

# **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 haplogroep hiërargieë 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 Afrikabevolking in geheel saam te stel nie, as gevolg van die afwesigheid van voldoende inligting oor mtDNS-variasie om die huidige Afrikabevolking 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 staving van die teenwoordigheid van Khoi-San maternale afkoms in die huidige Bantoe-sprekende bevolking van Suider-Afrika. Daarbenewens dra die waargenome volgorde-variasie van hierdie studie by tot die haplogroep hiërargie van Afrika-afstammeling 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 Afrikabevolking aan.



# TABLE OF CONTENTS

---

LIST OF ABBREVIATIONS AND SYMBOLS.....	i
LIST OF EQUATIONS.....	ix
LIST OF FIGURES.....	xi
LIST OF TABLES.....	xv
ACKNOWLEDGEMENTS.....	xix
<b>CHAPTER ONE</b>	
<b>INTRODUCTION.....</b>	<b>1</b>
<b>1.1 OBJECTIVES OF THIS RESEARCH STUDY.....</b>	<b>5</b>
<b>1.2 SPECIFIC AIMS OF THE PROJECT.....</b>	<b>6</b>
<b>CHAPTER TWO</b>	
<b>HUMAN EVOLUTIONARY GENETICS.....</b>	<b>9</b>
<b>2.1 THE FUNDAMENTALS OF EVOLUTIONARY THEORY.....</b>	<b>9</b>
<b>2.2 HUMAN EVOLUTIONARY GENETICS.....</b>	<b>11</b>
2.2.1 The extent and source of genetic variation.....	11
2.2.1.1 Mutation.....	12
2.2.1.2 Inheritance of genetic variation.....	14
2.2.2 Genetic drift.....	15
2.2.2.1 Effective population size.....	16
2.2.2.2 Population subdivision.....	16
2.2.2.3 Migration and gene flow.....	17
2.2.3 Natural selection.....	18
2.2.4 Genetic markers used to study genetic variation.....	20
<b>2.3 EVOLUTIONARY HISTORY OF MODERN HUMANS IN AFRICA.....</b>	<b>22</b>
2.3.1 Origin of modern humans in Africa.....	23
2.3.2 The distribution of genetic diversity.....	25
2.3.3 Effective population sizes of early human populations.....	26
2.3.4 Migrations and demographic changes of African populations.....	27
2.3.4.1 Migration out of Africa.....	27
2.3.4.2 Early migrations in Africa.....	29
2.3.4.3 The Bantu migrations.....	31
2.3.5 The prehistory of southern African Khoi-San-speaking populations.....	33
2.3.6 The prehistory of southern African Bantu-speaking populations.....	35
2.3.6.1 The history of the Tswana-speaking populations of southern Africa.....	36

**CHAPTER THREE**

<b>MITOCHONDRIAL DNA AND HUMAN EVOLUTION.....</b>	<b>39</b>
<b>3.1 HISTORY AND DEVELOPMENT.....</b>	<b>39</b>
<b>3.2 MITOCHONDRIAL STRUCTURE AND MORPHOLOGY.....</b>	<b>40</b>
<b>3.3 MITOCHONDRIAL FUNCTION.....</b>	<b>41</b>
<b>3.4 MITOCHONDRIAL DNA.....</b>	<b>45</b>
3.4.1 Mitonuclear interactions.....	49
<b>3.5 UNIQUE CHARACTERISTICS OF HUMAN MITOCHONDRIAL DNA.....</b>	<b>50</b>
3.5.1 Copy number of mitochondrial DNA.....	50
3.5.2 Mutation rate of mitochondrial DNA.....	51
3.5.3 Maternal inheritance.....	52
3.5.4 Lack of recombination.....	54
3.5.5 Homoplasmy and heteroplasmy.....	54
3.5.6 Effective population size.....	55
3.5.7 Neutrality versus selection.....	56
<b>3.6 MITOCHONDRIAL DNA VARIATION.....</b>	<b>58</b>
3.6.1 The nature of human mitochondrial DNA variation.....	59
3.6.2 Mitochondrial DNA variation in studies of human evolution.....	61
<b>3.7 MITOCHONDRIAL DNA HAPLOGROUPS.....</b>	<b>63</b>
3.7.1 Mitochondrial haplogroup dispersal in the world.....	65
3.7.1.1 Origin of anatomically modern humans.....	66
3.7.1.2 Out of Africa.....	67
3.7.1.3 Migration to Oceania and Australia.....	68
3.7.1.4 Migrations into Europe.....	68
3.7.1.5 Migration to the Americas.....	69
3.7.2 Mitochondrial haplogroup dispersal in Africa.....	70
3.7.2.1 Mitochondrial haplogroup L0.....	71
3.7.2.2 Mitochondrial haplogroup L1.....	72
3.7.2.3 Mitochondrial haplogroup L5.....	73
3.7.2.4 Mitochondrial haplogroup L2.....	73
3.7.2.5 Mitochondrial haplogroup L3.....	74
3.7.2.6 Mitochondrial haplogroup L4.....	75
3.7.2.7 Mitochondrial haplogroup L6.....	75
<b>3.8 MITOCHONDRIA AND DISEASE.....</b>	<b>76</b>
3.8.1 Mitochondrial haplogroups and diseases.....	78

**CHAPTER FOUR**

<b>PHYLOGENETIC ANALYSES.....</b>	<b>81</b>
<b>4.1 THE ROLE OF GENETIC DIVERSITY IN PHYLOGENETIC ANALYSES.....</b>	<b>81</b>
4.1.1 Homology and sequence alignment.....	82

<b>4.2</b>	<b>THE ROLE OF EVOLUTIONARY MODELS IN PHYLOGENETIC ANALYSES.....</b>	<b>86</b>
4.2.1	Modelling evolution.....	86
4.2.2	Base composition parameters in evolutionary models.....	88
4.2.3	Base substitution parameters in evolutionary models.....	88
4.2.4	Rate heterogeneity parameters in evolutionary models.....	90
<b>4.3</b>	<b>PHYLOGENETIC METHODS.....</b>	<b>91</b>
4.3.1	Basic principles of tree-building methods.....	92
4.3.1.1	Tree building by using distance or discrete data.....	92
4.3.1.2	Tree building by clustering or searching.....	93
4.3.2	Distance methods.....	94
4.3.2.1	Unweighted pair-group method .....	95
4.3.2.2	Neighbour-joining method.....	95
4.3.3	Discrete methods.....	96
4.3.3.1	Maximum parsimony .....	96
4.3.3.2	Maximum likelihood .....	98
4.3.4	Choosing a phylogenetic tree-building method.....	99
 <b>CHAPTER FIVE</b>		
	<b>MATERIALS AND METHODS.....</b>	<b>101</b>
<b>5.1</b>	<b>ETHICAL APPROVAL OF THE STUDY.....</b>	<b>101</b>
<b>5.2</b>	<b>SAMPLE DESIGN AND METHODS.....</b>	<b>101</b>
<b>5.3</b>	<b>DNA ISOLATION.....</b>	<b>103</b>
<b>5.4</b>	<b>POLYMERASE CHAIN REACTION.....</b>	<b>103</b>
5.4.1	PCR primers.....	103
5.4.2	PCR reaction.....	105
5.4.3	PCR conditions.....	105
<b>5.5</b>	<b>AGAROSE GEL ELECTROPHORESIS.....</b>	<b>106</b>
<b>5.6</b>	<b>DNA PURIFICATION.....</b>	<b>106</b>
<b>5.7</b>	<b>DNA QUANTIFICATION.....</b>	<b>107</b>
<b>5.8</b>	<b>AUTOMATED DNA SEQUENCING.....</b>	<b>108</b>
5.8.1	Sequencing strategy and primers.....	109
5.8.2	Cycle sequencing reaction protocol.....	109
5.8.3	Cycle sequencing reaction conditions.....	111
5.8.4	Post-sequencing treatment.....	111
5.8.5	Purification of sequence extension product.....	112
5.8.6	Capillary electrophoresis.....	112
<b>5.9</b>	<b>DATA ANALYSES.....</b>	<b>113</b>
5.9.1	Data review.....	113
5.9.2	Resequenced Cambridge Reference Sequence.....	114
5.9.3	Sequence alignment.....	114

<b>5.10</b>	<b>MITOCHONDRIAL GENOME REGIONS USED IN SEQUENCE DATASETS</b> .....	<b>114</b>
<b>5.11</b>	<b>MITOCHONDRIAL GENOME SEQUENCE DATASETS</b> .....	<b>115</b>
5.11.1	Global African dataset.....	116
5.11.2	All African dataset.....	117
5.11.3	Tswana dataset.....	117
5.11.4	Regional African datasets.....	117
5.11.5	Assignment of the subsets.....	118
5.11.6	Ethnicity of the individuals in the datasets.....	118
<b>5.12</b>	<b>DETERMINATION OF HAPLOGROUPS</b> .....	<b>119</b>
<b>5.13</b>	<b>PHYLOGENETIC ANALYSES</b> .....	<b>123</b>
5.13.1	Step 1: Data acquisition.....	123
5.13.2	Step 2: Sequence alignment.....	123
5.13.3	Step 3: Phylogenetic analyses.....	124
5.13.3.1	Transition:Transversion ratio calculation.....	126
5.13.3.2	Gamma-shaped parameter calculation.....	126
5.13.3.3	Rooting the phylogenetic trees.....	127
5.13.3.4	Tree-building methods used in this investigation.....	127
5.13.3.5	Neighbour-joining tree.....	127
5.13.3.6	Consensus trees.....	128
5.13.3.7	Maximum parsimony tree.....	129
<b>5.14</b>	<b>STATISTICAL ANALYSES</b> .....	<b>129</b>
5.14.1	Nucleotide composition.....	131
5.14.2	Nucleotide diversity.....	132
5.14.3	Population size.....	133
5.14.4	Selection.....	136
5.14.4.1	Gene-specific effects of selection.....	140
5.14.5	Population genetic structure.....	141
5.14.6	Coalescence-time estimation.....	143
5.15	<b>CONSTRUCTION OF A CONSENSUS SEQUENCE FOR THE TSWANA-SPEAKING COHORT OF THIS INVESTIGATION</b> .....	<b>145</b>
 <b>CHAPTER SIX</b>		
	<b>RESULTS AND DISCUSSION</b> .....	<b>147</b>
<b>6.1</b>	<b>POLYMERASE CHAIN REACTION</b> .....	<b>148</b>
6.1.1	Primers.....	149
6.1.2	PCR optimisation.....	149
6.1.3	PCR efficiency.....	150
6.1.4	Secondary PCR product.....	152
6.1.5	Primer-dimers.....	153
6.1.6	PCR product smearing.....	154



<b>6.2</b>	<b>AGAROSE GEL ELECTROPHORESIS.....</b>	<b>154</b>
<b>6.3</b>	<b>DNA PURITY AND QUANTITY.....</b>	<b>156</b>
<b>6.4</b>	<b>AUTOMATED DNA SEQUENCING OF THE FULL MITOCHONDRIAL GENOME.....</b>	<b>156</b>
6.4.1	Sequencing strategy.....	157
<b>6.5</b>	<b>DATA ANALYSIS RESULTS.....</b>	<b>160</b>
6.5.1	Sequence alignment.....	160
6.5.2	DNA contiguous sequences .....	161
6.5.3	Data quality.....	161
6.5.4	Sequencing errors and artefacts.....	162
6.5.4.1	Dye blobs.....	163
6.5.4.2	Weak signal.....	164
6.5.4.3	Trailing peaks.....	165
6.5.4.4	Truncated sequence.....	165
6.5.4.5	Signal loss at the end of the sequence.....	166
6.5.4.6	Sudden signal loss in the middle of the sequence.....	167
6.5.4.7	Poor resolution.....	167
6.5.4.8	Double sequence between np 16262 and 16282.....	168
6.5.4.9	N-5 peaks.....	169
6.5.4.10	Spikes.....	170
6.5.4.11	Homopolymeric tracks.....	171
6.5.4.12	Noisy data.....	174
6.5.4.13	Failed reactions.....	175
<b>6.6</b>	<b>SEQUENCING RESULTS OF ALL PCR REGIONS.....</b>	<b>176</b>
6.6.1	Primer region 1.....	179
6.6.1.1	Sequence alterations observed in primer region 1.....	182
6.6.2	Primer region 2.....	189
6.6.2.1	Sequence alterations observed in primer region 2.....	191
6.6.3	Primer region 3.....	195
6.6.3.1	Sequence alterations observed in primer region 3.....	197
6.6.4	Primer region 4.....	203
6.6.4.1	Sequence alterations observed in primer region 4.....	205
6.6.5	Primer region 5.....	216
6.6.5.1	Sequence alterations observed in primer region 5.....	218
6.6.6	Primer region 6.....	224
6.6.6.1	Sequence alterations observed in primer region 6.....	226
6.6.7	Primer region 7.....	236
6.6.7.1	Sequence alterations observed in primer region 7.....	237
6.6.8	Primer region 8.....	244
6.6.8.1	Sequence alterations observed in primer region 8.....	246

<b>6.7</b>	<b>HAPLOGROUP CLASSIFICATION OF THE TSWANA POPULATION.....</b>	<b>263</b>
6.7.1	The haplogroup classification systems used in this investigation.....	264
6.7.1.1	Overall comparison of the two haplogroup classification systems used in this investigation.....	266
6.7.2	Haplogroups of the Tswana-speaking individuals of this investigation.....	267
6.7.3	Haplogroup L0.....	269
6.7.4	Haplogroup L1.....	276
6.7.5	Haplogroup L2.....	277
6.7.6	Haplogroup L3.....	278
6.7.7	Overview of haplogroups present in the Tswana population.....	280
<b>6.8</b>	<b>PHYLOGENETIC ANALYSES.....</b>	<b>282</b>
6.8.1	NJ tree of the Global African dataset.....	285
6.8.2	MP tree of the Global African dataset.....	302
6.8.3	NJ tree of the All African dataset.....	318
6.8.4	MP tree of the All African dataset.....	330
6.8.5	NJ tree of the Tswana dataset.....	345
6.8.6	MP tree of the Tswana dataset.....	355
<b>6.9</b>	<b>STATISTICAL ANALYSES.....</b>	<b>364</b>
6.9.1	Nucleotide composition.....	366
6.9.2	Nucleotide diversity.....	369
6.9.3	Population size.....	374
6.9.4	Selection.....	390
6.9.5	Population genetic structure.....	400
6.9.6	Coalescence-time estimation.....	405
<b>6.10</b>	<b>TSWANA MTDNA CONSENSUS SEQUENCE.....</b>	<b>414</b>
 <b>CHAPTER SEVEN</b>		
	<b>CONCLUSIONS.....</b>	<b>417</b>
<b>7.1</b>	<b>STANDARDS USED FOR DETERMINATION OF GENETIC VARIATION....</b>	<b>419</b>
<b>7.2</b>	<b>MITOCHONDRIAL VARIATION IN THE TSWANA POPULATION DUE TO INDIVIDUAL FACTORS.....</b>	<b>421</b>
7.2.1	Sequence variation displayed in the mtDNA genomes.....	421
7.2.1.1	Novel sequence variants observed in this investigation.....	425
7.2.2	Sequence variation displayed in haplogroups observed in this investigation.....	427
<b>7.3</b>	<b>MITOCHONDRIAL VARIATION IN THE TSWANA POPULATION DUE TO POPULATION BEHAVIOUR.....</b>	<b>431</b>
7.3.1	Genetic diversity.....	431
7.3.2	Genetic drift.....	432
7.3.3	Population size and migration.....	433
7.3.4	Population genetic structure.....	437

7.3.5	Selection.....	441
<b>7.4</b>	<b>COALESCENCE-TIME ESTIMATIONS.....</b>	<b>444</b>
<b>7.5</b>	<b>MODEL OF GENETIC VARIATION.....</b>	<b>449</b>
<b>7.6</b>	<b>IMPLICATIONS OF THE MITOCHONDRIAL DNA CONSENSUS SEQUENCE OF THE TSWANA POPULATION.....</b>	<b>453</b>
<b>7.7</b>	<b>IMPLICATIONS OF MITOCHONDRIAL DIVERSITY OF THE TSWANA POPULATION FOR FUTURE STUDIES.....</b>	<b>454</b>
7.7.1	Mitochondrial disease.....	455
7.7.2	Contribution of genetic variation data of the South African Bantu Speakers to the global human phylogeny.....	456
7.7.3	Genetic diversity among Bantu-speaking populations of South Africa.....	457
7.7.4	Genetic diversity within the Tswana-speaking population of South Africa	459
7.7.5	Use of additional markers for the study of genetic diversity within and between South African Bantu-speaking populations.....	459
7.7.6	Future population genetic inferences from genomic sequence data.....	460
7.7.7	Investigation of the extent of the effects of selection on genetic variation...	460
 <b>CHAPTER EIGHT</b>		
	<b>REFERENCES.....</b>	<b>463</b>
 <b>APPENDIX A</b>		
	<b>HAPLOGROUP CLASSIFICATION OF THE MITOCHONDRIAL SEQUENCES OF THE TSWANA POPULATION.....</b>	<b>483</b>
 <b>APPENDIX B</b>		
	<b>GLOBAL AFRICAN MITOCHONDRIAL GENOME DATASET USED IN THIS STUDY.....</b>	<b>485</b>
 <b>APPENDIX C</b>		
	<b>LIST OF MITOCHONDRIAL DNA SEQUENCES EXCLUDED FROM THE GLOBAL AFRICAN DATASET TO COMPILE AN ALL AFRICAN DATASET.....</b>	<b>511</b>
 <b>APPENDIX D</b>		
	<b>REGIONAL MITOCHONDRIAL DNA GENOME DATASETS USED IN THIS STUDY.....</b>	<b>521</b>
 <b>APPENDIX E</b>		
	<b>GLOBAL AFRICAN AND ALL AFRICAN PHYLOGENETIC TREES OF THIS INVESTIGATION.....</b>	<b>533</b>
 <b>APPENDIX F</b>		
	<b>PHYLOGENETIC TREES OF THE TSWANA SPEAKING INDIVIDUALS OF THIS INVESTIGATION.....</b>	<b>543</b>

<b>APPENDIX G</b>	
<b>MITOCHONDRIAL CONSENSUS SEQUENCE FOR A TSWANA SPEAKING POPULATION OF SOUTH AFRICA.....</b>	<b>549</b>
<b>APPENDIX H</b>	
<b>TSWANA MTDNA CONSENSUS SEQUENCE VARIANTS THAT DIFFERED FROM THE rCRS.....</b>	<b>555</b>

# LIST OF ABBREVIATIONS AND SYMBOLS

---

Symbols and abbreviations are listed in alphabetical order:

## LIST OF SYMBOLS

$\alpha$	alpha used to indicate the gamma shape parameter
$\beta$	beta used to indicate the gamma shape parameter
$Cov(\hat{d}_{ij}, \hat{d}_{kl})$	covariance based on phylogenetic relationship among sequences
$\sigma_a^2$	covariance component due to differences among populations
$\sigma_b^2$	covariance component due to differences among haplotypes in different populations within a group
$\sigma_T^2$	total molecular variance
$\kappa$	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
$\epsilon$	epsilon
=	equal to
$\eta$	eta used to indicate the total number of mutations
$\gamma$	gamma
$\Gamma$	gamma shaped parameter
-	gap in DNA sequence
>	greater than
<	less than
$\lambda$	lamda used to indicate variation of the substitution rate
$\mu$	micro ( $10^{-6}$ )
$\mu\text{L}$	microlitre
x	mismatch distribution
$\mu$	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
●	<i>Pan troglodytes</i> outgroup position indicated with a red circle in phylogenetic trees
%	percent
$\pi$	pi, nucleotide diversity or the average number of pairwise differences between DNA sequences
$\pi_n$	the mean number of pairwise differences for n sequences
◆	rCRS position indicated with a green diamond in phylogenetic trees
®	registered trademark
$\rho$	rho, average number of nucleotide differences between a set of DNA sequences and a specified DNA sequence
$\hat{V}_s$	sampling variance
$\sigma^2$	sigma squared indicating variance
$V_{st}$	stochastic variance
$\sqrt{\quad}$	square root
$\theta_\pi$	the mean number of nucleotide differences between two sequences
$\theta$	theta, expected pairwise nucleotide site differences also referred to as the population parameter
$\theta_0$	initial population size
$\theta_1$	population size after expansion

## LIST OF ABBREVIATIONS AND SYMBOLS

$Q(i)$	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
$\kappa$	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	<i>Anno Domini</i> (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 <i>AluI</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
<i>ATP6</i>	ATP synthase F0 subunit 6 gene
<i>ATP8</i>	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) <sub>n</sub>	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 <sub>2</sub>	carbon dioxide ion
COI	cytochrome c oxidase subunit I
COII	cytochrome c oxidase subunit II
COIII	cytochrome c oxidase subunit III
<i>COI</i>	cytochrome c oxidase subunit I gene
<i>COII</i>	cytochrome c oxidase subunit II gene
<i>COIII</i>	cytochrome c oxidase subunit III gene
CoQ	coenzyme Q or ubiquinone
CoQH <sub>2</sub>	reduced coenzyme Q
Cov	covariance
COX	cytochrome c oxidase
CPEO	chronic progressive external ophthalmoplegia
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 <sub>B</sub>	copper B centre of the Q cycle in the mitochondria
Cys	cysteine amino acid
Cytb	cytochrome b
<i>Cytb</i>	cytochrome b gene
Cytc	cytochrome c
<i>Cytc</i>	cytochrome c gene
<i>d</i>	maximum number of nucleotide differences
<i>D</i>	Tajima's D test statistic
D	tRNA / amino acid aspartic acid
<i>D*</i>	Fu and Li's D* test statistic
dATP	2'-deoxyadenosine-5'-triphosphate
dCTP	2'-deoxycytidine-5'-triphosphate
ddH <sub>2</sub> O	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 <sup>-</sup>	electron
E	tRNA / amino acid glutamic acid
EDTA	ethylenediamine tetra-acetic acid
<i>et al.</i>	<i>et alia</i> : 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
<i>F</i>	fixation index
<i>F*</i>	Fu and Li's F test statistic
f <sub>0</sub>	probability of identity by descent of two different genes drawn from the same population
f <sub>1</sub>	probability of identity by descent of two genes drawn from two different populations
F <sub>1</sub> ATPase	F <sub>1</sub> subunit of adenosine tri-phosphate synthase
F81/FEL	Felsenstein's model
FADH <sub>2</sub>	Flavin Adenine Dinucleotide
FL	fragment length

## LIST OF ABBREVIATIONS AND SYMBOLS

F <sub>s</sub>	Fu's statistic
F <sub>ST</sub>	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 <sup>®1</sup>	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 <sup>®</sup>
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 <sup>+</sup>	hydrogen ion
h	gap-extension penalty
H <sub>2</sub> O	water
H <sub>2</sub> O <sub>2</sub>	hydrogen peroxide
HKA	Hudson-Kreitman-Aguadé
HKY85/HKY	Hasegawa, Kishino and Yano model of evolution
<i>Hpal</i>	restriction endonuclease isolated from a recombinant from <i>Haemophilus parainfluenzae</i>
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.	<i>id est</i> : 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
<i>l<sub>i</sub></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<sup>®</sup> is a registered trademark of the U.S. Department of Health and Human Services, Independence Avenue, S.W., Washington, DC, USA.



LIST OF ABBREVIATIONS AND SYMBOLS

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 <sup>2+</sup>	magnesium ion
MgCl <sub>2</sub>	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 <sup>+</sup>	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
<i>ND1</i>	NADH dehydrogenase subunit 1 gene
<i>ND2</i>	NADH dehydrogenase subunit 2 gene
<i>ND3</i>	NADH dehydrogenase subunit 3 gene
<i>ND4</i>	NADH dehydrogenase subunit 4 gene
<i>ND4L</i>	NADH dehydrogenase subunit 4L gene
<i>ND5</i>	NADH dehydrogenase subunit 5 gene
<i>ND6</i>	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

## LIST OF ABBREVIATIONS AND SYMBOLS

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 $i^{\text{th}}$ 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 hydrolysis of ATP and ADP
$P_L$	replication promotor 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

LIST OF ABBREVIATIONS AND SYMBOLS

S <sub>H</sub>	haplogroup associated synonymous substitutions
SINEs	short interspersed nuclear elements
SNP	single nucleotide polymorphism
S <sub>P</sub>	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
T <sub>s</sub>	transition
T <sub>a</sub>	annealing temperature
T-A	thymine paired to alanine in double-stranded DNA
Taq	Taq polymerase: DNA deoxynucleotidyltransferase from <i>Thermus aquaticus</i>
TBE	Tris-borate-EDTA
TCA	tricarboxylic acid cycle
Thr	threonine amino acid
T-G	thymine paired to guanine in double-stranded DNA
T <sub>m</sub>	melting temperature
TMRCA	time to most recent common ancestor
TrisCl	an organic compound known as tris(hydroxymethyl)aminomethane, with the formula (HOCH <sub>2</sub> ) <sub>3</sub> CNH <sub>2</sub> .
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

---

Equation No.	Title of Equation	Page
4.1	Gap penalty.....	84
4.2	Variation of substitution rate.....	90
5.1	Gamma value.....	126
5.2	Average number of nucleotide differences and nucleotide diversity.....	132
5.3	Sampling variance of nucleotide diversity.....	133
5.4	Fu's $F_s$ statistic.....	134
5.5	Mismatch distribution under constant population size and no recombination.....	134
5.6	Raggedness statistic.....	135
5.7	Ramos Onsins and Rozas $R_2$ statistic.....	135
5.8	Tajima's D statistic.....	137
5.9	Fu and Li's $D^*$ test statistic.....	138
5.10	Fu and Li's $F^*$ test statistic.....	139
5.11	Neutrality index.....	141
5.12	Fixation index.....	142
5.13	Determination of total sum of squares, F-statistics and covariance components for haplotype data within one group.....	143
5.14	Estimates for n and fixation index defined.....	143
5.15	Estimator of the genetic distance to the ancestral node of a haplogroup, subhaplogroup or lineage.....	144
5.16	Efficiency of coalescence time estimation of a sample.....	144



# LIST OF FIGURES

---

<b>Figure No.</b>	<b>Title of Figure</b>	<b>Page</b>
2.1	Demographic history of early human populations.....	28
2.2	Population migrations within Africa.....	33
3.1	OXPHOS system.....	42
3.2	Q cycle.....	43
3.3	Functional organisation of the human mitochondrial DNA.....	47
3.4	Global mitochondrial haplogroup hierarchy.....	64
3.5	MtDNA and the migration of world populations.....	66
3.6	Macrohaplogroup L hierarchy.....	71
4.1	Two approaches to construct evolutionary models.....	87
5.1	Wallace classification system of informative SNPs used to define macrohaplogroup L.....	121
5.2	Outline of the PhyloTree classification system for macrohaplogroup L.....	122
6.1	Photographic representation of secondary product of Region 7 optimisation at low $T_a$ .....	152
6.2	Photographic representation of primer-dimer product of Region 7 optimisation at low $T_a$ .....	153
6.3	Photographic representation of the smear found with optimisation of region 7 primers.....	154
6.4	Photographic image of the UV artefact spots observed in some of the gels.....	155
6.5	Example of distorted DNA fragments on a gel.....	156
6.6	Sequencing primers for the eight PCR regions.....	158
6.7	Dye blobs.....	164
6.8	Example of trailing peaks.....	165
6.9	Truncated sequence.....	166
6.10	Example of signal loss towards the end of the sequence.....	166
6.11	Example of loss of signal in the middle of the sequence.....	167
6.12	Example of poor peak resolution.....	168
6.13	Double sequence between nucleotide positions 16262 and 16282.....	168
6.14	Alignment of the sequence segment containing the artefact with the rCRS...	169
6.15	Example of N-5 artefact.....	170
6.16	Example of spike peaks.....	171
6.17	Homopolymer regions between nucleotide positions 303 and 315.....	171

LIST OF FIGURES

6.18	Homopolymer region between nucleotide positions 568 and 573.....	172
6.19	Homopolymer region between nucleotide positions 957 and 966.....	172
6.20	Homopolymer region between nucleotide positions 16184 and 16193.....	172
6.21	Sequence overlap of homopolymer region between forward and reverse primed sequences.....	173
6.22	Example of noisy data.....	174
6.23	Primer regions and a map of the functional areas of mitochondrial DNA.....	176
6.24	Photographic representation of the amplified mtDNA product of primer region 1.....	179
6.25	Representative electropherograms of the sequence generated for primer region 1 using the forward primers 1-4.....	180
6.26	Electropherograms of sequences that were sequenced by reverse primers in primer region 1.....	181
6.27	Representative electropherogram of the sequence data generated indicating a transversion at np 576.....	185
6.28	Representative electropherograms of the sequence data generated indicating transitions at np 211 and np 267.....	186
6.29	Length variation between np 309 and np 315.....	188
6.30	Structure of the tRNA phenylalanine (F) and observed sequence variation of the Tswana-speaking individuals of this investigation.....	188
6.31	Photographic representation of the amplified mtDNA product of primer region 2.....	189
6.32	Representative electropherograms of the sequence generated for primer region 2 using the forward primers 1-4.....	190
6.33	Locations of the sequence alterations within the 12S rRNA and 16S rRNA sequences of the Tswana cohort of this investigation.....	192
6.34	Representative electropherograms of the sequence data generated indicating transitions at np 980 and np 1415.....	193
6.35	Representative electropherograms of the sequence data generated indicating transitions at np 3202.....	194
6.36	Photographic representation of the amplified mtDNA product of primer region 3.....	196
6.37	Representative electropherograms of the sequence generated for primer region 2 using the forward primers 1-4.....	196
6.38	Representative electropherograms of the sequence data generated indicating a transition at np 3660 and a transversion at np 4048.....	200
6.39	Representative electropherograms of the sequence data generated indicating transitions at np 4011 and np 4023.....	201
6.40	Structure of the tRNA isoleucine (I) and observed sequence variation of the Tswana-speaking individuals of this investigation.....	202
6.41	Photographic representation of the amplified mtDNA product of primer region 4.....	204
6.42	Representative electropherograms of the sequence generated for primer region 4 using the forward primers 1-4.....	204



6.43	Representative electropherograms of the sequence data generated indicating transitions at np 4896, np 5782 and np 6083.....	210
6.44	Representative electropherograms of the sequence data generated indicating transitions at np 4814, np 4943 and np 7046 .....	211
6.45	Structure of the tRNA tryptophan, tRNA alanine, tRNA asparagine, tRNA cysteine and tRNA aspartic acid.....	213
6.46	Photographic representation of the amplified mtDNA product of primer region 5.....	216
6.47	Representative electropherograms of the sequence generated for primer region 5 using the forward primers 1-4.....	217
6.48	Representative electropherograms of the sequence data generated indicating transitions at np 7741, np 8014, np 8793, np 9039, np 9058 and np 9181.....	222
6.49	Photographic representation of the amplified mtDNA product of primer region 6.....	225
6.50	Representative electropherograms of the sequence generated for primer region 6 using the forward primers 1-4.....	225
6.51	Representative electropherograms of the sequence generated for the novel sequence alteration at np 9297.....	231
6.52	Structure of the tRNA arginine.....	232
6.53	Representative electropherograms of the sequence data generated indicating transitions at np 9278, np 10237 and np 10427 and a transversion at np 10128.....	233
6.54	Photographic representation of the amplified mtDNA product of primer region 7.....	236
6.55	Representative electropherograms of the sequence generated for primer region 7 using the forward primers 1-4.....	237
6.56	Representative electropherograms of the sequence generated of novel sequence alterations at np 10948 and np 12004.....	241
6.57	Representative electropherograms of the sequence data generated indicating a transition at np 10966 and a transversion at np 11557.....	242
6.58	Structure of the tRNA histidine and tRNA serine2.....	243
6.59	Photographic representation of the amplified mtDNA product of primer region 8.....	245
6.60	Representative electropherograms of the sequence generated for primer region 8 using the forward primers.....	245
6.61	Structure of the tRNA glutamic acid and tRNA threonine.....	253
6.62	Representative electropherograms of novel sequence alterations observed in the ND5 gene at np 12436, 13077, 13473, 13604 and 13767.....	254
6.63	Representative electropherograms of novel sequence alterations observed in the ND6 gene at np 14163 and np 14425.....	256
6.64	Representative electropherograms of a novel sequence alteration observed in the Cytb gene at np 15364.....	257
6.65	Representative electropherogram of a novel sequence alteration observed in the ND5 gene at np 12528.....	258

LIST OF FIGURES

6.66	Representative electropherogram of sequence alteration observed in the ND6 gene at np 14290.....	258
6.67	Representative electropherogram of sequence alterations observed in the Cytb gene at np 15140, np 15315 and np 15337.....	259
6.68	Wallace classification system of informative SNPs used to define macrohaplogroup L.....	264
6.69	Outline of the PhyloTree classification system for macrohaplogroup L.....	266
6.70	Haplogroup distribution of the Tswana-speaking population under investigation according to the Wallace classification system.....	268
6.71	Haplogroup distribution of the Tswana-speaking population under investigation according to the Phylotree classification system.....	269
6.72	Pie chart distribution of haplogroups observed in the Tswana population of this investigation.....	281
6.73	NJ tree of Global African dataset 1a.....	286
6.74	MP tree of the Global African dataset.....	303
6.75	NJ tree of All African dataset.....	319
6.76	MP tree of the All African dataset.....	331
6.77	NJ tree of the Tswana dataset.....	346
6.78	MP tree of the Tswana dataset.....	356
6.79	Mismatch distributions for the Global African and All African populations under the assumption of a sudden population expansion model.....	384
6.80	Mismatch distribution for the Eastern African population under the assumption of a sudden population expansion model.....	387
6.81	Mismatch distribution for the Western African population under the assumption of a sudden population expansion model.....	388
6.82	Mismatch distributions for the Southern African and Tswana populations under the assumption of a sudden population expansion model.....	389
7.1	Pie chart distribution of haplogroups observed in the Tswana population of this investigation.....	428
7.2	Model of genetic variation.....	450

# LIST OF TABLES

---

<b>Table No.</b>	<b>Title of Table</b>	<b>Page</b>
3.1	Functional organisation of human mitochondrial DNA.....	48
5.1	Sample identification numbers.....	102
5.2	Eight (8) primer pairs that were used to amplify the full mitochondrial genomes of the Tswana-speaking cohort of this investigation.....	104
5.3	Primers used to sequence the full mitochondrial genome.....	109
5.4	Description of reagents in BigDye <sup>®</sup> Terminator v3.1 Cycle Sequencing Kit.....	110
5.5	Amount of PCR product used for sequencing reactions in this investigation...	111
5.6	Technical specifications of genetic analysers used.....	113
5.7	Ethnicity, country and region of origin of the ethnic groups included in this investigation.....	119
5.8	Software programs used for statistical analyses.....	130
6.1	Primer pairs used in this investigation.....	149
6.2	Average DNA quantity for the different PCR regions under optimised conditions.....	151
6.3	Average DNA quantity obtained for the eight (8) different PCR regions.....	156
6.4	Overlap between PCR regions.....	161
6.5	Reverse primers used in this investigation.....	173
6.6	Functional locations of mitochondrial DNA.....	177
6.7	Sequence alterations observed between the complete mitochondrial DNA of the Tswana individuals included in this study and the rCRS in primer region 1	182
6.8	Observed sequence alterations in individuals that did not belong to the L macrohaplogroup.....	186
6.9	Sequence alterations observed between the complete mitochondrial DNA of the Tswana individuals included in this study and the rCRS in primer region 2	191
6.10	Reported mtDNA sequence alterations with pathological associations within primer region 2.....	194
6.11	Sequence alterations observed between the complete mitochondrial DNA of the Tswana individuals included in this study and the rCRS in primer region 3	198
6.12	Sequence variation within ND1 gene.....	198
6.13	Sequence alterations observed between the complete mitochondrial DNA of the Tswana individuals included in this study and the rCRS in primer region 4	205
6.14	Sequence variation within ND2 and COI genes .....	208
6.15	Reported mtDNA sequence alterations with disease associations within primer region 4.....	215

LIST OF TABLES

6.16	Sequence alterations observed between the complete mitochondrial DNA of the Tswana individuals included in this study and the rCRS in primer region 5	218
6.17	Sequence variation within COII, ATP8 and ATP6 genes.....	220
6.18	Sequence alterations observed between the complete mitochondrial DNA of the Tswana individuals included in this study and the rCRS in primer region 6	227
6.19	Sequence variation within COIII, ND3 and ND4L genes.....	228
6.20	Reported mtDNA sequence alterations within primer region 6 with disease associations.....	234
6.21	Sequence alterations observed between the complete mitochondrial DNA of the Tswana individuals included in this study and the rCRS in primer region 7	238
6.22	Sequence variation within the ND4 gene.....	239
6.23	Sequence alterations observed between the complete mitochondrial DNA of the Tswana individuals included in this study and the rCRS in primer region 8	247
6.24	Sequence variation within the ND5, ND6 and Cytb genes.....	250
6.25	Reported mtDNA sequence alterations with disease associations within primer region 8.....	261
6.26	Tswana-speaking individuals of this investigation assigned to haplogroup L0 by the Wallace classification system.....	270
6.27	Tswana-speaking individuals of this investigation assigned to haplogroup L0 by the PhyloTree classification system.....	270
6.28	Sequence alterations observed in the Tswana-speaking individual TS_5063..	272
6.29	Tswana-speaking individuals of this investigation assigned to haplogroup L1 by the Wallace classification system.....	276
6.30	Tswana-speaking individuals of this investigation assigned to haplogroup L1 by the PhyloTree classification system.....	276
6.31	Tswana-speaking individuals of this investigation assigned to haplogroup L2 by the Wallace classification system.....	277
6.32	Tswana-speaking individuals of this investigation assigned to haplogroup L2 by the PhyloTree classification system.....	278
6.33	Tswana-speaking individuals of this investigation assigned to haplogroup L3 by the Wallace classification system.....	279
6.34	Tswana-speaking individuals of this investigation assigned to haplogroup L3 by the PhyloTree classification system.....	279
6.35	Gamma shaped parameter values for datasets of this investigation.....	284
6.36	Haplogroups L4, 5 and 6.....	287
6.37	Bootstrap values for the Global African NJ tree of this investigation.....	292
6.38	Sequences belonging to a sub-clade of haplogroup L4 of the Global African MP tree.....	308
6.39	Bootstrap values for Global African NJ tree and All African NJ tree.....	322
6.40	Bootstrap values for Global African MP tree and All African MP tree.....	334
6.41	Composition of regional African mtDNA genome datasets of this investigation.....	365
6.42	Nucleotide composition of the Tswana population of this investigation.....	366
6.43	Nucleotide composition at codon positions for the Tswana dataset of this investigation.....	368

6.44	MtDNA coding region sequence diversity statistics of African populations of this investigation.....	369
6.45	MtDNA coding region sequence diversity statistics of global and African populations.....	371
6.46	Statistical measures of population growth for the datasets of this investigation.....	375
6.47	Mismatch distribution parameters estimated under a sudden expansion model.....	382
6.48	Tajima's D and Fu and Li's D* and F* test statistics.....	392
6.49	NS/SH and NS/SP ratios for the 13 protein coding genes of the mtDNA of African individuals that belonged to haplogroups L0, L1, L2 and L3.....	395
6.50	NI and P values for the 13 protein coding genes of the mtDNA of African individuals that belonged to haplogroups L0, L1, L2 and L3.....	397
6.51	Analysis of molecular variance (AMOVA) between populations of this investigation.....	401
6.52	Maternal lineages of the Tswana population of this investigation.....	405
6.53	Coalescent time estimates of the All African dataset of this investigation.....	408
6.54	Coalescent time estimates published for haplogroup L0.....	409
6.55	Coalescent time estimates published for haplogroup L0a and sub-haplogroups.....	410
6.56	Coalescent time estimates published for haplogroup L0d and sub-haplogroups.....	412
6.57	Coalescent time estimates published for haplogroup L2a and sub-haplogroups.....	413



# 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,

## ACKNOWLEDGEMENTS

**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.