Thesis for Doctor of Philosophy (Ph.D.), 2010

The mitochondrial DNA heritage of the Baganda, Lugbara and Acholi from Uganda

Dan Isabirye

North-west University (Potchefstroom Campus), South Africa

The mitochondrial DNA heritage of the Baganda, Lugbara and Acholi from Uganda

BY

DAN ISABIRYE, M.Sc.

Thesis submitted for the degree Philosophiae Doctor (Ph.D.)
in Biochemistry at the North-West University

PROMOTOR: Dr Gordon Wayne Towers
Centre of Excellence for Nutrition, North-West University (Potchefstroom Campus)

September 2010

Die mitokondriale DNS erfenis van die Baganda, Lugbara en Acholi van Uganda

DEUR

DAN ISABIRYE, M.Sc.

Proefskrif voorgelê vir die graad Philosophiae Doctor (Ph.D.)
in Biochemie aan die Noordwes-Universiteit

PROMOTOR: Dokter Gordon Wayne Towers
Sentrum vir Uitnemendheid in Voeding, Noordwes-Universiteit (Potchefstroom Kampus)

September 2010

This thesis is dedicated to Dan Waiswa (my son), Rebecca Babirye and the Family of Mrs Alice & Rev. Can. Capt. Samuel Isabirye.

ABSTRACT

The mtDNA genetic relatedness between and within 13 Baganda, 14 Lugbara and 13 Acholi individuals from Uganda was investigated in this research program. The complete mtDNA sequences of the 40 Ugandan samples were established and a phylogeographic analysis of these sequences was conducted using both a Neighbour-Joining and a Maximum Parsimony tree together with a global sample of 387 African sequences. Prior to this study, only two complete and six partial mtDNA sequences of Ugandans had been established.

A total of 563 polymorphisms were determined of which 276 were synonymous, 75 were nonsynonymous, 26 were novel and 208 occurred in the control region. The Lugbara sequences clustered more closely with the Acholi sequences than the Baganda sequences within the Neighbour-Joining and Maximum Parsimony tree. A phylogeographic analysis of the sequences demonstrated that the Acholi and Lugbara individuals in this investigation originated from Southern Sudan while the Baganda samples had a diversified origin which comprised of the Niger-Congo basin, Ethiopia and Sudan. Furthermore, the clustering of the Ugandan sequences with sequences from African American and Hispanic individuals was evidence of slave trade involving the shipping of people from Uganda to North America.

It was intriguing that the deepest branch in the phylogeny was L5 (instead of L0) suggesting that the Khoi-San may not be the ancestral origin of anatomically modern man. There was increased resolution of macrohaplogroup L (especially for the small haplogroups) as new branches and nodes were formed in the tree. The results also demonstrated that East Africa was the origin and source of dispersal of numerous small macrohaplogroup L haplogroups. These mtDNA sequences from Baganda, Acholi and Lugbara individuals have a potential for forensic, nutrigenomic and pharmacogenomic application and will serve as useful references in assessment of mtDNA sequences in other Ugandan and East African populations.

Key words: mitochondrial DNA, phylogenetics, human migration, Uganda, macrohaplogroup L, African.

OPSOMMING

Die genetiese verwantskap tussen en binne die mtDNS van 13 Baganda, 14 Lugbara en 13 Acholi individue van Uganda, was ondersoek in hierdie navorsings projek. Die volledige mtDNS volgorde van die 40 Ugandan deelnemers was vasgestel en ‘n filogeografiese analise van hierdie volgordes saam met ‘n wereldwye steekproef van 387 Afrikaanse volgordes was onderneem deur middel van “Neighbour-Joining” en “Maximum Parsimony” bome te teken. Voor hierdie ondersoek onderneem was, was daar net twee volledige en ses gedeeltelike mtDNS volgordes van Ugandans beskikbaar.

‘n Totaal van 563 polimorfismes was bevind, waarvan 276 sinoniem , 75 nie-sinoniem en 26 nuwe veranderinge was, en 208 in die beheer gebied van die mtDNS voor gekom het. Die Lugbara volgordes het nader aan die Acholi volgordes gekluster as die Baganda volgordes in die finale “Neighbour-Joining” en “Maximum Parsimony” bome. ‘n Filogeographiese analise van die volgordes het gewys dat die Acholi en Lugbara individue in hierdie ondersoek, van Suidelike Sudan ontstaan het, terwyl die Baganda individue ‘n afwisselende voorsprong gehad het wat uit die Niger-Kongo stroomgebied, Ethiopië en Sudan saamgestel is. Die klustering van die Ugandan volgordes met die van die Afrikaner-Amerikaner en Hispaniese individue was ‘n bewys van moontlike slawehandel tussen Uganda en Noord Amerika.

Dit was interesant, die diepste tak van die fillogenie was L5 (in stede van L0) wat demonstreer dat die Khoi-San miskien nie die voorvaderlike oorsprong van anatomiese moderne man was nie. Daar was groter resolusie van Makrohaplogroep L (hoofsaaklik die klein haplogroepe) omdat nuwe takke en knoppe in die bome gevorm het. Die uitslae van hierdie ondersoek het gewys dat Oos Afrika die oorsprong en die bron van verspreiding van meetalige klein Makrohaplogroep L haplogroepe was. Hierdie mtDNS volgordes van die Baganda, Acholi en Lugbara individue het ‘n potensiaal vir forensiese, nutrigenomiese en farmakogenomiese gebruik en sal as nuttige verwysings in die skatting van mtDNA volgordes in ander Ugandan en Oos Afrikaanse bevolkings gebruik kan word.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND SYMBOLS

Abbreviations and symbols are listed in alphabetical order.

List of symbols

&	and
#	number
β	beta
γ	gamma
°C	degrees Centigrade
%	percent
μ	micro: 10^{-6}
12S	12S ribosomal RNA
16S	16S ribosomal RNA

List of abbreviations

A or a	adenine
A_{260}/A_{280}	ratio of absorbance at 260 nm to 280 nm, measure of DNA purity
AD	*Anno Domino*
Af	African Americans
Ala	alanine
An	Angola
Arg	arginine
Asn	asparagine
Asp	aspartic acid
ATP	adenosine triphosphate
ATP6	gene encoding ATP synthase subunit 6
ATP8	gene encoding ATP synthase subunit 8
ATPase 6	gene encoding ATP synthase subunit 6
ATPase 8	gene encoding ATP synthase subunit 8
avg	average
Ba	Bakaka tribe from Cameroon
bp	base pairs
Bz	Brazil
C or c	cytosine
ca.	circa: approximately
Ca	Cameroon
CAR	Central African Republic
Cb	Cabinda
cm	centimetre
CO I – III	cytochrome oxidase subunits I to III
CoQ	coenzyme Q
COX	cytochrome oxidase
CRS	Cambridge Reference Sequence
CTP	cytidine triphosphate

Cys	cysteine
cyt b	cytochrome b
Da	Daba people from Cameroon
ddH_2O	double distilled water
dGTP	2'-deoxyguanosine 5'-triphosphate
dITP	2'-deoxyinosine 5'- triphosphate
Dk	Dokota tribe from Tanzania
D-loop	displacement loop
DNA	deoxyribonucleic acid
dNTPs	deoxyribonucleotide triphosphates
DRC	The Democratic Republic of Congo
dsDNA	double stranded DNA
EDTA	ethylenediamine tetraacetic acid: $C_{10}H_{16}N_2O_8$
e.g.	*exempli gratia:* for example
Eg	Egypt
Et	Ethiopia
et al.	*et alii*: Latin for "and others"
EtBr	ethidium bromide: $C_{21}H_{20}BrN_3$
Ew	Ewondo tribe from Cameroon
F	forward primer
Fa	Falis tribe from Cameroon
$FADH_2$	reduced flavin adenine dinucleotide
Fg	Fang tribe from Gabon
g	acceleration due to gravity
g	gram
G or g	guanine
gDNA	genomic DNA
GI	GenInfo Identifier sequence Identification number
Gl	Galoa tribe of Gabon
Gln	glutamine
Glu	glutamic acid
Gly	glycine
His	histidine
Hs	Hispanic
HSP	heavy strand promoter
H-strand	heavy strand of mtDNA
HV	hypervariable region
Hz	Hadza tribe from Tanzania
Ile	isoleucine
Is	Israel
Iw	Iraqw tribe from Tanzania
k years	kilo years implying 1,000 years
kb	kilo base pairs
KCl	potassium chloride
kDa	Kilodaltons
Ki	Kikuyu tribe from Kenya
Leu	leucine
LSP	light strand promoter
L-strand	light strand of mtDNA
Lys	lysine
M	molar
Ma	Mandara tribe from Cameroon
MEGA	Molecular Evolutionary Genetic Analysis Software

Met	methionine
mg	milligrams
$MgCl_2$	magnesium chloride
min	minutes
Mk	Makina tribe of Gabon
mL	millilitres
mM	millimolar
MM	molecular weight marker
Mo	Berber tribe from Morocco
MP	Maximum Parsimony
mRNAs	messenger RNAs
Mt	mount or mountain
mt	mitochondrial
mtDNA	mitochondrial DNA
mTERF	mitochondrial termination factor
mtRNApol	mitochondrial RNA polymerase
mtTFA	mitochondrial transcription factor A
N	any of the four bases in DNA sequence
Na	North America
NADH	nicotinamide adenine dinucleotide (reduced form)
NADH-Q	NADH coenzyme Q oxidoreductase complex
Na_2EDTA	di-sodium ethylenediamine tetraacetic acid
NCBI	National Centre for Biotechnology Information
NC-IUB	Nomenclature Committee of the International Union of Biochemistry
ND1-6	NADH-Q oxidoreductase subunits 1 to 6
nDNA	nuclear DNA
Neg	negative control sample
ng	nanogram
NJ	Neighbour-Joining
nm	nanometre
np	nucleotide position
NRF	nuclear respiratory factor
Nu	Nubia tribe of Sudan or Egypt
O_H	heavy strand origin of replication
O_L	light strand origin of replication
PCR	polymerase chain reaction
pH	potential of hydrogen ions
PHYLIP	Phylogeny Inference Package
Pi	nucleotide diversity
pI	Isoelectric point
Phe	phenylalanine
Po	Podowkos tribe of Cameroon
Pro	proline
Q	ubiquinone (coenzyme Q or CoQ)
R	reverse primer
rCRS	revised Cambridge reference sequence
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
ROS	reactive oxygen species
rRNA	ribosomal RNA
s	seconds
S	Svedberg unit
SDS	sodium dodecyl sulphate

Ser	serine
Sk	Sukuma tribe from Tanzania
SNP	single nucleotide polymorphism
Su	Sudan
Sy	Syria
Syn	synonymous
T or t	thymine
T_a	annealing temperature
Taq	thermostable enzyme isolated from *Thermus aquaticus* BM, recombinant (*Escherichia coli*)
TBE	89.15 mM Tris®[1] (pH 8.0), 88.95 mM boric acid, 2.498 mM Na_2EDTA
TCA	tricarboxylic acid cycle
TFB1M	mitochondrial transcription factor B1
TFB2M	mitochondrial transcription factor B2
Thr	threonine
Tk	Turkana tribe from Kenya
T_m	melting temperature
TMRCA	time to the most recent common ancestor
Tris®	tris(hydroxymethyl)aminomethane:2-amino-2-(hydroxymethyl)-1, 3-propanediol: $C_4H_{11}NO_3$
tRNA	transfer RNA
$tRNA^{Ala}$	transfer RNA for alanine
$tRNA^{Arg}$	transfer RNA for arginine
$tRNA^{Asn}$	transfer RNA for asparagine
$tRNA^{Gly}$	transfer RNA for glycine
$tRNA^{Leu(CUN)}$	transfer RNA for leucine specifically recognising the codon CUN
$tRNA^{Leu(UUR)}$	transfer RNA for leucine specifically recognising the codon UUR
$tRNA^{Lys}$	transfer RNA for lysine
$tRNA^{Met}$	transfer RNA for methionine
$tRNA^{Ser(AGY)}$	transfer RNA for serine specifically recognising the codon AGY
$tRNA^{Thr}$	transfer RNA for threonine
Trp	tryptophan
Tz	Tanzania
U	uracil
Ug or UG	Uganda
USA	United States of America
UV	ultraviolet light
V	volts
Val	valine
w/v	weight per volume
YBP	years before present

[1] Tris® is the registered trademark of the United States Biochemical Corporation, Cleveland, OH, USA.

LIST OF FIGURES

Figure no. **Title of Figure** **Page no.**

Figure no. **Title of Figure** **Page no.**

LIST OF TABLES

ACKNOWLEDGEMENT

This study was accomplished by the help of so many people and institutions, who financially, materially, socially or technically facilitated me. I am grateful to **Dr. Gordon Wayne Towers**, Centre of Excellence for Nutrition (CEN), North-West University (Potchefstroom Campus), South Africa, my supervisor, for the critical review of this research, from the time of development of the proposal through the experimental protocols for the lab work and write up of the thesis. Dr. Gordon Wayne Towers generated the pair wise differences amongst the samples used in this research and together with **Dr. J. Poole**, Centre for Molecular and Mitochondrial Medicine and Genetics, University of California, Irvine (UCI), USA and **Michelle Koekemoer**, a PhD student at the Centre for Genome Research (CGR), North-West University (Potchefstroom Campus), constructed the Neighbour-Joining and Maximum Parsimony trees. Dr. Gordon Wayne Towers also helped me a lot in the formatting of this thesis and secured me some reference articles and textbooks.

I am indebted to **Prof. Antonel Olckers**, the Director, DNAbiotec (Pty) Ltd, Lynnwood, Pretoria, South Africa, for the comments made on the experimental protocols and her review of this thesis during most of the stages when it was being written and through whom I was able to secure admission into CGR. **Prof**. Antonel Olckers also helped me secure some reference articles for this research. I am grateful to **Paul Olckers**, her husband, for the visits he made to me while in the lab and for the moral support and encouragement to work hard for long hours.

I am indebted to **Prof. Marlene Viljoen**, the Dean, Faculty of Healthy Sciences, North-West University (Potchefstroom Campus), South Africa; **Prof. Kobus Pienaar**, the Dean, Faculty of Natural Sciences, North-West University (Potchefstroom Campus), South Africa; **Ms Linda Grimbeek**, Administrator, Faculty of Healthy Sciences, North-West University (Potchefstroom Campus), South Africa; **Ms Monique van Deventer**, Registrar, Faculty of Natural Sciences, North-West University (Potchefstroom Campus), South Africa, for their excellent administration that enabled me to remain focussed in the studies. I could see it from a far that they wanted me to have a smooth ride through the studies. These administrators helped me overcome great challenges in the studies that had yoked me for a long time. When times became hard for me to finish the studies from South Africa, Prof. Marlene Viljoen readily granted me permission to finish them from Makerere University, Kampala, Uganda, while at home.

I am grateful to **CGR**, for funding the laboratory costs of this research and ensuring that chemicals were available on time. CGR also provided me with a bursary for tuition fee and upkeep for the first two years of this study. I am indebted to the **Ferdinand Postma library of North-West University and its staff** for facilitating my search for literature, and through which I was able to access **Sabinet** and **SACat**, which are links to all other South African Universities' libraries for scarce reference materials. Thanks to **DNAbiotec (Pty) Ltd**, for providing the infrastructure where this research was conducted. This study was conducted in the laboratories of DNAbiotec (Pty) Ltd and a number of DNAbiotec (Pty) Ltd equipment were used.

I am indebted to **Michelle Koekemoer**, for providing me with a downloaded sample of the African complete and coding region mtDNA sequences that were used in the phylogenetic

and phylogeographic analysis in the Neighbour-Joining and Maximum Parsimony trees. There were a number of discussions that I had with Michelle Koekemoer that helped me a lot in overcoming challenges in the study.

I am grateful to **Dr. Annelize van der Merwe**, Senior Research Officer, DNAbiotec (Pty) Ltd, who ensured that the lab supplies were readily available and when finished were ordered on time for the smooth running of the study. Her readiness to help me was a great inspiration in these studies.

I am indebted to the following who were colleagues and students at CGR: **Marco Alessandrini, Tumi Semete, Michelle Koekemoer, Desiré Dalton, Jake Darby, Scheán Babst, Delia Tanner** and **Estie Kotzé.** Through interactions with them I was able to make well thought out strategies to pursue the experimental parts of this research.

I am indebted to **Tshireletso Mataboge, Chricentia Khumalo, Leonard Mdluli, Kenneth Nkadimeng** and **Anri Raath**, research assistants at DNAbiotec (Pty) Ltd, for the assistance they rendered to me while pursuing the laboratory part of this study. The research assistants ensured that the lab was well organised, encouraged me to work hard, would readily help me and were great friends of mine. Thanks to **Martha Sebogoli** for maintaining a clean environment conducive for studying at CGR and DNAbiotec. (Pty) Ltd.

Thanks to **Prof. Douglas Wallace**, Center for Molecular and Mitochondrial Medicine and Genetics (MAMMAG), University of California, Irvine Irvine California, USA, for permission to use his figure on human mtDNA migrations. Thanks to **Dr Antonio Salas**, Unidad de Genética, Instituto de Medicina Legal, Universidad de Santiago de Compostela, A Coruña, Galicia, Spain, for permission to use the figure highlighting the phylogenetics and phylogeography of haplogroup L4g.

I am grateful to **Leon Breytenbach, Lois Breytenbach, Kevin Prozesky, Lyn Prozesky, Paul van Loosen, Shireen van Loosen, Annekie Brink** and **Tinah Nakajju Kyobe**, cell group members of Trinity Church, Lynnwood, Pretoria, for the social and spiritual support that they provided me. We studied together the Gospel of Mark, the Purpose driven life by Rick Warren and the book of Revelation, and the many discussions left a life long impact on me. The cell group members helped me obtain useful reading materials for spiritual growth and we had with them a relaxing and rewarding tour of the Pilanesberg National Park, Pilanesberg, South Africa. The discussions on, acceptance, forgiveness, responsibility, self control, perseverance, honesty, compassion and grace, as depicted in the Heartlines films (www.heartlines.org), were a vital life skill that helped me overcome many challenges that came my way, in a foreign country and will remain a vital tool for life. When I returned to Uganda, my stay in Uganda has been characterised by feelings of a partial vacuum in my life because of the benefits I derived from this group that I still long for. Thanks to **Hlengiwe Sehlapelo** who first introduced me to the Trinity Church community and who for a long time transported me from church back to home and for helping me cope up with social challenges that came my way. Thank you to **Nancy Herbert** of Trinity Church for helping me secure some references for this study, at such an elderly age. I am grateful to **Rev. Father Julian Kok**, the Pastor of Trinity Church, for the sermons, other Christian teachings and for the social and moral support that ensured that I lived happily in a foreign country. I am grateful to the entire **Trinity Church community** for their love, care and support that they rendered to me and through whose activities, Trinity Church was an inspiring community to associate with.

Thanks to **Dr Damalie Alirwa, Dr Lazarus Nyende, Geoffrey Kakaire, Moses Bukenya, Musa Ngobi Waiswa, Aggrey Muyinda, Leonard Kyeyune, Haliimah Nalyazi,**

Emmanuela Oryem, Florence Yiga and Maxima Assimwe, part of the Ugandan community I lived with in South Africa, for their moral support. Thank you to **Mbali,** Pretoria, South Africa, for rendering to me means of transport that I used in the latter part of this study. Thanks to **Colin Webster**, the President of the Mind Sports of South Africa (MSSA), for exposing me to a different environment in South Africa that enabled me to relax from the busy studies and helped secure a number of medals in International 10 x 10 draughts and Anglo-American draughts and through whom I was able to learn and introduce Morabaraba board game to the Ugandan community in Uganda.

I was not able to seek the opinions of **Dan Waiswa**, my son, as he was too young, but he was always inquisitive wanting to know what I was doing, opening books I was reading and asking to know what was attracting him such as pictures and a few words he could read from the pictures and this greatly refreshed my mind and enabled me to work for long hours. Thanks to **Rebecca Babirye**, my wife, for her prayers and keeping Dan safe. I am grateful to my parents, **Mrs Alice and Rev. Can. Capt. Samuel Isabirye**; brothers, **Joseph Musoke, James Nkumbo, Ben Kati, John Koosi, Moses Waiswa** and **Paul Butono; sisters, Sarah Nabirye, Esther Timutenda, Milly Namukose** and **Ever Nalumansi** and **Nephew, Cate Namuwaya** for their prayers and moral support. Great thanks to **Flavia Namuyomba** and **Ritah Nakibuule,** Kampala, Uganda, for the support and encouragement while pursuing these studies. I am grateful to **Rev. Can. Moses Isabirye**, Uganda Christian University, Kampala, Uganda, for the prayers, moral support and encouragement he provided. I appreciated the encouragement by **The Most Rt. Rev. Bishop Cyprian Bamwoze**, retired bishop of Busoga Diocese, Jinja, Uganda which helped me to persevere to the very end of the studies.

Thank you to **Dr. Joseph Hawumba**, Makerere University, Biochemistry Dept., Kampala, Uganda, through whom I got connected to Dr. Antonel Olckers who secured me a vacancy at North – West University (Potchefstroom Campus). Joseph Hawumba also helped me secure some clearances in Uganda and made many useful comments to this study and helped me overcome some of the challenges that I faced. I am grateful to **Prossy Namande**, Makerere University, Biochemistry Dept., Kampala, Uganda, **Simon Odokorach, Ezati Timothy, Immaculate Adrako and Daniel Nsimbe**, of Makerere University, **Francis Oryema**, of Bugema Adventist University, Kampala, Uganda for helping me with securing informed consent from Ugandan volunteers who donated blood used in this study. I am grateful to, **Posiano Mayambala**, Mengo Hospital, Kampala, the technician who drew blood from the study volunteers. I am grateful to the **volunteers** who consented to donating blood used in this study. I am grateful to **Rhona Baingana, Agnes Nandutu, Peter Vuzi, Apollo Balyeidhusa, John Enyaru, Denis Matovu Kasozi, Sam Wamutu, John Omara, James Ojambo, Fred Juuko, Pacific Owilla, Paul Kakande, Beatrice Nabisuubi** and **Florence Kabahenda**, members of Biochemistry Dept., Makerere University for encouraging me to work hard and for their prayers. I am grateful to **Dr Agnes Nandutu** and **Prof. John Enyaru** for taking up some of my teaching load at Makerere University so that I get through with the write up of the thesis.

I am grateful to Makerere University Staff Development Committee, for meeting part of the upkeep and tuition fee, for this study. I am grateful to **Prof. Nelson Sewankambo**, the Principal of Makerere University College of Healthy Sciences and **Prof. Sam Luboga**, Makerere University College of Healthy Sciences, through whom I was able to secure funding from Makerere University Staff Development Committee to supplement my upkeep while in South Africa. I am grateful to **Stephen Kateega** and **Kevin Nabiryo**, of Makerere University Human Resource Dept., for the cooperation they accorded me whenever I applied for Makerere University Staff Development funds, which had to be done on an annual basis. I am grateful to the Makerere University, Board of Research and

Publications, for meeting the costs of blood sample collection and printing of this thesis. I am grateful to **Prof. Livingstone Luboobi**, the former Vice Chancellor of Makerere University, by whose help I was able to secure funding from the Makerere University Board of Research and Publications, for collecting blood from the Ugandan volunteers. I am grateful to the Bibliographic Library of Makerere University for greatly aiding my literature search.

I am grateful to the Almighty for the time and wisdom He provided me with to plan, think and ponder over this study. Through your Son, you bridged for us the great divide and hatred which is ubiquitous on planet earth and by his blood, you saved us. Just like you favoured Noah, and Joseph the son of Jacob, it has been by your grace that I have come this far; just like you provided the eagle with strong wings to glide through storms, you took me through challenges. In the same fashion that you protected Daniel in a den of lions, you protected me in times of need and kept me safe during the times when I walked through the shadows of death. I experienced compassion in a foreign land that was beyond mere coincidences; in Trinity Church and by the various administrators of North-West University and Makerere University and by the family and friends, which I needed to get through unforeseen difficulties. For many times I registered beyond the normal deadlines but you still secured funding for my study. I did not respond well to South African drugs and diets but you ensured that the drugs I got from Uganda on an annual basis would see me through the year. Thanks too for the many nice moments and the learning experience that you provided. I was able to interact with all sorts of people in South Africa ranging from those who were so good and kind than I had ever seen before to those who were so cruel than I had ever met in life. It is my prayer that these experiences you have let me go through will build my character, generate endurance and help me serve the community I live in to the glory of your name.

CHAPTER ONE

Introduction

Mitochondrial deoxyribonucleic acid (mtDNA) is highly variable (De Benedictis *et al*., 1999), maternally inherited (Reich and Luck, 1966; Giles *et al*., 1980), does not recombine (Aquadro and Greenberg, 1983; Elson *et al*., 2001) and its molecules are present in great numbers in cells (Bogenhagen and Clayton, 1974; Luft, 1994; Wallace, 1995; Wallace *et al*., 1999; Fernandez-Silva *et al*., 2003). As humans evolved, and as our bodies interacted with different climates and diets on each of the continents, there was genetic drift, purifying selection and adaptive selection that resulted in these naturally occurring variants (Bogin and Rios, 2003; Mishmar *et al*., 2003; Ruiz-Pesini *et al*., 2004). These variations or polymorphisms are an inheritable indelible record of our past evolutionary history and are thus important in deciphering human evolution and origins (Brown *et al*., 1992; Wallace, 1995; Jorde *et al*., 1998; Wallace *et al*., 1999; Adcock *et al*., 2001; Coskun *et al*., 2003). Variations in the mtDNA sequences have been used to study major human demographic factors such as population migrations, bottlenecks and expansions, and the significance of mtDNA mutations in human disease (Brown *et al*., 1992; Wallace, 1995; Torroni *et al*., 1996; Jorde *et al*., 1998; Ruiz-Pesini *et al*., 2000; González *et al*., 2006).

The mutation rate of genes in mtDNA is higher than that of nuclear genes (Brown *et al*., 1979). Mutations that appreciably compromise mitochondrial energy production are generally lost due to negative selection (Ruizi-Pesini *et al*., 2004) but the mutations that have near-neutral or the neutral effects are not lost and it is such polymorphisms that accumulate over time (Wallace *et al*., 1995). Due to this process of bio-molecular differentiation taking place mostly during and after human colonization of the various continents, the populations of the world are currently divided into regional-specific mtDNA haplotypes, restricted to specific geographical areas (Mishmar *et al*., 2003). As these mtDNA variants tend to be restricted to specific geographic areas, thorough characterisation of mtDNA sequences of population groups can subsequently provide major insights about human origin and the demographic processes through which modern human populations have been shaped (Torroni *et al*., 1996; Herrnstadt *et al*., 2002; González *et al*., 2006). Consequently, mtDNA analysis has been an invaluable tool in

quantifying the evolutionary relationships of human ethnic groups (Brown, 1980; Denaro *et al*., 1981; Ballinger *et al*., 1992; Horai *et al*., 1993). The mtDNA sequences of a given population have been used to unravel the past records of the maternal line of that population (Jorde *et al*., 1998; Wallace *et al*., 1999), while Y-chromosome data have been used to recapitulate the historical events along the paternal line (Tarazona-Santos *et al*., 2001; Cruciani *et al*., 2002). Subsequently, mtDNA and Y-chromosome genetic data have been used to elucidate the relative contribution of females and males in shaping the history of humans (Passarino *et al*., 1998; Seielstad *et al*., 1998; Kalaydjieva *et al*., 2001; Wilson *et al*., 2001; Redd *et al*., 2002; Shen *et al*., 2004; Tambets, 2004).

Body size and the proportions of the human body parts have been used as indicators of one's ethnicity, nutritional history, socioeconomic status and a measure of adaptation to temperature of that individual (Bogin and Rios, 2003). Geographical distribution (Vigilant *et al*., 1989), similarities in languages (Raymond, 2005), skin colour (Parra *et al*., 2004) and surnames (Sykes and Irven, 2000; Jobling, 2001; Zei *et al*., 2003) have also been used to infer phylogenetic relationships among humans. However, the highest possible resolution for evolutionary analysis of populations is provided by nucleotide sequences (Vigilant *et al*., 1989) since they are the building blocks of the basic units of inheritance of any organism containing the instructions on the way each functional part was assembled and made to operate. Furthermore, the uniparental inheritance of mtDNA (Reich and Luck, 1966; Giles *et al*., 1980) and the relatively high mutation rate, make it ideal for evolutionary analysis of populations (Brown *et al*., 1979).

This study was an attempt to elucidate the genetic relatedness within and between the Baganda (Ganda), Acholi and Lugbara populations from Uganda and to outline the ancestral populations and ethnic groups that merit additional investigation. The study also assessed how the three Ugandan populations are related to other African populations. One population (Baganda) was relatively socially and geographically divergent from the other two populations (Acholi and Lugbara). This study has been the first to elucidate the phylogenetic relationship within and between Ugandan populations. The first mtDNA sequences characterising at least a Ugandan population have been established. Prior to this study, only isolated cases of two complete mtDNA sequences (Horai *et al*., 1995; Macaulay *et al*., 2005) and six partial control region sequences (Salas *et al*., 2004b) of Ugandans had been analysed. Furthermore, no archaeological data exists for the Ugandan population while Y-chromosome data for the country is available only for the

Karamojong Nilotes (Gomes *et al*., 2009). Moreover, East Africa is the most likely place of origin of modern humans (Were and Wilson, 1984; Knight *et al*., 2003; Liu *et al*. 2006; Gonder *et al*., 2007), therefore a characterisation of East African populations is useful in providing information about the human ancestral populations at the root of the global human phylogenetic tree (Gonder *et al*., 2007).

The results from the complete mtDNA sequencing of 13 Baganda, 14 Lugbara and 13 Acholi from Uganda belonging to Bantu, Moru-Madi and Nilotic ethnic groups respectively are reported here. The significance of the mtDNA sequences with respect to the origin of the tribes under study was assessed and the position of the identified haplotypes in the global phylogenetic tree was inferred.

The physical and genetic structure of the mitochondrion has been explained in Chapter Two. The mtDNA phylogenetic relationship amongst African populations, theories on the origin of man, and the effect of migrations of people in different continents as deciphered mainly from oral traditions and mtDNA polymorphisms have been explored in Chapter Three.

In Chapter Four, the materials and methods used in this study have been described. The mtDNA genome sequences of the 40 Ugandan individuals were determined by the technique of automated sequencing. The mtDNA was amplified in fragments ranging from 2049 to 3264 base pairs (bp) with neighbouring fragments exhibiting a 19 – 54% overlap. The primers for amplification of the polymerase chain reaction (PCR) fragments were 20 - 22 nucleotides in length and were designed according to the procedure by Maca-Meyer *et al*. (2001) to be complementary to sections of the light and heavy strands of the Revised Cambridge Reference Sequence (rCRS) of mtDNA of Andrews *et al*. (1999). The amplified PCR products were in each instance visualised as single fragments on ethidium bromide-stained agarose gels. The results of this investigation are presented and discussed in Chapter Five and the conclusions are drawn in Chapter Six. The polymorphisms were established with comparison to the rCRS as the reference mtDNA sequence (Andrews *et al*., 1999). The numerous citations made have been referenced in Chapter Seven and the relevant appendices are attached to the thesis after the reference chapter.

The study identified that the Acholi samples clustered more closely with the Lugbara samples than the Baganda samples. The clustering pattern also confirmed the origin of the Acholi and Lugbara to be from Sudan while the Baganda samples seemed to have had a diversified origin. The clustering pattern also demonstrated evidence for possible slave trade from Uganda to America and that East Africa was an important region of dispersal of many macrohaplogroup L small haplogroups. The deepest branch in the phylogeny happened to be that of Haplogroup L5 instead of Haplogroup L0. Further investigations involving mtDNA and Y-chromosome sequencing of more samples from the Acholi and Lugbara of Uganda together with samples from Khoi-San individuals and other East African populations needs to be undertaken, to establish the oldest lineage in the world as these have been previously reported to be the most ancestral of all world populations (Cruciani *et al*., 2002; Krings *et al*., 2003; Gonder *et al*., 2007; Behar *et al*., 2008). It is submitted that the Baganda, Acholi and Lugbara individuals used in this study have distinct mtDNA sequences as not a single haplotype was shared between any two of these three tribes.

CHAPTER TWO

The physical and genetic structure of the mitochondrion

The mitochondrion is a double-membraned organelle (Bauer *et al*., 1999; Berg *et al*., 2002; Gibson, 2005) that plays a role in a number of functions in the body (Gibson, 2005) and is the primary site where adenosine triphosphate (ATP), the immediate source of free energy in the body is manufactured (Wallace *et al*., 1999). The mitochondrion is estimated to consist of 1,000 - 2,000 proteins but so far only 685 have been identified and well characterised (Bauer *et al*., 1999; Taylor *et al*., 2003; Gibson, 2005). The identification of all the proteins localised to the mitochondrion will enhance a better understanding of its cellular functions and the role it plays in the regulation of complex processes such as aging and apoptosis (Gibson, 2005). The mitochondrion originated from a prokaryotic cell via endosymbiosis with a eukaryotic cell, and subsequently lost most of its genes to the nuclear chromosomes (Borst, 1977). The high number of mtDNA molecules in cells (Bogenhagen and Clayton, 1974) and its unique genetic features have made it a highly useful chromosome for the study of human population genetics and evolutionary processes (Brown *et al*., 1979; Brown and Simpson, 1982; Brown *et al*., 1982).

The formation of the mitochondrion was a huge initial evolutionary investment with long term rewards. The initiation of eukaryote formation involved the fusion of anaerobic archaeobacteria as the host with a proteobacteria as the competently respiring symbiont (Karlin *et al*., 1999; Gray *et al*., 2001). Due to increasingly oxidising atmospheric conditions of the earth, prokaryotes adapted quickly to the effects of the toxic oxygen to maintain normal cell function (Schon, 1993; Kurland and Andersson, 2000). Through endosymbiosis, the prokaryote (endosymbiont) enabled the complex eukaryote (host) to trap the toxic oxygen into forms that were benign and useful (Schon, 1993; Kurland and Andersson, 2000). The endosymbiont (prokaryote) detoxified the oxygen that was lethal to the host and provided useful by-products, while the host (eukaryote) provided the prokaryote with a free source of food and safety from external physical damage or mechanical injury (Schon, 1993; Kurland and Andersson, 2000). Subsequently, the bacteria transformed into the proto-mitochondrion as the host's viability became dependant on aerobic metabolism (Schon, 1993; Kurland and Andersson, 2000). There was a loss of

99% of genes from the proto-mitochondrion as it no longer needed to synthesise most of its metabolites since much of its nutrient requirements were being met by the host (Schon, 1993; Bauer *et al*., 1999; Kurland and Andersson, 2000; Gray *et al*., 2001; Wiedemann *et al*., 2004). Eventually, there was a change in the proto-mitochondrial genetic code due to the accumulation of mutations and consequently it lost the ability to translate any of the proteins needed for its viability (Kurland and Andersson, 2000). These transformations confined the prokaryote to the interior of the cells of eukaryotes until it eventually evolved into the mitochondrion (Kurland and Andersson, 2000).

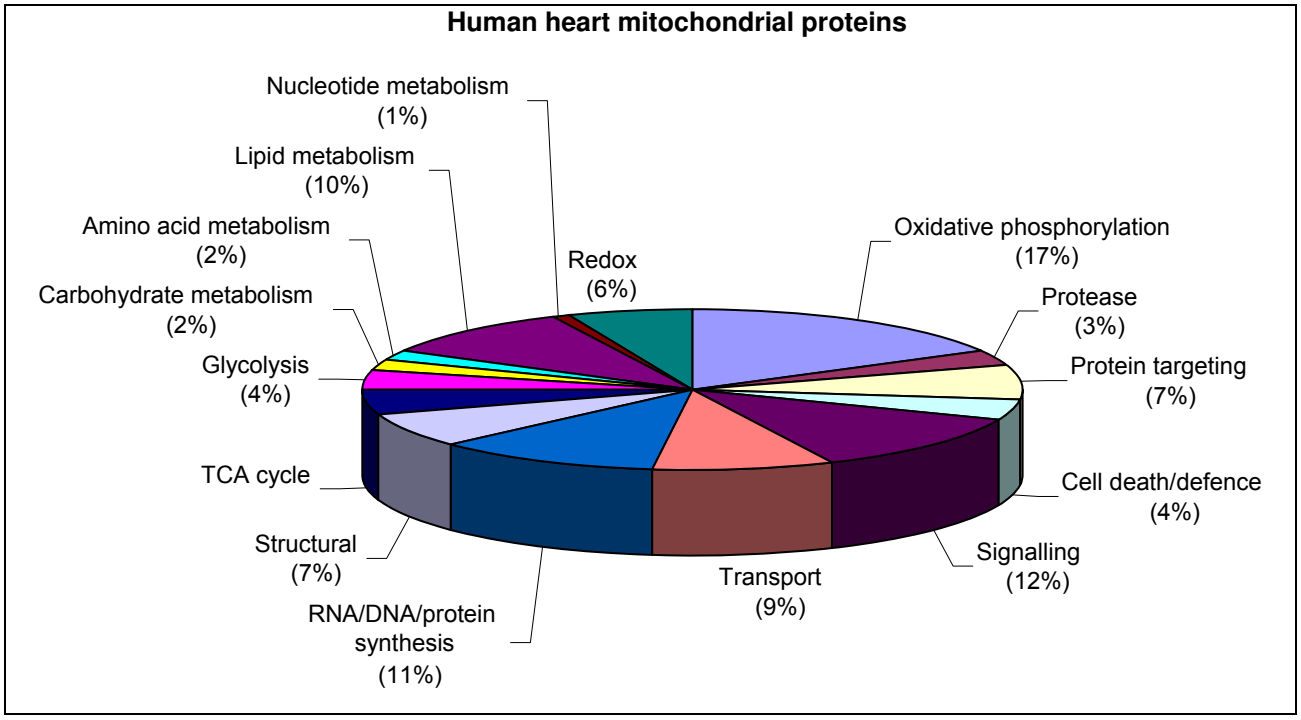
Only 37 genes are currently retained by the mitochondrion of which 13 are polypeptide coding genes while the remaining 24 are associated with mitochondrial ribosomal translation (Wallace, 1995; Kurland and Andersson, 2000). Thus, during the course of evolution, the mitochondrion exported the majority of its genes to the nucleus, and therefore, it currently has to re-import the functional proteins so that the organelle regains the products (Martin and Hermann, 1998; Kurland and Andersson, 2000; Gray *et al*., 2001; Wiedemann *et al*., 2004).

The main function of the mitochondrion is to produce ATP via oxidative phosphorylation (Scholte, 1988; Senior, 1988) with the aid of a proton-motive force (Mitchell, 1961; Berg *et al*., 2002). Oxidative phosphorylation consumes about 90% of the oxygen used by a eukaryotic cell and so the amount of mitochondria contained in a cell or tissue can serve as an index of the metabolic rate for the cell or tissue (Camougrand and Rigoulet, 2001). Other catabolic processes that take place in the mitochondrion include the conversion of pyruvate into lactate (Berg *et al*., 2002), the citric acid cycle (Berg *et al*., 2002) and the β-oxidation of fatty acids (Berg *et al*., 2002). The enzyme (pyruvate carboxylase) that catalyses the gluconeogenic reaction that converts pyruvate into oxaloacetate is also located in the mitochondrion (Berg *et al*., 2002).

Mitochondria serve as reservoirs for the storage of calcium ions and act as buffer sinks to avoid calcium overload (Ichas *et al*., 1997; Brustovetsky and Dubinsky, 2000). Mitochondria also play a role in apoptosis (Zamzam *et al*., 1996; Budd and Nicholls, 1998; Susin *et al*., 1998; Bauer *et al*., 1999; Nomura *et al*., 1999; Ferri *et al*., 2000; Harris and Thompson, 2000; Wang, 2001; Aoki *et al*., 2002; Turrens, 2003), excitotoxicity of neurons through glutamate mediation (Budd and Nicholls, 1998), cell signalling (Gibson, 2005), thermogenesis (Wallace, 1994; Mortola and Naso, 1998; Wagner *et al*., 1998; Wallace,

1999; Rippe *et al*., 2000; Berg *et al*., 2002; Zaninovich *et al*., 2002; Minorsky, 2003; Seymour, 2004; Fontanillas *et al*., 2005) and regulation of the redox state of cells (Scholte, 1988; Budd and Nicholls, 1998; Taylor *et al*., 2003; Gibson, 2005). This vital organelle also participates in the urea cycle and the biosynthesis of porphyrins, steroids and pyrimidines (Naviaux, 1997; Bauer *et al*., 1999; Taylor *et al*., 2003; Gibson, 2005). Figure 2.1 below is a pie-chart indicating the proportion of heart mitochondrial proteins performing different roles.

Figure 2.1 Distribution and proportion of functionally classified mitochondrial proteins



A functional classification of human heart mitochondrial proteins. Glycolysis, although a cytosolic process, was included due to its interaction with many of the processes that take place in the mitochondria. Adapted from Gibson (2005).

The distribution of the mitochondria in cells and tissues depends on the nature and energy demands of the respective cells and tissues (Naviaux, 1997; Díez-Sánchez *et al*., 2003; Piccoli *et al*., 2004). Mitochondria are numerous in tissues such as spermatozoa, flight muscles of birds, cardiac muscle and the cone cells of the eye (Naviaux, 1997; Berg *et al*., 2002; Díez-Sánchez *et al*., 2003). There are 1,000 - 2,000 mitochondria in a single liver cell, occupying ca. 20% of its total volume (Piccolli *et al*., 2004). Progressive human spermatozoa contain 700 mtDNA copies per cell while the non-progressive types contain 1,200 (Díez-Sánchez *et al*., 2003). Most of the nucleated cells in the human body contain 500 to 2,000 mitochondria, while in platelets this ranges from two to six mitochondria, while mature red blood cells have none (Naviaux, 1997).

2.1 STRUCTURE OF THE MITOCHONDRION

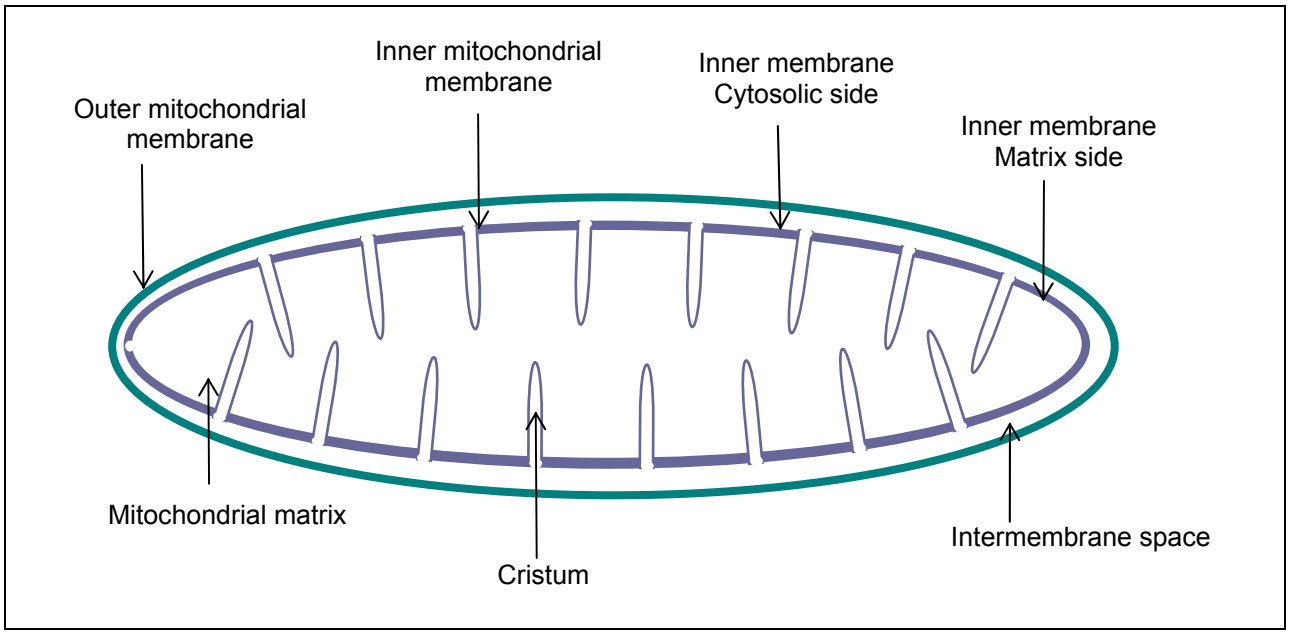
There are systems that in their complexity inspire awe (Aebersold, 2004) and the mitochondrion is such a system. The segregation of a portion of the genome of eukaryotic organisms into the mitochondria represents a unique phenomenon in nature (Attardi, 1985). The establishment of the complete sequence of the human mtDNA (Anderson *et al*., 1981), the unravelling of the genetic code of the mitochondria (Barrell *et al*., 1979; Barrell *et al*., 1980) and the detailed description of the structural and metabolic properties of the transcripts of mtDNA have provided great insight into the structure and principle of operation of the mitochondrial genome (Attardi, 1986).

2.1.1 Physical structure of the mitochondrion

The basic structure of the mitochondrion is illustrated in Figure 2.2. The mitochondrion consists of four main compartments – the outer mitochondrial membrane, the intermembrane space, the inner mitochondrial membrane and the matrix (Bauer *et al*., 1999; Harris and Thompson, 2000; Berg *et al*., 2002). The outer mitochondrial membrane is a relatively simple phospholipid bilayer, containing four types of proteins, of which porin (Berg *et al*., 2002; Wiedemann *et al*., 2004), a porous protein, renders it permeable to molecules that are at most 10 kiloDaltons (kDa) in weight (Sogo and Yaffe, 1994; Jänsch *et al*., 1998; Harris and Thompson, 2000; Ishikawa *et al*., 2004).

The majority of proteins in the mitochondria are situated within the matrix and the inner mitochondrial membrane (Taylor *et al*., 2003). The four enzyme complexes [nicotinamide adenine dinucleotide coenzyme Q oxidoreductase (NADH-Q reductase), succinate coenzyme Q oxidoreductase (succinate-Q reductase), cytochrome reductase and cytochrome oxidase] that catalyse the transfer of electrons, the two mobile electron carriers, ubiquinone and cytochrome c, as well as the ATP synthesising complex, are located in the inner mitochondrial membrane (Senior, 1988; Wallace, 1992; Wallace, 1994; Adams and Turnbull, 1996). The matrix contains mtDNA molecules, ribosomes, transfer ribonucleic acids (tRNAs) and the enzymes needed to catalyse the metabolic activities that take place in the mitochondria (Houshmand, 2003).

Figure 2.2 The physical structure of the mitochondrion



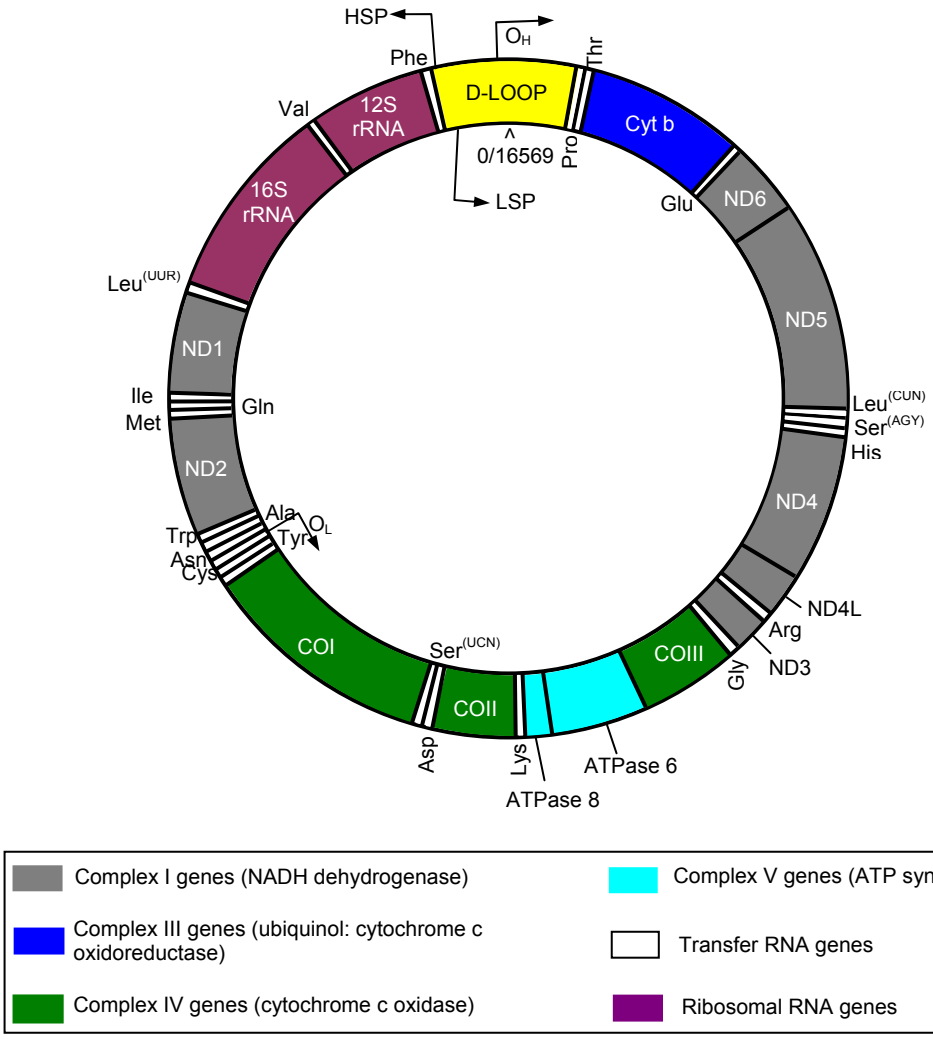
The number and extent of folding of the cristae in the diagram is smaller than in real life.

2.1.2 Genetic structure of the mitochondrion

The mitochondrion and the nucleus are the only cellular organelles in animals that contain DNA (Borst, 1977). The majority of mammalian cells contain hundreds or even thousands of mitochondria and the matrix of each mitochondrion contains 2 - 10 mtDNA molecules (Bogenhagen and Clayton, 1974; Luft, 1994; Fernandez-Silva *et al.*, 2003). The mtDNA as indicated in Figure 2.3 is a circular molecule of 16,569 base pairs (bp) and has two strands, a guanine-rich heavy (H) strand and a cytosine-rich light (L) strand (Zeviani and Antozzi, 1997). mtDNA encodes 13 polypeptides, two ribosomal ribonucleic acid molecules (rRNA) and 22 tRNA molecules (Anderson *et al.*, 1981; Andreas *et al.*, 1997; Zeviani and Antozzi, 1997). The H-strand encodes 2 rRNAs, 14 tRNAs and 12 protein subunits while the L-strand encodes no rRNAs but 8 tRNAs and 1 protein subunit i.e. subunit six (ND6) of the NADH-Q reductase complex (Wallace *et al.*, 1999). The mitochondrion has a different genetic code (Barrell *et al.*, 1979; Barrell *et al.*, 1980; Anderson *et al.*, 1981) and its genome exhibits high economy (Anderson *et al.*, 1981). Its genes are closely packed (some genes actually overlap) and in the majority of cases introns are lacking in the sequences of the coding regions (Anderson *et al.*, 1981). The only non-coding but functionally important parts of its genome are within the D-loop (Anderson *et al.*, 1981; Luft, 1994) and the region which serves as the origin of L-strand (O_L) replication (Zeviani and Antozzi, 1997) which is 30 nucleotides long (Zeviani and Antozzi, 1997). The regular

distribution of tRNAs in mtDNA molecules facilitates RNA processing (Fernandez-Silva *et al.*, 2003).

Figure 2.3 Map of the human mitochondrial genome



Outer circle = H strand, inner circle = L strand, O_H = origin of H-strand replication, O_L = origin of L-strand replication, HSP = H-strand promoter, LSP = L-strand promoter, rRNA = ribosomal RNA, ND 1 - 6 = genes encoding subunits 1 to 6 of NADH dehydrogenase, CO I - III = genes encoding subunits I to III of cytochrome c oxidase, ATPase 6 and 8 = genes encoding subunits 6 and 8 of ATP synthase, Cyt b = gene encoding cytochrome b, D-LOOP = displacement loop. The following three letter symbols of amino acids represent the tRNA for that amino acid: Ala = alanine, Asp = aspartic acid, Arg = arginine, Asn = asparagine, Cys = cysteine, Glu = glutamic acid, Gln = glutamine, Gly = glycine, His = histidine, Ile = isoleucine, Lys = lysine, Met = methionine, Phe = phenylalanine, Pro = proline, Thr = threonine, Trp = tryptophan, Tyr = tyrosine and Val = valine. 12S rRNA = ribosomal RNA of 12 Svedberg units, 16S rRNA = ribosomal RNA of 16 Svedberg units. The two tRNA genes for leucine are differentiated as $Leu^{(UUR)}$ and $Leu^{(CUN)}$ and the two tRNA genes for serine as $Ser^{(UCN)}$ and $Ser^{(AGY)}$. Adapted from Wallace *et al.*, 1999.

2.1.2.1 Inheritance pattern

Mitochondria are maternally inherited (Reich and Luck, 1966; Giles *et al.*, 1980; Schwartz and Vissing, 2002). The discrimination in the parental genotype is established at, or soon after, the formation of the zygote (Kaneda *et al.*, 1995; Shanske *et al.*, 2001). This is due to the sperm providing much fewer mitochondria than the ovum (Kaneda *et al.*, 1995), or the

mitochondria from the father's sperm not surviving (Kaneda *et al*., 1995; Adams and Turnbull, 1996). The sperm mitochondria undergo selective destruction aided by ubiquitin tagging (Sutovsky *et al*., 1999). A woman will transmit her mtDNA to all her male and female children, but only her daughters will in turn pass it on to their offspring (Shanske *et al*., 2001).

2.1.2.2 Replication, transcription and translation of the mitochondrion

Replication of mammalian mtDNA is catalysed by a DNA polymerase after processing of a short primer by mitochondrial RNA polymerase or mtRNApol (Fernández-Silva *et al*., 2003), in the presence of mitochondrial transcription factor A or mtTFA (Fernández-Silva *et al*., 2003). The regulation of replication is not clearly defined (Fernández-Silva *et al*., 2003) but is probably directed by the nucleus (Meirelles and Smith, 1998). Transcription of mtDNA is directed by the light strand promoter (LSP) and the heavy strand promoter or HSP (Anderson *et al*., 1981; Fernández-Silva *et al*., 2003). The transcription of mtDNA is catalysed by mtRNApol (Fernández-Silva *et al*., 2003), which requires mtTFA, and either mitochondrial transcription factor B1 (TFB1M) or mitochondrial transcription factor B2 (TFB2M) for initiation of transcription (Falkenberg *et al*., 2002; Gaspari *et al*., 2004) while a mitochondrial termination factor (mTERF) is required for the termination of transcription (Fernández-Silva *et al*., 2003). The transcription factors serve to enable mtRNApol to recognise the promoters (Garstka *et al*., 2003). Termination of transcription occurs when mtRNApol binds to mTERF, a protein of 34 kDa (Fernández-Silva *et al*., 2003). RNA processing and maturation occurs through 5' (catalysed by RNAse P) and 3' endonucleolytic cleavages of the tRNA, polyadenylation of rRNAs and mRNAs (catalysed by mitochondrial poly (A) polymerase) and linkage of CCA to the 3' end of tRNA, which is catalysed by an ATP(CTP): tRNA nucleotidyltransferase (Fernández-Silva *et al*., 2003). The process of polyadenylation does not only serve to stabilise RNAs but also facilitates the production of stop codons for certain mRNAs (Fernández-Silva *et al*., 2003).

The synthesis of proteins in the mitochondria is achieved through RNA that is synthesised in the mitochondrion and proteins imported from the cytoplasm (Barrell *et al*., 1980; Anderson *et al*., 1981; Larsson and Clayton, 1995). There are 22 tRNAs that the human mitochondrion uses for translation and it has four codons different from the universal code of nDNA as indicated in Table 2.1. AGA and AGG, which encode arginine in nDNA, are stop codons in mtDNA; AUA encodes methionine instead of isoleucine and UGA encodes

tryptophan rather than being a stop codon (Barrell *et al*., 1979; Barrell *et al*., 1980; Anderson *et al*., 1981).

Table 2.1 Differences between the genetic code of the mitochondrial genome and the universal code of nuclear DNA

Codon	Universal code	Mitochondrial code
UGA	Stop	Trp
AUA	Ile	Met
AGA	Arg	Stop
AGG	*Arg*	Stop

A = adenine, G = guanine, U = uracil, Arg = arginine, Ile = isoleucine, Met = methionine, Trp = tryptophan. Adapted from Berg *et al*., 2002).

Mammalian mtDNA is maintained as well as propagated by nuclear-encoded proteins (Larsson and Clayton, 1995). The nuclear genes encode the majority of subunits of the respiratory chain, all the proteins required for replication, transcription and translation of mtDNA transcripts, and all the proteins used for importation of mitochondrial proteins (Larsson and Clayton, 1995). The genes are transcribed and translated in the nucleus before the protein product is targeted into the mitochondrion (Blanchard and Lynch, 2000). The subunits encoded by mtDNA are indicated in Table 2.2. mtDNA also encodes 12S rRNA, 16S rRNA and 22 tRNAs (Wallace *et al*., 1999).

Table 2.2 Subunits of the respiratory enzyme complexes encoded by mitochondrial genes

Complex	Enzyme	Number of subunits	Subunits encoded by mitochondrial genes	Reference
I	NADH-Q reductase	46	ND1, ND2, ND3, ND4, ND4L, ND5 and ND6	Carroll *et al*., 2002
II	Succinate-Q reductase	4	None	Adams and Turnbull, 1996
III	Cytochrome reductase	11	Cyt b	Adams and Turnbull, 1996
IV	Cytochrome oxidase	13	COI, COII, COIII	Campbell and Smith, 1993
V	ATP synthase	16	ATPase 6 and ATPase 8	Walker *et al*., 1991

Q = ubiquinone, ND1-ND6 = NADH-Q reductase subunits 1 to 6, Cyt b = cytochrome b, COI-III = cytochrome oxidase subunits I to III, ATP = adenosine triphosphate, ATPase 6 and 8 = ATP synthase subunit 6 and 8.

2.2 MUTATION RATE OF MITOCHONDRIAL DNA

DNA damage can occur *in vivo* even at levels of normal oxidative metabolism (Beckman and Ames, 1999). The 8-oxo-2-deoxyguanosine lesion has been widely used as a marker

of oxidative damage (Hamilton *et al*., 2001). The mutation rate of mtDNA varies between tissues with tissues that have a high energy demand such as the brain and muscles being more affected than those with low energy demand (Hamilton *et al*., 2001). Transitions are much greater in number than transversions (Brown *et al*., 1982). It is estimated under steady-state conditions that the levels of 8-oxo-2-deoxyguanosine/10^5 deoxyguanosine in mtDNA, vary from 0.19 for liver tissue to 0.34 for brain tissue in the mouse genome (Hamilton *et al*., 2001) and the lowest level for human lymphocytes at 0.25 (Beckman and Ames, 1997).

The mtDNA mutation rate is ca. 10 - 17 times higher than that of nDNA (Brown *et al*., 1979; Wallace, 1994; Wallace *et al*., 1999; Hamilton *et al*., 2001). This higher mitochondrial mutation rate is due to a poor proofreading mechanism within mitochondria (Johnson and Johnson, 2001) and oxidation of the mtDNA by reactive oxygen radicals generated in the respiratory chain (Ames *et al*., 1993). One factor that leads to mtDNA having a higher mutation rate than nDNA is its vicinity to the site of reactive oxygen species (ROS) formation (Beckman and Ames, 1999). The lack of protective histone proteins and the high transcription rate also bring about the high mutation rate of mtDNA (Ames *et al*., 1993; Beckman and Ames, 1999). The high mutation rate of mtDNA renders it extremely useful for the assessment of the process of evolution (Brown *et al*., 1979; Springer *et al*., 1995; Kivisild *et al*., 2006). The high rate generates high signals that enable evolutionary analysis of even recently diverged populations in time and space (Elson *et al*., 2004; Kivisild *et al*., 2006).

Human populations, because of the high mtDNA mutation rate, possess population-specific polymorphisms that have facilitated their characterisation into specific haplogroups, which allows for the reconstruction of human-historical demographic events (Wallace *et al*., 1999). Since the mutations accumulate with time along female lineages, the mtDNA mutation rate can be used as a biological clock to monitor and date events in human pre-history (Wallace, 1994; Wallace *et al*., 1999).

Models of nucleotide substitutions have been used to convert sequence information into phylogenetic trees so as to make better predictions of population history (Felsenstein, 1988; Huelsenbeck, 1997; Felsenstein, 2008). The mutation rate of transitions and transversions and the measurement of the transition to transversion ratio have been used to improve on the correctness of phylogenetic trees and have led to a better understanding of the process of molecular evolution (Strandsberg and Salter, 2004). The distance-based

and the parsimony methods underestimate the transition to transversion ratio since no consideration of the effect of multiple substitutions at a given site in a specified time period is given (Strandsberg and Salter, 2004).

mtDNA is used as a tool in investigations of molecular phylogenetics (Elson *et al*., 2001) because it is inherited exclusively from the mother (Reich and Luck, 1966; Giles *et al*., 1980; Wallace *et al*., 1999) and does not undergo recombination (Aquadro and Greenberg, 1983; Elson *et al*., 2001). The lack of recombination and maternal inheritance rules out the complexities that can arise from bi-parental recombination and facilitates the deciphering of population history from the maternal perspective (Bonatto and Salvano, 1997; Maca-Meyer *et al*., 2001). The greater number of mtDNA molecules in cells (Bogenhagen and Clayton, 1974; Luft, 1994; Wallace, 1995; Wallace *et al*., 1999; Fernandez-Silva *et al*., 2003) coupled with its high mutation rate (Brown *et al*., 1979; Wallace, 1994; Wallace *et al*., 1999; Hamilton *et al*., 2001), has also made mtDNA a tool of choice for phylogenetics as compared to nDNA (Brown *et al*., 1979; Springer *et al*., 1995; Kivisild *et al*., 2006), as a relatively higher number of polymorphisms are used as signals of events that occurred in the past (Elson *et al*., 2004; Kivisild *et al*., 2006) which in turn leads to a higher resolution of population history.

CHAPTER THREE

Evolutionary and Phylogenetic Aspects of mtDNA

Migration is a major factor influencing mtDNA variation. Conversely, signals stored in the form of mtDNA polymorphisms can be used to decipher human demographic history such as population migrations, bottlenecks and expansions (Jorde *et al*., 1998). The analysis of mtDNA has elucidated the African origin of the human species (Vigilant *et al*., 1989; Hagelberg, 2003). mtDNA is commonly used to determine the time and route of the most important human demographic events in history, such as the migration into Europe of Neolithic farmers (Barbujani and Bertorelle, 2001; Hagelberg, 2003), and the settlement of people into the Pacific and the New World (Hagelberg, 2003). The study of the genetics and variation of human mtDNA has helped to define human evolution (Rogers and Jorde, 1995) and has been widely used in the phylogenetic analysis of human populations (Chen *et al*., 1995; Wallace *et al*., 1999).

3.1 HUMAN ORIGIN

The disciplines of human genetics and archaeology have been used to infer the geographical origin and timespan of anatomically modern humans (Cann *et al*., 1987; Stringer and Andrews, 1988; Nei, 1995; Wallace *et al*., 1999; Stringer, 2002). Studies of human genetics have been instrumental in inferring the geographical origin of an ancestral population from the distant past, due to the fact that varying human populations, large sample sizes and a variety of genetic markers can be used (Jorde *et al*., 1998; Takahata *et al*., 2001; Liu *et al*., 2006). However, two contrasting models - the Out-of-Africa hypothesis (Cann *et al*., 1987; Nei, 1995) and the multiregional evolution hypothesis (Nei, 1995; Tishkoff and Williams, 2002) have been traditionally used to explain the origin of *Homo sapiens sapiens*. Both models have been subjected to intense criticism, but most of the evidence that has emerged over the last 15 years has been supportive of the Out-of-Africa hypothesis (Jorde *et al*., 1995; Nei, 1995; Eswaran *et al*., 2005).

The Out-of-Africa hypothesis postulates that modern human beings originated in Africa and the original African population is the ancestor of all present human populations (Cann

et al., 1987; Stringer and Andrews, 1988; Jorde *et al*., 1998). By 1,000,000 years before present (YBP), archaic *hominids* had spread to most of the regions of the Old World (Nei, 1995; Tishkoff and Williams, 2002). According to the Out-of-Africa hypothesis, the first anatomically modern humans evolved from a small population of about 10,000 people in complete reproductive isolation, in Africa 100,000 - 200,000 YBP (Nei, 1995; Jorde *et al*., 1998; Harpending and Rogers, 2000). Liu *et al*. (2006) estimates the founding population to have been about 1,000 reproductively effective individuals. As the population evolved, it spread along the coastal zones of Africa into Asia and Europe. As this population spread further to other regions, it came into contact with the archaic beings already present in those areas but there is no evidence that they interbred with them (Cann *et al*., 1987; Aiello, 1993; Templeton, 1993; Nei, 1995; Jorde *et al*., 1998). All present day human genes are therefore descended from those of the isolated original population in Africa (Cann *et al*., 1987; Templeton, 1993; Tishkoff and Williams, 2002).

Furthermore, Templeton (1993) put forward a refined definition of the Out-of-Africa hypothesis which she termed the “Eve” hypothesis. The “Eve” hypothesis postulates that all the variation in mtDNA exhibited in modern human beings was derived from a single female predecessor. Since mtDNA does not recombine and is maternally inherited in primates, it can be hypothesised that all copies of human mtDNA can be traced to a common female ancestor (Cann *et al*., 1987; Templeton, 1993), and all other primates have their own mitochondrial “Eve” (Templeton, 1993). This common female ancestor inhabited the African continent, around 200,000 years ago (Cann *et al*., 1987; Templeton, 1993; Tishkoff and Williams, 2002). Furthermore, since all mtDNAs in modern humans are descended from a common female ancestor, the “Eve” hypothesis also implies that all modern humans in their entirety trace back to the same female ancestor and arose from the same geographical origin as the common ancestor (Templeton, 1993).

The multiregional evolution hypothesis holds that the transformation of anatomically modern humans from the archaic forms took place during the same period of time in different regions of the Old World (Cann *et al*., 1987; Stringer and Andrews, 1988; Jorde *et al*., 1998; Li and Su, 2000; Wolpoff *et al*., 2000; Tishkoff and Williams, 2002). Accordingly, the archaic *hominids (Homo erectus),* who were dispersed throughout Europe, East Asia, Australia and Africa, did not die out but simultaneously evolved independently in the different geographical regions to give rise to modern *Homo sapiens* (Nei and Roychoudhury, 1993; Nei, 1995; Tishkoff and Williams, 2002). This hypothesis

further claims that the extent of gene flow was so small that several regional anatomical features such as the incisors in north-eastern Asians that are shovel-shaped and the pronounced eyebrow ridge among the aborigines from Australia, have not changed since the time of the ancestral *Homo erectus* (Nei and Roychoudhury, 1993; Nei, 1995).

Multiregionalists contend that the replacement of *Homo erectus* by *Homo sapiens* without appreciable hybridisation as postulated by the Out-of-Africa hypothesis is unlikely since when people of varying ethnicities settle in the same geographical location, hybridisation between the different ethnic groups almost always takes place (Jorde *et al*., 1998). The Out-of-Africa hypothesis argues that replacement of *Homo erectus* by *Homo sapiens* without hybridisation is possible if the size of the *Homo erectus* population was small as compared to *Homo sapiens* around 100,000 years ago (Nei, 1995). An example to support this claim of the Out-of-Africa hypothesis is the Neanderthals in Europe who hit an evolutionary dead end, whereby their gene pool became extinct as it does not demonstrate continuity in present Europeans (Adcock *et al*., 2001; Relethford, 2001; Takahata, 2001; Caramelli *et al*., 2003; Currat and Excoffier, 2004).

Furthermore, the phenomenon of similar kinds of organisms existing in the same locality but leading to genetic discontinuity of one kind as put forward by the Out-of-Africa hypothesis (Nei, 1995) is demonstrated by the replacement in South and Central America of the European honey bees by the African killer bees (Nei, 1995; Schneider *et al*., 2004). In less than 50 years, the African honey bee subspecies *Apis mellifera scutellata* has had a remarkably outstanding biological invasion as it displaced the European honey bee subspecies in the New World (Schneider *et al*., 2004). Although there is substantial hybridisation that takes place when African bees invade European bees, over time, the European genetic and phenotypic characteristics tend to be lost (Schneider *et al*., 2004). A similar situation may have arisen in the case of the *Homo sapiens* and the ancestral hominid populations.

Analysis of ancient Australian human remains from Lake Mungo 3, Wallandra Lakes and Kow Swamp regions using molecular and morphological features has also provided insight into the origin of modern humans (Adcock *et al*., 2001). The differences in the morphology of the fossils of various ages were interpreted as being indicative of varying sources of migrants into Australia (Relethford, 2001). mtDNA sequences from remains at Lake Mungo 3, whose morphology resembled that of modern humans, belonged to a lineage that does

not demonstrate continuity in present humans (Relethford, 2001) while sequences from the Kow swamp remains that were morphologically different to living modern Australians, clustered with modern humans (Relethford, 2001). When comparing data of living humans with human fossils, the choice of the model applied for interpretation of the results affects the final outcome and therefore may be used to support the multiregional model of human evolution as well (Li and Su, 2000; Relethford, 2001). For example, if Neanderthals were taken as a species different from modern humans, the analysis of mtDNA sequences would generate information as to when they split away from modern humans (Relethford, 2001). However, if the Neanderthals were taken as belonging to the same species as modern humans, their mtDNA divergence times would be indicative of ancient patterns of gene flow and population expansion (Relethford, 2001). The extinction of human lineages as revealed by the study performed by Adcock *et al*. (2001) serves to demonstrate that the inference of Neanderthals and modern humans as separate species is debatable (Relethford, 2001).

Analysis of African, Asian, European and American mtDNA confirmed that there was appreciable correlation between mtDNA sequences of individuals and their ethnic and geographic origins (Wallace *et al*., 1999). Furthermore, the position of the root of the mtDNA phylogenetic tree of human continental populations (Wallace *et al*., 1999) and the degree of sequence diversity was consistent with an African origin for anatomically modern humans (Wallace *et al*., 1999). The Mbuti and Biaka pygmies, the !Kung, the Khwe, (Watson *et al*., 1996; Chen *et al*., 2000) and the Khoi-San from Tanzania (Gonder *et al*., 2007) may harbour the eldest sub-lineages of African mtDNA and could be the most ancient human population.

The argument that anatomically modern humans originated in Eurasia and some of the Eurasian populations immigrated into Africa (Denaro *et al*., 1981; Jorde *et al*., 1998; Stringer, 2002) does not hold because no Eurasian population bears a greater genetic distance from any other Eurasian population as has been determined for the African population (Cann *et al*., 1987; Jorde *et al*., 1998). Moreover, if the archaeological finding that the iron tool-making technology evolved much earlier in sub - Saharan Africa (90,000 YBP) than in Europe (40,000 YBP) holds, then the origin of modern humans from Africa is credited (Cann *et al*., 1987). Furthermore, the support of an East Asian origin of *Homo sapiens sapiens* on the basis of the many fossils discovered in China does not hold since the duration of time between 100,000 YBP (the age of the hominid remains that were last

discovered in China) and 40,000 YBP (the age of the *Homo sapiens sapiens* oldest fossils dated) indicates that there was a break in the continuity of the fossils (Li and Su, 2000). The duration of 60,000 years between the Hominid fossils being discovered before 100,000 YBP (Li and Su, 2000 and Wu, 2004) and the fossils discovered after 40,000 YBP can best be attributed to the extinction of archaic humans and a subsequent replacement by anatomically modern humans (Li and Su, 2000). If there was no time gap between fossils, this would have been evidence of genetic continuity between archaic and anatomically modern man which would in turn credit the multiregional model of evolution (Li and Su, 2000).

The elevated African genetic diversity, although consistent with the Out-of-Africa hypothesis, is not definitive that modern humans had an African origin (Jorde *et al*., 1998). For example, an Asian origin of anatomically modern humans may have occurred but a subsequent bottleneck could have reduced the genetic diversity of the Asian population to levels below that of the African population (Jorde *et al*., 1998). Moreover, it is possible that the high genetic diversity of Africans may simply be the consequence of a considerably larger effective population size as compared to other populations (Jorde *et al*., 1998). The star-like nature of the mtDNA phylogeny of non-Africans is suggestive of a population bottleneck and a recent sudden increase in population size (Bandelt *et al*., 1995; Jorde, 1998; Quintina-Murci *et al*., 1999; Ingman *et al*., 2000; Takahata *et al*., 2001; Torroni *et al*., 2006). All the mtDNA sequences of non-African populations demonstrate multiple origins which indicates that each of these areas have experienced repeated colonisation (Cann *et al*., 1987). Analysis of Y-chromosome (Takahata *et al*., 2001), X-chromosome and autosomal sequences (Takahata *et al*., 2001; Zhivotovsky *et al*., 2003) have, however, corroborated the mtDNA evidence supporting an African origin of *hominids.* Computer simulations using multilocus genetic data revealed that East Africa is the most likely place of origin for modern humans (Ray *et al*., 2005).

3.1.1 Origin of the people of East Africa

The long, epic and eventful evolutionary narrative of the emergence of mankind could have had its humble beginnings in East Africa (Were and Wilson, 1984). The Sandawe and Hadza of Tanzania of the Khoisan language family are believed to be the original ancestral population of anatomically modern humans, who moved out of Tanzania to populate the rest of Africa and thereafter the rest of the world (Were and Wilson, 1984; Cruciani *et al*., 2002; Salas *et al*., 2002; Gonder *et al*., 2007). It could have been from East Africa that

primitive man, while still a hunting hairy mammal, carried the technological skills of making stone and bone tools to Southern Africa, and into Asia and Europe during the era of the Stone Age (Were and Wilson, 1984). The second ethnic group to inhabit East Africa were the Caucasoid (Were and Wilson, 1984), Stone Age Southern Cushites from Ethiopia, who reared cattle and practised hunter-gathering (Were and Wilson, 1984). The Bantu migrated into East Africa beginning in the year 400 *Anno Domino* (AD) from the Niger-Congo basin (Were and Wilson, 1984) and comprises the biggest ethnic group in the region (Salas *et al*., 2002).

The first people to occupy Buganda probably came from the Mount (Mt) Elgon region (Were and Wilson, 1984; Ssekamwa, 1994), who moved to Bunyoro but broke off from the Banyoro to establish the kingdom of Buganda (Were and Wilson, 1984). A related account posits that the original Baganda were later mixed with immigrants from Bunyoro, Mt Elgon and the Sese Island (Were and Wilson, 1984). Another similar account posits that a group of people migrated from the northern areas around Madi to Chopi in Bunyoro where a section of the population migrated to establish the kingdom of Buganda after the collapse of the Bunyoro-Kitara Empire (Were and Wilson, 1984; Ssekamwa, 1994). Other accounts claim that the Baganda were part of the group of the Bantu that migrated from the Congo to Shungwaya in Tanzania (Were and Wilson, 1984; Ssekamwa, 1994). The origin and peopling of Buganda seems to be a contribution from different places involving individuals of diverse ethnicities (Were and Wilson, 1984). The Acholi trace their origin to Rumbek in Southern Sudan (Imperato, 1982; Were and Wilson, 1984; Ssekamwa, 1994) while the origin of the Lugbara is traced to Juba in Southern Sudan (Ssekamwa, 1994). There is little documentation on the origin of the Acholi and Lugbara because there is general consensus on the subject, both in written literature and oral tradition that they originated from Southern Sudan. The origin of the Baganda is however diversified (Were and Wilson, 1984; Ssekamwa, 1994) and remains debatable and therefore more literature have been written on it in an effort to seek consensus. This study will therefore be relevant in shedding light about the origin of the Baganda, Acholi and Lugbara tribes from Uganda from a genetic perspective.

3.2 HUMAN MIGRATION

mtDNA sequence analysis can result in the unravelling of prehistoric human demographic events such as ancient human migrations (Torroni *et al*., 1994; Salas *et al*., 2002). Tracing the evolutionary history of the human population can be achieved through analysis of the

information obtained from mtDNA nucleotide substitutions and length changes such as insertions and deletions (Redd *et al*., 1995). Culture and languages spread with the movement of people (Were and Wilson, 1984; Barbujani and Pilastro, 1993; Holden, 2002) and could therefore also be used to trace migration events that have occurred in the past (Were and Wilson, 1984; Barbujani and Pilastro, 1993; Holden, 2002; Pinhasi *et al*., 2005). Conversely, migrations could also be used to trace human culture and languages as well as the genetic make-up of humans (Armelagos and Harper, 2005; Pinhasi *et al*., 2005). The integration of results obtained independently or jointly from genetics, archaeology and linguistics will lead to a greater knowledge and understanding of human demographic prehistoric events (Weiss, 1994; Torroni *et al*., 1994; Blench and Dendo, 2004; Bolnick *et al*., 2004). It is envisaged that the knowledge of the migratory patterns of the Bantu, Nilotes and Moru-Madi will be of great value in accounting for the observed polymorphisms and phylogenetic relationship amongst the Baganda, Acholi and Lugbara samples that have been used in this project.

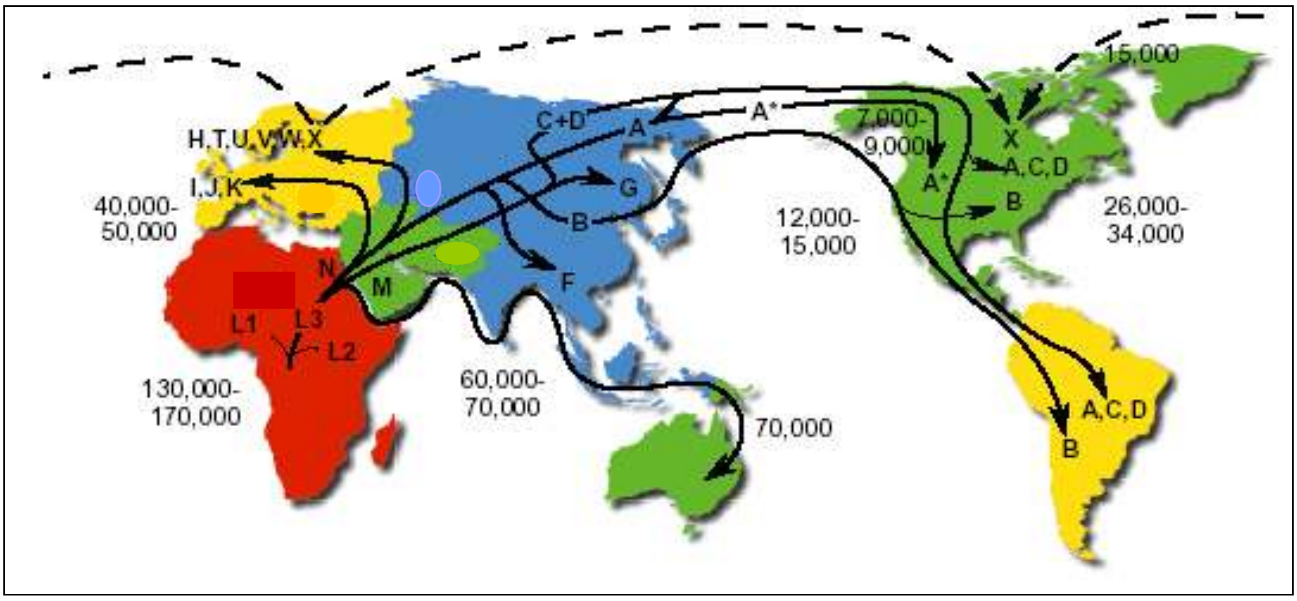
Human migrations have been greatly influenced by the landscape of their environment (Cordaux *et al*., 2004; Liu *et al*., 2006). For example, natural passageways such as the one formed by the Himalayas and the Bay of Bengal in India constitutes a unique corridor (as well as a barrier) that links the subcontinent of India to East and Southeast Asia (Cordaux *et al*., 2004). Other examples include the River Nile Valley through which Africa and Eurasia are connected (Krings *et al*., 1999; Cordaux *et al*., 2004), and geographical barriers such as mountains (Burchard *et al*., 2003; Cordaux *et al*., 2004; Liu *et al*., 2006), large water bodies (Burchard *et al*., 2003) and large deserts (Burchard *et al*., 2003; Were and Wilson, 1984). Other than landscape, there are other factors that influence human migration which include climate (Torroni *et al*., 1994; Forster *et al*., 1996; Richards *et al*., 2000; Forster *et al*., 2004; Achilli *et al*., 2004; Pereira *et al*., 2005), wars, overpopulation, need for cheap labour (through slavery), soil fertility as well as diseases that affect people and animals (Were and Wilson, 1984). Since migration is a major factor affecting gene flow among human beings (as outlined in the introduction to Chapter three), any factor affecting migration has therefore a major secondary effect on the genetic diversity of the mtDNA of the populations in this investigation.

3.2.1 Human migrations from and into Africa

According to the Out-of-Africa hypothesis, modern humans arose out of Africa about 150,000 YBP, migrated into Asia about 60,000 to 70,000 YBP and between 40,000 to

50,000 YBP they entered into Europe (Wallace *et al*., 1999; Maca-Meyer *et al*., 2001). Around 20,000 to 30,000 YBP, the anatomically modern humans migrated from Asia and Europe to colonise America (Wallace *et al*., 1999). All mtDNAs of non-Africans stem from African haplogroup L3, which gave rise to haplogroups M and N which are the ancestral haplotypes of the populations outside of Africa (Watson *et al*., 1997; Forster, 2004). It is postulated, as indicated in Figure 3.1, that two routes out of Africa were used with one being along the waters of the Nile River followed by land across the Sinai Peninsula while the other route is hypothesised to be across the Bab el Mandab Strait on sea to Yemen (Forster, 2004).

Figure 3.1 Migratory routes of mankind out of Africa



The migration of the oldest mtDNA macrohaplogroup L out of Africa into Asia, and into Europe and the Americas. Letters correspond to haplogroups and numbers are time estimates of occurrence of events in YBP. Adapted with permission from Mitomap (2002).

Haplogroup M and related clades spread from Africa into Asia following the southern route (Maca-Meyer *et al*., 2001). However, some of the haplogroup M individuals present in East Africa have been attributed to back migration from Asia (Maca-Meyer *et al*., 2001). The northern migration through the Levant gave rise to lineages X and A as the major clusters which are frequent among the Caucasoids in Asia, Europe and America (Maca-Meyer *et al*., 2001). The second but minor Levant route resulted in the distribution of haplogroups W, I and N1b to Europe, the Near East and the Caucasus with haplogroups I and N1 also being present in Egypt and Arabia (Maca-Meyer *et al*., 2001). The third cluster along the Levant route produced four haplogroups including haplogroup A which migrated into Japan and the south-eastern archipelagos in the Pacific (Maca-Meyer *et al*., 2001).

3.2.1.1 Human migrations from and into East Africa

As outlined in Section 3.1.1, anatomically modern humans had an East African origin. From East Africa, anatomically modern people moved to populate the rest of Africa and into Asia, Europe and the New World (Maca-Meyer *et al*., 2001) via the haplogroup L3 ancestral lineage (Wallace *et al*., 1999). Eastern Africa therefore harbours the ancestral haplotypes that served as precursors from which other global mtDNAs were derived (Chen *et al*., 2000). mtDNA haplogroups L3 is therefore postulated to have arisen in Eastern Africa, spread to Western and central Africa, and migrated out of Eastern Africa to yield the Asian macrohaplogroup M and N that populated Europe, Australia and America (Watson *et al*., 1997; Chen *et al*., 2000; Torroni *et al*., 2006). The lack of haplogroup M in the Levant in combination with its high occurrence in the South-Arabian Peninsula demonstrates that haplogroup M individuals migrated into Western India through Eastern Africa (Quintana-Murci *et al*., 1999). East Africa is therefore the geographical origin of haplogroup M, because in the analysis by Quintana-Murci *et al.* (1999) it was determined that the Asian haplogroup M bifurcated from the East-African M haplogroup more than 50,000 YBP.

Between 3,000 BC and 1,000 BC, the Nilotes from Sudan migrated into Ethiopia in two waves (Passarino *et al*., 1998). From 1,000 BC, the Arabians migrated into Ethiopia and intermarried with the local people to give rise to the Amhara, Tigrinya and Gurage (Passarino *et al*., 1998). The Bantu migrated into East Africa beginning from the year AD 400 from the Niger-Congo basin (Were and Wilson, 1984). Ethiopia also experienced the immigration of the Jews in the first century AD, Syrians in the 4^{th} to 6^{th} century AD and Arabs in the 11^{th} to 12^{th} century AD (Passarino *et al*., 1998).

Before AD 1,000, slaves had been transported from East Africa to countries such as Southern Persia, India and China in Asia (Were and Wilson, 1984). Trade between Eastern Africans and Arabs existed as far back as the 7^{th} millennium B.C. (Richards *et al*., 2003). These men did not move with their women, and so they married African women leading to an increase in the spread of the Afro-Arab culture and the origin of the Swahili people (Were and Wilson, 1984). Other countries that came to the East African coast and were involved in trade included Malaysia, the islands of Indonesia, Sri Lanka, Egypt, France, Syria and other Arabian states (Were and Wilson, 1984). These back migrations from the original radiation of anatomically modern humans from East Africa could have made a significant contribution to the African gene pool (Were and Wilson, 1984) and

altered the mtDNA landscape of the East African population. The African women taken by those who went back had their genes multiplied in these communities through marriage (Were and Wilson, 1984; Richards *et al*., 2003) and analysis of mtDNA of these communities would testify to the fact that Africans as a whole were not totally at a loss in the business of slave trade. Thus the African genetic heritage was dispersed all over the world in a manner wider than any other genetic grouping.

3.2.1.1.1 Migrations and settlements of the East African Bantu

The Bantu migrated from Central Africa, in the area between the Shaba province in the Democratic Republic of Congo and Cameroon (the Niger–Congo basin) and the biggest groups migrated into East Africa between AD 1,000 and AD 1,860 (Were and Wilson, 1984; Ssekamwa, 1994). The group that moved to East Africa reached the East Coast of Tanzania by AD 1,000 (Ssekamwa, 1994) and settled in Shungwaya but some dispersed further into the centre of East Africa (Ssekamwa, 1994). Due to attacks from the Galla, the remaining Bantu in Shungwaya were forcibly driven into the interior of East Africa (Ssekamwa, 1994). Other Bantu groups migrated to Central Africa and settled in countries such as Malawi, Zimbabwe, The Democratic Republic of Congo (DRC) and Zambia while others (the Ngoni) migrated to South Africa and back into Tanzania in East Africa (Were and Wilson, 1984; Ssekamwa, 1994).

According to this account, the Baganda belong to the Western or Interlacustrine Bantu that moved directly from the Congo and entered East Africa between Lake Albert and Lake Edward to settle in the region north-west of Lake Victoria (Ssekamwa, 1994). The first Bantu to inhabit Buganda settled in the counties of Busiro, Mawokota and Kyadondo (Were and Wilson, 1984; Ssekamwa, 1994). These original immigrants into Buganda were later joined by people from Bunyoro, the Sese Islands and Mt Elgon (Were and Wilson, 1984). The present Baganda are likely descendants of people from diverse origins that migrated and became absorbed by the Bantu (Were and Wilson, 1984; Ssekamwa, 1994). Furthermore, due to their interaction with the Chwezi, the Baganda may have genetic similarities with groups of people that interacted with the Chwezi (Were and Wilson, 1994; Ssekamwa, 1994). They could therefore be related to the Hima, Huma, Tutsi or Hinda populations that settled in Bunyoro, Ankole and Toro, in Uganda, and those who immigrated into Rwanda, Burundi, Tanzania (Kiziba in Bukoba district) and Kenya, specifically the Wanga Kingdom (Were and Wilson, 1984). The Chwezi are believed to be of Galla or Cushitic (Ethiopian) origin (Were and Wilson, 1984) and therefore some of the

Baganda and Ethiopians may have a close genetic relationship. The Baganda could also be related to the Luo Bito due to the Banyoro absorbing the Luo, the Nilotic immigrants of Sudanese origin who came from the North or Northeast beginning from the 16th century (Were and Wilson, 1984). These developments took place ca. between AD 1,490 and AD 1,733 (Were and Wilson, 1984).

3.2.1.1.2 Migrations and settlements of the East African Nilotes

The Nilotic ethnic group (River and Lake Nilotes) includes the Acholi, Luo, Jopadhola and Alur (Were and Wilson, 1984; Ssekamwa, 1994). One of the largest tribes in this group is the Luo (Raymond, 2005). The origin of the Luo was in Rumbek in Southern Sudan (Imperato, 1982) from where they spread to occupy the Bahr-el-Ghazal province (Imperato, 1982; Were and Wilson, 1984; Ssekamwa, 1994) and left this region for East Africa in AD 1,350 (Ssekamwa, 1994). In the 15th century, the Luo migrated to reach northern Uganda at Pakwach on the banks of River Nile in AD 1,480 (Were and Wilson, 1984; Ssekamwa, 1994). Some of the Luo in Pakwach migrated into Acholi land and intermarried with the Madi and the Langi to form the Luo-Acholi and Luo Langi respectively (Were and Wilson, 1984; Sssekamwa, 1994). A second migratory wave of the Luo moved from Pakwach and settled from Pakwir to Chope in Bunyoro, which ultimately led to the break up of the Bunyoro-Kitara Empire and the establishment of the Babito dynasty of the Kingdom of Bunyoro (Were and Wilson, 1984; Sssekamwa, 1994). The conquest and break up of the Bunyoro-Kitara Empire (Ssekamwa, 1994) by the Luo resulted in the formation of the Bunyoro, Ankole, Buganda, Rwanda and Karagwe (Tanzania) kingdoms as well as several Luo Babito chiefships in Busoga (Ssekamwa, 1994). The Luo also intermarried and influenced the culture of the people in Nebbi district in Uganda (Ssekamwa, 1994) and the Kumam in the Kaberamaido peninsula, also in Uganda (Ssekamwa, 1994).

There were three independent waves of migrations that formed the Kenyan Luo (Were and Wilson, 1984; Ssekamwa, 1994). The Joka-Jok were the first group that migrated from Acholi in Northern Uganda into the Ramogi Hills in Western Kenya, between AD 1,490 and AD 1,600 (Were and Wilson, 1984; Sssekamwa, 1994). The second group, the Joka-Owiny, came from Budama and reached Western Kenya between AD 1,590 and AD 1,670 (Were and Wilson, 1984; Ssekamwa, 1994). The third group, the Joka-Omolo, moved from Pawir in Bunyoro to Acholi and further to Tororo, Southern Busoga and Busia into Western Kenya via Samia and Bunyala at the start of the 17th century (Were and

Wilson, 1984). Some of the Joka-Jok were forced to move from Uyoma (Kenya) into South Nyanza (Kenya), in the second half of the 18th century (Were and Wilson, 1984; Ssekamwa, 1994) where they extended into the Mara District in Tanzania (Imperato, 1982).

3.2.1.1.3 Migrations and settlements of the East African Moru-Madi

The Moru-Madi include tribes such as the Lugbara, Metu, Okebu, Madi, Ndo and Aringa (Raymond, 2005). The Lugbara constitute the majority of the Moru-Madi of the Central Sudanic ethnic group (Raymond, 2005). The Lugbara migrated from Juba in Southern Sudan into Uganda between AD 1,600 and AD 1,650 in three independent waves (Ssekamwa, 1994). These migrations were due to invasions from the Bari and the Latuke between AD 1,000 and AD 1,500 (Ssekamwa, 1994). The first Lugbara immigrants came from Moru in the Juba region and settled south-east of Mt Liru, along the shores of the Nile River (Ssekamwa, 1994). The second group moved from Juba and settled in west Acholi where they were later forced by the Luo to move across the Nile River and occupy the area near Mt Wati (Ssekamwa, 1994). The third group moved from Juba and settled in the areas surrounding Mt Witu (Ssekamwa, 1994). Certain members of this group moved into DRC, where others returned to Vurra but there were contacts and intermarriage between the three Lugbara groups from AD 1,600 to AD 1,650 (Ssekamwa, 1994). Between AD 1,630 and AD 1,680, the Lugbara assimilated the Kebu who were living in the area (Ssekamwa, 1994) and between AD 1,790 and AD 1,850, the Lugbara absorbed and assimilated some of the Kakwa, Madi and Alur (Ssekamwa, 1994).

3.3 MITOCHONDRIAL PHYLOGENIES

Human beings can be classified into specific categories according to similarities and differences in their mtDNA sequences (Wallace *et al*., 1999; Kivisild *et al*., 2002). By using specific software programs, the mtDNA sequences can be used to construct phylogenetic trees that depict the genetic relatedness between the sequences (Wallace *et al*., 1999). The phylogenetic software can be used to convert the information contained in DNA or protein sequences into a tree that depicts the genetic relatedness of the individuals as well as the demographic history of the population, with specific evolutionary distances or time scales attached to the tree (Kivisild *et al*., 2002; Kumar *et al*., 2004). The relatedness between individuals can however also be obtained by use of language trees (Barbujani, 1997) or fossil records (Aiello, 1993) or migratory evidence (Were and Wilson, 1984;

Ssekamwa, 1994; Barbujani, 1997; Liu *et al*., 2006). Phylogenies deduced from linguistic comparisons are based on the assumption that a similar or related language signifies a common origin of the ancestors (Barbujani, 1997).

Since mtDNA is maternally inherited, mitochondrial phylogenies can be used to reconstruct the demographic history of the world populations from a female perspective (Bonatto and Salvano, 1997; Maca-Meyer *et al*., 2001). The characteristics of strict maternal inheritance and the lack of recombination make mtDNA based population phylogenetic analyses free of complexities brought about by recombination from both maternal and paternal parents (Elson *et al*., 2001; Eyre-Walker and Awadalla, 2001). The use of complete mtDNA sequences in the construction of phylogenetic networks is preferred since it increases the discrimination power of the sequences which in turn enables better resolution of population demographic and evolutionary history than the use of either restriction fragment length polymorphism (RFLP) data or sequences from the hypervariable segment I (HV I) or HV II regions (Bandelt *et al*., 1995; Finnilä *et al*., 2000; Álvarez-Iglesias *et al*., 2007). The set of enzymes usually used to perform RFLP experiments allow examination of only a small proportion of the mtDNA sequence, rendering a number of polymorphisms undetermined (Wallace, 1994; Finnilä *et al*., 2000).

The more related two languages are, the more recent their common ancestor (Barbujani, 1997) while differences and similarities in languages can be used to trace past demographic events (Barbujani, 1997; Holden, 2002; Tishkoff *et al*., 2009). The samples whose mtDNA were sequenced in this study were chosen on the basis of language since genetic relatedness between African tribes is similar to language classification in most instances (Tishkoff *et al*., 2009). Gene flow across and within languages is often regulated by the fact that human beings tend not to tread across languages when choosing a marriage partner (Barbujani, 1997). However, factors such as borrowing words from neighbouring or elite communities (Holden, 2002; Knight *et al*., 2003; Bolnick *et al*., 2004; Quintina-Murci *et al*., 2004) and assimilation of either the natives by the immigrants or immigrants by the natives (Were and Wilson, 1984; Ssekamwa, 1994; Holden, 2002; Knight *et al*., 2003) have made linguistic phylogenies substantially different from molecular phylogenies. The Baganda belong to the Bantu ethnic group, of the Niger-Congo language family. They speak Luganda (Ganda) and mainly occupy the south-central region of Uganda (Raymond, 2005). Lugbara as a language belongs to the Nilo-Saharan group (Boone and Watson, 1996; Raymond, 2005) and the Lugbara population currently inhabits

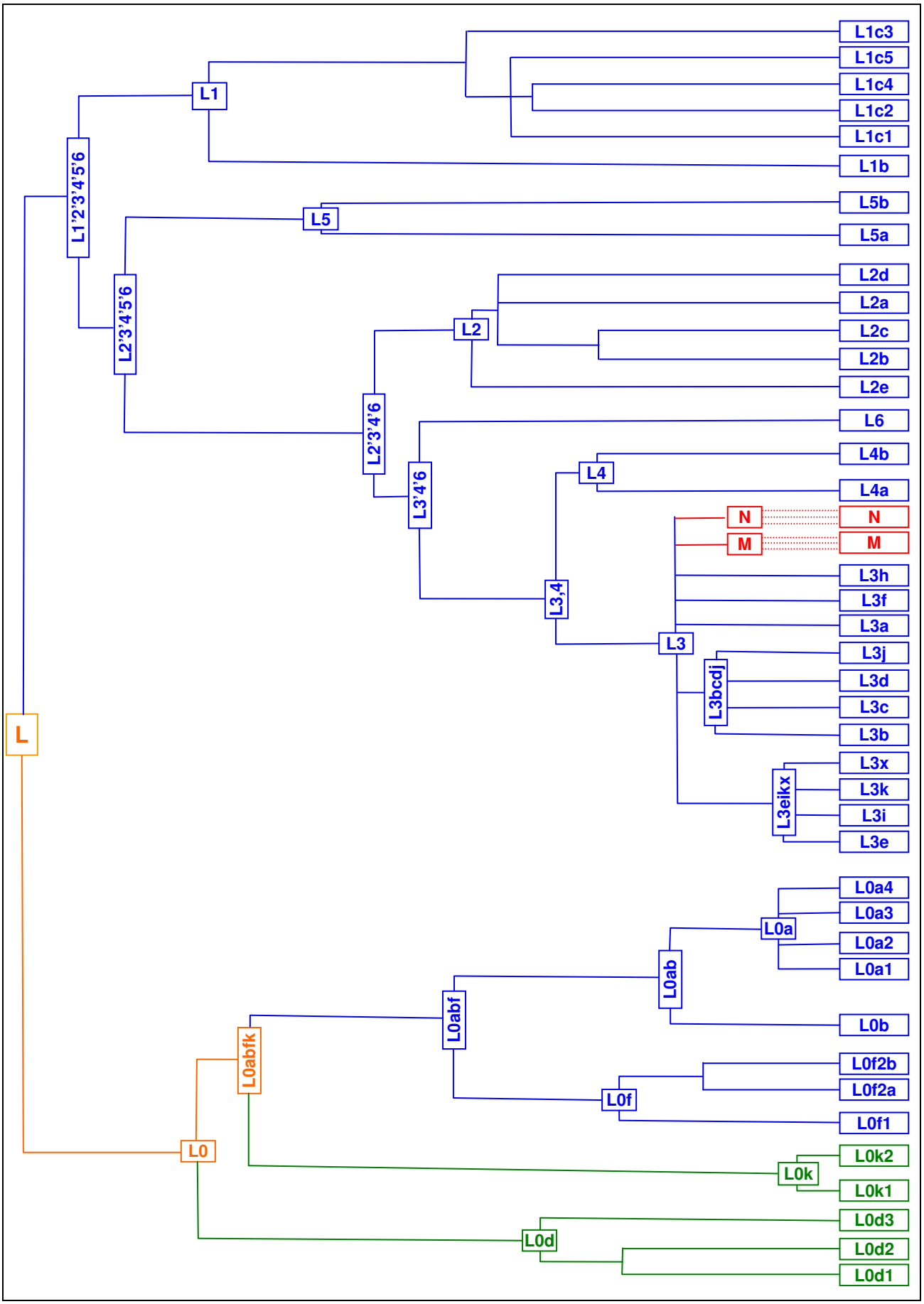
mainly Northwest Uganda and the Democratic Republic of Congo (Raymond, 2005). Acholi belongs to the Nilo-Saharan language family, as Lugbara does but the two tribes belong to different ethnic groups (Raymond, 2005). The Acholi presently occupy north-central Uganda and Southern Sudan (Raymond, 2005).

3.3.1 mtDNA phylogenetic analysis of African populations

A mtDNA phylogenetic analysis of African populations divided the population into two main branches as illustrated in Figure 3.2. African macrohaplogroup L is divided into an L0 haplogroup branch and a branch consisting of haplogroups L1, L2, L3, L4, L5 and L6 (Kivisild *et al*., 2004; Behar *et al*., 2008). L0 is defined by the T146C, A263A, C3516A, T5442C, C9042T, A9347G, G10589A, C10664T, T10915C, A12720G, A13276G, A16230G polymorphisms while L1 to L6 share the C182C, C1048T, C4312T, T6185C, G11914A and G12007A polymorphisms (Torroni *et al*., 2006; Behar *et al*., 2008). The L0 haplogroup bifurcates into two main branches namely L0abfk and L0d (Behar *et al*., 2008) with L0abfk characterised by the A189G, T4586C, C9818T, T16172C polymorphisms. L0d is predominant among the South African Khoi-San and is differentiated from the L0abfk cluster by harbouring the G1438A, T4232C, T6815C, C8113A, G8152A, G8251A, T12121C, G15466A, G15930A, T15941C, and the T16243C polymorphisms (Behar *et al*., 2008). The polymorphisms characterising the different haplogroups illustrated in Figure 3.2 have been outlined in Table A3 in Appendix C.

The South African Khoi-San have a high prevalence (60%) of L0d and L0k lineages (Salas *et al*., 2004a; Gonder *et al*., 2007; Behar *et al*., 2008). L1c1a is localised in Central Africa (Behar *et al*., 2008) while West-Central Africa has a higher proportion of L3e mtDNAs (Salas *et al*., 2004a). Haplogroup L3e belongs to the L3eikx cluster which is differentiated from related L3 clades by the C150T and A10819G polymorphisms (Behar *et al*., 2008). Haplogroups L2b, L2c and L2d are common in Western Africa but very rare in East Africa (Salas *et al*., 2004a). However, haplogroup L2 has five clades i.e. L2a, L2b, L2c, L2d and L2e (Behar *et al*., 2008). The five clades are divided into two main branches namely L2abcd and L2e (Behar *et al*., 2008).

Figure 3.2 mtDNA Phylogenetic relationship between African populations



● (blue) = non-Khoi-Sans, ● (green) = Khoi-Sans, ● (orange) = both Khoi-Sans and non-Khoi-Sans, ● (red) = haplogroups that moved out of Africa to repopulate Asia and the rest of the planet. The polymorphisms that characterise the different haplogroups were outlined in Appendix C Table A3. Adapted from Behar *et al.* (2008).

West Africa also harbours L3b and L3d mtDNAs but a few are in East Africa as a result of the Bantu and Khoi-San interactions (Salas *et al*., 2004a). Haplogroup L3e2 is more prevalent in Northern Africa but is also present in Western Africa as a result of migration from Northern Africa (Salas *et al*., 2004a). This haplogroup forms part of the L3eikx cluster which consists of four sub-haplogroup lineages that are defined by C150T and A10819G transition polymorphisms (Behar *et al*., 2008). Haplogroup L3e1 and L3e3 are widely distributed over Africa as a result of the Bantu migrations (Salas *et al*., 2002; 2004a). This cluster is differentiated from each other by the polymorphisms T2352C and T14212C for haplogroup L3e (Behar *et al*., 2008); T7645C for L3i; G3483A, 5899 Ins c, A6401G, T8311C, A8817G, G13708A, C16169T for L3x; A235G, C494T, G3918A, T6620G, T9467C, G13135A, C13992T and G15314A for L3k (Behar *et al*., 2008). An example of an L3e haplogroup sequence included a single sample from the Mandinga in Guinea Bissau (Behar *et al*., 2008).

Haplogroup L2a is the most frequent L2 haplogroup sub-clade in Africa and has a wide distribution all over the continent (Torroni *et al*., 2001). However, haplogroup L2 is divided into three major clusters namely L2a, L2bc and L2d (Behar *et al*., 2008). The polymorphisms characterising each of the L2 clades are indicated in Appendix C in Table A3. Haplogroup L1 is characterised by transitions at nucleotide positions G3666A, A7055G, T7389C, T13789C, T14178C, G14560A (Behar *et al*., 2008). Its branch, L1b is prevalent in Western Africa (Salas *et al*., 2002; Salas *et al*., 2004a). The L1c branch is prevalent in Angola but not common in Mozambique and came to southeastern Africa by the western Bantu immigrants (Salas *et al*., 2004a). Biaka pygmies were observed to have a high prevalence of L1c mtDNAs (Salas *et al*., 2004a). Haplogroup L1b differs from L1a at nucleotide position 9072 (Wallace, 2004). L1b has a G allele at the position while L1a has an A allele (Wallace, 2004). The polymorphisms further characterising haplogroup L1 are listed in Table A3 of Appendix C.

L0a1 and L0a2 haplogroups are present in significant proportions in Eastern and south-eastern Africa (Salas *et al*., 2004a). L0a1 differs from L0a2 in having the transitions T5096C and C16168T (Behar *et al*., 2008) whereas L0a2 is characterised by C64T, G185G, G5147A, A5711G, G6257A, 8281-9d, A8460G, A11172G, G16129G and T16519C polymorphisms (Behar *et al*., 2008).

3.3.1.1 mtDNA phylogenetic analysis of East African populations

East Africa is characterised by macrohaplogroup L lineages (Brandstatter *et al*., 2004; Gonder *et al*., 2007; Kivisild *et al*., 2004) similar to other African regions (Chen *et al*., 1995; Behar *et al*., 2008). Indeed, it is in East Africa that macrohaplogroup L originated (Salas *et al*., 2004a; Gonder *et al*., 2007). East Africa harbours haplogroup M sequences that are characterised by the T489C and T10873C polymorphisms (Quintina-Murci *et al*., 1999). The East African haplogroup M1 haplotypes are characterised by transitions at np G16129A, T16189C, A16249G, T16311C in the hypervariable region I and are divided into four clades (Quintina-Murci *et al*., 1999).

Haplogroups L0f, L3f, L4g (previously named L3g, Kivisild *et al*., 2004) and L5a (previously named L1e, Kivisild *et al*., 2004) are all of East African origin while the western African L3f1 is derived from East Africa (Salas *et al*., 2002; Salas *et al*., 2004a). It is postulated that haplogroup L4g originated from Eastern Africa and migrated to western and south-eastern Africa (Bortoloni *et al*., 2004; Kivisild *et al*., 2004). The ancestral L4g haplotype is characterised by the sequence motif C16223T-A16293T-T16311C-C16355T-T16362C (Salas *et al*., 2002; Bortoloni *et al*., 2004) and HVS II T146C and A244G transitions (Kivisild *et al*., 2004), and has been identified at a frequency of 52% but of very low haplotype diversity among the Tanzanian Hadzabe (Kivisild *et al*., 2004). The L4g haplotypes have been observed to exhibit a high diversity in Ethiopia (Kivisild *et al*., 2004). Ethiopia is also characterised by haplogroups HV, J, T and U as a result of the flow of genes from the Near East (Richards *et al*., 2003). Two Ethiopian haplogroup L3i samples were characterised by the control region polymorphism T7645C and the HVS I motif G16153A-T16223C-G16319A (Kivisild *et al*., 2004). The three Sudanese samples sequenced by Kring *et al.* in 1999, which possessed the G16129A-G16153A-T16223C HVS I motif should be close relatives of the L3i sequences (Kivisild *et al*., 2004). Haplogroup L3x has a 12% frequency among the Oromos from Ethiopia and is characterised by transitions at nps A6401G, G13708A and C16169T (Kivisild *et al*., 2004). L3x bifurcates into Lx1 defined by a transition at np C16278T and Lx2 which is defined by C16223T, C16193T and T16195C mtDNA polymorphisms (Kivisild *et al*., 2004). Haplogroup L3w and L4a have an East African geographical restriction (Kivisild *et al*., 2004). L3w is defined by the mtDNA mutations T15388C and C16260T while L4a is defined by G16388A (Kivisild *et al*., 2004).

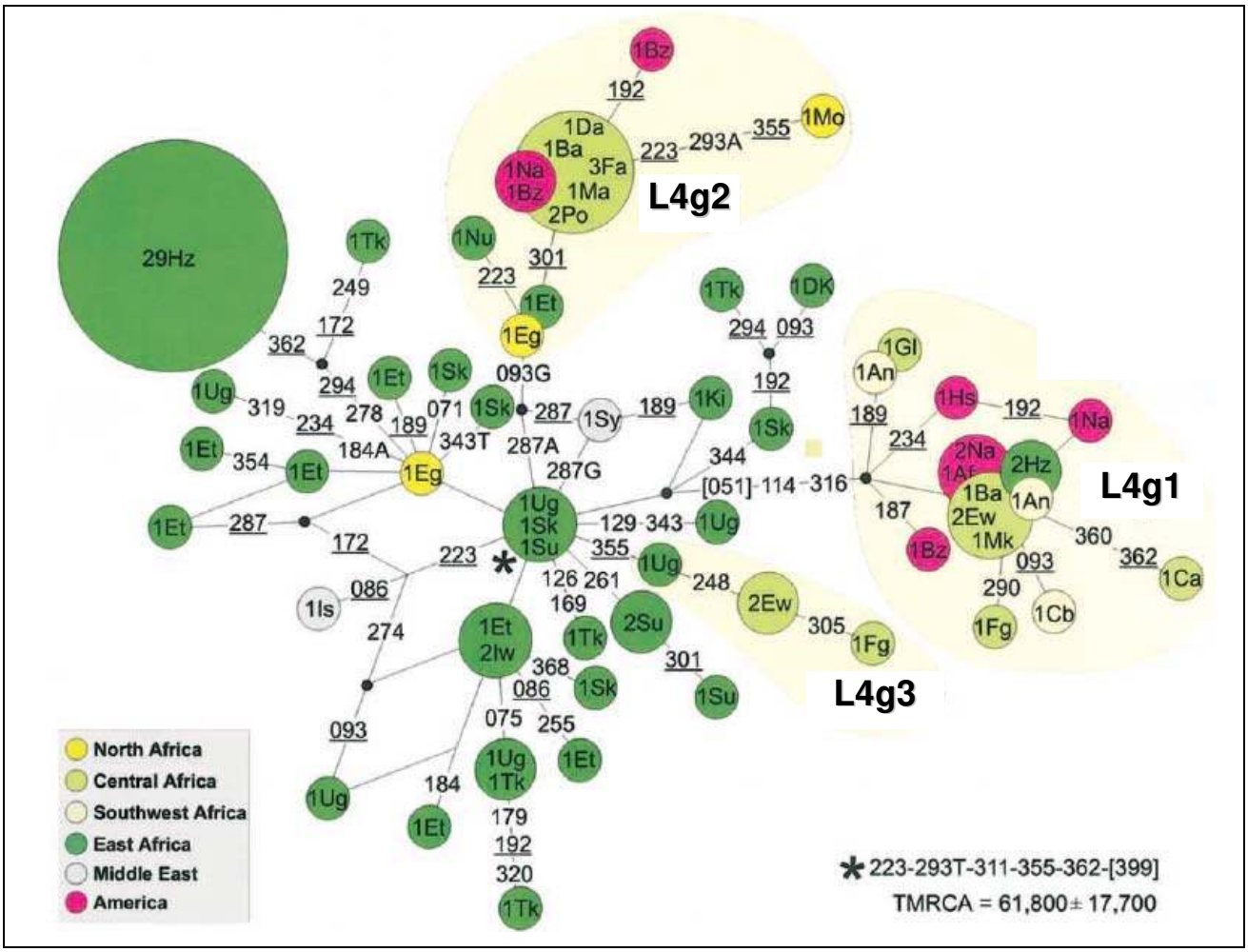
East Africa is also the ancestral origin of haplogroup L6 (Kivisild *et al*., 2004). The L6 haplogroup is characterised by the G709A, C770T, T961C, C13710T, T15289C, C15499T and T16224C mtDNA polymorphisms (Kivisild *et al*., 2004). It is closely related to haplogroup L2, L3 and L4 (Kivisild *et al*., 2004). The predominantly Western African L3d haplotypes have a HVS-1 sequence motif (T16124C-T16223C-G16319A) that was also harboured by samples from Tanzania (Sukuma) which normally exhibit a high number of L0a2 mtDNA sequences of Bantu speaking lineages (Kivisild *et al*., 2004). Ethiopian L3d2 sequences were distinguishable from the West African L3d2 samples by the T16368C transitional polymorphism (Kivisild *et al*., 2004). Haplogroup L4 shares transitions at np G709A and G1018A with L3 from which it was derived (Kivisild *et al*., 2004). Haplogroup L4a1 is characterised by the mtDNA polymorphisms T195C, C198T, C7376T, A16207G and C16260T, and has been observed in Ethiopia and Sudan (Kivisild *et al*., 2004). The Ethiopian haplogroup L2b samples resemble those from West Africa except that the Ethiopian samples have the G16145A polymorphism that is absent among West African samples (Kivisild *et al*., 2004). The L5 haplogroup is mainly common in East Africa and is divided into two main groups namely L5a and L5b (Kivisild *et al*., 2004; Behar *et al*., 2008). The characteristic polymorphisms for L5 include 459Ins C, T3423C, A7972G, C12432T, A12950G, C16148T and A16166G (Behar *et al*., 2008).

3.3.1.2 mtDNA phylogenetic analysis of Ugandan populations

Very little characterisation of the Ugandan population at the genetic level has been performed. Salas *et al*. (2004b) analysed the mtDNA variation in the control region among six Ugandans belonging to haplogroup L4g. The diversity of the L4g Ugandan sequences was as high as that exhibited by L4g sequences in Sudan, Tanzania and Ethiopia (Salas *et al*., 2004b). However, the ethnicity of these L4g sequences was not specified. The phylogenetic relationship between the Ugandan L4g sequences with other L4g sequences has been illustrated in Figure 3.3. Macaulay *et al.* (2005 Accession # AY963583) reports a Ugandan mtDNA sequence of haplogroup L0 with 98 alterations relative to the rCRS of mtDNA, but it is still of an unidentified ethnicity. An additional haplogroup L0 Ugandan sequence was characterised by Horai *et al*. (1995 Accession # D38112). Therefore only a total of six mtDNA control region sequences (Salas *et al*., 2004b) and two whole mtDNA sequences of the Ugandan population have been established to date (Horai *et al*. (1995 Accession # D38112); Macaulay *et al.* (2005 Accession # AY963583). Furthermore, phylogenetic analysis of the Ugandan population based on the Y-chromosome is lacking (Cruciani *et al*., 2002) except for the Karamojong nomads (Nilotic ethnic group) who were

established to be a distinct genetic grouping that fell between the Bantu and Ethiopians samples in a multidimensional scaling (MDS) plot (Gomes *et al*., 2009). There is therefore a necessity to characterise the mtDNA heritage of the Ugandan population in order to precisely define the evolutionary and phylogenetic relationships of its ethnic groups as well as the demographic history of the population. The root mtDNA L4g ancestral sequence harbours the C16223T-A16293T-T16311C-T16362C-A16399G polymorphic motif (Salas *et al*., 2004b). There are three distinct sublineages in haplogroup L4g namely L4g1, L4g2 and L4g3 (Salas *et al*., 2004b). Sublineage L4g1 harbours the A16051G-C16114T-T16189C-A16316G sequence motif, L4g2 is defined by T16093G and C16287A polymorphisms and L4g3 does not possess the C16355T transition when compared to the root (Salas *et al*., 2004b).

Figure 3.3 mtDNA phylogenetic relationship involving 6 Ugandan samples



Median – joining network of mtDNAs of haplogroup L4g. The frequency of the haplotypes is proportional to the area of the circles. The numbers represent nucleotide variants from the Revised Cambridge Reference Sequence less 16000. A transversion is indicated by a letter placed in front of the nucleotide position while an underlined number depicts a parallel mutation. Eastern Africa: Tz = Tanzania, Hz = Hadza (Tanzania), Dk = Dokota (Tanzania), Sk = Sukuma (Tanzania), Iw = Iraqw = (Tanzania), Ki = Kikuyu (Kenya), Tk = Turkana (Kenya), Ug = Uganda, Et = Ethiopia, Su = Sudan, Nu = Nubia (Sudan/Egypt). Northern Africa: Eg = Egypt, Mo = Berber (Morocco). Central Africa: Ew = Ewondo (Cameroon), Ba = Bakaka (Cameroon), Da = Daba (Cameroon), Fa = Falis (Cameroon), Ma = Mandara (Cameroon), Po = Podowkos (Cameroon), Ca = Cameroon, Fg = Fang (Gabon), Mk = Makina (Gabon), Gl = Galoa (Gabon). South-western Africa: Cb = Cabinda, An = Angola. America: Af = African American, Bz = Brazil, Hs = Hispanic (North America), Na = North America. Middle East: Sy = Syria, Is = Israel. TMRCA = time to the most recent common ancestor. Adapted with permission from Salas *et al*. (2004) with naming of haplogroup L3g as L4g as recommended by Kivisild *et al*. (2004).

3.4 OBJECTIVES OF THE STUDY

The major aim of this investigation was to establish the phylogenetic relationships among the Baganda, Lugbara and Acholi in Uganda via mtDNA sequencing and phylogenetic analysis. To date, only two complete mtDNA sequences of the Ugandan population have been established (Horai *et al.* (1995 Accession # D38112); Macaulay *et al.* (2005 Accession # AY963583) and so it was envisaged that this investigation will make a significant contribution to the human global mtDNA phylogenetic tree both in the context of human origin and disease.

3.4.1 Specific objectives

In order to achieve the specified overall objective of this project, three specific objectives were undertaken. The specific objectives were as follows:

3.4.1.1 Determine the complete mtDNA sequence of 13 Baganda, 14 Lugbara and 13 Acholi individuals in Uganda via an automated sequencing strategy.

3.4.1.2 Establish the phylogenetic relationship among the Baganda, Lugbara and Acholi from Uganda via phylogenetic tree construction.

3.4.1.3 Establish the phylogenetic relationship among the Baganda, Lugbara and Acholi from Uganda with other African populations sequenced thus far via phylogenetic tree construction.

CHAPTER FOUR

Materials and Methods

Only standardised protocols were used throughout this investigation. Chemicals and reagents of analar grade were used in this study. Unless otherwise stated, reagents were supplied by Promega®[1]. DNA was extracted from blood using the Flexigene™[2] DNA extraction kit prior to analysis via automated cycle sequencing. The generated sequences in turn were analysed for the phylogenetic significance of the observed polymorphisms.

4.1 ETHICS AND REGULATORY APPROVAL

Ethical approval for this project was granted by the Research and Ethics Committee of Makerere University Medical School, Uganda, with approval number REC REF No. 2005-035 and the Ethics Committee of the North-West University (Potchefstroom Campus), South Africa in 2005 with the title "mtDNA polymorphisms in three selected Ugandan populations". Informed consent was obtained from the individuals prior to the collection of blood samples. Permission to transfer the blood samples from Uganda to South Africa was granted by the Uganda National Council for Science and Technology under approval number MV 849. An import permit was obtained from the Department of Health, South Africa, with reference number J1/2/4/1 No 1.

4.2 SAMPLE POPULATION

A total of 40 blood samples from presumably healthy individuals (ascertained through oral interview) belonging to three ethnic groups were included in this investigation. Only healthy individuals were used since some diseases can have an effect on the mtDNA genotype and hence bias the phylogenetic relatedness of the individuals. The cohort consisted of 13 Baganda (Bantu), 14 Lugbara (Moru-Madi) and 13 Acholi (Nilotics) individuals from Uganda, specifically selected from Kampala but originating from diverse ancestral geographical sites. The selected individuals had to have maternal lineages which belonged to the same tribe for at least three generations while the paternal lineage did not

[1] Promega® is a registered trademark of Promega Corporation, Madison, WI, USA.
[2] Flexigene™ is a trademark of QIAGEN Corporation, Hilden, Germany.

necessarily have to be the same over this time since mtDNA is maternally inherited. Any individual who could not confirm his or her ancestral lineage was not enrolled in this project. Twenty millilitres (mL) of blood was collected from each of the study participants by venous puncture and stored in ethylenediamine tetraacetic acid (EDTA) Vacutainer®[1] tubes at 4 degrees Centigrade (ºC), for not more than three days, after which the blood samples were frozen at -80°C for longer storage prior to DNA extraction.

4.3 EXTRACTION OF DNA FROM WHOLE BLOOD

Extraction of DNA was performed using the Flexigene™ DNA extraction kit. Lyophilized QIAGEN®[2] protease was resuspended in 1.4 mL of hydration buffer (buffer FG3). Three hundred (300) μL of blood were used in each extraction. For every 300 μL of blood processed, 150 μL of buffer FG2 was mixed with 1.5 μL of QIAGEN® protease. The mixture of buffer FG2 and QIAGEN® protease was made within an hour of being used. The principle of the technique involved pelleting the mitochondria and cell nuclei from the blood by centrifugation, digesting the contaminating proteins using proteases, precipitating DNA by use of isopropanol, recovering the DNA by centrifugation, and washing the DNA using ethanol.

Frozen blood was thawed quickly at 37°C in a water bath while agitating the samples mildly. The thawed blood was thereafter placed on ice for ca. 15 min. Seven hundred fifty (750) μL of buffer FG1 was pipetted into a centrifuge tube of capacity 1.5 mL after which 300 μL of blood was added and the tubes were gently inverted five times to mix the solution. The sample was centrifuged at 10,000 x g for 60 s in a rotor at a fixed angle. The supernatant was discarded with care so that the pellet remained in the tube and the tube was left inverted on an absorbent paper for 2 min, to minimise the small traces of the supernatant that are left on the rim and side walls of the tube flowing back onto the pellet.

A volume of 150 μL of the buffer FG2/QIAGEN® protease mixture was added to the tube, the tube was closed and vortexed immediately so that the pellet was homogenised completely. When multiple samples were processed, each tube was vortexed immediately after the addition of the buffer FG2/QIAGEN® protease mixture. Generally 4 pulses each of ca. 5 s of high speed vortexing were sufficient to completely homogenise the pellet. If

[1] Vacutainer® is a registered trademark of Becton, Dickinson and Company, New Jersey, USA.
[2] QIAGEN® is a registered trademark of Qiagen Corporation, Hilden, Germany.

persistent jelly-like traces of pellet were observed, the sample was further homogenised by the addition of 30 μL of buffer FG2 and vortexed as before.

The tubes were briefly centrifuged for 5 s and thereafter incubated in a water bath at 65°C for 5 min. The red-coloured samples changed to olive green, as a sign that the digestion of proteins had taken place. The DNA was precipitated by the addition of 150 μL of 100% isopropanol. The samples were thoroughly mixed with isopropanol by inversion until the the precipitate of DNA could be seen in the form of threads or a clump. For samples in which the DNA was not visible, due to very low counts of white blood cells, the tubes were inverted at least 20 times.

The samples were thereafter centrifuged at 10,000 x *g* for 3 min. The supernatant was discarded and the tubes briefly inverted over a piece of clean absorbent paper. Care was taken to ensure that the pellet did not fall out of each of the tubes. One hundred fifty μL of 70% ethanol was added to each sample and the samples were vortexed for 5 s. The samples were centrifuged for 3 min at 10,000 x *g*, the supernatant was discarded and the tubes were inverted for 5 min on a piece of clean absorbent paper. The DNA was air-dried for 5 min and to each sample, 200 μL of buffer FG3 was added, followed by vortexing at a low speed for 5 s. The DNA was dissolved by incubating it in a 65°C water bath for 10 min. Thereafter the DNA was kept overnight at room temperature to ensure that complete dissolution had been attained. The concentration and purity of the DNA was thereafter measured as outlined in Section 4.4.

4.4 DETERMINATION OF DNA CONCENTRATION

The concentration of the isolated DNA was determined spectrophotometrically by measuring the absorbance at 260 nm (A_{260}) using an Eppendorf®[1] Biophotometer. The Biophotometer uses a ratio of one absorbance unit to 50 ng.μL^{-1} double stranded DNA (dsDNA) to calculate the DNA concentration of a specific sample. The DNA was diluted 1 in 5 before optical density was measured to achieve an absorbance value in the range of Beer-Lambert's law (Segel, 1976).

The absorbance at 280 nm (A_{280}) was also recorded to determine the level of protein contamination in the isolated DNA. The A_{260}/A_{280} ratio is a measure of the degree of

[1] Eppendorf® is a registered trademark of Eppendorf-Netheler-Hinz, Hamburg, Germany.

protein contamination, which is acceptable if the value is equal to or above 1.8. A DNA working dilution of 25 ng.μL^{-1} was prepared from the stock DNA solution and stored at 4°C while the rest of the stock DNA was placed at -20°C for longer preservation.

4.5 DNA AMPLIFICATION

The DNA was amplified via the polymerase chain reaction (PCR) using the primers and conditions listed in Table 4.1, in a Thermo Hybaid®[1] MBS®[2] thermocycler. The protocol used is a modified version of one described by Mullis *et al*. (1986). The conditions described below for amplification were subjected to optimisation. The sequences of the oligonucleotide primers used for amplification were according to Maca-Meyer *et al*. (2001).

Table 4.1 Primers used for amplification of the entire human mtDNA genome

Primer	Sequence (5'-3')	CRS reference	T_M	Avg T_M	Size (bp)
L15996	F: ctc cac cat tag cac cca aag c	15975-15996	64.5	64.5	2,103
H1487	R: gta tac ttg agg agg gtg acg g	1508-1487	64.5		
L923	F: gtc aca cga tta acc caa gtc a	902-923	60.8	60.7	2,789
H3670	R: ggc gta gtt tga gtt tga tgc	3690-3670	60.6		
L3073	F: aaa gtc cta cgt gat ctg agt tc	3051-3073	61.0	62.8	2,277
H5306	R: ggt gat ggt ggc tat gat ggt g	5327-5306	64.5		
L4750	F: cca ata cta cca atc aat act c	4729-4750	57.1	58.8	2,190
H6899	R: gca ctg cag cag atc att tc	6918-6899	60.4		
L6337	F: cct gga gcc tcc gta gac ct	6318-6337	66.6	63.7	2,544
H8861	R: gag cga aag cct ata atc act g	8882-8861	60.8		
L7882	F: tcc ctc cct tac cat caa atc a	7861-7882	60.8	60.0	2,090
H9928	R: aac cac atc tac aaa atg cca gt	9950-9928	59.2		
L8799	F: ctc gga ctc ctg cct cac tca	8779-8799	66.5	64.5	2,749
H11527	R: caa gga agg ggt agg cta tg	11546-11527	62.5		
L10403	F: aaa gga tta gac tga acc gaa	10383-10403	56.7	57.6	3,264
H13666	R: agg gtg ggg tta ttt tcg tt	13685-13666	58.4		
L12572	F: aca acc cag ctc tcc cta ag	12553-12572	62.5	62.6	2,153
H14685	R: cat tgg tcg tgg ttg tag tcc	14705-14685	62.6		
L14125	F: tct ttc ttc ttc cca ctc atc c	14104-14125	60.8	60.6	2,317
H16401	R: tga ttt cac gga gga tgg tg	16420-16401	60.4		

Avg = average, bp = base pairs, CRS = Cambridge reference sequence, F = forward primer, R = reverse primer, T_m = melting temperature. The primers used are those designed according to the procedure followed by Maca-Meyer *et al*. (2001).

The reaction constituents included double distilled water (ddH_2O), 1 x PCR buffer [50 mM potassium chloride (KCl), 10 mM Tris®-hydrochloride of pH 9.0 and 0.1% Triton® X-100],

[1] Thermo Hybaid® is a registered trademark of Hybaid Limited, Ashford, Middlesex, UK.
[2] MBS® is a registered trademark of Thermo Electron Corporation, Milford, MA, USA.

1.5 mM magnesium chloride ($MgCl_2$), 200 μM of each of the 2'-deoxyribonucleotide-5'-triphosphates (dNTPs), 5.0 picomoles of each of the respective forward and reverse primers, 0.5 units of Go *Thermus aquaticus* (Go*Taq*™) DNA polymerase (Glebs *et al*., 2003) and 50 ng gDNA in a volume totalling to 12.5 μL.

In the first stage of cycling, the samples were heated to 94°C for 10 min. As indicated in Table 4.2, denaturation of DNA took place at this stage. In the second stage, the samples were incubated at 94°C for 1 min, at the annealing temperature (Table 4.2) for 1 min and at 72°C for 2 or 3 min depending on the length of fragment to be amplified. This cycle was repeated a total of 30 times during which amplification occurred. In the third stage, the samples were incubated at 72°C for 10 min. This stage served to ensure that all PCR products were synthesised to their full length. In the last stage, the cooling of the samples was performed till 4°C and held at that temperature indefinitely until they were removed from the thermal cycler. The samples were kept at 4°C in a refrigerator until purified.

In general, the annealing temperature for amplification was optimised starting from a temperature 2°C below the calculated mean melting temperature (T_m) of the primer set for the specific DNA region of interest. The $MgCl_2$ concentration was optimised by varying its concentration until sufficient amplification, in terms of both yield and quality, had been achieved. The duration of the cycles was also varied one at a time until sufficient amplification was obtained.

Table 4.2 PCR conditions for amplification of mitochondrial DNA

PCR step	# of cycles	Purpose	Temperature	Duration
1	1	Denaturation	94°C	10 min
2	30	Denaturation	94°C	1 min
		Annealing	X°C	1 min
		Elongation	72°C	2 or 3 min
3	1	Elongation of incomplete fragments	72°C	10 min
4	1	Hold	4°C	Indefinite

= number, PCR = polymerase chain reaction, x = the annealing temperature for each pair of primers as indicated in Table 4.1.

The estimated annealing temperature (T_a) of the primer set served as the starting point from where it could be increased or decreased to achieve optimal amplification. The Oligonucleotide Properties Calculator software (Oligonucleotide Properties Calculator, 2007) was used by a contracted company to determine the T_m for each pair of primers

through the nearest neighbour and thermodynamic calculation methods by using the parameters of Sugimoto *et al*. (1996).

4.6 AGAROSE GEL ELECTROPHORESIS

A 0.9 percent (%) [w/v] agarose gel, in 1 x TBE buffer (8.915 mM Tris®, pH 8.1, 8.895 mM boric acid, 0.2498 mM Na_2EDTA) was used for electrophoresis at 10 volts per centimetre ($V.cm^{-1}$) for 40 min in order to verify the amplification of the mtDNA regions of interest prior to cycle sequencing. A volume of 2.5 μL of ethidium bromide (EtBr) of concentration 10 $mg.mL^{-1}$ was added to 40.0 mL of gel. A volume of ca. 2.5 μL of each amplified DNA sample was mixed with the 2 X loading buffer (0.04% orange G and 50% glycerol) and loaded onto the gel.

The remaining DNA samples were stored at 4°C for purification and sequencing. An amount of 70 ng of the FastRuler™[1] DNA ladder, high range molecular weight marker was mixed with 2.5 μL of the loading buffer and loaded onto the gel. Visualisation of the amplified DNA fragments was performed with an ultra-violet (UV) transilluminator. The gel images were captured by a camera and stored electronically.

4.7 PURIFICATION OF PCR PRODUCTS

Purification of the PCR products was performed using the DNA clean & concentrator - 5™[2] kit. The amplified fragments of DNA were purified from mineral oil, primers, nucleotides, polymerases and salts. All centrifugation steps were conducted at 17,900 x g in an Eppendorf™ centrifuge 5810 R. DNA, when in solution of a high salt concentration, undergoes selective adsorption to the uniquely designed silica membrane, while the contaminants, not adsorbed to the column, pass through it. The pure DNA was finally eluted from the column with ddH_2O.

[1] FastRuler™ is a trademark of Fermentas, Harrington Court, Burlington, Canada.
[2] DNA clean & concentrator – 5™ is a trademark of Zymo Research Corporation, California, USA.

Two volumes of DNA binding buffer were added to every volume of the PCR samples (excluding the volume of the mineral oil) in a microcentrifuge tube of 1.5 mL capacity. The contents were mixed by pipetting five times. The samples were applied to the Zymo-Spin™[1] columns and centrifuged for 1 min. The flow-through was disposed of and the Zymo-Spin™ columns placed back into the same tubes. To wash the DNA of impurities, 200 μL of wash buffer was added to each of the Zymo-Spin™ columns and centrifuged for 1 min. This step served to remove the impurities or unbound material. The supernatant was discarded again, and the Zymo-Spin™ columns placed back in the tubes they had been removed from. The wash buffer (200 μL) was pipetted into each of the columns and centrifuged for an additional 1 min to remove any impurities still in the column.

The column was placed in a microcentrifuge tube of capacity 1.5 mL and the DNA was eluted in 10 μL of ddH_2O via centrifugation for 1 min. Thirty μL of ddH_2O was added to each of the samples and the concentration of the purified DNA was determined by measuring its absorbance using an Eppendorf™ Biophotometer after diluting the DNA by a factor of five so as to obtain concentrations and absorbance values in the range of Beer-Lambert's law (Segel, 1976).

The sequencing reaction consisted of the following reagents which were pipetted in the order indicated: ddH_2O, 2.0 μL of 5 x sequencing buffer, 2.0 μL of ready reaction mix, 1.0 μL of primer (3.2 μM) and a specified amount of DNA into 0.2 mL tubes. The amount of DNA used in each sequencing reaction was 50 ng for templates that were 2000 – 2500 bp in length. For templates that were 2500 - 3264 bp in length, 60 ng of DNA was used. The total volume of DNA and water in each tube was 5.0 μL. All the other volumes in the protocol remained fixed.

4.8 AUTOMATED SEQUENCING ANALYSIS

Automated sequencing analysis (Wilson *et al*. 1990; Rosenblum *et al*., 1997) was used to identify the whole mtDNA genome sequence. Automated sequencing was performed using the ABI Prism®[2] BigDye™[3] Terminator version 3.1 (v3.1) Ready Reaction Cycle

[1] Zymo-Spin™ is a trademark of Zymo Research Corporation, California, USA.
[2] ABI Prism® is a registered trademark of Applera Corporation, Forster City, CA, USA.
[3] BigDye™ is a trademark of Applera Corporation, Forster City, CA, USA.

Sequencing Kit and the primers listed in Table 4.3 in a Thermo Hybaid® MBS® thermocycler using a modification of the procedure described by Sanger *et al*. (1977).

The samples were cycled at 94°C for 10 seconds, 50°C for 5 seconds and 60°C for 4 min for a total of 25 cycles. The samples were thereafter rapidly thermal-ramped to 4°C and held until ready to precipitate as outlined in Section 4.9.

Table 4.3 Primers used for sequencing the entire human mitochondrial genome

Primer	Sequence (5'-3')	CRS reference
L15996	F: ctc cac cat tag cac cca aag c	15975-15996
L16340	F: agc cat tta ccg tac ata gca ca	16318-16340
L382	F: caa aga acc cta aca cca gcc	362-382
L923	F: gtc aca cga tta acc caa gtc a	902-923
L1466	F: gag tgc tta gtt gaa cag ggc c	1445-1466
L2025	F: gcc tgg tga tag ctg gtt gtc c	2004-2025
L2559	F: cac cgc ctg ccc agt gac aca t	2538-2559
L3073	F: aaa gtc cta cgt gat ctg agt tc	3051-3073
L3644	F: gcc acc tct agc cta gcc gt	5851-5832
L4210	F: cca ctc acc cta gca tta ctt a	4189-4210
L4750	F: cca ata cta cca atc aat act c	4729-4750
L5278	F: tgg gcc att atg gaa gaa tt	5259-5278
L5781	F: agc ccc ggc agg ttt gaa gc	5762-5781
L6337	F: cct gga gcc tcc gta gac ct	6318-6337
L6869	F: ccg gcg tca aag tat tta gc	6850-6869
L7379	F: aga aga acc ctc cat aaa cct g	7358-7379
L7882	F: tcc ctc cct tac cat caa atc a	7861-7882
L8299	F: acc ccc tct aga gcc cac tg	8280-8299
L8799	F: ctc gga ctc ctg cct cac tca	8779-8799
L9362	F: ggc cta cta acc aac aca cta	9342-9362
L9886	F: tcc gcc aac taa tat ttc act t	9865-9886
L10403	F: aaa gga tta gac tga acc gaa	10383-10403
L10949	F: ctc cga ccc cct aac aac cc	10930-10949
L11486	F: aaa act agg cgg cta tgg ta	11467-11486
L12028	F: ggc tca ctc acc cac cac att	12008-12028
L12572	F: aca acc cag ctc tcc cta ag	12553-12572
L13088	F: agc cct act cca ctc aag cac	13068-13088
L13612	F: aag cgc cta tag cac tcg aa	13593-13612
L14125	F: tct ttc ttc ttc cca ctc atc c	14104-14125
L14650	F: ccc cat tac taa acc cac act c	14629-14650
L15162	F: ctc ccg tga ggc caa ata tc	15143-15162
L15676	F: tcc cca tcc tcc ata tat cc	15657-15676

CRS = Cambridge reference sequence, F = forward primer. The primers used were those designed by Maca-Meyer *et al*. (2001).

4.9 SODIUM DODECYL SULPHATE (SDS) CLEANUP OF THE SEQUENCED PRODUCTS

The protocol as proposed by the BigDye®[1] cycle sequencing kit (2002) was used for the clean up of sequenced products. To every 10 µL of the sequenced product, 1 µL of a 2.2% solution of sodium dodecyl sulphate (SDS) in deionised water was added. The samples were vortexed briefly at low speed. SDS eliminates excess or unbound nucleotides and reduces on the levels of background noise which improves the readability of sequence peaks (BigDye® cycle sequencing kit, 2002). The tubes were placed in a thermal cycler and cycled at 98°C for 5 min, 25°C for 10 min and held indefinitely at 4°C. The tubes were thereafter spun briefly before being transported to a contractor who performed the post sequencing electrophoresis of the samples.

4.10 PURIFICATION OF SEQUENCED PRODUCTS

Purification of the sequenced products was performed by the contractor electrophorising the sequenced samples, and was achieved using the Centri-Sep™[2] high throughput dye terminator purification kit (2007) with 96 well filter plates. The filter plates were pre-packed with hydrated, cross-linked gel. Eight µL of each sample was transferred to the centre of the gel bed within each well. The 96 well PCR plates were put into the deep well plates and the Centri-Sep™ plates were stacked on top. The samples were centrifuged at 1,500 x *g* for 2 min. The 96 well PCR plates were removed and dried in a speed-vacuum.

The method of gel filtration (rather than alcohol precipitation) was adopted in this investigation and served to purify the sequenced DNA products (Finn *et al.*, 2003). Gel filtration ensured selective removal of unreacted fluorescently-labelled dideoxyribonucleotides, excess salt and primers without loss of the shorter sequenced DNA fragments (Finn *et al.*, 2003). The excess dye-terminators and nucleotides, when not purified, co-migrate during electrophoresis with the labelled DNA products thereby reducing on the readability of sequences of the first 20 - 100 or more bases of the electropherogram (Finn *et al.*, 2003).

[1] BigDye® is a registered trademark of Applera Corporation, Forster City, CA, USA.
[2] Centri-Sep™ is a trademark of Princeton Separations Inc., Aldephia, USA.

At the contractor, the sequenced and purified product was injected into an Applied Biosystems®[1] 3130 xl genetic analyser for capillary gel electrophoresis and detection. Analysis of the electropherograms of the sequenced products was thereafter performed as discussed in Section 4.11.

4.11 ANALYSIS OF SEQUENCED DNA SAMPLES

After automated cycle sequencing (Section 4.8), the electropherograms of the individuals under study were analysed for single nucleotide polymorphisms (SNPs), deletions and insertions. The sequences were screened for readability or ambiguity by viewing the electropherograms using the BioEdit program for sequence alignment (Hall, 2001) and were in turn aligned with the rCRS (Andrews *et al*., 1999) still using the BioEdit software program for sequence alignment (Hall, 2001). Any sequence whose base recalling was poor was re-sequenced. A polymorphism was recorded to be present only if it was determined to be different from the rCRS upon alignment, and if it presented as an unambiguous peak upon review of the raw electropherogram. The significance of the polymorphisms was thereafter assessed as outlined in Section 4.11.3. As indicated in Sections 4.11.1 to 4.11.3, the NJ tree, MP tree and pairwise distances were calculated by external parties to ensure quality control and standardisation of the actual process within the department however all assertions and conclusions drawn from these outputs were determined by the author himself, thus requiring understanding and insight into the process of phylogenetic analysis.

4.11.1 Construction of phylogenetic trees

The NJ method (Saitou and Nei, 1987) and MP method (Saitou and Imanish, 1989) were used in this study for the construction and analysis of the evolutionary relationship within tribes and between tribes. The NJ tree was constructed by the use of the PHYLIP phylogenetic analysis package designed by Joseph Felsenstein (Tuimala, 2006; Felsenstein, 2008) while the MP tree was constructed by the use of the Molecular Evolutionary Genetic Analysis (MEGA 3.1) software version 3.1 (Kumar *et al*., 2004). The NJ tree and the MP tree[2] were constructed from the 40 Ugandan sequences, the African Reference sequence and the rCRS of mtDNA pooled together with a global sample[3] of

[1] Applied Biosystems® is a registered trademark of Applera Corporation, Forster City, CA, USA.
[2] The NJ tree and MP tree were constructed (computer processing) by Dr Poole, UCI, USA while the final formatting of the trees was performed by Dr G.W. Towers, Centre for Genome Research, North-West University (Potchefstroom Campus), South Africa.
[3] The global sample of African mtDNA coding region sequences was kindly provided by M. Koekemoer, Centre for Genome Research, North-West University (Potchefstroom Campus), South Africa.

African mtDNA sequences all belonging to macrohaplogroup L. All the sequences used in constructing the trees consisted of their mtDNA coding regions only. i.e. np 577 – 16023 (Achillis *et al*., 2004; Gonder *et al*., 2007), and were used to establish the genetic evolutionary relationship between the Baganda, Acholi and Lugbara as well as their genetic relationship with other African samples. Only the coding region was used to construct the phylogenetic trees so as to minimise the effect of back mutations which is high in the control region (Mishmar *et al*., 2003). The trees were rooted using a chimpanzee (*Pan troglodytes*) sequence as the closest non-human ancestor to human beings (Chen and Li, 2001; Graham *et al*., 2002).

The NJ tree method was selected as it was established to be an efficient distance method in retrieving the correct topology of the tree (Saitou and Imanishi, 1989; Gascuel and Steel, 2006). Moreover, the NJ method does not make any assumption of a constant evolutionary rate of lineages when constructing the phylogenetic tree (Rzhetsky and Nei, 1992). The MP method was chosen so as to compare its results with that of the NJ method since the former is a discrete method that constructs the tree based on the information stored in the sequences through an exhaustive search technique (Saitou and Imanish, 1989). The NJ method constructs the tree based on pair-wise distances between the different sequences and does not necessarily demonstrate the nature of the polymorphisms exhibited by the sequences (Gascuel and Steel, 2006). The NJ method however, does not give good branch length estimates (Tateno *et al*., 1994).

4.11.2 Construction of the NJ phylogenetic tree

The generated sequences were aligned to the rCRS using the BioEdit programme (Hall, 2001). Since the whole mitochondrial genome for each individual was sequenced in 32 overlapping fragments, contigs[1] were constructed using an assembly programme in BioEdit (Hall, 2001) after allowing for at least a 95% similarity between the sequences and an overlap of at least 10 bases. The non-coding regions and gaps were removed from the contigs before the sequences were aligned using ClustalX[2] multiple sequence alignment programme (Tuimala, 2004).

[1] The contigs were assembled by Dr G.W. Towers, Centre for Genome Research, North-West University (Potchefstroom Campus), South Africa.
[2] The sequences were aligned using ClustalX programme by Dr. G.W. Towers and M. Koekemoer, Centre for Genome Research, North-West University (Potchefstroom Campus), South Africa.

The model of nucleotide substitution used to calculate the NJ tree was the F84 model which assumes variation in the rate of occurrence of transitions and transversions and in the frequencies of the nucleotides (Felsenstein, 2008). The γ-shape parameter[1] (coefficient of variation) was measured using the GZ-Gamma PROGRAM to allow for the use of the model which assumes unequal rates of nucleotide substitution at different sites as is the case with mtDNA (Gu and Zhang, 1997). The transition to transversion ratio[2] was measured using the PHYLIP phylogenetic analysis package, since transitions occur more frequently than transversions, and therefore this ratio should never be assumed to be equal to 1.0 (Felsenstein, 2008). The following analyses were performed using the PHYLIP suite of programs[3]. A bootstrap analysis of the possible trees was conducted using SEQBOOT.exe program using 1,000 replicates (Tuimala, 2006; Felsenstein, 2008) so as to determine the statistical support for the tree branches that were to be obtained (Sitnikova *et al*., 1995). By the use of the established γ-shape parameter and the calculated transition to transversion ratio, pairwise distances were determined using the DNADIST.exe program and definition of the parameters was according to the F84 model which allows for differences in the rate of evolution at different sites of mtDNA via the gamma-shape parameter as well as variation in the occurrence of transitions and transversions in the sequences (Felsenstein, 2008). The PHYLIP program was used to measure the diversity of sequences within and between individuals which enabled the reconstruction of the genetic relatedness amongst individuals (Tuimala, 2006; Felsenstein, 2008).

The NJ tree was constructed using Neighbor.exe program of the PHYLIP package (Tuimala, 2006; Felsenstein, 2008). The majority rule consensus tree was thereafter constructed using consense.exe so as to generate a single tree from the variety of trees generated from the bootstrap analyses (Tuimala, 2006; Felsenstein, 2008).

4.11.3 Construction of the MP phylogenetic tree

The MP tree was constructed using MEGA 3.1 (Kumar *et al*., 2004). After assembling the contigs, the sequences (only coding region) were aligned using ClustalX[4], the pair wise

[1] The gamma shape parameter for this dataset was determined by Dr. G.W. Towers and M. Koekemoer, Centre for Genome Research, North-West University (Potchefstroom Campus), South Africa.
[2]The transition/transversion ratio was determined by Dr. G.W. Towers and M. Koekemoer, Centre for Genome Research, North-West University (Potchefstroom Campus), South Africa.
[3]Phylogenetic strategy for determination of the NJ tree and the MP designed by Michelle Koekemoer and performed (computational processing) by Dr J. Poole.
[4]The sequences were aligned using ClustalX program by Dr. G.W. Towers and M. Koekemoer, Centre for Genome Research, North-West University (Potchefstroom Campus), South Africa.

distances determined (Kumar *et al*., 2004) and the MP tree was constructed with a 1,000 bootstrap replicate analysis (Kumar *et al*., 2004).

Both the NJ and MP trees were manually assessed to reconstruct the evolutionary history of the individuals and the tribes. Both trees were furthermore rooted and formatted using TREEVIEW after which formatting of the edges and branches was performed using the MEGA 3.1.[1] The global African sample served to facilitate a detailed phylogeographical comparison of the Baganda, Acholi and Lugbara mtDNA sequences with their African counterparts and draw implications of their origin and peopling.

4.11.4 Comparisons of pairwise distances

The pairwise genetic differences[2] for the different samples used in this investigation were calculated using DnaSP (Librado and Rozas, 2009). Only the coding region sequences were used in generating the pairwise differences between the Baganda, Acholi and Lugbara samples in this investigation as well as the other macrohaplogroup L samples included in this study since it was the coding region used to construct the phylogenetic trees as well. The pairwise differences enabled a computation of the genetic variation amongst the Baganda, Acholi and Lugbara from Uganda and a retrieval of the demographic history of the population from the information stored in DNA (Librado and Rozas, 2009).

[1]Final formatting of the trees was performed by Dr G.W. Towers, Centre for Genome Research, North-West University (Potchefstroom Campus), South Africa.
[2] The computation of pairwise distances was performed by Dr W. Towers, North West University (Potchefstroom Campus), South Africa.

CHAPTER FIVE

Results and Discussion

The complete mtDNA sequences of 13 Baganda, 14 Lugbara and 13 Acholi Ugandan individuals belonging to the Bantu, Moru-Madi and Nilotic ethnic groups respectively were established and assessed. A phylogeny of the individuals has been inferred from their mtDNA sequences using the NJ and MP tree - generating methods. The implications of the findings have been presented and are discussed in the following sections.

5.1 OPTIMISATION OF EXPERIMENTAL PROCEDURES

The experimental procedures used in this study included DNA isolation from blood samples, PCR, agarose gel electrophoresis, purification of PCR products, DNA quantification via spectrophotometry, automated sequencing, precipitation and electrophoresis of the sequenced products. Apart from PCR, all the other experimental procedures involved the use of commercially available kits, which required minimal optimisation. However, for the PCR procedure, optimisation was performed with the use of control DNA samples.

5.1.1 Optimisation of DNA isolation

The Flexigene^{TM} DNA kit was used in the extraction of DNA from frozen blood samples as discussed in Section 4.3 of Materials and Methods. A time interval of about 15 min was instituted to thaw the frozen blood samples at 37°C. A final incubation time of 20 min instead of 10 min at 65°C was also used to ensure that the DNA pellet was fully dissolved in buffer FG3. Furthermore, the DNA was stored overnight at room temperature before absorbance was measured to ensure that full dissolution was achieved. The yield of DNA from the control blood samples (samples of blood drawn from anonymous control individuals that had been stored at 4°C for a few days or weeks) was 55.2 - 90.3 ng.μL^{-1} with the ratio of absorbance at 260 nm to 280 nm (A260/280) ranging from 1.64 to 2.01. The estimated DNA yield expected from the kit ranged from 55 - 70 ng.μL^{-1} and the optimal ratio of absorbance ranged from 1.70 – 1.90. However, the DNA yield from participants' blood samples ranged between 25.4 - 96.0 ng.μL^{-1}. The lower DNA yield in the study

participant samples was attributed to degradation of the DNA during initial storage, since the samples remained on ice for five days from the time that the blood was drawn in Uganda to the time it was transported to South Africa where it was aliquoted into collection tubes and stored at -80 °C. These DNA samples were however sufficient for PCR amplification and automated sequence analysis.

5.1.2 Optimisation of polymerase chain reactions

Three positive control samples of DNA with a concentration of 50 ng.μL^{-1} each (working DNA concentration which was obtained by appropriately diluting stock DNA) were first amplified using the set of conditions in Table 4.1. Primers synthesised according to those originally reported by Maca-Meyer *et al*. (2001) were used in amplification. Long fragments ranging from 2,090 to 3,264 bp were amplified in this study. The forward and reverse primers with the closest T_m range were selected for amplification of specific mtDNA fragments such that both primers anneal at nearly the same temperature (Viljoen *et al*., 2005). However, this optimum situation was not possible for the entire genome. Where the T_m of the primer pairs selected was nearly the same, a T_a of about 2°C below the mean T_m was chosen as the starting temperature for optimisation of PCR. Where the T_m varied more than 3 °C, the annealing temperature of the reaction was chosen starting 2°C below the melting temperature of the primer with the lowest T_m. The T_a was experimentally adjusted accordingly in order to generate a single specific fragment of sufficient concentration. Since the duration of extension varies with the length of the target sequence (Viljoen *et al*., 2005), an average time of 1 min per kb of fragment was generally used for the duration of the extension step during the PCR cycle. A $MgCl_2$ concentration of 1.5 mM was generally sufficient for successful amplification of all fragments while Go *Taq*^{TM} DNA polymerase was used as the polymerisation enzyme. Table 5.1 illustrates the optimised T_a used to amplify the DNA fragments of the entire mitochondrial genome. Fragments A, C, D, F, I and J were amplified as outlined in Section 4.5 using a duration of 2 min for extension of the PCR fragments at 72°C. The same cycle conditions were used for amplifying fragments B, E, G and H except that a time of 3 min was used for extension during the cycling steps, since these were longer fragments (Viljoen *et al*., 2005).

The entire genome was amplified in 10 overlapping fragments. The overlap, as indicated in Table 5.1, ranged from 405 to 600 bp for PCR fragments less than 2,500 bp and from 1,021 to 1,124 bp for overlaps involving fragments longer than 2,500 bp. A negative control having ddH_2O in place of DNA was included in each amplification phase and

absence of a visible fragment after electrophoresis of this sample ruled out any contamination of the samples. Presence of a visible fragment of expected length in the positive control confirmed that the experimental protocols were adhered to and the amplification process progressed as expected.

Table 5.1 Primers used for amplifying and sequencing of the entire human mt genome

PCR region	PCR primer	Overlap (bp)	Ta (°C)	Sequencing Primer
A	L15996 H1487	594	65.0	L15996
				L16340
				L382
B	L923 H3670	597	63.0	L923
				L1466
				L2025
				L2559
C	L3073 H5306	556	61.0	L3073
				L3644
				L4210
D	L4750 H6899	562	57.0	L4750
				L5278
				L5781
E	L6337 H8861	600	60.0	L6337
				L6869
				L7379
F	L7882 H9928	1021	58.5	L7882
				L8299
				L8799
G	L8799 H11527	1124	64.0	L9362
				L9886
				L10403
H	L10403 H13666	1094	56.5	L10949
				L11486
				L12028
				L12572
I	L12572 H14685	560	64.0	L13088
				L13612
J	L14125 H16401	405	60.5	L14125
				L14650
				L15162
				L15676

bp = base pairs, °C = degrees Centigrade, PCR = polymerase chain reaction, T_a = annealing temperature of the primer pairs used in PCR. All the fragments were amplified using Go *Taq*^{TM} DNA polymerase. Sequencing of fragment I using primer L12572 nearly always resulted in the loss of the first 250 bases (approximately) that was attributed to presence of a secondary product in fragment I as indicated in Figure 5.7 but the region was successfully sequenced using PCR fragment H.

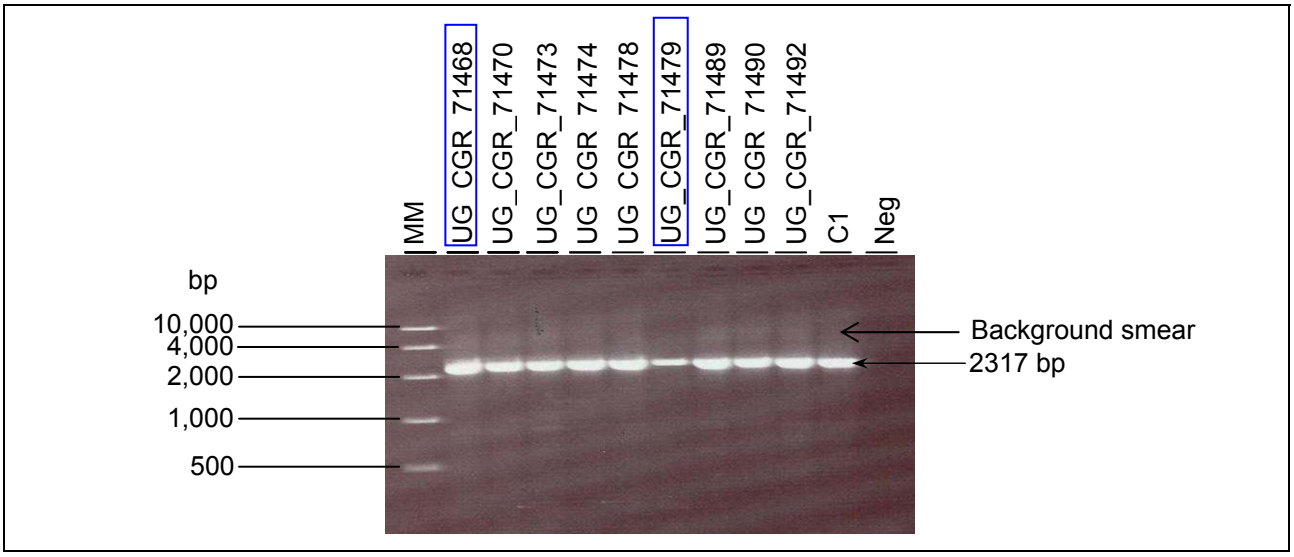
5.1.2.1 Artefacts observed in PCR amplified samples

Despite the fact that the PCR reaction should have high specificity, efficiency, sensitivity and fidelity (Cha and Thilly, 1993; Viljoen *et al*., 2005) even after observing the necessary precautions, a number of artefacts were generated in the PCR amplified samples. The PCR reaction was optimised by striking a compromise between specificity, efficiency and sensitivity (Cha and Thilly, 1993; Viljoen *et al*., 2005). However, in most instances, the artefacts that were obtained did not interfere with the downstream processes that were to be conducted on the amplified products. The different artefacts generated (both those that interfered and those that did not interfere with the downstream processes) in the course of PCR in this study are discussed in subsequent sections of this chapter.

5.1.2.1.1 Amplification efficiency

The efficiency of PCR amplification of the samples was not always the same. Figure 5.1 illustrates a marked decrease in the efficiency of the amplification of sample UG_CGR_71479 as compared to other samples, while sample UG_CGR_71468 had the highest efficiency. Since the annealing temperature, extension time, number of PCR cycles, and the concentration of dNTPs, Mg^{2+}, genomic DNA and Go *Taq*^{TM} DNA polymerase were kept constant for all samples, it was likely that differences in the purity of the genomic DNA resulted in the variation in amplification efficiency of the samples (Viljoen *et al*., 2005). In this investigation, samples, such as UG_CGR_71479 that had low PCR efficiency with others in the same run having good amplification efficiency had their DNA re-extracted so as to improve their DNA yield (Viljoen *et al*., 2005). In situations were amplification efficiency was low for all samples in an experiment, either the annealing temperature was lowered or the extension time increased or both (Viljoen *et al*., 2005). However, in nearly all instances, the variation in amplification efficiency did not affect the subsequent sequencing that was conducted on the samples. This was achieved by ensuring that the same concentration of PCR product was used as the sequencing DNA template for all the samples.

Figure 5.1 Photographic representation of the variation in amplification efficiency and background smear observed in PCR products amplified using primers L14125 and H16401



bp = base pairs, C1 = positive control DNA sample; MM = molecular weight marker; Neg = negative control sample. The numbers above the gel wells refer to study participants. Fragments were electrophoresed on a 0.9% agarose gel at 10 $V.cm^{-1}$ for 30 min.

5.1.2.1.2 Background smear

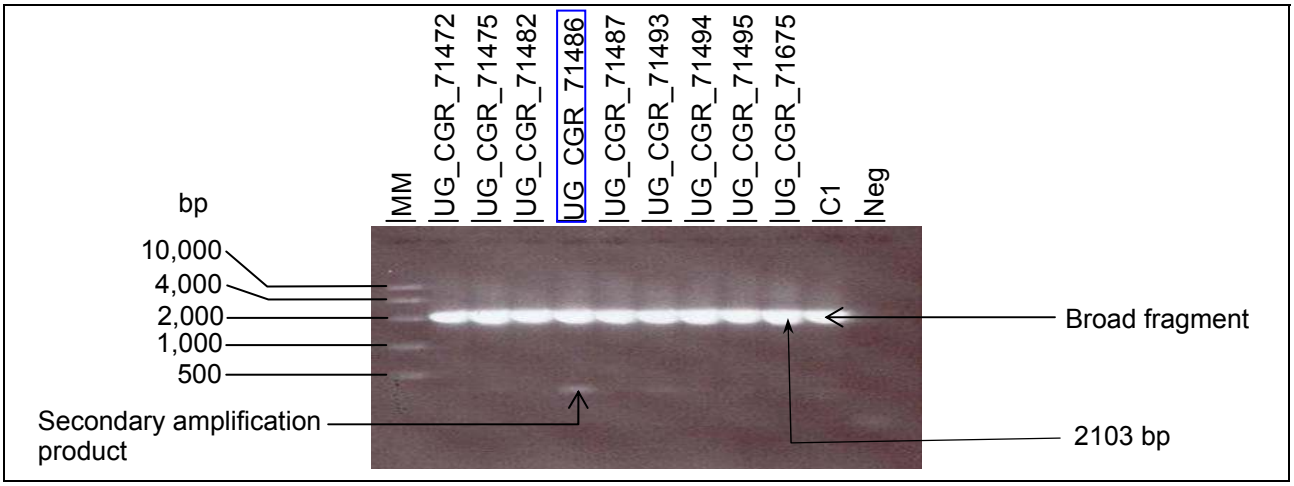
Background smears as depicted in Figure 5.1 are a general occurrence in PCR (Viljoen *et al*., 2005) and are due to amplification of non-specific DNA products of variable range (Andersson *et al*., 1993; Viljoen *et al*., 2005) or the non-amplified genomic DNA that is cleaved, due to the shearing forces it is exposed to during its extraction, and also due to, albeit minimal, the further fragmentation it experiences as it moves through the pores of the agarose gel. Background smears are frequent with high molecular weight DNA fragments, since they experience greater shearing forces due to their larger size which may be increased by overloading the DNA samples or using high voltages to electrophorese the samples (Viljoen *et al*., 2005). Smears can be minimised by optimising primer and $MgCl_2$ concentration, decreasing the number of cycles, decreasing the concentration of the template and decreasing the amount of enzyme in the reaction (Viljoen *et al*., 2005). Smears can also be avoided by the use of touchdown PCR which minimises background amplification or alternatively by using a lower voltage to electrophorese the samples (Viljoen *et al*., 2005). Smears were minimised in this study by reducing on the concentration of DNA loaded into the gel wells.

5.1.2.1.3 Secondary amplification

Secondary amplification or amplification of unwanted products, as illustrated in Figure 5.2 for sample UG_CGR_71486, was observed in certain PCR experiments. The phenomenon

of secondary amplification is due to either one or both of the primers annealing to the DNA template at more than one site (Cha and Thilly, 1993). Nucleotide alterations within the sequences could create secondary annealing sites for either or both of the primers, which result into secondary amplification. Furthermore, the binding of the primers to more than the intended recognition site is brought about by the fact that primers can still anneal to structures that contain a few mismatches (Cha and Thilly, 1993). Primers should be designed such that they are longer than 14 bp to reduce on the probability of having regions other than the target sequence being a perfect match (Cha and Thilly, 1993). Secondary amplification can be minimised by using different primers, decreasing primer concentration, decreasing the concentration of *Taq*^{TM} DNA polymerase, reducing $MgCl_2$ concentration, reducing the annealing time, reducing the extension time, increasing the annealing temperature to the optimum, decreasing the concentration of template or reducing on the number of PCR cycles so as to enhance specificity (Cha and Thilly, 1993; Viljoen *et al*., 2005). However, due to the lower yield of the secondary product compared to the primary product, successful sequencing of the primary product was attained in most instances where amplification resulted in the production of a secondary product as well. In this study, production of secondary product was minimised mainly by optimising the annealing temperature. In instances where altering the annealing temperature could not avoid secondary amplification, one or both of the primers were changed.

Figure 5.2 Photographic representation of secondary amplification observed in PCR products amplified using primers L15996 and H1487

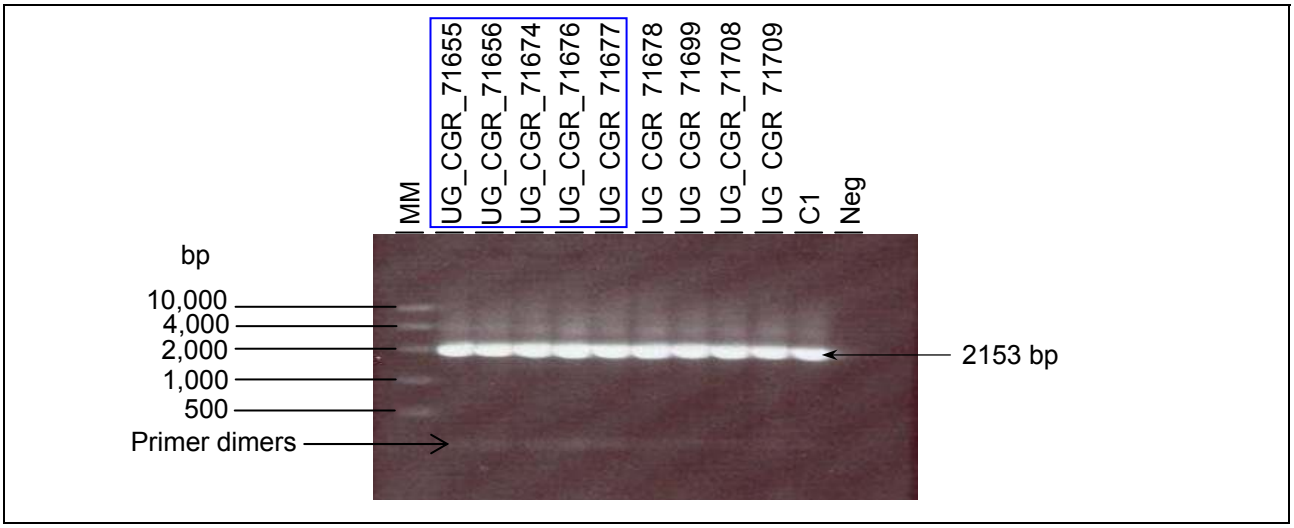


bp = base pairs, C1 = positive control DNA sample; MM = molecular weight marker; Neg = negative control sample. The numbers above the gel wells refer to study participants. Fragments were electrophoresed on a 0.9% agarose gel at 10 $V.cm^{-1}$ for 30 min.

5.1.2.1.4 Primer dimers

Figure 5.3 is an illustration of the phenomenon caused by the dimerisation of primers as observed in the lanes for samples UG_CGR_71655, UG_CGR_71656, UG_CGR_71674, UG_CGR_71676 and UG_CGR_71677. Primer dimer formation is a function of primer design and is brought about by complementary base pairing of the forward and reverse primers (Viljoen *et al.*, 2005). The phenomenon of primer-dimer formation is a common occurrence in PCR reactions (Viljoen *et al.*, 2005). Formation of primer-dimers can be minimised by avoiding self-annealing primers, reducing on the number of PCR cycles, raising the annealing temperature and decreasing the concentration of $MgCl_2$, primers and the polymerisation enzyme (Viljoen *et al.*, 2005). In this study, however, formation of primer-dimers did not affect the amplification and sequencing of the desired product and thus did not require optimisation.

Figure 5.3 Photographic representation of primer dimers observed in PCR products amplified using primers L12572 and H14685



bp = base pairs, C1 = positive control DNA sample; MM = molecular weight marker; Neg = negative control sample. The numbers above the gel wells refer to study participants. Fragments were electrophoresed on a 0.9% agarose gel at 10 $V.cm^{-1}$ for 30 min.

5.1.3 Electrophoresis

Agarose gel electrophoresis was used to determine whether there was successful amplification of DNA fragments via PCR before purification was performed (Viljoen *et al.*, 2005). The PCR products were electrophoresed on an ethidium bromide-stained 0.9% (w/v) agarose gel in 1 X TBE buffer at 10 $V.cm^{-1}$ for 30 min (Viljoen *et al.*, 2005). Visualisation, via UV light illumination, of fragments of the expected length as determined by comparison to a known molecular weight marker, indicated successful amplification (Viljoen *et al.*, 2005). All amplicons of the 10 pairs of primers used for amplification of the

entire mitochondrial genome in this study were photographed via a video documentation system to generate a photographic copy of the electrophoresis result. A high range molecular weight marker was electrophoresed alongside the amplified samples and was used to determine if the size of the observed fragments were as expected, while the absence of any fragment in the negative control ruled out that any contamination of the PCR reaction with foreign DNA had occurred (Viljoen *et al.*, 2005).

5.1.3.1 Artefacts observed on agarose gels

Despite observing all the necessary precautions, one kind of artefact was nonetheless seen in agarose gels. The one kind of artefact observed in agarose gels is discussed in Section 5.1.3.1.1. However, presence of this kind of artefact in agarose gels did not affect the interpretation of gel results and the subsequent sequencing of the observed products was successful.

5.1.3.1.1 Broad sample fragment

Figure 5.2 is an illustration of a broad sample fragment that was observed in agarose gels after electrophoresis. The existence of broad fragments was attributed to sample overload. The phenomenon of broad fragments was minimised by reducing the amount of amplified DNA product loaded into the gel wells.

5.1.4 Optimisation of PCR product purification

The protocol for PCR purification did not necessitate optimisation as the DNA yield and quality was sufficient in nearly all instances. In the few instances were the DNA yield and quality was wanting, the purification products were disposed off and the PCR amplification process repeated and the products purified. The amplified fragments were purified using the DNA clean & concentrator - 5^{TM} kit as outlined in Section 4.7. The concentration of the purified DNA was determined by measuring its absorbance with an Eppendorf^{TM} Biophotometer. The sample DNA concentrations varied between 7.7 and 35.9 ng.μL^{-1} for the different regions amplified. The variation in purified DNA concentration was attributed to variation in amplification efficiency during PCR (Cha and Thilly, 1993) and the efficiency of the purification of PCR products. In instances where the DNA quality was poor, cycle sequencing was still most often successful.

5.1.5 Optimisation of sequencing protocol

The quality of automated DNA sequencing depends upon the concentration and quality of the PCR product, the quality of the purified PCR product, the sequence precipitation process and the electrophoresis of the sequenced product (Automated DNA sequencing chemistry guide, 2000). The details of these factors are discussed in Sections 5.1.5.1.1 and 5.1.5.1.2. A range of 50 - 60 ng of purified DNA, depending on the length of the amplified template, was used for cycle sequencing in all the reactions (Automated DNA Sequencing Chemistry Guide, 2000).

Automated DNA sequencing is generally easier to optimise than PCR since, purification of the PCR products serves to enhance the specificity and efficiency of the subsequent sequencing reaction as the purification product has a high concentration of the specific DNA fragments targeted for sequencing. The entire mt genome, which was amplified in 10 overlapping fragments as indicated in Table 5.1, was sequenced in 32 overlapping fragments, due to the limitation in length of sequence that can be detected by final electrophoresis.

Only forward primers were used in sequencing, except for regions involving homopolymers. Homopolymer regions were sequenced using both forward and reverse primers as discussed in Section 5.1.5.1.8. After alignment, the processed control sequences were assessed to ensure that sufficient overlap was obtained with consecutive succeeding fragments. This was necessary and critical before study samples were sequenced. The entire genome was sequenced in 32 overlapping fragments (Maca-Meyer *et al.*, 2001). The same sequencing programme outlined in Section 4.8 was used for all 32 sequencing primers, unlike in PCR where annealing temperatures were varied between primers and the extension durations on the length of the fragments to be amplified. Electrophoresis of the sequenced products was performed by an external service provider and the amplitudes of the presented electropherograms in the following figures were not adjusted from those supplied by the provider.

5.1.5.1 Artefacts observed in sequences

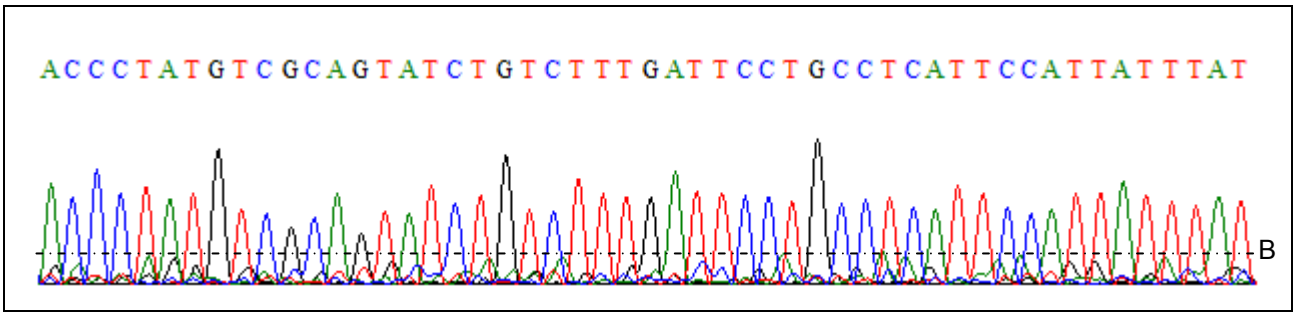
Although care was taken to ensure that well resolved sequences were obtained, artefacts were nevertheless observed in some of the generated sequences. These artefacts included background peaks, ambiguous bases, low signal intensity, dye blobs, band

compressions, secondary amplification to a specific point in a sequence, n - 1 primer and homopolymer tracts. The artefacts are discussed in Section 5.1.5.1.1 to Section 5.1.5.1.8.

5.1.5.1.1 Background peaks

Background peaks, as indicated in Figure 5.4, are undefined short peaks under the sequence peaks of interest which can interfere with the readability of the sequence of interest (Automated DNA Sequencing Chemistry Guide, 2000). A small amount of background noise is generally present in sequences but does not affect the readability of the sequences if the actual sequence peaks have high amplitude (Automated DNA Sequencing Chemistry Guide, 2000). Background peaks could be due to excess, insufficient or contaminated template (Automated DNA Sequencing Chemistry Guide, 2000). In this study, most of the background noise was attributed to insufficient template. Insufficient template generates low amplitude peaks which can be masked easily by noise in the background (Automated DNA Sequencing Chemistry Guide, 2000). When the template is in excess, there could be poor electrophoretic separation which generates broad peaks with poorly resolved individual bases or there could be a high signal intensity such that the detector gets fully saturated (Automated DNA Sequencing Chemistry Guide, 2000). The template should be cleaned or otherwise excess background noise is observed as well (Automated DNA Sequencing Chemistry Guide, 2000). SDS post-sequence clean up was performed to remove excess salts and dye terminators which improved on the quality of the sequenced product and the accuracy to which bases were resolved (BigDye® cycle sequencing kit, 2002). Excess salts serve to lower the intensity of the signal and bring about diminished base calling by competing with DNA for migration in the capillaries (Automated DNA Sequencing Chemistry Guide, 2000). Background noise can also be minimised by using nested or semi-nested PCR that greatly improves on the quality of the DNA and the specificity of the sequencing primer (Automated DNA Sequencing Chemistry Guide, 2000).

Figure 5.4 Representative electropherogram with background peaks

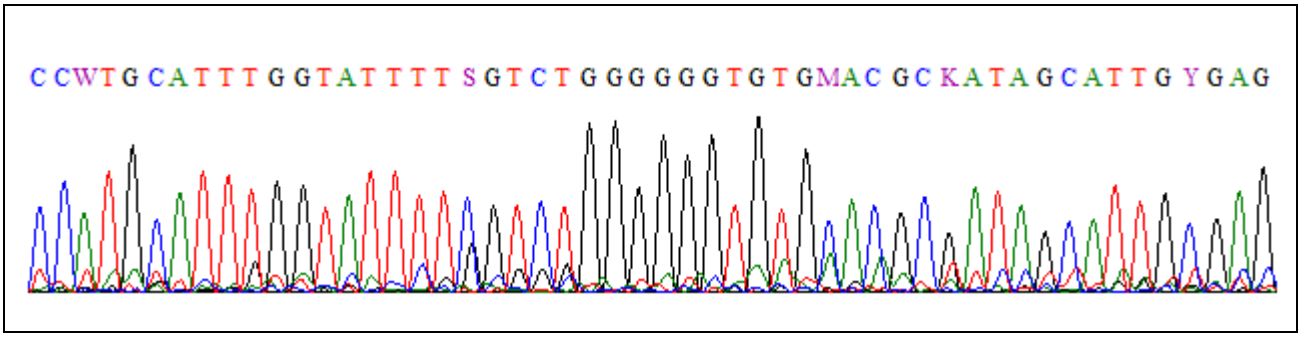


A = adenine; C = cytosine; G = guanine; T = thymine. Letter B indicates the level below which background peaks exist.

5.1.5.1.2 Ambiguous bases

In sequencing of samples, some of the generated electropherograms had ambiguous bases. In Figure 5.5, the ambiguous bases, W representing A or T; S representing G or C; M, representing A or T; K, representing G or T and Y, representing T or C (NC-IUB, 1985) were generated as the sequencer detector could not differentiate between the bases. In this study, ambiguous bases were observed in sequence electropherogragms with background noise implying that the noise interferes with the readability of the sequence. Upon visual inspection, the ambiguous bases in Figure 5.5 were resolved as follows: the W being an A; the S being a C; the M being a C; the K being a G and the Y as C. Other ambiguous bases (not diagrammatically illustrated) obtained in this study included N, representing A or C or G or T, and R, representing A or G (NC-IUB, 1985). The phenomenon of ambiguous bases was greatly reduced in this study by cleaning the sequenced DNA samples using SDS (BigDye® terminator cycle sequencing kit, 2002). However, most of the ambiguous bases generated in this study could be resolved by visual inspection of the electropherograms.

Figure 5.5 Representative electropherogragm with ambiguous bases

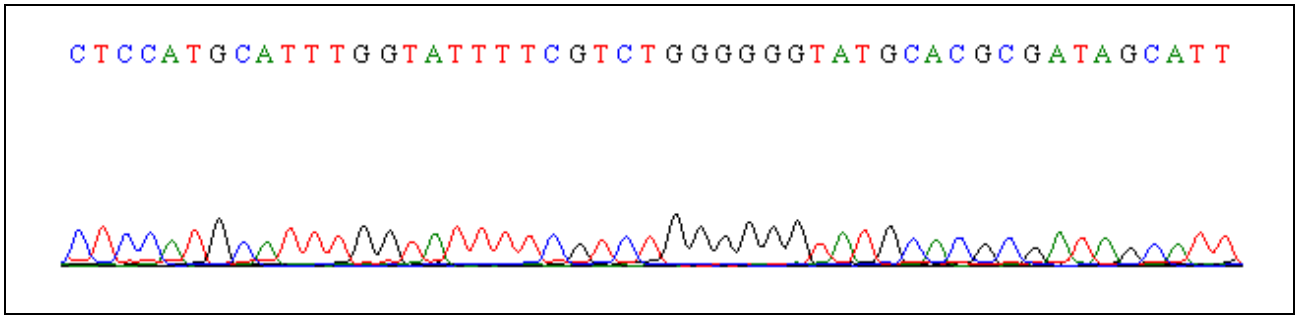


A = adenine, C = cytosine, G = guanine, T = thymine, W = A or T, S = G or C, M = A or T, K = G or T, Y = T or C (NC-IUB, 1985). In this particular case upon visual inspection, W = A, S = C, M = C, K = G and Y = C. The background noise in the Figure is discussed in Section 5.1.5.1.1.

5.1.5.1.3 Low signal intensity

Figure 5.6 is an electropherogram representative of low signal intensity. The low amplitude of the peaks was attributed to presence of the residual salts, unincorporated nucleotides and proteins in the DNA template (Automated DNA Sequencing Chemistry Guide, 2000) as nearly the same amount of DNA template was used in the cycle sequencing reactions per 1,000 bp of PCR product that was used in sequencing. The use of dye primers instead of dye terminators may minimise the artefact as the former generally gives rise to even peak intensities (Automated DNA Sequencing Chemistry guide, 2000). In this study, the problem of low signal intensity was circumvented by purifying a fresh sample of the amplified DNA so as to get rid of the excess salts and proteins (Automated DNA Sequencing Chemistry Guide, 2000).

Figure 5.6 Representative electropherogram with low signal intensity



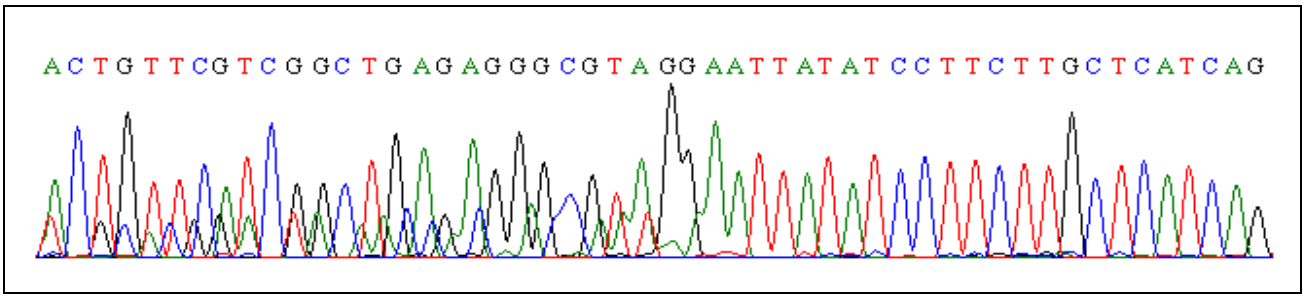
A = adenine, C = cytosine, G = guanine, T = thymine.

5.1.5.1.4 Secondary amplification product to a specific point in a sequence

Presence of secondary amplification products to fixed points in sequences as illustrated in Figure 5.7 were observed in the first 200 bases or so when sequencing amplified fragment I (Table 5.1) using primer L12572. All the sequences became uninterrupted (clean and well resolved) at nucleotide position 12778 of their mtDNA sequences. The presence of the unwanted secondary template was attributed to the presence of a secondary non-specific primer binding site within the template and attempts to optimise the PCR reactions did not yield meaningful results. During synthesis of the sequences, multiple sequences are synthesized when the DNA template harbours many primer binding sites (BigDye® cycle sequencing kit, 2002) including those with a few mismatches (Cha and Thilly, 1993). To generate this artefact, primer L12572 should have had a secondary binding site during PCR to give rise to the observed artefact. The design and use of primers of more than 18 bases in length minimises the presence of sites on the target DNA where non-specific hybridisation can take place (BigDye® cycle sequencing kit, 2002; Cha and Thilly, 1993)

but this practise was not sufficient to overcome the problem in this case. The problem was circumvented by sequencing the affected mtDNA region using amplified fragment H (amplified using L10403 and H13666 primers as indicated in Table 5.1) with primer L12572 as sufficient overlap was present with fragment I.

Figure 5.7 Representative electropherogram with secondary amplification product in sequence

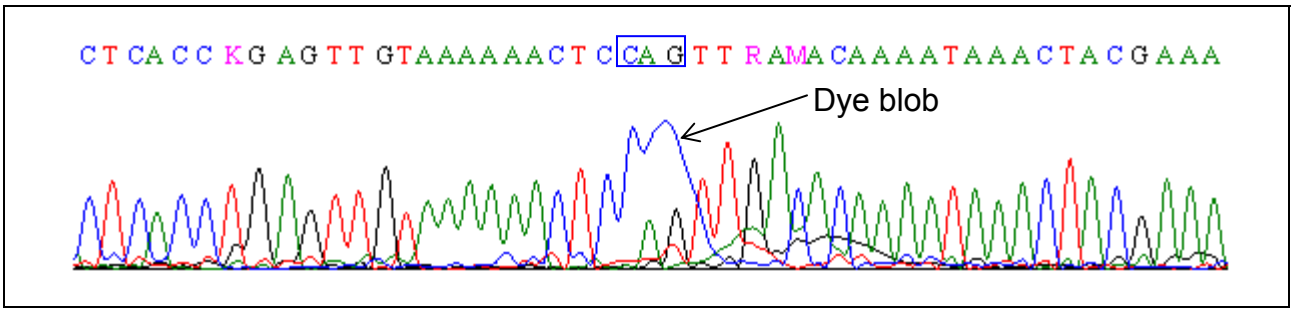


A = adenine, C = cytosine, G = guanine, T = thymine. The first half of the electropherogram consists of two sequence fragments.

5.1.5.1.5 Dye blobs

Dye blobs as illustrated in Figure 5.8 are brought about by unincorporated dye terminators and residual salts (Automated DNA Sequencing Chemistry Guide, 2000). The unincorporated dye terminators, if not removed, compete with the DNA extension fragments for migration in the capillaries thus obscuring peaks, which reduces the accuracy of base calling within the initial bases of the sequence (Harrold *et al.*, 2007). Dye blobs can be minimised by using SDS in the post sequencing clean up procedure as described in Section 4.9.

Figure 5.8 Representative electropherogram with dye blobs



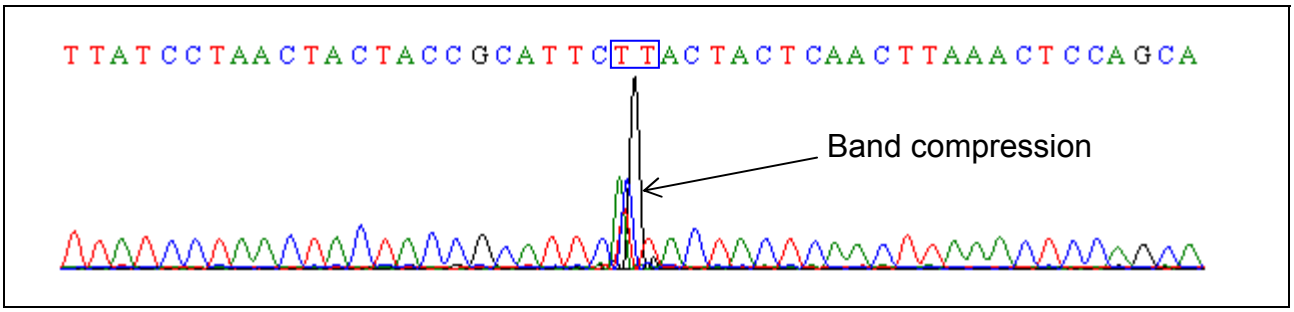
A = adenine, C = cytosine, G = guanine, T = thymine. The ambiguous bases K = T or G, R = A or G, M = C or A (NC-IUB, 1985) and have been discussed in Section 5.1.5.1.2. The dye blob affected readability at nucleotide positions 69, 70 and 71 of the sequence. Most of such dye blobs were eliminated by using SDS in post sequence clean up.

5.1.5.1.6 Band compressions

Figure 5.9 is an illustration of a band compression in a sequence. Their occurrence was attributed to the formation of unusual secondary structures in the DNA fragments that were

not eliminated by the denaturing conditions of the gel (Yamakawa *et al*., 1996; Mills and Kramer, 1979). The use of dITP in place of dGTP in dye terminator reactions eliminates most compressions since dITP forms less stable secondary structures (Mills and Kramer, 1979). However, this artefact can be avoided during electrophoresis by increasing the denaturing power of the slab gel by way of raising the temperature or concentration of the denaturant (Automated DNA sequencing chemistry guide, 1998).

Figure 5.9 Representative electropherogram with band compression

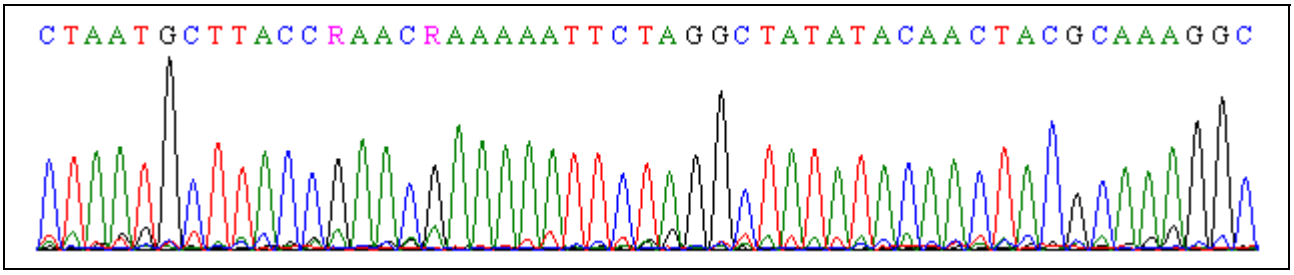


A = adenine; C = cytosine; G = guanine; T = thymine.

5.1.5.1.7 The n – 1 primer

The n – 1 primer artefact is brought about by the degradation of the primers by one base such that in addition to the primary product, a product which is one base shorter than the primary product also forms (Automated DNA Sequencing Chemistry Guide, 2000). The pattern of the sequence of the secondary sequencing product formed is similar to that of the primary product except that it is one base shorter (Automated DNA Sequencing Chemistry Guide, 2000). This phenomenon was observed in a few sequences such as that illustrated in Figure 5.10 in which the observed small peaks or bases of the secondary product were the same as that of the primary product one nucleotide ahead in the electropherogram. However, the magnitude of primer degradation was too small in the observed samples as to appreciably obscure the accuracy of base calling of the primary product. Degradation of primers is due to repeated freezing and thawing between storage and use which is more common for the primer working dilution than the stock solution (Automated DNA Sequencing Chemistry Guide, 2000). The n – 1 primer artefacts were minimised in this study by making a new dilution of the sequencing primers from the stock primer solutions (Automated DNA Sequencing Chemistry Guide, 2000). Both the stock and working dilutions of the primers were made in numerous small aliquots to minimise repeated freeze and thaw cycles and contamination that consequently happen between storage and use (Automated DNA Sequencing Chemistry Guide, 2000; Viljoen *et al*., 2005).

Figure 5.10 Representative electropherogram with n – 1 problem in a sequence

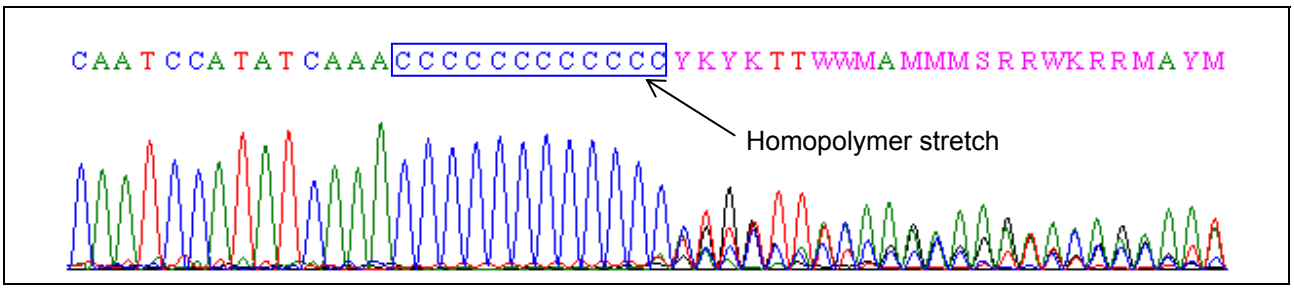


A = adenine; C = cytosine; G = guanine; T = thymine, R = A or G. The ambiguous base R was discussed in Section 5.1.5.1.2.

5.1.5.1.8 The homopolymer problem

It was difficult to sequence DNA over regions with stretches of the same bases due to slippage by the TaqDNA polymerase when reading the homopolymer stretch (Automated DNA Sequencing Chemistry Guide, 2000; Kieleczawa, 2005). The sequences fail to get resolved beyond the homopolymer stretch as illustrated in Figure 5.11. The difficulty was overcome by sequencing the homopolymer stretch in two directions by using both the reverse and forward primers (Automated DNA Sequencing Chemistry Guide, 2000). In this way sufficient overlap at the homopolymer stretch was obtained to resolve the base sequence over such stretches.

Figure 5.11 Representative electropherogram with a homopolymer sequence



A = adenine; C = cytosine; G = guanine; T = thymine, K = T or G, M = C or A, R = A or G, S = C or G, W = A or T, Y = C or T. The ambigious bases K, M, R, S, W and Y were discussed in Section 5.1.5.1.2. The sequence had a homopolymer stretch of 12 C.

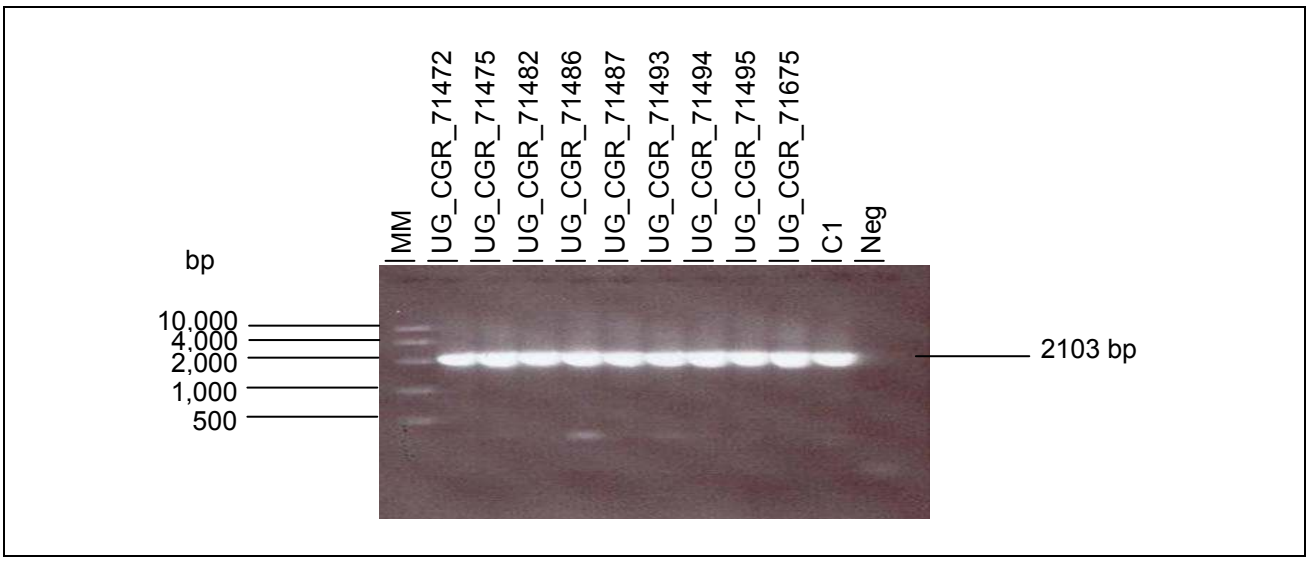
5.2 MITOCHONDRIAL DNA GENOME SEQUENCING

A total of 40 complete mtDNA genome sequences were determined from 13 Baganda, 14 Lugbara and 13 Acholi individuals respectively from the Bantu, Moru-Madi and Nilotic ethnic groups in Uganda. The complete mitochondrial genome was amplified in 10 overlapping segments as discussed in Section 5.1.2 and was sequenced in 32 overlapping fragments as discussed in Section 5.1.5. The results of the amplification of the 10 PCR fragments and the automated sequencing of the 32 sequencing fragments are discussed in the subsequent sections. The polymorphisms as indicated in Table A1 in Appendix A were established by comparison of the sequences to the rCRS mtDNA (Andrews *et al.*, 1999).

5.2.1 mtDNA region amplified using primers L15996 and H1487

The region amplified using primers L15996 and H1487, ranged from nucleotide position 15997 to 1486 (Maca-Meyer *et al.*, 2001). This region includes the D-loop (np 16024 - 576), the full coding sequence of the $tRNA^{Phe}$ gene (np 577 – 647) and the partial coding sequence for the $tRNA^{Pro}$ (np 15955 – 16023) and 12S rRNA (np 648 – 1601) genes [Mitomap, 2008a]. Figure 5.12 is a photographic representation of the PCR products generated from this region. Artefacts observed in Figure 5.12 included background smears (discussed in Section 5.1.2.1.2) and secondary amplification (discussed in Section 5.1.2.1.3).

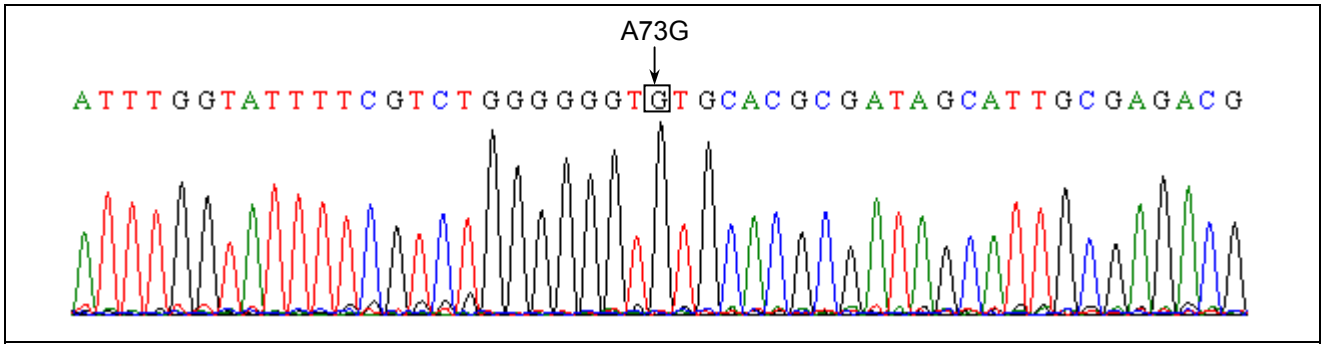
Figure 5.12 Photographic representation of PCR products amplified using primers L15996 and H1487



Fragments were electrophoresed on a 0.9% agarose gel at 10 $V.cm^{-1}$ for 30 min. bp = molecular weight base pairs, C1 = positive control DNA sample, MM = molecular weight marker, Neg = negative control sample. The numbers above the gel wells refer to study participants.

Figure 5.13 is a representative electropherogram of the fragments sequenced in this region. The A73G polymorphism in the non-coding region is a common polymorphism (Budowle *et al*., 1999; Brandstatter *et al*., 2004; Coble *et al*., 2004) and was observed in 32 out of the 40 individuals. This polymorphism occurred in all the Lugbara and Acholi (except individual UG_CGR_71468, who in all respects of the observed polymorphisms it harboured is an outlier among the Acholi) but was not observed in six out of the 13 Baganda sequences. This polymorphism could therefore be important in differentiating a certain section of the Baganda from the Lugbara and Acholi. i.e. the section of the Baganda (6/13) lacking the A73G polymorphism can be distinguished from the Lugbara and Acholi, which all have the A73G polymorphism. More Baganda, Acholi and Lugbara samples need to be analysed to verify this observed phenomenon as well as to further characterise the Baganda individuals lacking the A73G polymorphism.

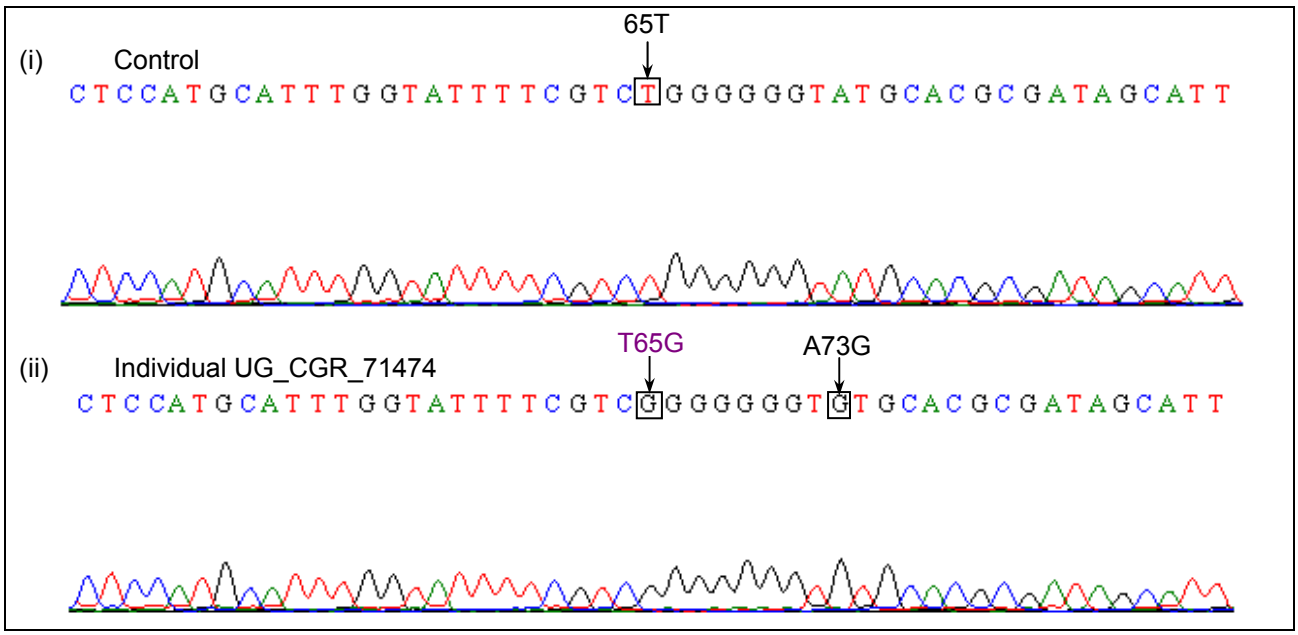
Figure 5.13 Representative electropherogram for DNA fragments sequenced within mtDNA nucleotide positions 15997 to 1486



A = adenine; C = cytosine; G = guanine; T = thymine. The reported alteration has been indicated in black. The background noise in the electropherogram has been discussed in Section 5.1.5.1.1.

A total of 145 polymorphisms were determined within this region as indicated in Table A1 in Appendix A. All the 145 polymorphisms, of which three (3) were novel, were in non-coding regions. The representative electropherograms of the novel polymorphisms (T65G, T650C and C16112T) are indicated in Figure 5.14 to Figure 5.16 while the phylogenetic significance, if any, of the 145 polymorphisms is discussed in Section 5.3.2 and 5.3.3. Figure 5.14 is an illustration of the T65G polymorphism. The T65G polymorphism was observed only in individual UG_CGR_71474, who happens to be an Acholi individual. This alteration occurs in the non-coding region resulting in a transversion within the HVII region of mtDNA (Mitomap, 2008a).

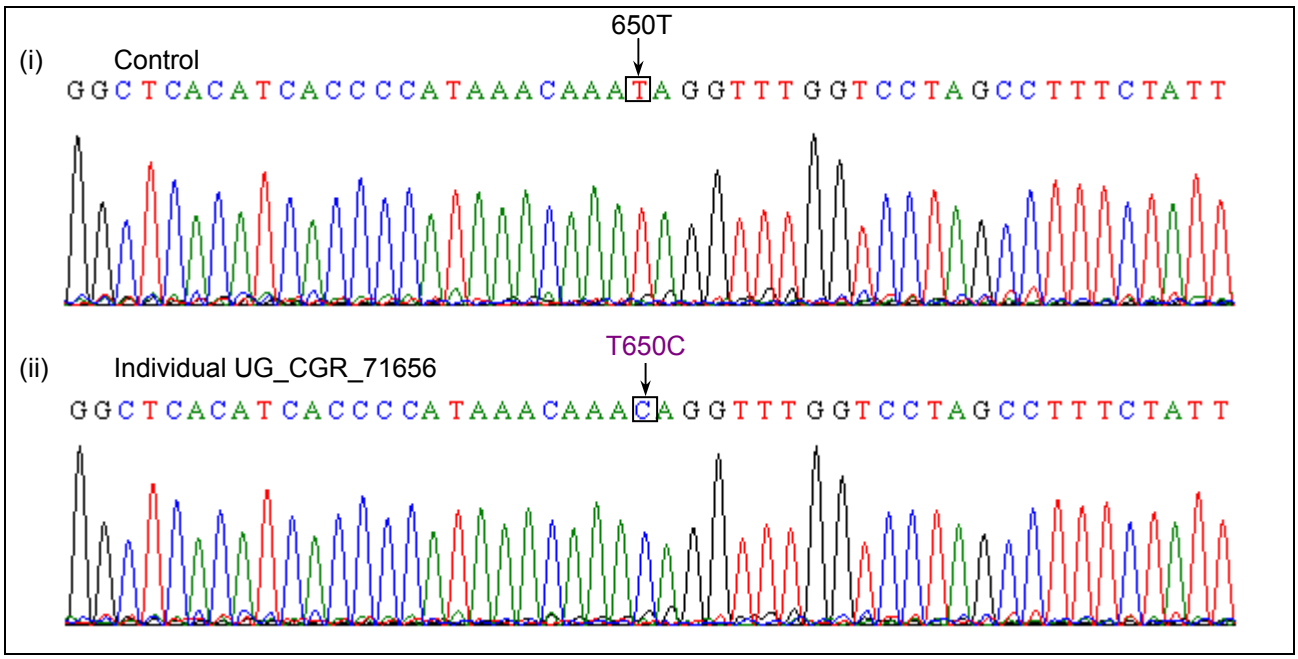
Figure 5.14 Representative electropherogram for the T65G novel polymorphism in the HVII region



Control = represents electropherogram of an individual with a sequence identical to the rCRS at nucleotide position 65 and 73. The nucleotide in violet text represents a novel alteration while the corresponding polymorphism in the control sample and the reported A73G alteration in individual UG_CGR_71474 are indicated in black. Arrows indicate the nucleotide (nt) positions at which the alterations occur. A = adenine; C = cytosine; G = guanine; T = thymine. The low signal intensity in the electropherogram was discussed in Section 5.1.5.1.3.

The novel T650C polymorphism in the 12S rRNA gene (Mitomap, 2008d) was harboured by only one individual belonging to the Baganda tribe (individual UG_CGR_71656). The polymorphism is a transition and it does not lead to a change in the protein synthesised as it occurs in the non-coding region (Mitomap, 2008e). However, mutations in the 12S rRNA gene have been identified to be associated with aminoglycoside induced hearing impairment (Pandya *et al.*, 1997; Ballana *et al.*, 2006) or gastric carcinomas (Han *et al.*, 2005). Further studies involving bigger populations will need to be performed to determine whether the T650C polymorphism has a pathogenic effect. A representative electropherogram for the T650C polymorphism has been presented in Figure 5.15.

Figure 5.15 Representative electropherogram for the T650C novel polymorphism in the 12S rRNA gene

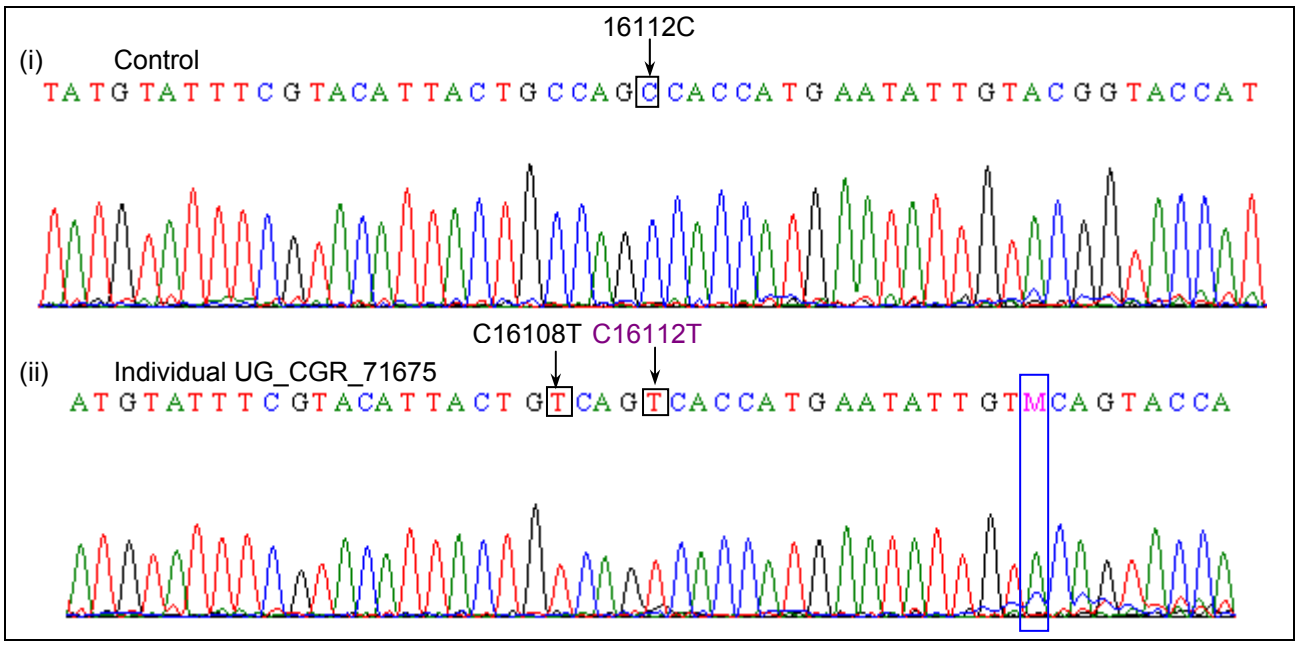


Control = represents electropherogram of an individual with a sequence identical to the rCRS at nucleotide position 650. The nucleotide in violet text represents a novel alteration while the corresponding polymorphism in the control sample is indicated in black. Arrows indicate the nucleotide (nt) positions at which the alterations occur. A = adenine; C = cytosine; G = guanine; T = thymine.

The novel transitional polymorphism, C16112T, was harboured by only one individual. This individual, UG_CGR_71675, was a Lugbara and harboured this mutation in the D-loop, specifically, in the 7 short DNA region (Mitomap, 2008a). The C16112T polymorphism is presented in an electropherogram in Figure 5.16. The ambiguous base M in the electropherogram (ii) in Figure 5.16 for individual UG_CGR_71675 was discussed in Section 5.1.5.1.2 under ambiguous bases in the sequencing artefacts section. In this particular case, M at np 16128 is an adenine nucleotide.

The C16108T transition polymorphism that occurs in the 7 short DNA strand (Mitomap, 2008a) was previously observed in an Ethiopian individual belonging to haplogroup L1b1a2 (Behar *et al.*, 2008) and Serere Senegalese individual of haplogroup M1 (Gonzalez *et al.*, 2007). The sharing of this polymorphism with an Ethiopian individual could be due to the close proximity between Northern Uganda (the geographical origin of individual UG_CGR_71675) and Ethiopia and there has been historical and geographical links between Northern Uganda, Southern Sudan and Ethiopia (Were and Wilson, 1984; Ssekamwa, 1994). The link between the Ugandan sample and the Senegalese sample as well as the Ethiopian sample could be due to the wide presence of haplogroup M1 in Africa (Gonzalez *et al.*, 2007).

Figure 5.16 Representative electropherogram for the C16112T novel polymorphism in the 7S DNA region

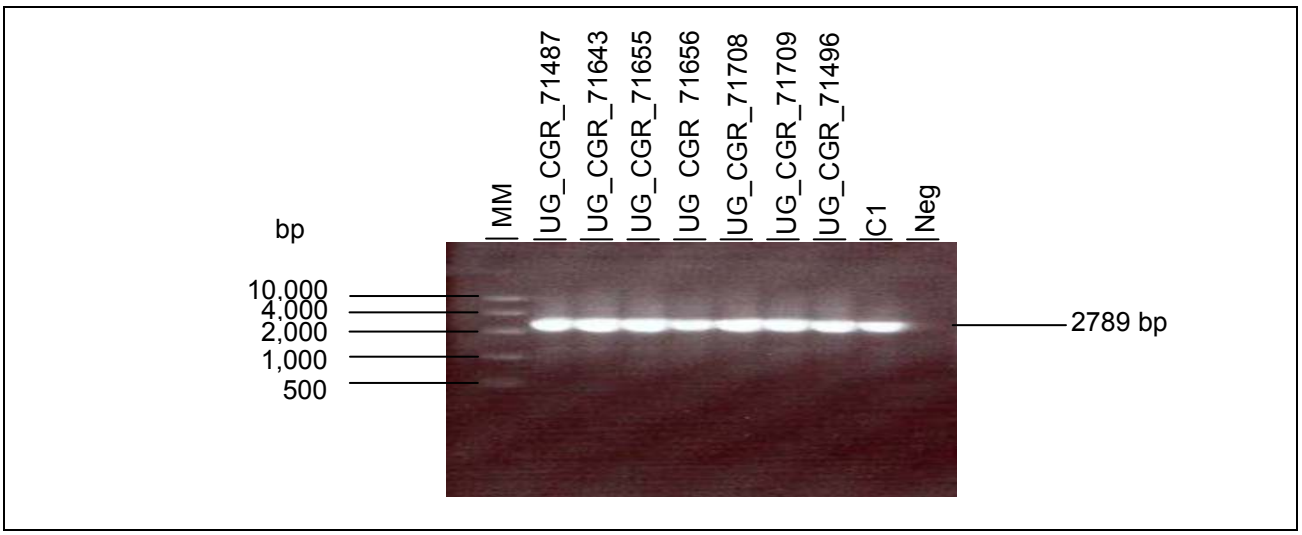


Control = represents electropherogram of an individual with a sequence identical to the rCRS at nucleotide position 16112. The nucleotide in violet text represents a novel alteration while the reported polymorphisms in are indicated in black. Arrows represent the nucleotide (nt) positions at which the alterations occur. A = adenine; C = cytosine; G = guanine; T = thymine; M = C or A (NC_IUB, 1985). The ambiguous bases have been discussed in Section 5.1.5.1.2 but in this instance, M was edited to A after visual inspection of the electropherogram.

5.2.2 mtDNA region amplified using primers L923 and H3670

The region amplified using primers L923 and H3670 ranged from nucleotide position 924 to 3669 (Maca-Meyer *et al.*, 2001). The region partially covers the genes for 12S rRNA (np 648 - 1601) and ND1 (np 3307 - 4262), and the complete coding sequence of the genes for $tRNA^{Val}$ (np 1602 - 1670), 16S rRNA (np 1671 - 3228), and $tRNA^{Leu}$ (np 3230 - 3304). The region also includes the transcription terminator site (np 3229 - 3256) and non-coding nucleotides at np 3305 - 3306 [Mitomap, 2008a]. Figure 5.17 is a photographic representation of the PCR products generated from this region. Artefacts observed in Figure 5.17 included background smears (discussed in Section 5.1.2.1.2).

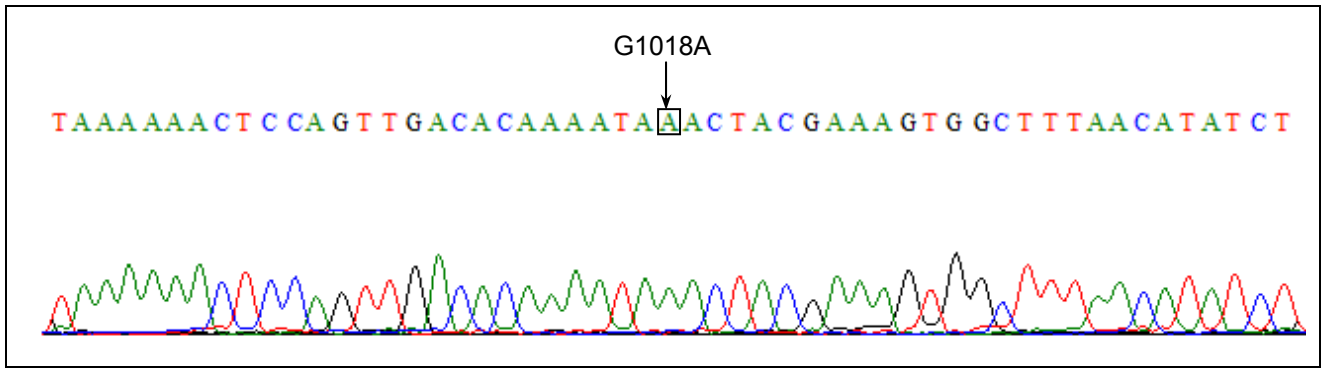
Figure 5.17 Photographic representation of PCR products amplified using primers L923 and H3670



Fragments were electrophoresed on a 0.9% agarose gel at 10 V.cm^{-1} for 30 min. bp = molecular weight in base pairs, C1 = positive control DNA sample, MM = molecular weight marker, Neg = negative control sample. The numbers above the gel wells refer to study participants.

Figure 5.18 is an electropherogram representing the fragments sequenced in this region. The G1018A polymorphism in the non-coding region was observed in 30 out of the 40 individuals. This polymorphism was characteristic of individuals belonging to haplogroup L3 (Kivisild *et al.*, 2004; Torroni *et al.*, 2006) which was in agreement with this study as discussed in Section 5.3.2 and 5.3.3.

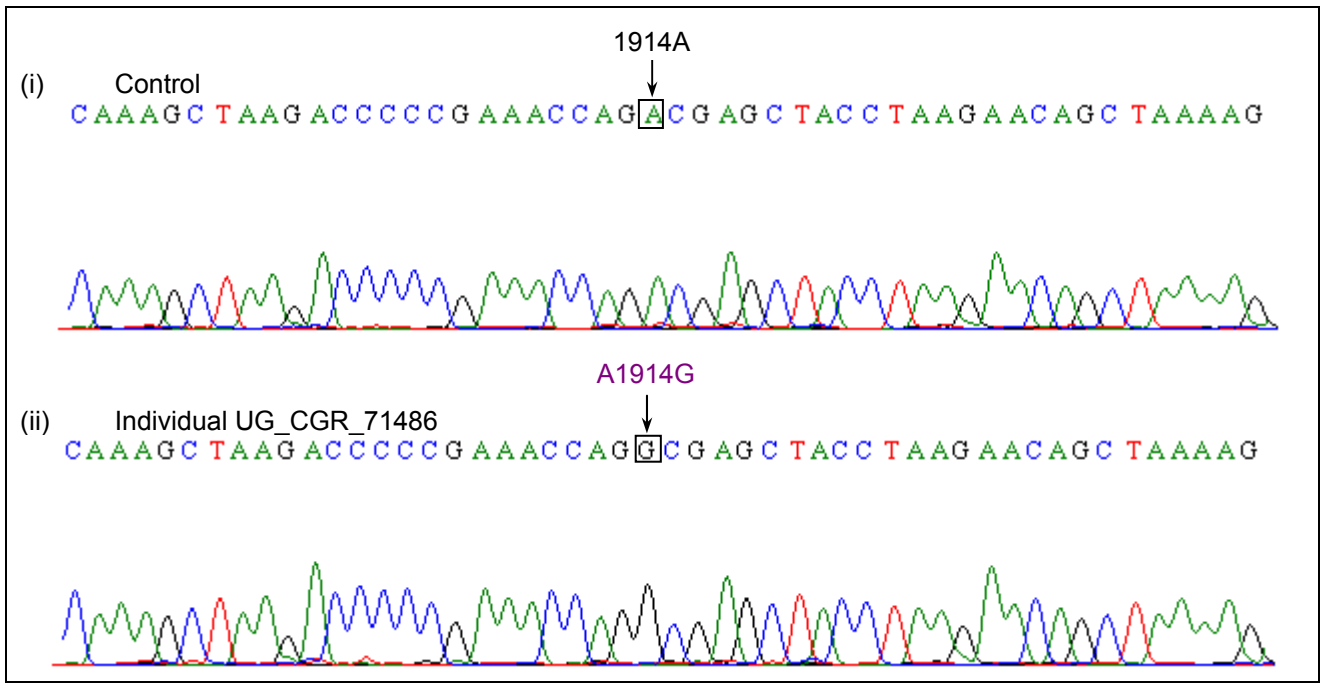
Figure 5.18 Representative electropherogram for DNA fragments sequenced within mtDNA nucleotide positions 924 to 3669



A = adenine; C = cytosine; G = guanine; T = thymine. The reported alteration has been indicated in black.

A total of 52 polymorphisms were determined within this region as indicated in Table A1 in Appendix A. Eleven (11) of these polymorphisms were synonymous, three (3) nonsynonymous and four (4) were novel while 38 were in the non-coding regions. The representative electropherograms for the novel polymorphisms (A1914G, T2385C, A2558G and C3321T) are indicated in Figure 5.19 to Figure 5.22. The phylogenetic significance of the polymorphisms, if any, is discussed in Section 5.3.2 and 5.3.3.

Figure 5.19 Representative electropherogram for the A1914G novel polymorphism in the 16S rRNA gene

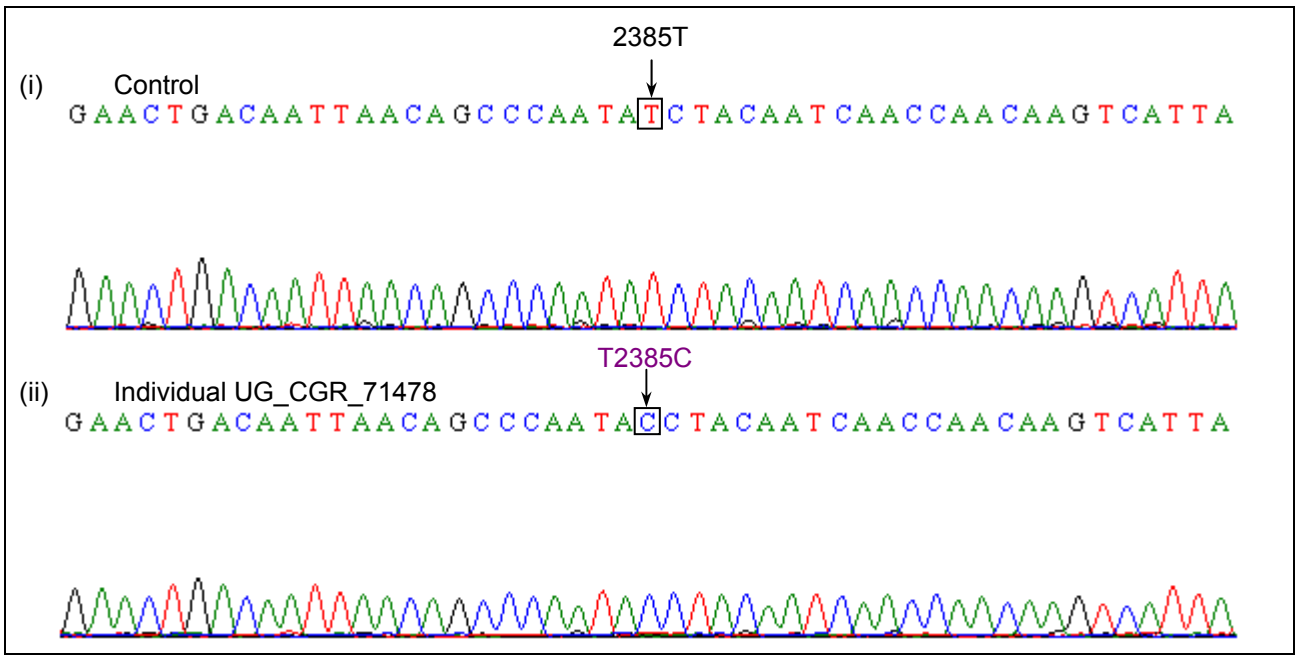


Control = represents electropherogram of an individual with a sequence identical to the rCRS at nucleotide position 1914. The nucleotide in violet text represents a novel alteration while the corresponding polymorphism for the control sample is indicated in black. Arrows represent the nucleotide (nt) positions at which the alterations occur. A = adenine; C = cytosine; G = guanine; T = thymine.

The A1914G novel polymorphism occurred in one Lugbara individual (individual UG_CGR_71486) in the 16S rRNA gene (Mitomap, 2008a). Polymorphisms in the 16S rRNA gene of mtDNA may affect termination of translation as they may alter the secondary structure of the gene (Abhyankar *et al*., 2009). The phylogenetic significance of the polymorphism, as outlined in Section 5.3.3, was limited to only defining a haplotype in haplogroup L3, since it was novel and occurred in only one individual.

The T2385C polymorphism occurred in a Lugbara individual (individual UG_CGR_71478) in the 16S rRNA gene (Mitomap, 2008c) where it does not code for any protein (Mitomap, 2008c). The polymorphism may alter the secondary structure of 16S rRNA and a change in termination of translation may occur (Göringer *et al*., 1991). The phylogenetic significance of this novel polymorphism remains to be identified in a study involving larger sample sizes since it occurred in only a single individual in the sample set investigated.

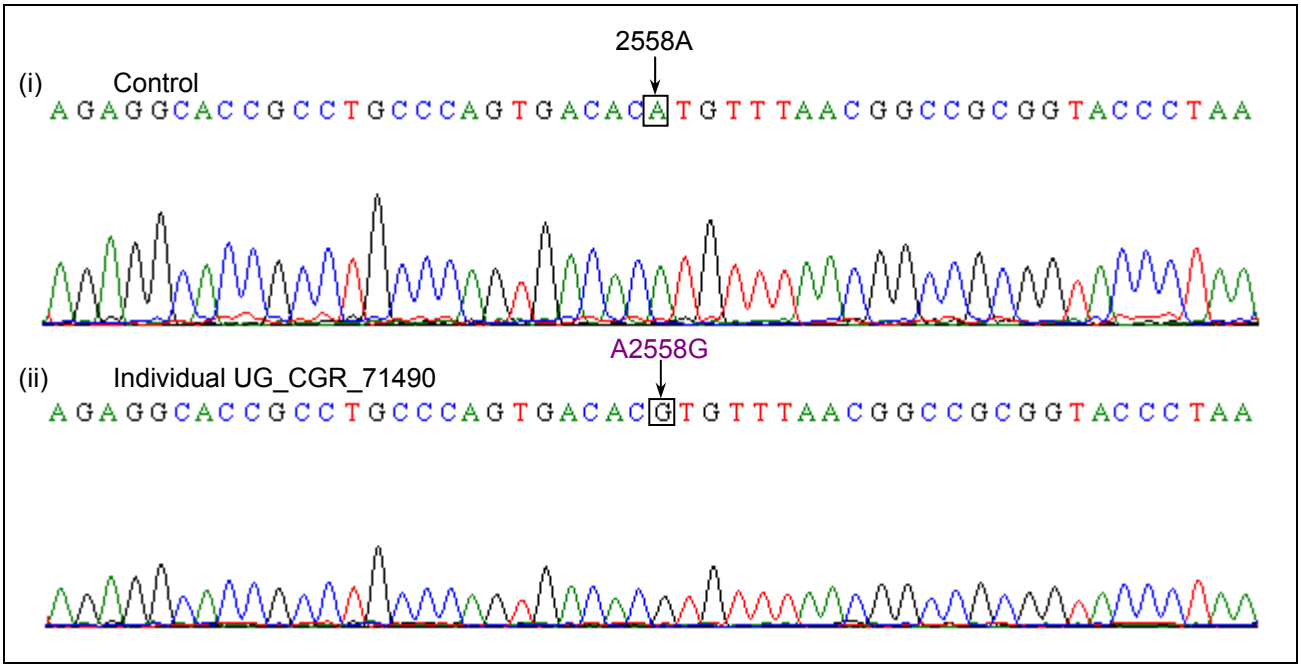
Figure 5.20 Representative electropherogram for the T2385C novel polymorphism in the 16S rRNA gene



Control = represents electropherogram of an individual with a sequence identical to the rCRS at nucleotide position 2385. The nucleotide in violet text represents a novel alteration while the corresponding polymorphism for the control sample is indicated in black. Arrows represent the nucleotide (nt) positions at which the alterations occur. A = adenine; C = cytosine; G = guanine; T = thymine.

The A2558G novel polymorphism, as indicated in Figure 5.21, was harboured by individual UG_CGR_71490. The phylogenetic significance of the polymorphism can not be defined in this study since only a single sample harboured the polymorphism.

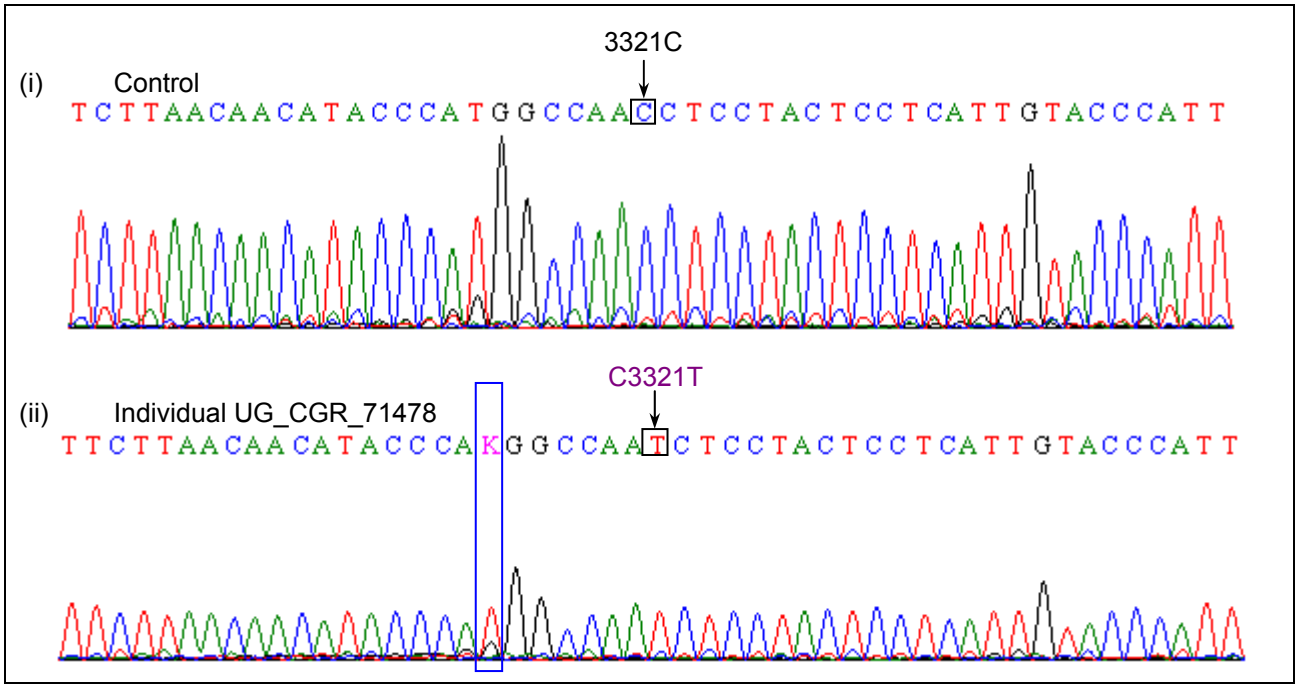
Figure 5.21 Representative electropherogram for the A2558G novel polymorphism in the 16S rRNA gene



Control = represents electropherogram of an individual with a sequence identical to the rCRS at nucleotide position 2558. The nucleotide in violet text represents a novel alteration while the corresponding polymorphism for the control sample is indicated in black. Arrows represent the nucleotide (nt) positions at which the alterations occur. A = adenine; C = cytosine; G = guanine; T = thymine. The low signal intensity in electropherogram (ii) was discussed in Section 5.1.5.1.3.

The C3321T polymorphism, as indicated in Figure 5.22, was harboured by two individuals, i.e. individual UG_CGR_71470 is an Acholi and individual UG_CGR_71478 is a Lugbara. The ambiguous base K has been discussed in Section 5.1.5.1.2 while background noise has been discussed in Section 5.1.5.1.1. By definition, K is a G or T nucleotide (NC_IUB, 1985) given that the sequencer could not distinguish between the true peak and the background peak. In this particular case therefore, K was edited to a thymine residue upon visual analysis of the sequence as it was clear that T was the true peak.

Figure 5.22 Representative electropherogram for the C3321T novel polymorphism in the NDI gene



Control = represents electropherogram of an individual with a sequence identical to the rCRS at nucleotide position 3321. The nucleotide in violet text represents a novel alteration while the corresponding polymorphism for the control sample is indicated in black. Arrows represent the nucleotide (nt) positions at which the alterations occur. A = adenine; C = cytosine; G = guanine; T = thymine; K = G or T (NC_IUB, 1985). The ambiguous bases have been discussed in Section 5.1.5.1.2 but in this instance, K was edited to a T after visual inspection of the electropherogram.

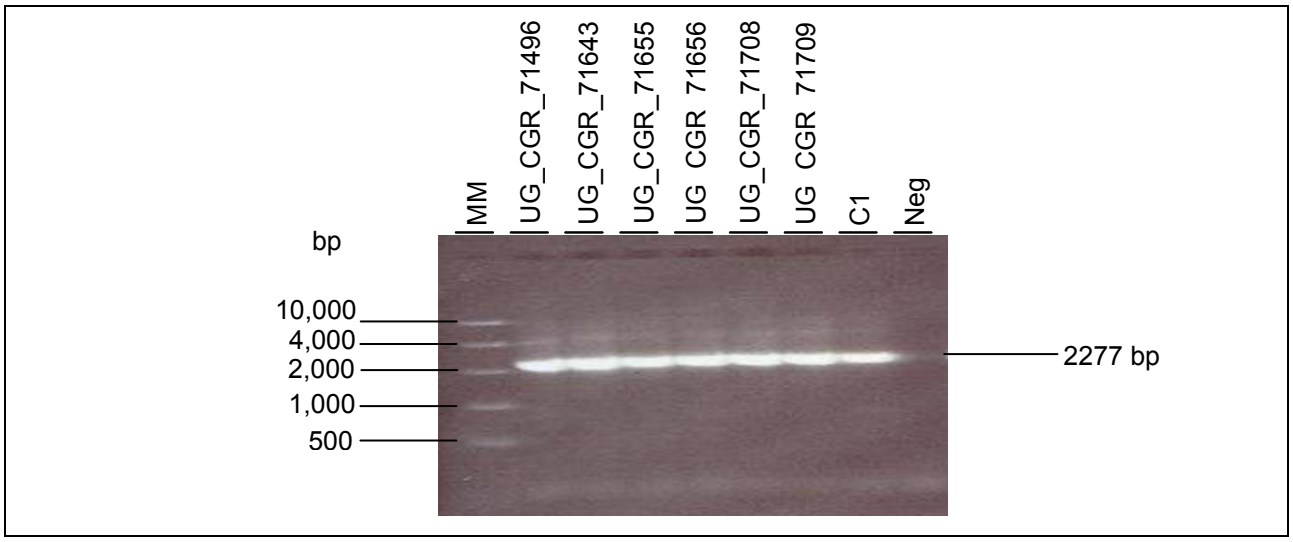
This synonymous polymorphism occurs in the ND1 gene (Mitomap, 2008c) and is therefore unlikely to have a functional effect on the mitochondria. This polymorphism, as outlined in Section 5.3.2 and 5.3.3, distinguishes members of haplogroup L2a1 from others in haplogroup L2, since only members in haplogroup L2a1 lacked the T allele at the np.

5.2.3 mtDNA region amplified using primers L3073 and H5306

The region amplified using primers L3073 and H5306 ranged from nucleotide position 3074 to 5305 (Maca-Meyer *et al.*, 2001). This region covers the partial coding regions of

the genes for 16S rRNA (np 1671 - 3228), ND2 (np 4470 - 5511), and the entire coding region for the $tRNA^{Leu}$ (np 3230 - 3304), ND1 (np 3307 - 4262), $tRNA^{Met}$ (np 4402 - 4469), $tRNA^{Ile}$ (np 4263 - 4331) and $tRNA^{Glu}$ (np 4329 - 4400) genes. This region also includes sites for the transcription terminator (np 3229 - 3256) as well as non-coding nucleotides at np 3305 – 3306 and at np 4401 [Mitomap, 2008a]. Figure 5.23 is a photographic representation of the PCR products generated from this region. Artefacts observed in Figure 5.23 included background smears (discussed in Section 5.1.2.1.2), secondary amplification (discussed in Section 5.1.2.1.3) and primer-dimers (discussed in Section 5.1.2.1.4).

Figure 5.23 Photographic representation of PCR products amplified using primers L3073 and H5306



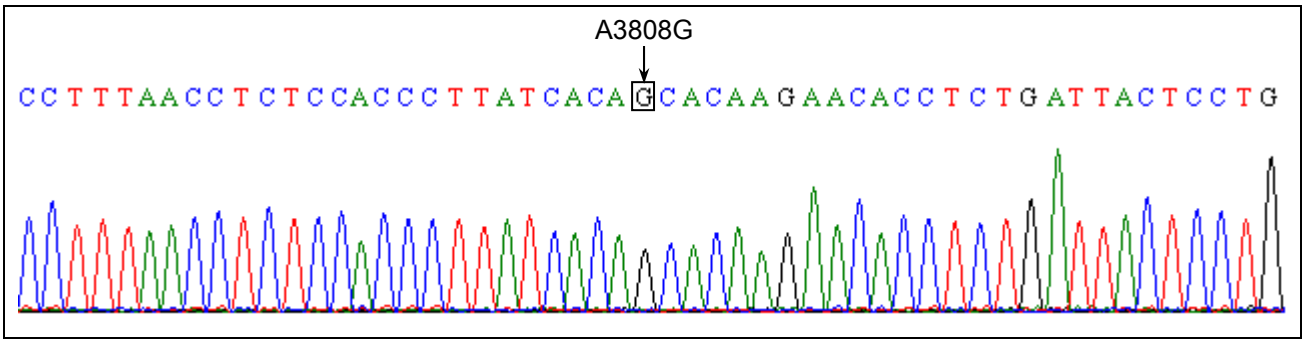
Fragments were electrophoresed on a 0.9% agarose gel at 10 $V.cm^{-1}$ for 30 min. bp = molecular weight in base pairs, C1 = positive control DNA sample, MM = molecular weight marker, Neg = negative control sample. The numbers above the gel wells refer to study participants.

Figure 5.24 is a representative electropherogram for fragments sequenced in this region. The A3808G polymorphism in the ND1 gene only occurred in individual UG_CGR_71674 in this cohort. This polymorphism was observed to be associated with L0a mtDNA profiles (Maca-Meyer *et al*., 2001). However, as discussed in Section 5.3.3, individual UG_CGR_71674 who harbours this polymorphism clustered with individuals of haplogroup L3h. The A3808G polymorphism could therefore be associated with a variety of haplogroups and can therefore not be used as a haplogroup specific marker.

A total of 57 polymorphisms were determined within this region as indicated in Table A1 in Appendix A. Forty-four (44) of these polymorphisms were synonymous, 10 were nonsynonymous, three (3) were novel (C3321T, C4032T, A4212G) while three (3) were within untranslated regions (Mitomap, 2008a). The representative electropherograms for

the novel polymorphisms are indicated in Figure 5.22, Figure 5.25 and Figure 5.26. The phylogenetically informative mutations have been outlined in Section 5.3.2 and 5.3.3.

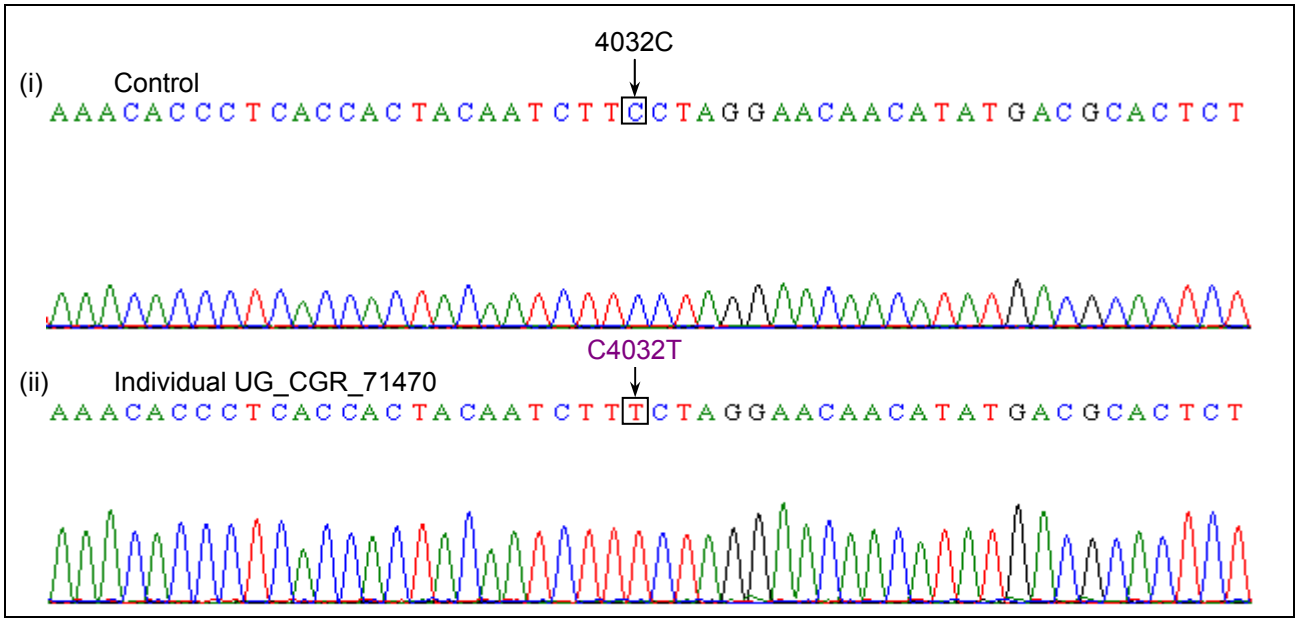
Figure 5.24 Representative electropherogram for DNA fragments sequenced within mtDNA nucleotide positions 3073 to 5306



A = adenine; C = cytosine; G = guanine; T = thymine. The reported alteration has been indicated in black.

The C4032T polymorphism, as illustrated in Figure 5.25, was harboured by two individuals i.e. UG_CGR_71470 (an Acholi) and UG_CGR_71478 (a Lugbara). The phylogenetic relevance of this polymorphism has been outlined in Section 5.3.2 and 5.3.3. The C4032T polymorphism occurs in the ND1 gene and does not lead to a change in the amino acid synthesised (Mitomap, 2008c). However, the two samples that possess the C4032T polymorphism should have had a near recent common ancestor as they cluster together in the global phylogenetic tree illustrated in Figure 5.59 and Figure 5.60, due to the polymorphisms they have in common.

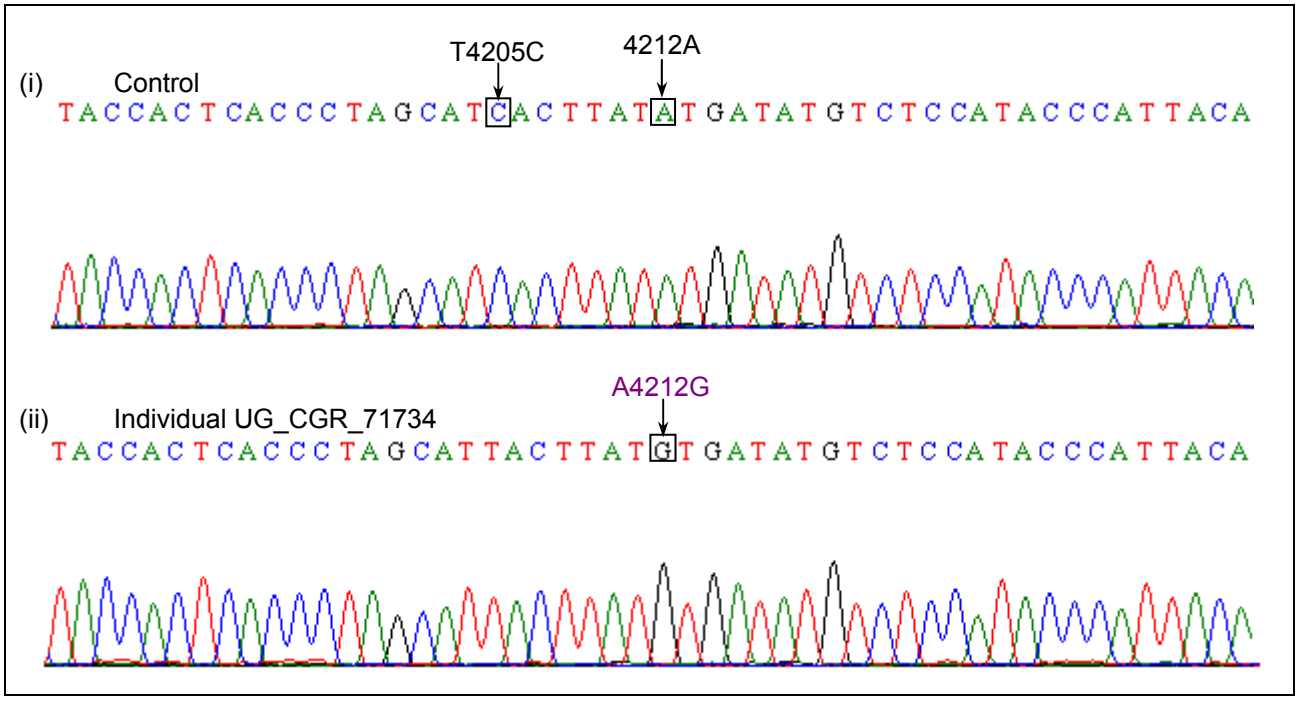
Figure 5.25 Representative electropherogram for the C4032T novel polymorphism in the NDI gene



Control = represents electropherogram of an individual with a sequence identical to the rCRS at nucleotide position 4032. The nucleotide in violet text represents a novel alteration while the corresponding polymorphism for the control sample is indicated in black. Arrows represent the nucleotide (nt) positions at which the alterations occur. A = adenine; C = cytosine; G = guanine; T = thymine.

The T4205C polymorphism, as illustrated in Figure 5.26, was harboured by only one individual namely UG_CGR_71471 (an Acholi). The T4205C transition occurred in the

Figure 5.26 Representative electropherogram for the A4212G novel polymorphism in the NDI gene



Control = represents electropherogram of an individual with a sequence identical to the rCRS at nucleotide position 4212. Nucleotides in violet text represent novel alterations while reported polymorphisms are indicated in black. Arrows indicate the nucleotide (nt) positions at which the alterations occur. A = adenine; C = cytosine; G = guanine; T = thymine.

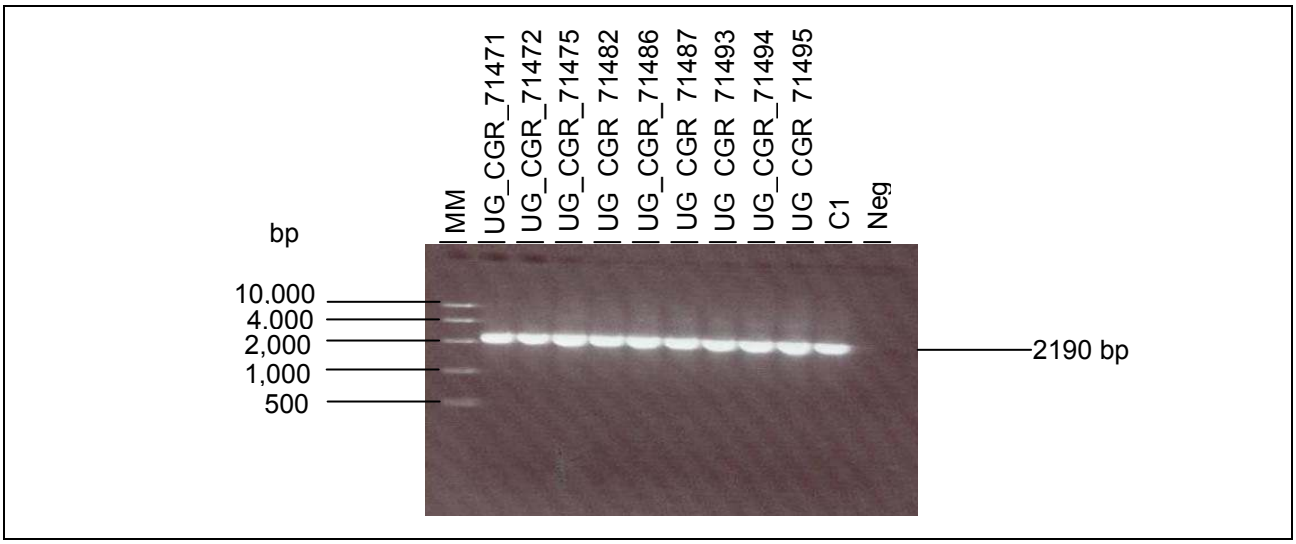
ND1 gene region of the mtDNA causing a Leu to Ser change in the amino acid encoded (Mitomap, 2008c). The T4205C polymorphism was observed to have previously been harboured by an individual of haplogroup L2e (Behar *et al.*, 2008). More samples bearing this polymorphism need to be sequenced to ascertain its phylogenetic relevance within the population. The transition polymorphism A4212G, as illustrated in Figure 5.26, was harboured by the Acholi individual UG_CGR_71734. The polymorphism occurred in the ND1 gene locus where it is synonymous and was therefore unlikely to lead to an alteration in the manner in which the first enzyme complex of the respiratory chain functions (Berg *et al.*, 2007). The phylogenetic relevance of this polymorphism needs to be ascertained by a study involving larger sample sizes.

5.2.4 mtDNA region amplified using primers L4750 and H6899

The region amplified using primers L4750 and H6899 ranged from nucleotide position 4751 to 6898 (Maca-Meyer *et al.*, 2001). This region partially covers the genes for ND2 (np

4470 - 5511), COI (np 5904 - 7445) and fully covers the $tRNA^{Trp}$ gene (np 5512 - 5579), non-coding nucleotides at position 5580 - 5586, $tRNA^{Ala}$ gene (np 5587 - 5655), non-coding nucleotide at position 5656, $tRNA^{Asn}$ gene (np 5657 - 5729), L-strand origin (np 5721 - 5798), $tRNA^{Cys}$ gene (np 5761 - 5826), $tRNA^{Tyr}$ gene (np 5826 - 5891) and non-coding nucleotides at position 5892 - 5903 [Mitomap, 2008a]. Figure 5.27 is a representative of the PCR products generated from this region. Artefacts observed in Figure 5.27 included background smears (discussed in Section 5.1.2.1.2).

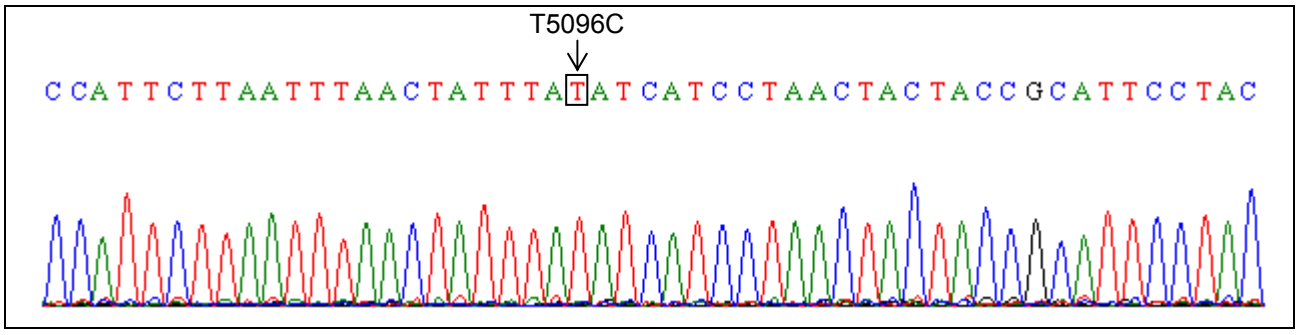
Figure 5.27 Photographic representation of PCR products amplified using primers L4750 and H6899



Fragments were electrophoresed on a 0.9% agarose gel at 10 $V.cm^{-1}$ for 30 min. bp = molecular weight in base pairs, C1 = positive control DNA sample, MM = molecular weight marker, Neg = negative control sample. The numbers above the gel wells refer to study participants. The gel wells were inclined and not horizontal otherwise the electrophoresis ran normally since the fragments were parallel to the gel wells.

Figure 5.28 is a representative electropherogram for the fragments sequenced in this region. The T5096C synonymous polymorphism in the ND2 gene (Mitomap, 2008c) was harboured by four individuals. The polymorphism is characteristic of individuals harbouring haplogroup L0c1 mtDNA sequences (Wallace, 2004).

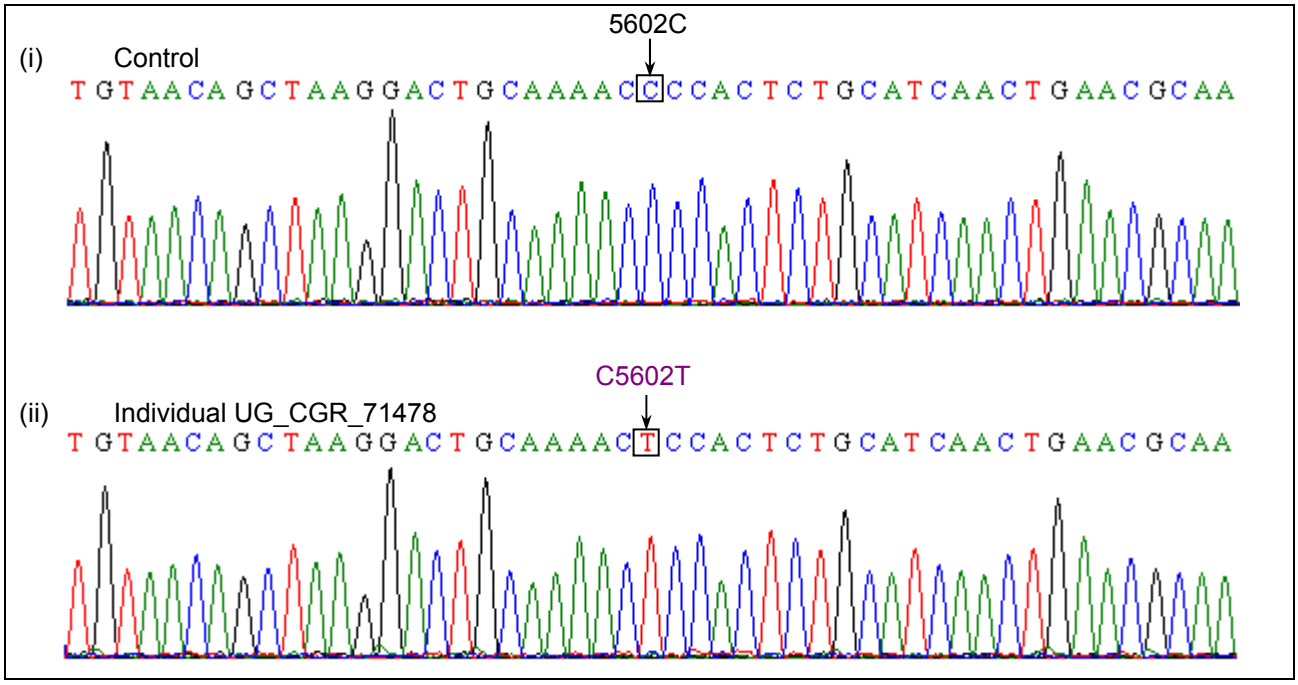
Figure 5.28 Representative electropherogram for DNA fragments sequenced within mtDNA nucleotide positions 4751 and 6898



A = adenine; C = cytosine; G = guanine; T = thymine. The reported alteration has been indicated in black.

A total of 68 polymorphisms were generated within this region as indicated in Table A1 in Appendix A. Forty five (45) of these polymorphisms were synonymous, 10 nonsynonymous, 13 non-coding and one (1) was novel (C5602T). The representative electropherograms for the novel polymorphism is indicated in Figure 5.29. The phylogenetically informative mutations will be outlined in Section 5.3.2 and 5.3.3.

Figure 5.29 Representative electropherogram for the C5602T novel polymorphism in the tRNA^{Ala} gene



Control = represents an electropherogram of an individual with a sequence identical to the rCRS at nucleotide position 5602. The nucleotide in violet text represents a novel alteration while the corresponding polymorphism for the control sample is indicated in black. Arrows represent the nucleotide (nt) positions at which the alterations occur. A = adenine; C = cytosine; G = guanine; T = thymine.

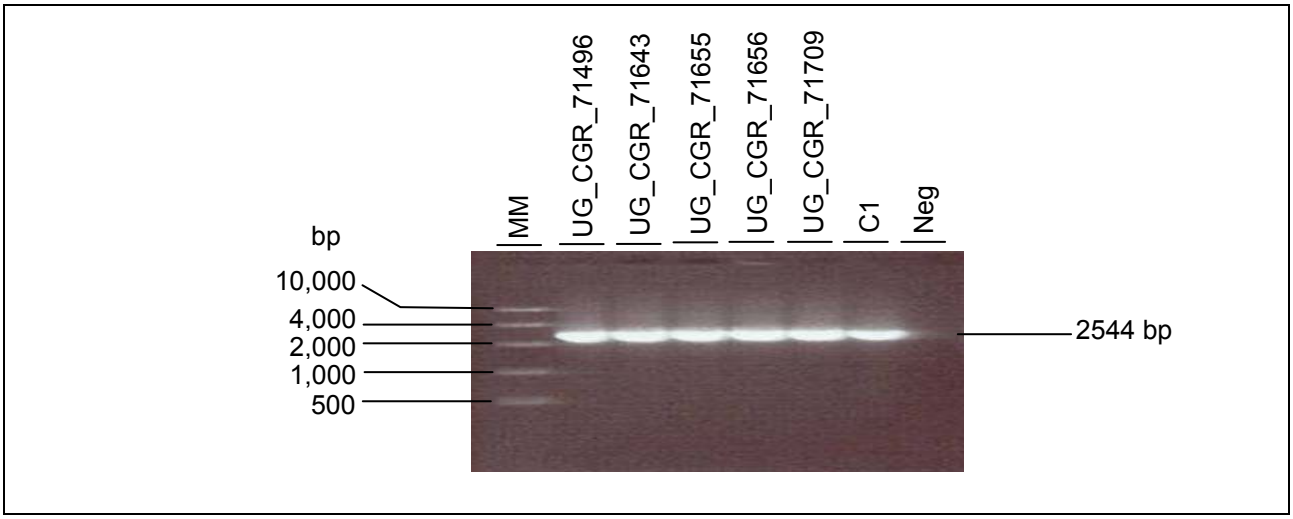
The C5602T polymorphism occurs in the gene tRNA^{Ala} of the mtDNA (Mitomap, 2008c). This polymorphism as outlined in Appendix A occurred in the Acholi individual UG_CGR_71470 and a Lugbara individual i.e. UG_CGR_71478. This polymorphism is harboured by a set of haplotypes that form a deep rooted independent branch in haplogroup L2 as indicated in Figure 5.59 and Figure 5.60. Therefore, it is likely that the polymorphism arose out of a common ancestor of the two novel haplotypes since the two haplotypes group together in the global African trees.

5.2.5 mtDNA region amplified using primers L6337 and H8861

The region amplified using primers L6337 and H8861 ranged from nucleotide position 6338 to 8860 (Maca-Meyer *et al.*, 2001). This region partially covers the genes for COI (np

5904 - 7445), ATP6 (np 8527 - 9207) and fully covers the genes for $tRNA^{Ser(UCN)}$ (np 7446 - 7516), $tRNA^{Asp}$ (np 7518 - 7585), COII (np 7586 - 8269), $tRNA^{Lys}$ (np 8295 - 8364) and ATP8 (np 8366 - 8572). It also includes the non-coding nucleotides at position 7517, position 8270 - 8294 and position 8365 [Mitomap, 2008a]. Figure 5.30 is a representative photograph of the PCR products generated from this region. Artefacts observed in Figure 5.30 included background smears (discussed in Section 5.1.5.1.2).

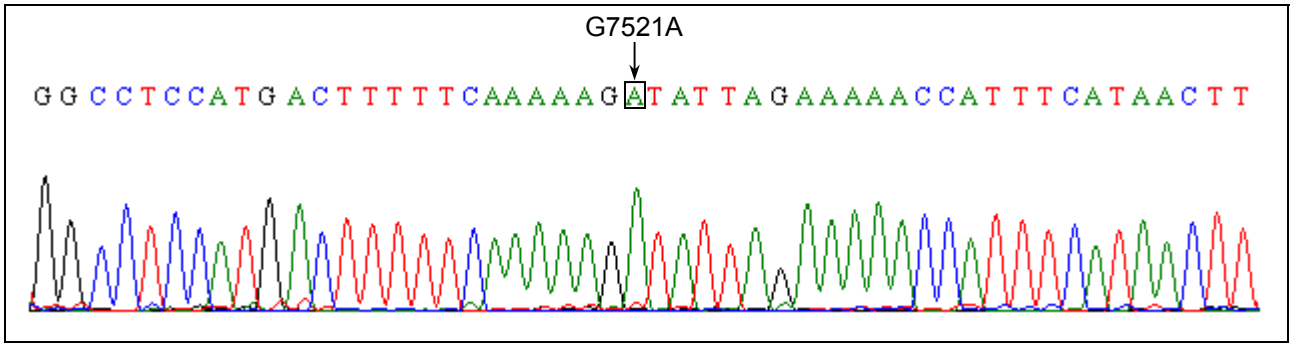
Figure 5.30 Photographic representation of PCR products amplified using primers L6337 and H8861



Fragments were electrophoresed on a 0.9% agarose gel at 10 $V.cm^{-1}$ for 30 min. bp = molecular weight in base pairs, C1 = positive control DNA sample, MM = molecular weight marker, Neg = negative control sample. The numbers above the gel wells refer to study participants.

Figure 5.31 is a representative electropherogram for fragments sequenced in this region. The G7521A polymorphism occurred in six (6) Acholi individuals, 10 Lugbara individuals and six (6) Baganda. This polymorphism is characteristic of mtDNA haplogroup L3, L4 and L6 sequence motifs amongst African populations (Kivisild *et al.*, 2004; Torroni *et al.*, 2006).

Figure 5.31 Representative electropherogram for DNA fragments sequenced within mtDNA nucleotide positions 6338 to 8860



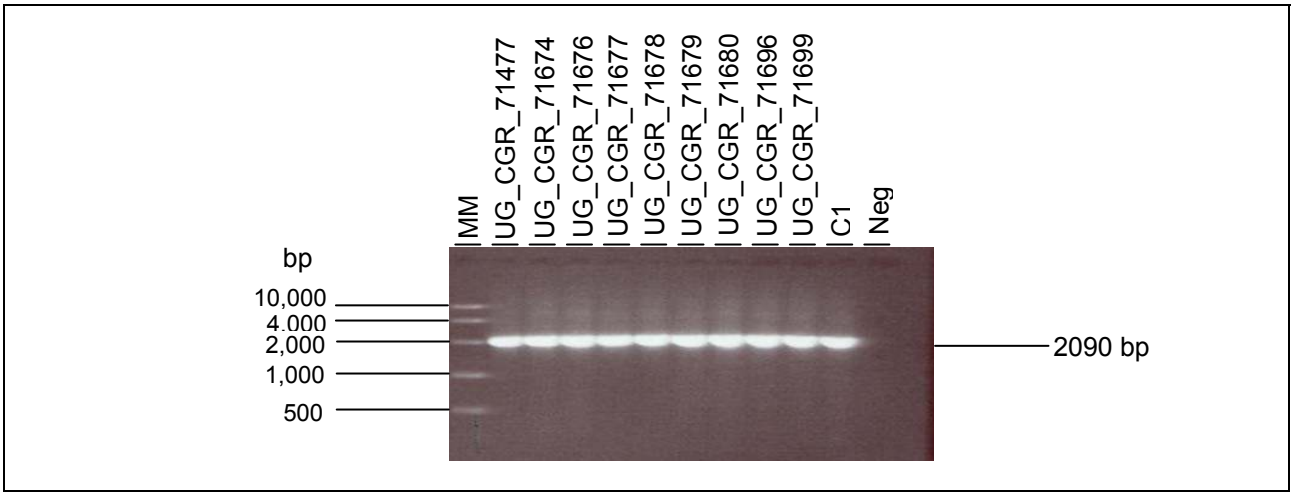
A = adenine; C = cytosine; G = guanine; T = thymine. The reported alteration has been indicated in black.

A total of 68 polymorphisms were determined within this region as indicated in Table A1 in Appendix A. Fifty two (52) of these polymorphisms were synonymous, 12 nonsynonymous and four (4) non-coding, however, none of these alterations were novel. The role of the phylogenetically important polymorphisms within this region is further discussed in Section 5.3.2 and 5.3.3.

5.2.6 mtDNA region amplified using primers L7882 and H9928

The region amplified using primers L7882 and H9928 ranged from nucleotide position 7883 to 9927 (Maca-Meyer *et al.*, 2001). This region fully covers the genes for ATP6 (np 8527 - 9207) and partially the COIII (np 9207 - 9990) gene [Mitomap, 2008a]. Figure 5.32 is a representative of the PCR products generated from this region. Artefacts observed in Figure 5.32 include background smears (discussed in Section 5.1.2.1.2).

Figure 5.32 Photographic representation of PCR products amplified using primers L7882 and H9928

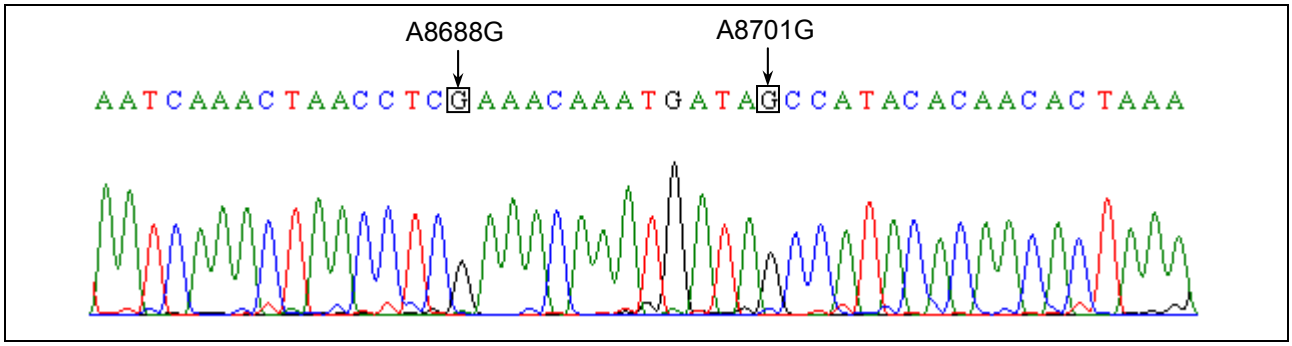


Fragments were electrophoresed on a 0.9% agarose gel at 10 $V.cm^{-1}$ for 30 min. bp = molecular weight in base pairs, C1 = positive control DNA sample, MM = molecular weight marker, Neg = negative control sample. The numbers above the gel wells refer to study participants.

Figure 5.33 is a representative electropherogram for fragments sequenced in this region. The A8688G polymorphism (Kivisild *et al.*, 2006) in the ATP6 gene (Mitomap, 2008c) was harboured by a single sequence from the Dominican Republic belonging to an unspecified L3 subhaplogroup. In this study the polymorphism was harboured by individual UG_CGR_71699 who clustered with haplogroup L3f and L3h sequences from the African pool, as indicated in Figure 5.59 and Figure 5.60. The phylogenetic relevance of this polymorphism requires further investigation by sequencing larger sample sets, especially those possessing haplogroup L3f and L3h mtDNA profiles. The A8701G non-synonymous polymorphism occurring in the ATP6 gene (Mitomap, 2008c) was harboured by all

individuals in this investigation. This polymorphism was harboured by African samples belonging to macrohaplogroup L and Asian haplogroup M samples (Herrnstadt *et al.*, 2002).

Figure 5.33 Representative electropherogram for DNA fragments sequenced within mtDNA nucleotide positions 7883 and 9927



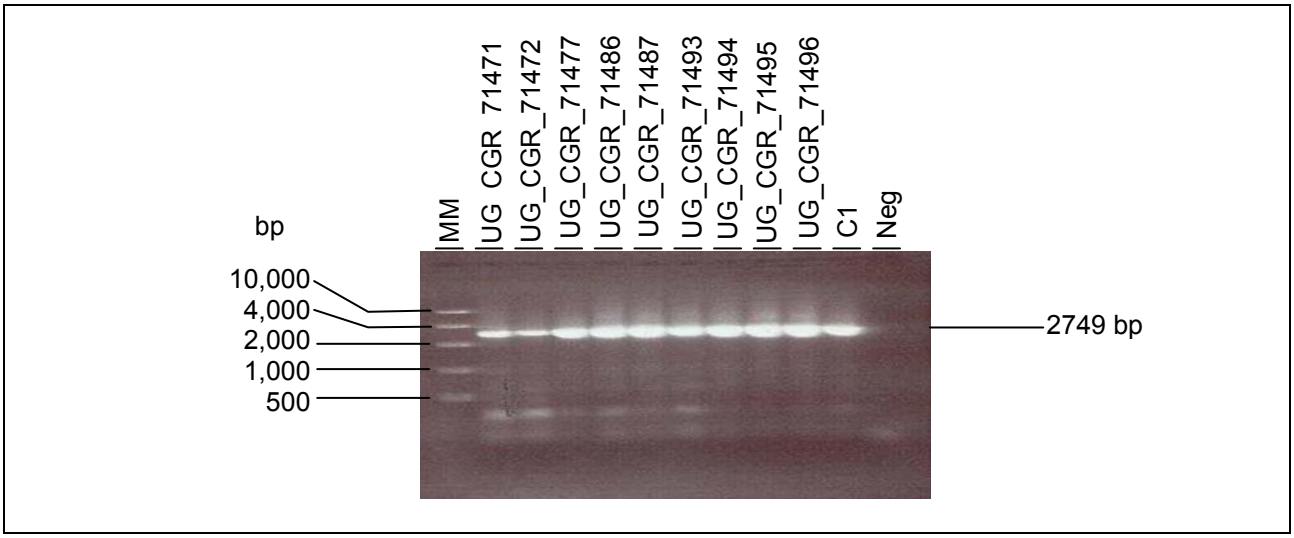
A = adenine; C = cytosine; G = guanine; T = thymine. Reported alterations have been indicated in black.

A total of 64 polymorphisms were defined within this region as indicated in Table A1 in Appendix A. Fifty one (51) of these polymorphisms were synonymous, 11 nonsynonymous, two (2) non-coding while none were novel. The role of the phylogenetically important mutations has been outlined in Section 5.3.2 and 5.3.3.

5.2.7 mtDNA region amplified using primers L8799 and H11527

The region amplified using primers L8799 and H11527 ranged from nucleotide position 8800 to 11526 (Maca-Meyer *et al.*, 2001). This region partially covers the genes for ATP6 (np 8527 - 9207) and ND4 (np 10760 - 12137) as well as including the entire genes for COIII (np 9207 - 9990), $tRNA^{Gly}$ (np 9991 - 10058), ND3 (np 10059 - 10404), $tRNA^{Arg}$ (10405 - 10469), ND4L (np 10470 - 10766) [Mitomap, 2008a]. Figure 5.34 is a representative electropherogram of the PCR products generated from this region. Artefacts observed in Figure 5.34 include background smears (discussed in Section 5.1.2.1.2) and wide variation in amplification efficiency (discussed in 5.1.2.1.1).

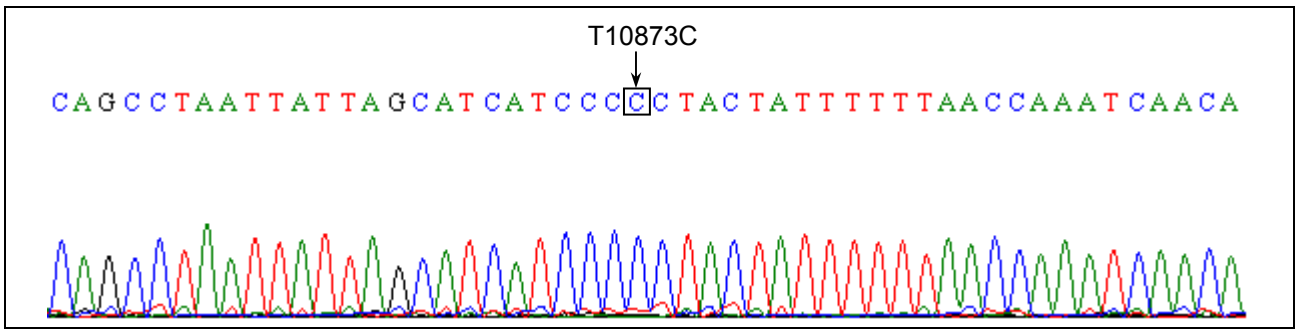
Figure 5.34 Photographic representation of PCR products amplified using primers L8799 and H11527



Fragments were electrophoresed on a 0.9% agarose gel at 10 V.cm^{-1} for 30 min. bp = molecular weight in base pairs, C1 = positive ontrol DNA sample, MM = molecular weight marker, Neg = negative control sample. The numbers above the gel wells refer to study participants.

Figure 5.35 is a representative electropherogram for the fragments sequenced in this region. The T10873C polymorphism in the ND4 gene (Mitomap, 2008c) was harboured by all individuals investigated in this study. This synonymous polymorphism occurs in all African samples belonging to macrohaplogroup L and all Asian samples belonging to macrohaplogroup M (Herrnstadt *et al.*, 2002).

Figure 5.35 Representative electropherogram for DNA fragments sequenced within mtDNA nucleotide positions 8799 and 11527

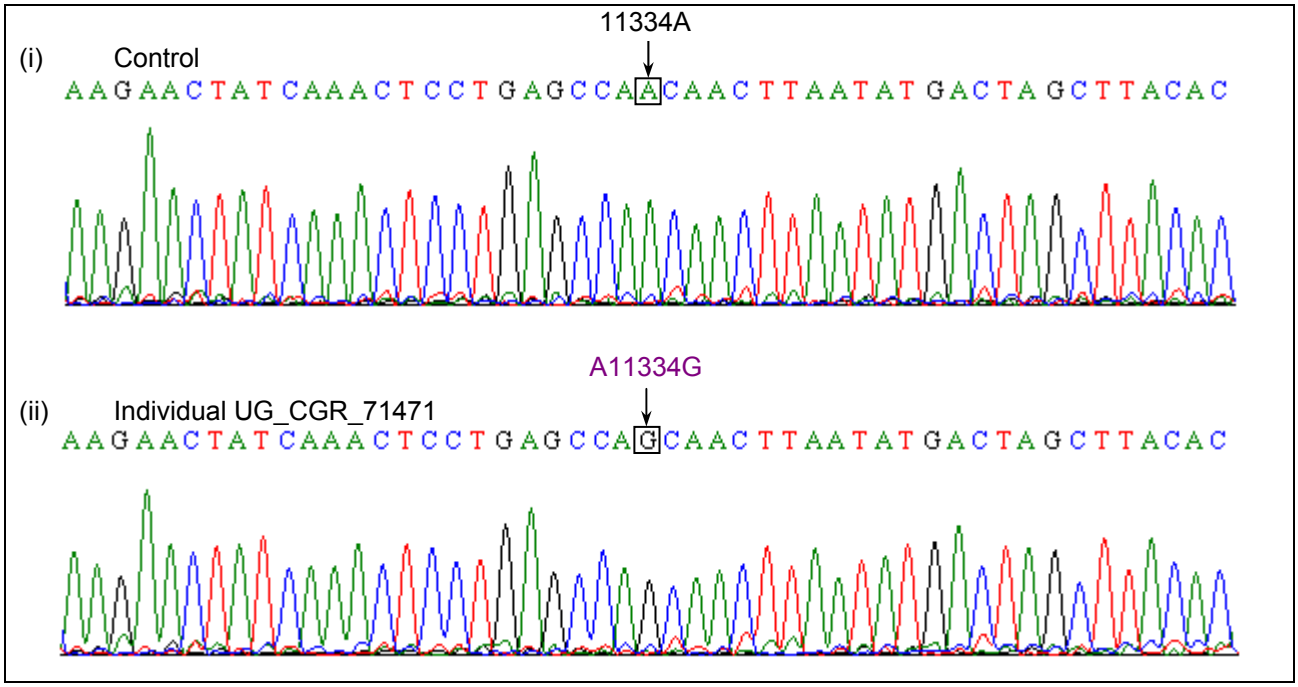


A = adenine; C = cytosine; G = guanine; T = thymine. The reported alteration has been indicated in black.

A total of 87 polymorphisms were generated within this region as indicated in Table A1 in Appendix A. Seventy (70) of these polymorphisms were synonymous, 15 nonsynonymous, two (2) non-coding and one (1) was novel i.e. A11334G. The representative electropherogram for the novel polymorphism is indicated in Figure 5.36. The polymorphism occurred in only one (1) individual in this investigation (Acholi individual, UG_CGR_71471) who happens to belong to haplogroup L2d which was very rare in the global African pool as indicated in Figure 5.59 and Figure 5.60 in Section 5.3.2 and 5.3.3.

Investigations involving many L2d samples need to be conducted to study the phylogenetic value of the A11334G polymorphism. The explanation of the role of the phylogenetically significant mutations in this amplified region is dealt with in Section 5.3.3.

Figure 5.36 Representative electropherogram for the A11334G novel polymorphism in the ND4 gene

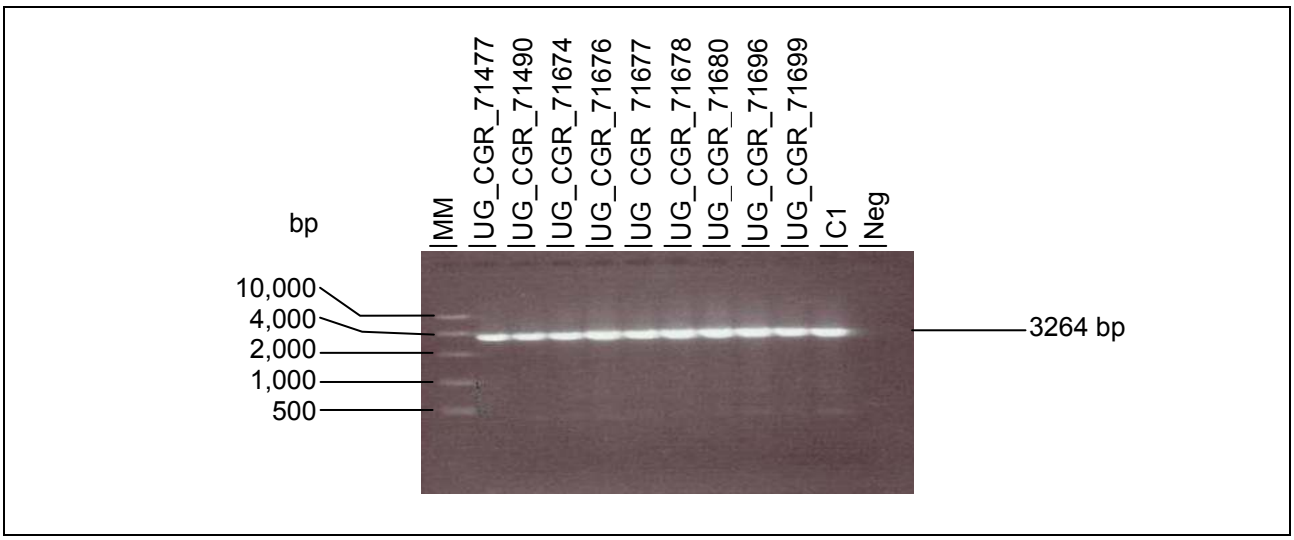


Control = represents electropherogram of an individual with a sequence identical to the rCRS at nucleotide position 11334. The nucleotide in violet text represents a novel alteration while the corresponding polymorphism for the control sample is indicated in black. Arrows represent the nucleotide (nt) positions at which the alterations occur. A = adenine; C = cytosine; G = guanine; T = thymine.

5.2.8 mtDNA region amplified using primers L10403 and H13666

The region amplified using primers L10403 and H13666 ranged from nucleotide position 10404 to 13665 (Maca-Meyer *et al.*, 2001). This region partially covers the genes for ND3 (np 10059 - 10404), ND5 (np 12337 - 14148) and fully includes the genes for $tRNA^{Arg}$ (np 10405 - 10469), ND4L (np 10470 - 10766), ND4 (np 10760 - 12137), $tRNA^{His}$ (np 12138 - 12206), $tRNA^{Ser(AGY)}$ (np 12207 - 12265) and $tRNA^{Leu(CUN)}$ (np 12266 - 12336) [Mitomap, 2008a]. Figure 5.37 is a representative electropherogram of the PCR products generated from this region. Artefacts observed in Figure 5.37 included background smears (discussed in Section 5.1.2.1.2) and secondary amplification (discussed in Section 5.1.2.1.3).

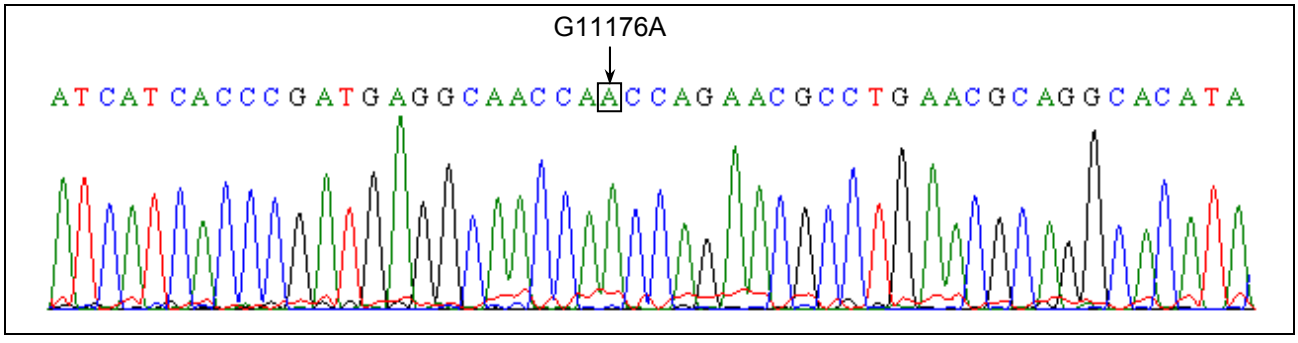
Figure 5.37 Photographic representation of PCR products amplified using primers L10403 and H13666



Fragments were electrophoresed on a 0.9% agarose gel at 10 $V.cm^{-1}$ for 30 min. bp = molecular weight in base pairs, C1 = positive control DNA sample, MM = molecular weight marker, Neg = negative control sample. The numbers above the gel wells refer to study participants.

Figure 5.38 is a representative electropherogram for the fragments sequenced in this region. The G11176A polymorphism occurred in the ND4 gene of the mtDNA (Mitomap, 2008c) in a single Acholi individual (UG_CGR_71468), one Lugbara individual (UG_CGR_71486) and five Baganda individuals. As outlined in Section 5.3.3, the G11176A polymorphism was harboured by individuals in this study possessing haplogroup L0c mtDNA sequences as well as individual UG_CGR_71486 who was designated as belonging to haplogroup L3i, as illustrated in Figure 5.59 and Figure 5.60.

Figure 5.38 Representative electropherogram for DNA fragments sequenced within mtDNA nucleotide positions 10404 and 13665

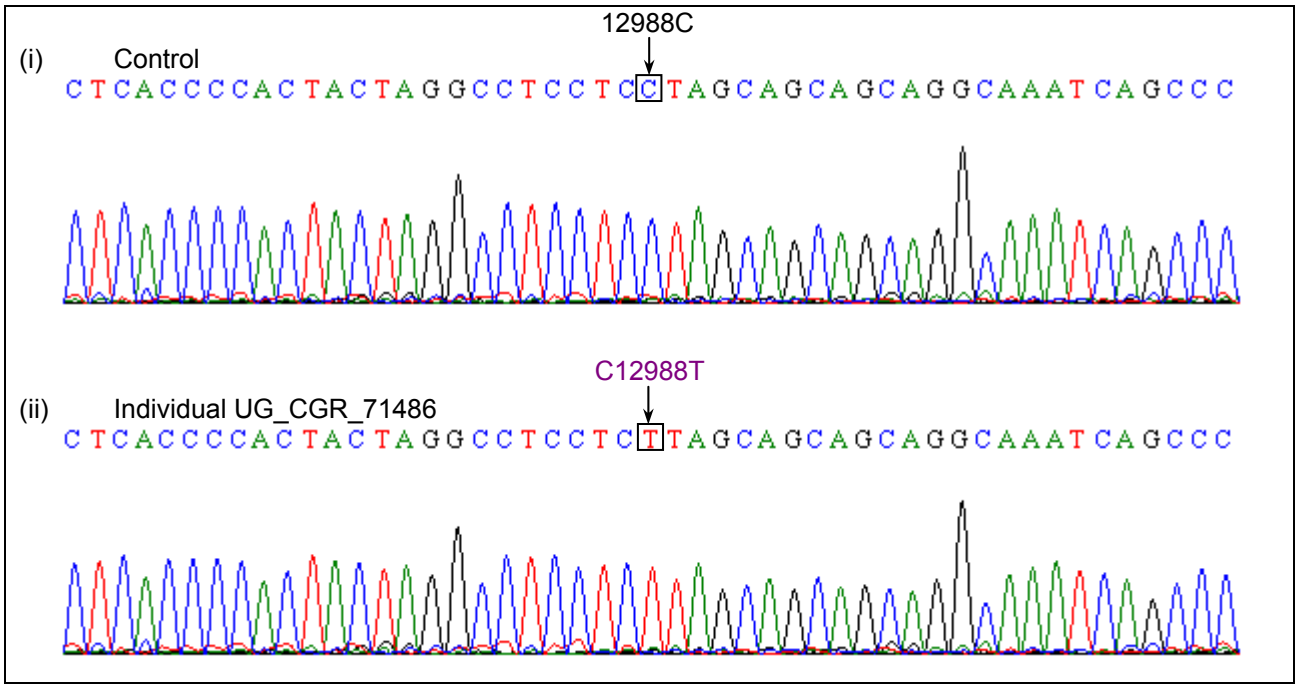


A = adenine; C = cytosine; G = guanine; T = thymine. The reported alteration has been indicated in black.

A total of 96 polymorphisms were determined within this region as indicated in Table A1 in Appendix A. Seventy-eight (78) of these polymorphisms were synonymous, 10 were nonsynonymous, eight (8) were non-coding and six (6) were novel (A11334G in overlap section as indicated above as well as C12988T, C13122A, C13125T). The representative

electropherograms for the novel polymorphisms are indicated in Figure 5.39 to Figure 5.41.

Figure 5.39 Representative electropherogram for the C12988T novel polymorphism in the ND5 gene

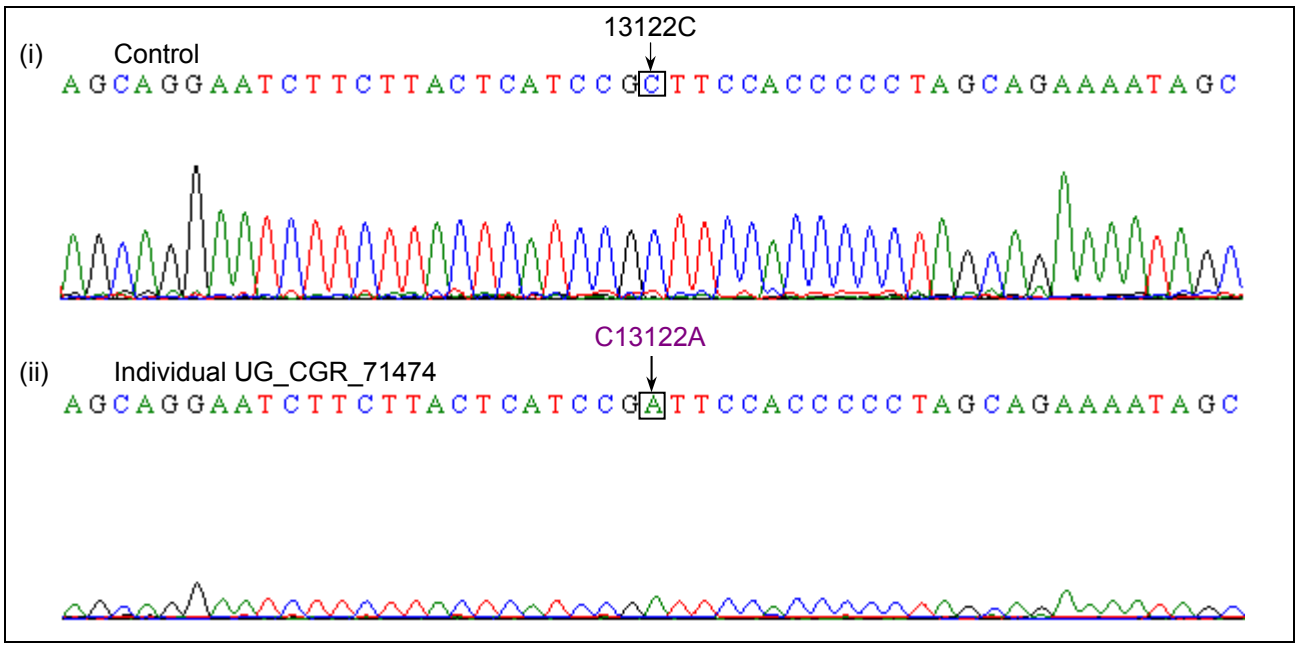


Control = represents electropherogram of an individual with a sequence identical to the rCRS at nucleotide position 12988. The nucleotide in violet text represents a novel alteration while the corresponding polymorphism for the control sample is indicated in black. Arrows represent the nucleotide (nt) positions at which the alterations occur. A = adenine; C = cytosine; G = guanine; T = thymine.

The C12988T polymorphism, as illustrated in Figure 5.39, was possessed by individual UG_CGR_71486 (a Lugbara), was synonymous and occurred in the ND5 gene region (Mitomap, 2008c). Being nonsynonymous (Mitomap, 2008c), the polymorphism is unlikely to have an effect on the proper functioning of the mitochondria. There is a need to sequence more samples belonging to haplogroup L4a (see Figure 5.59 and Figure 5.60) in order to explore the phylogenetic significance of this polymorphism.

The C13122A polymorphism, as illustrated in Figure 5.40, was harboured by individual UG_CGR_71474 belonging to the Acholi tribe. This polymorphism was a transversion and occurred in the ND5 gene of the mtDNA (Mitomap, 2008c). The polymorphism was synonymous (Mitomap, 2008c) and therefore was unlikely to lead to an alteration in the way the mitochondria functions. The phylogenetic significance of the polymorphism has to be assessed by investigating larger sample sets, belonging to haplogroup L4a as is the case for sample UG_CGR_71474 in Section 5.3.2 and 5.3.3.

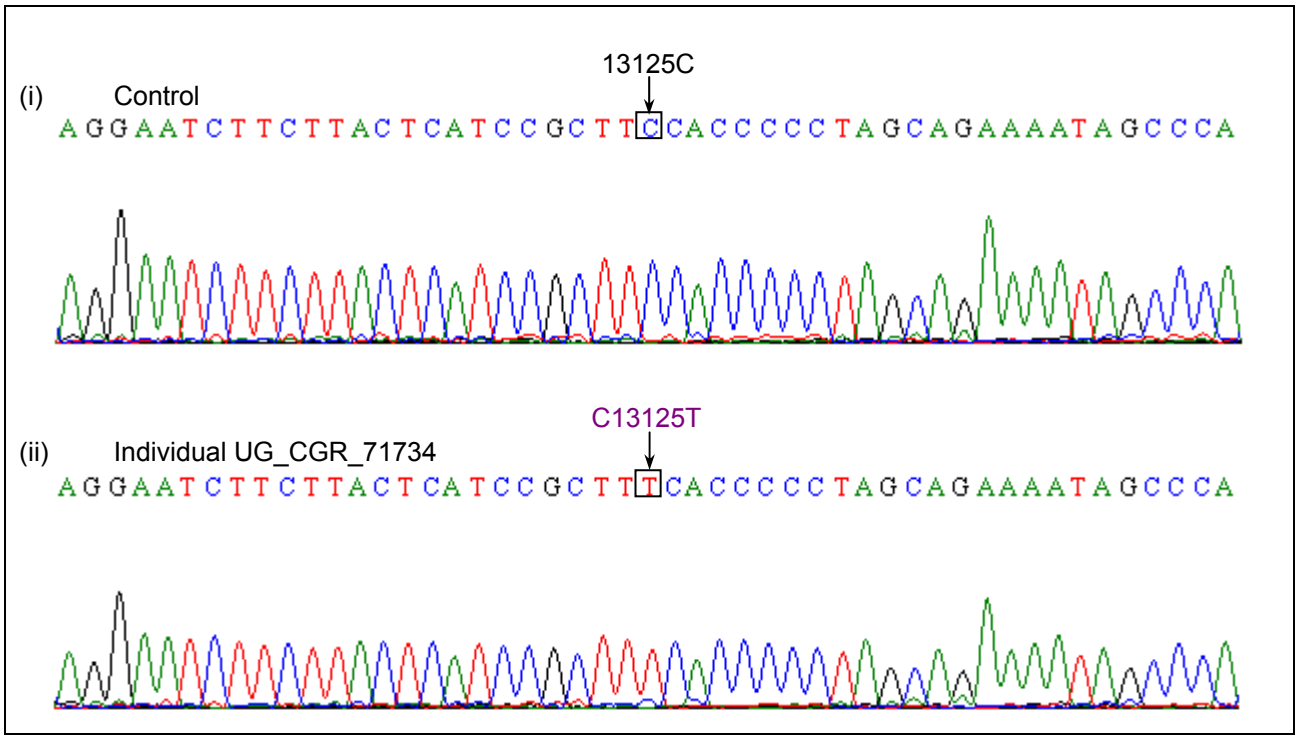
Figure 5.40 Representative electropherogram for the C13122A novel polymorphism in the ND5 gene



Control = represents electropherogram of an individual with a sequence identical to the rCRS at nucleotide position 13122. The nucleotide in violet text represents a novel alteration while the corresponding polymorphism for the control sample is indicated in black. Arrows represent the nucleotide (nt) positions at which the alterations occur. A = adenine; C = cytosine; G = guanine; T = thymine. The low signal intensity in the electropherogram was discussed in Section 5.1.5.1.3.

The C13125T polymorphism, as illustrated in Figure 5.41, was harboured by individual UG_CGR_71734 (an Acholi) in the ND5 gene of its mtDNA (Mitomap, 2008c). The

Figure 5.41 Representative electropherogram for the C13125T novel polymorphism in the ND5 gene



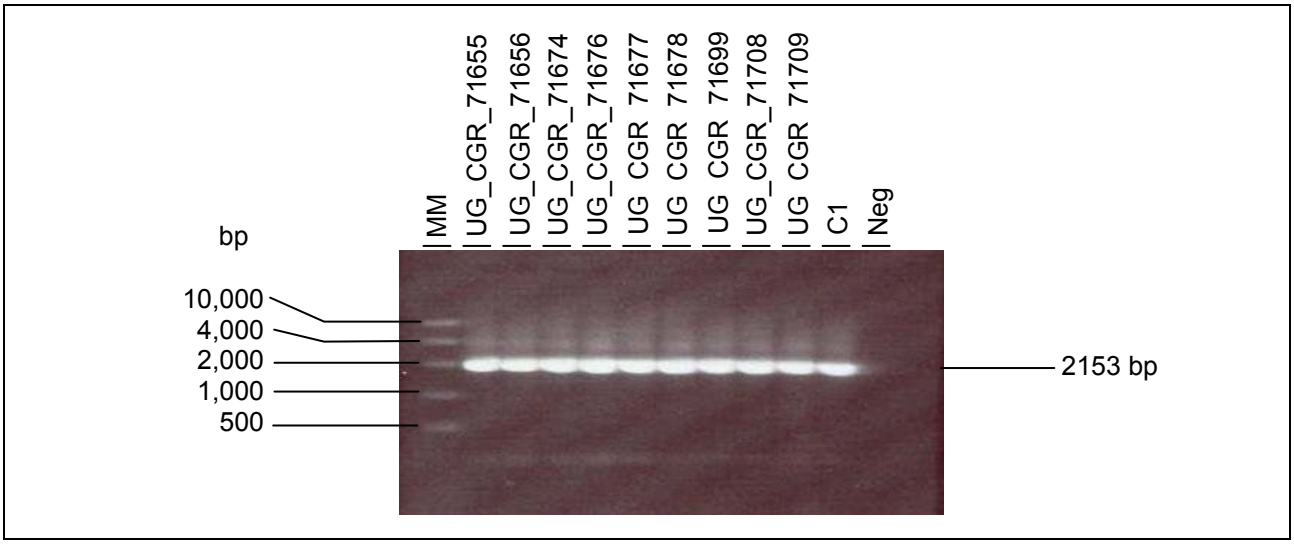
Control = represents electropherogram of an individual with a sequence identical to the rCRS at nucleotide position 13125. The nucleotide in violet text represents a novel alteration while the corresponding polymorphism for the control sample is indicated in black. Arrows represent the nucleotide (nt) positions at which the alterations occur. A = adenine; C = cytosine; G = guanine; T = thymine.

polymorphism was unlikely to have a functional effect on the mitochondria since it was a synonymous mutation (Mitomap, 2008c). Further investigations need to be performed on larger sample sets to assess the phylogenetic significance of this polymorphism amongst the Acholi from Uganda and the surrounding countries.

5.2.9 mtDNA region amplified using primers L12572 and H14685

The region amplified using primers L12572 and H14685 ranged from nucleotide position 12573 to 14684 (Maca-Meyer *et al.*, 2001). This region partially covers the genes for ND5 (np 12337 - 14148), and $tRNA^{Glu}$ (np 14674 - 14742) as well as fully includes the ND6 gene which occupies np 14149 - 14673 (Mitomap, 2008c). Figure 5.42 is representative of the PCR products generated from this region. Artefacts observed in Figure 5.42 included background smears (discussed in Section 5.1.2.1.2), secondary amplification (discussed in Section 5.1.2.1.3) and primer dimers (discussed in Section 5.1.2.1.4).

Figure 5.42 Photographic representation of PCR products amplified using primers L12572 and H14685

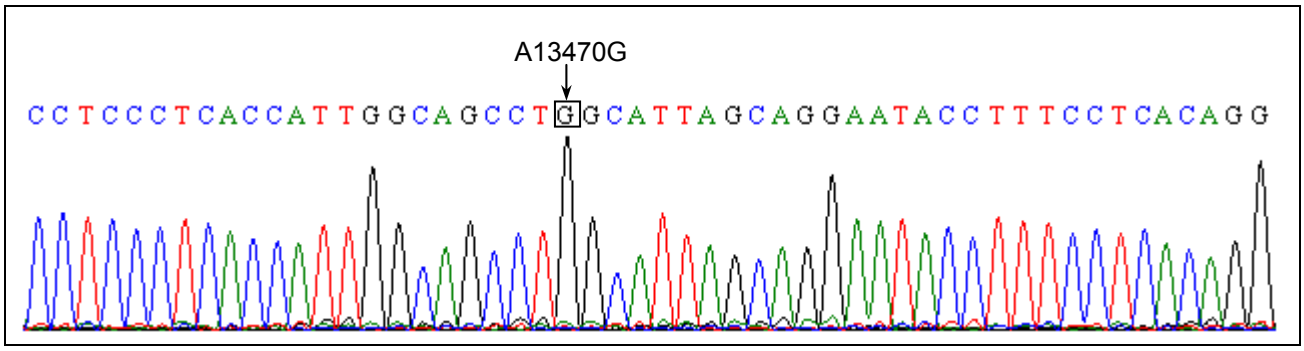


Fragments were electrophoresed on a 0.9% agarose gel at 10 $V.cm^{-1}$ for 30 min. bp = molecular weight in base pairs, C1 = positive control DNA sample, MM = molecular weight marker, Neg = negative control sample. The numbers above the gel wells refer to study participants.

Figure 5.43 is a representative electropherogram for fragments sequenced in this region. The A13470G polymorphism was observed among two Acholi and four Baganda individuals but not in any Lugbara samples in this investigation. The polymorphism does not cause a change in the amino acid encoded by the ND5 gene and is therefore unlikely to cause a functional change in the mitochondria. The A13470G polymorphism was harboured by individuals belonging to haplogroup L4 (Torroni *et al.*, 2006) which was

largely in agreement with this study where the individuals who harboured the polymorphism clustered in haplogroup L4. However, individual UG_CGR_71494 who was designated as belonging to haplogroup L0a also harboured the polymorphism. Further investigations need to be performed on large samples to ascertain the phylogenetic value of this polymorphism among haplogroup L0a samples.

Figure 5.43 Representative electropherogram for DNA fragments sequenced within mtDNA nucleotide positions 12572 to 14685

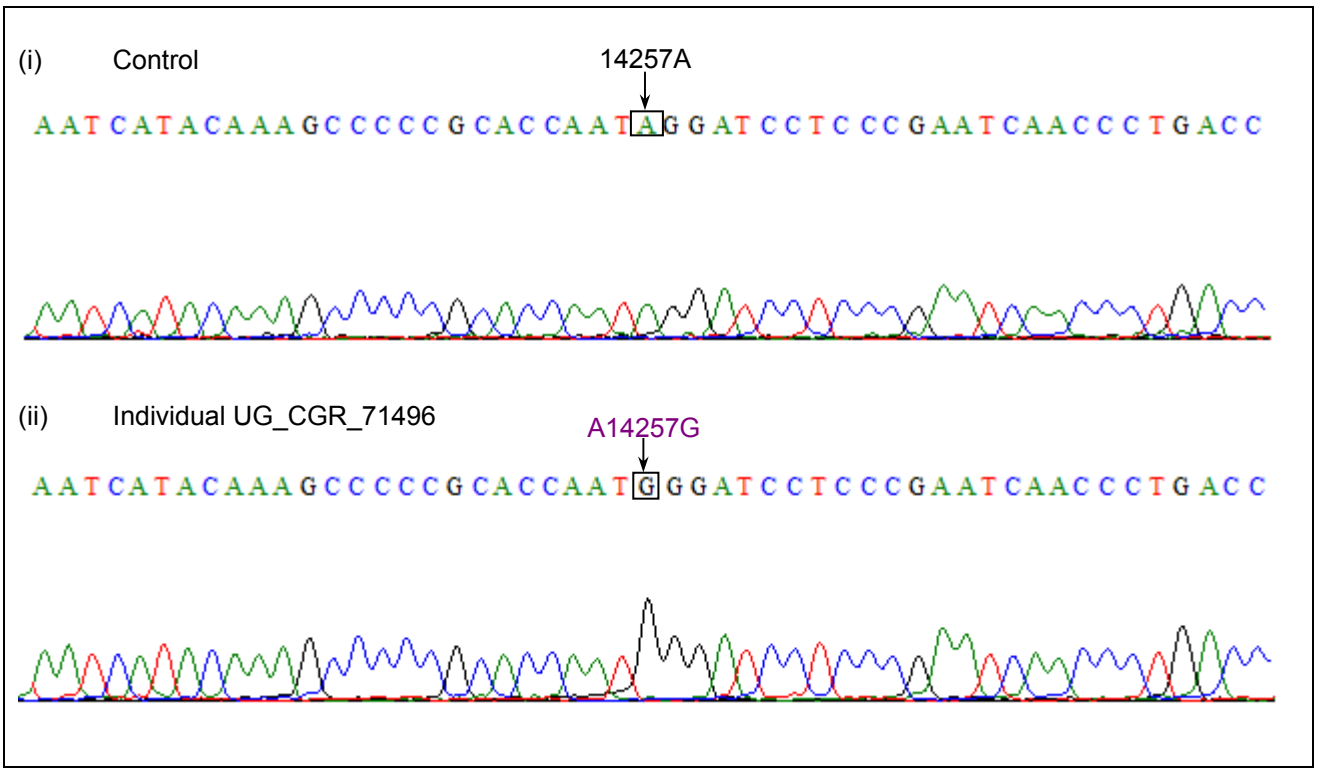


A = adenine; C = cytosine; G = guanine; T = thymine. The reported alteration has been indicated in black.

A total of 68 polymorphisms were generated within this region as indicated in Table A1 in Appendix A. Fifty three (53) of these polymorphisms were synonymous, 15 were nonsynonymous and five (5) were novel (C12988T, C13122A, C13125T [in overlap sections of neighbouring amplified fragments], A14257G, A14573G). All polymorphisms in this region were protein-coding. The representative electropherograms for the novel polymorphisms are indicated in Figure 5.44 and Figure 5.45.

The A14257G polymorphism, as illustrated in Figure 5.44, was harboured by two individuals, i.e. UG_CGR_71496 and UG_CGR_71657, both being Baganda. The phylogenetic significance of this polymorphism as outlined in Section 5.3.3 needs further investigation by sequencing and analysis of more Baganda samples. The polymorphism is unlikely to cause a change in the way that the mitochondria functions since it is synonymous (Mitomap, 2008c).

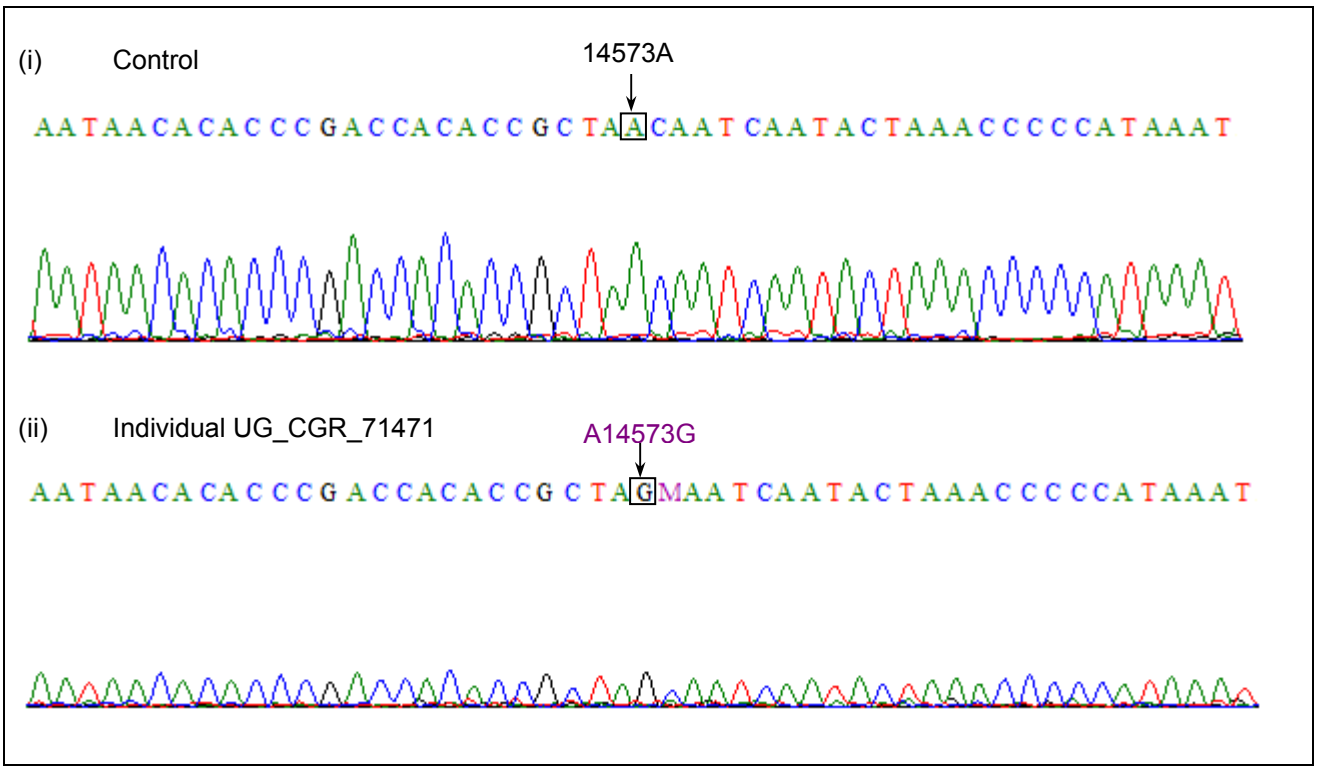
Figure 5.44 Representative electropherogram for the A14257G novel polymorphism in the ND6 gene



Control = represents electropherogram of an individual with a sequence identical to the rCRS at nucleotide position 14257. The nucleotide in violet text represents a novel alteration while the corresponding polymorphism for the control sample is indicated in black. Arrows represent the nucleotide (nt) positions at which the alterations occur. A = adenine; C = cytosine; G = guanine; T = thymine. The low signal intensity in the electropherogram was discussed in Section 5.1.5.1.3.

The A14573G nonsynonymous polymorphism in the ND6 gene (Mitomap, 2008c) was harboured by individual UG_CGR_71471, who was an Acholi. This polymorphism causes a change in the amino acid encoded from valine to alanine (Mitomap, 2008c), which are both hydrophobic and therefore unlikely to bring about a change in the functioning of the mitochondria. Further investigations involving greater numbers of Acholi and samples belonging to haplogroup L2d (See Figure 5.59 and Figure 5.60) have to be conducted to explore the phylogenetic relevance of this SNP.

Figure 5.45 Representative electropherogram for the A14573G novel polymorphism in the ND6 gene

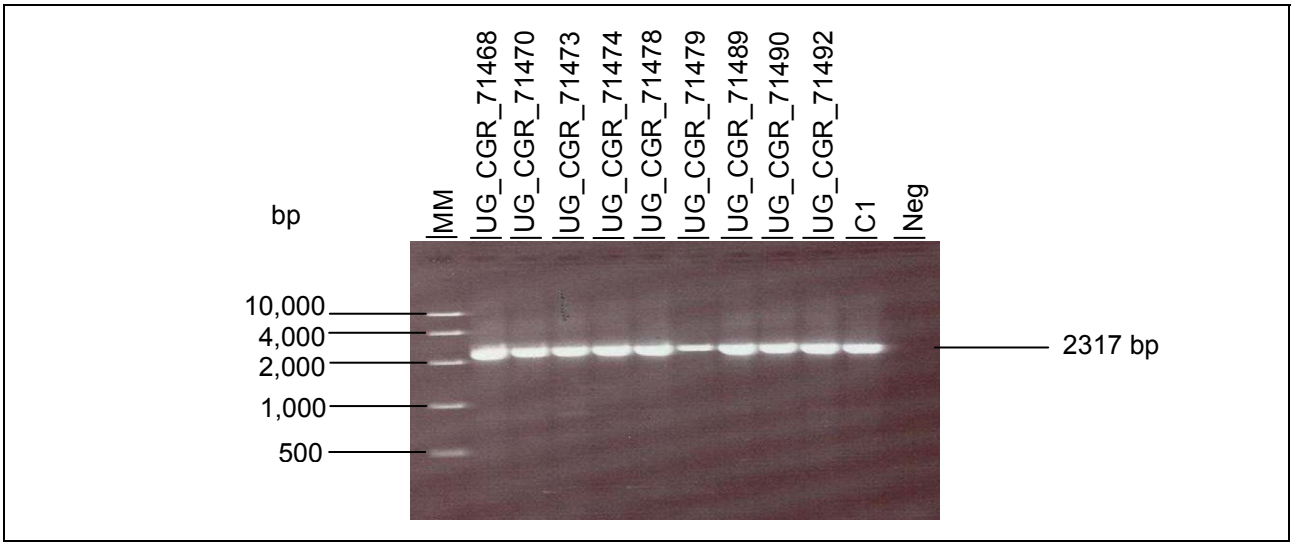


Control = represents electropherogram of an individual with a sequence identical to the rCRS at nucleotide position 14573. The nucleotide in violet text represents a novel alteration while the corresponding polymorphism for the control sample is indicated in black. Arrows represent the nucleotide (nt) positions at which the alterations occur. A = adenine; C = cytosine; G = guanine; T = thymine; M = C or A (NC_IUB, 1985). The ambiguous bases have been discussed in Section 5.1.5.1.2 but in this instance, M was edited to C after visual inspection of the electropherogram. The low signal intensity in the electropherogram was discussed in Section 5.1.5.1.3.

5.2.10 mtDNA region amplified using primers L14125 and H16401

The region amplified using primers L14125 and H16401 ranged from nucleotide position 14126 to 16400 (Maca-Meyer *et al*., 2001). This region partially covers the region of the ND5 gene (np 12337 - 14148), and the membrane attachment site (np 15925 - 499), and also fully includes the genes for ND6 (np 14149 - 14673), $tRNA^{Glu}$ (np 14674 - 14742), Cytb (np 14747 - 15887) $tRNA^{Thr}$ (np 15888 - 15953) and $tRNA^{Pro}$ (np 15955 - 16023). It also included the non-coding nucleotides at position 14743 - 14746, position 15954, and the HVI region (np 16024 - 16383) [Mitomap, 2008a]. Figure 5.46 is representative of the PCR products generated for this region. Artefacts observed in Figure 5.46 included background smears (discussed in Section 5.1.2.1.2) and a wide variation in amplification efficiency (discussed in Section 5.1.2.1.1).

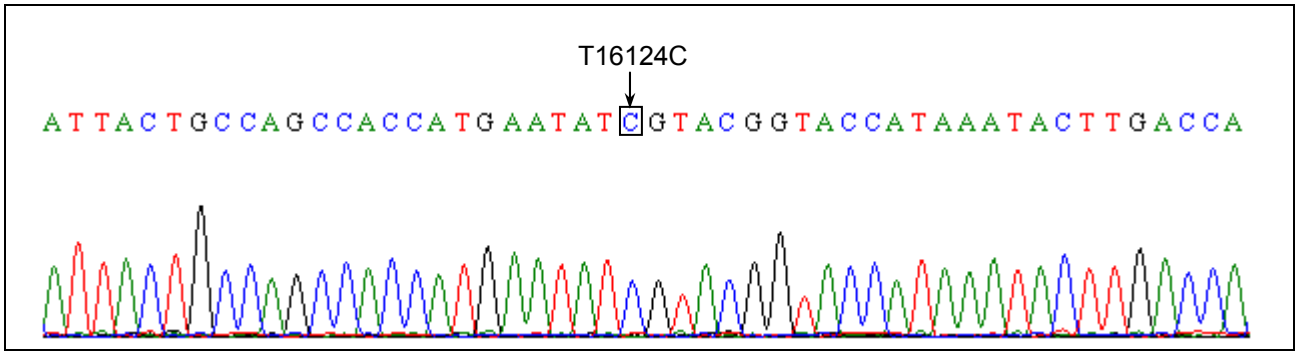
Figure 5.46 Photographic representation of PCR products amplified using primers L14125 and H16401



Fragments were electrophoresed on a 0.9% agarose gel at 10 V.cm^{-1} for 30 min. bp = molecular weight in base pairs, C1 = positive control DNA sample, MM = molecular weight marker, Neg = negative control sample. The numbers above the gel wells refer to study participants. Variation in amplification efficiency has been discussed in Section 5.1.2.1.1.

Figure 5.47 is a representative electropherogram for the fragments sequenced in this region. The T16124C polymorphism was harboured by two Acholi, two Baganda and one Lugbara individual that participated in this study. This polymorphism is harboured by individuals in haplogroups L3b, L3c and L3d (Kivisild *et al*., 2004; Jackson *et al*., 2005; Behar *et al*., 2008). However, none of the study participants belonged to haplogroup L3b and L3d.

Figure 5.47 Representative electropherogram for DNA fragments sequenced within mtDNA nucleotide positions 14125 to 16401

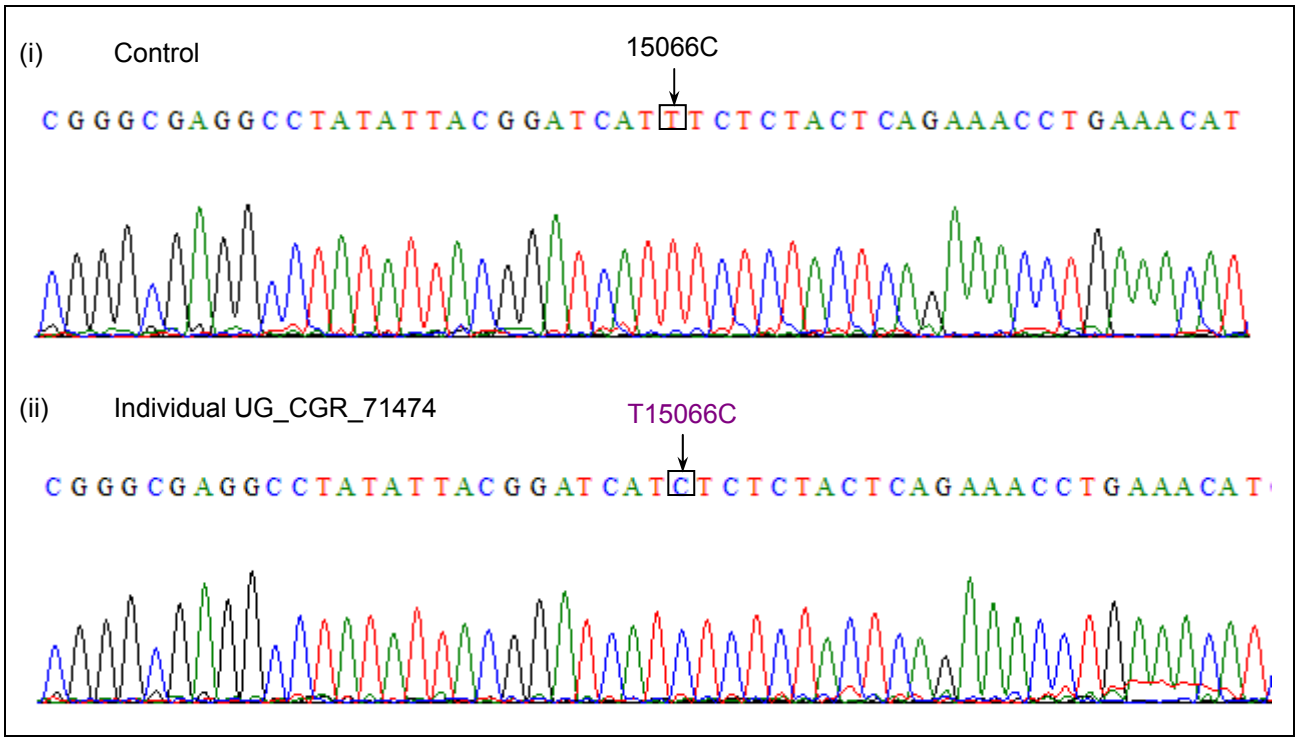


A = adenine; C = cytosine; G = guanine; T = thymine. The reported alteration has been indicated in black.

A total of 148 polymorphisms were determined within this region as indicated in Table A1 in Appendix A. Thirty nine (39) of these polymorphisms were synonymous, 21 were nonsynonymous, 88 were non-coding and nine (9) were novel (A14257G, A14573G [in overlap sections with neighbouring amplified fragments and previously discussed as indicated in Figure 5.44 and Figure 5.45 respectively], T15066C, A15328G, C15446T,

C15574T, A15655G, A15673G, [C16112T, in overlap]). The representative electropherograms for the novel polymorphisms are indicated in Figure 5.48 to Figure 5.53. The T15066C polymorphism, as discussed in Section 5.3.2 and 5.3.3, is expressed on a haplotype of haplogroup L3.

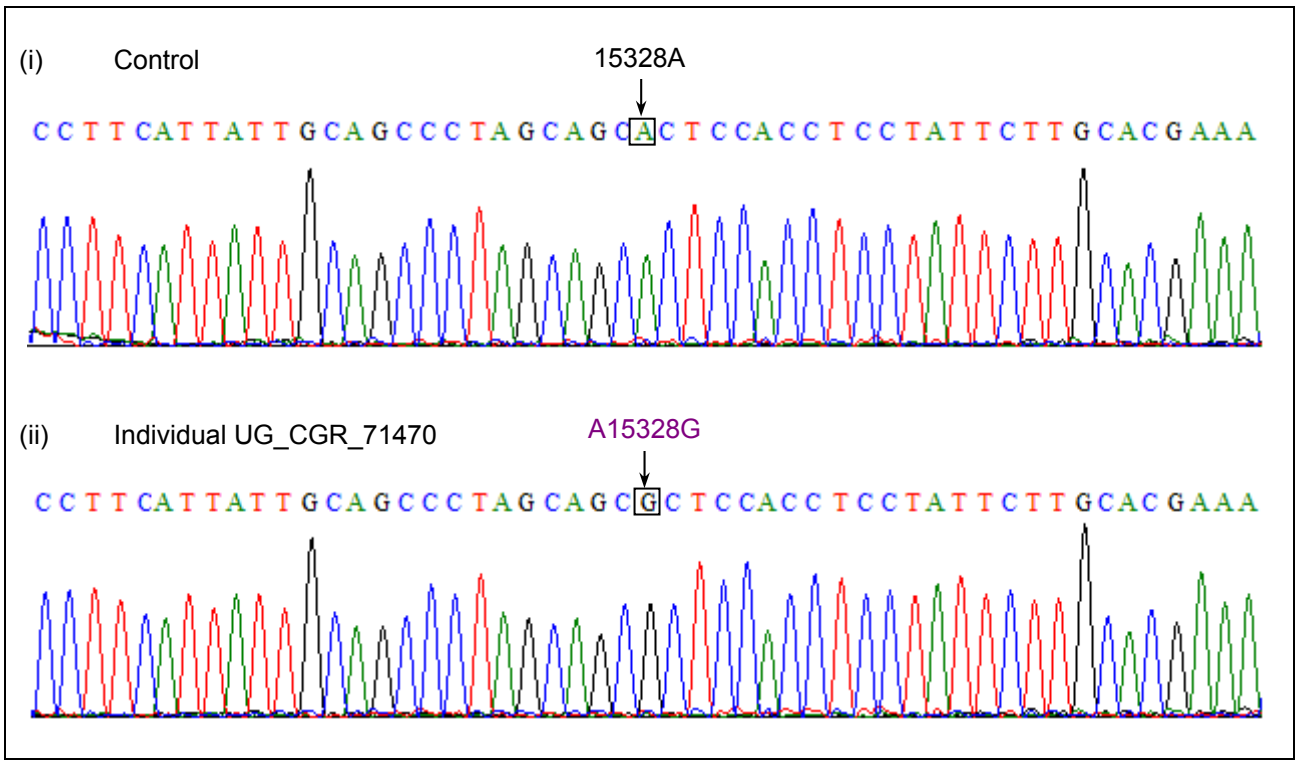
Figure 5.48 Representative electropherogram for the T15066C novel polymorphism in the Cytb gene



Control = represents electropherogram of an individual with a sequence identical to the rCRS at nucleotide position 15066. The nucleotide in violet text represents a novel alteration. Arrows represent the nucleotide (nt) positions at which the alterations occur. A = adenine; C = cytosine; G = guanine; T = thymine.

The T15066C polymorphism, as illustrated in Figure 5.48, was harboured by Acholi individual UG_CGR_71474. The phylogenetic significance of the polymorphism needs further investigation by sequencing more Acholi samples. The functional effect of the polymorphism in cytochrome b reductase complex (Mitomap, 2008c) of the respiratory chain also needs further assessment since the polymorphism causes a change of the amino acid to one which is more hydrophilic.

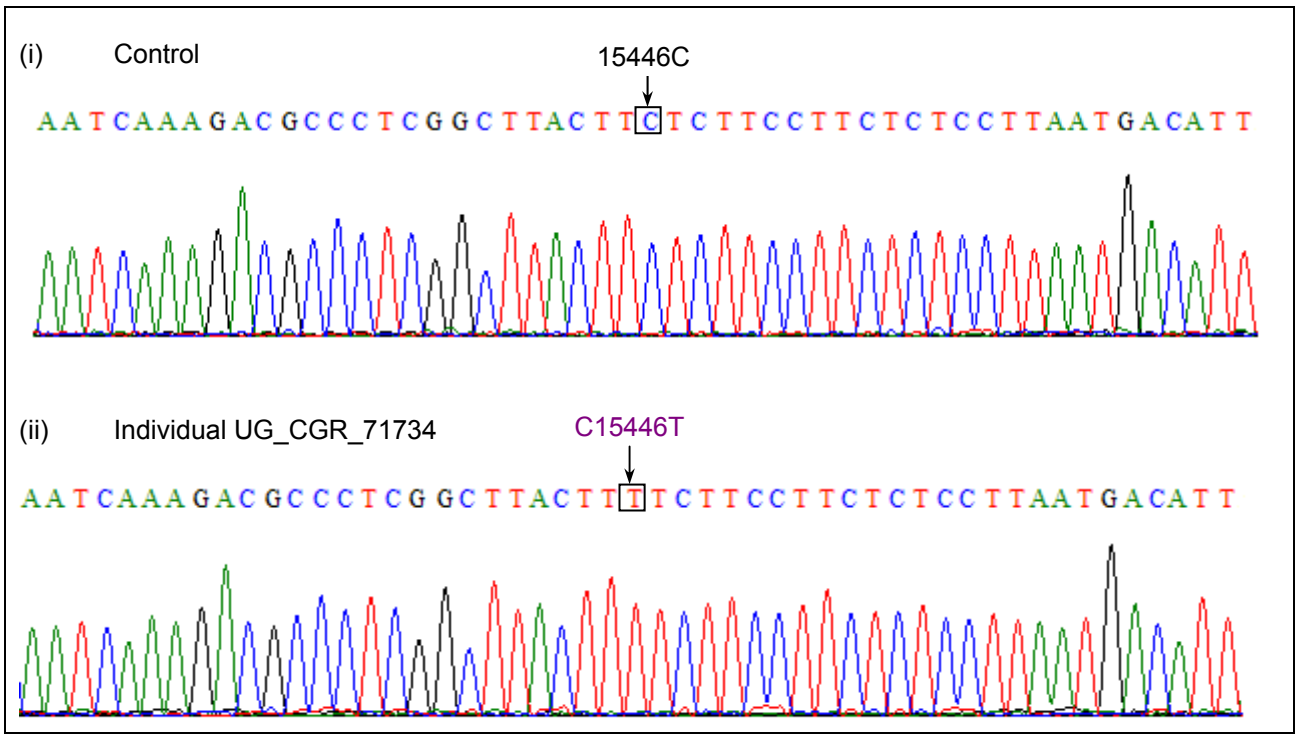
Figure 5.49 Representative electropherogram for the A15328G novel polymorphism in the Cytb gene



Control = represents electropherogram of an individual with a sequence identical to the rCRS at nucleotide position 15328. The nucleotide in violet text represents a novel alteration. Arrows represent the nucleotide (nt) positions at which the alterations occur. A = adenine; C = cytosine; G = guanine; T = thymine.

The A15328G polymorphism, as indicated in Figure 5.49, was harboured by individual UG_CGR_71470 (an Acholi). The polymorphism occurred in the Cytb gene of the mtDNA and was a synonymous mutation (Mitomap, 2008c). It was therefore unlikely to alter the way in which the mitochondria functions under normal conditions. The phylogenetic significance of the polymorphism needs to be elucidated further by analysis of larger Acholi samples.

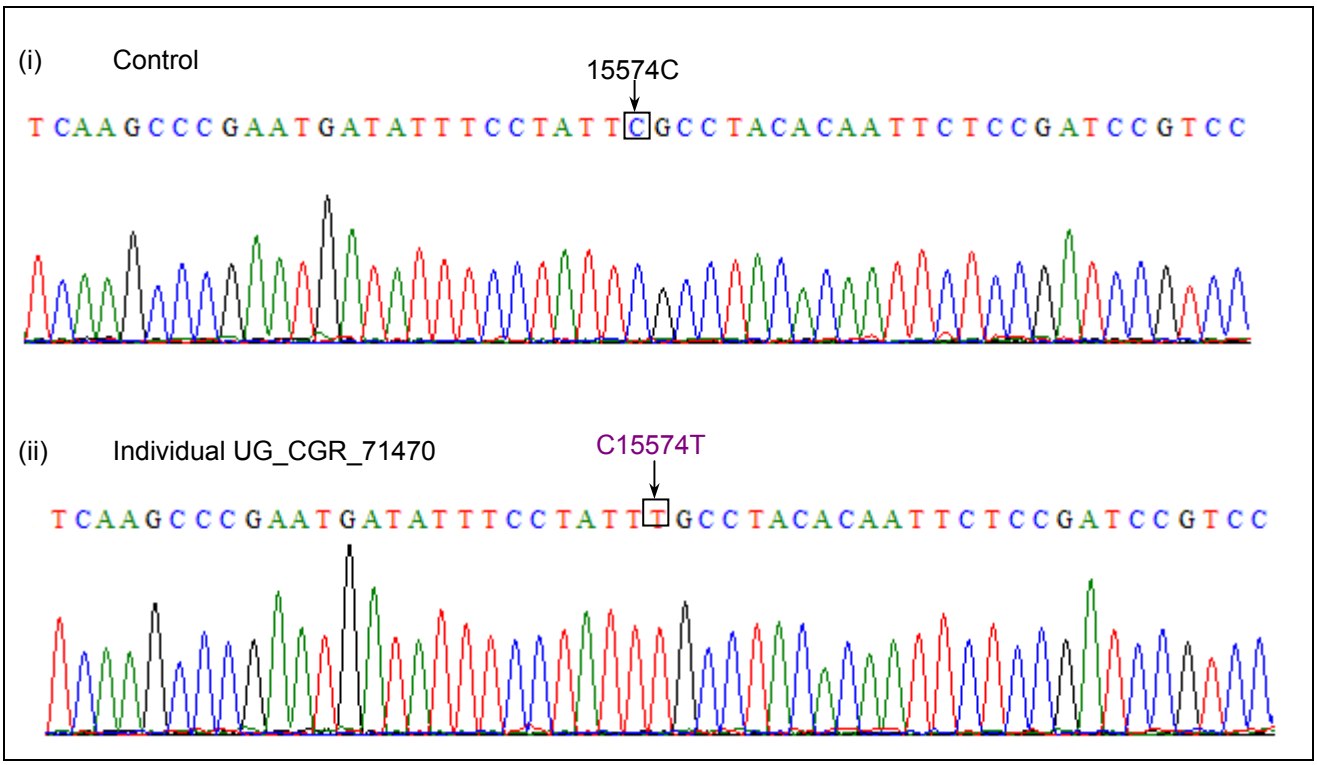
Figure 5.50 Representative electropherogram for the C15446T novel polymorphism in the Cytb gene



Control = represents electropherogram of an individual with a sequence identical to the rCRS at nucleotide position 15446. The nucleotide in violet text represents a novel alteration while the polymorphism in the control sample is indicated in black. Arrows represent the nucleotide (nt) positions at which the alterations occur. A = adenine; C = cytosine; G = guanine; T = thymine.

Figure 5.50 is an illustration of the C15446T novel polymorphism. The polymorphism was harboured by only one Acholi individual (UG_CGR_71734) and caused a change in the amino acid encoded from a Leu to a Phe in the cytb gene (Mitomap, 2008c). Additional studies conducted on larger samples are needed to investigate the phylogenetic value of the T15446C novel polymorphism. However, as outlined in Section 5.3.2 and 5.3.3, the T15446C polymorphism has been expressed on a haplotype belonging to haplgroup L2a.

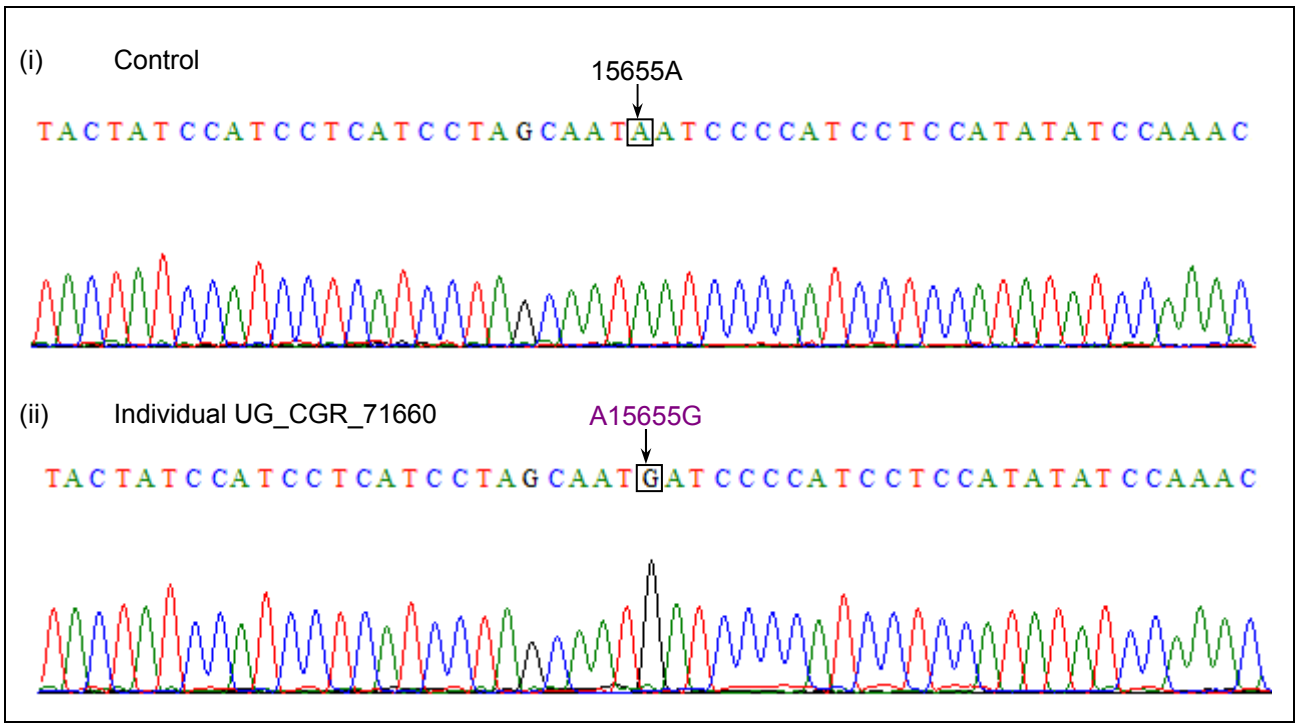
Figure 5.51 Representative electropherogram for the C15574T novel polymorphism in the Cytb gene



Control = represents electropherogram of an individual with a sequence identical to the rCRS at nucleotide position 15574. The nucleotide in violet text represents a novel alteration while the polymorphism in the control sample is indicated in black. Arrows represent the nucleotide (nt) positions at which the alterations occur. A = adenine; C = cytosine; G = guanine; T = thymine.

The C15574T polymorphism, as indicated in Figure 5.51, was harboured by individual UG_CGR_71470, who happened to have been an Acholi. The polymorphism is synonymous (Mitomap, 2008 c) and therefore does not lead to an alteration in the functioning of the mitochondria. The polymorphism, as outlined in Section 5.3.2 and 5.3.3, has been used to characterise a haplotype in haplogroup L2a. Further studies need to be conducted on larger samples of Acholi in haplogroup L2a, to explore the phylogenetic significance of this polymorphism.

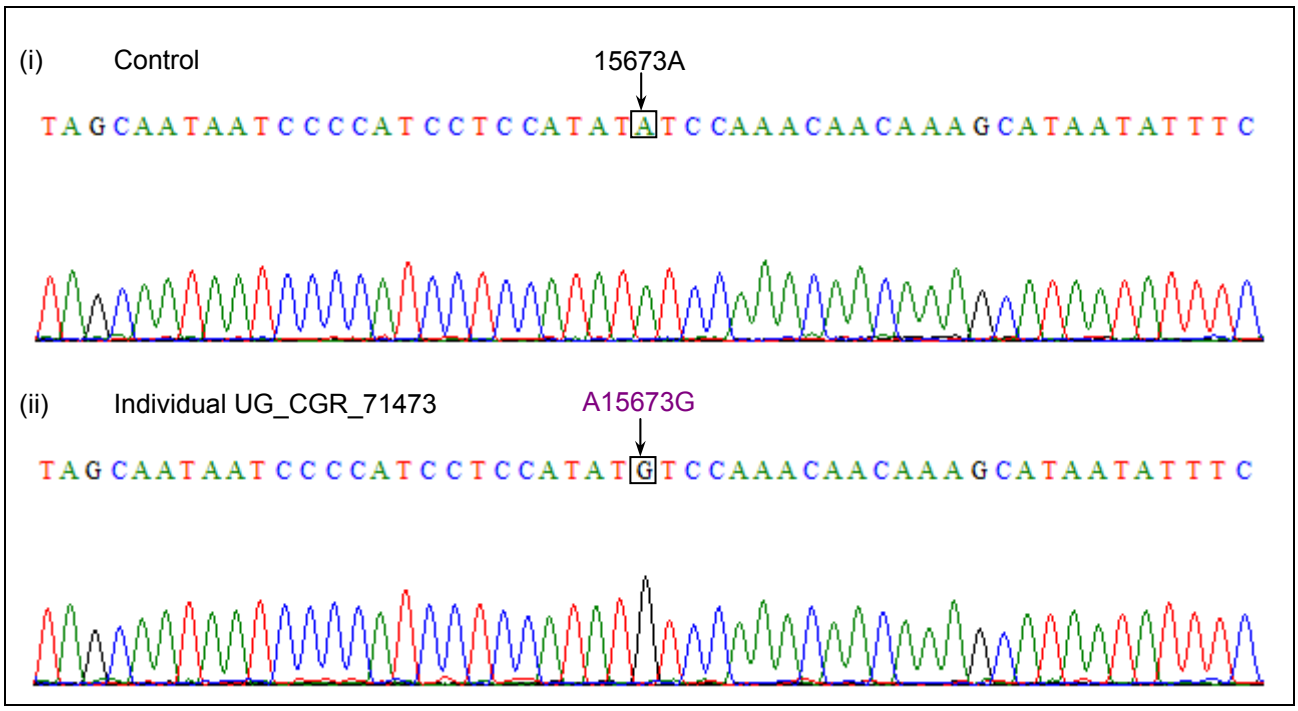
Figure 5.52 Representative electropherogram for the A15655G novel polymorphism in the Cytb gene



Control = represents electropherogram of an individual with a sequence identical to the rCRS at nucleotide position 15655. The nucleotide in violet text represents a novel alteration while the polymorphism in the control sample is indicated in black. Arrows represent the nucleotide (nt) positions at which the alterations occur. A = adenine; C = cytosine; G = guanine; T = thymine.

The A15655G polymorphism, as illustrated in Figure 5.52, was harboured by individual UG_CGR_71660 (a Muganda). The polymorphism occurred in the cytb gene (Mitomap, 2008c) causing no change in the amino acid encoded (Mitomap, 2008c) and therefore is unlikely to have a functional effect on the mitochondria. The phylogenetic significance of the polymorphism needs further assessment using larger samples of Baganda individuals or individuals in the same cluster in Figure 5.59 and Figure 5.60 as the Ganda individual UG_CGR_71660.

Figure 5.53 Representative electropherogram for the A15673G novel polymorphism in the Cytb gene



Control = represents electropherogram of an individual with a sequence identical to the rCRS at nucleotide position 15673. The nucleotide in violet text represents a novel alteration while the polymorphism in the control sample is indicated in black. Arrows represent the nucleotide (nt) positions at which the alterations occur. A = adenine; C = cytosine; G = guanine; T = thymine.

The A15673G polymorphism as indicated in Figure 5.53 was harboured by Acholi individual UG_CGR_71473. The polymorphism was synonymous (Mitomap, 2008c) and therefore unlikely to have an effect on the way the mitochondria functions. The relevance of this polymorphism in the phylogenetic context requires additional investigation through the sequencing of more Acholi samples.

5.2.11 Summary of all the alterations observed in this study

A total of 563 polymorphisms were identified in the entire mitochondrial genomes of the 40 individuals investigated in this study at 553 sites compared to the rCRS. Two hundred and seventy six (276) of these polymorphisms were synonymous, 75 were nonsynonymous, 208 in the non-coding regions and 26 were novel. Of the novel polymorphisms, seven (7) occurred in the non-coding regions, 15 were synonymous and four (4) were nonsynonymous. Of the novel polymorphisms, 15 were harboured by Acholi individuals, seven (7) by Lugbara individuals and four (4) by Baganda individuals.

The comparatively fewer novel polymorphisms amongst the Baganda in relation to the Acholi and Lugbara could be as a result of the greater amount of genome sequencing that has been performed on the African Bantu population to which the Baganda belong (Salas *et al*., 2002; Salas *et al*., 2003; Behar *et al*., 2008). However, there is a need to sequence more samples of the Baganda before firm conclusions can be drawn. The high prevalence of novel mutations amongst the Acholi could imply that such polymorphisms could have a phylogenetic significance and therefore more samples need to be sequenced to verify this. However, the Acholi as a tribe as well as the Nilotic ethnic group has been sequenced less than the Bantu. The sequencing of mtDNA involving people geographically and linguistically close to the Acholi and Lugbara in this study has been previously performed on a smaller sample set from the Nilotes of Sudanese descent (Krings *et al*., 1999; Kivisild *et al*., 2006).

Only 13 of the 40 individuals sequenced in this study harboured novel polymorphisms, of which five (5) were harboured by individual UG_CGR_71470. Individuals UG_CGR_71470 and UG_CGR_71478 shared three (3) novel polymorphisms and the two samples cluster together as represented in Figure 5.59 in Section 5.3.2 and Figure 5.60 in Section 5.3.3. The high proportion of the shared novel polymorphisms could signify a recent common origin for samples UG_CGR_71470 and UG_CGR_71478.

5.3 PHYLOGENETIC ANALYSIS OF THE BAGANDA, ACHOLI AND LUGBARA MITOCHONDRIAL DNA SEQUENCES

In order to compare the genetic characteristics of the 13 Baganda, 13 Acholi and 14 Lugbara individuals from Uganda, nucleotide diversities and polymorphisms determined between and within these individuals were assessed respectively. The NJ tree[1] (Saitou and Nei, 1987 as indicated in Figure 5.59 and the MP tree[2] (Saitou and Imanish, 1989) as indicated in Figure 5.60 were used to further investigate the genetic relationship between the samples of the three tribes and assess the phylogenetically important polymorphisms exhibited by the samples. The frequency distribution of the samples, if representative of their tribes, was used to estimate the extent to which the Baganda, Acholi and Lugbara underwent population expansion or contraction over the evolutionary times.

5.3.1 Comparison amongst the Baganda, Acholi and Lugbara mtDNA sequences

The Baganda sequences harboured a total of 846 polymorphisms at 249 sites when compared to the rCRS (Andrews *et al.*, 1999) but at the individual level, as indicated in Table A1 in Appendix A, the number of polymorphic sites per individual sequence ranged from 40 to 98. The Lugbara samples were observed to possess 832 polymorphisms at 318 distinct sites. The number of polymorphic sites harboured by individuals compared to the rCRS varied from 37 to 85 (see Table A1 in Appendix A). The Acholi samples possessed 738 polymorphisms occurring at 315 sites. As indicated in Table A1 in Appendix A, the individual with the least variation harboured 39 polymorphisms while the individual with the highest level had 89 polymorphisms. On average, the Baganda samples in this investigation had a higher ratio of number of polymorphisms to polymorphic sites (3.4) than the Lugbara (2.6) and the Acholi (2.3). The greater ratio for the Baganda was attributed to their having a high proportion of L0 haplogroup mtDNAs, which is an ancient haplogroup (Gonda *et al.*, 2007), as compared to the Lugbara and Acholi. Furthermore, the Baganda come from a diversified origin (Were and Wilson, 1984; Ssekamwa, 1994).

A number of polymorphisms were possessed by study participants belonging to single tribes, two tribes or to all individuals in the three tribes. These groupings excluded study

[1] The NJ tree was constructed by Dr G.W. Towers with the assistance of M. Koekemoer, from the Centre for Genome Research, North-West University (Potchefstroom Campus), South Africa and Dr J Poole from the Center for Molecular and Mitochondrial Medicine and Genetics, University of California, Irvine (UCI), USA.

[2] The MP tree was constructed by Dr G.W. Towers with the assistance of M. Koekemoer, from the Centre for Genome Research, North-West University (Potchefstroom Campus, South Africa) and and Dr J Poole from the Center for Molecular and Mitochondrial Medicine and Genetics, UCI, USA.

participant UG_CGR_71468 who by the nature of its exhibited polymorphisms was an outlier among the Acholi samples (see Table A1 in Appendix A). The greater number of unique polymorphisms among the Baganda samples when compared to those of the Acholi and Lugbara signified that the Baganda samples are genetically more distant from the Acholi and Lugbara samples than the Acholi individuals are from the Lugbara. The fewer number of unique polymorphisms amongst the Acholi and Lugbara samples was probably due to the greater genetic relatedness between the two tribes as compared to their relatedness to the Baganda samples. Furthermore, on the basis of the number of shared polymorphisms, the Baganda samples are genetically more related to the Acholi samples than they are to the Lugbara samples, in the sample set investigated.

All the Acholi (with the exception of individual UG_CGR_71468, who with respect to its observed polymorphisms is more closely related to the individuals from the Baganda samples) and Lugbara individuals harbour the A73G polymorphism, but the polymorphism is absent in six of the 13 Baganda samples. As discussed in Section 5.2.1 and Section 5.3.3, such a polymorphism may be of phylogenetic importance in differentiating a certain section of the Baganda from the Lugbara and Acholi as it divided the Baganda samples into two distinct genetic groups. In the sample set investigated, as observed in Table A1 in Appendix A, 55 polymorphisms were restricted to the Baganda samples, 15 to the Acholi samples and 10 to the Lugbara samples. The C182T polymorphisms occurred among five Acholi samples and six Lugbara samples while the G8206A and T10115C polymorphisms were shared amongst four Acholi and six Lugbara samples. Similar associations between polymorphisms and sections of ethnic groups have been observed in other populations (Brown, 1980; Denaro *et al.*, 1981; Bolnick *et al.*, 2004) while other studies (Barbujani, 1997; Edwin *et al.*, 2002) have established no link between specific polymorphisms and ethnic groups. The lack of tribal-specific or ethnic-specific polymorphisms may be attributed to inter-ethnic marriages or admixture (Kalaydjieva *et al.*, 2001) since the offspring adopt the ethnicity of the father according to cultural practice (Jobling, 2001). It may not be possible to identify if inter-ethnic marriages or admixture occurred many generations ago due to inevitable loss of information over extended periods of time especially where written historical records are lacking. Nevertheless, the undesirable effect of admixture was minimised in this study by having all the study participants fill in family pedigree forms up to two generations back (three generations in total, including that of the study participants) and ensuring that the maternal ancestors belonged to the same ethnic

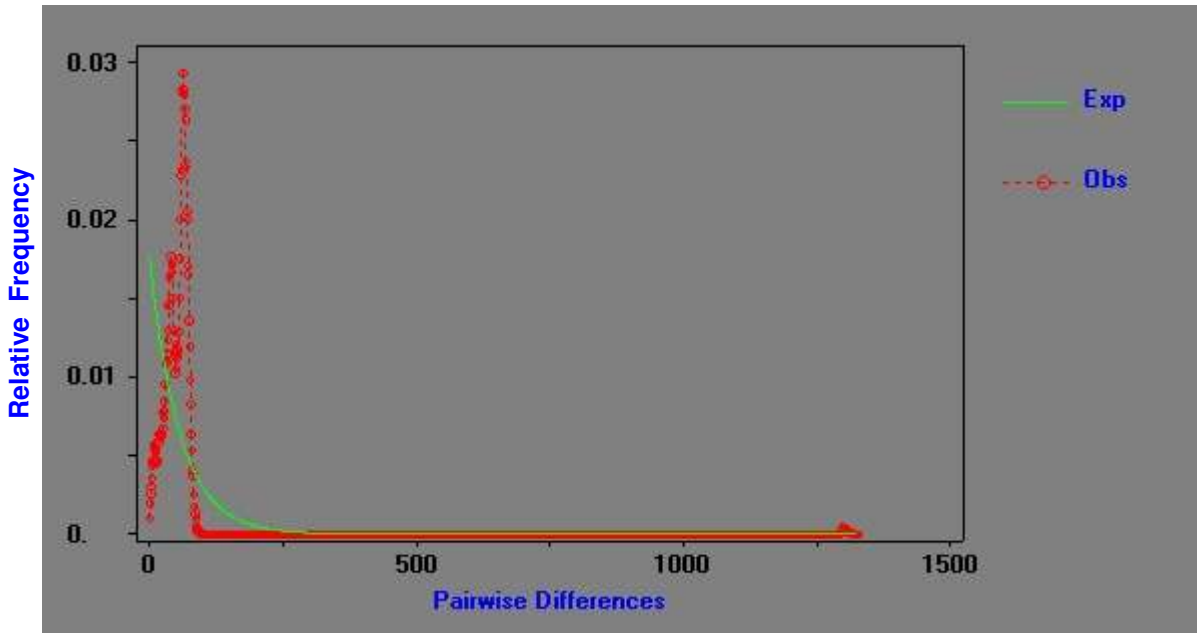
group. However, determination of the extent and effect of admixture was not a subject of interest in this investigation.

The mtDNA sequences obtained in this study, if representative of the larger tribes, could be used as controls when investigating potentially pathogenic mutations among suspected Baganda, Acholi and Lugbara patients. It is common practice to compare the mtDNA sequences of a diseased population with that of controls from the same ethnic group or haplogroup when screening for mutations of potential pathogenicity amongst populations (Torroni *et al*., 1996). By matching the control and diseased populations' ethnic groups, it makes it easy to differentiate mutations that are haplogroup-specific from those that could be pathogenic (Torroni *et al*., 1996). The use of these samples as controls will be rendered possible when large numbers of samples from these tribes are sequenced and analysed.

Figure 5.54 indicates the frequency distribution of the pairwise differences[1] between all the African haplogroup L samples used in this investigation. The frequency distribution of the pairwise differences is ragged and broad (The graph is broad and ragged except that a very wide scale has been used on the horizontal axis making the graph to occupy a very narrow part of the scale) indicating that the samples came from populations that have not underwent contractions and expansions for a long time which is a characteristic of African populations (Rogers and Harpending, 1992; Mishmar *et al*., 2003).

[1] The frequency distribution of the pairwise differences were generated by Dr G.W. Towers, Centre for Genome Research, North-West University (Potchefstroom Campus), South Africa.

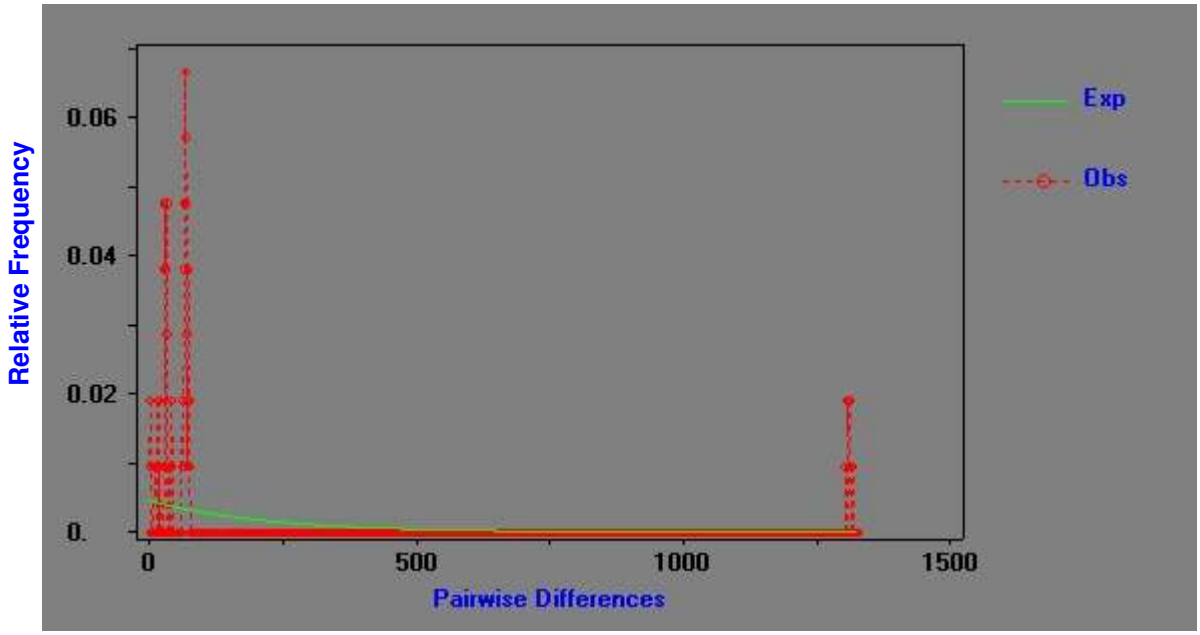
Figure 5.54 Frequency distribution of pair-wise differences within all the African haplogroup L samples



Exp = expected pairwise difference (indicated in green), Obs = observed pairwise difference (indicated in red). Number of sequences used = 428, average number of pairwise differences, k = 55.907. These pairwise differences were generated by Dr G.W. Towers, Centre for Genome Research, North-West University (Potchefstroom Campus), South Africa.

Figure 5.55 indicates the pairwise difference distribution of Ganda samples obtained from Uganda. The observed distribution of the pairwise differences is broad and ragged indicating that the samples came from populations at equilibrium (Rogers and Harpending, 1992; Mishmar *et al*., 2003). Populations in disequilibrium are characterised by bell-shaped pairwise difference distribution curves (Rogers and Harpending, 1992; Mishmar *et al*., 2003). As outlined in the paragraph above, such graphs are characteristic of African populations (Rogers and Harpending, 1992; Mishmar *et al*., 2003). Therefore, if the Ganda samples in this investigation are representative of the Ganda tribe, then the Baganda population has been a constant size for a very long period of time. The smaller peak on the right creating a bimodal pairwise distribution in Figure 5.55 is indicative of a mixed origin for the peopling of Buganda (Were and Wilson, 1984; Ssekamwa, 1993; Rogers and Jorde, 1995).

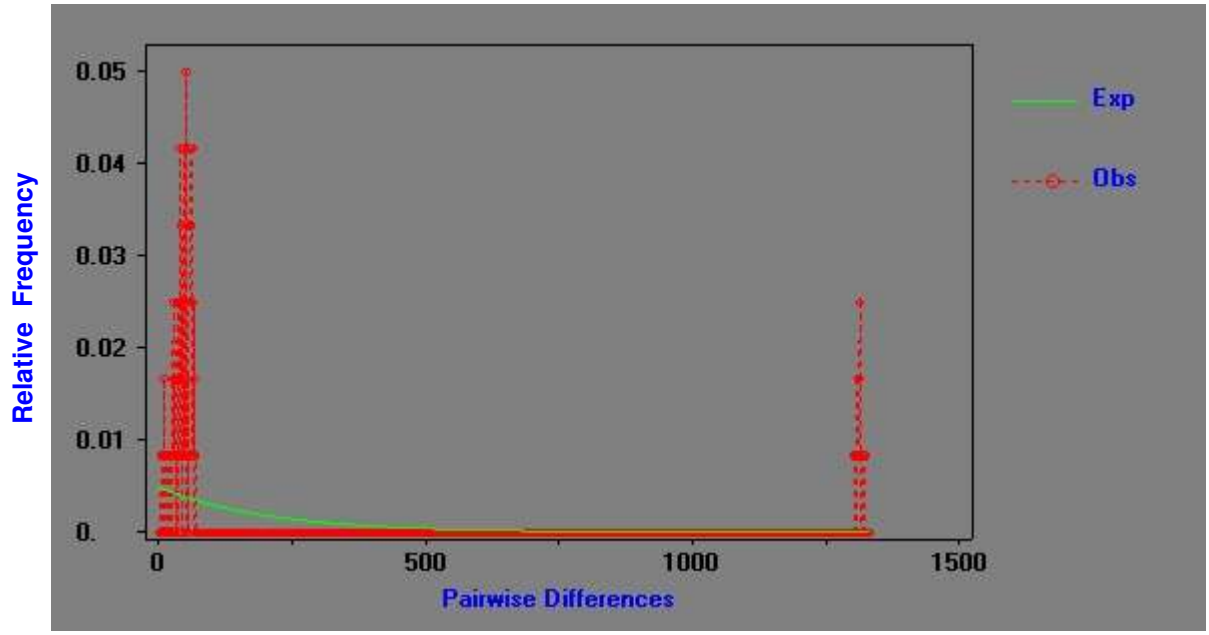
Figure 5.55 Frequency distribution of pair-wise differences within the Baganda from Uganda



Exp = expected pairwise difference (indicated in green), Obs = observed pairwise difference (indicated in red). Number of sequences used = 15, average number of pairwise differences, k = 216.076. These pairwise differences were generated by Dr G.W. Towers, Centre for Genome Research, North-West University (Potchefstroom Campus), South Africa.

The frequency distribution of the pairwise differences amongst the Lugbara samples as indicated in Figure 5.56 was broad and ragged indicating that if these samples were representative of the Lugbara population, such a population would have been at equilibrium for a long time (Rogers and Harpending, 1992; Mishmar *et al*., 2003). The Lugbara, like other African tribes have not experienced population bottlenecks for quite some time (Rogers and Harpending, 1992; Mishmar *et al*., 2003). The non-smoothness of the graph illustrates that the Lugbara samples if representative of the tribe, never underwent an ancient expansion (Rogers and Jorde, 1995). The peak at the extreme right of the graph indicates that the Lugbara population split up after the initial expansion into several divisions or otherwise it expanded prior to the split (Rodgers and Jorde, 1995). Historical records support the scenario of the Lugbara splitting up prior to their coming into Uganda (Ssekamwa, 1993).

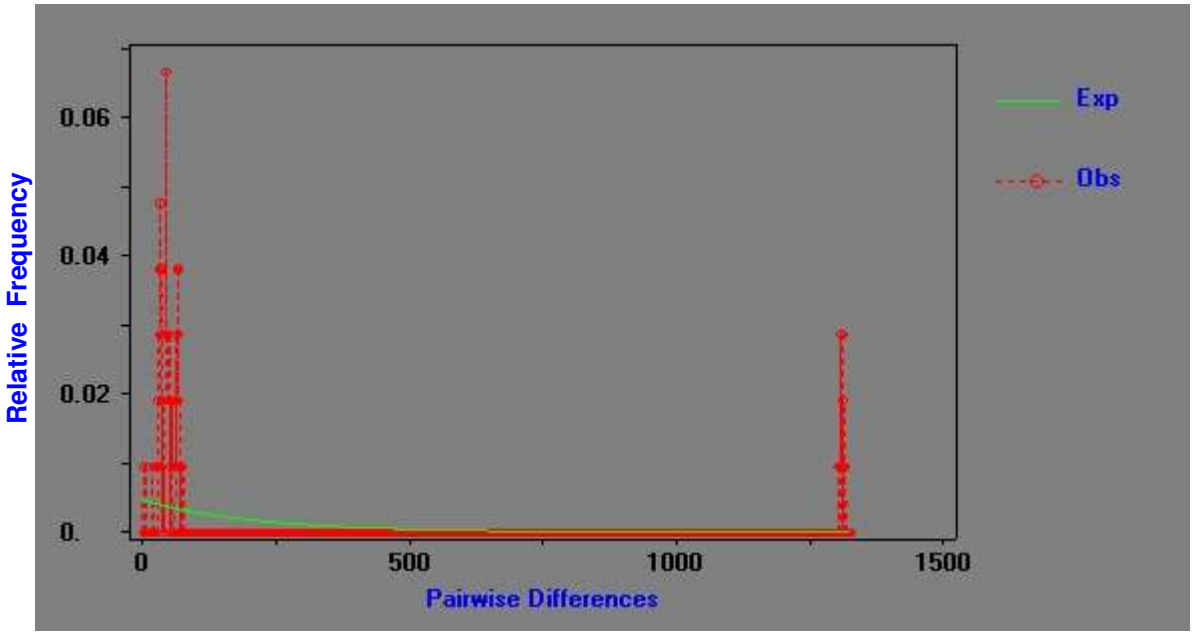
Figure 5.56 Frequency distribution of pair-wise differences within the Lugbara from Uganda



Exp = expected pairwise difference (indicated in green), Obs = observed pairwise difference (indicated in red). Number of sequences used = 16, average number of pairwise differences, k = 203.292. These pairwise differences were generated by Dr G.W. Towers, Centre for Genome Research, North-West University (Potchefstroom Campus), South Africa.

The distribution of pairwise differences for the Acholi samples as indicated in Figure 5.57 is also broad and ragged indicating that these samples belong to populations that have not experienced bottlenecks for a very long time (Rogers and Harpending, 1992; Mishmar *et al.*, 2003) as other African populations have been (Mishmar *et al.*, 2003). The pairwise distribution also reflects the fact that the duration between coalescent events was wide in the gene genealogy of the tribe (Harpending and Rogers, 2000). The exhibition of the smaller peak at the right of the mismatch distributions is in agreement with historical records that the Acholi underwent population splits after reaching Uganda from Sudan (Ssekamwa, 1993; Rogers and Jorde, 1995).

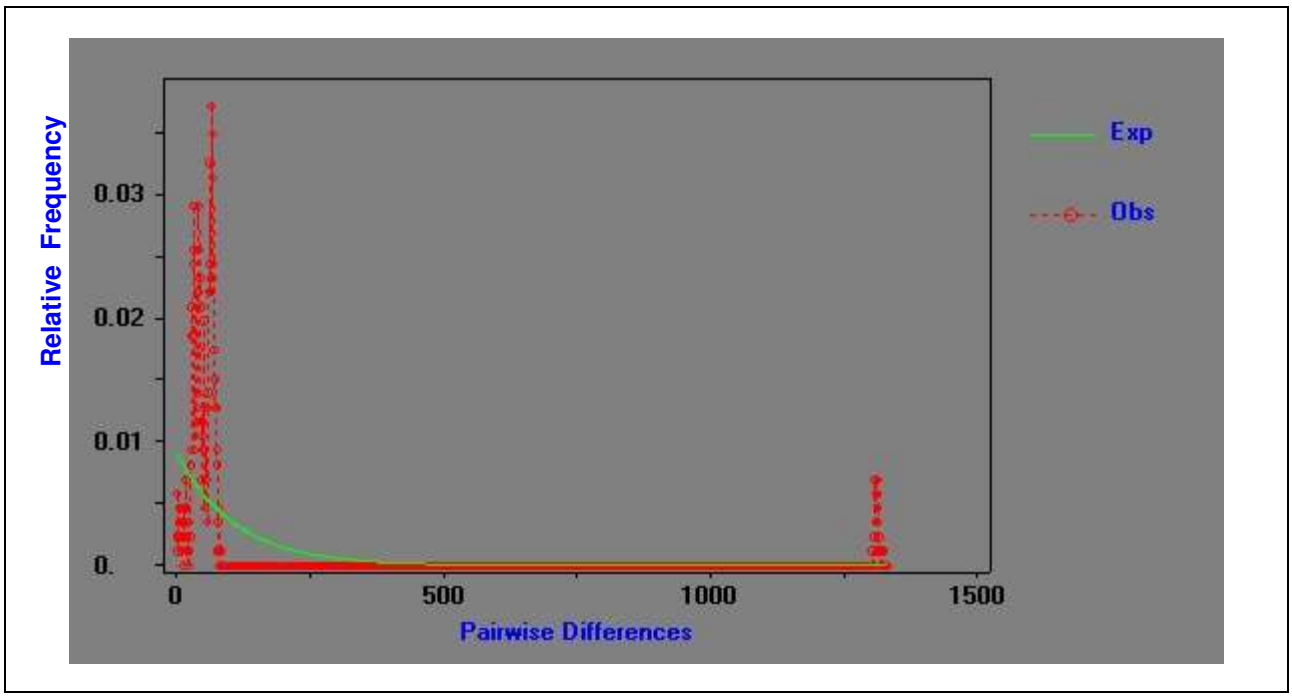
Figure 5.57 Frequency distribution of pair-wise differences within the Acholi samples from Uganda



Exp = expected pairwise difference (indicated in green), Obs = observed pairwise difference (indicated in red). Number of sequences used = 15, average number of pairwise differences, k = 215.095. These pairwise differences were generated by Dr G.W. Towers, Centre for Genome Research, North-West University (Potchefstroom Campus), South Africa.

Figure 5.58 illustrates how the pair-wise differences are distributed within the combined Baganda, Acholi and Lugbara samples in this investigation. The frequency distribution of the pair-wise differences within the samples of the three tribes is broad and ragged and in agreement with that of the individual tribal samples indicating that the Baganda, Acholi and Lugbara samples may have been at equilibrium (constant in size) for a very long time (Rogers and Harpending, 1992; Mishmar *et al*., 2003). The phenomenon of constant population size has been a characteristic of many African populations (Mishmar *et al*., 2003) while Asian, European and Native American populations are characterised by bell-shaped pair-wise frequency distribution curves signifying recent population expansion (Mishmar *et al*., 2003). However, larger samples need to be analysed before firm deductions can be drawn.

Figure 5.58 Frequency distribution of pair-wise differences within the Baganda, Acholi and Lugbara from Uganda



Exp = expected pairwise difference (indicated in green), Obs = observed pairwise difference (indicated in red). Number of sequences used = 42, average number of pairwise differences, k = 108.961. These pairwise differences were generated by Dr G.W. Towers, Centre for Genome Research, North-West University (Potchefstroom Campus), South Africa.

The nucleotide diversity (Pi) of the samples from the three tribes (Baganda samples, Pi = 0.01400; Acholi samples, Pi = 0.01393; Lugbara, Pi = 0.01317) was about the same indicating that if these samples were representative of the populations they were drawn from, such populations had nearly the same population size for a very long period of time if their mutation rate was taken as being very low (Harpending and Rogers, 2000). The nucleotide diversity for all 40 Ugandan samples i.e. 0.00706, is greater than that of the African macrohaplogroup L population, 0.00363, implying that the Ugandan population has had a longer genetic history that the general African population (Horai *et al.*, 1993; Harpending and Rogers, 2000). This is in line with the finding that East Africa is the focal origin of anatomically modern man, from where he spread to the rest of Africa and the rest of the world (Were and Wilson, 1984; Cruciani *et al.*, 2002; Salas *et al.*, 2002; Gonder *et al.*, 2007).

Of the 40 mtDNA sequences of the individuals investigated in this study, 39 different mtDNA haplotypes were revealed which constituted a mtDNA haplotype diversity of 97.5%. The percentage of different haplotypes obtained from the sequences of each of the three tribes ranged from 92.3% (Baganda samples) to 100% (Acholi and Lugbara samples). If these samples were large enough to be representative of the tribes from which they were drawn, such a high degree of haplotype diversity could be used in

forensics to improve on the discrimination between individuals more especially if the forensic investigation was to be performed on degraded samples where nDNA is scarce (Melton and Nelson, 2001) or in situations where nDNA is absent in the evidential material such as in shed hairs (Coble *et al*., 2004). However, no further assessment of the forensic potential of these samples was performed as it was not a major focus of this study.

5.3.2 The NJ tree of the Baganda, Acholi and Lugbara samples from Uganda

The phylogenetic relationship between and within the Baganda, Acholi and Lugbara samples was also assessed by the use of a NJ tree[1] as illustrated in Figure 5.59. The NJ method is a distance method that constructs a tree by the use of pair-wise distances obtained from sequences (Saitou and Imanish, 1989; Kumar *et al*., 2004; Gascuel and Steel, 2006). The information stored in the sequences is lost when distance-based methods are used to construct trees (Saitou and Nei, 1987). Alternatively, the MP tree is able to use information derived from insertions and deletions (Kumar *et al*., 1993; Kawakita *et al*., 2003). However, the MP method has poor consistency in estimating tree topology when there is a big difference in the rate of evolution across lineages (Kumar *et al*., 1993; Tateno *et al*., 1994) and the method is greatly inconsistent for trees with short branches in the interior (Kumar *et al*., 1993). Therefore construction of the tree was performed using the NJ method so as to maximise the generation of a tree of correct topology without the assumption that the rate of evolution was uniform (Kumar *et al*., 1993; Tateno *et al*., 1994; Gascuel and Steel, 2006) while the MP tree facilitated comparison so that information derived from insertions, deletions and actual changes in nucleotide sequences can be tracked (Rzhetsky and Nei, 1992; Kumar *et al*., 1993; Kawakita *et al*., 2003).

The individuals clustered within the NJ tree thus giving rise to distinctive patterns. Six of the 13 Baganda individuals clustered in the haplogroup L0 cluster. This cluster included one Acholi (UG_CGR_71468) individual who as discussed in Section 5.3.1 was more related to the Baganda samples in this investigation than the Acholi individuals with respect to its exhibited mtDNA polymorphisms. The clustering of this individual with the six Baganda is therefore in agreement with the manual assessment of the sequences as outlined in Section 5.3.1. The most likely explanation for the clustering of the one Acholi individual with the six Baganda is a near recent common ancestry for the six Baganda with

[1] The NJ tree was constructed by Dr G.W. Towers with the assistance of M. Koekemoer, from the Centre for Genome Research, North-West University (Potchefstroom Campus, South Africa) and Dr J Poole from the Center for Molecular and Mitochondrial Medicine and Genetics, University of California, Irvine (UCI), USA.

the Acholi individual UG_CGR_71468 in haplogroup L0. The clustering in haplogroup L1 was exclusive to individuals from the Lugbara tribe included in this investigation.

Six Lugbara and four Acholi individuals clustered in haplogroup L2 while no Muganda [singular for Baganda] belonged to this category. Two of the 13 Acholi individuals and three of the 13 Baganda grouped together in a cluster of haplogroup L4, demonstrating a closer genetic relationship between these individuals. The haplogroup L3 cluster encompassed samples from all three tribes. In one cluster, one Acholi individual (individual UG_CGR_71474) grouped together with one Lugbara individual (individual UG_CGR_71486). In the L3e cluster, one Acholi [UG_CGR_1739] did not group with any Ugandan samples while two Baganda samples (UG_CGR_1495 and UG_CGR_1656) clustered in haplogroup L3i. The haplogroup L3b cluster included two Lugbara, one Acholi and two Baganda individuals at the peripheries of the tree branches signifying that a sub-section of all three tribes under investigation could have had a recent common ancestor or else have undergone admixture.

The closer genetic relatedness of the Acholi and Lugbara samples in this investigation could have been as a result of the two tribes having their origin in Sudan (Ssekamwa, 1994) where admixture could have taken place. Alternatively, admixture between the Acholi and Lugbara could still have taken place after the two tribes had crossed into Uganda, a scenario which is favoured by their geographical proximity and closely related languages. The closer genetic relationship as observed in this sample set of the Acholi individual with Baganda and Lugbara individuals in haplogroup L3b could be attributed to admixture with the Luo-Babito of Bunyoro (Were and Wilson, 1984; Ssekamwa, 1994), the break away faction of the Luo from the Acholi, who migrated to Bunyoro and intermarried with the Banyoro (Were and Wilson, 1984; Ssekamwa, 1994) and the Baganda (Were and Wilson, 1984). The individuals in haplogroup L3b differed from each other only at one or two nucleotide positions of the entire mitochondrial genome, further illustrating how closely related they were. Samples from the Luo-Babito of Bunyoro need to be sequenced together with more samples from the Lugbara, Acholi and Baganda to further explore this assertion. One Lugbara individual clustered in haplogroup L3h while another Lugbara individual (UG_CGR_71674) clustered with the Acholi individual UG_CGR_71473, still in haplogroup L3h. The Lugbara individual UG_CGR_71486 and Acholi individual UG_CGR_71474 in haplogroup L4a clustered closest to the rCRS of mtDNA of European haplogroup H (Andrews *et al.*, 1999) implying that they are the most closely related

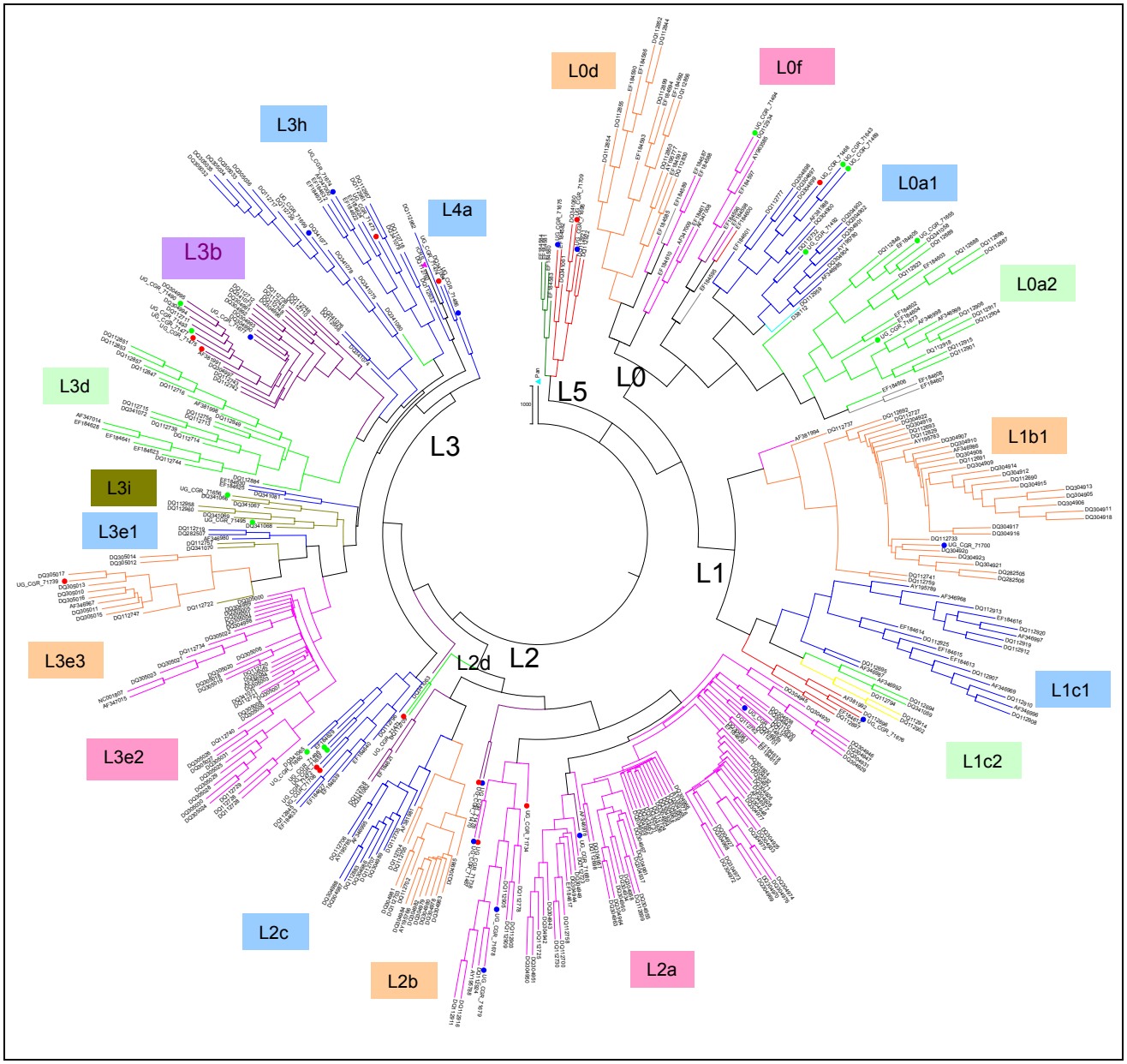
sequences to that of the rCRS. The clustering pattern also involved two Lugbara individuals and one Acholi individual in the L5 cluster, a deeper rooted branch than L0 of the phylogeny.

The NJ tree was divided into two main branches with the deepest branch involving the L5, L0 and L1 haplogroup branches and the other involving the L2, L3 and L4 haplogroup branches. The deepest rooted branch was L5 which gave rise to L0 and L1 haplogroup branches. This topology is in disagreement with other previously reported tree topologies where L0 is the deepest branch (Kivisild *et al.*, 2004; Gonder *et al.*, 2007; Kivisild *et al.*, 2006; Torroni *et al.*, 2006; Behar *et al.*, 2008). This branching pattern illustrates that L5 is the oldest lineage for anatomically modern human beings and the issue of the Khoi-San being the oldest human lineage stands to be challenged. However, it is also worthy noting that a good part of the Sandawe population of Tanzanian click language speakers is characterised by haplogroup L5 sequences (Gonder *et al.*, 2007). The additional haplogroup L5 sequences being added to the previously existing few before this investigation could have resulted in this finding i.e. that L5 is an older lineage than L0. More sequences especially those predicted in previous findings to be at the base of the trees such as L0k and L0d will have to be analysed alongside more sequences of L5 (especially those from Uganda, Sudan, Tanzania and Ethiopia) and analysed to verify this assertion using mtDNA and Y-chromosome. The L2 branch gave rise to L3 and L4 haplogroups in agreement with other tree topologies (Kivisild *et al.*, 2004; Kivisild *et al.*, 2006) while L3 which gave rise to the rest of the non-African mtDNAs (Mishmar *et al.*, 2003; Kivisild *et al.*, 2006; Torroni *et al.*, 2006) gave rise to the rCRS (Andrews *et al.*, 1999).

Overall, the 40 haplotypes that were generated in this study, as indicated in Figure 5.59 and Figure 5.60, improved on the resolution of macrohaplogroup L previously provided by the global African sample of 387 sequences. The resolution of macrohaplogroup L was greatly increased by the sequences included in this investigation that clustered in regions of small

African haplogroups such as L3h, L3i, L4 and L5. Samples UG_CGR_71678 (haplogroup L2a2) and UG_CGR_71471 (haplogroup L2d) also provided an improved resolution of macrohaplogroup L since they clustered in regions which have fewer existing complete sequences in the global African pool as indicated in Figure 5.59 and Figure 5.60.

Figure 5.59 The NJ tree of the Baganda, Acholi and Lugbara from Uganda



Blue circles = Lugbara, red circles = Acholi, green circles = Baganda, light blue triangle = *Pan troglodytes*, red triangle = Revised Cambridge Reference Sequence, rCRS.

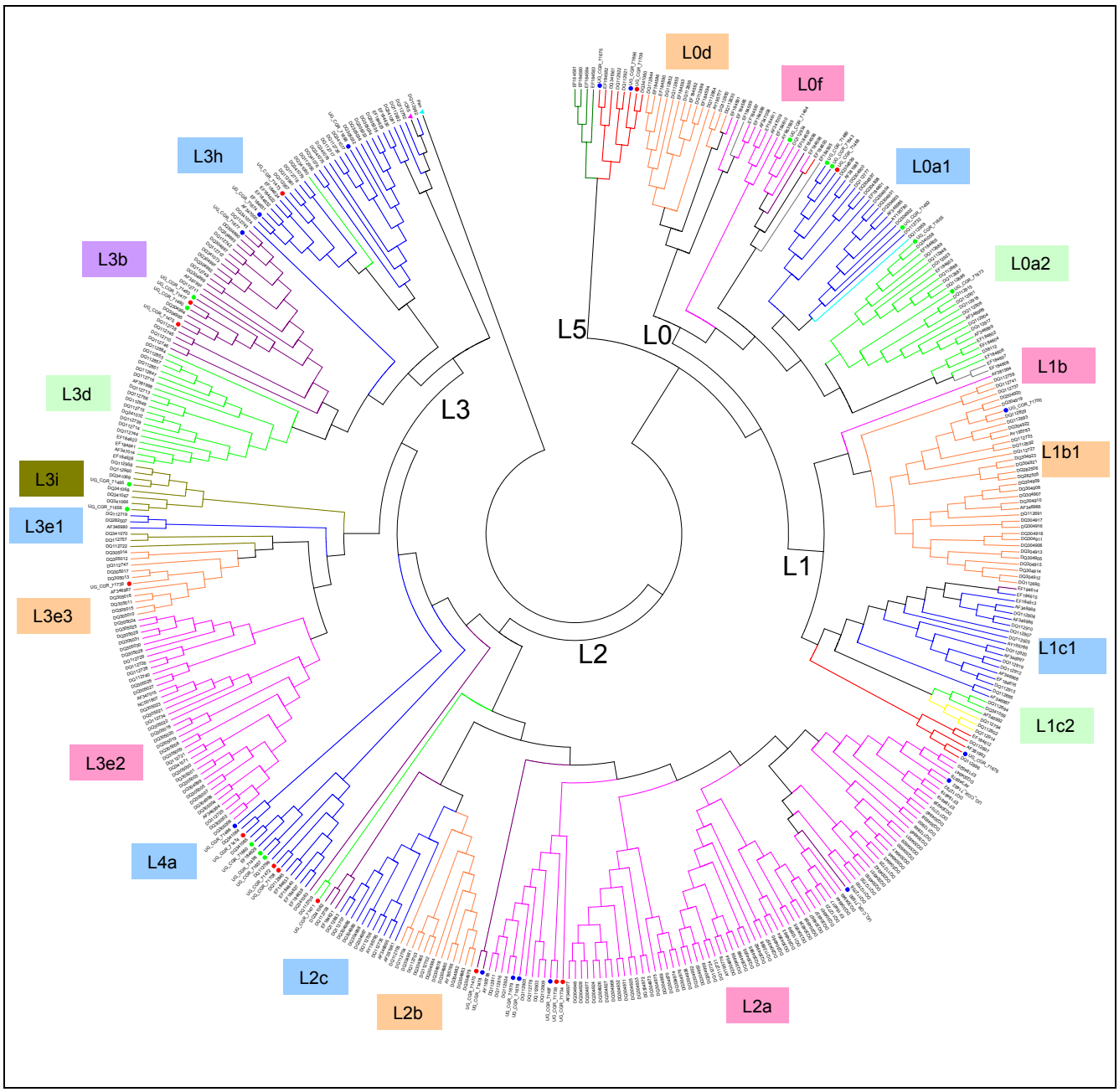
Novel branches in the tree were formed by samples that clustered in haplogroup L0a1 (samples UG_CGR_71468, UG_CGR_71643 and UG_CGR_71489), L2a (samples UG_CGR_71487 and UG_CGR_71738), haplogroup L3b (samples UG_CGR_71472 and UG_CGR_71708), and haplogroup L4 (UG_CGR_71496 and UG_CGR_71657). Furthermore, samples UG_CGR_71470 (Acholi) and UG_CGR_71478 (Lugbara) formed a distinct cluster or node of their own that signified a new evolutionary lineage amongst macrohaplogroup L haplotypes. The grouping together of these two samples indicates a closer genetic relatedness between the Lugbara and Acholi that could possibly be as a

result of a recent common ancestry since both tribes have Southern Sudan as their geographical origin (Were and Wilson, 1984; Ssekamwa, 1994).

5.3.3 The MP[1] tree of the Baganda, Acholi and Lugbara from Uganda

Figure 5.60 is a MP tree constructed to represent the evolutionary relationship within and between the Baganda, Acholi and Lugbara from Uganda to other African populations. Overall, the clustering pattern and topology of the 40 Ugandan individuals within the MP tree is in agreement with that of the NJ tree discussed in Section 5.3.2 as determined upon visual comparison, and therefore only the phylogenetic implications of the clustering pattern and topology are discussed. Unlike the NJ tree that uses pairwise-distances in constructing a tree (Saitou and Imanish, 1989; Kumar *et al.*, 2004), the MP uses discrete character states to infer phylogenetic relationships between organisms (Rzhetsky and Nei, 1992). The one Ugandan individual who lacked the A750G (Acholi individual UG_CGR_71473) clustered closest to the rCRS (Andrews *et al.*, 1999) which is of European haplogroup H indicating that this was the most closely related sample to haplogroup H. The same individual was also the only one, who like the rCRS, harboured the C allele at nucleotide position 16223. Individual UG_CGR_71473 must be a maternal descendant of the group of people who moved out of East Africa to populate the rest of the world, specifically those that generated haplogroup H in Europe.

Figure 5.60 The MP tree of the Baganda, Acholi and Lugbara from Uganda



Blue circles = Lugbara, red circles = Acholi, green circles = Baganda, light blue triangle = *Pan troglodytes*, red triangle = Revised Cambridge Reference Sequence, rCRS.

Both the NJ (Saitou and Nei, 1987) and MP (Saitou and Imanish, 1989) trees were used to assess the phylogenetically informative mtDNA mutations amongst the three populations. By manually examining the mtDNA mutations in Table A1 in Appendix A, it was possible to place particular polymorphisms in specific branches of the tree so as to generate the given tree topologies. Figure 5.61 illustrates the most probable placement of the polymorphisms in the branches of the tree and is a measure of the genetic relatedness of the different haplotypes identified in this study since the polymorphisms at the nodes are shared by the

[1] The MP tree was constructed by Dr G.W. Towers with the assistance of M. Koekemoer, from the Centre for Genome Research, North-West University (Potchefstroom Campus, South Africa) and Dr J. Poole from the Center for Molecular and Mitochondrial Medicine and Genetics, University of California, Irvine (UCI), USA.

haplotype derivatives (Torroni *et al*., 2006). The phylogenetic significance of polymorphisms which are unique to individuals in this study remains to be ascertained in a study involving larger sample sizes. In studies involving bigger samples, such polymorphisms may give rise to additional nodes, if they turned out to have phylogenetic significance, or otherwise they may merely serve to differentiate between individuals as they define unique branches or haplotypes. The novel polymorphisms identified in this study largely belong to this category of individual-level polymorphisms and therefore more samples need to be analysed in these populations to assess their phylogenetic value as they may re-define existing branches into clusters. The only novel polymorphism that was harboured by more than one individual is the C3321T polymorphism that was harboured by two individuals (UG_CGR_1470 and UG_CGR_1478) belonging to haplogroup L2.

Figure 5.61 Phylogenetically informative polymorphisms among the Baganda, Acholi and Lugbara from Uganda

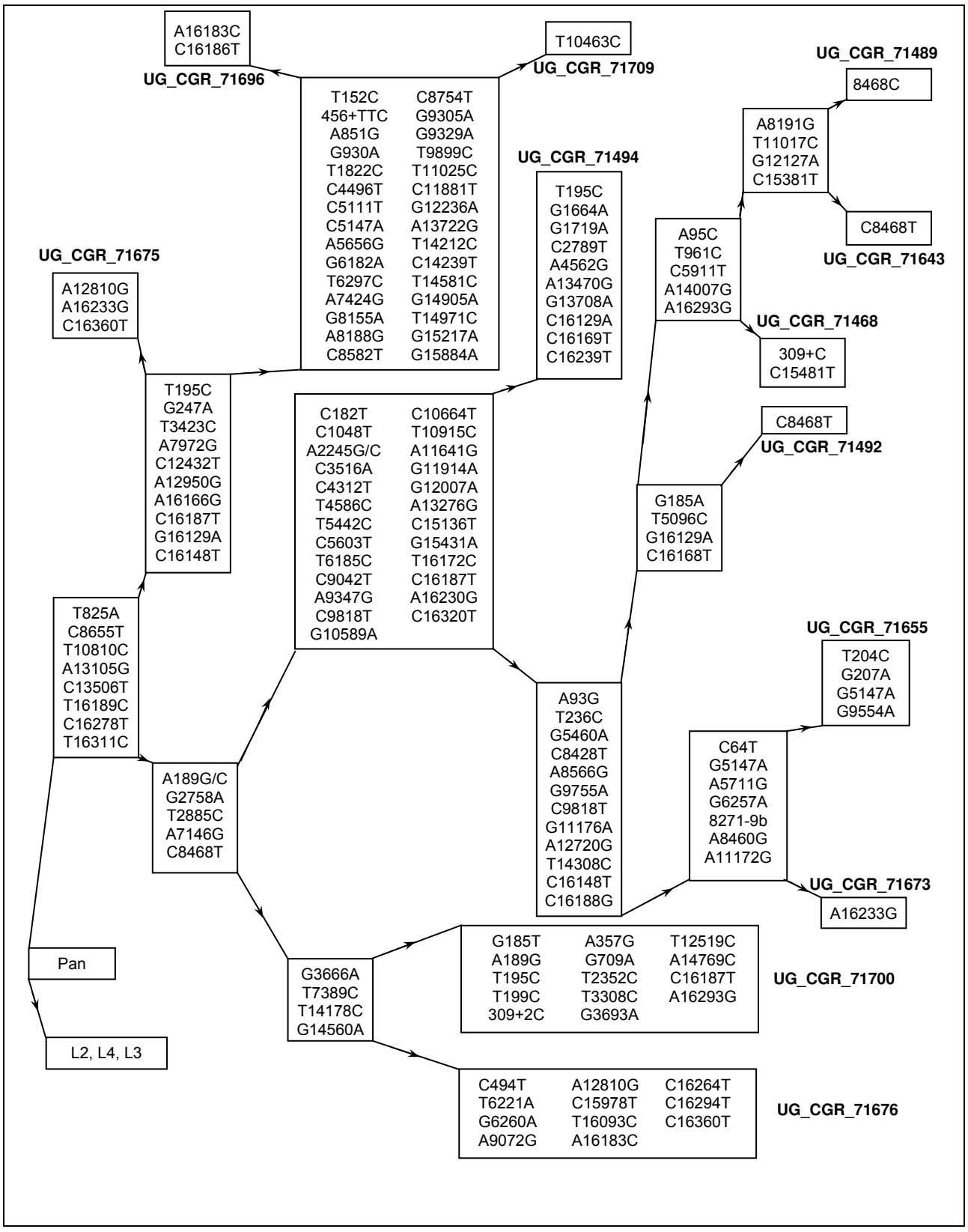


Figure 5.61 Continued …

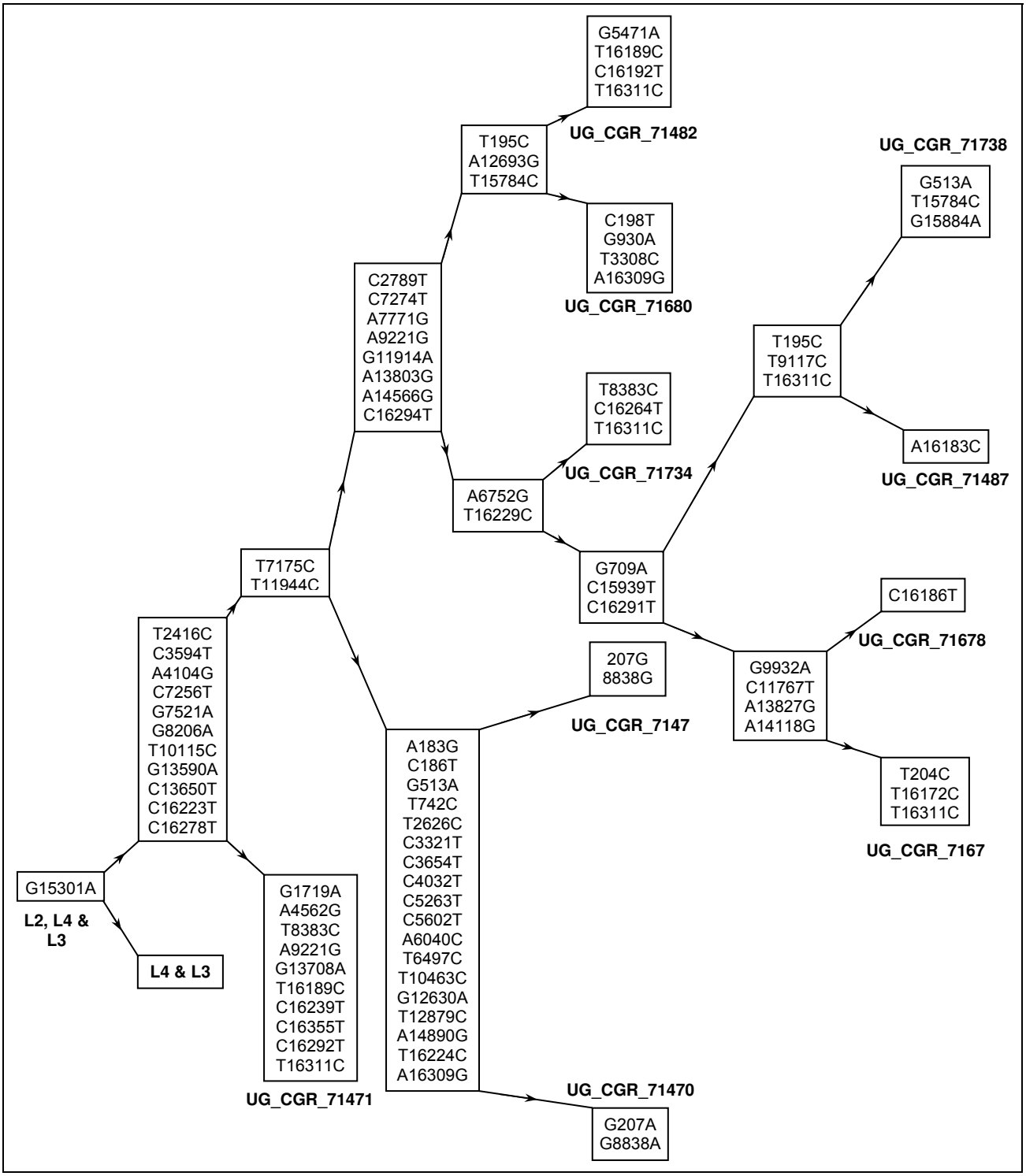
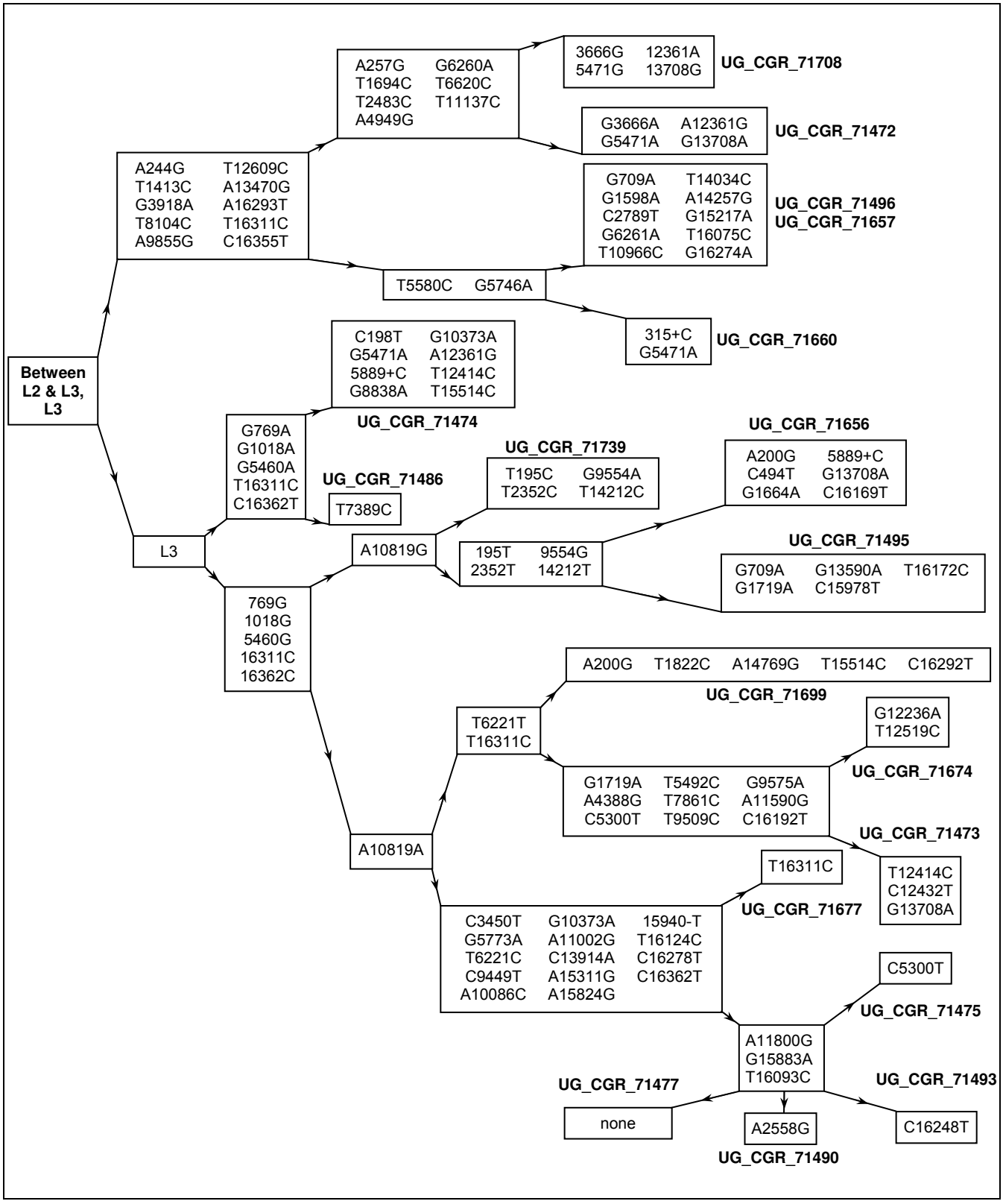


Figure 5.61 Continued …



A = adenine, C = cytosine, G = guanine, T = thymine, + = an insertion. The numbers in bold represent study participants. The phylogeny was reconstructed using both the coding and control region polymorphisms although the NJ and MP trees were constructed using sequences from only the coding regions. The polymorphisms indicated are with respect to the rCRS.

The C3594T, A4104G, C7256T, G7521A and C13650T polymorphisms allow for the division of the three tribes into two large groups with individuals harbouring the

polymorphisms clustering with either of the haplogroups L5, L0, L1 or L2. Except for individual UG_CGR_71472 (where the polymorphism is absent), the A244G polymorphism was present in all the individuals of the cluster between haplogroup L2 and L3. The absence of the A244G polymorphism in individual UG_CGR_71472 was attributed to the occurrence of back mutations, a phenomenon which is high in the control region (Aquadro and Greenberg, 1983). The G13590A polymorphism was only harboured by individuals who clustered in haplogroup L2a except for individual UG_CGR_71495 in haplogroup L3i. The phylogenetic significance of such a polymorphism in haplogroup L3i needs to be established in studies involving larger sample sizes to find out whether it occurred in the sample by chance or if it actually divides haplogroup L3i into sub-haplogroups. The G15301A polymorphism at one of the two innermost branches was harboured by individuals either in haplogroup L2, L3, or L4 as indicated in Figure 5.59 and Figure 5.60. All the individuals who lacked the A73G polymorphism clustered in haplogroup L0. As outlined in Section 5.2.1, this polymorphism served to divide the Baganda samples in this investigation into two major groups and was also lacking in the Acholi and Lugbara samples except for sample UG_CGR_71468 which was an outlier among the Acholi individuals. The absence of the A73G polymorphism should therefore be a characteristic of samples belonging to haplogroup L0 amongst macrohaplogroup L lineages.

5.3.4 Phylogeographic analysis of the Baganda, Acholi and Lugbara mtDNA sequences

A phylogeographic analysis of the Ugandan samples was performed by comparing their genetic relatedness to that of the global African pool of 387 sequences by the use of the NJ tree and the data compiled in Appendix B[1]. The Lugbara samples UG_CGR_71675, UG_CGR_71696 and the Acholi sample UG_CGR_71709 clustered with samples of haplogroup L5 from Tanzania (Gonder *et al.*, 2007), Pygmies from the Democratic Republic of Congo (Kivisild *et al.*, 2006) and Ethiopian samples (Torroni *et al.*, 2006). The clustering of the Acholi and Lugbara samples with the Ethiopian samples could arise as a result of the long-term historical interactions and geographical proximity of the ancestral origins of the Nilotes and related tribes (Were and Wilson, 1984; Ssekamwa, 1994). The clustering of Lugbara or Acholi individuals with the Tanzanian individuals is likely due to the presence of Nilotes who migrated into Tanzania from Sudan (Gonder *et al.*, 2007) which is the geographical area of origin of the Nilotes (Were and Wilson, 1984). The

[1] The information in Appendix B was kindly provided by M. Koekemoer, Centre for Genome Research, North-West University, Potchefstroom Campus, South Africa.

clustering of the Acholi sample and the Lugbara samples with the Democratic Republic of Congo samples could be attributed to the presence of the Moru-Madi in the Democratic Republic of Congo (Raymond, 2005) belonging to the same ethnic group as the Lugbara of Uganda.

The Ganda sample UG_CGR_71494 clustered with samples of haplogroup L0f from Tanzania (Gonder *et al*., 2007), a sample of haplogroup L0f from Ethiopia (Kivisild *et al*., 2006) and haplogroup L0f from a Ugandan sample of unspecified ethnicity (Macaulay *et al*., 2005). The Baganda samples (three) and Acholi sample (one) of haplogroup L0a1 clustered with L0a1 samples of African-Americans (Just *et al*., 2008) and one sample from Morocco of haplogroup L0a1 (Maca-Meyer *et al*., 2001). Sample UG_CGR_71492 clustered with an L0a1 sample from Burkina Faso (Kivisild *et al*., 2006). The L0a2 Ganda sample UG_CGR_71655 clustered with L0 samples i.e. two from the Dominican Republic (Kivisild *et al*., 2006, Torroni *et al*., 2006) and one from Tanzania (Gonder *et al*., 2007). The L0a2 Ganda sample UG_CGR_71673 formed an independent branch of its own indicating that it was not closely related to other samples that clustered closer to it. The clustering of the Ganda sample with the Tanzanian sample was attributed to the presence of the Bantu in Tanzania, some of whom migrated into Uganda as part of the western Bantu immigrants (Ssekamwa, 1994). The clustering of the Ganda and Acholi samples with African-American or Dominican-Republican samples was attributed to, albeit with difficulty, possible slave trade occurring between Uganda and America. The pedigree analysis performed on the individuals ruled out a possible clustering due to recent or near recent admixture of Ugandans with Americans brought about by increased mobility due to modern transport technology as the individuals included in the study came from pure lineages for at least three generations. Scanty historical records suggest that slave trade existed between East Africa and North and South America (Ssekamwa, 1994). It could be possible that Ugandans were shipped to other countries in Africa or Asia before their final destination to America but again no detailed historical records are available to support such a scenario (Ssekamwa, 1994). The lack of complete sequences in other African countries also hinders the proposal that a different African country was the source of the slaves but its gene pool was similar to that of the Ugandan individuals in this study. The clustering of the Ganda sample with samples from South Africa, Morocco, Hausa and Burkina Faso could be attributed to the Bantu migrations that were widespread across Africa (Were and Wilson, 1984; Ssekamwa, 1994; Cruciani *et al*., 2002) while the clustering of the Acholi and the Moroccan sample was attributed to admixture occurring

between the Baganda and Acholi who were likely to share a Luo-Babito heritage (Were and Wilson, 1984; Ssekamwa, 1994). In this case, the Acholi would have acquired the mtDNA of the Bantu (Ssekamwa, 1994).

The sample UG_CGR_71700 (Lugbara) grouped with the following L1b sequences: DQ304920, DQ304923 and DQ304921 who are all African-Americans (Just *et al*., 2008), DQ112733 from Burkina Faso (Kivisild *et al*., 2006), as well as DQ282506 and DQ282505 who are Hispanic (Just *et al*., 2008). The Lugbara sample UG_CGR_71676 had its closest match with one sample from the Dominican Republic (Kivisild *et al*., 2006) and another from Mauritania (Maca-Meyer *et al*., 2001). The matching of the Lugbara sequences with the Mossi sample from Burkina Faso could be attributed to the presence of the Zarma or Songhay in Burkina Faso, who belong to the same Nilo-Saharan language phyla as the Lugbara and who could have possibly had admixture with the Mossi (Raymond, 2005). The possibility of slave trade between Uganda and America as explained within the relationship between the Baganda and Acholi samples with the African-Americans in the second paragraph of Section 5.3.4 could have contributed to the matching of Lugbara sequences with sequences from the Dominican Republic and other Hispanic populations. The matching of the Lugbara sequences with those from Mauritania could be due to the high prevalence of L1b sequences from West Africa (Salas *et al*., 2002; Gonder *et al*., 2007).

The Lugbara and Acholi samples in haplogroup L2 clustered with haplogroup L2 samples from Tanzania (Gonder *et al*., 2007), Burkina Faso (Kivisild *et al*., 2006), African-Americans (Kivisild *et al*., 2006; Just *et al*., 2008), the Dominican Republic (Kivisild *et al*., 2006) and DQ112778 from Sudan (Kivisild *et al*., 2006). The UG_CGR_71478 (Lugbara) and UG_CGR_71470 (Acholi) samples formed a distinct branch in L2 indicating that they are quite a unique group of sequences. However, since they formed a distinct cluster of their own, more African samples need to be sequenced and analysed in order to define the origin of the two individuals in context. The Acholi individual UG_CGR_71471 clusters with sample DQ112709 only, which belongs to haplogroup L2 from the Dominican Republic (Kivisild *et al*., 2006). The two samples were rooted at the bottom of the haplogroup L2 cluster. The clustering of the Lugbara and Acholi samples with African-American samples or samples from the Dominican Republic was attributed to the possibility of slave trade occurring between Uganda and Dominican Republic as explained above in this section. The clustering of the Acholi and Lugbara samples with a sample

from Sudan served as further evidence to demonstrate that the Acholi and Lugbara originated from Sudan (Were and Wilson, 1984; Ssekamwa, 1993). The clustering of these samples with samples from Burkina Faso was attributed to possible admixture involving the Zarma or Songhay of the Nilo-Saharan language family (like the Lugbara and Acholi) with other ethnic groups in Burkina Faso (Raymond, 2005).

The Acholi Ugandan samples UG_CGR_71708 and UG_CGR_71472 cluster with Tanzanian haplogroup L4 samples EF184639, EF184627 and EF184640 (Gonder *et al*., 2007), EF184633 of haplogroup L4 from South Africa (Gonder *et al*., 2007), DQ112845 and DQ112796 of haplogroup L4 from Lissongo in the Central African Republic (Kivisild *et al*., 2006). The samples UG_CGR_71657 (Ganda), UG_CGR_71496 (Ganda) and UG_CGR_71660 (Ganda) clustered with samples EF184629 of haplogroup L4 from Tanzania (Gonder *et al*., 2007) and DQ341065 of haplogroup L4 from Ethiopia (Torroni *et al*., 2006). The matching of Acholi samples with Tanzanian samples was attributed to possible admixture involving other Tanzanian populations with Tanzanian Nilotes who migrated from Sudan which was their ancestral homeland (Were and Wilson, 1984; Ssekamwa, 1994; Gonder *et al*., 2007). The Acholi sample UG_CGR_71474 and the Lugbara sample UG_CGR_71486 form a distinct cluster with sample DQ341064 of haplogroup L7 from Ethiopia (Torroni *et al*., 2006), renamed haplogroup L4a by Behar *et al*., 2008. The addition of the two L4a sequences to the previous sequence resulted in the generation of two branches which improved on the resolution of macrohaplogroup L sequences. However, more samples belonging to haplogroup L4a need to be analysed if the haplogroup is to be finely resolved into sub-haplogroups.

The Acholi sample UG_CGR_71739 clusters with samples DQ305013, DQ305016, DQ305017, DQ305010, DQ305011 and DQ305015 which belong to African-American individuals of haplogroup L3e2 (Just *et al*., 2008) as well as AF346967 which is a sample from a Bamileke individual also belonging to haplogroup L3e2 (Ingman *et al*., 2000). This is the only Nilotic sample obtained in this study that clusters with haplogroup L3e, a haplogroup whose origin is postulated to have been from Southern Sudan (Bandelt *et al*., 2001). The presence of haplogroup L3e sequences among the Bamileke may be explained by the fact that it is widespread in Africa (Bandelt *et al*., 2001). The Ganda sample UG_CGR_71656 clustered with the Ethiopian samples DQ341066 and DQ341067 of haplogroup L3 and DQ341068 of haplogroup L3i, also from Ethiopia (Torroni *et al*., 2006). The Ganda sample UG_CGR_71495 also grouped with sample DQ341068 and

with samples DQ341069 of haplogroup L3i from Ethiopia (Torroni *et al*., 2006), as well as DQ112958 and DQ112960 of haplogroup L3i from Sudan (Kivisild *et al*., 2006). The clustering of the Ganda samples with Ethiopian samples could be a hint that Ethiopia may have played some role in the peopling of Buganda while the clustering with Sudanese sequences could be due to admixture involving the Baganda and the Luo-Babito of Bunyoro, an admixed population involving the Acholi and Banyoro which had historical interactions with the Kingdom of Buganda (Were and Wilson, 1984; Ssekamwa, 1994).

The haplogroup L3b samples UG_CGR_71475 (Acholi), UG_CGR_71490 (Ganda), UG_CGR_71477 (Acholi) and UG_CGR_71493 (Ganda) clustered with African-American haplogroup L3b samples DQ304995 and DQ304994 as well as haplogroup L3 samples DQ112711 from the Dominican Republic (Kivisild *et al*., 2006) and AF381991 from Mauritania (Maca-Meyer *et al*., 2001). The L3b Lugbara sample UG_CGR_71677 clustered with African-American haplogroup L3b samples DQ304990 and DQ304993 (Just *et al*., 2008). The clustering of these Ugandan samples with African-American and Dominican Republican samples has been attributed to possible slave trade as explained in the second paragraph of Section 5.3.4. The clustering of these samples with a sample from Mauritania was attributed to isolated cases of recent admixture involving female Mauritanians and Ugandan males aided probably due to modern transport technology.

The sample UG_CGR_71674 clustered with sample AF347000 of haplogroup L3h from the Mkamba (Ingman *et al*., 2000), EF184631 (haplogroup L3), EF184632 (haplogroup L3), EF184622 (haplogroup L3h) and EF184624 (haplogroup L3f) from Tanzania (Gonder *et al*., 2007). The Acholi sample, UG_CGR_71473 clusters with samples DQ112957 and DQ112961 of haplogroup L3h from Sudan (Kivisild *et al*., 2006), DQ341079 and DQ341080 of haplogroup L3h from Ethiopia (Torroni *et al*., 2006) and DQ112718 of haplogroup L3h from the Dominican Republic (Kivisild *et al*., 2006). The sample UG_CGR_71699 (Lugbara) clustered with samples DQ341077 of haplogroup L3f from Ethiopia (Torroni *et al*., 2006), sample DQ112717, of haplogroup L3f from the Dominican Republic (Kivisild *et al*., 2006), and African-American samples DQ305032, DQ305036, DQ305034 and DQ305035 which belong to haplogroup L3f (Just *et al*., 2008). The clustering of the Ugandan samples with African-Americans or Dominican Republican samples was explained in paragraph two of this section. The clustering of these samples with the Mkamba sample was likely due to admixture due to their geographical proximity since the Mkamba sample was obtained from East Africa (Behar *et al*., 2008). The

clustering of the Acholi sample UG_CGR_71473 with two Sudanese samples is evidence for the historical oral tradition that the Acholi and related tribes originated from Southern Sudan (Were and Wilson, 1984; Ssekamwa, 1994). The clustering of the Acholi and Lugbara samples with Ethiopian samples was attributed to the long term historical interactions (Were and Wilson, 1984; Ssekamwa, 1994) and geographical proximity (Ssekamwa, 1994) between Ethiopia and Sudan.

5.3.4.1 Impact of the study on the phylogeny of macrohaplogroup L

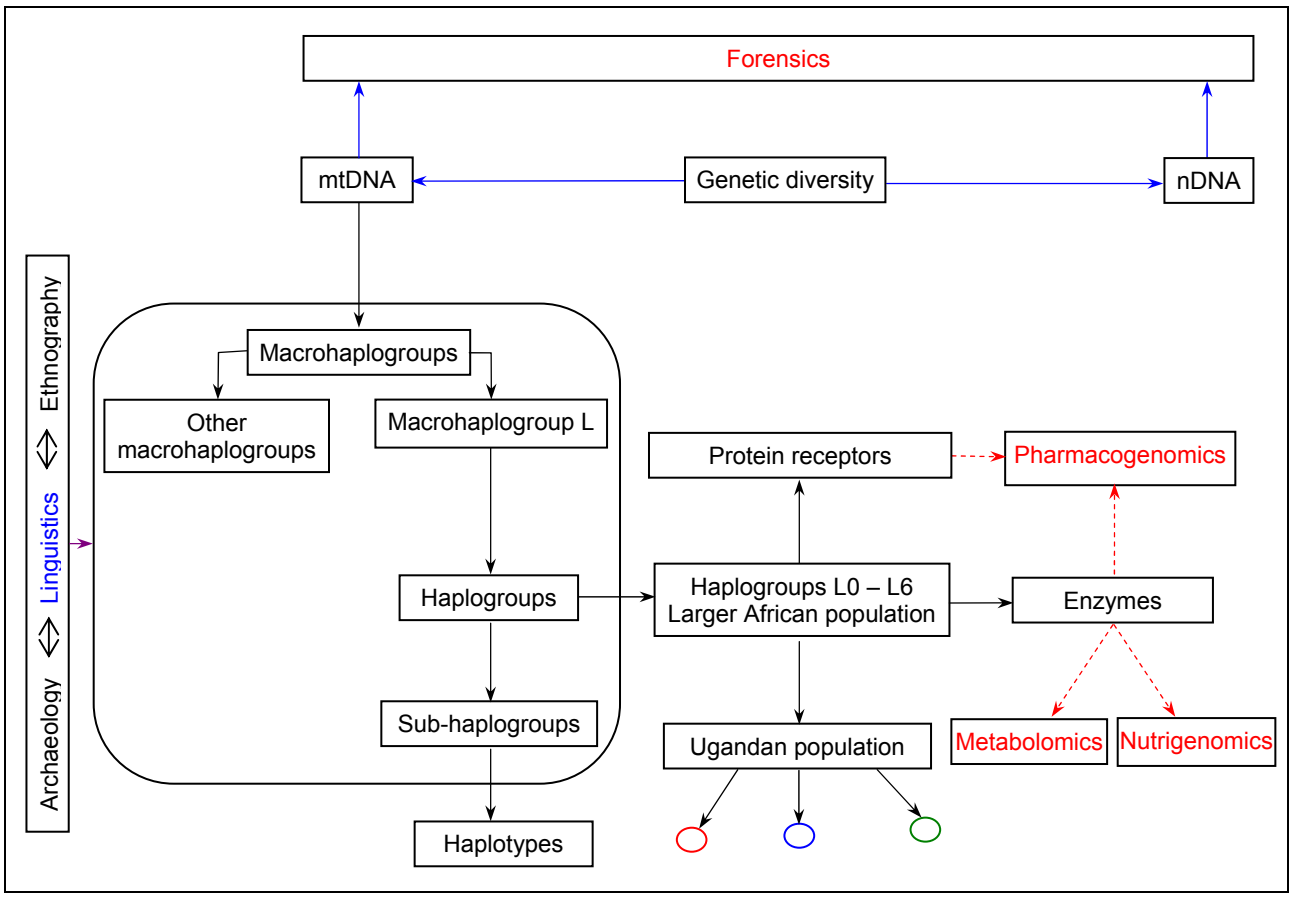
This study is in agreement with Gonder *et al*. (2007) where the mtDNA pool in Eastern Africa played a significant role as the source of many African mtDNA haplogroups. Due to their relatively high prevalence, it could be possible that Ugandans had a role in the origin and dispersal of small haplogroups such as L3i, L3h, L4 and L5. Larger samples need to be studied to verify this fact and to define further the genetic landscape of the Ugandan population. The study has also led to an improvement in the resolution of the small haplogroups, it has produced novel polymorphisms and a novel branch as well as suggesting the finding that the L5 haplogroup is the most ancestral of all African lineages. Both the NJ and MP tree are in agreement with these additional but provocative findings in human history. However, larger sample sizes need to be sequenced and analysed to confirm the fact that L5 is the most ancestral of all global lineages. The small number of available L5 sequences used in previous studies could account for the view that L0 was the oldest haplogroup.

CHAPTER SIX

Conclusions

In view of the scarcity of information on the mtDNA of Ugandan populations, the full mtDNA sequences of 40 individuals belonging to three ethnic groups were determined. This was performed with the aim of determining the genetic relatedness of the individuals within and between tribes and also to establish their evolutionary relationship on a global scale. The entire mtDNA sequences for 13 Baganda (Bantu ethnic group), 13 Acholi (Nilotic ethnic group) and 14 Lugbara (Moru-Madi ethnic group) individuals from Uganda were determined and the phylogenetic relationships of the observed polymorphisms established. This is the first ever attempt to sequence and assess the mtDNA heritage of identified ethnicities of the Ugandan population. Therefore, prior to this study, there was no knowledge at all of the mtDNA genetic relationship between any Ugandan ethnic groups, due to the unavailability of data. Furthermore, the origin of the Ugandan populations has largely been defined by oral tradition (Were and Wilson, 1984) while archaeological data for the population is lacking (Were and Wilson, 1984). In view of the absence or scarcity of archaeological and genetic information for Ugandans, and the available linguistic and ethnographic information on the population, the model as illustrated in Figure 6.1 has been developed in an attempt to define the genetic relatedness, origins and migrations of the Baganda, Acholi and Lugbara of Uganda.

Figure 6.1 The impact of the mtDNA sequences of samples from the Baganda, Acholi and Lugbara tribes from Uganda on gene expression



Small green circle = Baganda, small red circle = Acholi, small blue circle = Lugbara, blue arrows = data for the population is scarce, violet arrows = data for some of the population is available but lacking in others, dotted red arrows = data for the population is very scarce, words in blue colour indicate that there is sufficient information about the population, words in red colour indicate potential applications for the study when data from nDNA is available.

The model depicts a number of disciplines that are needed to precisely define the origins and migratory activities of the Baganda, Acholi and Lugbara from Uganda. The disciplines included archaeology, linguistics, ethnography and genetics. Both archaeological and genetic information for the populations under investigation were scant if not totally lacking. Moreover, only a few ethnographic values were useful in defining the relatedness of the ethnic groups in Africa. It is envisaged that there are sequence motifs that link geographically distant language speakers with each other and with their ancestral place of origin (Blench, 2004). This would help to confirm the relatedness of distant populations in space and time. Furthermore, the polymorphisms established in this investigation will aid in increasing on the resolution that can be attained for macrohaplogroup L.

6.1 PHARMACOGENOMIC, NUTRIGENOMIC, MEDICAL AND FORENSIC POTENTIAL OF THE SEQUENCES

The study identified 14 mtDNA clusters among the Baganda, Acholi and Lugbara samples by the haplotype comparison of 40 full mtDNA sequences. Sequence motifs unique to

clusters are of phylogenetic significance but can also be used in minimising genotyping errors and interpreting SNP profiles if the forensic investigations are to be performed in the context of a constructed mtDNA phylogeny (Álvares-Iglasius *et al*., 2007; Salas *et al*., 2007). These cluster-specific polymorphisms may be of importance in forensic applications as they can be critical in reducing the range of suspects in an investigation of potential criminals especially in situations where the material for evidence has been highly degraded (Melton and Nelson, 2001) or where the evidential material lacks nDNA such as is the case in telogen hair shafts (Melton and Nelson, 2001). The technique would also serve to identify missing persons such as corpses in mass graves for burial by the relatives who seek them. Furthermore, the haplotype diversity of 97.5% exhibited increases in the discriminatory power possessed by the sequences for forensic utility. However, larger samples need to be sequenced before firm deductions on the mtDNA forensic potential and the high haplotype diversity of these Ugandan populations are drawn.

Whereas no single mutation identified in this study, which consists of presumably healthy individuals, has been confirmed as pathogenic, the sequences generated will serve as a reference from which pathogenic mutations can be identified by comparison with the sequences from disease populations. The polymorphisms identified in diseased populations but not in controls, either as singletons or in specific combinations, could serve as signatures of disease that can be used to develop specific products tailored to particular ethnic groups. Particular mtDNA haplotypes when coupled with polymorphisms associated with diseases can predispose individuals to disease (Torroni *et al*., 1996; Wallace *et al*., 1999; Howell *et al*., 2003) while others may lower their risk of disease penetrance (Wallace *et al*., 1999). Thus investigating the genetic background of a specified alteration will enable a greater understanding of the pathophysiology of the diseases that affect the normal function of the mitochondria (Wallace *et al*., 1999).

As depicted in the model, information derived from the polymorphisms in combination with nDNA polymorphisms could be correlated to the underlying phenotypes or the metabolomic profiles of individuals or ethnic groups and used from a pharmacogenomic and nutrigenomic point of view to derive optimal pharmaceutical and nutritional treatments for such populations. The challenge would be to design appropriate diets as well as clinically useful drugs for the African populations that do not affect normal mitochondrial function or the proteome from the nuclear genome that interacts with the mitochondrion. Characterisation of populations at the genetic level could guide in predicting trends of

disease prevalence as well as enabling the informed prioritisation of health care services. The cluster-specific SNPs could be used as signatures to study variation in human response to drugs and nutrients since the alterations themselves could have an effect on the person's ability to metabolise drugs and nutrients. The knowledge of genetic diversity among human beings and the understanding of the manner in which human bodies vary in their response to drug treatment (Weber, 2001) and diet (Kaput *et al.*, 2005) suggests that the goal of pharmacogenomics (designing of personalised drugs) and nutrigenomics (designing of personalised diets) may be feasible and realistic (Weber, 2001; Kaput *et al.*, 2005). However, the effort to transduce the lineage-specific SNPs into pharmaceuticals and nutriceuticals which target specific lineages or individuals remains a formidable challenge. Molecular characterisation of enzymes, receptors, and other proteins that contribute to the variation in human drug and diet responses, increases the prospect that their polymorphisms may soon be catalogued for a variety of human populations (Weber, 2001) and used in the design of personalised or population-specific drugs and diets (Motulsky and Qi, 2006). However, the benefits of pharmacotherapy and personalised diets, though feasible, need to be tackled with care as there are underlying ethical considerations that can be violated with respect to ethnic-specific drugs and diets. Since self-identified ethnicity may be an inaccurate representation of genetic clusters (Wilson *et al.*, 2001), pharmacotherapy and personalised diets must involve genotyping at individual level whereby the costs of such undertakings is not affordable for many populations. The optimal global sharing of knowledge and use of systems biology techniques (Kaput *et al.*, 2005) is the ultimate way to deliver pharmacogenomics and nutrigenomics to the level of application.

6.2 GENETIC RELATEDNESS OF THE BAGANDA, ACHOLI AND LUGBARA FROM UGANDA

The pattern of the observed polymorphisms demonstrated that the tribal samples were genetically different from each other but more samples need to be sequenced to verify this assertion. The mtDNA sequence differences between the samples were further supported by the fact that no single haplotype out of the 40 sequences was shared between any two tribes. The only one haplotype shared and moreover only by two individuals was within a tribe. Therefore, a bigger sample will be required to establish the haplotype distribution among tribes and ethnic groups in Uganda. On the basis of the polymorphisms that were exclusive to two tribes only for the sample set studied, there seems to have been a closer genetic relatedness between Acholi and Lugbara individuals followed by that between the

Acholi and Baganda and the least between the Baganda and Lugbara individuals. The clustering patterns of the NJ and MP trees in Figure 5.59 and Figure 5.60 also demonstrated that the Acholi and Lugbara seem to be more closely related to each other than they are to the Baganda. Only the Baganda samples (46%) clustered in L0 haplogroup but none were determined in L1, L2 and L5 while the majority of the Lugbara samples (43%) and a high proportion of the Acholi samples (38%) clustered in haplogroup L2. The clustering of the Ganda L3 sequences with the Acholi individuals was attributed to admixture with the Luo-Babito of Bunyoro (Were and Wilson, 1984; Ssekamwa, 1994). Since it is a widely distributed haplogroup among African populations (Watson *et al*., 1997; Brandstätter *et al*., 2004; Jackson *et al*., 2005; Allard *et al*., 2005; Abu-Amero *et al*., 2007), the absence of haplogroup L2 among the Baganda samples is intriguing but firm deductions on this can only be drawn after analysis of more sequences.

6.3 IMPLICATIONS ON THE GLOBAL PHYLOGENETIC TREE

This study is the beginning of a larger project aimed at the complete sequencing of hundreds of mtDNAs belonging to haplogroups and sub-haplogroups that are representative of the whole range of the mtDNA variation in Uganda and its neighbouring countries. It was envisaged that this and future projects would make a unique contribution to the global human phylogenetic tree both in the context of human origins and disease. The mtDNA sequences generated in this study add to a growing volume of evidence characterising the structure and genetic diversity of human populations. These networks increase the bulk of sequences available to infer logical deductions on African mtDNA sequences and facilitated further resolution of macrohaplogroup L. Moreover, the amount of available complete mtDNA sequences from African populations is still limited (Behar *et al*., 2008). The 40 haplotypes generated in this study when combined with the global African sample of 387 sequences (as indicated in Figure 5.59 and Figure 5.60 increased the resolution of macrohaplogroup L as some represented novel haplotypes that produced new branches and nodes in the tree. Moreover, a number of these sequences clustered in regions of small African haplogroups such as L3h, L3i, L4 and L5 which markedly increased the resolution of those regions in the global tree. The resolution was also greatly increased by sample UG_CGR_71678 which clustered in a region of haplogroup L which had only had a few existing complete sequences and sample UG_CGR_71471 which formed a distinct cluster with only one other sample in the global African pool in haplogroup L2. The clustering of samples UG_CGR_71470 and UG_CGR_71478, where the two samples formed a distinct cluster of their own in the global tree, signified a

discovery of a new evolutionary lineage amongst macrohaplogroup L haplotypes and formed part of the area of the Ugandan sequences outside the region of intersection with other African samples in the model outlined in Figure 6.1. Thus, this was evidence that there may still be mtDNA lineages that are unique to Ugandan populations and raises the possibility of finding more new nodes in the macrohaplogroup L tree when more samples in these populations are sequenced. Sequencing and analysis of more samples from Southern Sudan and Ethiopia also needs to be performed to find out whether these novel lineages obtained and those still anticipated are unique to samples from Ugandan populations only.

6.4 IMPLICATIONS FOR THE ETHNOLINGUISTIC CLASSIFICATION OF THE BAGANDA, ACHOLI AND LUGBARA

This study was also an attempt to establish whether the genetic relationship between Baganda, Acholi and Lugbara tribes correlates with their linguistic relationship. Human populations worldwide can be compared and distinguished from each other based on the similarities and differences in their ancestral languages (Raymond, 2005). This study was therefore also designed to compare the Baganda, Acholi and Lugbara linguistic phyla, with the phylogeny derived from all L mtDNA dataset. In likely agreement with linguistic classification, where the Acholi and Lugbara belong to the Nilo-Saharan language phyla (Raymond, 2005), the mtDNA sequences of the Acholi samples were more closely related to the Lugbara samples compared to their relationship with the Baganda samples who belong to the Niger-Congo language phyla (Raymond, 2005) since more Lugbara and Acholi samples clustered together as compared to the Baganda - Acholi and Baganda - Lugbara sample clusters. More samples from the Baganda, Acholi and Lugbara should be sequenced to investigate further the linguistic - genetic relatedness of the three tribes.

There was a marked clustering of Acholi and Lugbara individuals in haplogroups L2, L3 and L5. The phylogenetic tree therefore illustrated that the Acholi and Lugbara samples were more closely related to each other than they were to the Baganda samples which was in agreement with linguistic classification and geographical proximity as well as their place of origin. However, the clustering of these samples in different branches of the tree was attributed to the samples having diversified origins from their ancestral place in Southern Sudan as the Bahr-el-Ghazal province is a wide area (Ssekamwa, 1994). The consequence of admixture of the Luo in Bunyoro with the Ganda after the collapse of the

Bunyoro-Kitara Empire as outlined in Section 3.2.1.1.2 (Were and Wilson, 1984; Ssekamwa, 1994) may have led to clustering of certain Baganda individuals with the Acholi as was observed in haplogroup L3b. If this is not the case, it may be due to the fact that the Baganda may have had a tendency of marrying from diverse origins/tribes more frequently compared to the Lugbara and Acholi. Furthermore, when the British colonised Uganda, they used the Baganda to colonise other areas of the country (Were and Wilson, 1984) and Luganda (Ganda) was used as the first language of administration. This could have served the Baganda men well, to break the language or communication barrier that limits intertribal marriages, and thus enabling interethnic marriages with the Baganda. Moreover, as the Baganda men moved with the British in colonising Uganda, they could have used the time and opportunity to marry women from other Ugandan tribes. Indeed there are many Baganda communities that are today situated in different parts of Uganda and this phenomenon may be a confounding variable in deciphering the origins of the Baganda from a genetic perspective. Admixture involving the Baganda, Acholi and Lugbara c'd make the individuals cluster in any haplogroup but in my study samples, only haplgroup L3b cluster had all the three tribes. Since no reason can be given as to why clustering of the samples did not happen in other haplogroup tree branches, other than L3b, it is likely that the phenomenon of clustering in L3b haplogroup was due to admixture involving Baganda men with the Luo-Babito from Bunyoro since these Luo-Babito were formed from a group of Luo who migrated from Pakwach to Bunyoro (Ssekamwa, 1994). These Luo should have had mainly haplogroup L3b mtDNAs.

It is recommended that Uganda would be a suitable country for studying the relationship between mtDNA and languages because a total of 54 indigenous languages/ethnic groups and at least two languages with origins elsewhere (Nubians from Sudan and Ugandan Asians) are spoken within it (Wairama, 2001). The Ugandan Asians include the Hindu, Gujarati and Pakistanis who came to Uganda between the end of the 19th and beginning of the 20th century (Wairama, 2001). Moreover, many of these languages have dialects such that gene flow across dialects and tribes could be studied. Furthermore, these languages are spoken in a small area of 244,400 km^2 (Young and Mottram, 1977), of which 17.2% is covered by water (Privatisation and Utility Sector Reform Project in Uganda, 2002) negating in most instances genetic variation due to geographical distances since tribes in geographical proximity tend to be more genetically related mostly through intermarriages. However, gene flow across geographical barriers such as the Nile River and Lake Victoria (which divides the country into two), Lake Kyoga, Lake George, Lake Edward and Lake

Albert and gene flow across large forests and large natural game reserves needs to be investigated in Uganda as in most instances, there is variation in the languages spoken between populations living on opposite sides of the bodies of water or large forests and large game reserves.

6.5 IMPLICATIONS FOR THE PEOPLING OF SOUTH-EASTERN, NORTHERN AND NORTH-WESTERN UGANDA

Due to the clustering of the Lugbara and Acholi samples with sequences of Nilotes from Tanzania (Gonder *et al*., 2007) who migrated from Sudan (Ssekamwa, 1994) and due to their clustering with some of the Sudanese sequence in the African dataset, this data is in agreement with the oral tradition that the Ugandan Nilotes and related tribes had their origin from Sudan (Were and Wilson, 1984; Ssekamwa, 1994). However, this assertion needs further support by the sequencing and analysis of more samples from Southern Sudan and Northern Uganda. Owing to it being a relatively recent and small haplogroup (Torroni *et al*., 2006), the origin of the L4a haplogroup (referred to as haplogroup L7 by Torroni *et al*., 2006) among the Acholi and Lugbara cannot be precisely delineated and its geographical link with the L4a sequence from Ethiopia is attributed to the long term historical interactions (Ssekamwa, 1994) and geographical proximity to Ethiopia. It was probable that Uganda, Ethiopia, Sudan and Kenya or all four countries could be the source of haplogroup L4a sequences. The presence and clustering of haplogroup L4 sequences of the Acholi and Ganda with Tanzanians and Ethiopians demonstrated that East Africa was the source and dispersal area for haplogroup L4 sequences.

The peopling of Buganda or Southeastern Uganda by the individuals harbouring the L0 haplogroup could have been by the Western Bantu immigrants who comprised the original inhabitants of Buganda, who migrated from the East Coast at Shungwaya and settled in the counties of Busiro, Kyadondo and Mawokota (Were and Wilson, 1984; Ssekamwa, 1994). The Baganda samples clustering with L4 sequences from the Acholi could have had a near recent ancestry with the Acholi. Since this section of the Baganda clustered at the terminal branches of the tree, it is highly probable that this section of the Baganda descended or split off from the Acholi after entering Uganda from Sudan. The records of the Acholi - Baganda split in the L4 haplogroup cluster were stored in the A257G, T1694C, T2483C, A4949G, G6260A, T6620C, T11137C polymorphisms among the Acholi and the T5580C, G5746A polymorphisms harboured by the Baganda as these are the only polymorphic differences between the haplogroup L4 Acholi and Baganda samples in this

investigation. Further investigations need to be performed on Acholi and Ganda samples to explore this assertion. Based on the greater genetic similarity between the Acholi and Ganda in haplogroup L3b, where sequences in this group differed at only one or two sites, it was highly probable that this L3b cluster of sequences as explained in Section 5.3.4 could be originating from the group of the Luo who conquered and led to the collapse of the Bunyoro-Kitara Empire (Ssekamwa, 1994) but also continually raided the Kingdom of Buganda (Were and Wilson, 1984). If this happens to be the case, these sequences should have had close matches with sequences from regions such as Bunyoro, Ankole, Busoga, Tooro in Uganda, Rwanda, Karagwe in Tanzania (Were and Wilson, 1984; Ssekamwa, 1994) and Wanga in Kenya (Were and Wilson, 1984) that came under the control of the Luo-Babito during the era of the Bunyoro-Kitara Empire. The mtDNA sequences of such populations have however not yet been established. In the same way, the absence of close matches of the L3b Ganda sequences with samples from Tanzania was attributed to a lack of sequence data amongst the inhabitants of Karagwe. The Ganda harbouring haplogroup L3i sequence motifs originated from Ethiopia and should belong to the descendants of the group of individuals that entered Buganda via Mt Elgon (Were and Wilson, 1984; Ssekamwa, 1994). It is therefore highly prudent that the females who accompanied these Ganda immigrants should have been harbouring L3i haplogroup sequence motifs.

To date, few mtDNA sequences from Sudan have been analysed. An analysis of mtDNA polymorphisms in Sudan and Northern Uganda will lead to assessment of the contribution of Sudan to the peopling of Uganda and vice-versa. As depicted in the model, it is recommended that a combination of linguistic, archaeological and genetic research (mtDNA and Y-chromosome sequencing) be undertaken on all Ugandan populations and the surrounding countries in order to define precisely the process that led to the peopling of Uganda by anatomically modern humans. Analysis of the Y-chromosome could be complementary in assessing the history of the Ugandan population from a male perspective.

6.6 IMPLICATIONS OF THE SLAVE TRADE

It is again envisaged that distantly related people have closely related mtDNA haplotypes that link them to their ancestral place of origin. The clustering of sequences of the Baganda, Lugbara and Acholi with sequences of individuals from Dominican Republic and the Island of Hispaniola, demonstrated that there was a possibility of slave trade involving

Ugandans being shipped into America by some means. Since there is no historical account of slave trade directly from Uganda to America and since it was the Arabs, Afro-Arabs and Egyptians who penetrated the interior of Uganda to deal in slavery (Ssekamwa, 1994; Were and Wilson, 1984), it is probable that the Arabs, Afro-Arabs or Egyptians sold some of the slaves to America. Alternatively, it could be that there was some undetectable gene flow from Uganda to certain countries in Africa and thereafter to America thus contributing to the transatlantic slave trade. Concrete evidence for this is however unavailable. Alternatively, there could have been genetic similarity between individuals in certain African countries that contributed to transatlantic slave trade with these Ugandan samples but no evidence of this is available. The characterisation of East African populations will therefore be of immense value to the African-Americans and African-Arabs to discover their heritage that was lost through slave trade, a goal that has been the primary objective of the DNA Roots projects (Alves-Silva *et al*., 2000; Bandelt *et al*., 2001; Salas *et al*., 2005; Abu-Amero *et al*., 2007).

6.7 RECOMMENDATIONS AND FINAL REMARKS

Inasmuch as the mtDNA sequences from samples belonging to three Ugandan tribes have been established, a full understanding of the genetic landscape and demography of the Ugandan population and ethnic groups requires the sequencing of the mtDNA of all tribes within each ethnic group for the Ugandan population. The stratification of the Baganda, Acholi, Lugbara and other East African populations at the genetic level will aid in defining the origins and migrations humans as well as allowing for the exploration of the forensic, nutrigenomic, metabolomic and pharmacogenomic aspects of the Ugandan population. This study advocates for collaboration of geneticists, historians, linguists and archaeologists to generate an accurate genetic and demographic overview of the population. The study illustrates that Uganda possesses populations harbouring many of the ancient haplogroups that occupy the basal branches of the human phylogenetic tree such as L0, L1 and L5. The study puts Sudan as the most likely geographical area anatomically modern man radiated from to populate East Africa, the rest of Africa and the world. Haplogroup L5, in contrast to studies prior to this, turns out to be the oldest of existing haplogroups as it was placed at the root of the human mtDNA macrohaplogroup L phylogenetic tree.

CHAPTER SEVEN

References

7.1 GENERAL REFERENCES

Achilli A., Rengo C., Magri C., Battaglia V., Olivieri A., Scozzari R., Cruciani F., Zeviani M., Briem E., Carelli V., Moral P., Dugoujon J-M., Roostalu U., Loogväli E-L., Kivisild T., Bandelt H-J., Richards M., Villems R., Santachiara-Benerecetti A.S., Semino O. and Torroni A. The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *Am. J. Hum. Genet.*, **75**, 910-918, 2004.

Achilli A., rengo C., Battaglia V., Pala M., Olivieri A., Fornarino S., Magri C., Scozzari R., Babudri N., Santachiara-Benerecetti A.S., Bandelt H-J., Semino O. and Torroni A. Saamis and Berbers – An unexpected mitochondrial DNA link. *Am. J. Hum. Genet.*, **76**, 883-886, 2005.

Adams P.L. and Turnbull D.M. Disorders of the electron transport chain. *J. Inher. Metab. Dis.*, **19**, 463-469, 1996.

Adcock G.J., Dennis E.S., Easteal S., Huttley G.A., Jermiin L.S., Peacock W.J. and Thorne A. Mitochondrial DNA sequences in ancient Australians: Implications for modern human origins. *Proc. Natl. Acad. Sci. USA*, **98**(2), 537-542, 2001.

Aiello L.C. The fossil evidence for modern human origins in Africa: A revised view. *Am. Anthropol.*, **95**(1), 73-76, 1993.

Allard M.W., Polanskey D., Miller K., Wilson M.R., Monson K.L. and Budowle B. Characterisation of human control region sequences of the African American SWGDAM forensic mtDNA data set. *Forensic Sci. Int.*, **148**, 169-179, 2005.

Álvarez-Iglesias V., Jaime J.C., Carracedo Á. and Salas A. Coding region mitochondrial DNA SNPs: Targeting East Asian and Native American haplogroups. *Forensic Sci. Int.*, **Genetics 1**, 44-55, 2007.

Alves-Silva J., da Silva S.M., Guimaraes P.E., Ferreira A.C., Bandelt H.J., Pena S.D. and Prado V.F. The ancestry of Brazilian mtDNA lineages. *Am. J. Hum. Genet.*, **67**(2), 444-461, 2000.

Ames B.N., Shigenaga M.K. and Hagen T.M. Oxidants, antioxidants and the degenerative diseases of aging. *Proc. Natl. Acad. Sci. USA*, **90**, 7915-7922, 1993.

Anderson S., Bankier A.T., Barrell B.G., de Bruijn M.H.L., Coulson A.R., Drouin J., Eperon I.C., Nierlich D.P., Roe B.A., Sanger F., Schreier P.H., Smith A.J.H., Staden R. and Young I.G. Sequence and organisation of the human mitochondrial genome. *Nature*, **290**, 457-465, 1981.

Andersson B., Jia-Hsu Y., Lewis D.E. and Gibbs R.A. Rapid characterisation of HIV-1 sequence diversity using denaturing gradient gel electrophoresis and direct automated DNA sequencing of PCR products. *PCR Methods and Appl.*,**2**, 293-300, 1993.

Andreas M.K., Lott M.T., Brown M.D., Navathe S.B. and Wallace D.C. Mitomap: an update on the status of the human mitochondrial genome database. *Nucleic Acids Res.,* **25**, 96-199, 1997.

Andrews R.M., Kubacka I., Chinnery P.F., Lightowers R.N., Turnbull D.M. and Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.*, **23**, 147, 1999.

Aoki H., Kang P.M., Hampe J., Yoshimura K., Noma T., Matsuzaki M. and Izumo S. Direct activation of mitochondrial apoptosis machinery by c-jun N-terminal kinase in adult cardiac myocytes. *J. Biol. Chem.*, **277**(12), 10244-10250, 2002.

Aquadro C.F. and Greenberg B.D. Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals. *Genetics*, **103**(2), 287-312, 1983.

Armelagos G.J. and Harper K.N. Genomics at the origins of agriculture, Part Two. *Evol. Anthropol.*, **14**, 109-121, 2005.

Attardi G. Animal mitochondrial DNA: An example of genetic economy. *Int. Rev. Cytol.*, **93**, 93-145, 1985.

Attardi G. The elucidation of the human mitochondrial genome: A historical perspective. *Bioassays*, **5**(1), 34-39, 1986.

Aubin N., Barneoud P., Carter C., Caille D., Sontag N., Marc C., Lolivier J., Gardes A., Perron C., Le Kim A., Charieras T., Pandini M., Burnier P., Puech F., Jegham S., George P., Scatton B. and Curet O. SL25.1131 [3(S), 3a(S)-3-methoxymethyl-7-[4,4,4-trifluorobutoxy]-3,3a,4,5-tetrahydro-1,3-oxazolo

[3,4-a]quinolin-1-one], a new, reversible, and mixed inhibitor of monoamine oxidase-A and monoamine oxidase-B: biochemical and behavioural profile. *J. Pharmacol. Exp. Ther.*, **310**(3), 1171-1182, 2004.

Ballana E., Morales E., Rabionet R., Montserrat B., Ventayol M., Bravo O., Gasparini P. and Estivill X. Mitochondrial 12S rRNA gene mutations affect RNA secondary structure and lead to variable penetrance in hearing impairment. *Biochem. Biophys. Res. Comm.*, **341**(4), 950-957, 2006.

Ballinger S.W., Schurr T.G., Torroni A., Gan Y.Y., Hodge J.A., Hassan K., Chen K.H. and Wallace D.C. Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient Mongoloid migrations. *Genetics*, **130**, 139-152, 1992.

Bandelt H-J., Alves-Silva J., Guimaraes P.E., Santos M.S., Brehm A., Pereira L., Coppa A., Larruga J.M., Rengo C., Scozzari R., Torroni A., Prata M.J., Amorim A., Prado V.F. and Pena S.D. Phylogeography of the human mitochondrial haplogroup L3e: a snapshot of African prehistory and Atlantic slave trade. *Ann. Hum. Genet.* **65**(6), 549-563, 2001.

Bandelt H-J., Forster P., Sykes B.C. and Richards M.B. Mitochondrial portraits of human populations using median networks. *Genetics*, **141**, 743-753, 1995.

Bandelt H-J., Salas A. and Bravi C.M. What is a novel mtDNA mutation – and does novelty really matter? *J. Hum. Genet.*, **51**, 1073-1082, 2006.

Barbujani G. DNA variation and language affinities. *Am. J. Hum. Genet.*, **61**, 1011-1014, 1997.

Barbujani G. and Bertorelle G. Genetics and the population history of Europe. *Proc. Natl. Acad. Sci. USA*, **98**(1), 22-25, 2001.

Barbujani G. and Pilastro A. Genetic evidence on origin and dispersal of human populations speaking languages of the Nostratic macrofamily. *Proc. Natl. Acad. Sci. USA*, **90**, 4670-4673, 1993.

Barrell B.G., Anderson S., Bankier A.T., De Bruijn M.H.L., Chen E., Coulson A.R., Drouin J., Eperon I.C., Nierlich D.P., Roe B.A., Sanger F., Schreier P.H., Smith A.J.H., Staden R. and Young I.G. Different pattern of codon recognition by mammalian mitochondrial tRNAs. *Proc. Natl. Acad. Sci. USA*, **77**(6), 3164-3166, 1980.

Barrell B.G., Bankier A.T. and Drouin J. A different genetic code in human mitochondria. *Nature*, **282**(8), 189-194, 1979.

Bauer M.F., Gempel K., Hofmann S., Jaksch M., Philbrook C. and Gerbitz K. Mitochondrial disorders. A diagnostic challenge in clinical chemistry. *Clin. Chem. Lab. Med.* **37**(7), 855-876, 1999.

Beal M.F. Bioenergetic approaches for neuroprotection in Parkinson's disease. *Ann. Neurol.*, **53**(3), S39-S48, 2003.

Beckman K.B. and Ames B.N. Oxidative decay of DNA. *J. Biol. Chem.*, **272**(32), 19633-19636, 1997.

Beckman K.B. and Ames B.N. Endogenous oxidative damage of mtDNA. *Mutat. Res.*, **424**, 51-58, 1999.

Behar D.M., 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.S., Rosset S. and the Genographic consortium. The dawn of the human matrilineal diversity. *Am. J. Hum. Genet.*, **82**, 1-11, 2008.

Berg J.M., Tymoczko J.L. and Stryer L. Biochemistry. 5th Ed. W.H. Freeman and Company, New York, 2002.

Blanchard J. and Lynch M. Organellar genes. Why do they end up in the nucleus? *TIG*, **16**(7), 315-320, 2000.

Bogenhagen D. and Clayton D.A. The number of mitochondrial deoxyribonucleic acid genomes in mouse L and human HeLa cells. *J. Biol. Chem.*, **249**(24), 7991-7995, 1974.

Bogin B. and Rios L. Rapid morphological change in living humans: implications for modern human origins. *Comp. Biochem. Physiol.*, **136**(1), 71-84, 2003.

Bolnick D.A., Shook B.A., Campbell L. and Goddard I. Problematic use of Greenberg's linguistic classification of the Americas in studies of native American genetic variation. *Am. J. Hum. Genet.*, **75**, 519-523, 2004.

Bonatto S.L. and Salzano F.M. A single and early migration for the peopling of the Americas supported by mitochondrial DNA sequence data. *Proc. Natl. Acad. Sci. USA*, **94**, 1866-1871, 1997.

Borst P. Structure and function of mitochondrial DNA. *Trends in Biochem. Sci.*, **2**, 31-34, 1977.

Bortoloni M.C.T., Da Silva W.A., Zago M.A. Elion J., Krishnamoorthy R., Goncalves and Pena S.D.J. The phylogeography of mitochondrial DNA Haplogroup L3g in Africa and the Atlantic Slave Trade. *Am. J. Hum. Genet.* **75**, 523–524, 2004.

Bowmaker M., Yang M.Y., Yasukawa T., Reyes A., Jacobs S.T., Huberman J.A. and Holt I.J. Mammalian mitochondrial DNA replicates bidirectionally from an initiation zone. *J. Biol. Chem.*, **278**(51), 50961-50969, 2003.

Bräuer G., Yokoyama Y., Falguères C. and Mbua E. Modern human origins backdated. *Nature*, **386**, 337-338, 1997.

Brandstatter A., Peterson C. T., Irwin J. A., Mpoke S., Koech D. K., Parson W., Parsons T. J. Mitochondrial DNA control region sequences from Nairobi (Kenya): inferring phylogenetic parameters for the establishment of a forensic database. *Int. J. Leg. Med.*, **118** (5), 294-306, 2004.

Brown G.G. and Simpson M.V. Novel features of animal mtDNA evolution as shown by sequences of two rat cytochrome oxidase subunit II genes. *Proc. Natl. Acad. Sci. USA*, **79**, 3246-33250, 1982.

Brown W.M. Polymorphisms of mitochondrial DNA of humans as revealed by restriction endonuclease analysis. *Proc. Natl. Acad. Sci. USA*, **77**(6), 3605-3609, 1980.

Brown M.D., Hosseini S.H., Torroni A., Bandelt H.J., Allen J.C., Schurr T., Scozzari R., Cruciani F. and Wallace D.C. mtDNA haplogroup X: An ancient link between Europe/Western Asia and North America? *Am. J. Hum. Genet.*, **63**(6), 1852-1861, 1998.

Brown W.M., Prager E.M., Wang A. and Wilson A.C. Mitochondrial DNA sequences of primates: tempo and mode of evolution. *J. Mol. Evol.*, **18**, 225-239, 1982.

Brown M.D., Voljavec A.S., Lott M.T., Torroni A., Yang C. and Wallace D.C. Mitochondrial DNA complex I and III mutations associated with Leber's hereditary optic neuropathy. *Genetics*, **130**, 163-173, 1992.

Brown W.M., George M. and Wilson A.C. Rapid evolution of animal mitochondrial DNA. *Proc. Natl. Acad. Sci. USA*, **76**(4), 1967-1971, 1979.

Brustovetsky N. and Dubinsky J.M. Dual responses of CNS mitochondria to elevated calcium. *J. Neurosci.*, **20**(1), 103-113, 2000.

Budd S.L. and Nicholls D.G. Mitochondria in the life and death of neurons. *Essays Biochem.*, **33**, 43-52, 1998.

Budowle B., Allard M.W., Wilson M.R. and Chakraborty R. Forensics and mitochondrial DNA: Applications, debates and foundations. *Annu. Rev. Genomics Hum. Genet.*, **4**, 119-141, 2003.

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

Burchard E.G., Ziv E., Coyle N., Gomez S.L., Tang H., Karter A.J., Mountain J.L., Pérez-Stable E.J., Sheppard D. and Risch N. The importance of race and ethnic background in biomedical research and clinical practice. *N. Engl. J. Med.*, **348**(12), 1170-1175, 2003.

Camougrand N. and Rigoulet M. Aging and oxidative stress: studies of some genes involved both in aging and in response to oxidative stress. *Respir. Physiol.*, **128**, 393-401, 2001.

Cann R.L., Stoneking M. and Wilson A.C. Mitochondrial DNA and human evolution. *Nature*, **325**, 31-36, 1987.

Caramelli D., Lalueza-Fox C., Vernesi C., Lari M., Casoli A., Mallegni F., Chiarelli B., Dupanloup I., Bertranpetit J., Barbujani G. and Bertorelle G. Evidence for a genetic discontinuity between Neanderthals and 24,000-year-old anatomically modern Europeans. *Proc. Natl. Acad. Sci. USA*, **100**(11), 6593-6597, 2003.

Carroll J., Shannon R.J., Fearnley I.M., Walker J.E. and Hirst J. Definition of the nuclear encoded protein composition of bovine heart mitochondrial complex I. *J. Biol. Chem.*, **277**, 50311-50317, 2002.

Cavalli-Sforza L.L., Minch E. and Mountain J.L. Coevolution of genes and languages revisited. *Proc. Natl. Acad. Sci. USA*, **89**, 5620-5624, 1992.

Chang D.D. and Clayton D.A. Priming of human mitochondrial DNA replication occurs at the light-strand promoter. *Proc. Natl. Acad. Sci. USA*, **82**, 351-355, 1985.

Cha R.S. and Thilly W.G. Specificity, efficiency, and fidelity of PCR. *PCR Methods Appl.*, **3**, 18- 29, 1993.

Chen F-C. and Li W-H. Genomic Divergences between Humans and Other Hominoids and the Effective Population Size of the Common Ancestor of Humans and Chimpanzees. *Am. J. Hum. Genet.*, **68**, 444-456, 2001.

Chen Y., Olckers A., Schurr T.G., Kogelnik A.M., Huoponen K. and Wallace D.C. mtDNA variation in the South African !Kung and Khwe and their genetic relationships to other African populations. Am. *J. Hum. Genet.*, **66**, 1362-1383, 2000.

Chen Y., Torroni A., Excoffier L., Santachiara-Benerecetti S.A. and Wallace D.C. Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. *Am. J. Hum. Genet.*, **57**, 133-149, 1995.

Chinnery P.F., Brown D.T., Andrews R.M., Singh-Kler R., Riordan-Eva P., Lindley J., Applegarth D.A., Turnbull D.M. and Howell N. The mitochondrial ND6 gene is a hot spot for mutations that cause Leber's hereditary optic neuropathy. *Brain*, **124**, 209-218, 2001.

Christianson T.W. and Clayton D.A. *In vitro* transcription of human mitochondrial DNA: accurate termination requires a DNA sequence that can function bidirectionally. *Proc. Natl. Acad. Sci. USA*, **83**, 6277-6281, 1986.

Coble M.D., Just R.S., O'Callaghan J.E., Letmanyi I.H., Peterson C.T., Irwin J.A. and Parsons T.J. Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of forensic testing. *Int. J. Legal Med.*, **118**, 137-146, 2004.

Comas D., Calafell F., Mateu E., Perez-Lezaun A., Bosch E., Martinez-Arias R., Clarimon J., Facchini F., Fiori G., Luiselli D., Pettener D. and Bertranpetit J. Trading genes along the silk road: mtDNA sequences and the origin of central Asian populations. *Am. J. Hum. Genet.*, **63**, 1824-1838, 1998.

Cordaux R., Weiss G., Saha N. and Stoneking M. The Northeast Indian Passageway: A barrier or corridor for human migrations? *Mol. Biol. Evol.*, 21(8), 1525-1533, 2004.

Cormio A., Mitella F., Vecchiet J., Felzani G., Gadaleta M.N. and Cantatore P. Mitochondrial DNA mutations in RRF of healthy subjects of different age. *Neurobiol. Aging,* **26**, 655-664, 2005.

Coskun P.N., Ruisi-Pesini E. and Wallace D.C. Control region mtDNA variants: Longevity, climatic adaptation, and a forensic conundrum. *Proc. Natl. Acad. Sci. USA*, **100**(5), 2174-2176, 2003.

Cruiciani F., Santolamazza P., Shen P., Moral P., Olckers A., Modiano D., Destro-Bisol G., Holmes S., Coia V., Wallace D.C., Oefner P.J., Torroni A., Cavalli-Sforza L.L., Scozzari R. and Underhill P.A. An Asia to sub-Sahara Africa back migration is supported by high-resolution analysis of human Y chromosome haplotypes. *Am. J. Hum. Genet.*, **70**(5), 1197-1214, 2002.

De Benedictis G., Rose G., Carrieri G., De Luca M., Falcone E., Passarino G., Bonafe M., Monti D., Baggio G., Bertolini S., Mari D., Mattace R. and Franceschi C. Mitochondrial DNA inherited variants are associated with successful aging and longevity in humans. *FASEB J.*, **13**, 1532-1536, 1999.

Del Bo R., Bordoni A., Boneschi F.M., Crimi M., Sciacco M., Bresolin N., Scarlato G. and Comi G.P. Evidence and age-related distribution of mtDNA D-loop point mutations in skeletal muscle from healthy subjects and mitochondrial patients. *J. Neurol. Sci.*, **202**, 85-91, 2002.

Denaro M., Blanc S., Johnson M.J., Chen K.H., Wilmsen E., Cavalli-Sforza L.L. and Wallace D.C. Ethnic variation in *Hpa* I endonuclease cleavage patterns of human mitochondrial DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **78**(9), 5768-5772, 1981.

Derenko M., Malyarchuk B., Grzybowski T., Denisova G., Dambueva I., Perkova M., Dorzhu C., Luzina F., Lee H.K., Vanecek T., Villems R. and Zakharov I. *Am. J. Hum. Genet.*, **81**, 1025-1041, 2007.

Díez-Sánchez C., Ruiz-Pesini E., Lapeña A.C., Montoya J., Pérez-Martos A., Enríquez and López-Pérez. J. Mitochondrial DNA content of human spermatozoa. *Biol. Reprod.*, **68**, 180-185, 2003.

Edwin D., Vishwanathan H., Roy S., Rani M.V.U. and Majunder P.P. Mitochondrial DNA diversity among five tribal populations of southern India. *Curr. Sci.*, **83**(2), 158-162, 2002.

Elson J.L., Andrews R.M., Chinnery P.F., Lightowlers R.N., Turnbull D.M. and Howell N. Analysis of European mtDNAs for recombination. *Am. J. Hum. Genet.*, **68**, 145-153, 2001.

Elson J.L., Turnbull D.M. and Howell N. Comparative genomics and the evolution of human mitochondrial DNA: assessing the effects of selection. *Am. J. Hum. Genet.*, **74**, 229-238, 2004.

Eswaran V., Harpending H. and Rogers A.R. Genomics refutes an exclusively African origin of humans. *J. Hum. Evol.*, **49**(1)1-18, 2005.

Eyre-Walker A. and Awadalla P. Does human mtDNA recombine? *J. Mol. Evol.*, **53**, 430-435, 2001.

Falkenberg M., Gaspari M., Rantanen A., Trifunovic A., Larsson N-G. and Gustafsson C.M. Mitochondrial transcription factors B1 and B2 activate transcription of human mtDNA. *Nat. Genet.*, 31, 289-294, 2002.

Felsenstein J. Phylogenies from molecular sequences: Inference and reliability. *Annu. Rev. Genet.*, **22**, 521-565, 1988.

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

Ferri K.F., Jacotot E., Blanco J., Este J.A. and Kroemer G. Mitochondrial control of cell death induced by HIV-1-encoded proteins. *Ann. NY Acad. Sci.*, **926**, 149-164, 2000.

Finnilä S., Hassinen I.E., Ala-Kokko L. and Majamaa K. Phylogenetic network of the mtDNA haplogroup U in Northern Finland based on sequence analysis of the complete coding region by conformation-sensitive gel electrophoresis. *Am. J. Hum. Genet.*, **66**, 1017-1026, 2000.

Finnilä S., Lehtonen M.S. and Majamaa K. Phylogenetic network for European mtDNA. *Am. J. Hum. Genet.*, **68**, 1475-1484, 2001.

Finn P.J., Bull M.G., Xiao H., Phillips P.D., Nelson J.R., Grossman G., Nampalli S., MaArdle B.F., Mamone J.A., Flick P.K., Fuller C.W. and Kumar S. Efficient incorporation of positively charged 2',3'-dideoxynucleoside-5'-triphosphates by DNA polymerases and their application in 'direct-load' DNA sequencing. *Nucleic Acids Res.*, **31**(16), 4769-4778, 2003.

Florentz C. and Sissler M. Disease-related versus polymorphic mutations in human mitochondrial tRNAs. *EMBO reports*, **21**(61), 481-486, 2001.

Fontanillas P., Dépraz A., Giorgi M.S. and Perrin N. Nonshivering thermogenesis capacity associated to mitochondrial DNA haplotypes and gender in greater white-toothed shrew, *Crocidura russula*. *Mol. Ecol.*, **14**, 661-670, 2005.

Forster P. Ice ages and the mitochondrial DNA chronology of human dispersals. *Phil. Trans. R. Soc. Lond.*, **359**, 255-264, 2004.

Forster P., Harding R., Torroni A. and Bandelt H. Origin and evolution of Native American mtDNA variation: A reappraisal. *Am. J. Hum. Genet.*, **59**, 935-945, 1996.

Garstka H., Schmitt W.E., Schultz J., Sogl B., Silakowski B., Pérez-Martos A., Montoya J. and Wiesner R.J. Import of mitochondrial transcription factor A (TFAM) into rat liver mitochondria stimulates transcription of mitochondrial DNA. *Nucleic acids Res.*, **31**(17), 5039-5047, 2003.

Gascuel O. and Steel M. Neighbor-joining revealed. Oxford University Press, 2006.

Gaspari M., Falkenberg M., Larsson N-G. and Gustafsson C.M. The mitochondrial RNA polymerase contributes critically to promoter specificity in mammalian cells. *EMBO J.*, **23**(23), 4606-4614, 2004.

Gaucher S.P., Taylor S.W., Fahy E., Zhang B., Warnock D.E., Ghosh S.S. and Gibson W.W. Expanded coverage of the human heart mitochondrial proteome using multidimensional liquid chromatography and tandem mass spectrometry. *J. Proteome Res.*, **3,** 495-505, 2004.

Gibson B.W. The human mitochondrial proteome: oxidative stress, protein modifications and oxidative phosphorylation. *Int. J. Biochem. Cell Biol.*, **37**, 927-934, 2005.

Giles R.E., Blanc H., Cann H.M. and Wallace D.C. Maternal inheritance of human mitochondrial DNA. *Proc. Natl. Acad. Sci. USA*, **77**(11), 6715-6719, 1980.

Gonder M.K., Mortensen H.M., Reed F.A., De Sousa A. and Tishkoff S.A. Whole mtDNA genome sequence analysis of ancient African lineages. *Mol. Biol. Evol.*, **24**(3), 757-768, 2007.

Göringer H.U., Huazi K.A., Murgola E.J. and Dahlberg A.E. Mutations in 16S rRNA that affect UGA (stop codon) – directed translation termination. *Proc. Natl. Acad. Sci. USA*, **88**, 6603-6607, 1991.

Graham S.W., Olmstead R.G. and Barrett S.C.H. Rooting Phylogenetic Trees with Distant Outgroups: A case Study from the Commelinoid Monocots. *Mol. Biol. Evol.* **19**(10), 1769–1781, 2002.

Graven L., Passarino G., Semino O., Boursot P., Santachiara-Benerecetti S., Langaney A. and Excoffier L. Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka sample. *Mol. Biol. Evol.*, **12**, 334-335, 1995.

Gray M.W., Burger G. and Lang B.F. The origin and early evolution of mitochondria. *Genome Biol.*, **2**(6), I0I8.I-I0I8.5, 2001.

Gu X. and Zhang J. A simple method for estimating the parameter of substitution rate variation among sites. *Mol. Biol. Evol.*, **14**(11), 1106-1113, 1997.

Hagelberg E. Recombination or mutation rate heterogeneity? Implications for mitochondrial Eve. *TIG.*, **19**(2), 84-90, 2003.

Hamilton M.L., Guo Z., Fuller C.D., Remmen H.V., Ward W.F., Austad S.N., Troyer D.A., Thompson I. and Richardson A. A reliable assessment of 8-oxo-2-deoxyguanosine levels in nuclear and mitochondrial DNA using the sodium iodide method to isolate DNA. *Nucleic Acids Res.*, **29**(10), 2117-2126, 2001.

Harpending H. and Rogers A. Genetic perspectives on human origins and differentiation. *Annu. Rev. Genomics Hum. Genet.*, **1**, 361-385, 2000.

Harris M.H. and Thompson C.B. The role of the Bcl-2 family in the regulation of outer mitochondrial membrane permeability. *Cell Death Diff.*, **7**, 1182-1191, 2000.

Harrold M., Pistacchi S. and Amparo G. A new purification method for DNA sequencing reactions. Applied Biosystems, 850 Lincoln Center Drive, Forster City, California, 94404, USA, 2007.

Herrnstadt C., Elson J.L., Fahy E., Gwen P., Turnbull D.M., Anderson C., Ghosh S S., Olefsky J.M., Beal M.F., Davis R.E. and Howell N. Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. *Am. J. Hum. Genet.*, **70**, 1152-1171, 2002.

Hirano M., Davidson M. and DiMauro S. Mitochondria and the heart. *Curr. Opin. Cardiol.*, **16**, 201-210, 2001.

Holden C.J. Bantu language trees reflect the spread of farming across sub-Saharan Africa: a maximum-parsimony analysis. *Proc. R. Soc. Lond. B*, **269**, 793-799, 2002.

Horai S., Kondo R., Nakagawa-Hattori Y., Hayashi S., Sonoda S. and Tajima K. Peopling of the Americas, founded by four major lineages of mitochondrial DNA. *Mol. Biol. Evol.*, **10**(1), 23-47, 1993.

Horai S. and Hayasaka K. Intraspecific nucleotide sequence differences in the major noncoding region of human mitochondrial DNA. *Am. J. Hum. Genet.*, **46**, 828-842, 1990.

Horai S., Hayasaka K., Kondo R., Tsugane K. and Takahata N. Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. *Proc. Natl. Acad. Sci. USA*, **92**, 532-536, 1995.

Houshmand M. Mitochondrial DNA mutations, pathogenicity and inheritance. *Iran J. Biotechnol.*, **1**(1), 1-18, 2003.

Howell N.,Oostra R-J., Bolhuis P.A., Spruijt L., Clarke L.A., Mackey D.A., Preston G. and Herrnstadt C. Sequence Analysis of the Mitochondrial Genomes from Dutch Pedigrees with Leber Hereditary Optic Neuropathy. *Am. J. Hum. Genet.*, **72**, 1460-1469, 2003.

Howell N., Elson J.L., Turnbull D.M. and Herrnstadt C. African haplogroup L mtDNA sequences show violations of clock-like evolution. *Mol. Biol. Evol.*, **21**(10), 1843-1854, 2004.

Hudjashov G., Kivisild T., Underhill P.A., Endicott P., Sanchez J.J., Lin A.A., Shen P., Oefner P., Renfrew C., Villems R. and Forster P. Revealing the prehistoric settlement of Australia by Y chromosome and mtDNA analysis. *Proc. Nat.Acad. Sci. USA*, **104**(21), 8726-8730, 2007.

Huelsenbeck J.P. Phylogeny estimation and hypothesis testing using maximum likelihood. *Annu. Rev. Eco. Syst.*, **28**, 437-466, 1997.

Ichas F., Jouaville L.S. and Mazat J.P. Mitochondria are excitable organelles capable of generating and conveying electrical and calcium signals. *Cell*, **89**, 1145-1153, 1997.

Imperato P.J. Luo shields from Tanzania. *African Arts*, **16**(1), 73-96, 1982.

Ingman M., Kaessmann H., Pääbo S. and Gyllesten U. Mitochondrial genome variation and the origin of modern humans. *Nature*, **408**, 708-713, 2000.

Ingman M. and Gyllensten U. Mitochondrial genome variation and evolutionary history of Australian and New Guinean aborigines. *Genome Res.*, **13**, 1600-1606, 2003.

Ishikawa D., Yamamoto H., Tamura Y., Moritoh K. and Endo T. Two novel proteins in the mitochondrial outer membrane mediate β-barrel protein assembly. *J. Cell Biol.*, **166**(5), 621-627, 2004.

Jackson B.A., Wilson J.L., Kirbah S., Sidney S.S., Rosenberger J., Bassie L., Alie J.A.D., McLean D.C., Garvey W.T. and Ely B. Mitochondrial DNA genetic diversity among four ethnic groups in Sierra Leone. *Am. J. Phys. Anthropol.*, **128**, 156-163, 2005.

Jänsch L., Kruft V., Schmitz U.K. and Braun H-P. Unique composition of the preprotein translocase of the outer mitochondrial membrane from plants. *J. Biol. Chem.*, **273**(27), 17251-17257, 1998.

Jha N., Jurma O., Lalli G., Liu Y., Pettue E.H., Greenamyre J.T., Liu R-M., Forman H.J. and Andersen, J.K. Glutathione depletion in PC12 results in selective inhibition of mitochondrial complex I activity. *J. Biol. Chem.*, **275**(34), 26096-26101, 2000.

Jin L. and Su B. Natives or immigrants: modern human origins in East Asia. *Nat. Rev. Genet.*, **1**, 126-133, 2000.

Jobling M.A. In the name of the father: surnames and genetics. *Trends Genet.*, **17**(6), 353-357, 2001.

Johnson J.A., Bootman J.L., Evans W.E., Hudson R.A., Knoell D., Simmons L., Straubinger R.M. and Meyer S.M. Pharmacogenomics: A scientific revolution in pharmaceutical sciences and pharmacy practice. Report of the 2001-2002 Academic Affairs Committee. *Am. J. Pharm. Educ.*, **66**, 12S-15S, 2002.

Johnson A.A. and Johnson K.A. Exonuclease proofreading by human mitochondrial DNA polymerase. *J. Biol. Chem.*, **276**(44), 38097-38107, 2001.

Jorde L.B., Bamshad M. and Rogers A.R. Using mitochondrial and nuclear DNA markers to reconstruct human evolution. *Bioessays*, **20**, 126-136, 1998.

Jorde L.B., Bamshad M., Watkins W.S., Zenger R., Fraley A.E., Krakowaik P.A., Carpenter K.D., Soodyall H., Jenkins T. and Rogers A.R. Origins and Affinities of modern humans: a comparison of mitochondrial and human genetic data. *Am. J. Hum. Genet.*, **57**, 523-538, 1995.

Just R.S., Diegoli T.M., Saunier J.L., Irwin J.A. and Parsons T.J. Complete mitochondrial genome sequences for 265 African American and U.S. "Hispanic" individuals. *Forensic Sci. Int. Genet.*, **2**, e45-e48, 2008.

Kalaydjieva L., Calafell F., Jobling M.A., Angelicheva D., De Knijff P., Rosser Z.H., Hurles M.E., Underhill P., Tournev I., Marushiakova E. and Popov V. Patterns of inter- and intra-group genetic diversity in the Viax Roma as revealed by Y chromosome and mitochondrial DNA lineages. *Eur. J. Hum. Genet.*, **9**, 97-104, 2001.

Kaneda H., Hayashi J., Takahama S., Taya C., Lindayl K.F. and Yonekawa H. Elimination of paternal mitochondrial DNA in intraspecific crosses during early mouse embryogenesis. *Proc. Natl. Acad. Sci. USA*, **92**, 4542-4546, 1995.

Kaput J., Ordovas J.M., Ferguson L., van Ommen B., Rodriguez R.L., Allen L., Ames B.N., Dawson K., German B., Krauss R., Malyj W., Archer M.C., Barnes S., Bartholomew A., Birk R., van Bladeren P., Bradford K.J., Brown K.H., Caetano R., Castle D., Chadwick R., Clarke S., Clément K., Cooney C.A., Corella D., Manica da Cruz I.B., Daniel H., Duster T., Ebbesson S.O.E., Elliott R., Fairweather-Tait S., Felton J., Fenech M., Finly J.W., Fogg-Johnson N., Gill-Garrison R., Gibney M.J., Gillies P.J., Gustafsson J-A., Hartman IV J.L., He L., Hwang J-K., Jais J-P., Jang Y., Joost H., Junien C., Kanter M., Kibbe W.A., Koletzko B., Korf B.R., Kornman K., Krempin D.W., Langin D., Lauren D.R., Lee J.H., Leveille G.A., Lin S-J., Mathers J., Mayne M., McNabb W., Milner J.A., Morgan P., Muller M., Nikolsky Y., van der Ouderaa F., Park T., Pensel N., Perez-Jimenez F., Poutanen K., Roberts M., Saris W.H.M., Schuster G., Shelling A.N., Simopoulos A.P., Southon S., Tai E.S., Towne B., Trayhurn P., Uauy R., Visek W.J., Warden C., Weiss R., Wiencke J., Winkler J., Wolff G.L., Zhao-Wilson X. and Zucker J-D. The case for strategic international alliances to harness nutritional genomics for public and personal health. *Br. J. Nutr.*, **94**, 623-632, 2005.

Karlin S., Brocchieri L., Mrázek J., Campbell A.M. and Spormann A.M. A chimeric eukaryotic ancestry of mitochondria and primitive eukaryotes. *Proc. Natl. Acad. Sci. USA*, **96**, 9190-9195, 1999.

Kato T., Kunugi H., Nanko S. and Kato N. *J. Affective Disorders*, **62**, 151-164, 2001.

Kawakita A., Sota T., Ascher J.S., Ito M., Tanaka H. and Kat M. Evaluation and forensic utility of alignment gaps within intron sequences of three nuclear genes in bumble bees (Bombus). *Mol. Biol. Evol.*, **20**(1), 87-92, 2003.

Kieleczawa J. DNA sequencing: Optimising the process and analysis. Jones and Bartlett publishers, 40 Tall Pine Drive, Sudbury, MA01776, 978-443-5000, USA, 2005.

Kivisild T., Reidla M., Metspalu E., Rosa A., Brehm A., Pennarun E., Parik J., Geberwihot T., Usanga E. and Villems R. Ethiopian mitochondrial DNA heritage: Tracking gene flow in and around the gate of tears. *Am. J. Hum. Genet.*, **75**, 752-770, 2004.

Kivisild T., Shen P., Wall D.P., Do B., Sung R., Davis K., Passarino G., Underhill P.A., Scharfe C., Torroni A., Scozzari R., Modiano D., Coppa A., De Knijff P., Feldman M., Cavalli-Sforza L.L. and Oefner P.J. The role of selection in the evolution of human mitochondrial genomes. *Genetics*, **172**, 373-387, 2006.

Kivisild T., Tolk H-V., Parik J., Wang Y., Papiha S.S., Bandelt H-J. and Villems R. The emerging limbs and twigs of the East Asian mtDNA tree. *Mol. Biol. Evol*., **19**(10), 1737-1751, 2002.

Kong Q., Yao Y., Sun C., Bandelt H., Zhu C. and Zhang Y. Phylogeny of East Asian mitochondrial DNA lineages inferred from complete sequences. *Am. J. Hum. Genet,*. **73**, 671-676, 2003.

Knight A., Underhill P.A., Mortenson H.M., Zhivotovsky L.A., Lin A.A., Henn B.M., Louis D., Ruhlen M. and Mountain J.L. African Y chromosome and mtDNA divergence provides insight into the history of click languages. *Curr. Biol*., **13**, 464-473, 2003.

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

Kumar S., Tamura K. and Nei M. MEGA: molecular evolutionary genetics analysis, version 1.01. The Pennsylvania State University, University Park, PA 16802, 1993.

Kumar S., Tamura K. and Nei M. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinform.*, **5**(2), 150-163, 2004.

Kurland C.G. and Andersson S.G.E. Origin and evolution of the mitochondrial proteome. *Microbiol. Mol. Biol. Rev.*, **64**(4), 786-820, 2000.

Larsson N. and Clayton D.A. Molecular genetic aspects of human mitochondrial disorders. *Annu. Rev. Genet.,* **29,** 151-178, 1995.

Legros F., Chatzoglou E., Frachon P., Ogier de Baulny H., Laforét P., Jardel C., Godnot C. and Lombès A. Functional characterisation of novel mutations in the human cytchrome b gene. *Eur. J. Hum. Genet*., **9**, 510-518, 2001.

Li J. and Su B. Natives or immigrants: Modern human origin in East Asia. *Nat. Rev. Genet*., **1**, 126-133, 2000.

Li W-H. and Sadler L.A. Low nucleotide diversity in man. *Genetics*, **129**, 513-523, 1991.

Liu H., Prugnolle F., Manica A. and Balloux F. A geographically explicit model of worldwide human-settlement history. *Am. J. Hum. Genet*., **79**, 230-237, 2006.

Liu Y., Fiskum G. and Schubert D. Generation of reactive species by the mitochondrial electron transport chain. *J. Neoruchem.*, **80**, 780-787, 2002.

Liu V.W.S., Shi H.H., Cheung A.N.Y., Chiu P.M., Leung T.W., Nagley P., Wong L.C. and Ngan H.Y.S. High incidence of somatic mitochondrial DNA mutations in human ovarian carcinomas. *Cancer Res*., **61**, 5998-6001, 2001.

Livshits G. and Nei M. Relationship between intrapopulational and interpopulational genetic diversity in man. *Ann. Hum. Biol*., **17**(6), 501-513, 1990.

Luft R. The development of mitochondrial medicine. *Proc. Natl. Acad. Sci.* USA, **91**, 8731-8738, 1994.

Macaulay V., Richards M., Hickey E., Vega E., Cruciani F., Guida V., Scozzari R., Bonne-Tamir B., Sykes B. and Torroni A. The emerging tree of West Eurasian mtDNAs: a synthesis of control region sequences and RFLP. *Am. J. Hum. Genet*. **64**, 232-239, 1999.

Macaulay V., Hill C., Achilli A., Rengo C., Clarke D., Meehan W., Blackburn J., Semino O., Scozzari R., Cruciani F., Taha A., Shaari N.K., Raja J.M., Ismail P., Zainuddin Z., Goodwin W., Bulbeck D., Bandelt H-J., Oppenheimer S., Torroni A. and Richards M. Singel, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science*, **308,** 1034-1036, 2005.

Mackey D. and Howell N. A variant of Leber heredirary optic neuropathy characterised by a recovery of vision and by an unusual mitochondrial genetic aetiology. *Am. J. Hum. Genet*., **51**, 1218-1228, 1992.

Marin-Garcia J., Goldenthal M.J., Ananthakrishnan R. and Pierpont M.E. The complete sequence of mtDNA genes in idiopathic dilated cardiomyopathy shows novel missense and tRNA mutations. *J. Cardiac Fail*., **6**(4), 321-329, 2000.

Martin W. and Herrmann R.G. Gene transfer from organelles to the nucleus: how much, what happens, and why? *Plant Physio*l., **118**, 9-17, 1998.

Meirelles F.V. and Smith L.C. Mitochondrial genotype segregation during preimplantation development in mouth heteroplasmic embryos. *Genetics*, **148**, 877-883, 1998.

Melton T. and Nelson K. Forensic mitochondrial DNA analysis: two years of commercial casework experience in the United States. *Croat. Med. J*., **42**(3), 298-303, 2001.

Mills D.R. and Kramer F.R. Structure-independent nucleotide sequence analysis (inosine substitution/nucleoside triphosphate terminators/slab gel electrophoresis/QfB replicase/MDV-1 RNA). *Proc. Natl. Acad. Sci. USA*, **76**(5), 2232-2235, 1979.

Minorsky P.V. The hot and the classic. Plant thermogenesis and thermoregulation. *Plant Physiol*., **132**, 25-26, 2003.

Mishmar D., Ruiz-Pesini E., Golik P., Macaulay V., Clark A.G., Hosseini S., Brandon M., Easley K., Chen E., Brown M.D., Sukernik R.I., Olckers A. and Wallace D.C. Natural selection shaped regional mtDNA variation in humans. *Proc. Natl. Acad. Sci.* USA, **100**(1), 171-176, 2003.

Mitchell P. Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. *Nature*, **191**, 144-148, 1961.

Moilanen J.S., Finnilä S. and Majamaa K. Lineage-specific selection in human mtDNA: Lack of polymorphisms in a segment of MTND5 gene in haplogroup J. *Mol. Biol. Evol.*, **20**(12), 2132-2142, 2003.

Moraes C.T., Ciacci F., Bonilla E., Jansen C., Hirano M., Rao N., Lovelace R.E., Rowland L.P., Schon E.A. and DiMauro S. Two novel pathogenic mitochondrial DNA mutations affecting organelle number and protein synthesis. Is the $tRNA^{Leu(UUR)}$ gene and etiological hotspot? *J. Clinic. Invest.*, **92**, 2906-2915, 1993.

Mortola J.P. and Naso L. Thermogenesis in newborn rats after prenatal or postnatal hypoxia. *J. Appl. Physiol.*, **85**(1), 84-90, 1998.

Motulsky A.G. and Qi M. Pharmacogenetics, pharmacogenomics and ecogenetics. *J. Zhejiang Univ. SCIENCE B*, **7**(2), 169-170, 2006.

Mullis K., Faloona F., Scharf S., Saiki R., Horn G. and Erlich H. Specific enzymatic amplification of DNA *In Vitro*: The polymerase chain reaction. *Quant. Biol.*, **51**, 263-273, 1986.

NC-IUB. Nomenclature committee of the International Union of Biochemistry (NC-IUB). Nomenclature for incompletely specified bases in nucleic acid sequences. Recommendations 1984. *Eur. J. Biochem.*, **150**, 1-5, 1985.

Nei M. Genetic support for the out-of-Africa theory of human evolution. *Proc. Natl. Acad. Sci.* USA, **92**, 6720-6722, 1995.

Nei M. and Roychoudhury A.K. Evolutionary relationships of human populations on a global scale. *Mol. Biol. Evol.*, **10**(5), 927-943, 1993.

Nishino I., Komatsu M., Kodama S., Horai S., Nonaka I. and Goto Y. The 3260 mutation in mitochondrial DNA can cause mitochondrial myopathy, encephalopathy, lactic acidosis, and stroke-like episodes (MELAS). *Muscle Nerve*, **19**, 1603-1604, 1996.

Nomura K., Imai H., Koumura T., Arai M. and Nakagawa Y. Mitochondrial phospholipids hydroperoxide glutathione peroxidase suppresses apoptosis mediated by a mitochondrial death pathway. *J. Biol. Chem.*, **274**(41), 29294-29302, 1999.

Ozawa T., Katsumata K., Hayakawa M., Tanaka M., Sugiyama S., Tanaka T., Itoyama S., Nunoda S. and Sekiguchi M. Genotype and phenotype of severe cardiomyopathy: a recipient of heart transplantation and the genetic control. *Biochem. Biophys. Res. Commun.*, **207**(2), 613-620, 1995.

Ozawa M., Nishino I., Horai S., Nonaka I. and Goto Y. Myoclonus epilepsy associated with ragged-red fibres: A G-to-A mutation at nucleotide pair 8363 in mitochondrial $tRNA^{Lys}$ in two families. *Muscle Nerve*, **20**, 271-278, 1997.

Palanichamy M.G., Sun C., Agrawal S., Bandelt H-J., Kong Q-P., Khan F., Wang C-Y., Chaudhuri T.K., Palla V. and Zhang Y-P. Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia. *Am. J. Hum. Genet.*, **75**, 966-978, 2004.

Parra E.J., Kittles R.A. and Shriver M.D. Implications of correlations between skin colour and genetic ancestry for biomedical research. *Nat. Genet.*, 36(11), S54-S60, 2004.

Parsons T.J. and Coble M.D. Increasing the forensic discrimination of mitochondrial DNA testing through analysis of the entire mitochondrial genome. *Croat. Med. J.*, **42**(3), 304-309, 2001.

Passarino G., Semino O., Quintana-Murci L., Excoffier L., Hammer M. and Santachiara-Benerecetti A.S. Different genetic components in the Ethiopian population, identified by mtDNA and Y-chromosome polymorphisms. *Am. J. Hum. Genet.*, **62**, 420-434, 1998.

Piccoli C., Boffoli D. and Capitanio N. Comparative analysis of mitochondria selective dyes in different cell types detected by confocal laser scanning microscopy: methods and applications. *Current Issues on Multidisciplinary Miscroscopy Research and Education*, 2004.

Pinhasi R., Fort J. and Ammerman A.J. Tracing the origin and spread of agriculture in Europe. *PloS Biol.*, **3**(12), 2220-2228, 2005.

Poetsch M., Wittig H., Krause D. and Lignitz E. The impact of mtDNA analysis between positions nt8306 and nt9021 for forensic casework. *Mitochondrion* **3,** 133-137, 2003.

Prezant T.R., Agapian J.V., Bolman M.C., Bu X., Oztas S., Qiu W., Arnos K.S., Cortopassi G.A., Jaber L., Rotter J.I., Shohat M. and Fischel-Ghodsian N. Mitochondrial ribosomal RNA mutation associated with both antibiotic-induced and non-syndromic deafness. *Nat. Genet.*, **4**, 289-294, 1993.

Qu J., Li R., Tong Y., Hu Y., Zhou X., Qian Y., Lu F. and Guan M-X. Only male matrilineal relatives with Leber's hereditary optic neuropathy in a large Chinese family carrying the mitochondrial DNA G11778A mutation. *Biochem. Biophys. Res. Comm.*, **328**, 1139-1145, 2005.

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

Quintana-Murci L., Chaix R., Wells R.S., Behar D.M., Sayar H., Scozzari R., Rengo C., Al-Zahery N., Semino O., Santachiara-Benerecetti, Coppa A., Ayub Q., Mohyuddin A., Tyler-Smith C., Mehdi Q., Torroni A. and McElreavey K. Where West meets East: The complex mtDNA landscape of the Southwest and Central Asian corridor. *Am. J. Hum. Genet.*, **74**, 827-845, 2004.

Rando J.C., Pinto F., Gonzalez A.M., Hernandez M., Larruga J.M., Cabrera V.M. and Bandelt H-J. Mitochondrial DNA analysis of Northwest African populations reveals genetic exchanges with European, Near-Eastern, and sub-Saharan populations. *Ann. Hum. Genet.*, **62**, 531-550, 1998.

Ray N., Currat M., Berthier P. and Excoffier L. Recovering the geographic origin of early modern humans by realistic and spatially explicit simulations. *Genome Res.*, **15**, 1161-1167, 2005.

Redd A.J., Roberts-Thompson J., Karafet T., Bamshad M., Jorde L.B., Naidu J.M., Walsh B. and Hammer M.F. Gene flow from the Indian subcontinent to Australia: Evidence from the Y chromosome. *Curr. Biol.*, **12**, 673-677, 2002.

Redd A.J. and Stoneking M. Peopling of Sahul: mtDNA variation in Aboriginal Australian and Papua New Guinean populations. *Am. J. Hum. Genet.*, **65**, 808-828, 1999.

Redd A.J., Takezaki N., Sherry S.T., McGarvey S.T., Sofro A.S.M. and Stoneking M. Evolutionary history of the COII/tRNALys intergenic 9 base pair deletion in human mitochondrial DNAs from the pacific. *Mol. Biol. Evol.*, **12**(4), 604-615, 1995.

Reich E. and Luck D.J.L. Replication and inheritance of mitochondrial DNA. *Proc. Natl. Acad. Sci. USA*, 55, 1600-1608, 1966.

Reider M.J., Taylor S.L., Tobe V.O. and Nickerson D.A. Automating the identification of DNA variations using quality-based fluorescence re-sequencing: analysis of the human mitochondrial genome. *Nucleic Acids Res.*, **26**(4), 967-973, 1998.

Relethford J.H. Ancient DNA and the origin of modern humans. *Proc. Natl. Acad. Sci. USA*, **98**(2), 390-391, 2001.

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., Calí 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. Tracing European founder lineages in the Near Eastern mtDNA pool. *Am. J. Hum. Genet.*, **67**, 1251-1276, 2000.

Richards M., Rengo C., Cruciani F., Gratrix F., Wilson J.F., Scozzari R., Macaulay V. and Torroni A. Extensive female-mediated gene flow from sub-Saharan Africa into Near Eastern Arab populations. *Am. J. Hum. Genet.*, **72**, 1058-1064, 2003.

Rickards O., Martinez-Labarga C., Lum J.K., De Stefano G.F. and Cann R.L. mtDNA history of the Capaya Amerinds of Equador: Detection of additional founding lineages for the Native American populations. *Am. J. Hum. Genet.*, **65**, 519-530, 1999.

Rippe C., Berger K., Böiers C., Ricquier D. and Erlanson-Albertsson C. Effect of high-fat diet, surrounding temperature, and enterostatin on uncoupling protein gene expression. *Am. J. Physiol. Endocrinol. Metab.*, **279**, E293-E300, 2000.

Rocha H., Flores C., Campos Y., Arenas J., Vilarinho L., Santorelli F.M. and Torroni A. About the "pathological" role of the mtDNA T3308C mutation …. *Am. J. Hum. Genet.*, **65**, 1457-1459, 1999.

Rogers A.R. and Jorde L.B. Genetic evidence on modern human origins. *Hum. Biol.*, **67**(1), 1-36, 1995.

Rogers A.R. and Harpending H. Population growth makes waves in the distribution of pairwise genetic differences. *Mol. Biol. Evol.*, **9**(3), 552-569, 1992.

Rosenblum B.B., Lee L.G., Spurgeon S.L., Khan S.H., Menchen S.M., Heiner C.R. and Chen S.M. New dye-labelled terminators for improved DNA sequencing patterns. Nucleic Acids Res., 25(22), 4500-4504, 1997.

Ruiz-Pesini E., Lapena A.C., Diez-Sanchez C., Perez-Martos A., Montoya J., Alvarez E., Diaz M., Urries A., Montoro L., Lopez-Perez M.J. and Enriquez J.A. Human mtDNA haplogroups associated with high or reduced spermatozoa motility. *Am. J. Hum. Genet.*, **67**(3), 682-696, 2000.

Ruiz-Pesini E., Mishmar D., Brandon M., Procaccio V. and Wallace D.C. Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science*, **303**, 223-226, 2004.

Ruiz-Pesini E., Lott M.T., Procaccio V., Poole J.C., Brandon M.C., Mishmar D., Yi C., Kreuziger J., Baldi P. and Wallace D.C. An enhanced mitomap with a global mtDNA mutational phylogeny. *Nucleic Acids Res.*, **35**, D823-D828, 2007.

Rzhetsky A. and Nei M. A simple method for estimating and testing minimum-evolution trees. *Mol. Biol. Evol.*, **9**(5), 945-967, 1992.

Saitou N. and Imanishi T. Relative efficiencies of the Fitch-Margoliash, Maximum-parsimony, maximum-likelihood, minimum-evolution and Neighbour-joining methods of phylogenetic tree construction in obtaining the correct tree. *Mol. Biol. Evol.*, **6**(5), 514-525, 1989.

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

Salas A., Bandelt H-J., Macaulay V. and Richards M.B. Phylogeographic investigations: The role of trees in forensic genetics. *Forensic Sci. Int.*, **168**, 1-13, 2007.

Salas A., Carracedo A., Richards M., and Macaulay V. Charting the Ancestry of African Americans. *Am. J. Hum. Genet.*, **77**, 676-680, 2005.

Salas A., Fé T.D., Sobrino B., Lareu M.V. and Carracedo A. Mitochondrial DNA variability patterns in Southeast Africa and forensic applications. *Int. Cong. Ser.*, **1239**, 541-545, 2003.

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

Salas A., Richards M., Lareu M., Scozzari R., Copper A., Torroni A., Macaulay V. and Carracedo A. The African diaspora: mitochondrial DNA and the Atlantic slave trade. *Am. J. Hum. Genet.*, **74**, 454-465, 2004a.

Salas A., Torroni A., Richards M., Quintana-Murci L., Hill C., Macaulay V. and Carracedo Á. *Am. J. Hum. Genet.*, **75**, 524-526, 2004b. Reply to Bortoloni M.C., Da Silva W.A., Zago M.A., Elion J., Krishnamoorthy J., Goncalves V.F. and Pena S.D.J. The phylogeography of mitochondrial DNA Haplogroup L3g in Africa and the Atlantic slave trade. *Am. J. Hum. Genet.*,**75**, 523-524, 2004b.

Sanger F., Nickelin S. and Coulson A.R. DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA*, 74(12), 5463-5467, 1977.

Saitou N. and Imanishi T. Relative efficiencies of the Fitch-Margoliash, maximum-parsimony, maximum-likelihood, minimum-evolution, and neighbour-joining methods of phylogenetic tree construction in obtaining the correct tree. *Mol. Biol. Evol.*, **6**(5), 514-525, 1989.

Saxena R., De Bakker P.I.W., Singer K., Mootha V., Burtt N., Hirschhorn J.N., Gaudet D., Isomaa B., Daly M.J., Groop L., Ardlie K.G. and Altshuler D. Comprehensive association testing of common mitochondrial DNA variation in metabolic disease. *Am.J. Hum. Genet.*, **79**, 54-61, 2006.

Schneider S.S., DeGrandi-Hoffman G. and Smith D.R. The African bee: factors contributing to a successful biological invasion. *Annu. Rev. Entomol.*, **49**, 351-376, 2004.

Scholte H.R. The biochemical basis of mitochondrial diseases. *J. Bioenerg. Biomemb.*, **20**(2), 161-191, 1988.

Schon E.A. Mitochondria in Mitochondrial DNA in Human pathology, edited by DiMauro S. and Wallace D.C. Raven Press, Ltd, New York, 1-7, 1993.

Schwartz M. and Vissing J. Paternal inheritance of mitochondrial DNA. *N. Eng. J. Med.*, **347**(8), 576-580, 2002.

Searcy D.G. Metabolic integration during the evolutionary origin of mitochondria. *Cell Res.*, **13**(4), 229-238, 2003.

Segel I.H. Biochemical calculations. 2 ed., **5**, 324-346, 1976. John Wiley & Sons, New York, USA.

Seielstad M.T., Minch E. and Cavalli-Sforza L.L Genetic evidence for a higher female migration rate in humans. *Nat. Genet.*, **20**, 278-280, 1998.

Senior A.E. ATP synthesis by oxidative phosphorylation. *Physiol. Rev.*, **68**(1), 177-231, 1988.

Seymour R.S. Dynamics and precision of thermoregulatory responses of eastern skunk cabbage *Symplocarpous foetidus*. *Plant Cell Environ.*, **27**(8), 1014, 2004.

Shadel G.S. and Clayton D.A. Mitochondrial transcription initiation. *J. Biol. Chem.*, **268**(2), 16083-16086, 1993.

Shanske A.L., Shanske S. and DiMauro S. The other human genome. *Arch. Pediatr. Adolesc. Med.*, **155**, 1210-1216, 2001.

Shen P., Lavi T., Kivisild T., Chou V., Sengun D., Gefel D., Shpirer I., Woolf E., Hillel J., Feldman M.W. and Oefner P.J. Reconstruction of patrilineages and matrilineages of Samaritans and other Israeli populations from Y-chromosome and mitochondrial DNA sequence variation. *Hum. Mutat.*, **24**, 248-260, 2004.

Shields G.F., Schmiechen A.M., Frazier B.L., Redd A., Voevoda M.I., Reed J.K. and Ward R.H. mtDNA sequences suggest a recent evolutionary divergence for Beringian and Northern North American populations. *Am. J. Hum. Genet.*, **53**, 549-562, 1993.

Silva W.A., Bonatto S.L., Holanda A.J., Ribeiro-dos-Santos A.K., Paixão B.M., Goldman G.H., Abe-Sandes K., Rodriguez-Delfin L., Barbosa M., Parcó-Larson M.L., Petzl-Erler M.L., Valente V., Santos S.E.B. and Zago M.A. Mitochondrial genome diversity of Native Americans supports a single early entry of founder populations into America. *Am. J. Hum. Genet.*, **71**, 187-192, 2002.

Sitnikova T., Rzhetsky A. and Nei M. Interior-branch and bootstrap tests of phylogenetic trees. Mol. Biol. Evol., 12(2), 319-333, 1995.

Sogo L.F. and Yaffe M.P. Regulation of mitochondrial morphology and inheritance by Mdm10p, a protein of the mitochondrial outer membrane, *J. Cell Biol.*, **126**(6), 1361-1373, 1994.

Soodyall H. and Jenkins T. Mitochondrial DNA polymorphisms in Negroid populations from Namibia: new light on the origins of the Dama, Herero and Ambo. *Ann. Hum. Biol.*, **20**(5), 477-485, 1993.

Springer M.S., Hollar L.J. and Burk A. Compensatory substitutions and the evolution of the mitochondrial 12S rRNA gene in mammals. *Mol. Biol. Evol.*, **12**(6), 1138-1150, 1995.

Ssekamwa J.C. A sketch map history of East Africa. 4th Ed. T & E publishers, Kampala, Uganda, 1994.

Starikovskaya Y.B., Sukernik R.I., Schurr T.G., Kogelnik A.M. and Wallace D.C. mtDNA diversity in Chukchi and Siberian Eskimos: Implications for the genetic history of ancient Beringia and the peopling of the New World. *Am. J. Hum. Genet*. **63**, 1473-1491, 1998.

Sternberg D., Danan C., Lombès A., Laforêt P., Girodon E., Goossens M. and Amselem S. Exhaustive scanning approach to screen all the mitochondrial tRNA genes for mutations and its application to the investigation of 35 independent patients with mitochondrial disorders. *Hum. Mol. Genet*., **7**(1), 33-42, 1998.

Stoneking M., Hedgecock D., Higuchi R.G., Vigilant L. and Erlich H.A. Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. Am. J. Hum. Genet., 48, 370-382, 1991.

Strandberg A.K.K. and Salter L.A. A comparison of methods for estimating the transition:transversion ratio from DNA sequences, *Mol. Phylogenet. Evol*., **32**, 495-503, 2004.

Stringer C. Modern human origins: progress and prospects. *Phil. Trans. R. Soc. Lond. B*., **357**, 563-579, 2002.

Stringer C.B. and Andrews P. Genetic and fossil evidence for the origin of modern humans. *Science*, **239**, 1263-1268, 1988.

Stryer L. Biochemistry. 3rd Ed. W.H. Freeman and Company, New York, 1988.

Sugimoto N., Nakano S., Yoneyama M. and Honda K. Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes. *Nucleic Acids Res*., **24**, 4501-4505, 1996.

Susin S.A., Zamzami N. and Kroemer G. Mitochondria as regulators of apoptosis: doubt no more. *Biochim. Biophys. Acta,* **1366**, 151-165, 1998.

Sutovsky P., Moreno R.D., Ramalho-Santos J., Dominko T., Simerly C. and Schatten G. Ubiquitin tag for sperm mitochondria. *Nature*, **402**, 371-372, 1999.

Sykes B. and Irven C. Surnames and the Y chromosome. *Am. J. Hum. Genet*., **66**, 1417-1419, 2000.

Taanman J. The mitochondrial genome: structure, transcription, translation and replication. *Biochim. Biophys. Acta*, **1410**, 103-123, 1999.

Takahata N., Lee S-H. and Satta Y. Testing multiregionality of modern human origins. *Mol. Biol. Evol*., **18**(2), 172-183, 2001.

Tambets K., Rootsi S., Kivisild T., Help H., Serk P., Loogväli E-L., Tolk H-V., Reidla M., Metspalu E., Pliss L., Balanovsky O., Pshenichnov A., Balanovska E., Gubina M., Zhadanov S., Osipova L., Damba L., Voevoda M., Kutuev I., Bermisheva M., Khusnutdinova E., Gusar V., Grechanina E., Parik J., Pennarun E., Richard C., Chaventre A., Moisan J-P., Barać L., Peričić M., Rudan P., Tervić R., Mikerezi I., Krumina A., Baumanis V., Koziel S., Rickards O., De Stefano G.F., Anagnou N., Pappa K.I., Michalodimitrakis E., Ferák V., Füredi S., Komel R., Beckman L. and Villems R. The western and eastern roots of the Saami - the story of genetic "outliers" told by mitochondrial DNA and Y chromosomes. *Am. J. Hum. Genet*., **74**, 661-682, 2004.

Tan D-J., Bai R-K. and Wong L-J.C. Comprehensive scanning of somatic mitochondrial DNA mutations in breast cancer. *Cancer Res*., **62**, 972-976, 2002.

Tanaka M., Cabrera V.M., González A.M., Larruga J.M., Takeyasu T., Fuku N., Guo L-J., Hirose R., Fujita Y., Kurata M., Shinoda K-I., Umetsu K., Yamada Y., Oshida Y., Sato Y., Hattori N., Mizuno Y., Arai Y., Hirose N., Ohta S., Ogawa O., Tanaka Y., Kawamori R., Shamoto-Nagai M., Maruyama W., Shimokata H., Suzuki R. and Shimodaira H. *Genome Res*., **14**, 1832-1850, 2004.

Tarazona-Santos E., Carvalho-Silva D.R., Pettener D., Luiselli D., De Stefano G.F., Labarga C.M., Richards O., Tyler-Smith C., Pena S.D.J. and Santos F.R. Genetic differentiation in South Amerindians is related to environmental and cultural diversity: evidence from the Y chromosome. *Am. J. Hum. Genet*., **68**, 1485-1496, 2001.

Tate S. and Goldstein D.B. Will tomorrow's medicines work for everyone? *Nat. Genet.,* **36**(11), S34-S42, 2004.

Tateno Y., Takezaki N. and Nei M. Relative efficiencies of the maximum-likelihood, neighbour-joining, and maximum parsimony methods when substitution rate varies with site. *Mol. Biol. Evol*., **11**(2), 261-277, 1994.

Taylor S.W., Fahy E., Zhang B., Glenn G.M., Warnock D.E., Wiley S., Murphy A.N., Gaucher S.P., Capaldi R.A., Gibson B.W. and Ghosh S.S. Characterisation of the human heart mitochondrial proteome, *Nat. Biotechnol*., **21**, 281-286, 2003.

Templeton A.R. The "Eve" hypothesis: a genetic critic and reanalysis. *Am. Anthropol*., **95**, 51-72, 1993.

Tishkoff S. and Williams S.M. Genetic analysis of African populations: human evolution and complex disease. *Genetics*, **3**, 611-621, 2002.

Tishkoff S.A., Reed F.A., Friedlaender F.R., Ehret C., Ranciaro A., Froment A., Hirbo J.B., Awomoyi A.A., Bodo J-M., Doumbo O., Ibrahim M., Juma A.T., Kotze M.J., Lema G., Moore J.H., Mortensen H., Nyambo T.B., Omar S.A., Powell K., Pretorius G.S., Smith M.W., Thera M.A., Wambebe C., Weber J.L. and Williams S.M. The genetic structure and history of Africans and African Americans. *Science*, **324**(5930), 1035-1044, 2009.

Torroni A., Achilli A., Macaulay V., Richards M. and Bandelt H-J. Harvesting the fruit of the human mtDNA tree. *Trends Genet.*, **22**(6), 339-345, 2006.

Torroni A., Rengo C., Guida V., Cruciani F., Sellito D., Coppa A., Calderon F.L., Simionati B., Valle G., Richards M., Macaulay V. and Scozarri R. Do the four clades of the mtDNA Haplogroup L2 evolve at different rates? *Am. J. Hum. Genet.*,**69**, 1348-1356, 2001.

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

Torroni A., Neel J.V., Barrantes R., Schurr T.G. and Wallace D.C. A mitochondrial DNA "clock" for the Amerinds and its implication for timing their entry into North America. *Proc. Natl. Acad. Sci. USA*, **91**(3), 1158-1162, 1994.

Torroni A., Schurr T.G., Cabell M.F., Brown M.D., Neel J.V., Larsen M., Smith D.G., Vullo C.M. and Wallace D.C. Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am. J. Hum. Genet.*, **53**, 563-590, 1993.

Torroni A. and Wallace D.C. Mitochondrial DNA variation in human populations and implications for detection of mitochondrial DNA mutations of pathological significance. *J. Bioenerg. Biomembr.*, **26**(3), 261-271,.1994.

Tuimala J. Using clustalX for multiple sequence alignment, 2004.

Tuimala J. A primer for phylogenetic analysis using the PHYLIP package. 5th Ed., 2006.

Turrens J.F. Mitochondrial formation of reactive oxygen species. *J. Physiol.*, **552**(2), 335-344, 2003.

Turrens J.F. and Boveris A. Generation of superoxide anion by the NADH dehydrogenase of bovine heart mitochondria. *Biochem. J.*, **191**, 421-427, 1980.

Uusimaa J., Finnilä S., Remes A.M., Rantala H., Vainionpää L., Hassinen I.E. and Majamaa K. Molecular epidemiology of childhood mitochondrial encephalomyopathies in a Finnish population: sequence analysis of entire mtDNA of 17 children reveals heteroplasmic mutations in $tRNA^{Arg}$, $tRNA^{Glu}$, and $tRNA^{Leu(UUR)}$ genes. *Pediatrics*, **114**, 443-450, 2004.

Van Brummelen A.C. Molecular analysis of the mitochondrial genome in South African patients with suspected mitochondrial disorders [dissertation]. North-West University, South Africa, 2003.

Vigilant L., Pennington R., Harpending H., Kocher T.D. and Wilson A.C. Mitochondrial DNA sequences in single hairs from a southern African population. *Proc. Natl. Acad. Sci. USA*, **86**, 9350-9354, 1989.

Viljoen G.J., Nel L.H. and Crowther J.R. Molecular diagnostics PCR handbook. Springer, Box 17, 3300 AA Dordrecht, The Netherlands, 2005.

Vilmi T., Moilanen J.S. Finnilä S. and Majamaa K. Sequence variation in the tRNA genes of human mitochondrial DNA. *J. Mol. Evol.*, **60**, 587-597, 2005.

Vives-Bauza C., Andreu A.L., Manifredi G., Beal M.F., Janetzsky B., Gruenwald T.H. and Lin M.T. Sequence analysis of the entire mitochondrial genome in Parkinson's disease. *Biochem. Biophys. Res. Comm.*, **290**, 1593-1601, 2002.

Wagner A.M., Wagner M.J. and Moore A.L. *In vivo* ubiquinone reduction levels during thermogenesis in Araceae. *Plant Physiol.*, **117**, 1501-1506, 1998.

Wairama G.B. Uganda: The marginalisation of minorities. A Minority Rights Group International Report, 2001. The British Library, UK.

Walker J.E., Lutter R., Dupuis A. and Runswick M.J. Identification of the subunits of F_1F_0-ATPase from bovine heart mitochondria. *Biochemistry*, **30**, 5369-5378, 1991.

Wallace D.C. Diseases of the mitochondrial DNA. *Annu. Rev. Biochem.*, **61**, 1175-1212, 1992.

Wallace D.C. Mitochondrial DNA sequence variation in human evolution and disease. *Proc. Natl. Acad. Sci. USA*, **91**, 8739-8746, 1994.

Wallace D.C. Mitochondrial DNA Variation in Human Evolution, Degenerative Disease, and Aging, *Am. J. Hum. Genet.*, **57**, 201-223, 1995.

Wallace D.C. A set of polymorphisms for haplogrouping people. Personal communication, 2004.

Wallace D.C., Brown M.D. and Lott M.T. Mitochondrial DNA variation in human evolution and disease. *Gene*, **238**, 211-230, 1999.

Wang X. The expanding role of mitochondria in apoptosis. *Genes & Dev.*, **15**, 2922-2933, 2001.

Ward R.H., Redd A., Valencia D., Frazier B. and Pääbo S. Genetic and linguistic differentiation in the Americas. *Proc. Natl. Acad. Sci. USA*, **90**, 10663-10667, 1993.

Watson E., Bauer K., Aman R., Weiss G., von Haesler A. and Pääbo S. mtDNA sequence diversity in Africa. *Am. J. Hum. Genet.*, **59**, 437-444, 1996.

Watson E., Forster P., Richards M. and Bandelt H. Mitochondrial footprints of human expansions in Africa. *Am. J. Hum. Genet.*, **61**, 691-704, 1997.

Weber W.W. The legacy of pharmacogenetics and potential applications. *Mutat. Res*., **459**, 1-18, 2001.
Weiss K.M. American origins. *Proc. Natl. Acad. Sci. USA*, **91**, 833-835, 1994.
Were G.S. and Wilson D.A. East Africa through a thousand years. Evans Brothers Ltd, @A Portman Mansions, Chiltern Street, London W1M 1LE, Britain, 1984.
Wiedemann N., Truscott K.N., Pfannschmidt S., Guiard B., Meisinger C. and Pfanner N. Biogenesis of the protein import channel Tom40 of the mitochondrial outer membrane. *J. Biol. Chem*., **279**(18), 18188-18194, 2004.
Wilson R.K., Chen C., Avdalovic N., Burns and Hood L. Development of an automated procedure for fluorescent DNA sequencing. *Genomics*, **6**, 626-634, 1990.
Wilson J.F., Weiss D.A., Richards M. Thomas M.G., Bradman N. and Goldstein D.B. Genetic evidence for different male and female roles during cultural transitions in the British Isles. *Proc. Natl. Acad. Sci. USA*, **98**(9), 5078-5083, 2001.
Wolpoff M.H., Hawks J. and Caspari R. Multiregional, not multiple origins. *Am. J. Phys. Anthropol*., **112**, 129-136, 2000.
Yamakata H., Nakajima D. and Ohara O. Identification of sequence motifs causing band compressions on Human cDNA sequencing. *DNA Res*., **3**, 81-86, 1996.
Yao Y-G. and Zhang Y-P. Phylogeographic analysis of mtDNA variation in four ethnic populations from Yunnan province: new data and a reappraisal. *J. Hum. Genet*., **47**, 311-318, 2002.
Young E.W. and Mottram B.H. A course in world geography, book 8, East Africa, physical, regional and human. 3rd Ed., 1977. Edward Arnold Publishers Ltd., 41 Bedford Square, London, WC1 B3D, Great Britain.
Zamzami N., Susin S.A., Marchetti P., Hirsch T., Gmmez-Monterrey I., Castedo M. and Kroemer G. Mitochondrial control of nuclear apoptosis. *J. Exp. Med*., **183**, 1533-1544, 1996.
Zaninovich A.A., Raíces M., Rebagliati I., Ricci C. and Hagmüller K. Brown fat thermogenesis in cold-acclimated rats is not abolished by the suppression of thyroid function. *Am. J. Physiol. Endocrinol. Metab*., **283**, E496-E502, 2002.
Zei G., Lisa A., Fiorani O., Magri C., Quintana-Murci L., Semino O. and Santachiara-Benerecetti A.S. *Eur. J. Hum. Genet*., **11**, 802-807, 2003.
Zeviani M. and Antozzi C. Mitochondrial disorders. *Mol. Hum. Reprod*., **3**(2), 133-148, 1997.
Zhao L., Wang Q., Qian Y., Li R., Cao J., Hart L.C., Zhai S., Han D., Young W-Y. and Guan M-X. Clinical evaluation and mitochondrial DNA sequence analysis in two Chinese families with aminoglycoside-induced and non-syndromic hearing loss. *Biochem. Biophys. Res. Comm*., *336*, 967-973, 2005.
Zhivotovsky L.A., Rosenberg N.A. and Feldman M.W. Features of evolution and expansion of modern humans, inferred from genomewide microsatellite markers. *Am. J. Hum. Genet*., **72**, 1171-1186, 2003.

7.2 ELECTRONIC REFERENCES

Abhyankar A., Park H-B., Tonolo G. and Luthman H. Comparative sequence analysis of the non-protein-coding mitochondrial DNA of inbred rat strains. www.plosone.org, 4(12), e8148, doi:10.1371/journal.pone.0008148, 2009.
Abu-Amero K.K., González A.M., Larruga J.M., Bosley T.M. and Cabrera V.M. Eurasian and African mitochondrial DNA influences in the Saudi Arabian population. http://www.biomedcentral.14l47l-2l48/7/32, 2007.
Aebersold R. http://www.sciencewatch.com/may-june2004/sw_may-june2004_page3.htm, 2004.
Automated DNA sequencing chemistry guide, Applied Biosystems, http://sequencingfacility.med.monash.edu.ac/pdf/chemistry guide.pdf, 2000.
Automated DNA sequencing chemistry guide, Applied Biosystems, http://www.ibt.lt/sc/files/DNAseqCG.pdf, 1998.
BigDye® terminator v3.1 cycle sequencing kit protocol, Applied Biosystems, http://sequencing facility.med.monash.edu.au/pdf/BDTv3.1protocol, 2002.
Blench R. and Dendo M. Stratification in the peopling of China: how far does the linguistic evidence match genetics and archaeology? http://homepage.ntlworld.com/roger_blench/RBOP.htm, Geneva symposium, June 10-13, 2004.
Boone D. and Watson R.L. Moru – Ma'di survey report, http://www.sil.org/silesr/1999/0001/srveyrep.html, 1996.
CentriSep 96 high throughput dye terminator purification kit, http://www.genaxxon.com/catalogue/DNA-Purification-Kits/CentriSep-96-High-Throughput-dye-terminator purification kit, 2007.
Currat M. and Excoffier L. Modern humans did not admix with Neanderthals during their range expansion into Europe. www.plosbiology.org, 2(12), e421, 2264-2274, 2004.

Ely B., Wilson J.L., Jackson F. and Jackson B.A. African-American mitochondrial DNAs often match mtDNAs found in multiple African ethnic groups. BMC Biol., 4(34), http://www.biomedcentral.com/1741-7007/4/34, 2006.

Felsenstein J. Phylip, http://evolution.genetics.washington.edu/phylip, 2008.

Glebs A.B.S., Stencel E.B.S. and Knoche K. Introducing GoTaq™ DNA polymerase: Improved amplification with a choice of buffers. www.promega.com, **83**, 21-24, 2003.

Gomes V., Alves C., Amorim A., Carracedo Á., Sánchez-Diz P. and Gusmão L. Nilotes from Karamoja, Uganda: haplotype data defined by 17 Y-chromosome STRs. Forensic Sci. Int. Genet., doi:10.1016/j.fsigen.2009.07.001, 2009.

González A.M., García O., Larruga J.M. and Cabrera V.M. The mitochondrial lineage U8a reveals a Paleolithic settlement in the Basque country. *BMC Genomics*, **7**, http://www.biomedcentral.com/1471-2164/7/124, 2006.

González A.M., Larruga J.M., Abu-Amero K.K., Shi2 Y., Pestano J. and Cabrera V.M. Mitochondrial lineage M1 traces an early human backflow to Africa. BMC Genomics, http://www.biomedcentral.com/1471-2164/8/223, 2007.

Hall T. BIOEDIT Sequence Alignment Editor, http://www.mbio.ncsu.edu/BioEdit/bioedit.html, 2001.

Librado P. and Rozas J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics***, 25**, 1451-1452 | doi: 10.1093/bioinformatics/btp187, 2009.

Maca-Meyer N., González A.M., Larruga J.M., Flores C. and Cabrera V.M. Major genomic mitochondrial lineages delineate early human expansions. *BMC Genetics*, **2**, http://www.biomedcentral.com/1471-2156/2/13, 2001.

Mitomap, A human mitochondrial genome database. The report of the committee on the human mitochondrial genome, http://www.mitomap.org, 2002a

Mitomap, A human mitochondrial genome database. Illustrations: World migrations, http://www.mitomap.org/world migrations, 2002b.

Mitomap, A human mitochondrial genome database. Mitochondrial DNA function locations, http://www.mitomap.org, 2008a.

Mitomap, A human mitochondrial genome database. Unpublished mitochondrial DNA polymorphisms, http://www.mitomap.org, 2008b.

Mitomap, A human mitochondrial genome database. mtDNA somatic mutations, http://www.mitomap.org, 2008c.

Mitomap, A human mitochondrial genome database. mtDNA coding region sequence polymorphisms, http://www.mitomap.org, 2008d.

Mitomap, A human mitochondrial genome database. mtDNA control region sequence polymorphisms, http://www.mitomap.org, 2008e.

Naviaux R.K. Mitochondrial and metabolic disorders. The spectrum of mitochondrial diseases. A primary care physician's guide, http://biochemgen.ucsd.edu/mmdc/ep-toc.htm, 1997.

OLIGONUCLEOTIDE PROPERTIES CALCULATOR, http://www.basic.nwu.edu/biotools/oligocalc.html, 2007.

Pereira L., Richards M., Goios A., Alonso A., Albarrán C., Garcia O., Behar D.M., Gölge M., Hatina J., Al-Gazali L., Bradley D.G., Macaulay V. and Amorim A. High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium, http://www.genome.org/cgi/doi/10.1101/gr.3182305, 2005.

Phan V.C., Nong V.H., Nguyen B.N., Tran T.M.N., Le T.B.T., Do Q.H., Nguyen N.L., Bui T.H., Pham D.M., Tran T.T., Tong Q.M., Nguyen T.T., Nguyen D.T., Le T.T.H., Nguyen D.C., Le Q.H., Dang D.H., Quyen D.T., Van D.H., Trinh V.B., Le B.Q. and Nguyen D.B. http://www.ncbi.nlm.nih.gov/nuccore/DQ834260.1, 2006.

Princeton Separations, Box 300, Adelphia NJ 07710, USA. http://www.prinsep.com/resources/pdf/centrisep96insert.pdf, 2006.

Privatisation and Utility Sector Reform Project in Uganda, Uganda Government Public Enterprise Reform and Divestiture Program, http://www.perds.go.ug/index.php, 2001.

Rajkumar R., Banerjee J., Gunturi H.B., Trevedi R. and Kashyap V. Phylogeny and antiquity of M macrohaplogroup inferred from complete mtDNA sequence of Indian specific lineages. *BMC Evol. Biol.*, **5**, http://www.biomedcentral.com/1471-2148/5/26, 2005.

Raymond G.G. Ethnologue: Languages of the World, 15th Ed., http://www.ethnologue.com/., 2005.

Tan D-J., Chang J., Liu L-L., Bai R-K., Wang Y-F., Yeh K-T. and Wong L-J.C. Significance of somatic mutations and content alteration of mitochondrial DNA in oesophageal cancer. *BMC Cancer*, http://www.biomedcentral.com/1471-2407-6-93, 2006.

Van Oven M. and Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum. Mutat., 30(**2**), E386-E394, http://www.phylotree.org, 2009.

Vladutiu G.D. On the diagnostic challenges of respiratory chain defects, http://www.rgbgl.org/research/respiratory chain.htm, 2005.

APPENDIX A

mtDNA polymorphisms among the Baganda, Acholi and Lugbara from Uganda

The mtDNA sequences of 13 Baganda, 13 Acholi and 14 Lugbara individuals from Uganda were established by automated sequencing. The sequences were aligned using the BioEdit programme (Hall, 2001) and the polymorphisms as indicated in Table A1 below were established by comparison with the rCRS of Andrews *et al*. (1999).

Table A1 mtDNA polymorphisms among the Baganda, Acholi and Lugbara from Uganda

Polymorphism	Individual																																			mtDNA region	Nature of polymorphism	Reference
	Acholi													Lugbara													Baganda												
	UG_CGR_71468	UG_CGR_71470	UG_CGR_71471	UG_CGR_71472	UG_CGR_71473	UG_CGR_71474	UG_CGR_71475	UG_CGR_71477	UG_CGR_71708	UG_CGR_71709	UG_CGR_71734	UG_CGR_71738	UG_CGR_71739	UG_CGR_71478	UG_CGR_71482	UG_CGR_71486	UG_CGR_71487	UG_CGR_71674	UG_CGR_71675	UG_CGR_71676	UG_CGR_71677	UG_CGR_71678	UG_CGR_71679	UG_CGR_71680	UG_CGR_71696	UG_CGR_71699	UG_CGR_71700	UG_CGR_71489	UG_CGR_71490	UG_CGR_71492	UG_CGR_71493	UG_CGR_71494	UG_CGR_71495	UG_CGR_71496	UG_CGR_71643	UG_CGR_71655	UG_CGR_71656	UG_CGR_71657	UG_CGR_71660	UG_CGR_71673			
C64T																																				1				1	HV2	non-coding	Rickards *et al*., 1999
T65G						1																																			HV2	non-coding	Not reported
A73G		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1		1		1	1			1	1	1		HV2	non-coding	Budowle *et al*., 1999
A93G	1																											1		1					1	1			1	1	HV2	non-coding	Budowle *et al*., 1999
A95C	1																											1							1				1		HV2	non-coding	Brandstatter *et al*., 2004
G143A															1									1								1									HV2	non-coding	Brandstatter *et al*., 2004
T146C		1	1	1					1					1	1			1						1								1		1				1	1		HV2	non-coding	Brandstatter *et al*., 2004
C150T													1			1																	1				1				HV2	non-coding	Budowle *et al*., 1999
C151T																	1			1																					HV2	non-coding	Brandstatter *et al*., 2004
T152C		1	1			1				1	1	1		1	1		1			1		1	1	1	1	1	1					1	1	1		1		1			HV2	non-coding	Brandstatter *et al*., 2004
C182T		1				1				1	1	1		1			1			1		1			1		1														HV2	non-coding	Kivisild *et al*., 2004

Table A1 mtDNA polymorphisms among the Baganda, Acholi and Lugbara from Uganda

Polymorphism	Individual																																							mtDNA region	Nature of polymorphism	Reference
	Acholi													Lugbara														Baganda														
	UG_CGR_71468	UG_CGR_71470	UG_CGR_71471	UG_CGR_71472	UG_CGR_71473	UG_CGR_71474	UG_CGR_71475	UG_CGR_71477	UG_CGR_71708	UG_CGR_71709	UG_CGR_71734	UG_CGR_71738	UG_CGR_71739	UG_CGR_71478	UG_CGR_71482	UG_CGR_71486	UG_CGR_71487	UG_CGR_71674	UG_CGR_71675	UG_CGR_71676	UG_CGR_71677	UG_CGR_71678	UG_CGR_71679	UG_CGR_71680	UG_CGR_71696	UG_CGR_71699	UG_CGR_71700	UG_CGR_71489	UG_CGR_71490	UG_CGR_71492	UG_CGR_71493	UG_CGR_71494	UG_CGR_71495	UG_CGR_71496	UG_CGR_71643	UG_CGR_71655	UG_CGR_71656	UG_CGR_71657	UG_CGR_71660	UG_CGR_71673			
A183G		1												1																											HV2	non-coding	Brandstatter *et al*., 2004
G185A	1																										1		1		1			1						HV2	non-coding	Salas *et al*., 2000
G185T																										1															HV2	non-coding	Graven *et al*., 1995
C186T		1												1																											HV2	non-coding	Stoneking *et al*., 1991
C186A																				1																					HV2	non-coding	Allard *et al*., 2005
A189G	1																									1	1	1		1			1		1	1	1				HV2	non-coding	Kivisild *et al*., 2004
A189C																				1																					HV2	non-coding	Brandstatter *et al*., 2004
T195C						1				1		1	1		1		1	1	1					1	1		1					1		1				1	1		HV2	non-coding	Kivisild *et al*., 2004
C198T						1																		1																	HV2	non-coding	Brandstatter *et al*., 2004
T199C																										1															HV2	non-coding	Ozawa *et al*., 1995
A200G																									1												1				HV2	non-coding	Kivisild *et al*., 2004
T204C																							1													1					HV2	non-coding	Howell *et al*., 2004
G207A		1																																		1					HV2	non-coding	Howell *et al*., 2004
T236C	1																										1		1					1	1				1	HV2	non-coding	Allard *et al*., 2005
A244G									1																								1				1	1		HV2	non-coding	Alves-Silva *et al*., 2000
G247A	1									1										1					1		1	1		1		1			1	1				1	HV2	non-coding	Brandstätter *et al*., 2004
247del G																			1																					HV2	non-coding	Brandstatter *et al*., 2004
249del A																																					1				HV2	non-coding	Bandelt *et al*., 2006
A257G				1					1																															HV2	non-coding	Poetsch *et al*., 2003
A263G	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	HV2	non-coding	Moraes *et al*., 1993
A302C																	1																							HV2	non-coding	Shields *et al*., 1993
309Ins C	1	1			1				1		1	1		1												1					1				1		1		HV2	non-coding	Brandstatter *et al*., 2004
310Ins C																										1														HV2	non-coding	Ozawa *et al*., 1995
315Ins C	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	HV2	non-coding	Coble *et al*., 2004
G316A																				1																					HV2	non-coding	Allard *et al*., 2005
A357G																										1															HV2	non-coding	Maca-Meyer *et al*, 2001

Table A1 mtDNA polymorphisms among the Baganda, Acholi and Lugbara from Uganda

Polymorphism	Individual																																						mtDNA region	Nature of polymorphism	Reference
	Acholi													Lugbara														Baganda													
	UG_CGR_71468	UG_CGR_71470	UG_CGR_71471	UG_CGR_71472	UG_CGR_71473	UG_CGR_71474	UG_CGR_71475	UG_CGR_71477	UG_CGR_71708	UG_CGR_71709	UG_CGR_71734	UG_CGR_71738	UG_CGR_71739	UG_CGR_71478	UG_CGR_71482	UG_CGR_71486	UG_CGR_71487	UG_CGR_71674	UG_CGR_71675	UG_CGR_71676	UG_CGR_71677	UG_CGR_71678	UG_CGR_71679	UG_CGR_71680	UG_CGR_71696	UG_CGR_71699	UG_CGR_71700	UG_CGR_71489	UG_CGR_71490	UG_CGR_71492	UG_CGR_71493	UG_CGR_71494	UG_CGR_71495	UG_CGR_71496	UG_CGR_71643	UG_CGR_71655	UG_CGR_71656	UG_CGR_71657	UG_CGR_71660	UG_CGR_71673			
T408A																				1																					LSP	non-coding	Del Bo *et al.*, 2002
456Ins TTC										1															1																HV3	non-coding	Kivisild *et al.*, 2006
459Ins C																			1																						HV3	non-coding	Brandstatter *et al.*, 2004
C481T												1																													HV3	non-coding	Tan *et al.*, 2006
C494T																				1																	1				HV3	non-coding	Behar *et al.*, 2008
G513A		1									1			1																											HV3	non-coding	Reider *et al.*, 1998
514del CA	1					1	1	1	1	1	1		1	1		1				1	1	1		1	1		1	1	1	1	1					1	1				HV3	non-coding	Abu-Amero *et al.*, 2007
C535T																			1																						mtTF1	non-coding	Brandstatter *et al.*, 2004
C549T																	1																								mtTF1	non-coding	Reider *et al.*, 1998
T593A											1																														$tRNA^{Phe}$	non-coding	Herrnstadt *et al.*, 2002
T650C																																					1				12S rRNA	non-coding	Behar *et al.*, 2008
G709A										1		1					1					1	1		1		1						1	1				1			12S rRNA	non-coding	Moraes *et al.*, 1993
T710C																											1														12S rRNA	non-coding	Kivisild *et al.*, 2006
G719A			1																																						12S rRNA	non-coding	Howell *et al.*, 2004
T742C		1												1																											12S rRNA	non-coding	Kivisild *et al.*, 2004
A750G	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	12S rRNA	non-coding	Torroni *et al.*, 2006
G769A	1	1	1	1		1			1	1	1	1		1	1	1	1		1	1		1	1	1	1		1	1		1		1		1	1	1		1	1	1	12S rRNA	non-coding	Kivisild *et al.*, 2004
T825A	1									1									1	1					1		1	1		1		1			1	1					1	12S rRNA	non-coding	Torroni *et al.*, 2006
A851G										1															1																	12S rRNA	non-coding	Reider *et al.*, 1998
G930A										1														1	1																	12S rRNA	non-coding	Kivisild *et al.*, 2006
T961C	1																										1							1							12S rRNA	non-coding	Ozawa *et al.*, 1995
A978G																															1										12S rRNA	non-coding	Shen *et al.*, 2004
G1018A	1	1	1	1		1			1	1	1	1		1	1	1	1		1	1		1	1	1	1		1	1		1		1	1	1	1	1		1	1	1	12S rRNA	non-coding	Kivisild *et al.*, 2004
C1048T	1																											1		1		1			1	1				1	12S rRNA	non-coding	Kivisild *et al.*, 2006
G1211A			1																																						12S rRNA	non-coding	Howell *et al.*, 2004
T1413C				1					1																								1				1	1		12S rRNA	non-coding	Torroni *et al.*, 2006

Table A1 mtDNA polymorphisms among the Baganda, Acholi and Lugbara from Uganda

Polymorphism	Individual																																					mtDNA region	Nature of polymorphism	Reference
	Acholi													Lugbara																Baganda												
	UG_CGR_71468	UG_CGR_71470	UG_CGR_71471	UG_CGR_71472	UG_CGR_71473	UG_CGR_71474	UG_CGR_71475	UG_CGR_71477	UG_CGR_71708	UG_CGR_71709	UG_CGR_71734	UG_CGR_71738	UG_CGR_71739	UG_CGR_71478	UG_CGR_71482	UG_CGR_71486	UG_CGR_71487	UG_CGR_71674	UG_CGR_71675	UG_CGR_71676	UG_CGR_71677	UG_CGR_71678	UG_CGR_71679	UG_CGR_71680	UG_CGR_71696	UG_CGR_71699	UG_CGR_71700	UG_CGR_71489	UG_CGR_71490	UG_CGR_71492	UG_CGR_71493	UG_CGR_71494	UG_CGR_71495	UG_CGR_71496	UG_CGR_71643	UG_CGR_71655	UG_CGR_71656	UG_CGR_71657	UG_CGR_71660	UG_CGR_71673			
A1438G	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1	1	1	1	1	1	1	1	12S rRNA	non-coding	Chinnery *et al.*, 2001
G1598A																																	1				1			12S rRNA	non-coding	Herrnstadt *et al.*, 2002
G1664A																															1				1					tRNA^{Val}	non-coding	Ruiz-Pesini *et al.*, 2007
T1694C				1					1																															16S rRNA	non-coding	Kivisild *et al.*, 2006
G1719A			1		1													1													1	1								16S rRNA	non-coding	Kivisild *et al.*, 2004
T1738C																										1														16S rRNA	non-coding	Kivisild *et al.*, 2006
T1822C										1															1	1														16S rRNA	non-coding	Prezant *et al.*, 1993
A1914G																1																								16S rRNA	non-coding	Not reported
C2000T													1																											16S rRNA	non-coding	Herrnstadt *et al.*, 2002
A2245C																														1										16S rRNA	non-coding	Behar *et al.*, 2008
A2245G	1																											1				1			1	1				1	16S rRNA	non-coding	Moraes *et al.*, 1993
T2352C													1													1															16S rRNA	non-coding	Herrnstadt *et al.*, 2002
T2385C														1																											16S rRNA	non-coding	Not reported
2394 del A																			1																						16S rRNA	non-coding	Maca-Meyer *et al.*, 2001
T2416C		1	1								1	1		1	1		1					1	1	1																	16S rRNA	non-coding	Torroni *et al.*, 2001
C2417G																			1																						16S rRNA	non-coding	Torroni *et al.*, 2006
T2483C				1					1																															16S rRNA	non-coding	Finnilä *et al.*, 2001
A2558G																													1												16S rRNA	non-coding	Not reported
T2626C		1												1																												16S rRNA	non-coding	Nishino *et al.*, 1996
A2706G	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16S rRNA	non-coding	Kivisild *et al.*, 2006
A2757G																																				1					16S rRNA	non-coding	Vives-Bauza *et al.*, 2002
G2758A	1																		1							1	1		1		1			1	1					1	16S rRNA	non-coding	Torroni *et al.*, 2001
A2768G																										1															16S rRNA	non-coding	Kivisild *et al.*, 2006
C2772T																															1										16S rRNA	non-coding	Tanaka *et al.*, 2004
A2775G																				1																					16S rRNA	non-coding	Behar *et al.*, 2008
C2780T											1																														16S rRNA	non-coding	Ruiz-Pesini *et al.*, 2007

Table A1 mtDNA polymorphisms among the Baganda, Acholi and Lugbara from Uganda

Polymorphism	Individual																																					mtDNA region	Nature of polymorphism	Reference
	Acholi													Lugbara														Baganda													
	UG_CGR_71468	UG_CGR_71470	UG_CGR_71471	UG_CGR_71472	UG_CGR_71473	UG_CGR_71474	UG_CGR_71475	UG_CGR_71477	UG_CGR_71708	UG_CGR_71709	UG_CGR_71734	UG_CGR_71738	UG_CGR_71739	UG_CGR_71478	UG_CGR_71482	UG_CGR_71486	UG_CGR_71487	UG_CGR_71674	UG_CGR_71675	UG_CGR_71676	UG_CGR_71677	UG_CGR_71678	UG_CGR_71679	UG_CGR_71680	UG_CGR_71696	UG_CGR_71699	UG_CGR_71700	UG_CGR_71489	UG_CGR_71490	UG_CGR_71492	UG_CGR_71493	UG_CGR_71494	UG_CGR_71495	UG_CGR_71496	UG_CGR_71643	UG_CGR_71655	UG_CGR_71656	UG_CGR_71657	UG_CGR_71660	UG_CGR_71673			
C2789T											1	1			1		1					1	1	1								1		1				1			16S rRNA	non-coding	Herrnstadt *et al.*, 2002
G2831A																			1																						16S rRNA	non-coding	Torroni *et al.*, 2006
T2885C	1																			1							1	1		1		1			1	1				1	16S rRNA	non-coding	Herrnstadt *et al.*, 2002
G3010A																								1																	16S rRNA	non-coding	Herrnstadt *et al.*, 2002
T3027C																			1																						16S rRNA	non-coding	Torroni *et al.*, 2006
T3308C																								1			1														ND1	Met-Thr	Rocha *et al.*, 1999
G3316A												1																													ND1	Ala-Thr	Herrnstadt *et al.*, 2002
C3321T		1												1																												ND1	syn	Not reported
G3357A						1																																			ND1	syn	Torroni *et al.*, 2006
T3396C																										1																ND1	syn	Torroni *et al.*, 2006
T3398C														1																												ND1	Met-Thr	Sternberg *et al.*, 1998
T3423C										1									1						1																	ND1	syn	Kivisild *et al.*, 2006
C3450T							1	1													1								1		1										ND1	syn	Kivisild *et al.*, 2006
G3483A																																						1			ND1	syn	Torroni *et al.*, 2006
C3516A	1																										1		1		1			1	1				1	ND1	syn	Torroni *et al.*, 2001
A3537G			1																																					ND1	syn	Howell *et al.*, 2004
C3594T	1	1	1							1	1	1		1	1		1		1	1		1	1	1	1		1	1		1		1			1	1				1	ND1	syn	Kivisild *et al.*, 2004
C3654T		1												1																											ND1	syn	Herrnstadt *et al.*, 2002
G3666A				1																1							1														ND1	syn	Torroni *et al.*, 2006
G3693A																											1															ND1	syn	Torroni *et al.*, 2006
A3808G																		1																								ND1	Thr-Ala	Maca-Meyer *et al.*, 2001
G3834A									1																																ND1	syn	Coble *et al.*, 2004
C3852T																																1									ND1	syn	Kivisild *et al.*, 2006
T3866C																													1												ND1	Ile-Thr	Uusimaa *et al.*, 2004
G3918A				1					1																								1					1	1		ND1	syn	Herrnstadt *et al.*, 2002
C4032T		1												1																												ND1	syn	Not reported

Table A1 mtDNA polymorphisms among the Baganda, Acholi and Lugbara from Uganda

Polymorphism	Individual																																						mtDNA region	Nature of polymorphism	Reference
	Acholi													Lugbara																Baganda													
	UG_CGR_71468	UG_CGR_71470	UG_CGR_71471	UG_CGR_71472	UG_CGR_71473	UG_CGR_71474	UG_CGR_71475	UG_CGR_71477	UG_CGR_71708	UG_CGR_71709	UG_CGR_71734	UG_CGR_71738	UG_CGR_71739	UG_CGR_71478	UG_CGR_71482	UG_CGR_71486	UG_CGR_71487	UG_CGR_71674	UG_CGR_71675	UG_CGR_71676	UG_CGR_71677	UG_CGR_71678	UG_CGR_71679	UG_CGR_71680	UG_CGR_71696	UG_CGR_71699	UG_CGR_71700	UG_CGR_71489	UG_CGR_71490	UG_CGR_71492	UG_CGR_71493	UG_CGR_71494	UG_CGR_71495	UG_CGR_71496	UG_CGR_71643	UG_CGR_71655	UG_CGR_71656	UG_CGR_71657	UG_CGR_71660	UG_CGR_71673			
A4104G	1	1	1							1	1	1		1	1		1		1	1		1	1	1	1		1	1		1		1			1	1				1	ND1	syn	Torroni *et al.*, 2006
C4194T																																1									ND1	syn	Kivisild *et al.*, 2006
T4205C			1																																					ND1	Leu-Ser	Behar *et al.*, 2008
A4212G											1																														ND1	syn	Not reported
T4218C																										1															ND1	syn	Kivisild *et al.*, 2004
C4312T	1																											1		1		1			1	1				1	$tRNA^{Ile}$	non-coding	Kivisild *et al.*, 2006
T4314C																																								1	$tRNA^{Ile}$	non-coding	Fraumene *et al.*, 2006
A4388G					1													1																							$tRNA^{Gln}$	non-coding	Kivisild *et al.*, 2006
G4481A			1																																					ND2	syn	Behar *et al.*, 2008
C4496T										1															1																ND2	syn	Torroni *et al.*, 2006
A4562G			1																													1								ND2	syn	Howell *et al.*, 2004
T4586C	1																											1		1		1			1	1				1	ND2	syn	Kivisild *et al.*, 2006
G4596A																																	1								ND2	Val-Met	Maca-Meyer *et al.*, 2001
G4655A													1																												ND2	syn	Kivisild *et al.*, 2006
T4688C																		1																							ND2	syn	Kivisild *et al.*, 2006
T4742C																		1																							ND2	syn	Ruiz-Pesini *et al.*, 2007
A4769G	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	ND2	syn	Herrnstadt *et al.*, 2002
T4937C																																						1		ND2	syn	Ruiz-Pesini *et al.*, 2007
A4949G				1					1																															ND2	syn	Ruiz-Pesini *et al.*, 2007
C4964T																																1									ND2	syn	Kivisild *et al.*, 2006
A4976G																			1																						ND2	syn	Ingman and Gyllensten, 2003
T4977C																1																									ND2	syn	Starikovskaya *et al.*, 2005
T5004C										1																															ND2	syn	Herrnstadt *et al.*, 2002
C5027T																																	1								ND2	syn	Herrnstadt *et al.*, 2002
A5036G																										1															ND2	syn	Kivisild *et al.*, 2006

Table A1 mtDNA polymorphisms among the Baganda, Acholi and Lugbara from Uganda

Polymorphism	Individual																																					mtDNA region	Nature of polymorphism	Reference
	Acholi													Lugbara																	Baganda											
	UG_CGR_71468	UG_CGR_71470	UG_CGR_71471	UG_CGR_71472	UG_CGR_71473	UG_CGR_71474	UG_CGR_71475	UG_CGR_71477	UG_CGR_71708	UG_CGR_71709	UG_CGR_71734	UG_CGR_71738	UG_CGR_71739	UG_CGR_71478	UG_CGR_71482	UG_CGR_71486	UG_CGR_71487	UG_CGR_71674	UG_CGR_71675	UG_CGR_71676	UG_CGR_71677	UG_CGR_71678	UG_CGR_71679	UG_CGR_71680	UG_CGR_71696	UG_CGR_71699	UG_CGR_71700	UG_CGR_71489	UG_CGR_71490	UG_CGR_71492	UG_CGR_71493	UG_CGR_71494	UG_CGR_71495	UG_CGR_71496	UG_CGR_71643	UG_CGR_71655	UG_CGR_71656	UG_CGR_71657	UG_CGR_71660	UG_CGR_71673			
G5046A																											1														ND2	Val-Ile	Herrnstadt *et al.*, 2002
A5069T			1																																					ND2	syn	Howell *et al.*, 2004
T5096C	1																											1		1					1					ND2	syn	Maca-Meyer *et al.*, 2001
T5108C																		1																						ND2	syn	Starikovskaya *et al.*, 2005
C5111T										1															1															ND2	syn	Torroni *et al.*, 2006
G5147A										1															1											1				1	ND2	syn	Kivisild *et al.*, 2006
C5213T																			1																					ND2	syn	Moilanen *et al.*, 2003
G5231A	1																											1		1					1	1				1	ND2	syn	Herrnstadt *et al.*, 2002
G5262A													1																												ND2	Ala-Thr	Bandelt *et al.*, 2001
C5263T		1												1																											ND2	Ala-Val	Saxena *et al.*, 2006
C5300T					1		1											1																							ND2	syn	Kivisild *et al.*, 2006
C5321T																																1									ND2	syn	Kivisild *et al.*, 2006
T5333C						1																																			ND2	syn	Not reported
T5393C																											1														ND2	syn	Torroni *et al.*, 2006
T5442C	1																											1		1		1			1	1				1	ND2	Phe-Leu	Kivisild *et al.*, 2006
C5459T															1																										ND2	syn	Herrnstadt *et al.*, 2002
G5460A	1					1										1												1		1					1	1				1	ND2	Ala-Thr	Torroni *et al.*, 2006
G5471A				1		1									1																								1		ND2	syn	Howell *et al.*, 1995
T5492C					1													1																							ND2	syn	Hudjashov *et al.*, 2007
A5539G																											1														$tRNA^{trp}$	non-coding	Kato *et al.*, 2001
T5553C																					1																				$tRNA^{trp}$	non-coding	Ruiz-Pesini *et al.*, 2007
T5580C																																		1				1	1		NC3	non-coding	Ruiz-Pesini *et al.*, 2007
C5601T																										1															$tRNA^{Ala}$	non-coding	Torroni *et al.*, 2006
C5602T		1												1																											$tRNA^{Ala}$	non-coding	Not reported
C5603T	1																											1		1		1			1	1				1	$tRNA^{Ala}$	non-coding	Torroni *et al.*, 2006

Table A1 mtDNA polymorphisms among the Baganda, Acholi and Lugbara from Uganda

Polymorphism	Individual																																			mtDNA region	Nature of polymorphism	Reference
	Acholi													Lugbara														Baganda											
	UG_CGR_71468	UG_CGR_71470	UG_CGR_71471	UG_CGR_71472	UG_CGR_71473	UG_CGR_71474	UG_CGR_71475	UG_CGR_71477	UG_CGR_71708	UG_CGR_71709	UG_CGR_71734	UG_CGR_71738	UG_CGR_71739	UG_CGR_71478	UG_CGR_71482	UG_CGR_71486	UG_CGR_71487	UG_CGR_71674	UG_CGR_71675	UG_CGR_71676	UG_CGR_71677	UG_CGR_71678	UG_CGR_71679	UG_CGR_71680	UG_CGR_71696	UG_CGR_71699	UG_CGR_71700	UG_CGR_71489	UG_CGR_71490	UG_CGR_71492	UG_CGR_71493	UG_CGR_71494	UG_CGR_71495	UG_CGR_71496	UG_CGR_71643	UG_CGR_71655	UG_CGR_71656	UG_CGR_71657	UG_CGR_71660	UG_CGR_71673			
T5655C																											1														tRNA^{Ala}	non-coding	Kivisild *et al.*, 2006
A5656G										1															1															NC4	non-coding	Herrnstadt *et al.*, 2002
A5711G																																				1				1	tRNA^{Asn}	non-coding	Kivisild *et al.*, 2006
G5746A																																	1				1	1		O_L	non-coding	Cormio *et al.*, 2005
G5773A							1	1													1								1		1									tRNA^{Cys}	non-coding	Kivisild *et al.*, 2006
C5790A			1																																					tRNA^{Cys}	non-coding	Palanichamy *et al.*, 2004
5889Ins C						1																														1				tRNA^{Tyr}	non-coding	Torroni *et al.*, 2006
C5911T	1																											1						1						COI	Cys-Thr	Van Brummelen, 2003
A5951G																				1																				COI	syn	Torroni *et al.*, 2006
T6014C			1																																					COI	syn	Howell *et al.*, 2004
A6040C		1												1																											COI	Asn-Ser	Mitomap, 2008b
T6071C																				1																				COI	syn	Torroni *et al.*, 2006
C6164T																								1																COI	syn	Ruiz-Pesini *et al.*, 2007
G6182A										1															1															COI	syn	Kivisild *et al.*, 2006
T6185C	1																											1		1		1			1	1				1	COI	syn	Kivisild *et al.*, 2006
T6221C							1	1													1								1		1									COI	syn	Kivisild *et al.*, 2006
T6221A																				1																				COI	syn	Kivisild *et al.*, 2006
G6257A																																				1				1	COI	syn	Kivisild *et al.*, 2006
G6260A				1					1											1																				COI	syn	Reider *et al.*, 1998
G6261A													1																					1			1	1			COI	Ala-Thr	Bandelt *et al.*, 2006
T6297C										1															1															COI	syn	Ruiz-Pesini *et al.*, 2007
C6311T																								1																COI	syn	Herrnstadt *et al.*, 2002
A6338G																							1																	COI	syn	Zhao *et al.*, 2005
A6401G																																				1					COI	syn	Torroni *et al.*, 2006
T6497C		1												1																											COI	syn	Phan *et al.*, 2006
C6524T													1																												COI	syn	Herrnstadt *et al.*, 2002

Table A1 mtDNA polymorphisms among the Baganda, Acholi and Lugbara from Uganda

Polymorphism	Individual																																					mtDNA region	Nature of polymorphism	Reference
	Acholi													Lugbara														Baganda													
	UG_CGR_71468	UG_CGR_71470	UG_CGR_71471	UG_CGR_71472	UG_CGR_71473	UG_CGR_71474	UG_CGR_71475	UG_CGR_71477	UG_CGR_71708	UG_CGR_71709	UG_CGR_71734	UG_CGR_71738	UG_CGR_71739	UG_CGR_71478	UG_CGR_71482	UG_CGR_71486	UG_CGR_71487	UG_CGR_71674	UG_CGR_71675	UG_CGR_71676	UG_CGR_71677	UG_CGR_71678	UG_CGR_71679	UG_CGR_71680	UG_CGR_71696	UG_CGR_71699	UG_CGR_71700	UG_CGR_71489	UG_CGR_71490	UG_CGR_71492	UG_CGR_71493	UG_CGR_71494	UG_CGR_71495	UG_CGR_71496	UG_CGR_71643	UG_CGR_71655	UG_CGR_71656	UG_CGR_71657	UG_CGR_71660	UG_CGR_71673			
C6546T												1																												COI	Leu-Phe	Achilli *et al.*, 2005
C6548T																										1														COI	syn	Kivisild *et al.*, 2006
T6620C				1					1																														COI	syn	Liu *et al.*, 2001
T6641C																					1																		COI	syn	Hirano *et al.*, 1997
A6663G																								1																COI	Ile-Val	Bandelt *et al.*, 2006
C6713T											1																												COI	syn	Herrnstadt *et al.*, 2002
A6752G											1	1					1					1	1																COI	syn	Kivisild *et al.*, 2006
T6827C																										1														COI	syn	Kivisild *et al.*, 2006
G6917A																			1																				COI	syn	Kivisild *et al.*, 2006
C6960T			1																																				COI	syn	Qu *et al.*, 2005
A6989G																										1														COI	syn	Kivisild *et al.*, 2006
T7022C																			1																				COI	syn	Herrnstadt *et al.*, 2002
C7028T	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	COI	syn	Herrnstadt *et al.*, 2002
A7055G																										1														COI	syn	Kivisild *et al.*, 2006
A7091G																																					1		COI	syn	Rajkumar *et al.*, 2005
A7146G	1																		1							1	1		1		1			1	1				1	COI	Thr-Ala	Torroni *et al.*, 2006
T7148C																															1								COI	syn	Kivisild *et al.*, 2006
T7175C		1									1	1		1	1		1					1	1	1															COI	syn	Herrnstadt *et al.*, 2002
C7256T	1	1	1							1	1	1		1	1		1		1	1		1	1	1	1	1	1		1		1			1	1				1	COI	syn	Herrnstadt *et al.*, 2002
C7274T											1	1			1		1					1	1	1															COI	syn	Herrnstadt *et al.*, 2002
T7389C																1				1						1													COI	Tyr-His	Torroni *et al.*, 2006
G7419A											1																												COI	Glu-Lys	Ruiz-Pesini *et al.*, 2007
A7424G										1															1														COI	syn	Torroni *et al.*, 2006
G7498A																			1																				$tRNA^{Ser(UCN)}$	non-coding	Maca-Meyer *et al.*, 2001
G7521A	1	1	1							1	1	1		1	1		1		1	1		1	1	1	1		1	1		1		1			1	1			1	$tRNA^{Asp}$	non-coding	Torroni *et al.*, 2006
T7645C																																1							COII	syn	Kivisild *et al.*, 2006

Table A1 mtDNA polymorphisms among the Baganda, Acholi and Lugbara from Uganda

Polymorphism	Individual																																			mtDNA region	Nature of polymorphism	Reference
	Acholi													Lugbara													Baganda												
	UG_CGR_71468	UG_CGR_71470	UG_CGR_71471	UG_CGR_71472	UG_CGR_71473	UG_CGR_71474	UG_CGR_71475	UG_CGR_71477	UG_CGR_71708	UG_CGR_71709	UG_CGR_71734	UG_CGR_71738	UG_CGR_71739	UG_CGR_71478	UG_CGR_71482	UG_CGR_71486	UG_CGR_71487	UG_CGR_71674	UG_CGR_71675	UG_CGR_71676	UG_CGR_71677	UG_CGR_71678	UG_CGR_71679	UG_CGR_71680	UG_CGR_71696	UG_CGR_71699	UG_CGR_71700	UG_CGR_71489	UG_CGR_71490	UG_CGR_71492	UG_CGR_71493	UG_CGR_71494	UG_CGR_71495	UG_CGR_71496	UG_CGR_71643	UG_CGR_71655	UG_CGR_71656	UG_CGR_71657	UG_CGR_71660	UG_CGR_71673			
G7664A																							1																		COII	Ala-Thr	Ruiz-Pesini *et al.*, 2007
T7741C												1																												COII	syn	Coble *et al.*, 2004
A7771G											1	1			1		1					1	1	1																COII	syn	Herrnstadt *et al.*, 2002
G7789A																				1																				COII	syn	Reider *et al.*, 1998
C7819A																										1														COII	syn	Torroni *et al.*, 2006
C7849T																1																								COII	syn	Macaulay *et al.*, 2005
T7861C					1													1																						COII	syn	Kivisild *et al.*, 2006
C7867T																											1													COII	syn	Kivisild *et al.*, 2006
A7972G										1									1						1															COII	syn	Kivisild *et al.*, 2006
G8027A																				1																				COII	Ala-Thr	Torroni *et al.*, 2006
T8047C					1																																		COII	syn	Kivisild *et al.*, 2006
T8104C				1					1																									1				1	1		COII	syn	Kivisild *et al.*, 2006
G8155A										1															1															COII	syn	Kivisild *et al.*, 2006
A8188G										1															1															COII	syn	Ruiz-Pesini *et al.*, 2007
A8191G	1																										1							1					COII	syn	Maca-Meyer *et al.*, 2001
G8206A		1	1								1	1		1	1		1					1	1	1															COII	syn	Torroni *et al.*, 2006
A8248G																										1														COII	syn	Kivisild *et al.*, 2006
8271 9b-del																																			1				1	NC7	non-coding	Torroni *et al.*, 2006
T8311C																																				1				tRNA^{Lys}	non-coding	Torroni *et al.*, 2006
T8383C			1								1																													ATP8	syn	Howell *et al.*, 2004
G8392A																1																								ATP8	syn	Torroni *et al.*, 2006
C8428T	1																										1		1					1	1				1	ATP8	syn	Kivisild *et al.*, 2006
A8460G																																			1				1	ATP8	Asn-Ser	Maca-Meyer *et al.*, 2001
C8468T	1																			1						1			1		1			1	1				1	ATP8	syn	Torroni *et al.*, 2006
G8485A					1																																		ATP8	syn	Kivisild *et al.*, 2006
A8527G																										1														ATP6/ATP8	Met-Val/syn	Torroni *et al.*, 2006

Table A1 mtDNA polymorphisms among the Baganda, Acholi and Lugbara from Uganda

Polymorphism	Individual																																						mtDNA region	Nature of polymorphism	Reference
	Acholi													Lugbara													Baganda														
	UG_CGR_71468	UG_CGR_71470	UG_CGR_71471	UG_CGR_71472	UG_CGR_71473	UG_CGR_71474	UG_CGR_71475	UG_CGR_71477	UG_CGR_71708	UG_CGR_71709	UG_CGR_71734	UG_CGR_71738	UG_CGR_71739	UG_CGR_71478	UG_CGR_71482	UG_CGR_71486	UG_CGR_71487	UG_CGR_71674	UG_CGR_71675	UG_CGR_71676	UG_CGR_71677	UG_CGR_71678	UG_CGR_71679	UG_CGR_71680	UG_CGR_71696	UG_CGR_71699	UG_CGR_71700	UG_CGR_71489	UG_CGR_71490	UG_CGR_71492	UG_CGR_71493	UG_CGR_71494	UG_CGR_71495	UG_CGR_71496	UG_CGR_71643	UG_CGR_71655	UG_CGR_71656	UG_CGR_71657	UG_CGR_71660	UG_CGR_71673			
A8566G	1																											1		1					1	1				1	ATP6/ATP8	Thr-Ala/syn	Torroni *et al*., 2006
C8582T										1															1																ATP6	Ala-Val	Prezant *et al*., 1993
T8604C																								1																	ATP6	syn	Herrnstadt *et al*., 2002
C8655T	1									1									1	1					1		1	1		1		1			1	1				1	ATP6	syn	Torroni *et al*., 2006
A8688G																										1																ATP6	syn	Kivisild *et al*., 2006
A8701G	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	ATP6	Thr-Ala	Herrnstadt *et al*., 2002
A8721G																						1																				ATP6	syn	Kivisild *et al*., 2006
C8754T										1															1																ATP6	syn	Ruiz-Pesini *et al*., 2007
C8781A																		1																								ATP6	syn	Ingman *et al*., 2000
A8817G																																						1				ATP6	syn	Torroni *et al*., 2006
G8838A		1				1																																				ATP6	syn	Starikovskaya *et al*., 2005
A8860G	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	ATP6	Thr112Ala	Herrnstadt *et al*., 2002
T8928C																1										1																ATP6	syn	Torroni *et al*., 2006
C8932T																										1																ATP6	Pro-Ser	Torroni *et al*., 2006
C8943T																		1																								ATP6	syn	Poetsch *et al*., 2003
C9042T	1																											1		1		1			1	1				1	ATP6	syn	Kivisild *et al*., 2006
A9051G																																							1		ATP6	syn	Palanichamy *et al*., 2004
G9053A			1																																						ATP6	Ser-Asn	Ruiz-Pesini *et al*., 2007
G9055A																																							1		ATP6	Ala-Thr	Herrnstadt *et al*., 2002
A9072G																			1																							ATP6	syn	Torroni *et al*., 2006
C9081T						1																																				ATP6	syn	Tanaka *et al*., 2004
T9117C												1					1																									ATP6	syn	Herrnstadt *et al*., 2002
C9168T																1																										ATP6	syn	Kivisild *et al*., 2006
A9181G																			1																								ATP6	Ser-Gly	Marzuki *et al*., 1991
A9221G			1								1	1			1		1					1	1	1																		COIII	syn	Torroni *et al*., 2006
A9254G																																				1						COIII	syn	Torroni *et al*., 2006

Table A1 mtDNA polymorphisms among the Baganda, Acholi and Lugbara from Uganda

Polymorphism	Individual																																						mtDNA region	Nature of polymorphism	Reference
	Acholi													Lugbara														Baganda														
	UG_CGR_71468	UG_CGR_71470	UG_CGR_71471	UG_CGR_71472	UG_CGR_71473	UG_CGR_71474	UG_CGR_71475	UG_CGR_71477	UG_CGR_71708	UG_CGR_71709	UG_CGR_71734	UG_CGR_71738	UG_CGR_71739	UG_CGR_71478	UG_CGR_71482	UG_CGR_71486	UG_CGR_71487	UG_CGR_71674	UG_CGR_71675	UG_CGR_71676	UG_CGR_71677	UG_CGR_71678	UG_CGR_71679	UG_CGR_71680	UG_CGR_71696	UG_CGR_71699	UG_CGR_71700	UG_CGR_71489	UG_CGR_71490	UG_CGR_71492	UG_CGR_71493	UG_CGR_71494	UG_CGR_71495	UG_CGR_71496	UG_CGR_71643	UG_CGR_71655	UG_CGR_71656	UG_CGR_71657	UG_CGR_71660	UG_CGR_71673			
G9305A										1															1																COIII	syn	Kivisild *et al.*, 2006
G9329A										1															1																COIII	syn	Kivisild *et al.*, 2006
A9347G	1																											1		1		1			1	1				1	COIII	syn	Kivisild *et al.*, 2006
A9377G			1																																						COIII	syn	Howell *et al.*, 2004
G9380A			1																																						COIII	syn	Torroni *et al.*, 2006
G9438A											1																														COIII	syn	Ruiz-Pesini *et al.*, 2007
C9449T							1	1													1								1		1										COIII	syn	Kivisild et al., 2006
T9509C					1													1																							COIII	syn	Kivisild *et al.*, 2006
T9540C	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	COIII	syn	Herrnstadt *et al.*, 2002
A9545G																																			1						COIII	syn	Torroni *et al.*, 2006
G9554A													1																						1						COIII	syn	Kivisild *et al.*, 2006
G9575A					1													1																							COIII	syn	Kivisild *et al.*, 2004
T9581C																															1										COIII	syn	Shen *et al.*, 2004
C9620T																															1										COIII	syn	Shen *et al.*, 2004
G9755A	1																											1		1					1	1				1	COIII	syn	Kivisild *et al.*, 2006
T9758C						1																																			COIII	syn	Rajkumar *et al.*, 2005
A9809C																			1																							COIII	syn	Torroni *et al.*, 2006
C9818T	1																											1		1		1			1	1				1	COIII	syn	Kivisild *et al.*, 2006
A9855G				1					1																									1				1	1		COIII	Ile-Val	Torroni *et al.*, 2006
T9899C										1															1																COIII	syn	Coble *et al.*, 2004
G9932A																					1	1																			COIII	syn	Kivisild *et al.*, 2006
A9941G																																				1					COIII	syn	Torroni *et al.*, 2006
T9950C																										1															COIII	syn	Torroni *et al.*, 2006
G9966A																				1																					COIII	Val-Ile	Maca-Meyer *et al.*, 2001
C9971T			1																																						COIII	syn	Howell *et al.*, 2004
A10086G							1	1													1								1		1										ND3	Asn-Asp	Kivisild *et al.*, 2006

Table A1 mtDNA polymorphisms among the Baganda, Acholi and Lugbara from Uganda

Polymorphism	Individual																																					mtDNA region	Nature of polymorphism	Reference
	Acholi													Lugbara														Baganda													
	UG_CGR_71468	UG_CGR_71470	UG_CGR_71471	UG_CGR_71472	UG_CGR_71473	UG_CGR_71474	UG_CGR_71475	UG_CGR_71477	UG_CGR_71708	UG_CGR_71709	UG_CGR_71734	UG_CGR_71738	UG_CGR_71739	UG_CGR_71478	UG_CGR_71482	UG_CGR_71486	UG_CGR_71487	UG_CGR_71674	UG_CGR_71675	UG_CGR_71676	UG_CGR_71677	UG_CGR_71678	UG_CGR_71679	UG_CGR_71680	UG_CGR_71696	UG_CGR_71699	UG_CGR_71700	UG_CGR_71489	UG_CGR_71490	UG_CGR_71492	UG_CGR_71493	UG_CGR_71494	UG_CGR_71495	UG_CGR_71496	UG_CGR_71643	UG_CGR_71655	UG_CGR_71656	UG_CGR_71657	UG_CGR_71660	UG_CGR_71673			
T10115C		1	1								1	1		1	1		1					1	1	1																ND3	syn	Kivisild *et al.*, 2006
G10143A																															1									ND3	Gly-Ser	Prezant *et al.*, 1993
G10373A						1	1	1													1								1		1									ND3	syn	Torroni *et al.*, 2006
A10398G	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	ND3	Thr-Ala	Herrnstadt *et al.*, 2002
T10410C																			1																					tRNA^{Arg}	non-coding	Chen *et al.*, 1995
T10463C		1								1				1																										tRNA^{Arg}	non-coding	Macaulay *et al.*, 1999
T10481C																	1																							ND4L	syn	Kivisild *et al.*, 2006
T10493C																			1																					ND4L	syn	Gonder *et al.*, 2007
G10586A																				1																					ND4L	syn	Maca-Meyer *et al.*, 2001
G10589A	1																											1		1		1			1	1				1	ND4L	syn	Kivisild *et al.*, 2006
C10664T	1																											1		1		1			1	1				1	ND4L	syn	Kivisild *et al.*, 2006
T10667C													1																											ND4L	syn	Herrnstadt *et al.*, 2002
A10679G																																1								ND4L	syn	Kivisild *et al.*, 2006
G10688A	1									1									1	1					1		1	1		1		1			1	1				1	ND4L	syn	Torroni *et al.*, 2006
A10783C									1																															ND4	syn	Gonder *et al.*, 2007
T10790C																																1								ND4	syn	Kivisild *et al.*, 2006
T10810C	1									1									1	1					1		1	1		1		1			1	1				1	ND4	syn	Torroni *et al.*, 2006
C10813T																1																								ND4	syn	Behar *et al.*, 2008
A10816G													1																											ND4	syn	Kivisild *et al.*, 2006
A10819G													1																				1				1			ND4	syn	Torroni *et al.*, 2006
T10873C	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	ND4	syn	Herrnstadt *et al.*, 2002
T10909C																																							1	ND4	syn	Gonder *et al.*, 2007
T10915C	1																											1		1		1			1	1				1	ND4	syn	Kivisild *et al.*, 2006
T10966C																																	1				1			ND4	syn	Mackey and Howell, 1992
A11002G							1	1												1									1		1									ND4	syn	Chen *et al.*, 1995
T11017C																											1							1						ND4	syn	Ruiz-Pesini *et al.*, 2007

Table A1 mtDNA polymorphisms among the Baganda, Acholi and Lugbara from Uganda

Polymorphism	Individual																																							mtDNA region	Nature of polymorphism	Reference
	Acholi													Lugbara																Baganda													
	UG_CGR_71468	UG_CGR_71470	UG_CGR_71471	UG_CGR_71472	UG_CGR_71473	UG_CGR_71474	UG_CGR_71475	UG_CGR_71477	UG_CGR_71708	UG_CGR_71709	UG_CGR_71734	UG_CGR_71738	UG_CGR_71739	UG_CGR_71478	UG_CGR_71482	UG_CGR_71486	UG_CGR_71487	UG_CGR_71674	UG_CGR_71675	UG_CGR_71676	UG_CGR_71677	UG_CGR_71678	UG_CGR_71679	UG_CGR_71680	UG_CGR_71696	UG_CGR_71699	UG_CGR_71700	UG_CGR_71489	UG_CGR_71490	UG_CGR_71492	UG_CGR_71493	UG_CGR_71494	UG_CGR_71495	UG_CGR_71496	UG_CGR_71643	UG_CGR_71655	UG_CGR_71656	UG_CGR_71657	UG_CGR_71660	UG_CGR_71673			
T11025C										1															1																	ND4	Leu-Pro	Prezant *et al.*, 1993
T11137C				1					1																															ND4	syn	Behar *et al.*, 2008
C11143T																																			1					ND4	syn	Kivisild *et al.*, 2006
T11152C																									1																ND4	syn	Palanichamy *et al.*, 2004
A11172G																																			1				1	ND4	Asn-Ser	Kivisild *et al.*, 2006
G11176A	1															1												1		1					1	1				1	ND4	syn	Kivisild *et al.*, 2006
T11253C						1																																		ND4	Ile-Thr	Herrnstadt *et al.*, 2002
T11260C																																1								ND4	syn	Kivisild *et al.*, 2006
C11269T																																							1	ND4	syn	Kivisild *et al.*, 2006
T11287C																															1									ND4	syn	Kivisild *et al.*, 2006
T11299C																															1									ND4	syn	Kivisild *et al.*, 2006
C11302T																				1																					ND4	syn	Kivisild *et al.*, 2006
A11334G			1																																					ND4	Asn-Ser	Not reported
A11344G						1																																		ND4	syn	Torroni *et al.*, 2006
G11377A																																1								ND4	syn	Herrnstadt *et al.*, 2002
G11440A																										1															ND4	syn	Torroni *et al.*, 2006
T11485C						1																																		ND4	syn	Coble *et al.*, 2004
A11590G					1													1																							ND4	syn	Kivisild *et al.*, 2006
A11641G	1																											1		1		1			1	1				1	ND4	syn	Kivisild *et al.*, 2006
T11701C																			1																						ND4	syn	Behar *et al.*, 2008
G11719A	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	ND4	syn	Herrnstadt *et al.*, 2002
C11767T																						1	1																	ND4	syn	Kivisild *et al.*, 2006
A11800G							1	1																					1		1									ND4	Ile-Val	Herrnstadt *et al.*, 2002
C11881T										1															1															ND4	syn	Kivisild *et al.*, 2006
G11914A	1										1	1			1		1					1	1					1		1		1			1	1				1	ND4	syn	Herrnstadt *et al.*, 2002
T11935C			1																																					ND4	syn	Howell *et al.*, 2004

Table A1 mtDNA polymorphisms among the Baganda, Acholi and Lugbara from Uganda

Polymorphism	Individual																																						mtDNA region	Nature of polymorphism	Reference
	Acholi													Lugbara														Baganda														
	UG_CGR_71468	UG_CGR_71470	UG_CGR_71471	UG_CGR_71472	UG_CGR_71473	UG_CGR_71474	UG_CGR_71475	UG_CGR_71477	UG_CGR_71708	UG_CGR_71709	UG_CGR_71734	UG_CGR_71738	UG_CGR_71739	UG_CGR_71478	UG_CGR_71482	UG_CGR_71486	UG_CGR_71487	UG_CGR_71674	UG_CGR_71675	UG_CGR_71676	UG_CGR_71677	UG_CGR_71678	UG_CGR_71679	UG_CGR_71680	UG_CGR_71696	UG_CGR_71699	UG_CGR_71700	UG_CGR_71489	UG_CGR_71490	UG_CGR_71492	UG_CGR_71493	UG_CGR_71494	UG_CGR_71495	UG_CGR_71496	UG_CGR_71643	UG_CGR_71655	UG_CGR_71656	UG_CGR_71657	UG_CGR_71660	UG_CGR_71673			
T11944C		1									1	1		1	1		1					1	1	1																	ND4	syn	Torroni *et al*., 2006
G12007A	1																											1		1		1			1	1				1	ND4	syn	Herrnstadt *et al*., 2002
C12019T																			1																						ND4	syn	Maca-Meyer *et al*., 2001
G12127A	1																											1							1						ND4	syn	Herrnstadt *et al*., 2002
A12172G																																								1	$tRNA^{His}$	non-coding	Kivisild *et al*., 2006
C12175T																		1																							$tRNA^{His}$	non-coding	Florentz and Sissler, 2001
T12188C																			1																						$tRNA^{His}$	non-coding	Vilmi *et al*., 2005
T12189C			1																																					$tRNA^{His}$	non-coding	Howell *et al*., 2004
G12236A										1								1							1																$tRNA^{Ser(AGY)}$	non-coding	Prezant *et al*., 1993
A12248G													1																												$tRNA^{Ser(AGY)}$	non-coding	Kivisild *et al*., 2006
A12361G				1		1																																		ND5	Thr-Ala	Starikovskaya *et al*., 2005
T12414C					1	1																																		ND5	syn	Torroni *et al*., 2006
C12432T					1					1									1						1																ND5	syn	Kivisild *et al*., 2006
T12471C																										1																ND5	syn	Herrnstadt *et al*., 2002
G12501A																				1																						ND5	syn	Maca-Meyer *et al*., 2001
T12519C																		1								1																ND5	syn	Kivisild *et al*., 2006
A12546T																			1																						ND5	syn	Torroni *et al*., 2006
T12609C				1					1																									1				1	1		ND5	syn	Kivisild *et al*., 2006
G12630A		1												1																												ND5	syn	Herrnstadt *et al*., 2002
C12633T																					1																					ND5	syn	Marin-Garcia *et al*., 2000
G12651A					1																																				ND5	syn	Palanichamy *et al*., 2004
A12693G															1									1																		ND5	syn	Herrnstadt *et al*., 2002
C12705T	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	ND5	syn	Herrnstadt *et al*., 2002
T12714C																			1																						ND5	syn	Torroni *et al*., 2006
A12720G	1																											1		1					1	1				1	ND5	syn	Herrnstadt *et al*., 2002

Table A1 mtDNA polymorphisms among the Baganda, Acholi and Lugbara from Uganda

Polymorphism	Individual																																						mtDNA region	Nature of polymorphism	Reference
	Acholi													Lugbara													Baganda														
	UG_CGR_71468	UG_CGR_71470	UG_CGR_71471	UG_CGR_71472	UG_CGR_71473	UG_CGR_71474	UG_CGR_71475	UG_CGR_71477	UG_CGR_71708	UG_CGR_71709	UG_CGR_71734	UG_CGR_71738	UG_CGR_71739	UG_CGR_71478	UG_CGR_71482	UG_CGR_71486	UG_CGR_71487	UG_CGR_71674	UG_CGR_71675	UG_CGR_71676	UG_CGR_71677	UG_CGR_71678	UG_CGR_71679	UG_CGR_71680	UG_CGR_71696	UG_CGR_71699	UG_CGR_71700	UG_CGR_71489	UG_CGR_71490	UG_CGR_71492	UG_CGR_71493	UG_CGR_71494	UG_CGR_71495	UG_CGR_71496	UG_CGR_71643	UG_CGR_71655	UG_CGR_71656	UG_CGR_71657	UG_CGR_71660	UG_CGR_71673			
A12723G																			1																						ND5	syn	Herrnstadt *et al.*, 2002
A12753G					1																																			ND5	syn	Kivisild *et al.*, 2006
A12810G																			1	1																					ND5	syn	Coble *et al.*, 2004
T12879C		1												1																											ND5	syn	Kivisild *et al.*, 2006
A12950G										1									1						1																ND5	Asn-Ser	Prezant *et al.*, 1993
C12988T																1																									ND5	syn	Not reported
A13068G																							1																		ND5	syn	Mishmar *et al.*, 2003
A13101C													1																												ND5	syn	Herrnstadt *et al.*, 2002
A13105G	1						1	1		1										1	1				1		1	1	1	1	1	1			1	1				1	ND5	Ile-Val	Prezant *et al.*, 1993
C13116T																																					1				ND5	syn	Torroni *et al.*, 2006
C13122A						1																																			ND5	syn	Not reported
C13125T											1																														ND5	syn	Not reported
T13174C						1																																			ND5	syn	Torroni *et al.*, 2006
T13191C												1																													ND5	syn	Kivisild *et al.*, 2006
C13197T													1																													ND5	syn	Herrnstadt *et al.*, 2002
A13276G	1																											1		1		1			1	1				1	ND5	Met-Val	Kivisild *et al.*, 2006
A13470G				1					1																							1		1				1	1		ND5	syn	Kivisild *et al.*, 2006
A13485G																			1																						ND5	syn	Torroni *et al.*, 2006
C13506T	1									1									1	1					1		1	1		1		1			1	1				1	ND5	syn	Torroni *et al.*, 2006
T13569C																			1																						ND5	syn	Torroni *et al.*, 2006
G13590A		1	1								1	1		1	1		1					1	1	1									1								ND5	syn	Kivisild *et al.*, 2006
C13650T	1	1	1							1	1	1		1	1		1		1	1		1	1	1	1		1	1		1		1			1	1				1	ND5	syn	Torroni *et al.*, 2006
A13651G													1																												ND5	Thr-Ala	Herrnstadt *et al.*, 2002
C13680T																																1									ND5	syn	Kivisild *et al.*, 2006
G13708A			1	1	1																										1					1				ND5	Ala-Thr	Wallace *et al.*, 2004
A13722G										1															1																ND5	syn	Torroni *et al.*, 2006

Table A1 mtDNA polymorphisms among the Baganda, Acholi and Lugbara from Uganda

Polymorphism	Individual																																					mtDNA region	Nature of polymorphism	Reference
	Acholi													Lugbara														Baganda													
	UG_CGR_71468	UG_CGR_71470	UG_CGR_71471	UG_CGR_71472	UG_CGR_71473	UG_CGR_71474	UG_CGR_71475	UG_CGR_71477	UG_CGR_71708	UG_CGR_71709	UG_CGR_71734	UG_CGR_71738	UG_CGR_71739	UG_CGR_71478	UG_CGR_71482	UG_CGR_71486	UG_CGR_71487	UG_CGR_71674	UG_CGR_71675	UG_CGR_71676	UG_CGR_71677	UG_CGR_71678	UG_CGR_71679	UG_CGR_71680	UG_CGR_71696	UG_CGR_71699	UG_CGR_71700	UG_CGR_71489	UG_CGR_71490	UG_CGR_71492	UG_CGR_71493	UG_CGR_71494	UG_CGR_71495	UG_CGR_71496	UG_CGR_71643	UG_CGR_71655	UG_CGR_71656	UG_CGR_71657	UG_CGR_71660	UG_CGR_71673			
T13789C																				1							1														ND5	Tyr-His	Torroni *et al.*, 2006
C13800A																																1								ND5	Asn-Lys	Kivisild *et al.*, 2006
A13803G											1	1			1		1					1	1	1																ND5	syn	Herrnstadt *et al.*, 2002
G13813A															1																										ND5	Val-Ile	Herrnstadt *et al.*, 2002
A13827G																						1	1																		ND5	syn	Uusimaa *et al.*, 2004
T13830C																			1																						ND5	syn	Torroni *et al.*, 2006
C13914A							1	1													1								1		1										ND5	syn	Kivisild *et al.*, 2006
T13926C					1																																				ND5	syn	Kivisild *et al.*, 2006
G13928C																																1								ND5	Ser-Thr	Howell *et al.*, 2004
T14000A																				1																					ND5	Leu-Gln	Torroni *et al.*, 2006
A14007G	1																											1							1						ND5	syn	Maca-Meyer *et al.*, 2001
T14034C																																		1				1			ND5	syn	Torroni *et al.*, 2006
T14106C																																							1		ND5	syn	Herrnstadt *et al.*, 2002
C14109T																																1								ND5	syn	Kivisild *et al.*, 2006
T14110C					1																																			ND5	Phe-Leu	Ruiz-Pesini *et al.*, 2007
A14118G																						1	1																		ND5	syn	Mishmar *et al.*, 2003
T14178C																			1							1															ND6	Ile-Val	Torroni *et al.*, 2006
A14203G																										1															ND6	syn	Kivisild *et al.*, 2006
T14212C										1			1												1															ND6	syn	Torroni *et al.*, 2006
C14239T										1															1															ND6	syn	Torroni *et al.*, 2006
G14249A																																				1					ND6	Ala-Val	Mitomap, 2008b
T14256C																																				1					ND6	Ile-Val	Vives-Bauza *et al.*, 2002
A14257G																																	1				1				ND6	syn	Not reported
T14299C			1																																						ND6	syn	Howell *et al.*, 2004
T14302C						1																																			ND6	syn	Torroni *et al.*, 2006
T14308C	1																											1		1					1	1				1	ND6	syn	Herrnstadt *et al.*, 2002

Table A1 mtDNA polymorphisms among the Baganda, Acholi and Lugbara from Uganda

Polymorphism	Individual																																					mtDNA region	Nature of polymorphism	Reference
	Acholi													Lugbara																Baganda												
	UG_CGR_71468	UG_CGR_71470	UG_CGR_71471	UG_CGR_71472	UG_CGR_71473	UG_CGR_71474	UG_CGR_71475	UG_CGR_71477	UG_CGR_71708	UG_CGR_71709	UG_CGR_71734	UG_CGR_71738	UG_CGR_71739	UG_CGR_71478	UG_CGR_71482	UG_CGR_71486	UG_CGR_71487	UG_CGR_71674	UG_CGR_71675	UG_CGR_71676	UG_CGR_71677	UG_CGR_71678	UG_CGR_71679	UG_CGR_71680	UG_CGR_71696	UG_CGR_71699	UG_CGR_71700	UG_CGR_71489	UG_CGR_71490	UG_CGR_71492	UG_CGR_71493	UG_CGR_71494	UG_CGR_71495	UG_CGR_71496	UG_CGR_71643	UG_CGR_71655	UG_CGR_71656	UG_CGR_71657	UG_CGR_71660	UG_CGR_71673			
T14374C										1																														ND6	syn	Kivisild *et al.*, 2006
T14422C				1																1																				ND6	syn	Behar *et al.*, 2006
G14560A																			1							1														ND6	syn	Torroni *et al.*, 2006
A14566G											1	1			1		1					1	1	1																ND6	syn	Herrnstadt *et al.*, 2002
A14573G			1																																					ND6	Val-Ala	Not reported
T14581C										1															1																ND6	syn	Torroni *et al.*, 2006
A14587G																		1																						ND6	syn	Torroni *et al.*, 2006
C14620T																																1								ND6	syn	Kivisild *et al.*, 2006
A14755G																																			1						Cytb	syn	Kivisild *et al.*, 2006
C14766T	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Cytb	Thr-Ile	Herrnstadt *et al.*, 2002
A14769G																										1	1													Cytb	Asn-Ser	Kivisild *et al.*, 2006
A14793G																1																								Cytb	His-Arg	Uusimaa *et al.*, 2004
C14862T																		1																						Cytb	Ala-Val	Abu-Amero and Bosley, 2006
A14890G		1												1																										Cytb	syn	Hudjashov *et al.*, 2007
G14905A										1															1															Cytb	syn	Kivisild *et al.*, 2006
C14911T																				1																				Cytb	syn	Torroni *et al.*, 2006
T14971C										1															1															Cytb	syn	Shen *et al.*, 2004
A15061G																														1										Cytb	syn	Kivisild *et al.*, 2006
T15066C						1																																		Cytb	Phe-Ser	Not reported
G15077A																										1														Cytb	Glu-Lys	Ingman and Gyllensten, 2003
C15103T																										1														Cytb	syn	Ruiz-Pesini *et al.*, 2007
T15115C																										1														Cytb	syn	Kivisild *et al.*, 2006
C15136T	1																											1		1		1			1	1			1	Cytb	syn	Herrnstadt *et al.*, 2002
G15148A																			1																					Cytb	syn	Legros *et al.*, 2001
G15217A										1															1								1				1			Cytb	syn	Herrnstadt *et al.*, 2002

Table A1 mtDNA polymorphisms among the Baganda, Acholi and Lugbara from Uganda

Polymorphism	Individual																																						mtDNA region	Nature of polymorphism	Reference
	Acholi													Lugbara														Baganda														
	UG_CGR_71468	UG_CGR_71470	UG_CGR_71471	UG_CGR_71472	UG_CGR_71473	UG_CGR_71474	UG_CGR_71475	UG_CGR_71477	UG_CGR_71708	UG_CGR_71709	UG_CGR_71734	UG_CGR_71738	UG_CGR_71739	UG_CGR_71478	UG_CGR_71482	UG_CGR_71486	UG_CGR_71487	UG_CGR_71674	UG_CGR_71675	UG_CGR_71676	UG_CGR_71677	UG_CGR_71678	UG_CGR_71679	UG_CGR_71680	UG_CGR_71696	UG_CGR_71699	UG_CGR_71700	UG_CGR_71489	UG_CGR_71490	UG_CGR_71492	UG_CGR_71493	UG_CGR_71494	UG_CGR_71495	UG_CGR_71496	UG_CGR_71643	UG_CGR_71655	UG_CGR_71656	UG_CGR_71657	UG_CGR_71660	UG_CGR_71673			
A15226G																				1																					Cytb	syn	Kivisild *et al.*, 2006
T15299C												1																													Cytb	syn	Torroni *et al.*, 2006
G15301A		1	1	1	1	1	1	1	1		1	1	1	1	1	1	1	1			1	1	1	1		1			1		1		1	1			1	1	1		Cytb	syn	Torroni *et al.*, 2006
A15311G							1	1													1								1		1										Cytb	Ile-Val	Kivisild *et al.*, 2006
T15313C																																1									Cytb	syn	Kivisild *et al.*, 2006
A15326G	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Cytb	Thr194Ala	Herrnstadt *et al.*, 2002
A15328G		1																																							Cytb	syn	Not reported
C15381T																											1							1							Cytb	Thr-Ile	Behar *et al.*, 2006
G15431A	1																										1		1		1			1	1				1	Cytb	Ala-Thr	Maca-Meyer *et al.*, 2001
C15446T											1																														Cytb	Leu-Phe	Not reported
C15481T	1																																								Cytb	syn	Mitomap, 2008b
T15514C						1																				1															Cytb	syn	Torroni *et al.*, 2006
C15574T		1																																							Cytb	syn	Not reported
C15646T																		1																							Cytb	syn	Behar *et al.*, 2008
A15655G																																						1			Cytb	syn	Not reported
A15673G					1																																			Cytb	syn	Not reported
T15697C			1																																					Cytb	syn	Howell *et al.*, 2004
G15734A			1																																					Cytb	Ala-Thr	Howell *et al.*, 2004
A15758G																																					1			Cytb	Ile-Val	Kivisild *et al.*, 2006
T15784C												1			1									1																Cytb	syn	Herrnstadt *et al.*, 2002
G15812A													1																												Cytb	Val-Met	Kivisild *et al.*, 2006
A15824G							1	1													1								1		1										Cytb	Thr-Ala	Torroni *et al.*, 2006
T15852C																																1									Cytb	Ile-Thr	Gonder *et al.*, 2007
G15883A							1	1																					1		1										Cytb	syn	Gonzalez *et al.*, 2006
G15884A										1		1													1																Cytb	Ala-Thr	Prezant *et al.*, 1993
T15889C			1																																					$tRNA^{Thr}$	non-coding	Ruiz-Pesini *et al.*, 2007

Table A1 mtDNA polymorphisms among the Baganda, Acholi and Lugbara from Uganda

Polymorphism	Individual																																			mtDNA region	Nature of polymorphism	Reference
	Acholi													Lugbara														Baganda											
	UG_CGR_71468	UG_CGR_71470	UG_CGR_71471	UG_CGR_71472	UG_CGR_71473	UG_CGR_71474	UG_CGR_71475	UG_CGR_71477	UG_CGR_71708	UG_CGR_71709	UG_CGR_71734	UG_CGR_71738	UG_CGR_71739	UG_CGR_71478	UG_CGR_71482	UG_CGR_71486	UG_CGR_71487	UG_CGR_71674	UG_CGR_71675	UG_CGR_71676	UG_CGR_71677	UG_CGR_71678	UG_CGR_71679	UG_CGR_71680	UG_CGR_71696	UG_CGR_71699	UG_CGR_71700	UG_CGR_71489	UG_CGR_71490	UG_CGR_71492	UG_CGR_71493	UG_CGR_71494	UG_CGR_71495	UG_CGR_71496	UG_CGR_71643	UG_CGR_71655	UG_CGR_71656	UG_CGR_71657	UG_CGR_71660	UG_CGR_71673			
T15905C																				1																				tRNA^{Thr}	non-coding	Maca-Meyer *et al.*, 2001
G15927A																			1																					tRNA^{Thr}	non-coding	Brown *et al.*, 1998
G15928A																																					1				tRNA^{Thr}	non-coding	Macaulay *et al.*, 1999
C15939T												1					1					1	1																		tRNA^{Thr}	non-coding	Mishmar *et al.*, 2003
15940del T							1	1													1								1		1									tRNA^{Thr}	non-coding	Kivisild *et al.*, 2006
15944del T																										1														tRNA^{Thr}	non-coding	Herrnstadt *et al.*, 2002
C15978T																				1													1							tRNA^{Pro}	non-coding	Kivisild *et al.*, 2006
T16075C																																		1				1			HVI	non-coding	Torroni *et al.*, 1998
T16086C																								1																	HVI	non-coding	Allard *et al.*, 2005
T16093C							1	1												1									1		1									HVI	non-coding	Kivisild *et al.*, 2004
C16107T																			1																					HVI	non-coding	Quintana-Murci *et al.*, 2004
C16108T																			1																					HVI	non-coding	Behar *et al.*, 2008
C16111A			1																																					HVI	non-coding	Kivisild *et al.*, 2004
C16112T																			1																					HVI	non-coding	Not reported
T16124C							1	1													1								1		1									HVI	non-coding	Kivisild *et al.*, 2004
T16126C																											1													HVI	non-coding	Kivisild *et al.*, 2004
G16129A	1									1									1	1					1			1		1		1			1					HVI	non-coding	Budowle *et al.*, 1999
T16136C																											1													HVI	non-coding	Kivisild *et al.*, 2004
G16145A			1																																					HVI	non-coding	Jackson *et al.*, 2005
C16147T																	1																							HVI	non-coding	Tan *et al.*, 2002
C16148T	1									1									1						1			1		1					1	1			1	HVI	non-coding	Salas *et al.*, 2002
G16153A																																	1							HVI	non-coding	Kivisild *et al.*, 2004
A16165G																		1																						HVI	non-coding	Kivisild *et al.*, 2004
A16166G										1									1						1															HVI	non-coding	Kivisild *et al.*, 2004
C16168T	1																				1							1		1					1					HVI	non-coding	Salas *et al.*, 2002

Table A1 mtDNA polymorphisms among the Baganda, Acholi and Lugbara from Uganda

Polymorphism	Individual																																							mtDNA region	Nature of polymorphism	Reference
	Acholi													Lugbara														Baganda														
	UG_CGR_71468	UG_CGR_71470	UG_CGR_71471	UG_CGR_71472	UG_CGR_71473	UG_CGR_71474	UG_CGR_71475	UG_CGR_71477	UG_CGR_71708	UG_CGR_71709	UG_CGR_71734	UG_CGR_71738	UG_CGR_71739	UG_CGR_71478	UG_CGR_71482	UG_CGR_71486	UG_CGR_71487	UG_CGR_71674	UG_CGR_71675	UG_CGR_71676	UG_CGR_71677	UG_CGR_71678	UG_CGR_71679	UG_CGR_71680	UG_CGR_71696	UG_CGR_71699	UG_CGR_71700	UG_CGR_71489	UG_CGR_71490	UG_CGR_71492	UG_CGR_71493	UG_CGR_71494	UG_CGR_71495	UG_CGR_71496	UG_CGR_71643	UG_CGR_71655	UG_CGR_71656	UG_CGR_71657	UG_CGR_71660	UG_CGR_71673			
C16169T																							1									1					1				HVI	non-coding	Kivisild *et al.*, 2004
A16170G											1																														HVI	non-coding	Yao *et al.*, 2002
T16172C	1																					1					1		1		1	1		1	1				1	HVI	non-coding	Salas *et al.*, 2002
C16173T																							1									1									HVI	non-coding	Kivisild *et al.*, 2004
C16174T																																	1								HVI	non-coding	Stoneking *et al.*, 1991
C16176T																	1																								HVI	non-coding	Bandelt *et al.*, 2001
C16179T																							1																		HVI	non-coding	Kivisild *et al.*, 2004
16180 del A										1																															HVI	non-coding	Horai and Hayasaka, 1990
A16183C																	1			1					1																HVI	non-coding	Jackson *et al.*, 2005
C16184T			1																																					HVI	non-coding	Jackson *et al.*, 2005
C16186T																					1				1																HVI	non-coding	Alves-Silva *et al.*, 2000
C16187T	1									1									1							1					1				1	1				1	HVI	non-coding	Batini *et al.*, 2007
C16188A																																								1	HVI	non-coding	Kivisild *et al.*, 2004
C16188G	1																											1		1					1	1					HVI	non-coding	Salas *et al.*, 2002
T16189C	1		1							1	1	1			1		1		1	1		1	1		1		1		1		1				1	1				1	HVI	non-coding	Salas *et al.*, 2002
C16192T					1										1			1																							HVI	non-coding	Jackson *et al.*, 2005
C16193T																																						1				HVI	non-coding	Kivisild *et al.*, 2004
T16195C																																						1				HVI	non-coding	Kivisild *et al.*, 2004
T16209C																										1																HVI	non-coding	Salas *et al.*, 2002
A16212G		1																																							HVI	non-coding	Horai and Hayasaka, 1990
C16214T																																								1		HVI	non-coding	Torroni *et al.*, 2001
A16215G																				1																						HVI	non-coding	Salas *et al.*, 2002
C16223T	1	1	1	1		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	HVI	non-coding	Torroni *et al.*, 2001
T16224C		1												1																												HVI	non-coding	Kivisild *et al.*, 2004
T16229C											1	1					1					1	1																			HVI	non-coding	Watson *et al.*, 1997

Table A1 mtDNA polymorphisms among the Baganda, Acholi and Lugbara from Uganda

Polymorphism	Individual																																							mtDNA region	Nature of polymorphism	Reference
	Acholi													Lugbara														Baganda														
	UG_CGR_71468	UG_CGR_71470	UG_CGR_71471	UG_CGR_71472	UG_CGR_71473	UG_CGR_71474	UG_CGR_71475	UG_CGR_71477	UG_CGR_71708	UG_CGR_71709	UG_CGR_71734	UG_CGR_71738	UG_CGR_71739	UG_CGR_71478	UG_CGR_71482	UG_CGR_71486	UG_CGR_71487	UG_CGR_71674	UG_CGR_71675	UG_CGR_71676	UG_CGR_71677	UG_CGR_71678	UG_CGR_71679	UG_CGR_71680	UG_CGR_71696	UG_CGR_71699	UG_CGR_71700	UG_CGR_71489	UG_CGR_71490	UG_CGR_71492	UG_CGR_71493	UG_CGR_71494	UG_CGR_71495	UG_CGR_71496	UG_CGR_71643	UG_CGR_71655	UG_CGR_71656	UG_CGR_71657	UG_CGR_71660	UG_CGR_71673			
A16230G	1																											1		1		1			1	1				1	HVI	non-coding	Salas *et al*., 2002
A16233G																			1																						1	HVI	non-coding	Jackson *et al*., 2005
A16235G																																					1				HVI	non-coding	Rando *et al*., 1998
C16239T			1																												1									HVI	non-coding	Salas *et al*., 2002
T16243C																																					1				HVI	non-coding	Salas *et al*., 2002
C16248T																														1											HVI	non-coding	Salas *et al*., 2004b
A16254G																			1																							HVI	non-coding	Kivisild *et al*., 2004
C16260T						1																																			HVI	non-coding	Kivisild *et al*., 2004
C16261T									1																																	HVI	non-coding	Jackson *et al*., 2005
C16264T						1					1																1															HVI	non-coding	Bandelt *et al*., 2001
A16265G				1																																					HVI	non-coding	Allard *et al*., 2005
A16265T													1																													HVI	non-coding	Bandelt *et al*., 2001
C16270T																										1																HVI	non-coding	Salas *et al*., 2002
G16274A																																		1				1			HVI	non-coding	Salas *et al*., 2002
C16278T	1	1	1				1	1		1	1	1		1	1		1		1	1	1	1	1	1	1	1	1	1		1	1			1						HVI	non-coding	Salas *et al*., 2002
A16284G		1																																							HVI	non-coding	Kivisild *et al*., 2004
C16291T												1					1					1	1																		HVI	non-coding	Salas *et al*., 2002
C16292T			1																							1															HVI	non-coding	Salas *et al*., 2002
A16293G	1																										1	1							1						HVI	non-coding	Salas *et al*., 2002
A16293T				1					1																									1				1	1		HVI	non-coding	Kivisild *et al*., 2004
C16294T											1	1			1		1			1		1	1	1																	HVI	non-coding	Salas *et al*., 2002
T16298C					1																																				HVI	non-coding	Alves-Silva *et al*., 2000
C16301T											1																														HVI	non-coding	Kivisild *et al*., 2004
A16309G		1												1										1																	HVI	non-coding	Salas *et al*., 2002
T16311C	1		1	1	1	1			1	1	1	1			1	1	1	1		1	1		1		1	1	1	1		1		1		1	1	1		1	1	1	HVI	non-coding	Salas *et al*., 2002
A16317T			1																																						HVI	non-coding	Starikovskaya *et al*., 2005

Table A1 mtDNA polymorphisms among the Baganda, Acholi and Lugbara from Uganda

Polymorphism	Individual																																							mtDNA region	Nature of polymorphism	Reference
	Acholi													Lugbara														Baganda														
	UG_CGR_71468	UG_CGR_71470	UG_CGR_71471	UG_CGR_71472	UG_CGR_71473	UG_CGR_71474	UG_CGR_71475	UG_CGR_71477	UG_CGR_71708	UG_CGR_71709	UG_CGR_71734	UG_CGR_71738	UG_CGR_71739	UG_CGR_71478	UG_CGR_71482	UG_CGR_71486	UG_CGR_71487	UG_CGR_71674	UG_CGR_71675	UG_CGR_71676	UG_CGR_71677	UG_CGR_71678	UG_CGR_71679	UG_CGR_71680	UG_CGR_71696	UG_CGR_71699	UG_CGR_71700	UG_CGR_71489	UG_CGR_71490	UG_CGR_71492	UG_CGR_71493	UG_CGR_71494	UG_CGR_71495	UG_CGR_71496	UG_CGR_71643	UG_CGR_71655	UG_CGR_71656	UG_CGR_71657	UG_CGR_71660	UG_CGR_71673			
G16319A																																1									HVI	non-coding	Salas *et al*., 2002
C16320T	1																											1		1					1	1				HVI	non-coding	Salas *et al*., 2002
C16327T																															1									HVI	non-coding	Salas *et al*., 2002
C16355T			1	1					1	1														1									1				1			HVI	non-coding	Salas *et al*., 2002
C16360T																			1	1																				HVI	non-coding	Salas *et al*., 2002
T16362C				1		1	1	1	1	1						1					1				1				1		1			1				1	1		HVI	non-coding	Salas *et al*., 2002
T16368C																															1									HVI	non-coding	Jackson *et al*., 2005
G16390A		1	1								1	1		1	1		1					1	1	1						1										D-loop	non-coding	Chen *et al*., 1995
G16398A																									1									1				1			D-loop	non-coding	Melton and Nelson, 2001
A16399G		1	1	1					1					1				1																1				1	1		D-loop	non-coding	Jackson *et al*., 2005
C16400T			1																																					D-loop	non-coding	Jackson *et al*., 2005
T16519C		1	1	1		1	1	1	1		1	1	1	1			1			1	1	1	1			1	1		1		1	1	1			1	1			1	D-loop	non-coding	Ozawa *et al*., 1995
Total	89	64	72	47	40	51	40	39	47	83	64	63	39	61	54	37	61	45	69	83	40	58	62	58	81	39	85	88	40	82	40	98	41	48	89	91	48	48	46	87			

Ala = alanine, Asn = asparagine, ATP6 = gene encoding ATP synthase subunit 6, ATP8 = gene encoding ATP synthase subunit 8, CO I - III = cytochrome oxidase subunits I to III, Cytb = cytochrome b, D-loop = displacement loop, Gln = glutamine, His = histidine, HV 1 - 3 = hypervariable region 1 to 3, Ile = isoleucine, Leu = leucine, LSP = light strand promoter, Met = methionine, mtTF1 = mitochondrial transcription factor 1, NC3 to 4 = noncoding nucleotides 3 to 4, ND 1 - 6 = NADH-Q oxidoreductase subunits 1 to 6, O_L = origin of replication, Phe = phenylalanine, Pro = proline, RNA = ribonucleic acid, Ser = serine, 12S rRNA = 12S ribosomal RNA, 16S rRNA = 16S ribosomal RNA, syn = synonymous, Thr = threonine, $tRNA^{Ala}$ = transfer RNA for alanine, $tRNA^{Arg}$ = transfer RNA for arginine, $tRNA^{Asn}$ = tranfer RNA for asparagine, $tRNA^{Asp}$ = transfer RNA for aspartate, $tRNA^{Cys}$ = transfer RNA for cysteine, $tRNA^{Gln}$ = transfer RNA for Glutamine, $tRNA^{His}$ = transfer RNA for histidine, $tRNA^{Ile}$ = transfer RNA for isoleucine, $tRNA^{Lys}$ = transfer RNA for lysine, $tRNA^{Phe}$ = transfer RNA for phenyalanine, $tRNA^{Ser(AGY)}$ = transfer RNA for serine recognising the codon AGY, $tRNA^{Ser(CUN)}$ = transfer RNA for serine recognising the codon CUN, $tRNA^{Thr}$ = transfer RNA for threonine, $tRNA^{Trp}$ = transfer RNA for tryptophan, $tRNA^{Tyr}$ = transfer RNA for tyrosine, $tRNA^{Val}$ = transfer RNA for valine, Tyr = tyrosine, Val = valine.

APPENDIX B

mtDNA genome sequence data for the African sequences used in the construction of Neighbour-Joining and Maximum Parsimony trees

The complete mtDNA sequences of the African individuals included in this study, together with those of only the coding region were searched from Genbank and assembled for use in making a phylogeographic analysis of the 40 sequences from the Ugandan individuals. The list of the macrohaplogroup L sequences[33] was assembled and the bibliographic information is indicated in 0.

Table A2 mtDNA genome sequence data for the sequences used in the construction of Neighbour-Joining and Maximum Parsimony trees

GenBank® accession number	GI number	Sequence name/ isolate	Ethnicity/ geographic origin[a]	Haplogroup	Number of base pairs	Reference
AF346967	13272612	bam17b	Bamileke/Africa	L3e3b	16,567	Ingman *et al.*, 2000
AF346968	13272626	biak757	Biaka/Africa	L1c	16,567	Ingman *et al.*, 2000
AF346969	13272640	biak776	Biaka/Africa	L1c	16,567	Ingman *et al.*, 2000
AF346976	13272738	eff 004	Effik/Africa	L2a1i	16,570	Ingman *et al.*, 2000
AF346977	13272752	eff007	Effik/Africa	L2a1a2	16,569	Ingman *et al.*, 2000
AF346980	13272794	ewon w77	Ewondo/Africa	L3e3	16,570	Ingman *et al.*, 2000
AF346985	13272864	haus 004	Hausa/Africa	L0a1	16,567	Ingman *et al.*, 2000
AF346986	13272878	ibo 006	Ibo/Africa	L1b	16,567	Ingman *et al.*, 2000
AF346987	13272892	ibo 12	Ibo/Africa	L1c1d	16,566	Ingman *et al.*, 2000
AF346992	13272962	kiku dbs	Kikuyu/Africa	L1c2a1a	16,571	Ingman *et al.*, 2000
AF346994	13272990	Liso f30	Lisongo[e]/Africa	L3e1	16,569	Ingman *et al.*, 2000
AF346995	13273004	mand 318	Mandenka/Africa	L2c3	16,568	Ingman *et al.*, 2000
AF346996	13273018	mbe 225	Mbenzele/Africa	L1c	16,566	Ingman *et al.*, 2000
AF346997	13273032	mbe 240	Mbenzele/Africa	L1c1a1a1a	16,567	Ingman *et al.*, 2000
AF346998	13273046	mbu 1057	Mbuti/Africa	L0a2b	16,561	Ingman *et al.*, 2000
AF346999	13273060	mbu 1058	Mbuti/Africa	L0a	16,562	Ingman *et al.*, 2000
AF347000	13273074	mkam oao	Mkamba/Africa	L3h1a2a	16,569	Ingman *et al.*, 2000
AF347008	13273186	san 1	San/Africa	L0k1	16,569	Ingman *et al.*, 2000
AF347009	13273200	san 2	San/Africa	L0k1	16,567	Ingman *et al.*, 2000
AF347014	13273270	yoru 002	Yoruba/Africa	L3d	16,567	Ingman *et al.*, 2000
AF347015	13273284	yoru 013	Yoruba/Africa	L3e2b1a	16,571	Ingman *et al.*, 2000
AF381981	17985375	441	Mauritania	L2c	16,568	Maca-Meyer *et al.*, 2001
AF381988	17985473	271	Morocco	L0a1b1	16,567	Maca-Meyer *et al.*, 2001
AF381991	17985515	430	Mauritania	L3b	16,567	Maca-Meyer *et al.*, 2001

[33] The bibliographic information pertaining to the sequences was compiled by M. Koekomoer, Centre for Genome Research, North-West University (Potchefstroom Campus), South Africa, and used here with permission.

GenBank® accession number	GI number	Sequence name/ isolate	Ethnicity/ geographic origin[a]	Haplogroup	Number of base pairs	Reference
AF381992	29294531	432	Mauritania	L1c	16,566	Maca-Meyer *et al.*, 2001
AF381994	29294532	451	Mauritania	L1b	16,567	Maca-Meyer *et al.*, 2001
AF381998	17985613	800	Jordan	L3d3	16,567	Maca-Meyer *et al.*, 2001
AY195766	29690686	A11L2b	Africa	L2b	16,567	Mishmar *et al.*, 2003
AY195776	48596191	A9L2a	Africa	L2a1f	16,569	Mishmar *et al.*, 2003
AY195777	29690840	A10L1A2	Africa	L0d2a1	16,557	Mishmar *et al.*, 2003
AY195780	29690882	A2L1	Africa	L1a	16,567	Mishmar *et al.*, 2003
AY195783	48596203	A4L1B2	Africa	L1b1a4	16,567	Mishmar *et al.*, 2003
AY195785	48596208	A6L2C	Africa	L2c	16,567	Mishmar *et al.*, 2003
AY195788	48596215	A5L2A1	Africa	L2a2b1	16,570	Mishmar *et al.*, 2003
AY195789	48596217	A3L1B1	Africa	L1b1	16,567	Mishmar *et al.*, 2003
AY963585	75905883	Tor65	Uganda	L0f2a1	16,569	Macaulay *et al.*, 2005
D38112[b]	644480	SB17	Uganda	L0a2d	16,559	Horai *et al.*, 1995
DQ112686	70955706	AD1	Dominican Republic	L0a2a1a1	15,574	Kivisild *et al.*, 2006
DQ112687	70955708	AD2	Dominican Republic	L0a2a	15,576	Kivisild *et al.*, 2006
DQ112688	70955710	AD3	Dominican Republic	L0a2a1a1	15,574	Kivisild *et al.*, 2006
DQ112689	70955712	AD4	Dominican Republic	L0a2	15,574	Kivisild *et al.*, 2006
DQ112690	70955714	AD5	Dominican Republic	L1b1	15,583	Kivisild *et al.*, 2006
DQ112691	70955716	AD6	Dominican Republic	L1b1	15,583	Kivisild *et al.*, 2006
DQ112692	70955718	AD7	Dominican Republic	L1b1	15,583	Kivisild *et al.*, 2006
DQ112693	70955720	AD8	Dominican Republic	L1b	15,583	Kivisild *et al.*, 2006
DQ112694	70955722	AD9	Dominican Republic	L1c2	15,584	Kivisild *et al.*, 2006
DQ112695	70955724	AD10	Dominican Republic	L1c1b	15,574	Kivisild *et al.*, 2006
DQ112696	70955726	AD11	Dominican Republic	L1c3	15,582	Kivisild *et al.*, 2006
DQ112697	70955728	AD12	Dominican Republic	L1c3	15,582	Kivisild *et al.*, 2006
DQ112698	70955730	AD13	Dominican Republic	L2a1	15,585	Kivisild *et al.*, 2006
DQ112699	70955732	AD14	Dominican Republic	L2a1	15,585	Kivisild *et al.*, 2006
DQ112700	70955734	AD15	Dominican Republic	L2a1c2	15,585	Kivisild *et al.*, 2006
DQ112701	70955736	AD16	Dominican Republic	L2a1i1	15,585	Kivisild *et al.*, 2006
DQ112702	70955738	AD17	Dominican Republic	L2b	15,583	Kivisild *et al.*, 2006
DQ112703	70955740	AD18	Dominican Republic	L2b	15,583	Kivisild *et al.*, 2006
DQ112704	70955742	AD19	Dominican Republic	L2b2	15,585	Kivisild *et al.*, 2006
DQ112705	70955744	AD20	Dominican Republic	L2b	15,585	Kivisild *et al.*, 2006
DQ112706	70955746	AD21	Dominican Republic	L2c	15,583	Kivisild *et al.*, 2006
DQ112707	70955748	AD22	Dominican Republic	L2c	15,583	Kivisild *et al.*, 2006
DQ112708	70955750	AD23	Dominican Republic	L2d	15,585	Kivisild *et al.*, 2006
DQ112709	70955752	AD24	Dominican Republic	L2d	15,591	Kivisild *et al.*, 2006
DQ112710	70955754	AD25	Dominican Republic	L3b	15,582	Kivisild *et al.*, 2006
DQ112711	70955756	AD26	Dominican Republic	L3b	15,582	Kivisild *et al.*, 2006
DQ112712	70955758	AD27	Dominican Republic	L3b	15,582	Kivisild *et al.*, 2006
DQ112713	70955760	AD28	Dominican Republic	L3d	15,583	Kivisild *et al.*, 2006
DQ112714	70955762	AD29	Dominican Republic	L3d1d	15,583	Kivisild *et al.*, 2006
DQ112715	70955764	AD30	Dominican Republic	L3d1	15,583	Kivisild *et al.*, 2006
DQ112716	70955766	AD31	Dominican Republic	L3d3a	15,583	Kivisild *et al.*, 2006
DQ112717	70955768	AD32	Dominican Republic	L3f1	15,582	Kivisild *et al.*, 2006
DQ112718	70955770	AD33	Dominican Republic	L3h	15,583	Kivisild *et al.*, 2006
DQ112719	70955772	AD34	Dominican Republic	L3e1	15,585	Kivisild *et al.*, 2006
DQ112720	70955774	AD35	Dominican Republic	L3e1	15,585	Kivisild *et al.*, 2006
DQ112721	70955776	AD36	Dominican Republic	L3e1	15,585	Kivisild *et al.*, 2006
DQ112722	70955778	AD37	Dominican Republic	L3e4	15,583	Kivisild *et al.*, 2006

GenBank® accession number	GI number	Sequence name/ isolate	Ethnicity/ geographic origin[a]	Haplogroup	Number of base pairs	Reference
DQ112723	70955780	AF83	Mossi/Burkina Faso	L2a1	15,583	Kivisild *et al.*, 2006
DQ112724	70955782	AF84	Mossi/Burkina Faso	L2a1	15,585	Kivisild *et al.*, 2006
DQ112725	70955784	AF82	Mossi/Burkina Faso	L2a1	15,583	Kivisild *et al.*, 2006
DQ112726	70955786	AF65	Rimaibe/Burkina Faso	L3e1	15,585	Kivisild *et al.*, 2006
DQ112727	70955788	AF91	Rimaibe/Burkina Faso	L1b	15,583	Kivisild *et al.*, 2006
DQ112728	70955790	AF66	Rimaibe/Burkina Faso	L3e1	15,585	Kivisild *et al.*, 2006
DQ112729	70955792	AF63	Rimaibe/Burkina Faso	L3e1	15,585	Kivisild *et al.*, 2006
DQ112730	70955794	AF79	Rimaibe/Burkina Faso	L2a1c2	15,585	Kivisild *et al.*, 2006
DQ112731	70955796	AF80	Rimaibe/Burkina Faso	L2a1	15,585	Kivisild *et al.*, 2006
DQ112732	70955798	AF92	Rimaibe/Burkina Faso	L0a1	15,583	Kivisild *et al.*, 2006
DQ112733	70955800	AF90	Mossi/Burkina Faso	L1b1	15,583	Kivisild *et al.*, 2006
DQ112734	70955802	AF67	Rimaibe/Burkina Faso	L3e2b1a	15,585	Kivisild *et al.*, 2006
DQ112735	70955804	AF81	Rimaibe/Burkina Faso	L2c	15,583	Kivisild *et al.*, 2006
DQ112736	70955806	AF86	Rimaibe/Burkina Faso	L3f1	15,582	Kivisild *et al.*, 2006
DQ112737	70955808	AF87	Foulbe/Burkina Faso	L1b	15,583	Kivisild *et al.*, 2006
DQ112738	70955810	AF68	Foulbe/Burkina Faso	L3b1b	15,582	Kivisild *et al.*, 2006
DQ112739	70955812	AF69	Foulbe/Burkina Faso	L3d1b	15,583	Kivisild *et al.*, 2006
DQ112740	70955814	AF64	Foulbe/Burkina Faso	L3e1	15,585	Kivisild *et al.*, 2006
DQ112741	70955816	AF88	Foulbe/Burkina Faso	L1b1a6	15,583	Kivisild *et al.*, 2006
DQ112742	70955818	AF70	Foulbe/Burkina Faso	L3b	15,582	Kivisild *et al.*, 2006
DQ112743	70955820	AF71	Foulbe/Burkina Faso	L3b	15,582	Kivisild *et al.*, 2006
DQ112744	70955822	AF72	Foulbe/Burkina Faso	L3d1	15,583	Kivisild *et al.*, 2006
DQ112745	70955824	AF73	Foulbe/Burkina Faso	L3b	15,582	Kivisild *et al.*, 2006
DQ112746	70955826	AF74	Mossi/Burkina Faso	L3b2	15,584	Kivisild *et al.*, 2006
DQ112747	70955828	AF75	Mossi/Burkina Faso	L3e3	15,585	Kivisild *et al.*, 2006
DQ112748	70955830	AF76	Mossi/Burkina Faso	L3b	15,582	Kivisild *et al.*, 2006
DQ112756	70955846	AF77	Mossi/Burkina Faso	L3d2	15,583	Kivisild *et al.*, 2006
DQ112757	70955848	AF78	Mossi/Burkina Faso	L3e	15,583	Kivisild *et al.*, 2006
DQ112758	70955850	AF85	Mossi/Burkina Faso	L2a1	15,582	Kivisild *et al.*, 2006
DQ112759	70955852	AF89	Foulbe/Burkina Faso	L1b	15,583	Kivisild *et al.*, 2006
DQ112777	70955888	AF16	Nuba/Sudan	L0a1b	15,583	Kivisild *et al.*, 2006
DQ112778	70955890	AF14	Nuba/Sudan	L2a2a	15,585	Kivisild *et al.*, 2006
DQ112782	70955898	AF10	Berta/Ethiopia	L2a1	15,585	Kivisild *et al.*, 2006
DQ112792	70955918	AF41	Eritrea	Pre-HV1	15,585	Kivisild *et al.*, 2006
DQ112794	70955922	AF06	Lissongo[e]/Africa	L1c2	15,589	Kivisild *et al.*, 2006
DQ112796	70955926	AF07	Lissongo/Africa	L4b2b	15,585	Kivisild *et al.*, 2006
DQ112829	70955992	AF53	Ghana	L1b	15,583	Kivisild *et al.*, 2006
DQ112830	70955994	AF54	Zulu/Africa	L0d2	15,582	Kivisild *et al.*, 2006
DQ112844	70956022	AF17	NaSan/Africa	L0d1	15,584	Kivisild *et al.*, 2006
DQ112845	70956024	AF18	Khoisan/Africa	L4g	15,585	Kivisild *et al.*, 2006
DQ112847	70956028	AF19	Herero/Africa	L3d3a	15,583	Kivisild *et al.*, 2006
DQ112848	70956030	AF20	Sotho/Africa	L0a2a2	15,574	Kivisild *et al.*, 2006
DQ112849	70956032	AF21	Pedi/Africa	L2a1	15,585	Kivisild *et al.*, 2006
DQ112850	70956034	AF22	Tswana/Africa	L0d2	15,584	Kivisild *et al.*, 2006
DQ112851	70956036	AF23	Ambo/Africa	L3d3a	15,583	Kivisild *et al.*, 2006
DQ112852	70956038	AF24	San/Namibia	L0d1	15,584	Kivisild *et al.*, 2006
DQ112853	70956040	AF26	Herero/Africa	L3d3a	15,583	Kivisild *et al.*, 2006
DQ112854	70956042	AF25	Zulu/Africa	L0d1	15,584	Kivisild *et al.*, 2006
DQ112855	70956044	AF27	Tswana/Africa	L0d1	15,583	Kivisild *et al.*, 2006
DQ112856	70956046	AF55	San/Namibia	L0d1	15,582	Kivisild *et al.*, 2006

GenBank® accession number	GI number	Sequence name/ isolate	Ethnicity/ geographic origin[a]	Haplogroup	Number of base pairs	Reference
DQ112857	70956048	AF28	Herero/Africa	L3d3a	15,583	Kivisild *et al.*, 2006
DQ112883	70956100	AF29	Mandenka/Africa	L2c	15,583	Kivisild *et al.*, 2006
DQ112884	70956102	AF30	Mandenka/Africa	L3d	15,583	Kivisild *et al.*, 2006
DQ112899	70956132	AF09	Khoisan/Africa	L0d1	15,582	Kivisild *et al.*, 2006
DQ112900	70956134	AF08	Khoisan/Africa	L2a1b1	15,585	Kivisild *et al.*, 2006
DQ112901	70956136	AF102	Pygmy/Zaire	L0a2b	15,576	Kivisild *et al.*, 2006
DQ112902	70956138	AF02	Biaka Pygmy/Africa	L1c4	15,587	Kivisild *et al.*, 2006
DQ112903	70956140	AF31	Pygmy/Zaire	L2a2	15,585	Kivisild *et al.*, 2006
DQ112904	70956142	AF05	Pygmy/Zaire	L0a2b	15,574	Kivisild *et al.*, 2006
DQ112905	70956144	AF32	Pygmy/Zaire	L2a2b	15,585	Kivisild *et al.*, 2006
DQ112906	70956146	AF33	Pygmy/Zaire	L0a2b	15,576	Kivisild *et al.*, 2006
DQ112907	70956148	AF34	Pygmy/CAR	L1c1a2b	15,582	Kivisild *et al.*, 2006
DQ112908	70956150	AF35	Pygmy/CAR	L1c1a	15,582	Kivisild *et al.*, 2006
DQ112909	70956152	AF03	Pygmy/Zaire	L2a2a1	15,585	Kivisild *et al.*, 2006
DQ112910	70956154	AF93	Pygmy/CAR	L1c1a	15,582	Kivisild *et al.*, 2006
DQ112911	70956156	AF52	Pygmy/Zaire	L2a2	15,585	Kivisild *et al.*, 2006
DQ112912	70956158	AF94	Pygmy/CAR	L1c1a	15,582	Kivisild *et al.*, 2006
DQ112913	70956160	AF95	Pygmy/CAR	L1c1a	15,582	Kivisild *et al.*, 2006
DQ112914	70956162	AF01	Biaka Pygmy/Africa	L1c2	15,587	Kivisild *et al.*, 2006
DQ112915	70956164	AF97	Pygmy/Zaire	L0a2b1	15,576	Kivisild *et al.*, 2006
DQ112916	70956166	AF36	Pygmy/Zaire	L2a2	15,585	Kivisild *et al.*, 2006
DQ112917	70956168	AF37	Pygmy/Zaire	L0a2b	15,576	Kivisild *et al.*, 2006
DQ112918	70956170	AF98	Pygmy/Zaire	L0a2b1	15,576	Kivisild *et al.*, 2006
DQ112919	70956172	AF38	Pygmy/CAR	L1c1a	15,582	Kivisild *et al.*, 2006
DQ112920	70956174	AF39	Pygmy/CAR	L1c1a	15,582	Kivisild *et al.*, 2006
DQ112921	70956176	AF99	Pygmy/Zaire	L5a1c	15,586	Kivisild *et al.*, 2006
DQ112922	70956178	AF100	Pygmy/Zaire	L5a1c	15,586	Kivisild *et al.*, 2006
DQ112923	70956180	AF40	Pygmy/CAR	L0a2a1	15,574	Kivisild *et al.*, 2006
DQ112924	70956182	AF101	Pygmy/Zaire	L2a2	15,585	Kivisild *et al.*, 2006
DQ112925	70956184	AF96	Pygmy/Zaire	L1c1a2a	15,582	Kivisild *et al.*, 2006
DQ112926	70956186	AF15	Ethiopia	M1	15,585	Kivisild *et al.*, 2006
DQ112932	70956198	AF42	Ethiopia	U1a	15,584	Kivisild *et al.*, 2006
DQ112933	70956200	AF43	Ethiopia	M1	15,585	Kivisild *et al.*, 2006
DQ112934	70956202	AF44	Ethiopia	L0f	15,585	Kivisild *et al.*, 2006
DQ112949	70956232	AF50	Sudan	L3d5	15,585	Kivisild *et al.*, 2006
DQ112953	---	AS46	Asian	L0a2a	15,574	Kivisild *et al.*, 2006
DQ112956	70956246	AF45	Sudan	L3j	15,583	Kivisild *et al.*, 2006
DQ112957	70956248	AF46	Sudan	L3h1a1	15,585	Kivisild *et al.*, 2006
DQ112958	70956250	AF47	Sudan	L3i1a	15,585	Kivisild *et al.*, 2006
DQ112959	70956252	AF48	Sudan	L0a1	15,583	Kivisild *et al.*, 2006
DQ112960	70956254	AF49	Sudan	L3i1a	15,585	Kivisild *et al.*, 2006
DQ112961	70956256	AF51	Sudan	L3h1a1	15,585	Kivisild *et al.*, 2006
DQ112962	70956258	AF13	Tuareg/Africa	H1	15,585	Kivisild *et al.*, 2006
DQ282505	82494118	L1b-3-01	Hispanic	L1b1a7	16,567	Just *et al.*, 2008
DQ282506	82494132	L1b-3-02	Hispanic	L1b	16,567	Just *et al.*, 2008
DQ282507	82494146	L3e-5-01	Hispanic	L3e1e	16,563	Just *et al.*, 2008
DQ304897	83265615	L0a-1-01	African American	L0a1b2	16,567	Just *et al.*, 2008
DQ304898	83265629	L0a-1-02	African American	L0a	16,567	Just *et al.*, 2008
DQ304899	83265643	L0a-1-03	African American	L0a1b1a	16,567	Just *et al.*, 2008
DQ304900	83265657	L0a-1-04	African American	L0a1b1	16,567	Just *et al.*, 2008

GenBank® accession number	GI number	Sequence name/ isolate	Ethnicity/ geographic origin[a]	Haplogroup	Number of base pairs	Reference
DQ304901	83265671	L0a-2-01	African American	L0a	16,568	Just *et al*., 2008
DQ304902	83265685	L0a-2-02	African American	L0a	16,567	Just *et al*., 2008
DQ304903	83265699	L0a-2-03	African American	L0a1a2	16,567	Just *et al*., 2008
DQ304904	83265713	L0a-2-04	African American	L0a	16,567	Just *et al*., 2008
DQ304905	83265727	L1b-1-01	African American	L1b1a3a	16,569	Just *et al*., 2008
DQ304906	83265741	L1b-1-02	African American	L1b1a3a1	16,567	Just *et al*., 2008
DQ304907	83265755	L1b-1-03	African American	L1b	16,567	Just *et al*., 2008
DQ304908	83265769	L1b-1-04	African American	L1b1a3	16,568	Just *et al*., 2008
DQ304909	83265783	L1b-1-05	African American	L1b	16,568	Just *et al*., 2008
DQ304910	83265797	L1b-1-06	African American	L1b	16,568	Just *et al*., 2008
DQ304911	83265811	L1b-1-07	African American	L1b1a3a1	16,567	Just *et al*., 2008
DQ304912	83265825	L1b-1-08	African American	L1b	16,568	Just *et al*., 2008
DQ304913	83265839	L1b-1-09	African American	L1b	16,569	Just *et al*., 2008
DQ304914	83265853	L1b-1-10	African American	L1b1a3a	16,568	Just *et al*., 2008
DQ304915	83265867	L1b-1-11	African American	L1b	16,567	Just *et al*., 2008
DQ304916	83265881	L1b-1-12	African American	L1b1a3	16,568	Just *et al*., 2008
DQ304917	83265895	L1b-1-13	African American	L1b	16,568	Just *et al*., 2008
DQ304918	83265909	L1b-1-14	African American	L1b	16,567	Just *et al*., 2008
DQ304919	83265923	L1b-2-01	African American	L1b	16,567	Just *et al*., 2008
DQ304920	83265937	L1b-2-02	African American	L1b	16,567	Just *et al*., 2008
DQ304921	83265951	L1b-2-03	African American	L1b	16,572	Just *et al*., 2008
DQ304922	83265965	L1b-2-04	African American	L1b	16,567	Just *et al*., 2008
DQ304923	83265979	L1b-2-05	African American	L1b1a7	16,567	Just *et al*., 2008
DQ304924	83265993	L2a-1-01	African American	L2a1a	16,570	Just *et al*., 2008
DQ304925	83266007	L2a-1-02	African American	L2a1a1	16,574	Just *et al*., 2008
DQ304926	83266021	L2a-1-03	African American	L2a	16,571	Just *et al*., 2008
DQ304927	83266035	L2a-1-04	African American	L2a	16,569	Just *et al*., 2008
DQ304928	83266049	L2a-1-05	African American	L2a	16,568	Just *et al*., 2008
DQ304929	83266063	L2a-1-06	African American	L2a	16,567	Just *et al*., 2008
DQ304930	83266077	L2a-1-07	African American	L2a1e1	16,570	Just *et al*., 2008
DQ304931	83266091	L2a-1-08	African American	L2a	16,567	Just *et al*., 2008
DQ304932	83266105	L2a-1-09	African American	L2a	16,569	Just *et al*., 2008
DQ304933	83266119	L2a-1-10	African American	L2a1a1	16,571	Just *et al*., 2008
DQ304934	83266133	L2a-2-01	African American	L2a	16,573	Just *et al*., 2008
DQ304935	83266147	L2a-2-02	African American	L2a	16,571	Just *et al*., 2008
DQ304936	83266161	L2a-2-03	African American	L2a	16,570	Just *et al*., 2008
DQ304937	83266175	L2a-2-04	African American	L2a	16,572	Just *et al*., 2008
DQ304938	83266189	L2a-3-01	African American	L2a	16,560	Just *et al*., 2008
DQ304939	83266203	L2a-3-02	African American	L2a	16,570	Just *et al*., 2008
DQ304940	83266217	L2a-3-03	African American	L2a	16,559	Just *et al*., 2008
DQ304941	83266231	L2a-3-04	African American	L2a	16,569	Just *et al*., 2008
DQ304942	83266245	L2a-4-01	African American	L2a	16,568	Just *et al*., 2008
DQ304943	83266259	L2a-4-02	African American	L2a	16,568	Just *et al*., 2008
DQ304944	83266273	L2a-4-03	African American	L2a	16,569	Just *et al*., 2008
DQ304945	83266287	L2a-4-04	African American	L2a1e	16,570	Just *et al*., 2008
DQ304946	83266301	L2a-4-05	African American	L2a1e1	16,570	Just *et al*., 2008
DQ304947	83266315	L2a-4-06	African American	L2a	16,570	Just *et al*., 2008
DQ304948	83266329	L2a-4-07	African American	L2a	16,569	Just *et al*., 2008
DQ304949	83266343	L2a-4-09	African American	L2a	16,568	Just *et al*., 2008
DQ304950	83266357	L2a-4-10	African American	L2a	16,567	Just *et al*., 2008

GenBank® accession number	GI number	Sequence name/ isolate	Ethnicity/ geographic origin[a]	Haplogroup	Number of base pairs	Reference
DQ304951	83266371	L2a-4-11	African American	L2a1c4	16,568	Just *et al*., 2008
DQ304952	83266385	L2a-5-01	African American	L2a	16,570	Just *et al*., 2008
DQ304953	83266399	L2a-5-02	African American	L2a	16,571	Just *et al*., 2008
DQ304954	83266413	L2a-5-03	African American	L2a	16,571	Just *et al*., 2008
DQ304955	83266427	L2a-5-04	African American	L2a1f1	16,570	Just *et al*., 2008
DQ304956	83266441	L2a-5-05	African American	L2a	16,570	Just *et al*., 2008
DQ304957	83266455	L2a-5-06	African American	L2a1f	16,568	Just *et al*., 2008
DQ304958	83266469	L2a-5-07	African American	L2a	16,570	Just *et al*., 2008
DQ304959	83266483	L2a-5-08	African American	L2a	16,571	Just *et al*., 2008
DQ304960	83266497	L2a-5-09	African American	L2a	16,570	Just *et al*., 2008
DQ304961	83266511	L2a-5-10	African American	L2a1f1	16,571	Just *et al*., 2008
DQ304962	83266525	L2a-5-11	African American	L2a	16,571	Just *et al*., 2008
DQ304963	83266539	L2a-5-12	African American	L2a	16,570	Just *et al*., 2008
DQ304964	83266553	L2a-5-13	African American	L2a	16,570	Just *et al*., 2008
DQ304965	83266567	L2a-5-14	African American	L2a	16,568	Just *et al*., 2008
DQ304966	83266581	L2a-5-15	African American	L2a	16,570	Just *et al*., 2008
DQ304967	83266595	L2a-5-16	African American	L2a	16,571	Just *et al*., 2008
DQ304968	83266609	L2a-6-01	African American	L2a	16,570	Just *et al*., 2008
DQ304969	83266623	L2a-6-02	African American	L2a	16,569	Just *et al*., 2008
DQ304970	83266637	L2a-6-03	African American	L2a	16,569	Just *et al*., 2008
DQ304971	83266651	L2a-6-04	African American	L2a	16,571	Just *et al*., 2008
DQ304972	83266665	L2a-6-05	African American	L2a1a2b	16,571	Just *et al*., 2008
DQ304973	83266679	L2a-6-06	African American	L2a	16,570	Just *et al*., 2008
DQ304974	83266693	L2a-6-07	African American	L2a	16,569	Just *et al*., 2008
DQ304975	83266707	L2a-6-08	African American	L2a1a2a	16,570	Just *et al*., 2008
DQ304976	83266721	L2a-6-09	African American	L2a1a2a1	16,569	Just *et al*., 2008
DQ304977	83266735	L2a-6-12	African American	L2a	16,569	Just *et al*., 2008
DQ304978	83266749	L2b-1-01	African American	L2b1a3	16,568	Just *et al*., 2008
DQ304979	83266763	L2b-1-02	African American	L2b	16,567	Just *et al*., 2008
DQ304980	83266777	L2b-1-03	African American	L2b	16,567	Just *et al*., 2008
DQ304981	83266791	L2b-1-04	African American	L2b1a	16,567	Just *et al*., 2008
DQ304982	83266805	L2b-1-05	African American	L2b	16,567	Just *et al*., 2008
DQ304983	83266819	L2b-1-07	African American	L2b	16,567	Just *et al*., 2008
DQ304984	83266833	L2b-1-08	African American	L2b	16,567	Just *et al*., 2008
DQ304985	83266847	L2b-1-09	African American	L2b1a2	16,567	Just *et al*., 2008
DQ304986	83266861	L2c-1-01	African American	L2c2	16,567	Just *et al*., 2008
DQ304987	83266875	L2c-1-02	African American	L2c	16,568	Just *et al*., 2008
DQ304988	83266889	L2c-1-03	African American	L2c2a	16,568	Just *et al*., 2008
DQ304989	83266903	L2c-1-04	African American	L2c	16,567	Just *et al*., 2008
DQ304990	83266917	L3b-1-01	African American	L3b1a3	16,566	Just *et al*., 2008
DQ304991	83266931	L3b-1-02	African American	L3b1a	16,566	Just *et al*., 2008
DQ304992	83266945	L3b-1-03	African American	L3b	16,566	Just *et al*., 2008
DQ304993	83266959	L3b-2-01	African American	L3b	16,566	Just *et al*., 2008
DQ304994	83266973	L3b-2-02	African American	L3b	16,566	Just *et al*., 2008
DQ304995	83266987	L3b-2-03	African American	L3b1a1	16,566	Just *et al*., 2008
DQ304996	83267001	L3b-2-04	African American	L3b	16,566	Just *et al*., 2008
DQ304997	83267015	L3b-2-05	African American	L3b	16,566	Just *et al*., 2008
DQ304998	83267029	L3e-1-01	African American	L3e	16,570	Just *et al*., 2008
DQ304999	83267043	L3e-1-02	African American	L3e	16,572	Just *et al*., 2008
DQ305000	83267057	L3e-1-03	African American	L3e	16,571	Just *et al*., 2008

GenBank® accession number	GI number	Sequence name/ isolate	Ethnicity/ geographic origin[a]	Haplogroup	Number of base pairs	Reference
DQ305001	83267071	L3e-1-04	African American	L3e	16,571	Just *et al.*, 2008
DQ305002	83267085	L3e-1-05	African American	L3e	16,571	Just *et al.*, 2008
DQ305003	83267099	L3e-1-06	African American	L3e	16,570	Just *et al.*, 2008
DQ305004	83267113	L3e-1-07	African American	L3e	16,570	Just *et al.*, 2008
DQ305005	83267127	L3e-1-08	African American	L3e	16,571	Just *et al.*, 2008
DQ305006	83267141	L3e-1-09	African American	L3e	16,570	Just *et al.*, 2008
DQ305007	83267155	L3e-1-10	African American	L3e	16,572	Just *et al.*, 2008
DQ305008	83267169	L3e-1-11	African American	L3e2b3	16,572	Just *et al.*, 2008
DQ305009	83267183	L3e-1-12	African American	L3e2b3	16,571	Just *et al.*, 2008
DQ305010	83267197	L3e-2-01	African American	L3e	16,567	Just *et al.*, 2008
DQ305011	83267211	L3e-2-02	African American	L3e3b	16,567	Just *et al.*, 2008
DQ305012	83267225	L3e-2-03	African American	L3e3a	16,572	Just *et al.*, 2008
DQ305013	83267239	L3e-2-04	African American	L3e3b1	16,568	Just *et al.*, 2008
DQ305014	83267253	L3e-2-05	African American	L3e2	16,572	Just *et al.*, 2008
DQ305015	83267267	L3e-2-06	African American	L3e	16,567	Just *et al.*, 2008
DQ305016	83267281	L3e-2-07	African American	L3e	16,568	Just *et al.*, 2008
DQ305017	83267295	L3e-2-08	African American	L3e3b1	16,567	Just *et al.*, 2008
DQ305018	83267309	L3e-3-01	African American	L3e	16,572	Just *et al.*, 2008
DQ305019	83267323	L3e-3-02	African American	L3e2b	16,571	Just *et al.*, 2008
DQ305020	83267337	L3e-3-03	African American	L3e	16,572	Just *et al.*, 2008
DQ305021	83267351	L3e-4-01	African American	L3e	16,570	Just *et al.*, 2008
DQ305022	83267365	L3e-4-02	African American	L3e2b1	16,572	Just *et al.*, 2008
DQ305023	83267379	L3e-4-03	African American	L3e	16,568	Just *et al.*, 2008
DQ305024	83267393	L3e-6-01	African American	L3e	16,570	Just *et al.*, 2008
DQ305025	83267407	L3e-6-02	African American	L3e2a1b	16,570	Just *et al.*, 2008
DQ305026	83267421	L3e-6-03	African American	L3e2a1a	16,570	Just *et al.*, 2008
DQ305027	83267435	L3e-6-04	African American	L3e2a1a	16,570	Just *et al.*, 2008
DQ305028	83267449	L3e-6-05	African American	L3e	16,569	Just *et al.*, 2008
DQ305029	83267463	L3e-6-06	African American	L3e2a1b	16,569	Just *et al.*, 2008
DQ305030	83267477	L3e-6-07	African American	L3e2a1b1	16,569	Just *et al.*, 2008
DQ305031	83267491	L3e-6-08	African American	L3e2a1b1	16,569	Just *et al.*, 2008
DQ305032	83267505	L3f-1-01	African American	L3f	16,568	Just *et al.*, 2008
DQ305033	83267519	L3f-1-02	African American	L3f	16,569	Just *et al.*, 2008
DQ305034	83267533	L3f-1-03	African American	L3f	16,569	Just *et al.*, 2008
DQ305035	83267547	L3f-1-04	African American	L3f	16,569	Just *et al.*, 2008
DQ305036	83267561	L3f-1-05	African American	L3f1b1	16,568	Just *et al.*, 2008
DQ341058	84682334	Tor29	Dominican Republic	L0a2a2a	16,558	Torroni *et al.*, 2006
DQ341059	84682348	Tor7	Dominican Republic	L1c2b1a	16,584	Torroni *et al.*, 2006
DQ341060	84682362	Tor74	Ethiopia	L5a1a	16,570	Torroni *et al.*, 2006
DQ341061	84682376	Tor68	Ethiopia	L5c1	16,571	Torroni *et al.*, 2006
DQ341062	84682390	Tor38	Dominican Republic	L2d1a	16,569	Torroni *et al.*, 2006
DQ341063	85541074	Tor39	Ethiopia	L6b	16,572	Torroni *et al.*, 2006
DQ341064	84682418	Tor66	Ethiopia	L4a1a	16,569	Torroni *et al.*, 2006
DQ341065	84682432	Tor71	Ethiopia	L4b2a1	16,567	Torroni *et al.*, 2006
DQ341066	84682446	Tor72	Ethiopia	L3x2a1	16,574	Torroni *et al.*, 2006
DQ341067	84682460	Tor82	Ethiopia	L3x1	16,570	Torroni *et al.*, 2006
DQ341068	84682474	Tor70	Ethiopia	L3i2	16,566	Torroni *et al.*, 2006
DQ341069	84682488	Tor69	Ethiopia	L3i1b	16,571	Torroni *et al.*, 2006
DQ341070	84682502	Tor79	Nigeria	L3e5	16,567	Torroni *et al.*, 2006
DQ341071	84682516	Tor5	Dominican Republic	L3e	16,570	Torroni *et al.*, 2006

GenBank® accession number	GI number	Sequence name/ isolate	Ethnicity/ geographic origin[a]	Haplogroup	Number of base pairs	Reference
DQ341072	84682530	Tor4	Dominican Republic	L3d1b1	16,567	Torroni *et al.*, 2006
DQ341073	84682544	Tor6	Dominican Republic	L3b	16,567	Torroni *et al.*, 2006
DQ341074	84682558	Tor84	Ethiopia	L3c	16,568	Torroni *et al.*, 2006
DQ341075	84682572	Tor80	Nigeria	L3f	16,568	Torroni *et al.*, 2006
DQ341076	84682586	Tor83	Ethiopia	L3f2	16,569	Torroni *et al.*, 2006
DQ341077	84682600	Tor75	Ethiopia	L3f1b	16,569	Torroni *et al.*, 2006
DQ341078	84682614	Tor81	Nigeria	L3f1a	16,568	Torroni *et al.*, 2006
DQ341079	84682628	Tor31	Ethiopia	L3h1b1a	16,568	Torroni *et al.*, 2006
DQ341080	84682642	Tor67	Ethiopia	L3h2	16,567	Torroni *et al.*, 2006
DQ341081	84682656	Tor73	Ethiopia	L3a	16,567	Torroni *et al.*, 2006
EF184580	133854527	TZSW053_L5	Sandawe/Tanzania	L5	16,559	Gonder *et al.*, 2007
EF184581	133854541	TZSW087_L5	Sandawe/Tanzania	L5	16,560	Gonder *et al.*, 2007
EF184582	133854555	TZMW009_L5	Mbugwe/Tanzania	L5	16,570	Gonder *et al.*, 2007
EF184583	133854569	TZSW086_L5	Sandawe/Tanzania	L5	16,559	Gonder *et al.*, 2007
EF184584	133854583	TZSW055_L5	Sandawe/Tanzania	L5	16,560	Gonder *et al.*, 2007
EF184585	133854597	Tzsw109_L0d	Sandawe/Tanzania	L0d	16,567	Gonder *et al.*, 2007
EF184586	133854611	Tzsw084_L0d	Sandawe/Tanzania	L0d	16,568	Gonder *et al.*, 2007
EF184587	133854625	Tzsw068_L0d	Sandawe/Tanzania	L0d	16,569	Gonder *et al.*, 2007
EF184588	133854639	Tzsw010_L0d	Sandawe/Tanzania	L0d	16,569	Gonder *et al.*, 2007
EF184589	133854653	Tzbg031_L0d	Burunge/Tanzania	L0d	16,569	Gonder *et al.*, 2007
EF184590	133854667	sanC5_L0d	San/South Africa	L0d	16,568	Gonder *et al.*, 2007
EF184591	133854681	San_54_L0d	San/South Africa	L0d	16,568	Gonder *et al.*, 2007
EF184592	133854695	sanC6_L0d2	San/South Africa	L0d2	16,566	Gonder *et al.*, 2007
EF184593	133854709	sanC2_L0d2	San/South Africa	L0d2	16,566	Gonder *et al.*, 2007
EF184594	133854723	San_107_L0d2	San/South Africa	L0d2	16,571	Gonder *et al.*, 2007
EF184595	133854737	TZWF021_L0f	Wafiome/Tanzania	L0f	16,568	Gonder *et al.*, 2007
EF184596	133854751	TZAK002_L0f	Akie/Tanzania	L0f	16,568	Gonder *et al.*, 2007
EF184597	133854765	TZTR029_L0f	Turu/Tanzania	L0f	16,570	Gonder *et al.*, 2007
EF184598	133854779	TZBG034_L0f	Burunge/Tanzania	L0f	16,566	Gonder *et al.*, 2007
EF184599	133854793	TZDT068_L0f	Datog/Tanzania	L0f	16,570	Gonder *et al.*, 2007
EF184600	133854807	TZGG017_L0f	Gogo/Tanzania	L0f	16,568	Gonder *et al.*, 2007
EF184601	133854821	TZDT067_L0a	Datog/Tanzania	L0a	16,568	Gonder *et al.*, 2007
EF184602	133854835	TZSW038_L0a	Sandawe/Tanzania	L0a	16,559	Gonder *et al.*, 2007
EF184603	133854849	TZPR030_L0a	Pare/Tanzania	L0a	16,559	Gonder *et al.*, 2007
EF184604	133854863	Tzsw050_L0a	Sandawe/Tanzania	L0a	16,557	Gonder *et al.*, 2007
EF184605	133854877	tztr002_L0a	Turu/Tanzania	L0a	16,559	Gonder *et al.*, 2007
EF184606	133854891	Tzsw005_L0a	Sandawe/Tanzania	L0a	16,559	Gonder *et al.*, 2007
EF184607	133854905	tzsw015_L0a	Sandawe/Tanzania	L0a	16,558	Gonder *et al.*, 2007
EF184608	133854919	TZTR001_L0a	Turu/Tanzania	L0a	16,560	Gonder *et al.*, 2007
EF184609	133854933	San_102_L0k	San/South Africa	L0k	16,568	Gonder *et al.*, 2007
EF184610	133854946	San_11_L0k	San/South Africa	L0k	16,564	Gonder *et al.*, 2007
EF184611	133854959	sanC3_L0k	San/South Africa	L0k	16,568	Gonder *et al.*, 2007
EF184612	133854973	TZTR031_L1c	Turu/Tanzania	L1c	16,567	Gonder *et al.*, 2007
EF184613	133854987	CAPL081_L1c	Bakola/Cameroon	L1c	16,567	Gonder *et al.*, 2007
EF184614	133855001	CAPL079_L1c	Bakola/Cameroon	L1c	16,567	Gonder *et al.*, 2007
EF184615	133855015	CAPL029_L1c	Bakola/Cameroon	L1c	16,567	Gonder *et al.*, 2007
EF184616	133855029	CAPL006_L1c	Bakola/Cameroon	L1c	16,567	Gonder *et al.*, 2007
EF184617	133855043	Tzbg037_L2a	Burunge/Tanzania	L2a	16,570	Gonder *et al.*, 2007
EF184618	133855057	Tzhz037_L2a	Hadza/Tanzania	L2a	16,570	Gonder *et al.*, 2007
EF184619	133855070	Tzhz077_L2a	Hadza/Tanzania	L2a	16,570	Gonder *et al.*, 2007

GenBank® accession number	GI number	Sequence name/ isolate	Ethnicity/ geographic origin[a]	Haplogroup	Number of base pairs	Reference
EF184620	133855083	tztr037_L2a	Turu/Tanzania	L2a	16,571	Gonder *et al.*, 2007
EF184621	133855097	tzms014_L2	Maasai/Tanzania	L2	16,569	Gonder *et al.*, 2007
EF184622	133855111	TZHZ251_L3h	Hadza/Tanzania	L3h	16,570	Gonder *et al.*, 2007
EF184623	133855125	Tzbg004_L3	Burunge/Tanzania	L3	16,568	Gonder *et al.*, 2007
EF184624	133855139	Tzhz032_L3f	Hadza/Tanzania	L3f	16,570	Gonder *et al.*, 2007
EF184625	133855153	Tzhz061_L3f	Hadza/Tanzania	L3f	16,572	Gonder *et al.*, 2007
EF184627	133855181	Tzsw011_L3g	Sandawe/Tanzania	L4g	16,570	Gonder *et al.*, 2007
EF184628	133855195	tztr007_L3d	Turu/Tanzania	L3d	16,568	Gonder *et al.*, 2007
EF184629	133855209	tztr023_L3g	Turu/Tanzania	L4g	16,570	Gonder *et al.*, 2007
EF184630	133855223	Tztr024_L3	Turu/Tanzania	L3	16,568	Gonder *et al.*, 2007
EF184631	133855237	TZRG044_L3	Rangi/Tanzania	L3	16,569	Gonder *et al.*, 2007
EF184632	133855251	TZIQ006_L3	Iraqw/Tanzania	L3	16,569	Gonder *et al.*, 2007
EF184633	133855265	sanC1_L3	San/South Africa	L3	16,570	Gonder *et al.*, 2007
EF184639	133855349	TZIQ053_L3	Iraqw/Tanzania	L3	16,570	Gonder *et al.*, 2007
EF184640	133855363	TZHZ081_L3	Hadza/Tanzania	L3	16,569	Gonder *et al.*, 2007
EF184641	133855376	TZMW031_L3	Mbugwe/Tanzania	L3	16,568	Gonder *et al.*, 2007
EF657286	151328456	MTDNA149	African descent	L1a	15,437	Herrnstadt *et al.*, 2002
EF657743	151334843	MTDNA560	African descent	L0a1b2a	15,446	Herrnstadt *et al.*, 2002
NC012920	113200490	rCRS	European	H	16,569	Andrews *et al.*, 1999
NC001643[d]	5835121	Pan trog	Pan troglodytes	---	16,554	Horai *et al.*, 1995
NC001807[c]	17981852	---	Yoruba/Africa	---	16,571	Ingman *et al.*, 2000

[a] = Zaire was the name given to the present Democratic Republic of the Congo between 27 October 1971 and 17 May 1997. The geographic region in Africa from which the individual sequenced originated, are indicated using colour. = Northern Africa, = Western Africa, = Central Africa, = Eastern Africa, = Southern Africa. [b] = This sequence, D38112, is listed on GenBank® as one of a list of other reference human mtDNAs and represents mtDNA from an African (Ugandan) individual. [c] = This sequence, NC001807, is an NCBI Reference sequence representing mtDNA from an African (Yoruba) individual and was derived from AF347015. [d] = This sequence, NC001643, is an NCBI Reference sequence representing mtDNA from a chimpanzee, *Pan troglodytes*, and was derived from D38113. Abbreviations: GI = GenInfo Identifier sequence identification number; CAR = Central African Republic. [e] = alternate names for the same ethnic group. The information in table A2 together with the footnote was compiled by M. Koekemoer, Centre for Genome Research, North-West University (Potchefstroom Campus) and used here with permission. The haplogroups assigned to the individuals were updated to reflect the classification system in Appendix A3 adapted from Behar *et al.* (2008) and that of Van Oven and Kayser (2009).

APPENDIX C

Polymorphisms characterising African macrohaplogroup L

0 is a list of polymorphisms characterising macrohaplogroup L haplogroups. The phylogenetic relationship between the haplogroups is as outlined in Figure 3.2.

Table A3 Polymorphisms defining macrohaplogroup L haplogroups

Haplogroup	Defining polymorphisms
L1'2'3'4'5'6	C182C!, C1048T, C4312T, T6185C, G11914A, G12007A
L1	G3666A, A7055G, T7389C, T13789C, T14178C, G14560A
L1b	G185T, A357G, G709A, T710C, A1438A!, T1738C, T2352C, T3308C, G3693A, A5036G, G5046A, T5655C, C6548T, T6827C, A6989G, C7867T, A8248G, T12519C, C13880A, A14203G, A14769G, T15115C, T16126C, G16129G!, C16264T, C16270T, A16293G
L1b1	A2768G
L1b1a	T5393C,
L1b1a1	C264T, T3396C
L1b1a2	T7954C
L1b1a3	A189G, G13980A
L1b1a4	G8790A, G13194A, C16114A
L1b1a5	T1406C, C14812T
L1b1b	G9755A, T14110C
L1c	C151T, C186A, A189C, G316A, 2395d, A5951G, T6071C, G8027A, A9072G, G10586A, A12810G, A13485G, T14000A, C14911T, C16294T, C16320T
L1c1'2'4'5	A297G, T10321C
L1c1	A3796T, A3843G, T11899C, A14148G, A16293G
L1c1a	T4454A, T8087C, T14088C
L1c1a1	A10398A!, T14034C, C16223C!
L1c1a1a	44 Ins C, C11257T, C16214T, C16234T, C16249T,
L1c1a1a1	C182C!, C467T, A2308G, A5984G, A11167G, A12930T, G16274A
L1c1a1a1a	T204C, G6182A, T8928C, T9311C, T15663C, A16293A!
L1c1a1a1b	A16051G, A16258G
L1c1a2	A93G, A95C, C151C!, T236C, A2755G, A2863G, C3513T, A3927G, A4506G, 7202G, 9647C, 12768G, 16274A
L1c1a2a	T4634C, A9336G, T12477C
L1c1a2b	T152T!, T7660C, C7693T, C9272T
L1c1bcd	C198T, T16086C
L1c1b	T1291C, T4688C, A4824G, T5553C, C8619T, T9861C, T10084C, T12681C, A14393G, C15025T, A16241G, C16291T, A16293A!
L1c1cd	G11914A
L1c1c	C198C!, 247del G, C534T, G3705A, G4991A, G5054A, A5984G, G6267A, G8387A, A8389G, T9233C, T9497C, C11335T, T11899T!, A12026G, C12545T, T12879C, A12961G, A13134G, C14911C!, C16169T, T16172C, A16183C, C16187C!, C16234T, C16261T, C16294C!
L1c1d	T5108C, G5460A, C7948T, G15301A, C15626T, A16038G

Haplogroup	Defining polymorphisms
L1c2'4	C12049T, A13149G
L1c2	C198T, G6150A, T6253C, A7076G, G7337A, A8784G, T8877C, A10792G, C10793T, A11654G, A16265c, C16286G, C16527T
L1c2a	T1420C, 2156 Ins A, C15016T, T15784C
L1c2a1	5899 Ins C, G16145A, G16213A
L1c2b	A11164G
L1c2b1	523 Ins CA, T10031C
L1c4	C151C!, C280T, 523 Ins CA, T5196C, 5899 Ins C, A6629G, A7673G, A7960G, G9266A, A9545G, A13741G, T13879C, A16037G, T16093C, T16140C, C16184T, C16301T, G16390A
L1c5	C198T, T455C, C514T, T593C, G930A, A2251G, A2315G, A4562G, T5074C, G5231A, G5460A, T5465C, C8829T, C9611T, T13659C, A13933G, T14470C, T14798C, C15700T, C16234T, C16256T, G16274A, C16527T
L1c3	T195T!, T6221A, G6917A, A7055A!, C11302T, A15226G, T15905C, C15978T
L1c3a	G6260A, G7498A, G7789A, G9966A, C12019T, G12501A, C16187C!, A16215G
L1c3b	G8251A, C13981T, C14794T, A16293G
L1c3b1	T629C, C2283T, C3210T, A3434G, T4755C, C8417T, A12400G, C12542T, C16017T, A16163G
L1c3b2	523 Ins CA, C2000T, C3096T, T6297C, A6353G, G7805A, A7844G, T9758C, A14128G, G14831A, A15244G, A15924G, T16086C, C16185T
L2'3'4'5'6	T152C, G2758A, T2885C,A 7146G, C8468T, T16519C
L5	459 Ins C, T3423C, A7972G, C12432T, A12950G, C16148T, A16166G
L5a	T152C, 455 Ins T, G709A, A851G, T1822C, T5111C, G5147A, A5656G, G6182A, T6297C, A7424G, G8155A, A8188G, C8582T, G9305A, G9329A, T11025C, C11881T, G12236A, A13722G, T14212C, C14239T, T14581C, G14905A, T14971C, G15217A, G15884A, C16355T, T16362C
L5a1	455 Ins TT, G930A, C4496T, C8754T
L5a1a	T5004C,T 9899C
L5a1b	C14688T, T14819C, A16183C
L5a1c	G513A, C4137T, C12303T, T13752A
L5a2	A444G, C527T, A5582G, T5892C, G8392A, T8588C, G8856A, T11318C, C12022T, A12654G, 16183del A, C16192T
L5b	247del G, 523 Ins CA, C535T, C2417G, T3027C, A3720G, A4976G, C5213T, A9809C, T10493C, T11701C, T12188C, A12546T, T12714C, A12810G, A13105A!, T13569C, T13830C, C16111T, A16254G, C16360T
L5b1	G8152A, A11065G, T11260C, T12215C
L5b2	C182C!, 455 Ins T, C2911T, T3801C, T4117C, T14581C, C16311T, A16355G, T16519C
L2'3'4'6	T195C, G247A, T825A, C8655T, G10688A, T10810C, C13506T, A13105G, G15301G!, G16129A, C16187T, T16189C
L2	T146C, C150T, T152C, T2416C, G8206A, A9221G, T10115C, G13590A, T16311T!, G16390A
L2abcd	T195C, T11944C
L2a	C150C!, C2789T, T7175C, C7274T,A 7771G, G11914A, A13803G, A14566G, C16294T
L2a1	G143A, C182C!, A12693G, T15784C, A16309G
L2a1c	G3010A, A 6663G
L2a1c1	C198T,G 930A, T3308C, T8604C, T16086C
L2a1c2	G513A, C10903T, A15924G, C16193T, G16213A, C16239T, T16519C
L2a1c3	G9932A
L2a1d	T5196C, T9530C, T11386C, G12007A, A12612G, A13395G, C13394T, T16209C, C16301T, A16309A!, C16354T
L2a2	T146T!, G709A, A6752G, G9932A, C15939T, T16189C, T16229C, C16291T, C16311T, T16519C

Haplogroup	Defining polymorphisms
L2a2a	T9083C, G9438A, G15803A
L2a2b	C11767T, A14118G, T16172C, T16183C, 16188 Ins C
L2a1abf	G143G!
L2a1a	G3918A, A5285G, A15244G, T15629C, T16519C
L2a1a1	523 Ins CA, T6152C, C15391T, T16368C
L2a1a2	T10454C, C15211T, A15421G
L2a1f	A5581G, T16189C, T16519C
L2a1b	G10143A, C15735T, A16182C, A16183C, T16189C, C16290T
L2a1e	C3495A,G 8790A, G12630A
L2a1h	A3505G, T4772C, C12976T, A16309A!
L2a1i	T15229C, A16309A!, T16362C
L2a1j	G8764A, A14464G
L2bc	C198T, G1442A,C 2332T, T7624A, G12236A, G15110A, G15217A
L2b	T204C, C1706T, A2358G, A4158G, T4370C, A4767G, C5027T, C5331A, T5814C, C6713T, C8080T, G8387A, A12948G, A14059G, C16114A, G16129A, G16213A
L2b1	C418T, G6026A, T10828C, C13924T, T16362C
L2b2	A183G, T6614C, A6806G, T8503C, G16274A
L2b3	15944 Ins T
L2c	A93G, C325T, T680C, G709A, T3200A, G13928C, G13958C, C15849T
L2c2	T1040C, C16264T
L2d	C182C!, C456T, C870T, T2159C, C2332T, C3254A, A3434G, G3693A, C6231T, G9554A, A9941G, C10955T, T11353C, C14845T, G15777C, T16093C, G16129A, T16189C, C16223C!, A16300G, C16354T, A16399G, T16519C
L2d1	T152T!,G 8856A, A10700G, C15263T, T15458C, A15703G, T16311C
L2e	A479G, G719A, G1211A, A3537G, A4562G, A5069T, T6014C, T8383C, A9377G, C9971T, T11935C, T12189C, G13708A, T14299C, T15697C, G15734A, T15889C, C16111A, G16145A, C16184T, C16239T, C16292T, C16355T, A16399G, C16400T, T16519C
L3'4'6	A4104G, G7521A
L6	T146C, T152C, G185C, G709A, C770T, T961C, A1461G, C4964T, T5267C, A6002G, A6284G, C9332T, A10978G, T11116C, C11743T, G12771A, A13710G, C14791T, A14959G, A15244G, T15289C, C15499T, G16048A, T16224C, T16519C
L6a	G207A, T265C
L6b	C14533A
L3'4	C182T, C3594T, C7256T, C13650T, C16278T
L4	C16362C
L4a	T195C, G3357A, G5460A, G10373A, T11253C, A11344G, T11485C, T12414C, T13174C, T14302C, C16260T, T16519C
L4a1	C325T, T6167C,C 7376T,G 7762A, G7775A, T8473C, A8631G, A11653G, T14000A, A16207T
L4a2	C198T, A12361G, C16264T
L4b	G3918A
L4b1	C150T, T152C, T199C, T204A, C496T, G513A, G709A, A1804G, G3010A, A3505G, C4017T, C4029A, T4216C, T4232C, T4977C, G5460A, A6710G, T7624C, 8289 Ins CCCCCTCTA, T8614C, T8937C, C8974G, C9248T, G9986A, C10813T, A12661T, A13497G, G14016A, C14791T, G14905A, G16153A, C16179T, A16183C, T16189C, C16239T, C16320T, C16365T, T16519C
L4b2	T146C, T8104C, A9855G, T12609C, A13470G, A16293T, C16355T, A16399G
L4b2a	T1413C
L4b2a1	T152C, T195C, A244G, T471C, A547G, G5471A,T 5580C, G5746A, G16274A
L4b2a2	T1694C, A4949G, G6260A, A10873C
L4b2b	G513A, G709A, T4500C, A5128G, G6260A, G7805A, T8227C, T10265C,

Haplogroup	Defining polymorphisms
	T10389C, G11914A, T12354C, T12438C, T12903C, G15217A, G15883A, G15930A
L3	G769A, G1018A, T16311C
L3a	T152C, C12816T, A16316G
L3f	T3396C, T4218C, T15514C, 15944del T, T16209C, T16519C
L3f1	C5601T, T9950C
L3f1a	G3693A, C4350T, C5194T, A14148G, G15106A
L3f1b	A189G, A200G, T1822C, C7819A, A8527G, C8932T, G11440A, A14769G, C16292T, T16311C
L3f1b1	C8410T, C10070T, G16129A, C16295T
L3f1b2	C7235A
L3f1b3	T711C, A6806G, A7158G
L3f1b4	A3505G
L3f2	745 Ins T, T16311C
L3f2b	374 Ins A, T2442C, C5342T, G11016A, T12338C, G15466A, C15658T
L3f3	A189G, T318C, C959T, G4643A, A5181G, C6602T, A8158G, G8251A, G9932A, T10604C, T11770C, T15940C, C16176T, C16234T
L3h	T7861C, G9575A
L3h1	G1719A, A4388G, C5300T, T9509C, A11590G, T16311C
L3h1a	T5492C
L3h1a1	C516T, G709A, T4117C, G5316A, T8047C, G8485A, A9080G, C12432T, G13708A, T13926C, T14110C
L3h1a2	C8781A
L3h1a2a	T146C, T4688C, T4742C, C8943T, T12175C, G12236A, T12519C, A14587G, C14862T, C15646T, A16399G
L3h1b	A189C, T195C, A10044G, G14410A, C16256A, A16284G
L3h1b1	T3777C
L3h2	C150T, T195C, T318C, G513A, A2707G, A3879G, A4122G, G5147A, G5460A, T5567C, A5813G, T5930C, G8020A, T9098C, A9254G, G9380A, T9965C, G11440A, A12469G, C13080T, C13755T, C16111T, C16184T, T16304C, T16519C
N	A8701G, T9540C, A10398G, T10873C, G15301A
M	T489C, C10400T, T14783C, G15043A
L3bcdj	A13105G
L3b	C3450T, G5773A, T6221C, C9449T, A10086G, C13914A, A15311G, A15824G, 15944del T, T16124C, C16278T, T16362C
L3b1	G10373A, T16519C
L3b1a	A11002G
L3b1b	T152C, A9079G, C15664T
L3b2	C3420T, A16183C, T16189C
L3cdj	T152C
L3c	T195C, 498 Ins C, T678C, C3582T, G4491A, T5393C, A7394G, T9337C, T9682C, A12373G, T14221C, T14371C, G14560A, A14587G, T16311C, T16362C
L3d	G5147A, A7424G, T8618C, T13886C, C14284T, T16124C
L3d1’2’3	T921C
L3d1	T6680C
L3d1a	G4048A, C7648T, G11887A
L3d1a1	G1503A, G16319A
L3d1b	G5046A
L3d1c	A3203G, T9111C, A11239G, C12870T, A13542G, A16166G
L3d1d	A7765G, A9151G, T16086C, C16256T, T16368C, T16519C
L3d2	T3394C, G11518A, T12873C, C14272T, T14584C, T15115C

Haplogroup	Defining polymorphisms
L3d3	T4688C
L3j	G5821A, G6182A, G6722A, C8676T, C9365T, C9731A, A12280G, A12534G, A14260G, G15314A, T15479C, T15514C
L3eikx	C150T, A10819G
L3e	T2352C, T14212C
L3e1	A189G, A200G, T6221C, C6587T, A14152G, T15670C, T15942C, C16327T
L3e1a	C16185T
L3e1a1	G185A, C8650T, T16311C
L3e1a2'3	T152C
L3e1a2	T195C, G207A, T5774A, A9254G, C11024T, G14569A, T16209C
L3e1a3	A10398A!
L3e1b	A14926G, 16325del T
L3e1c	A3675G, G5460A, 8289 Ins CCCCCTCTA, A8701A!, G14323A
L3e1d	T152C, C8703T, G9300A, T12738C, C16176T
L3e1e	T10370C, T14571A
L3e2	T195C, G14905A, C16320T, C16519T
L3e2a	T4823C, A13105G, G14869A
L3e2b	T16172C, A16183C, T16189C
L3e2b1	T408A, T2483C, A9377G, 16188 Ins C
L3e2b2	G11377A, A16182C
L3e3'4'5	A750A!, T16519C
L3e3'4	G5262A
L3e3	T195C, C2000T, T6524C, G9554A, T10667C, A10816G, A13101C, A16265T
L3e3a	A10286G, A12397G
L3e3b	G4655A, A12248G, C13197T, A13651G, G15812A
L3e4	G3915A, A5584G, C11257T, C13749T, A16051G
L3e5	T398C, G8392A, A16041G
L3i	T7645C
L3i1	A10679G, T11260C, C13800A
L3i1a	C13967T, T14060C, A15758G
L3i1b	T2158C, C13687T, G16153A
L3i2	T152C, A189G, T504C, A5441G, T8222C, G12630A, A14818G, T15388C, 15944del T, C16260T, T16311C
L3k	A235G, C494T, G3918A, T6620G, T9467C, G13135A, C13992T, G15314A
L3x	G3483A, 5899 Ins C, A6401G, T8311C, A8817G, G13708A, C16169T
L3x1	T204C, G15172A, C16278T
L3x2	249del A, C494T, A9941G, T16195C
L3x2a	C3435T, C16193T, C16223C!
L3x2b	A188G, T650C, C4230T, C8476T, A9254G, G14569A, A15758G, G15928A, T16189C, G16213A, T16243C
L0	T146C, A263A!, C3516A, T5442C, C9042T, A9347G, G10589A, C10664T, T10915C, A12720G, A13276G, A16230G
L0abfk	A189G, T4586C, C9818T, T16172C
L0abf	A73A!, G185A, T195T!, A263G, A2245G, C5603T, A11641G, C15136T, G15431A
L0ab	A93G, T236C, C8428T,A 8566G, G9755A, C16148T
L0a	T146T!, G5231A, G5460A, G11176A, T14308C, C16188g, C16278C!, C16320T, C16519T!
L0a1	T5096C, C16168T
L0a1a	A200G, T3866C
L0a1bcd	A16293G
L0a1b	A95C, C5911T, A14007G, C16278T

Haplogroup	Defining polymorphisms
L0a1b1	152T!, A8191G
L0a1b2	G5563A, T14106C
L0a1c	A8188G, G14569A, G16129G!, C16287T, T16519C
L0a1d	C553T, C12557T, C16320C!
L0a2	C64T, G185G!, G5147A, A5711G, G6257A, 8281-9d, A8460G, A11172G, G16129G!, T16519C
L0a2a	C11143T, A14755G
L0a2a1	T14182C
L0a2a2	T204C, G207A, A9545G, G9554A
L0a2b	T152T!, A189A!, 523 Ins CA, T3372C, G5237A, C11269T, A12172G, T13281C, C16188A, C16242T, G16390A
L0a2c'd	A95C
L0a2c	C194T, G3693A, C3738T, A4012G, A8389G, T11009C, G11887A, C13984T, G14560A, C16188A, C16214T, C16234T, C16320C!
L0a2d	T146C, A234G, 573 Ins C, A5581G, A5954G, A9261G, C9860T, C15250T, G15317A, A16037G, G16129A
L0a3	A95C, G207A, T318C, G796A, T1193C, T6050C, G6261A, C6689a, A8659G, G10993A, A12082G, A13276A!, T14581C, G15323A, C15418T, T15941C
L0a4	T152T!, T2887C, G3010A, C5060T, G7830A, G8545A, T8870C, A11380G, T11809C, A12780G, A14001G, C14142T, C16168T, C16192T, C16234T, C16259A, G16319A
L0b	C64T, A95C, A1536G, G1664A, T4679C, T6719C, A7874G, T13965C, G15106A, G15148A, T15262C, T15622C, A16051G, A16164G, A16182C, A16183C, C16187C!, T16209C, C16223C!, C16287T, C16291T
L0f	G207A, C4964T, T7148C, T9581C, C9620T, G13145A, A13470G, C14109T, C14620T, T15852C, C16169T, C16327T, C16354T, T16368C
L0f1	T146T!, C151T, G185G!, T482C, T3644C, C3699G, T5108C, T5492C, A5984G, A7364G, A7543G, T8911C, G10310A, C10692T, A10694G, A11084G, A11314G, C12346T, A14091G, T14178C, G14364A, A15236G, C16187C!, C16174T, A16182C, A16183C
L0f2	C4194T, T11287C, C13680T, G13928C
L0f2a	523 Ins CA, G1719A, A4562G, T10790C, G13145G!, C16354C!
L0f2b	T152T!, A183G, G184A, A783G, T3027C, G3531A, G3592A, T6152C, T6176C, G7419A, G8020A, C8748T, T9725C, A11974G, A12855G, G13317A, A14683G, G15884A, G15927A, G16129G!, C16218T, C16291T
L0k	T850C, T1243C, G4541A, T4907C, A5811G, A7257G, T8911C, G8994A, A9136G, A10499G, A10876G, C10920T, C11296T, T11299C, A11653G, G13590A, T13819C, G13928C, T14020C, T14182C, T14371C, T14374C, A16166C, C16214T, C16291G
L0k1	C198T, C2836A, C6938T, C10939T, G12070A, A12720A!, T13020C, G16129G!, T16209C
L0k2	T199C, T204C, A215G, T294C, 523 Ins CA, G3882A, A4767G, C6053T, C8922T, C9884T, C11061T, A13535G, G15221A, C16223C!, C16287T, C16291A
L0d	A1438A!, T4232C, T6815C, C8113A, G8152A, G8251A, T12121C, G15466A, G15930A, T15941C, T16243C,
L0d1'2	498del C, A3756G, G9755A, C16278C!
L0d1	G719A, A2706A!, G3438A, A6266G, G13759A
L0d1a'b	523 Ins CA
L0d1a	T199C, T1243C, T9950C, A12141G, C12798T, C16223C!, C16234T, C16266G
L0d1b	T3618C, T7283C, C14315T, C14659T, C16239T, C16294T
L0d1c	C456T, C4197T, A9150G, T11437C, T12235C, C13129T, A15951G, G16129G!, C16234T
L0d1c1	C198T, A7828G, G9438A, C15550T, C16167T, C16242T
L0d2	A3981G, C4025T, A4044G, A7154G, T11854C, A15766G

Haplogroup	Defining polymorphisms
L0d2a'b	A16212G
L0d2a	C198T, C597T, A4225G, A5153G, G8392A, A12172G, A12234G, A12810G, T14221C, G16390A
L0d2b	T794C, T1386C, G3918A, G5054A, G9932A, T10084C, A11101G, T13768C, T14326C, A14605G, C15194T, T15511C, C16069T, C16169T
L0d2c	T294a, A4038G, T4204C, T4937C, G6249A, C6644T, C8284T, A8420G, T9230C, T9305A, A11974G, A13827G, A14007G, G15346A
L0d3	C150T, G316A, 523 Ins CA, T721C, T1243C, A2755G, G5460A, G5773A, C6377T, A8459G, T8598C, C9027T, C9488T, C11061T, G13359A, A15236G, T15312C, C16290T, A16300G

A deletion is indicated by del after the nucleotide position followed by the type of nucleotide deleted while insertions are designated by Ins followed by the number and kind of nucleotide inserted. A reversion of the polymorphism to the ancestral state with respect to the rCRS is illustrated by an exclamation mark (!). The phylogeny excluded indels at nucleotide positions 309, 315, 522 and 16189 because they are hyper variable. Adapted from Behar *et al.* (2008).

To where were the slave ancestors of the Baganda (back cover photo above, the entrance to the Kasubi tombs, the burial site for the kings of the Buganda Kingdom), the Lugbara and the Acholi (front cover photo, Acholi engaged in the Larakarak cultural dance) taken? Why are the history books silent on the final destination of the slaves from Uganda? Could there have been no slavery in Uganda? Where did the Baganda, Acholi and Lugbara come from? How are they related to each other? Where is the oldest human population in the world located? The indelible records of mitochondrial DNA are filling the gaps on these and many other unanswered questions on the Ugandan and World history. The first complete mitochondrial DNA sequences of a Ugandan population have been determined. Can these sequences be of value in pharmacotherapy, nutrigenomics and forensics?