Mitochondrial DNA consensus sequence for the Tswana population of South Africa

BY

SCHEÁN BABST, B.Sc. (Hons)

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

PROMOTOR: Professor Antonel Olckers
Centre for Genome Research, North-West University (Potchefstroom Campus)

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

June 2012

Mitokondriale DNS-konsensusvolgorde vir die Tswanabevolking van Suid-Afrika

DEUR

SCHEÁN BABST, B.Sc. (Hons)

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

PROMOTOR: Professor Antonel Olckers
Sentrum vir Genomiese Navorsing, Noordwes-Universiteit (Potchefstroom-kampus)

MEDEPROMOTOR: Doktor Wayne Towers
Sentrum van Uitnemendheid vir Voeding, Noordwes-Universiteit (Potchefstroom-kampus)

Junie 2012

This thesis is dedicated to my husband,
Neels Babst,
and to my sons,
Karl, Marco and Alec Babst

The question of questions for mankind — the problem which underlies all others, and is more deeply interesting than any other — is the ascertainment of the place which Man occupies in nature and of his relations to the universe of things.

— Thomas Henry Huxley, 1894

ABSTRACT

Evolutionary studies are critical in eliciting the fundamental phylogeny within and among populations of living organisms. Genetic diversity is displayed in human mitochondrial DNA (mtDNA) as haplogroups that consist of shared mutations, which are carried to the following generation through the maternal lineage. The current haplogroup hierarchies commonly used to describe and compare the genetic diversity of global human populations are based on the available mtDNA sequence variation datasets of numerous continent-specific populations. The description of mtDNA variation in human populations is furthermore of importance, as it allows the identification of population-specific genetic variation that has an effect on gene function, as well as on adaptation and susceptibility to disease. Owing to the limited amount of available mtDNA variation data from the numerous African populations currently residing in Africa, a lack of genetic diversity data exists for the determination of a sufficient baseline standard sequence representing the genetic variation present in African populations and thus also for a representative African haplogroup hierarchy.

In this study, the mtDNA variation of 50 Tswana-speaking individuals from South Africa was determined and a novel Tswana consensus sequence was constructed to contribute to the urgent need for information of the mtDNA variation present in African populations. The consensus mtDNA sequence variation data obtained through this analysis should be regarded as a baseline for the observed sequence variance and genetic diversity of the maternal ancestral genetic pool of a Bantu-speaking population of South Africa.

This study therefore contributes novel information regarding the mitochondrial genetic diversity of a South African Tswana-speaking population to the current body of literature. The results of this study provide strong evidence to support the ancient nature of African haplogroups and also provide evidence in support of the presence of Khoi-San maternal ancestry in the origins of the current Bantu-speaking populations of southern Africa. In addition, the observed sequence variation contributes to the current haplogroup hierarchy of African lineages and provides information in support of the previously reported distinct phylogenetic relationship between individuals of African and non-African origin, thereby explaining the high level of genetic diversity among and between African populations.

OPSOMMING

Evolusionêre studies is noodsaaklik in die verkryging van die fundamentele filogenie van lewende organismes. Genetiese diversiteit van die menslike mitokondriale DNS (mtDNS) word uitgedruk as haplogroepe wat bestaan uit gedeelde mutasies wat deur die moeder na die volgende generasies oorgedra word . Die haplogroephiërargieë wat tans algemeen gebruik word om die genetiese diversiteit van globale menslike bevolkings uit te beeld en met mekaar te vergelyk, is gebaseer op die beskikbare mtDNS-variasie wat opgeteken is vir verskillende bevolkings. MtDNS-variasie in menslike bevolkings is ook van kritiese belang in studies rakende siekte en gesondheidsorg omdat genetiese mutasie geenfunksionering affekteer en dus menslike aanpasbaarheid by die omgewing en vatbaarheid vir siektes kan bepaal. Dit is tans nie moontlik om ’n basislynstandaard van genetiese variasie of volledige filogenetiese hiërargie vir Afrikabevolkings in geheel saam te stel nie, as gevolg van die afwesigheid van voldoende inligting oor mtDNS-variasie om die huidige Afrikabevolkings te verteenwoordig.

In hierdie studie is die mtDNS-variasie van 50 Tswana-sprekende individue uit Suid-Afrika bepaal en ’n unieke Tswana-konsensusvolgorde daaruit saamgestel as bydrae tot die bestaande inligting oor mtDNS-variasie in Afrika. Die konsensus- mtDNS-volgorde variasie-data wat verkry is deur middel van hierdie ontledings, kan beskou word as ’n basislyn van die volgorde-variasie en genetiese diversiteit van die maternale voorouer genetiese poel van ’n Bantoe-sprekende bevolking van Suid-Afrika.

Hierdie studie dra dus unieke en nuwe inligting oor die mitokondriale genetiese diversiteit van ’n Suid-Afrikaanse Tswana-sprekende bevolking by tot die huidige kennis soos opgeteken in die literatuur. Die resultate van hierdie studie voorsien sterk bewyse om die antieke aard van Afrika-haplogroepe te ondersteun en verskaf ook bewys ter stawing van die teenwoordigheid van Khoi-San maternale afkoms in die huidige Bantoe-sprekende bevolkings van Suider-Afrika. Daarbenewens dra die waargenome volgorde-variasie van hierdie studie by tot die haplogroephiërargie van Afrika-afstammelinge en verskaf inligting ter ondersteuning van die filogenetiese verhouding tussen individue van Afrika en diegene van nie-Afrika-oorsprong, en dui as sulks ook die hoë vlak van genetiese diversiteit onder en tussen Afrikabevolkings aan.

TABLE OF CONTENTS

CHAPTER ONE

CHAPTER TWO

CHAPTER THREE

CHAPTER FOUR

CHAPTER FIVE

CHAPTER SIX

CHAPTER SEVEN

CHAPTER EIGHT

APPENDIX A

APPENDIX B

APPENDIX C

APPENDIX D

APPENDIX E

APPENDIX F

APPENDIX G

APPENDIX H

LIST OF ABBREVIATIONS AND SYMBOLS

Symbols and abbreviations are listed in alphabetical order:

LIST OF SYMBOLS

α	alpha used to indicate the gamma shape parameter
β	beta used to indicate the gamma shape parameter
$Cov(\hat{d}_{ij},\hat{d}_{kl})$	covariance based on phylogenetic relationship among sequences
σ_a^2	covariance component due to differences among populations
σ_b^2	covariance component due to differences among haplotypes in different populations within a group
σ_T^2	total molecular variance
κ	average number of nucleotide differences
k_{ij}	the number of nucleotide differences between sequence i and j
$	dollar sign used to indicate incompletely classified sequences
$\hat{d}_{ij}$	the number of nucleotide substitutions per site between sequence i and j
ε	epsilon
=	equal to
η	eta used to indicate the total number of mutations
γ	gamma
Γ	gamma shaped parameter
–	gap in DNA sequence
>	greater than
<	less than
λ	lamda used to indicate variation of the substitution rate
µ	micro (10^{-6})
µL	microlitre
x	mismatch distribution
μ	mutation rate
n	number of DNA sequence samples from a population
nd	number of nucleotide substitutions
$\binom{n}{2}$	the total number of sequence comparisons
●	*Pan troglodytes* outgroup position indicated with a red circle in phylogenetic trees
%	percent
π	pi, nucleotide diversity or the average number of pairwise differences between DNA sequences
π_n	the mean number of pairwise differences for n sequences
◆	rCRS position indicated with a green diamond in phylogenetic trees
®	registered trademark
ρ	rho, average number of nucleotide differences between a set of DNA sequences and a specified DNA sequence
$\hat{V}_s$	sampling variance
σ^2	sigma squared indicating variance
V_{st}	stochastic variance
√	square root
θ_π	the mean number of nucleotide differences between two sequences
θ	theta, expected pairwise nucleotide site differences also referred to as the population parameter
θ_0	initial population size
θ_1	population size after expansion

$Q(i)$	the number of differences between a pair of genes where *i* is the number of different genes
$\bar{t}_0$	the mean coalescence time of two genes drawn from the same population
$\bar{t}_1$	the mean coalescence times of two genes drawn from two different populations
U_i	the number of singletons in sequence *i*.
$\hat{V}$	total variance
™	trademark
κ	transversion rate
▲ or ★	Tswana mtDNA sequence positions of this investigation are indicated with a blue triangle or blue star
$V(\hat{\pi})$	variance of nucleotide diversity

LIST OF ABBREVIATIONS

12S	12 Svedberg units
12S rRNA	12S ribosomal RNA
16S	16 Svedberg units
16S rRNA	16S ribosomal RNA
A or a	adenine nucleobase in DNA sequence (in DNA context)
A	tRNA / amino acid alanine (in amino acid context)
A_{260}	absorbance of samples at 260 nm
A_{280}	absorbance of samples at 280 nm
A_{260}/A_{280}	absorbance ratio measured at 260 nm and 280 nm
ABO gene	gene that codes for the histo-blood group ABO system transferase enzyme with glycosyltransferase activity which in humans determines the ABO blood group of an individual
acetyl-CoA	acetyl-coenzyme A
ac-CoA	acetyl-coenzyme
acc-stem	tRNA acceptor stem
ac-stem	tRNA anticodon stem
AD	Alzheimer disease (in disease context)
AD	*Anno Domini* (in date context)
ADP	adenosine diphosphate
Ala	alanine amino acid
Alu element	DNA fragments that are approximately 300 bp in length with a single recognition site for the restriction enzyme *Alu*I located near the middle of the Alu element
AMH	anatomically modern humans
AMOVA	analysis of molecular variance
anticd-loop	tRNA anticodon loop
AP	among populations
Arg	arginine amino acid
Asn	asparagine amino acid
Asp	aspartic acid amino acid
ATP	adenosine-5-triphosphate
ATP6	ATP synthase F0 subunit 6
ATP8	ATP synthase F0 subunit 8
ATP9	ATP synthase F0 subunit 9
ATP6	ATP synthase F0 subunit 6 gene
ATP8	ATP synthase F0 subunit 8 gene
ATT	membrane attachment site
bp	base pair
C or c	cytosine nucleobase (in DNA context)
C	tRNA / amino acid cysteine (in amino acid context)
°C	degrees Celsius
C-A	cytosine paired to alanine in double-stranded DNA
Ca^{2+}	calcium ion
(CA)n	cytosine and adenine nucleotide repeat stretch
CAR	Central African Republic
C-G	cytosine paired to guanine in double stranded DNA
CGR	Centre for Genome Research
CI	confidence intervals
CM	cardiomyopathy

CNI	Close-neighbour interchange
CO_2	carbon dioxide ion
COI	cytochrome c oxidase subunit I
COII	cytochrome c oxidase subunit II
COIII	cytochrome c oxidase subunit III
COI	cytochrome c oxidase subunit I gene
COII	cytochrome c oxidase subunit II gene
COIII	cytochrome c oxidase subunit III gene
CoQ	coenzyme Q or ubiquinone
$CoQH_2$	reduced coenzyme Q
Cov	covariance
COX	cytochrome c oxidase
CPEO	chronic progressive external opthalmoplegia
CpG	cytosine and guanine separated by only one phosphate; used to distinguish the linear sequence from the CG base-pairing of cytosine and guanine
CR	control region
CRS	Cambridge Reference Sequence
C-T	cytosine paired to thymine in double-stranded DNA
Cu_B	copper B centre of the Q cycle in the mitochondria
Cys	cysteine amino acid
Cytb	cytochrome b
Cytb	cytochrome b gene
Cytc	cytochrome c
Cytc	cytochrome c gene
d	maximum number of nucleotide differences
D	Tajima's D test statistic
D	tRNA / amino acid aspartic acid
*D**	Fu and Li's D* test statistic
dATP	2'-deoxyadenosine-5'-triphosphate
dCTP	2'-deoxycytidine-5'-triphosphate
ddH_2O	double distilled water
ddNTP	2',3' dideoxynucleotide triphosphates
DEAF	maternally inherited DEAFness or aminoglycoside-induced DEAFness
del	deletion
DGGE	denaturing gradient-gel electrophoresis
dGTP	2'-deoxyguanosine-5'-triphosphate
dHPLC	denaturing high pressure liquid chromatography
D loop	d-loop, non-coding region of the mitochondrial DNA; between nucleotide positions 16024-576 also referred to as the control region
DNA	deoxyribonucleic acid
Dnapars	DNA parsimony programme
DNS	deoksiribonukleïensuur
dNTP	2'-deoxynucleotide triphosphates
dsDNA	Double-stranded DNA
dTTP	2'-deoxythymidine-5'-triphosphate
e^-	electron
E	tRNA / amino acid glutamic acid
EDTA	ethylenediamine tetra-acetic acid
et al.	*et alia*: and other people
EtBr	2,7-diamino-10-ethyl-9-phenyl-phenanthridinium bromide (ethidium bromide)
ETC	electron transport chain
EtOH	ethanol
f	Farris's statistic
F	forward primer
F	tRNA / amino acid phenylalanine
F	fixation index
*F**	Fu and Li's F test statistic
f_0	probability of identity by descent of two different genes drawn from the same population
f_1	probability of identity by descent of two genes drawn from two different populations
F_1 ATPase	F_1 subunit of adenosine tri-phosphate synthase
F81/FEL	Felsenstein's model
$FADH_2$	Flavin Adenine Dinucleotide
FL	fragment length

F_S	Fu's statistic
F_{ST}	Wright’s F statistic or fixation index
G	guanine nucleobase in DNA sequence(in DNA context)
G	tRNA / amino acid glycine (in amino acid context)
g	gap opening penalty
G + C	G + C refers to the cytosine and guanine composition of DNA
G-A	guanine paired to alanine in double- stranded DNA
gDNA	genomic DNA
GenBank®[1]	A public DNA sequence database maintained by the National Center for Biotechnology Information (NCBI).
GI	GenInfo Identifier sequence identification number used for identification in GenBank®
GLP	Good laboratory practice
Glu	glutamate or glutamic acid amino acid
Gly	glycine amino acid
G-T	guanine paired to thymine in double-stranded DNA
GP	gap penalty
H	haplogroup associated substitutions(in mutation context)
H (in DNA context)	heavy strand of the mitochondrial DNA
H (in aminoacid context)	tRNA / amino acid histidine
H^+	hydrogen ion
h	gap-extension penalty
H_2O	water
H_2O_2	hydrogen peroxide
HKA	Hudson-Kreitman-Aguadé
HKY85/HKY	Hasegawa, Kishino and Yano model of evolution
*Hpa*I	restriction endonuclease isolated from a recombinant from *Haemophilus parainfluenzae*
HR-RFLP	high resolution restriction fragment length polymorphism
HV1	hypervariable segment 1
HV2	hypervariable segment 2
HVR	hypervariable regions of the mtDNA
HVS	hypervariable segment
HVS1	hypervariable segment 1
HVS2	hypervariable segment 2
HVS3	hypervariable segment 3
HVS-I	hypervariable segment 1
HVS-II	hypervariable segment 2
I	tRNA / amino acid isoleucine
Ile	isoleucine amino acid
i.e.	*id est*: that is to say
indels	sequence characters that have been deleted or inserted
ins	insertion
JC	Jukes-Cantor
K	lysine amino acid (in amino acid context)
K	tRNA lysine (in tRNA context)
K2P	Kimura 2 parameter model
kb	kilobase
KS	prefix used to indicate mtDNA samples from Khoi-San-speaking individuals from southern Africa
KSS	Kearns-Sayre syndrome
kya	thousand years ago
L	light strand of the mitochondrial DNA
l	length of the gap
l_i	observed substitutions
L(CUN)	tRNA / amino acid leucine 2
L(UUA/G)	tRNA / amino acid leucine 1
LD	linkage disequilibrium
Leu	leucine amino acid
LGAM	last glacial aridity maximum

---

[1] GenBank® is a registered trademark of the U.S. Department of Health and Human Services, Independence Avenue, S.W., Washington, DC, USA.

LGM	last glacial maximum
LHON	Leber's hereditary optic neuropathy
LINEs	long interspersed nuclear elements
LS	Leigh syndrome
LVNC	left ventricular noncompaction syndrome
Lys	lysine amino acid
M	cytosine or adenine (in DNA context)
M	tRNA / amino acid methionine (in amino acid context)
m	milli (10^{-3})
mA	milli Amperes
MBS 0.5S	Multiblock System 0.5 Satellite
MICM	maternally inherited cardiomyopathy
ME	minimum evolution phylogenetic tree-building method
MEGA	Molecular Evolutionary Genetics Analysis
MELAS	mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes
MERRF	myoclonic epilepsy with ragged-red fibres
mg	milligram
Mg^{2+}	magnesium ion
$MgCl_2$	magnesium chloride
min	minute
MITOMAP	Human Mitochondrial Genome Database
MK	McDonald-Kreitman
mL	millilitre
ML	maximum likelihood
mm	millimetre
mM	millimolar
MnSOD	Mn superoxide dismutase
MP	maximum parsimony phylogenetic tree-building method
MR	multiregion model
MRCA	most recent common ancestor
mRNA	messenger ribonucleic acid
MS	multiple sclerosis
MSD	mean squared deviations
mtDNA	mitochondrial DNA
mtDNS	mitokondriale DNS
mtPTP	permeability transition pore in the mitochondrial membrane
mtTFA	mitochondrial transcription factor A
mya	a million years ago
N_e	effective population size
N	tRNA / amino acid asparagine
n	nano (10^{-9})
NAD^+	oxidised nicotinamide adenine dinucleotide
NADH	reduced nicotinamide adenine dinucleotide
NCBI	National Centre for Biotechnology Information
ND1	NADH dehydrogenase subunit 1
ND2	NADH dehydrogenase subunit 2
ND3	NADH dehydrogenase subunit 3
ND4	NADH dehydrogenase subunit 4
ND4L	NADH dehydrogenase subunit 4L
ND5	NADH dehydrogenase subunit 5
ND6	NADH dehydrogenase subunit 6
ND1	NADH dehydrogenase subunit 1 gene
ND2	NADH dehydrogenase subunit 2 gene
ND3	NADH dehydrogenase subunit 3 gene
ND4	NADH dehydrogenase subunit 4 gene
ND4L	NADH dehydrogenase subunit 4L gene
ND5	NADH dehydrogenase subunit 5 gene
ND6	NADH dehydrogenase subunit 6 gene
nDNA	nuclear DNA
ng	nanogram
NI	index of neutrality
NJ	Neighbour-joining tree-building method
N-J	Neighbour-joining
nm	nanometre

np	nucleotide position
n_s	number of singleton mutations
NS	nonsynonymous substitutions
NS_H	haplogroup associated nonsynonymous substitutions
NS_P	private nonsynonymous substitutions
numts	nuclear inserts of mitochondrial DNA
NWU	North-West University
O.D.	optical density
O_2	oxygen
O^{2-}	superoxide anions
O_H	mitochondrial H-strand origin of replication
O_L	mitochondrial L-strand origin of replication
OOA	Out of Africa model
OTU	operational taxonomic unit
OXPHOS	oxidative phosphorylation
P	statistical significance (in enzyme context)
P	private substitution (in mutation context)
P	tRNA / amino acid proline
P	total number of populations
p	pico (10^{-12})
p_i	the demographically unbiased estimator of the average genetic distance to the root of a node in the ith haplogroup, sub-haplogroup or lineage
PAUP	Phylogenetic Analysis Using Parsimony software
PCR	polymerase chain reaction
PD	Parkinson’s disease
PDH	pyruvate dehydrogenase
Phe	phenylalanine amino acid
P_H	replication promotor of heavy strand
PHYLIP	Phylogeny Inference Package software
P_i	orthophosphate byproduct produced by the hydrolosis of ATP and ADP
P_L	replication promoter of light strand
pmol	picomole
POP™	Performance Optimised Polymer
PR2	Parity Rule type 2
PRIMER	Profiles of Resistance to Insulin in Multiple Ethnicities and Regions
Pro	proline amino acid
Q	ubiquinone
Q^-	ubisemiquinone
Q	tRNA / amino acid glutamine
R	purine (adenine or guanine)
R	reverse primer (in DNA context)
R	tRNA / amino acid arginine(in amino acid context)
R(t)	dispersion index
R_2	Ramos-Onsins and Rozas test
RAO	Recent African Origin hypothesis
rCRS	revised Cambridge Reference Sequence
REV	general reversible model
RFLP	restriction fragment length polymorphism
rfu	relative fluorescent unit
rg	raggedness statistic
ROS	reactive oxygen species
rpm	rotation per minute
rRNA	ribosomal ribonucleic acid
RRSS	Reduced-representation shotgun sequencing
S	cytosine or guanine (in DNA context)
S	segregating site
S	synonymous substitutions
S(AGY)	tRNA / amino acid serine 2
S(UCN)	tRNA / amino acid serine 1
SAM	S-adenosyl-methionine
SD	standard deviation
SDH	succinate dehydrogenase
SDS	sodium dodecyl sulphate
Ser	serine amino acid

S_H	haplogroup associated synonymous substitutions
SINEs	short interspersed nuclear elements
SNP	single nucleotide polymorphism
S_P	private synonymous substitutions
SSCP	single-strand conformational polymorphism
SSD	sum of squared deviations
SSD (*AP*)	sum of squared deviations among populations
SSD (*WP*)	sum of squared deviations within populations
ssDNA	single-stranded DNA
STR	short tandem repeat
STRP	short tandem repeat polymorphism
T	tRNA threonine (in tRNA context)
T or t	thymine nucleobase (in DNA sequence context)
T	tRNA / amino acid threonine
TS	prefix used to indicate mtDNA samples from the Tswana-speaking individuals of this investigation
T-loop	telomere-loop
Ts	transition
T_a	annealing temperature
T-A	thymine paired to alanine in double-stranded DNA
Taq	*Taq* polymerase: DNA deoxynucleotidyltransferase from *Thermus aquaticus*
TBE	Tris-borate-EDTA
TCA	tricarboxylic acid cycle
Thr	threonine amino acid
T-G	thymine paired to guanine in double-stranded DNA
T_m	melting temperature
TMRCA	time to most recent common ancestor
TrisCl	an organic compound known as tris(hydroxymethyl)aminomethane, with the formula $(HOCH_2)_3CNH_2$.
tRNA	transfer ribonucleic acid
Trp	tryptophan amino acid
Tv	transversion
Tyr	tyrosine amino acid
UG	prefix used to indicate mtDNA samples of Ugandian origin
UK	United Kingdom
UPGMA	unweighted pair-group phylogenetic tree-buidling method
USA	United States of America
US	University of Stellenbosch
UV	ultraviolet
UVIvue	ultraviolet transilluminator
V	volts
V	tRNA / amino acid valine (in amino acid context)
Val	valine amino acid
W	tRNA / amino acid tryptophan
WP	within populations
WPGMA	weighted-pair group method with arithmetic means
Y	pyrimidine (cytosine, thymine, or urasil)
Y	tRNA / amino acid tyrosine
ybp	years before present

LIST OF EQUATIONS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

This study would not have been possible without the kind participation of the **Tswana-speaking people** of the Ikageng and Sonderwater urban areas and the rural areas of Ganyesa and Tklagameng in the North-West province of South Africa and their generous donations of DNA under the Profiles of Resistance to Insulin in Multiple Ethnicities and Regions (PRIMER) study conducted by the **North-West University (Potchefstroom Campus)**. My heartfelt thanks and acknowledgement go to these volunteers for their contribution to the current body of knowledge with regard to African maternal lineages of current Bantu-speaking populations and to the **North-West University (Potchefstroom Campus)** and all persons who made this project possible.

I would like to express my sincere acknowledgement to my supervisor, **Prof. Antonel Olckers**, for her assistance with the preparation and completion of this thesis. I have sincere gratitude for her help and support, not to mention the advice given to me based on her unsurpassed knowledge of the field of human phylogenetics. I was privileged to have had the opportunity to study under someone of her stature and will forever be thankful for the opportunities with which she provided me.

I would also like to acknowledge the contributions that were made by my co-supervisor, **Dr Wayne Towers**. My deepest thanks and gratitude go to him for his patience and assistance with the practical aspects of my research work and his unfailing willingness to provide me with excellent guidance, not only on the content of my thesis, but also on the structuring of my thoughts and the quality of my writing.

I would further like to thank the **North-West University (Potchefstroom Campus)**, the **Centre for Genome Research (CGR)**, **DNAbiotec (Pty) Ltd.** and the **Central Analytical Facility of the University of Stellenbosch** for providing excellent laboratory facilities and services and financial assistance towards the research costs of this project. I am indebted to a fellow student at the CGR, **Dr Desiré Dalton**, and my co-supervisor, **Dr Wayne Towers**, for the isolation and preparation of the mtDNA from the samples of the Tswana-speaking cohort. I am also greatly indebted to my fellow students,

Dr Michelle Koekemoer and **Dr Dan Isabirye**, for providing me with the mtDNA genome datasets of a Khoi-San-speaking cohort and Bantu-speaking cohort from Uganda for inclusion in the mtDNA datasets used in this study. A special word of thanks goes to my then colleague at DNAbiotec (Pty) Ltd., **Dr Annelize van der Merwe**, for her superb management of the laboratory, reagents and instruments and for her contributions to troubleshooting my laboratory results when I struggled to find answers. In my work relating to the construction of the phylogenetic trees, I am particularly indebted to **Dr Michelle Koekemoer** and **Dr Wayne Towers** for providing me with protocols of the tree-building methods I employed.

My colleagues at **DNAbiotec (Pty) Ltd**, **Dr Annelize van der Merwe**, **Ms Anri Raath, Mr Kenneth Nkadimeng** and **Mr Leonard Mdluli,** made an immense contribution during my study period as sources of friendship and professional support. A special word of thanks goes to **Dr Annelize van der Merwe** for her constant support and words of encouragement in the times that I felt overwhelmed by everything.

This thesis would not have been possible without the love and support of my friends and family. In particular, I want to acknowledge the unwavering support and encouragement from my lifelong friend, **Mari Campbell**, who has always shared my dreams with me since our childhood together. I am grateful that she can now share the realisation of this one. Many thanks to my other dear friends who have stood beside me and often pushed me along on this journey. You will forever be remembered for your love and support.

My deepest gratitude and appreciation goes to my family who never stopped believing in me and stood firmly behind me during this time. A special acknowledgment goes to my mother, **Nyn Benadie**, for her unconditional love and support and many hours of assistance, without which I surely would not have been able to work as many hours as I did. Special thanks also to my brothers, **Arno** and **Rohan Benadie**, for the sheer belief in their sister's abilities. I also gratefully acknowledge the continuous encouragement and support I received from the Babst family, especially from my parents-in-law, **Hans** and **Carine Babst**.

Most of all, my deepest and everlasting gratitude goes to my husband, **Neels Babst**. He was the sole reason behind the completion of this thesis and my most loyal supporter throughout these long and hard years. I will forever be thankful for his love and encouragement, without which I would never have succeeded.

And lastly, I would like to acknowledge, with the deepest love, my dearest, brave sons, **Karl, Marco** and **Alec**, who have and always will inspire me to great and sometimes unthinkable heights. Special thanks goes to **Karl and Marco**, for the grown-up way in which they coped with their pre-occupied mother, especially during the last phase of writing, and for their honest encouragement and love. And special thanks goes to my challenged child, **Alec**, who has given me a brave heart and the clarity to see what is important and to do what is right. Thank you.