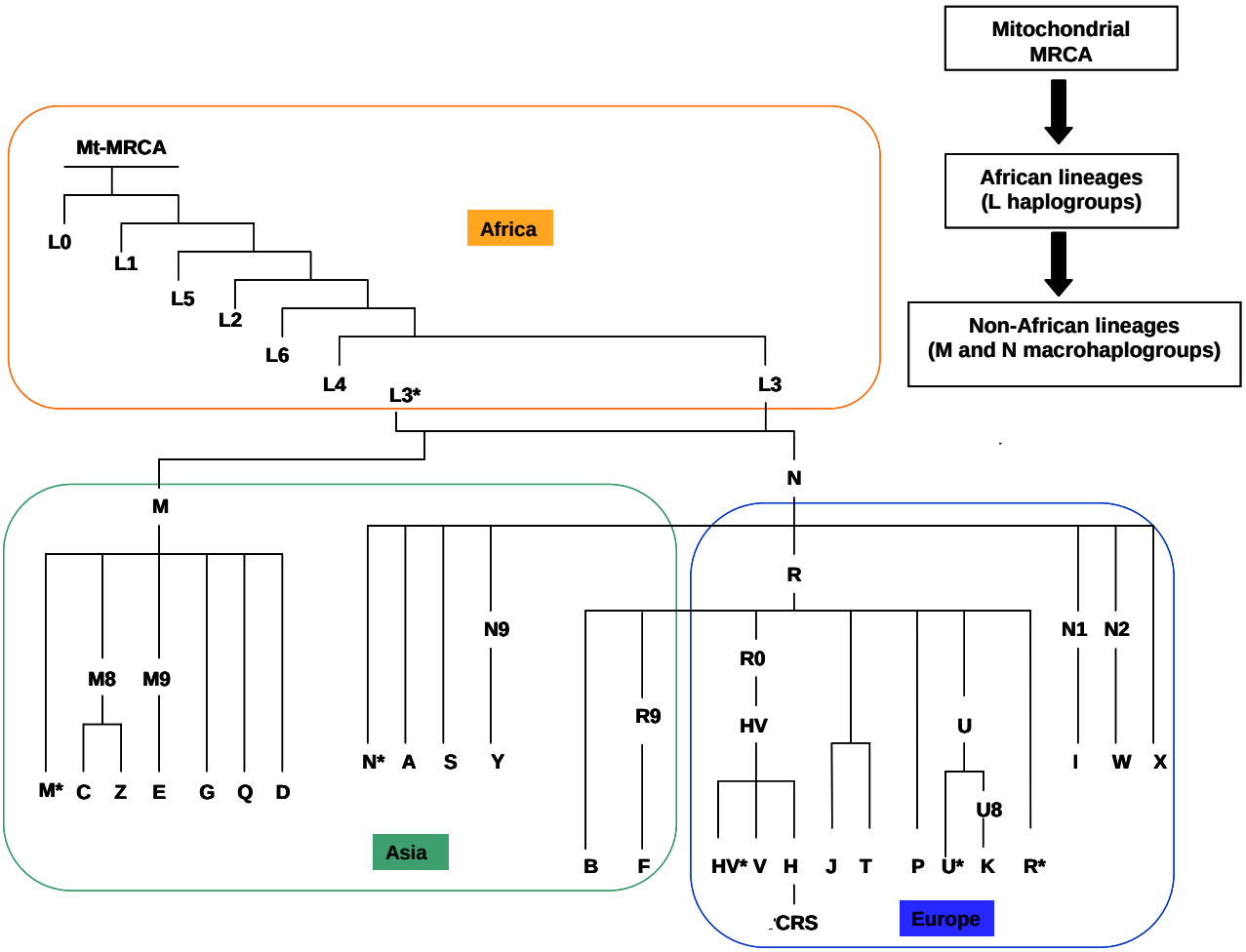



**Figure 1.10: A simplified phylogenetic tree of the mitochondrial DNA haplogroups.** Different mitochondrial lineages were represented by different letters. The L haplogroups represent the most deep rooting branches of the tree originating from the most recent common matrilineal ancestor (MRCA) and are African specific. All the other mitochondrial lineages are derived from the two major macrohaplogroups M and N which are subclades of haplogroup L3. Reproduced from Van Oven and Kayser (2009).

### 1.4.3 Distribution of the mitochondrial DNA haplogroups around the world

#### *1.4.3.1 African specific mitochondrial haplogroups*

There are three major haplogroups (L1, L2 and L3) that represent the most ancient and diverse mtDNA haplogroups, and these are largely restricted to the sub-Saharan African lineages, apart from the L3 haplogroup which is also present in non-Africans, and is considered as the progenitor of the two Eurasian macrohaplogroups (M and N). Dispersal of haplogroup L3 lineages out of Africa has been estimated to have occurred 60,000 – 80,000 years ago (Watson et al. 1997; Salas et al. 2007). The proposed radiation ages for the two macrohaplogroups M and N are 30,000 to 57,600 years ago and 43,000 to 53,000 years ago, respectively (Maca-Meyer et al. 2001).

The Levantine Corridor and the Horn of Africa are thought to be the two main routes for early human migrations out of Africa. Through analysis of mtDNA from ten African and Middle Eastern populations, Rowold et al. (2007) identified the Levantine Corridor as the more likely route. In contrast, other archaeological and genetic evidence supports a single route of human dispersal out of Africa through the Horn of Africa into South Asia and Australia, followed by a second wave of migration into Europe (Mellars 2006; Kumar et al. 2009).

#### *1.4.3.2 Asian specific mitochondrial haplogroups*

Mitochondrial lineages in Asia appear to be a mixture of subclades from both macrohaplogroups M and N. These subclades clearly show a geographic localization, with macrohaplogroup M and its subclades concentrated in the Southern area and South West Asia, contributing to the great majority of the Indian and Mongoloid lineages. Macrohaplogroup N and its subclades are localised in Northern Asia with only a few representatives in South Asia (Tanaka et al. 2004; Thangaraj et al. 2006).

Both M and N derived lineages are found among Eastern Asian populations, resulting from separate waves of modern human dispersal from Northern and Southern Asia (Yao et al. 2002; Allard et al. 2004; Tanaka et al. 2004; Asari et al. 2007; Li et al. 2007; Asari et al. 2008).

Analysis of Chinese, Korean, Japanese and Thai populations have shown heterogeneous structures harbouring both M derived lineages (D4 and D4a subclades) and N derived lineages (haplogroups A and R, and subclades F1a and B4a). The most commonly observed Chinese haplogroup was B4a followed by haplogroups A, F1a, and D4, whereas haplogroups A, D4 and D4a were the most common haplogroups among Korean and Japanese populations. In Thai populations, mitochondrial haplogroups B4a, F1a, and R have the highest frequencies **Figure 1.12** (Allard et al. 2004).

Haplogroup M which originated in South-Western Asia has given rise to the subclade M1 which appears to re-migrated into Africa following the northern coastal route about 40,000 – 45,000 years ago (Olivieri et al. 2006). However, a younger coalescence age for the whole M1 subclade of 20,000 – 30,000 years ago was estimated by others (Gonzalez et al. 2007).

#### *1.4.3.3 European Caucasian mitochondrial haplogroups*

There are 9 major Western European haplogroups ; H, I, J, K, T, U, V, W, and X derived from the macrohaplogroup N (**Figure 1.13**) (Maca-Meyer et al. 2001; Budowle et al. 2003; Torroni et al. 2006).

According to the Scientific Working Group on DNA Analysis Methods (SWGDAM) European Caucasian data, haplogroup H was the most commonly observed mtDNA haplogroup (45.7%), followed by haplogroups U (15.6%), T (10.5%), K (10.5%), and J (10%) (Macaulay et al. 1999; Finnila et al. 2001b; Achilli et al. 2004).

#### *1.4.3.4 Native American mitochondrial haplogroups*

The American continent is known to be the last land to be colonised by humans, however, estimates of the timing of this colonization vary from 11,000 to 40,000 years ago. Until recently, the maternal ancestry of Native Americans was traced back to one of four major mitochondrial haplogroups (A2, B2, C1 and D1). These haplogroups originated in Asia from the two macrohaplogroups N and M (**Figures 1.11 and 1.13**) and then migrated to the Americas via Beringia. Nevertheless, later studies have identified five other minor mitochondrial lineages (subclades) as Native American haplogroups. These lineages are X2a, D2a, D3, D4h3 and C4c (Tamm et al. 2007; Achilli et al. 2008).

All the major African haplogroups (L1, L2, and L3) have been detected in the SWGDAM African-American dataset, primarily the L2a cluster (18.9%), but haplogroups L1c, L1b, L3e2 and L3b are also common (**Table 1.1**). Overall, the African mitochondrial lineages constitute approximately 90% of the mtDNA in Afro-descendents Americans. Several European Caucasian and East Asian haplogroups are also present, with the most common of these being haplogroup H (2.8%) (Allard et al. 2005).

**Table 1.1: Frequency of the African mitochondrial lineages in the Afro-descendents Americans.** According to the SWGDAM African-American dataset.

| **Haplogroup** | L2a | L1c | L1b | L3e2 | L3b | L2c | L3d | L3e1 | L2b | L3f |
|---|---|---|---|---|---|---|---|---|---|---|
| **Frequency (%)** | 18.9 | 11.0 | 9.1 | 9.0 | 8.1 | 6.2 | 6.0 | 4.9 | 4.7 | 4.6 |

### 1.4.4 Definitive markers for the identification of mitochondrial DNA haplogroups

Analysis of either the whole mitochondrial genome or the hypervariable regions has been considered as the main focus of mitochondrial population studies. Examination of distinct populations has improved both the quantity and the quality of the mtDNA database, enabling the use of mtDNA in forensic identification, medical genetics and studies of human evolution.

In order to facilitate and standardise the process of haplogroup assignment of mitochondrial DNA sequences, specific markers for each haplogroup were identified. These markers are variations in the whole mitochondrial genome relative to the CRS. These variations represent mutations occurring in the mtDNA throughout the human evolution and are shared by closely related mitochondrial lineages. These mutations are known as SNPs. Phylogenetic networks of the mitochondrial DNA sequences were constructed with specification of the diagnostic SNPs for each haplogroup (**Figures 1.11-1.13**).

**Phylogenetic tree of the major African mitochondrial haplogroups**

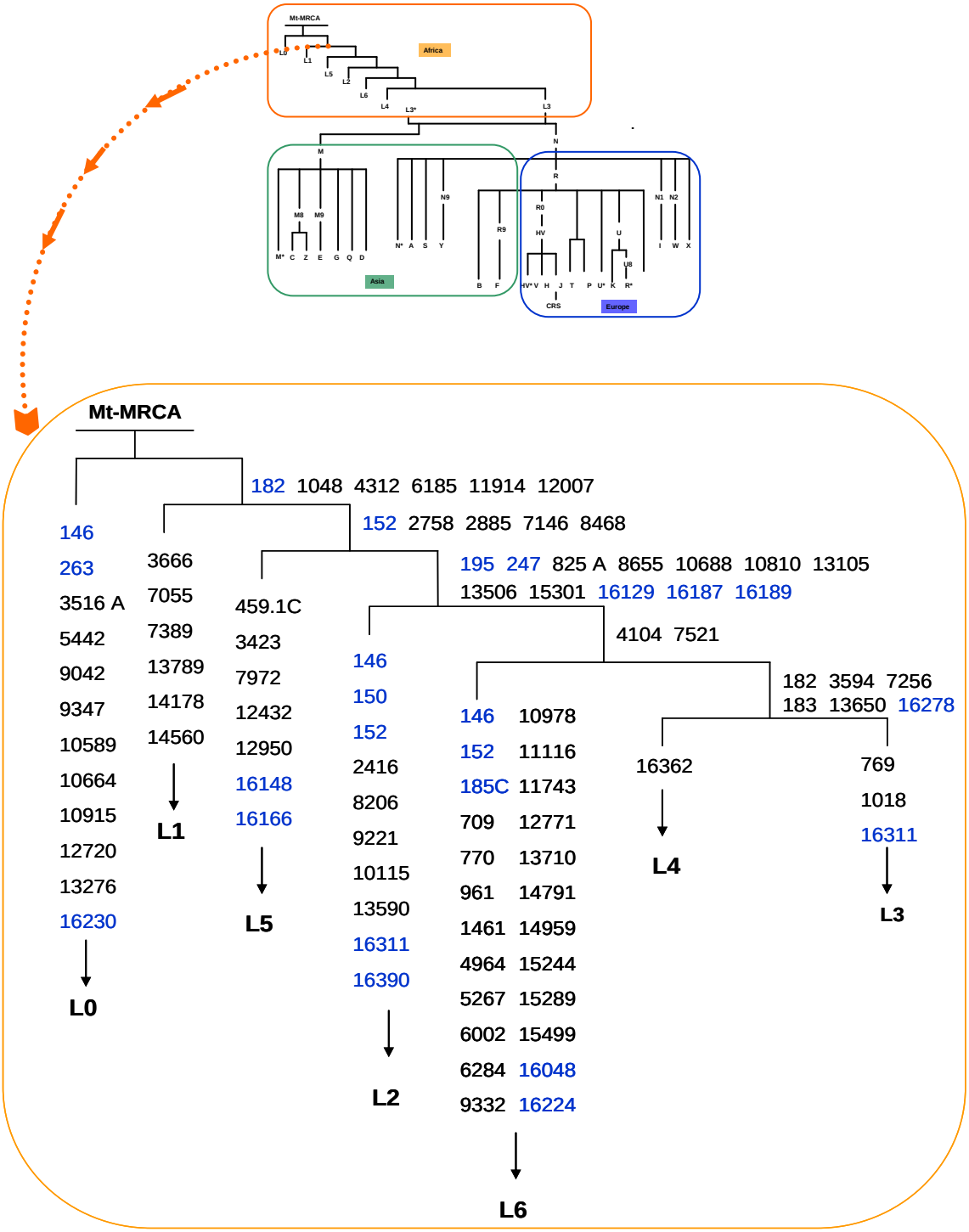



**Figure 1.11: Human mitochondrial DNA phylogeny - major African mitochondrial haplogroups.** Along the links of the tree are the mutations that are known to be shared between different haplotypes within the same haplogroups. Nucleotide positions are relative to the CRS (Gene Bank accession NC_012920). Mutations are transitions unless an exact base change is specified. Insertions are indicated by a dot followed by the number and type of inserted nucleotide(s). Mutations in the control region are shown in blue and mutations in the coding region are shown in black (Palanichamy et al. 2004; Bandelt et al. 2006; Torroni et al. 2006; Behar et al. 2008b; Malyarchuk et al. 2008).

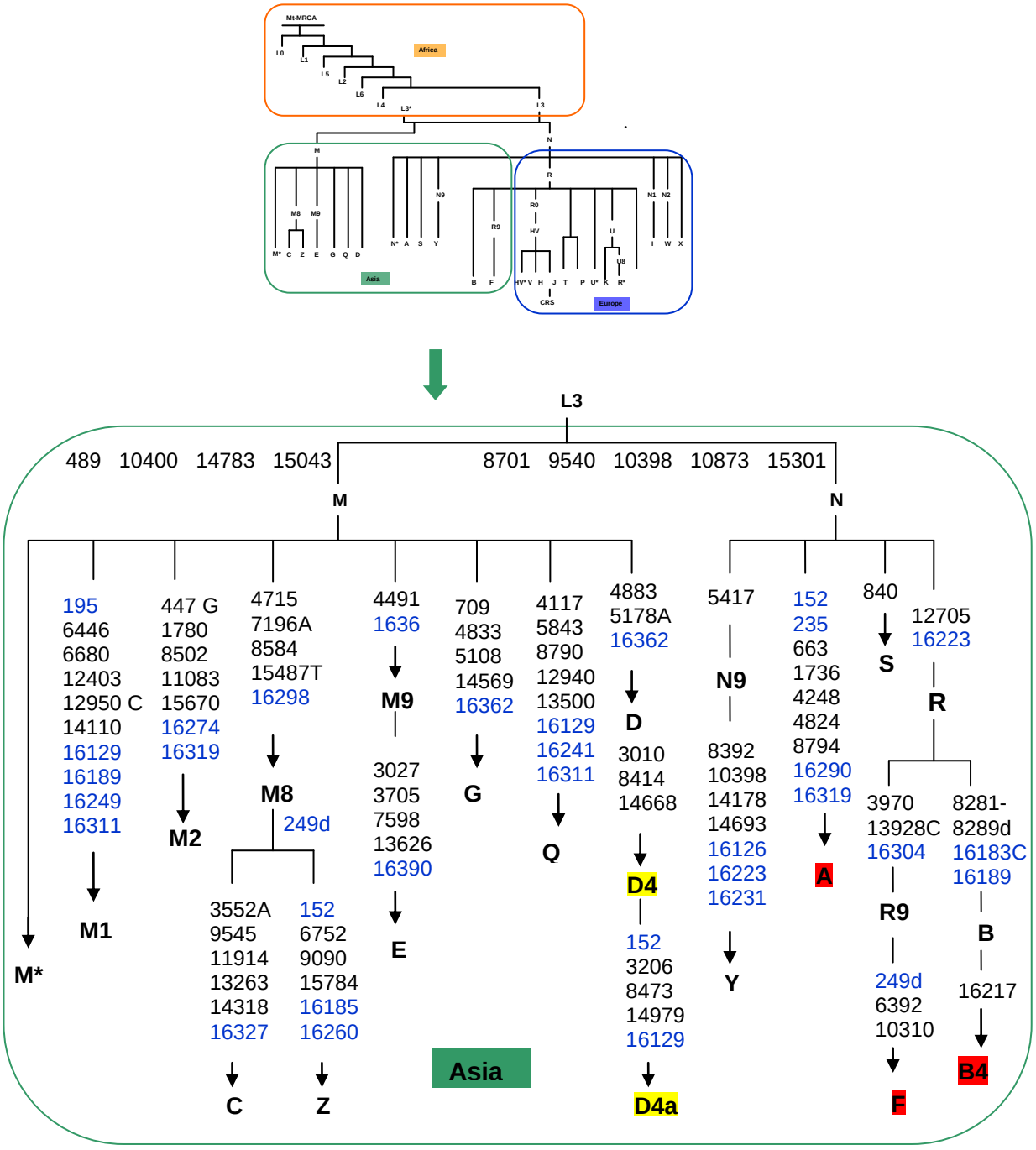



**Figure 1.12: Human mitochondrial DNA phylogeny - M and N derived mitochondrial lineages reported in Asian populations.** Along the links of the tree are the mutations that are known to be shared between different haplotypes within the same haplogroups. Nucleotide positions are relative to the CRS (Gene Bank accession NC_012920). Mutations are transitions unless an exact base change is specified. d indicates a deletion. Control region mutations are shown in blue and coding region mutations are shown in black. D4 and D4a (highlighted in yellow) are the most commonly observed M derived lineages among east Asian populations whereas A, B4a and F1a (highlighted in red) are the most commonly observed N derived lineages among east Asian populations (Palanichamy et al. 2004; Tanaka et al. 2004; Bandelt et al. 2006; Kong et al. 2006; Metspalu et al. 2006; Olivieri et al. 2006; Thangaraj et al. 2006; Derenko et al. 2007; Gonzalez et al. 2007; Achilli et al. 2008; Kumar et al. 2008).

**Phylogenetic tree of the major European mitochondrial DNA haplogroups**

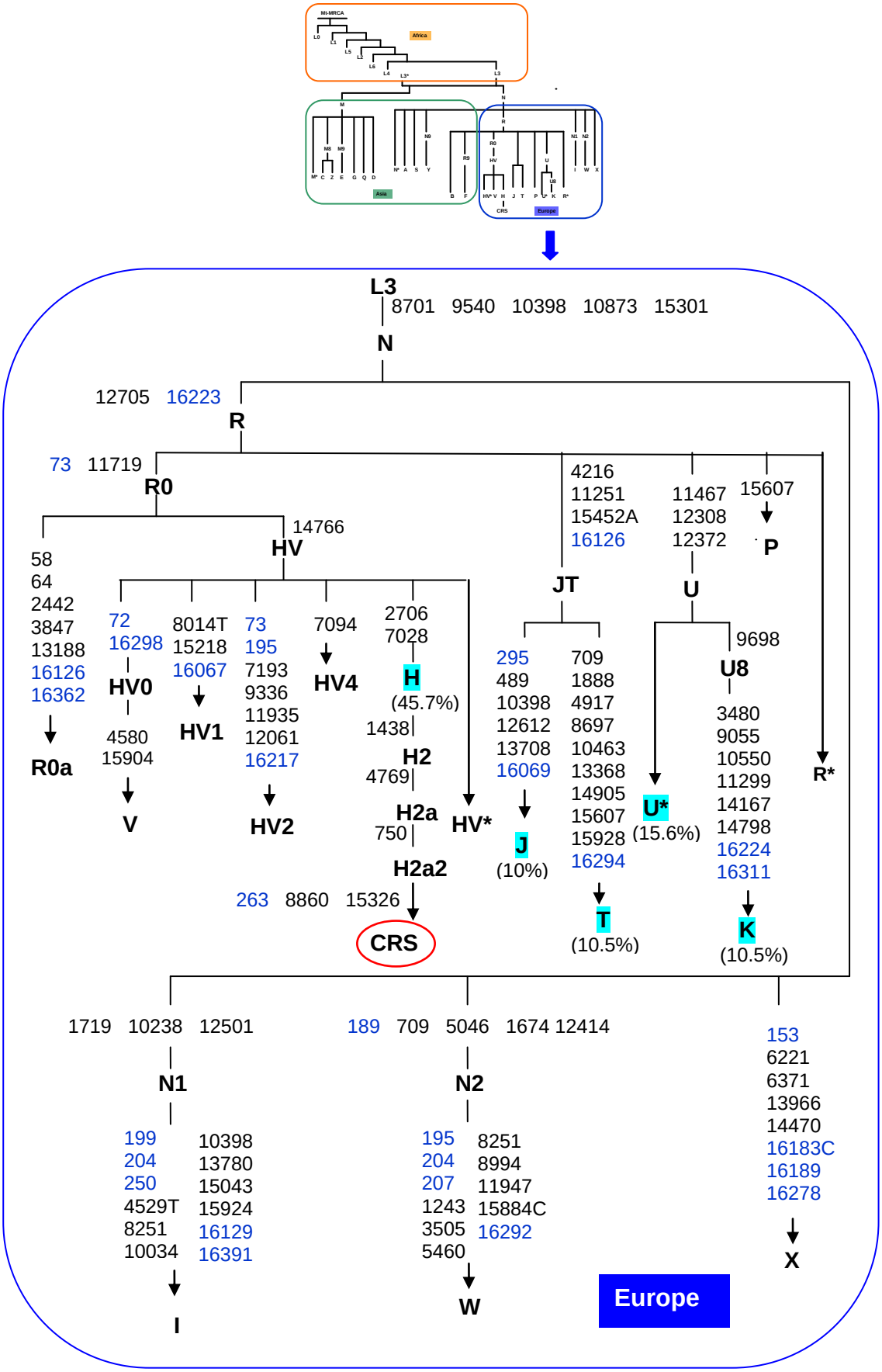


**Figure 1.13: Human mitochondrial DNA phylogeny - major European mitochondrial haplogroups.** Along the links of the tree are the mutations that are known to be shared between different haplotypes within the same haplogroups. Nucleotide positions are relative to the CRS (Gene Bank accession NC_012920) which represents a descendant of H2a2 subclade. Mutations are transitions unless an exact base change is specified. Control region mutations are shown in blue and coding region mutations are shown in black. The most commonly observed European haplogroups are highlighted in turquoise with the frequency of each shown between brackets. (Macaulay et al. 1999; Finnila et al. 2001b; Reidla et al. 2003; Achilli et al. 2004; Loogvali et al. 2004; Shen et al. 2004; Achilli et al. 2005; Behar et al. 2006; Carelli et al. 2006; Gonzalez et al. 2006; Kivisild et al. 2006; Torroni et al. 2006; Abu-Amero et al. 2007; Roostalu et al. 2007; Tamm et al. 2007; Behar et al. 2008a; Brandstaetter et al. 2008; Alvarez-Iglesias et al. 2009; Logan 2009).

## 1.5 Impact of mitochondrial DNA on the forensic domain

### 1.5.1 Medico legal peculiarities of the mitochondrial genetic system

Mitochondrial DNA has multiple unique characteristic features that drive its direct impact on different forensic applications:

- MtDNA is a small genome, present in high copy number per somatic cell (up to several thousands of copies) compared to the two copies of the nuclear genome, in addition to the fact that they are encapsulated in a double-walled organelle, mtDNA is highly recoverable from severely degraded forensic samples, whereas nuclear DNA is not (Bandelt et al. 2006). Moreover, the circular nature of the mitochondrial genome provides a protective mechanism against the breakdown by exonucleases and hence increases its survival potential (Butler 2005b).

- Mitochondrial DNA is inherited down the maternal line according to the most widely accepted model of inheritance, and hence mtDNA passes unchanged from mother to daughters and sons, from generation to generation apart from spontaneous mutations. Thus maternal relatives can provide clues for solving missing persons and mass disaster investigations even in the absence of the direct mother (Giles et al. 1980; Lightowlers et al. 1997; Primorac and Schanfield 2000).

- MtDNA is characterised by rapid evolutionary rate with 5-10 times higher mutation rate compared to the nuclear genome. This high mutation rate is more evident in the hypervariable regions (Howell et al. 2007). There are local differences in the mutation rate among different nucleotide positions within the control region. There are certain nucleotide positions which are particularly highly mutating, showing high level of nucleotide variability and so referred to as “hot spots” (e.g. positions 146, 150, 152, 195, 16189, 16311, 16362, 16519) while others are known to be more stable (e.g.

positions 477, 493, 16108, 16219) (Van Oven and Kayser 2009). This differential rate of mutation might be explained by the presence of certain sequence background of some sites in the control region which in turn leads to a higher mutation rate at these sites than at others (Howell et al. 2003) or alternatively, these sites might be evolving under positive selection conditions (Howell et al. 2007).

Fortunately, this high rate of nucleotide variability and polymorphisms do not have major consequences on the human health since these regions do not code for essential substances. On the other aspect, these nucleotide polymorphisms provide a valuable tool for human identification and differentiation between different maternal lineages as well as tracing of modern human migration and colonisation around the world (Kogelnik et al. 1998; Ingman and Gyllensten 2001).

### 1.5.2 Mitochondrial DNA as a useful tool in human identification (case studies)

#### *1.5.2.1 Identification of historical cases by mitochondrial DNA analysis*

**A royal maternity kinship cleared by ancient DNA analysis**

There have been doubts among historians whether the woman entombed at Roskilde Cathedral (during the Scandinavian Viking era) was indeed Estrid, the mother of the last Danish Viking King, Sven Estridsen or not. Analysis of the mitochondrial DNA from dental pulps of the two royals was able to assign the king's sequence to haplogroup H, whereas the mitochondrial sequence of the woman's sample was assigned to subclade H5a and differed from the king's sequence at two nucleotide positions in HVI. Therefore, it was highly unlikely that the entombed woman was the mother of the king (Dissing et al. 2007).

**Identification of the remains of the Romanov Nicholas II royal family**

The controversy surrounding the fate of the last Russian Emperor, Nicholas II and his family after the Russian Civil War in 1918 was solved partly by mtDNA analysis. In early 1990s, a first grave was found in the Ural region of Russia which contained the remains of the Romanov Imperial family. Mitochondrial DNA analysis of both skeletal remains and an archival blood sample (~ 117 years old) of Nicholas II has confirmed his identity after using his brother's mtDNA as a reference sample. Following analysis of the mtDNA from the other skeletal remains with the aid of maternal reference samples (collected from descendants of the same maternal lineage as the Emperor's wife),

identity of the last Russian Emperor's wife Empress Alexandra and three of their children was verified (Gill et al. 1994; Ivanov et al. 1996). The fate of the two youngest children of the Romanov family was unknown till 2007 when a second grave was discovered in the same Ural region containing bone fragments of two burned skeletons. Following DNA extraction from the severely damaged bone fragments, mtDNA analysis was successfully used to verify the kinship relationship between these remains and those exhumed from the first grave, confirming the identity of the two skeletons as being the two youngest children of the Romanov family (Rogaev et al. 2009).

There are several other historical cases in which mtDNA analysis has been successfully applied and has greatly enabled verification of the human identity; (**Table 1.2**).

**Table 1.2: Historical cases identified by mitochondrial DNA analysis.** Herein listed some cases in which mtDNA analyses have been used to verify the identity of the putative body remains.

| Genetic identification of | Reference |
|---|---|
| 7500-year-old Neolithic skeletons of the first European farmers | Haak et al. (2005) |
| A 5000-year-old mummified human body found in the Tyrolean Alps | Handt et al. (1994) |
| The Early Christian writer evangelist Luke | Vernesi et al. (2001) |
| Francesco Petrarch (1304-1374), the Italian poet and scholar | Caramelli et al. (2007) |
| The famous astronomer Nicolaus Copernicus (1473-1543) | Bogdanowicz et al. (2009) |
| Louis XVII (1785-1795), son of Louis XVI, King of France and Marie-Antoinette | Jehaes et al. (2001) |
| The American outlaw Jesse James (1847-1882) | Stone et al. (2001) |
| The prominent Nazi official Martin Bormann (1900-1945?) | Anslinger et al. (2001) |

#### *1.5.2.2 Genetic identification of human skeletal remains from mass graves*

During repatriation and identification of Finnish World War II soldiers, 167 (out of 181) bone samples were successfully sequenced for mitochondrial hypervariable regions and were found to belong to a minimum of 104 individuals. Identity of 79 individuals (76%) was confirmed by comparing their mitochondrial DNA sequences against the corresponding family reference sequences. The majority of the reference samples were obtained from direct maternal relatives, meanwhile more distant relatives were helpful in certain cases. There was an extreme case, where a positive match for a soldier was found to the grandson of his great-grandmother's sister (Palo 2007).

During the identification of the human remains after the (1991-1995) Serbian aggression on Croatia and Bosnia and Herzegovina (BH), DNA analysis was requested for identification of cases in which the standard forensic methods failed. However, in certain cases, the quantity and /or the quality of the nuclear DNA was not satisfactory and so, the mtDNA analysis was performed (Andelinovic et al. 2005).

In the Armed Forces DNA identification Laboratory (AFDIL), United States, it was possible to amplify a 410bp segment of the mtDNA control region from extremely degraded DNA samples, which were extracted from 125 years old skeletal remains of U.S. Civil War soldiers (Fisher et al. 1993). A success rate of 83.6% was reported for obtaining a mitochondrial DNA profile either partial or complete from 116 unidentified skeletal remains submitted to the Mitotyping Technologies laboratory, State College, between February 1999 and May 2005 (Nelson and Melton 2007). Amongst all skeletal remains, femurs and teeth were the most successful samples in providing either full or partial profiles.

In Germany, mitochondrial hypervariable regions analysis has been successfully applied at the Institute of Legal Medicine for the identification of victims of crime cases or missing persons whose bodies remained undiscovered for many years (up to 9 years) where STR typing was unsuccessful (Bender et al. 2000).

#### *1.5.2.3 Utility of mitochondrial DNA analysis in mass disasters*

**DNA identification of Tsunami victims in Thailand**

After the 2004 Ocean earthquake Tsunami at Southeast Asia that killed nearly 5,400 people in Southern Thailand including a large number of foreign tourists and local residents. There have been extensive efforts conducted by the forensic scientists aiming to reveal the identity of these victims. Unfortunately, serious decomposition of bodies caused failure in recovery of enough DNA for nuclear marker typing, however, forensic scientists succeeded in obtaining fully informative results for mitochondrial markers (HVI and HVII) from 258 tooth samples with a success rate of 51% (258/ 507) (Deng et al. 2005).

**DNA identification of the World Trade Centre (WTC) disaster victims**

The World Trade centre attacks in New York, in September, 2001 which resulted in more than 2,700 human losses of varied racial and ethnic background, represented a great challenge for DNA identification efforts which were confronted with a massive number (several thousands) of skeletal remains. These remains were subjected to severe DNA deteriorating conditions including the explosion itself, intense fire and heat, and the subsequent water exposure by the fire extinguishers. These adverse environmental conditions did not facilitate the human identification using the standard STR profiling techniques. In an effort to develop a high quality, high throughput DNA analysis procedure, supplementation of the STR profiling results with the mitochondrial DNA data has improved the success rate of victims' identification (Holland et al. 2003).

**The use of mitochondrial DNA typing in the identification of air crash victims**

Mitochondrial DNA analysis has played an important role in the identification of victims after several aircraft accidents, particularly when complete STR typing profiles were not available. In the experience of the air plane crash in the Philippines, in 1998, which killed all 104 passengers and crew, the success rate of victims' identification after using the STR typing was low (50%), due to extreme degradation of the DNA extracted from the victims' remains. However, the use of mtDNA typing combined with the STR profiling raised the success rate up to 95% (Goodwin et al. 1999). In another air accident, two victims were crushed into small pieces, and the investigators were provided with 33 fragmentary tissue samples recovered from under the sea and suspected to belong to the victims. Identification of the victims was enabled by applying ABO genotyping, STR profiling and mtDNA analysis altogether (Mukaida et al. 2000). In the efforts conducted for the identification of 10 individuals died after an aircraft accident in the Madeira Island north cost, in 2003, autosomal STR typing of nine biological samples resulted in five different profiles. However, the STR profiles obtained from three of these fragments were incomplete and so the accurate identification of those individuals was completed via analysis of Y- chromosome STRs and mitochondrial hypervariable regions (Anjos et al. 2004).

### *1.5.2.4 A case of exchanged body remains solved by mtDNA analysis*

In a case of homicide involving five Portuguese citizens whose identification was suspected to have been exchanged, mitochondrial hypervariable regions analysis confirmed the exchange of body remains. MtDNA sequence of each body always matched the alleged mother. Moreover, a kinship relation was suggested by detecting similar mitochondrial substitutions in two bodies and one of the presumptive mothers (**Table 1.3**) (Cruz et al. 2004).

**Table 1.3: HVI and HVII mtDNA Sequences of the five individuals and four alleged mothers compared to the Cambridge Reference Sequence (CRS).** A letter in each position indicates a nucleotide substitution relative to the CRS and the absence means that the sequence at that nucleotide position is identical to the reference sequence. A nucleotide position followed by a dot and a number indicates inserted nucleotide which is not present in CRS. Each pair of matched sequences is highlighted in the same colour. Kinship relation was evident between samples from body 2, body 5 and mother 3 by having the same HVI and HVII sequences (Cruz et al. 2004).

| | Nucleotide position | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | HVI | | | | | | | | | | | | | | | | | | |
| Sample origin | 16069 | 16126 | 16129 | 16148 | 16168 | 16172 | 16187 | 16188 | 16189 | 16193 | 16223 | 16224 | 16230 | 16264 | 16270 | 16278 | 16293 | 16311 | 16320 |
| CRS | C | T | G | C | C | T | C | C | T | C | C | T | A | C | C | C | A | T | C |
| Body 1 | . | . | A | T | T | C | T | G | C | . | T | . | G | . | . | T | G | C | T |
| Body 2 | . | . | . | . | . | . | . | . | . | . | . | C | . | . | . | . | . | C | . |
| Body 3 | . | C | . | . | . | . | T | . | C | . | T | . | . | T | T | T | G | C | . |
| Body 4 | T | C | . | . | . | . | . | . | . | T | . | . | . | . | . | T | . | . | . |
| Body 5 | . | . | . | . | . | . | . | . | . | . | . | C | . | . | . | . | . | C | . |
| Mother 1 | . | . | A | T | T | C | T | G | C | . | T | . | G | . | . | T | G | C | T |
| Mother 2 | . | C | . | . | . | . | T | . | C | . | T | . | . | T | T | T | G | C | . |
| Mother 3 | . | . | . | . | . | . | . | . | . | . | . | C | . | . | . | . | . | C | . |
| Mother 4 | T | C | . | . | . | . | . | . | . | T | . | . | . | . | . | T | . | . | . |

| | HVII | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 73 | 93 | 95 | 114 | 150 | 152 | 182 | 185 | 189 | 195 | 198 | 136 | 247 | 263 | 295 | 309.1 | 315.1 | 316 |
| CRS | A | A | A | C | C | T | C | G | A | T | C | T | G | A | C | -- | -- | G |
| Body 1 | . | G | C | . | . | . | . | . | G | . | . | C | A | G | . | C | C | . |
| Body 2 | G | . | . | T | . | . | . | . | . | . | . | . | . | G | . | C | C | A |
| Body 3 | G | . | . | . | . | C | T | T | . | C | T | . | A | G | . | . | C | . |
| Body 4 | G | . | . | . | T | . | . | . | . | . | . | . | . | G | T | . | C | . |
| Body 5 | G | . | . | T | . | . | . | . | . | . | . | . | . | G | . | C | C | A |
| Mother 1 | . | G | C | . | . | . | . | . | G | . | . | C | A | G | . | C | C | . |
| Mother 2 | G | . | . | . | . | C | T | T | . | C | T | . | A | G | . | . | C | . |
| Mother 3 | G | . | . | T | . | . | . | . | . | . | . | . | . | G | . | C | C | A |
| Mother 4 | G | . | . | . | T | . | . | . | . | . | . | . | . | G | T | . | C | . |

### *1.5.2.5 Identification of burned human remains using mtDNA analysis*

In a case of homicide in Italy, in which burned body remains were found in a wood, the putative mother was the only relative available. STR typing analysis was performed and provided an incompatibility at VWA locus, meanwhile, analysis of the mitochondrial hypervariable regions revealed a perfect match between the sequences obtained from the burned remains and the putative mother in both HVI and HVII regions which raised

the likelihood of the maternal relationship between the victim remains and the putative mother (Cerri et al. 2004).

In a relatively similar case of homicide in Japan, a skeletonised female body, which was found in a forest, was identified a year and eight months later using both DNA analysis and dental characteristics in a Chest X-ray photograph. The image of the Chest X-ray included frontal parts of the upper and lower jaws which enabled a comparison between the dental characteristics of the candidate and the photographs and the dental X-ray images that were taken when the unknown body was found. The comparison of the dental characteristics revealed a similarity in the patterns of fillings, in addition to the characteristic protrusion of the left upper central incisor. DNA analysis involved both mitochondrial haplotype analysis and ABO genotyping. Mitochondrial haplotype obtained from the body remains (teeth) showed an exact match with the haplotype obtained from the putative mother and her daughter, and thus, supported the mutual blood relationship (Minaguchi et al. 2005).

#### *1.5.2.6 A rape case solved by mtDNA mixture analysis*

Using the differential lysis-based DNA extraction and cloning techniques, it was possible to obtain the mitochondrial hypervariable regions HVI and HVII from the male fraction of the DNA extracted from the vaginal swab in a case of rape. Following sequencing of clones, the suspect's haplotype was detected despite being in a very low ratio compared to the female fraction (Hatsch et al. 2007). However, detection of the suspect's haplotype is not a sufficient evidence to prove his direct involvement in the rape accident particularly that the male mtDNA in this case was only detected at a very low ratio compared to the female one. Thus it is absolutely difficult to exclude the presence of a third mtDNA type at a non-detectable level using this differential-lysis protocol that might belong to the actual rape perpetrator.

## 1.6 The association of mitochondrial DNA haplogroups with the health status

A current area of interest for many researchers is the role of mitochondrial haplogroup specific sequence variations in the pathogenesis of complex diseases and neurodegenerative disorders (Torroni and Wallace 1994). On top of the list of disorders that have been extensively investigated is Leber Hereditary Optic Neuropathy (LHON). Analysis of the mitochondrial DNA in LHON European patients has reported increased risk of disease expression of the LHON mutation (11778/ND4) in presence of either a J1c or J2b mtDNA. And so, these mitochondrial subclades may enhance the effect of the LHON mutation on some functional parameters of the respiratory-chain complexes I and III (Carelli et al. 2006). In LHON patients of Asian origin, there was a significant increase in the risk of hearing loss in patients carrying mitochondria haplogroup M7b1'2, whereas haplogroup M8a has been shown to provide a protective effect (Bilal et al. 2008).

LHON is not the only disorder in which mitochondrial haplogroup J may be involved. In a study of 1833 American HIV-1 patients of European origin, a provisional association of haplogroups J and U5a with accelerated progression to AIDS and death was reported (Hendrickson et al. 2008). Haplogroup J may also be associated with optic neuritis (Reynier et al. 1999). Higher frequencies of haplogroups J, L and T were reported among 52 Iranian Multiple Sclerosis (MS) patients. It was also noted in the same study that a high proportion of haplogroup H was associated with optic nerve involvement, whereas , haplogroup A was associated with a younger age of onset (Ghabaee et al. 2009).

Mitochondrial DNA has always been considered to play a role in several disorders that include stroke as a component of the symptoms. It was suggested that mitochondrial haplogroup H1 has a protective effect against ischemic stroke (Rosa et al. 2008), whereas, R0a, HV, and U haplogroups were proposed as potential factors for increasing risk of stroke. Mitochondrial haplogroup U has also been implicated in increasing the risk of many other serious conditions, like occipital stroke in patients with migraine (Majamaa et al. 1998), Alzheimer's disease in men (Van der Walt et al. 2004), and severe Knee Osteoarthritis (Rego-Perez et al. 2008). Despite the apparent protective effect against ischemic stroke, haplogroup H was proposed to interfere with the

therapeutic effect of high doses of riboflavin in migraine patients (Di Lorenzo et al. 2009).

Haplogroup T has gained a particular attention in the last few years among the expanding list of association studies. It has been suggested to be more prevalent among patients with Coronary artery disease and type 2 Diabetes. The association between mitochondrial haplogroup T and type 2 Diabetes was more pronounced among patients with diabetic retinopathy (Kofler et al. 2009). In addition, haplogroup T2 mutational motif was linked to Age-Related Macular Degeneration (AMD) (SanGiovanni et al. 2009). Several studies have suggested implication of the mitochondrial DNA in male infertility (Folgero et al. 1993; Kao et al. 1995; Zeviani and Antozzi 1997; Moore and Reijo-Pera 2000; Shamsi et al. 2008; Venkatesh et al. 2009). There was a proposed association between mitochondrial haplogroup T and reduced sperm motility (asthenozoospermia) (Ruiz-Pesini et al. 2000).

Longevity in the Japanese population has been investigated with regards to the possible involvement of mitochondrial DNA. An association between haplogroup D4a and longevity in Japan was reported (Bilal et al. 2008). The African mitochondrial DNA lineages have been underrepresented in the growing series of association studies, nonetheless, an association was reported between African ancestry and increased risk of asthma (Vergara et al. 2009).

Despite the big number of association studies that have been carried out during the last few years, aiming to unravel the role of mtDNA sequence variation in various disorders, they all (apart from LHON) have to be considered provisional. This is due to presence of several limitations in these studies, including small number of samples, restriction to certain population groups, non availability of replicated studies (Carelli et al. 2006). Many of these studies require to be validated by examining different populations to rule out the possible associations due to diet, life style or environmental factors (Bilal et al. 2008).

Normal mitochondrial function requires both mtDNA and nuclear DNA encoded factors, and so nuclear DNA markers should not be overlooked when considering associations (Suissa et al. 2009).

## 1.7 Amelogenin

The amelogenin gene belongs to a family of extracellular matrix protein, involved in enamel mineralization. It is encoded by two single-copy genes located on the short arm of both the X- chromosome (AMELX, on Xp22.1-Xp22.3) and the Y- chromosome (AMELY, on Yp11.2) (Nakahori et al. 1991). The two amelogenin homologues differ in size and sequence, and hence amelogenin is commonly used as a genetic marker for sex determination in forensic casework, archaeological specimens, preimplantation and prenatal diagnoses (Kashyap et al. 2006; Cadenas et al. 2007).

The commonly used DNA-based gender test targets a conserved region of the amelogenin gene in intron 1 which includes a 6 bp deletion on the X- homologue. Therefore, PCR amplification of this region produces two different amplicons in male samples that differ in size by 6 bp and can be easily distinguished by gel electrophoresis. In female samples PCR amplification of the same region results only in one product (Nakahori et al. 1991; Sullivan et al. 1993).

The amelogenin gene is included as a component of several commercial multiplexing kits that are commonly used for DNA profiling such as PowerPlex®16 Bio and PowerPlex® ES manufactured by Promega and AmpF/STR® Identifiler™, AmpF/STR® Profiler Plus™ ID and AmpF/STR® SEfiler™ manufactured by Applied Biosystems. Amplification failure can cause misidentification of a male sample as a female, with possibly severe consequences in criminal investigations and prenatal diagnosis of X-linked recessive disorders (Shadrach et al. 2004; Kashyap et al. 2006; Mitchell et al. 2006).

The male specific part of the Y-chromosome is well known to possess massive palindromes and inverted repeats, and hence has long been described as a candidate for high degree of structural polymorphisms and genomic rearrangements. These variations include heterochromatin and euochromatin length variations, pericenteric inversion and neutral translocations with autosomes (Skaletsky et al. 2003; Jobling et al. 2007; Yong et al. 2007).

In the last ten years, a few cases of amelogenin drop-out were reported in certain populations causing misinterpretation of sex testing. Failure of amelogenin typing could be a result of either mutations in the primer binding sites or a deletion of the AMELY locus on the Y-chromosome (Kashyap et al. 2006). While establishing a national DNA database in Australia, amelogenin Y-homologue drop-out was reported (Roffey et al. 2000). Similarly, a rare mutation in the primer binding region in intron 1 on the X-homologue (AMELX) was reported in a Caucasian male (Shadrach et al. 2004). However, the extreme rarity of such events, postulate that these alterations are resulting from individual mutations rather than polymorphisms.

Failure to amplify AMELY can be due to an interstitial deletion of a relatively large segment involving the pericentromic region of the short arm of the Y-chromosome. These deletions have arisen through non-allelic homologous recombination (NAHR) between the proximal major repeat array of TSPY gene and the distal copy of TSPY immediately downstream of IR3 inverted repeat (**Figure 1.14)** (Santos et al. 1998; Lattanzi et al. 2005; Jobling et al. 2007; Yong et al. 2007).

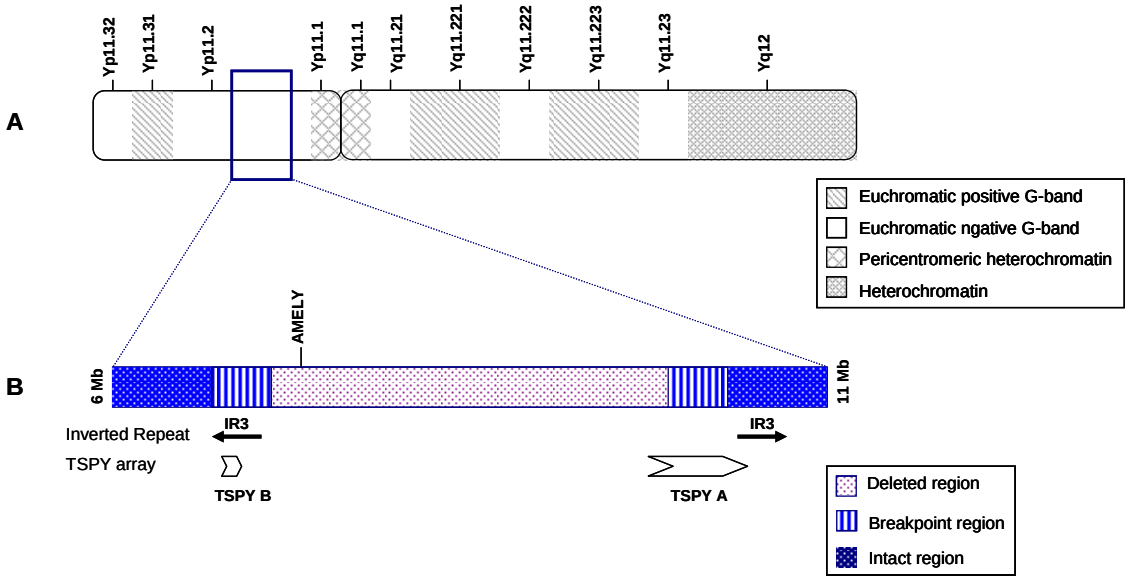



**Figure 1.14: Schematic representation of the human Y-chromosome. A** is an ideogram of the Y-chromosome with the cytogenetic designations for the bands. The blue box outlines the deleted region. **B** illustrates the deleted region of Yp11.2 on the short arm of the Y-chromosome. The breakpoints reside within the TSPY array proximally and inverted repeat IR3 distally. AMELY locus is included in the deleted region. Reproduced from Yong et al. (2007).

Several studies have been carried out using Y-STRs, STS (Sequence-tagged site mapping) and TSPY copy estimation in order to analyse and characterise as accurate as possible the exact length of the deleted regions on the Y-chromosome short arm. A deletion of about 2.5 Mb on the pericentromic region of the short arm was reported in two Italian case studies (Lattanzi et al. 2005). Similarly, an approximately 2.56-Mb deletion in the Yp11.2 was reported in a Japanese case study (Kumagai et al. 2008). A 2.3-Mb deletion was reported in 5 Kathmandu (Nepal) phenotypically normal males (Cadenas et al. 2007). A longer deletion was also reported in Singapore and Malaysia populations with approximate length of 3-3.7 Mb (Yong et al. 2007). Interestingly, a Japanese case report has shown three separate deleted regions in the Yp11.2 with estimated lengths of 2.51 Mb, 25 Kb and 834 bases (Takayama et al. 2009).

The reported frequencies of AMELY drop-out (**Table 1.4**) clearly illustrates an association with people originating from the Indian subcontinent (Yong et al. 2007). Nevertheless, the variation in the characteristics of the reported deletions (length and pattern) suggests an independent occurrence of deletions through independent evolutionary events among different ethnic groups (Mitchell et al. 2006; Kumagai et al. 2008).

The highest frequencies of the AMELY drop-out were reported among the haplogroup J2 Y-chromosome lineages. Nonetheless, fewer AMELY null cases were also reported in the Y-chromosome haplogroups F* and D*. Therefore, these findings suggest that the deletion events, despite showing relatively similar patterns, have occurred independently at least three times throughout the human evolution and arguing against a common ancestry (Kashyap et al. 2006; Cadenas et al. 2007; Chang et al. 2007; Yong et al. 2007).

The reported variation in the frequency of AMELY deletion among different Y-chromosome haplogroups (with the highest frequency among J2 lineages) was suggested to be due to the difference in the age of the deletion event, and hence the older age of the deletion event correlates to the higher frequency (Yong et al. 2007).

**Table 1.4: Summary of percentages of the AMELY drop-out reported cases in different populations' studies.** The data did not include reported cases of AMELY drop-out resulting from point mutations in the primer binding sites. Reproduced from Yong et al. (2007).

| Country | Population | No. of studied individuals [Ref] | Frequency (%) |
|---|---|---|---|
| Sri Lanka | Sri Lankan | 24[a] | 8.3 |
| India | Indian | 270[b] | 1.9 |
| India (whole) | Indian (caste and tribes) | 4,257[c] | 0.23 |
| Nepal | Nepalese | 77[d] | 6.5 |
| | Nepalese | 769[e] | 1.2 |
| Austria | Austrian | 28,182[f] | 0.018 |
| Italy | Italian | 13,000[g] | 0.008 |
| Spain | Spanish | 768[h] | 0.13 |
| Israel | | 96[i] | 1.0 |
| Australia | Mixed | 109,000[j] | 0.02 |
| Malaysia | Indian | 315[k, l] | 3.2 |
| | Malay | 334[k, l] | 0.6 |
| | Chinese | 331[k, l] | 0 |
| Singapore | Indian | 175[l] | 1.76 |
| | Malay | 182[l] | 0.6 |
| | Chinese | 210[l] | 0 |
| Japan | | 500[m] | 0.2 |

[a] (Santos et al. 1998), [b] (Thangaraj et al. 2002), [c] (Kashyap et al. 2006) [d] (Cadenas et al. 2007) [e] (Parkin et al. 2007), [f] (Steinlechner et al. 2002), [g] (Lattanzi et al. 2005), [h] (Bosch et al. 2002), [i] (Michael and Brauner 2004), [j] (Mitchell et al. 2006) [k] (Steemers et al. 2006), [l] (Yong et al. 2007), [m] (Takayama et al. 2009).

Owing to the extreme importance of accurate gender determination in many crucial situations as in rape cases, mass disasters and others, it is of utmost necessity to include other sexing markers for gender determination particularly in certain regional populations (Yong et al. 2007).

## Technical issues

## 1.8 MtDNA sample sources

Mitochondrial DNA is well known to be very resistant to extreme environmental conditions, and as a consequence, it has an utmost importance in many forensic applications, particularly in such circumstances where the nuclear DNA is limited or degraded (Bender et al. 2000; Cruz et al. 2004). Dealing with degraded DNA samples, as in cases of mass disasters, is always considered as a challenge due to the absolute difficulty in obtaining useful information via typing nuclear markers, however it is not always the case for mtDNA markers (Palo 2007; Marjanović 2007 ).

Due to its presence in high copy number per somatic cell, mitochondrial DNA is easily recoverable from any kind of biological samples like blood, saliva, and semen. Specimens such as skin, hairs (as little as 1-2 cm of a single shaft) and bones are always good sources for mtDNA (Graffy and Foran 2005; Palo 2007; Marjanović 2007 ).

MtDNA has also been extracted from teeth and used for forensic DNA identification purposes, this is particularly important as nuclear DNA is usually degraded in teeth specimens (Sampietro et al. 2006; Tsutsumi et al. 2006).

Interestingly, fingerprints have been shown to provide a convenient source of mtDNA. Thus, mtDNA markers retrieved from fingerprints can be used for forensic identification purposes. This is particularly important where the prints are smeared or partial in a way that hinder the potential for identification using the traditional fingerprint analysis (Andreasson et al. 2006).

## 1.9 Different methods of DNA extraction

There are three main techniques for DNA extraction that can be used in forensic DNA laboratories:-

### 1.9.1 Organic Extraction (Phenol-chloroform extraction).

It involves serial addition of several chemicals to break open the cell walls and to break down the proteins to be followed by separation of DNA from proteins. It is the most effective method used to obtain high molecular weight DNA, which is essential for Restriction Fragment Length Polymorphism (RFLP) analysis. It has been in use for the longest period of time. However, this method of extraction is time consuming and

involves the use of multiple hazardous chemicals. In addition, it necessitates transfer of samples between different tubes that increases the liability of contamination (Butler 2001).

### 1.9.2 Chelex Extraction

It has become a popular alternative method of DNA extraction among forensic scientists. It entails the use of a chelating-resin suspension that can be added directly to the samples, followed by two steps of heating at different temperatures. Then centrifugation is performed followed by direct application of the supernatant into the PCR reaction. Chelex acts by binding to the polyvalent metal ions (such as magnesium ions ($Mg^{2+}$)) required as co factors for nucleases to work. Thus, removing $Mg^{2+}$ ions inactivates nucleases, and hence protects the DNA molecules. This method of extraction is simple and more rapid than the organic extraction one. In addition, it involves fewer steps, and thus reduces the potential for contamination compared to the organic method. However, it involves denaturation of the DNA, and thus produces single-stranded DNA which is only useful for PCR-based analysis (Kain and Lanar 1991; Walsh 1991; Sweet et al. 1996; Butler 2001).

### 1.9.3 FTA ™ Cards

FTA (Flinders Technology Associates) Cards (Whatman) is a relatively novel method designed to simplify the collection, shipment, archiving and purification of nucleic acids from a wide variety of biological sources.

#### *1.9.3.1 Definition of FTA cards*

They are absorbent cellulose-based cards, impregnated with a patented chemical formula that lyses cell membranes and denatures proteins upon contact. DNA is physically entrapped, immobilized within the matrix and stabilized for storage at room temperature over a period of several years (Fujita and Kubo 2006).

### *1.9.3.2 Characteristics of FTA cards*

FTA cards are considered advantageous over the standard methods of DNA collection and purification in many aspects;

- FTA cards are considered unique in their ability to store DNA samples at room temperature and hence no need for a precious freezer space with no risk of freeze-thaw degradation of samples. Additionally, they do not require special storage conditions apart from keeping the cards away from light and humidity.

- They have the ability to protect DNA from nucleases, oxidation, and UV damage as well as microbial and fungal attacks. Thus, FTA cards play an important niche in preserving the quality of DNA samples, which in turn enable long term storage of samples over several years without the threat of DNA degradation (He et al. 2007).

- Infectious pathogens (viruses, bacteria, etc.) in samples applied to FTA cards are rendered inactive upon contact. Which subsequently guard against the risk of infections' transmission (Moscoso et al. 2004; Muthukrishnan et al. 2008).

- FTA cards can be loaded with minimal volumes of DNA-containing samples which is particularly important when dealing with precious limited volume forensic samples.

- The easy and simple way of applying FTA cards render them more suitable for self-collection and field sample collection than any other methods (Muthukrishnan et al. 2008). Additionally, FTA technology facilitates collection of samples from geographically isolated populations and gets them transported by mail for accurate molecular diagnosis (Guio et al. 2006)

- Use of FTA cards simply involves direct application of a sample spot on the card that is left to dry and then a small disc (1.2mm or 2-mm) is punched out of that card and subjected to serial wash processes using different solutions aiming to purify DNA from inhibitors. The PCR reaction mix is then added directly to the clean dried punch. Therefore, no hassle of contacting multiple hazardous chemicals (Smith and Burgoyne 2004).

- Processing of FTA cards does not necessitate transfer of samples between different tubes that in turn reduces the potential of contamination.
- Purifying the DNA from FTA cards does not involve its denaturation and hence it produces double-stranded DNA that can undergo various analyses.

#### *1.9.3.3 Different applications of FTA cards*

FTA cards have seen the light in the last few years and have been involved in many areas of research. That could be attributed to their enormous potential for storing a wide range of biological samples including blood (Mai et al. 2004; Sultan et al. 2009; Untoro et al. 2009), saliva (Beckett et al. 2008; Park et al. 2008; Johanson et al. 2009), sputum (Guio et al. 2006), semen (Fujita and Kubo 2006), faecal material (Jaravata et al. 2006; Nechvatal et al. 2008), nasal swab (Mai et al. 2004), plant material (Tsukaya 2004; Ndunguru et al. 2005; Owor et al. 2007), bacterial suspension (Moscoso et al. 2004; Rajendram et al. 2006), viral suspension (Brassard et al. 2009), yeast (Borman et al. 2006), cell culture (Muthukrishnan et al. 2008), Bone marrow and skin smears (Jamjoom and Sultan 2009), solid tissue impression (Moscoso et al. 2007) as well as tumour cell suspension (Galaal et al. 2007) and many other kinds of biological samples (Adams et al. 2006; Moscoso et al. 2006; Perozo et al. 2006; Purvis et al. 2006; Sudhakaran et al. 2009).

FTA cards have recently gained increasing popularity in the domain of forensic investigations and herein mentioned few examples of studies where FTA cards have played a pivotal role:

- FTA cards were used for collection of blood samples from 638 individuals (representing the three ethnic groups of the Malaysian population) that have been analysed to determine the allele frequencies of the 15 STR loci in the AmpFISTR Identification kit (Seah et al. 2003). Similarly, the allele frequency of CODIS 13 was determined through analysis of 402 blood samples (from western and eastern Indonesia) collected on FTA cards (Untoro et al. 2009).

- Following the use of buccal cells collected from long term Betel quid chewers, FTA cards have shown promising results for STR markers typing with success rate of 89 % as compared to 7 % using the standard organic extraction procedures (Salvador and De Ungria 2003).

- Several studies have recommended FTA cards as a convenient method for DNA storage and purification to be used for WGA (Whole Genome Amplification), SNP typing and PCR-based genotyping (Lema et al. 2006; Beckett et al. 2008; Tomas et al. 2008).

- Using FTA technology, Fujita and Kubo (2006) were able to develop a method for extraction of sperm DNA from mixed body fluids containing semen.

- Based on non-invasive characteristics, Beckett et al. (2008) have suggested FTA cards as a convenient tool for collection of buccal samples from paediatric populations for molecular epidemiologic studies.

Apart from the forensic applications, FTA cards were adopted for the detection of viral infections as follow:

- Porcine reproductive and respiratory syndrome virus was detected in pig whole blood on FTA cards (Kobayashi et al. 2003).
- Foot and mouth disease virus was retrieved from tongue epithelial cells from infected animals applied on FTA cards (Muthukrishnan et al. 2008).
- Owor et al (2007) were able to clone complete Maize streak virus genome using FTA cards for DNA purification from leaf samples of infected maize plant.
- Infectious bursal disease virus in poulty was detected in bursal samples applied on FTA cards (Moscoso et al. 2006; Purvis et al. 2006).

Not only viruses but also bacteria and parasites were detected by analysis of various samples on FTA cards. Listed below are examples of these:

- *Mycobacterium tuberculosis* was detected by PCR analysis of sputum specimens on FTA cards as a tool for molecular testing of TB (Guio et al. 2006).
- Bovine manure on FTA cards was successfully used for detection of *Mycobacterium avium* (Jaravata et al. 2006).
- *Trypanosoma brucei* were identified in blood samples collected from human, sheep and mouse on FTA cards (Becker et al. 2004; Gonzales et al. 2006; Morrison et al. 2007).
- Using FTA technology, Jamjoom and Sultan (2009) were able to diagnose leishmaniasis in bone marrow samples and skin smears.

- Blood samples on FTA cards were successfully used for diagnosis of Maralria (Sultan et al. 2009).

## 1.10 RFLP Technology

### 1.10.1 Definition of RFLP

RFLP (Restriction Fragment Length Polymorphism) was one of the first techniques to be widely used in the forensic DNA analysis as well as medical genetics. In the early 1980s, the RFLP technique was adapted by Sir Alec Jeffreys for the purpose of DNA profiling through isolation of minisatellites (VNTRs). And hence it has been established as a standard DNA identification protocol that enables differentiation between samples of different sources (Jeffreys and May 2003).

It is based on the use of a set of restriction endonucleases that cleave the double-stranded DNA at specific sequence patterns which are known as cut sites or recognition sites. Thereafter, RFLP analysis results in digestion of the double-stranded DNA into fragments of varying lengths which are separated according to their molecular weights using gel electrophoresis. The analysis is then completed by performing Southern blotting and hybridization of the fragments to radioactive labelled DNA complementary probes to be visualized by X-ray exposure (Botstein et al. 1980; Rudin and Inman 2002b). Based upon the presence or absence of these recognition sites resulting from the base substitution events, variations in the fragment lengths may generate and hence may help in the differentiation between DNA samples derived from different sources (Habano et al. 1999).

### 1.10.2 Characteristics of RFLP technology

RFLP analysis is characterised by a high power of discrimination (~1 in 1 billion using 6 loci) due both to the hypervariability of each locus and the ability to look at many loci. In addition, it has the capability of handling sample mixtures using single-locus probes. Nevertheless, it has many drawbacks compared to the other available techniques including the time factor since it is a long tedious process. In addition, RFLP analysis requires a relatively large volume of high quality (high molecular weight) double-stranded DNA to start with, which is not always available in forensic investigations that are occasionally confronted by degraded DNA. Therefore, RFLP is not the most preferred technique by the forensic DNA laboratories nowadays, and subsequently it has

almost been usurped by the more recent more efficient PCR-based DNA analysis techniques (Rudin and Inman 2002a).

### 1.10.3 Applications of RFLP analysis

RFLP has played an important role in forensic as well as medical genetics. DNA profiling using RFLP analysis was the main tool for DNA-based human identification system that has been successfully used to aid forensic investigations in both criminal cases and disputed paternity. Furthermore, RFLP markers have been used for construction of the first-generation genetic linkage maps of human chromosomes and performing the genetic linkage analysis by determination of genetic loci responsible for genetic diseases such as cystic fibrosis, multiple endocrine neoplasia type I, familial polyposis coli and familial breast cancer (Botstein et al. 1980; Habano et al. 1999).

However, the more recent PCR-based analyses have almost replaced the traditional RFLP typing system. PCR-RFLP analysis, which removes the need for southern blotting and hybridization, has become a very popular technique in many laboratories and is currently adapted for several purposes such as screening of the human genome for mutations associated with resistance to viral infections (Krings et al. 1999), species identification of different animals and micro-organisms (Malathi et al. 2009; Simon et al. 2010), prediction of the treatment outcome of certain diseases (Kazibwe et al. 2009).

Several polymorphic sites in the coding and non-coding regions of the human mitochondrial DNA were screened for using PCR-RFLP technique for the purpose of haplotype analysis and haplogroup assignment among various populations as a tool for studying human history and population dynamics (Andelinovic et al. 2005; Park et al. 2008; Vidrova et al. 2008).

## 1.11 Aims and objectives of the project

The presented project was mainly concerned with mitochondrial molecular analysis of the Egyptian population.

The primary aims included:

1. Evaluation of the currently available techniques, namely FTA cards, for mitochondrial DNA testing and assessment of their efficiency for collection, storage and transportation of the DNA containing samples which would further aid the development of forensic investigations.
2. Assessment of the nucleotide diversity and heterogeneity among the Egyptian population via analysis of the mitochondrial hypervariable regions (HVI and HVII) sequence genotypes. That would lead to enrichment of the mtDNA database, which by the time that project was launched was lacking sufficient information about the molecular structure of this population group.
3. Assignment of the Egyptian mitochondrial haplotypes to known haplogroups. The frequency of different haplogroups was to be concluded and correlated with the published data, particularly of those haplogroups that were thought to show certain geographic localisation.
4. Screening of the coding region of the mtDNA among the Egyptian samples for sets of mutations that were thought to be definitive markers for mitochondrial DNA haplogroups. For that, a novel strategy for PCR-RFLP analysis of the mtDNA was to be developed.
5. The data obtained via hypervariable regions analysis was to be correlated with that obtained by PCR-RFLP strategy. Efficiency of both analyses was to be evaluated for more accurate assignment of the mitochondrial DNA to known haplogroups.

It was hoped that this investigation would yield important insights into the molecular diversity of the Egyptian population that would further enhance our understanding of the original structure of this population group. Furthermore, it was anticipated that this study would enrich the mtDNA database that would further aid the development of the global mitochondrial phylogenetic network.

In addition, this project was concerned with Y-chromosome analysis through screening of 131 Egyptian males for amelogenin Y gene deletion, which was previously reported at very low frequencies in certain populations as a component of a large interstitial deletion on the short arm of the Y-chromosome (Yp11.2).

# 2. MATERIALS AND METHODS

## 2.1 Use of FTA cards for collection and processing of blood samples

### 2.1.1 Collection and storage of samples

- Whole blood samples were collected from 261 adult Egyptians (131 males and 130 females) with consent. Ethical permission for this study was granted by Ethics Committee, University of Alexandria, Egypt (Appendix I)[a].

- Samples were mainly collected from those individuals belonging to a lower economic class in the community who used to receive free medical care at the main university hospital. This population group was mainly chosen due to the fact that they have less potential for travelling abroad with minimal chances of admixture and thus should be a good representative model of the residential Egyptian population.

- A simple questionnaire was applied where all the participants confirmed being born to Egyptian mothers.

- Based on our previous medical knowledge, samples were collected by venipuncture and each individual sample was split into two; a small volume (~150-200µl) was applied as fresh blood directly to an FTA (Flinders Technology Associates) micro-card (WB120310, Whatman) (**Figure 2.1**) using a micropipette and left to dry at room temperature and the remaining volume of the sample (5-ml) was placed in a vacuum blood collection tube containing EDTA and stored at -20$^{o}$C for one year before application to an FTA micro-card.

- FTA cards were stored and transported from Egypt to the UK at room temperature.

[a] The letter of approval provided by the Ethics Committee has included the provisional title of the project which was later on modified to the current title.

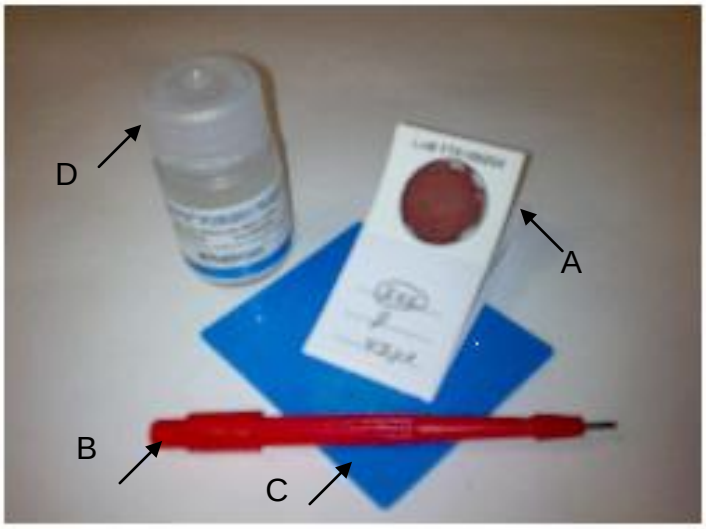



**Figure 2.1: Whatman FTA kit. A** is a whole blood sample on FTA micro-card (Cat.No.WB120310), **B** is a Harris Uni-Core Micro-Punch with 1.2 mm steel tip (Cat. No. WB1000028), **C** is a cutting mat, and **D** is the FTA Purification Reagent.

### 2.1.2 Processing of blood samples on FTA cards for DNA purification

Processing of blood stained FTA micro-cards involved three main steps:-

1- A small disc was manually punched out of the card, using a Harris Uni-Core Micro-Punch fitted with 1.2 mm steel tip, and placed in a sterile 500µl PCR microcentrifuge tube. To avoid cross contamination between samples, discs from non-stained areas were punched out of the cards in between different samples.

2- The discs were then washed in the PCR tube according to the manufacturer's instructions (Whatman) in order to remove any non-DNA material that might inhibit the PCR reaction. Variations on the recommended washing protocol were performed to maximise the removal of inhibitors in the blood samples (**Table 2.1**). The waste wash was discarded after each washing time. Washes were continued until the discs were blank and then discs were transferred to fresh sterile 500µl PCR tubes.

3- Discs were then allowed to dry either at room temperature for 1 hour or at 56$^{o}$C for 10 minutes, according to Whatman's recommendations. Finally, the dried discs were transferred to fresh sterile 500µl PCR tubes to which the PCR master mix was then added.

Whenever processing of blood samples on FTA cards was performed, a non-stained disc was included and taken through the whole process of washing and drying in a separate sterile eppendorf tube, to be used as a negative control in the subsequent analysis.

**Table 2.1: Summary of different methods used for washing blood stained discs.**

| Method | First Wash | | | Intermediate Wash | | | Final wash | | |
|---|---|---|---|---|---|---|---|---|---|
| | Reagent (volume) | Duration (min) | Number of times | Reagent (volume) | Duration (min) | Number of times | Reagent (volume) | Duration (min) | Number of times |
| 1 | FTA Purification Reagent[(a)] (200µl) | 5 | 3 | / | / | / | $TE^{-1}$ Buffer[(b)] (200 µl) | 5 | 2 |
| 2 | " | 10 | 4 | / | / | / | " | 10 | 4 |
| 3 | " | " | " | 0.3 M $Na_2CO_3$[(c)] (200µl) | 5 | 1 | " | " | " |
| 4 | " | " | " | 5% SDS[(d)] (500µl) | " | " | " | " | " |

[(a)] WB120204, Whatman.
[(b)] 10mM Tris, 0.1mM EDTA, pH 8.0.
[(c)] Sodium Carbonate, Fisher Scientific.
[(d)] Sodium Dodecyl Sulfate, Sigma-Aldrich.

Preliminary experiments were performed using both 2-mm and 1.2-mm discs. However comparative results revealed no significant difference between both of them, therefore 1.2-mm disc was considered as the optimum disc size to be used all through the current study.

## 2.2 Analysis of the mitochondrial hypervariable regions (HVI and HVII)

### 2.2.1 Polymerase Chain Reaction (PCR)

Following processing of the blood samples on FTA micro-cards, dried discs were then subjected to PCR amplification of mitochondrial hypervariable regions (HVI and HVII) either in separate reactions or simultaneously. The experiment was repeated in an identical manner for all the Egyptian blood samples (n=261).

#### *2.2.1.1 Design of primers*

Two sets of primer pairs have been used to amplify mitochondrial hypervariable regions HVI and HVII between nucleotide positions 16024-16365 and 73-340 respectively (**Table 2.2**):

- A set of short primers (~20mer) previously designed by Gabriel et al. (2003) was used in conjunction with the conventional thermal cyclic conditions.
- A second set of Long (~30mer) oligonucleotides was designed and synthesized during the current study to be used with fast PCR conditions.

**Table 2.2: Details of the primers for PCR amplification of the mitochondrial hypervariable regions (HVI and HVII).** Sequences, length, nucleotide position, GC content and melting temperature of primers were mentioned. Sequences of short primers were obtained from Gabriel et al. (2003). Sequences of long primers were designed based on the CRS of the human mitochondrial genome (accession number NC_012920).

| Name | Primer sequence (5′ to 3′) | Length (bases) | Nucleotide position [(a)] | GC content (%) | Tm ($^{o}$C) |
|---|---|---|---|---|---|
| Short primers | | | | | |
| HVI F Short | CTCCACCATTAGCACCCAA | 19 | 15975-15993 | 52.6 | 63.9 |
| HVI R Short | ATTTCACGGAGGATGGTG | 18 | 16418-16401 | 50.0 | 61.5 |
| HVII F Short | CACCCTATTAACCACTCACG | 20 | 15-34 | 50.0 | 60.0 |
| HVII R Short | CTGTTAAAAGTGCATACCGC | 20 | 429-410 | 45.0 | 60.2 |
| Long primers | | | | | |
| HVI F Long | AAGTCTTTAACTCCACCATTAGCACCCAAAGC | 32 | 15965-15996 | 43.8 | 72.9 |
| HVI R Long | GGATATTGATTTCACGGAGGATGGTGGTC | 29 | 16426-16398 | 48.3 | 73.6 |
| HVII F Long | GGTCTATCACCCTATTAACCACTCACGGGAG | 31 | 8-38 | 51.6 | 72.9 |
| HVII R Long | GGGTGACTGTTAAAAGTGCATACCGCCAAAAG | 32 | 435-404 | 46.9 | 73.7 |

[(a)] Nucleotide positions according to the Cambridge Reference Sequence.

Combinations of primers were used (**Table 2.3**) in order to develop an appropriate protocol for the analysis of mitochondrial hypervariable regions that would be consistently applied for all the Egyptian blood samples:

- HVI F and HVI R primers were used for PCR amplification of hypervariable region I.
- HVII F and HVII R primers were used for amplification of hypervariable region II.
- All four primers; HVI F, HVI R, HVII F, HVII R, were used together for multiplex amplification of the two hypervariable regions simultaneously in the same reaction as two separate fragments.
- HVI F and HVII R primers were used for amplification of both regions together as one fragment.

**Table 2.3: Combinations of primers used for PCR amplification of the mitochondrial hypervariable regions.** The expected PCR product size of each combination is indicated.

| Primers | Amplification of | PCR product size (bp) |
|---|---|---|
| HVI F and HVI R (Short) | HVI | 444 |
| HVII F and HVII R (Short) | HVII | 415 |
| HVI F and HVI R (Long) | HVI | 462 |
| HVII F and HVII R (Long) | HVII | 428 |
| HVI F, HVI R, HVII F and HVII R (Short) | Multiplex of HVI and HVII | 444 and 415 |
| HVI F, HVI R, HVII F, HVII R (Long) | " " " | 462 and 428 |
| HVI F and HVII R (Short) | HVI and HVII as one fragment | 1024 |
| HVI F and HVII R (Long) | " " " | 1039 |

### *2.2.1.2 Constituents of PCR master mix*

The reaction mixtures were prepared in sterile nuclease-free 500µl PCR tubes and contained the following constituents for each reaction volume of 50µl per sample:

- 5x GoTaq colourless Flexi buffer (Promega) 10.0µl
- $MgCl_2$ (25mM; Promega) 5.0µl
- DeoxyNTP mixture (10mM; Promega) 2.0µl
- Forward primer x (100pmol/µl) 0.5µl
- Reverse primer y (100pmol/µl) 0.5µl
- GoTaq DNA polymerase (5u/µl; Promega) 0.25µl
- Nuclease free water to final volume of 50µl 31.75µl

GoTaq Flexi buffer and $MgCl_2$ solution were vortexed briefly following complete thawing. The listed components were added to a sterile thin-walled 500µl eppendorf tube contained the processed dried disc.

Positive and negative controls were also included. The positive control was 1.0µl of human placental DNA (10ng/µl; Sigma) and the negative control was a processed dried blank disc. All the other constituents of the PCR master mix were unchanged except for the nuclease free water which was readjusted in the positive control reaction.

Following the results of preliminary experiments, the constituents of the PCR master mix included a modification to reduce the effects of inhibitor(s) in blood that was achieved by adding acetylated Bovine Serum Albumin (BSA) (Promega).

Modified constituents for a 50µl reaction mix were as follows:

- 5x GoTaq colourless Flexi buffer (Promega) 10.0µl
- $MgCl_2$ (25mM; Promega) 5.0µl
- DeoxyNTP mixture (10mM; Promega) 2.0µl
- Bovine Serum Albumin (BSA) (10 mg/ml; Promega) 2.0µl
- Forward primer x (100pmol/µl) 0.5µl
- Reverse primer y (100pmol/µl) 0.5µl
- GoTaq DNA polymerase (5u/µl; Promega) 0.25µl
- Nuclease free water to final volume of 50µl 29.75µl

### *2.2.1.3 PCR Thermal Cyclic Conditions*

All PCR reactions were performed in a thermocycler (TECHNE GENUIS). Conventional cyclic conditions were used when short primers (~20mer) have been included in the PCR reaction mix as follows:

| | | | |
|---|---|---|---|
| Initial Denaturation | 94°C | 5 min | 1 cycle |
| Denaturation | 92°C | 45 sec | 30 cycles |
| Annealing | 59°C | 1 min | |
| Extension | 72°C | 1 min | |
| Final Extension | 72°C | 10 min | 1 cycle |
| Soak | 4°C | Hold | 1 cycle |

Optimised PCR thermal cyclic conditions were as follows;

| | | | |
|---|---|---|---|
| Initial Denaturation | 98°C | 30 sec | 1 cycle |
| Denaturation | 92°C | 5 sec | 30 cycles |
| Annealing/ Extension | 72°C | 30 sec | |
| Final Extension | 72°C | 1 min | 1 cycle |
| Soak | 4°C | Hold | 1 cycle |

A high initial denaturation temperature at 98°C has been used to ensure full denaturation of the initial genomic template in a short period of time (30 sec) as recommended by Sullivan et al. (2006) at Bio-Rad laboratories. Despite being specific recommendations for the hot-start polymerases, this modification was applicable to GoTaq DNA polymerase as well without any reduction in the efficiency of amplification.

### *2.2.1.4 Gel electrophoresis of the PCR products*

PCR products were analysed by agarose gel electrophoresis. Small volumes (5-10µl) of the PCR products were separated using 2% (w/v) agarose (Bioline) gels containing 1x SYBR Safe DNA gel stain (stock of 10,000x in DMSO, Invitrogen) in 1x TAE buffer (40mM Tris-base, 20mM glacial acetic acid, and 5mM EDTA; pH 8). PCR products were mixed with 6x loading dye (0.25% (w/v) bromophenol blue (Sigma-Aldrich), 40% (w/v) sucrose (Fisher Scientific), 4ml (v/v) TE buffer and 10.8ml of 0.1% (v/v) DEPC-treated water). A 100bp DNA ladder (Promega) (5µl ladder + 1µl loading dye) was used as a molecular marker to monitor the migration of samples and determine the size of

different PCR products. Gels were run at 80-100V for 45 min to 1 hour. The bands were then visualised using a UV transilluminator within the Syngene gel documentation system.

#### *2.2.1.5 Purification of the PCR products*

PCR products were purified to remove excess nucleotides and primers using Wizard® SV Gel and PCR Clean-Up Purification System (Promega) according to the manufacturer's instructions. [This system was based on the selective adsorption of DNA to Silica-gel membranes present at the bottom of the SV Minicolumn under controlled ionic conditions using Membrane Binding Solution, whilst all impurities were washed out using micro-centrifugation and the Membrane Wash Solution.] DNA was eluted in 50µl of nuclease-free water. Purified samples (3µl) were checked by agarose gel electrophoresis.

#### *2.2.1.6 Ethanol precipitation*

For those samples that have shown poor DNA yield after purification of the PCR products as evidenced by faint bands on the second gel, Ethanol precipitation was considered in order to increase the DNA concentration. For this, 2.5 vol. of 100% (v/v) ethanol and 1/10 vol. of 3M sodium acetate (Sigma-Aldrich) (pH 5.2) were added to each sample, after which the samples were vortexed and incubated at −20$^{o}$C for 2 hours. The tubes were then centrifuged at 13,000rpm at 4$^{o}$C for 20 min. The supernatant was discarded and 150µl of 75% (v/v) ethanol was added. The tubes were briefly vortexed and centrifuged at 13000rpm at 4$^{o}$C for a further 5min. The supernatant was removed and the pellets were air-dried. Pellets were resuspended in 10µl nuclease-free water.

### 2.2.2 Sequencing of the mitochondrial hypervariable regions (HVI and HVII)

Sequencing reactions were performed and analysed at the Functional Genomics and Proteomics Laboratories at University of Birmingham. Samples were prepared for sequencing by adding 5µl of the purified PCR products and 4µl of the sequencing primer (at a final concentration of 0.8pmol/µl). Both forward and reverse short primers (**Table 2.2**) have been used in two separate sequencing reactions for each sample.

### 2.2.3 Data analysis

#### *2.2.3.1 Mitochondrial haplotype analysis*

Nucleotide diversity and heterogeneity among the Egyptian population was assessed via analysis of the sequencing results of HVI and HVII regions of the mitochondrial genome. The obtained nucleotide sequences were aligned against the Cambridge Reference Sequence (CRS) (accession number NC_012920) using CLUSTALW analysis tool of Biology Workbench at (www5). Instances of point heteroplasmy were evaluated on the proportion of minor versus major component. Variations from the CRS represented mutations, nucleotide positions of these mutations were determined and subsequently mitochondrial haplotypes were concluded and correlated to the published databases.

#### *2.2.3.2 Mitochondrial haplogroup assignment*

Assignment of the Egyptian mitochondrial hypervariable region haplotypes to the well established mitochondrial haplogroups was performed via correlation of the reported haplotypes to the published databases as well as the electronic databases including the global human mitochondrial DNA phylogenetic tree (www6). This global phylogenetic tree is regularly updated to maintain high resolution of the phylogenetic network of the mitochondrial DNA (Van Oven and Kayser 2009).

Following analysis of a considerable representative number of complete Egyptian HVI and HVII sequences between nucleotide positions 16024-16365 and 73-340 respectively, and depending on the usefulness of each hypervariable region solely in haplogroup assignment, a modification was considered which included analysis of HVI (between nucleotide position 16024-16365) together with nucleotide position 73 in HVII.

#### *2.2.3.3 Statistical analysis*

The overall number of nucleotide variations in the mitochondrial HVI region was determined and those occurring at a high frequency were identified. The total number of different haplotypes was determined, the frequency of each haplotype was estimated and the most commonly observed haplotypes were identified. The probability of randomly selecting two unrelated individuals with the same mitochondrial haplotype (random match probability, RMP) was calculated using the equation $p=\sum x^2$, where $x$ is

the frequency of each mtDNA haplotype (Stoneking et al. 1991). Genetic diversity of the mtDNA hypervariable regions in the studied population group was calculated using the algorithm $h= n(1-\sum x^2 )/(n-1)$, where $n$ is the sample size and $x$ is the frequency of each mtDNA haplotype (Nei and Tajima 1981).

## 2.3 PCR-RFLP screening of the whole mitochondrial genome among the Egyptian population

To further aid in the mtDNA analysis and haplogroup assignment of the Egyptian population, the mitochondrial genome was screened for 26 RFLP (Restriction Fragment Length Polymorphism) markers that were described as haplogroup specific. For this, mitochondrial genome was amplified by PCR as nine non-overlapping fragments, which were subsequently digested using restriction endonucleases. Amplification of the whole mitochondrial genome (~16Kb) as one PCR product was considered at the beginning of that part of the study, however, it was shown to be inconvenient due to the extreme difficulty in the analysis of digestion patterns produced by different restriction enzymes particularly *AluI* (64 cut sites) and *DdeI* (72 cut sites) endonucleases using the standard gel electrophoresis techniques. On the other hand, amplification of 26 regions of the mitochondrial genome in the entire Egyptian population sample (n=261) to screen for the determined 26 RFLP markers was expected to be tedious and time consuming. Therefore, a decision was made to minimize the number of PCR amplicons as much as possible that would be further used in the FRLP analysis. Hence, the nine fragments protocol was designed to include the total number of RFLP markers. Those fragments were subsequently digested using different endonucleases where the digestion patterns could be easily analysed using the standard gel electrophoresis techniques.

### 2.3.1 Design of primers for PCR amplification of the nine targeted regions of the mitochondrial genome

In order to achieve the target of the mitochondrial genome amplification, nine pairs of primers (**Table 2.4**) were designed to amplify nine non-overlapping fragments (**Figure 2.2**). Sequences of primers were based on the Cambridge Reference Sequence (CRS) of the human mitochondrial genome (accession number NC_012920).

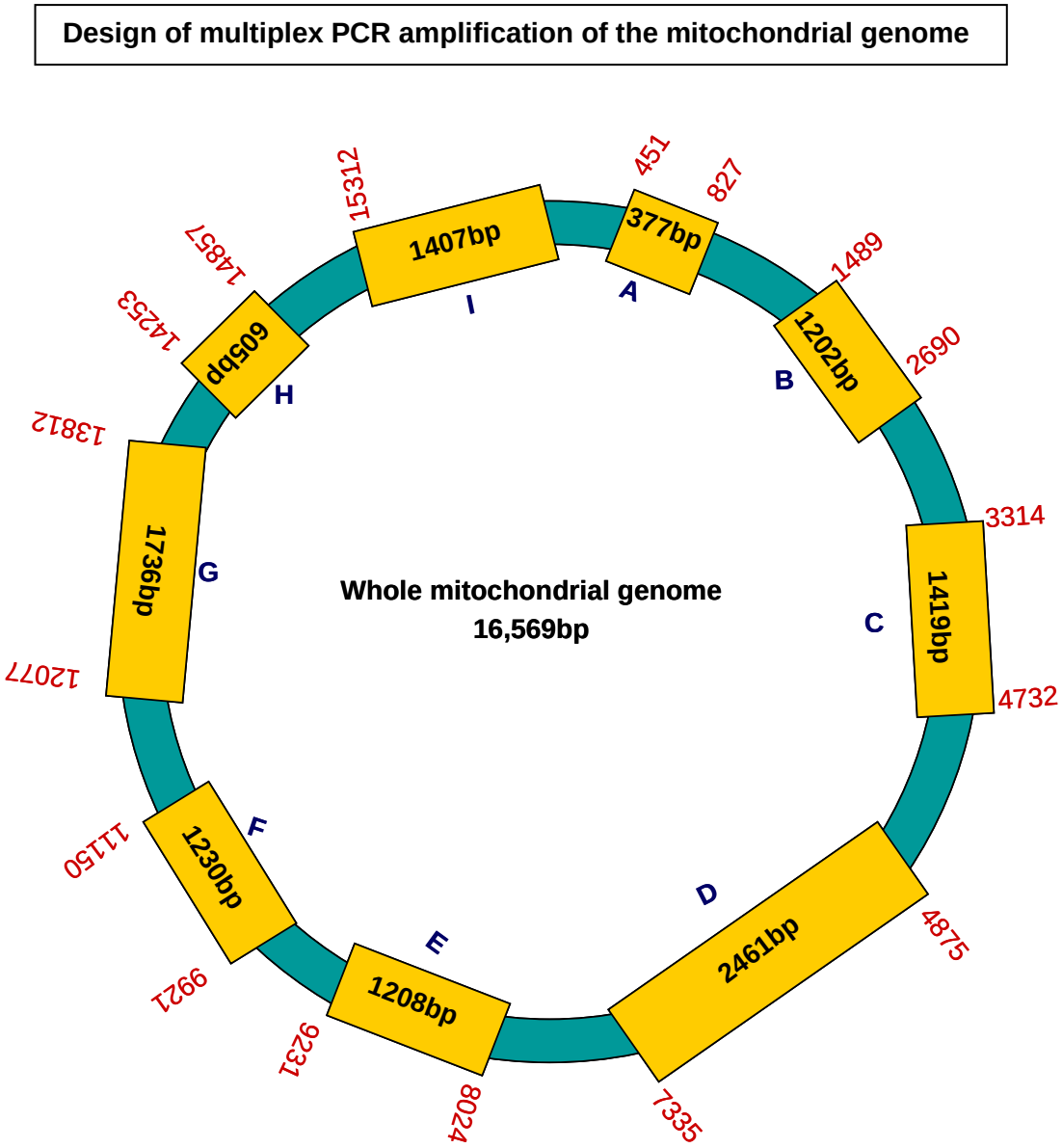



**Figure 2.2: Schematic representation of the nine targeted regions of the whole mitochondrial genome.** The nine targeted regions are represented by the yellow boxes with the length of each fragment inside. Fragments are designated by letters from A-I. The nucleotide positions of the different fragments are shown in red. The nine fragments altogether constituted 11645bp of the whole mitochondrial genome (16569bp). Each fragment was designed to have one or more of the RFLP definitive markers of the different mitochondrial DNA haplogroups.

**Table 2.4: Primers pairs used for PCR amplification of nine targeted regions of the mitochondrial genome.** All primers were based on the Cambridge Reference Sequence of the human mitochondrial genome (accession number NC_012920).

| Primer name | Primer sequence (5′ to 3′) | Primer position[(a)] | Length (bases) | GC content (%) | Tm (°C) | PCR product (Size) |
|---|---|---|---|---|---|---|
| Mt A F<br>Mt A R | TTATTTTCCCCTCCCACTCC<br>AATCACTGCTGTTTCCCGTG | 451-470<br>827-808 | 20<br>20 | 50<br>50 | 63.8<br>65.0 | Fragment A<br>(377bp) |
| Mt B F<br>Mt B R | CCGTCACCCTCCTCAAGTAT<br>GGCAGGTCAATTTCACTGGT | 1489-1508<br>2690-2671 | 20<br>20 | 55<br>50 | 62.9<br>63.9 | Fragment B<br>(1202bp) |
| Mt C F<br>Mt C R | ATGGCCAACCTCCTACTCCT<br>TGGTTATGGTTCATTGTCCG | 3314-3333<br>4732-4713 | 20<br>20 | 55<br>45 | 63.7<br>63.2 | Fragment C<br>(1419bp) |
| Mt D F<br>Mt D R | ACTAGCCCCCATCTCAATCA<br>AGCGAAGGCTTCTCAAATCA | 4875-4894<br>7335-7316 | 20<br>20 | 50<br>45 | 63.4<br>63.9 | Fragment D<br>(2461bp) |
| Mt E F<br>Mt E R | TGAAGCCCCCATTCGTATAA<br>ATAGGCATGTGATTGGTGGG | 8024-8043<br>9231-9212 | 20<br>20 | 45<br>50 | 64.0<br>64.5 | Fragment E<br>(1208bp) |
| Mt F F<br>Mt F R | CGCCTGATACTGGCATTTTG<br>CAAGGTGGGGATAAGTGTGG | 9921-9940<br>11150-11131 | 20<br>20 | 50<br>55 | 65.3<br>64.1 | Fragment F<br>(1230bp) |
| Mt G F<br>Mt G R | CACCTATCCCCCATTCTCCT<br>GCGAGGGCTGTGAGTTTTAG | 12077-12096<br>13812-13793 | 20<br>20 | 55<br>55 | 63.9<br>63.8 | Fragment G<br>(1736bp) |
| Mt H F<br>Mt H R | CCAATAGGATCCTCCCGAAT<br>AGGAGTGAGCCGAAGTTTCA | 14253-14272<br>14857-14838 | 20<br>20 | 50<br>50 | 63.8<br>63.9 | Fragment H<br>(605bp) |
| Mt I F<br>Mt I R | ATTGCAGCCCTAGCAGCACT<br>ATGAGGCAGGAATCAAAG | 15312-15331<br>149-132 | 20<br>18 | 55<br>44.4 | 66.1<br>57.7 | Fragment I<br>(1407bp) |

[(a)] Nucleotide positions according to the Cambridge Reference Sequence.

### 2.3.2 Multiplex PCR amplification of the nine fragments of the mitochondrial DNA

Amplification of 11645bp of the mitochondrial genome was performed in a two- stage PCR; the first stage was a multiplex amplification of all the nine non-overlapping fragments for 10 cycles using all nine pairs of primers, followed by the second-stage PCR in which the PCR product of the first reaction was used as a template. The second-stage PCR involved amplification of each fragment solely using the specific pair of primers (25 cycles). This protocol of two-stages PCR was designed aiming to reduce the initial volume of the DNA template that would be required for performing the nine separate PCR reactions, which would have a major impact on the forensic investigations where the amount of a sample may be an issue.

#### *2.3.2.1 PCR using Phusion High-Fidelity DNA polymerase*

The reaction mixtures for the multiplex amplification (first-stage PCR) were prepared in 500µl sterile nuclease-free PCR tubes (Eppendorf) as follows:

- 5x Phusion HF buffer (Finnzymes) 10.0µl
- $MgCl_2$ (25mM; Promega) 2.0µl
- DeoxyNTP mixture (10mM; Promega) 2.0µl
- Bovine Serum Albumin (BSA) (10 mg/ml; Promega) 2.0µl
- Nine forward primers (Mt A-I F) (100pmol/µl) 9.0µl (1.0µl of each primer)
- Nine reverse primers (Mt A-I R) (100pmol/µl) 9.0µl (1.0µl of each primer)
- Phusion High-Fidelity DNA polymerase (2u/µl; Finnzymes) 0.5µl
- Nuclease free water to final volume of 50µl

The listed components were then added to a sterile thin-walled 500µl eppendorf tube contained the processed dried disc (as previously described in section 2.1.2), ensuring that all solutions were kept on ice and keeping the reactions capped as often as possible. A negative control was run throughout the whole experiment using a processed non-stained disc replacing the sample disc.

Phusion DNA polymerase is a magnesium dependent enzyme and Phusion buffer contains $MgCl_2$ at a concentration of 7.5mM giving a final concentration of 1.5mM in the 50µl reaction. However, 1.5mM $MgCl_2$ was not the optimum concentration for

Phusion polymerase to work efficiently with these primers. Therefore, 2.0µl of 25mM $MgCl_2$ stock was added to the reaction which together with Phusion buffer produced the overall final concentration of $Mg^{2+}$ at 2.5mM.

Following the first-stage PCR for 10 cycles, the reaction mixtures for the second-stage PCR were prepared for each target region of the mitochondrial DNA separately in a sterile thin-walled 500µl PCR tube as follows:

- 5x Phusion HF buffer (Finnzymes) 10.0µl
- $MgCl_2$ (25mM; Promega) 2.0µl
- DeoxyNTP mixture (10mM; Promega) 2.0µl
- Bovine Serum Albumin (BSA) (10 mg/ml; Promega) 2.0µl
- Forward primer x (100pmol/µl) 1.0µl
- Reverse primer y (100pmol/µl) 1.0µl
- Multiplex (first-stage) PCR product (as template) 2.0µl
- Phusion High-Fidelity DNA polymerase (2u/µl; Finnzymes) 0.5µl
- Nuclease free water to final volume of 50µl

Negative control was run by replacing the multiplex PCR product as the template with nuclease-free water. All reactions were placed in a thermocycler (TECHNE GENIUS) for 25 cycles.

#### *2.3.2.2 PCR conditions for amplification of 11Kb of the human mitochondrial genome*

The thermal cyclic conditions for the *multiplex first-stage PCR* were as follows:

| | |
|---|---|
| Initial Denaturation (1 cycle) | at 98°C, for 30 sec |
| 10 cycles of | |
| Denaturation | at 98°C, for 10 sec |
| Annealing | at 64°C, for 20 sec |
| Extension | at 72°C, for 1 min 10 sec |
| 1 cycle of | |
| Final Extension | at 72°C, for 3 min |

The thermal cyclic conditions that were used for second-stage PCR amplification of the nine targeted regions of the mitochondrial genome were summarised in **Table 2.5**.

**Table 2.5: PCR conditions used for the second-stage PCR amplification of the nine targeted regions of the mitochondrial genome.** The duration of extension in each cycle was directly related to the expected size of the PCR product.

| Name of PCR product | Size of PCR product | Initial denaturation (1 cycle) | 25 cycles of | | | Final extension (1 cycle) |
|---|---|---|---|---|---|---|
| | | | Denaturation | Annealing | Extension | |
| A | 377bp | 98°C for 30 sec | 98°C for 10 sec | 64°C for 10 sec | 72°C for 15 sec | 72°C for 3 min |
| B<br>C<br>E<br>F<br>G<br>I | 1202bp<br>1419bp<br>1208bp<br>1230bp<br>1736bp<br>1407bp | " | " | 64°C for 20 sec | 72°C for 1 min and 10 sec | 72°C for 5 min |
| D | 2461bp | " | " | " | 72°C for 1 min and 30 sec | " |
| H | 605bp | " | " | " | 72°C for 40 sec | " |

#### *2.3.2.3 Gel electrophoresis of PCR fragments*

Small volumes (5µl) of all PCR products were separated using 1% and 2% (w/v) agarose (Bioline) gels containing 1x SYBR Safe DNA gel stain (Invitrogen) (in 1x TAE buffer), along with 1kb and 100bp DNA ladders (Promega), respectively. Electrophoresis was performed at 100V for 45 min. DNA fragments were visualised with an ultraviolet (UV) transilluminator and gel documentation system (Syngene).

### 2.3.3 Restriction digestion of the PCR products (RFLP analysis)

#### *2.3.3.1 Constituents and conditions of the various restriction digestion reactions*

All products of the second-stage PCR were digested using restriction endonucleases for detection of the RFLP markers that were specific for the different mitochondrial DNA haplogroups (Bailliet et al. 1994; Macaulay et al. 1999; Alves-Silva et al. 2000; Finnila et al. 2001b; Tolk et al. 2001). A set of 15 endonucleases (Promega and Biolabs) were used. Restriction digestions were prepared in 500µl sterile PCR tubes as follows:

| Restriction digestion: | |
|---|---|
| Second-stage PCR product | 14.0µl |
| Restriction enzyme 10x Buffer (**Table 2.6**) | 2.0µl |
| Acetylated BSA (10 µg/µl; Promega) | 0.2µl |
| Nuclease-free water | 3.4µl |
| Restriction enzyme (10u/µl) | 0.4µl |
| Final volume | 20.0µl |

The restriction enzyme was the last component to be added to the digestion reaction and then all the constituents were mixed by gentle pipetting up and down. Reactions were incubated at 37$^{o}$C for 2 hours except for *TaqI* restriction enzyme where the incubation was performed at 65$^{o}$C for 2 hours according to the optimum assay temperature given by Promega.

Digestion products were separated in agarose gels ranging from 1% − 3.5% containing SYBR Safe stain (in 1x TAE buffer), along with 1 Kb DNA ladder (Promega) or hyper ladder IV (Bioline) for 45min at 100V. DNA fragments were visualised using UV transilluminator and gel documentation system (Syngene).

**Table 2.6: Restriction enzymes and the corresponding buffers used in the RFLP analysis of the mitochondrial genome.**

| Restriction enzyme* | Buffer |
|---|---|
| *AccI* | 10x Buffer G (500mM Tris-HCl (pH 8.2) and 50mM $MgCl_2$) |
| *AluI* | 10x Buffer B (60mM Tris-HCl (pH 7.5), 500mM NaCl, 60mM $MgCl_2$ and 10mM DTT) |
| *AvaII* | 10x Buffer C (100mM Tris-HCl (pH 7.9), 500mM NaCl,100mM $MgCl_2$ and 10mM DTT) |
| *BamHI* | 10x Buffer E (60mM Tris-HCl (pH 7.5), 1M NaCl, 60mM $MgCl_2$ and 10mM DTT) |
| *DdeI* | 10x Buffer D (60mM Tris-HCl (pH 7.9), 1.5M NaCl,60mM $MgCl_2$ and 10mM DTT) |
| *EcoRI* | 10x Buffer H (900mM Tris-HCl (pH 7.5), 500mM NaCl and 100mM $MgCl_2$) |
| *HaeII* | 10x Buffer B |
| *HaeIII* | 10x Buffer C |
| *HincII* | 10x Buffer B |
| *HinfI* | 10x Buffer B |
| *HpaI* | 10x Buffer J (100mM Tris-HCl (pH 7.5), 500mM KCl, 70mM $MgCl_2$ and 10mM DTT) |
| *MboI* | 10x Buffer C |
| *MseI* | 1x NEBuffer 4 (50mM potassium acetate (pH7.9), 20mM Tris-acetate, 10mM magnesium acetate and 1mM DTT) |
| *NlaIII* | 1x NEBuffer 4 |
| *TaqI* | 10x Buffer E |

***** All restriction enzymes were purchased from Promega apart from *MseI* and *NlaIII* that were purchased from Biolabs and they all were in the concentration of 10u/µl.

#### *2.3.3.2 HaeIII restriction digestion of fragment A of the mitochondrial genome (nt451-827)*

The PCR product containing fragment A (377bp) which extended between mitochondrial nucleotide positions 451 and 827 was digested with *HaeIII* endonuclease. The Cambridge Reference Sequence of the mitochondrial DNA (accession number NC_012920) lacks the recognition sequence for *HaeIII* restriction enzyme (5′…GG▼CC…3′) between those two nucleotide positions. Whereas, mitochondrial DNA belonging to the native American mitochondrial DNA haplogroup A has an A663G mutation (Bailliet et al. 1994) which produces a novel *HaeIII* cut site at that position along fragment A (**Figure 2.3**).

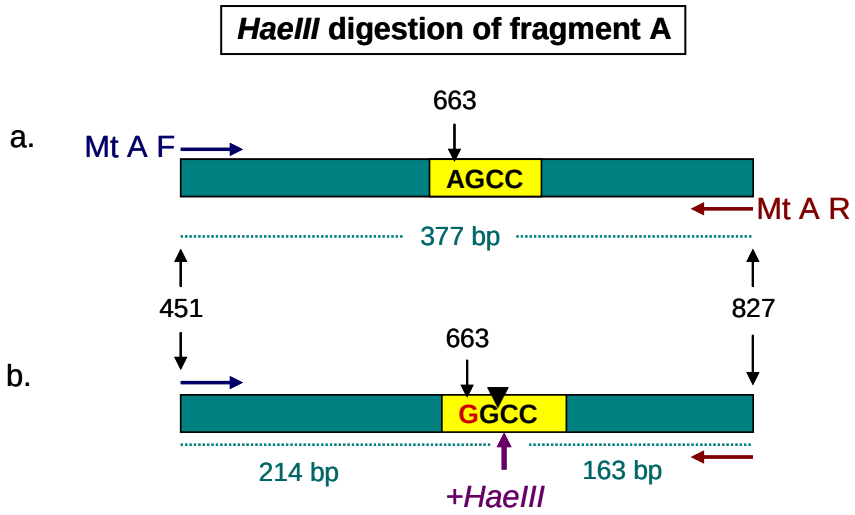



**Figure 2.3: *HaeIII* digestion of mitochondrial fragment (A) nt451−827.** (a) represents the Cambridge Reference Sequence of the mitochondrial DNA amplified using Mt A forward primer (blue arrow) and Mt A reverse primer (brown arrow) which lacks *HaeIII* recognition site and (b) mitochondrial sequence belonging to haplogroup A which has an A663G mutation shown in red (Bailliet et al. 1994) producing a *HaeIII* cut site and the digestion products are two fragments of sizes 214 and 163 bp. Sequences of (a) and (b) were obtained from NCBI database (accession number NC_012920) and MITOMAP database at (www7) respectively.

#### *2.3.3.3 DdeI and MboI restriction digestions of fragment B of the mitochondrial genome (nt1489-2690)*

Products of the PCR amplification of fragment B of the mitochondrial genome (1202bp) (nucleotide position 1489 − 2690) were digested with *DdeI* and *MboI* endonucleases in separate reactions. These restriction digestions aimed to identify two sets of mitochondrial DNA haplogroups via detection of certain RFLP markers.

The CRS of the mitochondrial genome between nucleotide positions 1489 and 2690 has six *DdeI* recognition sites (5′…C▼TNA_G…3′) according to Andrews et al.(1999) including one site at nucleotide position 1715 (**Figure 2.4.a**). The recognition sequence at that position is lost following a G1719A substitution which occurs in the mitochondrial DNA sequences belonging to haplogroups N and I and subclade X2 (Palanichamy et al. 2004; Kivisild et al. 2006; Tamm et al. 2007; Achilli et al. 2008), and hence loss of that cut site is diagnostic of these lineages (**Figure 2.4.b**).

The same region of the mitochondrial DNA (nt1489-2690) in the CRS lacks a recognition sequence for *MboI* restriction enzyme (5′…▼GATC…3′); (**Figure 2.4.c**). However, the African mitochondrial subclades L1b and L3e were described to have a T2352C mutation (Rando et al. 1998; Herrnstadt et al. 2002; Behar et al. 2008b; Malyarchuk et al. 2008) that produces an *MboI* cut site at position 2349 which is considered as a definitive marker for these two African lineages (**Figure 2.4.d**).

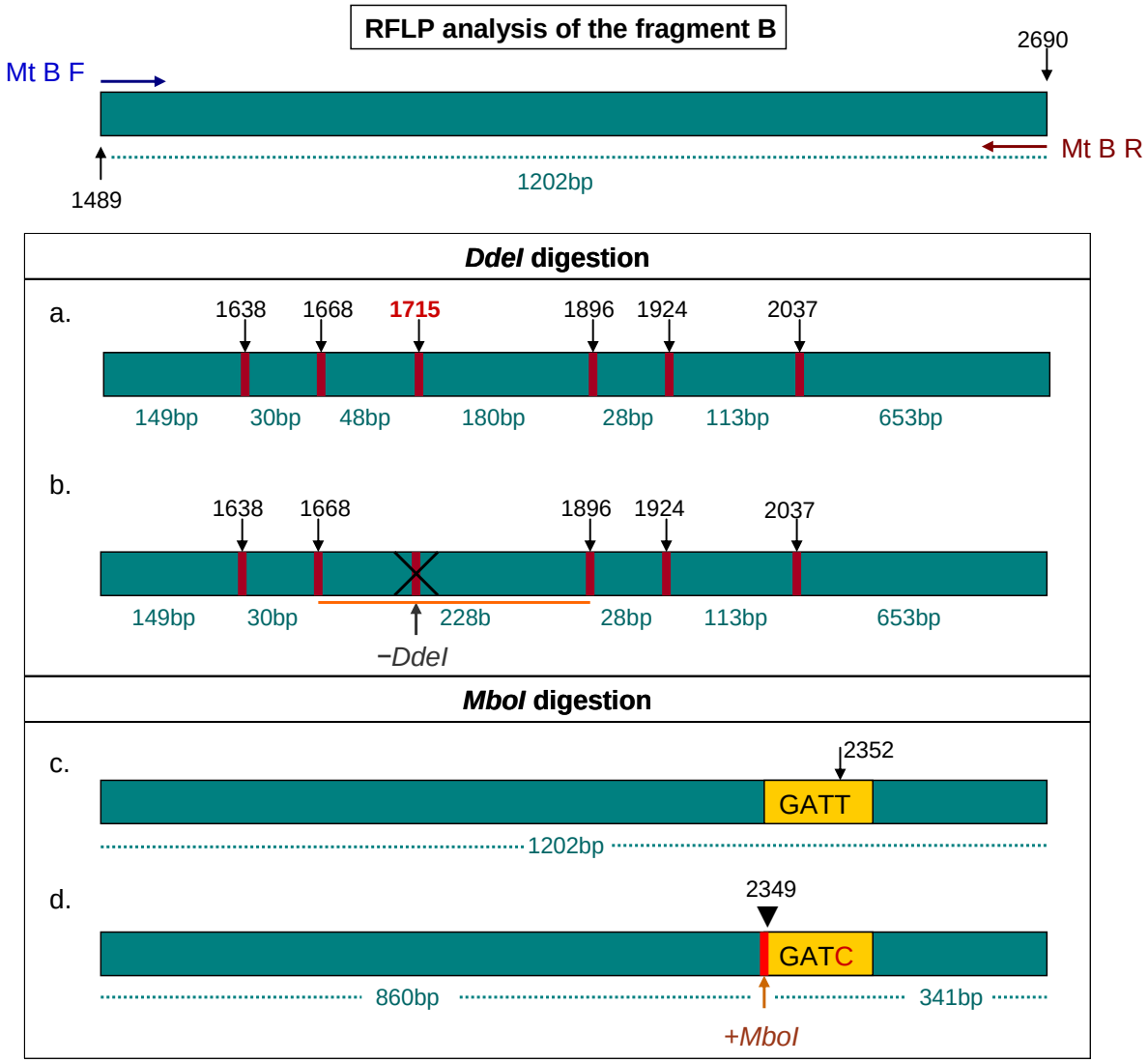



**Figure 2.4: Restriction digestion of the fragment B of the mitochondrial genome (nt1489 −2690.** Fragment B was amplified using Mt B forward primer (blue arrow) and Mt B reverse primer (brown arrow) and it involved 1202bp of the mitochondrial genome. Numbers above the fragment represented the nucleotide positions and numbers below were the expected lengths of the restriction digestion products. Longitudinal red bars represented the restriction cut sites. A plus sign (+) indicates a new cut site and a minus sign (−) indicates loss of a cut site. The top half of the figure shows the results of *DdeI* digestion of fragment B and the bottom one shows the products of *MboI* digestion of the same fragment. (a) and (c) represent the digestion results of the Cambridge Reference sequence of the mtDNA (accession number NC_012920), the length of each segment was mentioned below its corresponding bar. (a) shows the expected pattern of *DdeI* digestion of CRS with seven fragments of varying sizes ranging from 28bp up to

653bp. (b) represents mitochondrial DNA sequence belonging to hapolgroups N, I or X2 in which the *DdeI* cut site at position 1715 was lost resulting in merging the two segments of sizes 48 and 180bp into one segment of 228bp. (c) shows the absence of *MboI* recognition sequence along this region of the CRS. (d) represents an example of mitochondrial DNA sequence belonging to subclades L1b or L3e in which T2352C mutation occurred resulting in an *MboI* cut site gain at position 2349.

#### *2.3.3.4 HpaI and NlaIII restriction digestions of fragment C of the mitochondrial genome (nt3314-4732)*

PCR products containing fragment C (1419bp) (nucleotide position 3314 − 4732) were analysed with *HpaI* and *NlaIII* endonucleases separately. *HpaI* restriction digestion was carried out to screen for an African specific marker characterised by a C3594T mutation (Chen et al. 1995b; Salas et al. 2002) which produces an *HpaI* recognition sequence (5′…GTT▼AAC…3′) (**Figure 2.5**). This cut site gain is a diagnostic marker of the major African mitochondrial haplogroups L0, L1, L2, L5 and L6.

For *NlaIII* restriction digestion, the CRS of the mitochondrial DNA includes two internal cut sites (5′…CATG▼…3′) between nucleotide positions 3314 and 4732 according to the NCBI database (**Figure 2.5.c**). Mitochondrial DNA belonging to haplogroup JT (a descendant of the macro-haplogroup R, which originated in Southwest Euroasia) has a T4216C mutation (Torroni et al. 1996; Macaulay et al. 1999; Finnila et al. 2001a; Pike 2006; Logan 2009) which results in an additional recognition sequence for *NlaIII* endonuclease (**Figure 2.5**.**d**).

Additionally, *NlaIII* digestion of the fragment C of the mitochondrial genome helped in the identification of the mitochondrial haplogroup V (a sister to haplogroup H, originated from the haplogroup HV0a), in which an G4580A mutation (Torroni et al. 2001; Achilli et al. 2004; Palanichamy et al. 2004; Torroni et al. 2006) causes loss of *NlaIII* recognition sequence (**Figure 2.5.e**).

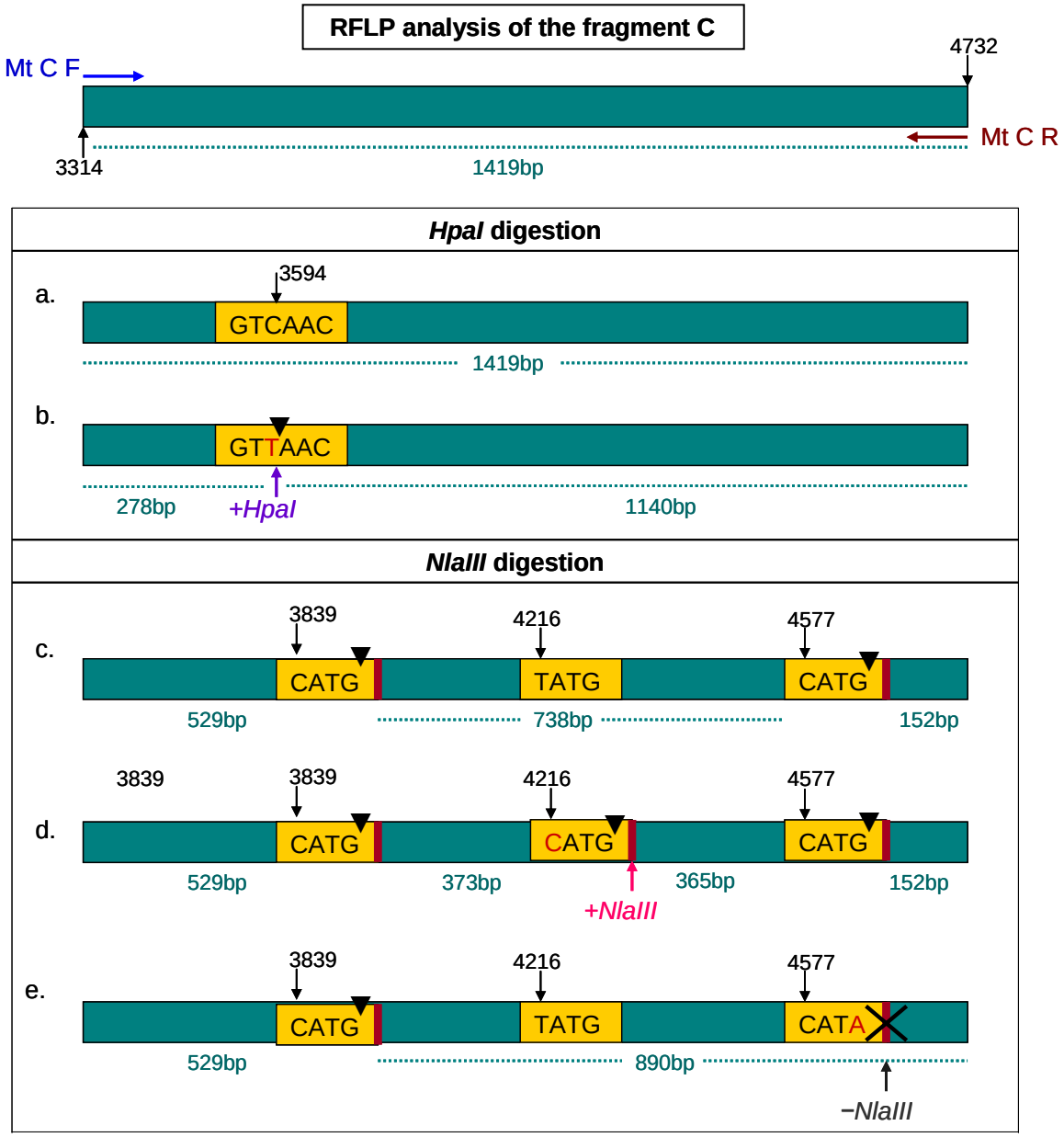



**Figure 2.5: Restriction digestion analysis of the fragment C of the mitochondrial genome (nt3314 and 4472).** Fragment C was the product of the PCR amplification of mtDNA using Mt C forward primer (blue arrow) and Mt C reverse primer (brown arrow) and it represented 1419bp of the mitochondrial genome. Numbers shown above the horizontal bars represent the nucleotide positions and numbers below are the expected lengths of the restriction digestion products. Longitudinal red bars represent the restriction cut sites. A plus sign (+) indicates a new cut site and a minus sign (−) indicates loss of a cut site. (a) and (b) show the expected results of *HpaI* digestion and (c), (d) and (e) show *NlaIII* digestion products. (a) represents the sequence of fragment C as in the CRS which lacks *HpaI* recognition sequence, and (b) represents an example of mitochondrial DNA sequence belonging to one of the major African mitochondrial haplogroups (L0, L1, L2, L5 or L6) which has a C3594T mutation producing a *HpaI* cut site which is considered a specific marker to the African lineages. (c) illustrates the sequence of the fragment C as in the CRS, having two *NlaIII* recognition sites. (d) represents an example of the mitochondrial haplogroup JT having an additional *NlaIII* cut site at nucleotide position 4216 resulting from T4216C mutation. This mutation is a diagnostic marker of the mitochondrial haplogroup JT and its descendants. (e) illustrates mitochondrial DNA sequence belonging to haplogroup V with the G4580A mutation causing loss of *NlaIII* recognition site at 4577.

#### *2.3.3.5 AluI restriction digestion of the fragment D of the mitochondrial genome (nt4875-7335)*

The fragment D was produced by PCR amplification of the mitochondrial genome using Mt D forward and Mt D reverse primers (**Table 2.4**). It included 2461bp extending between the mitochondrial nucleotide positions 4875 and 7335. The fragment D was digested with *AluI* restriction enzyme to further aid in the assignment of the mtDNA samples to well established mitochondrial haplogroups.

The CRS of the mtDNA belongs to the mitochondrial haplogroup H (the most common European mitochondrial haplogroup; 40-45% (Macaulay et al. 1999; Finnila et al. 2001b; Achilli et al. 2004)). According to Andrews et al. (1999), the nucleotide sequence between positions 4875 and 7335 contains fifteen *AluI* recognition sites (5′…AG▼CT…3′). However, the CRS lacks an *AluI* recognition sequence at position 7025 due to the presence of a cytosine base at nt7028 which is a specific marker to mitochondrial haplogroup H **(Figures 2.6.a and b)** (Loogvali et al. 2004; Roostalu et al. 2007).

Mitochondrial haplogroup D (a descendant of the macro-haplogroup M, originating in central Asia) is characterised by the C5178A mutation (Tanaka et al. 2004; Kong et al. 2006; Metspalu et al. 2006; Derenko et al. 2007; Achilli et al. 2008) leading to loss of *AluI* recognition sequence at that position **(Figure 2.6.c)** which is considered as a definitive marker for mitochondrial haplogroup D.

For better separation of the products of *AluI* restriction digestion, DNA fragments were separated on an 8% neutral polyacrylamide gel in 1xTBE buffer (89mM Tris-base, 89mM boric acid, 2mM EDTA; pH 8.3), along with hyper ladder IV (Bioline) as a molecular marker. Samples were mixed with 6x loading dye (0.25% (w/v) bromophenol blue (Sigma-Aldrich), 40% (w/v) sucrose (Fisher Scientific), 4ml (v/v) TE buffer and 10.8ml of 0.1% (v/v) DEPC-treated water) and electrophoresis was performed at 40V for 90min to 2 hours. After which, the gel was stained in 1xTBE buffer containing 30% (v/v) SYBR Safe DNA gel stain (Invitrogen) for 15 min. DNA fragments were visualised using ultraviolet transilluminator and gel documentation system (Syngene). Preparation of the polyacrylamide gel will be discussed in section 2.6.2.1.

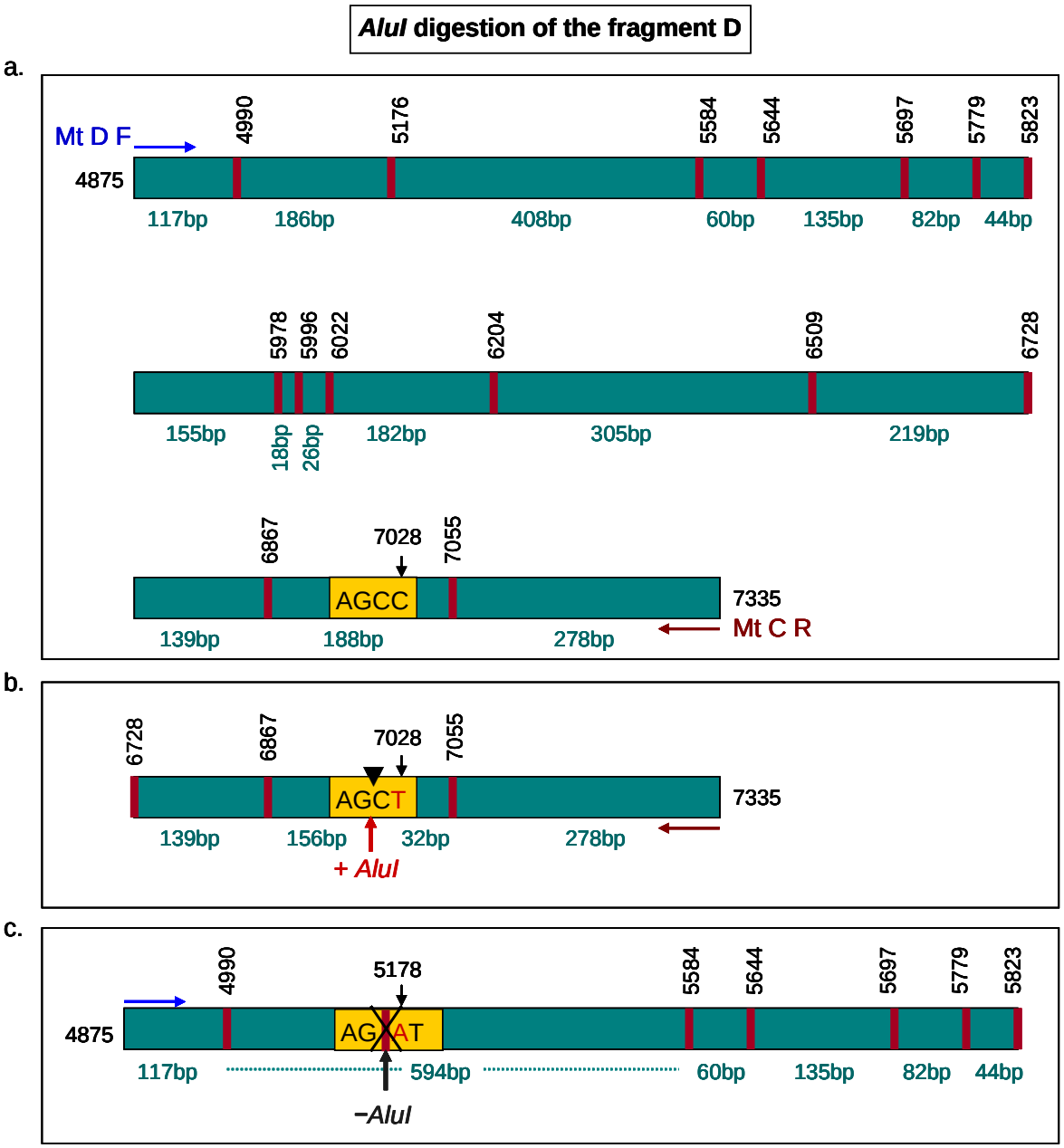



**Figure 2.6: *AluI* restriction digestion of the fragment D of the mitochondrial genome (nt4875 − 7335).** Blue and brown horizontal arrows represent forward and reverse primers respectively. Numbers shown above the horizontal bars represents the nucleotide positions and numbers below are the expected lengths of the restriction digestion products. Longitudinal red bars represent the restriction cut sites. A plus sign (+) indicates a new cut site and a minus sign (−) indicates loss of a cut site. (a) represents the Cambridge Reference Sequence between the nucleotide positions 4875 and 7335 showing lack of *AluI* recognition site at nucleotide position 7025 due to the presence of a cytosine base at 7028. (b) represents mitochondrial DNA between nucleotide positions 6728 and 7335 not belonging to haplogroup H having an *AluI* recognition site at position 7025 due to presence of a thymine base at nucleotide position 7028. (c) illustrates an example of mitochondrial DNA belonging to haplogroup D between nucleotide positions 4875 and 5823 with its characteristic C5178A mutation leading to loss of the *AluI* recognition site at that position producing a 594bp fragment instead of the two fragments of 186bp and 408bp in the CRS.

#### *2.3.3.6 RFLP analysis of the fragment E of the mitochondrial genome (nt8024-9231)*

Mitochondrial DNA fragment E (1208bp) (nucleotide position 8024 − 9231) (**Table 2.4**) was analysed by restriction digestion with four restriction enzymes individually; *AvaII, HaeIII, MboI* and *HaeII*. Collectively, these restriction analyses provide a highly specific set of RFLP markers that greatly enhance the mitochondrial haplogroup assignment.

*AvaII* digestion was used to differentiate the mitochondrial haplogroups (N1b, N1e, I and W) from all the other lineages including haplogroup H to which the CRS belongs. These four mitochondrial lineages represent direct descendants (non-R derived) of the macro-haplogroup N and they are characterised by a G8251A mutation (Palanichamy et al. 2004; Behar et al. 2006; Kivisild et al. 2006; Derenko et al. 2007). This mutation produces an *AvaII* cut site (5′…G▼G (A/T) C…3′) which is not present in the CRS (**Figure 2.7.a and b**). Therefore, it is considered as a characteristic marker for this set of closely related mitochondrial haplogroups.

Following the *AvaII* restriction digestion of the fragment E, *HaeIII* digestion was performed to confirm the assignment of the previously mentioned four mitochondrial haplogroups since the G8251A mutation causes loss of *HaeIII* recognition sequence (5′…GG▼CC…3′) which is linked to the *AvaII* gain at the same nucleotide position in these haplogroups. In addition, *HaeIII* digestion of the fragment E was used to differentiate haplogroup W from all the other mitochondrial lineages including sister haplogroups N1b, N1e and I. Loss of *HaeIII* recognition sequence at the nt8996 due to the presence of a G8994A mutation in the mtDNA delineates haplogroup W (Torroni et al. 1996; Finnila et al. 2001b; Palanichamy et al. 2004), and is considered as a specific marker for this haplogroup (**Figure 2.7.c and d**).

For better resolution of the African mitochondrial DNA belonging to the haplogroup L3, *MboI* digestion of the fragment E was applied to identify the subclade L3d which was previously described to have a T8618C mutation (Salas et al. 2002). This mutation causes loss of *MboI* recognition sequence (5′…▼GATC…3′) which is present in the CRS (accession NC_ 012920) (**Figure 2.8.a and b**).

In order to recognise mtDNA belonging to the haplogroup K and to differentiate it from all the other mitochondrial lineages including its ancestral haplogroup U, *HaeII* restriction digestion of the fragment E was performed. A definitive RFLP marker for haplogroup K and its sister clade U8 was evidenced by loss of *HaeII* cut site (5′. (A/G) GCGC▼ (T/C).3′) at the nucleotide position 9052 resulting from a G9055A mutation, which was described earlier as being specific to the mitochondrial haplogroup K and its sister clade U8 (Macaulay et al. 1999; Derbeneva et al. 2002; Herrnstadt et al. 2002; Achilli et al. 2005; Gonzalez et al. 2006) (**Figure 2.8.c and d**).

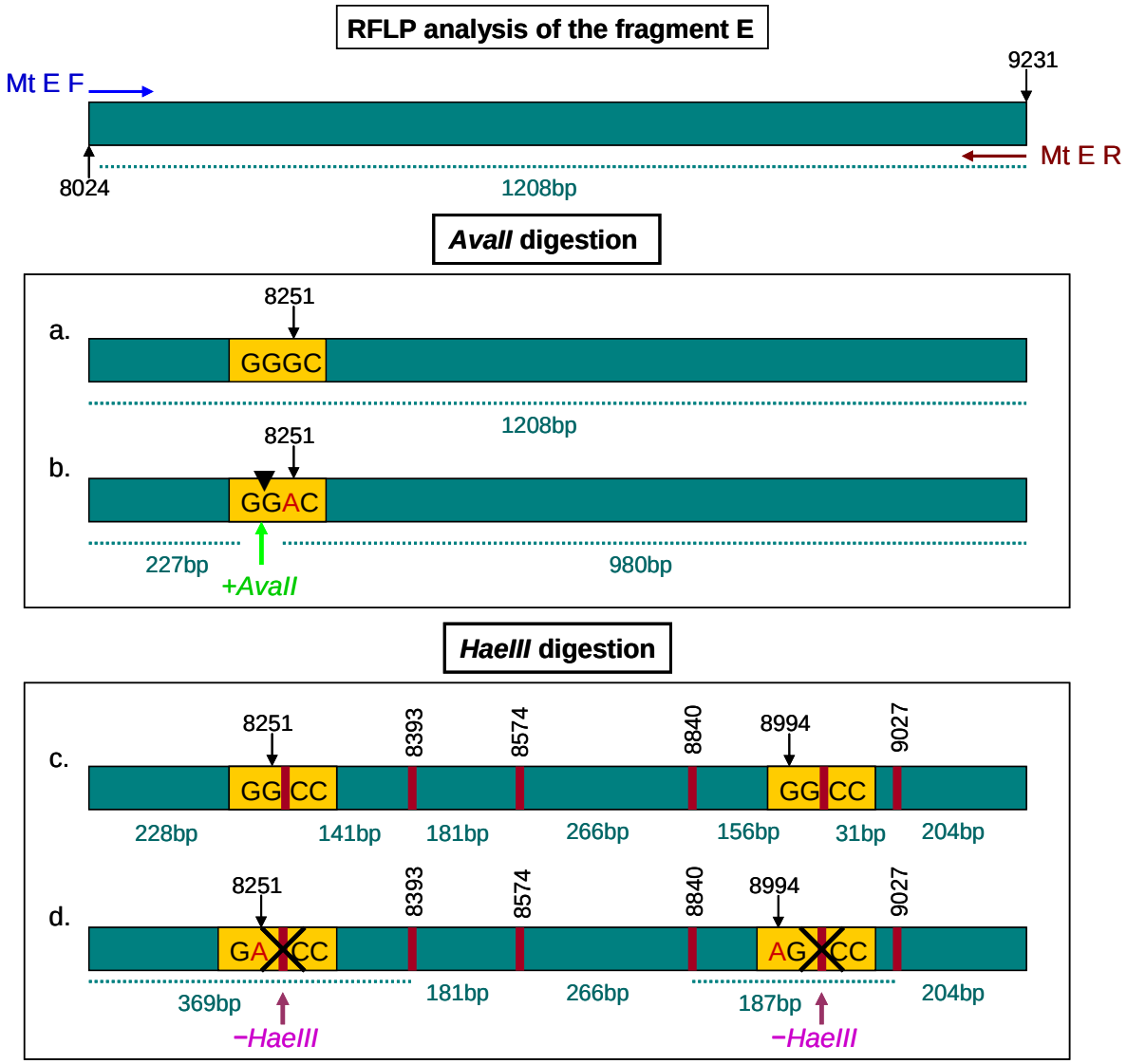



**Figure 2.7: Restriction digestion analysis of the fragment E of the mitochondrial genome (nt8024 and 9231) (part I).** Blue and brown horizontal arrows represent forward and reverse primers respectively. Numbers shown above the horizontal bars represent the nucleotide positions and numbers below are the expected lengths of the restriction digestion products. Longitudinal red bars represent the restriction cut sites. A plus sign (+) indicates a new cut site and a minus sign (−) indicates loss of a recognition sequence. Models a. and b. represent products of *AvaII* restriction digestion and models c. and d. represent *HaeIII* digestion products.

(a) illustrates the CRS between nucleotide positions 8024 and 9231 showing no recognition sequences for *AvaII* restriction enzyme, and (b) represents an example of mtDNA belonging to any of the four mitochondrial haplogroups (N1b, N1e, I and W) with their characteristic G8251C mutation leading to the occurrence of an *AvaII* recognition site. (c) illustrates the same region of the CRS having six *HaeIII* cut sites at the given nucleotide positions, and (d) represents an example of mtDNA belonging to haplogroup W which is identified by loss of *HaeIII* cut sites at 8252 and 8996 resulting from G8251A and G8994A mutations respectively.

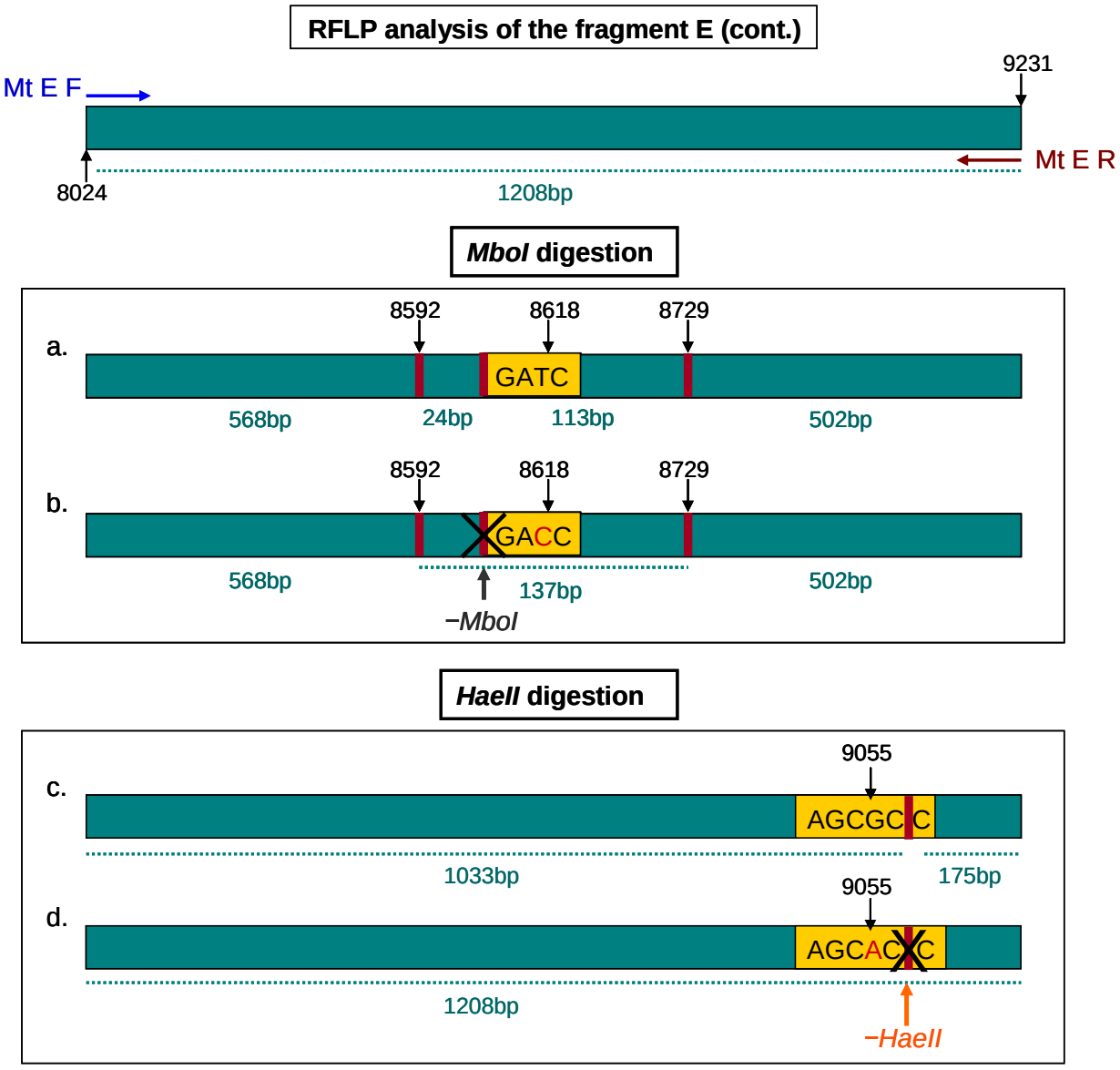



**Figure 2.8: Restriction digestion analysis of the fragment E of the mitochondrial genome (nt8024 and 9231) (part II).** Blue and brown horizontal arrows represent forward and reverse primers respectively. Numbers shown above the horizontal bars represent the nucleotide positions and numbers below are the expected lengths of the restriction digestion products. Longitudinal red bars represent the restriction cut sites. A plus sign (+) indicates a new cut site and a minus sign (−) indicates loss of a recognition sequence. Models a. and b. represent products of *MboI* restriction digestion and models c. and d. represent *HaeII* digestion products. (a) represents the CRS between the nucleotide positions 8024 and 9231 having three internal cut sites for *MboI* endonuclease, and (b) illustrates an example of mtDNA belonging to L3d subclade with its specific mutation at position 8618 by having a cytosine base rather than a thymine one causing loss of *MboI* recognition sequence. For *HaeII* digestion, (c) represents the CRS of the same region having one internal cut site at 9052 which is lost in (d) as an example

of mtDNA belonging to haplogroup K or its sister clade U8 with its characteristic G9055A mutation.

#### *2.3.3.7 RFLP analysis of the fragment F of the mitochondrial genome (nt9921-11150)*

Mitochondrial DNA fragment F (1230bp) (nt9921- 11150) was amplified using Mt F forward and Mt F reverse primers (**Table 2.4**). This region of the mitochondrial genome includes several important RFLP diagnostic markers for different mitochondrial haplogroups, and so fragment F was subjected to a series of individual restriction digestions with *DdeI, AluI, HinfI* and *TaqI* restriction enzymes.

*DdeI* restriction digestion splits the mitochondrial DNA haplogroups into two big sets according to the type of the nucleotide base at position 10398, which can produce a *DdeI* recognition sequence (5′…C▼TNA_G…3′) by being guanosine, or alternatively it can be an adenosine base causing loss of that cut site (**Figure 2.9.a and b**). Gain of 10398 *DdeI* cut site occurs in most of the relatively old mitochondrial DNA haplogroups including the African haplogroup L (Torroni and Wallace 1994; Chen et al. 1995b), the two macro-haplogroups M and N (Richards et al. 2003; Nelson et al. 2007; Maji et al. 2009), in addition to a number of N derived lineages either as a direct descendant like haplogroup I or indirect descendants as J and K1 haplogroups (Macaulay et al. 1999; Tolk et al. 2001). All the other mitochondrial haplogroups (including haplogroup H, to which the CRS belongs) lack this *DdeI* cut site (Finnila et al. 2001b).

For more specific identification of the mitochondrial haplogroups, *AluI* restriction digestion was performed on the same region of the mtDNA. The CRS of that region according to Andrews et al. (1999) had three internal cut sites for *AluI* (5′…AG▼CT…3′), however, mtDNA belonging to the haplogroup M is characterised by a C10400T mutation (Yao et al. 2002; Richards et al. 2003; Li et al. 2007; Maji et al. 2009) which in conjunction with the previously mentioned A10398G mutation produce an additional *AluI* cut site (**Figure 2.9.c and d**). A different *AluI* restriction digestion pattern was detected in mtDNA belonging to haplogroup I due to the occurrence of the diagnostic T10034C mutation (Palanichamy et al. 2004; Derenko et al. 2007; Van Oven and Kayser 2009) producing an additional *AluI* recognition sequence (**Figure 2.9.e**).

To further aid the assignment of mtDNA to the African haplogroup L5, *HinfI* restriction digestion of fragment F was performed. The digestion pattern of the mtDNA belonging to haplogroup L5 differs from that of the CRS due to the presence of the characteristic T10810C mutation (Chen et al. 1995b; Alves-Silva et al. 2000) which produces an additional *HinfI* cut site (5′…G▼ANT_C…3′) at nt10806 (**Figure 2.10.a and b)**.

For fine resolution of the mitochondrial haplogroup L3 and defining its subclade L3b, another RFLP marker was screened for by performing *TaqI* restriction digestion of the fragment F. An extra *TaqI* cut site (5′…T▼CG_A…3′) occurs at position 10084 in the mitochondrial subclade L3b compared to the CRS resulting from the characteristic A10086G mutation (Chen et al. 1995a; Alves-Silva et al. 2000; Richards et al. 2003; Rosa et al. 2004) (**Figure 2.10.c and d**).

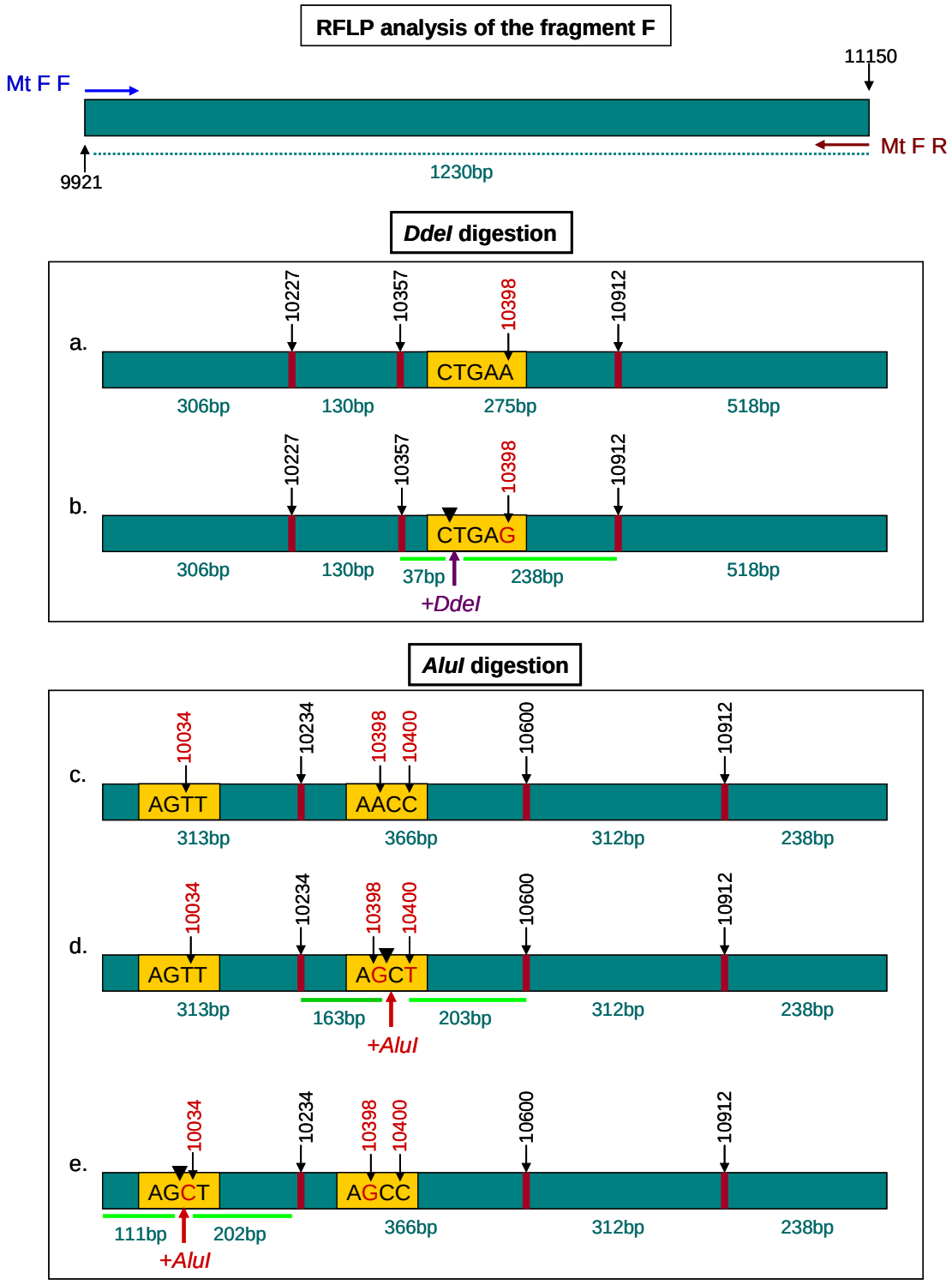



**Figure 2.9: Restriction digestion of mitochondrial DNA fragment F (nt9921-11150) (part I).** Blue and brown horizontal arrows represent forward and reverse primers respectively. Numbers shown above the horizontal bars represent the nucleotide positions and numbers below are the expected lengths of the restriction digestion products. Longitudinal red bars represent the restriction cut sites. A plus sign (+) indicates a new cut site and a minus sign (−) indicates loss of a recognition sequence. (a) and (b) represent products of *DdeI* restriction digestion and (c), (d) and (e) represent the *AluI* digestion patterns. (a) illustrates the CRS with the three characteristic *DdeI* cut sites along the fragment F together with an adenosine base at 10398, and (b) shows the digestion pattern of the old mitochondrial haplogroups (L, M, N, I, J and K1)in which a guanosine base is present at nt10398 leading to an additional *DdeI* recognition

sequence. *AluI* digestion pattern of the CRS between nt9921 and nt11150 is illustrated in (c) showing three internal cut sites**.** (d) highlights the different digestion pattern shown in the mitochondrial haplogroup M resulting from the characteristic C10400T mutation which produces a new cut site. (e) represents mtDNA belonging to haplogroup I with its characteristic T10034C mutation resulting in a novel *AluI* recognition sequence.

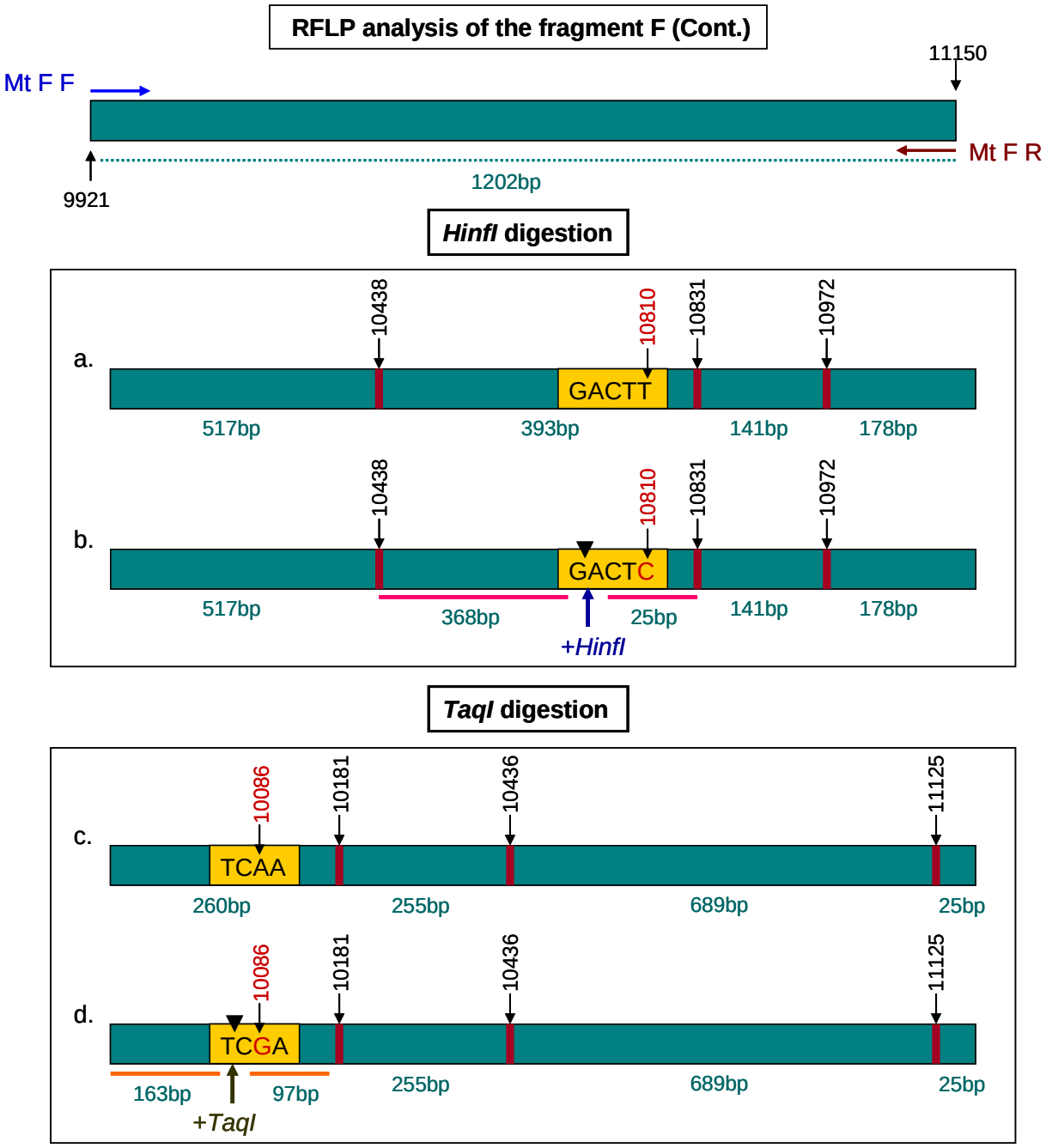



**Figure 2.10: Restriction digestion of mitochondrial DNA fragment F (nt9921-11150) (part II).** Blue and brown horizontal arrows represent forward and reverse primers respectively. Numbers shown above the horizontal bars represent the nucleotide positions and numbers below are the expected lengths of the restriction digestion products. Longitudinal red bars represent the restriction cut sites. A plus sign (+) indicates a new cut site and a minus sign (−) indicates loss of a recognition sequence. (a) and (b) represent the different *HinfI* digestion patterns and (c) and (d) illustrate the different *TaqI* digestion products. (a) shows the pattern resulting from *HinfI* digestion of the CRS along this region of the mitochondrial genome, and (b) represents an example of mtDNA belonging to haplogroup L5 with the characteristic T10810C mutation producing an additional *HinfI* recognition site. (c) illustrates the products of *TaqI*

restriction digestion of the CRS between nt9921 and nt11150 showing three internal cut sites, and (d) represents an example of mitochondrial subclade L3b sequence with the characteristic A10086G mutation producing an additional *TaqI* recognition sequence.

#### *2.3.3.8 RFLP analysis of the fragment G of the mitochondrial genome (nt12077-13812)*

The product of the PCR amplification of the mitochondrial genome using Mt G forward and Mt G reverse primers (**Table 2.4**), fragment G (1736bp) (nt12077-13812) was digested with *BamHI, AluI* and *HincII* restriction enzymes, separately. Digestion patterns were analysed and compared to those patterns resulting from the corresponding digestions of the CRS of the mitochondrial genome according to Andrews et al. (1999).

*BamHI* restriction digestion was used to specifically identify the mitochondrial haplogroup T [a European lineage, originated in the Near East approximately 46,500 years ago and is currently most prevalent in Europe (~10.5%) (Richards et al. 1998; Helgason et al. 2001)]. Haplogroup T is characterised by a G13368A mutation producing a novel *BamHI* recognition sequence (5′…G▼GATC_C…3′) at position 13366 which is not present in the CRS (**Figure 2.11.a and b**); (Torroni et al. 1996; Macaulay et al. 1999; Finnilä and Majamaa 2001).

Both *AluI* and *HincII* restriction digestions of the fragment G were performed to identify the mitochondrial haplogroup C, a descendant of the macro-haplogroup M which is found in Northeast Asia and is considered as one of the main mitochondrial haplogroups found in the indigenous peoples of the Americas (Tamm et al. 2007; Achilli et al. 2008). Identification of the haplogroup C was based on the detection of the characteristic A13263G mutation which is considered as a diagnostic marker for this haplogroup according to the MITOMAP database at (www7). Resulting from the A13263G mutation, a novel *AluI* recognition sequence (5′…AG▼CT…3′) is produced (**Figure 2.11.c and d**) and a previously known *HincII* cut site (5′.GT (T/C) ▾ (A/G) AC.3′) is lost (**Figure 2.11.e and f**) compared to the corresponding digestion patterns of the CRS.

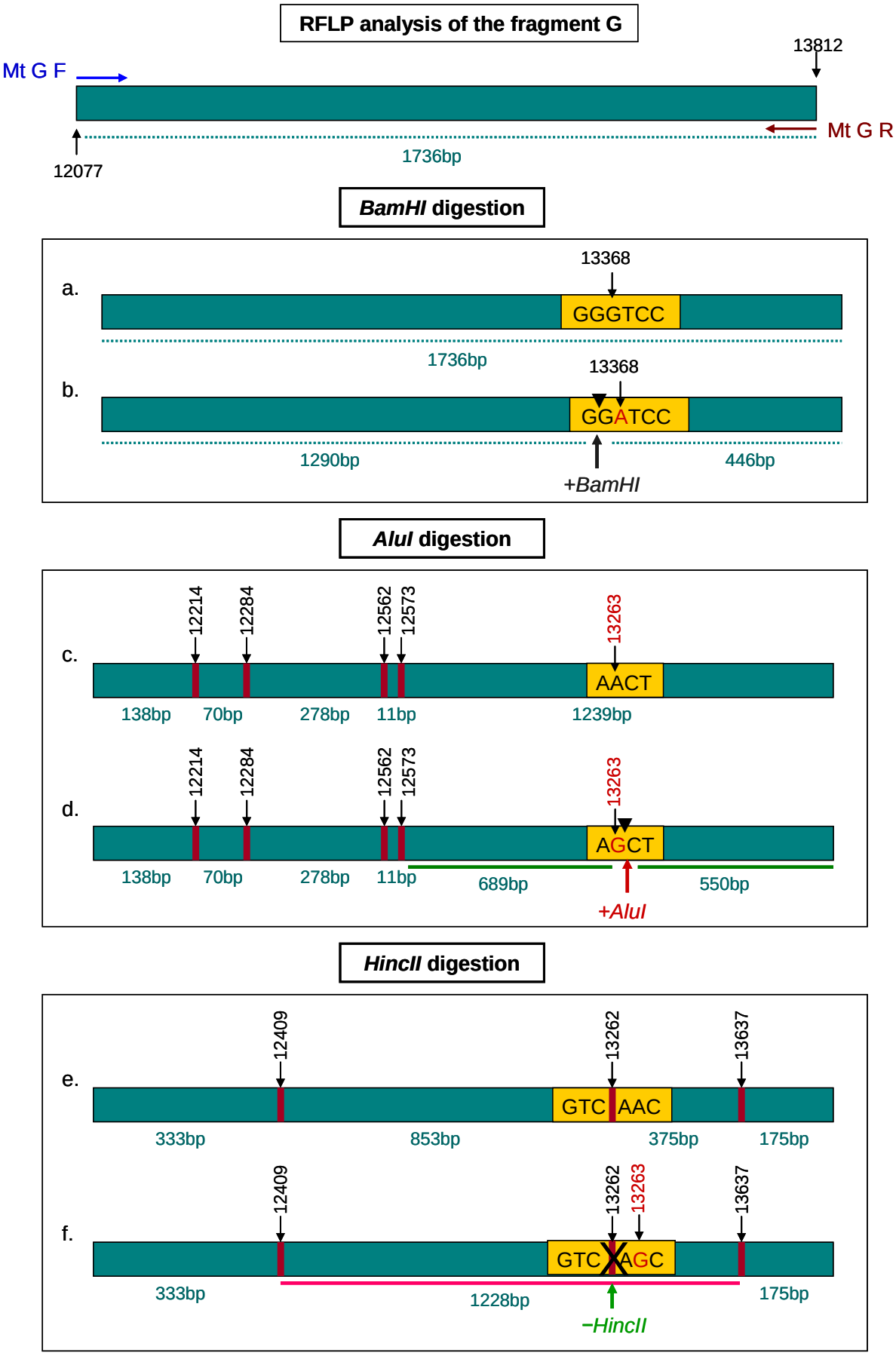

RFLP analysis of the fragment G
Mt G F
13812
Mt G R
12077
1736bp
BamHI digestion
a.
13368
GGGTCC
1736bp
b.
13368
GGATCC
1290bp
446bp
+BamHI
AluI digestion
c.
12214
12284
12562
12573
13263
AACT
138bp
70bp
278bp
11bp
1239bp
d.
12214
12284
12562
12573
13263
AGCT
138bp
70bp
278bp
11bp
689bp
550bp
+AluI
HincII digestion
e.
12409
13262
13637
GTC
AAC
333bp
853bp
375bp
175bp
f.
12409
13262
13263
13637
GTC
AGC
333bp
1228bp
175bp
−HincII

**Figure 2.11: Restriction digestion of mitochondrial DNA fragment G (nt12077-13812).** Blue and brown horizontal arrows represent forward and reverse primers respectively. Numbers shown above the horizontal bars represent the nucleotide positions and numbers below are the expected lengths of the restriction digestion products. Longitudinal red bars represent the restriction cut sites. A plus sign (+) indicates a new cut site and a minus sign (−) indicates loss of a recognition sequence. (a) illustrates *BamHI* digestion products of the CRS with no recognition sequence for *BamHI* , and (b) represents an example of mtDNA belonging to haplogroup T with its characteristic G13368A mutation producing a novel *BamHI* cut site. Shown in (c) is the CRS (nt12077-13812) as digested by the *AluI* restriction enzyme with four cut sites. (d) represents an example of mtDNA belonging to haplogroup C which has a different *AluI* digestion pattern resulting from the diagnostic A13263G mutation that produces an additional *AluI* recognition sequence. Represented in (e) is the digestion pattern of the CRS produced by *HincII* restriction enzyme with three internal cut sites, and (f) illustrates an example of mtDNA haplogroup C with its specific A13263G mutation causing loss of *HincII* at that position.

### *2.3.3.9 AccI and MseI restriction digestions of the fragment H of the mitochondrial genome (nt14253-14857)*

Mitochondrial DNA fragment H (605bp) was amplified using Mt H forward and Mt H reverse primers (**Table 2.4**). The nucleotide extension of fragment H was designed so as to include two important RFLP markers that were detected with *AccI* and *MseI* restriction enzymes.

The CRS of the mtDNA (accession number NC_012920) lacks *AccI* recognition sequence (5′.GT$^{\blacktriangledown}$ (A/C) (T/G) AC.3′) in the specified region of the fragment H (nt14253 − 14857), and hence the digestion reaction does not cut the original PCR product (**Figure 2.12.a**). Mitochondrial DNA belonging to the haplogroup X (a direct descendant of the macro-haplogroup N) was reported to have a T14470C mutation which produces a novel *AccI* recognition sequence (**Figure 2.12.b**) (Reidla et al. 2003; Kivisild et al. 2006; Derenko et al. 2007).

*MseI* restriction digestion of the fragment H was used to differentiate the mitochondrial haplogroup HV and all its descendants (including haplogroup H) from all the other mitochondrial lineages. That is based on the presence of a cytosine base at nt14766 in the haplogroup HV instead of a thymine base in all the other elderly derived mitochondrial lineages, which causes loss of the *MseI* cut site (5′…T$^{\blacktriangledown}$TAA...3′) (**Figure 2.12.c and d**) (Macaulay et al. 1999; Maca-Meyer et al. 2001; Achilli et al. 2004; Malyarchuk et al. 2005).

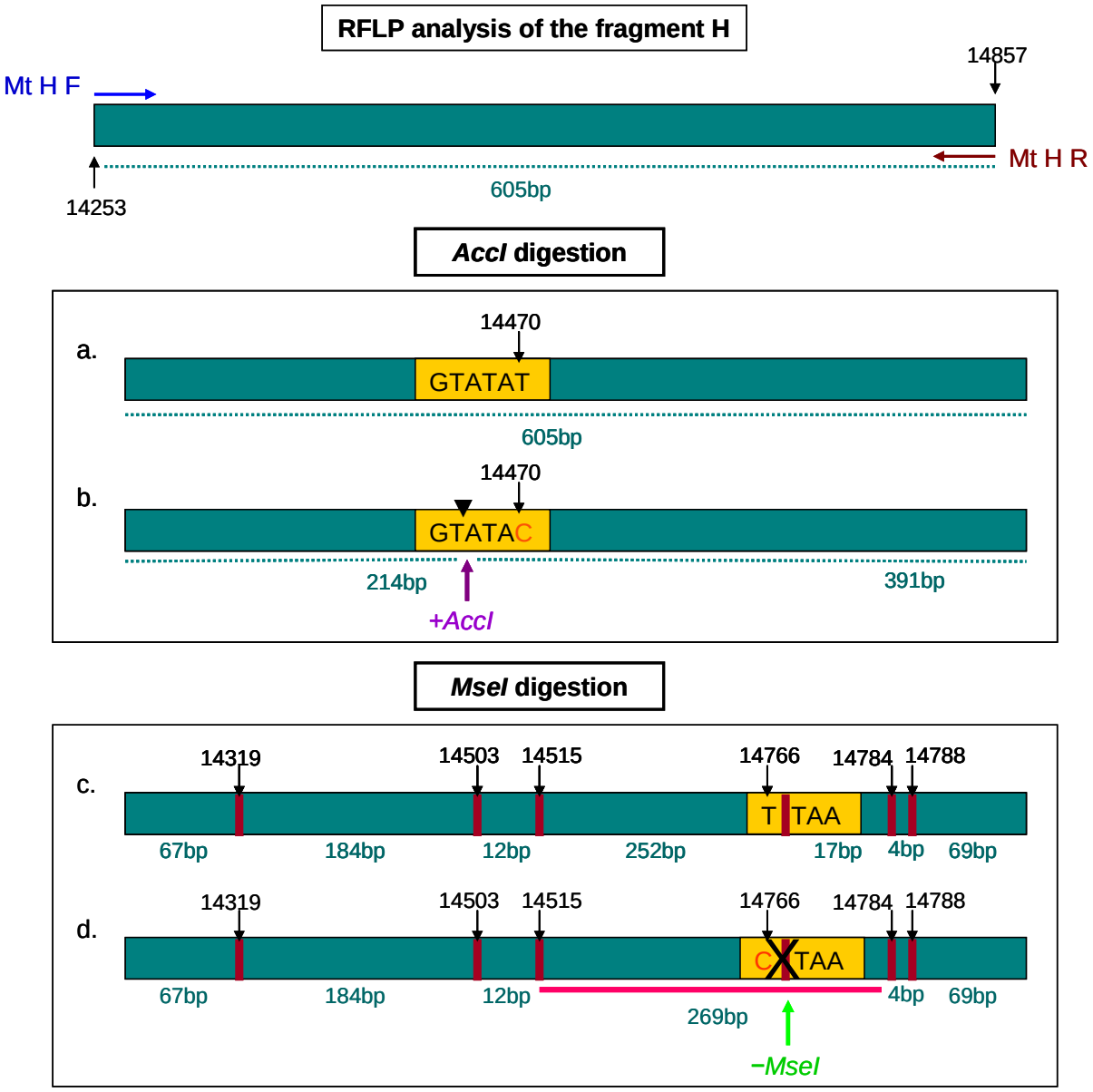



**Figure 2.12: Restriction digestion of mitochondrial DNA fragment H (nt14253-14857).** Blue and brown horizontal arrows represent forward and reverse primers respectively. Numbers shown above the horizontal bars represent the nucleotide positions and numbers below are the expected lengths of the restriction digestion products. Longitudinal red bars represent the restriction cut sites. A plus sign (+) indicates a new cut site and a minus sign (−) indicates loss of a recognition sequence. (a) and (b) represent *AccI* digestion products of the fragment H of the mitochondrial genome and (c) and (d) illustrate the products of *MseI* restriction digestion of the same region. (a) represents the CRS (nt14253 − nt14857) with no *AccI* recognition sequence, and (b) illustrates an example of mtDNA belonging to the haplogroup X with the characteristic T14470C mutation which produces a new *AccI* cut site. Shown in (c) is a representation of mtDNA belonging to any mitochondrial haplogroup other than HV and its descendants (including haplogroup H) in which a thymine base is present at nucleotide position 14766 producing an *MseI* recognition sequence. (d) illustrates the mitochondrial haplogroups HV and its descendants with a cytosine base at nt14766 causing loss of *MseI* cut site at that position.

#### *2.3.3.10 AluI and HinfI restriction digestions of the fragment I of the mitochondrial genome (nt15312-149)*

Mitochondrial DNA fragment I (1407bp) (nt15312 − 149) was amplified using Mt I forward and Mt I reverse primers (**Table 2.4**). This region of the mtDNA contains two important RFLP markers that further aid the assignment of two closely related mitochondrial lineages, haplogroups T and J. To screen for these markers, the fragment I was digested with *AluI* and *HinfI* restriction enzymes individually.

*AluI* restriction digestion was used to confirm the assignment of the mitochondrial haplogroup T, which was previously obtained through *BamHI* digestion of the fragment G (**Figure 2.11.b**). Mitochondrial DNA belonging to haplogroup T is characterised by an A15607G mutation producing a novel *AluI* cut site (5′…AG▼CT…3′) which is not present in the CRS (**Figure 2.13.a and b**); (Torroni et al. 1996; Macaulay et al. 1999; Finnilä and Majamaa 2001). Therefore, gain of *AluI* and *BamHI* at nucleotide positions 15606 and 13366, respectively are considered as definitive RFLP markers for the haplogroup T.

*HinfI* digestion of the fragment I was performed to confirm the assignment of the mitochondrial haplogroup J and to differentiate it from its sister clade, haplogroup T and the others. Mitochondrial haplogroup J is characterised by a C16069T mutation which disrupts the *HinfI* recognition sequence (5′…G▼ANT_C…3′) that is present in the CRS (**Figure 2.13.c and d**); (Macaulay et al. 1999; Maca-Meyer et al. 2001; Logan 2009; Van Oven and Kayser 2009). Therefore, loss of *HinfI* cut site at nt16066 is considered as a diagnostic marker for the haplogroup J.

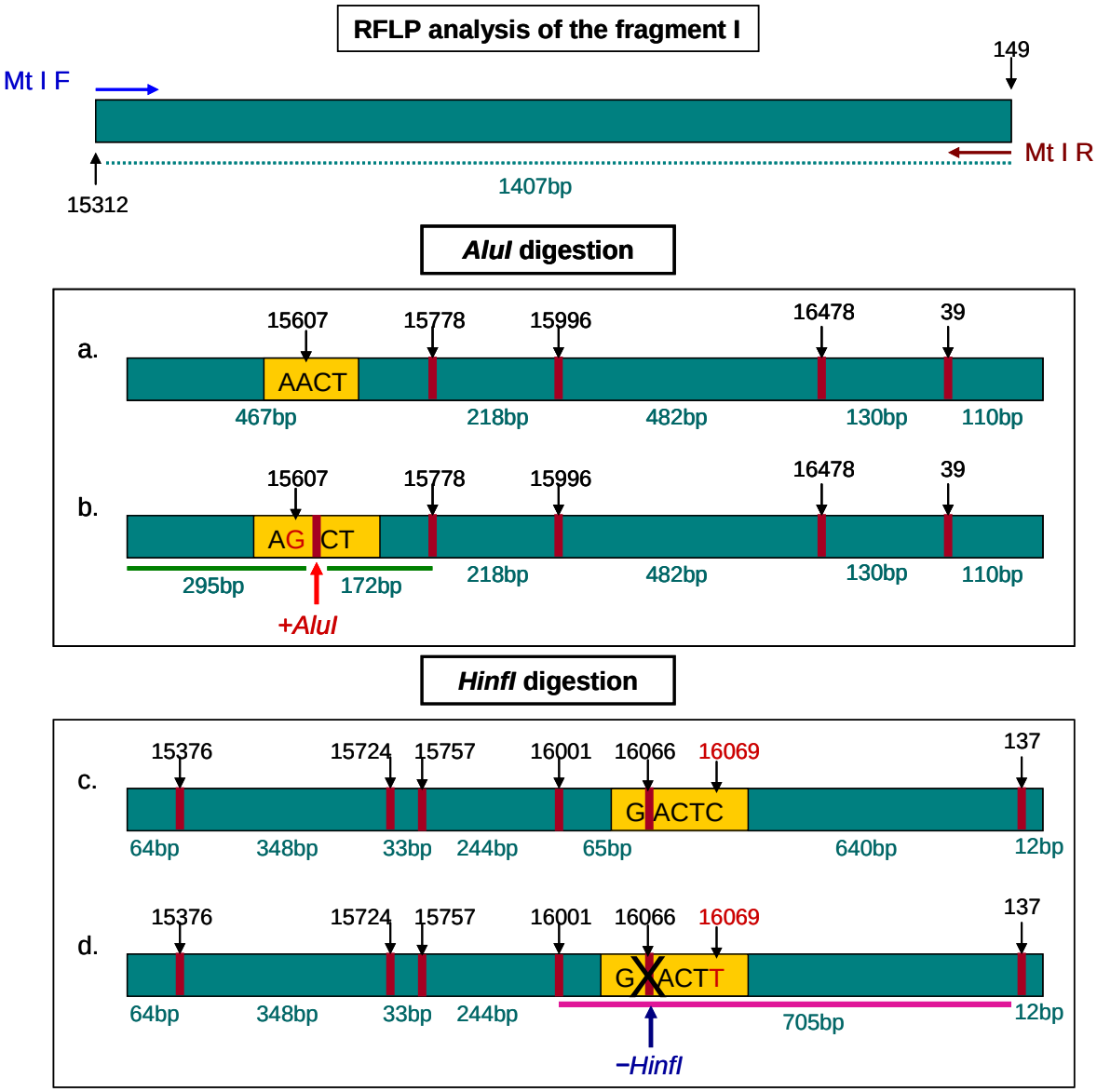



**Figure 2.13: Restriction digestion of mitochondrial DNA fragment I (nt15312-149).** Blue and brown horizontal arrows represent forward and reverse primers respectively. Numbers shown above the horizontal bars represent the nucleotide positions and numbers below are the expected lengths of the restriction digestion products. Longitudinal red bars represent the restriction cut sites. A plus sign (+) indicates a new cut site and a minus sign (−) indicates loss of a recognition sequence. (a) represents the *AluI* digestion products of the CRS, and (b) illustrates an example of mtDNA belonging to haplogroup T with its specific A15607G mutation that produces an additional *AluI* cut site. Represented by (c) is the CRS as digested by *HinfI* restriction enzyme showing six internal cut sites. (d) illustrates mtDNA belonging to haplogroup J with its characteristic C16069T mutation causing disruption of the *HinfI* recognition sequence at nt16066.

### 2.3.4 Screening for A12308G mutation as a diagnostic marker for the mitochondrial haplogroups U and K

#### *2.3.4.1 Mispairing PCR*

Mitochondrial haplogroups U and K have been described to have a characteristic A12308G mutation which specifically enables the differentiation of these mitochondrial lineages from many others (Torroni et al. 1996; Achilli et al. 2005). This mutation was screened for by the mispairing PCR technology, in which the mitochondrial DNA was amplified between the nucleotide positions 12221 and 12334 using a mismatched reverse primer. The mispairing PCR product size was 114bp. The reverse primer bound its 3′ end next to the nucleotide position 12308 and included a nonpairing nucleotide G at position 12312. When the A12308G mutation is present, in conjunction with the mismatched primer sequence, a new *EcoRI* restriction site (5′…G▼AATT_C…3′) is generated (**Figure 2.14**). Sequences of the primers were obtained from Lauber et al. (1991) whose study was concerned with mutations in the mitochondrial tRNA genes as a frequent cause of the neuromuscular diseases.

The primer sequences were as follows;
Ser-For: 5′-GAACTGCTAACTCATGCCCCCCATGT-3′
(Corresponding to the mitochondrial nucleotide positions 12221 to 12245)

Leu-Rev: 5′-CTTTTATTTGGAGTTGCACCA**G**AATT-3′
(Corresponding to the mitochondrial nucleotide positions 12309 to 12334)

The PCR reaction mixtures were prepared in 500µl sterile nuclease-free PCR tubes (Eppendorf) as follows:

- 5x Phusion HF buffer (Finnzymes) 10.0µl
- $MgCl_2$ (25mM; Promega) 2.0µl
- DeoxyNTP mixture (10mM; Promega) 2.0µl
- Bovine Serum Albumin (BSA) (10 mg/ml; Promega) 2.0µl
- Ser-For primer (100pmol/µl) 1.0µl
- Leu-Rev primer (100pmol/µl) 1.0µl
- Phusion High-Fidelity DNA polymerase (2u/µl; Finnzymes) 0.5µl
- Nuclease free water to final volume of 50µl

All the constituents were then added to the processed dried discs of the blood samples on FTA cards (as previously described in section 2.1.2) in sterile 500µl PCR tubes after which the tubes were placed into the thermal cycler.

The PCR conditions were as follows:

| | | | |
|---|---|---|---|
| Initial Denaturation | 98°C | 30 sec | 1 cycle |
| Denaturation | 98°C | 10 sec | 30 cycles |
| Annealing | 65°C | 20 sec | |
| Extension | 72°C | 30 sec | |
| Final Extension | 72°C | 5 min | 1 cycle |

Small volumes (5µl) of the PCR products were separated using a 2.5% (w/v) agarose (Bioline) gel containing 1x SYBR Safe DNA gel stain (Invitrogen) (in 1x TAE buffer), along with hyper ladder V (Bioline) as a molecular marker and run for 45min at 100V. 114bp bands were then visualised using an ultraviolet (UV) transilluminator and gel documentation system (Syngene).

#### *2.3.4.2 EcoRI restriction digestion of the mispairing PCR product*

To determine whether the mtDNA sample contained the definitive mutation (A12308G) for the mitochondrial haplogroups U and K, the products of the mispairing PCR were digested with *EcoRI* restriction enzyme (Promega). MtDNA belonging to haplogroups U and K were cut by *EcoRI* due to presence of a guanosine base at nt12308 which in conjunction with the mispairing reverse primer (Leu-Rev) produce an *EcoRI* restriction site that cuts the PCR product (114bp) into two fragments of 88bp and 26bp (**Figure 2.14.a**). The PCR products of samples not belonging to haplogroups U and K lack *EcoRI* cut sites due to presence of an adenosine base at the nucleotide position 12308 which does not constitute an *EcoRI* recognition sequence even in conjunction with the mispairing primer sequence (**Figure 2.14.b**).

The *EcoRI* digestion reaction was prepared as described early in section 2.3.4, and all the constituents were incubated at 37°C for 2 hours. The products of digestion were separated on a 3.5% (w/v) agarose gel stained with SYBR Safe, along with hyper ladder V as a molecular marker and DNA fragments were visualised using the UV transilluminator and gel documentation system (Syngene).

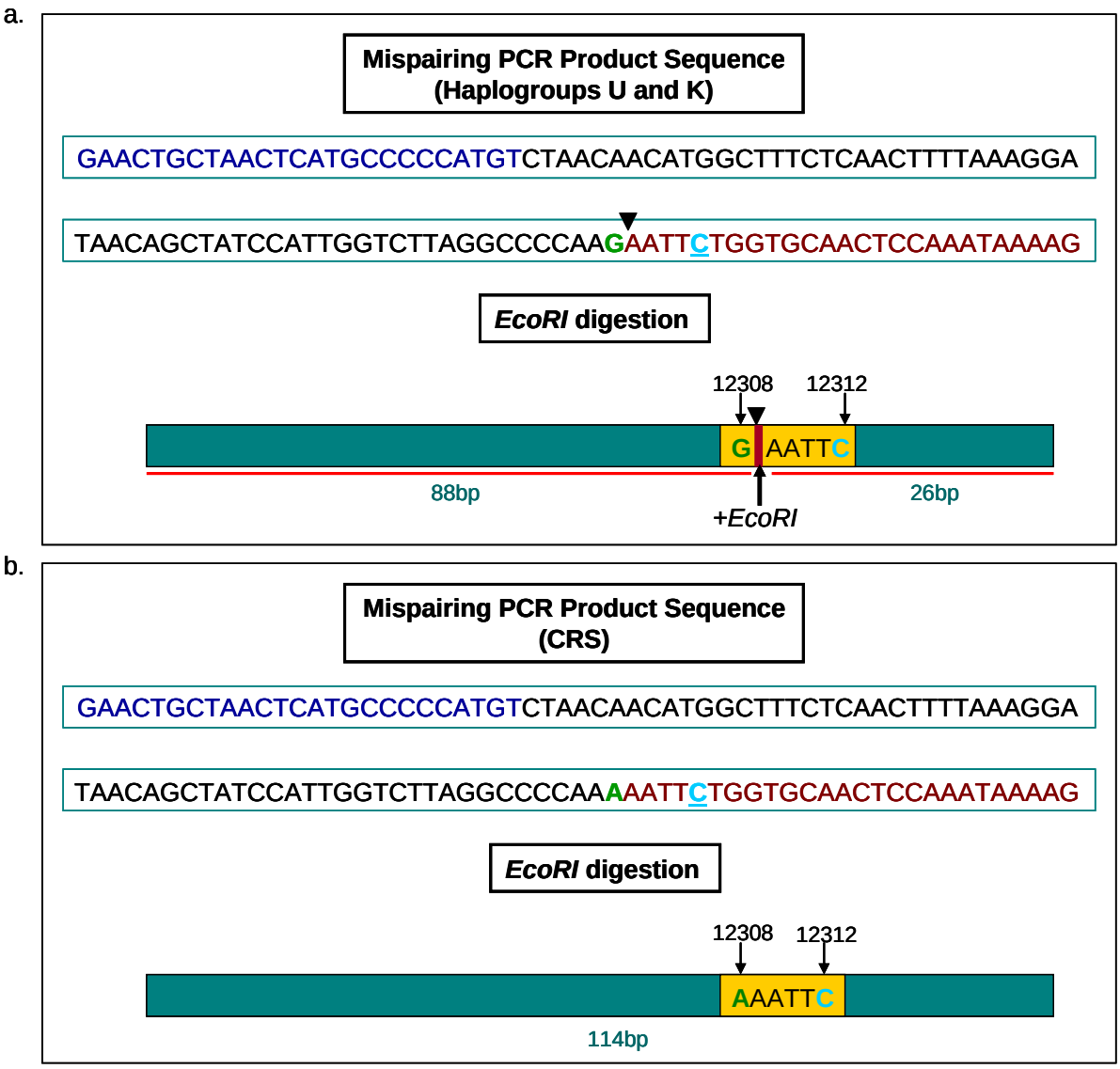



**Figure 2.14: Sequences and *EcoRI* digestion of the mitochondrial DNA mispairing PCR products (nt12221-12334).** Shown in dark blue is the sequence of the forward primer (Ser-For) and in brown is the complementary sequence of the reverse primer (Leu-Rev) with the mispairing nucleotide in turquoise. The nucleotide position 12308 is shown in green. (a) shows an example of mtDNA belonging to the haplogroup U or K (nt12221 − 12334) with the characteristic A12308G mutation in conjunction with the mismatched primer sequence producing an *EcoRI* cut site. Shown in (b) is the corresponding region of the CRS of the mitochondrial genome with an adenosine base at nt12308 and absence of *EcoRI* recognition sequence in conjunction with the mismatched primer.

### 2.3.5 Summary of the mitochondrial haplogroup specific RFLP markers

The total of 26 diagnostic RFLP markers that were used in the current study to enable the assignment of the mitochondrial DNA haplogroups in the Egyptian population is summarised in **Table 2.7**.

**Table 2.7: Diagnostic RFLP markers applied for the assignment of the mitochondrial DNA haplogroups.** Nucleotide positions for RFLP markers represent the first base of the recognition sequence on the heavy strand and not the actual cut site. A plus sign (+) indicates gain of an enzyme cut site and a minus sign (−) indicates loss of an enzyme cut site. All diagnostic markers are indexed to the CRS (accession number NC_012920).

| PCR product | Nucleotide position | Length (bp) | Diagnostic SNPs | RFLP marker(s) | Haplogroup(s) |
|---|---|---|---|---|---|
| Fragment A | 451-827 | 377 | A663G | + 663 *HaeIII* | A |
| Fragment B | 1489-2690 | 1202 | G1719A | − 1715 *DdeI* | N, I and X2 |
| | | | T2352C | + 2349 *MboI* | L1b and L3e |
| Fragment C | 3314-4732 | 1419 | C3594T | + 3592 *HpaI* | L0, L1, L2, L5 and L6 |
| | | | T4216C | + 4216 *NlaIII* | JT |
| | | | G4580A | − 4577 *NlaIII* | V |
| Fragment D | 4875-7335 | 2461 | C5178A | − 5176 *AluI* | D |
| | | | C7028C | − 7025 *AluI* | H |
| Fragment E | 8024-9231 | 1208 | G8251A | + 8251 *AvaII* / − 8251 *HaeIII* | N1b, N1e, I and W |
| | | | T8618C | − 8616 *MboI* | L3d |
| | | | G8994A | − 8994 *HaeIII* | W |
| | | | G9055A | − 9055 *HaeII* | K |
| Fragment F | 9921-11150 | 1230 | T10034C | + 10032 *AluI* | I |
| | | | A10086G | + 10084 *TaqI* | L3b |
| | | | A10398G | + 10394 *DdeI* | L, M, N, I, J and K1 |
| | | | C10400T/ A10398G | + 10397 *AluI* | M |
| | | | T10810C | + 10806 *HinfI* | L5 |

**Table 2.7: continue**

| PCR product | Nucleotide position | Length (bp) | Diagnostic SNPs | RFLP marker(s) | Haplogroup(s) |
|---|---|---|---|---|---|
| Fragment G | 12077-13812 | 1736 | A12308G | + 12308 *EcoRI** | U and K |
| | | | A13263G | − 13259 *HincII* | C |
| | | | A13263G | + 13262 *AluI* | C |
| | | | G13368A | + 13366 *BamHI* | T |
| Fragment H | 14253-14857 | 605 | T14470C | + 14465 *AccI* | X |
| | | | C14766C | − 14766 *MseI* | HV** |
| Fragment I | 15312-149 | 1407 | A15607G | + 15606 *AluI* | T |
| | | | C16069T | − 16066 *HinfI* | J |

* *EcoRI* digestion was performed following mispairing PCR amplification of the mtDNA as will be discussed in section 2.3.4.
** Loss of *MseI* cut site is considered as a diagnostic marker for haplogroup HV and all its descendants including haplogroup H to which the CRS belongs.

### 2.3.6 Development of a new strategy for RFLP analysis of the human mitochondrial genome

Following screening of the whole mitochondrial genome for the complete set of the 26 haplogroup specific RFLP markers in a representative number of the Egyptian blood samples on the FTA cards (n= 64) and based on the results obtained, eighteen of these markers were selected for screening of the remaining samples (n=197). A highly efficient specific scheme was designed to enable rapid and successful assignment of the Egyptian blood samples to well established mitochondrial DNA haplogroups (**Figure 2.15**).

The scheme was primarily based on splitting the mitochondrial DNA samples into two big sets of mitochondrial lineages based on presence or absence of a *DdeI* cut site at position 10394 which differentiates the old mitochondrial lineages, mainly the African L haplogroups and the two macro-haplogroups M and N from the more recently derived mitochondrial lineages, including the haplogroup H to which the CRS belongs. Following that, a certain sequence of highly specific restriction analyses was performed to enable a definitive assignment of the mitochondrial DNA haplogroups (**Figure 2.15**).

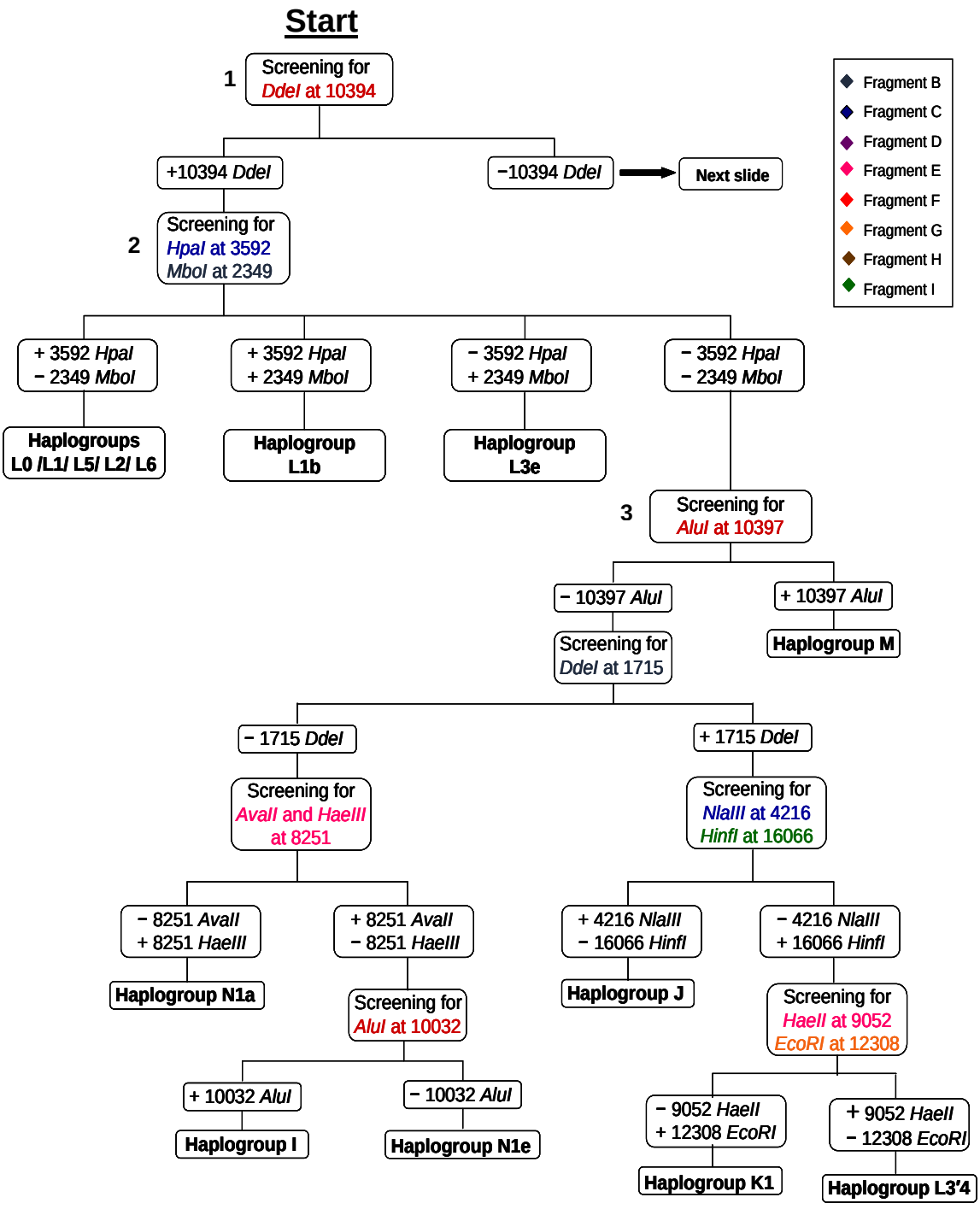
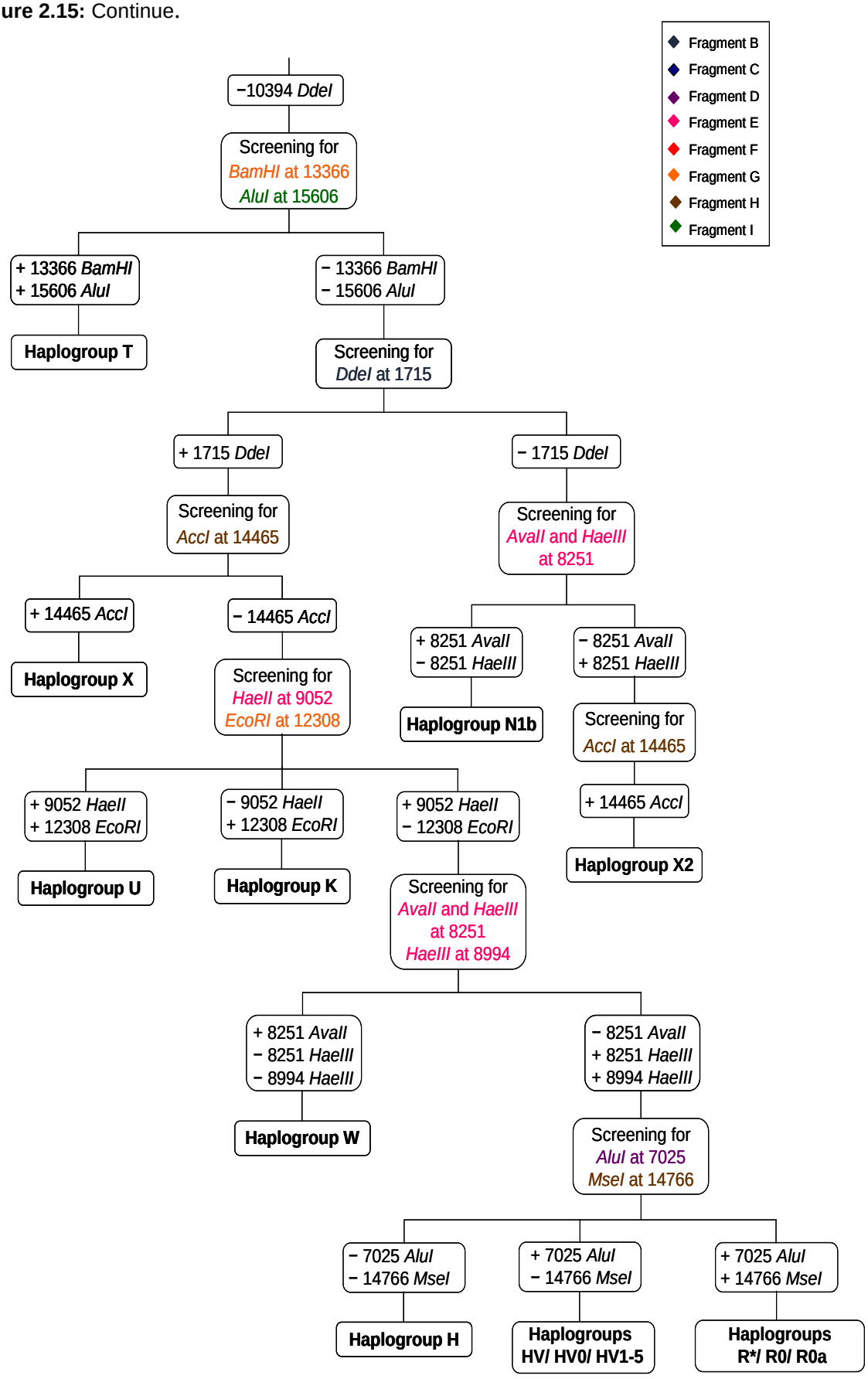



**Figure 2.15: The workflow of the RFLP analysis of the human mitochondrial genome.** The flowchart shows the sequence of the restriction digestion analyses that was followed to screen the Egyptian mtDNA samples for haplogroup specific RFLP markers aiming at fulfilment of rapid and successful haplogroup assignment.

**Figure 2.15:** Continue.



### 2.3.7 Frequency distribution of the mitochondrial DNA haplogroups in the Egyptian population sample

The frequency distribution of the different mitochondrial DNA haplogroups in the Egyptian population sample (n= 261) was determined. The frequency of each haplogroup was calculated by dividing the number of individuals assigned to this haplogroup by the total number of samples. Finally, the maternal lineages of the modern Egyptian population was concluded and correlated to the previously reported mitochondrial genetic structure of the surrounding populations.

## 2.4 Screening for the amelogenin Y (AMELY) deletion among the Egyptian population

### 2.4.1 PCR amplification of AMELY

The Egyptian male blood samples on FTA cards (n=131) were screened for AMELY deletion by PCR amplification of the amelogenin gene which is present on both the X- (AMELX, GenBank accession number M55418) and the Y- (AMELY, GenBank accession number M55419) chromosomes (Nakahori et al. 1991). The PCR was performed using the primers that were described by Sullivan and his colleagues and have been incorporated into most of the commonly used commercial STR multiplex kits (Sullivan et al. 1993; Butler 2005a; Goodwin et al. 2007). These primers were designed to bind to conserved regions on both homologues of the amelogenin gene and to delimit a 6bp deletion on the X- chromosome. Accordingly, the PCR of the two versions of the amelogenin gene produces fragments of 112bp and 106bp for AMELY and AMELX, respectively and subsequently allows the differentiation between the two homologues which is the basis for the DNA-based gender testing (**Figure 2.16**).

The primer sequences were as follows:

AMEL-For: 5′-CCCTGGGCTCTGTAAAGAATAGTG-3′

(Corresponding to the nucleotide positions 637553 to 637576 on the Y-chromosome and the nucleotide positions 2885578 to 2885601 on the X-chromosome)

AMEL-Rev: 5′-ATCAGAGCTTAAACTGGGAAGCTG-3′

(Corresponding to the nucleotide positions 637664 to 637641 on the Y-chromosome and the nucleotide position 2885683 to 2885660 on the X-chromosome)

Nucleotide positions were derived from the whole genome sequences of the Y-chromosome (NCBI reference sequence: NW_001842425.2) and the X-chromosome (NCBI reference sequence: NW_001842360.1)

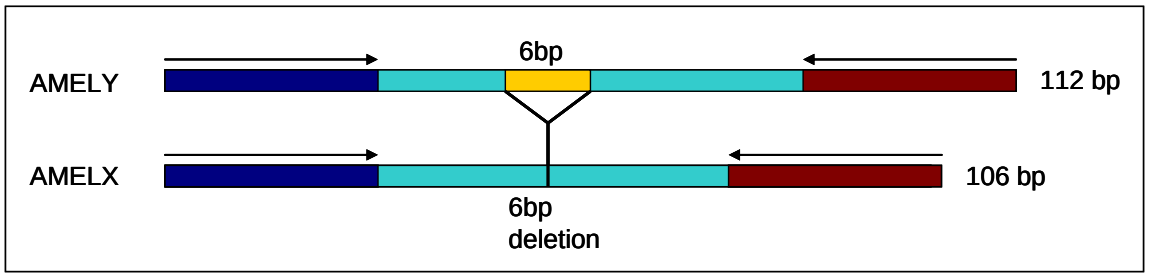



**Figure 2.16: A Schematic of the PCR amplification of a specific region of the amelogenin gene on the Y and X chromosomes.** The blue and the brown bars represent the AMEL-forward and reverse primers respectively. Shown at the top is an illustration of the PCR product of AMELY having 6bp (represented by the yellow bar) more than that of AMELX. The PCR product sizes of AMELY and AMELX were 112bp and 106bp, respectively.

The constituents of the PCR reaction mixture were prepared in 500µl sterile PCR tubes as follows:

- 5x Phusion HF buffer (Finnzymes) 10.0µl
- $MgCl_2$ (25mM; Promega) 2.0µl
- DeoxyNTP mixture (10mM; Promega) 2.0µl
- Bovine Serum Albumin (BSA) (10 mg/ml; Promega) 2.0µl
- AMEL-For primer (100pmol/µl) 0.5µl
- AMEL-Rev primer (100pmol/µl) 0.5µl
- Phusion High-Fidelity DNA polymerase (2u/µl; Finnzymes) 0.5µl
- Nuclease free water to final volume of 50µl

The listed components were added to sterile thin-walled 500µl eppendorf tubes contained the processed dried discs (as previously described in section 2.1.2). A positive control was always included in the experiment in which a processed female blood sample on FTA micro-card was used as a template. A negative control was also taken throughout the whole experiment with a processed blank disc replacing the blood stained disc.

The PCR conditions for AMELY amplification were as follows:

| | | | |
|---|---|---|---|
| Initial Denaturation | 98°C | 30 sec | 1 cycle |
| Denaturation | 98°C | 10 sec | |
| Annealing | 65°C | 20 sec | 30 cycles |
| Extension | 72°C | 30 sec | |
| Final Extension | 72°C | 5 min | 1 cycle |

### 2.4.2 Polyacrylamide gel electrophoresis for AMELY PCR products

For good separation of the AMELY PCR products and accurate discrimination between the normal males and the AMELY-null males which give the same PCR product as females, 5% polyacrylamide gels were used. The main principle of the gel electrophoresis involves separation of the nucleic acid molecules according to their sizes (molecular weights) that influence the speed by which the molecules migrate through the gel having the smaller ones running faster. The polyacrylamide gel is characterised by having much smaller pore sizes than the agarose gel which is an important factor in determining the ability of the gel to resolve two similarly sized DNA fragments. Therefore, the main purpose behind the use of the polyacrylamide gel in this experiment was to enable the differentiation between the expected two fragments pattern of the normal male samples having two bands only 6bp apart and the single fragment pattern of the AMELY-null males simulating females.

#### *2.4.2.1 Preparation of the polyacrylamide gel*

Mini-PROTEAN Tetra Electrophoresis System (Bio-Rad) was used for the preparation of the polyacrylamide gel according to the manufacturers' instructions, which involved assembly of two clean and completely dry glass plates with 1.5mm vertical spacers in between for each gel. These assembled plates were then placed in a casting frame which was clamped into a casting stand ready for the separating gel solution to be poured in. The acrylamide separating gel mix was then prepared (**Table 2.8**) and was poured immediately between the two glass plates while trying not to trap any air bubbles as much as possible. A 10-well comb was then placed on the top between the two glass plates and the gel was then allowed to polymerize at room temperature for 30 to 45min.

**Table 2.8: Preparation of the acrylamide separating gels.**

| Constituents | 5% gel | 8% gel[e] |
|---|---|---|
| Acrylamide-bisacrylamide (29:1)[a] | 1.992ml | 3.192ml |
| 5x TBE buffer, pH 8.3[b] | 2.4ml | 2.4ml |
| dd$H_2O$ | 7.512ml | 6.312ml |
| 10% APS (Sigma-Aldrich) [c] | 84µl | 84µl |
| TEMED (Sigma-Aldrich) [d] | 12µl | 12µl |

[a] the 40% acrylamide stock (Sigma-Aldrich) contained N, N'-methylenebisacrylamide in a final ratio of 29:1 giving the suitable cross-linking ratio between the polyacrylamide chains. [b] 5xTBE buffer contained 445mM Tris-base, 445mM boric acid and10mM EDTA. [c] 10% (w/v) ammonium per sulphate was used to initiate the polymerization of acrylamide and bisacrylamide monomers. [d] *N*, *N*, *N′*, *N′*-tetramethylethylenediamine was the last constituent to be added to the polyacrylamide separating gel mix as a great enhancer and stabilizer to the polymerization reaction. [e] 8% polyacrylamide gel was used for separation of *AluI* restriction digestion products of fragment D of the mitochondrial genome (**Figure 2.6.a**).

### *2.4.2.2 Loading of samples and running the gel*

Once the separating gel had polymerized, the two glass plates with the gel in between were placed in the electrophoresis running chamber and submerged with the running buffer (1x TBE buffer; 89mM Tris-base, 89mM boric acid, 2mM EDTA; pH 8.3) and the comb was then taken out. Pre-stained hyper ladder V (Bioline) was first loaded onto the gel as a molecular marker to determine the size of the PCR fragments. A volume of 7µl of the AMELY PCR product was mixed with 1.4µl 6x loading dye (0.25% (w/v) bromophenol blue, 40% (w/v) sucrose, 4ml (v/v) TE buffer and 10.8ml of 0.1% (v/v) DEPC-treated water) and loaded onto the gel. A female AMELY PCR product was run along the male samples on each gel as a control. In addition, a negative control was also run to rule out any possibility of contamination which might affect the results. Electrophoresis was performed at 40V for 90min-2hrs after which the dye front almost ran off the gel to ensure the adequate separation of the bands.

### *2.4.2.3 Staining of the polyacrylamide gel and visualisation of the AMELY PCR products*

Once the electric current was stopped, the running buffer was then discarded and the glass plates were taken out. After which, the two plates were detached using a plastic wedge and the gel was ready for staining. The acrylamide gel was then immersed in 1xTBE buffer containing 3x SYBR Safe DNA gel stain (Invitrogen) for 15-20min. DNA fragments were visualised using UV transilluminator and gel documentation system (Syngene).

# 3. THE UTILITY OF FTA CARDS IN FORENSIC INVESTIGATIONS

## 3.1 Introduction

Following the growing demand for the DNA analysis in many areas, such as diagnostic screening, drug monitoring, medical genetics and forensic investigations, various methods of DNA extraction have been adopted, including the standard Phenol-chloroform organic extraction, Chelex extraction, and FTA™ cards (Butler 2005c). Each technique has several advantages and disadvantages regarding the time and the effort needed to complete the analysis, the cost and the risk of health hazards, as well as the quality and the quantity of the DNA yield. Among these techniques, FTA™ cards have lately gained popularity among researchers due to their particular advantages over the other traditional methods of biological sample storage and DNA extraction.

FTA ™ cards are chemically treated paper, characterised by their ability to store a wide range of biological samples at room temperature for many years, on which the DNA remains stable and protected against degradation and microbial infections (Fujita and Kubo 2006). FTA cards can be loaded with a minimal volume of DNA-containing sample which is particularly important when dealing with precious, limited volume forensic samples. FTA cards provide a valuable tool for non-invasive sample collection that is particularly useful in high-throughput population studies which involve analysis of a large number of samples (Seah et al. 2003; Wong et al. 2007; Saunier et al. 2009). Owing to the non-invasive characteristics, FTA cards are more suitable for molecular epidemiologic studies involving paediatric populations (Beckett et al. 2008).

Sample collection and storage using FTA cards is a simple process which does not require sophisticated environmental conditions, and hence they can be used for self collection, field sample collection (Smith and Burgoyne 2004) and crime scene sample retrieval, ensuring safe delivery of samples to the laboratory. The ease of transportation of the DNA samples using FTA cards has a great impact on the population studies which deal with samples from different populations, and hence samples can be easily and safely transported between countries. FTA technology facilitates collection of samples from geographically isolated populations and enables transport by mail for molecular analysis (Guio et al. 2006).

Processing of DNA samples on FTA cards is a simple procedure which involves punching a small disc (1.2 or 2mm) out of the card, which then undergoes a simple washing *in situ* while the impurities being washed away. The DNA is retained on the disc and is ready to be included in the subsequent analysis with no need to transfer samples between tubes, and thus the risk of contamination is minimised.

The aim of this part of the study was to evaluate FTA cards for mitochondrial DNA testing and to assess their efficiency for collection, storage and transportation of DNA containing samples which would further aid the mitochondrial DNA analysis in the Egyptian population. The long term DNA storage capabilities of the FTA cards were investigated by performing DNA purification of blood samples stored on FTA cards for varying length of times (ranging from 6 months up to 4 years). In addition, the standard PCR protocol was modified to enable fast mitochondrial DNA analysis which would impose a direct impact on various applications.

## 3.2 Results

### 3.2.1 DNA purification of blood samples on FTA cards

#### *3.2.1.1 Evaluation of the DNA quantity contained in one processed disc of blood stained FTA cards*

The DNA quantity contained in one disc area of the blood loaded FTA cards was assessed (**Figure 3.1**). A good DNA yield (as evidenced by PCR amplification and gel electrophoresis) (**Figure 3.2**) was obtained after processing 1.2-mm discs of the stained areas of the FTA cards using method-1 of the washing procedures (**Methods; Table 2.1**), which followed the manufacturer's instructions, by performing three washes with FTA purification reagent and two washes with $TE^{-1}$ buffer. Washing was considered successful when the processed disc turned clear.

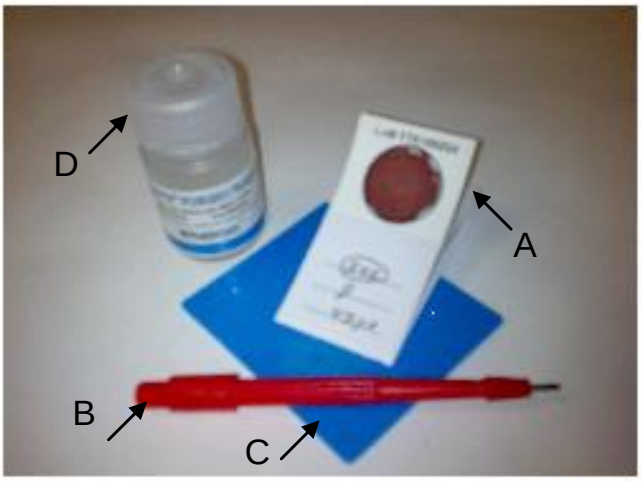



**Figure 3.1: Whatman FTA kit. A** is a whole blood sample on FTA micro-card (Cat.No.WB120310), **B** is a Harris Uni-Core Micro-Punch with 1.2 mm steel tip (Cat. No. WB1000028), **C** is a cutting mat, and **D** is the FTA Purification Reagent.

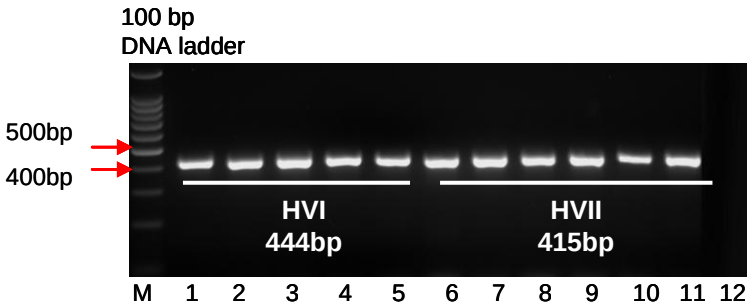



**Figure 3.2: PCR products of mitochondrial hypervariable regions (HVI and HVII) amplification from whole blood samples on FTA micro-cards.** PCR products were separated by electrophoresis on 2% agarose gel. A100 bp DNA ladder was used as a molecular marker (Lane M). Labelled lanes 1-5 represent HVI PCR products (444bp) of five Egyptian blood samples, while 6-10 represent HVII PCR products (415bp) of the same blood samples. Lanes 11 and 12 represent positive and negative control samples respectively.

This washing protocol, however, did not produce consistent results with all blood samples, which means that not all samples gave a good DNA yield after being processed using method-1. Similar results were obtained when bigger discs (2-mm) were considered (**Figure 3.3**).

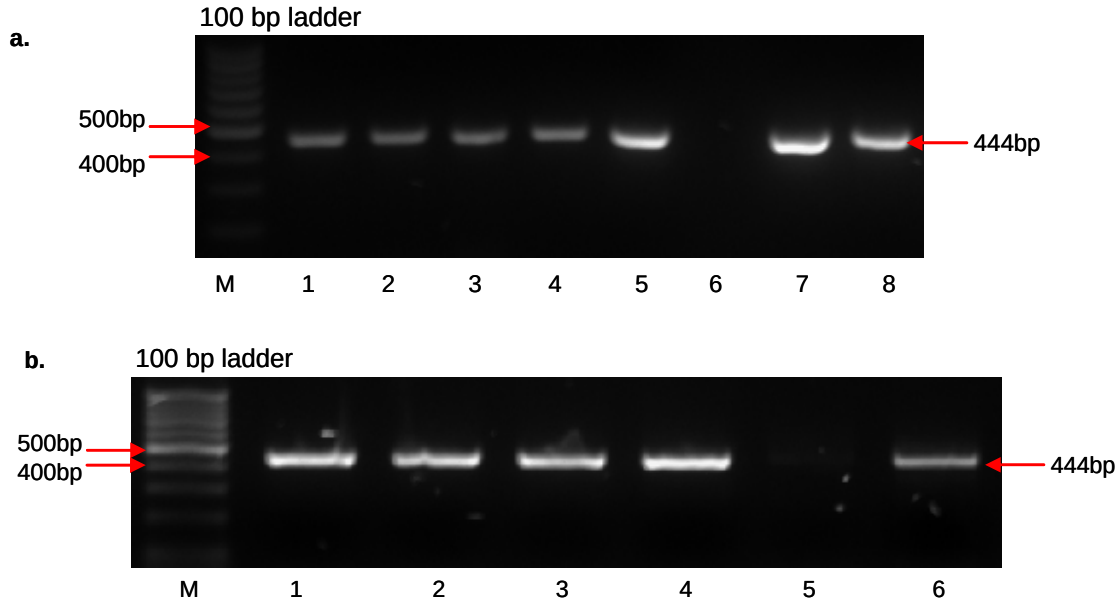



**Figure 3.3: PCR products of mitochondrial HVI region amplification following the use of one processed disc of the blood stained FTA cards as DNA template.** Samples were separated by electrophoresis on 2% agarose gel. A 100 bp DNA ladder was used as a molecular marker (Lane M). Gel (a) shows PCR products of HVI amplification following the use of 1.2-mm disc of each sample as DNA template. Lanes 1-7 contained PCR products of different Egyptian blood samples and Lane 8 contained a positive control. Gel (b) shows HVI PCR products following the use of 2-mm disc as a template. Lanes 1-5 contained HVI PCR products of different blood samples on FTA cards and lane 6 represents a positive control. Both gels showed failed PCR amplification of one sample which means that DNA purification of these samples from the FTA cards were not successful.

### *3.2.1.2 Evaluation of the DNA yield following the use of multiple processed blood stained discs as DNA template in one PCR reaction*

Following failure to amplify from some blood samples, and as a trial to increase the available DNA, varying numbers of 1.2-mm discs (2, 3 or 4 discs) were processed together (using method 1 for washing procedures) in a single tube and used collectively as a template in one PCR reaction. This approach showed an improvement in the DNA yield of some, but not all samples (**Figure 3.4**). These results proved that the previous unsuccessful PCR amplification of some samples was not due to limited DNA concentration on the disc but caused by presence of inhibitors in blood.

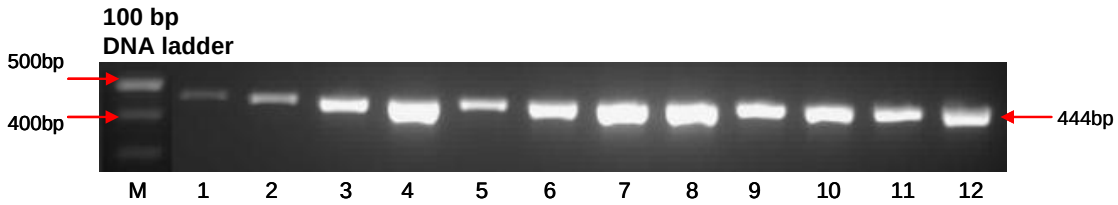



**Figure 3.4: PCR products of mitochondrial HVI region amplification following the use of three 1.2-mm discs of each sample in one PCR reaction.** PCR products of mitochondrial HVI amplification of 11 different blood samples (Lanes; 1-11) were separated by electrophoresis on 2% agarose gel. A 100bp DNA ladder was used as a molecular marker (Lane M) and a human placenta DNA was amplified as a positive control (Lane 12). Both the brightness and thickness of the DNA bands clearly demonstrated the variability of the DNA yield between different samples.

### *3.2.1.3 Investigation of the homogeny of the DNA distribution on the FTA cards following direct blood application*

In order to investigate the DNA distribution across the stained area of the card, 10 discs (1.2-mm) were randomly punched out of the same card and placed in separate PCR tubes. Discs were processed separately (using method-1) and then were placed in separate PCR reactions. The obtained results revealed that the DNA was almost equally distributed all over the FTA card, as discs of the same sample showed an equivalent DNA yield, as evidenced by PCR amplification of the mitochondrial hypervariable region (HVI) (**Figures 3.5)**. This approach also showed that the previous unsuccessful PCR amplification of some samples was not due to variable distribution of the blood samples on the cards, as the failure was consistent for all discs from the same card **(Figure 3.6**).

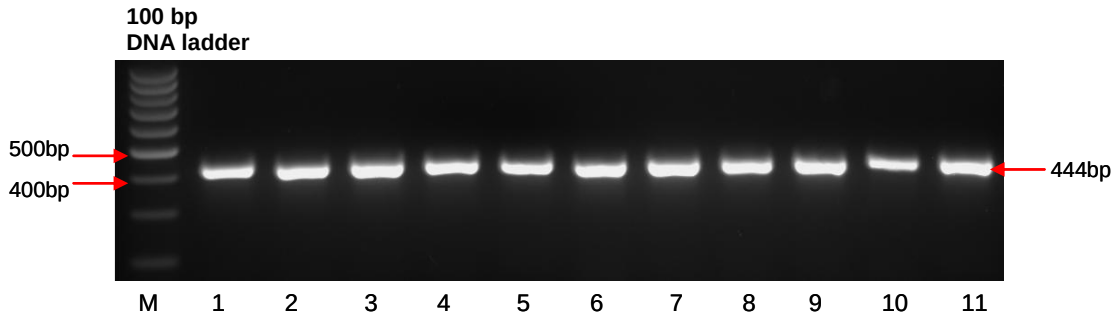



**Figure 3.5: PCR products of mitochondrial HVI amplification of 10 discs of the same blood sample on an FTA card.** HVI PCR products (444bp) were separated by electrophoresis on 2% agarose gel. Lane M contained the molecular marker (100 bp DNA ladder). Lanes 1-10 contained the PCR products obtained following the use of 10 discs of the same blood sample on the FTA card. A positive control sample was run in lane 11. A relatively same quality of the DNA yield was obtained from all discs as evidenced by the PCR amplification products.

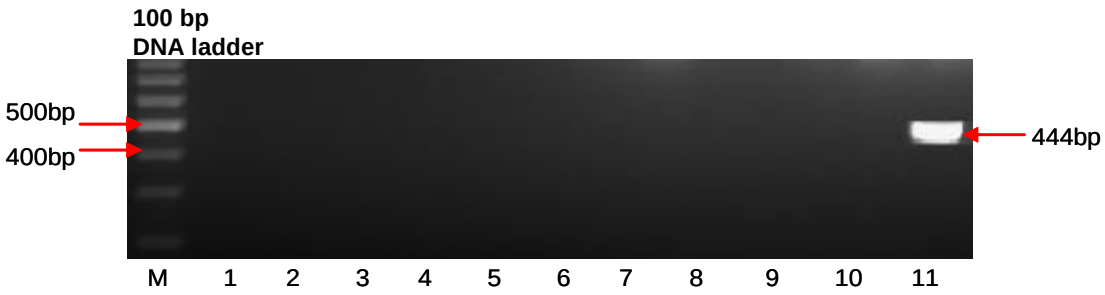



**Figure 3.6: PCR amplification products of 10 discs of a previously unsuccessfully tested blood sample on an FTA card.** HVI PCR products (444bp) were separated by electrophoresis on 2% agarose gel. Lane M contained the molecular marker (100 bp DNA ladder). Lanes 1-10 contained the PCR products obtained following the use of 10 discs of the same blood sample on the FTA card. A positive control sample was run in lane 11. None of the discs showed evidence of DNA amplification.

### 3.2.2 Modifications of the washing protocol for DNA purification of blood samples on the FTA cards

After short term storage of whole blood samples on the FTA cards (~6 months), method-1 of the washing procedures was used for DNA purification, according to the manufacturer's instructions, by performing three washes with FTA purification reagent and two washes with $TE^{-1}$ buffer, but was no longer effective in purifying DNA from cards, and discs did not become clear after these series of washes. PCR reactions almost always failed after this period of time. This failure was mainly attributed to the continuous presence of inhibitors in blood, such as haematin and any other non-DNA materials, which over time became fixed to the FTA cards and resisted all the washing steps. To overcome this increasing difficulty in removing all non-DNA materials that inhibited the PCR reactions, several modifications of the manufacturer's washing protocol were performed for DNA purification of these stored blood stained cards with the aim of removing such PCR inhibitors (in consultation with Whatman technical support).

These modified protocols for processing the cards (**Methods; Table 2.1**) are briefly mentioned below:

- ***Method 2***; an increase in the number and the duration of washes using the same purifying solutions.
- ***Method 3***; an intermediate washing step using 200-µl of 0.3M $Na_2CO_3$ for 5 minutes.
- ***Method 4***; an intermediate washing step using 500-µl of 5% SDS (Sodium Dodecyl Sulfate) for 5 minutes.

Unfortunately, all these different processing methods that were used for DNA purification from whole blood on FTA cards were not sufficient for elimination of haematin and other possible non-DNA materials that were consistently causing inhibition of the PCR.

### 3.2.3 Effect of modification of the PCR reaction mix in elimination of persistent PCR inhibition

Bovine Serum Albumin (BSA) is effective in overcoming inhibition in enzymatic reactions. Addition of BSA to the PCR reaction mix was shown to be effective in elimination of the PCR inhibition caused by the presence of haematin and/or other inhibitor(s). To further validate the effective role of BSA in eliminating such inhibitors, 6 months-36 months old samples on FTA cards were processed using the four different washing protocols and subjected to PCR amplification using the modified PCR master mix. Methods 2, 3, and 4 showed evidence of successful DNA amplification (and thus purification) without any significant difference between them. However, method-1 which was originally recommended by Whatman, failed to yield DNA from these stored samples **(Figure 3.7)**. Therefore, it was concluded that the longer the samples were being stored on the FTA cards, more stringent processing was required, and BSA became an essential component of the PCR reaction mix in such circumstances.

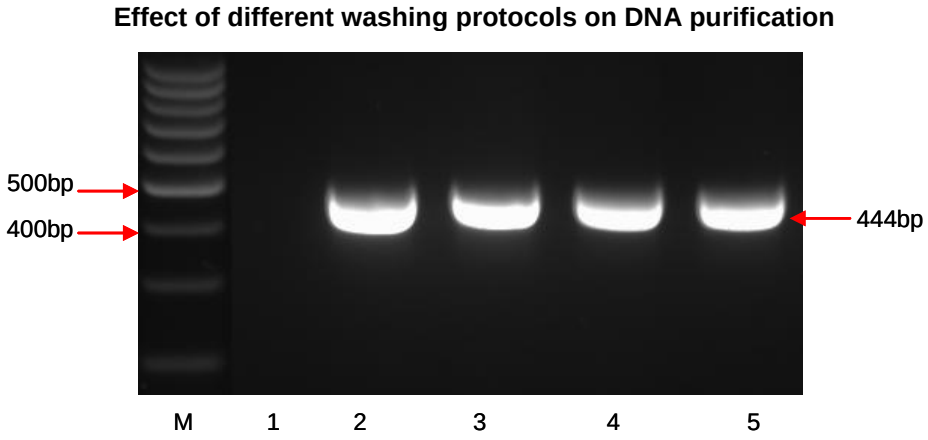



**Figure 3.7: PCR products of mitochondrial HVI amplification of whole blood samples on FTA cards following DNA purification using different washing protocols.** HVI PCR products were separated on 2% agarose gel. A 100 bp DNA ladder was used as a molecular marker (Lane M). Human placental DNA was amplified as a positive control and was loaded in lane 5. Lane 1 contained the HVI PCR product of a 36 months old blood sample on an FTA card processed using method-1 of the washing protocols. No evidence of DNA amplification was detected. Lanes 2-4 contained HVI PCR products (444bp) of the same long term stored blood sample after using methods 2, 3 and 4 of DNA purification respectively. The strong band intensities of all the three PCR products which are very similar to the positive control indicated the preservation of highly concentrated DNA in such long term stored samples.

### 3.2.4 Evaluation of the role of FTA cards in preservation of the DNA integrity

With the successful DNA purification following long term storage (up to 4 years) of blood samples on FTA cards, and verification of the effective role of such cards in preservation of the DNA quantity, further analysis was performed in order to assess their capabilities of preservation of the DNA. For this, the complete mitochondrial hypervariable region was amplified in one PCR reaction using HVI Forward and HVII Reverse primers (**Methods; Table 2.2**). The obtained results revealed successful amplification of 1024bp of the mitochondrial hypervariable region denoting preservation of good quality DNA of the blood samples on FTA cards following long term storage at room temperature (**Figure 3.8**).

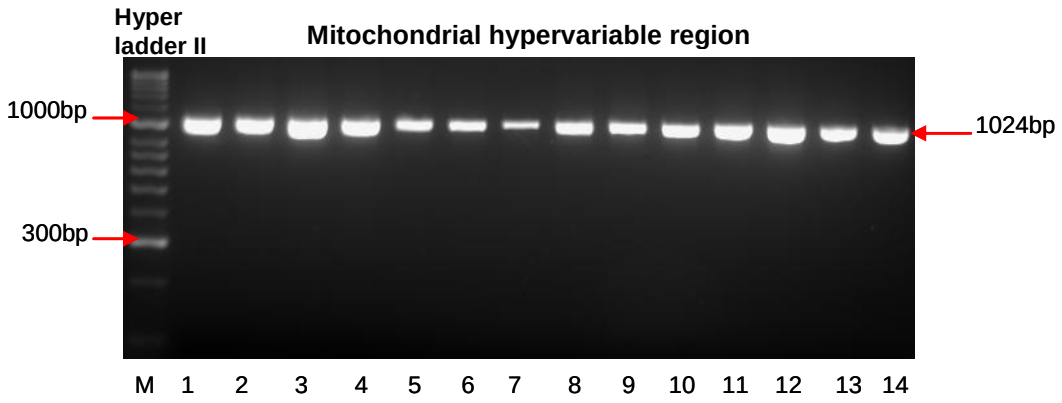



**Figure 3.8: PCR products of the whole mitochondrial hypervariable region as an indication of a good quality DNA preservation.** PCR products of mitochondrial hypervariable region amplification of 13 long term stored blood samples on FTA cards were separated by electrophoresis on 1% agarose gel. Hyper ladder II was used as a DNA molecular marker. Human placental blood was amplified as a positive control and was run in lane 14. Variation in the band intensities was observed between samples that might be attributed to the variation of blood sample volume applied to the FTA cards which had a direct impact on the presence and concentration of inhibitors.

## 3.3 Discussion

### 3.3.1 FTA™ cards as a method for biological sample collection and DNA purification

Following the successful use of FTA cards for blood sample collection in previous published studies (Mai et al. 2004; Park et al. 2008; Tomas et al. 2008; Sultan et al. 2009; Untoro et al. 2009), the current study supported the use of such technology as a simple and easy method for blood sample collection, storage, shipment, and DNA extraction. However, effort conducted in this study to counteract the continuous presence of PCR inhibitors in blood, highlights the guidelines of loading blood samples on FTA cards including (i) the sample volume should not exceed 200µl and (ii) the importance of the application of blood to the card to ensure even distribution over the defined sample area using a micropipette. These two key points are also essential to avoid overloading the matrix at any site, where the integrated chemicals will not be sufficient for preservation of nucleic acids. These crucial aspects of sample application to the FTA cards have never been addressed in previous studies which involved the use of FTA cards for collection of various range of biological samples. From the obtained results, it was evident that DNA purification could not be considered complete until the discs turned blank at the end of the washing steps.

### 3.3.2 FTA punch size as a template in PCR amplification

In the current study, the use of 1.2-mm and 2-mm FTA discs as DNA templates in PCR amplifications was investigated using different reaction volumes (25µl, 50µl and 100µl). The results shown indicate that 1.2-mm blood stained discs contain sufficient quantities of genomic DNA for PCR reaction volumes up to 100µl. These results are consistent with other studies which involved successful use of 1.2-mm discs of blood samples spotted on FTA cards (Seah et al. 2003; Suematsu and Isohashi 2006), whereas, others have used the bigger (2-mm) discs in their studies (Becker et al. 2004; Morrison et al. 2007; Pezzoli et al. 2007; Park et al. 2008). According to the results obtained throughout this study, it was clearly evident that the use of bigger discs did not improve the quality of the PCR products; meanwhile, the risk of involving more potential inhibitors in the reaction was increased. In addition, the use of bigger discs of samples stored on FTA cards in a single experiment means wasting more of the available samples. This is particularly important in forensic investigations which in many

circumstances deal with limiting DNA samples. Therefore, 1.2-mm discs are considered advantageous for forensic applications.

### 3.3.3 Cross contamination hazards when using FTA™ cards technology

According to the manufacturer's instructions, it was recommended to punch blank discs from non-stained areas of the cards in between different samples to avoid cross contamination hazards. In the current study, no DNA contamination was detected in the negative control punch discs, whereas others have shown punch crossover contamination when DNA was amplified from negative control punch discs (Jefferies et al. 2007). Accordingly, they recommended punching out six negative control discs using a sterile filter paper between each two different samples. That was not essentially required in the current work, since no DNA contamination was detected in any of the negative control samples throughout all the experiments.

### 3.3.4 Relief of persistent PCR inhibition

Several compounds in blood were suggested to be PCR inhibitors, namely, heme, leukocytes, and anticoagulants (Akane et al. 1994; Morata et al. 1998). Haemoglobin and lactoferrin are the major PCR inhibitors in erythrocytes and leukocytes, respectively (Al-Soud and Radstrom 2001). In this study, amplification inhibition was a persistent problem, particularly after long term storage of the blood samples on FTA cards. Several modifications of the washing protocol were performed; however, they did not solve the issue of PCR inhibition. The PCR facilitator Bovine Serum Albumin (BSA) is characterised by high binding potentials which enable it to bind to the possible PCR inhibitors in blood such as heme, which would rather bind to the DNA polymerase causing its inactivation and thus PCR inhibition (Kreader 1996). Based on the results obtained by this study as well as others (Akane et al. 1994; Powell et al. 1994; Al-Soud and Radstrom 2001), BSA was shown to be highly effective in reducing PCR inhibition. Subsequently, BSA was considered as an essential constituent of the PCR master mix when using Whatman FTA method for DNA extraction from blood.

### 3.3.5 Long term DNA storage potentials of the FTA™ cards

Following addressing the issue of persistent PCR inhibition, mitochondrial hypervariable regions were successfully amplified from samples stored on FTA cards over long periods of time (6 - 48 months), thus demonstrating that FTA cards are a suitable medium for long term storage and preservation of the nucleic acids at room temperature. This finding is in agreement with the data obtained by a previous study which involved direct STR typing of 1-2 year old blood samples stored on FTA cards (Park et al. 2008). Similarly, no variation of genomic DNA quantification was detected by analysis of the same blood samples over a period of one year after being stored on FTA cards (Pezzoli et al. 2007). Others have shown contradictory results when they could not process blood stained FTA discs after being stored for more than 9 months (Becker et al. 2004). Indeed, the current study faced a real difficulty in DNA purification from blood samples stored on FTA cards for longer than 6 months, nonetheless that did not hinder the progress of the project particularly after inclusion of the BSA as an essential constituent of the PCR reaction mix. Hence, it is concluded that the longer the samples are stored on the FTA cards, the more difficult it will be to purify DNA from them.

### 3.3.6 Conclusion

FTA cards were successfully used for blood sample collection, storage, shipment and DNA extraction. Several guidelines for the use of FTA cards for DNA purification were approached. This study justified and extended the utility of FTA cards to the routine forensic case work that would save a lot of labour for genomic DNA extraction, which was inevitable in conventional genetic screening. This relatively new method was shown to be rapid, simple and inexpensive, and it conferred a reduced risk for cross-contamination due to minimum manipulation of samples during extraction. Moreover, since FTA cards have the advantage of using minimal amounts of blood samples, they are recommended as a standard tool for genetic identification studies, especially when high throughput analysis is required.

# 4. ANALYSIS OF THE MITOCHONDRIAL HYPERVARIABLE REGIONS IN AN EGYPTIAN POPULATION SAMPLE

## 4.1 Introduction

The genetic sequence of the human mitochondrial DNA is of particular interest to forensic analyses that are concerned with human identification and population genetics (Budowle et al. 2002). For identification purposes, mtDNA is advantageous over the nuclear genome due to the fact that it is present in high copy number in each somatic cell, and thus is highly recoverable from degraded samples from which nuclear DNA is not (Robin and Wong 1988). This is particularly important in victim identification following mass disasters. In addition, mtDNA is inherited down the maternal line with non-recombination characteristics, and thus is treated as a single locus or a haplotype (Giles et al. 1980; Lightowlers et al. 1997; Primorac and Schanfield 2000). The primary sequence of the human mitochondrial genome (16.5 kb) was first determined in 1981 and was known as the Anderson Reference Sequence (Anderson et al. 1981). However, following improvement in DNA sequencing technology, in 1999, several minor errors were identified and corrected to derive the Cambridge Reference Sequence (CRS) (Andrews et al. 1999). This sequence was primarily derived from an individual of European descent and has been used as the standard reference sequence in all the mitochondrial DNA population studies. The majority of these population studies involved analysis of two hypervariable regions of the mitochondrial genome, HVI and HVII. These regions reside in the control region of the mitochondrial genome and have the highest mutation rate within the mitochondrial DNA, and thus more likely to differ between individuals (Howell et al. 2007). Variations in the hypervariable regions are mainly attributed to mutations and have been largely used to assess the diversity and heterogeneity among and within different population groups (Asari et al. 2007; Mabuchi et al. 2007; Gonzalez et al. 2008; Turchi et al. 2009).

The MtDNA database is growing rapidly, with more than six thousand complete mitochondrial DNA sequences available (www9). The frequency of different mitochondrial DNA haplotypes within a population group can be used to assess the weight of evidence of matching between an evidence sample and a known reference sample (Budowle et al. 1999; Holland and Parsons 1999). Analysis of the mtDNA haplotypes within population groups gives information about the maternal ancestry of these populations and the possible admixture between different populations irrespective

of the geographic barriers. Despite being continuously growing, the high quality mtDNA database is mainly constituted by mitochondrial sequences originating from the West Eurasia and the Asian sub-continent, whereas populations from the African continent are still underrepresented (Turchi et al. 2009).

The aim was to analyse the mitochondrial DNA hypervariable regions of 261 adult Egyptians, and to establish a database of Egyptian mitochondrial DNA haplotypes. These mitochondrial haplotypes can then be used for assessment of the mitochondrial nucleotide diversity, and heterogeneity among the studied population group. For this, the mitochondrial HVI and HVII regions were amplified from blood samples on FTA cards and sequenced between the nucleotide positions 16024-16365 and 73-340, respectively. Mitochondrial haplotypes were determined and statistical measures including nucleotide diversity and random match probability were calculated. The most common polymorphic sites in the HVI region were identified and the frequency of each SNP was calculated. The utility of the mitochondrial hypervariable region data for forensic DNA analysis was investigated. The Egyptian mitochondrial haplotypes were assigned to known haplogroups which served to identify the maternal ancestry of the modern Egyptian population and to uncover their genetic relationships to other neighbour populations.

## 4.2 Results

### 4.2.1 PCR amplification of the mitochondrial hypervariable regions (HVI and HVII)

DNA was successfully purified from all the Egyptian blood samples (n=261) that were previously stored on FTA cards. A 1.2-mm processed dried disc of each blood sample was used as a template in all the subsequent analysis. The present study involved analysis of blood samples stored on FTA cards from 6 months up to 48 months old. Mitochondrial hypervariable regions (HVI and HVII) were PCR amplified between the nucleotide positions 16024-16365 and 73-340, respectively according to the Cambridge Reference Sequence (CRS) (accession number NC_012920) using primers indicated in Materials and Methods.

#### *4.2.1.1 Conventional PCR*

Mitochondrial hypervariable regions were amplified separately using the modified PCR master mix together with the traditional short primers (~20mer) (**Methods; Table 2.2**). HVI F and HVI R primers were used for HVI amplification producing a product size of 444bp and for HVII amplification; HVII F and HVII R primers were used producing a 415bp PCR product. Conventional thermal cyclic conditions, as described in section 2.2.1.3, were used and the run was completed in two hours. PCR products were analysed by agarose gel electrophoresis (**Figure 4.1**).

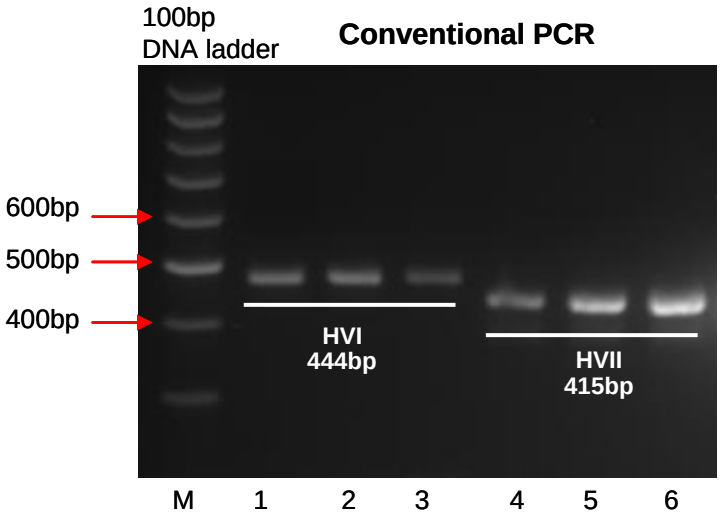



**Figure 4.1: Conventional PCR products of the mitochondrial hypervariable regions amplification.** PCR products were separated by electrophoresis on 2% agarose gel. 100 bp DNA ladder was used as a molecular marker; Lane M. HVI PCR products (444bp) of three Egyptian blood samples on FTA cards were separated in lanes 1-3 and HVII PCR products (415bp) of the same samples were separated in lanes 4-6.

Following successful DNA purification from blood samples on FTA cards and amplification of the mitochondrial hypervariable regions in two separate PCR reactions using a duplicate set of processed discs of each sample, both mitochondrial hypervariable regions were simultaneously amplified in one PCR reaction using one disc of each sample instead of two. The use of a single disc for mitochondrial hypervariable regions amplification aimed at reduction of the initial volume of the DNA template that would be required for the analysis, which would have a major impact on forensic investigations where the amount of a sample may be an issue. The two mitochondrial hypervariable regions were amplified as a multiplex using the two region specific pairs of primers (HVI F, HVI R, HVII F and HVII R; **Table 2.2**). The results of the multiplex PCR showed preferential amplification of the HVI region over the HVII region which was less successfully amplified (**Figure 4.2**). In order to assess the quality of the DNA in samples stored on FTA cards, the whole mitochondrial hypervariable region between the mitochondrial nucleotide positions 16024 and 340 was amplified as one fragment (1024bp) using HVI F and HVII R primers (**Figure 4.3**).

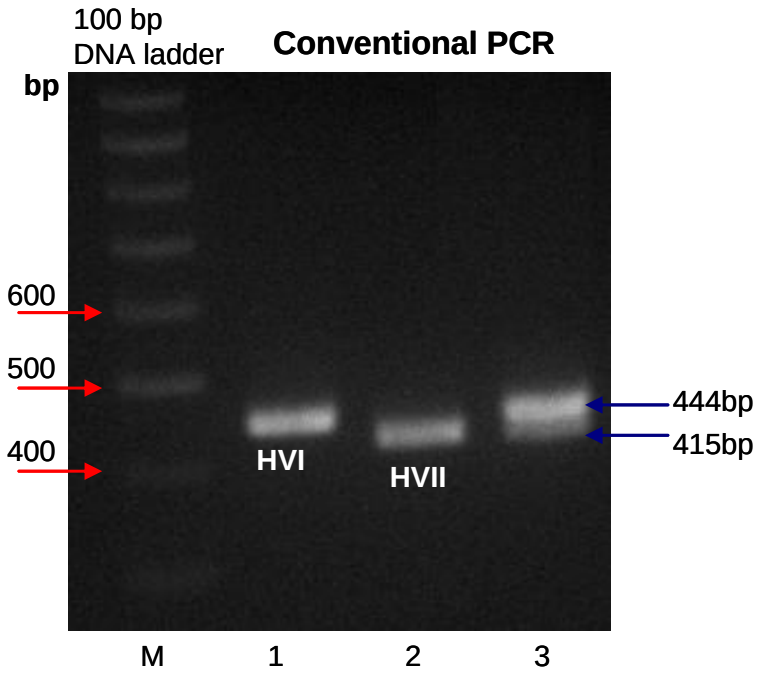



**Figure 4.2: PCR products of multiplex and individual amplification of mitochondrial HVI and HVII regions.** PCR products were separated by electrophoresis on 2% agarose gel. A 100 bp DNA ladder was used as a molecular marker. Lanes 1 and 2 contained PCR products of HVI (444bp) and HVII (415bp) respectively. Lane 3 contained products of a multiplex PCR in which the two pairs of short primers were used resulting in two bands of sizes 444 and 415bp corresponding to HVI and HVII regions respectively.

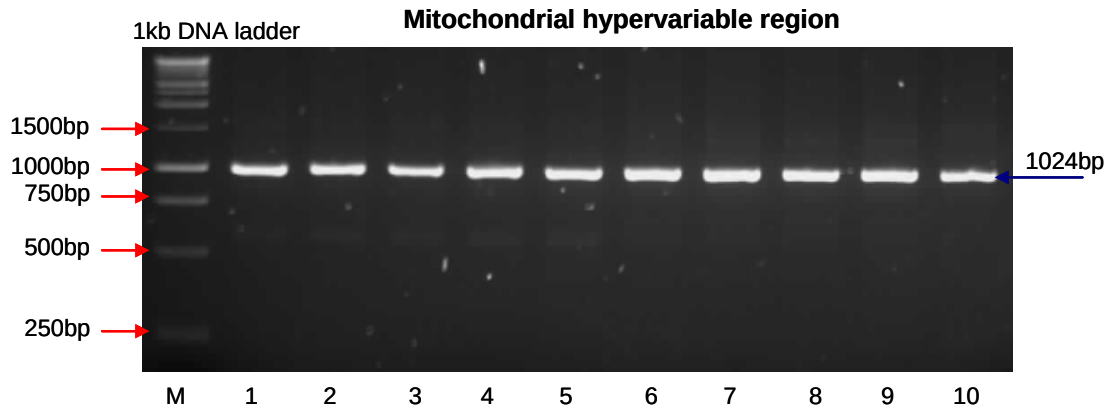



**Figure 4.3: Analytical gel showing products of PCR amplification of the mitochondrial control region as one fragment.** The whole mitochondrial hypervariable region was PCR amplified as one fragment and the PCR products were separated on 1% agarose gel. 1Kb DNA ladder was used as a molecular marker in lane M. Lanes 1-9 contained the PCR amplification products (1024bp) of nine Egyptian blood samples. Lane 10 contained the positive control in which human placental DNA was used as a template.

#### *4.2.1.2 Rapid PCR amplification of mitochondrial hypervariable regions*

Following successful PCR amplification of the mitochondrial hypervariable regions using the conventional thermal cyclic conditions which required approximately two hours to be completed, there was a demand to modify these cyclic conditions and to develop a rapid method for amplification of the mitochondrial hypervariable regions which would have a direct impact on forensic investigations. For this, two pairs of newly designed long primers (~30mer) which were characterised by high Tm values were used (**Methods; Table 2.2**). These high Tm values enable the primers to anneal at high temperature at which the DNA polymerase was highly active, therefore it was possible to combine the annealing and the extension steps at 72$^{o}$C, and hence to shift from three-step PCR protocol to two-step PCR protocol. This results in considerable reduction in the total run time. In addition, the use of high initial denaturation temperature at 98$^{o}$C as recommended by Sullivan et al.(2006), helped to reduce the initial denaturation time down to 30 seconds instead of 5 minutes, while ensuring proper denaturation of DNA template on FTA cards. Three more parameters of the PCR cyclic conditions were optimised to achieve the target of rapid amplification of the mitochondrial hypervariable regions including, the length of time for denaturation at the beginning of each cycle, the number of cycles, and the duration of the final elongation.

#### 4.2.1.2.1 Optimising the length of time for denaturation at the beginning of each cycle

Modification of the length of time for denaturation was performed via amplification of the mitochondrial hypervariable region I (HVI) of the same blood sample on FTA card in six different PCR tubes (1.2-mm processed disc in each tube as DNA template) with identical PCR constituents. These amplification reactions were performed under the same thermal cyclic conditions except for the cyclic denaturation time which was subjected to a gradient reduction from 45 seconds down to 5 seconds while keeping all the other thermal cyclic conditions unchanged. The results showed successful amplification of the mitochondrial HVI in all the six PCR reactions, without any significant difference between the products of any of them **(Figure 4.4)**. Therefore, a decision was made to use 5 seconds as the appropriate time for denaturation.

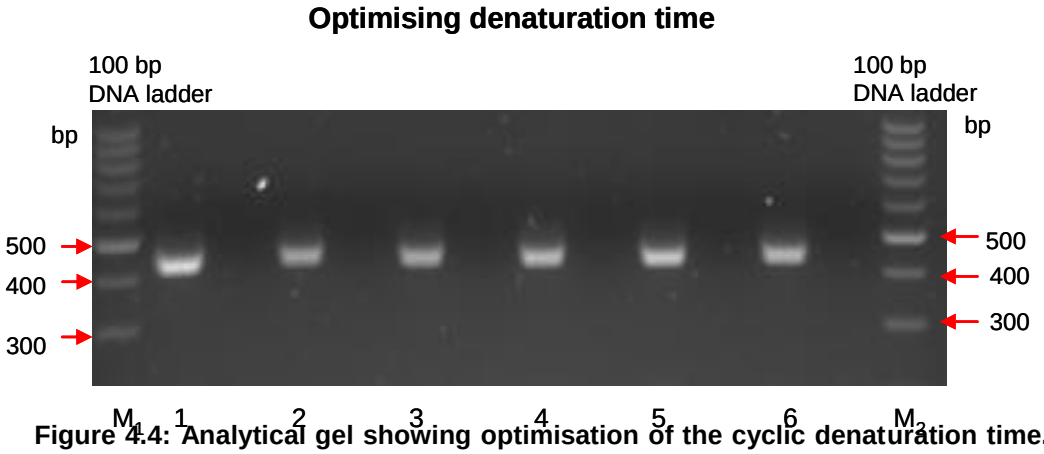



**Figure 4.4: Analytical gel showing optimisation of the cyclic denaturation time.** HVI PCR products were separated by electrophoresis on 2% agarose gel. A 100 bp DNA ladder was used as a molecular marker (Lanes; $M_1$and $M_2$). Lanes 1-6 show HVI PCR products (size= 444bp) following the use of different cyclic denaturation times as follows; lane 1 (45 seconds), lane 2 (30 seconds), lane 3 (20 seconds), lane 4 (15 seconds), lane 5 (10 seconds) and lane 6 (5 seconds). No significant difference was detected among the different PCR products.

#### 4.2.1.2.2 Optimising the number of cycles

To further reduce the total run time of the PCR program, different numbers of cycles were tested. Mitochondrial HVII region of one blood sample on FTA card was amplified in five different PCR tubes (1.2-mm processed disc in each tube as DNA template) with identical PCR constituents. The thermal cyclic conditions of all the five amplification reactions were kept the same except for the number of cycles. Different numbers of cycles were tested from 28 cycles up to 40 cycles in order to determine the minimal number of cycles that would give good quality PCR products and hence the

shortest possible run time of the PCR. The obtained results showed that 30 cycles were appropriate to obtain an adequate amplification product without smearing (**Figure 4.5**).

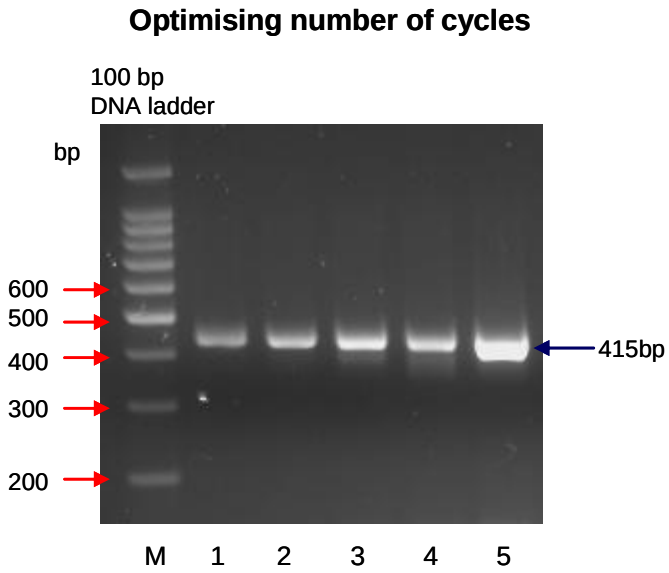



**Figure 4.5: Analytical gel showing optimisation of the number of cycles.** HVII PCR products were separated by electrophoresis on 2% agarose gel. A 100 bp DNA ladder was used as a molecular marker; lane M. Lanes 1-5 show HVII PCR products (size= 415bp) following the use of different number of cycles as follows; lane 1 (28 cycles), lane 2 (30 cycles), lane 3 (32 cycles), lane 4 (35 cycles) and lane 5 (40 cycles). The bigger the number of cycles used, the more smearing was observed on the gel. The use of 28 cycles produced a relatively less bright band than all the others reflecting less DNA concentration. The use of 30 cycles produced a relatively bright band (lane 2) reflecting a good DNA concentration with no smearing.

#### 4.2.1.2.3 Optimising the duration of final extension

As a final step to further reduce the PCR run time, the duration of final elongation was assessed, by performing two different PCR reactions for multiplex amplification of the mitochondrial hypervariable regions (HVI and HVII) of a blood sample on FTA card using a 1.2-mm processed disc in each tube. All the PCR parameters were kept the same for both reactions, including the thermal cyclic conditions as well as the master mix constituents which included the whole set of the four long primers (**Methods; Table 2.2**). The only difference between the two reactions was the time of final extension which was tested at 1 and 5 minutes. The obtained results showed that there was no significant difference between the PCR products of the two reactions (**Figure 4.6**), and hence it was concluded that 1 minute was an adequate time for the final extension.

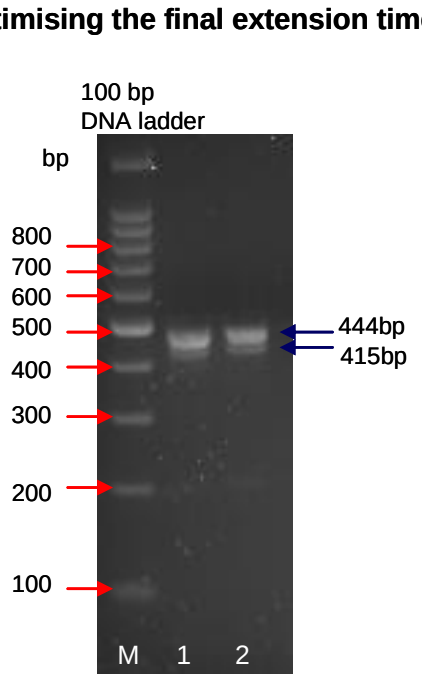



**Figure 4.6: Analytical gel showing optimisation of the final extension time.** PCR products of multiplex amplification of the mitochondrial hypervariable regions HVI and HVII were separated by electrophoresis on 2% agarose gel. A 100 bp DNA ladder was used as a molecular marker; lane M. PCR products show 2 bands of sizes; 444bp and 415bp corresponding to HVI and HVII regions respectively. Lane 1 shows multiplex PCR products following 1 minute of final extension and lane 2 shows multiplex PCR products following 5 minutes of final extension. No significant difference was detected between the products of the two reactions.

#### 4.2.1.2.4 Final protocol of rapid PCR

From the previously mentioned experimental work, a new protocol for rapid PCR amplification of the mitochondrial hypervariable regions, HVI and HVII was developed. This protocol enabled amplification of the mitochondrial hypervariable regions from blood samples on FTA cards within a period of 30 minutes. It is also worth noting that this rapid protocol was developed using Gotaq DNA polymerase which is commonly used in the conventional PCR programs. This rapid protocol was reproducible using different thermal cyclers (TECHNE GENUIS TC 3000X, TECHNE GENUIS TC-512, LABNET, and PIKO®) demonstrating that the protocol was not machine dependent. Optimised thermal cyclic conditions for the rapid PCR were summarised in **Table 4.1**. It is now possible to amplify mitochondrial HVI and HVII regions in a short period of time which will aid forensic investigations as well as the development of protocols for undergraduate teaching.

**Table 4.1: A summary of rapid PCR protocol.**

| PCR stage | Temperature | Duration |
|---|---|---|
| 1 cycle of<br>Initial denaturation | <br>98°C | <br>30 sec |
| 30 cycles of<br>Denaturation<br>Annealing/ Extension | <br>92°C<br>72°C | <br>5 sec<br>30 sec |
| 1 cycle of<br>Final Extension | <br>72°C | <br>1 min |
| Total run time | | ~ 30 min |

### *4.2.1.3 DNA purification of the PCR products*

Following successful amplification of the mitochondrial hypervariable regions of all the Egyptian blood samples (n=261), PCR products were purified using Wizard® SV Gel and PCR Clean-Up Purification System (Promega) (Methods, section 2.2.1.5). Success of the purification process was verified for all samples by electrophoresis on a 1% agarose gel (**Figure 4.7**).

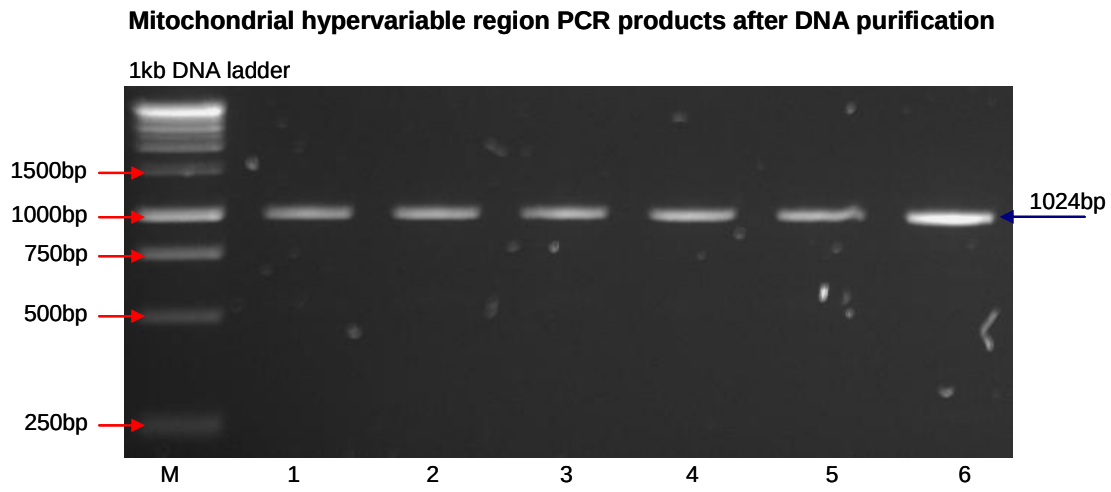



**Figure 4.7: Mitochondrial hypervariable region PCR products after DNA purification.** PCR products of the whole hypervariable region were purified using Wizard® SV Gel and PCR Purification System (Promega) and 3µl of each sample was separated by electrophoresis on 1 % agarose gel. 1Kb DNA ladder was used as a molecular marker; lane M. Lanes 1-6 contained purified hypervariable region PCR products (1024bp) of six Egyptian blood samples. The band intensities revealed successful DNA purification.

### 4.2.2 Mitochondrial hypervariable region sequence analysis and haplotype determination

#### *4.2.2.1 Sequencing results of the mitochondrial hypervariable regions*

Purified PCR products of the mitochondrial hypervariable regions from all the Egyptian blood samples were sequenced using the two pairs of short primers (HVI F, HVI R, HVII F and HVII R) that have been used in the conventional PCR (**Methods, Table 2.2**) and samples were run on an ABI 3730 capillary sequencer. The sequencing data was obtained using Applied Biosystems Sequence Scanner Software v1.0. It was possible to obtain complete HVI sequences between the mitochondrial nucleotide positions 16024 and 16365 for all samples (n=261). However, a reliable sequence of the complete HVI specified region was occasionally difficult to obtain via one sequencing reaction, in which case a partial sequence was the result. This partial sequence was always obtained whenever a length polymorphism was detected at the poly cytosine stretch spanning the nucleotides 16184-16188 and 16190-16193. In the Cambridge Reference Sequence a thymine base is present at nt16189 in the middle of the poly C stretch separating 5Cs before and 4Cs after. When a T16189C transition occurs, it results in a stretch of 10 or more cytosines in a row which further predisposes to length polymorphism due to replication slippage. Such length polymorphism results in a mixture of length variants (mtDNA copies of varying lengths) which always disrupt the sequence downstream. These variants are either already present in the original DNA sample or generated in the sequencing reaction itself. Two other mutations (A16182C and A16183C) may occasionally be present that further increases the length of the poly C stretch and hence augment the sequencing difficulty (**Figure 4.8a and 4.8b**). To tackle this problem, sequencing of the two strands was always required. Moreover, double sequencing of the same strand in two separate sequencing reactions was occasionally performed to ensure that the mtDNA sequence information was reliable.

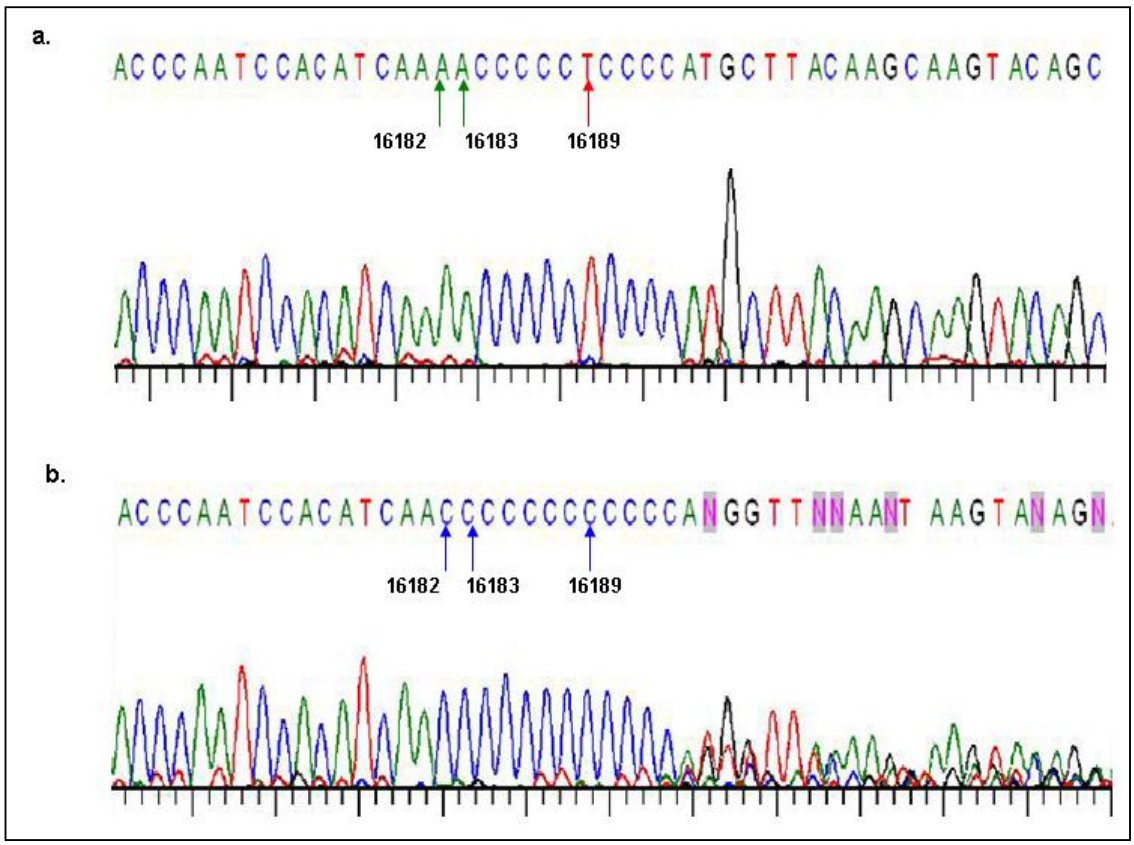



**Figure 4.8a: Snapshots of HVI electropherograms between the nucleotide positions 16166 and 16214 showing the difference between the Cambridge Reference Sequence and sequences with length polymorphisms.** Electropherogram (a) represents the CRS between the nucleotide positions 16166 and 16214 having 16182A, 16183A and 16189T and showing a good quality sequence after the C-stretch zone. Electropherogram (b) represents an example of HVI sequences having A16182C, A16183C and T16189C mutations and showing a quick drop of the sequence quality immediately after the poly C stretch due to presence of a mixture of length variants. Electropherograms were extracted form the sequencing results obtained during the present study.

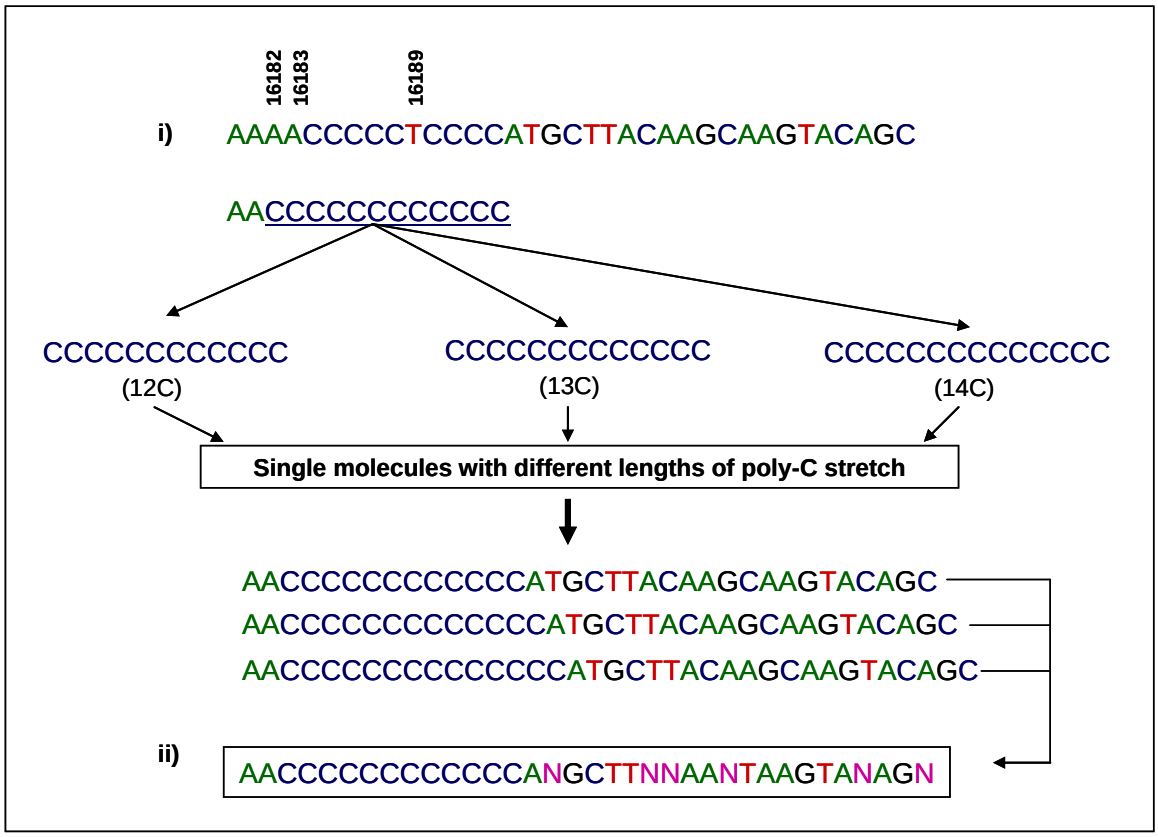



**Figure 4.8b: Variations in the poly-C stretch of the mitochondrial hypervariable region I causing sequencing difficulty.** i) is the Cambridge Reference Sequence between nucleotides 16180-16214. i) is not a problem with sequencing, and ii) represents 1 variant, which then does cause difficulties for the sequencing. These sequencing difficulties are due to mitochondrial mutations at positions 16182, 16183, and 16189 that lead to a long poly-C stretch (≥ 12C). Different molecules of the same DNA sequence show different lengths of poly-C either produced by length heteroplasmy or generated during the sequencing reaction itself. The resulting sequence ii) is a mixture of these varying-length molecules with a remarkable reduction in the sequence quality immediately after the poly-C stretch.

For the mitochondrial hypervariable region II analysis, mtDNA was sequenced between the nucleotide positions 73 − 340 in approximately 30% of samples (n=78). Sequencing of the two strands was performed to ensure a reliable sequencing data especially for samples with length polymorphism at the poly C stretch region spanning the nucleotides 303-309 and 311-315 which always caused sequencing difficulty (**Figure 4.9**). Based on the analysed sequencing data (see section 4.2.2.4), the calculated random match probability of the HVII haplotypes and the actual role that HVII haplotypes played in the mitochondrial haplogroup assignment (which was an integral part of this project), the protocol of the hypervariable region analysis was modified to only include nucleotide position 73 instead of the whole extension of the HVII region. In the CRS there is an adenosine base at nucleotide position 73 (Andrews et al. 1999), which

represents a G to A mutation that occurred during modern human evolution from the most recent maternal ancestor. 73A is considered as a specific characteristic for the mitochondrial major haplogroup R0 and all its descendants including haplogroup H to which the CRS belongs (Brandstaetter et al. 2008; Alvarez-Iglesias et al. 2009). Interestingly, a similar G to A mutation was reported in one of the most ancient sub-Saharan African lineages, L0a'b'f (Behar et al. 2008b). Hence, the nucleotide position 73 was determined for all the Egyptian individuals (n=261) and used in conjunction with the HVI haplotypes for the mitochondrial haplogroup assignment and the nucleotide diversity assessment in the Egyptian population.

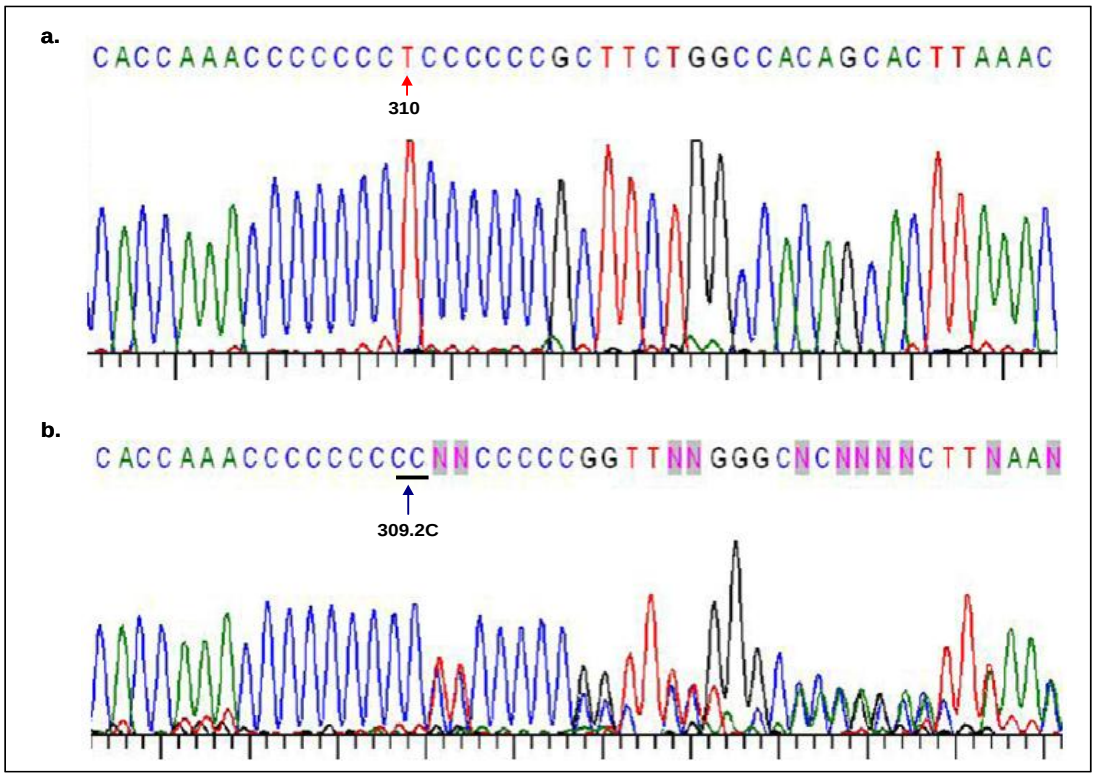



**Figure 4.9: Snapshots of HVII electropherograms between the nucleotide positions 296 and 338 showing the impact of length polymorphisms on the sequencing results.** Electropherogram (a) represents an example of HVII sequence between the nucleotide positions 296 and 338 showing an identical sequence to the CRS between the nucleotides 303-310 with an overall good quality sequence. Electropherogram (b) represents an example of samples with two cytosine insertions after nt309 leading to a complete disruption of the sequence beyond that poly C stretch due to presence of a mixture of length variants. Electropherograms were extracted from the sequencing results obtained during the present study.

#### *4.2.2.2 Analysis of the sequencing results*

Mitochondrial HVI and HVII sequencing results were aligned against the Cambridge Reference Sequence. The CRS is the primary sequence of the human mitochondrial genome which was first determined in 1981 (Anderson et al. 1981) and following the advancement of the DNA sequencing technology, minor errors were identified and corrected in 1999 (Andrews et al. 1999). This reference sequence belongs to an individual of European descent, and has been used as a standard reference in all the mitochondrial DNA population studies. The sequence alignment was performed using the CLUSTALW analysis tool of Biology Workbench (www5). Mutations were identified whenever variations from the CRS were detected (**Figure 4.10**), and the mutation sites were numbered according to the revised version of the CRS obtained from the NCBI database (accession number NC_012920) and hence the mitochondrial haplotypes were concluded. HVI haplotypes and the nucleotide base at position 73 were determined for all samples that were unambiguously sequenced during the present study (n=261), whereas, complete HVII haplotypes were only determined for 30% of samples (n=78) **(Table 4.4)**.

**CLUSTALW analysis of Egyptian mitochondrial HVI sequences relative to the CRS (TEXSHADE)**

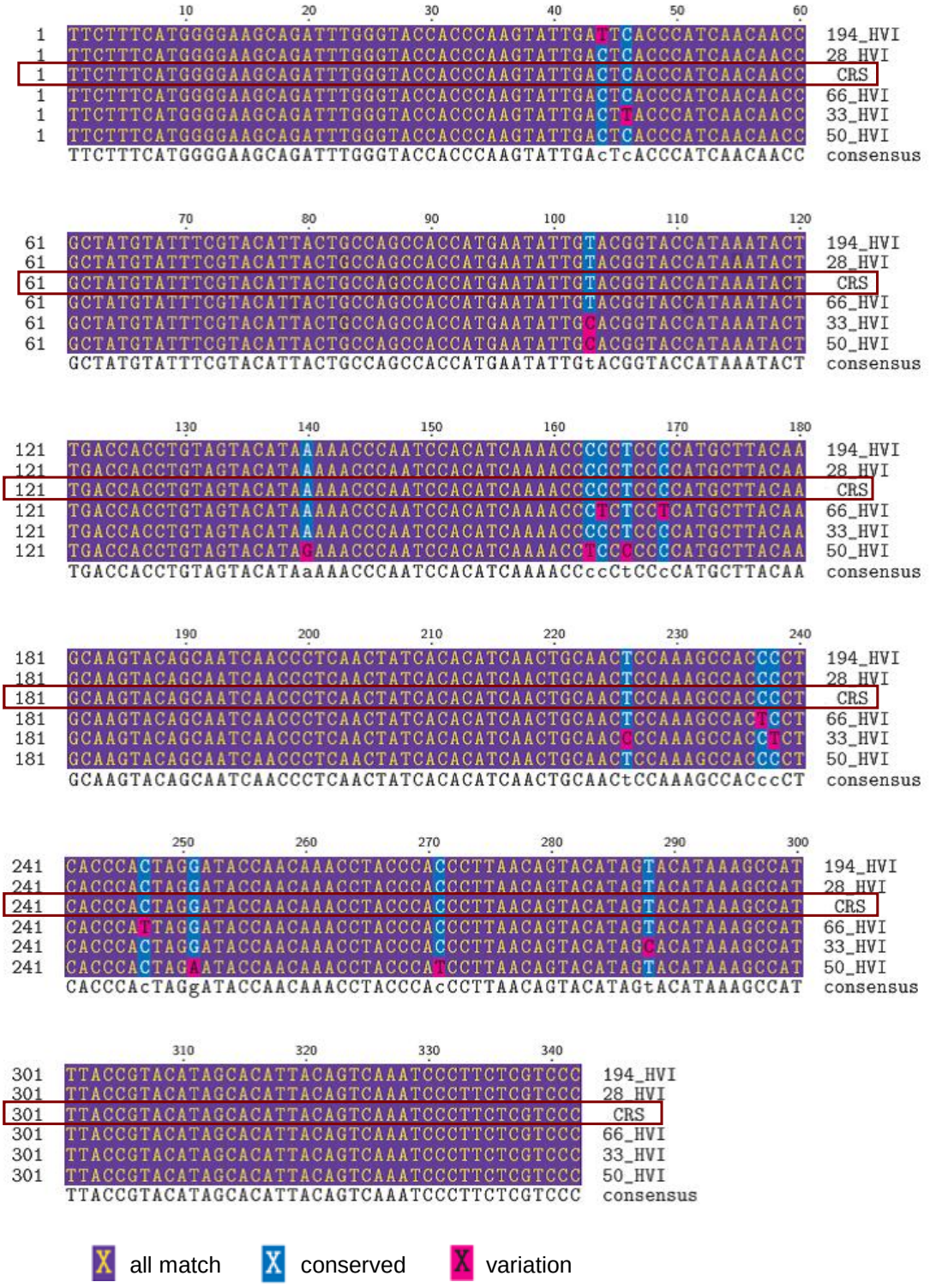



**Figure 4.10: CLUSTALW analysis of five Egyptian mtDNA HVI sequences relative to the CRS using TEXSHADE tool at the Biology Workbench.** Five Egyptian mtDNA sequences (sample numbers 28, 33, 50, 66, and 194) between the nucleotides 16024−16365 were aligned against the revised version of the Cambridge Reference Sequence. Dark brown boxes outline the CRS. The consensus line is shown at the bottom with an upper case character in case of all sequences are matching to the CRS and a lower case character indicates a conserved base with few variations between sequences.

#### *4.2.2.3 Sequence variations in the mitochondrial hypervariable regions*

Overall, a total of 113 polymorphic sites were reported in the mitochondrial HVI region (16024-16365) which characterised a total of 187 different Egyptian HVI haplotypes. These polymorphic sites varied greatly in their frequency in the analysed mtDNA sequences, ranging from occurring in 1 individual to 99 individuals. The average number of times for a variable character to occur was 9.2 times. Sites that showed the highest variability in the HVI region of the Egyptian mtDNA sequences, found in 20 individuals or more, included 16126, 16129, 16172, 16183, 16189, 16223, 16249, 16278, 16292, 16294, 16311, and 16362. Sites that showed lower rate of variability but still reported in 10 individuals or more included 16069, 16093, 16145, 16163, 16182, 16186, 16187, 16192, 16209, 16210, 16224, 16256, 16270, 16296, and 16309 (**Table 4.2**). In the HVII analysed sequences, a total of 42 polymorphic sites were reported which enabled identification of 62 different haplotypes. An A263G mutation was found in all sequences. The next common polymorphic sites in the HVII region were 73, 146, 152, 195, and 303-315 occurring in 20 individuals or more. All singletons (a polymorphic site represented by a single variant sequence) were checked carefully to exclude sequencing errors.

**Table 4.2: A full list of the polymorphic sites reported in the mitochondrial HVI sequences of 261 Egyptian individuals.** Sites were numbered according to the Cambridge Reference Sequence numbering system (Andrews et al. 1999). N is the number of occurrences of the polymorphic site in the analysed mtDNA sequences (n=261) with the frequency percentage shown between brackets. Sites were ordered according to their frequency starting by the most common. Polymorphic sites reported in 10 individuals or more are shown in bold and italic.

| N (Frequency %) | Polymorphic site | N (Frequency %) | Polymorphic site | N (Frequency %) | Polymorphic site |
|---|---|---|---|---|---|
| ***99 (37.93)*** | ***16223*** | 5 (1.92) | 16124 | 2 (0.77) | 16248 |
| ***86 (32.95)*** | ***16189*** | 5 | 16153 | 2 | 16260 |
| ***71 (27.20)*** | ***16311*** | 5 | 16167 | 2 | 16298 |
| ***65 (24.90)*** | ***16126*** | 5 | 16188 | 2 | 16299 |
| ***50 (19.16)*** | ***16294*** | 5 | 16219 | 2 | 16300 |
| ***45 (17.24)*** | ***16183*** | 5 | 16230 | 2 | 16301 |
| ***41 (15.70)*** | ***16278*** | 5 | 16261 | 2 | 16356 |
| ***29 (11.11)*** | ***16129*** | 5 | 16274 | 1 (0.38) | 16037 |
| ***28 (10.73)*** | ***16292*** | 5 | 16324 | 1 | 16048 |
| ***22 (8.43)*** | ***16172*** | 5 | 16343 | 1 | 16081 |
| ***22*** | ***16249*** | 4 (1.53) | 16104 | 1 | 16092 |
| ***21 (8.05)*** | ***16362*** | 4 | 16166 | 1 | 16096 |
| ***19 (7.28)*** | ***16192*** | 4 | 16169 | 1 | 16107 |
| ***19*** | ***16270*** | 4 | 16181 | 1 | 16131 |
| ***18 (6.90)*** | ***16186*** | 4 | 16265 | 1 | 16138 |
| ***17 (6.51)*** | ***16163*** | 4 | 16271 | 1 | 16150 |
| ***16 (6.13)*** | ***16209*** | 4 | 16291 | 1 | 16156 |
| ***14 (5.36)*** | ***16182*** | 4 | 16304 | 1 | 16173 |
| ***14*** | ***16224*** | 3 (1.15) | 16051 | 1 | 16201 |
| ***12 (4.60)*** | ***16187*** | 3 | 16139 | 1 | 16221 |
| ***11 (4.21)*** | ***16069*** | 3 | 16218 | 1 | 16222 |
| ***11*** | ***16093*** | 3 | 16234 | 1 | 16231 |
| ***11*** | ***16256*** | 3 | 16266 | 1 | 16233 |
| ***11*** | ***16309*** | 3 | 16318 | 1 | 16254 |
| ***10 (3.83)*** | ***16145*** | 3 | 16325 | 1 | 16269 |
| ***10*** | ***16210*** | 2 (0.77) | 16086 | 1 | 16286 |
| ***10*** | ***16296*** | 2 | 16134 | 1 | 16287 |
| 9 (3.45) | 16168 | 2 | 16136 | 1 | 16288 |
| 9 | 16176 | 2 | 16147 | 1 | 16305 |
| 9 | 16320 | 2 | 16164 | 1 | 16313 |
| 8 (3.07) | 16327 | 2 | 16174 | 1 | 16316 |
| 8 | 16355 | 2 | 16179 | 1 | 16317 |
| 7 (2.68) | 16148 | 2 | 16207 | 1 | 16319 |
| 7 | 16293 | 2 | 16213 | 1 | 16352 |
| 7 | 16359 | 2 | 16214 | 1 | 16354 |
| 6 (2.30) | 16067 | 2 | 16215 | 1 | 16357 |
| 6 | 16264 | 2 | 16239 | 1 | 16365 |
| 5 (1.92) | 16111 | 2 | 16240 | | |

#### *4.2.2.4 Mitochondrial DNA haplotype analysis*

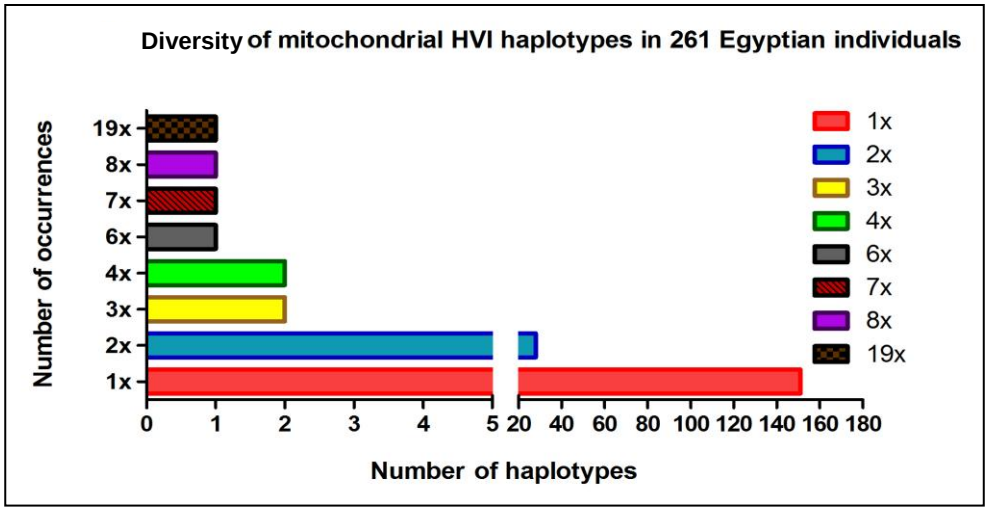
Following determination of the mitochondrial DNA sequence variations in the hypervariable regions, HVI haplotypes were analysed both individually and in combination with the available HVII data. Among the 261 analysed sequences and using the HVI data alone, a total of 187 different haplotypes were observed, of which 151 (~81%) were seen only once during this study. Such a large number of unique mitochondrial haplotypes obtained via analysis of the mitochondrial HVI region alone indicates high mtDNA diversity and supports the use of the mitochondrial HVI region analysis for the identification purposes. The other 36 HVI haplotypes were shared between multiple individuals, of which 28 types were observed twice (**Table 4.3**) and (**Figure 4.11a**). The most commonly observed HVI haplotype, seen in 19 individuals, was identical to the CRS between the nucleotide positions 16024-16365. Among the 78 analysed HVII sequences, a total of 62 different haplotypes were reported, of which 51 (~82%) were unique to single individuals. The remaining 11 haplotypes were shared between multiple individuals. The most frequent HVII haplotypes, seen in 4 individuals, were (73G, 143A, 146C, 152C, 195C, 263G, 309.1C, 315.1C) and (73G, 263G, 309.1C, 315.1C) (**Tables 4.3 and 4.4**) (**Figure 4.11a**). Combination of 261 HVI sequences and 78 HVII sequences revealed a total of 207 different mitochondrial haplotypes, of which 183 (~88%) were unique to single individuals and 24 haplotypes were observed more than once. In 78 individuals for whom both HVI and HVII sequences were available, no shared haplotypes were detected (**Table 4.3**) and (**Figure 4.11b**). Hence, it was concluded that the HVII data when combined with the HVI data increased the power of discrimination and enabled unique identification in 78 individuals.

**Table 4.3: Diversity of mtDNA haplotypes in an Egyptian population sample.** Number and frequency of the haplotypes were calculated for HVI and HVII data sets individually and in combination. Numbers of unique haplotypes in each category was shown in bold and italic.

| Number of times observed | HVI (16024−16365) (n=261) | | HVII (73-340) (n=78) | | HVI + HVII (n=78) | | HVI +HVII (n=261)* | |
|---|---|---|---|---|---|---|---|---|
| | Number of haplotypes | Total | Number of haplotypes | Total | Number of haplotypes | Total | Number of haplotypes | Total |
| 1 | ***151*** | 151 | ***51*** | 51 | ***78*** | 78 | ***183*** | 183 |
| 2 | 28 | 56 | 8 | 16 | 0 | 0 | 15 | 30 |
| 3 | 2 | 6 | 1 | 3 | 0 | 0 | 4 | 12 |
| 4 | 2 | 8 | 2 | 8 | 0 | 0 | 0 | 0 |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 15 |
| 6 | 1 | 6 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7 | 1 | 7 | 0 | 0 | 0 | 0 | 1 | 7 |
| 8 | 1 | 8 | 0 | 0 | 0 | 0 | 0 | 0 |
| 14** | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 14 |
| 19** | 1 | 19 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Total** | 187 | 261 | 62 | 78 | 78 | 78 | 207 | 261 |

* This data set involved HVI sequences and nucleotide position 73 for all samples (n=261) combined with 30% HVII data coverage (n=78).
** The only HVI haplotype which was observed in more than 7 individuals was identical to the CRS with no mutations between the nucleotide positions 16024-16365.

**a.**




**b.**

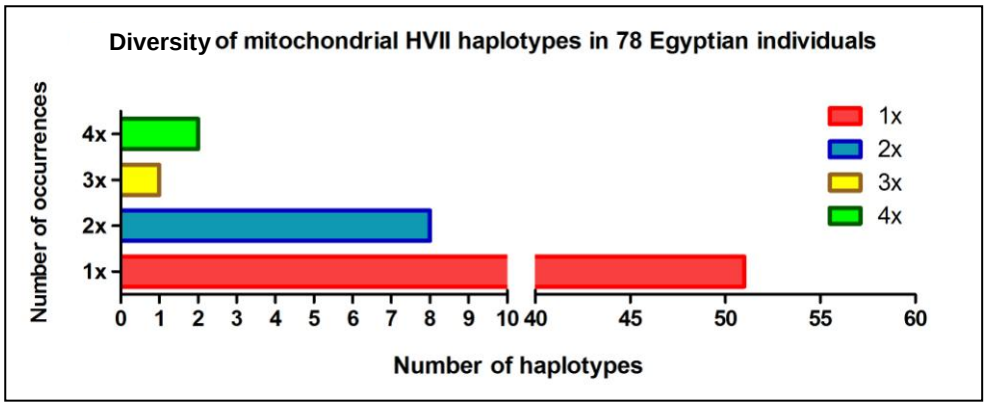
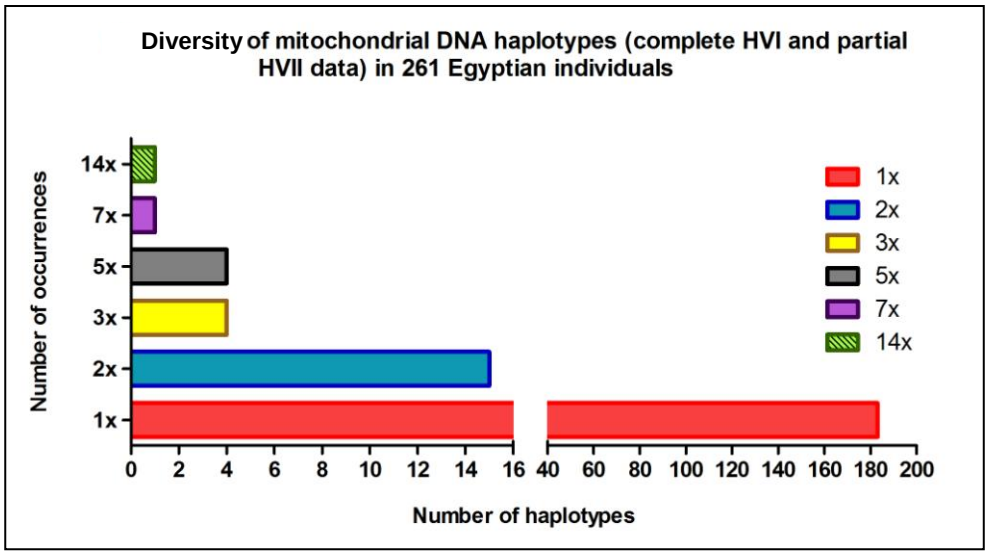



**Figure 4.11a: Diversity of mitochondrial HVI and HVII haplotypes in an Egyptian population sample.** a) represents distribution of mitochondrial HVI haplotypes (nt16024-16365) in 261 adult Egyptians demonstrating high mtDNA diversity with approximately 81% of haplotypes unique to single individuals. The most frequent HVI haplotype was found in 19 individuals and was identical to the CRS in the specified HVI region. b) represents distribution of mitochondrial HVII haplotypes (nt73-340) in 78 adult Egyptians with approximately 82% of haplotypes unique to single individuals. The most frequent haplotypes were shared by four individuals.

**a.**




**b.**

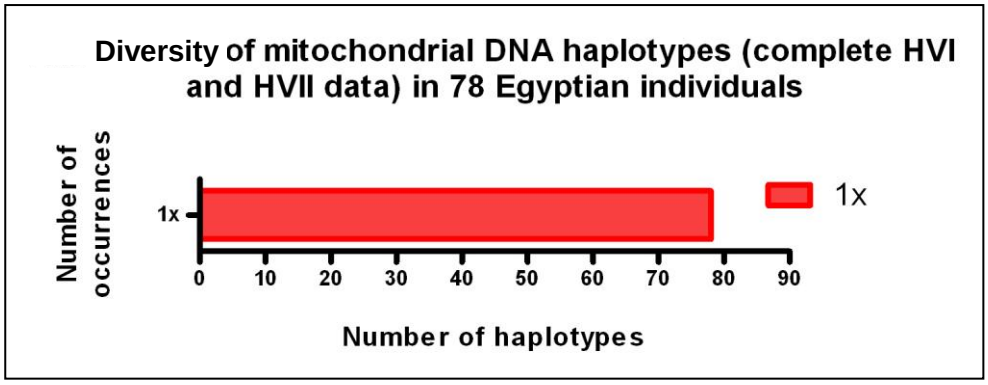



**Figure 4.11b: Diversity of mitochondrial DNA haplotypes in an Egyptian population sample following combination of HVI and HVII data.** a) represents distribution of mitochondrial DNA haplotypes following combination of 261 HVI sequences (nt16024-16365) and 78 HVII sequences (nt73-340). Approximately 88% of haplotypes were unique to single individuals in the analysed population sample. The most frequent haplotype was found in 14 individuals and was identical to the CRS in the specified HVI region. b) represents distribution of mitochondrial DNA haplotypes in 78 individuals for whom complete HVI (16024-16365) and HVII (73-340) sequences were determined. No shared haplotypes were detected and hence unique identification in 78 individuals was achieved via combination of HVI and HVII sequencing data.

**Table 4.4: Mitochondrial DNA sequence polymorphisms in the hypervariable region and haplogroup associations in a population sample of 261 Egyptian individuals.**

| # Individuals[(a)] | Haplogroup[(b)] | HVI haplotype (16024−16365)[(c)] | Nucleotide at postion 73 | HVII haplotype (74−340)[(c,d)] |
|---|---|---|---|---|
| 14 | R0 | CRS | ***A*** | ND |
| 7 | U5a | 16192T, ***16256T, 16270T*** | G | ND |
| 5 | R* | CRS | G | ND |
| 5 | N1b | ***16145A, 16176G***, 16223T, ***16390A*** | G | ND |
| 5 | L3f1b | ***16209C,*** 16223T, ***16292T, 16311C*** | G | ND |
| 3 | T1a | ***16126C, 16163G, 16186T, 16189C***, 16274A, ***16294T*** | G | ND |
| 3 | T1a | ***16126C, 16163G, 16186T, 16189C, 16294T*** | G | ND |
| 3 | M1 | ***16129A***, 16183C, ***16189C***, ***16249C, 16311C*** | G | ND |
| 3 | L1b | ***16126C***, 16187T, 16189C, 16223T, ***16264T, 16270T***, 16278T, ***16293G***, 16311C | G | ND |
| 2 | H5 | 16166G, 16189C,16304C | ***A*** | ND |
| 2 | R0 | 16129A | ***A*** | ND |
| 2 | R0 | 16356C | ***A*** | ND |
| 2 | R0a1a | ***16126C, 16355T, 16362C*** | ***A*** | ND |
| 2 | HV1 | ***16067T*** | ***A*** | ND |
| 2 | HV1 | ***16067T***, 16183G, 16327A | ***A*** | ND |
| 2 | J | ***16069T, 16126C*** | G | ND |
| 2 | J1b | ***16069T, 16126C***, 16249C, ***16261T***, 16311C | G | ND |
| 2 | T1a | 16111A, ***16126C, 16163G, 16186T, 16189C, 16294T*** | G | ND |
| 2 | T2c | ***16126C, 16292T***, 16294T | G | ND |
| 2 | M1 | ***16129A***, 16183C, ***16189C***, 16223T, ***16249C, 16311C***, 16359C | G | ND |
| 2 | M1 | 16183C, ***16189C, 16249C, 16311C***, 16327T | G | ND |
| 2 | L3f1b | 16172C, ***16209C***, 16223T, ***16292T, 16311C*** | G | ND |
| 2 | L3e2 | 16223T, ***16320T*** | G | ND |
| 2 | R11'B | ***16183C, 16189C***, 16223T, ***16311C*** | G | ND |
| 1 | R0 | 16167T, 16269C, 16313T | ***A*** | 85A, 263G, 309.1C, 315.1C |
| 1 | R0 | 16187A | ***A*** | 99G, 241C, 263G, 309.1C, 315.1C |
| 1 | R0 | 16271C, 16311C | ***A*** | 146C, 263G, 315.1C |
| 1 | R0 | 16210G, 16224C, 16292T | ***A*** | 263G, 309.1C, 315.1C |
| 1 | R0 | 16126C, 16168T, 16172C, 16325C | ***A*** | 263G, 315.1C |
| 1 | R0 | 16233G | ***A*** | ND |
| 1 | R0 | 16219G | ***A*** | ND |
| 1 | R0 | 16164G | ***A*** | ND |
| 1 | R0 | 16167T | ***A*** | ND |
| 1 | R0 | 16210G, 16224C | ***A*** | ND |
| 1 | R0 | 16147A | ***A*** | ND |
| 1 | R0 | 16172C | ***A*** | ND |
| 1 | R0 | 16051G | ***A*** | ND |
| 1 | R0 | 16183G, 16327A | ***A*** | ND |
| 1 | R0 | 16355 | ***A*** | ND |
| 1 | R0 | 16210G, 16224C, 16292T | ***A*** | ND |

**Table 4.4:** (Continued)

| # Individuals[a] | Haplogroup[b] | HVI haplotype (16024−16365)[c] | Nucleotide at postion 73 | HVII haplotype (74−340)[c,d] |
|---|---|---|---|---|
| 1 | R0 | 16129A, 16265C, 16362C | ***A*** | ND |
| 1 | R0 | 16270T | ***A*** | ND |
| 1 | H20 | ***16218T***, 16316G | ***A*** | ND |
| 1 | H2a2b | 16218T, ***16291T*** | ***A*** | 143A, 263G, 309.1C, 315.1C |
| 1 | H20 | ***16218T*** | ***A*** | 263G, 315.1C |
| 1 | H14a | ***16256T***, 16311C, ***16352C*** | ***A*** | ND |
| 1 | H6b | 16234T, ***16300G, 16362C***. | ***A*** | ND |
| 1 | R0a | ***16126C, 16362C*** | ***A*** | ND |
| 1 | R0a | ***16126C***, 16150T, 16168T, 16266T, ***16362C*** | ***A*** | 99G, 263G, 309.1C, 315.1C |
| 1 | R0a | ***16126C***, 16271C, 16291T, ***16362C*** | ***A*** | 204C, 263G, 315.1C |
| 1 | R0a | ***16126C***, 16271C, 16291T, ***16362C*** | ***A*** | ND |
| 1 | R0a | 16107T, ***16126C***, 16172C, 16288C, 16291T, 16301T, ***16362C*** | ***A*** | ND |
| 1 | R0a | ***16126C***, 16192A, 16278T, ***16362C*** | ***A*** | ND |
| 1 | R0a1a | ***16126C, 16355T, 16362C*** | ***A*** | 146C, 263G, 309.1C, 315.1C |
| 1 | HV0 | 16093T, 16126C, 16221T, ***16298C*** | ***A*** | ***72C,*** 195C, 263G, 3069.1C, 315.1C |
| 1 | HV1 | ***16067T*** | ***A*** | 263G, 309.1C, 315.1C |
| 1 | HV1 | ***16067T***, 16129A | ***A*** | ND |
| 1 | R11'B | 16181C, ***16182C***, ***16183C, 16189C*** | G | 146C, 153G, 263G, 309.1C, 315.1C |
| 1 | R11'B | 16093C, 16104T, 16181C, **16182C, *16183C, 16189C*** | G | ND |
| 1 | R11'B | 16104C, 16181C, ***16182C, 16183C,16189C*** | G | 146C, 153G, 228A, 263G, 309.1C, 315.1C |
| 1 | R11'B | ***16182C, 16183C, 16189C***, 16223T | G | ND |
| 1 | R11'B | 16104T, 16136C, ***16183C, 16189C***, 16223T | G | ND |
| 1 | R* | 16169T, 16183C, 16189C, 16223T, 16278T | ***G*** | 146C, 153G, 263G, 315.1C |
| 1 | R* | 16210G, 16224C, 16292T | ***G*** | ND |
| 1 | R* | 16096d, 16270T | ***G*** | ND |
| 1 | R | 16124C, 16209C, ***16223T*** | ***G*** | ND |
| 1 | R | 16092C, 16138G, ***16223T*** | ***G*** | ND |
| 1 | R | 16126C, ***16223T***, 16240 | ***G*** | 152C, 195C, 262T, 263G, 315.1C |
| 1 | R | ***16223T*** | ***G*** | 146C, 244G, 263G, 315.1C |
| 1 | R | ***16223T*** | ***G*** | 152C, 189G, 195C, 263G, 309.1C, 315.1C |
| 1 | R | 16111A, 16210G, ***16223T*** | G | ND |
| 1 | R | 16166G, 16172C, ***16223T*** | G | ND |
| 1 | R | 16167T, ***16223T*** | ***G*** | 150T, 189G, 200G, 263G, 315.1C |
| 1 | R | ***16223T,*** 16264T, 16299G | ***G*** | 150T, 254C, 263G, 309.1C, 315.1C |
| 1 | R | 16182A, 16183A | ***G*** | 143A, 146C, 152C, 195C, 263G, 315.1C |
| 1 | R | 16081G, 16270T | ***G*** | ND |
| 1 | T | ***16126C, 16294T*** | G | 263G, 309.1C, 315.1C |
| 1 | T | ***16126C***, 16136C, ***16294T***, 16324C | G | ND |
| 1 | T | ***16126C. 16294T*** | G | ND |
| 1 | T1 | 16051G, ***16126C***, ***16189C***, **16294T** | G | ND |
| 1 | T1a | 16037G, 16093C, ***16126C, 16163G, 16186T, 16189C*** | G | 152C, 263G, 309.1C, 315.1C |
| 1 | T1a | ***16126C, 16186T, 16189C, 16294T*** | G | 200G, 263G, 309.1C, 315.1C |
| 1 | T1a | ***16126C***, 16139, ***16163G, 16186T, 16189C, 16294T*** | G | 263G, 309.1C, 315.1C |

**Table 4.4:** (Continued)

| # Individuals[a] | Haplogroup[b] | HVI haplotype (16024−16365)[c] | Nucleotide at postion 73 | HVII haplotype (74−340)[c,d] |
|---|---|---|---|---|
| 1 | T1a | ***16126C***, 16139, ***16163G, 16186T, 16189C, 16294T*** | G | 152C, 263G, 309.1C, 315.1C |
| 1 | T1a | ***16126C, 16163G, 16186T, 16189C, 16294T*** | G | 263G, 309.1C, 315.1C |
| 1 | T1a | ***16126C, 16163G***, 16172C, ***16186T, 16189C, 16298C*** | G | 263G, 272G, 309.1C, 315.1C |
| 1 | T1a | ***16126C,*** 16139, ***16163G, 16186T, 16189C***, 16270T, 16271C, ***16294T*** | G | 227, 263G, 309.1C, 315.1C |
| 1 | T1a | ***16126C, 16163G, 16186T, 16189C*** | G | ND |
| 1 | T1a | ***16126C, 16163G***, 16172C, ***16186T, 16189C,*** 16274A, ***16294T*** | G | ND |
| 1 | T1a | ***16126C, 16163G***, 16172C, ***16186T, 16189C,16294T*** | G | ND |
| 1 | T2 | ***16126C, 16294T, 16296T*** | G | ND |
| 1 | T2 | ***16126C, 16294T, 16296T,*** 16324C | G | ND |
| 1 | T2 | ***16126C***, 16168T, ***16294T, 16296T,*** 16311C. | G | ND |
| 1 | T2 | **16126C,** 16183C, 16189C, ***16294T, 16296T*** | | ND |
| 1 | T2b | ***16126C***, 16189C, ***16294T, 16296T, 16304C***, 16365T | G | ND |
| 1 | T2b3a | ***16126C, 16292T, 16294T, 16296T, 16304C*** | G | 263G, 309.1C, 315.1C |
| 1 | T2c | ***16126C, 16292T, 16294T, 16296T*** | G | ND |
| 1 | T2c | ***16126C,*** 16174T, ***16292T, 16294T*** | G | ND |
| 1 | T2e | ***16126C, 16153A,*** 16192T, ***16294T, 16296T,*** 16311C | G | 150T, 185A, 199C, 263G, 309.1C, 315.1C |
| 1 | T2e | ***16126C, 16153A, 16294T, 16296T,*** 16324C | G | ND |
| 1 | T2e | ***16126C, 16153A, 16294T, 16296T*** | G | ND |
| 1 | T2e | ***16126C, 16153A,*** 16278T, ***16294T*** | G | ND |
| 1 | J | ***16069T, 16126C,*** 16169T | G | 146C, 150T, 195C, 235G, 263G, 295T, 309.1C, 315.1C |
| 1 | J | ***16069T, 16126C,*** 16311C | G | ND |
| 1 | J | ***16069T, 16126C,*** 16362C | G | ND |
| 1 | J | ***16069T, 16126C,*** 16248T, 16362C | G | ND |
| 1 | J | ***16069T, 16126C,*** 16193T, 16239T, 16300G, 16309G | G | ND |
| 1 | J1b | ***16069T, 16126C***, 16145A, 16222T, ***16261T,*** 16274A, 16311C | G | ND |
| 1 | J1b | ***16069T, 16126C,*** 16145A, ***16261T*** | G | ND |
| 1 | U1b | 16111T, ***16249C, 16311C, 16327T*** | G | ND |
| 1 | U1a'c | 16183C, ***16189C,16249C***, 16292T, 16319A, 16357C | G | 199C, 263G, 285T, 309.1C, 315.1C |
| 1 | U3 | 16168T, 16192T, ***16343G*** | G | ***150T***, 263G, 309.1C, 315.1C |
| 1 | U3 | ***16343G*** | G | ***150T,*** 263G, 315.1C |
| 1 | U3 | ***16343G*** | G | ND |
| 1 | U3 | 16131C, 16168T, ***16343G*** | G | ND |
| 1 | U3 | 16210G, 16224C, 16292T, ***16343G*** | G | ND |
| 1 | U5 | 16167C, ***16270T***, 16311C | G | ND |
| 1 | U5 | 16187T, 16192T, 16260T, ***16270T*** | G | ND |
| 1 | U5a | 16192T, ***16256T, 16270T*** | G | 263G, 315.1C |
| 1 | U6a | ***16172C, 16219G***, ***16278T*** | G | 103A, 195C, 263G, 3096.1C, 315.1C |
| 1 | U6a | 16189C, ***16219G, 16278T*** | G | 103A, 146C, 195C, 207A, 263G, 309.1C, 315.1C |
| 1 | U6a1a | ***16172C***, ***16183C, 16189C***, ***16219G, 16278T*** | G | ND |
| 1 | U6a4 | ***16134G, 16172C, 16219G, 16278T*** | G | ND |
| 1 | U7 | 16309G, ***16318T***, 16362C | G | 152C, 263G, 309d |

**Table 4.4:** (Continued)

| # Individuals[a] | Haplogroup[b] | HVI haplotype (16024−16365)[c] | Nucleotide at postion 73 | HVII haplotype (74−340)[c,d] |
|---|---|---|---|---|
| 1 | U7 | 16126C, 16148T, 16309G, ***16318T*** | G | 99G, 146C, 151T, 152C, 195C, 263G, 315.1C |
| 1 | U7 | ***16318T*** | G | ND |
| 1 | U8b | 16179T, 16183C, ***16189C, 16234T***, 16324C | G | 195C, 263G, 309.1C, 315.1C, 324G |
| 1 | U8b | 16179T, 16183C, ***16189C, 16234T***, 16324C | G | ND |
| 1 | K | ***16224C, 16311C*** | G | 152C, 185A, 263G, 295A, 315.1C |
| 1 | K | ***16224C, 16311C*** | G | 263G, 315.1C |
| 1 | K | 16174T, ***16224C, 16311C*** | G | 146C, 263G, 315.1C |
| 1 | K | 16124C, 16172C, 16187T, ***16224C, 16311C*** | G | 152C, 195C, 263G, 309.1C, 315.1C |
| 1 | K | 16176T, 16210G, ***16224C*** | G | ND |
| 1 | K | 16172C, ***16224C***, 16256T, 16261T, ***16311C*** | G | ND |
| 1 | K1a1a | ***16093C, 16224C, 16311C*** | G | 195C, 263G, 315.1C |
| 1 | K1a1a | ***16093C, 16224C, 16311C*** | G | 263G, 315.1C |
| 1 | N/ L3* | 16210G, ***16223T***, 16270T, ***16311C*** | G | ND |
| 1 | N/ L3* | ***16223T, 16311C*** | G | ***199C, 204C***, 214G, 250C, 263G, 315.1C |
| 1 | N1a | ***16172C,*** 16223T, 16293T, 16311C, ***16355T,*** 16362C | G | 146C, 244G, 263G, 315.1C |
| 1 | N1a | ***16147G, 16172C***, 16223T, ***16248T, 16355T*** | G | ND |
| 1 | N1b | ***16145A, 16176G***, 16223T, ***16390A*** | G | ***152C***, 204C, 263G, 315.1C |
| 1 | N1b | ***16145A, 16176G,*** 16223T, ***16390A*** | G | ***152C***, 263G, 309.1C, 315.1C |
| 1 | N1b | ***16145A, 16176G***, 16210G, 16223T, ***16390A*** | G | ND |
| 1 | N1c | ***16201T,*** 16223T, ***16265G*** | G | ND |
| 1 | I1 | ***16129A***, 16223T, 16311C, ***16391A*** | G | 152C, ***199C***, 204C, ***250C***, 263G, 315.1C |
| 1 | W | ***16223T, 16292T*** | G | ND |
| 1 | W6 | ***16223T, 16292T***, ***16325C*** | G | ND |
| 1 | W6 | 16192T, ***16223T, 16292T, 16325C*** | G | ND |
| 1 | X | ***16183C, 16189C***, 16223T, 16254G, 16266T, ***16278T*** | G | ND |
| 1 | X | ***16183C, 16189C***, 16223T, ***16278T***, 16305G | G | ND |
| 1 | X | 16134T, ***16189C***, 16223T, ***16278T*** | G | ***153G***, 189G, 195C, 225A, 263G, 309.1C, 309.1C |
| 1 | X | 16183C, ***16189C,*** 16223T, ***16278T*** | G | 193C, 195C, 198A, 225A, 263G, 315.1C |
| 1 | X2a1 | ***16093C***, ***16189C***, 16223T, 16265G, ***16278T***, 16311C | G | ND |
| 1 | M1a1 | ***16129A,*** 16183C, ***16189C, 16249C, 16311C, 16359C*** | G | ***195C,*** 263G, 315.1C |
| 1 | M1a1 | ***16129A,*** 16183C, ***16189C, 16249C, 16311C, 16359C*** | G | ND |
| 1 | M1 | ***16129A***, 16156A, 16183C, ***16189C,*** 16223T, ***16249C, 16311C*** | G | ***195C***, 263G, 309.1C, 315.1C |
| 1 | M1 | ***16129A***, 16183C, ***16189C,*** 16223T, ***16249C, 16311C*** | G | 152C, 183G, ***195C***, 258T, 263G, 315.1C |
| 1 | M1 | ***16129A***, 16182C, 16183C, ***16189C***, 16223T, ***16249C, 16311C*** | G | ***195C,*** 263G, 309.1C, 315.1C |
| 1 | M1 | ***16129A,*** 16182C, 16183C, ***16189C,*** 16223T, ***16249C, 16311C*** | G | ND |
| 1 | M1 | ***16129A***, 16182C, 16183C, ***16189C, 16249C, 16311C*** | G | ND |
| 1 | M1 | ***16129A***, 16182C, 16183C, ***16189C,*** 16223T, 16240G, ***16249C, 16311C*** | G | 150T, ***195C,*** 263G, 315.1C |
| 1 | M1 | ***16129A,*** 16183C, ***16189C***, 16223T, 16264T, ***16311C***, 16359C | G | ND |
| 1 | M1 | 16183C, ***16189C***, 16223T, ***16249C***, 16260T, ***16311C***, 16320T | G | ND |
| 1 | M1a1 | ***16129A,*** 16183C***, 16189C,*** 16223T***, 16311C, 16359C*** | G | ND |
| 1 | M1a1 | ***16129A***, 16153A, 16183C, ***16189C***, 16214A, 16223T, ***16249C, 16311C, 16359C*** | G | 152C, ***195C,*** 235G, 263G, 309.1C, 315.1C |

**Table 4.4:** (Continued)

| # Individuals[a] | Haplogroup[b] | HVI haplotype (16024−16365)[c] | Nucleotide at postion 73 | HVII haplotype (74−340)[c,d] |
|---|---|---|---|---|
| 1 | M9 | 16223T, ***16362C*** | G | ND |
| 1 | L3b | ***16124C,*** 16223T***, 16278T, 16362C*** | G | ND |
| 1 | L3c | 16223T, ***16311C, 16362C*** | G | ND |
| 1 | L3e1 | 16210G, 16223T, 16292T, ***16327T*** | G | ND |
| 1 | L3e1 | 16111T, 16167T, 16223T, ***16327T*** | G | ND |
| 1 | L3e2 | 16093C, 16126C, 16223T, 16299G, ***16320T*** | G | ND |
| 1 | L3e2 | 16223T, 16286T, ***16320T*** | G | ND |
| 1 | L3f1b | ***16209C,*** 16223T, ***16292T, 16311C*** | G | ***189G, 200G***, 207A, 263G, 309.1C, 315.1C |
| 1 | L3f1b | ***16189C, 16209C***, 16223T, ***16292T, 16311C*** | G | ***189G, 200G***, 263G, 315.1C |
| 1 | L3f1b | ***16209C,*** 16214T, 16223T, ***16292T, 16311C*** | G | ***189G,*** 195C, ***200G,*** 263G, 309.1C, 315.1C |
| 1 | L3f1b | 16093C, ***16129A***, 16148T, ***16209C***, 16223T, ***16292T, 16311C***, 16317 | G | ***189G, 200G***, 215G, 263G, 315.1C |
| 1 | L3f1b | ***16209C,*** 16223T, ***16311C*** | G | ND |
| 1 | L3f1b | 16124C, 16172C, ***16209C***, 16223T, 16266A, ***16292T, 16311C*** | G | ND |
| 1 | L3x1 | ***16169T***, ***16189C***, 16223T, 16256T, ***16278T***, 16311C | G | ND |
| 1 | L2a | ***16189C***, 16192T, 16223T, 16278T, ***16294T***, 16362C | G | 143A, ***146C, 152C***, 195C, 263G, 309.1C, 315.1C |
| 1 | L2a | ***16189C***, 16223T, 16278T, 16292T, ***16294T*** | G | 143A, ***146C, 152C***, 195C, 263G, 309.1C, 315.1C |
| 1 | L2a | 16209C, 16223T, 16278T, ***16294T***, 16301T, ***16311C***, 16354T | G | 143A, ***146C, 152C,*** 189G, 195C, 263G, 315.1C |
| 1 | L2a | 16093C, 16223T, 16278T, ***16294T*** | G | 143A, ***146C, 152C,*** 195C, 263G, 315.1C |
| 1 | L2a | ***16189C,*** 16192T, 16223T, 16278T, ***16294T*** | G | 143A, ***146C, 152C,*** 195C, 263G, 309.1C, 315.1C |
| 1 | L2a | 16213A, 16223T, 16249C, 16265G, 16278T, ***16294T*** | G | ND |
| 1 | L2a | ***16129A,*** 16182C, 16183C***, 16189C***, 16215G, 16223T, 16231C, 16278T, ***16294T, 16311C,*** 16355T | G | ND |
| 1 | L2a | ***16223T***, 16278T, ***16294T***, ***16390A*** | G | ND |
| 1 | L2a1 | 16169T, 16223T, 16239T, 16278T, ***16294T, 16309G*** | G | 143A, ***146C, 152C***, 189G, 195C, 263G, 315.1C |
| 1 | L2a1 | 16223T, 16278T, ***16294T, 16309G*** | G | ND |
| 1 | L2a1 | ***16189C,*** 16192T**,** 16223T, 16278T, 16292T, **16294T, 16309G** | G | 143A, 146C, 152C, 195C, 263G, 315.1C |
| 1 | L2a1 | ***16189C***, 16192T, 16223T, 16278T, 16292T, ***16294T, 16309G*** | G | ND |
| 1 | L2a1 | ***16189C***, 16192T, 16223T, 16278T, ***16294T, 16309G*** | G | ND |
| 1 | L2a1 | 16093C, ***16189C***, 16223T, 16278T, ***16294T, 16309G*** | G | ND |
| 1 | L2a1 | 16093C, 16124C, ***16189C,*** 16223T, 16278T, ***16294T***, 16309G | G | ND |
| 1 | L2a1c1 | ***16086C, 16189C***, 16192T, 16223T, 16278T, ***16294T*** | G | ND |
| 1 | L2a1c1 | ***16086C, 16189C***,16223T, 16278T, ***16294T*** | G | ND |
| 1 | L2b | ***16129A***, 16189C, ***16213G***, 16215A, ***16223T***, ***16278T, 16311C*** | G | ND |
| 1 | L4 | ***16129A***, 16207T, 16166G, 16183d, ***16187C***, 16188T, 16223T, ***16278T, 16355T, 16362C.*** | G | ND |
| 1 | L6 | ***16048A***, 16223T, ***16224C,*** 16278T, 16311C | G | ND |
| 1 | L2 | 16182C, 16183C, ***16189C***, ***16223T***, ***16278T*** | G | 146C, 153G, 156T, 263G, 315.1C |
| 1 | N3* | 16182C, 16183C, 16189C, 16210G, 16224C, 16309G | G | ND |

**Table 4.4:** (Continued)

| # Individuals[(a)] | Haplogroup[(b)] | HVI haplotype (16024−16365)[(c)] | Nucleotide at postion 73 | HVII haplotype (74−340)[(c,d)] |
|---|---|---|---|---|
| 1 | L1b | ***16126C***, 16187T, 16189C, 16223T, ***16264T, 16270T***, 16278T, ***16293G***, 16311C | G | 152C, 182T, ***185T,***189G,195C, 263G, 3096.1C, 315.1C |
| 1 | L0a | 16129A, ***16148T***, ***16168T, 16172C***, 16187T, ***16188G,*** 16189C, 16223C, ***16230G,*** 16311C, ***16320T*** | ***A*** | ***93G,*** 99G, 152C, ***185A, 189G, 236C***, 247A, 263G, 315.1C |
| 1 | L0a | ***16129A***, ***16148T, 16168T***, ***16172C,*** 16187T, ***16188G***, 16189C, 16223T, ***16230G***, ***16278T,*** 16293G, 16311C, ***16320T*** | ***A*** | ***93G,*** 95C, ***185A***, 189G, ***236C,*** 247A, 263G, 315.1C |
| 1 | L0a | ***16129A, 16148T, 16168T, 16172C***, 16187T, ***16188G,*** 16189C, 16223T, ***16230G***, 16311C, ***16320T***, 16362C | ***A*** | ND |
| 1 | L0a1c | ***16148T, 16168T, 16172C***, 16187T, ***16188G***, 16189C, 16223T, ***16230G, 16287T***, 16293G, 16311C, ***16320T*** | ***A*** | ND |
| 1 | L0b | ***16051G***, 16129A, 16148T, ***16164G***, 16172C, 16173T, ***16182T, 16183C***, 16189C, 16207G, ***16209C,*** 16230G, 16278T, 16311C | ***A*** | ND |

[(a)] Number of individuals that shared the same mitochondrial haplotype within the studied population group.
[(b)] Mitochondrial haplogroup assignment was performed primarily using the simple tree of the Genebase mtDNA Haplogroup Reference Guide_ April 2008 (V2.6) available at (www8) and then higher resolution of the haplogroup assignment was obtained via the Global human mtDNA phylogenetic tree_ Build 8 available at (www6). Haplogroup R0 was diagnosed whenever an adenosine base was detected at nt73 with absence of all other haplogroup specific mutations.
[(c)] Haplotypes were annotated following the phylogenetic alignment method and the polymorphic sites were numbered according to the revised version of the CRS (accession number NC_012920). The suffixes A, G, C, and T indicate substitutions and d indicates a deletion. Insertions are indicated by a dot followed by the number and type of inserted nucleotide. Shown in blue are substitutions observed outside the specified HVI region. Shown in green is a substitution detected outside the specified HVII extension. Sequences that had no mutations in the HVI sequenced region compared with the reference sequence are labelled as CRS. Characteristic mutations for each haplogroup are shown in bold and italic.
[(d)] ND means complete HVII sequence between the nucleotides 73−340 was not determined for that sample.

### *4.2.2.5 Statistical analysis*

To estimate the mitochondrial diversity and heterogeneity among the studied population group, statistical analysis of the hypervariable regions data was performed. Diversity indices included number of haplotypes, number of polymorphic sites, random match probability (RMP) and genetic diversity. The probability of randomly selecting two unrelated individuals with the same mitochondrial haplotype (RMP) was calculated using the equation $p=\sum x^2$, where $x$ is the frequency of each mtDNA haplotype (Stoneking et al. 1991). It was found that the random match probability among the Egyptian population sample was relatively high when HVI and HVII data were analysed individually giving values of 2.34% and 3.22%, respectively, whereas combination of the two data sets significantly reduced the RMP to 1.28%. The genetic diversity in the mitochondrial hypervariable regions was calculated using the algorithm $h= n(1-\sum x^2)/(n-1)$, where $n$ is the sample size and $x$ is the frequency of each mtDNA haplotype (Nei and Tajima 1981). High genetic diversity (0.9911) was reported in the Egyptian population sample (n=261) when combined HVI and HVII data were analysed. Slightly lower diversities (0.9804 and 0.9803) were reported when HVI and HVII data were analysed individually. When complete HVI and HVII data were available (n=78), the genetic diversity was high (1.0000) (**Table 4.5**).

**Table 4.5: Diversity indices of the mitochondrial hypervariable region obtained by analysis of mtDNA sequences in an Egyptian population sample (n=261).**

| Population Statistic | HVI (16024-16365) (n=261) | HVII (73-340) (n=78) | HVI+HVII (n=78) | Combined data (n=261)* |
|---|---|---|---|---|
| Number of haplotypes | 187 (36 shared) | 62 (11 shared) | 78 (0 shared) | 207 (24 shared) |
| Number of polymorphic sites | 113 | 42 | 120 | 155 |
| Random match probability | 2.34% | 3.22% | 1.28% | 1.28% |
| Genetic diversity | 0.9804 | 0.9803 | 1.0000 | 0.9911 |

* Included complete HVI sequences (16024-16365) and the nucleotide position 73 in all samples (n=261) combined with HVII sequences (73-340) in 30% of samples (n=78).

### 4.2.3 Mitochondrial haplogroup designation

Mitochondrial DNA sequences (n=261) were all assigned to known mitochondrial haplogroups. The assignment was primarily performed according to the most updated global human mtDNA phylogenetic tree Build 8 (Van Oven and Kayser 2009) and guided by the haplogroup reference guide provided by the Genebase at (www8). Some haplogroups were easily diagnosed using HVI data (16024-16365) obtained by direct sequencing (particularly haplogroups L, M, N1a, N1b, U, K, J, T, R0a, and HV1), while in others, HVII information was required to confirm the haplogroup designation. Sequences that did not show any of the previously reported haplogroup specific mutations in the HVI region were allocated to haplogroup R0 or haplogroup R* according to the nucleotide base at position 73 either adenosine or guanosine, respectively (**Table 4.4**). Classified sequences were checked against the previously reported data sets from different population studies (Watson et al. 1997; Maca-Meyer et al. 2001; Kivisild et al. 2004; Salas et al. 2004; Behar et al. 2008a; Saunier et al. 2009). The haplogroup assignment of the mtDNA sequences based on the hypervariable region sequencing information was further confirmed and refined into sub-haplogroups via PCR-RFLP analysis of the mitochondrial genome, after which the frequency distribution of the different mitochondrial haplogroups in the studied population group was concluded (see Chapter 5).

### 4.2.4 Meta-analysis

HVI sequencing data obtained from the present study was combined with two previously published Egyptian mtDNA sequence databases of 58 individuals from Gurna and Upper Egypt (Stevanovitch et al. 2004) and 277 individuals from Alexandria (Saunier et al. 2009). A database of 596 Egyptian HVI sequences (nt16024-16365) was generated. Overall, a total of 139 polymorphic sites were reported at the specified mitochondrial HVI region in the combined dataset. All the detected polymorphisms were previously reported in the mitochondrial DNA databases. These polymorphic sites varied greatly in their frequencies in the analysed sequences, ranging from occurring in 1 individual to 234 individuals. Sites that showed the highest variability in the HVI region of the Egyptian mtDNA sequences, found in more than 10% of individuals, included 16126, 16129, 16183, 16189, 16223, 16278, 16294, and 16311. Approximately 23% of the reported polymorphisms (n=32) were found only once in the combined database of 596 Egyptian mtDNA sequences, and thus can be considered haplotype specific. Such unique polymorphisms increase the power of discrimination and provide a useful tool for individual identification (**Table 4.6**).

**Table 4.6: Mitochondrial DNA polymorphic sites and frequencies reported in a combined dataset of 596 Egyptian mitochondrial HVI sequences (nt16024-16365).** Sites were numbered according to the Cambridge Reference Sequence numbering system (Andrews et al. 1999). N is the number of individuals in which the polymorphisms were detected and the frequency percentages are shown between brackets. Meta-analysis involved 261 HVI sequences from the present study, 277 sequences from Saunier et al. (2009), and 58 sequences from Stevanovitch et al. (2004).

| N (Frequency %) | Polymorphic site | N (Frequency %) | Polymorphic site | N (Frequency %) | Polymorphic site |
|---|---|---|---|---|---|
| 234 (39.26) | 16223 | 10 (1.68) | 16291 | 3 (0.50) | 16239 |
| 182 (30.54) | 16189 | 10 | 16304 | 3 | 16240 |
| 176 (29.53) | 16311 | 10 | 16360 | 3 | 16254 |
| 129 (21.64) | 16126 | 9 (1.51) | 16086 | 3 | 16287 |
| 102 (17.11) | 16294 | 9 | 16111 | 3 | 16288 |
| 94 (15.77) | 16129 | 9 | 16166 | 3 | 16290 |
| 90 (15.10) | 16183 | 9 | 16274 | 2 (0.34) | 16164 |
| 77 (12.92) | 16278 | 9 | 16300 | 2 | 16221 |
| 56 (9.40) | 16362 | 9 | 16324 | 2 | 16295 |
| 52 (8.72) | 16172 | 9 | 16356 | 2 | 16299 |
| 52 | 16292 | 8 (1.34) | 16147 | 2 | 16316 |
| 48 (8.05) | 16249 | 8 | 16219 | 2 | 16344 |
| 37 (6.21) | 16093 | 8 | 16234 | 2 | 16357 |
| 37 | 16209 | 8 | 16264 | 1 (0.17) | 16048 |
| 36 (6.04) | 16069 | 8 | 16266 | 1 | 16066 |
| 36 | 16187 | 8 | 16298 | 1 | 16081 |
| 32 (5.37) | 16145 | 7 (1.17) | 16265 | 1 | 16085 |
| 30 (5.03) | 16224 | 7 | 16319 | 1 | 16092 |
| 28 (4.70) | 16192 | 7 | 16354 | 1 | 16096 |
| 27 (4.53) | 16163 | 6 (1.01) | 16167 | 1 | 16107 |
| 27 | 16186 | 6 | 16179 | 1 | 16114 |
| 26 (4.36) | 16176 | 6 | 16215 | 1 | 16120 |
| 26 | 16182 | 6 | 16222 | 1 | 16123 |
| 26 | 16296 | 6 | 16260 | 1 | 16131 |
| 26 | 16320 | 6 | 16271 | 1 | 16137 |
| 24 (4.03) | 16148 | 6 | 16291 | 1 | 16138 |
| 24 | 16270 | 6 | 16301 | 1 | 16140 |
| 23 (3.86) | 16309 | 6 | 16318 | 1 | 16156 |
| 23 | 16359 | 5 (0.84) | 16051 | 1 | 16173 |
| 21 (3.52) | 16168 | 5 | 16214 | 1 | 16201 |
| 19 (3.19) | 16067 | 5 | 16231 | 1 | 16229 |
| 19 | 16256 | 5 | 16248 | 1 | 16233 |
| 19 | 16343 | 5 | 16286 | 1 | 16235 |
| 17 (2.85) | 16355 | 4 (0.67) | 16136 | 1 | 16243 |
| 16 (2.68) | 16188 | 4 | 16146 | 1 | 16245 |
| 16 | 16261 | 4 | 16150 | 1 | 16259 |
| 15 (2.52) | 16230 | 4 | 16184 | 1 | 16269 |
| 15 | 16327 | 4 | 16213 | 1 | 16305 |
| 13 (2.18) | 16193 | 4 | 16325 | 1 | 16313 |
| 13 | 16293 | 3 (0.50) | 16037 | 1 | 16317 |
| 12 (2.01) | 16104 | 3 | 16041 | 1 | 16321 |
| 12 | 16124 | 3 | 16134 | 1 | 16335 |
| 12 | 16153 | 3 | 16139 | 1 | 16337 |
| 12 | 16169 | 3 | 16171 | 1 | 16352 |
| 12 | 16218 | 3 | 16174 | 1 | 16365 |
| 11 (1.85) | 16210 | 3 | 16195 | | |
| 10 (1.68) | 16181 | 3 | 16207 | | |

The aforementioned mitochondrial DNA polymorphisms identified 384 different mitochondrial HVI haplotypes in the combined dataset, of which 306 (~80%) were unique to single individuals. Thus, analysis of the mitochondrial HVI sequences (nt16024-16365) solely enabled unique identification in 306 individuals. Such a large number of unique haplotypes indicates high mtDNA diversity in the Egyptian population and validates the utility of the mitochondrial HVI region analysis for human identification purposes. The remaining 78 haplotypes were reported in multiple individuals, of which 43 haplotypes were shared by 2 individuals and 11 haplotypes were shared by 3 individuals (**Table 4.7**). The most frequently reported mitochondrial DNA haplotype, seen in 40 Egyptians, was identical to the CRS in the specified HVI region. Mitochondrial HVI haplotypes shared by multiple individuals (n=78) were identified and their frequencies were determined in the three Egyptian mitochondrial DNA datasets individually and in combination (**Table 4.8)**. The estimated random match probability and genetic diversity in the combined Egyptian dataset of 596 HVI haplotypes (nt16024-16365) were 2.08% and 0.9808, respectively.

**Table 4.7: Distribution of the mitochondrial DNA HVI haplotypes (nt16024-16365) in a combined Egyptian dataset of 596 individuals.** HVI sequences were obtained from the present study (n=261) and two previously published Egyptian datasets of 58 individuals from Gurna and Upper Egypt (Stevanovitch et al. 2004) and 277 individuals from Alexandria (Saunier et al. 2009).

| Number of HVI haplotypes | Number of individuals sharing the same haplotype | Total number of individuals |
|---|---|---|
| 306 | 1 | 306 |
| 43 | 2 | 86 |
| 11 | 3 | 33 |
| 11 | 4 | 44 |
| 4 | 5 | 20 |
| 1 | 6 | 6 |
| 3 | 7 | 21 |
| 2 | 8 | 16 |
| 1 | 9 | 9 |
| 1 | 15 | 15 |
| 1 | 40 | 40 |
| Total= 384 | | Total= 596 |

**Table 4.8: Mitochondrial DNA HVI haplotypes repeatedly reported in a combined dataset of 596 Egyptian mitochondrial HVI sequences (nt16024-16365).** Data were obtained from the present study (n=261) and two previously published mitochondrial DNA datasets of 277 (Saunier et al. 2009) and 58 (Stevanovitch et al. 2004) Egyptian individuals. Numbers of individuals sharing the same haplotypes were determined in both the individual databases as well as the combined dataset.

| Haplogroup[a] | HVI haplotype (16024−16365)[b] | Present study (n=261) | Alexandria (n=277)[c] | Gurna and Upper Egypt (n=58)[d] | Combined dataset (n=596) |
|---|---|---|---|---|---|
| R0 | CRS | 19 | 19 | 2 | 40 |
| N1b | 16145A, 16176G, 16223T | 7 | 5 | 3 | 15 |
| L3f1b | 16209C, 16223T, 16292T, 16311C | 6 | 3 | 0 | 9 |
| HV1 | 16067T | 3 | 5 | 0 | 8 |
| U5a | 16192T, 16256T, 16270T | 8 | 0 | 0 | 8 |
| R0 | 16104T, 16181C, 16182C, 16183C, 16189C | 2 | 5 | 0 | 7 |
| U3 | 16343G. | 2 | 4 | 1 | 7 |
| M1 | 16129A, 16183C, 16189C, 16223T, 16249C, 16311C, 16359C | 2 | 5 | 0 | 7 |
| T1a | 16126C, 16163G, 16186T, 16189C, 16294T | 4 | 2 | 0 | 6 |
| J | 16069T, 16126C | 2 | 2 | 1 | 5 |
| M1 | 16129A, 16183C, 16189C, 16249C, 16311C | 3 | 2 | 0 | 5 |
| K1a1a | 16093C,16224C, 16311C | 2 | 2 | 1 | 5 |
| I | 16129A, 16223T, 16391A | 0 | 4 | 1 | 5 |
| H6 | 16362C | 0 | 4 | 0 | 4 |
| T1a | 16126C, 16163G, 16186T, 16189C, 16274A, 16294T | 3 | 1 | 0 | 4 |
| T | 16126C, 16294T | 2 | 2 | 0 | 4 |
| J | 16069T, 16126C, 16311C | 1 | 3 | 0 | 4 |
| K | 16224C, 16311C | 2 | 2 | 0 | 4 |
| M1a1 | 16129A, 16183C, 16189C, 16249C, 16311C, 16359C | 2 | 2 | 0 | 4 |
| M1 | 16129A, 16182C, 16183C, 16189C, 16223T, 16249C, 16311C | 2 | 2 | 0 | 4 |
| M1 | 16129A, 16189C, 16223T, 16249C, 16311C, 16359C | 0 | 3 | 1 | 4 |
| L0a1b | 16129A, 16148T, 16168T, 16172C, 16187T, 16188, 16189C, 16223C, 16230G, 16311C, 16320T | 1 | 0 | 3 | 4 |
| L1b | 16126C, 16187T, 16189C, 16223T, 16264T, 16270T, 16278T, 16293G, 16311C | 4 | 0 | 0 | 4 |
| L2a1 | 16189C, 16192T, 16223T, 16278T, 16292T, 16294T, 16309G | 2 | 2 | 0 | 4 |
| R0 | 16145A, 16150T, 16218T | 0 | 3 | 0 | 3 |
| R0 | 16218T | 1 | 2 | 0 | 3 |
| R | 16223T | 2 | 1 | 0 | 3 |
| HV1 | 16067T, 16183G, 16327A | 2 | 1 | 0 | 3 |

**Table 4.8: (Continued).**

| Haplogroup[a] | HVI haplotype (16024−16365)[b] | Present study (n=261) | Alexandria (n=277)[c] | Gurna and Upper Egypt (n=58)[d] | Combined dataset (n=596) |
|---|---|---|---|---|---|
| T1a | 16111A, 16126C, 16163G, 16186T, 16189C, 16294T | 2 | 1 | 0 | 3 |
| T2 | 16126C, 16294T, 16296T, 16324C | 1 | 1 | 1 | 3 |
| N1a | 16147G, 16172C, 16223T, 16248T, 16311C, 16355T | 0 | 2 | 1 | 3 |
| M1 | 16183C, 16189C, 16223T, 16249C, 16260T, 16311C, 16320T | 1 | 2 | 0 | 3 |
| M1a1 | 16129A, 16183C, 16189C, 16223T, 16311C, 16359C | 1 | 1 | 1 | 3 |
| L3c | 16223T, 16311C, 16362C | 1 | 1 | 1 | 3 |
| L3f1b | 16189C, 16209C, 16223T, 16292T, 16311C | 1 | 2 | 0 | 3 |
| H2a1 | 16354T | 0 | 2 | 0 | 2 |
| HV1 | 16067T, 16356C | 0 | 2 | 0 | 2 |
| R0 | 16355T, 16362C | 0 | 2 | 0 | 2 |
| R0 | 16166G, 16189C,16304C | 2 | 0 | 0 | 2 |
| R0 | 16129A | 2 | 0 | 0 | 2 |
| R0 | 16356C | 2 | 0 | 0 | 2 |
| R0 | 16210G, 16224C, 16292T | 2 | 0 | 0 | 2 |
| R0 | 16218T, 16291T | 1 | 1 | 0 | 2 |
| R0a | 16126C, 16362C | 1 | 1 | 0 | 2 |
| R0a | 16126C, 16192A, 16278T, 16362C | 1 | 1 | 0 | 2 |
| R0a | 16126C, 16271C, 16291T, 16362C | 2 | 0 | 0 | 2 |
| R0a | 16126C, 16304C, 16362C | 0 | 2 | 0 | 2 |
| R0a1a | 16126C, 16355T, 16362C | 2 | 0 | 0 | 2 |
| R11'B | 16183C, 16189C, 16223T, 16311C. | 2 | 0 | 0 | 2 |
| R* | 16129A, 16183C, 16189C, 16215G | 0 | 2 | 0 | 2 |
| J | 16069T, 16126C, 16169T. | 1 | 1 | 0 | 2 |
| J1 | 16069T, 16126C, 16249C, 16261T, 16311C | 2 | 0 | 0 | 2 |
| J1 | 16069T, 16126C, 16193T, 16300G | 0 | 2 | 0 | 2 |
| J1b1 | 16069T, 16126C, 16145A, 16172C, 16222T, 16261T | 0 | 2 | 0 | 2 |
| T | 16126C, 16294T | 0 | 2 | 0 | 2 |
| T2 | 16126C, 16146G, 16147T, 16171T, 16292T, 16294T, 16296T | 0 | 2 | 0 | 2 |
| T2 | 16126C, 16153A, 16294T, 16296T | 0 | 2 | 0 | 2 |
| T2c | 16126C, 16292T, 16294T, 16296T | 1 | 1 | 0 | 2 |
| T2c | 16126C, 16292T, 16294T | 2 | 0 | 0 | 2 |
| T1a | 16126C, 16139, 16163G, 16186T, 16189C, 16294T. | 2 | 0 | 0 | 2 |

**Table 4.8: (Continued).**

| Haplogroup[a] | HVI haplotype (16024−16365)[b] | Present study (n=261) | Alexandria (n=277)[c] | Gurna and Upper Egypt (n=58)[d] | Combined dataset (n=596) |
|---|---|---|---|---|---|
| U3 | 16184T, 16319A, 16343G | 0 | 2 | 0 | 2 |
| U4 | 16261, 16356 | 0 | 0 | 2 | 2 |
| U6a | 16172C, 16219G, 16278T. | 1 | 1 | 0 | 2 |
| U7 | 16309G, 16318T, 16362C | 1 | 1 | 0 | 2 |
| U8b | 16179T, 16183C, 16189C, 16234T, 16324C | 2 | 0 | 0 | 2 |
| K | 16174T, 16224C, 16311C | 1 | 1 | 0 | 2 |
| K1a | 16176T, 16224C, 16311C | 0 | 2 | 0 | 2 |
| I | 16093, 16129, 16223 | 0 | 0 | 2 | 2 |
| I1 | 16129A, 16223T, 16311C | 1 | 1 | 0 | 2 |
| M1 | 16183C, 16189C, 16249C, 16311C, 16327T | 2 | 0 | 0 | 2 |
| L1c2 | 16093C, 16129A, 16187T, 16189C, 16223T, 16265C, 16278T, 16286G, 16294T, 16311C, 16320T, 16360T | 0 | 2 | 0 | 2 |
| L1c3 | 16093C, 16129A, 16183C, 16189C, 16215G, 16223T, 16278T, 16294T, 16311C, 16360T | 0 | 2 | 0 | 2 |
| L2a | 16189C, 16192T, 16223T, 16278T, 16294T | 1 | 1 | 0 | 2 |
| L2a1 | 16189C, 16223T, 16278T, 16294T, 16309G | 0 | 2 | 0 | 2 |
| L3e2 | 16223T, 16320T | 2 | 0 | 0 | 2 |
| L3f1b | 16172C, 16209C, 16223T, 16292T, 16311C | 2 | 0 | 0 | 2 |
| L3d | 16111T, 16124C, 16223T | 0 | 2 | 0 | 2 |
| L3* | 16169, 16193, 16195, 16266, 16301 | 0 | 0 | 2 | 2 |

[a] Mitochondrial haplogroup assignment was performed primarily using the simple tree of the Genebase mtDNA Haplogroup Reference Guide_ April 2008 (V2.6) available at (www8) and then higher resolution of the haplogroup assignment was obtained via the Global human mtDNA phylogenetic tree_ Build 8 available at (www6). Haplogroup R0 was diagnosed whenever an adenosine base was detected at nt73 with absence of all other haplogroup specific mutations.
[b] Haplotypes were annotated following the phylogenetic alignment method and the polymorphic sites were numbered according to the revised version of the CRS (accession number NC_012920). The suffixes A, G, C, and T indicate substitutions.
(c) Data obtained from Saunier et al. (2009).
(d) Data obtained from Stevanovitch et al. (2004) in which the polymorphic sites were reported without specifying the nucleotide variations.

## 4.3 Discussion

Mitochondrial DNA has long been used for the human identification and population genetics due to its peculiar characteristic advantages over the nuclear genome, including the high copy number in each somatic cell, the maternal mode of inheritance with non-recombination characteristics and the high mutation rate. The mitochondrial DNA database is currently enriched with over six thousand complete mtDNA sequences and is continuously growing. However, the current database does not include equivalent representatives of all populations and ethnic groupings, those of African descent are currently underrepresented compared to European and Asian populations. Egypt has a characteristic geographic location among three continents: Africa, Asia, and Europe and thus its population structure might have been affected by several waves of modern human migration in and out of Africa (Coudray et al. 2009). In addition, Egypt has been ruled by several populations throughout its vast history including the Greeks, Romans, Arabs, Turks, French and British (Saunier et al. 2009). Therefore, admixture between the original Egyptian population and all the previous ruling populations was expected to occur and subsequently to affect the genetic structure of present day Egyptians. The present study investigated the mitochondrial diversity and heterogeneity among 261 adult Egyptians (randomly selected and therefore thought to be maternally unrelated) through analysis of the mitochondrial hypervariable regions which have been extensively investigated in many other population groups.

### 4.3.1 Rapid PCR

The present study developed an optimised laboratory protocol for rapid PCR amplification of the mitochondrial hypervariable regions from blood samples stored on FTA cards (Whatman). This protocol involved the use of newly designed long oligonucleotides (~30mer) which enabled combination of annealing and extension steps at 72$^{o}$C and thus to shift from a three-step PCR protocol to a two-step one. The PCR thermal cyclic conditions were also optimised to further reduce the amplification time. The final optimised protocol allowed the amplification reaction to be completed in a short period of time (30 mins) compared to the 2-3 hours conventional PCR program. This rapid PCR amplification has proven to be a reliable and easily reproducible procedure for dealing with a large number of samples in a short period of time and thus it has a direct application in population studies which require high throughput analyses. The robustness of the rapid PCR technique can be verified by comparing sequence

quality of amplicons produced by this protocol to those produced via conventional PCR programs. Such a rapid PCR protocol has a direct impact on forensic investigations in which the time is an important factor. It can also play an effective role in forensic science undergraduate laboratories.

### 4.3.2 Sequence variations in the mitochondrial hypervariable regions

The present study involved analysis of the mitochondrial DNA variations in an Egyptian population sample (n=261) collected from Alexandria, which is the second largest city in Egypt residing on Egypt's north coast bordering the Mediterranean Sea and representing the major seaport connecting the country to Europe and Asia. Analysis of the mitochondrial hypervariable region sequences has revealed a large number of sequence variations relative to the Cambridge Reference Sequence (CRS) in both the HVI and HVII regions of the mitochondrial genome. Overall, a total of 113 polymorphic sites were reported in 261 mitochondrial HVI sequences (16024-16365) at varying frequencies. The average number of times for a variable character to occur was 9.2, which is higher than the average reported in other populations such as the U.S and European Caucasians (6.7) (Allard et al. 2002) and East Asian populations (4.5) (Allard et al. 2004). Sites that showed the higher rate of variations, seen in 20 individuals or more, included nucleotides 16126, 16129, 16172, 16183, 16189, 16223, 16249, 16278, 16292, 16294, 16311, and 16362. All these sites (except 16126, 16223, 16249, and 16294) were previously reported with a high rate of variability in the SWGDAM mtDNA dataset of 1771 U.S and European Caucasians (Allard et al. 2002). All these sites (except 16249) were previously described as mutational hotspots in the human evolutionary studies (Stoneking et al. 1991; Wakeley 1993; Meyer et al. 1999). Among 78 HVII sequences (73-340), a total of 42 polymorphic sites were reported with varying frequencies. An A to G mutation at position 263 was detected in all HVII sequences, this finding is not surprising as occurrence of adenosine at that position is a specific character to the CRS (Finnila et al. 2001b; Maca-Meyer et al. 2001). The next common polymorphic sites reported in HVII region were 73, 146, 152, 195, and 303-315 occurring in 20 individuals or more. These findings are fully consistent with other population studies which involved analysis of the mitochondrial HVII sequence variations (Allard et al. 2002; Poetsch et al. 2003; Barbosa et al. 2008).

### 4.3.3 Mitochondrial haplotype analysis

In the present study, a database of 261 Egyptian mitochondrial hypervariable region sequences was established. A total of 187 HVI haplotypes were identified with a percentage number of haplotypes K (number of haplotypes/number of samples) 72%. This percentage of haplotypes based on HVI analysis is slightly low compared to previous studies which involved 68 Egyptians from Mansoura and Asuit (K=85%) (Krings et al. 1999) and 34 from Gurna (K=88%) (Stevanovitch et al. 2004). However, this difference can be easily attributed to the small sample size in the previous studies. Other population studies which involved HVI analysis in a relatively similar sample size have reported much lower haplotype percentages, 64% in Singapore Malays (Wong et al. 2007) and 63% in Japanese population (Asari et al. 2007). Analysis of HVII in 78 individuals revealed 62 haplotypes with a percentage (K) 79%. When combined with HVI data for those individuals, the total number of haplotypes increased to 78 giving a percentage (K) 100%. Such a high haplotype percentage was not previously reported in any other population studies and it indicates a very high level of diversity in the studied population group. Merging of the available HVII data (n=78) with the full set of HVI sequences (n=261) increased the overall number of mtDNA haplotypes characterised in the present study to 207 giving a haplotype percentage (K) of 79%. However, this percentage is not to be compared to other studies because the HVII sequences were not determined for all samples. Approximately 81% (151) and 82% (51) of the reported HVI and HVII haplotypes, respectively, were unique to single individuals. The other haplotypes were shared between multiple individuals. The most frequently reported HVI haplotype, seen in 19 individuals out of 261, was identical to the CRS between the nucleotides (16024-16365). Consistent results were reported by a coincident study, which involved analysis of the control region in 277 Egyptians, where they reported the same highly frequent haplotype (identical to the CRS in the specified HVI region) in 19 individuals (Saunier et al. 2009). A similar finding was also previously reported in the North of Portugal (Lima et al. 2006).

### 4.3.4 Diversity indices

Diversity measures were estimated based on the HVI and HVII data individually and in combination. The random match probability of the HVI and HVII haplotypes were estimated to be 2.34% and 3.22%, respectively. Similar values were reported in a previous study which involved analysis of the mtDNA variations in 124 Japanese

individuals, where the estimated RMP was 2.24% and 2.68% for HVI and HVII, respectively (Mabuchi et al. 2007). A higher RMP (4.2%) was reported among 142 HVI haplotypes from the North of Portugal (Lima et al. 2006), whereas lower values (1.25% and 1.57%) were reported in 167 HVI and HVII Brazilian haplotypes, respectively (Barbosa et al. 2008). In the present study, combination of HVI and HVII sequencing data of 78 individuals reduced the RMP to 1.28%, which is in the average of many other population studies that reported values of 1% (Banerjee et al. 2005; Egyed et al. 2007) and 1.29% (Lima et al. 2006). Compared to other North African populations, the estimated RMP in the present study is significantly lower than the those reported in 64 Tunisians (2.1%) and 56 Moroccans (2.1%) (Turchi et al. 2009). However, a much lower RMP value was recently reported in 277 Egyptians (0.49%) which can be attributed to the analysis of the whole extension of the control region between the nucleotides 16024-576 compared to analysis of the individual HVI and HVII regions performed in the present study (Saunier et al. 2009). The genetic diversity *h* was estimated as 0.9804 and 0.9803 for HVI and HVII, respectively. These values are very similar to other population studies in which the statistical analysis of the HVI and HVII data was performed individually (Asari et al. 2007; Mabuchi et al. 2007). When HVI and HVII data were combined in the present study (n=78), the genetic diversity was estimated to be 1.0000 which is comparable to the diversity reported in the previous Egyptian database of 277 individuals which was estimated to be 0.9986 (Saunier et al. 2009). Similar sequence diversity values were reported in Jordanian population (Gonzalez et al. 2008), European populations (Poetsch et al. 2003; Lima et al. 2006), Asian populations (Imaizumi et al. 2002; Asari et al. 2007; Wong et al. 2007) and Brazil (Barbosa et al. 2008). For the combined HVI database (n=596), the estimated genetic diversity was 0.9808 which is very similar to the values reported in the Japanese population (Asari et al. 2007) and in the North of Portugal (Lima et al. 2006). Based on the general agreement regarding the high mtDNA diversity in Africa compared to other areas of the world (Jorde et al. 2000), it was expected to detect higher mitochondrial DNA diversity indices in the Egyptian population compared to European and Asian populations, but that did not occur. Unfortunately, most of the mitochondrial DNA population studies which involved characterisation of nucleotide variations in the mitochondrial hypervariable regions and estimation of the genetic diversity were more concerned with populations outside Africa, which did not allow a comparison between the mitochondrial genetic diversity in the Egyptian population and other populations in Africa.

### 4.3.5 Conclusion

An optimised laboratory protocol for rapid PCR amplification of the mitochondrial hypervariable regions was developed. This rapid PCR has a direct impact on forensic investigations as well as forensic science undergraduate laboratories. A database of 261 Egyptian mitochondrial HVI sequences (nt16024-16365) was established. This database contained a total of 113 polymorphic sites, of which those with higher frequencies were determined. A total of 187 HVI haplotypes were identified, of which 151 (~81%) were unique to single individuals. The most frequently reported HVI haplotype was identical to the Cambridge Reference Sequence at the specified HVI region. For a subset of 78 Egyptian individuals, both HVI (nt16024-16365) and HVII (nt73-340) haplotypes were determined and no shared mitochondrial haplotypes were reported. These findings confirm that sequence polymorphism of the control region of the human mtDNA could be highly informative in forensic casework, particularly when HVI and HVII data combined together which increases the power of discrimination. The genetic diversity *h* was estimated to be 0.9804 and 0.9803 for HVI and HVII respectively. The probability of two random individuals to have the same HVI haplotypes or HVII haplotypes were determined at 2.34% and 3.22% respectively. Overall, a diverse genetic structure of the Egyptian population was reported via detection of a large number of unique HVI haplotypes.

# 5. PCR-RFLP SCREENING OF THE WHOLE MITOCHONDRIAL GENOME IN THE EGYPTIAN POPULATION SAMPLE

## 5.1 Introduction

Maternally inherited mitochondrial DNA has been extensively used as a powerful tool for understanding the demographic history of human populations. The rapidly growing mtDNA database with more than six thousand complete mitochondrial DNA sequences available enabled further investigation of the mitochondrial nucleotide variations and better understanding of the phylogenetic trees of the human mitochondrial DNA. Numerous studies have revealed the link between certain sets of mutations and specific mitochondrial lineages that were found to be continental specific. These mitochondrial mutations represent variations relative to the Cambridge Reference Sequence of the mitochondrial genome which belongs to an individual of European descent, and has been used as the standard reference sequence in all the mitochondrial DNA population studies. These mutations are distributed throughout the mitochondrial genome (~16Kb) including both coding and non-coding regions with particular accumulation in the non-coding hypervariable regions. Different studies focussed on different parts of the molecule with little or no attention paid to the rest of it. Several mitochondrial DNA studies were more concerned with the nucleotide variations in the mitochondrial hypervariable regions, HVI and HVII and their application for identification of the different mtDNA lineages and their maternal ancestry. Despite being commonly used for differentiation between mitochondrial lineages, hypervariable region data do not always provide sufficient information for haplogroup identification. The nucleotide polymorphisms in the coding region play an essential role in the mitochondrial haplogroup assignment, particularly for those lineages which do not have specific diagnostic mutations in their hypervariable region sequences. For example, the main European haplogroup H can only be identified by detection of a 7028C mutation causing loss of *AluI* cut site, due to lack of any other specific markers in the hypervariable region. Thus, in studies which involved analysis of the hypervariable region sequences solely, haplogroup H was mainly identified by the absence of the other haplogroup specific mutations. Similarly, identification of the African subclade L3e can only be confirmed via detection of a T2352C mutation which produces a novel *MboI* cut site. Polymorphisms in the coding region of the mitochondrial genome are as important as the hypervariable region variations for accurate and reliable identification

of the mitochondrial DNA haplogroups. Therefore, the aim of that part of the project was to develop a new strategy for screening of the mitochondrial genome for the most informative and specific mutations using the PCR-RFLP technology for the purpose of mitochondrial haplogroup designation. This strategy was to be used for haplogroup identification of the Egyptian population (n=261) and to be validated by correlating the results of the FRLP screening with those obtained by hypervariable region analysis. The genetic structure of the Egyptian population was to be concluded based on the combined data from both hypervariable region analysis and mitochondrial genome RFLP screening. This genetic structure was to be correlated to the published population databases. The relative matrilineal contributions of Europeans, Asians and Africans to the present day Egyptians were to be determined, and thus the genetic relationship between Egypt and the surrounding populations was to be concluded.

## 5.2 Results

### 5.2.1 Multiplex PCR amplification of 11Kb of the mitochondrial genome as nine fragments

The specified nine regions (A-I) of the mitochondrial genome (**Chapter 2, Figure 2.2**) were successfully amplified from all the Egyptian blood samples (n=261) that were previously stored on FTA cards and processed using method 3 of the washing procedures (**Chapter 2, Table 2.1**). The amplification reaction was performed via a multiplex PCR which was performed in two stages: the first stage reaction involved the use of a 1.2-mm dried processed disc of each sample as a DNA template and nine pairs of primers and the second stage involved the use of the first-stage PCR product as a template and the region specific pair of primers (**Figure 5.1**). Products of the second - stage PCR were all separated by gel electrophoresis (**Figures 5.2-5.6**).

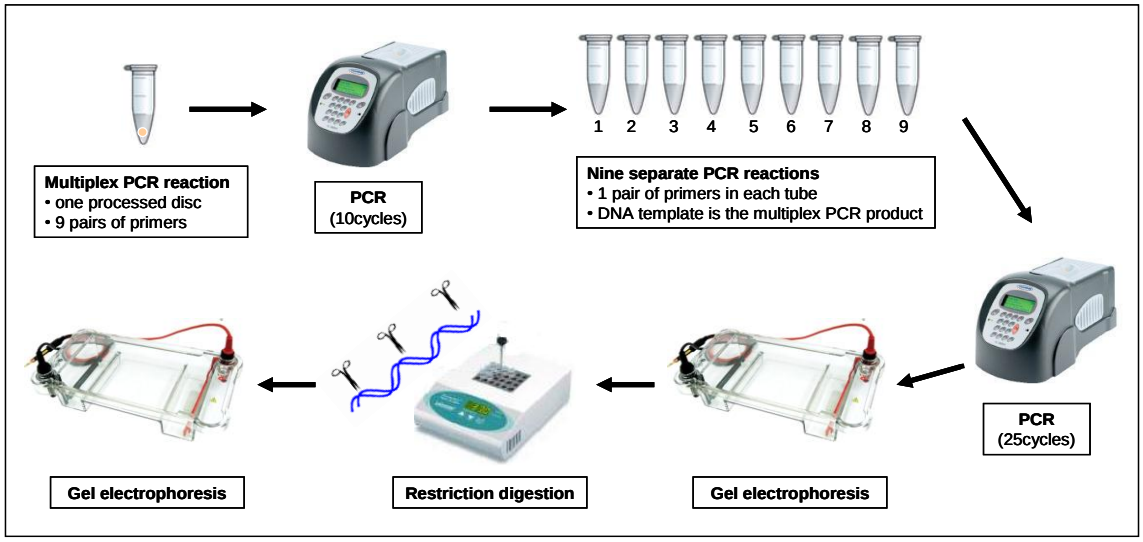



**Figure 5.1: Steps of multiplex PCR amplification and RFLP analysis of the mitochondrial genome.**

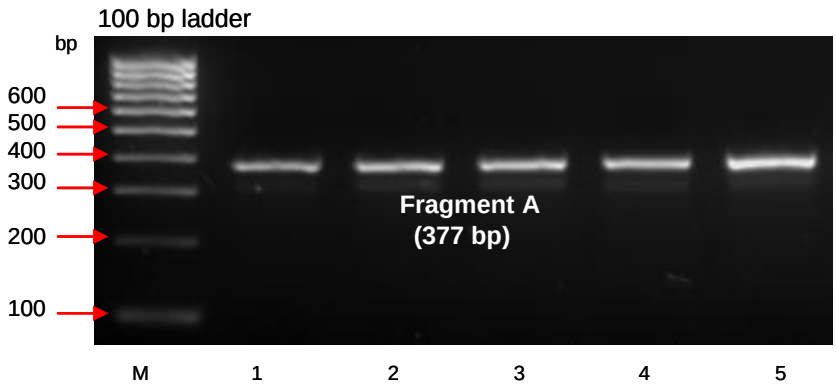



**Figure 5.2: PCR amplification products of the fragment A of the mitochondrial genome (nt451-827).** PCR products of the fragment A (377 bp) were separated on 2% agarose gel. A 100 bp DNA ladder was used as a molecular marker (Lane M). Lanes 1-5 show PCR products of five Egyptian samples.

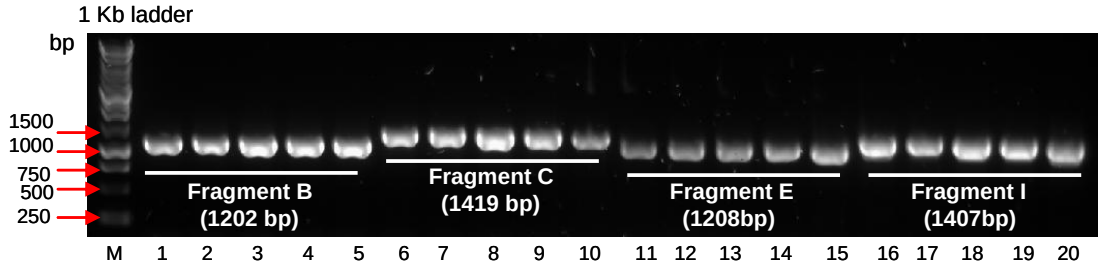



**Figure 5.3: PCR amplification products of four regions (B, C, E, and I) of the mitochondrial genome.** PCR products were separated by electrophoresis on 1% agarose gel. A 1Kb DNA ladder was used as a molecular marker (Lane M). Lanes 1-5 show the PCR products of the fragment B (nt1489-2690) of five Egyptian samples with a product size of 1202bp. Lanes 6-10 show the PCR products of the fragment C (nt3314-4732) of the same Egyptian samples with a product size of 1419bp. Lanes 11-15 show the PCR products of the fragment E (nt8024-9231) with a product size of 1208bp. Lanes 16-20 show the PCR products of the fragment I (nt15312-149) with a product size of 1407bp.

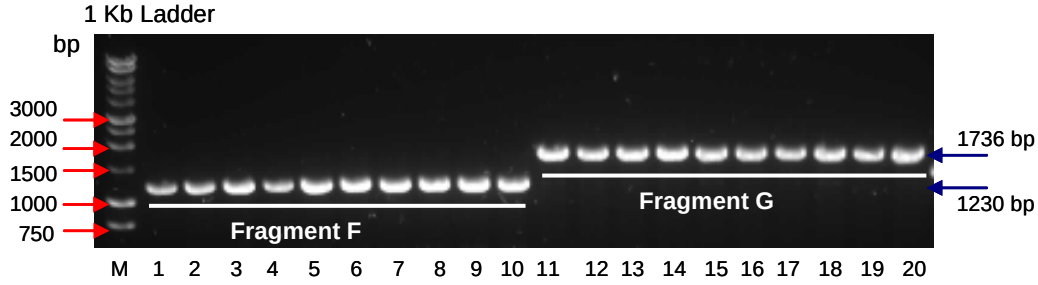



**Figure 5.4: PCR amplification products of fragment F and Fragment G of the mitochondrial genome.** PCR products were separated by electrophoresis on 1% agarose gel. A 1Kb DNA ladder was used as a molecular marker (Lane M). Lanes 1-10 show the PCR products of the fragment F (nt9921-11150) of 10 Egyptian samples with a product size of 1230bp. Lanes 11-20 show the PCR products of the fragment G (nt12077-13812) of the same samples with a product size of 1736bp.

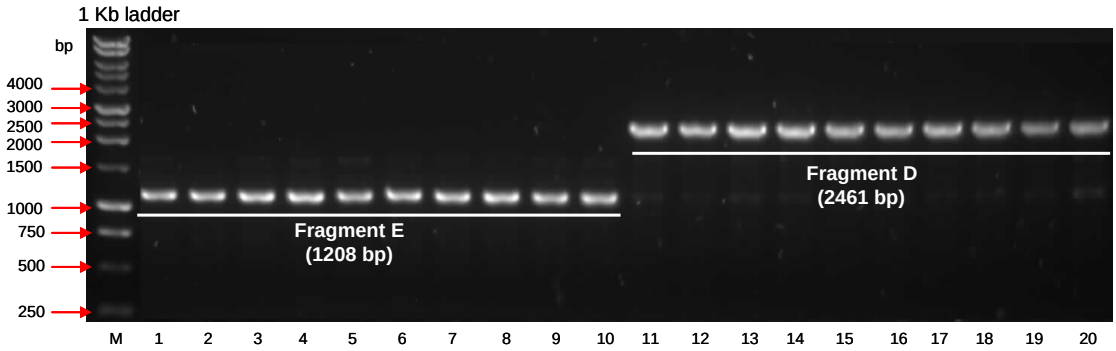



**Figure 5.5: PCR amplification products of fragment E and Fragment D of the mitochondrial genome.** PCR products were separated by electrophoresis on 1% agarose gel. A 1Kb DNA ladder was used as a molecular marker (Lane M). Lanes 1-10 show the PCR products of the fragment E (nt8024-9231) of 10 Egyptian samples with a product size of 1208bp. Lanes 11-20 show the PCR products of the fragment D (nt4875-7335) of the same samples with a product size of 2461bp.

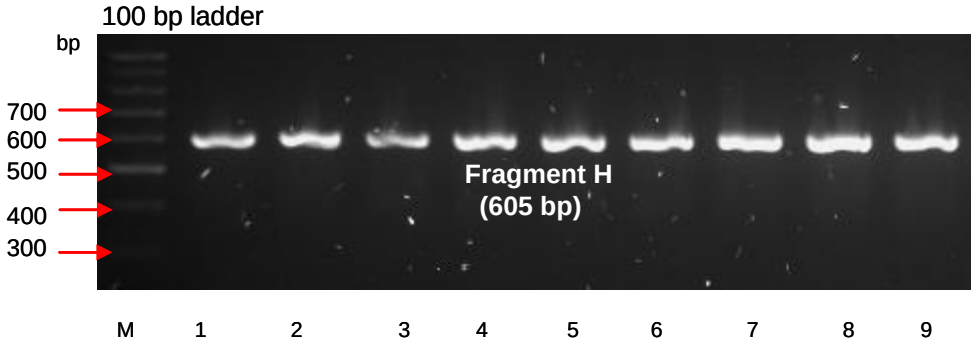



**Figure 5.6: PCR amplification products of the fragment H of the mitochondrial genome (nt14253-14857).** PCR products of the fragment H (605 bp) were separated on 2% agarose gel. A 100 bp DNA ladder was used as a molecular marker (Lane M). Lanes 1-9 show PCR products of nine Egyptian samples.

### 5.2.2 Restriction digestions of the PCR products (RFLP analysis)

A pilot study was performed in order to identify what mitochondrial DNA haplogroups were in the Egyptian population. These preliminary experiments involved screening of 64 Egyptian individuals for all the previously described haplogroup specific RFLP markers (Bailliet et al. 1994; Macaulay et al. 1999; Alves-Silva et al. 2000; Finnila et al. 2001b; Tolk et al. 2001), which meant that the PCR products of the specified nine regions of the mitochondrial genome were subjected to a total of 26 restriction digestions for each sample.

#### *5.2.2.1 HaeIII restriction digestion of the fragment A of the mitochondrial genome (nt451-827)*

The PCR product of the fragment A (377bp) was digested with *HaeIII* endonuclease to screen for an A663G mutation which is specific to the mitochondrial haplogroup A and leads to gain of a new *HaeIII* recognition sequence. All the screened samples lacked that mutation (**Figure 5.7**).

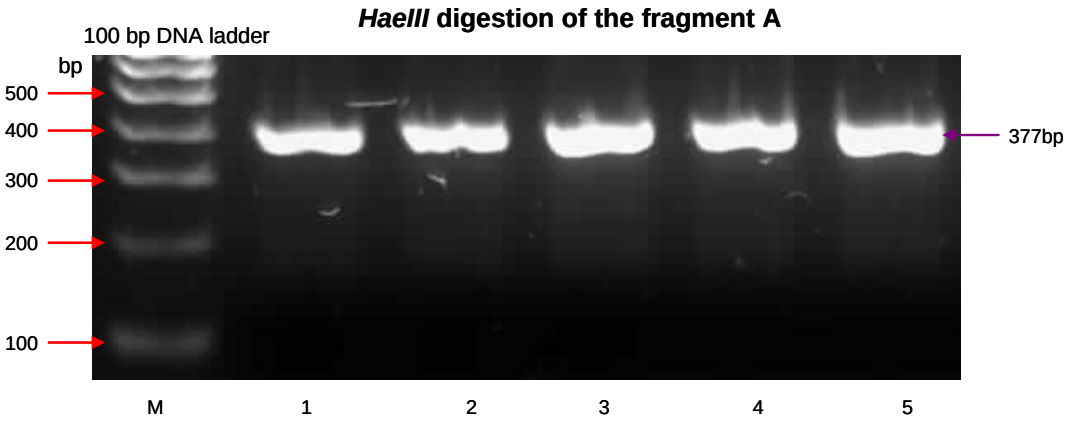



**Figure 5.7: *HaeIII* digestion products of the fragment A of the mitochondrial genome (nt451-827).** Digestion products were separated on 2% agarose gel by electrophoresis. A 100 bp DNA ladder was used as a molecular marker (Lane M). Lanes 1-5 show *HaeIII* digestion products of five Egyptian samples. None of samples contained a *HaeIII* cut site and thus the PCR products remained at the same size of 377bp.

### *5.2.2.2 DdeI and MboI restriction digestions of the fragment B of the mitochondrial genome (nt1489-2690)*

The PCR products of the fragment B (1202bp) were digested with *DdeI* endonuclease to screen for a G1719A substitution which characterises the mitochondrial haplogroups N, I and X2. This mutation disrupts the *DdeI* recognition sequence at that position and hence a *DdeI* cut site is lost (**Figure 5.8**). The same PCR products were digested with *MboI* endonuclease to screen for a T2352C mutation which specifies the African mitochondrial subclades L1b and L3e. This mutation produces a novel *MboI* cut site (**Figure 5.9**).

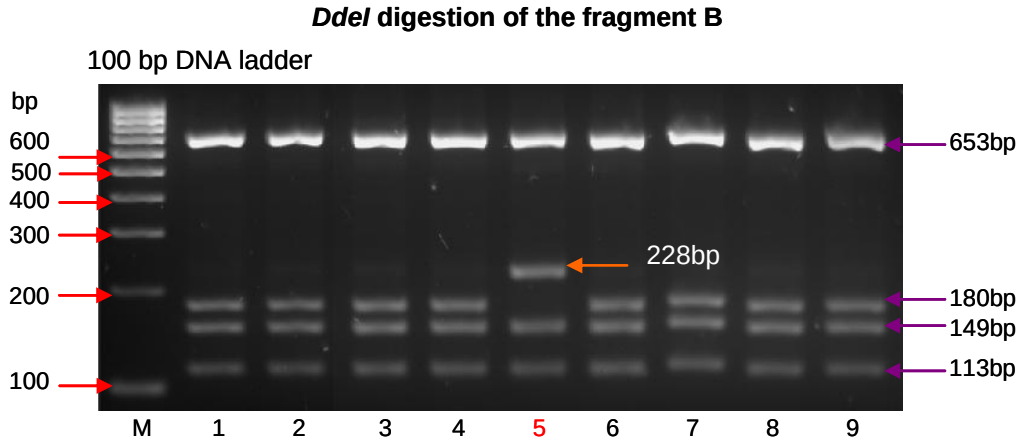



**Figure 5.8: *DdeI* digestion products of the fragment B of the mitochondrial genome (nt1489-2690).** Digestion products were separated on 3.5% agarose gel. A 100 bp DNA ladder was used as a molecular marker (Lane M). Lanes 1-9 show *DdeI* digestion products of nine Egyptian samples. All samples (except sample in lane 5) showed identical digestion pattern to that of the CRS. Lane 5 showed a sample belonging to mitochondrial haplogroup N, I or X2 due to loss of a *DdeI* cut site produced by a G1719A mutation.

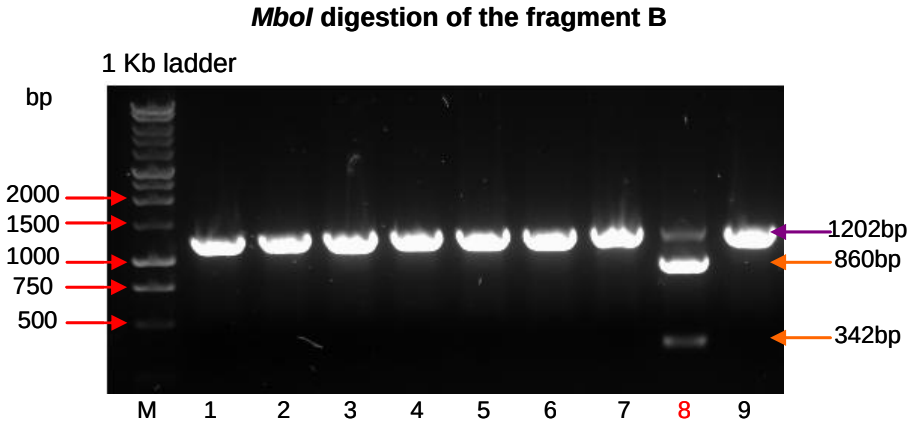



**Figure 5.9: *MboI* digestion products of the fragment B of the mitochondrial genome (nt1489-2690).** Digestion products were separated on 1% agarose gel. A 1Kb DNA ladder was used as a molecular marker (Lane M). Lanes 1-9 show *MboI* digestion products of nine Egyptian samples. All samples (except sample in lane 8) did not show any evidence of digestion and retained the product size of 1202bp. Lane 8 contained a sample belonging to mitochondrial haplogroup L1b or L3e as evidenced by a new *MboI* cut site resulting from a T2352C mutation and producing two bands of sizes 860 and 342 bp. Note full length DNA fragment in lane 8 most likely indicates incomplete digestion of PCR product.

### *5.2.2.3 HpaI and NlaIII restriction digestions of the fragment C of the mitochondrial genome (nt3314-4732)*

PCR products of the fragment C (1419bp) were digested with *HpaI* endonuclease to screen for the African specific C3594T mutation which produces a new *HpaI* recognition sequence. This cut site gain is a diagnostic marker of the major African mitochondrial haplogroups L0, L1, L2, L5 and L6 (**Figure 5.10**). The same PCR products were digested with *NlaIII* endonuclease to screen for a T4216C mutation which is specific to mitochondrial haplogroup JT and its descendants. This mutation leads to a new *NlaIII* cut site which characteristically identifies the haplogroup JT (**Figure 5.11**). *NlaIII* digestion of the fragment C was also used to screen for a G4580A mutation which causes loss of an *NlaIII* cut site. This mutation is diagnostic to mitochondrial haplogroup V (**Figure 5.11**).

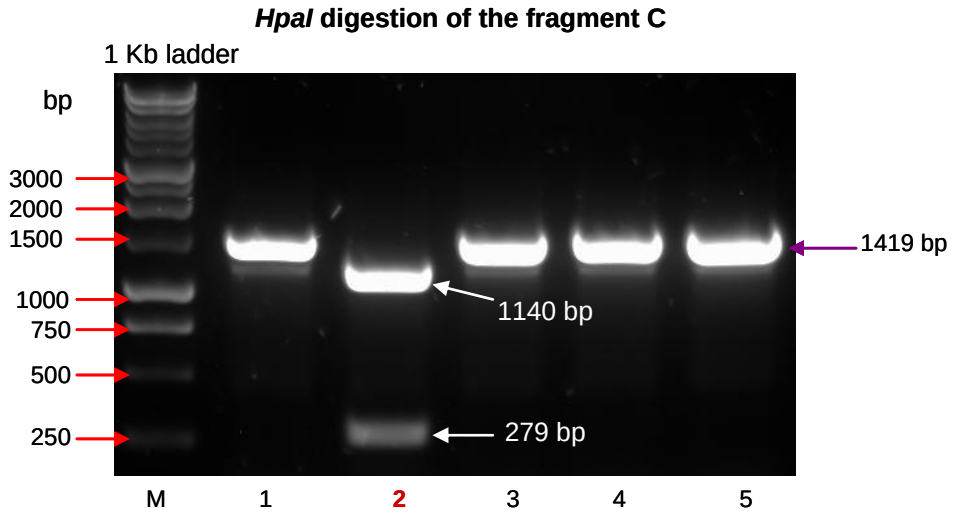



**Figure 5.10: *HpaI* digestion products of the fragment C of the mitochondrial genome (nt3314-4732).** Digestion products were separated on 1% agarose gel. A 1Kb DNA ladder was used as a molecular marker (Lane M). Lanes 1-5 show *HpaI* digestion products of five Egyptian samples. All samples (except sample in lane 2) lacked an *HpaI* cut site and retained the PCR product size of 1419bp. Sample in lane 2 belonged to one of the major African mitochondrial DNA haplogroups as evidenced by gain of an *HpaI* cut site resulting from the African specific C3594T mutation leading to cutting of the PCR product into two bands of sizes 1140 and 279 bp.

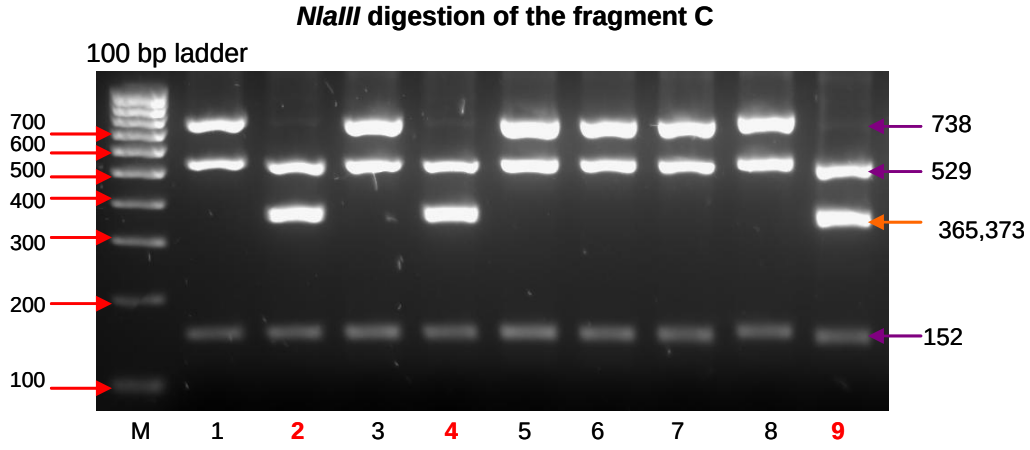



**Figure 5.11: *NlaIII* digestion products of the fragment C of the mitochondrial genome (nt3314-4732).** Digestion products were separated on 2.5% agarose gel. A 100 bp DNA ladder was used as a molecular marker (Lane M). Lanes 1-9 show *NlaIII* digestion products of nine Egyptian samples. All samples (except samples in lanes 2, 4, and 9) showed a digestion pattern identical to that of the CRS. Samples in lanes 2, 4, and 9 showed an additional *NlaIII* cut site resulting from a T4216C mutation which is diagnostic to the mitochondrial haplogroup JT. None of samples showed loss of *NlaIII* cut site relative to the CRS digestion pattern which is a diagnostic marker of the mitochondrial haplogroup V.

### *5.2.2.4 AluI restriction digestion of the fragment D of the mitochondrial genome (nt4875-7335)*

The PCR products of the fragment D (2461bp) were digested with *AluI* endonuclease to screen for loss of *AluI* cut site at position 7028 which is a specific diagnostic RFLP marker to the mitochondrial haplogroup H to which the CRS belongs. *AluI* digestion was also used to identify haplogroup D which is characterised by a C5178A mutation leading to loss of *AluI* cut site at that position, however, none of the analysed Egyptian mtDNA samples showed that marker (**Figure 5.12**).

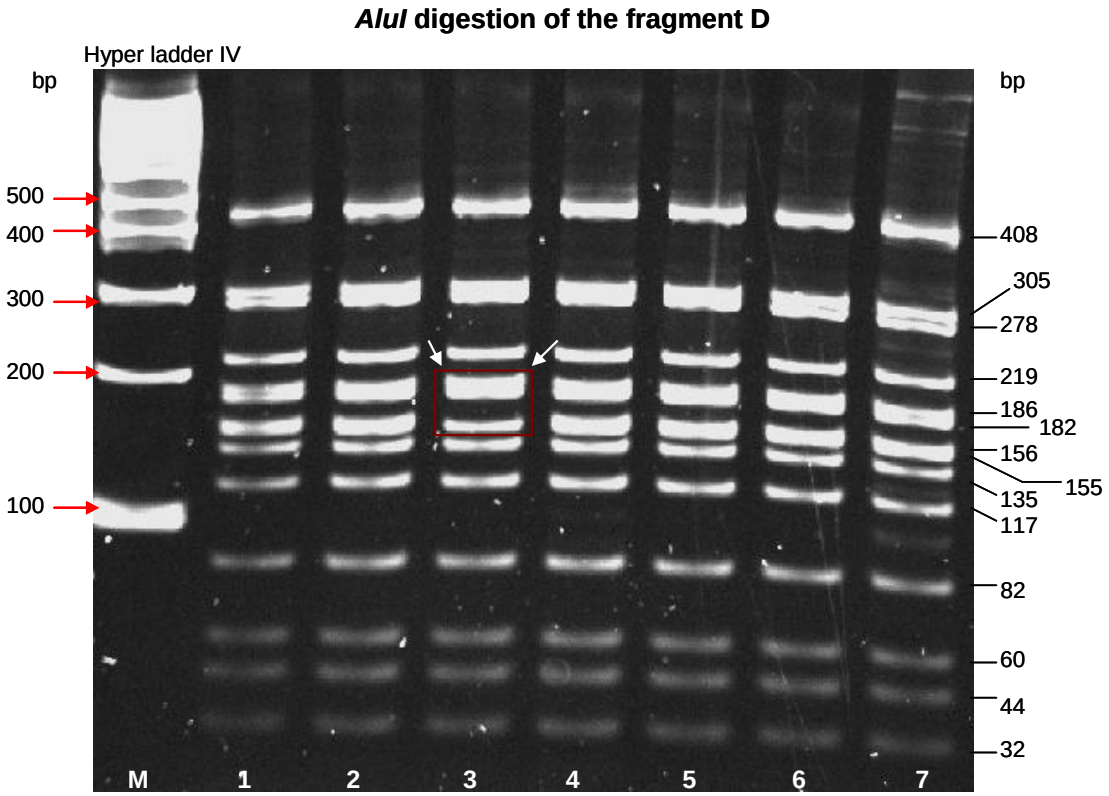



**Figure 5.12: Analytical gel of *AluI* digestion products of the fragment D of the mitochondrial genome (nt4875-7335).** Digestion products were separated on 8% polyacrylamide gel. Hyper ladder IV was used as a molecular marker (Lane M). Lanes 1-7 showed digestion products of 7 Egyptian samples. Sample in lane 3 showed a different digestion pattern at the region outlined by a red box, where the 156bp band was lost and a new 188bp band appeared due to loss of *AluI* cut site at 7028 and thus this sample was assigned to mitochondrial haplogroup H. None of samples showed the marker of haplogroup D which would be merging of the 408 and 186bp bands into a 594bp band.

Due to occurrence of several fragments that are very similar in size, the analysis of *AluI* digestion products of the fragment D was difficult and considered uncertain. To tackle this difficulty, a new primer was designed (Mt6067 F) and together with the original Mt D R reverse primer was used to amplify the modified version of the fragment D between the nucleotides 6067 and 7335 giving a product size of 1269bp (**Figure 5.13**). *AluI* digestion products of the modified fragment D were clearly separated by electrophoresis and unambiguously interpreted and hence, the Egyptian mtDNA samples were assigned to the mitochondrial haplogroup H with confidence (**Figure 5.14**).

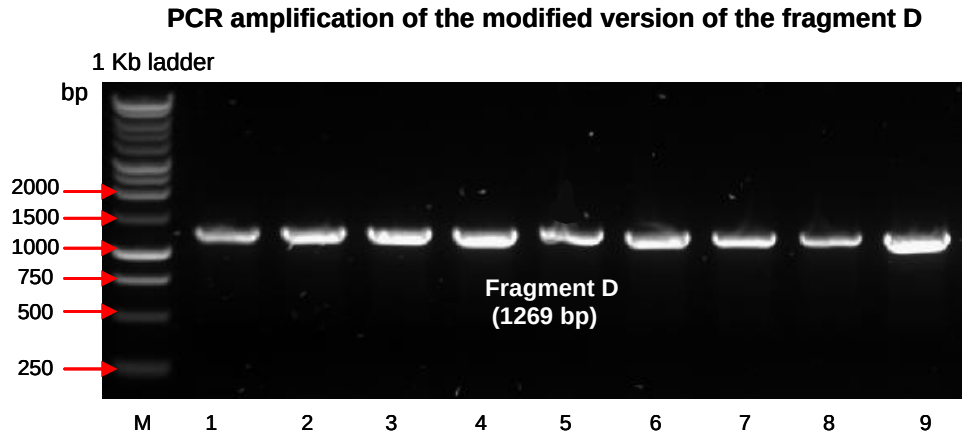



**Figure 5.13: PCR amplification products of the modified version of the fragment D of the mitochondrial genome (nt6067-7335).** PCR products of size 1269bp were separated on 1% agarose gel. A 1Kb DNA ladder was used as a molecular marker (Lane M). Lanes 1-9 show PCR products of nine Egyptian samples.

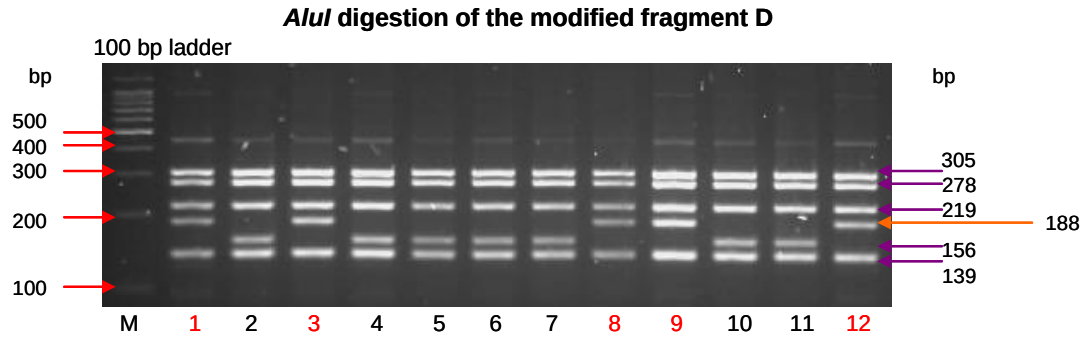



**Figure 5.14: Analytical gel of *AluI* digestion products of the modified version of the fragment D of the mitochondrial genome (nt6067-7335).** Digestion products were separated on 3.5% agarose gel. A 100 bp DNA ladder was used as a molecular marker (Lane M). Lanes 1-12 show digestion products of twelve Egyptian samples. Samples labelled in red show a digestion pattern similar to the CRS by having the 188bp band which indicates loss of *AluI* cut site at position 7028 and thus these samples were assigned to the mitochondrial haplogroup H. The other samples (labelled in black) show the 156bp band which indicates that the *AluI* enzyme cut the 188bp fragment at position 7028.

### *5.2.2.5 RFLP analysis of the fragment E of the mitochondrial genome (nt8024-9231)*

The PCR products of the fragment E (1208bp) contained several diagnostic RFLP markers. In order to screen for them all, the fragment E was individually digested with four endonucleases; *AvaII, HaeII, HaeIII,* and *MboI*. *AvaII* and *HaeIII* restriction digestions were successfully used to screen for a G8251A mutation which characterises a group of mitochondrial lineages that represent direct descendants of the major macro-haplogroup N, including N1b, N1e, I, and W. Egyptian individuals belonging to these mitochondrial lineages showed gain of a new *AvaII* cut site and loss of *HaeIII* cut site at the specified mutation position (**Figures 5.15 and 5.16**). *HaeIII* digestion also helped to specifically identify mitochondrial haplogroup W which is characterised by a G8994A mutation causing additional loss of *HaeIII* cut site at that position (**Figure 5.16**). In one individual (sample ID 217), the *HaeIII* cut site at 9025 was lost, which could be due to any mutation at nt9025-9028 (**Figure 5.16**, lane 3). A G to A mutation was previously reported at 9025 in the Polish population (Piechota et al. 2004) and a C9027G mutation was previously described in patients with MELAS syndrome (Mitochondrial Myopathy, Encephalopathy, Lactic Acidosis, and Strokelike Episodes) (Kobayashi et al. 1991; Eyre-Walker A et al. 1999).

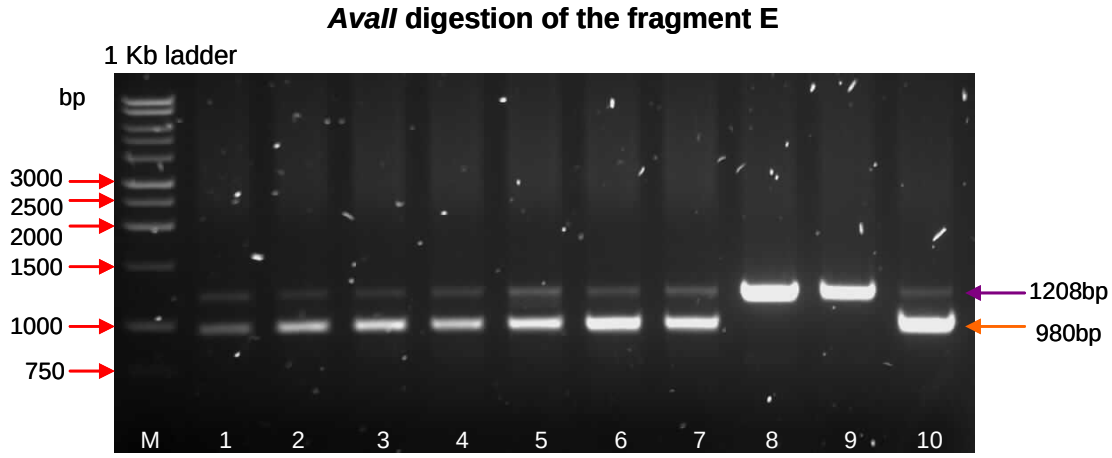



**Figure 5.15: *AvaII* digestion products of the fragment E of the mitochondrial genome (nt8024-9231).** Digestion products were separated on 1% agarose gel. A 1Kb DNA ladder was used as a molecular marker (Lane M). Lanes 1-10 show digestion products of ten Egyptian individuals. All samples (except 8 and 9) showed evidence of an internal *AvaII* cut site in the fragment E which produced two bands of sizes 980 and 227bp and thus all these eight individuals were assigned to a group of N derived mitochondrial lineages including N1b, N1e, I, and W. The faint 1208bp bands seen in all digested samples represent some uncut products due to the presence of too much DNA in the PCR products as evidenced by the brightness of the other undigested PCR products in lanes 8 and 9. The 227bp bands ran out of the gel.

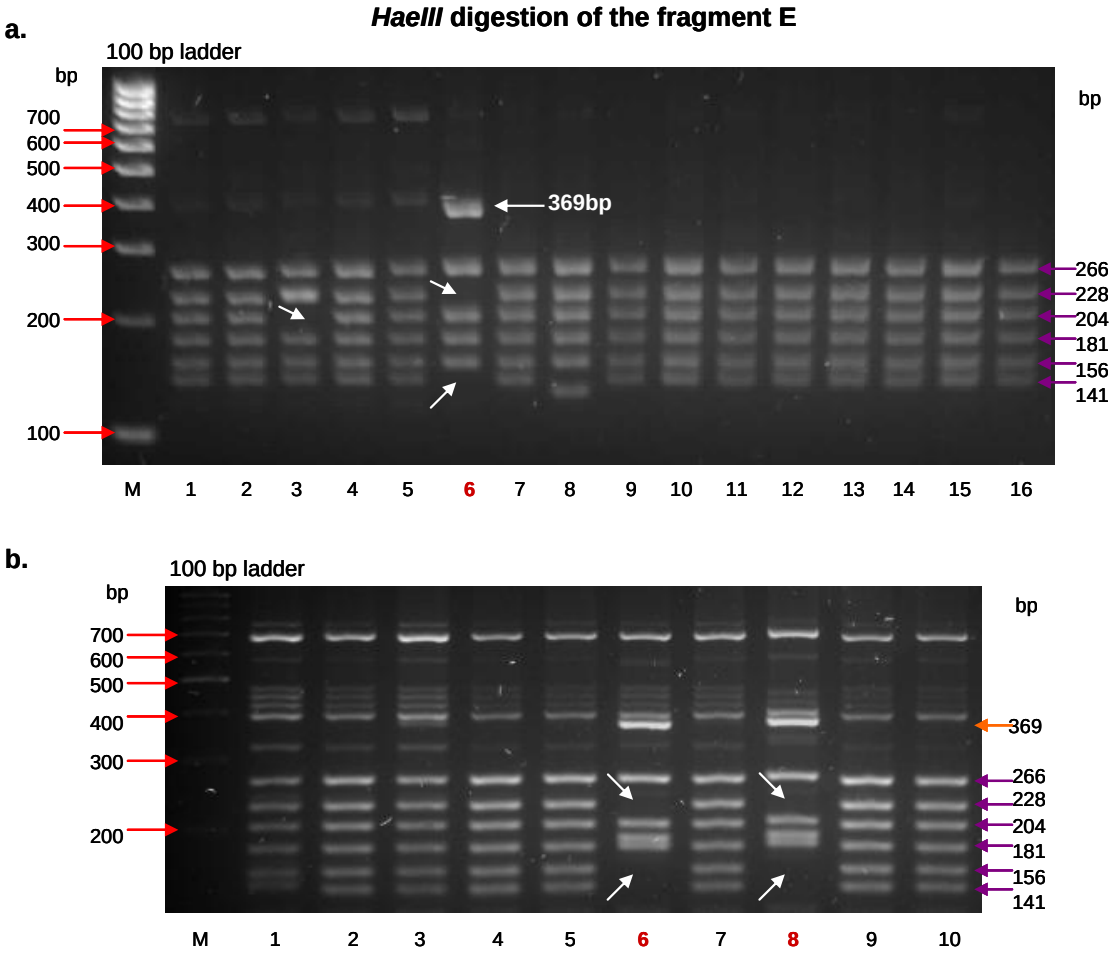



**Figure 5.16: *HaeIII* digestion products of the fragment E of the mitochondrial genome (nt8024-9231).** Digestion products were separated on 3.5% agarose gel. A 100bp DNA ladder was used as a molecular marker (Lane M). In gel (a), all samples (except 3, 6, and 8) showed a digestion pattern identical to that of the CRS. Sample 6 showed a loss of the 228 and 141bp fragments which merged in a new 369bp fragment indicating loss of *HaeIII* cut site at position 8251 which is a marker for haplogroups N1b, N1e, I and W. Sample 3 showed loss of the 204bp fragment which indicates loss of another *HaeIII* cut site at position 9025 but this was not a haplogroup specific marker. Also for sample 8 which showed an additional cut site between the nucleotides 8251 and 8393 by having a fragment smaller than 141bp, but this cut site again was not a haplogroup specific marker. In gel (b), all samples (except 6, and 8) showed a digestion pattern identical to that of the CRS. For samples 6 and 8, three bands (228, 156, and 141bp) were lost indicating loss of 2 *HaeIII* cut sites at 8251 and 8994 which identified those samples as belonging to mitochondrial haplogroup W.

*MboI* digestion of the fragment E was performed to screen for a T8618C mutation which identifies mitochondrial haplogroup L3d by causing loss of *MboI* cut sit at that position, however, none of the screened Egyptian samples belonged to this haplogroup. *HaeII* digestion of the fragment E was performed to differentiate between the mitochondrial haplogroup K and its ancestral haplogroup U. A characteristic G9055A mutation occurred in the mitochondrial haplogroup K and caused loss of *HaeII* cut site at that position which was considered as a diagnostic marker of this haplogroup in the present study (**Figure 5.17**).

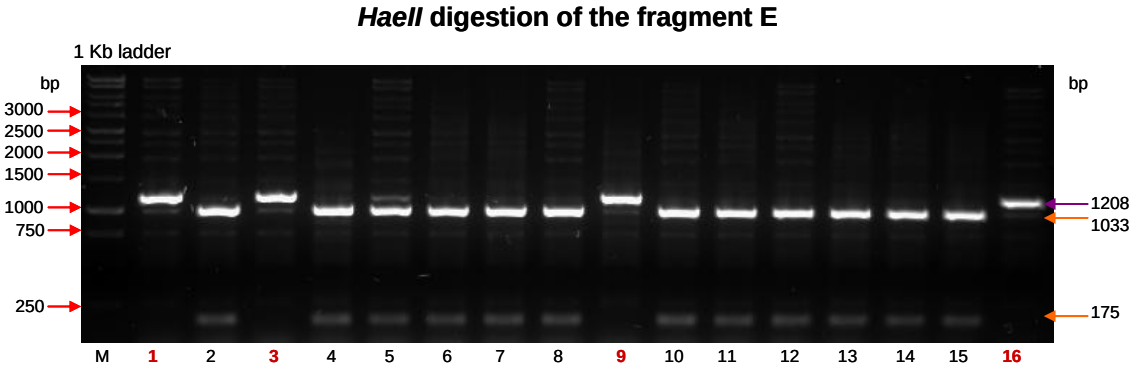



**Figure 5.17: *HaeII* digestion products of the fragment E of the mitochondrial genome (nt8024-9231).** Digestion products were separated on 1% agarose gel. A 1Kb DNA ladder was used as a molecular marker (Lane M). Lanes 1-16 showed digestion products of 16 Egyptian mtDNA samples. All samples (Except 1, 3, 9, and 16) showed an internal *HaeII* cut site in the fragment E as evidenced by cutting the PCR product into 2 fragments of sizes 1033 and 175bp. Samples 1, 3, 9, and 16 lost this *HaeII* cut site which was due to a G9055A mutation and hence they were assigned to mitochondrial haplogroup K.

### *5.2.2.6 RFLP analysis of the fragment F of the mitochondrial genome (nt9921-11150)*

The PCR products of the fragment F of the mitochondrial genome (1230bp) contained numerous RFLP markers which were proved to be very useful in designation of the different mitochondrial DNA haplogroups. Therefore, this region of the mtDNA was digested with four restriction enzymes individually; *AluI*, *DdeI*, *HinfI*, and *TaqI*. *DdeI* restriction digestion of the fragment F was used to classify the Egyptian mtDNA samples into two main categories of the mitochondrial haplogroups, one category included the old mitochondrial DNA lineages (African L, M, N, J and K1) and the other category included all the other more recently derived mitochondrial haplogroups. The old category was identified by having a *DdeI* cut site at position 10398 which was caused by an A10398G mutation, while those lacking this cut site were classified under the recent category of mitochondrial haplogroups (**Figure 5.18**).

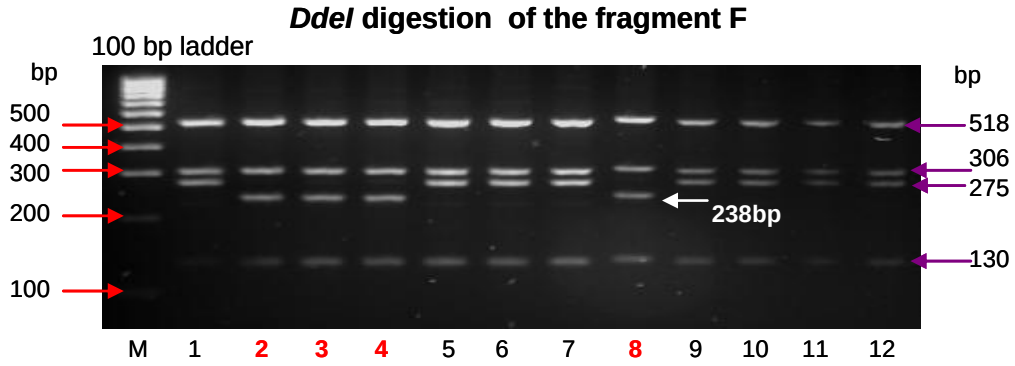



**Figure 5.18: *DdeI* digestion products of the fragment F of the mitochondrial genome (nt9921-11150).** Digestion products were separated on 3.5% agarose gel. A 100bp DNA ladder was used as a molecular marker (Lane M). Lanes 1-12 show digestion products of twelve Egyptian mtDNA samples. Samples labelled in black showed a digestion pattern identical to that of the CRS, whereas samples in red showed an additional *DdeI* cut site as evidenced by the formation of a 238bp fragment, which was caused by an A10398G mutation, and thus were assigned to the category of old derived mitochondrial lineages (L, M, N, J, or K1).

*AluI* digestion of the same region of the mtDNA was used to screen for a C10400T mutation which specifically characterises haplogroup M, and thus Egyptian individuals were assigned to mitochondrial haplogroup M when having both a *DdeI* cut site at 10398 and an *AluI* cut site at 10400 (**Figure 5.19**). In addition, *AluI* digestion helped in the identification of mitochondrial haplogroup I which is characterised by a T10034C mutation that produces an additional diagnostic *AluI* cut site (**Figure 5.19**).

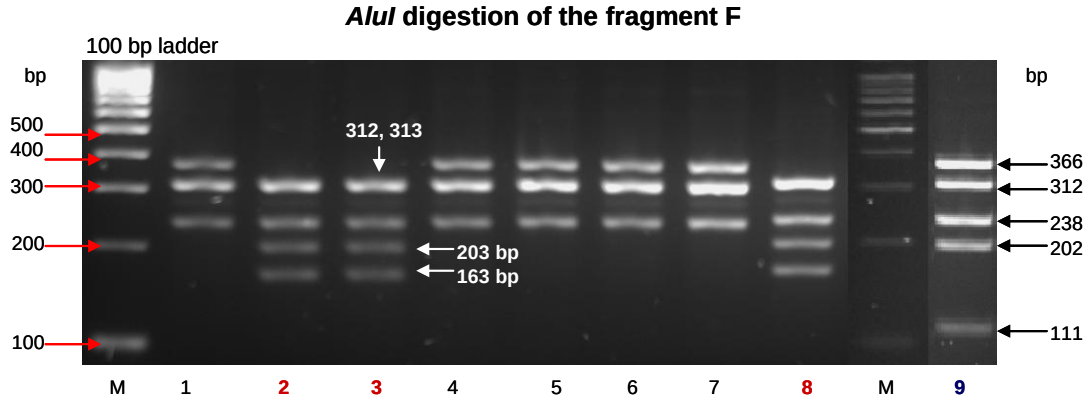



**Figure 5.19: *AluI* digestion products of the fragment F of the mitochondrial genome (nt9921-11150).** Digestion products were separated on 3.5% agarose gel. A 100bp DNA ladder was used as a molecular marker (Lanes M). Lanes 1-8 and 9 show digestion products of nine Egyptian mtDNA samples. Samples 1 and 4-7 showed identical digestion pattern to that of the CRS while samples 2, 3, and 8 showed gain of a new *AluI* cut site relative to the CRS as evidenced by loss of the 366bp band and appearance of two new bands of sizes 203 and 163bp. That was caused by a C10400T mutation, and thus these samples were assigned to mitochondrial haplogroup M. Sample 9 showed a different digestion pattern resulting from a T10034C mutation which produced a new *AluI* cut site that split the 313bp band into 202 and 111bp bands and thus, it was assigned to haplogroup I.

To further aid the haplogroup assignment of the Egyptian mtDNA samples, the PCR products containing fragment F was digested with *HinfI* endonuclease. The African mitochondrial haplogroup L5 was identified when an additional *HinfI* cut site relative to the CRS was detected at position 10806 which was caused by a T10810C mutation (**Figure 5.20**). *TaqI* restriction digestion of the fragment F was performed to screen for an A10086G mutation which characterises mitochondrial haplogroup L3b, however, none of the screened Egyptian individuals showed that marker.

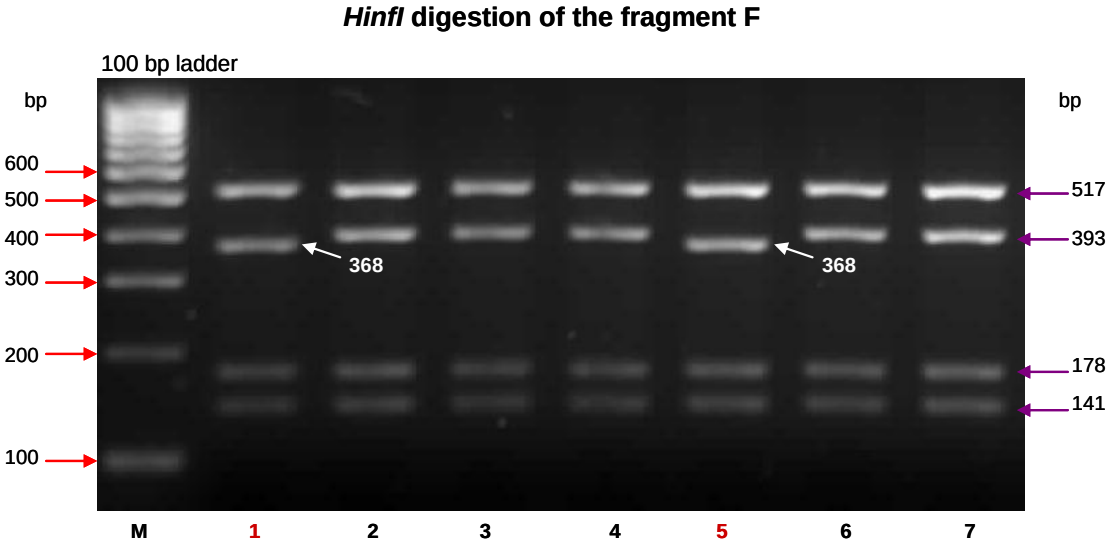



**Figure 5.20: *HinfI* digestion products of the fragment F of the mitochondrial genome (nt9921-11150).** Digestion products were separated on 2.5% agarose gel. A 100bp DNA ladder was used as a molecular marker (Lane M). Lanes 1-7 show digestion products of seven Egyptian samples. All samples (except 1 and 5) showed a digestion pattern identical to that of the CRS. Samples 1 and 5 showed an additional *HinfI* cut site as evidenced by the production of a 368bp fragment which was caused by a T10810C mutation. Hence, these two samples were assigned to mitochondrial haplogroup L5.

### *5.2.2.7 RFLP analysis of the fragment G of the mitochondrial genome (nt12077-13812)*

The PCR products of the fragment G (1736bp) was digested with three restriction enzymes individually; *AluI*, *BamHI*, and *HincII*. To screen for a G13368A mutation which characterises mitochondrial haplogroup T, *BamHI* digestion was performed. Samples that showed an evidence of *BamHI* cut were assigned to haplogroup T (**Figure 5.21**). Both *AluI* and *HincII* restriction digestions were performed to screen for an A13263G mutation which identifies mitochondrial haplogroup C. Such a mutation produces a novel *AluI* cut site and causes loss of a *HincII* cut site, however these markers were not detected among the screened Egyptian samples.

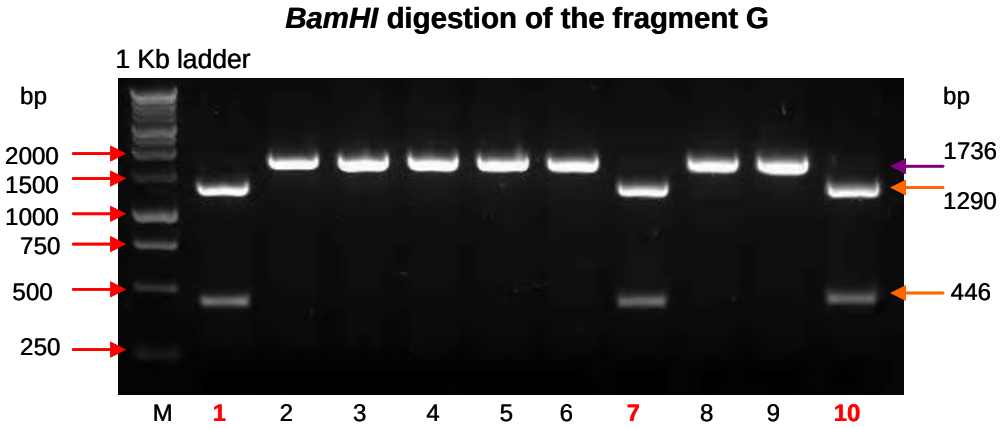



**Figure 5.21: *BamHI* digestion products of the fragment G of the mitochondrial genome (nt12077-13812).** Digestion products were separated on 1% agarose gel. A 1Kb DNA ladder was used as a molecular marker (Lane M). Lanes 1-10 show digestion products of 10 Egyptian samples. All samples (except 1, 7, and 10) did not show evidence of *BamHI* cut. The region G of samples 1, 7, and 10 contained an internal *BamHI* cut site as evidenced by having two bands of sizes 1290 and 446bp instead of having a single band of 1736bp which was caused by a G13368A mutation. Thus, these samples were assigned to mitochondrial haplogroup T.

### *5.2.2.8 RFLP analysis of the fragment H of the mitochondrial genome (nt14253-14857)*

The PCR products of the fragment H of the mitochondrial genome (605bp) were digested with *AccI* and *MseI* restriction enzymes which served to identify mitochondrial haplogroups X and HV, respectively. The Egyptian mtDNA samples were assigned to haplogroup X when they showed evidence of *AccI* cut in the fragment H which was caused by a haplogroup specific T to C mutation at position 14470 (**Figure 5.22**). *MseI* digestion was performed to identify mitochondrial haplogroup HV and all its descendants including haplogroup H to which the CRS belongs. Egyptian mtDNA samples were assigned to haplogroup HV and its descendants whenever they showed loss of an *MseI* cut site, which was caused by a C14766C mutation (**Figure 5.23**).

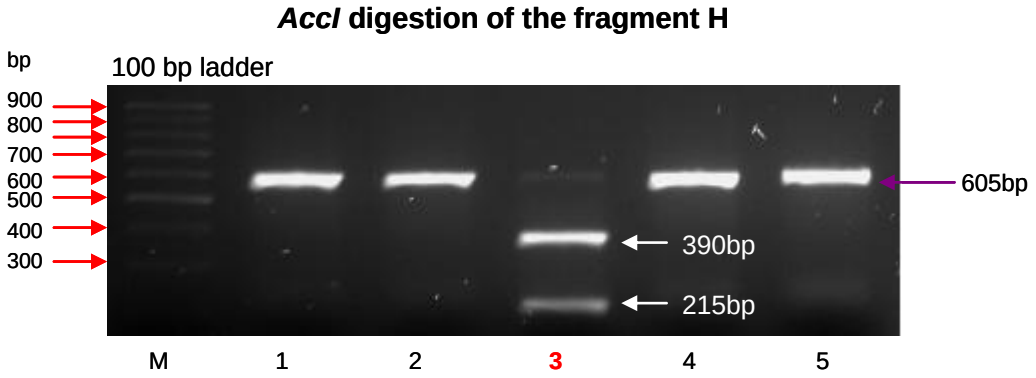



**Figure 5.22: *AccI* digestion products of the fragment H of the mitochondrial genome (nt14253-14857).** Digestion products were separated on 2% agarose gel. A 100bp DNA ladder was used as a molecular marker (Lane M). Lanes 1-5 show digestion products of five Egyptian samples. All samples (except sample 3) did not show evidence of *AccI* cut. Sample 3 showed cutting of the PCR product (605bp) into two fragments of sizes 390 and 215bp which was caused by a T14470C mutation and therefore, it was assigned to mitochondrial haplogroup X.

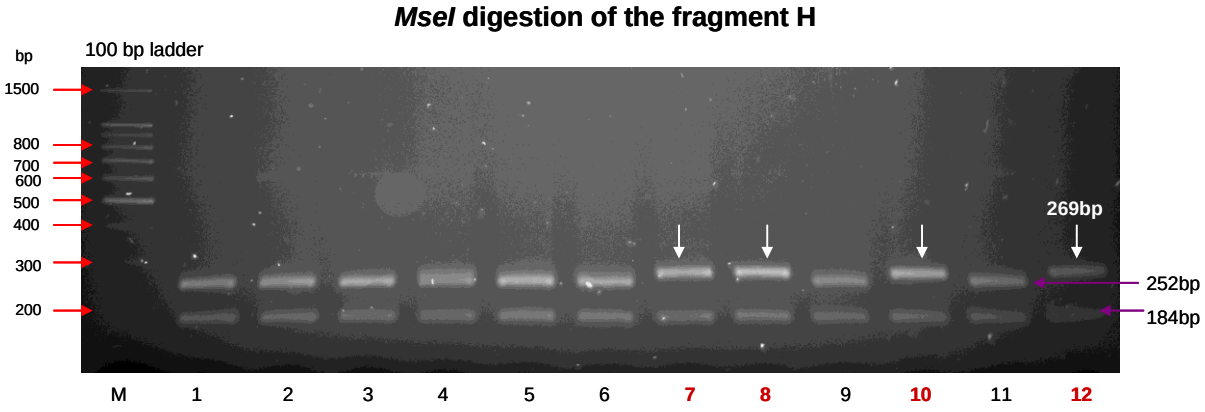



**Figure 5.23: *MseI* digestion products of the fragment H of the mitochondrial genome (nt14253-14857).** Digestion products were separated on 2.5% agarose gel. A 100bp DNA ladder was used as a molecular marker (Lane M). Lanes 1-12 show digestion products of twelve Egyptian samples. Evidence of *MseI* cut site loss was detected in samples 7, 8, 10 and 12 by showing a fragment of 269bp instead of 252bp which was caused by a C14766C mutation, while all other samples did not show a similar *MseI* cut site loss. Therefore, these four samples were assigned to mitochondrial haplogroup HV and its descendants.

### *5.2.2.9 RFLP analysis of the fragment I of the mitochondrial genome (nt15312-149)*

The PCR products of the fragment I of the mitochondrial genome (1407bp) were digested with *AluI* and *HinfI* restriction enzymes to differentiate between two closely related mitochondrial DNA haplogroups, T and J. *AluI* digestion aimed to screen for an A15607G mutation which characterises the haplogroup T and was detected by detection of a novel *AluI* cut site relative to the CRS digestion pattern (**Figure 5.24**). *HinfI* digestion was performed to confirm assignment of the Egyptian mtDNA samples to haplogroup J through detection of a C16069T mutation which is specific to this haplogroup. This mutation causes loss of a *HinfI* cut site relative to the CRS digestion pattern (**Figure 5.25**).

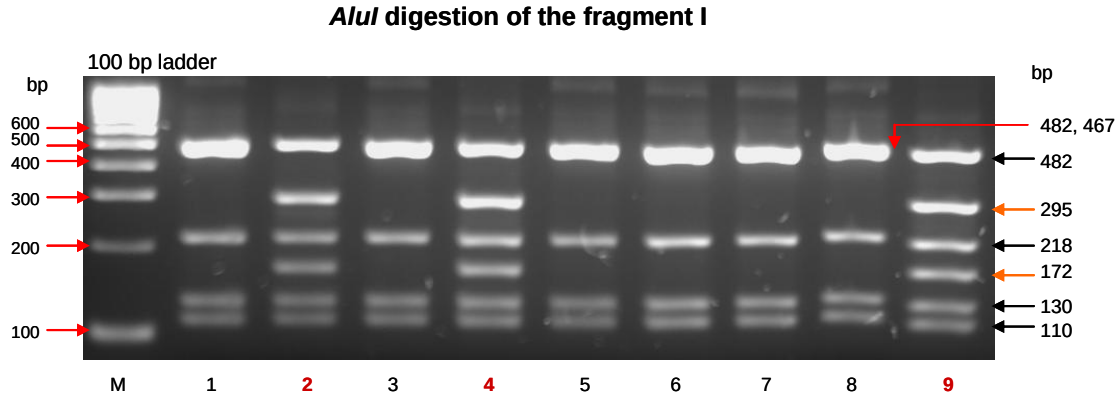



**Figure 5.24: *AluI* digestion products of the fragment I of the mitochondrial genome (nt15312-149).** Digestion products were separated on 3.5% agarose gel. A 100 bp DNA ladder was used as a molecular marker (Lane M). Lanes 1-9 show digestion products of nine Egyptian samples. All samples (except 2, 4, and 9) showed a digestion pattern identical to that of the CRS. Samples 2, 4, and 9 showed an additional *AluI* cut site which was evidenced by cutting the 467bp band into two bands of sizes 295 and 172bp. This new cut site was caused by an A15607G mutation and thus those three samples were assigned to mitochondrial haplogroup T.

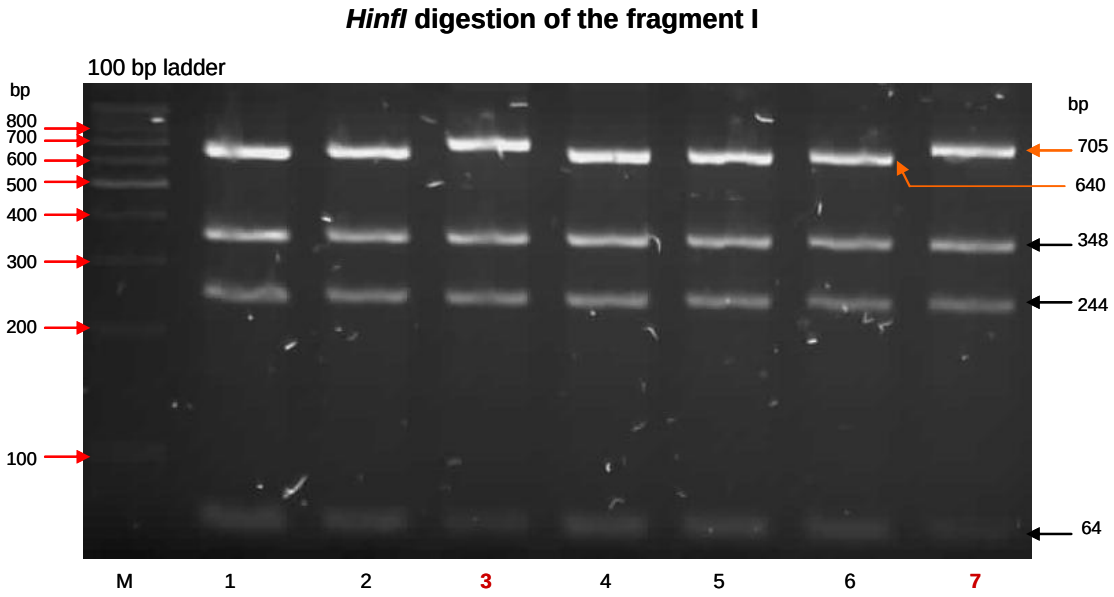



**Figure 5.25: *HinfI* digestion products of the fragment I of the mitochondrial genome (nt15312-149).** Digestion products were separated on 3.5% agarose gel. A 100 bp DNA ladder was used as a molecular marker (Lane M). Lanes 1-7 show digestion products of seven Egyptian samples. All samples (except 3 and 7) showed a digestion pattern identical to that of the CRS. Samples 3 and 7 showed loss of a *HinfI* cut site as evidenced by absence of the 640bp fragment and occurrence of a 705bp fragment. This cut site was lost due to the presence of a C16069T mutation which is specific to haplogroup J. Therefore, samples 3 and 7 were assigned to the mitochondrial haplogroup J.

### 5.2.3 Screening for an A12308G mutation as a diagnostic marker to mitochondrial haplogroups U and K

A new fragment of the mitochondrial genome (114bp) was amplified using mispairing PCR (**Figure 5.26**) aiming to screen for an A12308G mutation which in conjunction with the mispairing primer generates a new *EcoRI* restriction cut site. This mutation is specific to the mitochondrial haplogroups U and K. Therefore, the Egyptian individuals whose mtDNA showed an *EcoRI* cut site at that position were confidently assigned to haplogroups U and K (**Figure 5.27**).

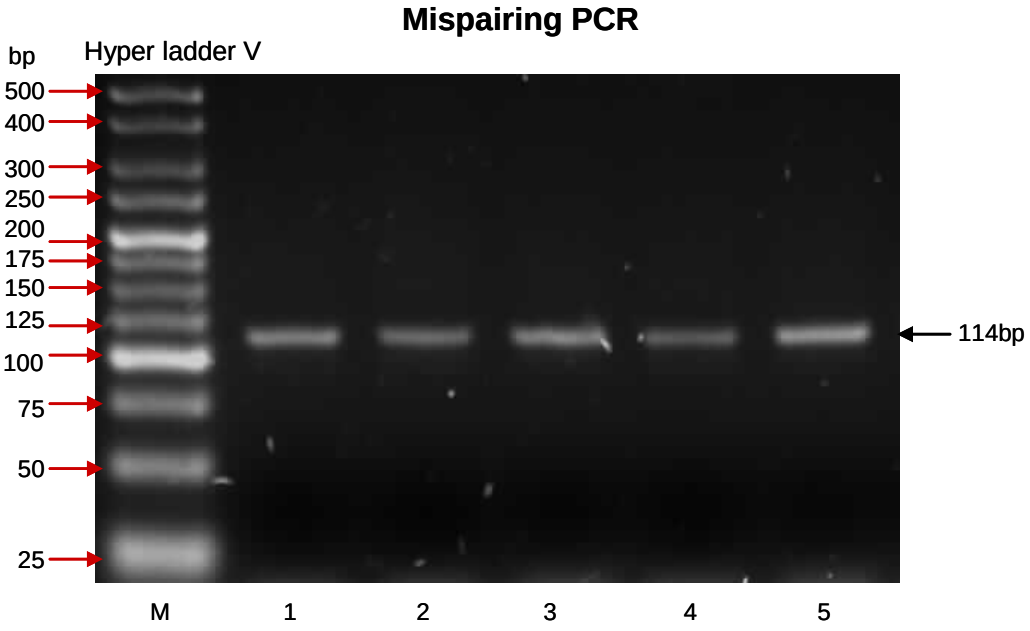



**Figure 5.26: Mispairing PCR amplification products of the mitochondrial DNA (nt12221-12334).** PCR products (114bp) were separated on 2.5% agarose gel. Hyper ladder V was used as a molecular marker (Lane M). Lanes 1-5 show PCR products of five Egyptian mtDNA samples.

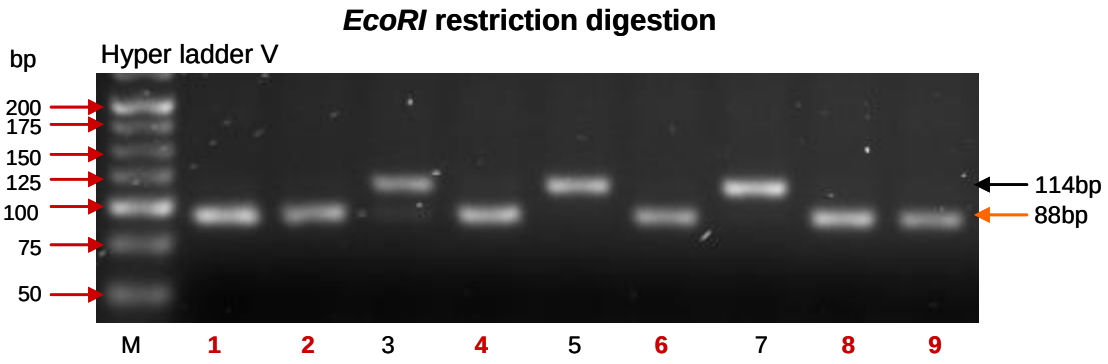



**Figure 5.27: *EcoRI* digestion products of the mispairing PCR (nt12221-12334).** Digestion products were separated on 3.5% agarose gel. Hyper ladder V was used as a molecular marker (Lane M). Lanes 1-9 show digestion products of nine Egyptian samples. All samples (except 3, 5, and 7) showed an *EcoRI* cut of the PCR product as evidenced by loss of the 114bp band and appearance of a new 88bp band. This cut site was produced by an A12308G mutation in conjunction with the mispairing primer, therefore, these Egyptian mtDNA samples were assigned to mitochondrial haplogroup U and K. The PCR products of samples 3, 5, and 7 lacked an *EcoRI* cut as evidenced by having the same size band as the PCR product (114bp).

### 5.2.4 Development of a new strategy for RFLP analysis of the mitochondrial genome

Following completion of the pilot study which involved screening of 64 Egyptian individuals for all the 26 haplogroup specific RFLP markers and consideration of the results obtained, a new strategy for RFLP analysis was developed. This new strategy involved screening of samples for only 18 RFLP markers in a specific sequence, which enabled rapid and successful assignment of all the Egyptian individuals to well established mitochondrial DNA haplogroups. The scheme was primarily based on splitting the mitochondrial DNA samples into two big sets of mitochondrial lineages based on presence or absence of a *DdeI* cut site at 10394 which differentiates the old mitochondrial lineages (haplogroups L, M, N, J and K1) from the more recently derived mitochondrial lineages (haplogroups H, HV, R0, R0a, T, U, K, X, and W). This scheme involved a certain sequence of restriction digestions in which the results of each digestion determined the following restriction digestion till identification of a certain mitochondrial DNA haplogroup was achieved (**Figure 5.28**). For example, step 1 was digestion of the PCR product of fragment F (nt9921-11150) of the mtDNA with *DdeI* restriction enzyme, if the sample had a *DdeI* cut site at 10394, then step 2 was digestion of fragment C (nt3314-4732) with *HpaI* restriction enzyme and fragment B (nt1489-2690) with *MboI* restriction enzyme. Based on the digestion results, if the sample lacked an *HpaI* recognition sequence at 3592 and an *MboI* recognition sequence at 2349, then step 3 was digestion of fragment F with *AluI* restriction enzyme, if the sample had an *AluI* recognition sequence at 10397, then mitochondrial haplogroup M was identified (**Figure 5.28**). The developed strategy was validated by comparing the haplogroup assignment results obtained by the RFLP analysis to those obtained by mitochondrial hypervariable region analysis and it was proved to be a highly efficient scheme.

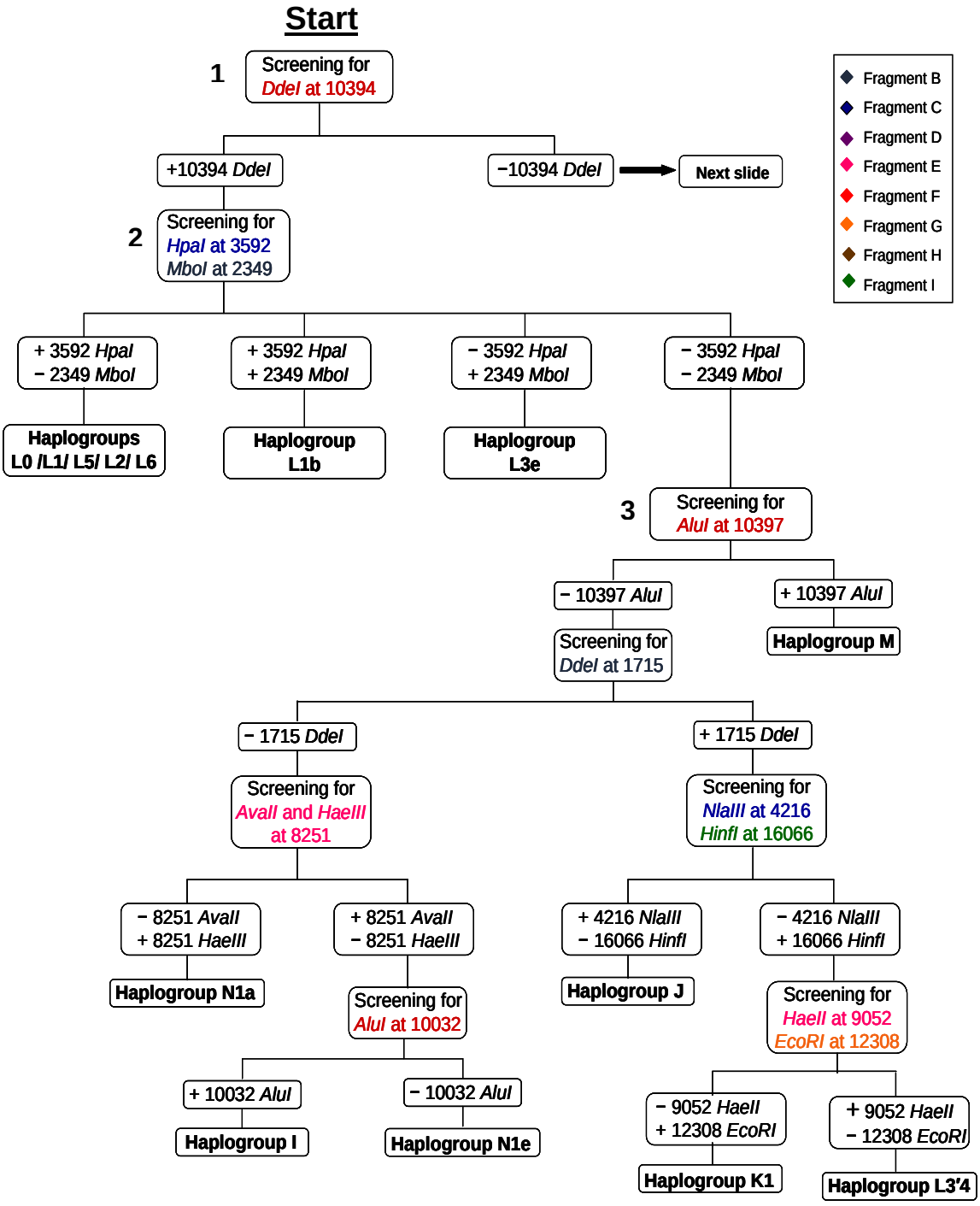
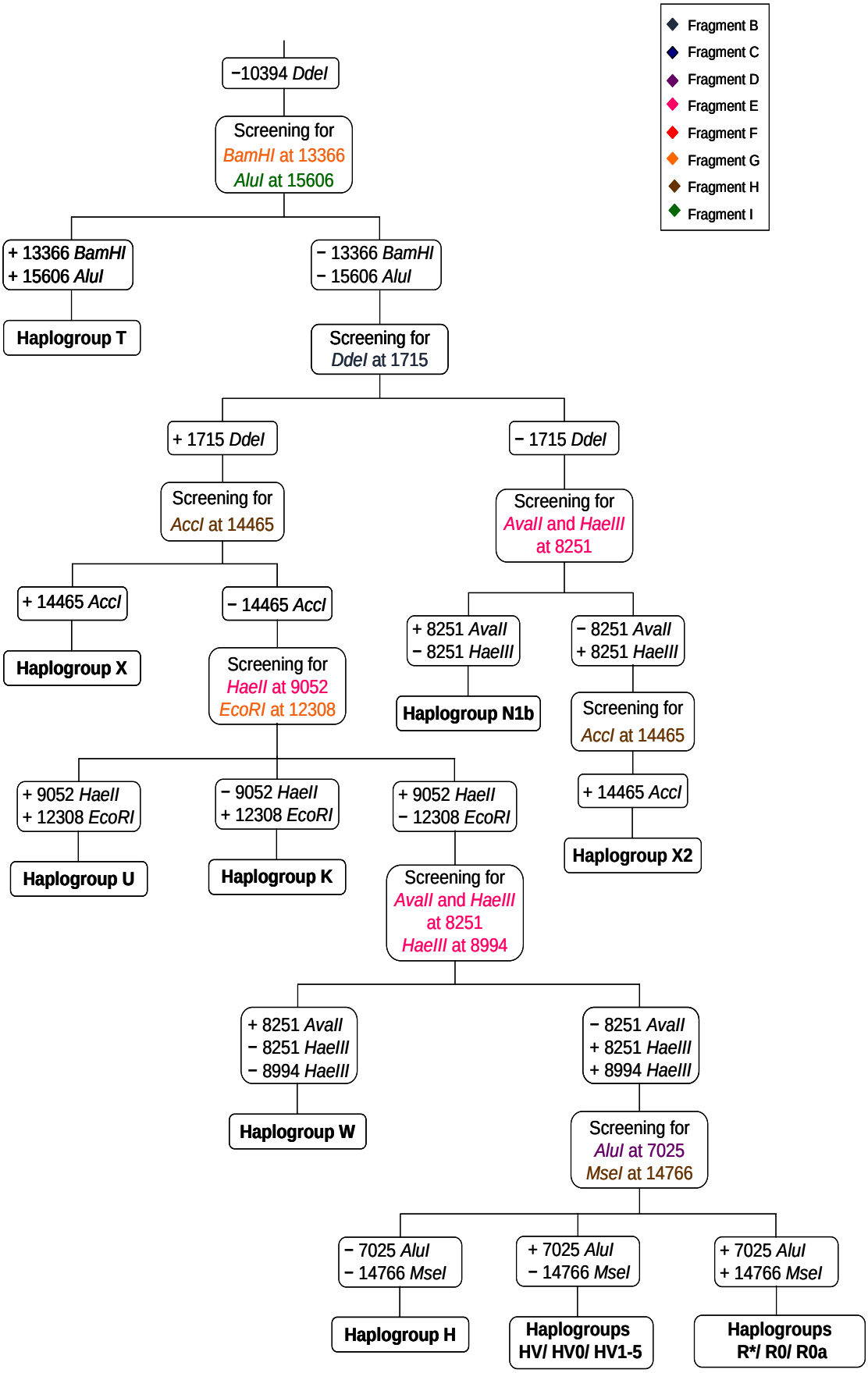



**Figure 5.28: A schematic representation of the workflow of the RFLP analysis of the whole mitochondrial genome.** The flowchart shows the sequence of the restriction digestion analyses that was followed to screen the Egyptian mtDNA samples for haplogroup specific RFLP markers aiming at fulfilment of rapid and successful haplogroup assignment.

**Figure 5.28:** Continue.



### 5.2.5 Frequency distribution of the mitochondrial DNA haplogroups in the Egyptian population

All the Egyptian individuals (n=261) were successfully assigned to well established mitochondrial DNA haplogroups using the developed strategy of RFLP analysis of the mitochondrial genome. To validate this newly developed scheme, the results obtained by RFLP analysis were compared to those obtained via mitochondrial hypervariable region sequencing. The final haplogroup identification of the Egyptian individuals was based on combined data from both analyses. The European lineages predominated and constituted around 62.5% of the Egyptian population sample, whereas the African origin was only reported in approximately 25%. A few representatives of the mitochondrial DNA Asian lineages were also detected (12.5%) (**Table 5.1**) (**Figures 5.29 and 5.30**). The genetic structure of the modern Egyptians revealed a wide variety of mtDNA lineages represented by more than 20 different mitochondrial haplogroups with varying frequencies. This large number of mitochondrial DNA lineages denotes a high level of diversity in the studied population group. By estimation of the individual haplogroups frequencies, it was found that the European haplogroup T was the most frequent mitochondrial DNA lineage in the Egyptian population occurring at a frequency of 13.8%, followed by haplogroups L3 (12.6%), H (12.3%), U (~10%) and M (8.4%) (**Table 5.1**).

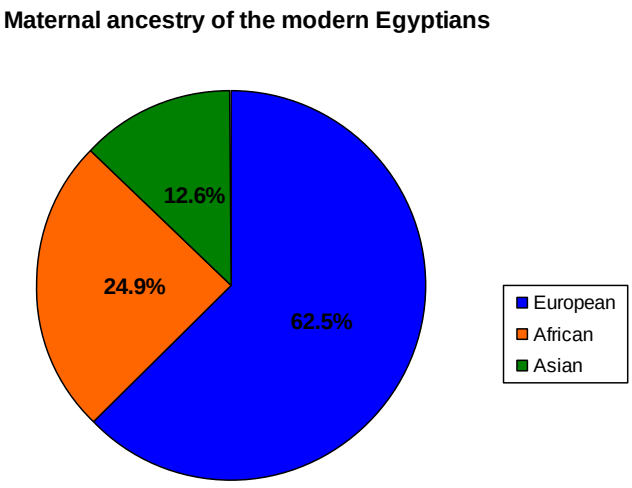



**Figure 5.29: Geographic distribution of the maternal ancestry of the modern Egyptians.** The European mitochondrial DNA lineages were found in 62.5% of the Egyptian population sample while the African origin constituted less than 25%. Few Asian representatives were also reported.

**Table 5.1: Mitochondrial DNA haplogroup frequencies in 261 Egyptian individuals and their geographic origin.** Mitochondrial haplogroups are named according to the universal agreed nomenclature system (Van Oven and Kayser 2009). n is the number of individuals designated to each haplogroup.

| Haplogroup | n | Frequency | Haplogroup | n | Frequency |
|---|---|---|---|---|---|
| **European** | | | **African** | | |
| H | 32 | 0.123 | L3 | 33 | 0.126 |
| HV | 15 | 0.057 | L4 | 1 | 0.004 |
| R0 | 3 | 0.011 | L6 | 1 | 0.004 |
| R0a | 9 | 0.034 | L2 | 20 | 0.077 |
| J | 11 | 0.042 | L1 | 5 | 0.019 |
| T | 36 | 0.138 | L0 | 5 | 0.019 |
| U | 26 | 0.1 | Total | 65 | 0.249 |
| K | 8 | 0.031 | **Asian** | | |
| R* | 13 | 0.05 | M | 22 | 0.084 |
| W | 3 | 0.011 | N1b | 8 | 0.031 |
| X | 6 | 0.023 | N1a | 2 | 0.008 |
| I | 1 | 0.004 | N1e | 1 | 0.004 |
| Total | 163 | 0.624 | Total | 33 | 0.127 |

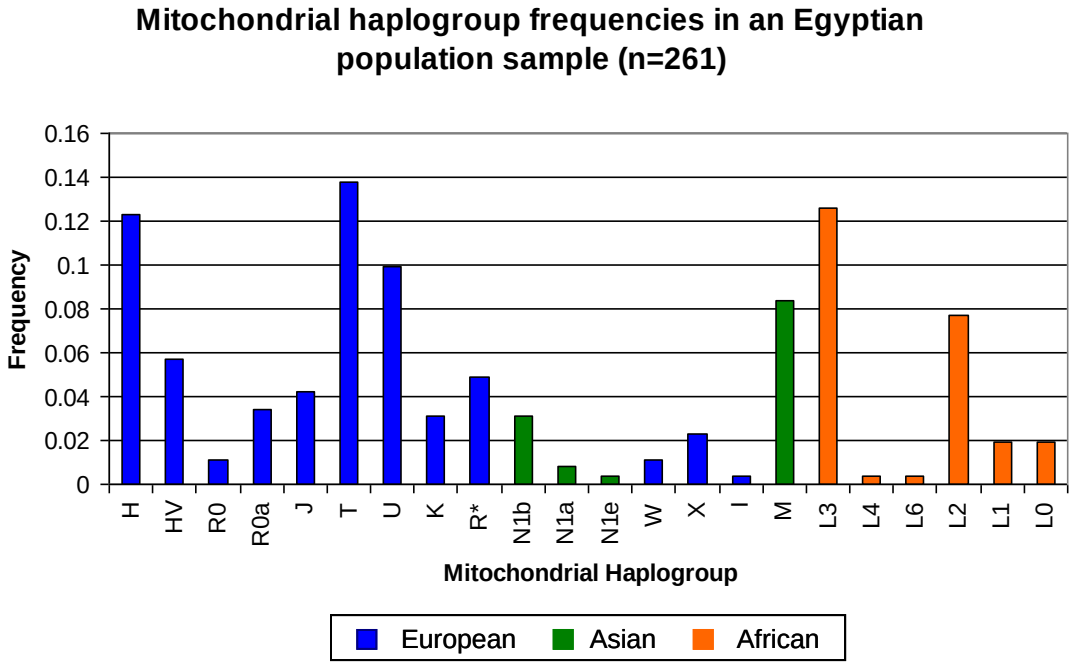



**Figure 5.30: MtDNA haplogroup frequencies reported in an Egyptian population sample (n=261).** Letters refer to mitochondrial haplogroups according the universal agreed nomenclature system (Van Oven and Kayser 2009). Identification of haplogroups was based on the hypervariable region data and RFLP markers. A diverse genetic structure was shown through detection of a large number of mitochondrial haplogroups with varying geographical origin. Continental specific haplogroups were labelled in different colours. The European mitochondrial haplogroup T was the most frequent lineage in the reported dataset followed by the haplogroups L3, H, U and M.

Data obtained from the present study was compared to haplogroup frequencies previously reported by analysis of the mitochondrial control region sequences of 277 Egyptians, where haplogroup identification was only based on hypervariable region data (Saunier et al. 2009). There was a close consistency between the two datasets regarding the different geographic contributions to the maternal ancestry of the modern Egyptian population (**Figure 5.31**). Most of the mitochondrial DNA haplogroups were detected at relatively similar frequencies in the two datasets. Different frequencies were reported for haplogroups H, R0, T, J and I (**Figure 5.32**). These frequency differences can be partly explained by relying on the hypervariable region data solely for the haplogroup assignment in the previous study, which caused a statistical bias towards the frequency of haplogroup R0 against haplogroup H whose identification can be confirmed via detection of a C7028C mutation which was not available in the previous Egyptian population study. Haplogroup I was only reported once in the present study with a frequency of 0.4%, whereas it showed a higher frequency (3.2%) in the other dataset reported by Saunier and colleagues (2009). Haplogroup T was observed at a relatively high frequency (13.8%) compared to the earlier study (9.3%) and the opposite was true for haplogroup J which was slightly less frequent in the present study (4.2%) compared to the other dataset (7.2%) (**Table 5.2**).

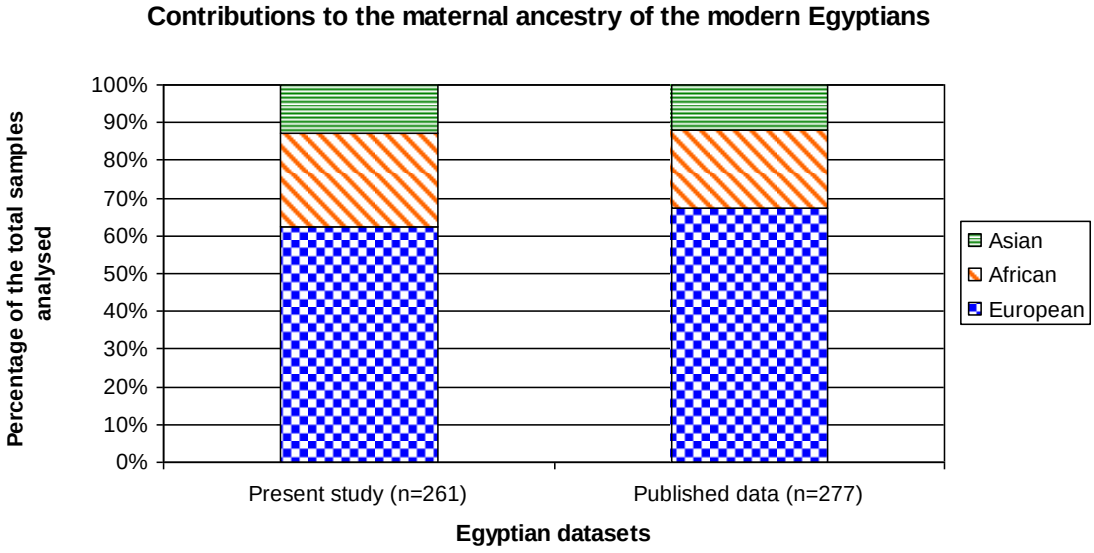



**Figure 5.31: Comparison of the relative contributions of geographic specific mtDNA lineages observed in two Egyptian datasets.** The European lineages included haplogroups H, HV, R0, R0a, J, T, U, K, X, W, I and R*. The African specific category included L0 and L1-6. The Asian origin was mainly represented by M1 and N1b haplogroups, in addition to few representatives of other N derived haplogroups; N1a and N1e.

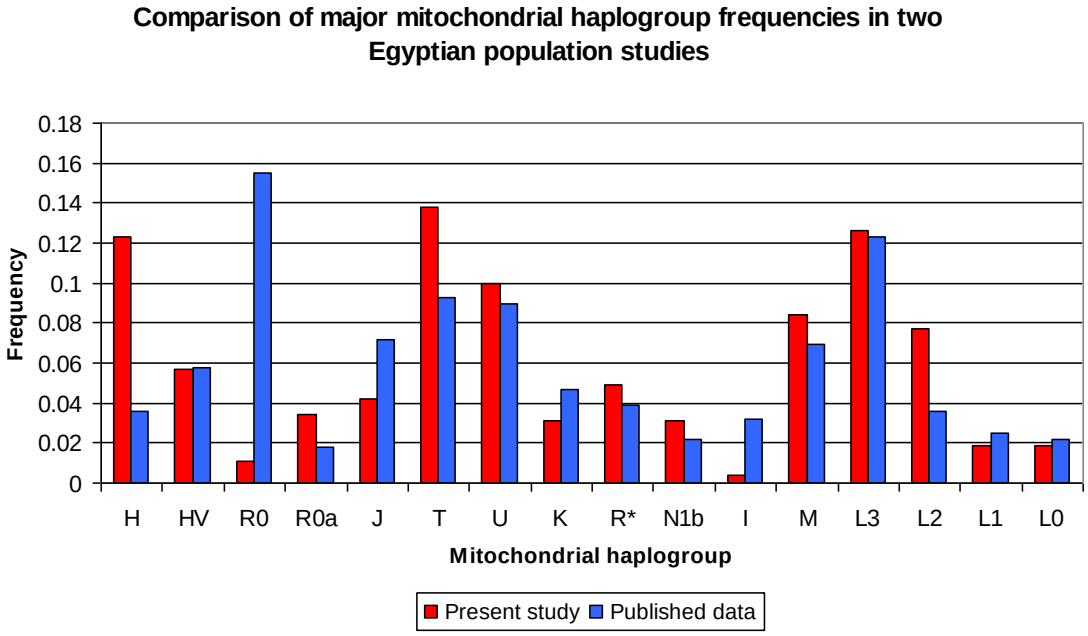



**Figure 5.32: Major haplogroup frequencies observed in the present study (n=261) compared to a previously published Egyptian dataset of similar sample size (n=277).** Letters refer to mitochondrial haplogroups according the universal agreed nomenclature system (Van Oven and Kayser 2009). Haplogroup identification in the present study was based on both hypervariable region sequencing data and coding region data via RFLP analysis while haplogroups obtained from Saunier et al.(2009) was only based on the control region sequencing data.

Data obtained from the two Egyptian population studies were combined and the overall haplogroup frequencies were estimated for the total of 538 Egyptians (**Table 5.2**). The combined dataset was predominated by the European lineages which constituted 65% of the combined Egyptian population (n=350), followed by the African lineages forming approximately 23% (n=122), and lastly showed a smaller contribution from Asian origin representing 12% (n=66). The most frequent haplogroup was L3 (12.5%) followed by haplogroups T (11.5%), U (9.5%), R0 (8.6%) and H (7.8%) (**Figure 5.33**). These combined dataset was then used for assessment of the genetic relationship between the Egyptian population and geographically related populations.

**Table 5.2: Mitochondrial DNA haplogroup frequencies in two Egyptian population studies.** Haplogroup frequencies were calculated for the two datasets individually and in combination.

| Haplogroup | Present study (n=261) | Published data* (n=277) | Combined data (n=538) |
|---|---|---|---|
| H | 0.123 (32) | 0.036 (10) | 0.078 (42) |
| HV | 0.057 (15) | 0.058 (16) | 0.058 (31) |
| R0 | 0.011 (3) | 0.155 (43) | 0.086 (46) |
| R0a | 0.034 (9) | 0.018 (5) | 0.026 (14) |
| JT | 0.000 (0) | 0.004(1) | 0.002 (1) |
| J | 0.042 (11) | 0.072 (20) | 0.058 (31) |
| T | 0.138 (36) | 0.093 (26) | 0.115 (62) |
| U | 0.100 (26) | 0.09 (25) | 0.095 (51) |
| K | 0.031 (8) | 0.047 (13) | 0.039 (21) |
| R* | 0.050 (13) | 0.039 (11) | 0.048 (26) |
| N1b | 0.031 (8) | 0.022 (6) | 0.026 (14) |
| N1a | 0.008 (2) | 0.007 (2) | 0.007 (4) |
| N1e | 0.004 (1) | 0.000 (0) | 0.002 (1) |
| N | 0.000 (0) | 0.022 (6) | 0.011 (6) |
| W | 0.011 (3) | 0.007 (2) | 0.009 (5) |
| X | 0.023 (6) | 0.014 (4) | 0.019 (10) |
| I | 0.004 (1) | 0.032 (9) | 0.019 (10) |
| M | 0.084 (22) | 0.069 (19) | 0.076 (41) |
| L3 | 0.126 (33) | 0.123 (34) | 0.125 (67) |
| L4 | 0.004 (1) | 0.000 (0) | 0.002 (1) |
| L6 | 0.004 (1) | 0.000 (0) | 0.002 (1) |
| L2 | 0.077 (20) | 0.036 (10) | 0.056 (30) |
| L1 | 0.019 (5) | 0.025 (7) | 0.022 (12) |
| L0 | 0.019 (5) | 0.022 (6) | 0.020 (11) |

* Data obtained from Saunier et al. (2009).

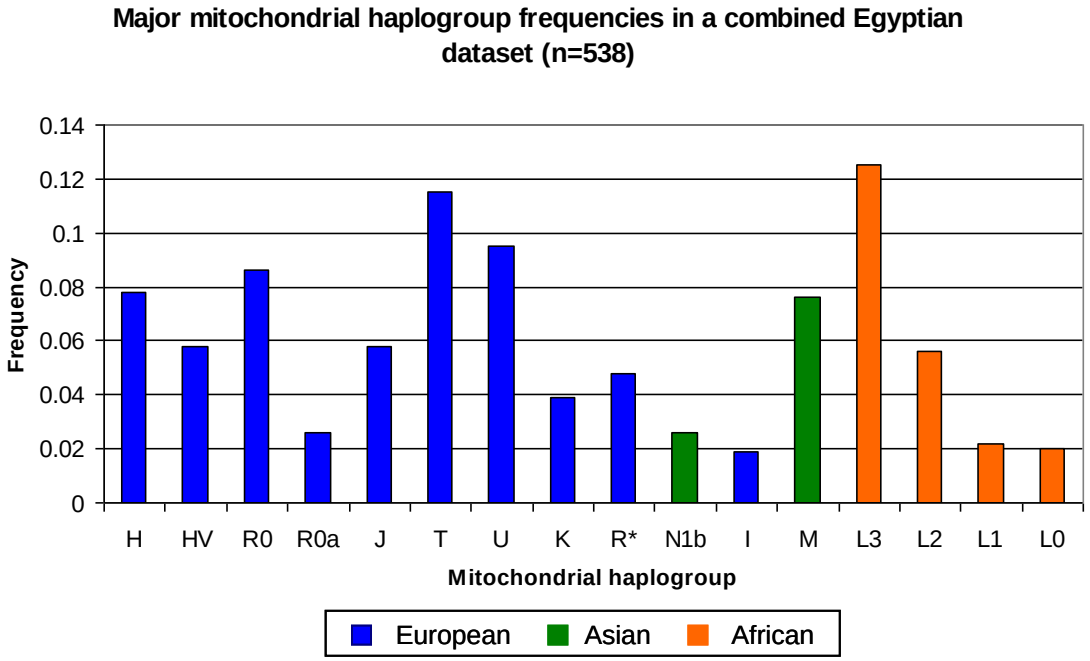



**Figure 5.33: Graphical representation of major mitochondrial haplogroup frequencies in a combined data set of 538 Egyptians.** Data obtained from the present study (n=261) was combined with data obtained by analysis of the mitochondrial control region sequences of 277 Egyptians performed by another scientific group. Letters refer to mitochondrial haplogroups according to the universal agreed nomenclature (Van Oven and Kayser 2009).

### 5.2.6 Mitochondrial genetic relationships between modern Egyptians and their neighbours

The genetic relationship between modern Egyptians and geographically related populations was assessed by comparing their mitochondrial genetic structures to the Egyptian one. Populations involved in the comparison study included East Africa (Mozambique and Ethiopia), North West Africa (Tunisia, Algeria, Morocco and Mauritania), Near East (Jordan, Syria and Palestine) and South Europe (South Portugal, Spain and Italy). Using the previously published mtDNA datasets of these populations (**Figure 5.34**) (**Table 5.3**), the relative frequencies of the continental specific major mitochondrial DNA lineages were determined and compared to those reported in the combined Egyptian dataset. It was found that the Egyptian genetic make-up was unsurprisingly very similar to that of the North West African populations (Tunis, Algeria and Morocco) with relatively similar proportions of the major mitochondrial lineages L, M, N and R except for Mauritania which did not show Asian derived lineages (**Figure 5.35**).

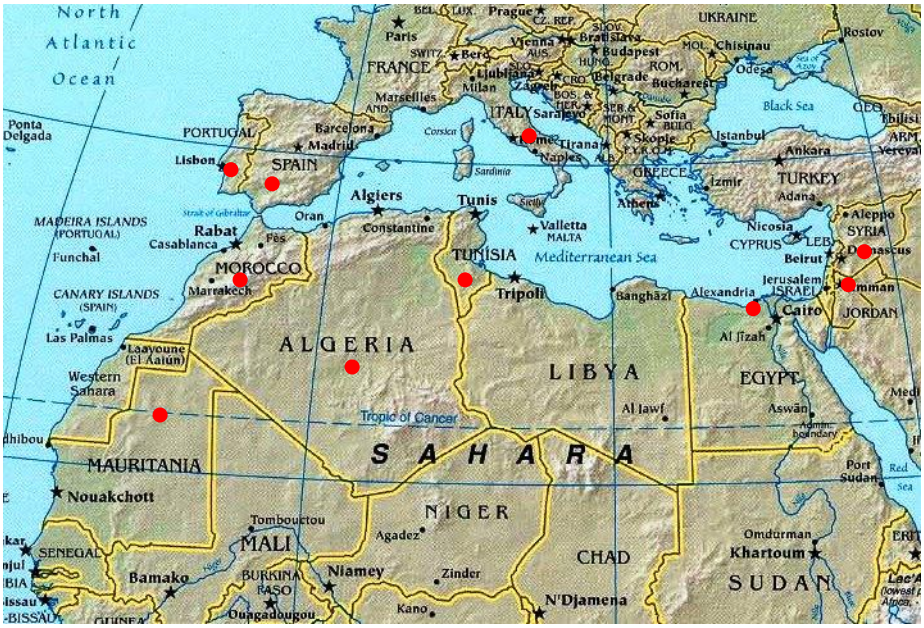



**Figure 5.34: Map of North Africa, South West Europe, and Near East showing the geographic relationship between Egypt and the populations included in the analysis.** The geographic location of the different populations involved in the analysis is indicated by red circles.

**Table 5.3: Frequencies (%) of major mitochondrial haplogroups in the Egyptian population and geographically related populations.** The (n) refers to the number of individuals involved in the study. For Egypt, Tunisia, and Morocco, frequencies were calculated based on combined datasets from two studies.

| | | Haplogroup Frequencies (%) | | | | |
|---|---|---|---|---|---|---|
| **Population** | **(n)** | **R** | **N** | **M** | **L** | **References** |
| East Africa | | | | | | |
| Mozambique | 416 | 0 | 0 | 0.2 | 99.5 | Pereira et al. (2001)<br>Salas et al. (2002) |
| Ethiopia | 270 | 15.4 | 6 | 17 | 41.2 | Kivilsid et al. (2004) |
| North West Africa | | | | | | |
| Egypt | 538 | 55.7 | 9.3 | 7.6 | 22.3 | Present study<br>Saunier et al. (2009) |
| Tunisia | 111 | 53.1 | 1.8 | 2.7 | 39.6 | Plaza et al. (2003)<br>Turchi et al. (2009) |
| Algeria | 47 | 57.3 | 2.1 | 12.8 | 27.7 | Plaza et al 2003 |
| Morroco | 106 | 61.1 | 2.8 | 2.8 | 28.3 | Plaza et al. (2003)<br>Turchi et al. (2009) |
| Mauritania | 30 | 56.5 | 0 | 0 | 43.5 | Rando et al. (1998) |
| Near East | | | | | | |
| Palestine | 118 | 74 | 8 | 2 | 14 | Gonzalez et al. (2008) |
| Jordan | 101 | 75 | 9 | 5 | 10 | Gonzalez et al. (2008) |
| Syria | 119 | 87 | 6 | 0 | 7 | Gonzalez et al. (2008) |
| South Europe | | | | | | |
| South Portugal | 59 | 88.3 | 3.4 | 0 | 3.4 | Plaza et al.(2003) |
| Spain | 312 | 84.5 | 8 | 0.3 | 2.9 | Alvarez et al. (2007) |
| Italy | 411 | 83.9 | 1.1 | 0.6 | 2.9 | Plaza et al. (2003) |

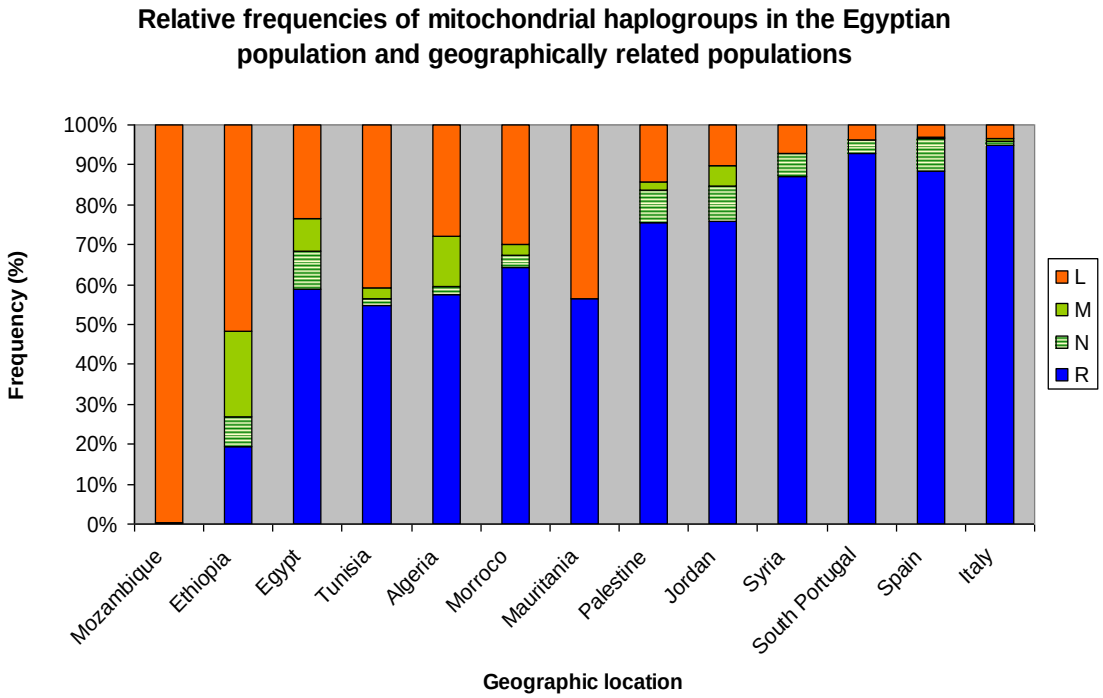



**Figure 5.35: Frequency distribution of major mitochondrial DNA haplogroups in representative populations of North West Africa, East Africa, Near East and South West Europe.** Major mitochondrial haplogroup R included all individuals belonging to haplogroups H, HV, R0, R0a, J, T, U, K, and R*. Macro haplogroup N included N1b, N1a, X, W, and I. Macro haplogroup M included M1. The African L haplogroup included members of haplogroups L0 and L1-6. The sample sizes are Mozambique (n=416), Ethiopia (n=270), Egypt (n=538), Tunisia (n=111), Algeria (n=47), Morocco (n=106), Mauritania (n=30), Palestine (n=118), Jordan (n=101), Syria (n=119), South Portugal (n=59), Spain (n=312), and Italy (n=411).

To further assess the genetic variation between the North African populations, a more detailed comparison was performed to assess the frequency distribution of the mitochondrial DNA haplogroups in four population groups: Egyptians (n=538), Tunisians (n=111), Algerians (n=47), and Moroccans (n= 106) (**Figure 5.36**). Haplogroups H, HV, and R0 occurred at similar frequencies in the Egyptians and Tunisians with values of 25% and 27%, respectively, while higher values were reported in Algerians and Moroccans, 34% and 33.7%, respectively. The highest frequency of the European mitochondrial haplogroup T (11.5%) was reported in the Egyptian population and the lowest was reported in Moroccans (3.8%). Both Moroccans and Egyptians showed high frequencies of the European haplogroup U (excluding the subclade U6) with values of 8.5% and 8.2%, respectively, whereas Algerians showed the lowest frequency (2.1%). The European Haplogroup I was only found in the Egyptians and Tunisians at frequencies of 1.9% and 0.9%, respectively. Haplogroup X was shown in all North African populations at low frequencies with the highest at 2.8%

reported in Moroccans. The European haplogroup W was not found in North African populations except in the Egyptian group where it was reported at a frequency of 0.9%. Regarding the Asian derived mitochondrial lineages, haplogroup M1 was found in the Egyptian population at a frequency of 7.6% which is higher than the values reported in Tunisians and Moroccans (2.7% and 5.7%, respectively), but still lower than the reported frequency in Algerians (12.8%). Amongst all North Africans, the Egyptian population is the only group that showed Asian haplogroup N1 which was reported at a frequency of 4.6%. This can be attributed to the geographic location of Egypt near the Levantine Corridor which was thought to be an important route of modern human migrations from and to Africa (Rowold et al. 2007; Gonzalez et al. 2008). Among African specific haplogroups, L0 was only reported in the Egyptians and Tunisians at frequencies of 2% and 0.9%, respectively. Haplogroup L1 was less represented in the Egyptian population (2.2%) compared to the other North African groups which had a low of 6.3% in Tunisians and a high of 9.4% in Moroccans. Haplogroup L2 was reported at a much lower frequency in the Egyptian population (5.6%) compared to the frequency reported in Tunisians (14.4%). Similarly, haplogroup L3 was reported in 12.5% of the Egyptian population which is lower than values reported in Tunisians and Algerians (18% and 14.9%, respectively). Overall, the African L haplogroups were less represented in the Egyptian population compared to the other North Africans.

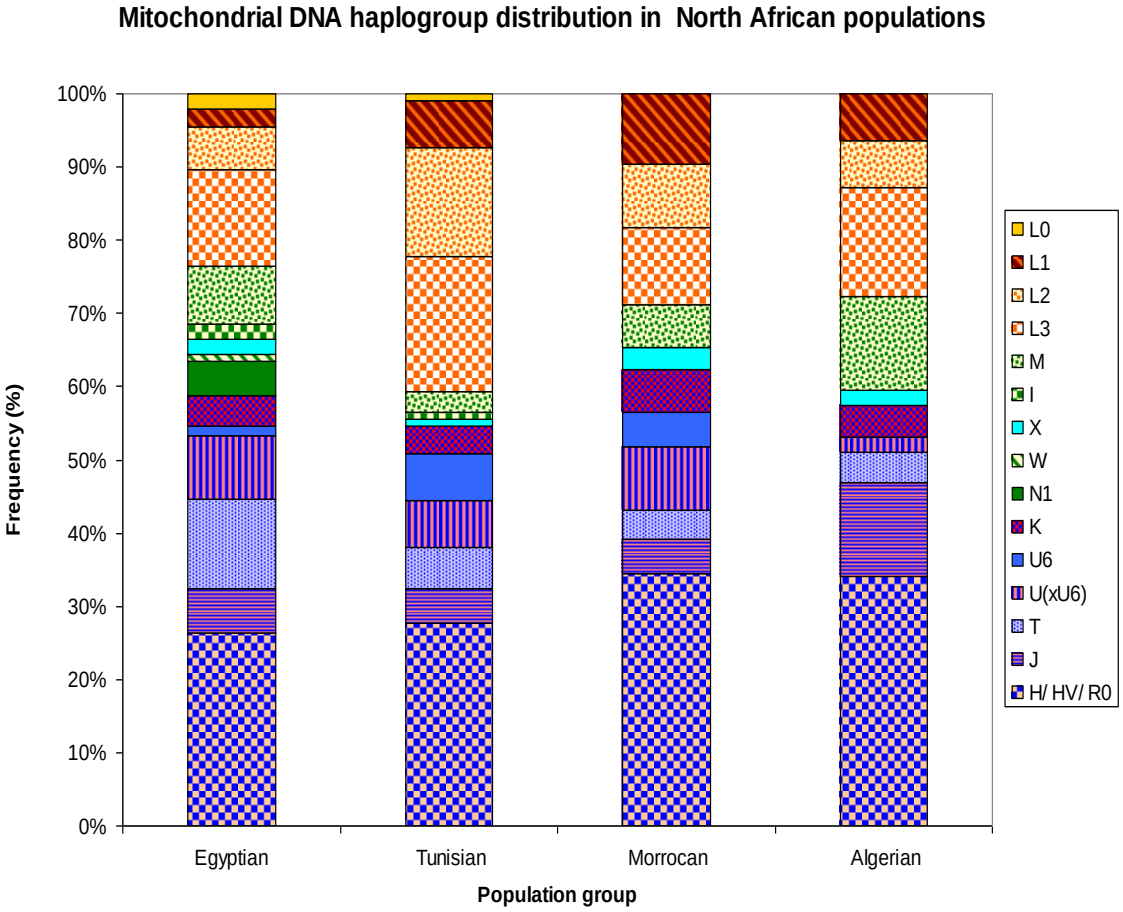



**Figure 5.36: Frequency distribution of mitochondrial DNA haplogroups in North African populations.** The frequency of each mitochondrial haplogroup was calculated as a percentage considering the total for each population group as 100%. Sample sizes are Egypt (n=538), Tunisia (n=111), Morocco (n=106), and Algeria (n=47). For simplification, mitochondrial haplogroups H, HV, R0 and R0a were included in one big group at the bottom of each bar. The category of haplogroup U included all sub-haplogroups U1-U8 except U6 which was considered separately. Haplogroup identification was mainly based on hypervariable region sequencing data for all the studied populations except for the Egyptian population group (n=261) included in the present study in which haplogroup assignment was based on combined data from hypervariable region sequences and RFLP markers.

## 5.3 Discussion

The main aim of this part of the study was to develop a new reliable strategy for screening of the mitochondrial genome for haplogroup specific variations using PCR-RFLP technology. That strategy was to be used to further confirm the haplogroup assignment of the Egyptian mtDNA types that were previously identified using hypervariable region sequencing data. The overall mitochondrial genetic diversity of the Egyptian population was to be assessed and correlated to the previously published datasets. The geographic origin(s) of the maternal ancestry of present day Egyptians was (or were) to be concluded. The genetic relationship between modern Egyptians and their neighbours was to be determined.

### 5.3.1 Development of a new strategy for analysis of mitochondrial nucleotide variations as a useful tool for haplogroup identification

Mitochondrial hypervariable region data have been used for haplogroup identification in different population studies (Alvarez et al. 2007; Egyed et al. 2007; Barbosa et al. 2008; Saunier et al. 2009), however the use of hypervariable region sequences solely can result in an ambiguous mitochondrial phylogeny (Bandelt et al. 2000; Torroni et al. 2000; Kivisild et al. 2002; Yao et al. 2002; Li et al. 2007). Therefore, combinations of hypervariable region sequencing data with coding region polymorphisms data can enhance the identification of mitochondrial DNA haplogroups with fine resolution and provide sufficient information for better understanding of the mitochondrial genetic backgrounds of different population groups.

Many of the previous population studies which involved restriction analyses of the mitochondrial genome relied on amplification of the mtDNA as several fragments in different PCR reactions which necessitate an abundance of the DNA template (Chen et al. 1995b; Macaulay et al. 1999; Tolk et al. 2001; Li et al. 2007). In the present study, a new multiplex PCR amplification protocol was developed and successfully used to amplify 11Kb of the mitochondrial genome as 9 non-overlapping fragments in one reaction, using only a 1.2-mm processed disc of the blood stained FTA cards as DNA template. Using this protocol, the required volume of DNA samples for the whole analysis was significantly reduced and this enabled preservation of samples for subsequent investigations, a character which has a great value in forensic investigations which occasionally deal with limited sample volumes.

The newly developed strategy of RFLP analysis of the mitochondrial genome was primarily designed to suit any scientific laboratory even those lacking the highly advanced sequencing technologies. Thus, it can be easily used by all researchers, particularly in the developing countries, for analysis of mtDNA in different population groups. Despite the relatively rapid advancement in the sequencing technologies, many researchers groups all over the world, who are interested in mtDNA population studies, are still restricting the use of direct sequencing to the mitochondrial hypervariable regions while they are using different SNP detection techniques for analysis of the coding region of the mitochondrial genome (Kohnemann et al. 2008; Grignani et al. 2009; Schlebusch et al. 2009).

### 5.3.2 Selection of the most informative RFLP markers

The present study included an important aspect which was selection of the most informative RFLP markers that could be effectively used for mitochondrial haplogroup identification in the Egyptian population. Several population studies have used lists of haplogroup specific RFLP markers (Chen et al. 1995b; Alves-Silva et al. 2000; Richards et al. 2000; Finnila et al. 2001b; Tolk et al. 2001; Salas et al. 2002; Richards et al. 2003; Stevanovitch et al. 2004) and others (Rowold et al. 2007; Shlush et al. 2008) have followed the basic classification scheme of Macaulay et al. (1999) which involved analysis of a very long list of RFLP markers, some of which were more informative than others. Some studies however did not include important haplogroup differentiating markers in their analysis, for example, in the study conducted to construct a phylogenetic network for the European mtDNA (Finnila et al. 2001b), identification of haplogroup W was based on two RFLP markers, +8251*AvaII* and −10394*DdeI*, which are not exclusive to haplogroup W since they also occur in its sister clades N1b, N1e, and I. A loss of *HaeIII* at position 8994 is more specific to haplogroup W as described earlier (Torroni et al. 1996). Similarly, Herrnstadt et al. (2002) reported a reticulation in the phylogenetic network among sequences belonging to haplogroups I, W, and X due to shared polymorphisms at positions 1719 and 8251, whereas they did not mention the polymorphism at postion 10032 which is specific to haplogroup I as described elsewhere (Chen et al. 1995b; Alves-Silva et al. 2000). Therefore, it was required to perform a preferential selection of the most informative RFLP markers to be included in the present analysis which would enable identification of mitochondrial haplogroups with fine resolution. Thus, a general list of 26 RFLP markers was developed and

applied to a pilot study which involved screening of 64 Egyptian individuals. Based on the findings obtained from the preliminary analysis, a smaller list of 18 RFLP markers was tailored to perform haplogroup assignment of the Egyptian individuals as accurately and as quick as possible through performing the restriction analyses in a defined order. The new developed scheme for RFLP analysis was proved to be reliable through comparison of results obtained by RFLP analysis to those obtained by hypervariable region sequencing.

All the previous mitochondrial DNA studies which involved identification of haplogroup U and its descendant K via RFLP analysis used mispairing PCR technology to identify a characteristic A12308G mutation as described earlier (Torroni et al.1996). This mutation was described to constitute a new *HinfI* cut site in conjunction with the mismatched primer, however, by reviewing the mispairing primer sequence and the Cambridge Reference Sequence of the mtDNA, it was found that the produced restriction cut site using this particular mismatched primer was an *EcoRI* cut site instead of *HinfI*. Thus, the present study reported an error in the literature which has not been addressed by any other researchers before.

### 5.3.3 Mitochondrial DNA haplogroup distribution in the Egyptian population

Using the newly developed scheme of RFLP analysis, mtDNA of 261 Egyptian individuals were confidently assigned to established mitochondrial haplogroups. The overall genetic structure of present day Egyptians seems to be greatly influenced by the admixture between past Egyptians and all the previous ruling populations, as evidenced by a heterogeneous composition, with a mixture of more than 20 mtDNA haplogroups with varying racial backgrounds. This mixture of varying mitochondrial lineages reflects a mixed maternal ancestry of the Egyptian population which can also be a result of the peculiar geographic location of Egypt next to the Levantine Corridor. This Corridor was thought to be an important route of modern human migration from and to Africa, and subsequently Egypt might have been influenced by several waves of human migration (Rowold et al. 2007; Gonzalez et al. 2008).

The European lineages constituted the main component of the genetic make-up of the Egyptian population since it was present in 62.5% of the analysed individuals, whereas, the African lineages were only present in approximately 25%. In addition, Asian derived lineages were also reported in 12.5% of the population. These figures are

comparable to the results obtained by a recent analysis of mitochondrial hypervariable region sequences of 277 Egyptians, where the European lineages constituted 67.5% and the African lineages were found in 20.5% of the population (Saunier et al. 2009). Amongst all the mitochondrial DNA lineages, the European mitochondrial haplogroup T was the most frequently reported (13.8%) in the present study. This haplogroup originated in the Near East approximately 46,500 years ago but was reported to be most prevalent in Europe with an average of 10.5% or even higher in some European populations (Richards et al. 1998; Helgason et al. 2001). Such a high frequency of haplogroup T in the Egyptian population raises a question regarding the universal agreed consideration of haplogroup T as being European. The distinction is that haplogroup T might probably be considered European due to its high frequencies in the current European populations, as opposed to its origin. The next common European haplogroup was H which represents the most frequent European haplogroup with an average of 45% or even more in some European populations. In the present study, haplogroup H accounted for 12.3% of the Egyptian lineages which is very similar to the average reported in Saudi Arabia (13%) (Achilli et al. 2004; Abu-Amero et al. 2007). The European Haplogroup U accounted for 10% of the Egyptian lineages, of which only 1.5% were members of haplogroup U6, which is an established North African lineage that was previously reported in Tunisians and Moroccans with frequencies of 6.3 % and 4.7%, respectively (Rando et al. 1998; Macaulay et al. 1999; Plaza et al. 2003; Turchi et al. 2009). The overall European haplogroups accounted for 62.5% of the Egyptian lineages which is close to the average in the Middle East where values of 74 % and 75% were reported in Palestinians and Jordanians, respectively (Gonzalez et al. 2008). Representation of the European lineages in East African populations was significantly low ranging from 15.4% in Ethiopia (Kivisild et al. 2004) down to 0% in Mozambique (Pereira et al. 2001; Salas et al. 2002). Therefore, the presence and the frequencies of European haplogroups in the Egyptian population sample clearly place Egypt distinct from sub-Saharan African populations, a finding which is consistent with another recent population study which considered Egypt as the closest African population to the Middle East (Gonzalez et al. 2008).

The N derived mitochondrial haplogroups (N1a, N1b, N1e, W, X, and I) were found in a total of 8.1% of the Egyptian haplotypes that reported in the present study and in 9.3% in the combined Egyptian dataset (n=538). These figures are very similar to the frequencies reported in the Middle East, 9% in Palestine and 8% in Jordan (Gonzalez et

al. 2008). In contrast, haplogroups N1a, N1b, N1e, and W were not detected in North West Africa; Morocco, Tunisia, Algeria and Mauritania (Plaza et al. 2003; Turchi et al. 2009). All these N derived lineages were also not found in Central and South East Africa (Kivisild et al. 2004; Gonzalez et al. 2008). Thus, the presence of these N derived lineages in the Egyptian population at such a high frequency clearly distinguishes Egypt from all the other African populations including those in the North West of Africa, and supports the presumed genetic relationship between Egypt and the Middle East. On the other aspect, presence of these haplogroups in Egypt proves that the maternal lineages originated in Asia have found a route to Africa either through the Mediterranean or the Levantine Corridor.

Haplogroup M was found at a frequency of 8.4 % of the Egyptian lineages which was slightly higher than the frequency reported by a recent Egyptian population study (6.96%) (Saunier et al. 2009). In both studies, all the reported haplogroup M lineages belonged to the sub-haplogroup M1 except one case in each study which belonged to a different M lineage. These findings are in contrast to an earlier study which reported a low frequency (3%) of M1 in Egypt and Israeli Druze (Rowold et al. 2007). The most accepted hypothesis regarding the origin of haplogroup M1 states that this haplogroup originated in Asia but soon afterwards migrated back to Africa (Passarino et al. 1998; Maca-Meyer et al. 2001; Gonzalez et al. 2007). Therefore, the presence of M1 lineages at a high frequency in the present study was not surprising, since samples were collected from Alexandria city which is located on the Mediterranean coast. These findings are in support of the earlier suggestions of the preferred usage of the Levantine Corridor over the Horn of Africa for intercontinental gene flow. However, the previously reported high frequency of haplogroup M1 in Ethiopia (17%) was behind the early assumption that M1 lineage migrated out of Asia through the Horn of Africa (Kivisild et al. 2004; Rowold et al. 2007). The low frequency of M1 in Nubia and Sudan (3%) (Gonzalez et al. 2008) and the high frequency in Egypt (8.4%) and Algeria (12.8%) go with the alternative route of migration through the Levantine Corridor or the Mediterranean as suggested elsewhere (Olivieri et al. 2006).

According to the present study, the general sub-Saharan Africa maternal gene flow to Egypt was around 25% which is in the range of other North African populations, Algeria (27.7 %) (Plaza et al. 2003) and Morocco (26.8 %) (Turchi et al. 2009), but is significantly lower than the African gene flow to Tunisia (48.4%) (Turchi et al. 2009)

and Mauritania (43.5%) (Rando et al. 1998). Dissection of the African lineages in the Egyptian population revealed that L3 was the most frequent haplogroup occurring at a frequency of 12.6 %. This is consistent with the recently reported frequency by another group (12.3%) (Saunier et al. 2009). This frequency falls in North African populations, 14.9% in Algeria and 16% in Morocco (Plaza et al. 2003) and East African populations, 16.4% in Ethiopia (Kivisild et al. 2004). But, a higher frequency of haplogroup L3 was reported in Tunisia (20.3%) suggesting a stronger African gene flow to Tunisia compared to Egypt (Turchi et al. 2009). In the present study, L3 haplogroup was mainly represented by L3f1 and L3e clades. L3f is a rare mitochondrial lineage originated in East Africa and gave rise to L3f1 subclade which has spread at an early date into West Africa (Watson et al. 1997; Salas et al. 2002). L3e is the most ancient and widespread L3 lineage, originated ~ 45,000 years ago in Central Africa and it accounts for almost one third of all L3 types in sub-Saharan Africa (Bandelt et al. 2001; Salas et al. 2002).

The next common African haplogroup was L2 with a frequency of 7.7%, all of which belonged to L2a subclade. Similar frequencies of L2 were reported in Algerians and Moroccans (6.4% and 6%, respectively) (Plaza et al. 2003) and higher frequencies were reported in Tunisians and Mauritanians (15.6% and 13.4%, respectively) (Rando et al. 1998; Turchi et al. 2009). The most accepted hypothesis is that L2a subclade diverged out of the main L2 haplogroup ~55,000 years ago in West Africa and then dramatically expanded into South East Africa. This hypothesis was supported by the high frequencies of L2a reported in Mozambique (33%) (Pereira et al. 2001; Salas et al. 2002), Sudan (24%) (Gonzalez et al. 2008) and Ethiopia (13%) (Kivisild et al. 2004).

Haplogroup L1 was only found in 1.9% of the Egyptian population, which is a very low frequency compared to North West Africa where frequencies of 10% , 10.9% and 23.4% were reported in Moroccans, Tunisians, and Mauritanians, respectively (Rando et al. 1998; Plaza et al. 2003; Turchi et al. 2009). All the reported L1 lineages in the present study were L1b types. Such a low frequency of L1 in Egypt compared to the North West African populations supports the early assumption that L1 haplogroup originated in West Africa and gave rise to subclade L1b around 30,000 years ago which then diffused into North Africa through the coastal pathway (Salas et al. 2002). Haplogroup L0 is the earliest offshoot of the most recent common ancestor (MRCA) of the mtDNA tree in Africa which appears as a sister group to the branch that gives rise to all other haplogroups. L0 was reported in the present study at a frequency of 1.9%

which mainly belonged to L0a subclade. Consistent results were reported in Ethiopians where L0a constituted the majority of L0 lineages, however the frequency of L0 was much higher (7.8%) (Kivisild et al. 2004). A much higher frequency of L0 was reported in Mozambicans (30.5%) (Pereira et al. 2001; Salas et al. 2002). L0 was not found in Algerians and Moroccans (Plaza et al. 2003). The overall high diversity found in the Egyptian population agrees with the early human mtDNA studies which demonstrated that the highest mtDNA diversity in the world was found within Africa (Cann et al. 1987; Vigilant et al. 1991; Chen et al. 1995a).

### 5.3.4 Conclusion

A new multiplex PCR protocol for amplification of 11Kb of the mitochondrial genome was developed and was used successfully to analyse a very restricted volume of DNA samples (1.2-mm disc). A new strategy was designed to facilitate PCR-RFLP analysis of the mitochondrial genome for the purpose of haplogroup assignment and was successfully applied to the Egyptian population sample (n=261). A heterogeneous origin of maternal lineages in the Egyptian population was reported via detection of a mixture of mitochondrial DNA haplogroups with varying racial backgrounds. Overall, a high diversity of the mitochondrial DNA was reported as evidenced by reporting over 20 different mitochondrial DNA haplogroups. The genetic relationship between Egypt and its surrounding populations was uncovered and it was found that present day Egyptians were the closest African population to the Middle East and Europe with an overall 62.5% European, 25% African, and 12.5% Asian mitochondrial lineages.

# 6. SCREENING FOR AMELOGENIN Y (AMELY) DELETION IN THE EGYPTIAN POPULATION

## 6.1 Introduction

Accurate gender determination is crucial in many scientific disciplines with particular interest to forensic investigations and prenatal diagnosis of X-linked recessive disorders. The amelogenin gene is the main tool in DNA-based gender tests and it represents an integral component of most PCR multiplexing kits that are routinely used for DNA profiling (for example, PowerPlex®16 Bio, PowerPlex® ES, AmpF/STR® Identifiler™, AmpF/STR® Profiler Plus™ ID, and AmpF/STR® SEfiler™). Gender determination is based on the differences between the two copies of the amelogenin gene on the X- and Y- chromosomes. Numerous length and sequence differences exist between the two homologues but the commonly used gender test is based on a 6bp deletion in intron 1 of the X-homologue. This difference can be easily detected using PCR amplification and gel electrophoresis (Kashyap et al. 2006; Cadenas et al. 2007). Despite being the most commonly used test for sex determination for many years, the last two decades have witnessed a failure of amelogenin typing in a small number of cases in certain populations leading to misidentification of males as females. These cases have raised a suspicion regarding the absolute reliability of the amelogenin test. Further investigations of the amelogenin null males revealed that amelogenin gene deletion was included in larger interstitial deletions on the short arm of the Y-chromosome ranging from several hundreds of bases up to 3.7Mb (Cadenas et al. 2007; Yong et al. 2007; Takayama et al. 2009). These deletions were mainly explained by non-allelic homologous recombination (NAHR) between the proximal major repeat array of TSPY gene and the distal copy of TSPY immediately downstream of an IR3 inverted repeat (Lattanzi et al. 2005; Jobling et al. 2007). These reported cases of amelogenin gene deletion were mainly associated with males originating from the Indian subcontinent, particularly those belonging to the haplogroup J2 Y-chromosome lineages (Yong et al. 2007). However, occurrence of amelogenin gene deletions through independent evolutionary events was suggested (Chang et al. 2007; Yong et al. 2007) and thus occurrence of amelogenin gene deletion in other ethnic groupings can not be excluded. Therefore, the aim of this part of the study was to screen 131 Egyptian males for amelogenin gene deletion on the Y-chromosome.

## 6.2 Results

The amelogenin gene was PCR amplified in 131 males from Alexandria, Egypt using the standard pair of primers that are routinely used in PowerPlex™ 16 Kit (Promega). PCR products were separated by electrophoresis on 5% polyacrylamide gel (**Figure 6.1**). The amelogenin gene copy on the X-chromosome has a 6bp deletion in the amplified region compared to its homologue on the Y-chromosome. Thus, the expected PCR products in males were two bands of sizes 106 and 112bp. If amelogenin gene was deleted on the Y-chromosome, then the PCR products would be simulating a female pattern with one band of 106bp.

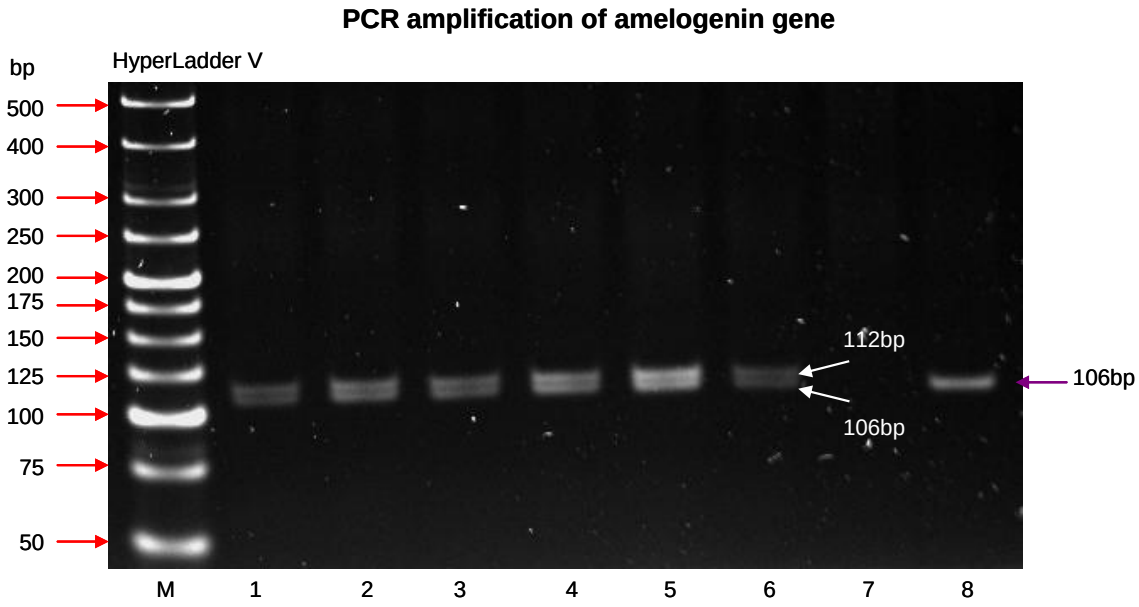



**Figure 6.1: Products of PCR amplification of amelogenin gene.** PCR products were separated by electrophoresis on 5% polyacrylamide gel. HyperLadder V was used as a molecular marker in lane M. Lanes 1-6 show PCR products of amelogenin gene amplification of 6 Egyptian males. Lane 7 contained a negative control where no DNA was added to the PCR mix. Lane 8 shows PCR products of amelogenin gene amplification of a female blood sample. The male pattern is clearly distinguished from the female one by having two bands of sizes 106 and 112bp whereas the female shows only one band of 106bp.

By the end of the study, no evidence of amelogenin gene deletion was detected in the studied population group.

## 6.3 Discussion

In the present study, no cases of amelogenin gene deletion were reported in 131 males from Northern Egypt as evidenced by PCR amplification of a specified region of the amelogenin gene on the short arm of the Y-chromosome. A previous study involved typing of 17 Y-STR loci in 208 males from Upper (Southern) Egypt, four of which (DYS393, DYS456, DYS458, DYS19) are located on the short arm of the Y-chromosome, no drop out was reported of any of these loci in the studied males, which demonstrates that at least there was no large interstitial deletion on the short arm of the Y-chromosome that might have included the amelogenin gene (Omran et al. 2008). However, the fact that the previously reported cases of amelogenin gene deletion occurred at a very low frequency, together with the small sample size included in the present study did not lead to an absolute exclusion of the existence of amelogenin gene deletion in the Egyptian population. Despite the fact that the majority of amelogenin gene deletion cases were reported in people originating from the Indian subcontinent (Yong et al. 2007), the very few cases of AMELY deletion reported in Spain (Bosch et al. 2002), Italy (Lattanzi et al. 2005), Austria (Steinlechner et al. 2002), and Israel (Michael and Brauner 2004), together with the suggested link between the North Mediterranean and the South Mediterranean coasts raises a suspicion of the possibility of existence of AMELY deletion in North Africa. Thus, a larger sample size of Egyptian males will be required to give a better overview of this issue.

## 6.4 Conclusion

To conclude, no amelogenin gene deletion was reported in 131 Egyptian males from Alexandria city as evidenced by PCR amplification and gel electrophoresis. The obtained results did not rule out the possibility of occurrence of amelogenin gene deletion in the Egyptian population and a larger sample size will be required for better estimates.

# 7. GENERAL DISCUSSION AND FUTURE WORK

The main aims of the present study were to assess the mitochondrial nucleotide diversity and heterogeneity in the Egyptian population, and to establish a database of mitochondrial hypervariable region sequences in 261 adult Egyptians from Alexandria city, which would further enrich the global human mitochondrial DNA database. The present study also aimed to assign Egyptian mitochondrial DNA sequences to well known mitochondrial haplogroups, aided by PCR-RFLP analysis of the mtDNA, for which a novel strategy was to be developed. Other aims of the study included evaluation of the currently available techniques, namely FTA cards, for mitochondrial DNA testing and assessment of their efficiency for collection, storage and transportation of DNA-containing samples which would further aid the development of forensic investigations. In addition, the present study aimed to screen 131 Egyptian males for amelogenin gene deletion on the Y-chromosome which was previously reported in certain populations.

## 7.1 The utility of FTA cards in forensic investigations

There are several methods for biological sample collection and DNA extraction, and each method has its own advantages and disadvantages. Of these, FTA card technology was extensively investigated during the present study. FTA cards were successfully used for human blood sample collection, storage and transportation. Several guidelines for the DNA purification using FTA cards technology were addressed. A 1.2-mm disc of blood stained FTA cards was shown to contain a sufficient quantity of DNA for PCR analysis, which means that a very limited sample volume is required, and thus the rest of the DNA sample can be saved for further investigations. Persistent PCR inhibition produced by unwashed inhibitor(s) in blood was eliminated through addition of Bovine Serum Albumin as a PCR enhancer. It was possible to amplify mitochondrial DNA from blood samples after being stored on FTA cards for up to 48 months. In the present study which involved analysis of several hundred blood samples on FTA cards, no cross contamination was detected in the negative control discs. Therefore, it was concluded that FTA card technology was highly advantageous over all the other methods of DNA extraction due to its special ability to amplify DNA from a very limited sample volume, its potential capability to preserve the integrity of DNA for a long period of time at room temperature, and fewer sample manipulation requirements with minimal or even no cross contamination risks. All these aforementioned features justify and extend the

utility of FTA cards to the routine forensic case work and genetic identification studies where high throughput analysis is required.

## 7.2 Rapid PCR

A new laboratory protocol for rapid PCR amplification of the hypervariable regions of mtDNA was developed. This protocol was optimised to enable the amplification reaction to be completed in a very short period of time using Gotaq DNA polymerase (Promega) without adversely affecting the quality of the PCR product. This protocol involved the use of newly designed long oligonucleotides (~30mer) which enabled combination of annealing and extension steps at 72$^{o}$C, and thus a shift from the standard three-step PCR protocol to a two-step one. Reduction of the duration of each stage of the thermal cyclic conditions together with reduced number of cycles enabled a total reduction of the PCR run time from 2-3 hours of the conventional PCR protocol down to 30 minutes. This fast PCR protocol will have a direct impact on forensic investigations, forensic science undergraduate laboratories and population genetic studies which are usually dealing with a huge number of samples that usually take a long time to be processed, particularly when automation facilities are not available.

## 7.3 Mitochondrial hypervariable region analysis and diversity measures in the Egyptian population

Analysis of mitochondrial DNA has gained increasing popularity in the last two decades for its use in human identification purposes and population genetics. Mitochondrial nucleotide variations are currently widely used for tracing maternal lineages and determination of populations' diversities. The mtDNA database is growing rapidly since the first complete human mitochondrial genome sequence was reviewed in 1999 (Andrews et al. 1999). However, mtDNA sequences of African origin are still under represented in the current database compared to those of European or Asian origin. When the present study was launched, there was little data about mtDNA variation in the Egyptian population. Therefore, analysis of the mitochondrial hypervariable region in 261 adult Egyptians was performed. A database of 261 mitochondrial HVI sequences, between nucleotides 16024-16365 and 78 HVII sequences between the nucleotides 73-340 was established. A total of 113 polymorphic sites were reported in the specified HVI region with varying frequencies. Those sites that showed the higher rate of variations agree with the mutational hotspots that were previously described in

the human evolution studies (Finnila et al. 2001b; Maca-Meyer et al. 2001). A total of 187 different HVI haplotypes were identified with a haplotype percentage K 72% which is higher than the previously reported figures in Asian populations, 63% in Japan (Asari et al. 2007) and 64% in Singapore (Wong et al. 2007) and European populations, 59% in the North of Portugal (Lima et al. 2006) and 44% in Finland (Hedman et al. 2007). Analysis of 78 HVII sequences revealed 62 different haplotypes that were identified by 42 polymorphic sites. Combination of HVII sequences with HVI sequences resulted in a total of 207 Egyptian mitochondrial haplotypes with a haplotype percentage K 79%. Of these mitochondrial types, 183 (~88%) were unique to single individuals, whereas the remaining 24 haplotypes were shared by multiple individuals. Such a large number of unique mitochondrial haplotypes indicates high mtDNA diversity and supports the use of the mitochondrial hypervariable region analysis for identification purposes. Of the shared mitochondrial haplotypes, fifteen were observed twice in the Egyptian population sample and 4 types were shared by three individuals. The remaining 5 haplotypes were shared by more than three individuals. The most commonly observed mitochondrial haplotype in the Egyptian population was identical to the CRS in the specified HVI region (nt16024-16365). It is worth noting that for all the shared mitochondrial haplotypes, HVII sequence was not available and this match was only based on the 342bp of the HVI region. Among the unique mitochondrial haplotypes, 23 haplotypes primarily shared their HVI sequences with one or more individuals, nonetheless, analysis of HVII sequences of these individuals discriminated between them. Hence, it was concluded that combination of HVI and HVII sequences increased the power of discrimination and showed an overall high level of mitochondrial diversity in the Egyptian population. These findings together with the characteristic features of the mtDNA regarding the high copy number per somatic cell and the maternal mode of inheritance support the use of mitochondrial hypervariable region analysis in the human identification purposes.

The probability of randomly selecting two unrelated Egyptians with the same mitochondrial HVI or HVII haplotypes in the analysed population sample was estimated to be 2.34% and 3.22% respectively. These Random Match Probabilities were calculated based on HVI (n=261) and HVII (n=78) sequences individually. Whereas, when HVI and HVII data were combined together, the RMP was reduced to 1.28%, which again confirmed the value having both HVI and HVII sequences in order to achieve a high power of discrimination. The RMP reported herein is in the average of

many other populations, 1.29% in the North of Portugal (Lima et al. 2006) and 1% in Romania and Japan (Egyed et al. 2007; Mabuchi et al. 2007), putting in consideration that HVII sequences were only determined in 78 individuals and all the diversity measures were mainly influenced by HVI sequences solely. Compared to other North African populations, the estimated RMP in the present study is significantly lower than the those reported in 64 Tunisians (2.1%) and 56 Moroccans (2.1%) (Turchi et al. 2009). The genetic diversity *h* was estimated as 0.9804 and 0.9803 for HVI and HVII respectively. These values are comparable to the diversities reported in other population studies which considered HVI and HVII data individually (Asari et al. 2007; Mabuchi et al. 2007). In the subset of 78 Egyptians for which both HVI and HVII sequences were analysed, the genetic diversity was estimated as 1.000 which is comparable to the diversity recently reported in 277 Egyptians based on analysis of the whole control region (Saunier et al. 2009).

## 7.4 Multiplex PCR amplification of 11kb of the human mitochondrial genome in one PCR reaction

A new laboratory protocol for PCR amplification of nine fragments of the human mitochondrial genome was developed. This protocol was mainly designed to enable amplification of 11Kb of the mitochondrial genome from a limited DNA sample volume to be used for RFLP screening of the mtDNA. This newly developed protocol involved the use of one processed 1.2-mm blood stained FTA disc as DNA template. The nine targeted regions of the mitochondrial genome varied in length depending on the nucleotide positions of the selected mitochondrial haplogroup specific polymorphisms. The amplification was performed in two stages: the first stage involved multiplex amplification of the nine targeted regions using nine pairs of oligonucleotides which were newly designed in the present study, and the second stage involved amplification of the nine regions individually using the product of the first stage PCR as DNA template. This newly developed protocol was mainly designed to save DNA samples for further analysis which is particularly important in forensic investigations which not uncommonly deal with restricted sample volumes. This protocol was validated and proved to be reliable via its application to an Egyptian population sample of 261 individuals.

## 7.5 A new scheme for mitochondrial haplogroup assignment via PCR-RFLP analysis

There are numerous mitochondrial DNA polymorphisms that were previously described and have been used for haplogroup identification (Macaulay et al. 1999; Finnila et al. 2001b), however, some of these nucleotide variations are less specific than others. In addition, the global human mitochondrial DNA phylogenetic tree is continuously updated with better resolution classification of the mitochondrial lineages (Van Oven and Kayser 2009). Therefore, the list of haplogroup specific markers had to be reviewed accordingly. A general list of 26 RFLP markers was developed for the preliminary analysis which involved screening of 64 Egyptian individuals. Based on the results obtained by the preliminary work, a more specific list of 18 RFLP markers was developed to be used for mitochondrial haplogroup assignment in the Egyptian population. These markers were screened for in all the remaining Egyptian individuals separately using thirteen different restriction enzymes in eighteen digestion reactions. These digestion reactions were performed in certain order to enable haplogroup assignment as accurate and as quick as possible. This newly developed scheme for RFLP analysis was proved to be reliable through comparison of the given results to those obtained via hypervariable region sequencing.

## 7.6 Mitochondrial DNA haplogroup distribution in the Egyptian population

Using the newly developed scheme of PCR-RFLP analysis of the mitochondrial genome together with the hypervariable region sequencing data, mtDNA was confidently assigned to well known mitochondrial haplogroups in all the Egyptian population sample (n=261). The overall genetic structure of present day Egyptians was found to be a mixture of different maternal lineages with varying racial backgrounds. The major component of the Egyptian mtDNA dataset was European and constituted around 62.5% of the whole population. These European lineages were mainly represented by haplogroups T, H, U, and HV. The African origin was found in 25% leaving the remaining 12.5% for Asian representatives. The African lineages were mainly identified as haplogroups L3 and L2, and the Asian derived lineages were mainly M1 and N1b. The most frequently reported mitochondrial lineage was the European haplogroup T which occurred at a frequency of 13.8%. This frequency is significantly higher than any other previously reported figures in North West or East Africa and even higher than the

reported frequencies in many European populations where the average was around 10.5%. This striking finding doubts the early hypothesis which considered haplogroup T as European in origin because such a very high frequency in the Egyptian population can not be easily explained by the admixture between populations. The next common haplogroups were L3, H, U, and M. The African haplogroup L3 was found in 12.6% of the Egyptian population and was mainly represented by L3f1 and L3e clades. These findings are highly consistent with the results obtained via high-resolution analysis of sub-Saharan African mitochondrial DNA lineages. Distribution of which was highly affected by the trans-Saharan slave trade (Harich et al. 2010). The European haplogroup H was found in 12.3% which is very similar to the frequency reported in Saudi Arabia (13%) (Achilli et al. 2004; Abu-Amero et al. 2007) but significantly lower than the frequencies reported in North West Africa (Plaza et al. 2003; Turchi et al. 2009). This remarkable differential distribution of the European haplogroups T and H in Egypt and North West Africa demonstrates that those two lineages found their ways to Africa through different routes of human migration. This finding supports the aforementioned doubts regarding the origin of haplogroup T. The European Haplogroup U accounted for 10% of the Egyptian lineages, of which only 1.5% belonged to haplogroup U6 which is an established North African lineage. This frequency of U6 is significantly lower than the frequencies reported in North West Africa, 6.3 % in Tunisians and 4.7% in Moroccans (Plaza et al. 2003; Turchi et al. 2009). Mitochondrial macro-haplogroup M was mainly represented by M1 lineage which accounted for 8.4% of the Egyptian population. This relatively high frequency of haplogroup M1 disagree with a previous study which reported a low frequency of 3% in Egypt and Israeli Druze (Rowold et al. 2007). The Asian derived macro-haplogroup N was found in a total of 8.1% and was represented by haplogroups N1a, N1b, N1e, W, X, and I. Similar frequencies were reported in the Middle East, 9% in Palestine and 8% in Jordan (Gonzalez et al. 2008), whereas, most of these N derived lineages were not found in North West Africa. Thus, a close genetic relationship between Egypt and the Middle East was not excluded. Overall, the genetic structure of present day Egyptians was found to be heterogeneous and consisted of a mixture of over twenty different mitochondrial lineages with varying racial backgrounds. This mixed maternal ancestry could be the result of admixture between past Egyptians and the past ruling populations and/or might have been influenced by several waves of modern human migration facilitated by the peculiar geographic location of Egypt next to the Levantine Corridor. The frequency distribution

of mitochondrial haplogroups showed that Egypt is clearly distinguished from all the other African populations including North West Africa. It was concluded that the general sub-Saharan maternal gene flow to Egypt was minimal and present day Egyptians are genetically related to the Middle East.

## 7.7 Future work on mitochondrial DNA

The current study has developed an easy and economic method for the analysis of mitochondrial DNA without the use of highly advanced technology. This method can be applied in mtDNA studies involving other populations in North Africa, which is an important geographic area currently lacking the attention of researchers, since not much data was published in the last 6 months. Hence, this study helped to enrich the global mitochondrial DNA database, not only by adding information about the Egyptian population but also by providing a suitable, easy and economic method to be further used for the analysis of mtDNA of other North African populations.

Analysis of the mitochondrial HVII region in those Egyptian individuals for whom the HVII sequences were not determined in the present study will increase the diversity estimates and reduce the number of shared mitochondrial haplotypes and thus will demonstrate a higher power of discrimination of the mitochondrial hypervariable regions. In addition, the mtDNA database is lacking sufficient information about the genetic structure and the haplogroup distribution in the Libyan population which will be the future point of interest, due to the long lasting strong relationship between the Egyptians and the Libyans through the geographic neighbourhood.

## 7.8 Screening of amelogenin Y deletion in the Egyptian population

Amelogenin gene is the routinely used locus for gender determination in forensic investigations and it is an integral part of the most common multiplexing PCR kits used for DNA profiling. However, the lately reported few cases of amelogenin gene deletion on the short arm of the Y-chromosome in certain populations (Chang et al. 2007; Yong et al. 2007; Kumagai et al. 2008) raised a doubt regarding the absolute validity of using amelogenin gene as the only marker for sex determination, particularly in those populations. Therefore, amelogenin gene was typed in 131 Egyptian males from Alexandria city using PCR amplification and gel electrophoresis and no deletion was reported in the studied individuals. However, a larger sample size will be required to

provide a better conclusion regarding this issue as the previously reported cases in other populations occurred at very low frequencies.

# REFERENCES


Abu-Amero K, Gonzalez A, Larruga J, Bosley T and Cabrera V (2007). Eurasian and African mitochondrial DNA influences in the Saudi Arabian population. BMC Evol Biol **7**(1):32.

Achilli A, Perego U, Bravi C, Coble M, Kong Q, Woodward S, Salas A, Torroni A and Bandelt H (2008). The phylogeny of the pour Pan-American mtDNA haplogroups: implications for evolutionary and disease studies. PLoS One **3**(3):e1764.

Achilli A, Rengo C, Battaglia V, Pala M, Olivieri A, Fornarino S, Magri C, Scozzari R, Babudri N, Santachiara-Benerecetti A, Bandelt H, Semino O and Torroni A (2005). Saami and Berbers-an unexpected mitochondrial DNA link. Am J Hum Genet **76**:883-6.

Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, Scozzari R, Cruciani F, Zeviani M, Briem E, Carelli V, Moral P, Dugoujon J, Roostalu U, Loogväli E, Kivisild T, Bandelt H, Richards M, Villems R, Santachiara-Benerecetti A, Semino O and Torroni A (2004). The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian Glacial refuge was a major source for the European gene pool. The American Journal of Human Genetics **75**(5):910-8.

Adams E, Malele I, Msangi A and Gibson W (2006). Trypanosome identification in wild tsetse populations in Tanzania using generic primers to amplify the ribosomal RNA ITS-1 region. Acta Tropica **100**(1-2):103-9.

Akane A, Matsubara K, Nakamura H, Takahashi S and Kimura K (1994). Identification of the heme compound copurified with deoxyribonucleic acid (DNA) from bloodstains, a major inhibitor of polymerase chain reaction (PCR) amplification. J Forensic Sci **39**(2):362-72.

Al-Soud W and Radstrom P (2001). Purification and characterization of PCR-inhibitory components in blood cells. J Clin Microbiol **39**(2):485-93.

Allard M, Miller K, Wilson M, Monson K and Budowle B (2002). Charaterisation of the Caucasian haplogroups present in the SWGDAM forensic mtDNA dataset for 1771 human control region sequences. J Forensic Sci **47**(6):1215-23.

Allard M, Polanskey D, Miller K, Wilson M, Monson K and Budowle B (2005). Characterization of human control region sequences of the African American SWGDAM forensic mtDNA data set. Forensic Science International **148**(2-3):169-79.

Allard M, Wilson M, Monson K and Budowle B (2004). Control region sequences for East Asian individuals in the Scientific Working Group on DNA Analysis Methods forensic mtDNA data set. Legal Medicine **6**(1):11-24.

Alvarez-Iglesias V, Mosquera-Miguel A, Cerezo M, Quintans B, Zarrabeitia M, Cusco I, Lareu M, Garcia O, Perez-Jurado L, Carracedo A and Salas A (2009). New population and phylogenetic features of the internal variation within mitochondrial DNA macro-haplogroup R0. PLoS One **4**(4):e5112.

Alvarez JC, Diana LEJ, Jose AL, Esther M-e, Luis JM-G, Marc A, Mark RW and Bruce B (2007). Characterization of human control region sequences for Spanish individuals in a forensic mtDNA data set. Legal medicine (Tokyo, Japan) **9**(6):293-304.

Alves-Silva J, da Silva Santos M, Guimaraes P, Ferreira A, Bandelt H, Pena S and Prado V (2000). The ancestry of Brazilian mtDNA lineages. Am J Hum Genet **67**(2):444-61.

Andelinovic S, Sutlovic D, Erceg Ivkosic I, Skaro V, Ivkosic A, Paic F, Rezic B, Definis-Gojanovic M and Primorac D (2005). Twelve-year experience in identification of skeletal remains from mass graves. Croat Med J **46**(4):530-9.

Anderson S, Bankier A, Barrell B, de Bruijn M, Coulson A, Drouin J, Eperon I, Nierlich D, Roe B, Sanger F, Schreier P, Smith A, Staden R and Young I (1981). Sequence and organization of the human mitochondrial genome. Nature **290**(5806):457-65.

Andreasson H, Nilsson M, Budowle B, Lundberg H and Allen M (2006). Nuclear and mitochondrial DNA quantification of various forensic materials. Forensic Sci Int **164**(1):56-64.

Andrews R, Kubacka I, Chinnery P, Lightowlers R, Turnbull D and Howell N (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet **23**(2):147.

Anjos M, Carvalho M, Andrade L, Lopes V, Serra A, Batista L, Oliveira C, Tavares C, Balsa F, Corte-Real F, Vieira D and Vide M (2004). Individual genetic identification of biological samples: a case of an aircraft accident. Forensic Sci Int **146** S115-7.

Ankel-Simons F and Cummins J (1996). Misconceptions about mitochondria and mammalian fertilization: implications for theories on human evolution. Proc Natl Acad Sci U S A **93**(24):13859-63.

Anslinger K, Weichhold G, Keil W, Bayer B and Eisenmenger W (2001). Identification of the skeletal remains of Martin Bormann by mtDNA analysis. Int J Legal Med **114**(3):194-6.

Asari M, Azumi J, Shimizu K and Shiono H (2008). Differences in tissue distribution of HV2 length heteroplasmy in mitochondrial DNA between mothers and children. Forensic Sci Int **175**(2-3):155-9.

Asari M, Umetsu K, Adachi N, Azumi J, Shimizu K and Shiono H (2007). Utility of haplogroup determination for forensic mtDNA analysis in the Japanese population. Leg Med (Tokyo) **9**(5):237-40.

Bailliet G, Rothhammer F, Carnese F, Bravi C and Bianchi N (1994). Founder mitochondrial haplotypes in Amerindian populations. Am J Hum Genet **55**(1):27-33.

Bandelt H, Alves-Silva J, Guimaraes P, Santos M, Brehm A, Pereira L, Coppa A, Larruga J, Rengo C, Scozzari R, Torroni A, Prata M, Amorim A, Prado V and Pena S (2001). Phylogeography of the human mitochondrial haplogroup L3e: a snapshot of African prehistory and Atlantic slave trade. Ann Hum Genet **65**(Pt6):549-63.

Bandelt H, Forster P, Sykes B and Richards M (1995). Mitochondrial portraits of human populations using median networks. Genetics **141**(2):743-53.

Bandelt H, Kong Q, Richards M and Macaulay V (2006). Estimation of mutation rates and coalescence times: some caveats. Human Mitochondrial DNA and the Evolution of Homo sapiens. Heidelberg, Springer-Verlag. 18**:** 47-89.

Bandelt H, Macaulay V and Richards M (2000). Median networks: speedy construction and greedy reduction, one simulation, and two case studies from human mtDNA. Mol Phylogenet Evol **16**(1):8-28.

Banerjee J, Trivedi R and Kashyap V (2005). Mitochondrial DNA control region sequence polymorphism in four indigenous tribes of Chotanagpur plateau, India. Forensic Sci Int **149**(2-3):271-4.

Barbosa A, da Silva L, Azevedo D, Balbino V and Mauricio-da-Silva L (2008). Mitochondrial DNA control region polymorphism in the population of Alagoas State, North-Eastern Brazil. J Forensic Sci **53**(1):142-6.

Becker S, Franco J, Simarro P, Stich A, Abel P and Steverding D (2004). Real-time PCR for detection of Trypanosoma brucei in human blood samples. Diagn Microbiol Infect Dis **50**(3):193-9.

Beckett S, Laughton S, Pozza L, McCowage G, Marshall G, Cohn R, Milne E and Ashton L (2008). Buccal swabs and treated cards: methodological considerations for molecular epidemiologic studies examining pediatric populations. Am J Epidemiol **167**(10):1260-7.

Behar D, Metspalu E, Kivisild T, Achilli A, Hadid Y, Tzur S, Pereira L, Amorim A, Quintana-Murci L, Majamaa K, Herrnstadt C, Howell N, Balanovsky O, Kutuev I, Pshenichnov A, Gurwitz D, Bonne-Tamir B, Torroni A, Villems R and Skorecki K (2006). The matrilineal ancestry of Ashkenazi Jewry: portrait of a recent founder event. Am J Hum Genet **78**(3):487-97.

Behar D, Metspalu E, Kivisild T, Rosset S and Tzur S (2008a). Counting the founders: the matrilineal genetic ancestry of the Jewish Diaspora. PLoS One **3**(4):e2062.

Behar D, Villems R, Soodyall H, Blue-Smith J, Pereira L, Metspalu E, Scozzari R, Makkan H, Tzur S, Comas D, Bertranpetit J, Quintana-Murci L, Tyler-Smith C, Wells R and Rosset S (2008b). The dawn of human matrilineal diversity. Am J Hum Genet **82**(5):1130-40.

Bender K, Schneider P and Rittner C (2000). Application of mtDNA sequence analysis in forensic casework for the identification of human remains. Forensic Sci Int **113**(1-3):103-7.

Bilal E, Rabadan R, Alexe G, Fuku N, Ueno H, Nishigaki Y, Fujita Y, Ito M, Arai Y, Hirose N, Ruckenstein A, Bhanot G and Tanaka M (2008). Mitochondrial DNA haplogroup D4a is a marker for extreme longevity in Japan. PLoS ONE **3**(6):e2421.

Bogdanowicz W, Allen M, Branicki W, Lembring M, Gajewska M and Kupiec T (2009). Genetic identification of putative remains of the famous astronomer Nicolaus Copernicus. Proc Natl Acad Sci U S A **106**(30):12279-82.

Bogenhagen D (1999). Repair of mtDNA in vertebrates. Am J Hum Genet **64**(5):1276-81.

Borman A, Linton C, Miles S, Campbell C and Johnson E (2006). Ultra-rapid preparation of total genomic DNA from isolates of yeast and mould using Whatman FTA filter paper technology - a reusable DNA archiving system. Med Mycol **44**(5):389-98.

Bosch E, Lee A, Calafell F, Arroyo E, Henneman P, de Knijff P and Jobling M (2002). High resolution Y chromosome typing: 19 STRs amplified in three multiplex reactions. Forensic Sci Int **125**(1):42-51.

Botstein D, White R, Skolnick M and Davis R (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet **32**(3):314-31.

Brandstaetter A, Zimmermann B, Wagner J, Goebel T, Roeck A, Salas A, Carracedo A and Parson W (2008). Timing and deciphering mitochondrial DNA macro-haplogroup R0 variability in Central Europe and Middle East. BMC Evol Biol **8**:191.

Brassard J, Lamoureux L, Gagne M, Poitras E, Trottier Y and Houde A (2009). Comparison of commercial viral genomic extraction kits for the molecular detection of foodborne viruses. Can J Microbiol **55**(8):1016-9.

Brown T, Cecconi C, Tkachuk A, Bustamanate C and Clayton D (2005). Replication of mitochondrial DNA occurs by strand displacement with alternative light-strand origins, not via a strand-coupled mechanism. Genes Dev. **19**(20):2466-76.

Budowle B, Allard M, Fisher C, Isenberg A, Monson K, JE S, Wilson M and Miller K (2002). HVI and HVII mitochondrial DNA data in Apaches and Navajos. Int J Legal Med **116**(4):212-5.

Budowle B, Allard M, Wilson M and Chakraborty R (2003). Forensics and mitochondrial DNA: applications, debates, and foundations. Annu Rev Genomics Hum Genet **4**:119-41.

Budowle B, Wilson M, DiZinno J, Stauffer C, Fasano M, Holland M and Monson K (1999). Mitochondrial DNA regions HVI and HVII population data. Forensic Sci Int **103**(1):23-35.

Butler J (2001). Additional DNA markers: Amelogenin, Y-chromosome STRs, mtDNA, SNPs, Alu repeats. Forensic DNA typing: biology and technology behind STR arkers Amsterdam, Boston, London, New York, Oxford, Paris, San Diedo, San Francisco, Singapore, Sydney, Tokyo, Elsevier Academic Press**:** 117-34.

Butler J (2005a). Commonly used short tandem repeat markers and commercial kits. Forensic DNA typing- biology, technology and genetics of STR markers Amsterdam, Boston, Heidelberg, London, Oxford, Paris, San Diego, San Francisco, Singapore, Sydney, Tokyo, Elsevier**:** 85-121.

Butler J (2005b). Mitochondrial DNA analysis. Forensic DNA typing-Biology, Technology and genetics of STR markers. Amsterdam, Boston, Heidelberg, London, Oxford, Paris, San Diego, San Francisco, Singapore, Sydney, Tokyo, Elsevier**:** 241-98.

Butler J (2005c). Sample collection, DNA extraction and DNA quantitation. Forensic DNA typing: Biology, technology, and genetics of STR markers. Amsterdam, Boston, Heidelberg, London, New York, Oxford, Paris, San Diego, San Francisco, Singapore, Sydney, Tokyo, Elsevier**:** 33-62.

Cadenas A, Regueiro M, Gayden T, Singh N, Zhivotovsky L, Underhill P and Herrera R (2007). Male amelogenin dropouts: phylogenetic context, origins and implications. Forensic sci int **166**(2):155-63.

Calloway C, Reynolds R, Herrin Jr G and Anderson W (2000). The frequency of heteroplasmy in the HVII region of mtDNA differs across tissue types and increases with age. Am J Hum Genet **66**(4):1384-97.

Cann R, Stoneking M and Wilson A (1987). Mitochondrial DNA and human evolution. Nature **325**(6099):31-6.

Caramelli D, Lalueza-Fox C, Capelli C, Lari M, Sampietro M, Gigli E, Milani L, Pilli E, Guimaraes S, Chiarelli B, Marin V, Casoli A, Stanyon R, Bertranpetit J and Barbujani G (2007). Genetic analysis of the skeletal remains attributed to Francesco Petrarca. Forensic Sci Int **173**(1):36-40.

Carelli V, Achilli A, Valentino M, Rengo C, Semino O, Pala M, Olivieri A, Mattiazzi M, Pallotti F, Carrara F, Zeviani M, Leuzzi V, Carducci C, Valle G, Simionati B, Mendieta L, Salomao S, Belfort R, Sadun A and Torroni A (2006). Haplogroup effects and recombination of mitochondrial DNA: novel clues from the analysis of Leber Hereditary Optic Neuropathy pedigrees. Am J Hum Genet. **78**(4):564-74.

Carracedo A, Bar W, Lincoln P, Mayr W, Morling N, Olaisen B, Schneider P, Budowle B, Brinkmann B, Gill P, Holland M, Tully G and Wilson M (2000). DNA commission of the international society for forensic genetics: guidelines for mitochondrial DNA typing. Forensic Sci Int **110**(2):79-85.

Carracedo A, D'Aloja E, Dupuy B, Jangblad A, Karjalainen M, Lambert C, Parson W, Pfeiffer H, Pfitzinger H, Sabatier M, Syndercombe Court D and Vide C (1998). Reproducibility of mtDNA analysis between laboratories: a report of the European DNA Profiling Group (EDNAP). Forensic Sci Int **97**(2-3):165-70.

Cavelier L, Jazin E, Jalonen P and Gyllensten U (2000). MtDNA substitution rate and segregation of heteroplasmy in coding and noncoding regions. Hum Genet **107**(1):45-50.

Cerri N, Verzeletti A and De Ferrari F (2004). MtDNA: a successful identification case in burned remains. International Congress Series **1261**:425-7.

Chang Y, Perumal R, Keat P, Yong R, Kuehn D and Burgoyne L (2007). A distinct Y-STR haplotype for amelogenin negative males characterized by a large Yp11.2 (DYS458-MSY1-AMEL-Y) deletion. Forensic sci int **166**(2):115-20.

Chen X, Prosser R, Simonetti S, Sadlock J, Jagiello G and Schon E (1995a). Rearranged mitochondrial genomes are present in human oocytes. Am J Hum Genet **57**(2):239-47.

Chen Y, Torroni A, Excoffier L, Santachiara-Benerecetti A and Wallace D (1995b). Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. Am J Hum Genet. **57**(1):133-49.

Ciborowski K, Consuegra S, Garcia de Leaniz C, Beaumont M, Wang J and Jordan W (2007). Rare and fleeting: an example of interspecific recombination in animal mitochondrial DNA. Biol Lett. **3**(5):554-7.

Clayton D (1991). Nuclear gadgets in mitochondrial DNA replication and transcription. Trends Biochem Sci. **16**(3):107-11.

Coudray C, Olivieri A, Achilli A, Pala M, Melhaoui M, Cherkaoui M, El-Chennawi F, Kossmann M, Torroni A and Dugoujon J (2009). The complex and diversified mitochondrial gene pool of Berber populations. Ann Hum Genet **73**(2):196-214.

Cruz C, Ribeiro T, Vieira-Silva C, Lucas I, Geada H and Espinheira R (2004). Identification by mtDNA of exchanged human body remains. International Congress Series **1261**:374-76.

D'Aurelio M, Gajewski C, Lin M, Mauck W, Shao L, Lenaz G, Moraes C and Manfredi G (2004). Heterologous mitochondrial DNA recombination in human cells. Hum Mol Genet **13**(24):3171-9.

Deng Y, Li Y, Yu X, Li L, Wu D, Zhou J, Man T, Yang G, Yan J, Cai D, Wang J, Yang H, Li S and Yu J (2005). Preliminary DNA identification for the tsunami victims in Thailand. Genomics Proteomics Bioinformatics **3**(3):143-57.

Derbeneva O, Starikovskaya E, Wallace D and Sukernik R (2002). Traces of early Eurasians in the Mansi of northwest Siberia revealed by mitochondrial DNA analysis. Am J Hum Genet **70**(4):1009-14.

Derenko M, Malyarchuk B, Grzybowski T, Denisova G, Dambueva I, Perkova M, Dorzhu C, Luzina F, Lee H, Vanecek T, Villems R and Zakharov I (2007). Phylogeographic analysis of mitochondrial DNA in Northern Asian populations. Am J Hum Genet **81**(5):1025-41.

Di Lorenzo C, Pierelli F, Coppola G, Grieco G, Rengo C, Ciccolella M, Magis D, Bolla M, Casali C, Santorelli F and Schoenen J (2009). Mitochondrial DNA haplogroups influence the therapeutic response to riboflavin in migraineurs. Neurology **72**(18):1588-94.

Dissing J, Binladen J, Hansen A, Sejrsen B, Willerslev E and Lynnerup N (2007). The last Viking king: A royal maternity case solved by ancient DNA analysis. Forensic Sci Int **166**(1):21-7.

Dixon L, Dobbins A, Pulker H, Butler J, Vallone P, Coble M, Parson W, Berger B, Grubwieser P, Mogensen H, Morling N, Nielsen K, Sanchez J, Petkovski E, Carracedo A, Sanchez-Diz P, Ramos-Luis E, Brion M, Irwin J, Just R, Loreille O, Parsons T, Syndercombe-Court D, Schmitter H, Stradmann-Bellinghausen B, Bender K and Gill P (2006). Analysis of artificially degraded DNA using STRs

and SNPs-results of a collaborative European (EDNAP) exercise. Forensic Sci Int **164**(1):33-44.

Edwards A, Civitello A, Hammond H and Caskey C (1991). DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Am J Hum Genet. **49**(4):746-56.

Egyed B, Brandstätter A, Irwin J, Padar Z, Parsons T and Parson W (2007). Mitochondrial control region sequence variations in the Hungarian population: Analysis of population samples from Hungary and from Transylvania (Romania). Forensic Sci Int **1**(2):158-62.

Elson J, Andrews R, Chinnery P, Lightowlers R, Turnbull D and Howell N (2001). Analysis of European mtDNAs for recombination. Am J Hum Genet **68**(1):145-53.

Enriquez J, Cabezas-Herrera J, Bayona-Bafaluy M and Attardi G (2000). Very rare complementation between mitochondria carrying different mitochondrial DNA mutations points to intrinsic genetic autonomy of the organelles in cultured human cells. J Biol Chem. **275**(15):11207-15.

Eyre-Walker A, Smith NH and JM. S (1999). How clonal are human mitochondria? Proc Biol Sci. **266**(1418):477-83.

Fernandez-Silva P, Enriquez J and Montoya J (2003). Replication and transcription of mammalian mitochondrial DNA. Exp Physiol **88**(1):41-56.

Finnila S, Hassinen I and Majamaa K (2001a). Phylogenetic analysis of mitochondrial DNA in patients with an occipital stroke. Evaluation of mutations by using sequence data on the entire coding region. Mut Res **458**(1-2):31-9.

Finnila S, Lehtonen M and Majamaa K (2001b). Phylogenetic network for European mtDNA. Am J Hum Genet **68**(6):1475-84.

Finnilä S and Majamaa K (2001). Phylogenetic analysis of mtDNA haplogroup TJ in a Finnish population. J Hum Genet **46**(2):64-9.

Fish J, Raule N and Attardi G (2004). Discovery of a major D-Loop replication origin reveals two modes of human mtDNA synthesis. Science **306**(5704):2098-101.

Fisher D, Holland M, Mitchell L, Sledzik P, Wilcox A, Wadhams M and Weedn V (1993). Extraction, evaluation, and amplification of DNA from decalcified and undecalcified United States Civil War bone. J Forensic Sci **38**(1):60-8.

Folgero T, Bertheussen K, Lindal S, Torbergsen T and Oian P (1993). Andrology: Mitochondrial disease and reduced sperm motility. Hum. Reprod. **8**(11):1863-8.

Frudakis T, Thomas M, Gaskin Z, Venkateswarlu K, Chandra K, Ginjupalli S, Gunturi S, Natrajan S, Ponnuswamy V and Ponnuswamy K (2003a). Sequences associated with human iris pigmentation. Genetics **165**(4):2071-83.

Frudakis T, Venkateswarlu K, Thomas M, Gaskin Z, Ginjupalli S, Gunturi S, Ponnuswamy V, Natarajan S and Nachimuthu P (2003b). A classifier for the SNP-based inference of ancestry. J Forensic Sci **48**(4):771-82.

Fujita Y and Kubo S (2006). Application of FTA® technology to extraction of sperm DNA from mixed body fluids containing semen. Legal Med **8**(1):43-7.

Gabriel M, Calloway C, Reynolds R and Primorac D (2003). Identification of human remains by immobilized sequence-specific oligonucleotide probe analysis of mtDNA hypervariable regions I and II. Croat Med J **44**(3):293-8.

Galaal K, Meirovitz M, Hussain R, Allcroft L, Sullivan N, Lopes A and Edmondson R (2007). The feasibility of storing ovarian tumor cells on databasing paper: establishing a library of ovarian cancer DNA. Int J Gynecol Cancer **17**(1):94-100.

Ghabaee M, Omranisikaroudi M, Amrisaroukolaei S, Meysamie A, Sahraian M, Bayati A, Sanati M, Houshman M, Sadeghian H and Vajihazaman K (2009).

Mitochondrial mutation in Iranian patients with multiple sclerosis, correlation between haplogroups H, A and clinical manifestations. Cell Mol Neurobiol. **29**(3):341-6.

Giles R, Blanc H, Cann H and Wallace D (1980). Maternal inheritance of human mitochondrial DNA. Proc Natl Acad Sci U S A **77**(11):6715-9.

Gill P, Ivanov P, Kimpton C, Piercy R, Benson N, Tully G, Evett I, Hagelberg E and Sullivan K (1994). Identification of the remains of the Romanov family by DNA analysis. Nat Genet **6**(2):130-5.

Gonzales J, Loza A and Chacon E (2006). Sensitivity of different Trypanosoma vivax specific primers for the diagnosis of livestock trypanosomosis using different DNA extraction methods. Veterinary Parasitol **136**(2):119-26.

Gonzalez A, Garcia O, Larruga J and Cabrera V (2006). The mitochondrial lineage U8a reveals a Paleolithic settlement in the Basque country. BMC Genomics **7**:124.

Gonzalez A, Karadsheh N, Maca-Meyer N, Flores C, Cabrera V and Larruga J (2008). Mitochondrial DNA variation in Jordanians and their genetic relationship to other Middle East populations. Ann Hum Biol **35**(2):212-31.

Gonzalez A, Larruga J, Abu-Amero K, Shi Y, Pestano J and Cabrera V (2007). Mitochondrial linegae M1 traces an early human backflow to Africa. BMC Genomics **8**(1):223.

Goodwin W, Linacre A and Hadi S (2007). The analysis of short tandem repeats. An intoduction to Forensic Genetics. West Sussex, John Wiley & Sons, Ltd**:** 51-63.

Goodwin W, Linacre A and Vanezis P (1999). The use of mitochondrial DNA and short tandem repeat typing in the identification of air crash victims. Electrophoresis **20**(8):1707-11.

Graffy E and Foran D (2005). A simplified method for mitochondrial DNA extraction from head hair shafts. J Forensic Sci. **50**(5):1119-22.

Gray M (1989). Origin and evolution of mitochondrial DNA. Ann Rev Cell Biol **5**(1):25-50.

Grignani, P, Turchi, C, Achilli, A, Peloso, G, Alù, M, Ricci, U, Robino, C, Pelotti, S, Carnevali, E, Boschi, I, Tagliabracci, A and Previderè, C (2009). Multiplex mtDNA coding region SNP assays for molecular dissection of haplogroups U/K and J/T. Forensic Sci Int Genet 4(1): 21-5.

Grimes E, Noake J, Dixon L and Urquhart A (2001). Sequence polymorphism in the human melanocortin 1 receptor gene as an indicator of the red hair phenotype. Forensic sci int **122**(2):124-9.

Guio H, Okayama H, Ashino Y, Saitoh H, Xiao P, Miki M, Yoshihara N, Nakanowatari S and Hattori T (2006). Method for efficient storage and transportation of sputum specimens for molecular testing of tuberculosis. Int J Tuberc Lung Dis **10**(8):906-10.

Gyllensten U, Wharton D and Wilson A (1985). Maternal inheritance of mitochondrial DNA during backcrossing of two species of mice. J Hered. **76**(5):321-4.

Haak W, Forster P, Bramanti B, Matsumura S, Brandt G, Tanzer M, Villems R, Renfrew C, Gronenborn D, Alt K and Burger J (2005). Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. Science **310**(5750):1016-8.

Habano W, Sugai T, Yoshida T and Nakamura S (1999). Mitochondrial gene mutation, but not large-scale deletion, is a feature of colorectal carcinomas with mitochondrial microsatellite instability. Int J Cancer **83**(5):625-9.

Handt O, Richards M, Trommsdorff M, Kilger C, Simanainen J, Georgiev O, Bauer K, Stone A, Hedges R, Schaffner W and et al. (1994). Molecular genetic analyses of the Tyrolean Ice Man. Science **264**(5166):1775-8.

Harich, N, Costa, M, Fernandes, V, Kandil, M, Pereira, J, Silva, N and Pereira, L (2010). The trans-Saharan slave trade - clues from interpolation analyses and high-resolution characterization of mitochondrial DNA lineages. BMC Evol Biol 10(1): 138.

Hatsch D, Amory S, Keyser C, Hienne R and Bertrand L (2007). A rape case solved by mitochondrial DNA mixture analysis. J Forensic Sci **52**(4):891-4.

He H, Argiro L, Dessein H and Chevillard C (2007). Improved technique that allows the performance of large-scale SNP genotyping on DNA immobilized by FTA® technology. Infect Genet Evol **7**(1):128-32.

Hedman M, Brandstatter A, Pimenoff V, Sistonen P, Palo J, Parson W and Sajantila A (2007). Finnish mitochondrial DNA HVS-I and HVS-II population data. Forensic Sci Int **172**(2-3):171-8.

Helgason A, Hickey E, Goodacre S, Bosnes V, Stefansson K, Ward R and Sykes B (2001). MtDNA and the islands of the North Atlantic: estimating the proportions of Norse and Gaelic ancestry. Am J Hum Genet **68**(3):723-37.

Hendrickson S, Hutcheson H, Ruiz-Pesini E, Poole J, Lautenberger J, Sezgin E, Kingsley L, Goedert J, Vlahov D, Donfield S, Wallace D and O'Brien S (2008). Mitochondrial DNA haplogroups influence AIDS progression. AIDS **22**(18):2429-39.

Herrnstadt C, Elson J, Fahy E, Preston G, Turnbull D, Anderson C, Ghosh S, Olefsky J, Beal M, Davis R and Howell N (2002). Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. Am. J Hum Genet **70**(5):1152 - 71.

Holland M, Cave C, Holland C and Bille T (2003). Development of a quality, high throughput DNA analysis procedure for skeletal samples to assist with the identification of victims from the World Trade Center attacks. Croat Med J **44**(3):264-72.

Holland M and Parsons T. (1999). Mitochondrial DNA sequence analysis-- validation and use of forensic casework. Central Police University Press.

Hoppins S, Lackner L and Nunnari J (2007). The machines that divide and fuse Mitochondria. Ann Rev Biochem **76**(1):751-80.

Howell N (1997). MtDNA recombination: what do in vitro data mean? Am J Hum Genet **61**(1):19-22.

Howell N, Elson J, Howell C and Turnbull D (2007). Relative rates of evolution in the coding and control regions of African mtDNAs. Mol Biol Evol **24**(10):2213-21.

Howell N, Smejkal C, Mackey D, Chinnery P, Turnbull D and Herrnstadt C (2003). The pedigree rate of sequence divergence in the human mitochondrial genome: there is a difference between phylogenetic and pedigree rates. Am J Hum Genet **72**(3):659-70.

Imaizumi K, Parsons T, Yoshino M and Holland M (2002). A new database of mitochondrial DNA hypervariable regions I and II sequences from 162 Japanese individuals. Int J Legal Med **116**(2):68-73.

Ingman M. (2001). Mitochondrial DNA clarifies human evolution ActionBioscience Retrieved May 2008, from www.actionbioscience.org/evolution/Ingman.

Ingman M and Gyllensten U (2001). Analysis of the complete human mtDNA genome: methodology and inferences for human evolution. J Hered. **92**(6):454-61.

Ivanov P, Wadhams M, Roby R, Holland M, Weedn V and Parsons T (1996). Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II. Nat Genet **12**(4):417-20.

Jacobs H, Lehtinen SK and JN S (2000). No sex please, we are mitochondria: a hypothesis on the somatic unit of inheritance of mammalian mtDNA. Bioassays **22**(6):564-72.

Jamjoom M and Sultan A (2009). Diagnosis of clinical samples spotted on FTA cards using PCR-based methods. J Egypt Soc Parasitol. **39**(1):227-46.

Jaravata C, Smith W, Rensen G, Ruzante J and Cullor J (2006). Detection of Mycobacterium avium subsp. paratuberculosis in bovine manure using Whatman FTA card technology and Lightcycler real-time PCR. Foodborne Pathog Dis **3**(2):212-5.

Jefferies R, Ryan U and Irwin P (2007). PCR-RFLP for the detection and differentiation of the canine piroplasm species and its use with filter paper-based technologies. Vet Parasitol **144**(1-2):20-7.

Jeffreys AJ and May CA (2003). DNA enrichment by allele-specific hybridization (DEASH): a novel method for haplotyping and for detecting low-frequency base substitutional variants and recombinant DNA molecules. Genome Res **13**(10):2316-24.

Jehaes E, Pfeiffer H, Toprak K, Decorte R, Brinkmann B and Cassiman J (2001). Mitochondrial DNA analysis of the putative heart of Louis XVII, son of Louis XVI and Marie-Antoinette. Eur J Hum Genet **9**(3):185-90.

Jin L and Su B (2000). Natives or immigrants: modern human origin in east Asia. Nat Rev Genet. **1**(2):126-33.

Jobling M, Lo I, Turner D, Bowden G, Lee A, Xue Y, Carvalho-Silva D, Hurles M, Adams S, Chang Y, Kraaijenbrink T, Henke J, Guanti G, McKeown B, van Oorschot R, Mitchell R, de Knijff P, Tyler-Smith C and Parkin E (2007). Structural variation on the short arm of the human Y chromosome: recurrent multigene deletions encompassing Amelogenin Y. Hum Mol Genet **16**(3):307-16.

Johanson H, Hyland V, Wicking C and Sturm R (2009). DNA elution from buccal cells stored on Whatman FTA classic cards using a modified methanol fixation method. Biotechniques **46**(4):309-11.

Jorde L, Watkins W, Bamshad M, Dixon M, Ricker C, Seielstad M and Batzer M (2000). The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. Am J Hum Genet **66**(3):979-88.

Kain K and Lanar D (1991). Determination of genetic variation within Plasmodium falciparum by using enzymatically amplified DNA from filter paper disks impregnated with whole blood. J Clin Microbiol **29**(6):1171-4.

Kao S, Chao H and Wei Y (1995). Mitochondrial deoxyribonucleic acid 4977-bp deletion is associated with diminished fertility and motility of human sperm. Biol Reprod **52**(4):729-36.

Kashyap V, Sahoo S, Sitalaximi T and Trivedi R (2006). Deletions in the Y-derived amelogenin gene fragment in the Indian population. BMC Med Genet. **7**(1):37.

Kazibwe A, Nerima B, de Koning H, M[illegible]ser P, Barrett M and Matovu E (2009). Genotypic status of the TbAT1/P2 adenosine transporter of Trypanosoma brucei gambiense isolates from Northwestern Uganda following melarsoprol withdrawal. PLoS Negl Trop Dis **3**(9):e523.

Kirches E, Michael M, Warich-Kirches M, Schneider T, Weis S, Krause G, Mawrin C and Dietzmann K (2001). Heterogeneous tissue distribution of a mitochondrial DNA polymorphism in heteroplasmic subjects without mitochondrial disorders. J Med Genet **38**(5):312-7.

Kivisild T, Reidla M, Metspalu E, Rosa A, Brehm A, Pennarun E, Parik J, Geberhiwot T, Usanga E and Villems R (2004). Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. Am J Hum Genet **75**(5):752-70.

Kivisild T, Shen P, Wall D, Do B, Sung R, Davis K, Passarino G, Underhill P, Scharfe C and Torroni A (2006). The role of selection in the evolution of human mitochondrial genomes. Genetics **172**(1):373-87.

Kivisild T, Tolk H, Parik J, Wang Y, Papiha S, Bandelt H and Villems R (2002). The emerging limbs and twigs of the East Asian mtDNA tree. Mol Biol Evol **19**(10):1737-51.

Kobayashi S, Inoue N, Ohashi Y, Terashima M, Matsui K, Mori T, Fujita H, Awano K, Kobayashi K, Azumi H, Ejiri J, Hirata K, Kawashima S, Hayashi Y, Yokozaki H, Itoh H and Yokoyama M (2003). Interaction of oxidative stress and inflammatory response in coronary plaque instability: important role of C-reactive protein. Arterioscler Thromb Vasc Biol **23**(8):1398-404.

Kobayashi Y, Momoi MY, Tominaga K, Shimoizumi H, Nihei K, Yanagisawa M, Kagawa Y and Ohta S (1991). Respiration-deficient cells are caused by a single point mutation in the mitochondrial tRNA-Leu (UUR) gene in mitochondrial myopathy, encephalopathy, lactic acidosis, and strokelike episodes (MELAS). Am J Hum Genet **49**(3):590-9.

Kobilinsky L, Liotti T and Oeser-Sweat J (2004). DNA: Forensic and Legal Applications Wiley-Interscience

Kofler B, Mueller E, Eder W, Stanger O, Maier R, Weger M, Haas A, Winker R, Schmut O, Paulweber B, Iglseder B, Renner W, Wiesbauer M, Aigner I, Santic D, Zimmermann F, Mayr J and Sperl W (2009). Mitochondrial DNA haplogroup T is associated with coronary artery disease and diabetic retinopathy: a case control study. BMC Med Genet. **10**(1):35.

Kogelnik A, Lott M, Brown M, Navathe S and Wallace D (1998). MITOMAP: a human mitochondrial genome database_1998 update. Nucleic Acids Res **26**(1):112-5.

Kohnemann, S, Sibbing, U, Pfeiffer, H and Hohoff, C (2008). A rapid mtDNA assay of 22 SNPs in one multiplex reaction increases the power of forensic testing in European Caucasians. Int J Legal Med. 122(6): 517-23.

Kong Q, Bandelt H, Sun C, Yao Y, Salas A, Achilli A, Wang C, Zhong L, Zhu C and Wu C (2006). Updating the East Asian mtDNA phylogeny: a prerequisite for the identification of pathogenic mutations. Hum Mol Genet **15**(13):2076-86.

Kreader C (1996). Relief of amplification inhibition in PCR with bovine serum albumin or T4 gene 32 protein. Appl Environ Microbiol **62**(3):1102-6.

Krings M, Halim Salem A, Bauer K, Geisert H, Malek A, Chaix L, Simon C, Welsby D, Di Rienzo A, Utermann G, Sajantila A, Paabo S and Stoneking M (1999). MtDNA analysis of Nile valley populations: A genetic corridor or a barrier to migration? Am J Hum Genet **64**(4):1166-76.

Kumagai R, Sasaki Y, Tokuta T, Biwasaka H and Aoki Y (2008). DNA analysis of family members with deletion in Yp11.2 region containing amelogenin locus. Legal Med (Tokyo) **10**(1):39-42.

Kumar S, Padmanabham P, Ravuri R, Uttaravalli K, Koneru P, Mukherjee P, Das B, Kotal M, Xaviour D, Saheb S and Rao V (2008). The earliest settlers' antiquity and evolutionary history of Indian populations: evidence from M2 mtDNA lineage. BMC Evol Biol **8**(1):230.

Kumar S, Ravuri R, Koneru P, Urade B, Sarkar B, Chandrasekar A and Rao V (2009). Reconstructing Indian-Australian phylogenetic link. BMC Evol Biol **9**(1):173.

Kvist L, Martens J, Nazarenko A and Orell M (2003). Paternal leakage of mitochondrial DNA in the great tit (Parus major). Mol Biol Evol **20**(2):243-7.

Ladoukakis E and Zouros E (2001). Direct evidence for homologous recombination in mussel (Mytilus galloprovincialis) mitochondrial DNA. Mol Biol Evol **18**(7):1168-75.

Lattanzi W, Di Giacomo M, Lenato G, Chimienti G, Voglino G, Resta N, Pepe G and Guanti G (2005). A large interstitial deletion encompassing the amelogenin gene on the short arm of the Y chromosome. Hum Genet **116**(5):395-401.

Lauber J, Marsac C, Kadenbach B and Seibel P (1991). Mutations in mitochondrial tRNA genes: a frequent cause of neuromuscular diseases. Nucl. Acids Res. **19**(7):1393-97.

Lee HY, Chung U, Yoo JE, Park MJ and Shin KJ (2004). Quantitative and qualitative profiling of mitochondrial DNA length heteroplasmy. Electrophoresis **25**(1):28-34.

Lema C, Kohl-White K, Lewis L and Dao D (2006). Optimised pH method for DNA elution from buccal cells collected in Whatman FTA cards. Genet Test **10**(2):126-30.

Li B, Zhong F, Yi H, Wang X, Li L, Wang L, Qi X and Wu L (2007). Genetic polymorphism of mitochondrial DNA in Dong, Gelao, Tujia, and Yi ethnic populations from Guizhou, China. J Genet Genomics **34**(9):800-10.

Lightowlers R, Chinnery P, Turnbull D and Howell N (1997). Mammalian mitochondrial genetics: heredity, heteroplasmy and disease. Trends Genet **13**(11):450-5.

Lima G, Pontes M, Abrantes D, Cainé L, Pereira M, Matos P and Pinheiro M (2006). HVI and HVII sequence polymorphisms of the human mtDNA in the North of Portugal: population data and maternal lineages Int Cong Ser. **1288**:133-5.

Logan J (2009). A refined phylogeny for mtDNA haplogroup J. J Genet Genealogy **5**(1):16-22.

Loogvali E, Roostalu U, Malyarchuk B, Derenko M, Kivisild T, Metspalu E, Tambets K, Reidla M, Tolk H, Parik J, Pennarun E, Laos S, Lunkina A, Golubenko M, Barac L, Pericic M, Balanovsky O, Gusar V, Khusnutdinova E, Stepanov V, Puzyrev V, Rudan P, Balanovska E, Grechanina E, Richard C, Moisan J, Chaventre A, Anagnou N, Pappa K, Michalodimitrakis E, Claustres M, Golge M, Mikerezi I, Usanga E and Villems R (2004). Disuniting uniformity: a pied cladistic canvas of mtDNA haplogroup H in Eurasia. Mol Biol Evol **21**(11):2012-21.

Lutz-Bonengel S, Sänger T, Pollak S and Szibor R (2004). Different methods to determine length heteroplasmy within the mitochondrial control region. Int J Legal Med **118**(5):274-81.

Mabuchi T, Susukida R, Kida A and Oya M (2007). Typing the 1.1 kb control region of human mitochondrial DNA in Japanese individuals. J Forensic Sci **52**(2):355-63.

Maca-Meyer N, Gonzalez A, Larruga J, Flores C and Cabrera V (2001). Major genomic mitochondrial lineages delineate early human expansions. BMC Genet **2**(1):13-20.

Macaulay V, Richards M, Hickey E, Vega E, Cruciani F, Guida V, Scozzari R, Bonne-Tamir B, Sykes B and Torroni A (1999). The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. Am J Hum Genet **64**(1):232-49.

Mai M, Hoyer J and McClure R (2004). Use of multiple displacement amplification to amplify genomic DNA before sequencing of the haemoglobin genes. J Clin Pathol **57**(6):637-40.

Majamaa K, Finnila S, Turkka J and Hassinen IE (1998). Mitochondrial DNA haplogroup U as a risk factor for occipital stroke in migraine. Lancet **352**(9126):455-6.

Maji S, Krithika S and Vasulu T (2009). Phylogeographic distribution of mitochondrial DNA macrohaplogroup M in India. J Genet. **88**(1):127-39.

Malathi J, Sowmiya M, Margarita S, Madhavan H and Therese K (2009). Application of PCR based-RFLP for species identification of ocular isolates of methicillin resistant staphylococci (MRS). Indian J Med Res **130**(1):78-84.

Malik S, Sudoyo H, Pramoonjago P, Sukarna T, Darwis D and Marzuki S (2002a). Evidence for the de novo regeneration of the pattern of the length heteroplasmy associated with the T16189C variant in the control (D-loop) region of mitochondrial DNA. J Hum Genet **47**(3):122-30.

Malik S, Sudoyo H, Pramoonjago P, Suryadi H, Sukarna T, Njunting M, Sahiratmadja E and Marzuki S (2002b). Nuclear mitochondrial interplay in the modulation of the homopolymeric tract length heteroplasmy in the control (D-loop) region of the mitochondrial DNA. Hum Genet **110**(5):402-11.

Malyarchuk B, Derenko M, Perkova M, Grzybowski T, Vanecek T and Lazur J (2008). Reconstructing the phylogeny of African mitochondrial DNA lineages in Slavs. Eur J Hum Genet **16**(9):1091-6.

Malyarchuk B, Grzybowski T, Derenko M, Czarny J and Miscicka-Sliwka D (2005). Mitochondrial DNA Diversity in the Polish Roma. Ann Hum Genet **70**(2):195-206.

Marjanović D, Durmić-Pašić A, Bakal N, Haverić S, Kalamujić B, Kovačević L, Ramić J, Pojskić N, Škaro V, Projić P, Bajrović K, Hadžiselimović R, Drobnič K, Huffine ED, Davoren J, Primorac D. (2007 ). DNA identification of skeletal remains from World War II Mass Graves uncovered in Slovenia. Croat Med J **48**(4):513-9.

Martin W and Herrmann R (1998). Gene transfer from organelles to the nucleus: how much, what happens, and why? Plant Physiol **118**(1):9-17.

Mellars P (2006). Going East: new genetic and archaeological perspectives on the modern human colonization of Eurasia. Science **313**(5788):796-800.

Metspalu M, Kivisild T, Bandelt H, Richards M and Villems R (2006). The pioneer settlement of modern humans in Asia. Human mitochondrial DNA and the evolution of Homo sapiens. M. V. Bandelt HJ, Richards M Heidelberg, Springer-Verlag**:** 181-99.

Meyer S, Weiss G and von Haeseler A (1999). Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. Genetics **152**(3):1103-10.

Michael A and Brauner P (2004). Erroneous gender identification by the amelogenin sex test. J Forensic Sci **49**(2):258-9.

Minaguchi K, Maruyama S, Kasahara I, Nohira C, Hanaoka Y, Tsai T, Kiriyama H and Takahashi N (2005). Identification of unknown body using DNA analysis and dental characteristics in chest X-ray photograph. Bull Tokyo Dent Coll **46**(4):145-53.

Mitchell R, Kreskas M, Baxter E, Buffalino L and Van Oorschot R (2006). An investigation of sequence deletions of amelogenin (AMELY), a Y-chromosome locus commonly used for gender determination. Ann Hum Biol **33**(2):227-40.

Moore F and Reijo-Pera R (2000). Male sperm motility dictated by mother's mtDNA. Am J Hum Genet **67**(3):543-8.

Morata P, Queipo-Ortuno M and de Dios Colmenero J (1998). Strategy for optimizing DNA amplification in a peripheral blood PCR assay used for diagnosis of human brucellosis. J Clin Microbiol **36**(9):2443-6.

Morrison L, McCormack G, Sweeney L, Likeufack A, Truc P, Turner C, Tait A and Macleod A (2007). Use of multiple displacement amplification to increase the detection and genotyping of Trypanosoma species samples immobilized on FTA filters. Am J Trop Med Hyg **76**(6):1132-7.

Moscoso H, Alvarado I and Hofacre C (2006). Molecular analysis of infectious bursal disease virus from bursal tissues collected on FTA filter paper. Avain Dis **50**(3):391-6.

Moscoso H, Bruzual J, Sellers H and Hofacre C (2007). FTA liver impressions as DNA template for detecting and genotyping fowl adenovirus. Avain Dis **51**(1):118-21.

Moscoso H, Thayer S, Hofacre C and Kleven S (2004). Inactivation, storage, and PCR detection of Mycoplasma on FTA filter paper. Avain Dis **48**(4):841-50.

Mukaida M, Kimura H, Takada Y, Masuda T and Nakata Y (2000). The personal identification of many samples recovered from under the sea. Forensic Sci Int **113**(1-3):79-85.

Muthukrishnan M, Singanallur N, Ralla K and Villuppanoor S (2008). Evaluation of FTA® cards as a laboratory and field sampling device for the detection of foot-and-mouth disease virus and serotyping by RT-PCR and real-time RT-PCR. J Virol Methods **151**(2):311-6.

Nakahori Y, Takenaka O and Nakagome Y (1991). A human X-Y homologous region encodes "amelogenin". Genomics **9**(2):264-9.

Ndunguru J, Taylor N, Yadav J, Aly H, Legg J, Aveling T, Thompson G and Fauquet C (2005). Application of FTA technology for sampling, recovery and molecular characterization of viral pathogens and virus-derived transgenes from plant tissues. Virol J **2**(1):45.

Nechvatal J, Ram JL, Basson M, Namprachan P, Niec S, Badsha K, Matherly L, Majumdar A and Kato I (2008). Fecal collection, ambient preservation, and DNA extraction for PCR amplification of bacterial and human markers from human feces. J Microbiol Methods **72**(2):124-32.

Nei M and Tajima F (1981). DNA polymorphism detectable by restriction endonucleases. Genetics **97**(145-63).

Nelson K and Melton T (2007). Forensic mitochondrial DNA analysis of 116 casework skeletal samples. J Forensic Sci **52**(3):557-61.

Nelson T, Just R, Loreille O, Schanfield M and Podini D (2007). Development of a multiplex single base extension assay for mitochondrial DNA haplogroup typing. Croat Med J **48**(4):460-72.

Olivieri A, Achilli A, Pala M, Battaglia V, Fornarino S, Al-Zahery N, Scozzari R, Cruciani F, Behar D, Dugoujon J, Coudray C, Santachiara-Benerecetti A, Semino O, Bandelt H and Torroni A (2006). The mtDNA legacy of the Levantine early Upper Palaeolithic in Africa. Science **314**(5806):1767-70.

Omran G, Rutty G and Jobling M (2008). Diversity of 17-locus Y-STR haplotypes in Upper (Southern) Egyptians. For Sci Int Genet Suppl Ser **1**(1):230-2.

Owor B, Shepherd D, Taylor N, Edema R, Monjane A, Thomson J, Martin D and Varsani A (2007). Successful application of FTA® Classic Card technology and use of bacteriophage [phi]29 DNA polymerase for large-scale field sampling and cloning of complete maize streak virus genomes. J Virol Methods **140**(1-2):100-5.

Palanichamy M, Sun C, Agrawal S, Bandelt H, Kong Q, Faisal K, Wang C, Chaudhuri T, Palla V and Zhang Y (2004). Phylogeny of mitochondrial DNA

macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia. Am J Hum Genet **75**(6):966-78.

Palo J, Hedman M, Soderholm N, Sajantila A (2007). Repatriation and Identification of Finnish World War II Soldiers. Croat Med J **48**(4):528-35.

Paneto G, Martins J, Longo L, Pereira G, Freschi A, Alvarenga V, Chen B, Oliveira R, Hirata M and Cicarelli R (2007). Heteroplasmy in hair: differences among hair and blood from the same individuals are still a matter of debate. Forensic Sci Int **173**(2-3):117-21.

Park S, Kim J, Yang Y and Lee S (2008). Direct STR amplification from whole blood and blood- or saliva-spotted FTA without DNA purification. J Forensic Sci **53**(2):335-41.

Parkin E, Kraayenbrink T, Opgenort J, van Driem G, Tuladhar N, de Knijff P and Jobling M (2007). Diversity of 26-locus Y-STR haplotypes in a Nepalese population sample: isolation and drift in the Himalayas. Forensic Sci Int **166**(2-3):176-81.

Passarino G, Semino O, Quintana-Murci L, Excoffier L, Hammer M and Santachiara-Benerecetti A (1998). Different genetic components in the Ethiopian population, identified by mtDNA and Y-chromosome polymorphisms. Am J Hum Genet **62**(2):420-34.

Pereira L, Macaulay V, Torroni A, Scozzari R, Prata M and Amorim A (2001). Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade. Ann Hum Genet **65**(Pt 5):439-58.

Perozo F, Villegas P, Estevez C, Alvarado I and Purvis L (2006). Use of FTA® filter paper for the molecular detection of Newcastle disease virus. Avian Pathol **35**(2):93-8.

Pezzoli N, Silvy M, Woronko A, Le Treut T, Levy-Mozziconacci A, Reviron D, Gabert J and Picard C (2007). Quantification of mixed chimerism by real time PCR on whole blood-impregnated FTA cards. Leukemia Res **31**(9):1175-83.

Piechota J, Tonska K, Nowak M, Kabzinska D, Lorenc A and Bartnik E (2004). Comparison between the Polish population and European populations on the basis of mitochondrial morphs and haplogroups. Acta Biochim Pol **51**(4):883-95.

Pike D (2006). Phylogenetic networks for human mtDNA haplogroup T. J Genet Genealogy **2**:1-11.

Plaza S, Calafell F, Helal A, Bouzerna N, Lefranc G, Bertranpetit J and Comas D (2003). Joining the pillars of Hercules: mtDNA sequences show multidirectional gene flow in the western Mediterranean. Ann Hum Genet **67**:312-28.

Poetsch M, Wittig H, Krause D and Lignitz E (2003). Mitochondrial diversity of a northeast German population sample. Forensic Sci Int **137**(2-3):125-32.

Powell H, Gooding C, Garrett S, Lund B and Mckee R (1994). Proteinase inhibition of the detection of Listeria monocytogenes in milk using the polymerase chain reaction. Lett Appl Microbiol **18**(1):59-61.

Primorac D and Schanfield M (2000). Application of forensic DNA testing in the legal system. Croat Med J **41**(1):32-46.

Purvis L, Villegas P and Perozo F (2006). Evaluation of FTA® paper and phenol for storage, extraction and molecular characterization of infectious bursal disease virus. J Virol Methods **138**(1-2):66-9.

Rajendram D, Ayenza R, Holder F, Moran B, Long T and Shah H (2006). Long-term storage and safe retrieval of DNA from microorganisms for molecular analysis using FTA matrix cards. J Microbiol Methods **67**(3):582-92.

Rand D (2001). The units of selection of mitochondrial DNA. Ann Rev Ecol Syst **32**(1):415-48.

Rando J, Pinto F, Gonzalez A, Hernandez M, Larruga J, Cabrera V and Bandelt H (1998). Mitochondrial DNA analysis of Northwest African populations reveals genetic exchanges with European, Near-Eastern, and sub-Saharan populations. Ann Hum Genet **62**(Pt 6):531-50.

Rego-Perez I, Fernandez-Moreno M, Fernandez-Lopez C, Arenas J and Blanco F (2008). Mitochondrial DNA haplogroups: role in the prevalence and severity of knee osteoarthritis. Arthritis Rheum. **58**(8):2387-96.

Reidla M, Kivisild T, Metspalu E, Kaldma K, Tambets K, Tolk H, Parik J, Loogvali E, Derenko M, Malyarchuk B, Bermisheva M, Zhadanov S, Pennarun E, Gubina M, Golubenko M, Damba L, Fedorova S, Gusar V, Grechanina E, Mikerezi I, Moisan J, Chaventre A, Khusnutdinova E, Osipova L, Stepanov V, Voevoda M, Achilli A, Rengo C, Rickards O, De Stefano G, Papiha S, Beckman L, Janicijevic B, Rudan P, Anagnou N, Michalodimitrakis E, Koziel S, Usanga E, Geberhiwot T, Herrnstadt C, Howell N, Torroni A and Villems R (2003). Origin and diffusion of mtDNA haplogroup X. Am J Hum Genet **73**(5):1178-90.

Reynier P, Penisson-Besnier I, Moreau C, Savagner F, Vielle B, Emile J, Dubas F and Malthiery Y (1999). MtDNA haplogroup J: a contributing factor of optic neuritis. Eur J Hum Genet. **7**(3):404-6.

Richards M, Macaulay V, Bandelt H and Sykes B (1998). Phylogeography of mitochondrial DNA in western Europe. Ann Hum Genet **62**(Pt 3):241-60.

Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C, Sellitto D, Cruciani F, Kivisild T, Villems R, Thomas M, Rychkov S, Rychkov O, Rychkov Y, Gölge M, Dimitrov D, Hill E, Bradley D, Romano V, Calì F, Vona G, Demaine A, Papiha S, Triantaphyllidis C, Stefanescu G, Hatina J, Belledi M, Di Rienzo A, Novelletto A, Oppenheim A, Nørby S, Al-Zaheri N, Santachiara-Benerecetti S, Scozzari R, Torroni A and Bandelt H (2000). Tracing European founder lineages in the Near Eastern mtDNA Pool. Am J Hum Genet **67**(5):1251-76.

Richards M, Rengo C, Cruciani F, Gratrix F, Wilson J, Scozzari R, Macaulay V and Torroni A (2003). Extensive female-mediated gene flow from sub-Saharan Africa into Near Eastern Arab populations. Am J Hum Genet **72**(4):1058-64.

Rispe C and Moran N (2000). Accumulation of deleterious mutations in Endosymbionts: Muller's Ratchet with two levels of selection. Am Naturalist **156**(4):425-41.

Robin E and Wong R (1988). Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. J cell Physiol **136**(3):507-13.

Roffey P, Eckhoff C and Kuhl J (2000). A rare mutation in the amelogenin gene and its potential investigative ramifications. J Forensic Sci **45**(5):1016-9.

Rogaev E, Grigorenko A, Moliaka Y, Faskhutdinova G, Goltsov A, Lahti A, Hildebrandt C, Kittler E and Morozova I (2009). Genomic identification in the historical case of the Nicholas II royal family. Proc Natl Acad Sci U S A **106**(13):5258-63.

Roostalu U, Kutuev I, Loogvali E, Metspalu E, Tambets K, Reidla M, Khusnutdinova E, Usanga E, Kivisild T and Villems R (2007). Origin and expansion of haplogroup H, the dominant human mitochondrial DNA lineage in West Eurasia: the Near Eastern and Caucasian perspective. Mol Biol Evol **24**(2):436-48.

Rosa A, Brehm A, Kivisild T, Metspalu E and Villems R (2004). MtDNA profile of West Africa Guineans: towards a better understanding of the Senegambia region. Ann Hum Genet **68**(Pt 4):340-52.

Rosa A, Fonseca B, Krug T, Manso H, Gouveia L, Albergaria I, Gaspar G, Correia M, Viana-Baptista M, Simoes R, Pinto A, Taipa R, Ferreira C, Fontes J, Silva M, Gabriel J, Matos I, Lopes G, Ferro J, Vicente A and Oliveira S (2008). Mitochondrial haplogroup H1 is protective for ischemic stroke in Portuguese patients. BMC Med Genet. **9**(1):57.

Rowold D, Luis J, Terreros M and Herrera R (2007). Mitochondrial DNA geneflow indicates preferred usage of the Levant Corridor over the Horn of Africa passageway. J Hum Genet **52**(5):436-47.

Rudin N and Inman K (2002a). An overview of forensic DNA typing systems. An introduction to forensic DNA analysis. London, CRC press**:** 41-2.

Rudin N and Inman K (2002b). Procedures for forensic DNA analysis. An introduction to forensic DNA analysis. London, CRC press**:** 71-6.

Ruiz-Pesini E, Lapena A, Diez-Sanchez C, Perez-Martos A, Montoya J, Alvarez E, Diaz M, Urries A, Montoro L, Lopez-Perez M and Enriquez J (2000). Human mtDNA haplogroups associated with high or reduced spermatozoa motility. Am J Hum Genet. **67**(3):682-96.

Salas A, Bandelt H, Macaulay V and Richards M (2007). Phylogeographic investigations: the role of trees in forensic genetics. Forensic Sci Int **168**(1):1-13.

Salas A, Lareu M and Carracedo A (2001). Heteroplasmy in mtDNA and the weight of evidence in forensic mtDNA analysis: a case report. Int J Legal Med **114**(3):186-90.

Salas A, Richards M, De la Fe T, Lareu M, Sobrino B, Sanchez-Diz P, Macaulay V and Carracedo A (2002). The making of the African mtDNA landscape. Am J Hum Genet **71**(5):1082-111.

Salas A, Richards M, Lareu M, Scozzari R, Coppa A, Torroni A, Macaulay V and Carracedo Á (2004). The African diaspora: mitochondrial DNA and the Atlantic slave trade. Am J Hum Genet **74**(3):454-65.

Salvador J and De Ungria M (2003). Isolation of DNA from saliva of betel quid chewers using treated cards. J Forensic Sci **48**(4):794-7.

Sampietro M, Gilbert M, Lao O, Caramelli D, Lari M, Bertranpetit J and Lalueza-Fox C (2006). Tracking down human contamination in ancient human teeth. Mol Biol Evol **23**(9):1801-7.

SanGiovanni J, Arking D, Iyengar S, Elashoff M, Clemons T, Reed G, Henning A, Sivakumaran T, Xu X, DeWan A, Agr³ín E, Rochtchina E, Sue C, Wang J, Mitchell P, Hoh J, Francis P, Klein M, Chew E and Chakravarti A (2009). Mitochondrial DNA variants of respiratory complex I that uniquely characterize haplogroup T2 are associated with increased risk of age-related macular degeneration. PLoS One **4**(5):e5508.

Santos F, Pandya A and Tyler-Smith C (1998). Reliability of DNA-based sex tests. Nat Genet **18**(2):103.

Saunier J, Irwin J, Strouss K, Ragab H, Sturk K and Parsons T (2009). Mitochondrial control region sequences from an Egyptian population sample. Forensic Sci Int Genet **3**(3):e97-103.

Schlebusch, C, Naidoo, T and Soodyall, H (2009). SNaPshot minisequencing to resolve mitochondrial macro-haplogroups found in Africa. Electrophoresis 30(21): 3657-64.

Schwartz M and Vissing J (2002). Paternal inheritance of mitochondrial DNA. N Engl J Med **347**(8):576-80.

Seah L, Jeevan N, Othman M, Jaya P, Ooi Y, Wong P and Kee S (2003). STR data for the AmpFlSTR Identifiler loci in three ethnic groups (Malay, Chinese, Indian) of the Malaysian population. Forensic Sci Int **138**(1-3):134-7.

Shadel G and Clayton D (1997). Mitochondrial DNA and maintenance in vertebrates. Ann Rev Biochem **66**(1):409-35.

Shadrach B, Commane M, Hren C and Warshawsky I (2004). A rare mutation in the primer binding region of the amelogenin gene can interfere with gender identification. J Mol Diagn **6**(4):401-5.

Shamsi M, Kumar R, Bhatt A, Bamezai R, Kumar R, Gupta N, Das T and Dada R (2008). Mitochondrial DNA mutations in etiopathogenesis of male infertility. Indian J Urol **24**(2):150-4.

Shen P, Lavi T, Kivisild T, Chou V, Sengun D, Gefel D, Shpirer I, Woolf E, Hillel J, Feldman M and Oefner P (2004). Reconstruction of patrilineages and matrilineages of Samaritans and other Israeli populations from Y-chromosome and mitochondrial DNA sequence variation. Hum Mutat **24**(3):248-60.

Shlush L, Atzmon G, Weisshof R, Behar D, Yudkovsky G, Barzilai N and Skorecki K (2008). Ashkenazi Jewish centenarians do not demonstrate enrichment in mitochondrial haplogroup J. PLoS One **3**(10):e3425.

Simon S, Veron V and Carme B (2010). Leishmania spp. identification by polymerase chain reaction-restriction fragment length polymorphism analysis and its applications in French Guiana. Diagnostic Microbiology and Infectious Disease **66**(2):175-80.

Skaletsky H, Kuroda-Kawaguchi T, Minx P, Cordum H, Hillier L, Brown L, Repping S, Pyntikova T, Ali J, Bieri T, Chinwalla A, Delehaunty A, Delehaunty K, Du H, Fewell G, Fulton L, Fulton R, Graves T, Hou S, Latrielle P, Leonard S, Mardis E, Maupin R, McPherson J, Miner T, Nash W, Nguyen C, Ozersky P, Pepin K, Rock S, Rohlfing T, Scott K, Schultz B, Strong C, Tin-Wollam A, Yang S, Waterston R, Wilson R, Rozen S and Page D (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature **423**(6942):825-37.

Smith L and Burgoyne L (2004). Collecting, archiving and processing DNA from wildlife samples using FTA databasing paper. BMC Ecol **4**:4.

Sobrino B, Brion M and Carracedo A (2005). SNPs in forensic genetics: a review on SNP typing methodologies. Forensic Sci Int **154**(2-3):181-94.

Steemers F, Chang W, Lee G, Barker D, Shen R and Gunderson K (2006). Whole-genome genotyping with the single-base extension assay. Nat Methods **3**(1):31-3.

Steinlechner M, Berger B, Niederstatter H and Parson W (2002). Rare failures in the amelogenin sex test. Int J Legal Med **116**(2):117-20.

Stevanovitch A, Gilles A, Bouzaid E, Kefi R, Paris F, Gayraud R, Spadoni J, El-Chenawi F and Béraud-Colomb E (2004). Mitochondrial DNA sequence diversity in a sedentary population from Egypt. Ann Hum Genet. **68**(Pt 1):23-39.

Stewart J, Fisher C, Aagaard P, Wilson M, Isenberg A, Polanskey D, Pokorak E, DiZinno J and Budowle B (2001). Length variation in HV2 of the human mitochondrial DNA control region. J Forensic Sci **46**(4):862-70.

Stone A, Starrs J and Stoneking M (2001). Mitochondrial DNA analysis of the presumptive remains of Jesse James. J Forensic Sci **46**(1):173-6.

Stoneking M, Hedgecock D, Higuchi R, Vigilant L and Erlich H (1991). Population variation of human mtDNA control region sequences detected by enzymatic

amplification and sequence-specific oligonucleotide probes. Am J Hum Genet **48**(2):370-82.

Sudhakaran R, Mekata T, Kono T, Supamattaya K, Linh N, Suzuki Y, Sakai M and Itami T (2009). A simple non-enzymatic method for the preparation of white spot syndrome virus (WSSV) DNA from the haemolymph of Marsupenaeus japonicus using FTA matrix cards. J Fish Dis **32**(7):611-7.

Suematsu N and Isohashi F (2006). Rapid and simple screening of transgenic mice: novel extraction-free, filter-based PCR genotyping from blood samples. Acta Biochim Pol **53**(3):613-6.

Suissa S, Wang Z, Poole J, Wittkopp S, Feder J, Shutt T, Wallace D, Shadel G and Mishmar D (2009). Ancient mtDNA genetic variants modulate mtDNA transcription and replication. PLoS Genet **5**(5):e1000474.

Sullivan D, Fahey B and Titus D (2006). Fast PCR: general considerations for minimizing run times and maximizing throughput. Bio Rad Laboratories:Technical note 5362.

Sullivan K, Mannucci A, Kimpton C and Gill P (1993). A rapid and quantitative DNA sex test: fluorescence- based PCR analysis of X-Y homologous gene amelogenin. Biotechniques **15**(4):636-41.

Sultan D, Khalil M, Abdouh A, Doleh W and Al Muthanna A (2009). Imported Malaria in United Arab Emirates: evaluation of a new DNA extraction technique using nested PCR. Korean J Parasitol **47**(3):227-33.

Sutovsky P, Van Leyen K, McCauley T, Day B and Sutovsky M (2004). Degradation of paternal mitochondria after fertilization: implications for heteroplasmy, assisted reproductive technologies and mtDNA inheritance. Reprod Biomed Online. **8**(1):24-33.

Sweet D, Lorente M, Valenzuela A, Lorente J and Alvarez J (1996). Increasing DNA extraction yield from saliva stains with a modified Chelex method. Forensic Sci Int **83**(3):167-77.

Takayama T, Takada N, Suzuki R, Nagaoka S, Watanabe Y, Kumagai R, Aoki Y and Butler J (2009). Determination of deleted regions from Yp11.2 of an amelogenin negative male. Legal med (Tokyo) **11**(Suppl 1):S578-80.

Tamm E, Kivisild T, Reidla M, Metspalu E and Smith D (2007). Beringian standstill and spread of Native American founders. PLoS One **2**(9):e829.

Tanaka M, Cabrera V, Gonzalez A, Larruga J, Takeyasu T, Fuku N, Guo L, Hirose R, Fujita Y and Kurata M (2004). Mitochondrial genome variation in Eastern Asia and the peopling of Japan. Genome Res **14**(10A):1832-50.

Taylor R and Turnbull D (2005). Mitochondrial DNA mutations in human disease. Nat Rev Genet **6**(5):389-402.

Thangaraj K, Chanbey G, Singh V, Vanniarajan A, Thanseem I, Reddy A and Singh L (2006). In situ origin of deep rooting lineages on Indian mitochondrial macrohaplogroup M in India. BMC Genomics **7**:151.

Thangaraj K, Reddy A and Singh L (2002). Is the amelogenin gene reliable for gender identification in forensic casework and prenatal diagnosis? Int J Legal Med **116**(2):121-3.

Thompson W, Ramalho-Santos J and Sutovsky P (2003). Ubiquitination of prohibitin in mammalian sperm mitochondria: possible roles in the regulation of mitochondrial inheritance and sperm quality control. Biol Reprod **69**(1):254-60.

Thyagarajan B, Padua R and Campbell C (1996). Mammalian mitochondria possess homologous DNA recombination activity. J Biol Chem. **271**(44):27536-43.

Tolk H, Barac L, Pericic M, Klaric I, Janicijevic B, Campbell H, Rudan I, Kivisild T, Villems R and Rudan P (2001). The evidence of mtDNA haplogroup F in a

European population and its ethnohistoric implications. Eur J Hum Genet **9**(9):717-23.

Tomas C, Sanchez J, Barbaro A, Brandt-Casadevall C, Hernandez A, Ben Diva M, Ramon M and Morling N (2008). X-chromosome SNP analyses in 11 human Mediterranean populations show a high overall genetic homogeneity except in North-west Africans (Moroccans). BMC Evol Biol **8**:75.

Torroni A, Achilli A, Macaulay V, Richards M and Bandelt H (2006). Harvesting the fruit of the human mtDNA tree. Trends Genet **22**(6):339-45.

Torroni A, Bandelt H, Macaulay V, Richards M, Cruciani F, Rengo C, Martinez-Cabrera V, Villems R, Kivisild T, Metspalu E, Parik J, Tolk H, Tambets K, Forster P, Karger B, Francalacci P, Rudan P, Janicijevic B, Rickards O, Savontaus M, Huoponen K, Laitinen V, Koivumaki S, Sykes B, Hickey E, Novelletto A, Moral P, Sellitto D, Coppa A, Al-Zaheri N, Santachiara-Benerecetti A, Semino O and Scozzari R (2001). A signal, from human mtDNA, of postglacial recolonization in Europe. Am J Hum Genet **69**(4):844-52.

Torroni A, Huoponen K, Francalacci P, Petrozzi M, Morelli L, Scozzari R, Obinu D, Savontaus M and Wallace D (1996). Classification of European mtDNAs from an analysis of three European populations. Genetics **144**(4):1835-50.

Torroni A, Lott M, Cabell M, Chen Y, Lavergne L and Wallace D (1994). mtDNA and the origin of Caucasians: identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-loop region. Am J Hum Genet **55**(4):760-76.

Torroni A, Richards M, Macaulay V, Forster P, Villems R, Norby S, Savontaus M, Huoponen K, Scozari R and Bandelt H (2000). MtDNA haplogroups and frequency patterns in Europe. Am J Hum Genet **66**(3):1173-7.

Torroni A and Wallace D (1994). Mitochondrial DNA variation in human populations and implications for detection of mitochondrial DNA mutations of pathological significance. J Bioenerg Biomembr **26**(3):261-71.

Tsukaya H (2004). Gene flow between Impatiens radicans and I. javensis(Balsaminaceae) in Gunung Pangrango. Am J Bot **91**(12):2119-23.

Tsutsumi H, Komuro T, Mukoyama R and Nogami H (2006). Hypervariable region structure and polymorphism of mtDNA from dental pulp and a family analysis J Oral Sci **48**(3):145-52.

Tully G, Barritt S, Bender K, Brignon E, Capelli C, Dimo-Simonin N, Eichmann C, Ernst C, Lambert C, Lareu MV, Ludes B, Mevag B, Parson W, Pfeiffer H, Salas A, Schneider P and Staalstrom E (2004). Results of a collaborative study of the EDNAP group regarding mitochondrial DNA heteroplasmy and segregation in hair shafts. Forensic Sci Int **140**(1):1-11.

Turchi C, Buscemi L, Giacchino E, Onofri V, Fendt L, Parson W and Tagliabracci A (2009). Polymorphisms of mtDNA control region in Tunisian and Moroccan populations: an enrichment of forensic mtDNA databases with Northern African data. Forensic Sci Int Genet **3**(3):166-72.

Ujvari B, Dowton M and Madsen T (2007). Mitochondrial DNA recombination in a free-ranging Australian lizard. Biol Lett **3**(2):189-92.

Untoro E, Atmadja D, Pu C and Wu F (2009). Allele frequency of CODIS 13 in Indonesian population. Legal Med **11**(Supplement 1):S203-5.

Van der Walt J, Dementieva Y, Martin E, Scott W, Nicodemus K, Kroner C, Welsh-Bohmer K, Saunders A, Roses A, Small G, Schmechel D, Murali Doraiswamy P, Gilbert J, Haines J, Vance J and Pericak-Vance M (2004). Analysis of European mitochondrial haplogroups with Alzheimer disease risk. Neurosci Lett **365**(1):28-32.

Van Oven M and Kayser M (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mutat **30**(2):E386-94.

Venkatesh S, Deecaraman M, Kumar R, Shamsi M and Dada R (2009). Role of reactive oxygen species in the pathogenesis of mitochondrial DNA (mtDNA) mutations in male infertility. Indian J Med Res **129**(2):127-37.

Vergara C, Caraballo L, Mercado D, Jimenez S, Rojas W, Rafaels N, Hand T, Campbell M, Tsai Y, Gao L, Duque C, Lopez S, Bedoya G, Ruiz-Linares A and Barnes K (2009). African ancestry is associated with risk of asthma and high total serum IgE in a population from the Caribbean Coast of Colombia. Hum Genet **125**(5):565-79.

Vernesi C, Di Benedetto G, Caramelli D, Secchieri E, Simoni L, Katti E, Malaspina P, Novelletto A, Marin V and Barbujani G (2001). Genetic characterization of the body attributed to the evangelist Luke. Proc Natl Acad Sci U S A **98**(23):13460-3.

Vidrova V, Tesarova M, Trefilova E, Honzik T, Magner M and Zeman J (2008). Mitochondrial DNA haplogroups in the Czech population compared to other European countries. Hum Biol **80**(6):669-74.

Vigilant L, Stoneking M, Harpending H, Hawkes K and Wilson A (1991). African populations and the evolution of human mitochondrial DNA. Science **253**(5027):1503-7.

Wakeley J (1993). Subsititution rate variation among sites in hypervariable region I of human mitochondrial DNA. J Mol Evol **37**(6):613-23.

Wallace D, Brown M and Lott M (1999). Mitochondrial DNA variation in human evolution and disease. Gene **238**(1):211-30.

Walsh P, Metzger, DA, Higuchi, R (1991). Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. Biotechniques. **10**(4):506-13.

Watson E, Forster P, Richards M and Bandelt H (1997). Mitochondrial footprints of human expansions in Africa. Am J Hum Genet **61**(3):691-704.

Westermann B (2008). Molecular machinery of mitochondrial fusion and fission. J Biol Chem **16,283**(20):13501-5.

Westermann B (2009). Mitochondrial dynamics in model organisms: what yeasts, worms and flies have taught us about fusion and fission of mitochondria. Semin Cell Dev Biol **In Press, Corrected Proof**.

White D, Wolff J, Pierson M and Gemmell N (2008). Revealing the hidden complexities of mtDNA inheritance. Mol Ecol **17**(23):4925-42.

Wolff J and Gemmell N (2008). Lost in the zygote: the dilution of paternal mtDNA upon fertilization. Heredity **101**(5):429-34.

Wolpoff M, Hawks J and Caspari R (2000). Multiregional, not multiple origins. Am J Phys Anthropol **112**(1):129-36.

Wong H, Tang J, Budowle B, Allard M, Syn C, Tan-Siew W and Chow S (2007). Sequence polymorphism of the mitochondrial DNA hypervariable regions I and II in 205 Singapore Malays. Legal Med (Tokyo) **9**(1):33-7.

Yao Y, Kong Q, Bandelt H, Kivisild T and Zhang Y (2002). Phylogeographic differentiation of mitochondrial DNA in Han Chinese. Am J Hum Genet **70**(3):635-51.

Yong R, Gan L, Chang Y and Yap E (2007). Molecular characterization of a polymorphic 3-Mb deletion at chromosome Yp11.2 containing the AMELY locus in Singapore and Malaysia populations. Hum Genet **122**(3):237-49.

Zeviani M and Antozzi C (1997). Mitochondrial disorders. Mol Hum Reprod **3**(2):133-48.

Zhao X, Li N, Guo W, Hu X, Liu Z, Gong G, Wang A, Feng J and Wu C (2004). Further evidence for paternal inheritance of mitochondrial DNA in the sheep (Ovis aries). Heredity **93**(4):399-403.

# WEB RESOURCES

www1: http://www.mda.org/Publications/mitochondrial_myopathies.html
(Accessed on 25/03/2010)
www2: http://freepages.genealogy.rootsweb.ancestry.com/~ncscotts/GG/mt_DNA.htm
(Accessed on 25/03/2010)
www3: http://www.phylotree.org/mtDNA_seqs.htm
(Accessed on 26/03/2010)
www4: http://www.ncbi.nlm.nih.gov/nuccore/NC_012920
(Accessed on 26/03/2010)
www5: http://workbench.sdsc.edu/
(Accessed on 26/03/2010)
www6: http://www.phylotree.org/
(Accessed on 26/03/2010)
www7: http://www.mitomap.org/bin/view/MITOMAP/HaplogroupMarkers
(Accessed on 26/03/2010)
www8: http://www.genebase.com/doc/mtdnaHaplogroup Tree_Ref.pdf
(Accessed on 25/05/2010)
www9: http://www.mitomap.org/MITOMAP/MitoSeqs
(Accessed on 29/05/2010)

ETHICS COMMITTEE
FACULTY OF MEDICINE
ALEXANDRIA UNIVERSITY

## Letter of approval

This is to certify that the **research proposal** of the Thesis of Doctor Degree in Forensic Medicine provided by **Rania Abd El Moneim El Saiid** (MS) , Assistant Lecturer of Forensic Medicine & Clinical Toxicology , titled **" Development of Human DNA Profiling Techniques for Analysis of Egyptian Population"** is **accepted** by the Ethics Committee , Faculty of Medicine , Alexandria University **provided that the biological samples taken for the purpose of this study will not be used for other research except after taking the necessary approvals of the Ethics Committee and the candidate is responsible for storage of these biological samples .**

**Chief of Ethics Committee**

**Prof. Dr. Amal Mashali**

17 / 6 / 2006

A. Mashal

Vice Dean
Hind H