

Molecular Genetic Variation in the Malagasy

Cindy Dorf

**A dissertation submitted in full fulfilment of the
requirements for the degree of Master of Science in
Human Genetics at the University of the
Witwatersrand, Johannesburg**

September, 2001

ABSTRACT

The island of Madagascar has been separated from Africa for almost 70 million years and was the last of the major land masses to be colonised by humans. Several theories concerning Malagasy origins have been proposed, but the prehistory of the Malagasy is still unresolved. Historical, archaeological, linguistic and genetic data suggest two major contributors to the Malagasy, namely African and Indonesian populations, with additional contributions from Arab, Indian, Chinese and various other Asian sources. All the islanders speak a common language called Malagasy, and although there are dialectical differences, these are not sufficient to elucidate the geographic origins of the proto-Malagasy. Some studies have reported that differences in physical appearance and culture between populations occupying different regions of the island may suggest different ancestral histories. In this study, we made use of information from three different genetic systems, namely mitochondrial DNA (mtDNA), the Y chromosome and several autosomal DNA markers to trace the origins of the males and females who contributed to the gene pool of the present-day Malagasy. We found that the mtDNA 9-bp deletion was present, throughout the island, at a total frequency of 22.5%. The 1.1 kb mtDNA control region was sequenced from 228 Malagasy individuals from which one hundred and twenty four unique mtDNA types, defined by 151 nucleotide changes were derived. Studies have shown that when the 9-bp deletion is used in conjunction with mtDNA control region sequence variation, the 9-bp deletion is a useful marker to trace population affinities. We were able to demonstrate that an overall Asian to African contribution of approximately 2:1 is indicated in present-day Malagasy populations. We used the Y *Alu* polymorphism

(YAP) in conjunction with M9 to distinguish between African and non-African Y chromosomes. Overall, an African contribution of 78.2% and a non-African contribution of 21.85% was calculated. To gain insight into the autosomal DNA found in the Malagasy, we made use of three loci with population specific alleles (Sb19.3, AT3id and APO-4) and the human mannose binding protein. Again, we observed both African and Asian affinities in the various Malagasy ethnic groups. By jointly analysing all the marker systems that were used, we were able to demonstrate a close relationship between ethnic groups from the different geographic regions on the island, and found the Malagasy to be a relatively homogeneous population. To test the ethnic and geographical structure present on the island, we conducted an analysis of molecular variance (AMOVA). These tests showed little biological support in favour of such hierarchies. A correlation analysis showed that very weak correlations exist between the data generated by the different genetic systems. Furthermore, by vividly demonstrating the differential input from males and females to the present-day Malagasy, this study also illustrates the advantages of combining mtDNA data with Y chromosomal results from the same individuals.

DECLARATION

I declare that this is my own unaided work. It is being submitted for the degree of Master of Science in Human Genetics at the University of the Witwatersrand, Johannesburg. It has not been submitted for any other degree at any other university.

Cindy Dorf

8th day of September, 2001

DEDICATION

To my parents and husband, for their love and support in so many ways.

ACKNOWLEDGEMENTS

The success of this project would not have been possible without the help and support of a number of people, whom I take this opportunity to thank:

The Malagasy individuals who consented to participate in the study.

Professor Trefor Jenkins, Doctors Thomy de Ravel, Pascale Willem, Gwyn Campbell, A. Rakouth, Stan Quambek and Jean Roux for their assistance in the collection of blood specimens, and our colleagues in the Department of Human Genetics, South African Institute for Medical Research (SAIMR) for help in processing them.

My supervisor, Dr Himla Soodyall, for affording me the opportunity and the use of facilities in her laboratory (Population and Evolutionary Genetics Laboratory, SAIMR) to advance my research goals, and also for reading drafts of this thesis and for her advice and suggestions.

Professor Trefor Jenkins for taking the time to read this dissertation.

To all the members of the Department of Human Genetics, SAIMR for their support, and companionship throughout my stay at the SAIMR.

The University of the Witwatersrand, the Medical Research Council and the SAIMR for funding.

My family and friends for their support, encouragement and understanding.

TABLE OF CONTENTS

ABSTRACT.....	I
DECLARATION.....	III
DEDICATION.....	ERROR! BOOKMARK NOT DEFINED.
ACKNOWLEDGEMENTS	V
TABLE OF CONTENTS	VI
LIST OF FIGURES	IX
LIST OF TABLES	XII
ABBREVIATIONS AND SYMBOLS.....	XV
CHAPTER 1: INTRODUCTION.....	1
1.1 AIMS OF THE STUDY.....	1
1.2 MADAGASCAR.....	2
1.2.1 GEOGRAPHY AND NATURAL HISTORY OF MADAGASCAR.....	3
1.2.2 THE PEOPLES OF MADAGASCAR	5
1.2.3 THE ANCESTRY OF THE MALAGASY	8
1.3 GENETIC MARKERS USED IN THIS STUDY	20
1.3.1 HAPLOID DNA MARKERS.....	21
1.3.2 AUTOSOMAL DIPLOID MARKERS	30

CHAPTER 2: MATERIALS AND METHODS 36

2.1	SUBJECTS AND SAMPLE COLLECTION	36
2.2	MOLECULAR METHODS.....	39
2.2.1	MtDNA.....	39
2.2.2	Y CHROMOSOME.....	42
2.2.3	AUTOSOMAL DNA MARKERS.....	46
2.3	COMPUTATIONAL ANALYSES	50
2.3.1	MtDNA ANALYSES	50
2.3.2	Y CHROMOSOME ANALYSES	52
2.3.3	AUTOSOMAL DNA MARKERS.....	52
2.3.4	CORRELATION ANALYSIS	53

CHAPTER 3: RESULTS 54

3.1	MtDNA RESULTS.....	54
3.1.1	FREQUENCY OF THE 9-BP DELETION	54
3.1.2	MtDNA CONTROL REGION SEQUENCING.....	57
3.1.3	SEQUENCE TYPES ASSOCIATED WITH THE 9-BP DELETION	62
3.1.4	IDENTIFYING ORIGINS OF mtDNA TYPES NOT ASSOCIATED WITH THE 9-BP DELETION IN THE MALAGASY.....	65
3.1.5	RELATIONSHIPS BETWEEN MALAGASY ETHNIC GROUPS AND GEOGRAPHIC REGIONS	73
3.1.6	PAIRWISE DIFFERENCE DISTRIBUTIONS.....	78
3.1.7	HIERARCHICAL STRUCTURE IN THE MALAGASY	81
3.1.8	SUMMARY OF MtDNA RESULTS IN THE MALAGASY	84
3.2	Y CHROMOSOME RESULTS.....	85
3.3	AUTOSOMAL DNA RESULTS	96
3.3.1	PSA RESULTS.....	96
3.3.2	MBP RESULTS.....	109
3.4	CORRELATION ANALYSIS	118

CHAPTER 4: DISCUSSION 120

4.1	WHAT IS THE EXTENT OF GENETIC DIVERSITY IN THE PRESENT-DAY MALAGASY POPULATIONS?.....	121
4.1.1	MITOCHONDRIAL DNA.....	121
4.1.2	Y CHROMOSOME.....	123
4.1.3	AUTOSOMAL DNA MARKERS.....	124
4.2	WHAT ARE THE RELATIONSHIPS BETWEEN THE DIFFERENT ETHNIC GROUPS THROUGHOUT THE ISLAND?	128
4.3	ARE THERE DIFFERENCES IN THE PATTERNS OF VARIATION IN HIGHLAND VERSUS LOWLAND POPULATION GROUPS?	130
4.4	TO WHAT EXTENT HAVE FEMALES AND MALES CONTRIBUTED TO THE GENE POOL OF THE MALAGASY.....	131

4.5	WHAT ARE THE GENETIC AFFINITIES OF MALAGASY, AFRICAN AND INDONESIAN POPULATIONS?.....	132
4.6	IS IT POSSIBLE TO TRACE THE GENETIC TRAIL OF MIGRATION TO MADAGASCAR?.....	135
	CONCLUSION	137
	APPENDIX 1: ETHICS CLEARANCE CERTIFICATE	139
	APPENDIX 2: SOLUTIONS USED.....	140
	APPENDIX 3: 1 KB MOLECULAR SIZE STANDARD.....	146
	APPENDIX 4: THE 124 UNIQUE MTDNA TYPES FOUND IN THE MALAGASY, THE NUMBER OF INDIVIDUALS WITH EACH TYPE, THE ETHNIC GROUP(S) EACH TYPE IS FOUND IN AS WELL AS THE CONTROL REGION MUTATIONS CHARACTERISING EACH TYPE.....	147
	REFERENCES.....	154

LIST OF FIGURES

Figure 1: Map of the world indicating the size and location of Madagascar with respect specifically to Africa and Asia.....	4
Figure 2: Map of Madagascar indicating the 5 geographic regions (Central highland, Eastern Central, Northern, South Western and Western) and the location of the 19 Malagasy ethnic groups.....	6
Figure 3: A diagrammatic representation of a Mitochondrial DNA molecule.....	23
Figure 4: Electrophoretic separation of PCR products with the 9-bp deletion on a 2% Metaphor agarose gel.	55
Figure 5: Electrophoretic separation of 1.1-kb mtDNA control region PCR products, prior to excision, on a 2% NuSieve agarose gel	59
Figure 6: A CR1 sequence produced by electrophoresis on a 4.3% denaturing polyacrylamide gel run on an ABI 377 automated DNA sequencer.....	60
Figure 7: A neighbor-joining tree (constructed using the complete deletion method) indicating the relationship between the 124 unique mtDNA types found in the Malagasy.	61
Figure 8: Map of Madagascar indicating the distribution of the Asian form of the 9-bp deletion (including the Polynesian motif) and the African form of the 9-bp deletion throughout Madagascar.	64
Figure 9: Neighbor Joining tree (constructed by the complete deletion method) showing the relationship between the 124 unique mtDNA sequence types found in the Malagasy with various Asian and African populations.	68

Figure 10: Map of Madagascar indicating the distribution of mtDNA types associated with African and Asian types throughout the island.....	71
Figure 11: Neighbour joining tree (constructed by the complete deletion method) showing which of the unique mtDNA types found in the Malagasy are associated with types found in Africans and Asians.....	72
Figure 12: Principal Component plot showing the relationship of ethnic groups in Madagascar using mtDNA data.	74
Figure 13: Principal component plot showing the relationships between the Malagasy ethnic groups, and various Asian and African populations using mtDNA data.	77
Figure 14: Mismatch distribution of mtDNA types with the 9-bp deletion found in the Malagasy.	79
Figure 15: Mismatch and intermatch distribution of the African and Asian forms of the deletion in the Malagasy.	79
Figure 16: Mismatch and intermatch distributions of the Asian (A) and African (B) forms of the deletion in the Malagasy compared with Asians and Africans, respectively.....	80
Figure 17: Electrophoretic separation of PCR products with and without the Y <i>Alu</i> polymorphism on a 2% HGT agarose gel.....	87
Figure 18: Electrophoretic separation of the 340-bp M9 PCR product on a 3% HGT agarose gel.....	87
Figure 19: RFLP patterns produced by the digestion of the 340-bp M9 PCR product with <i>HinF1</i> on a 2% HGT agarose gel.....	88
Figure 20: Map of Madagascar showing the distribution of the YAP and M9 throughout the island.	90

Figure 21: Principal component plot showing the relationship between the Malagasy ethnic groups based on Y chromosome data.....	94
Figure 22: Electrophoretic separation of Sb19.3 PCR products on a 3% HGT agarose gel.....	97
Figure 23: Electrophoretic separation of AT3id PCR products on a 4% composite agarose gel.....	97
Figure 24: Electrophoretic separation of APO-4 PCR products on a 2% HGT agarose gel.....	101
Figure 25: Principal component plot showing the relationship between the Malagasy ethnic groups and various populations from the literature, using PSA data.....	106
Figure 26: Electrophoretic separation of the 349-bp MBP PCR product on a 2% HGT agarose gel.....	110
Figure 27: Electrophoretic separation MBP PCR products after digestion with <i>BanI</i> on a 2% Metaphor Agarose gel.....	110
Figure 28: Electrophoretic separation MBP PCR products after digestion with <i>MboI</i> on a 2% Metaphor Agarose gel.....	111

LIST OF TABLES

Table 1: Table showing reported frequencies of the “1” allele from various populations for each PSA system used in this study, together with the respective reference.....	32
Table 2 : Table showing reported frequencies of the B and C alleles in various populations around the world.	35
Table 3 : Malagasy ethnic groups examined for mtDNA, Y chromosome DNA and autosomal DNA variation.....	38
Table 4: Primers used in the amplification of the 9-bp deletion, the 1-kb control region of mtDNA and cycle sequencing (Hertzberg <i>et al.</i> , 1989; Vigilant, 1990).....	43
Table 5: Primers used in the YAP and M9 PCR reactions and their relevant sequences (Underhill <i>et al.</i> , 1997).	43
Table 6: PCR components, PCR conditions, gel conditions and PCR product sizes for the Sb19.3, AT3id, APO and MBP systems.....	48
Table 7: Primers used in the autosomal DNA PCR reactions and their sequences	49
Table 8 :Number of individuals and frequency of the 9-bp deletion in Malagasy ethnic groups in the different regions.	56
Table 9: African and Asian populations (with respective sample size and reference) used for comparative purposes.	67
Table 10 : MtDNA types found in the various ethnic groups and regions traced to African and Asian origins, as well as those types which couldnot be unambiguously resolved.	69
Table 11: Hierarchical structure in the Malagasy based on mtDNA variation.....	83

Table 12: Number of individuals and frequency of the YAP and M9 in Malagasy ethnic groups in the different regions.	89
Table 13: STR haplotypes (constructed in the order DYS393, DYS19, DYS390, and DYS391) occurring on a YAP- background (Morar, 2000) that we were able to resolve based on our M9 data.....	92
Table 14: Hierarchical structure in the Malagasy based Y Chromosome variation.	95
Table 15: Ethnic groups included in the Sb19.3 screen, the respective sample sizes as well as the results obtained.	98
Table 16: Ethnic groups included in the AT3id screen, the respective sample sizes as well as the results obtained.	99
Table 17: Table showing the ethnic groups included in the APO-4 screen, the respective sample sizes as well as the results obtained.	102
Table 18: Chi-squared (χ^2) and “p values” obtained for each Malagasy ethnic group with each of the PSAs included in the study.	103
Table 19: Number of individuals and frequency of the “1” allele in each ethnic group and geographic region for each PSA system.....	105
Table 20: Hierarchical structure in the Malagasy based on variation contributed by PSAs.....	108
Table 21: Ethnic groups, respective sample sizes and results obtained after digestion with <i>BanI</i>	112
Table 22: Ethnic groups, respective sample sizes and results obtained after digestion with <i>MboI</i>	113
Table 23: Table showing the frequencies of the B and C alleles for each of the groups and regions included in the MBP study.	115

Table 24: Hierarchical structure in the Malagasy based on variation contributed by the MBP.	117
Table 25: Summary Statistics generated in the correlation analysis.....	119

ABBREVIATIONS AND SYMBOLS

A	Adenine
ATP	Adenosine Triphosphate
APS	Ammonium Persulphate
AD	Anno Domini
ABI 377	Applied Biosystems 377
bp	Base Pair
BC	Before Christ
BSA	Bovine Serum Albumin
CR1	Control Region 1
CR2	Control Region 2
COII	Cytochrome Oxidase Subunit II
C	Cytosine
Da	Daltons
°	Degrees
°C	Degrees Centigrade
DNA	Deoxyribonucleic Acid
dNTPs	Dioxynucleotide Triphosphates
ddNTPs	Dideoxynucleotide Triphosphates
D-loop	Displacement Loop
dH ₂ O	Distilled Water
ddH ₂ O	Distilled Deionised Water
EtBr	Ethium Bromide
EDTA	Ethylenediaminetetra-acetic Acid
e.g.	Example
5'	Five Prime End
FREQ	Frequency
g	Grams
G6PD	Glucose-6-phosphate Dehydrogenase

G	Guanine
H strand	Heavy Strand
HGT	High Gelling Temperature
HCl	Hydrochloric Acid
HVR1	Hypervariable Region 1
HVR2	Hypervariable Region 2
HVR 3	Hypervariable Region 3
kb	Kilobase
KDa	Kilodaltons
km	Kilometers
km ²	Kilometers Squared
L strand	Light Strand
q	Long Arm of the Chromosome
LGT	Low Gelling Temperature
tRNA ^{Lys}	Lysine tRNA gene
MBP	Mannose Binding Protein
µg	Micrograms
µl	Microlitres
mA	MilliAmps
mg	Milligrams
ml	Millitres
mM	Millimolar
MEBOOT	Minimum Evolution Bootstrapping
mtDNA	Mitochondrial DNA
M	Molar
MEGA	Molecular Evolutionary Genetic Analysis
ng	Nanograms
nm	Nanometres
NJ	Neighbor Joining
OD	Optical Density
%	Percentage

PCR	Polymerase chain Reaction
PSAs	Population Specific Alleles
PCA	Principle Component Analysis
RFLP	Restriction Fragment Length Polymorphism
RNA	Ribonucleic Acid
rRNA	Ribosomal Ribonucleic Acid
N	Sample Size
SINEs	Short Interspersed Repeats
p	Short Arm of the Chromosome
STR	Short Tandem Repeat
NaCl	Sodium Chloride
SDS	Sodium Dodecyl Sulphate
TAS	Termination Associated Sequences
3'	Three Prime End
T	Thymine
TF	Transcription Factor
tRNA	Transfer Ribonucleic Acid
X	Times
TRIS	Tris-(hydromethyl)-amino methane
TAE	Tris Acetic Acid EDTA
TBE	Tris Borate EDTA
TE	Tris-EDTA
T20E5	20mM Tris 5mM EDTA
TEMED	N,N,N',N'-methylenebisacrylamide
U	Unit
UV	Ultraviolet
VNTR	Variable Number Tandem Repeat
V	Volts
YAP	Y <i>Alu</i> Polymorphism

CHAPTER 1: INTRODUCTION

The island of Madagascar is biologically diverse, both in its human and non-human inhabitants. Much of the history of the human colonisation of the island is unknown and the origins of the people of Madagascar, the Malagasy or *Malagache*, are uncertain. Malagasy origins have been a subject of debate over the years and we hope to use genetic studies to help resolve the ancestry of the Malagasy.

The specific aims of this study are discussed in **section 1.1**. Theories concerning the prehistory of the Malagasy proposed on the basis of other studies will be discussed in **section 1.2** while the advantage of the genetic markers used in the present study will be elaborated on in **section 1.3**.

1.1 Aims of the study

The goal of this study was to gain new insights into the genetic structure and evolutionary history of the Malagasy through the use of various types of molecular genetic markers. We hope to achieve these objectives by examining the degree of concordance of the results of haploid (mitochondrial DNA and the Y chromosome) and autosomal DNA markers (population specific alleles, including those alleles for the mannose binding protein) to investigate the relative influences of the hypothesised ancestral populations on the genetic structure of the Malagasy.

Specifically, this research addresses the following **questions**:

1. What is the extent of genetic diversity in present-day Malagasy populations?
2. What are the relationships between the different ethnic groups throughout the island?
3. Are there differences in the patterns of variation in highland versus lowland population groups?
4. To what extent have females and males contributed to the gene pool of the Malagasy?
5. What are the genetic affinities of Malagasy, African and Indonesian populations?
6. Is it possible to trace the genetic trail of migration to Madagascar?

Different genetic markers (discussed in **section 1.3**) were used to examine the level of variation in Malagasy individuals who were drawn randomly from different ethnic groups. The resulting data were then analysed using different statistical methods (discussed in **section 2.3**) to shed more light on these questions. The mtDNA and Y chromosome results were compared to results given by the autosomal markers used in order provide answers to question 3. Question 4 was addressed by comparing the results obtained from mtDNA (female contribution) and Y chromosome (male contribution) data. Altogether, these data have enabled us to refine theories concerning the peopling of Madagascar and hence address questions 5 and 6.

1.2 Madagascar

Madagascar was “officially” discovered on 18 August 1500 when Diogo Dias, a Portuguese navigator landed on its coast. He was the first European to visit the Island, but

noted that there were already people of Javanese, Hindu, African and Arab origin on the island (Uzoigwe, 1995). In 1642, Le Cardinal Richelieu ordered that Madagascar be taken over by France and in 1643 they established a trading post at Fort-Dauphin. During the late sixteenth century, the Merina Kingdom was founded, and they continued to control the island until the end of the nineteenth century. In 1960 the Malagasy Republic became an independent nation under the rule of President Philibert Tsiranana who was of Merina origin (Uzoigwe, 1995).

1.2.1 Geography and natural history of Madagascar

The island of Madagascar separated from Africa at least 70-120 million years ago, before the evolutionary origin of the monkeys and great apes (Cole, 1992). Madagascar is located in the western Indian Ocean between latitudes 12 and 26 degrees south and longitudes 43 and 47 degrees east (**Figure 1**). It is separated from the southeastern coast of Africa by the 400km wide Mozambique Channel to the west and by almost 6 400km from Indonesia to the east (Brown, 1978).

When compared with Africa, Madagascar appears to be very small. However, it is the world's fourth largest island (after Greenland, New Guinea and Borneo). Madagascar is 1 580km long and 570km wide at its widest point, and spans an area of approximately 590 000km² (Brown, 1978; Bradt, 1993).



Figure 1: Map of the world indicating the size and location of Madagascar with respect specifically to Africa and Asia.

A series of mountain ranges that run from north to south divide the island into different geographic and climatic regions (**Figure 2**): a narrow eastern section referred to by the islanders as the **lowlands** or coastal regions, and a wider western section referred to as the plateau or the high plateau (Singer *et al.*, 1957; Buettner-Janusch and Buettner-Janusch, 1964). The lowland regions include the **Eastern Central, Northern, South Western and Western** regions. The mountainous region forms the **highlands** of the central part of the island (Bradt, 1993).

1.2.2 The peoples of Madagascar

The Malagasy have ancestors from all the shores of the Indian Ocean (apart from Australia) and have thus been described as the ‘**progeny of the Indian Ocean**’ (Dewar and Wright, 1993). Although no precise historical record can be reconstructed of the manner in which Madagascar was settled, nor the order in which each ancestral population group arrived, there are sufficient data available around which several theories have been proposed.

The present population of Madagascar is estimated to be approximately 11 million (Brown, 1978). The Malagasy government recognises 18 loosely defined ethnic groups or *fuka* on the island. The criteria for the ethnic, tribal or cultural differentiation of the Malagasy have been mainly geographic or political and may have no biological basis. However, they have been used to identify Malagasy kingdoms (or ethnic groups) over the years (Brown, 1978; Bradt, 1993).



Figure2 : Map of Madagascar indicating the 5 geographic regions (Central highland, Eastern Central, Northern, South Western and Western) and the location of the 19 Malagasy ethnic groups.

Using geographic criteria, the Malagasy ethnic groups may be subdivided into the highland peoples (those living on the plateau) and the lowland or coastal peoples (Brown, 1978) and these are distributed on the island as follows (**figure 2**):

- **Highland groups:** Betsileo, Bezanozano, Merina, Sihanaka and Tanala
- **Lowland and coastal groups:** Antaifasy, Antaimoro, Antaisaka, Antambahoaka, Antandroy, Antankarana, Antanosy, Bara, Betsimisiraka, Mahafaly, Masikoro, Mikea, Sakalava and Tsimihety.

There are also two “unofficial” ethnic groups on the island namely the Vakinankaratra (who are of Merina and Betsileo descent) and the Zafisoro (Singer *et al.*, 1957; Brown, 1978; Bradt, 1993).

The relative sizes of the various ethnic groups found on Madagascar vary tremendously (Verin *et al.*, 1970) but a number of the larger groups are generally divided into three castes (Dewar, 1992): the nobles (caste I), the commoners or freeman (caste II), and the former slaves and their descendants (caste III). Tribes living in the lowland or peripheral areas are collectively referred to as *côtiers*; the Betsimisiraka, Tsimihety and Antandroy are the largest of this group (Mack, 1986).

These social distinctions are no longer very strictly adhered to, but they do form the basis of other social and economic distinction. This population structure is seen as a possible problem for sampling the Malagasy population according to ethnic groups, and differentiating racial traits (Dewar, 1995) as these distinctions may be purely geographic rather than biological.

1.2.3 The ancestry of the Malagasy

Europeans have attempted to elucidate the ancestry of the Malagasy since the sixteenth century AD (Mourant *et al.*, 1976), but have been hindered by the absence of written records. The main sources of historical information about Malagasy history are the verbal traditions carried down by the island's inhabitants (especially those of the Merina) from rulers since 1 000AD (Mourant *et al.*, 1976). Arab colonies were founded in the northwest of the island in the thirteenth century (Brown, 1978) and they contributed the ancient Malagasy texts (called the *Sorabe*) written in Arabic scripts which document the history of an alleged Arabian ancestor. Other texts include the nineteenth century writings of a Jesuit priest, Père Callet and the Malagasy writings that date to the middle of the nineteenth century when the European missionaries taught writing (Singer *et al.*, 1957). Also noteworthy is the "History, Annals and Journal" of the Malagasy historian, Raombana (Mack, 1986).

A tradition amongst the Antaimoro states that they descended from 30 men who sailed directly to Madagascar from Mecca during the fourteenth century (Campbell, 1996). However, there is no evidence to suggest direct migrations to Madagascar from either Arabia or India. Island traditions also speak of an early race of dwarfs. Again, there is no evidence in Madagascar to support this (Campbell, 1996). Grandidier (1908) suggested that the Malagasy might be descendants of Indonesian settlers and a few other minor groups while Ferrand (1908) and Verin (1986) favored an African ancestry from Swahili and Indonesian traders.

Campbell (1996) reviewed the various hypotheses regarding the origins of the Malagasy. The first hypothesis proposed that the population of Madagascar was autochthonous or “native”. This suggestion was only seriously entertained until 1880 when studies showed that the earliest human inhabitants arrived to the island in the early part of the Christian era. The second hypothesis suggested that the original colonisation of Madagascar was by peoples of African origin who were later conquered by Indonesian immigrants. This may be feasible if one considers the proximity of Africa to Madagascar (illustrated in **Figure 1**), but lacks support, however, because there is no evidence of long-distance maritime techniques or tradition of voyages amongst the East African peoples (Brown, 1978). It was then proposed that the initial settlers of Madagascar were either of Malayo-Indonesian-Polynesian origin, with few migrations of Indian and Semitic peoples and a minor contribution from Africa (third hypothesis), or that the various migrations to Madagascar took place via Africa and comprised mixed groups of Indonesians and Africans (fourth hypothesis). Historical evidence of far ranging Indonesian voyages into the Indian Ocean lends support to both of these theories (Ferrand, 1908; Grandidier, 1908) but it remains unclear how the proto-Malagasy reached Madagascar.

Grandidier (1908) suggested that the proto-Malagasy traveled in outrigger canoes directly across the southern Indian Ocean from Java to the eastcoast of Madagascar as early as the sixth century BC (Campbell, 1996). It was suggested that the Malay-like highland peoples arose primarily from Indonesian peoples and that the Negroid-like features of the lowlanders can be attributed to an ancient direct migration by Negroids, as well as being supplemented by 650 000 slaves from Africa who were liberated in 1877 and 1896 (Campbell, 1989).

This theory implies a migration of a substantial number of individuals across the Indian Ocean. By sailing a distance of 4 300 nautical miles across the Pacific Ocean in a balsa log raft over a period of three months (Heyerdahl, 1973), Thor Heyerdahl proved that given the wind and ocean currents, such a journey may have been possible. However, taking into account the suggested time period together with the enormous distance and the hazardous ocean and weather conditions, it seems unlikely.

Ferrand (1908) suggested that the peoples of Madagascar originated from both African and Indonesian populations. The Indonesian people are thought to have followed trade routes along the northern and western shores of the Indian Ocean where they set up trading posts, establishing a proto-Malagasy colony (reviewed by Campbell, 1996). These Indonesian settlements in Africa may have been gradually displaced and forced on to the island in successive waves, first by the Bantu expansion (Brown, 1978), and then from northern East Africa around the ninth century by Muslim Arabs and Swahili (Bantu) traders (Brown, 1978). The first settlers are thought to have been more Indonesian-like and subsequent waves more African-like. It is further suggested that the initial colonists on Madagascar were forced further and further inland by the subsequent waves. The result was that the first Indonesian-like immigrants settled on the high plateau and the more African-like peoples inhabited the lowland (Campbell, 1996).

This hypothesis seems more likely. Not only was the trip made up of shorter land-hugging trips, but it may explain how Asian influences reached East Africa and how Africans reached Madagascar. Furthermore, the Malayo-Polynesians are known to have been active maritime traders between the first and fifth centuries AD, and Arab and Chinese

accounts indicate that Indonesians were trading and living in East Africa and Madagascar during this period (Campbell, 1996).

Although the origin of the Malagasy is only partially resolved, two conclusions can be drawn with reasonable confidence: the first inhabitants of the island originally came from Indonesia, and the first permanent occupation of Madagascar did not take place until some time after the Christian era (Brown, 1978). Colonisation and settlement of the island is thought to have taken place in multiple waves, probably originating in the South East Asian archipelago in the early part of the first millennium AD (Campbell, 1989). The Malagasy were initially exporters of slaves but in the nineteenth century, a market for imported slaves developed (Deschamps, 1965; Campbell, 1989). It is thought that this continuous movement of people to and from the island may have substantially influenced the present day Malagasy gene pool (Campbell, 1989).

1.2.3.1 Theories concerning Malagasy origins

1.2.3.1.1 Evidence based on physical appearance

Since both Africans and Indonesians are thought to have contributed to the gene pool of present-day Malagasy, it is not surprising that “racial” differences are apparent among the Malagasy. Some populations appear “Malay” in stature, skin colour and hair texture whereas other groups, in contrast, appear to be “Negroid” and are taller, darker and have curly hair. In 1550, Nagnort stated, “There are two types of inhabitants; the ones black

with frizzy hair who are the original people of the country; the other white with long hair who came from the coast of Persia 500 years ago.” In 1658, Flacourt reported that the inhabitants in the southeast of the island were divided into two sorts: “les Blancs” and “les Noirs”. Singer *et al* (1957) found that the Malagasy physical characteristics were consistent with a hypothesis that the Malagasy are composed of about two-thirds African influences and one-third Indonesian influences.

The present-day inhabitants of Madagascar are very heterogeneous in physical appearance with features suggesting varying degrees of admixture between African and Asian populations (Singer *et al.*, 1957). According to Singer *et al.*, (1957) the highland people generally appear to have lighter skins, straight hair and resemble Malaysians, Indonesians or Indo-Chinese. It has been suggested that the Merina (with their light skin and straight hair) are clearly different from the other groups, though with a strong “mixture” of African physical traits (Singer *et al.*, 1957). On the other hand, in the coastal areas, especially in the west and south, the people have darker skins, short crinkly hair and Negroid features (Singer *et al.*, 1957).

1.2.3.1.2 Linguistic evidence

Despite the apparent variation in the physical characteristics of the Malagasy, almost the entire population speaks different dialects of the same language, “**Malagasy**”. Malagasy is a Malayo-Polynesian language belonging to the **Austronesian** (from the Latin root *austro* meaning “southern,” and the Greek word *nesis* meaning “island”) family of

languages. This language family includes between 600 and 1 000 languages including languages of Malaysia, Indonesia, the Pacific islands (Melanesia, Micronesia and Polynesia), the Philippines, southern Vietnam, Cambodia, Taiwan and Madagascar (Grimes, 1992).

Similarities between Malagasy and languages spoken in Indonesia have been recognised since 1603 (Verin *et al.*, 1970). However, the identity of these Indonesian languages has not yet been discovered. Dahl (1951, 1977) suggested that the surviving languages most closely related to Malagasy was Maanyaan and Ngaju spoken in the Barito River Valley in the basin of the river Pati situated 90 miles north of Bandjarmasin in southeast Borneo. Dahl (1951) proposed that proto-Malagasy speakers diverged or separated from people speaking Indonesian languages around 400AD on the basis of the limited number of Sanskrit loan words in Malagasy.

By means of lexicostatistics and glottochronology, Verin *et al.* (1970) have proposed that the different dialects of Malagasy have been diverging for as much as 2000 years. Based on the observation that a certain percentage of words in a language would be dropped in a given number of years, they concluded that a single group of people gave rise to the Malagasy between 100BC and 225AD. They went on to suggest that soon after the colonisation of Madagascar, the ancestral group split into three groups: one group giving rise to the Tsimihety, one to the Antankarana and the third to the rest of the dialects. The proto-Tsimihety and proto-Antankarana are thought to have settled in the north of Madagascar, and members of the third group eventually split, one division giving rise to the dialects of the east and centre, and the other to the dialects of the south and west (Verin

et al., 1970; Dewar, 1995). In 1998 Greunier challenged this scenario. He pointed out that the island's speech patterns reflect various clines which may not necessarily follow the same geographical ones. In addition, he suggested that the similarity of neighbouring groups may be a result of either common ancestry or due to the convergence of groups of diverse ancestry over time (Greunier, 1998).

Despite there being dialectical differences between the different ethnic groups of Madagascar, the dialects are reasonably mutually intelligible (Brown, 1978). This suggests that the original settlers were a relatively unified group originating in south east Asia, regardless of where they were living immediately before they colonised Madagascar (Beuttner-Janusch and Beuttner-Janusch, 1964).

It is important to mention that there are minor exceptions to the uniformity of language spoken throughout the island. In 1613, Luis Mariano described two languages in Madagascar: an African language spoken along the northwest coast and a language similar to Malay on other coasts and in the interior (Mariano, 1904). In addition, in the eighteenth century, Robert Drury (1729) reported non-Malagasy languages along the west coast. A further complication is the influence of Arabic, French, English and, in particular, Bantu languages on Malagasy. Today, there remain a few bilingual communities along the west coast where Makoa, a Bantu language of Mozambique, is spoken by families of recent descent. In both the northwest and the island of Nosy-be, there are bilingual Swahili speakers. It is not yet known when or how these influences occurred (Dewar, 1995).

1.2.3.1.3 Malagasy culture

Malagasy culture shows even more regional variation than their language. This may be as a result of differing geographical conditions as well as the presence of certain ethnic influences in some areas.

Much of the way of living of the Malagasy is that of the Eastern archipelagos, mainly Indonesia. These include slash and burn rice cultivation, utensils, basket weaving, iron smelting (using the double tuyere bellows), clothing, musical instruments (notably the tube zither, the conch and the cone drum), totem poles, beliefs, and superstitions (Singer *et al.*, 1957; Mack, 1986). Further indicators of Indonesian ancestry include the typical Indonesian tombs and rectangular wooden huts of the highlands which are built on stilts and face specific directions because of the attached symbolic significance (Brown, 1978; Dewar and Wright, 1993). The Arab influence (tombs, forts, mosques and astrology) seen on the eastcoast of the island appears to be pre-Islamic (Mack, 1986). However, this appears to be a relatively recent influence (Brown, 1978).

The presence of great herds of hump-backed zebu cattle provides evidence of an African culture on Madagascar. Cattle raising, which is the chief activity of the pastoral people (Sakalava, Bara and Antandroy) is also of mainland African origin (Brown, 1978). In addition, fibre weaving, some musical instruments and certain funeral customs are also similar to those in Africa (Brown, 1978).

In some cases, it is difficult to conclusively identify the sources of the different elements in the culture since the patterns of material culture do not always correlate with other cultural markers. An example is the outrigger canoe which is undoubtedly of Indonesian origin but is predominantly found along the west coast where the African influence is said to be the strongest. This may indicate that the Malagasy living in different parts of the island adapted to their respective environments by selecting different available traits (Brown, 1978).

1.2.3.1.4 Archaeological evidence

Although Madagascar has fascinated archaeologists for a long time, very few archaeological studies have been carried out there until fairly recently. Unreliable and expensive dating techniques, together with Malagasy cultural resistance to the opening of tombs, have hampered archaeological progress on the island.

No Stone Age archaeological sites have been found on the island, and even the earliest known settlers were apparently reliant only on iron tools. The only stone tools found on the island are net sinkers, weights and gunflints (Dewar, 1988). This suggests that human habitation on the island occurred relatively recently.

The first traces of human activity in Madagascar appear early in the first millennium AD along the southwest coast at two coastal marshes (Dewar and Wright, 1993). The evidence is linked to extinct pygmy hippopotamus bones that clearly bear the marks of

butchery with metal tools (Dewar, 1995). These bones have since been given radiocarbon dates in the first to fourth centuries AD (MacPhee and Burney, 1991). In 1986 a rock shelter (indicating permanent occupation on the island) as well as pottery, iron blades, stone weights and animal bones were discovered in the extreme north of the island. The earliest occupation date estimated is about 700AD (Dewar and Wright, 1993).

Tombs have been found containing Persian glassware, bead necklaces, bronze mirrors, Middle Eastern ceramics and fragments of late Song Dynasty Chinese ceramics along the north east coast as far back as the eleventh century and suggest the presence of Arab settlements in Madagascar (Cole, 1992; Dewar and Wright, 1993). Trading sites along the east African coast from Mozambique to the Persian Gulf are known to have been in existence since the seventh century AD. These sites have yielded tradeware (iron smelting evidence and jewels of silver, gold, carnelian and glass beads) indicating that the sites formed part of a trade network. Food vessels carved of soapstone-like rock are commonly found along such sites and date to the tenth century AD. Indications of cattle herding settlements and agricultural villages have been found in the south and southeastern parts of Madagascar and date to the eleventh century (Dewar and Wright, 1993). In addition, tombs found along the westcoast of the island contain traditional bird-carvings which are similar to soapstone carvings found in Zimbabwe (Brown, 1978) indicating an African heritage.

1.2.3.1.5 Genetic studies

Early studies making use of blood groupings and the sickle cell trait confirm the presence of a significant African contribution to the founding Malagasy population (Singer *et al.*, 1957; Beuttner-Janusch and Beuttner-Janusch, 1964; Beuttner-Janusch *et al.*, 1973).

Rhesus blood group data clearly indicates the presence of a typical African marker (the cDe chromosome) in the Malagasy suggesting that the African contribution is greater than the Indonesian contribution. Using these data Singer *et al.*, (1957) estimated that the Malagasy gene pool was derived from one third Indonesian (33-38%) and two thirds African (62-67%) contribution (Singer *et al.*, 1957).

Singer *et al.* (1957) were the first to suggest that the sickle cell gene was brought to Madagascar from Africa (and not from India or other parts of the East) at the same time as slaves and zebu cattle were introduced to the island. The sickle cell trait was found to occur at a higher frequency in the lowland groups than in the highland groups (Beuttner-Janusch and Beuttner-Janusch, 1964). Although this could be interpreted as indicating a larger proportion of African admixture, at least two other explanations are possible. The first is that since the sickle cell trait provides a selective advantage in the heterozygote state against malaria; the higher frequency in the lowland area could be the effect of natural selection. Second, the highland groups (especially the Merina) may have maintained strict rules against marriages with the coastal tribes and slaves from Africa indicating that they may have fewer affinities with African populations. In a recent study of haplotypes associated with the sickle cell mutation, the “Bantu” haplotype, noted only

in Bantu-speaking Africa, was found on 25 of 28 chromosomes studied (Hewitt *et al.*, 1996). This study has provided the first definitive evidence that the major African component of the Malagasy ancestry is derived from Bantu-speaking people.

By making use of mitochondrial DNA (mtDNA), Soodyall *et al.* (1995) found the 9-bp deletion marker to be present in 75 out of 280 (26.8%) Malagasy individuals. By studying control region sequence variation in this sample, it was estimated that 70.7% of the Malagasy individuals with the 9-bp deletion were derived from an Asian source, while 29.3% were derived from an African source. Of those with the Asian form of the deletion, 96% were found to have the “Polynesian motif” (see **section 1.3.1.1**).

Dangerfield (1997) made use of the variation at seven restriction fragment length polymorphisms (RFLPs) in the glucose-6-phosphate dehydrogenase (G6PD) gene to study Malagasy ancestry. The G6PD A⁻ variant, is associated with partial enzyme deficiency and is also thought to offer protection against malaria. It is a useful marker to use since it is almost exclusive to Africa. The A⁻ variant is associated with a particular seven-site RFLP haplotype and this A⁻ haplotype was found at high frequencies in G6PD-deficient Malagasy (Dangerfield, 1997). Haplotypes associated with G6PD deficiency in the Mediterranean region have also been detected in the Malagasy suggesting a possible Middle Eastern contribution.

Morar (2000) used thirteen loci (eight of which were autosomal and five of which were on the Y chromosome) to examine genetic variation in eleven of the Malagasy ethnic groups. In this study, it was reaffirmed that both African and South East Asian populations have

contributed significantly to the Malagasy gene pool. An African contribution of 50% was found with autosomal data and an African contribution of at least 60% was found with Y chromosome data. In addition, a higher African contribution to the lowland populations was evident. Specifically, with the CD4 locus, a significant African contribution to the people of Madagascar was seen. However, most of the data showed that all the Malagasy ethnic groups are similar to each other. Lastly, the colonisation of Madagascar was estimated to have occurred within the last 3000 years, and it was suggested that the proto-Malagasy were probably already admixed before arriving to the island.

It is our hope that the current study will add to this growing body of knowledge and thus enable us to further test and refine existing theories concerning the origins of the Malagasy.

1.3 Genetic markers used in this study

DNA markers have been proven to be very powerful for distinguishing between individuals (Schriver *et al.*, 1997). It is well documented that the majority of the genetic variation among humans is inter individual and only 10% - 20% of the total variation is due to population differences (Cavalli-Sforza *et al.*, 1994; Deka *et al.*, 1995). Most populations share alleles at any given locus and the alleles that are most frequent in one geographically or ethnically defined population are also found commonly in other populations (Shriver *et al.*, 1997). This high degree of genetic similarity reflects the recent divergence of the human species into continental groups – there are marked physical and

physiological differences among human racial groups which reflect long-term adaptation to unique ecological conditions, random genetic drift and sexual selection (Shriver *et al.*, 1997).

To shed more light on the questions asked in **section 1.1**, a molecular genetic approach incorporating the use of both haploid and diploid genetic markers was adopted. The advantages and disadvantages of each of these markers for the purposes outlined earlier (**section 1.1**) are explored in the next section.

1.3.1 Haploid DNA markers

1.3.1.1 MtDNA

MtDNA has several unique properties that make it a popular tool for population and evolutionary studies. Consequently, it has been used widely in the study of human evolution and in reconstructing population history.

Mitochondria are double-membraned cytoplasmic organelles that are found only in eukaryotes and have a strict maternal inheritance (Giles *et al.*, 1980). Each human mitochondrion is approximately 0.5–1.0 μ m by 5–10 μ m in size (Spuhler, 1988) and contains several mtDNA molecules that are independent of the nuclear genome and are self-replicating (Wallace, 1995). MtDNA accumulates mutations at a rate of 5–10 times faster than nuclear DNA (Wilson *et al.*, 1985). This together with the absence of

recombination (Olivo *et al.*, 1983) are extremely valuable properties because any new variation can be assumed to be due to mutation only. In addition, the fact that mtDNA is maternally inherited makes it possible to study the maternal history of a species (Stoneking, 1993). Other features making mtDNA a useful marker are its small size and location and high copy number within a cytoplasmic organelle. These features make it easily accessible, even when small amounts of DNA are present (Strachan and Read, 1996).

Each mtDNA is a closed circular molecule consisting of 16 596 base pairs (bp) (**Figure 3**). Each molecule consists of two complementary single DNA strands that differ in their guanine and thymine (G + T) content and can be separated into the heavy (H) and light (L) strands by alkaline caesium chloride gradients (Brown *et al.*, 1980). The genetic information contained in human mtDNA is organised in a highly economical fashion. Unlike the nuclear genome, approximately 90% of the mitochondrial genome has a coding function with no introns. The complete mtDNA nucleotide sequence of one individual was first characterised by Anderson *et al.* (1981) and is referred to as the reference sequence. All new mutations can be mapped to a specific position on the reference sequence making it possible to determine the exact nature of a mutation (and its consequence) and to examine evolutionary history (Cann *et al.*, 1984).

One section of the mtDNA molecule (**Figure 3**) contains almost all (~90%) of the noncoding DNA within the mitochondrial genome and is referred to as the **control region** (Anderson *et al.*, 1981). This is the most rapidly evolving and polymorphic region of the mtDNA genome and is thus useful when used for sequence analysis to distinguish closely

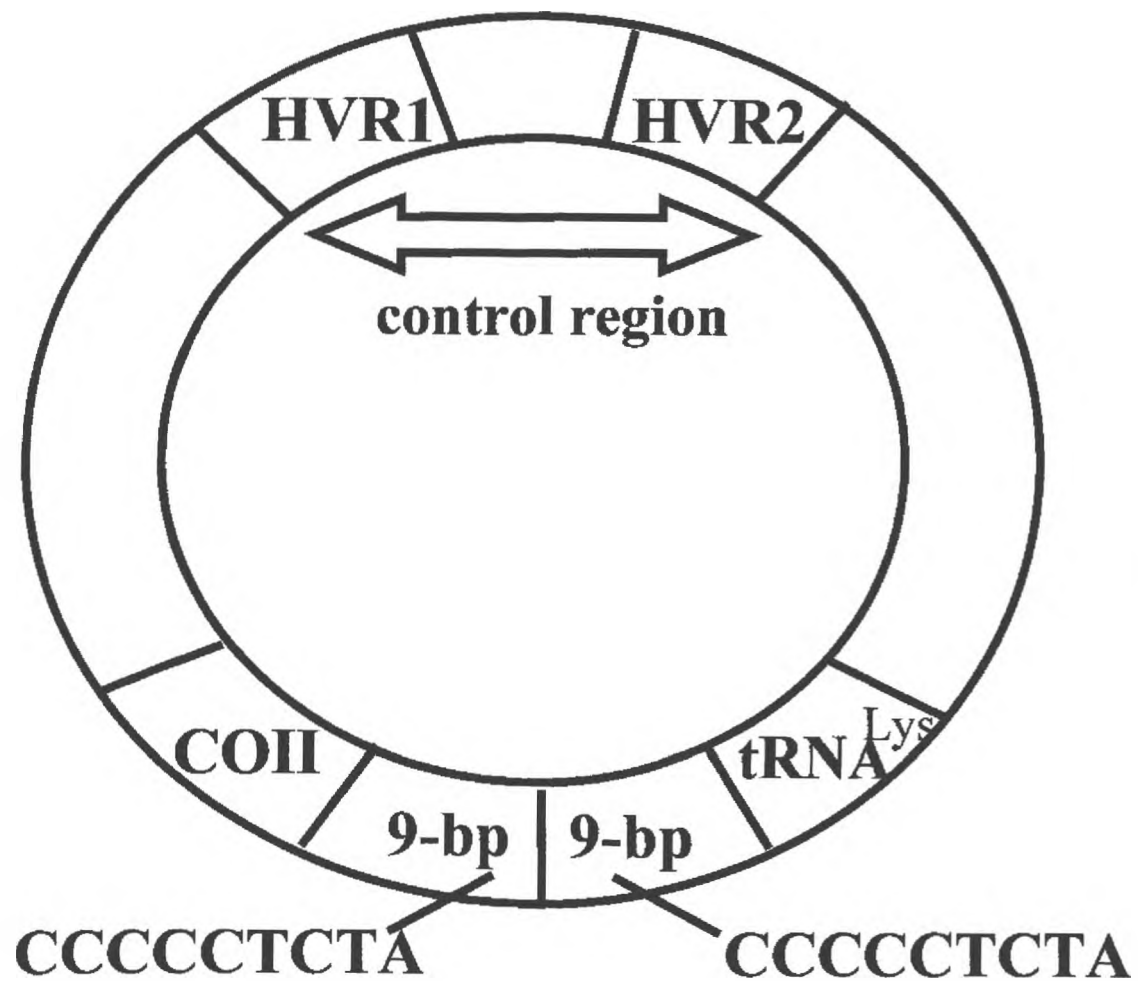


Figure3 : A diagrammatic representation of a Mitochondrial DNA molecule, showing the position of the control region and the 9-bp deletion.

related mtDNAs (Vigilant *et al.*, 1991). The control region consists of 1 112-bp and is believed to be involved in functions associated with transcription and replication (Walberg *et al.*, 1981; Chang and Clayton, 1984; Hixson and Clayton, 1985). Within the control region are three hypervariable (HV) or control regions (CR) (Anderson *et al.*, 1981):

- **Hypervariable region 1 (HVR1 or CR1)** between positions 16 024 – 16 365 according to Anderson *et al.* (1981).
- **Hypervariable region 2 (HVR2 or CR2)** between positions 73 – 340 according to Greenberg *et al.* (1983) and Wilson *et al.* (1993).
- **Hypervariable region 3 (HVRIII)** located between positions 438 – 574 (Lutz *et al.*, 1997).

Most studies to date have focused on variation in HVR1 and HVR2 to elucidate population prehistory and human origins.

Another region on the molecule that has been examined widely is the small intergenic region between the cytochrome oxidase subunit II gene and the lysine tRNA gene (COII/tRNA^{Lys}) (**Figure 3**). Most humans have two copies of a 9-bp tandem repeat (CCCCCTCTA). However, in some humans, one copy of the repeat is deleted, probably as a result of DNA slippage (Cann and Wilson, 1984). This is referred to as the ‘**9-bp deletion**’ (Hertzberg *et al.*, 1989).

Originally, phylogenetic analysis based on high resolution mapping suggested that the deletion had occurred only once in human evolution because it was restricted to mtDNA types from peoples of East Asian origin and was not found in Caucasian or African

populations (Wrischnik *et al.*, 1987). The deletion was found in nearly every Asian population examined including populations from Asia, Polynesia, and the New World (Horai *et al.*, 1987; Stoneking and Wilson, 1989; Torroni *et al.*, 1992; Passarino *et al.*, 1993; Redd *et al.*, 1995) and was thus referred to as an “Asian-specific” marker. The 9-bp deletion has subsequently been found at varying frequencies throughout central and southern Africa (Vigilant, 1990; Soodyall *et al.*, 1996). The deletion was found to be rare in east-central Africa and was absent from western African populations (Soodyall *et al.*, 1996). It has also been found sporadically in some European (Torroni *et al.*, 1996; Alves-Silva *et al.*, 1998) and Australian populations (Hertzberg *et al.*, 1989).

Phylogenetic analysis of control-region sequence data indicated that African and Asian mtDNAs with the deletion were not closely related, thus suggesting separate origins for the deletion in Africans and Asians (Vigilant, 1990). This was further supported by Redd *et al.* (1995) and Soodyall *et al.* (1996). The different control region sequence profiles associated with the presence of the 9-bp deletion are referred to as the “**African form**” and the “**Asian form**” of the deletion (Soodyall *et al.*, 1996).

Moreover, it has been shown that the deletion arose at least three times in Africa. The majority of mtDNAs associated with the **African form** of the deletion have a similar profile that was referred to by Soodyall *et al.* (1996) as the **major African form** of the deletion. These authors suggested that the major form of the deletion in Africa could have originated in central Africa and spread to other parts of Africa as a consequence of the Bantu expansion.

Previous studies of control region variation in ancient Polynesians have shown that the deletion is frequently associated with three substitutions (Hagelberg *et al.*, 1994; Lum *et al.*, 1994): 16 217 (T→C), 16 247 (A→G), and 16 261 (C→T) relative to the published reference sequence (Anderson *et al.*, 1981). These mutations have been shown to have evolved in the order 16247, 16217 and 16261 (Melton *et al.*, 1995). The presence of the 9-bp deletion together with all three substitutions is referred to as the ‘**Polynesian motif**’, but is found commonly throughout southeast Asia and the Pacific (Melton *et al.*, 1995; Redd *et al.*, 1995).

Since the 9-bp deletion in conjunction with the mtDNA control region sequence data is useful in distinguishing between African and Indonesian mtDNA types, mtDNA studies on the Malagasy ought to be extremely valuable in elucidating the African and Indonesian mtDNA types in the present-day population.

It is necessary to emphasise that mtDNA is a haploid marker and thus represents a single genetic locus. Thus, the information obtained from such studies reflects only the maternal history of the species. Therefore, results from work done on mtDNA should be interpreted cautiously and should be supplemented with data from other markers (Poloni *et al.*, 1995; Stoneking and Soodyall, 1996).

1.3.1.2 The Y Chromosome

The human Y chromosome is inherited solely by males in the hemizygous state and, with the exception of genes in the pseudo-autosomal region, it does not undergo homologous recombination at meiosis. Thus, every chromosome preserves a unique record of the mutational events that have occurred in its ancestors and each male individual has the same Y chromosome profile as his father, brothers, paternal grandfather, paternal uncles, etc. (Malaspina *et al.*, 1990; Santos *et al.*, 1996).

Significant variability has been found at certain STR loci on the human Y chromosome, including both a minisatellite and several trinucleotide and tetranucleotide repeat microsatellites (Cooper *et al.*, 1996; Deka, 1996). In addition, many single nucleotide polymorphisms (SNPs) as well as a polymorphism of the insertion/deletion of member of the *Alu* family of repeated DNA elements have been identified. The 455-bp *Alu* element is contained in a 2.6-kb fragment on the long arm of the human Y chromosome, Yq11 and is referred to as the Y *Alu* polymorphic (YAP) element. It is present at a specific site on the Y chromosome in some humans and is absent in others (Hammer, 1995).

The YAP element is a useful marker as it has a single origin in human evolution, and its frequency varies among human populations (Hammer, 1994). It is thought that the absence of the YAP element is likely to be the ancestral state and that the insertion most likely occurred after the divergence of humans from modern apes (Jakubicza *et al.*, 1989; and Hammer, 1994) and is believed to be the result of a unique mutational event (Spurdle *et al.*, 1994). The frequency of the YAP insertion in sub-Saharan Africa is approximately

ten-fold higher than in Europe, Oceania (India, China etc.) and most regions of the Far East. It is widespread in Africa with the highest frequency in South African Bantu-speakers and West Africans (78-86%) (Hammer, 1994). In Europeans, the frequency is much lower (4-11%), and it is absent in most Asian and Oceanic populations (Hammer, 1994).

If a sample is shown to have the *Alu* insertion (YAP+), there is a high possibility that it is of African origin. However, this is not the case for YAP- samples. These samples may be further resolved with the use of the M9 system. The M9 polymorphism is associated with a C → G substitution. The “G” allele is found in all geographic regions except Africa (Underhill *et al.*, 1997). Its absence in Africa and extensive distribution elsewhere suggests that it occurred initially outside of Africa and then dispersed widely, possibly as a result of genetic drift during an early modern human bottleneck (Underhill *et al.*, 1997). The “C” allele is ancestral and is thought to be of African origin, but has also been noted at low levels in other populations such as Europeans (Underhill *et al.*, 1997).

Individuals who are found to be YAP+ have a high probability of being of African origin. Those who are found to be YAP- and also have the M9 C allele, are also of African ancestry. Those who are found to have the M9 G allele remain unresolved as these are known only to be “non-African”. Resolution with additional markers would be necessary to resolve these further. Thus, the YAP in conjunction with M9 is useful in distinguishing between African and non-African Y chromosomes. Therefore, such studies on the Malagasy would be valuable in elucidating the male African and non-African contribution to the present-day Malagasy.

Male lineages are likely to show different patterns of evolution to those provided by mtDNA analyses because of the potentially greater variance in male reproductive success, hitchhiking associated with selective sweeps and, most importantly, sexual differences in patterns of migration and population mixing (Cooper *et al.*, 1996; Deka, 1996; Mitchell and Hammer, 1996). As a result, markers situated on the non-recombining region of the Y chromosome could provide a view of evolution which is complementary to that provided by mtDNA (Jobling and Tyler-Smith, 1995; Cooper *et al.*, 1996).

There are a few possible disadvantages of using the Y chromosome as a tool to study evolutionary relationships. The Y chromosome gene pool has a smaller effective size than that of other chromosomes because there is a lower proportion of Y chromosomes in any given human population (roughly a 1:3 ratio of Y chromosomes to autosomes). In addition, the effective size of the Y gene pool is influenced by male cultural and behavioural patterns such as polygyny; the levirate (whereby widows marry their brothers-in-law); male migrations; as well as wars and invasions (Mitchell and Hammer, 1996). Furthermore, since male lineages are not shuffled by recombination, a selected allele may 'drag' with it all the rest of the non-recombining part of the Y chromosome in potentially gigantic selective sweeps, thereby homogenising the gene pool within a population (Pääbo, 1995).

1.3.2 Autosomal diploid markers

1.3.2.1 Population specific alleles

Neel (1973) was the first to suggest that some genetic markers are polymorphic in one population and not in others. He referred to these markers as “private” and used them to estimate mutation rates. Reed (1973) used the term “ideal” to describe hypothetical genetic marker loci at which different alleles are fixed in different populations. Chakraborty *et al* (1991) used the term “unique alleles” to describe those variants that are found only in one population. Shriver *et al.* (1997) used the designation “population specific alleles” (PSAs) to describe genetic markers with large allele frequency differentials between populations.

We have chosen a panel of three dimorphic genetic markers/PSAs that have been typed in ethnically defined populations and that have been reported to have large allele-frequency differentials between the hypothesised Malagasy parental populations (**Table 1**). The markers that were used in this study are as follows:

- **Sb19.3** – This marker is an *Alu* insertion polymorphism on chromosome 19 (Batzer *et al.*, 1996). The insertion is commonly found in Africa and appears to be absent from Asian populations (Batzer *et al.*, 1996, Shriver *et al.*, 1997).
- **AT3id** – This marker corresponds to a length polymorphism of the 5' untranslated region of exon 1 of the antithrombin III (ATIII) gene (Liu *et al.*, 1995). The ATIII gene is approximately 19-kb long and has been localised to chromosome 1q22-25, spreading over seven exons. It is a member of the serine proteinase inhibitor family

and functions irreversibly inactivating several coagulation proteinases, mainly in the liver (Thaler and Lechner, 1981). The “1 allele” is present in both Africa and Asia but is found at significantly higher frequencies in African populations (Liu *et al.*, 1995; Batzer *et al.*, 1996).

- **APO 4** - This marker is an *Alu* insertion polymorphism on chromosome 11 (Batzer *et al.*, 1996). This *Alu* insertion is commonly found in Africa and appears to be absent from Asian populations (Batzer *et al.*, 1996, Shriver *et al.*, 1997).

PSAs are the ideal markers for accurate population and individual admixture estimation in populations that are expected to be of African or Asian ancestry. They will be useful in this study because they will enable us to compare the frequencies of these markers in the Malagasy gene pool to those found in various populations throughout the world (specifically African and Asian populations), and thereby gain some insight into the origins of the Malagasy.

Table 1: Table showing reported frequencies of the “1” allele from various populations for each PSA system used in this study, together with the respective reference.

POPULATION	Sb19.3	AT3id	APO	REFERENCE
CHINESE	-	0.627	-	Liu <i>et al.</i> , 1995
INDIANS	-	0.346	-	Liu <i>et al.</i> , 1995
MALAYS	-	0.461	-	Liu <i>et al.</i> , 1995
CAUCASIANS	-	0.250	-	Bock <i>et al.</i> , 1983 ¹
EUROPEANS	0.910	0.279	0.927	Batzer <i>et al.</i> , 1996
NIGERIA	0.455-0.457	0.875-0.889	0.409-0.480	Batzer <i>et al.</i> , 1996
CENTRAL AFRICAN REPUBLIC (CAR)	0.364	0.859	0.435	Batzer <i>et al.</i> , 1996
PYGMIES	-	-	0.794	Batzer <i>et al.</i> , 1996

¹cited in Batzer *et al.*, 1996

1.3.2.2 Mannose binding protein

Human mannose binding protein (MBP) is a calcium dependent lectin that plays an important role in innate immunity (Lipscombe *et al.*, 1996; Summerfield *et al.*, 1997). MBP is a protein 600Kda in size and is composed of multimers of 32Kda subunits which are synthesised by hepatocytes and secreted into serum (Arai *et al.*, 1993). The MBP gene is located on the long arm of chromosome 10 at 10q11.2-q21, and encodes four exons: a cysteine-rich region (exon 1), a glycine-rich/collagenous region (exons 1 and 2), a “neck” region (exon 3), and a carbohydrate-binding domain (exon 4) (Lipscombe *et al.*, 1992).

Since MBP plays an important role in host defence, low serum levels are associated with severe infection, diarrhoea and failure to thrive in infants (Garred *et al.*, 1992; Lipscombe *et al.*, 1992; Arai *et al.*, 1993). The molecular basis for low serum concentrations of the protein is the result of three independent nucleotide substitutions in exon 1 of the MBP gene (Lipscombe *et al.*, 1992; Arai *et al.*, 1993). According to Lipscombe *et al.* (1996) these are:

- the **B allele** which is due to a point mutation (GGC → GAC) at **codon 54** and results in a change of glycine to aspartic acid;
- the **C allele** which is due to a point mutation (GGA → GAA) in **codon 57** and results in a change of glycine to glutamic acid;
- and a third rare allele known as the **D allele** which results in a change of arginine to cystine at **codon 52**.

The wild type (‘normal’) allele is referred to as **allele A** (Lipscombe *et al.*, 1996).

The B variant is distributed widely across Europe and high frequencies are found in Caucasians, Asians, and Eskimos (Garred *et al.*, 1992 and Lipscombe *et al.*, 1996). The frequency of the B allele was found to range from 0.11 to 0.16 in non-African populations (Garred *et al.*, 1992; Lipscombe *et al.*, 1996). In contrast, Africans (specifically the Sekele San) were found to have either very low frequencies (0.009) of the B allele (Garred *et al.*, 1992), or it was found to be absent altogether (Lipscombe *et al.*, 1996). The C variant is found exclusively in Africa with high frequencies (0.27) in Negroids (Gambians, Kenyans, and Xhosa). The D variant is rare but is found in low frequencies in Africans, Asians, and Europeans (Lipscombe *et al.*, 1996). Please see **Table 2**.

These mutations result in low levels of the protein and an opsonic deficiency, but the frequencies also suggest that some selective advantage for low MBP levels may exist (Garred *et al.*, 1992; Lipscombe *et al.*, 1996). Garred *et al.* (1992) hypothesised that MBP deficiency is maintained in populations because MBP deficiency decreases the infectivity of some intracellular microorganisms which are dependent on opsonisation.

By comparing the frequencies of the B and C variants in the Malagasy gene pool to those found in various populations throughout the world (specifically African and Asian populations), we ought to gain insight into the origins of the Malagasy.

Table 2 : Table showing reported frequencies of the mannose binding protein B and C alleles in various populations around the world.

POPULATION	B ALLELE	C ALLELE	REFERENCE
CAUCASIANS	0.11 - 0.17 ¹	0.02 ¹¹	¹ Garred <i>et al.</i> , 1992 ¹¹ Madsen <i>et al.</i> , 1994
DANISH	0.13		Garred <i>et al.</i> , 1992
ASIANS	0.11 - 0.17		Garred <i>et al.</i> , 1992
CHINESE	0.11		Lipscombe <i>et al.</i> , 1992
VANUATU (S. PACIFIC)	0.01	0	Lipscombe <i>et al.</i> , 1992
PAPUA NEW GUINEA	0.07	0	Lipscombe <i>et al.</i> , 1992
ESKIMOS	0.11 - 0.17 ¹	0 ¹¹	¹ Garred <i>et al.</i> , 1992 ¹¹ Madsen <i>et al.</i> , 1994
GENERAL NON-AFRICAN	0.11 - 0.16		Lipscombe <i>et al.</i> , 1992
AFRICANS	0.03	0.23	Madsen <i>et al.</i> , 1994
KENYA		0.23	Madsen <i>et al.</i> , 1994
GAMBIANS (W. AFRICA)	0.003	0.23 - 0.29	Lipscombe <i>et al.</i> , 1992
XHOSA (SOUTH AFRICA)	0	0.27	Lipscombe <i>et al.</i> , 1992
SAN BUSHMEN (NAMIBIA)	0.009	0.07	Lipscombe <i>et al.</i> , 1992

CHAPTER 2: MATERIALS AND METHODS

2.1 *Subjects and sample collection*

To achieve the objectives of this study outlined in **section 1.1**, we examined Malagasy individuals from 15 of the 19 Malagasy ethnic groups. **Table 3** gives a list of the ethnic groups included in the study, the corresponding sample sizes and the polymorphic systems used.

Blood samples were obtained during six field trips to Madagascar between 1992 and 1996. Samples were obtained by Professor Trefor Jenkins and co-workers in the department of Human Genetics, SAIMR with approval from the ethics committee of the University of the Witwatersrand, and informed consent from volunteers.

Approximately 2 500 healthy Malagasy individuals from various regions of Madagascar were sampled, including 1 425 unrelated Malagasy individuals who claimed to have the same patrilineal and matrilineal ethnic groups. In addition, information regarding name, gender, age, ethnicity, mother's ethnicity and father's ethnicity was noted. Approximately 30 – 50ml of whole blood were drawn from each individual. Total genomic DNA was extracted from stored buffy coats by means of the salting out method (Miller *et al.*, 1988).

For the purposes of this study, further approval was obtained from the ethics committee of the University of the Witwatersrand (protocol number: M980606; **Appendix 1**). All

samples selected for mtDNA and autosomal studies were from unrelated individuals with matched maternal and paternal ethnicities, whereas the paternal ethnicity alone guided selection of males for Y chromosome studies. Where possible, the same individuals were used for all systems. For comparative purposes, data from the literature has been utilized and is referenced appropriately.

Table 3 : Malagasy ethnic groups examined for mtDNA, Y chromosome DNA and autosomal DNA variation

REGION / ETHNIC GROUP	HAPLOID SYSTEMS					DIPLOID SYSTEMS			
	MtDNA			Y Chromosome		PSAs			MBP
	9-bp Deletion	Control Region Sequencing		YAP	M9	Sb19.3	AT3id	APO-4	
		CR1	CR2						
CENTRAL HIGHLAND									
BETSILEO	66	36	34	28	3	30	21	22	17
MERINA	85	36	29	42	15	37	32	29	31
EASTERN CENTRAL									
ANTAIMORO	44	7	7	56	24	-	-	-	-
ANTAISAKA	79	46	37	41	26	46	34	28	24
ANTAMBAHOAKA	12	-	-	-	-	-	-	-	-
ANTAIFASY	15	1	1	-	-	-	-	-	-
BETSIMISARAKA	12	2	1	55	19	-	-	-	-
NORTHERN									
ANTANKARANA	-	3	3	13	4	-	-	-	-
TSIMIHETY	65	40	37	46	22	42	40	35	36
SOUTH WESTERN									
ANTANDROY	5	3	3	46	12	-	-	-	-
BARA	28	2	2	-	-	-	-	-	-
MAHAFALY	42	31	26	23	7	33	19	-	30
MIKIA	-	-	-	34	8	-	-	-	-
VEZO	20	11	10	2	-	15	11	15	9
WESTERN									
SAKALAVA	42	38	38	43	19	35	31	28	27
TOTALS:	515 ⁱ	256 ⁱⁱ	228 ⁱⁱⁱ	429 ^{iv}	159	211	188	157	174

ⁱ 280 of these samples were typed by Soodyall *et al.* (1996)

ⁱⁱ 55 of these samples were sequenced by Soodyall, personal communication

ⁱⁱⁱ 49 of these samples were sequenced by Soodyall, personal communication

^{iv} 380 of these samples were typed by Morar (2000)

2.2 Molecular methods

Specific regions of the genome were amplified using the polymerase chain reaction (PCR) which results in the production of an exponential amount of duplicate copies of the DNA region of interest (Saiki *et al.*, 1985). In all cases, these products were then resolved using agarose gel electrophoresis. Where necessary, PCR products were digested using the appropriate restriction enzyme or gel-purified and prepared for sequencing. The recipes for all solutions used are given in **Appendix 2**.

2.2.1 MtDNA

Screen for the 9-bp deletion

The 9-bp deletion was amplified by means of PCR following the method described by Wrischnick *et al.*, 1987. A typical 25µl reaction consisted of 1X *Taq* DNA polymerase buffer (already containing 1.5mM MgCl₂), 0.1mM dNTPs, 1M Bovine Serum Albumin (BSA), 0.4µM of each primer, 1U *Taq* DNA polymerase and 1µl DNA (approximately 1-2µg/µl) made up to 25µl with ddH₂O. The primers L8215 and H8297 were used in this reaction and their sequences are shown in **Table 4**. The PCR reaction was performed in a thermocycler where each cycle consisted of a denaturing step at 94°C for 1min, an annealing step at 56°C for 1min, and an extension step at 72°C for 1min. This process was repeated for 30 cycles and was followed by an extension step at 74°C for 10min (Hammer and Horai, 1995).

PCR products were resolved on 2% Metaphor high gelling temperature (HGT) agarose gels stained with 3µl 10mg/µl Ethidium Bromide (EtBr) per 100ml of gel. A 1-kb molecular weight marker (Gibco BRL) was used in order to size the fragments (**Appendix 3**). Five µl of Ficoll dye were added to 5µl of the 9-bp PCR product, and electrophoresed for 1.5hrs at 100mA in 1X TBE buffer. PCR products were visualised on an UV light box (260nm) and each gel was photographed to keep a record of the electrophoresed products.

The presence of two copies of the 9-bp repeat produced a PCR fragment 121-bp in size, while a fragment of 112-bp in size was obtained if only one copy of the repeat was detected.

MtDNA Control Region Sequencing

The sequencing procedure began by amplifying the 1.1-kb non-coding control region of the mtDNA molecule using primers H408 and L15996 (**Table 4**). PCR was conducted in a total volume of 50µl and each reaction contained 1X *Taq* DNA polymerase buffer (already containing 1.5mM MgCl₂), 0.1mM dNTPs, 1M BSA, 0.4µM of each primer, 2U *Taq* DNA polymerase and 1µl DNA (approximately 1-2µg/µl) made up to the total volume with 32.6µl ddH₂O. PCR was performed in a thermocycler where each cycle consisted of a denaturation step at 94°C for 1min, an annealing step at 56°C for 1min, and an extension step at 74°C for 1min. This process was repeated for a total of 30 cycles.

Following PCR, 5µl aliquots of PCR products were mixed with 5µl Ficoll loading dye and resolved on 1% HGT agarose gels stained with 3µl 10mg/µl EtBr per 100ml of gel. Electrophoresis was carried out at 100mA for approximately 1hr in 1X TBE buffer. A 1-kb molecular weight marker (**Appendix 3**) was used in order to size the fragments.

Once amplification had been confirmed, 35µl of the PCR product was electrophoresed in preparative 2% NuSieve low gelling temperature (LGT) agarose gels. Gels were prepared with 1X TAE buffer stained with 3µl 10mg/µl EtBr per 100ml of gel and were electrophoresed for 1.5hrs at 120mA in 1X TAE buffer. The products were viewed on an UV transilluminator and the intensity of each product was noted thus providing an indicator as to how much DNA should be used in the cycle sequencing reaction. Following this, the band of interest was excised. DNA in the excised gels were purified using the Wizard Column system (Promega, Whitehead Scientific) following the manufacturer's recommendation.

The purified DNA was then amplified by cycle sequencing using the terminator AmpliTaq FS premix kit (Perkin Elmer). The cycle sequencing step allows for the incorporation of fluorescent-labelled terminator dideoxynucleotide triphosphates (ddNTPs) and normal dioxynucleotide triphosphates (dNTPs) into the newly synthesised DNA strands. The primer used in each sequencing reaction was chosen according to the strand being sequenced: for the L strand, L15996 was used to amplify CR1 and L29 was used to amplify CR2; for the H-strand, H16401 was used to amplify CR1 and H408 was used to amplify CR2 (**Table 4**). Each reaction consisted of 4µl AmpliTaq FS premix kit, 0.33µM

of the appropriate primer, 5-8µl 1-kb PCR product, made up with ddH₂O to a total volume of 20µl. Cycle sequencing was carried out in a thermocycler where each amplification cycle consisted of a denaturation step at 96°C for 30s, an annealing step at 50°C for 15s and an extension step at 60°C for 4min. Twenty-five cycles were performed.

The cycle sequencing products were then purified using CENTRI-CEP columns (Princeton Separations, Inc), and then vacuum-dried for an hour in a vacuum dryer with no heat. Dried samples were resuspended in 1.5µl of dextran-formamide loading dye, were denatured by heating at 90°C for 2min, cooled on ice, and then loaded onto 4.3% denaturing polyacrylamide gels. Electrophoresis was conducted in 1X TBE buffer for seven hours using an automated ABI 377 DNA sequencer. Sequences were analysed using ABI Prism 377 sequencing software.

2.2.2 Y Chromosome

The Y *Alu* Polymorphism (YAP)

The protocol for PCR detection of the Y *Alu* insert was first described by Hammer and Horai in 1995. Slight modifications to the original protocol were necessary to optimise the PCR reaction.

Table 4: Primers used in the amplification of the 9-bp deletion, the 1-kb control region of mtDNA and cycle sequencing (*Hertzberg *et al.*, 1989; **Vigilant, 1990)

mtDNA REGION	PRIMER NAME	STRAND	POSITION OF 3' END	SEQUENCE (5'-3')
COII/tRNA ^{Lys}	L8215**	Light	8215	ACA GTT TCA TGC CCA TCG TC
	H8297**	Heavy	8297	ATG CTA AGT TAG CTT TAC AG
CR1	L15996*	Light	15 996	CTC CAC CAT TAG CAC CCA AAG C
	H16401*	Heavy	16 401	TGA TTT CAC GGA GGA TGG TG
CR2	L29*	Light	29	GGT CTA TCA CCC TCT TAA CCA C
	H408*	Heavy	408	CTG TTA AAA GTG CAT ACC GCC A

Table 5: Primers used in the YAP and M9 PCR reactions and their sequences (Underhill *et al.*, 1997)

SYSTEM	PRIMER NAME	SEQUENCE (5'-3')
YAP	YAP Forward	CAG GGG AAG ATA AAG AAA TA
	YAP Reverse	ACT GCT AAA AGG GGA TGG AT
M9	M9 Forward	GCA GCA TAT AAA ACT TTG AGG
	M9 Reverse	AAA ACC TAA CTT TGC TCA AGC

The YAP region was amplified using a total reaction volume of 15.0 μ l. A typical reaction consisted of 1X *Taq* DNA polymerase buffer (without MgCl₂), 0.1mM dNTPs, 3mM MgCl₂, 0.4 μ M of each primer, 0.4U *Taq* DNA polymerase and 1 μ l DNA (approximately 1-2 μ g/ μ l) made up to 25 μ l with ddH₂O. The primers (Hammer and Horai, 1995; Jobling and Tyler-Smith, 1995) used for this system are shown in Table 5

The PCR reaction was performed in a thermocycler where each cycle consisted of a denaturing step at 94°C for 1min, an annealing step at 51°C for 1min, and an extension step at 72°C for 1min. This process was repeated for 30 cycles. Each run was preceded by an initial denaturation step at 94°C for 2min. The final holding temperature was set to 26°C.

The PCR products were run on 2% HGT agarose gels stained with 3 μ l 10mg/ μ l EtBr per 100ml of gel in order to size the fragments. This was achieved by using 10 μ l of each sample mixed with 5 μ l of Ficoll loading dye, and electrophoresed at 100mA for 45min in 1X TBE buffer. A 1-kb molecular size standard (Appendix 3) was used in order to determine the molecular sizes of each fragment. Each gel was irradiated with UV light of 260nm and photographed in order to keep a record of the electrophoresed products.

This system has two polymorphic alleles: an allele 455-bp in size that represents the presence of the *Alu* insertion (called YAP+ or the ‘insertion allele’), and an allele 150-bp in size that represents the absence of the insertion (called YAP- or the ‘ancestral allele’).

M9

The samples that were found to be YAP- were further screened to assess their M9 status. A typical 25µl reaction consisted of 1X *Taq* DNA polymerase buffer (already containing 1.5mM MgCl₂), 0.2mM dNTPs, 0.2µM of each primer, 0.5U *Taq* DNA polymerase and 1µl DNA (approximately 1-2µg/µl) made up to 25µl with ddH₂O. The primers used for this system are shown in **Table 5**.

The PCR reaction was performed in a thermocycler where each cycle consisted of a denaturing step at 94°C for 45s, an annealing step at 54°C for 45s and an extension step at 72°C for 45s. This process was repeated for 30 cycles and was preceded by an initial denaturing step at 94°C for a period of 2min.

PCR products were resolved using 3% HGT agarose gels stained with 3µl 10mg/µl EtBr per 100ml of gel. Five µl of Ficoll dye were added to 5µl of the M9 PCR product, and electrophoresed for 1.5hrs at 100mA in 1X TBE buffer. A 1-kb molecular weight marker (**Appendix 3**) was used in order to size the fragments. Each gel was irradiated on an UV light box (260nm) and photographed in order to keep a record of the electrophoresed products. Amplification was confirmed when a 340-bp product was visualised.

The 340-bp product was digested using *Hinf*I (Boehringer Mannheim) for 2hrs at 37°C. Each digestion contained 10µl PCR product, 0.1U *Hinf*I, 1X Buffer H (Boehringer Mannheim), made up to a total volume of 20µl. Digestion products were resolved using

2% HGT agarose gels stained with 3µl 10mg/µl EtBr per 100ml of gel. A 1-kb molecular weight marker (Appendix 3) was used in order to size the fragments. Five µl of Ficoll dye were added to 15µl of the digested M9 PCR product, and electrophoresed for approximately 1hr at 100mA in 1X TBE buffer. This system has two restriction fragment length polymorphism (RFLP) patterns. If the “C” allele is present, digestion with *HinF1* produces a 245-bp and a 95-bp fragment; and if the “G” allele is present, 3 fragments of 181-bp, 95-bp and 64-bp in size are produced.

2.2.3 Autosomal DNA markers

Four autosomal gene markers (Sb19.3, AT3id, APO 4 and MBP) were used to examine the Asian and African contribution in the Malagasy.

All four loci were amplified in a total reaction volume of 25µl. In all cases, 1X *Taq* DNA polymerase buffer (already containing 1.5mM MgCl₂) and 1µl DNA (approximately 1-2µg/µl) was used. All PCR reactions were made up to the required total volume with ddH₂O. Differences between systems as well as PCR conditions are outlined in Table 6. Primers used in each system are detailed in Table 7.

In all cases, PCR products were resolved using HGT agarose gels stained with 3µl 10mg/µl EtBr per 100ml of gel in order to size the fragments. Five µl of Ficoll loading dye were added to each sample preceding electrophoresis at 100mA for 1hr in 1X TBE buffer. A 1-kb molecular size standard (Appendix 3) was used in order to determine the

molecular sizes of each fragment. Each gel was viewed on an UV light box (260nm) and photographed. Again, differences between systems are outlined in **Table 6**.

Only the MBP system required restriction endonuclease (RE) digestion. The 349-bp product was digested using the restriction enzymes *BanI* (Boehringer Mannheim) and *MboII* (Boehringer Mannheim). The *BanI* digestion detects allele B, which if present remains uncut and the wildtype A allele is cleaved to produce fragments of sizes 260-bp and 89-bp. Ten μ l of PCR product were digested with 5.04U *BanI* and 1X Buffer T (Boehringer Mannheim), made up with ddH₂O to a total volume of 20 μ l for 1hr at 50°C. The *MboII* digestion detects allele C, which if present cleaves the 349-bp PCR product into fragments of sizes 279-bp and 70-bp. The wildtype allele A remains uncut. Ten μ l of PCR product were digested with 3.48U *MboII* and 1X Buffer L (Boehringer Mannheim), made up with ddH₂O to a total volume of 20 μ l for 90min at 37°C. Products from each digestion were resolved using 2% HGT Metaphor agarose gels stained with 3 μ l 10mg/ μ l EtBr per 100ml of gel. A 1-kb molecular weight marker (**Appendix 3**) was used in order to size the fragments. Five μ l of Ficoll dye were added to 15 μ l of the digested PCR products, and electrophoresed for approximately 1hr at 100mA in 1X TBE buffer. Each gel was irradiated with UV light of 260nm and photographed to keep a record of the electrophoresed products.

Table 6: PCR components, PCR conditions, gel conditions and PCR product sizes for the Sb19.3, AT3id, APO and MBP systems.

	Sb19.3	AT3id	APO 4	MBP
PCR Components	0.1mM dNTPs, 0.4μM of each primer and 1U <i>Taq</i> DNA polymerase	0.1mM dNTPs, 0.4μM of each primer and 1U <i>Taq</i> DNA polymerase	0.1mM dNTPs, 0.4μM of each primer; 0.15mM MgCl ₂ and 1U <i>Taq</i> DNA polymerase,	0.125mM dNTPs, 0.2μM of each primer; 0.15mM MgCl ₂ and 0.5U <i>Taq</i> DNA polymerase
PCR conditions:				
Initial step	-	-	-	94°C, 2min
Denaturation	94°C, 1min	94°C, 90s	94°C, 1min	94°C, 30s
Annealing	60°C, 1min	55°C, 30s	50°C, 1min	61°C, 1min
Extension	72°C, 1min	72°C, 30s	72°C, 1min	72°C, 90s
Number of cycles	30 cycles	35 cycles	35 cycles	35 cycles
Final step	-	72°C, 10min	-	72°C, 5min
Gel Conditions	3% HGT agarose gel; 15μl PCR product used in electrophoresis	4% composite gel (1% NuSieve and 1% HGT agarose); 10μl PCR product used in electrophoresis	2% HGT agarose gel; 20μl PCR product used in electrophoresis	2% HGT agarose gel; 5μl PCR product used in electrophoresis; this system also requires enzyme digestion
PCR Product Sizes	allele 1: 450-bp allele 2: 122-bp	allele 1: 572-bp allele 2: 496-bp	allele 1: 437-bp allele 2: 122-bp	349-bp

Table 7: Primers used in the autosomal DNA PCR reactions and their sequences

PSA SYSTEM	PRIMER NAME	SEQUENCE (5'-3')
Sb19.3ⁱ	Sb19.3 forward	TCT AGG CCC AGA TTT ATG GTA ACT G
	Sb19.3 reverse	AAG CAC AAT TGG TTA TTT TCT GAC
AT3idⁱ	AT3id forward	CCA CAG GTG TAA CAT TGT GT
	AT3id reverse	GAG ATA GTG TGA TCT GAG GC
APO 4ⁱ	APO 4 forward	AAG TGC TGT AGG CCA TTT AGA TTA G
	APO 4 reverse	AGT CTT CGA TGA CAG CGT ATA CAG A
MBPⁱⁱ	M9 Forward	GCA GCA TAT AAA ACT TTG AGG
	M9 Reverse	AAA ACC TAA CTT TGC TCA AGC

ⁱ Schriver et al., 1997

ⁱⁱ Lipscombe *et al.*, 1994

2.3 Computational analyses

2.3.1 MtDNA analyses

Sequence data from printed results were compared with the published reference sequence (Anderson *et al.*, 1981) and recorded using the GELIN programme (Sherry, 1991). These files were then edited by means of an editor programme (PCWRITE) in order to create input files compatible for further analysis using programmes such as the Molecular Evolutionary Genetics Analysis (MEGA) version 1.01 (Kumar *et al.*, 1993) and Minimum Evolution Bootstrapping (MEBOOT) software (Kumar, personal communication). These programmes contained algorithms to construct phylogenetic trees and estimate evolutionary distances. Specifically, MEGA was used to generate pairwise comparisons of sequences in order produce a table of variant sites and MEBOOT was used to draw Neighbor-joining (NJ) gene trees (Saitou and Nei, 1987) through the use of a matrix of the proportion of nucleotide differences (p-distance). A bootstrap test based on 1 000 replications was applied to estimate the confidence of the branching patterns of NJ trees (Felsenstein, 1975).

The demographic history (past population growth or decline, population separations and bottlenecks) of the Malagasy was then investigated using mismatch (within a population) and intermatch (between populations) distributions which are based on the frequency distributions of the number of pairwise differences from the mtDNA sequences (Rogers and Harpending, 1992). Mismatch distributions reveal episodes of past population

expansions that can be dated using the mutation rate of mtDNA and can thus be said to preserve a record of past population dynamics (Harpending *et al.*, 1993). Both mismatch and intermatch distributions were constructed using the IWAVE programme (Sherry, personal communication) and a substitution rate of 16.6% per million years (Soodyall *et al.*, 1996). The mean (μ) and variance (δ) of the distribution, as well as estimates of Tau (τ), a parameter that was shown to characterise mismatch distributions in Rogers and Harpending (1992) model of sudden population growth, were computed also using IWAVE. Estimates of past population expansions and separations are measured as $\tau=2\mu t$ units of mutational time, in which μ is the total mutation rate over all sites in the sample and t is time in generations (Rogers and Harpending, 1992). The mean expansion time in years or time of divergence of two populations is computed by the IWAVE programme using the following formula: $(\tau/\text{effective sequence length})(1^6/2\mu)/2$. Harpending's raggedness statistic (r), which can be used to differentiate between expanding and constant population size (Tajima, 1989; Harpending 1994) was calculated using ARLEQUIN (Schneider *et al.*, 1997).

To facilitate comparison of data with that of the hypothesised ancestral Malagasy populations (Africans and Asians), we made use of principle component analysis (PCA) (Hotelling, 1993) and constructed principal component plots (PC plots) using both ARLEQUIN and NTSYS (Slatkin, 1995). ARLEQUIN was used to calculate Slatkin's linearized F_{st} matrices which were used as the input data for NTSYS. To test various relationships, the analysis of molecular variance (AMOVA) was calculated using ARLEQUIN. By estimating the percentage of variation between certain groups, AMOVA

enabled us to test whether or not the biological data were in accordance with historical and geographical hypotheses (such as whether there is a difference between individuals from the various ethnic groups and geographic regions) as well as determine the percentage of variation between Madagascar and populations from Africa and Asia.

2.3.2 Y Chromosome analyses

Allele frequencies were determined for each Malagasy ethnic group, geographic region, and for the whole population. The frequency of YAP⁺ chromosomes suggested an African contribution to the Malagasy gene pool. The frequency of the M9C and M9G alleles were used in conjunction with YAP⁻ chromosomes to infer the proportion of non-African Y chromosomes in the present-day Malagasy. Furthermore, hierarchical structure among the Malagasy was assessed through the use of AMOVA calculated with ARLEQUIN (Schneider *et al.*, 1997) as well as PCA (Hotelling, 1993).

2.3.3 Autosomal DNA markers

For all four autosomal markers (Sb19.3, AT3id, APO-4 and MBP), allele frequencies were determined for each ethnic group, geographic region, and the total Malagasy population. For the three PSAs, results were tested for goodness of fit to Hardy-Weinberg by means of the Chi-squared test.

Hierarchical structure among the Malagasy ethnic groups was assessed in an AMOVA calculated with the ARLEQUIN package (Schneider *et al.*, 1997). Population affinities were examined by using comparative data from the literature (Tables 1 and 2) by means of PCA.

2.3.4 Correlation analysis

To establish whether or not the results obtained with mtDNA, Y chromosome and autosomal DNA markers were in agreement, Pearson's correlation coefficient (r) was utilised. Using this method, we were able to estimate the magnitude (i.e. the degree to which the variables move in unison or opposition) and direction (positive or negative) of each of the following relationships: mtDNA versus Y chromosome markers, mtDNA versus autosomal DNA markers and Y chromosome markers versus autosomal DNA markers.

The correlation analysis was performed using a computer programme supplied by Dr Lane (personal communication) in which Pearson's correlation coefficient was calculated through the use of the following formula: $\Sigma(X-\bar{X})(Y-\bar{Y}) / (N-1)s_x s_y$, where $\bar{X}$ is the mean of X , $\bar{Y}$ is the mean of Y , N is the number of pairs of cases, and s_x and s_y are the standard deviations for X and Y respectively.

CHAPTER 3: RESULTS

3.1 *MtDNA results*

3.1.1 Frequency of the 9-bp deletion

When two copies of the 9-bp repeat were present, we obtained a PCR fragment of 121-bp in size, while a deletion of one repeat produced a PCR fragment of 112-bp in size (**Figure 4**). Altogether we found the deletion in 116 out of the 515 Malagasy individuals screened (**Table 8**). The frequency of the deletion throughout the island is therefore 22.5%.

The deletion was found in all ethnic groups but was seen to be more frequent in the highland region (25.8%) than in the lowland regions (21.1%), although these differences were not significant at a 95% significance level ($p=0.28$). Furthermore, the 9-bp deletion was found at a higher frequency in inhabitants from the Western and Eastern Central regions compared with groups from the Northern and South Western regions.

In a study conducted by Soodyall *et al.* (1997), it was found that the 9-bp deletion was present in 75 out of 280 (26.8%) Malagasy individuals. In the present study, we have refined these estimates using a larger sample size and a more thorough sampling throughout all regions of the island.

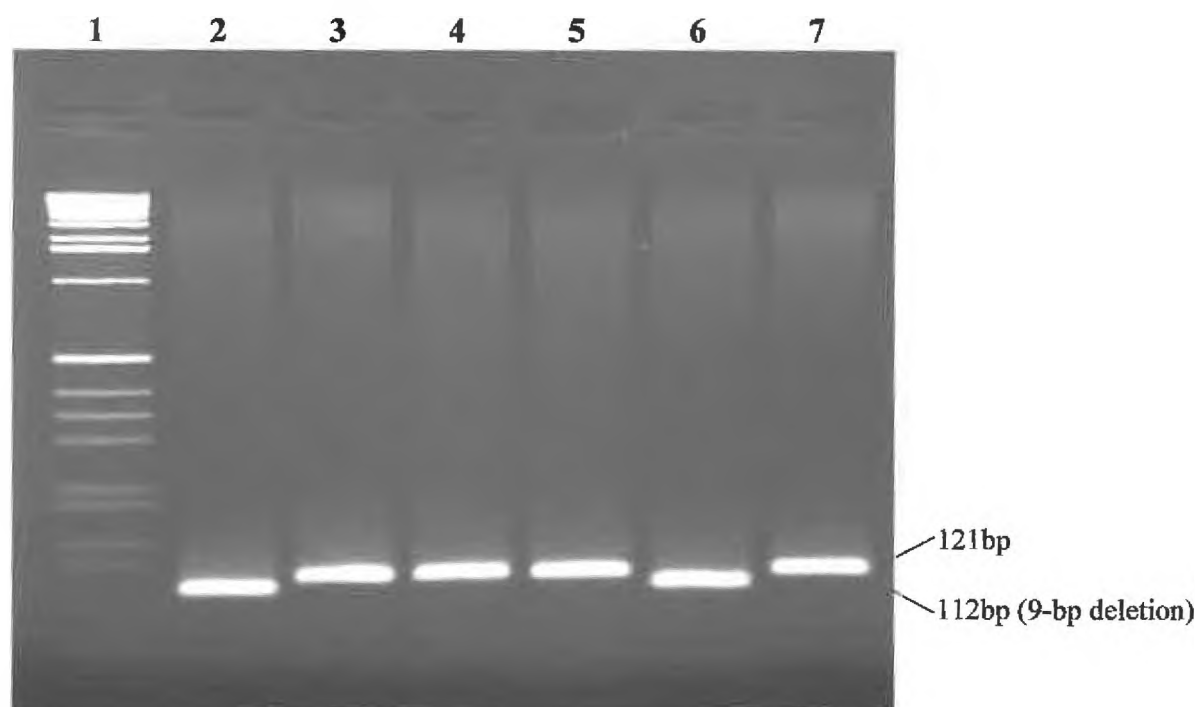


Figure4 : Electrophoretic separation of PCR products with (lanes 2 and 6) and without (lanes 3, 4, 5 and 7) the 9-bp deletion on a 2% Metaphor agarose gel. We were able to size the fragments through the use of a 1-kb molecular marker (lane 1).

Table 8 :Number of individuals and frequency of the 9-bp deletion in Malagasy ethnic groups in the different regions.

REGION / ETHNIC GROUP	NUMBER SCREENED	NUMBER WITH THE DELETION	FREQUENCY OF THE DELETION
CENTRAL HIGHLAND	151	39	0.258
BETSILEO	66	18	0.273
MERINA	85	21	0.247
EASTERN CENTRAL	162	42	0.259
ANTAIMORO	44	14	0.318
ANTAISAKA	79	17	0.215
ANTAMBAHOAKA	12	3	0.250
ANTAIFASY	15	5	0.333
BÊTSIMISARAKA	12	3	0.250
NORTHERN	65	13	0.200
TSIMIHETY	65	13	0.200
SOUTH WESTERN	95	10	0.105
ANTANDROY	5	2	0.400
BARA	28	3	0.107
MAHAFALY	42	4	0.095
VEZO	20	1	0.050
WESTERN	42	12	0.286
SAKALAVA	42	12	0.286
TOTAL:	515	116	0.225

3.1.2 MtDNA control region sequencing

Several lines of evidence based on historic, linguistic and archaeological data have indicated that the present-day Malagasy are derived from varying degrees of Indonesian and African admixture. However, the source of the African and Asian components has not yet been defined. To elucidate the African and Asian contribution to the present-day Malagasy, we amplified the 1.1-kb control region of the mtDNA genome by PCR (**Figure 5**) and sequenced it on an automated ABI 377 DNA sequencer from 228 Malagasy individuals. Individual sequences (**Figure 6**) were compared to the published reference sequence (Anderson *et al.*, 1981) and all substitutions are described relative to the positions in the latter.

From pairwise comparisons of sequences incorporating data from both CR1 and CR2, 124 unique mtDNA types were derived (**Appendix 4**). Using 777-bp of sequence data (402-bp in CR1 and 375-bp in CR2), 151 nucleotide changes (98 in CR1 and 53 in CR2) were identified. Transition mutations accounted for substitutions at 133 sites whilst transversions were noted at 27 sites, resulting in an overall transition to transversion ratio of approximately 5:1. Transversions occurred at positions 16017, 16064, 16070, 16141, 16170, 16171, 16188, 16207, 16220, 16242, 16247, 16248, 16249, 16261, 16262 and 16355 in CR1, and 58, 91, 95, 109, 186, 189, 249, 259, 277, 289 and 345 in CR2. Ten sites were found to be affected by multiple state substitutions: 16070 (A→T/G), 16141 (A→G/T), 16242 (C→G/A), 16247 (A→G/T/C), 16248 (C→T/G), 16249 (C→T/G),

16261 (A→G/T), 16262 (C→T/G) and 16355 (C→T/G) in CR1 and 189 (A→G/C) in CR2.

We used the NJ method to examine the relationship between the 124 unique mtDNA types found in the Malagasy (Figure 7). We found that mtDNA types from the various ethnic groups clustered together on branches throughout the tree and that in some instances, individuals from different ethnic groups shared the same mtDNA type (Appendix 4).

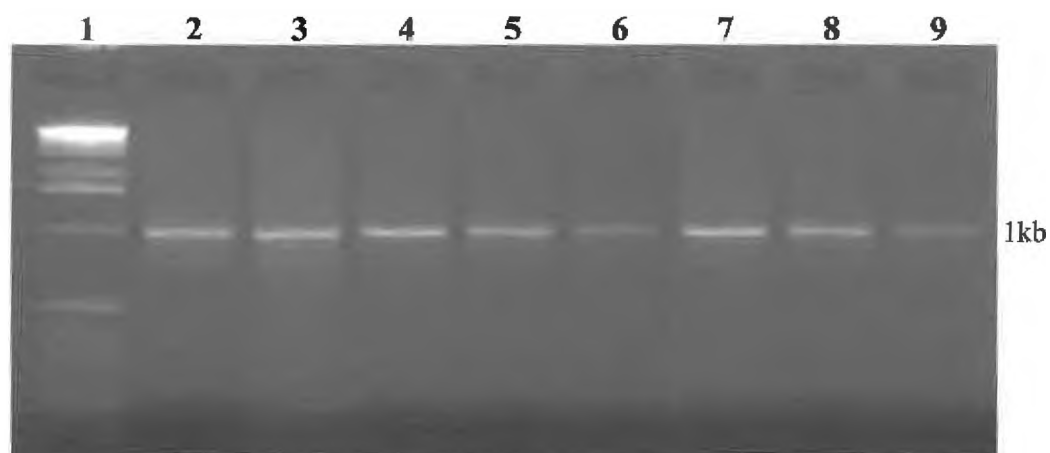


Figure5 : Electrophoretic separation of 1.1-kb mtDNA control region PCR products, prior to excision, on a 2% NuSieve agarose gel. A 1-kb molecular marker (lane 1) was used to size the fragments.



Model 377
Version 3.3
ABI100
Version 3.2

35-MA117/L2
MA117/L2
Lane 35

Signal G:403 A:431 T:297 C:245
DT4%Ac(A Set-AnyPrimer)
MAT2-95110930
Points 984 to 8256 Pk 1 Loc: 984

Page 2 of 4
Wed, Jun 23, 1999 8:56 am
Tue, Jun 22, 1999 5:05 pm
Spacing: 11.54(11.54)

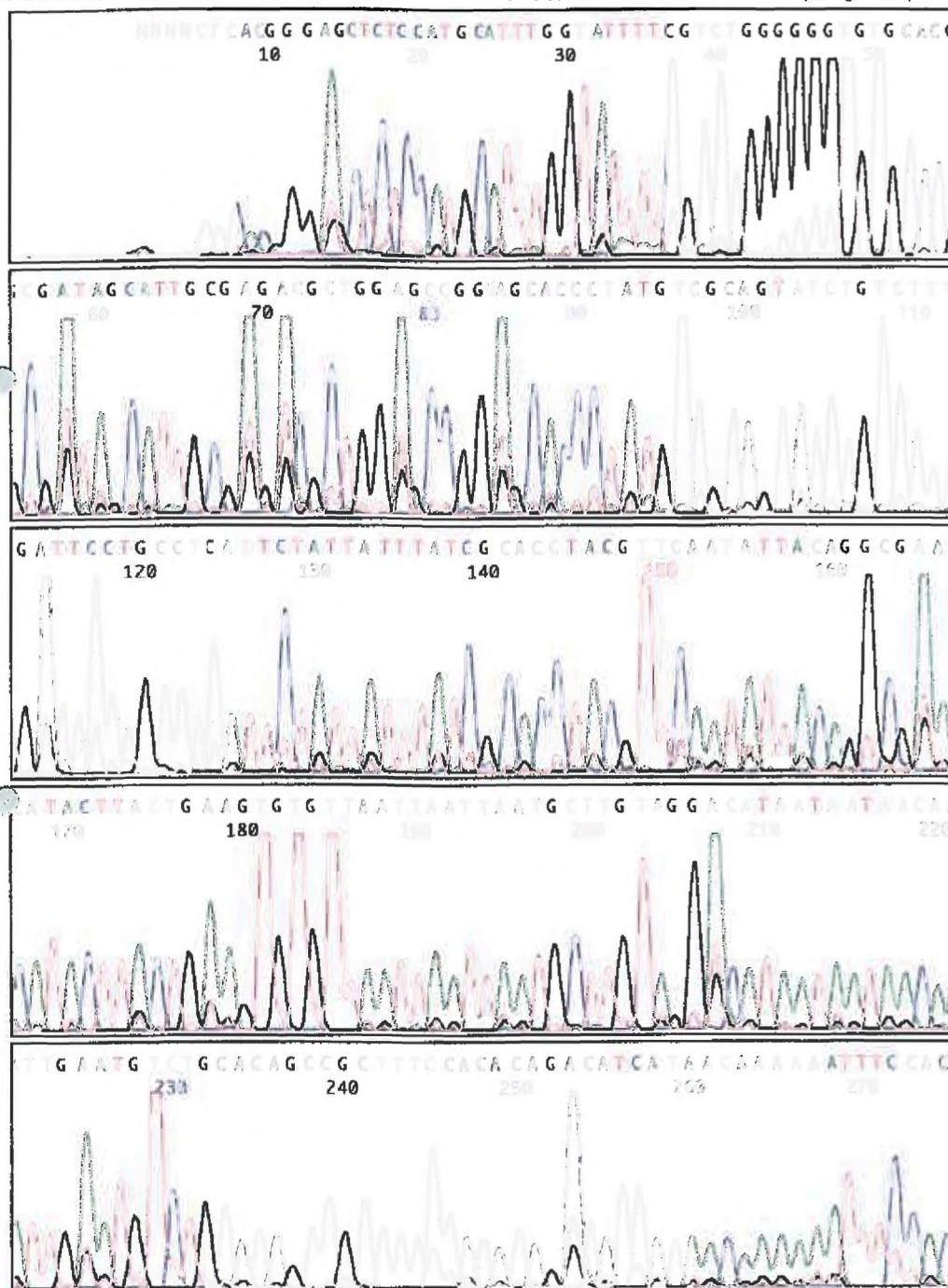


Figure 6: A CR1 sequence produced by electrophoresis on a 4.3% denaturing polyacrylamide gel run on an ABI 377 automated DNA sequencer.

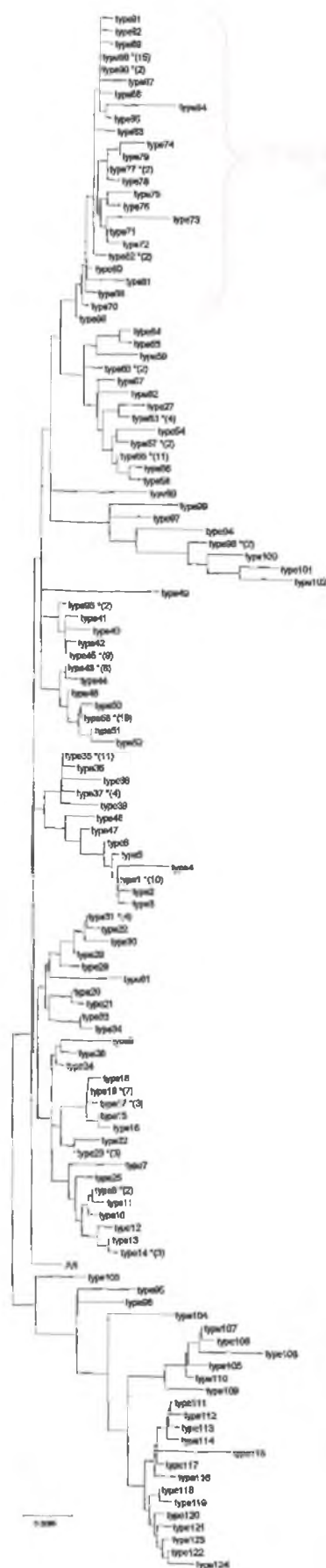


Figure 7: A neighbor-joining tree (constructed using the complete deletion method) indicating the relationship between the 124 unique mtDNA types found in the Malagasy. Shared types are illustrated by an asterisk (*), followed by a number in brackets indicating the number of individuals with that type. The types indicated in this figure correspond to those described in Table 9.

3.1.3 Sequence types associated with the 9-bp deletion

Since the 9-bp deletion has independent origins in Asians and Africans, and have different sequence profiles (Redd *et al.*, 1995; Soodyall *et al.*, 1996), mtDNA types associated with the deletion in the Malagasy can be traced to their parental populations (Soodyall *et al.*, 1996). From our survey of mtDNA type variation, we found that of the 124 unique mtDNA types in the Malagasy (**Appendix 4**), types 68-92 and 104-124 were found only in individuals with the 9-bp deletion. Types 68-92 have the sequence profile similar to that found in Asians (Redd *et al.*, 1995) and we refer to these mtDNA types as “9-bp deletion cluster 1”, while types 104-124 are closely related to those mtDNAs found in Africans and we refer to these types as “9-bp deletion cluster 2” (**Figure 7**). The frequencies of the Asian and African forms of the deletion in the total Malagasy sample were estimated to be 14.8% and 7.7%, respectively (**Figure 8**).

The Asian form of the deletion was found at highest frequency in the Western (83.3%) and Central Highland (81.5%) regions. In these regions, the proportion of the Asian to African form of the deletion is approximately 4:1, whereas this proportion is approximately 3:1 in the South Western region, and 1:1 in the Northern and Eastern Central regions.

The Polynesian motif, defined by the 9-bp deletion and the presence of three substitutions at nucleotide positions 16 247 (A→G), 16 217 (T→C), and 16 261 (C→T), was found in 51 out of 58 (87.9%) Malagasy with the Asian form of the deletion. Those individuals who do not have the Polynesian motif (12.1%) have the sequence profile similar to that

found in Southeast Asians and Indonesians (Melton *et al.*, 1995). All individuals with the African form of the deletion were found to have the common profile in Africa, referred to by Soodyall *et al.* (1996) as ‘the major form of the deletion in Africa’. The frequencies of the African and Asian forms of the deletion (including the Polynesian motif) among ethnic groups in the five regions in Madagascar are shown in **Figure 8**.

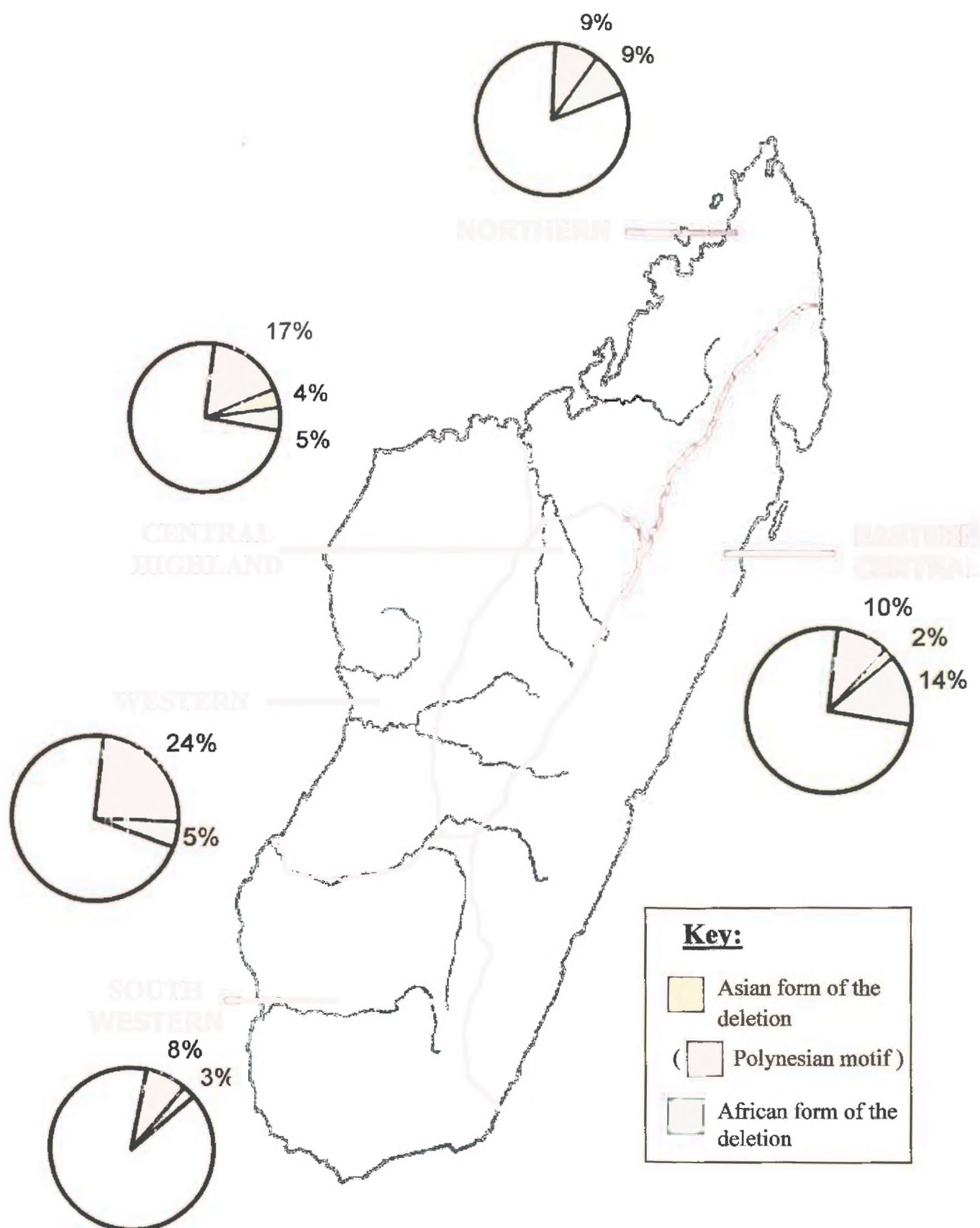


Figure 8: Map of Madagascar indicating the distribution of the Asian form of the 9-bp deletion (including the Polynesian motif) and the African form of the 9-bp deletion throughout Madagascar. See key for details.

3.1.4 Identifying origins of mtDNA types not associated with the 9-bp deletion in the Malagasy

In attempting to further resolve the ancestry of mtDNA types not associated with the deletion, we compared the mtDNA types found in the Malagasy with types found in Africans (n=165) and Asians (n=216) since these are the major parental contributors to the Malagasy gene pool (Table 9).

The NJ tree featured in **Figure 9** reveals that many of the mtDNA types found in Africans and Asians cluster on different branches of the tree. Furthermore, mtDNA types associated with the African and Asian forms of the deletion are located on different branches of the tree, but show close affinity to mtDNA types without the deletion in Africans and Asians, respectively. MtDNA types found in the Malagasy are located on branches throughout the tree. The NJ tree is able to resolve some of the mtDNA types found in Asians and Africans into separate branches, but the ancestry of some types could not be resolved (**Figure 9**; **Table 10**). The latter are closely associated in some branches of the tree and it is possible that these types belong to haplogroup L3 which is thought to be the lineage that left Africa to give rise to other lineages outside of Africa (Watson *et al.*, 1997). Haplotypes belonging to haplogroup L3 are widely dispersed in eastern Africa but are less prevalent in western and southern Africa. Watson *et al.* (1997) proposed that a subset of L3 mtDNAs moved out of eastern Africa to found the Eurasian populations about 60 000 years ago. On the basis of these observations, it is possible that mtDNA types belonging to haplogroup L3 first arose in sub-Saharan Africa, moved upward into

eastern Africa and then out of eastern Africa into the Middle East to yield Asian and European haplogroups (Chen *et al.*, 2000), before appearing in Madagascar.

Table 9: African and Asian populations (with respective sample size and reference) used for comparative purposes.

POPULATION		SAMPLE SIZE	REFERENCE
AFRICAN DATA	Central African Republic (CAR)	30	Barkhan, D (Personal Communication)
	Uganda	20	Barkhan, D (Personal Communication)
	Zambia	31	Barkhan, D (Personal Communication)
	Democratic Republic of Congo (DRC)	27	Manga, S(Personal Communication)
	Mozambique	31	Eddy, C (Personal Communication)
	Venda	26	Eddy, C (Personal Communication)
ASIAN DATA	Indian	30	Mountain et al, (1995) <i>Am. J. Hum. Genet.</i> 56 : 979-992
	China	10	Vigilant et al., (1991). <i>Science</i> 253 : 1503-1507
		15	Redd et al., (1995). <i>MBE</i> 12 : 604-625
	Australian Aboriginal	25	Redd et al., (1995). <i>MBE</i> 12 : 604-625
	Indonesia	20	Redd et al., (1995). <i>MBE</i> 12 : 604-625
		20	Vigilant et al., (1991). <i>Science</i> 253 : 1503-1507
	Papua New Guinea (PNG) Highland	15	Redd et al., (1995). <i>MBE</i> 12 : 604-625
	PNG Coast	25	Redd et al., (1995). <i>MBE</i> 12 : 604-625
	Moluccas	16	Redd et al., (1995). <i>MBE</i> 12 : 604-625
	Nusa Tenggaraans	20	Redd et al., (1995). <i>MBE</i> 12 : 604-625
	Polynesia	20	Redd et al., (1995). <i>MBE</i> 12 : 604-625

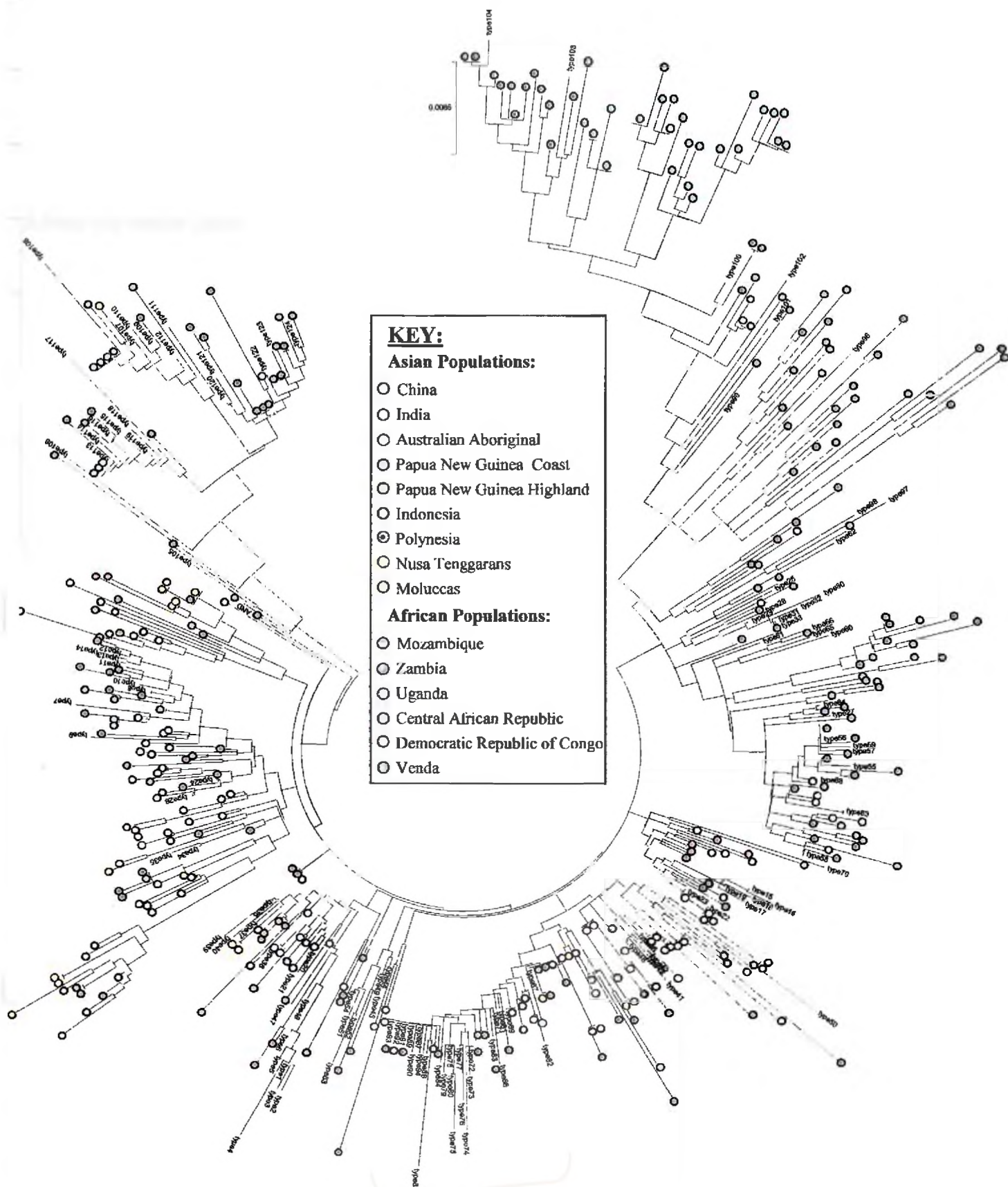


Figure 9: Neighbor Joining tree (constructed by the complete deletion method) showing the relationship between the 124 unique mtDNA sequence types found in the Malagasy with various Asian and African populations.

Table 10 : MtDNA types found in the various ethnic groups and regions traced to African and Asian origins, as well as those types which could not be unambiguously resolved.

REGION / ETHNIC GROUP	AFRICAN		ASIAN		UNRESOLVEABLE	
	# types	# individuals	# types	# individuals	# types	# individuals
CENTRAL HIGHLAND	8	10 (4.39%)	24	42 (18.42%)	8	9 (3.95%)
BETSILEO	2	2	15	27	3	4
MERINA	6	8	9	15	5	5
EASTERN CENTRAL	15	16 (7.02%)	21	32 (14.04%)	7	8 (3.51%)
ANTAIMORO	2	2	3	11	1	1
ANTAISAKA	9	10	17	20	6	7
ANTAKARANA	3	3	-	-	-	-
ANTAIFASY	-	-	1	1	-	-
BETSIMISARAKA	1	1	-	-	-	-
NORTHERN	16	16 (7.02%)	12	16 (7.02%)	4	4 (1.75%)
TSIMIHETY	16	16	12	16	4	4
SOUTH WESTERN	12	15 (6.58%)	15	23 (10.09%)	3	3 (1.32%)
ANTANDROY	2	2	1	1	-	-
BARA	-	-	2	2	-	-
MAHAFALY	6	9	9	16	1	1
VEZO	4	4	3	4	2	2
WESTERN	14	15 (6.58%)	7	16 (7.02%)	3	3 (1.32%)
SAKALAVA	14	15	7	16	3	3
TOTAL:	65*	72(31.58)	79*	129 (56.58%)	25*	27 (11.84%)

=number

* = The number of mtDNA types appears to be more than 124. This is because several types exist in more than one ethnic group thus inflating the numbers.

Overall, it was found that 55 of the 124 unique mtDNA types (accounting for 73 individuals) found in the Malagasy are associated with mtDNA types found in Africans, 53 types (accounting for 128 individuals) are derived from an Asian source and 16 types (found in 27 individuals) could not be resolved. Of the 156 individuals who do not have the 9-bp deletion, 44 were associated with Africans, 85 were associated with Asians, and 27 were unresolvable. Thus, a larger Asian than African contribution (approximately 2:1) is evident in the Malagasy (**Figure 9; Table 10**).

The mtDNA types that were unresolvable are so due to the close affinities between several types found in Africa and Asia on some branches of the tree (**Figure 9**), and also because those types may have been contributed by parental populations not included in this analysis.

MtDNA types found in both Africans and Asians are found throughout the island (**Figure 10**). The proportion of mtDNA types that were traced to Asian and African populations in the Central Highland region is approximately 4½:1. With respect to the lowland regions, the Asian contribution was greatest in the Eastern Central region where the Asian to African proportion was 2:1. This proportion was approximately 1½:1 in the South Western region and 1:1 in the Northern and Western regions.

The NJ tree presented in **Figure 7** was adapted to show the distribution of mtDNA types found in Asian and African populations in the Malagasy (**Figure 11**). Again we see that types found in Africans and Asians cluster together on branches throughout the tree, and in some instances are closely related to each other.

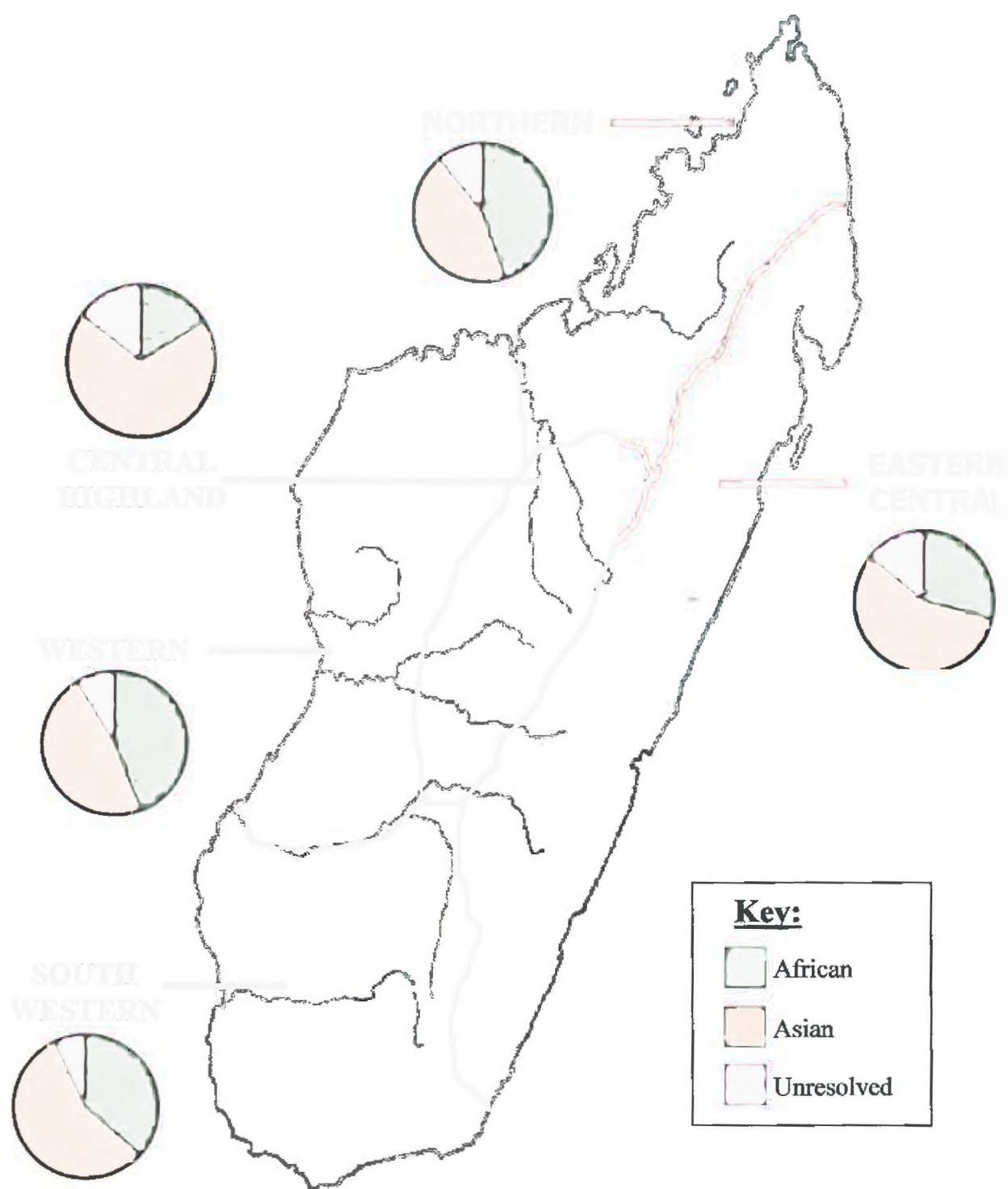


Figure10: Map of Madagascar indicating the distribution of mtDNA types associated with African and Asian types throughout the island. See key for details.

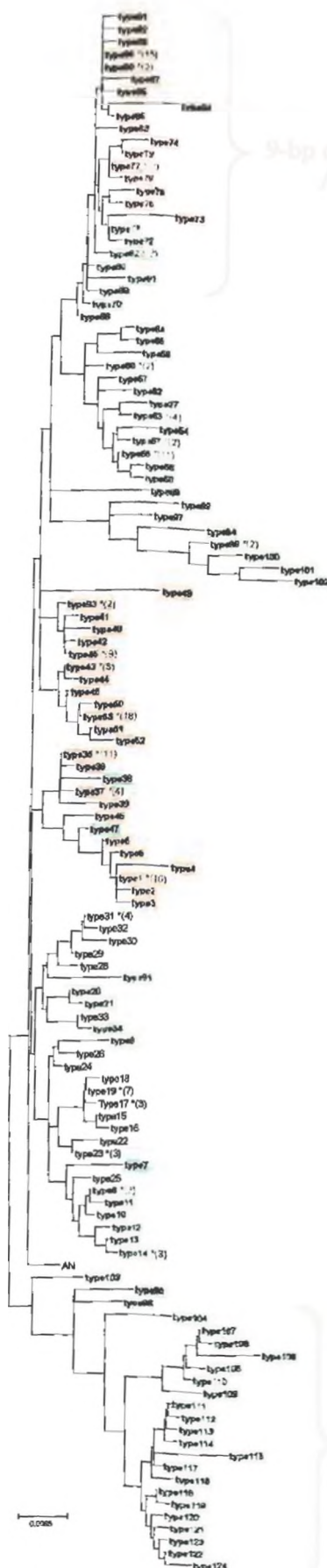


Figure11: Neighbour joining tree (constructed by the complete deletion method) showing which of the unique mtDNA types found in the Malagasy are associated with types found in Africans (green) and Asians (orange). Those sequence types that were unresolvable have not been coloured. This tree has been adapted from the one shown in Figure 7.

3.1.5 Relationships between Malagasy ethnic groups and geographic regions

We were able to draw conclusions about the relationships between ethnic groups occupying different regions of Madagascar by examining the mtDNA types found in the Malagasy (**Figure 7** and **Appendix 4**), and through the use of PCA (**Figure 12** and **13**). We found that individuals from different Malagasy ethnic groups shared the same mtDNA types and clustered together on branches throughout the tree (**Figure 7** and **Appendix 4**). This finding indicates a close relationship between ethnic groups and suggests that the population is relatively homogenous.

The PC plot (**Figure 12**) shows a close association between the Sakalava, Antaisaka and Betsileo groups who occupy different geographic regions on the island, namely the Western, Eastern Central and Central Highland regions, respectively. One would, however, expect to see a close affinity between the Merina and the Betsileo who occupy the same region, but this is not evident. It can therefore be said that the PC plot shows a lack of concordance between ethnic and geographical classification. Specifically, the Merina were not associated with the other Malagasy ethnic groups. This supports the historical data which indicates that the Merina have separate origins, presumably similar to that of Indonesians (Bradt, 1993).

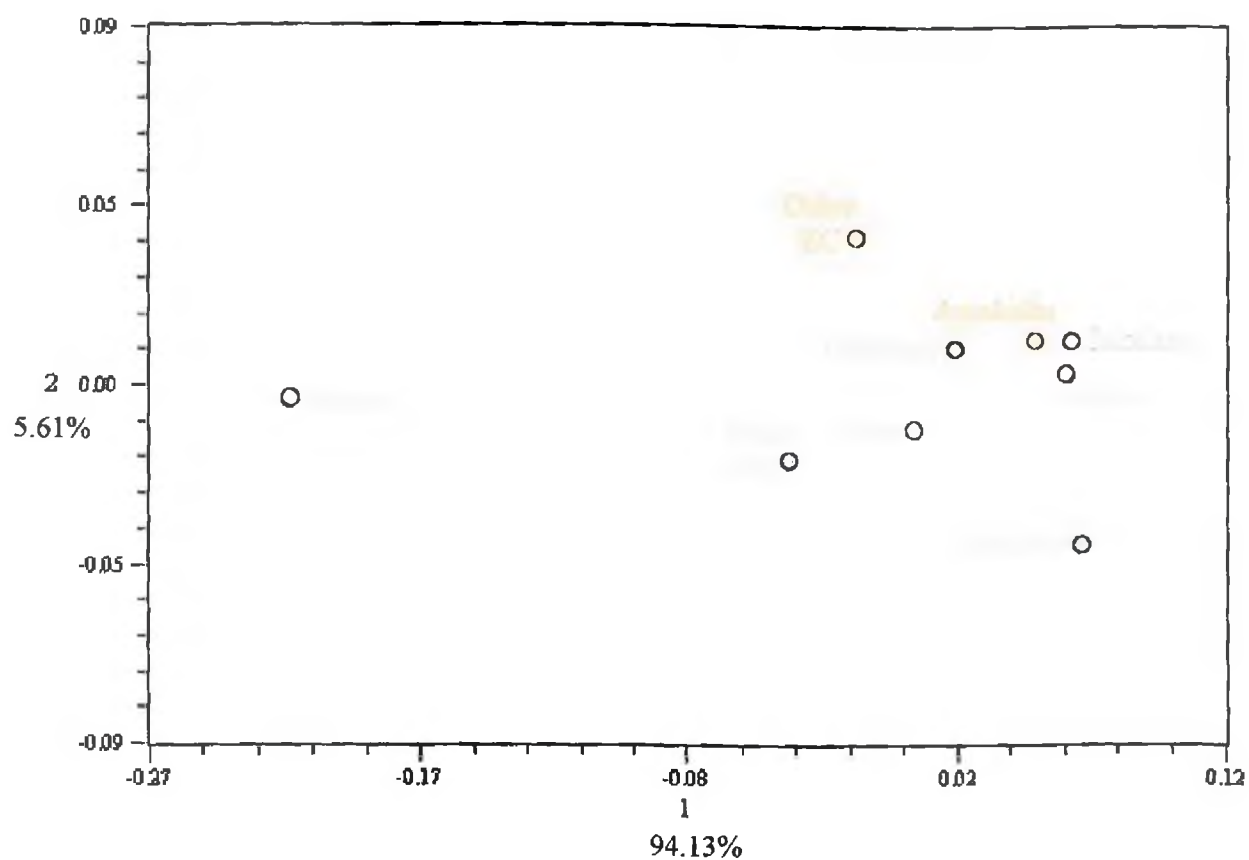


Figure12: Principal Component plot showing the relationship of ethnic groups in Madagascar using mtDNA data. Orange text represents ethnic groups in the Eastern Central region, green text represents the Northern region, red text the Central Highland region, pink text the Western region, and blue text the South Western region. The populations referred to as 'Other EC' and 'Other SW' consist of the smaller ($n < 10$) Eastern Central and South Western groups, respectively that have been pooled.

The association between the various Malagasy ethnic groups with populations from Africa and Asia reveals that some groups have a closer affinity with African populations while others are more closely associated with Asian populations (**Figure 13**).

The first principal component (the x-axis) appears to resolve the mtDNA types found in populations of Asian ancestry. The 'Other SW' and 'Other EC' populations are situated towards the left of the plot, and were not closely associated with the other Malagasy ethnic groups. These samples have been pooled and consist of the populations with small sample sizes ($n < 10$) that fall into the South Western and Eastern Central regions, respectively. It is possible that pooling these populations resulted in large variances within these groups thereby biasing their placement. Similarly, the Polynesian population is separated from the other populations of Asian origin. Presumably this is a consequence of the high frequency of the mtDNA 9-bp deletion harbouring the Polynesian motif, or it may be the result of genetic drift.

The second principal component (the y-axis) appears to carry information showing the separation of African and non-African populations, with the most amount of variation between the Central African Republic (CAR) and Papua New Guinea (PNG) Coast group. The African populations represented here belong to the Niger-Kordofanian family and, with the exception of the population from the Central African Republic (CAR), cluster in close proximity to each other. This result is due to the Pygmies that were included in the CAR sample.

The Malagasy ethnic groups are closely associated with each other in the plot (**Figure 13**). There is no concordance between ethnic groups and geographic region; the Sakalava from the Western region, the Tsimihety from the Northern region and the Antaisaka from the Eastern Central region cluster together in close proximity to the Australian Aboriginal group. The two highland groups (the Merina and Betsileo) show different affinities, with only the Merina showing a closer affinity with Indonesians. The Betsileo show a closer affinity with the Mahafaly who occupy the South Western region. The Vezo, thought to be of African ancestry (Bradt, 1993), show affinities with the Indian and PNG Highland populations.

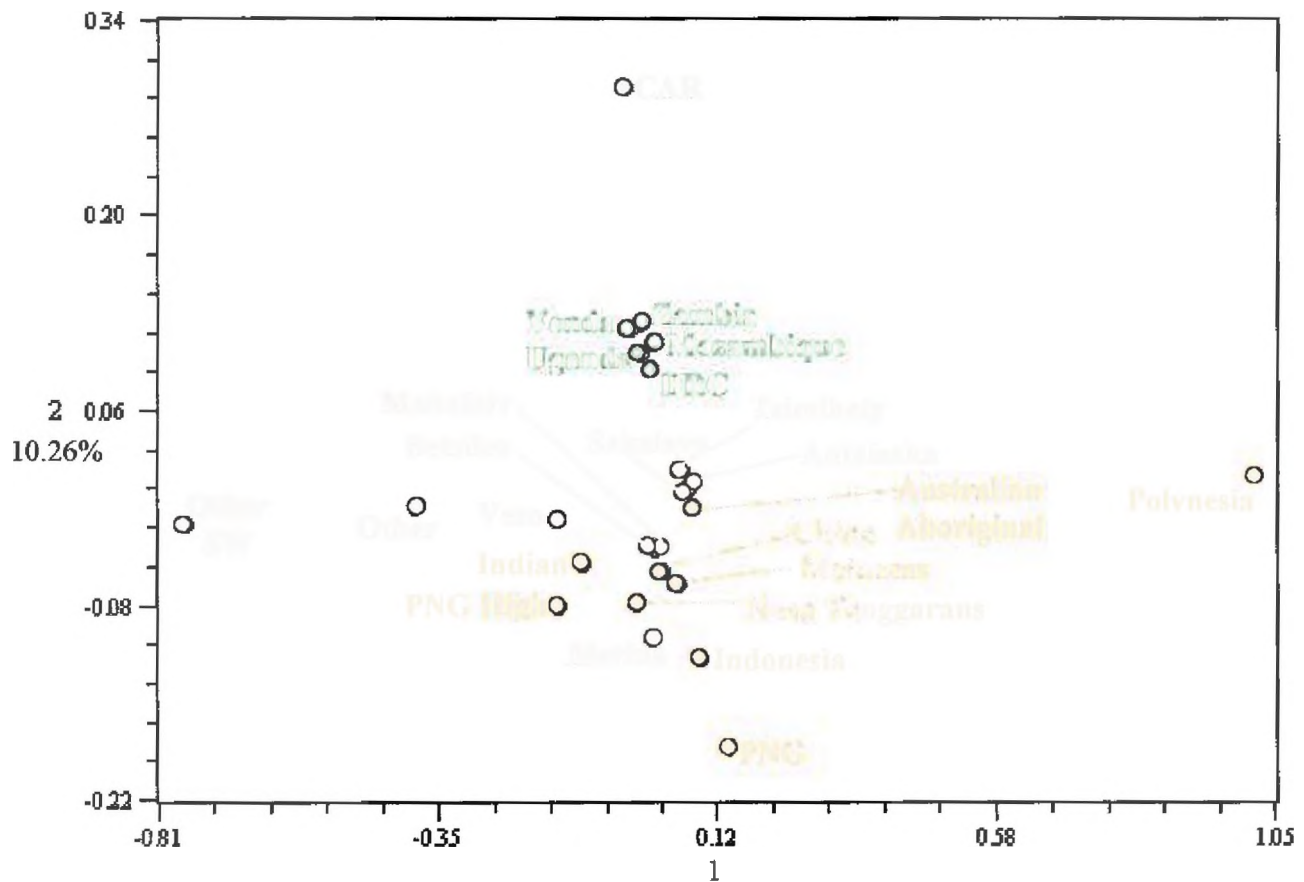


Figure13: Principal component plot showing the relationships between the Malagasy ethnic groups (red), and various Asian (orange) and African populations (Green) using mtDNA data.

3.1.6 Pairwise difference distributions

Mismatch and intermatch distributions were used to examine the evolution of the 9-bp deletion in the Malagasy.

In the first analysis, we examined the relationship of all mtDNA types with the 9-bp deletion (**Figure 14**). The wave profile characterising these mtDNA types is bimodal, with peaks at three and 13 differences, respectively. Secondly, we compared the mismatch and intermatch differences of individuals with the African and Asian forms of the deletion (**Figure 15**). This exercise has enabled us to resolve the wave profile characterising the entire 9-bp sample: the first peak (“peak 1” in **Figure 14**) is the result of the comparison of differences within each group whereas the second peak (“peak 2” in **Figure 14**) corresponds to the intermatch distribution which captures differences between the two groups.

In the third analysis, we compared the mismatch and intermatch distributions of the Asian and African forms of the deletion in the Malagasy with Asian (**Figure 16A**) and African (**Figure 16B**) forms of the deletion. We were able to calculate the divergence time (D) and calculate the time of separation (t) from the ancestral population ($D=2\mu t$). The divergence times computed for the Asian and African forms of the deletion were 8 752 years and 5 079 years, respectively. This suggests that the Malagasy were separated from Asian populations approximately 26 361 years ago, and the separation from African populations occurred about 15 298 years ago.

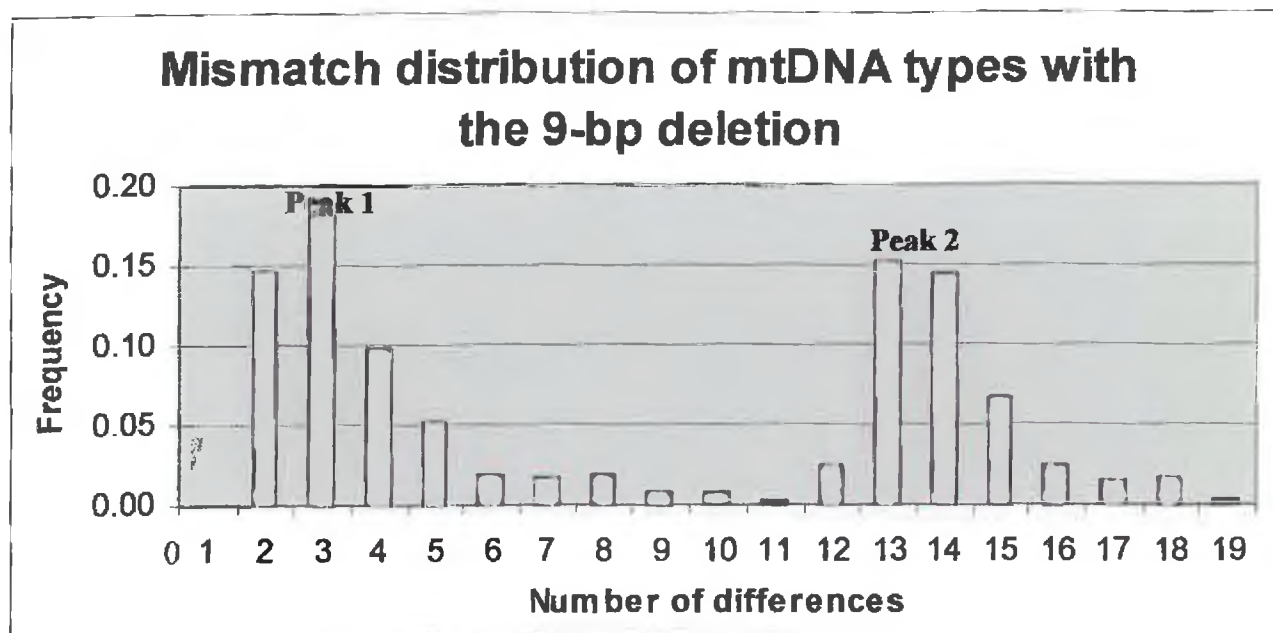


Figure14: Mismatch distribution of mtDNA types with the 9-bp deletion found in the Malagasy.
 mean (μ) = 6.454; variance (δ) = 29.925; Tau (τ) = 1.576; Harpending's raggedness statistic (r) = 0.036

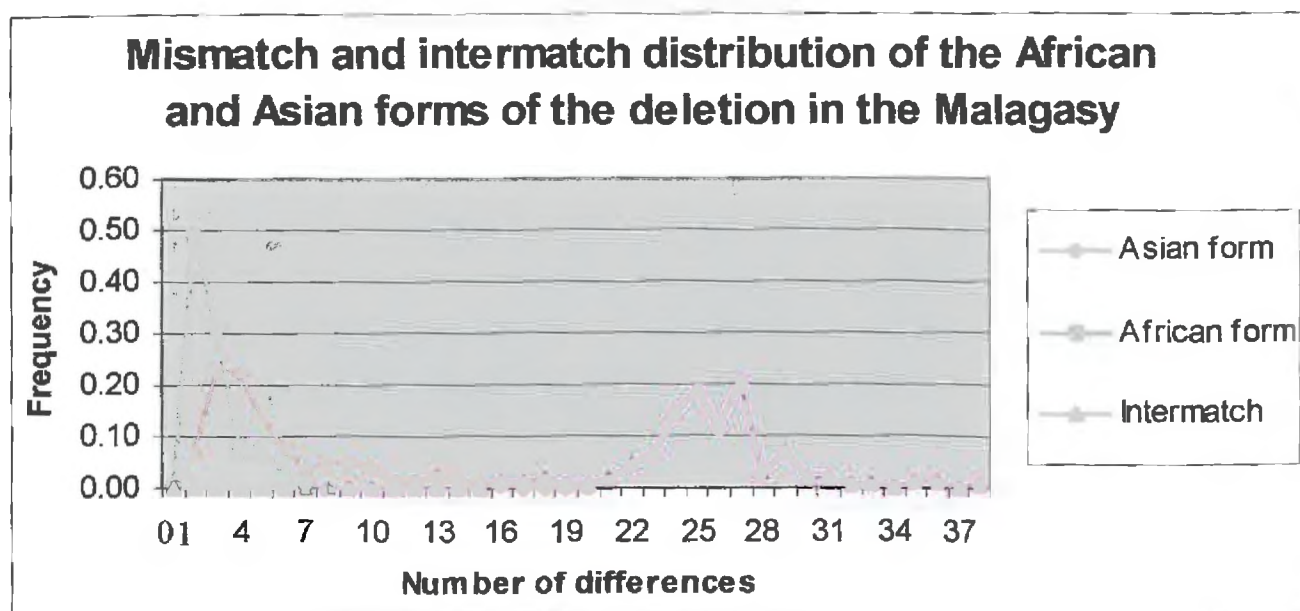


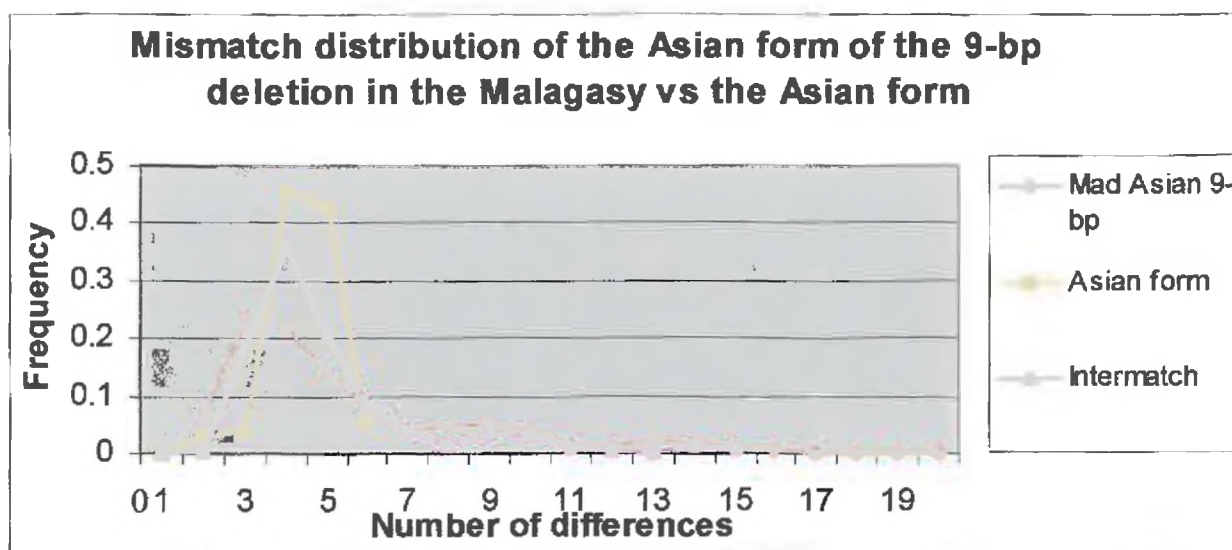
Figure15: Mismatch and intermatch (purple) distribution of the African (green) and Asian (orange) forms of the deletion in the Malagasy.

African curve: $\mu = 0.897$; $\delta = 1.392$; $\tau = 0.193$; $r = 0.092$

Asian curve: $\mu = 3.49$; $\delta = 10.14$; $\tau = 0.91$; $r = 0.197$

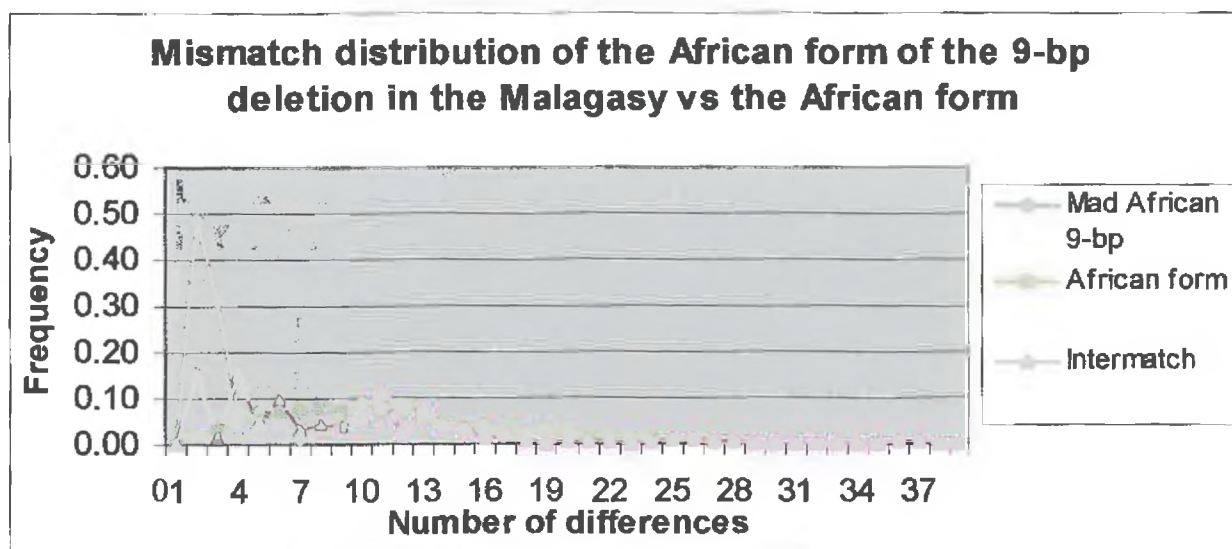
Intermatch curve: $\mu = 24.59$; $\delta = 11.33$; $\tau = 24.59$; $r = 0.244$

A



* Intermatch curve: $\mu = 3.08$; $\delta = 5.35$; $\tau = 1.57$; $r = 0.325$; divergence time = 8 752 years

B



* Intermatch curve: $\mu = 7.07$; $\delta = 44.99$; $\tau = 0.91$; $r = 0.271$; divergence time = 5 079 years

Figure 16: Mismatch and intermatch distributions of the Asian (A) and African (B) forms of the deletion in the Malagasy compared with Asians and Africans, respectively.

3.1.7 Hierarchical structure in the Malagasy

The data presented thus far reveals that there is no concordance between ethnicity and geographical placement of present-day Malagasy groups. This was further supported by AMOVA testing (Table 11).

In the first instance, we considered all ethnic groups together (Test 1). The AMOVA test revealed that 2.53% of the mtDNA variation found in the Malagasy is due to differences between groups whereas 97.47% of the variation was due to differences within groups. This result indicates that much more variation occurs within each ethnic group than between them and therefore suggests that the ethnic groups are relatively homogeneous.

In the second analysis we tested the validity of grouping the ethnic groups by regions (Test 2a). Firstly, we grouped the ethnic groups according to the geographic region that they occupy (i.e., Central Highland, Eastern Central, Northern, South Western and Western region). A percentage of variation of 2.63% among populations was calculated, compared with 0.06% among populations within groups. This means that there are few differences between the groups in each region. Furthermore there is more variation within each region than between them ($p = 0.09482$). Then we compared the highland groups (the Merina and Betsileo) to populations that occupy the lowland regions (Test 2b). In this case, a percentage of variation of 96.67% within each region was calculated. The intra group variance was 1.61% and the variance among populations within groups was 1.72%, again indicating that more variation exists within these regions than between them ($p = 0.13685$).

Therefore, through the use of AMOVA, we have shown that there is little validity, based on mtDNA, to support the ethnic and geographical distinctions employed on the island.

Finally, AMOVA was used to test the results obtained from the comparative component of the study (Test 3). The PC plot (**Figure 13**) indicated that the Merina are distinct from the remainder of the Malagasy groups. When this pattern was tested, we found that only 2.87% of the variance was due to inter population differences, whilst 1.57% was contributed by differences among populations within groups and 95.56% was due to variation within populations ($p=0.13490$). We found that the “within population” value was slightly lowered and this suggests that the mtDNA pool in the Merina is different to that of the other ethnic groups.

Table 11: Hierarchical structure in the Malagasy based on mtDNA variation.

TEST	SOURCE OF VARIATION		
	Among groups	Among populations within groups	Within populations
Test 1: all ethnic groups together	-	2.53%	97.47%
Test 2a: geographic regions	2.63%	0.06%	97.31%
Test 2b: highland vs. lowland	1.61%	1.72%	96.67%
Test 3: Merina vs. others	2.87%	1.57%	95.56%

3.1.8 Summary of MtDNA results in the Malagasy

The 9-bp deletion was present, throughout the island, at a total frequency of 22.5% (**Table 8**). The 1.1 kb mtDNA control region was sequenced from 228 Malagasy individuals. One hundred and twenty four unique mtDNA types, defined by 151 nucleotide changes were derived (**Appendix 4**).

Studies have shown that when the 9-bp deletion is used in conjunction with mtDNA control region sequence variation, the 9-bp deletion is a useful marker to trace population affinities. We were able to demonstrate that the frequencies of the Asian and African forms of the deletion in the total Malagasy sample were 14.8% and 7.7% respectively. The Polynesian motif accounted for 87.9% of those individuals with the Asian form (**Figure 8**). Furthermore, by making use of comparative data from the literature, we were able to demonstrate that of the 156 individuals without the deletion, 28.2% had an African ancestry, 54.5% had an Asian ancestry, and 17.3% remained unresolved. Therefore, the proportion of the Asian to African contribution in the Malagasy is approximately 2:1 (**Table 10**).

Through the use of NJ trees (**Figure 7, 9 and 11**) and PC plots (**Figure 12 and 13**), we found that there was a close genetic relationship between ethnic groups from the different geographic regions on the island. To test the ethnic and geographical structure employed on the island, we used AMOVA. These tests showed little biological support in favour of such hierarchies.

3.2 Y Chromosome results

Male lineages are likely to show different patterns of evolution from those provided by mtDNA analyses. Markers situated on the non-recombining region of the Y chromosome could provide a view of evolution which is complementary to that provided by mtDNA (Jobling and Tyler-Smith, 1995; Cooper *et al.*, 1996).

The YAP element is a useful marker as it has a single origin in human evolution, and its frequency varies among human populations (Hammer, 1994). The presence of the YAP insertion (YAP +) is indicative of an African ancestry (Hammer, 1994). Furthermore, this allele is not present in South East Asia (Hammer, 1994). However, a chromosome that does not have the insertion (YAP-) requires resolution with additional markers in order to establish its origins (Hammer, 1994). When the *Alu* insertion was present, we obtained a PCR fragment of 455-bp in size, while the absence of the insertion (YAP-) produced a fragment of 150-bp (**Figure 17**). Altogether, we found the YAP insertion (African) in 259 out of the 429 Malagasy individuals screened (**Table 12**). Thus, the insertion is present throughout the island at a frequency of 60.4%.

Of the individuals who were found to be YAP- (n=170), 159 were typed using M9 to resolve the origins of these chromosomes further. The M9 C allele is indicative of an African ancestry and the M9 G allele is found in all geographic regions except Africa (Hammer, 1994). When the M9 C allele was present, digestion of the 340-bp PCR product (**Figure 18**) with *HinfI* produced products of 245-bp and 95-bp in size (**Figure**

19). When the M9 G allele was present, products of 181-bp, 95-bp and 64-bp in size were produced (**Figure 19**). We found that 68 individuals (42.8%) had the M9 C allele and can thus be considered to be African in origin. The remaining 91 chromosomes (57.2%) had the M9 G allele and were therefore of non-African ancestry (**Table 12**).

Using the YAP in conjunction with M9, a total African contribution of 78.2% and a non-African contribution of 21.8% was calculated (**Table 12, Figure 20**).

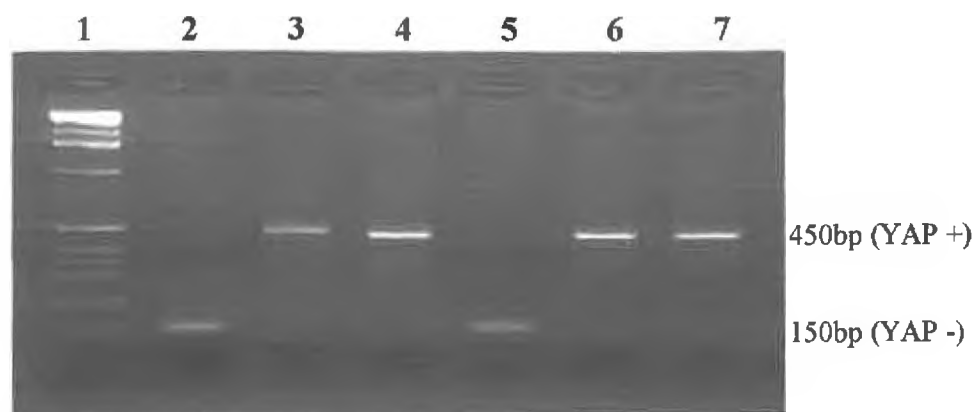


Figure 17: Electrophoretic separation of PCR products with (lanes 3, 4, 6 and 7) and without (lanes 2 and 5) the Y *Alu* sequence on a 2% HGT agarose gel. A 1-kb molecular marker was used to size the fragments (lane 1).

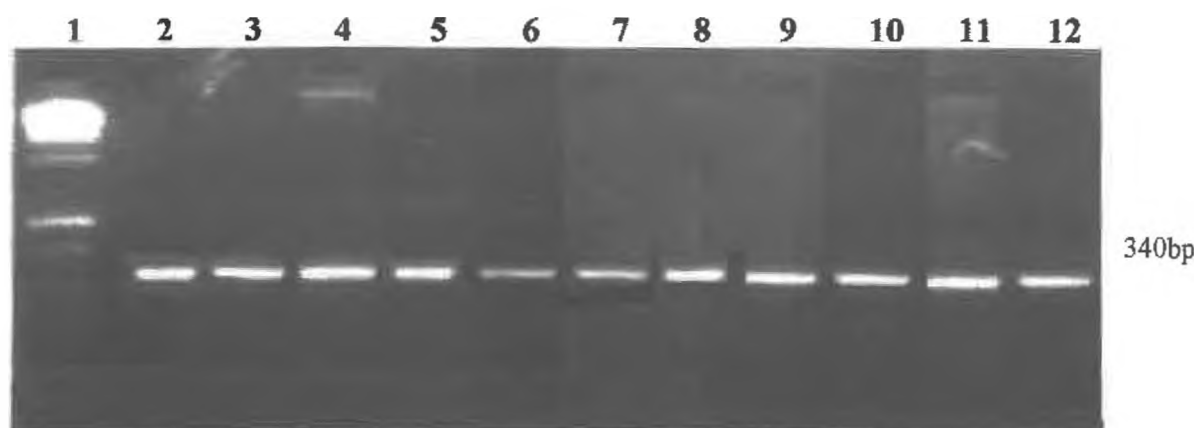


Figure 18: Electrophoretic separation of the 340-bp M9 PCR product on a 3% HGT agarose gel. A 1-kb molecular marker was used to size the fragments (lane 1).

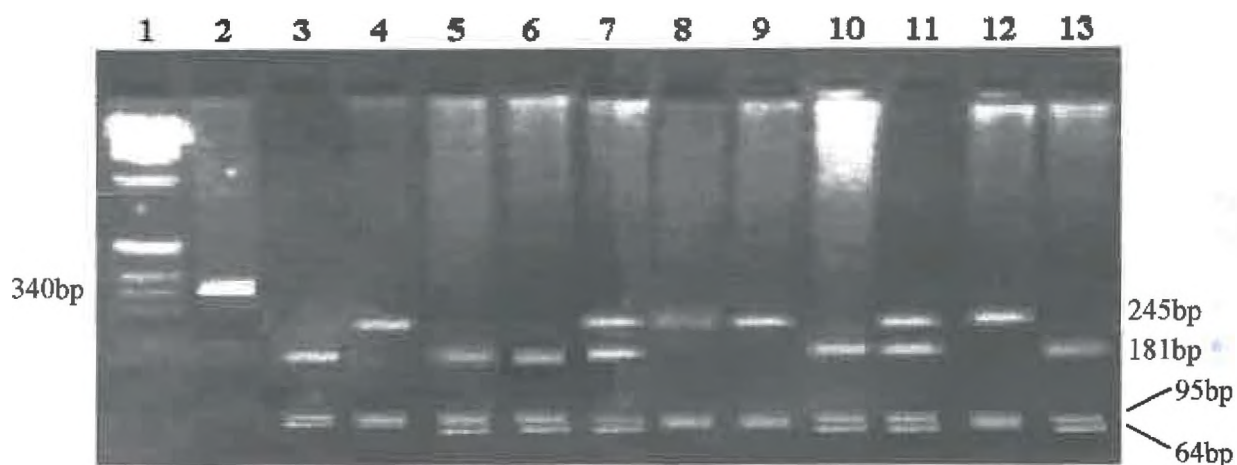


Figure 19: RFLP patterns produced by the digestion of the 340-bp M9 PCR product (lane 2) with *HinF1* on a 2% HGT agarose gel. A 'GG homozygote' produces fragments of sizes 181-bp, 95-bp and 64-bp in size (lanes 3, 5, 6, 10, and 13) and a 'CC homozygote' produces fragments of sizes 245-bp and 95-bp (lanes 4, 8, 9, and 12). Control samples are illustrated in lanes 7 and 11. A 1-kb molecular marker was used to size the fragments (lane 1).

Table 12: Number of individuals and frequency of the YAP and M9 in Malagasy ethnic groups in the different regions.

REGION / ETHNIC GROUP	YAP					M9				
	SAMPLE SIZE	YAP -		YAP +		SAMPLE SIZE	M9 C		M9 G	
		#	FREQ	#	FREQ		#	FREQ	#	FREQ
CENTRAL HIGHLAND	70	23	0.329	47	0.671	18	6	0.333	12	0.667
BETSILEO	28	6	.214	22	.786	3	0	0	3	1.000
MERINA	42	17	.405	25	.595	15	6	.400	9	.600
EASTERN CENTRAL	152	71	0.467	81	0.533	69	22	0.319	47	0.681
ANTAIMORO	56	25	.446	31	.554	24	6	.250	18	.750
ANTAISAKA	41	27	.659	14	.341	26	6	.231	20	.769
BETSIMISARAKA	55	19	.345	36	.655	19	10	.526	9	.474
NORTHERN	59	28	0.475	31	0.525	26	15	0.577	11	0.423
ANTANKARANA	13	4	.308	9	.692	4	1	.250	3	.750
TSIMIHETY	46	24	.522	22	.478	22	14	.636	8	.364
SOUTH WESTERN	105	29	0.276	76	0.724	27	11	0.407	16	0.593
ANTANDROY	46	12	.261	34	.739	12	6	.500	6	.500
MAHAFALY	23	8	.348	15	.652	7	4	.571	3	.429
MIKEA	34	9	.265	25	.735	8	1	.125	7	.875
VEZO	2	0	0	2	1	0	0	0	0	0
WESTERN	43	19	0.442	24	0.558	19	14	0.737	5	0.263
SAKALAVA	43	19	.442	24	.558	19	14	.737	5	.263
TOTALS:	429	170	.396	259	.604	159	68	.428	91	.572

= number; FREQ = frequency

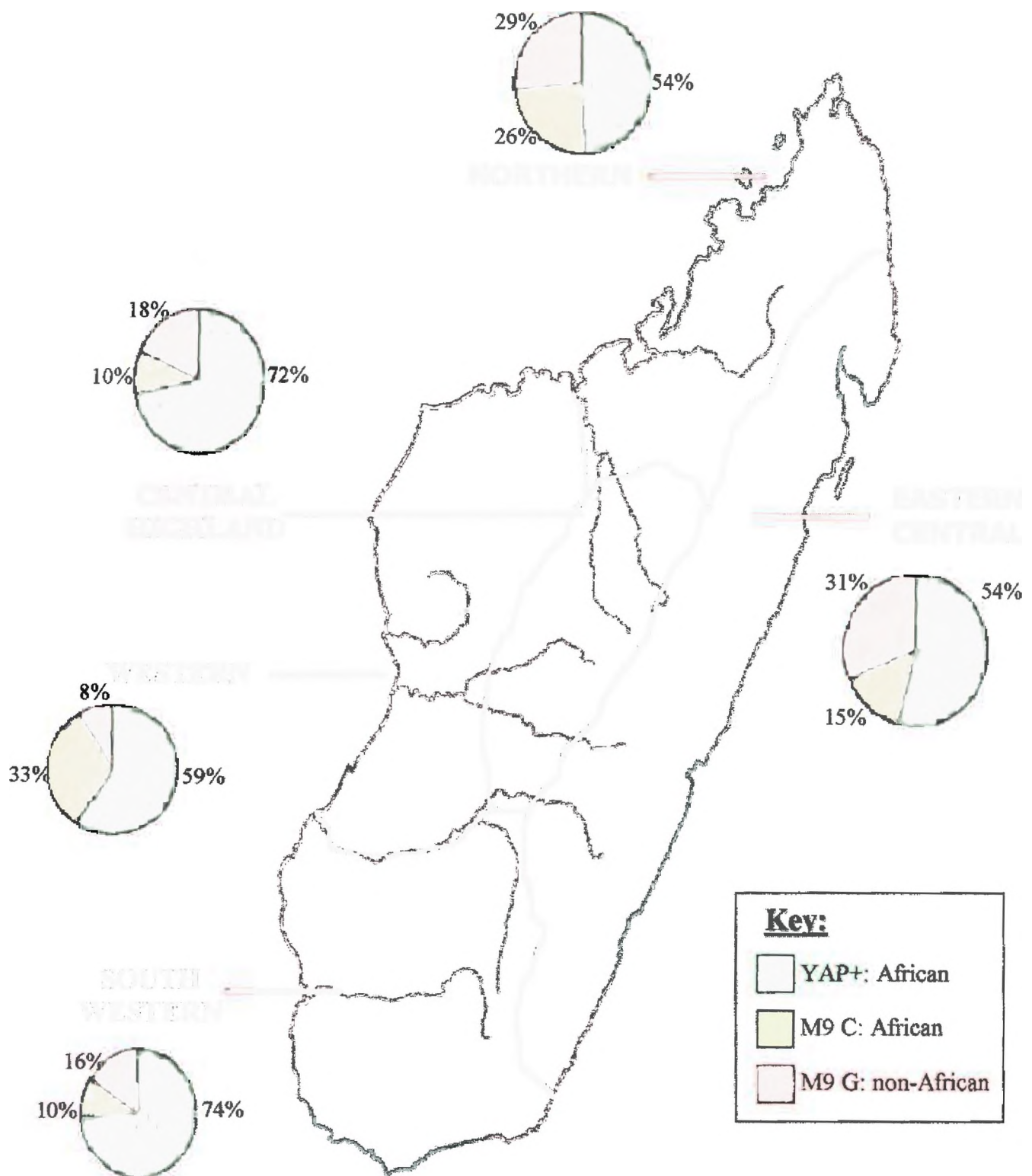


Figure 20: Map of Madagascar showing the distribution of the YAP and M9 throughout the island. Note that all M9C and M9G alleles are also YAP-. See key for details.

Morar (2000) used several Y chromosome markers to study the parental contribution to the Malagasy. Amongst the markers used were four short tandem repeats (STRs) namely DYS393, DYS19, DYS390, and DYS391. Using the M9 data obtained in the current study, we were able to resolve those haplotypes (based on Morar's STR data) that occurred on a YAP- background (**Table 13**). Out of a total of 57 YAP- haplotypes, 23 (40.35%) were found to occur on an African background, 31 (54.39%) were of non-African origin, and 4 haplotypes (7.02%) were found to occur both in Africans and non-Africans.

Table 13: STR haplotypes (constructed in the order DYS393, DYS19, DYS390, and DYS391) occurring on a YAP- background (Morar, 2000) that we were able to resolve based on our M9 data.

ETHNIC GROUP/ REGION	M9 C (African)	M9 G (Non-African)
CENTRAL HIGHLAND		
BETSILEO		FCHC
MERINA	FDEC, ECHC, DDGC	EDHC, ECGC, FCHC, EEHC, EEGD
EASTERN CENTRAL		
ANTAIMORO	ECEC, DCGC, DBED	EDGC, ECGC, ECID, FCID, FCHC, ECGD, EEGC, DBHD
BETSIMISARAKA	ECHC, DBED, EBGC, DDGB, ECGC, ECGD, EDIC	FCIC, ECHD, ECGC, EDGC, EBHD, EEFC, CAFD, CAFC
NORTHERN		
ANTANKARANA	DCHC	FBFB, EDGC, EDIIG
TSIMIHETY	EBHD, DBGC, ECEC, ECHC, EBGD, EAGC, EDHC, ECGC	FCID, EEGD, FCHC, GCID, FDHC, EEGC, GCID
SOUTH WESTERN		
ANTANDROY	ECHD, DBHD, FCID, DCHC, EDHC, DDHB, ECHC	DBHD, EDHC, CBFB, EDHC, FCHC
MAHAFAHY	DBED, ECHC, FCGD, EAGC	EEGC
MIKEA	DDHB	FCHC, EEGC, EDIC, ECID, GCHC, EDGC
WESTERN		
SAKALAVA	EEHC, ECHD, ECHC, EDIC, ECIC, FBHC, DBFC	GCID, EAIC, ECHC, FEGC, FCID, EEGC, EDGC, EEGD, DBGD, FCHD, EDFC

By combining our YAP and M9 data with Morar's (2000) STR data, we were able to examine the relationship between different ethnic groups through the use of PCA (**Figure 21**). We found no specific clustering patterns; ethnic groups from different geographic regions were located amongst each other on the plot suggesting that there is no concordance between ethnic groups and geographic region. The majority of ethnic groups clustered tightly towards the centre of the plot, but some (Betsileo, Antandroy, Antaisaka and Antaimoro) appeared to show some differences.

We used AMOVA to verify our results statistically (**Table 14**). In the first test, when all ethnic groups were considered together, a percentage of variation between populations of 2.60% and a percentage of variation within populations of 97.40% was calculated. This result indicates that there is more variation within each ethnic group than between them. In the second analysis, ethnic groups were grouped according to region (i.e., Central Highland, Eastern Central, Northern, South Western and Western region). We found that 1.56% of the variance was due to inter population differences, whilst 1.28% was contributed by differences among populations within groups and 97.16% was due to variation within populations ($p = 0.07331$). Furthermore, when highland groups (Betsileo and Merina) were compared to lowland groups (Test 3), we found that 96.54% of the variation could be attributed to inter group differences, 2.21% was due to differences among populations within groups and 1.25% was contributed by variation among groups ($p = 0.12610$). Again, these results suggest that more variation occurs within each region than between them indicating that Y chromosome data offer no support in favour of ethnic and geographic distinctions.

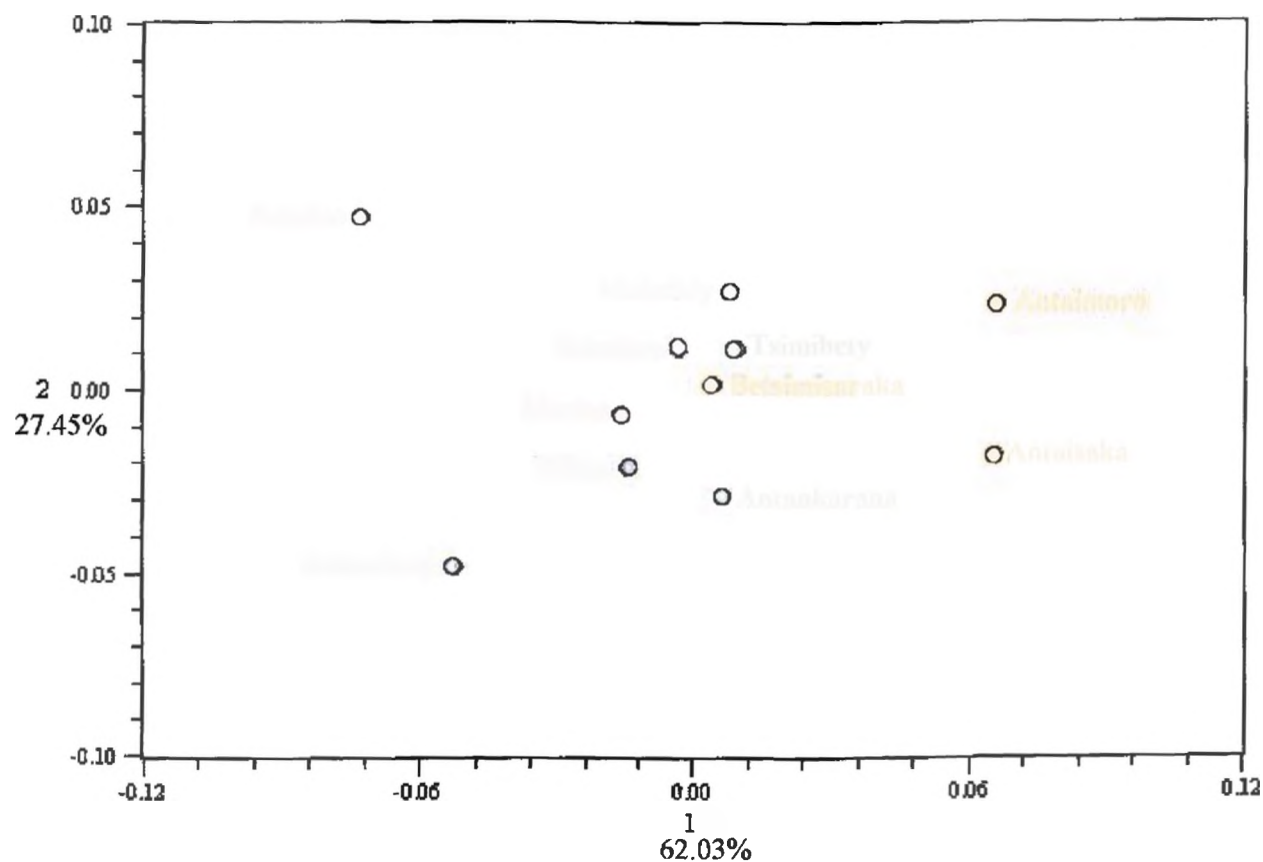


Figure21: Principal component plot showing the relationship between the Malagasy ethnic groups based on Y chromosome data. Orange text represents ethnic groups in the Eastern Central region, green text represents the Northern region, red text the Central Highland, pink text the Western region, and blue text the South Western region.

Table 14: Hierarchical structure in the Malagasy based Y Chromosome variation.

TEST	SOURCE OF VARIATION		
	Among groups	Among populations within groups	Within populations
Test 1: all ethnic groups together	-	2.60%	97.40%
Test 2: geographic regions	1.56%	1.28%	97.16%
Test 3: highland vs. lowland	1.25%	2.21%	96.54%

3.3 Autosomal DNA results

We used two forms of autosomal DNA markers (PSAs and the MBP) to examine nuclear DNA variation in the Malagasy. As noted earlier (Section 1.3.2), PSAs are markers that have varying frequencies in global populations. In this study, we made use of Sb19.3, AT3id and APO-4. The MBP, on the other hand, has implications in innate immunity (Section 1.3.2).

3.3.1 PSA results

When the SB19.3 insertion was present (allele 1), we obtained a PCR fragment 450-bp in size, while the absence of the insertion (allele 2) produced a PCR fragment of 122-bp (Figure 22). Altogether we found that 49 out of the 238 individuals screened (20.6%) were homozygous for the presence of the insertion, 134 (56.3%) lacked the insertion, and 55 individuals (23.1%) were heterozygous (Table 15). Therefore the frequency of the insertion throughout the island is 32.1%.

When the AT3id allele 1 was present, we obtained a PCR fragment 572-bp in size, while the presence of allele 2 yielded a fragment 496-bp in size (Figure 23). We found that out of a total of 188 individuals, 96 (51.1%) were homozygous for allele 1, 16 (8.5%) were homozygous for allele 2, and 76 (40.4%) were heterozygous (Table 16).

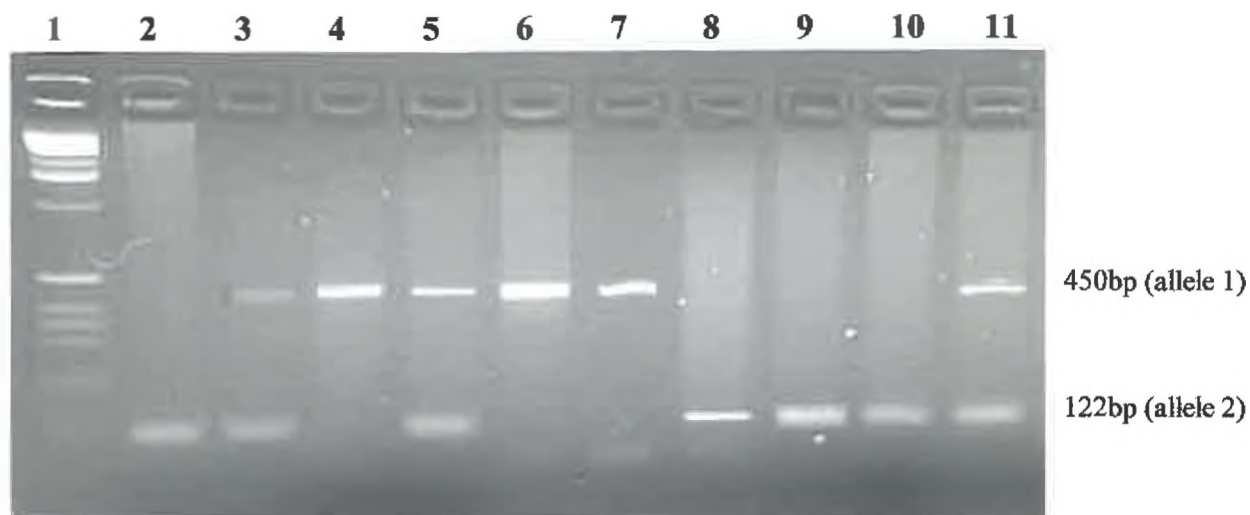


Figure22: Electrophoretic separation of Sb19.3 PCR products on a 3% HGT agarose gel. A '1/1 homozygote' produces fragments 450-bp in size (lanes 4, 6 and 7) and a '2/2 homozygote' produces fragments 122-bp in size (lanes 2, 8, 9, and 10). Heterozygotes are shown in lanes 3, 5 and 11. A 1-kb molecular marker (lane 1) was used to size the fragments.

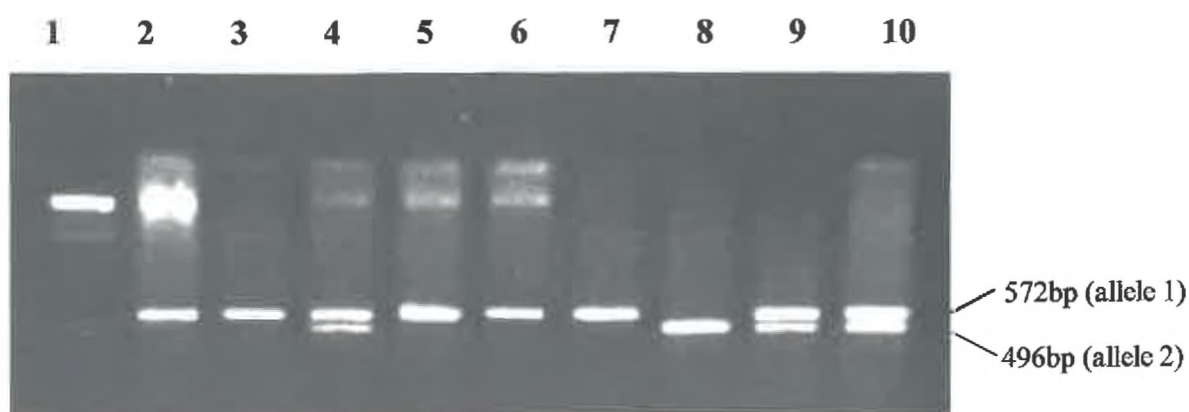


Figure23: Electrophoretic separation of AT3id PCR products on a 4% composite agarose gel. A '1/1 homozygote' produces fragments 572-bp in size (lanes 2, 3, 5, 6, and 7) and a '2/2 homozygote' produces fragments 496-bp in size (lane 8). Heterozygotes are shown in lanes 4, 9 and 10. A 1-kb molecular marker (lane 1) was used to size the fragments.

Table 15: Ethnic groups included in the Sb19.3 screen, the respective sample sizes as well as the genotypic results obtained.

REGION / ETHNIC GROUP	SAMPLE SIZE	HOMOZYGOUS 1/1		HETEROZYGOUS 1 /2		HOMOZYGOUS 2/2	
		#	FREQ	#	FREQ	#	FREQ
CENTRAL HIGHLAND	67	15	0.224	14	0.209	38	0.567
BETSILEO	30	9	0.300	10	0.330	11	0.370
MERINA	37	6	0.162	4	0.108	27	0.730
EASTERN CENTRAL	46	15	0.326	14	0.304	17	0.370
ANTAISAKA	46	15	0.326	14	0.304	17	0.370
NORTHERN	42	4	0.095	10	0.238	28	0.667
TSIMIHETY	42	4	0.095	10	0.238	28	0.667
SOUTH WESTERN	48	5	0.104	8	0.167	35	0.729
MAIAFALY	33	3	0.091	6	0.182	24	0.727
VEZO	15	2	0.133	2	0.133	11	0.734
WESTERN	35	10	0.286	9	0.257	16	0.457
SAKALAVA	35	10	0.286	9	0.257	16	0.457
TOTALS:	238	49	0.206	55	0.231	134	0.563

= number; FREQ = frequency

Table 16: Ethnic groups included in the AT3id screen, the respective sample sizes as well as the genotypic results obtained.

REGION / ETHNIC GROUP	SAMPLE SIZE	HOMOZYGOUS 1/1		HETEROZYGOUS 1 / 2		HOMOZYGOUS 2/2	
		#	FREQ	#	FREQ	#	FREQ
CENTRAL HIGHLAND	53	22	0.415	24	0.453	7	0.132
BETSILEO	21	6	0.285	14	0.667	1	0.048
MERINA	32	16	0.500	10	0.312	6	0.188
EASTERN CENTRAL	34	20	0.588	12	0.353	2	0.590
ANTAISAKA	34	20	0.588	12	0.353	2	0.590
NORTHERN	40	21	0.525	18	0.450	1	0.025
TSIMIHETY	40	21	0.525	18	0.450	1	0.025
SOUTH WESTERN	30	15	0.500	11	0.367	4	0.133
MAHAFALY	19	10	0.526	5	0.263	4	0.211
VEZO	11	5	0.455	6	0.545	0	0
WESTERN	31	18	0.580	11	0.355	2	0.065
SAKALAVA	31	18	0.580	11	0.355	2	0.065
TOTALS:	188	96	0.511	76	0.404	16	0.085

= number; FREQ = frequency

At the APO-4 locus, the presence of the insertion (allele 1) produced a PCR product of 437-bp, and in the absence of an insertion (allele 2) a fragment 122-bp in size was produced (**Figure 24**). When 157 Malagasy individuals were screened, we found that 81 individuals (51.6%) were homozygous for the presence of the insertion, 24 (15.3%) lacked the insertion, and 52 (33.1%) were heterozygous (**Table 17**). The frequency of the insertion throughout the island is therefore 68.2%.

Chi-squared goodness-of-fit tests (**Table 18**) were performed (for all three PSA systems) to assess whether the frequencies of alleles in the populations were in Hardy-Weinberg (HW) equilibrium. At a 95% significance level, it was found that, with the exception of the Vezo and the Tsimihety, all ethnic groups deviated from expected HW proportions with respect to Sb19.3; all groups were at HW equilibrium with respect to AT3id; and only the Tsimihety deviated from expected HW proportions with respect to APO. Several reasons (such as sampling bias, non-random mating, migration, mutation, genetic drift or selective advantage) may explain why some groups were seen to deviate from expected HW proportions.

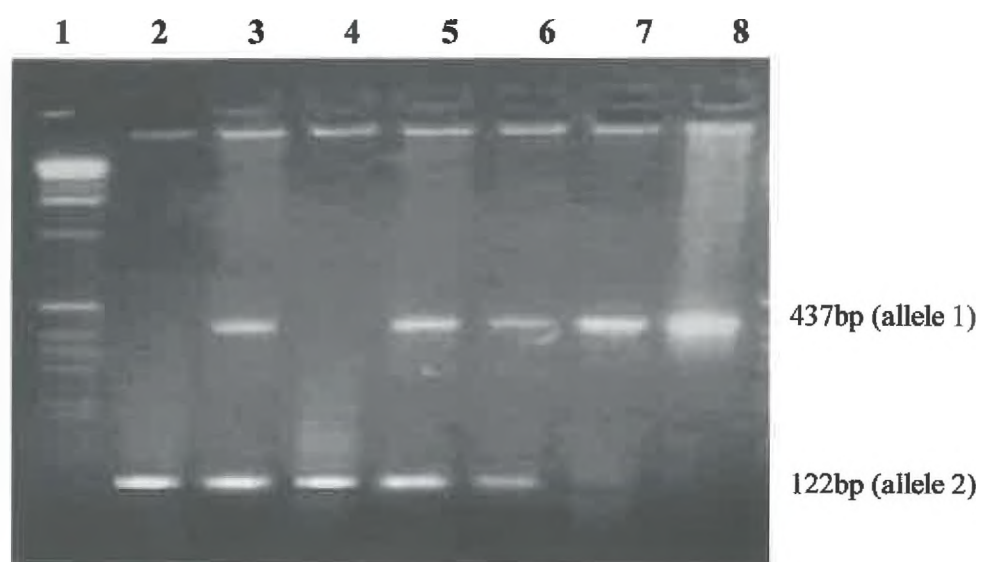


Figure24: Electrophoretic separation of APO-4 PCR products on a 2% HGT agarose gel. A '1/1 homozygote' produces fragments 437-bp in size (lanes 7 and 8) and a '2/2 homozygote' produces fragments 122-bp in size (lanes 2 and 4). Heterozygotes are shown in lanes 3, 5 and 6. A 1-kb molecular marker (lane 1) was used to size the fragments.

Table 17: Table showing the ethnic groups included in the APO-4 screen, the respective sample sizes as well as the genotypic results obtained.

REGION / ETHNIC GROUP	SAMPLE SIZE	HOMOZYGOUS 1/1		HETEROZYGOUS 1 /2		HOMOZYGOUS 2/2	
		#	FREQ	#	FREQ	#	FREQ
CENTRAL HIGHLAND	51	31	0.608	14	0.274	6	0.118
BETSILEO	22	12	0.545	6	0.273	4	0.182
MERINA	29	19	0.655	8	0.276	2	0.069
EASTERN CENTRAL	28	18	0.643	8	0.286	2	0.071
ANTAISAKA	28	18	0.643	8	0.286	2	0.071
NORTHERN	35	13	0.372	11	0.314	11	0.314
TSIMIHETY	35	13	0.372	11	0.314	11	0.314
SOUTH WESTERN	15	6	0.400	6	0.400	3	0.200
VEZO	15	6	0.400	6	0.400	3	0.200
WESTERN	28	13	0.464	13	0.464	2	0.072
SAKALAVA	28	13	0.464	13	0.464	2	0.072
TOTALS:	157	81	0.516	52	0.331	24	0.153

= number; FREQ = frequency

Table 18: Chi-squared (χ^2) and “p values” obtained for each Malagasy ethnic group with each of the PSAs included in the study.

REGION / ETHNIC GROUP	PSA SYSTEM					
	Sb19.3		AT3id		APO-4	
	χ^2	P	χ^2	P	χ^2	P
CENTRAL HIGHLAND						
BETSILEO	3.600	0.049	4.714	0.080*	10.364	0.076*
MERINA	46.568	0.0001	10.750	0.061*	25.759	0.259*
EASTERN CENTRAL						
ANTAISAKA	7.217	0.006	22.000	0.353*	23.429	0.271*
NORTHERN						
TSIMIHETY	38.952	0.057*	20.400	0.185*	5.057	0.021
SOUTH WESTERN						
MAHAFALY	40.091	0.042	8.053	0.069*	-	-
VEZO	18.867	0.055*	4.636	0.396*	1.800	0.311*
WESTERN						
SAKALAVA	10.314	0.005	19.129	0.354*	8.786	0.329*

* indicates groups that are in Hardy-Weinberg equilibrium

For analysis purposes, and to enable comparison with reported frequencies (**Table 1**), the frequency of allele 1 was calculated for each PSA system (**Table 19**). These data were used in conjunction with the data from the literature to produce the PC plot shown in **Figure 25**. On the first principal component (the x-axis), the Malagasy ethnic groups show close affinities with populations from Africa. The only exception to this are the Antaisaka who cluster between African and Caucasian populations. On the second principal component (the y-axis), the Mahafaly, Vezo, Sakalava and Merina appear to be very closely related to Asians. The Betsileo, however, show close affinities to Caucasians and the Antaisaka cluster more closely with Africans, as on the first axis. The Tsimihety, on the other hand, cluster midway between Asians and Africans. On the x-axis, there is no distinction between highland and lowland populations. But, on the y-axis, the highland groups clustered to the left of the distribution. Slight differences in the positioning of geographic regions was noted, but overall, no distinct clustering patterns were apparent indicating that the ethnic groups and geographic regions are very closely related.

Table 19: Number of individuals and frequency of the “1” allele in each ethnic group and geographic region for each PSA system.

REGION/ ETHNIC GROUP	Sb19.3			AT3id			APO-4		
	N	#	FREQ	N	#	FREQ	N	#	FREQ
CENTRAL HIGHLAND	134	44	0.328	106	68	0.642	102	76	0.745
BETSILEO	60	28	0.467	42	26	0.619	44	30	0.682
MERINA	74	16	0.216	64	42	0.656	58	46	0.793
EASTERN CENTRAL	92	44	0.478	68	52	0.765	56	44	0.786
ANTAISAKA	92	44	0.478	68	52	0.765	56	44	0.786
NORTHERN	84	18	0.214	80	60	0.750	70	37	0.529
TSIMIHETY	84	18	0.214	80	60	0.750	70	37	0.529
SOUTH WESTERN	96	18	0.188	60	41	0.683	30	18	0.600
MAHAFALY	66	12	0.182	38	25	0.658	-	-	-
VEZO	30	6	0.200	22	16	0.727	30	18	0.600
WESTERN	70	29	0.414	62	47	0.758	56	39	0.696
SAKALAVA	70	29	0.414	62	47	0.758	56	39	0.696
TOTAL ISLAND	476	153	0.321	376	268	0.712	314	214	0.682

N = sample size; # = number; FREQ = frequency

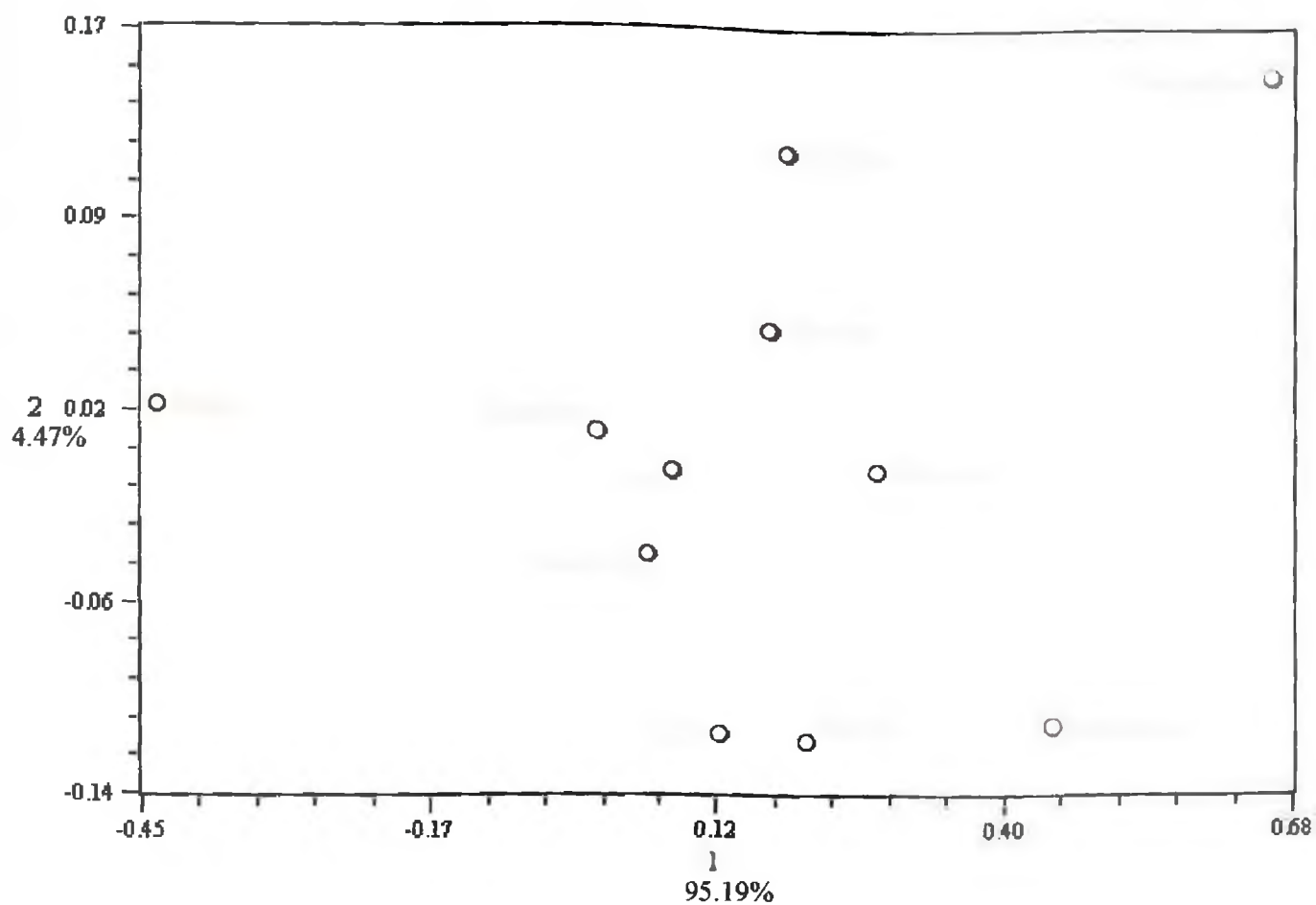


Figure 25: Principal component plot showing the relationship between the Malagasy ethnic groups (red) and various populations from the literature (Africa depicted in green, Asia in orange and Caucasians in blue), using PSA data.

We used AMOVA to test these findings statistically (**Table 20**). When all ethnic groups were considered (Test 1), a percentage of variation among populations of 3.14% compared with a percentage of variation within populations of 96.86% was calculated suggesting that the majority of variation occurs within each population rather than between them. When we grouped ethnic groups according to the geographic region (i.e., Central Highland, Eastern Central, Northern, South Western and Western region) they occupy (Test 2), we found that inter groups differences result in 2.83% of variation and that intra group variance accounts for 97.17% ($p = 0.88954$). Furthermore, when we compared highland (Betsileo and Merina) and lowland regions (Test 3), AMOVA indicated a percentage of variation of 2.97% among populations within groups and 97.03% within populations ($p = 0.43109$). These results confirm statistically that there are no significant differences between the ethnic groups or the regions they occupy.

Table 20: Hierarchical structure in the Malagasy based on variation contributed by PSAs.

TEST	SOURCE OF VARIATION		
	Among groups	Among populations within groups	Within populations
Test 1: all ethnic groups together	-	3.14%	96.86%
Test 2: geographic regions	0%	2.83%	97.17%
Test 3: highland vs. lowland	0%	2.97%	97.03%

3.3.2 MBP results

When the 349-bp MBP PCR product (**Figure 26**) was digested using *BanI*, allele A was detected, producing fragments of sizes 260-bp and 89-bp (**Figure 27**). Digestion with *MboII* detected allele C, which if present, produced fragments 279-bp and 70-bp in size (**Figure 28**). When *BanI* was used, we found that out of the 173 individuals screened, 162 individuals (93.1%) had chromosomes that remained undigested, 4 (2.3%) had chromosomes that were digested, and 8 (4.6%) were heterozygous (**Table 21**). When *MboII* was used, however, we found that out of 174 individuals 1 (0.6%) had chromosomes that remained undigested, 157 (90.7%) had chromosomes that were digested, and 15 (8.7%) were heterozygous (**Table 22**).

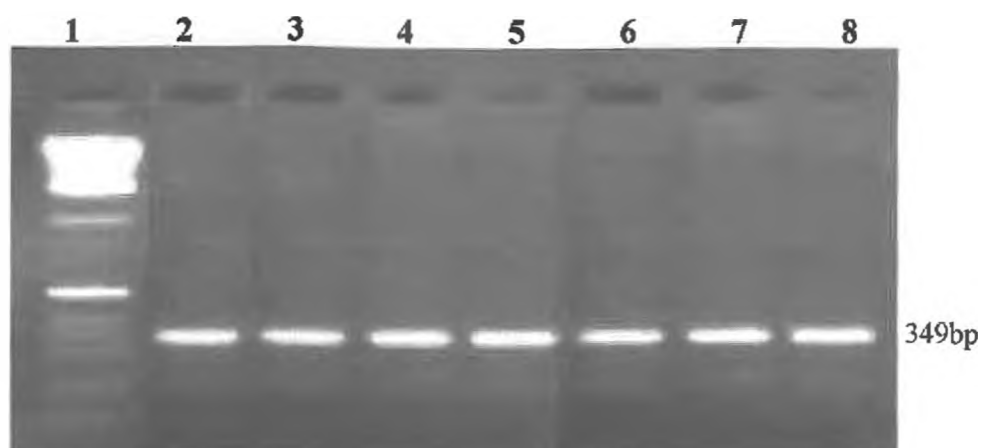


Figure 26: Electrophoretic separation of the 349-bp MBP PCR product on a 2% HGT agarose gel. A 1-kb molecular marker (lane 1) was used to size the fragments.

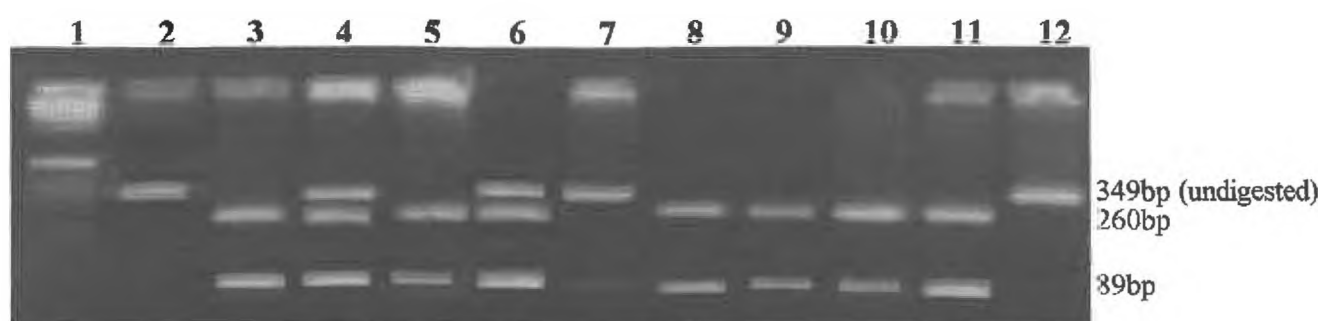


Figure 27: Electrophoretic separation MBP PCR products after digestion with *BanI* on a 2% Metaphor Agarose gel. The B allele remains uncut (lanes 2, 7, and 12) whereas the A allele produces fragments of sizes 260-bp and 89-bp (lanes 3, 5, 8, 9, 10, and 11). Heterozygotes are shown in lanes 4 and 6. A 1-kb molecular marker (lane 1) was used to size the fragments.

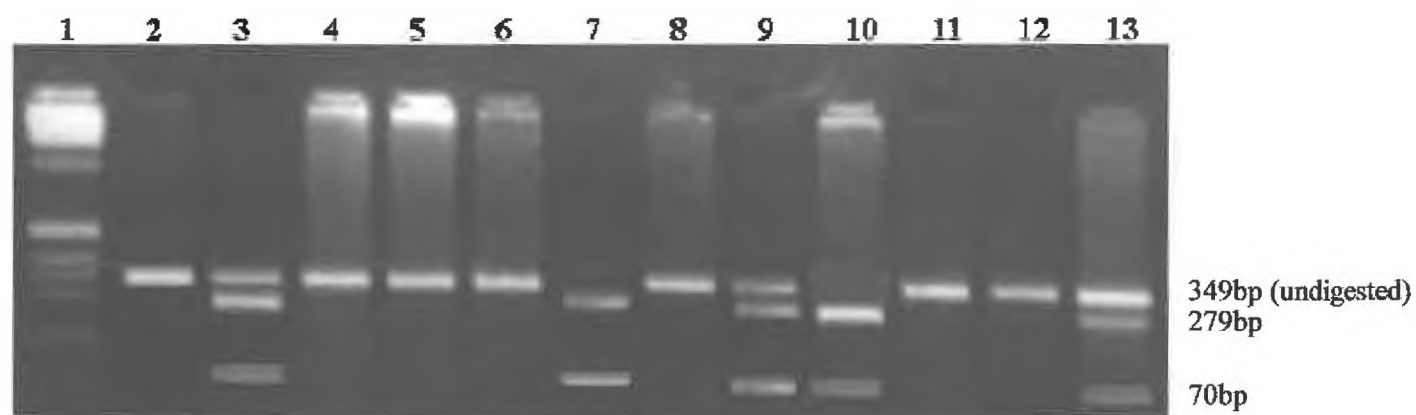


Figure 28: Electrophoretic separation MBP PCR products after digestion with *MboII* on a 2% Metaphor Agarose gel. The A allele remains uncut (lanes 2, 4, 5, 6, 8, 11, and 12) whereas the C allele produces fragments of sizes 279-bp and 70-bp (lanes 7 and 10). Heterozygotes are shown in lanes 3, 9, and 13. A 1-kb molecular marker (lane 1) was used to size the fragments.

Table 21: Ethnic groups, respective sample sizes and results obtained after digestion with *BanI*.

“C” = cut (allele A), “U” = uncut

REGION / ETHNIC GROUP	SAMPLE SIZE	HOMOZYGOUS C/C		HETEROZYGOUS C/U		HOMOZYGOUS U/U	
		#	FREQ	#	FREQ	#	FREQ
CENTRAL HIGHLAND	48	46	0.958	1	0.021	1	0.021
BETSILEO	17	17	1.000	0	0	0	0
MERINA	31	29	0.936	1	0.032	1	0.032
EASTERN CENTRAL	24	21	0.875	1	0.420	2	0.083
ANTAISAKA	24	21	0.875	1	0.042	2	0.083
NORTHERN	36	33	0.916	2	0.056	1	0.028
TSIMIHETY	36	33	0.916	2	0.056	1	0.028
SOUTH WESTERN	39	39	1.000	0	0	0	0
MAHAFALY	30	30	1.000	0	0	0	0
VEZO	9	9	1.000	0	0	0	0
WESTERN	27	23	0.852	4	0.148	0	0
SAKALAVA	27	23	0.852	4	0.148	0	0
TOTALS:	174	162	0.931	8	0.046	4	0.023

= number; FREQ = frequency

Table 22: Ethnic groups, respective sample sizes and results obtained after digestion with *Mbo*II.
“C” = cut (allele C), “U” = uncut

REGION / ETHNIC GROUP	SAMPLE SIZE	HOMOZYGOUS C/C		HETEROZYGOUS C/U		HOMOZYGOUS U/U	
		#	FREQ	#	FREQ	#	FREQ
CENTRAL HIGHLAND	47	0	0	3	0.064	44	0.936
BETSILEO	17	0	0	2	0.118	15	0.882
MERINA	30	0	0	1	0.033	29	0.967
EASTERN CENTRAL	24	1	0.042	0	0	23	0.958
ANTAISAKA	24	1	0.042	0	0	23	0.958
NORTHERN	36	0	0	5	0.139	31	0.861
TSIMIHETY	36	0	0	5	0.139	31	0.861
SOUTH WESTERN	39	0	0	2	0.051	37	0.949
MAHAFALY	30	0	0	2	0.067	28	0.933
VEZO	9	0	0	0	0	9	1.000
WESTERN	27	0	0	5	0.185	22	0.815
SAKALAVA	27	0	0	5	0.185	22	0.815
TOTALS:	173	1	0.006	15	0.087	157	0.907

= number; FREQ = frequency

To enable comparison with frequencies reported in the literature (Table 2), the frequency of the B allele (uncut by *BanI*) and the C allele (cut by *MboII*) were calculated (Table 23) for each group and region that was included in the study. The frequency of the B allele ranges from 0.11 to 0.16 in non-African populations (Garred *et al.*, 1992; Lipscombe *et al.*, 1996) but is either absent or present at very low frequencies (~0.009) in African populations (Garred *et al.*, 1992). We found that the Betsileo, Mahafaly and Vezo did not have the B allele at all indicating African affinities. However it was present in the other ethnic groups at frequencies ranging from 0.031 – 0.104 thereby indicating that these groups have non-African affinities. Furthermore, the C variant is found exclusively in Africa (Garred *et al.*, 1992) and we found that all ethnic groups (except the Vezo) have the C allele at low frequencies again suggesting an African ancestry.

Table 23: Table showing the frequencies of the B and C alleles for each of the groups and regions included in the MBP study.

REGION / ETHNIC GROUP	B ALLELE			C ALLELE		
	N	#	FREQ	N	#	FREQ
CENTRAL HIGHLAND	96	3	0.031	94	3	0.032
BETSILEO	34	0	0	34	2	0.059
MERINA	62	3	0.048	60	1	0.017
EASTERN CENTRAL	48	5	0.104	48	2	0.042
ANTAISAKA	48	5	0.104	48	2	0.042
NORTHERN	72	4	0.056	72	5	0.069
TSIMIHETY	72	4	0.056	72	5	0.069
SOUTH WESTERN	78	0	0	78	2	0.026
MAHAFALY	60	0	0	60	2	0.033
VEZO	18	0	0	18	0	0
WESTERN	54	4	0.074	54	5	0.093
SAKALAVA	54	4	0.074	54	5	0.093
TOTAL ISLAND	348	16	0.046	346	17	0.049

= number of chromosomes; FREQ = frequency

AMOVA was used to test the relationships between ethnic groups when different levels of grouping were applied (**Table 24**). When all ethnic groups were considered (Test 1), a percentage of variation among populations of 4.81% and a percentage of variation within populations of 95.196% was calculated. This result indicates that there is a lot more variation within each group than between them ($p = 0.98925$). When we grouped ethnic groups according to the geographic region they occupy (Test 2), 4.31% of variation could be attributed to differences among populations within groups and 95.69% was due to within population differences ($p = 0.98925$). Similarly when we compared highland and lowland regions (Test 3), we found that 0.45% of variation contributed by the MBP was due to variation among populations within groups that that 99.55% was contributed by differences within populations ($p = 0.90714$). These results show, once again, that more variation occurs within each ethnic group or region than between them indicating that the Malagasy are a homogeneous population.

Table 24: Hierarchical structure in the Malagasy based on variation contributed by the MBP.

TEST	SOURCE OF VARIATION		
	Among groups	Among populations within groups	Within populations
Test 1: all ethnic groups together	-	4.81%	95.19%
Test 2: geographic regions	0%	4.31%	95.69%
Test 3: highland vs. lowland	0%	0.45%	99.55%

3.4 *Correlation analysis*

For simplicity, the correlation between mtDNA and the Y chromosome markers is referred to as “correlation 1”, the correlation between mtDNA and autosomal DNA markers is referred to as “correlation 2”, and the correlation between the Y chromosome markers and autosomal DNA markers is referred to as “correlation 3” (Table 25).

Both correlation 1 and correlation 2 have a positive direction indicating that in each case, both marker systems increase or decrease in unison. The marker systems in correlation 3, however, have a negative direction indicating that these systems are in opposition to each other. The magnitude of each of these correlations is small: 0.136, 0.133 and 0.118 respectively for correlation 1, 2 and 3. This indicates that there is virtually no relationship between the different marker systems under comparison. In all cases, the probability of a chance deviation from a population probability of zero was positive further suggesting that the coefficient does not represent a relationship between the variables being correlated, but rather that the appearance of a relationship occurs by chance. The low correlation between haploid and biparental genetic markers from the same individuals may reflect differences in genome-effective population sizes or in sex-biased gene flow (Lum et al., 1998). It can be concluded that there is no significance between the results offered by the various haploid and diploid markers used in this study with respect to the present-day Malagasy; rather each of these markers suggest either a different history or is indicative of different maternal and paternal contributions.

Table 25: Summary Statistics generated in the correlation analysis

COMPARISON BETWEEN	SLOPE	Y INTERCEPT	r	P
mtDNA versus Y Chromosome	0.346788	0.038205	0.136734	0.755
mtDNA versus Autosomal DNA	0.1314735	0.17502	0.133592	0.697
Autosomal DNA versus Y Chromosome	-0.030446	0.038205	-0.118142	0.317

CHAPTER 4: DISCUSSION

The origins of the people of Madagascar, the Malagasy, are controversial and have still not been completely resolved. In the absence of written records, studies have been based on physical appearance, linguistics, ethnology, anthropology and archaeology. Such studies have indicated two major sources of ancestry to the Malagasy people, namely the African and Indonesian populations, with additional contributions from Arabian, Indian, Chinese and other Asian sources. However, due to the linguistic unity present on the island, these contributions are not easily distinguished. It was our intention to use genetic analyses to aid in elucidating the geographical origins of the parental populations of the Malagasy.

The goal of this study was to gain detailed knowledge of the molecular genetic variation of various kinds of DNA markers from the same individuals, thereby gaining new insights into the genetic structure and evolutionary history of human populations in Madagascar. These objectives have been achieved by examining the degree of concordance of the results of haploid (mtDNA and Y-chromosome) and autosomal (population specific alleles and the mannose binding protein) DNA markers. Following, is a discussion of the results obtained in this study. The arguments that follow have been debated in terms of the questions asked in **section 1.1**.

4.1 What is the extent of genetic diversity in the present-day Malagasy populations?

4.1.1 Mitochondrial DNA

The intergenic COII/tRNA^{Lys} 9-bp deletion in human mtDNA was found in 116 out of 515 Malagasy individuals (**Table 8**). Thus, the 9-bp deletion occurs at an observed frequency of approximately 22.5% in this population. In addition, the control region of the mtDNA genome from 228 Malagasy individuals was sequenced. From pairwise comparisons of both CR1 and CR2, 124 unique mtDNA types, defined by 151 nucleotide changes relative to the reference sequence (Anderson *et al.*, 1981), were derived (**Appendix 4**). Overall, the majority of the mutations were transitions (83.13%) while the remainder were transversions (16.87%).

Of those individuals with the 9-bp deletion, 14.8% had the Asian form of the deletion (of which 87.9% had the Polynesian motif) and 7.7% had the African form of the deletion. Of the 165 individuals who do not have the 9-bp deletion, 83 were associated with African types from the literature whilst 82 were associated with Asian types. Thus, an almost equal African and Asian contribution is indicated.

MtDNA variation within groups is influenced by the number of mtDNA types found within each group, and by the amount of sequence divergence between types within these groups (Soodyall *et al.*, 1996). By examining **Figure 7** and **Appendix 4**, it was found that mtDNA types from all ethnic groups were interspersed throughout the NJ tree and that

molecule to reflect a much more ancient history (Wilson *et al.*, 1985). Therefore results from work done on mtDNA should be interpreted cautiously and should be supplemented by data obtained from nuclear loci (Stoneking and Soodyall, 1996).

Madagascar is a perfect example of a hybrid population in which to study the evolution of the 9-bp deletion. This is due to the diverse nature of the population which has large contributions from both Africa and Asia, populations that have been separated from each other for thousands of years. This statement is supported by the fact that the mismatch distribution characterising mtDNA types with the 9-bp deletion (**Figure 14**) forms both the mismatch (“peak 1”) and intermatch distribution (“peak 2”).

4.1.2 Y Chromosome

Y chromosome variation in the Malagasy was studied by making use of the Y *Alu* polymorphism together with M9. Individuals who are found to be YAP+ have a high probability of being of African origin, and those who are found to be YAP- and also have the M9 C allele, can be presumed to be of African origin too. Thus, the YAP in conjunction with M9 is useful in distinguishing between African and non-African Y chromosomes. Those who are found to have the M9 G allele can be presumed to be non-African and remain unresolved without the use of additional markers.

The YAP insertion was found to occur at a frequency of 60.4% in 429 Malagasy individuals (**Table 12**). Of the chromosomes that were typed to be YAP-, 42.8% had the

M9 C allele (African) and 57.2% had the M9 G allele (non-African). An African contribution of 78.2% and a non-African contribution of 21.8% was calculated.

The African contribution was found to occur at highest frequency in the Western region. This observation is not surprising as the Sakalava who occupy that region are considered to be mainly of African origin (Brown, 1978). In contrast, the non-African contribution is strongest in ethnic groups occupying the Eastern Central region. Again, there is historical evidence to support this: the Antaimoro are considered to have a strong Arab/Islam influence and to this day use a form of *Sorabe* (Bradt, 1993), and the Betsimisaraka are thought to have obtained an ancient European influence from pirates (Bradt, 1993).

4.1.3 Autosomal DNA markers

Since both mtDNA and the Y chromosome have been utilised, both of which are haploid, three population specific alleles (Sb19.3, AT3id and APO-4) and the human mannose binding protein were used to provide an insight into the autosomal DNA found in the Malagasy population.

A total of 238, 188 and 157 individuals were screened with the Sb19.3, the AT3id and the APO systems respectively and 174 individuals were screened with the MBP system. Results in the form of genotype frequencies, obtained with each PSA system, are presented in **Tables 15-17** and the results obtained with the MBP system are presented in **Tables 21-22**. The comparative data used for these analyses are indicated in **Tables 1-2**.

At the Sb19.3 locus, the majority of the population (56.3%) were homozygous for the absence of the *Alu* insertion, with a relatively larger proportion of the population being heterozygous (23.2%) than homozygous for the presence of the insertion (20.6%). At the AT3id locus, the majority of the population was found to be either homozygous for allele 1 (51.1%) or heterozygous (40.4%), with only 8.5% being homozygous for allele 2. Similarly at the APO-4 locus, 52.6% of the individuals screened were homozygous for the presence of the *Alu* insertion, 33.1% were heterozygous and 25.3% were homozygous for the absence of the insertion.

Through the use of PCA (Figure 25), we were able to demonstrate that genetic affinities exist between all Malagasy ethnic groups and populations from Africa. In addition, the Mahafaly, Vezo, Sakalava and Merina also showed affinities with populations from Asia. The Mahafaly, Vezo and Sakalava are all thought to have a stronger African ancestry (Bradt, 1993; Brown, 1978) but the Merina are considered to be of Malayo-Polynesian origin (Singer *et al.*, 1957; Bradt, 1993). The Tsimihety, however, appeared to be equally related to Asians and Africans. This is interesting because the Tsimihety have been hypothesised as being the “original African-Indonesian hybrid group” who settled Madagascar (Brown, 1978).

The PSA results obtained in this study reinforce that these loci are population specific, with some alleles occurring more often in one population than in another. Further studies may indicate how powerful these PSAs could be in ethnic affiliation estimations. Owing to their frequency differentials between major geographic and ethnic groups, they have also become useful tools in admixture, anthropological, forensic and epidemiological

studies. Furthermore, the study of hybrid populations, such as the Malagasy, can be a valuable resource for mapping disease genes through a process referred to as “mapping by admixture linkage disequilibrium” (Shriver *et al.*, 1997).

Chi-squared goodness-of-fit tests were performed for each of the PSAs to assess whether the frequencies that were observed for each population are in Hardy-Weinberg (HW) equilibrium. The HW rule predicts the distribution of observed phenotypes in a diploid organism given the presence of certain conditions including random mating, absence of selection, mutation, migration and genetic drift. At a 95% significance level, it was found that, with the exception of the Vezo and the Tsimihety, all ethnic groups deviated from expected HW proportions with respect to Sb19.3; all groups were at HW equilibrium with respect to AT3id; and only the Tsimihety deviated from expected HW proportions with respect to APO.

Several reasons may explain why some groups were seen to deviate from expected HW proportions. Small sample size as well as sampling bias are recognised as being problematic. Also, there may be some as yet unknown, selective advantage to having a particular allele state at these loci. In addition, factors such as non-random mating, migration, mutation or genetic drift may lead to deviation from expected HW equilibrium.

The MBP component of the study revealed that the majority of the individuals typed are homozygous for the wildtype A allele both at codon 54 (93.1%) and at codon 57 (90.7%). Furthermore, it was observed that the MBP B and C alleles are present at a total frequency of 4.6% (out of 348 chromosomes) and 4.9% (out of 346 chromosomes), respectively.

These results indicate that although the majority of the population does not suffer from MBP deficiency, a small proportion of the sample are either heterozygous (4.6% for the B allele and 8.7% for the C allele) or homozygous (2.3% for the B allele and 0.6% for the C allele) for a deficiency causing mutation.

To date, MBP deficiency is the most frequent genetic form of immunodeficiency known (Garred *et al.*, 1992; Lipscombe *et al.*, 1996). The high prevalence of MBP deficiency among healthy individuals indicates that MBP deficiency also confers some selective advantages. Garred *et al.*, (1992) proposed that MBP deficiency is maintained in populations due to a positive selection pressure because MBP deficiency decreases the infectivity of some intracellular micro-organisms. The opposing negative effect of MBP deficiency (due to increased susceptibility to infection early in life) prevents the deficiency from reaching high frequencies (Garred *et al.*, 1992). As previously discussed, this phenomenon is evident in the Malagasy since a small proportion of the population, although otherwise healthy, do suffer from MBP deficiency.

Through the use of comparative data, it was found that with respect to the B allele, the Betsileo, Mahafaly and Vezo had African affinities, while the remainder of the ethnic groups were seen to have non-African affinities. In contrast to this, we found that all ethnic groups (except the Vezo) have the C allele at low frequencies suggesting an African ancestry. The Betsileo and Mahafaly were associated predominantly with Africans. The Betsileo have been described as being intermediate between Negroes and Malays (Brown, 1978; Campbell, 1996) and it is therefore expected that they should be associated with both populations. The Mahafaly are amongst the least known of the Malagasy ethnic

groups, but are thought to be African-like in appearance (Bradt, 1993); our results confirm this. As with the PSA studies, the Vezo were found to be more closely associated with non-Africans than with Africans. This is in contrast with historical data that suggests that the Vezo are a clan of the Sakalava and should therefore be associated with an African heritage (Brown, 1978). The Sakalava and Tsimihety are both thought to have African origins (Brown, 1978) and it is therefore surprising that non-African affinities were noted. The Merina, however, are thought to have an Asian ancestry (Brown, 1978) and it is therefore not surprising that non-African affinities were noted.

4.2 What are the relationships between the different ethnic groups throughout the island?

Based on analyses of mtDNA control region sequence variation, we have concluded that all ethnic groups as well as geographic regions are very closely related. This was indicated by the tight clustering of mtDNA types that was observed on all levels (**Figure 7** and **Appendix 3**), as well as the high degree of type-sharing between individuals from different ethnic groups and geographic regions. This close relationship was confirmed through the use PCA analysis and AMOVA, which indicated a higher degree of variation within each group than between them. This result was illustrated in **Figure 12** where the Sakalava, Antaisaka and Tsimihety – ethnic groups occupying different geographic regions - were seen to be very closely associated.

However, when one compares the frequency differentials that were obtained from both the haploid (9-bp, YAP and M9) and diploid markers used in the study (PSAs and MBP), slight differences between the groups are seen. Nevertheless, through the use of AMOVA and PCA, we were able to demonstrate that these markers also indicated more variation within each group than between groups.

When examined together, these results suggest that the Malagasy constitute a homogeneous population. Furthermore, they do not lend support to the ethnic and geographical distinctions employed on the island. While studies based on parameters such as physical appearance and anthropology do support such differences, there appears to be little genetic evidence in support thereof.

From **Appendix 4** and the NJ tree shown in **Figure 7**, we see a very close relationship between the ethnic groups from the different geographic regions. This is indicated by the interspersal of mtDNA types from all regions throughout the tree, as well as by the number of mtDNA types shared by individuals from different regions. This close relationship is emphasised by the PC plot shown in **Figure 12** where a close relationship between the ethnic groups from different regions was noted. One striking example is provided by the Sakalava, Antaisaka and Tsimihety who occupy different geographical regions on the island, but cluster to almost the same point on the plot.

4.3 Are there differences in the patterns of variation in highland versus lowland population groups?

Non-genetic studies have shown that Highland populations tend to be more ‘Indonesian-like’ whereas those living in the Lowland are more ‘African-like’. In order to test this hypothesis, a panel of genetic markers (Section 1.3) was used.

When looking at mtDNA sequence types in entirety, an approximately equal African to Asian contribution is noted in both the Highland and the Lowland regions. When one considers the 9-bp deletion types only, African and Asian contributions are also seen in both regions. Similarly, Y chromosome studies indicated that both regions showed high frequencies of the African YAP+ allele. Infact, the Highland people were seen to have a higher frequency of the YAP+ (67.1%) than people occupying the Lowland regions (58.5%). These findings are in contrast to what we would expect to find based on other types of studies. Analyses with autosomal markers confirmed these results by also displaying African and Asian affinities in both populations. It can be concluded that both regions posses both African and Asian contributions, and that biologically, neither region is specifically ‘African-like’ nor ‘Indonesian-like’.

When one draws comparison between the Highland and Lowland regions as a whole, slight differences are noted. These regions were found to be significantly different with regards to AT3id, APO and MBP. Higher frequencies of the 9-bp deletion, including a greater Polynesian contribution, were seen in the Highland regions. However, when one considers the results generated by AMOVA for each of the marker systems used, again

significantly more variation within each group than between groups was noted. This indicates that populations from the Highland and Lowland regions are in fact, very closely related.

4.4 To what extent have females and males contributed to the gene pool of the Malagasy?

Using the 9-bp deletion together with mtDNA control region sequence variation, we were able to estimate the proportion of African and Asian mtDNA types in the present-day **female** Malagasy gene pool. Of those individuals with the 9-bp deletion, an Asian to African contribution of 2:1 was derived. When the non 9-bp deletion mtDNA types were compared to Asian and African types from the literature, an almost equal contribution was noted.

Using the YAP in conjunction with M9, we were able to estimate the proportion of African and non-African contribution to the present-day **male** Malagasy gene pool. A total male African contribution of 78.2% and a non-African contribution of 21.8% was calculated thereby indicating that the Malagasy have a predominantly African (possibly sub-Saharan) ancestry rather than an Asian or Oceanic heritage.

Thus, we do see differences in the contribution made to the present-day Malagasy by their maternal and paternal ancestors. Maternally, it appears as if the Asian to African contribution is approximately 2:1. Of the Asian contribution, it appears as if a major

proportion was contributed by Indonesians. Paternally, we resolved a different picture; male ancestry shows that the African to non –African contribution is about 4:1.

These differences may be due to the potentially greater variance in male reproductive success, hitchhiking associated with selective sweeps and, most importantly, sexual differences in patterns of migration and population mixing (Cooper *et al.*, 1996; Dekka, 1996; Mitchell and Hammer, 1996). It has also been suggested that mtDNA data are expected to show less influence from recent historic male-mediated admixture (Redd *et al.*, 1999) than the Y chromosome.

4.5 What are the genetic affinities of Malagasy, African and Indonesian populations?

Since the 9-bp deletion has independent origins in Asians and Africans (Soodyall *et al.*, 1996; Redd *et al.*, 1995), the associated mtDNA types which are exclusive to Asians and Africans were used to estimate the proportion of mtDNA genes contributed by Asians and Africans respectively to the maternal gene pool of the Malagasy. The occurrence of the Asian and African form of the deletion was estimated to be approximately 2:1. Similarly, the Y Chromosome was used as an index of African and non-African contribution in the male Malagasy gene pool. In contrast to mtDNA studies, Y Chromosome markers indicated an overall African to non-African contribution of about 4:1.

The Polynesian motif was found to be present in 87.9% of the Malagasy individuals who have the Asian form of the 9-bp deletion (**Figure 8**). The Polynesian motif occurs in some parts of Borneo and East Indonesia at low frequencies, but it is not found in Indonesians from the Barito River area (Melton *et al.*, in press). This suggests a lack of concordance between the linguistic and genetic evidence concerning the origins of the Malagasy. A likely explanation for these results is that the Polynesian motif could have been introduced into the Malagasy either directly from Polynesia or from the founding population of both Madagascar and Polynesia which originated in Indonesia. From these data, it is not possible to establish which migration did in fact take place, but the latter is more likely because the mtDNA types that are not associated with the deletion are similar to some mtDNA types found in Indonesia.

The Eastern Central region was seen to have an equal Asian (93.4% Polynesian) and African contribution with regards to the 9-bp deletion, and an African to non-African contribution of approximately two thirds to one third with regards to the Y chromosome markers. Furthermore, the MBP B allele indicated Chinese affinities for the Antaisaka. We would expect to find a greater African ancestry for this region (Bradt, 1993) although certain ethnic groups who live there (such as the Antaimoro) have been shown to have strong Arab/Islam influences (Singer *et al.*, 1957).

In the Western region, the 9-bp deletion indicated a Polynesian to African contribution of 4:1 whereas the YAP and M9 together indicated an African to non-African contribution of approximately 9:1. This region is occupied by the Sakalava who would be expected to exhibit a larger African affinity (Bradt, 1993). We found this to be true with respect to the

SB19.3 locus where frequencies within the African range were noted. We found, however, that for the MBP B allele, the Sakalava had frequencies within the range expected for Papua New Guinea. This finding lends support to Bradt's (1993) observation that "Polynesian" features have been noted in these people.

The South Western ethnic groups were found to have a 9-bp deletion Polynesian to African contribution of about 3:1, and with regards to the Y chromosome markers, the African contribution was five and a half times greater than the non-African contribution. Furthermore, both the MBP B and C allele frequencies were found to be within the African range. This is to be expected because both the Mahafaly and Vezo who occupy this region are thought to be more African-like (Bradt, 1993).

With regards to the Northern region, an equal Polynesian and African contribution was indicated by the 9-bp deletion screen. The Tsimihety who live in this region have been hypothesised as being the "original African-Indonesian hybrid group" and so this finding is not surprising. Y chromosome studies showed an African to non-African contribution of 4:1. The Tsimihety were found to have MBP C allele frequencies comparable to those of the San Bushmen (Lipscombe *et al.*, 1992). Again, an African affinity is not surprising.

In the Highland region, the 9-bp deletion indicated an Asian to African contribution of 4:1 whereas the YAP and M9 together indicated an African to non-African contribution of approximately 3½: 1. This is consistent with the theory of a stronger Indonesian influence in the Merina and Betsileo who occupy the Central Highland. This was also supported by the data obtained from the MBP component of the study. However, the Betsileo showed

some affinities with the San Bushmen. With the PSAs, both the Merina and Betsileo showed affinities with African pops. As mentioned, this result is surprising as it contradicts non-genetic studies, but again lends support to our findings that both African and Asian contributions are present in all ethnic groups on the island.

Overall, the observation that both African and Asian ('non-African' in the case of the Y chromosome) forms of all markers included in this study are found throughout the island suggests that both the African and Asian contributions have co-existed in Madagascar for many centuries.

4.6 Is it possible to trace the genetic trail of migration to Madagascar?

The results obtained from the 9-bp deletion screen show a lack of concordance between linguistic and genetic data. This contradiction occurs as a result of the high frequency of the Polynesian motif in the Malagasy. The Polynesian motif is not found in Indonesians from the Barito River area (Redd *et al.*, 1995) which, based on linguistic evidence (Dahl, 1951), is the hypothesised geographic origin of the Malagasy. However, the Polynesian motif does occur elsewhere in Borneo as well as in east Indonesia at low frequencies (Redd *et al.*, 1995). From these data, it is not possible to determine whether the Polynesian motif was introduced to Madagascar directly from Polynesia or from an ancestral Polynesian population in Indonesia.

The Polynesians were skillful ocean navigators and prior to 1 500 AD they constituted the most geographically widespread people on earth (Bellwood, 1979). MtDNA types related to the Polynesian motif are highest in frequency in the corridor from Taiwan south through the Philippines and east Indonesia (Melton *et al.*, 1995). This finding is consistent with linguistic evidence of a Taiwanese origin for the proto-Polynesian expansion, which spread throughout Oceania by way of Indonesia (Melton *et al.*, 1995). This information lends support to our second proposition of an ancestral Polynesian population in Indonesia.

The absence of the Polynesian motif and mtDNA types found in peoples of Asian descent from Indian (Melton *et al.*, 1995) and African populations (Soodyall *et al.*, 1995) lends support to the hypothesis that Madagascar was colonised directly from the South Pacific region (Verin, 1986) rather than to the suggestion of an indirect route via India and Africa (Deschamps, 1963). Furthermore, several studies have documented gene flow from southeast Asia into Oceanic gene pools where Austronesian languages are spoken (Lum *et al.*, 1998), lending further support to a route of colonisation from the south Pacific Region.

Y chromosome results lend support to the hypothesised indirect route of colonisation via Africa (Deschamps, 1963), thus reflecting a remarkably different picture to that presented by mtDNA. A remarkable influence on the male gene pool is likely due to the importation of male African slaves that continued until slavery was abolished in 1886 (Heseltine, 1971). This discrepancy emphasises the fact that significant differences exist between male and female contributions to the Malagasy gene pool.

CONCLUSION

Through the use of data gathered from several different genetic systems (mtDNA, the Y chromosome and a variety of autosomal DNA markers), we have been able to test and refine hypotheses regarding Malagasy origins.

We have demonstrated that all Malagasy ethnic groups, as well as people living in all the geographic regions, are very closely related and showed that there is no concordance between ethnic and geographical stratification. Furthermore, we were able to show that the Malagasy are a relatively homogeneous population. Similarly, we showed that highland (the Merina and Betsileo) and lowland populations are also closely related. Although non-genetic studies have been claimed to show that differences do exist between these groups, we found little evidence in support of such claims.

We also showed that differences exist between maternal and paternal contributions to the present-day Malagasy: mtDNA data indicated an Asian to African contribution of 2:1 whereas the Y chromosome suggested a non-African to African contribution of 1:4. These results illustrate the advantages of combining mtDNA data with Y chromosomal results from the same individuals.

Finally, we were able to refine theories concerning the trail of migration to Madagascar. MtDNA analyses indicated support for the hypothesis that Madagascar was colonised by an ancestral Polynesian population in Indonesia (Verin, 1986). However, Y chromosome

studies lend support to the hypothesised indirect route of colonisation via Africa (Deschamps, 1963). These results reflect the differing male and female contributions to the present-day Malagasy. However, due to the fact that both African and Asian forms of all markers included in the study are found throughout the island, we can conclude that both African and Asian contributions have co-existed on the island for a many centuries.

APPENDIX 1: Ethics clearance certificate

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

Division of the Deputy Registrar (Research)

COMMITTEE FOR RESEARCH ON HUMAN SUBJECTS (MEDICAL)

Ref: R14/49 Dorf

CLEARANCE CERTIFICATE

PROTOCOL NUMBER M980606

PROJECT

Molecular Genetic Variation In The Malagasy

INVESTIGATORS

Miss C Dorf

DEPARTMENT

Dept of Human Genetics, SAIMR

DATE CONSIDERED

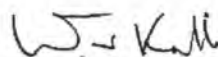
980628

DECISION OF THE COMMITTEE *

Approved unconditionally

DATE 980629

CHAIRMAN



(Professor P E Cleaton-Jones)

* Guidelines for written "Informed consent" attached where applicable.

c c Supervisor: Prof T Jenkins

Dept of Dept of Human Genetics, SAIMR

Works2\ainc0015U-umEth97.wdb\M 980606

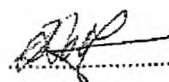
DECLARATION OF INVESTIGATOR(S)

To be completed in duplicate and **ONE COPY** returned to the Secretary at Room 10001, 10th Floor, Senate House, University.

I/we fully understand the conditions under which I am/we are authorized to carry out the abovementioned research and I/we guarantee to ensure compliance with these conditions. Should any departure to be contemplated from the research procedure as approved I/we undertake to resubmit the protocol to the Committee.

DATE 6/6/98

SIGNATURE



PROTOCOL NO.: M 980606

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

APPENDIX 2: Solutions used

x% Agarose Gel (100ml)

Mix: xg Agarose powder

100ml 1X TBE

Stir and boil simultaneously until the mixture is completely dissolved and appears translucent. After cooling the solution, add 3 μ l Ethidium Bromide.

10% Ammonium Persulphate (APS) (10ml)

Add 1g APS to 10ml dH₂O, at room temperature, and stir until dissolved. Store in a darkened container.

10mg/ml Bovine Serum Albumin (BSA) (10ml)

Dissolve 1g BSA in 10ml dH₂O. Make 1ml aliquots and store at -20°C.

40% Deionised Acrylamide-bisacrylamide Stock Solution (200ml)

Mix: 76g 40% Acrylamide

4g N,N'-methylene bisacrylamide

5g amberlite MB-6 ionic resin

~ 100ml ddH₂O

Stir at room temperature for 45 minutes to dionize. Filter through Watman No.1 filter paper to remove the amberlite. Make up to 200ml with ddH₂O. Store in a darkened container at 4°C for up to two weeks.

4.3 Denaturing Polyacrylamide gel (100ml)

Mix: 10.6ml 40% deionised acrylamide/N,N'-methylene bisacrylamide

10ml 10X TBE

36g urea

~50ml ddH₂O

Stir while applying heat until the urea has dissolved and the solution is translucent. Filter using the Nelgene® filtration system to remove any solid particles. Store at 4°C for up to one week in a darkened container. Just before use, add 30µl TEMED and 250µl APS to 50ml of the gel mixture.

1.25mM Deoxyribonucleic Triphosphate (dNTP) Solutions

Each 100mM commercial dNTP stock solution was diluted to 10mM to make a working stock solution. To make 320µl of a 1.25mM dNTP mix, 40µl of each 10mM solution (for dATP, aGTP, dCTP, dTTP) was added to 320µl ddH₂O. The solution was stored at -20°C.

Dextran Formamide Dye (1ml)

Dissolve 20mg Dextran in 1ml Formamide. Store at 4°C.

Ethidium Bromide (EtBr) (10mg/ml)

Add 100mg of EtBr to 10ml dH₂O and shake overnight to dissolve. Store in a darkened container at 4°C.

Ethylenediaminetetra-acetic acid (EDTA) (500ml)

Add 96.03g EDTA (disodium salt) into 470ml dH₂O and stir at room temperature. pH with 5M NaOH to pH 7.0 / 8.0 as needed. When fully dissolved, make up to 500ml with dH₂O and autoclave.

Ficoll Bromophenol Blue Loading Dye (100ml)

Mix: 50g sucrose

10ml 0.5M EDTA pH 7.0

0.1g Bromophenol blue

10g Ficoll

Allow to dissolve overnight and then filter through Watman No. 1 filter paper. Make up to 100ml and autoclave. Store at 4°C for a few weeks, or at -20°C for longer.

Formamide (100ml)

Stir 0.5g amberlite MB-6 resin and 100ml formamide at room temperature for 30 minutes.

Filter twice through Whatman No. 1 filter paper. Store at -20°C.

80% Isopropanol (100ml)

Add 80ml 100% Isopropanol to 20ml dH₂O and mix thoroughly.

1M MgCl₂ (250ml)

Add 50.83g MgCl₂ to 250ml dH₂O and autoclave.

1kb Molecular Weight Marker

The Boehringer Mannheim standard stock solution of DNA molecular weight marker (shown in **Appendix 1**) contained 0.25 μ g/ μ l DNA in 10mM Tris-HCl and 1mM EDTA (pH 8.0). To prepare a working solution, 10.9 μ l stock solution, 10 μ l Ficoll-bromophenol blue dye and 84.0 μ l 1 X TBE buffer were mixed together. The solution was stored at 4°C. When required for agarose gel electrophoresis, 5 μ l was loaded into one well to provide a visible size marker.

Primer Mixes

10 μ M primers: Determine the concentration of stock primers spectrophotometrically and calculate its concentration in μ M. Dilute with ddH₂O and store at -20°C. Store concentrated primers at -70°C.

3.3 μ M primers: Add 66 μ l of 10 μ M primer to 134 μ l ddH₂O to produce 200 μ l of 3.3 μ M primer.

Proteinase-K Solution (for 16 extractions)

Mix: 800 μ l 10% SDS

32 μ l 0.5M EDTA

5.6ml ddH₂O

Mix ingredients and add 1600 μ l Proteinase-K (2mg/ml) just before use.

Saturated Sodium Chloride (100ml)

Add 40g NaCl slowly to 100ml sterile H₂O, while stirring, until the mixture becomes saturated. Agitate before use and allow NaCl to precipitate out. Use the aqueous component.

10% Sodium Dodecyl Sulphate (SDS) (1l)

Add 100g SDS to 1l autoclaved dH₂O and stir at room temperature.

Sucrose Triton X (1l)

Mix: 10ml 1M Tris-EDTA (pH 8.0)

5ml 1M MgCl₂

10ml Triton X-100

Add ingredients and make up to 1l with dH₂O, autoclave and store at 4°C. Just before use, add 109.5g sucrose. The solution should not be stored once the sucrose has been added.

1 X TE Buffer (500ml)

Add: 5ml 1M Tris-HCl (pH 8.0)

1ml 0.5M EDTA

Make up to 500ml with dH₂O and autoclave.

1 X TAE buffer (1 000ml)

Add 20ml 50 X TAE to 980ml dH₂O.

50 X Tris Acetate EDTA (TAE) Buffer (500ml)

Mix: 121g Tris

28.55ml glacial acetic acid

50ml 0.5M EDTA, pH8.0

Allow to dissolve and make up to 500ml with dH₂O.

10 X Tris Borate EDTA (TBE) Buffer (1l)

Mix: 108g Tris base

55g Boric acid

9.3g Na₂EDTA

Dissolve in 1l dH₂O and autoclave.

1 X TBE buffer (1l)

Add 100ml 10 X TBE to 900ml dH₂O.

1M Tris-HCL (pH 8.0) (1l)

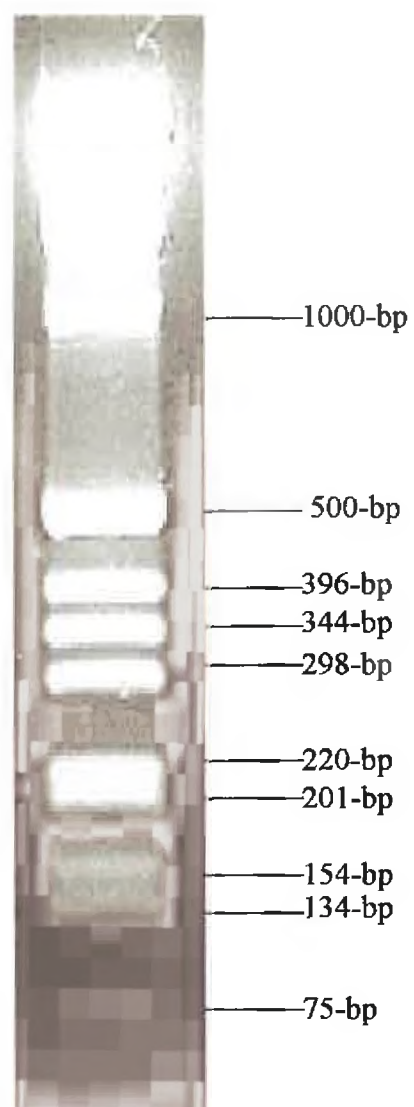
Add 121.1g Tris base to 800ml dH₂O. While stirring, adjust the pH using 1M HCl. Make up to 1 litre.

T20:E5 (pH 8.0) (1l)

Add: 20ml 20mM Tris-HCl

10ml 5mM EDTA (pH 8.0).

Make up to 1l with dH₂O and autoclave.

APPENDIX 3: 1 kb molecular size standard

APPENDIX 4: The 124 unique mtDNA types found in the Malagasy, the number of individuals with each type, the ethnic group(s) each type is found in as well as the control region mutations characterising each type.

MIDNA TYPE	# WITH TYPE	ETHNIC GROUPS FOUND IN	CR1 MUTATIONS	CR2 MUTATIONS
Type 1	10	Antaimoro (1), Antaisaka (2), Bara (1), Betsileo (5), Sakalava (1)	16086 T→C; 16148 C→T; 16223 C→T; 16259 C→T; 16278 C→T; 16319 G→A	73 A→G; 150 C→T; 200 A→G; 263 A→G
Type 2	1	Betsileo	16086 T→C; 16148 C→T; 16223 C→T; 16259 C→T; 16278 C→T; 16319 G→A	73 A→G; 141 C→T; 150 C→T; 200 A→G; 263 A→G
Type 3	1	Tsimihety	16086 T→C; 16148 C→T; 16223 C→T; 16250 C→T; 16259 C→T; 16278 C→T; 16319 G→A; 16399 A→G	73 A→G; 150 C→T; 200 A→G; 263 A→G
Type 4	1	Betsileo	16086 T→C; 16148 C→T; 16223 C→T; 16236 C→T; 16259 C→T; 16278 C→T; 16280 A→G; 16319 G→A; 16399 A→G	73 A→G; 85 G→T; 150 C→T; 200 A→G; 259 A→T; 263 A→G
Type 5	1	Merina	16086 T→C; 16148 C→T; 16223 C→T; 16259 C→T; 16278 C→T; 16319 G→A	73 A→G; 150 C→T; 152 T→C; 200 A→G; 263 A→G
Type 6	1	Bara	16086 T→C; 16148 C→T; 16223 C→T; 16278 C→T; 16319 G→A	73 A→G; 150 C→T; 200 A→G; 263 A→G
Type 7	1	Merina	16176 C→T; 16223 C→T; 16327 C→T	73 A→G; 150 C→T; 152 T→C; 200 A→G; 214 A→G
Type 8	2	Merina (2)	16223 C→T; 16327 C→T	73 A→G; 150 C→T; 189 A→G; 200 A→G; 263 A→G
Type 9	1	Tsimihety	16023 T→C; 16265 A→T; 16292 C→T	73 A→G; 150 C→T; 189 A→G; 200 A→G; 204 T→C; 263 A→G
Type 10	1	Sakalava	16185 C→T; 16223 C→T; 16327 C→T	73 A→G; 150 C→T; 200 A→G; 263 A→G
Type 11	1	Antaimoro	16185 C→T; 16223 C→T; 16327 C→T	73 A→G; 150 C→T; 189 A→G; 200 A→G; 263 A→G
Type 12	1	Mahafaly	16185 C→T; 16223 C→T; 16311 T→C; 16327 C→T	73 A→G; 150 C→T; 185 G→A; 200 A→G; 263 A→G
Type 13	1	Vezo	16185 C→T; 16223 C→T; 16311 T→C; 16327 C→T	73 A→G; 150 C→T; 185 G→A; 189 A→G; 263 A→G
Type 14	3	Sakalava (3)	16172 T→C; 16185 C→T; 16223 C→T; 16311 T→C; 16327 C→T	73 A→G; 150 C→T; 185 G→A; 189 A→G; 263 A→G
Type 15	1	Antaisaka	16220 A→C; 16265 A→G; 16298 T→C; 16362 T→C	73 A→G; 150 C→T; 152 T→C; 263 A→G; 297 A→G
Type 16	1	Antaisaka	16141 A→G; 16220 A→C; 16265 A→G; 16298 T→C; 16362 T→C	73 A→G; 150 C→T; 152 T→C; 263 A→G; 297 A→G; 379 A→G
Type 17	3	Betsileo (2), Sakalava (1)	16265 A→G; 16298 T→C; 16362 T→C	73 A→G; 150 C→T; 152 T→C; 200 A→G; 249 A→T; 263 A→G
Type 18	1	Antaisaka	16220 A→C; 16265 A→G; 16298 T→C; 16362 T→C	73 A→G; 150 C→T; 152 T→C; 263 A→G; 289 A→T
Type 19	7	Antaimoro (2), Antaisaka (1), Merina (2), Tsimihety (1), Vezo (1)	16265 A→G; 16298 T→C; 16362 T→C	73 A→G; 150 C→T; 152 T→C; 263 A→G
Type 20	1	Sakalava	16124 T→C; 16223 C→T; 16278 C→T; 16311 T→C; 16362 T→C	73 A→G; 152 T→C; 263 A→G
Type 21	1	Sakalava	16124 T→C; 16149 A→G; 16223 C→T; 16278 C→T; 16311 T→C; 16362 T→C	73 A→G; 152 T→C; 263 A→G
Type 22	1	Betsileo	16124 T→C; 16223 C→T; 16319 G→A	73 A→G; 85 G→T; 150 C→T; 152 T→C; 263 A→G
Type 23	3	Antaisaka (2), Merina (1)	16124 T→C; 16223 C→T	73 A→G; 150 C→T; 152 T→C; 263 A→G
Type 24	1	Merina	16223 C→T	73 A→G; 150 C→T; 195 T→C; 263 A→G
Type 25	1	Tsimihety	16185 C→T; 16223 C→T; 16327 C→T	73 A→G; 150 C→T; 195 T→C; 263 A→G
Type 26	1	Betsileo	16223 C→T; 16265 A→T	73 A→G; 150 C→T; 195 T→C; 263 A→G
Type 27	1	Tsimihety	16223 C→T; 16278 C→T; 16286 C→T; 16294 C→T; 16309 A→G; 16357 T→C; 16373 G→A; 16390 G→A	73 A→G; 150 C→T; 152 T→C; 195 T→C; 263 A→G
Type 28	1	Tsimihety	16223 C→T; 16265 A→T	73 A→G; 152 T→C; 195 T→C; 204 T→C; 263 A→G

MITOCHONDRIAL DNA TYPE	# WITH TYPE	ETHNIC GROUPS FOUND IN	CR1 MUTATIONS	CR2 MUTATIONS
Type 29	1	Merina	16223 C→T; 16263 T→C; 16311 T→C	73 A→G; 152 T→C; 195 T→C; 263 A→G
Type 30	1	Antaisaka	16141 A→G; 16223 C→T; 16263 T→C; 16311 T→C	73 A→G; 152 T→C; 195 T→C; 204 T→C; 263 A→G; 297 A→G
Type 31	4	Merina (1), Sakalava (1), Tsimihety (1), Vezo (1)	16223 C→T; 16263 T→C; 16311 T→C	73 A→G; 152 T→C; 195 T→C; 204 T→C; 263 A→G
Type 32	1	Sakalava	16223 C→T; 16263 T→C; 16274 G→A; 16311 T→C	73 A→G; 152 T→C; 195 T→C; 204 T→C; 263 A→G
Type 33	1	Mahafaly	16218 C→T; 16255 G→A; 16304 T→C; 16311 T→C	73 A→G; 152 T→C; 263 A→G
Type 34	1	Antaimoro	16218 C→T; 16241 A→G; 16255 G→A; 16304 T→C; 16311 T→C	73 A→G; 152 T→C; 263 A→G
Type 35	11	Antaisaka (1), Antaisaka (2), Betsileo (5), Merina (1), Tsimihety (2)	16093 T→C; 16223 C→T; 16278 C→T; 16362 T→C	73 A→G; 263 A→G
Type 36	1	Betsileo	16093 T→C; 16223 C→T; 16266 C→T; 16278 C→T; 16362 T→C	73 A→G; 263 A→G
Type 37	4	Mahafaly (3), Sakalava (1)	16093 T→C; 16124 T→C; 16223 C→T; 16278 C→T; 16362 T→C	73 A→G; 263 A→G
Type 38	1	Antaisaka	16093 T→C; 16223 C→T; 16278 C→T; 16362 T→C	73 A→G; 223 T→C; 263 A→G; 290 A→G; 291 A→G
Type 39	1	Antaisaka	16093 T→C; 16223 C→T; 16278 C→T; 16362 T→C	73 A→G; 243 A→G; 263 A→G; 297 A→G; 341 A→G
Type 40	1	Antaisaka	16104 C→T; 16221 C→T; 16223 C→T; 16291 C→T; 16362 T→C; 16390 G→A	73 A→G; 263 A→G; 297 A→G; 372 T→C
Type 41	1	Mahafaly	16221 C→T; 16223 C→T; 16270 C→T; 16291 C→T; 16362 T→C	73 A→G; 263 A→G; 372 T→C
Type 42	1	Merina	16221 C→T; 16223 C→T; 16291 C→T; 16362 T→C; 16390 G→A	73 A→G; 93 A→G; 263 A→G; 372 T→C
Type 43	8	Mahafaly (5), Sakalava (3)	16223 C→T; 16295 C→T; 16362 T→C	73 A→G; 263 A→G
Type 44	1	Tsimihety	16156 G→A; 16223 C→T; 16295 C→T; 16362 T→C	73 A→G; 263 A→G
Type 45	9	Antaimoro (1), Betsileo (3), Merina (2), Tsimihety (1), Vezo (2)	16221 C→T; 16223 C→T; 16291 C→T; 16362 T→C	73 A→G; 263 A→G; 372 T→C
Type 46	1	Betsileo	16223 C→T; 16259 C→T; 16278 C→T; 16319 G→A	73 A→G; 109 G→T; 146 T→C; 263 A→G
Type 47	1	Betsileo	16259 C→T; 16278 C→T; 16319 G→A	73 A→G; 200 A→G; 263 A→G
Type 48	1	Antaisaka	16223 C→T; 16295 C→T; 16362 T→C	73 A→G; 146 T→C; 263 A→G
Type 49	1	Mahafaly	16179 C→T; 16189 T→C; 16223 C→T; 16239 C→T; 16270 C→T; 16311 T→C; 16320 C→T; 16352 T→C; 16362 T→C	73 A→G; 103 G→A; 150 C→T; 199 T→C; 204 T→C; 263 A→G
Type 50	1	Tsimihety	16223 C→T; 16295 C→T; 16362 T→C	73 A→G; 146 T→C; 199 T→C; 204 T→C; 263 A→G
Type 51	1	Antaisaka	16223 C→T; 16295 C→T; 16362 T→C; 16375 C→T	73 A→G; 146 T→C; 199 T→C; 204 T→C; 263 A→G; 297 A→G
Type 52	1	Antaisaka	16093 T→C; 16223 C→T; 16295 C→T; 16362 T→C	73 A→G; 146 T→C; 199 T→C; 263 A→G; 290 A→G; 297 A→G
Type 53	18	Antandroy (1), Betsileo (3), Mahafaly (2), Merina (5), Sakalava (2), Tsimihety (4), Vezo (1)	16223 C→T; 16295 C→T; 16362 T→C	73 A→G; 146 T→C; 199 T→C; 263 A→G
Type 54	1	Merina	16189 T→C; 16192 C→T; 16223 C→T; 16256 C→T; 16278 C→T; 16294 C→T; 16309 A→G; 16344 C→T	73 A→G; 146 T→C; 152 T→C; 195 T→C; 263 A→G

MtDNA TYPE	# WITH TYPE	ETHNIC GROUPS FOUND IN	CR1 MUTATIONS	CR2 MUTATIONS
Type 55	11	Antaisaka (1), Antandroy (1), Mahafaly (4), Merina (1), Sakalava (2), Tsimihety (1), Vezo (1)	16189 T→C; 16223 C→T; 16278 C→T; 16290 C→T; 16294 C→T	73 A→G; 146 T→C; 152 T→C; 195 T→C; 263 A→G
Type 56	1	Mahafaly	16189 T→C; 16223 C→T; 16248 C→T; 16278 C→T; 16290 C→T; 16294 C→T; 16309 A→G	73 A→G; 146 T→C; 152 T→C; 195 T→C; 263 A→G
Type 57	2	Sakalava (1), Tsimihety (1)	16189 T→C; 16223 C→T; 16278 C→T; 16294 C→T; 16309 A→G	73 A→G; 146 T→C; 152 T→C; 195 T→C; 198 C→T; 263 A→G
Type 58	1	Merina	16189 T→C; 16207 C→G; 16223 C→T; 16278 C→T; 16290 C→T; 16294 C→T; 16309 A→G	51 T→C; 73 A→G; 146 T→C; 152 T→C; 195 T→C; 263 A→G
Type 59	1	Mahafaly	16141 A→G; 16189 T→C; 16223 C→T; 16224 T→C; 16270 C→T; 16278 C→T; 16309 A→G	73 A→G; 146 T→C; 152 T→C; 195 T→C; 263 A→G
Type 60	2	Antaimoro (1), Antaisaka (1)	16093 T→C; 16223 C→T; 16278 C→T; 16362 T→C	73 A→G; 146 T→C; 152 T→C; 195 T→C; 263 A→G
Type 61	1	Sakalava	16070 A→T; 16170 A→T; 16172 T→C; 16185 C→T; 16223 C→T; 16311 T→C; 16327 C→T	73 A→G; 146 T→C; 152 T→C; 195 T→C; 263 A→G
Type 62	1	Tsimihety	16223 C→T; 16234 C→T; 16249 T→C; 16278 C→T; 16294 C→T; 16295 C→T; 16390 G→A	73 A→G; 146 T→C; 152 T→C; 195 T→C; 263 A→G
Type 63	4	Mahafaly (1), Merina (1), Tsimihety (1), Vezo (1)	16223 C→T; 16278 C→T; 16286 C→T; 16294 C→T; 16309 A→G	73 A→G; 146 T→C; 152 T→C; 195 T→C; 263 A→G
Type 64	1	Mahafaly	16017 T→G; 16189 T→C; 16278 C→T; 16309 A→G; 16385 A→G	73 A→G; 143 G→A; 146 T→C; 152 T→C; 182 C→T; 183 A→G; 263 A→G
Type 65	1	Vezo	16189 T→C; 16223 C→T; 16224 T→C; 16278 C→T; 16309 A→G	73 A→G; 146 T→C; 152 T→C; 182 C→T; 183 A→G; 263 A→G
Type 66	1	Antankarana	16189 T→C; 16217 T→C; 16261 C→T	73 A→G; 146 T→C; 263 A→G
Type 67	1	Merina	16189 T→C; 16223 C→T; 16278 C→T; 16290 C→T; 16294 C→T	73 A→G; 146 T→C; 195 T→C; 263 A→G
Type 68	1	Tsimihety	16189 T→C; 16217 T→C; 16247 A→G; 16261 C→T	73 A→G; 146 T→C; 195 T→C; 263 A→G
Type 69	1	Antaisaka	16189 T→C; 16217 T→C; 16242 C→G; 16247 A→G; 16261 C→T	73 A→G; 150 C→T; 152 T→C; 253 C→T; 254 T→C; 263 A→G; 277 C→A; 278 A→G
Type 70	1	Tsimihety	16189 T→C; 16217 T→C; 16247 A→G; 16261 C→T	73 A→G; 146 T→C; 152 T→C; 263 A→G
Type 71	1	Antaisaka	16189 T→C; 16217 T→C; 16247 A→G; 16261 C→T	73 A→G; 146 T→C; 263 A→G; 310 T→C
Type 72	1	Mahafaly	16189 T→C; 16217 T→C; 16247 A→G; 16261 C→T	73 A→G; 91 C→A; 146 T→C; 263 A→G; 310 T→C
Type 73	1	Mahafaly	16189 T→C; 16207 C→G; 16217 T→C; 16247 A→G; 16248 C→G; 16261 C→T; 16262 C→T; 16358 C→T; 16360 C→T	73 A→G; 146 T→C; 263 A→G; 310 T→C; 320 C→T
Type 74	1	Antaisaka	16189 T→C; 16217 T→C; 16242 C→A; 16247 A→G; 16261 C→T; 16270 C→G; 16399 A→G	73 A→G; 146 T→C; 263 A→G; 310 T→C
Type 75	1	Vezo	16189 T→C; 16217 T→C; 16247 A→G; 16261 C→T; 16270 C→T	73 A→G; 146 T→C; 263 A→G; 310 T→C; 317 C→T
Type 76	1	Betsileo	16189 T→C; 16217 T→C; 16247 A→G; 16261 C→T; 16357 T→C	73 A→G; 146 T→C; 263 A→G; 317 C→T
Type 77	2	Antaisaka (1), Tsimihety (1)	16189 T→C; 16217 T→C; 16242 C→A; 16247 A→G; 16261 C→T	73 A→G; 146 T→C; 263 A→G

MtDNA TYPE	# WITH TYPE	ETHNIC GROUPS FOUND IN	CR1 MUTATIONS	CR2 MUTATIONS
Type 78	1	Antaisaka	16189 T→C; 16217 T→C; 16242 C→A; 16243 T→C; 16247 A→G; 16261 C→T	73 A→G; 146 T→C; 263 A→G
Type 79	1	Betsileo	16189 T→C; 16217 T→C; 16242 C→A; 16247 A→G; 16399 A→G	73 A→G; 146 T→C; 263 A→G
Type 80	1	Betsileo	16217 T→C; 16247 A→G; 16261 C→T	73 A→G; 146 T→C; 263 A→G
Type 81	1	Tsimihety	16093 T→C; 16179 C→T; 16189 T→C; 16217 T→C; 16247 A→G; 16261 C→T	73 A→G; 146 T→C; 263 A→G; 341 A→G
Type 82	2	Antaisaka (1), Betsileo (1)	16189 T→C; 16217 T→C; 16247 A→G; 16261 C→T	73 A→G; 146 T→C; 199 T→C; 263 A→G
Type 83	1	Tsimihety	16189 T→C; 16217 T→C; 16247 A→G; 16261 C→T	73 A→G; 146 T→C; 236 T→C; 247 G→A; 263 A→G
Type 84	1	Sakalava	16141 A→T; 16171 A→T; 16189 T→C; 16217 T→C; 16226 A→G; 16247 A→G; 16272 A→G; 16280 A→T; 16355 C→T	73 A→G; 146 T→C; 263 A→G
Type 85	1	Antaisaka	16141 A→G; 16189 T→C; 16217 T→C; 16247 A→G; 16261 C→T	73 A→G; 146 T→C; 263 A→G
Type 86	1	Mahafaly	16189 T→C; 16217 T→C; 16228 C→T; 16247 A→G; 16261 C→T	73 A→G; 146 T→C; 263 A→G
Type 87	1	Mahafaly	16092 T→C; 16145 G→A; 16189 T→C; 16217 T→C; 16247 A→G; 16261 C→T	73 A→G; 146 T→C; 263 A→G
Type 88	15	Antaisaka (2), Betsileo (1), Merina (3), Sakalava (9)	16189 T→C; 16217 T→C; 16247 A→G	73 A→G; 146 T→C; 263 A→G
Type 89	1	Merina	16189 T→C; 16217 T→C; 16246 A→G; 16247 A→G; 16261 C→T	73 A→G; 146 T→C; 263 A→G
Type 90	2	Betsileo (2)	16189 T→C; 16217 T→C; 16247 A→G; 16261 C→T	73 A→G; 146 T→C; 263 A→G
Type 91	1	Tsimihety	16189 T→C; 16217 T→C; 16247 A→G; 16261 C→T	73 A→G; 146 T→C; 263 A→G; 345 C→G
Type 92	1	Betsileo	16162 A→G; 16189 T→C; 16217 T→C; 16247 A→G; 16261 C→T	73 A→G; 146 T→C; 263 A→G
Type 93	2	Antaisaka (1), Sakalava (1)	16189 T→C; 16221 C→T; 16223 C→T; 16291 C→T; 16362 T→C	73 A→G; 263 A→G; 372 T→C
Type 94	1	Antaisaka	16038 A→G; 16093 T→C; 16129 G→A; 16187 C→T; 16189 T→C; 16223 C→T; 16271 T→C; 16278 C→T; 16293 A→G; 16294 C→T; 16311 T→C; 16360 C→T	73 A→G; 150 C→T; 200 A→G; 263 A→G; 297 A→G
Type 95	1	Tsimihety	16093 T→C; 16148 C→T; 16168 C→T; 16172 T→C; 16187 C→T; 16188 C→G; 16189 T→C; 16223 C→T; 16230 A→G; 16287 C→T; 16293 A→G; 16311 T→C; 16320 C→T	73 A→G; 146 T→C; 263 A→G
Type 96	1	Tsimihety	16093 T→C; 16148 C→T; 16167 C→T; 16172 T→C; 16187 C→T; 16188 C→G; 16189 T→C; 16223 C→T; 16230 A→G; 16311 T→C; 16320 C→T	73 A→G; 150 C→T; 152 T→C; 195 T→C; 263 A→G
Type 97	1	Tsimihety	16129 G→A; 16189 T→C; 16207 C→G; 16215 A→G; 16223 C→T; 16311 T→C	73 A→G; 152 T→C; 182 C→T; 186 C→A; 247 G→A; 263 A→G; 316 G→A
Type 98	2	Antaisaka (2)	16129 G→A; 16187 C→T; 16189 T→C; 16223 C→T; 16278 C→T; 16293 A→G; 16294 C→T; 16311 T→C; 16360 C→T	73 A→G; 152 T→C; 182 C→T; 186 C→A; 189 A→C; 195 T→C; 247 G→A; 248 A→G; 263 A→G; 316 G→A
Type 99	1	Tsimihety	16189 T→C; 16217 T→C; 16242 C→A; 16247 A→G; 16261 C→T	73 A→G; 151 C→T; 152 T→C; 186 C→A; 189 A→C; 195 T→C; 198 C→T; 247 G→A; 263 A→G; 297 A→G; 316 G→A
Type 100	1	Antaisaka	16189 T→C; 16228 C→T; 16271 T→C; 16278 C→T; 16293 A→G; 16294 C→T; 16311 T→C; 16360 C→T	73 A→G; 151 C→T; 152 T→C; 182 C→T; 186 C→A; 189 A→C; 195 T→C; 198 C→T; 263 A→G; 297 A→G; 316 G→A

MITDNA TYPE	# WITH TYPE	ETHNIC GROUPS FOUND IN	CR1 MUTATIONS	CR2 MUTATIONS
Type 101	1	Sakalava	16038 A→G; 16093 T→C; 16129 G→A; 16187 C→T; 16189 T→C; 16223 C→T; 16271 T→C; 16278 C→T; 16293 A→G; 16294 C→T; 16311 T→C; 16360 C→T	73 A→G; 151 C→T; 152 T→C; 182 C→T; 186 C→A; 189 A→C; 195 T→C; 198 C→T; 247 G→A; 263 A→G; 297 A→G; 316 G→A
Type 102	1	Sakalava	16071 C→T; 16093 T→C; 16129 G→A; 16145 G→A; 16187 C→T; 16189 T→C; 16213 G→A; 16223 C→T; 16234 C→T; 16265 A→C; 16278 C→T; 16286 C→G; 16294 C→T; 16311 T→C; 16360 C→T	73 A→G; 151 C→T; 152 T→C; 182 C→T; 186 C→A; 189 A→C; 195 T→C; 198 C→T; 247 G→A; 263 A→G; 297 A→G; 316 G→A
Type 103	1	Tsimihety	16223 C→T; 16265 A→T	93 A→G; 95 A→C; 185 G→A; 236 T→C; 247 G→A; 263 A→G
Type 104	1	Antaisaka	16148 C→T; 16187 C→T; 16188 C→G; 16189 T→C; 16223 C→T; 16311 T→C	64 C→T; 89 T→C; 93 A→G; 95 A→C; 152 T→C; 189 A→G; 236 T→C; 247 G→A; 263 A→G; 274 A→G; 290 A→G; 297 A→G
Type 105	1	Sakalava	16093 T→C; 16129 G→A; 16148 C→T; 16168 C→T; 16172 T→C; 16187 C→T; 16188 C→G; 16189 T→C; 16223 C→T; 16230 A→G; 16278 C→T; 16311 T→C; 16320 C→T	93 A→G; 95 A→C; 185 G→A; 189 A→G; 236 T→C; 247 G→A; 263 A→G
Type 106	1	Sakalava	16064 T→A; 16070 A→G; 16071 C→T; 16129 G→A; 16148 C→T; 16168 C→T; 16172 T→C; 16187 C→T; 16188 C→G; 16189 T→C; 16223 C→T; 16225 C→T; 16230 A→G; 16278 C→T; 16293 A→G; 16311 T→C; 16320 C→T	93 A→G; 95 A→C; 185 G→A; 189 A→G; 236 T→C; 247 G→A; 263 A→G
Type 107	1	Antaisaka	16129 G→A; 16148 C→T; 16168 C→T; 16172 T→C; 16187 C→T; 16188 C→G; 16189 T→C; 16223 C→T; 16230 A→G; 16278 C→T; 16293 A→G; 16311 T→C; 16320 C→T	93 A→G; 95 A→C; 185 G→A; 236 T→C; 247 G→A; 263 A→G
Type 108	1	Tsimihety	16129 G→A; 16148 C→T; 16168 C→T; 16172 T→C; 16187 C→T; 16188 C→G; 16189 T→C; 16223 C→T; 16230 A→G; 16278 C→T; 16293 A→G; 16311 T→C; 16320 C→T	89 T→C; 93 A→G; 95 A→C; 185 G→A; 236 T→C; 247 G→A; 263 A→G
Type 109	1	Merina	16129 G→A; 16148 C→T; 16172 T→C; 16187 C→T; 16188 C→G; 16189 T→C; 16230 A→G; 16242 C→G; 16270 C→T; 16311 T→C; 16320 C→T	93 A→G; 95 A→C; 185 G→A; 189 A→G; 236 T→C; 247 G→A; 263 A→G
Type 110	1	Tsimihety	16129 G→A; 16148 C→T; 16168 C→T; 16187 C→T; 16188 C→G; 16189 T→C; 16223 C→T; 16230 A→G; 16278 C→T; 16293 A→G; 16311 T→C; 16320 C→T	93 A→G; 95 A→C; 185 G→A; 236 T→C; 247 G→A; 263 A→G
Type 111	1	Antaisaka	16148 C→T; 16172 T→C; 16187 C→T; 16188 C→G; 16189 T→C; 16223 C→T; 16230 A→G; 16311 T→C; 16320 C→T	64 C→T; 93 A→G; 152 T→C; 189 A→G; 204 T→C; 207 G→A; 236 T→C; 247 G→A; 263 A→G
Type 112	1	Antankarana	16093 T→C; 16148 C→T; 16172 T→C; 16187 C→T; 16188 C→G; 16189 T→C; 16223 C→T; 16230 A→G; 16311 T→C; 16320 C→T	64 C→T; 93 A→G; 152 T→C; 189 A→G; 204 T→C; 207 G→A; 236 T→C; 247 G→A; 263 A→G
Type 113	1	Tsimihety	16148 C→T; 16172 T→C; 16187 C→T; 16188 C→G; 16189 T→C; 16223 C→T; 16230 A→G; 16311 T→C; 16320 C→T	64 C→T; 93 A→G; 152 T→C; 189 A→G; 195 T→C; 204 T→C; 207 G→A; 236 T→C; 247 G→A; 263 A→G
Type 114	1	Antankarana	16148 C→T; 16172 T→C; 16187 C→T; 16188 C→G; 16189 T→C; 16223 C→T; 16230 A→G; 16311 T→C; 16320 C→T	64 C→T; 93 A→G; 150 C→T; 152 T→C; 189 A→G; 204 T→C; 207 G→A; 236 T→C; 247 G→A; 263 A→G
Type 115	1	Betsileo	16148 C→T; 16172 T→C; 16187 C→T; 16188 C→G; 16189 T→C; 16226 A→C; 16230 A→G; 16242 C→G; 16270 C→T; 16311 T→C; 16320 C→T; 16343 A→G; 16355 C→G	64 C→T; 93 A→G; 150 C→T; 152 T→C; 189 A→G; 204 T→C; 236 T→C; 247 G→A; 263 A→G

MtDNA TYPE	# WITH TYPE	ETHNIC GROUPS FOUND IN	CR1 MUTATIONS	CR2 MUTATIONS
Type 116	1	Sakalava	16148 C→T; 16172 T→C; 16187 C→T; 16188 C→G; 16189 T→C; 16223 C→T; 16230 A→G; 16266 C→T; 16311 T→C; 16320 C→T	64 C→T; 93 A→G; 152 T→C; 204 T→C; 207 G→A; 236 T→C; 247 G→A; 263 A→G
Type 117	1	Antaisaka	16148 C→T; 16172 T→C; 16187 C→T; 16188 C→G; 16189 T→C; 16223 C→T; 16230 A→G; 16311 T→C; 16320 C→T	64 C→T; 93 A→G; 152 T→C; 189 A→G; 204 T→C; 207 G→A; 236 T→C; 263 A→G
Type 118	1	Tsimihety	16148 C→T; 16172 T→C; 16187 C→T; 16188 C→G; 16189 T→C; 16223 C→T; 16230 A→G; 16311 T→C; 16320 C→T	64 C→T; 93 A→G; 95 A→C; 152 T→C; 189 A→G; 236 T→C; 247 G→A; 263 A→G
Type 119	1	Sakalava	16148 C→T; 16172 T→C; 16187 C→T; 16188 C→G; 16189 T→C; 16223 C→T; 16230 A→G; 16311 T→C; 16320 C→T	64 C→T; 89 T→C; 93 A→G; 95 A→C; 152 T→C; 189 A→G; 236 T→C; 247 G→A; 263 A→G
Type 120	1	Antandroy	16093 T→C; 16148 C→T; 16172 T→C; 16187 C→T; 16188 C→G; 16189 T→C; 16223 C→T; 16230 A→G; 16311 T→C; 16320 C→T	64 C→T; 93 A→G; 152 T→C; 189 A→G; 195 T→C; 236 T→C; 247 G→A; 263 A→G
Type 121	1	Tsimihety	16093 T→C; 16148 C→T; 16167 C→T; 16172 T→C; 16187 C→T; 16188 C→G; 16189 T→C; 16223 C→T; 16230 A→G; 16311 T→C; 16320 C→T	64 C→T; 93 A→G; 152 T→C; 189 A→G; 236 T→C; 247 G→A; 263 A→G
Type 122	1	Betsimisaraka	16093 T→C; 16148 C→T; 16172 T→C; 16187 C→T; 16188 C→G; 16189 T→C; 16223 C→T; 16230 A→G; 16311 T→C; 16320 C→T	64 C→T; 93 A→G; 95 A→C; 152 T→C; 189 A→G; 236 T→C; 247 G→A; 263 A→G
Type 123	1	Sakalava	16093 T→C; 16148 C→T; 16172 T→C; 16187 C→T; 16188 C→G; 16189 T→C; 16210 A→G; 16214 C→T; 16223 C→T; 16230 A→G; 16311 T→C; 16320 C→T	64 C→T; 93 A→G; 152 T→C; 189 A→G; 207 G→A; 236 T→C; 247 G→A; 263 A→G
Type 124	1	Tsimihety	16093 T→C; 16148 C→T; 16172 T→C; 16187 C→T; 16188 C→G; 16189 T→C; 16214 C→T; 16223 C→T; 16230 A→G; 16289 A→G; 16311 T→C; 16320 C→T	64 C→T; 93 A→G; 95 A→C; 152 T→C; 189 A→G; 207 G→A; 236 T→C; 247 G→A; 263 A→G

REFERENCES

- Alves-Silva, J.; Guimaraes, P.E.M; Rocha, J.; Pena, S.D.J. and Prado, V.F. (1998). Identification in Portugal and Brazil of a mtDNA Lineage Containing a 9-bp Triplication of the Intergenic COII/tRNA^{Lys} Region. Human Heredity 49: 56-58.
- Anderson, S.; Bankier, .T.; Barrell, B.G.; de Bruijn, M.H.; Coulson, A.R.; Drouin, J.; Eperon, I.C.; Nierlich, D.P.; Roe, B.A.; Sanger, F.; Schreier, P.H.; Smith, A.J.; Straden, R. and Young, I.G. (1981). Sequence and Organisation of the Human Mitochondrial Genome. Nature 290: 457-465.
- Arai, T.; Tabona, P. and Summerfield, J.A. (1993). Human mannose-binding protein gene is regulated by interleukins, dexamethasone and heat shock. Quarterly Journal of Medicine 86: 575-582.
- Bamshad, M.; Fraley, A.E.; Crawford, M.H.; Cann, R.L.; Busi, B.R.; Naidu, J.M. and Jorde, L.B. (1996). MtDNA Variation in Caste Populations of Andhra Pradesh, India. Human Biology 68: 1-28.
- Batzer, M.A.; Arcot, S.S.; Phinney, J.W.; Alegria-Hartman, M.; Kass, D.H.; Milligan, S.M.; Kimpton, C.; Gill, P.; Hochmeister, M.; Ioannou, P.A.; Herrera, R.J.; Boudreau, D.A.; Scheer, W.D.; Keats, B.J.B.; Deininger, P.L. and Stoneking, M. (1996). Genetic Variation of Recent Alu Insertions in Human Populations. Journal of Molecular Evolution, 42:22-29.
- Beuttner-Janusch, J. and Beuttner-Janusch, V. (1964). Hemoglobins, haptoglobins and transferrins in the people of Madagascar. American Journal of Physical Anthropology, 22: 163-169.
- Beuttner-Janusch, J.; Beuttner-Janusch, V.; Coppenhaver, D. and Rcisman, R. (1973). Transferrins, haptoglobins and ceruloplasmins among the Merina and Sakalava. Arch Inst Pasteur Madagascar, 42: 193-200.
- Bradt, H. (1993). Guide to Madagascar. UK: Bradt Publishing, and USA: Hunter Publishing.
- Brown, M. (1978). Madagascar Rediscovered London: Damien Tunnacliffe.

Brown, W.M.; George, M. and Wilson, A.C. (1979). Rapid evolution of animal mitochondrial DNA. Proceedings of the National Academy of Science USA, **76**: 1967-1971.

Brown, W.M.; Prager, E.M.; Wang, A. and Wilson, A. (1982). Mitochondrial DNA sequences of primates: tempo and mode of evolution. Journal of Molecular Evolution, **18**: 225-239.

Campbell, G. (1989). The East African Slave Trade, 1861-1895. The “southern” complex. International Journal of African Historical Studies, **22**: 1-26.

Campbell, G. (1996). Theories of the origins of the Malagasy: the nineteenth century British school. Australes – Études historiques aixoises sur l’Afrique australe et l’océan Indien occidental. Paris. L’Harmattan.

Cann, R.L. and Wilson, A.C. (1984). Polymorphic sites and the mechanism of evolution in human mitochondrial DNA. Genetics, **104**: 699-711.

Cavalli-Sforza, L.L.; Menozzi, P. and Piazza, A. (1994). The History and Geography of Human Genes. Princeton University Press, Princeton, New Jersey.

Chakraborty, R.; Kamboh, M.I. and Ferrell, R.E. (1991). “Unique” alleles in admixed populations: a strategy for determining hereditary population differences of disease frequencies. Ethn Dis, **1**:245-256.

Chang, D.D. and Clayton, D.A. (1984). Precise identification of individual promoters for transcription of each strand of human mitochondrial DNA. Cell, **36**: 635-643.

Chen, Y.; Torroni, A.; Excoffier, L.; Santachiara-Benerecetti, A.S. and Wallace, D.C. (1995). Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. American Journal of Human Genetics **57**: 133-149.

Chen, Y.; Olekers, A.; Schurr, T.G.; Kogelnik, A.M.; Huoponen, K. and Wallace, D. (2000). MtDNA Variation in the south African Kung and Khwe – and Their Genetic Relationships to Other African Populations. American Journal of Human Genetics **66**: 1362 – 1383.

- Cole, M. (1992). Return to Madagascar. New Scientist, 22: 40-43.
- Cooper, G. Amos, W.; Hoffman, D. and Rubinsztein, D.C. (1996). Network analysis of human Y microsatellite haplotypes. Human Molecular Genetics, **5**(11): 1759-1766.
- Dahl, O.C. (1951). Malagache et Maanvan: Une comparaison linguistique. Oslo: Egede Institute.
- Dahl, O.C. (1977). La subdivision de la famille Barito et la place du Malagache. Acta Orientalia, **38**: 77-134.
- Dangerfield, B. (1997). The characterisation of Glucose-6-Phosphate Dehydrogenase (G6PD) variation in the Malagasy. MSc dissertation, Faculty of Medicine, University of the Witwatersrand, Johannesburg.
- Deininger, P.L.; Batzer, M.A.; Hutchinson, C.A.; and Edgell, M.H. (1992). Master Genes in Mammalian Repetitive DNA Amplification. Trends In Genetics, **8**: 307-311.
- Deka, R.; Jin, L.; Shriver, M.D.; Yu, L.M.; DeCoo, S.; Hundrieser, J.; Bunker, C.H. *et al.* (1995). Population genetics of dinucleotide (dC-dA)_n/(dG-dT)_n polymorphisms in world populations. American Journal of Human Genetics, **56**: 461-474.
- Deka, R.; Jin, L.; Shriver, M.D.; Yu, L.M.; Saha, N.; Barrantes, R.; Chakraborty, R. and Ferrell, R.E. (1996). Dispersion of Human Y Chromosome Haplotypes Based on Five Microsatellites in Global Populations. Genome Research, **6**: 1 177-1 184.
- Deschamps, H. (1965). Histoire de Madagascar. Editions Berger-Levrault, Paris. Cited in Dewar and Wright, 1993.
- Dewar, R.E. (1995). Of nets and trees: untangling the reticulate and dendritic in Madagascar's prehistory. World Archaeology, **26**(3): 301-318.
- Dewar, R.E. and Wright, H.T. (1993). The Cultural History of Madagascar. Journal of World Prehistory, **7**: 417-466.

Drury, R. (1729). The Adventures of Robert Drury during Fifteen Years of Captivity in the Island of Madagascar. London [1890 edition reprinted by Negro Universities Press, New York, 1969]. Quoted in Heseltine, N. (1971). Madagascar. Praeger Publishers, Inc., London.

Felsenstein, J. (1975). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**: 783-791.

Ferrand, G. (1908). L'origine africaine des Malagaches. *Journal Asiatique*, ser10, **15(3)**: 353-500.

Flacourt, E. (1658). Historie de la grande isle Madagascar. Pierre L'amy, Paris.

Garred, P.; Madsen, H.O.; Kurtzhals, J.A.L.; Lamm, L.U.; Thiel, S.; Hey, A.S. and Svejgaard, A. (1992). Diallelic Polymorphism may explain variations of the blood concentration of mannose-binding protein in Eskimos, but not in black Africans. *European Journal of Immunogenetics* **19**: 403-412.

Giles, R.E.; Blanc, H.; Cann, H.M. and Wallace, D.C. (1980). Maternal inheritance of human mitochondrial DNA. *Proceedings of the National Academy of Science USA*, **77**: 6715-6719.

Grandidier, A. (1908). Ethnographie de Madagascar. Imprimerie Nationale, Paris.

Greenberg, S.; Bankier, A.T.; Barrell, B.G.; 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. (1981). Sequence and organization of the human mitochondrial genome. *Nature*, **290**: 457-464.

Greunier, N.J. (1998). Dialectologie et lexicostatistique: cas du dialecte malagache de Mayotte (Comores). *Etudes Océan Indien*, **9**: 143-167.

Grimes, B.F. (1992). Ethnologue: languages of the world. 12th edition. Summer Institute of Linguistics, Dallas.

Hagelberg, E.; Quevedu, S.; Turbon, D. and Clegg, J.B. (1994). DNA from Easter Islanders. *Nature* **269**: 25-26.

Hagelberg, E.; Kayser, M.; Nagy, M.; Roewer, L.; Zimdahl, H.; Krawczak, M.; Lio, P. and Schiefenhövel, W. (1999). Molecular genetic evidence for the human settlement of the Pacific: analysis of mitochondrial DNA, Y chromosome and HLA markers. Philosophical Transactions of the Royal Society of London **354**: 141-152.

Hammer, M.F. (1994). A Recent Insertion of an Alu Family Member on the Y Chromosome is a Useful Marker for Human Population Studies. Molecular Biology and Evolution **11**: 749-761.

Hammer, M.F. and Horai, S. (1995). Y Chromosome DNA Variation and the Peopling of Japan. American Journal of Human Genetics **56**: 951-962.

Harpending, H. C.; Sherry, S.T.; Rogers, A.R. and Stoneking, M. (1993). The Genetic Structure of Ancient Human populations. Current Anthropology **34**: 483-496.

Hertzberg, M.; Mickelson, K.N.P.; Serjeantson, S.W.; Prior, J.F. and Trent, R.J. (1989). An Asian-specific 9-bp Deletion of Mitochondrial DNA is Frequently Found in Polynesians. American Journal of Human Genetics, **44**: 504-510.

Herseltine, N. (1971). Madagascar. Praeger Publishers, Inc., London.

Heyerdahl, T. (1973). Kon-Tiki: Across the Pacific by Raft. Washington Square Press, New York.

Hewitt, R.; Krause, A.; Goldman, A.; Campbell, G. and Jenkins, T. (1996). β -globin haplotype analysis suggests that a major source of Malagasy ancestry is derived from Bantu-speaking Negroids. American Journal of Human Genetics, **58**: 1303-1308.

Hixson, J.E. and Clayton, D.A. (1985). Initiation of transcription from each of the two human mitochondrial promoters requires unique nucleotides at the transcriptional start sites. Proceedings of the National Academy of Science USA, **82**: 2660-2664.

Horai, S.; Gojobori, T. and Matsunaga, E. (1987). Evolutionary implications of mitochondrial DNA polymorphisms in human populations. In Vogel, F. and Sperling, K. (editors). Human Genetics: Proceedings of the 7th International Congress, 177-181. Springer, Heidelberg.

Hurles, M.E.; Irvén, C.; Nicholson, J.; Taylor, P.G.; Santos, F.R.; Loughlin, J.; Jobling, M.A. and Sykes, B.C. (1998). European Y-Chromosomal Lineages in Polynesians: A Contrast to the Population Structure Revealed by mtDNA. American journal of Human Genetics **63**: 1793-1806.

Jakubicza, S.; Arnemann, J.; Cooke, H.J.; Krawczak, M.; and Schmidtke, J. (1989). A Search for Restriction Fragment Length Polymorphism on the Human Y Chromosome. Human Genetics **84**: 86-88.

Jorde, L.B.; Bamshad, M.J.; Watkins, W.S.; Zenger, R.; Fraley, A.E.; Kralowiak, P.A.; Carpenter, K.D.; Soodyall, H.; Jenkins, T. and Rogers, A.R. (1995). Origins and affinities of modern humans: a comparison of mitochondrial and nuclear genetic data. American Journal of Human Genetics, **57**: 523-538.

Kanada, H.; Hayashi, J.; Takahama, S.; Taya, C.; Lindahl, K.F. and Yonekawa, H. (1995). Elimination of paternal mitochondrial DNA in intraspecific crosses during early mouse embryogenesis. Proceedings of the National Academy of Science USA, **92**: 4542-4546.

Kimpton, C.P.; Gill, P.; Walton, A.; Urquhart, A.; Millican, E.S. and Adams, M. (1993). Automated DNA Profiling Employing Multiplex Amplification of Short Tandem Repeat Loci. PCR Methods and Applications **3**: 13-22.

Knight, A.; Batzer, M.A.; Stoneking, M.; Tiwari, H.K.; Scheer, D.W.; Herrera, R.; and Deininger, P.L. (1996). DNA Sequences of Alu Elements Indicate a Recent Replacement of the Human Autosomal Genetic Complement. Proceedings of the National Academy of Sciences, USA **93**: 4 360-4 364.

Kumar, S.; Tamura, K. and Nei, M. (1993). MEGA: Molecular Evolutionary Genetics Analysis, version 1.01. University Park, Pennsylvania State University.

Labuda, D. and Striker, G. (1989). Sequence Conservation in *Alu* Evolution. Nucleic Acids Research 17: 2 477-2 491.

Lipscombe, R.J.; Sumiya, M.; Hill, A.V.S.; Lau, Y.L.; Levinsky, R.J.; Summerfield, J.A. and Turner, M.V. (1992). High frequencies in African and non-African populations of independent mutations in the mannose binding protein gene. Human Molecular Genetics 1(9): 709-715.

Lipscombe, R.J.; Beatty, D.W.; Ganczakowski, M.; Goddard, E.A.; Jenkins, T.; Lau, Y.L.; Spurdle, A.B.; Sumiya, M.; Summerfield, J.A. and Turner, M.W. (1996). Mutations in the Human Mannose-Binding Protein Gene: Frequencies in Several Population Groups. European Journal of Human Genetics 4: 13-19.

Liu, Y.; Saha, N.; Low, P.S.; and Tay, J.S.H. (1995). Linkage Disequilibrium between Two Loci (5' Untranslated Exon 1 and Intron 5-Ddel) of the Antithrombin III Gene in Three Ethnic Groups in Singapore. Human Heredity, 45:192-198.

Lum, J.K.; Rickards, O.; Ching, C. and Cann, R.C. (1994). Polynesian mitochondrial DNAs reveal three deep maternal lineage clusters. Human Biology 66: 567-590.

Lum, J.K.; Cann, R.L.; Martinson, J.J. and Jorde, L.B. (1998). Mitochondrial and nuclear Genetic Relationships among Pacific Island and Asian Populations. American Journal of Human Genetics 63: 613-624.

Lutz, S.; Weissner, H.J.; Heizmann, J. and Pollak, S. (1997). A third hypervariable region in the human mitochondrial D-loop. Human Genetics 101: 384.

Mack, J. (1986). Madagascar: Island of Ancestors. London: British Museum Publications Ltd.

MacPhee, R.D.E. and Burney, D.A. (1991). Dating of modified femora of extinct dwarf Hippopotamus from southern Madagascar: implications for constraining human colonization and vertebrate extinction events. Journal of Archaeological Science, 18: 695-706.

Madsen, H.O.; Garred, P.; Kurtzhals, J.A.L.; Lamm, L.U.; Ryder, L.R.; Thiel, S. and Svejgaard, A. (1994). A new frequent allele is the missing link in the structural polymorphism of the human mannan-binding protein. Immunogenetics 40: 37-44.

Malaspina, P.; Persichetti, F.; Novelletto, A.; Iodice, C.; Terrenato, L.; Wolfe, J.; Ferraro, M. and Prantera, G. (1990). The Human Y Chromosome Shows a Low Level of DNA Polymorphism. *Annals of Human Genetics* **54**: 297-305.

Mariano, L. (1904). Relation du voyage de découverte fait a l'Ile Saint-Laurent dans les années 1613 –1614. In *Collection des ouvrages anciens concernant Madagascar*, 2: 1-64. Union Coloniale, Paris.

Matera, G.A.; Hellman, U.; Hintz, M.F.; and Schmid, C.W. (1990). Recently Transposed *Alu* Repeats Result from Multiple Source Genes. *Nucleic Acids Research* **18**: 6 019-6 023.

Melton, T.; Peterson, R.; Redd, A.J.; Saha, N; Sofro, A.S.M.; Martison, J. and Stoneking, M. (1995). Polynesian Genetic Affinities with Southeast Asian Populations as Identified by mtDNA Analysis. *American Journal of human Genetics* **57**: 403-414.

Mountain J.L.; Hebert, J.M.; Bhattacharyya, S.; Underhill, P.A.; Ottolenghi, C.; Gadgil, M. and Cavalli-Sforza, L.L. (1995). Demographic History of India and mtDNA-Sequence Diversity. *American Journal of Human Genetics* **56**: 979-992.

Mourant, A.; Kopec, A. and Domaniewska-Sobczak, K. (1976). The distribution of the human blood groups and other polymorphisms. Oxford University Press, London.

Nagnort (1550). Quoted by Sibree, J. (1880). Cited in Singer, R., Budtz-Olsen, O.E., Brain, P. and Saugrain, J. (1957). Physical features, sickling and serology of the Malagasy of Madagascar. *American Journal of Physical Anthropology*, **15**: 91-124.

Neel, J.V. (1973). "Private" genetic variants and the frequency of mutation among South American Indians. *Proceedings of the National Academy of Sciences, USA*, **70**:3311-3315.

Olivo, P.D.; Van de Walle, M.J.; Lapis, P.J. and Hauswirth, W.W. (1983). Nucleotide sequence evidence for rapid genotypic shifts in the bovine mitochondrial D-loop. *Nature*, **306**: 400-402.

Pääbo, S. (1995). The Y Chromosome and the Origin of All of Us (Men). Science **268**: 1141-1142.

Passarino, G.; Semino, O.; Modiano, G. and Santachiara-Benerecetti, A.S. (1993). COII/tRNA^{Lys} intergenic 9-bp deletion and other mtDNA markers clearly reveal that the Tharus (southern Nepal) have Oriental affinities. American Journal of Human Genetics **53**: 609-618.

Poloni, E.S.; Excoffier, L.; Mountain, J.L.; Langaney, A. and Cavalli-Sforza, L.L. (1995). Nuclear DNA polymorphism in a Mandenka population from Senegal: comparison with eight other human populations. Annals of Human Genetics, **59**: 43-61.

Redd, A.J.; Takezaki, N.; Sherry, S.T.; McGarvey, T.; Sofro, A.S.M. and Stonking, M. (1995). Evolutionary history of the COII/tRNA^{Lys} intergenic 9-bp deletion in human mitochondrial DNAs from the Pacific. Molecular Biology and Evolution **12**: 604-615.

Reed, T.E. (1973). Number of gene loci required for accurate estimation of ancestral population proportions in individual human hybrids. Nature, **244**:575-576.

Rogers, A. and Harpending, H. (1992). Population growth makes waves in the distribution of pairwise differences. Molecular Biology and Evolution **9**: 552-569.

Saccone, C.; Pescle, G. and Sbisà, E. (1991). The main regulatory region of mammalian mitochondrial DNA: Structure-function model of an evolutionary pattern. Journal of Molecular Evolution, **33**:83-91.

Saiki, R.K.; Scharf, S.; Faloona, F.; Mullis, K.B.; Horn, G.T.; Erlich, H.A. and Arnheim, N.A. (1985). Enzymatic amplification of β -globin genomic sequences and restriction site analysis for sickle cell anemia. Science **230**: 1350-1354.

Saitou, N. and Nei, M (1987). The Neighbor-joining method: a new method for constructing phylogenetic trees. Molecular Biology and Evolution **4**: 406-425.

Santos, F.R.; Gerelsaikhan, T.; Munkhtuja, B.; Oyunsuren, T.; Epplen, J.T. and Pena, S.D.J. (1996). Geographic Differences in the Allele Frequencies of the Human Y-Linked Tetranucleotide Polymorphism DYS19. *Human Genetics* **97**: 309-313.

Sawada, I.; Willard, C.; Shen, C.K.J.; Chapman, B.; Wilson, A.C.; and Schmid, C.W. (1985). Evolution of *Alu* Family Repeats Since the Divergence of Human and Chimpanzee. *Journal of Molecular Evolution* **22**: 316-322.

Shriver, M.D.; Smith, M.W.; Jin, L.; Marcini, A.; Akey, J.M.; Deka, R. and Ferrell, R. E. (1997). Ethnic-Affiliation Estimation by Use of Population-Specific DNA Markers. *American Journal of Human Genetics* **60**: 957-964.

Shen, M.R.; Batzer, M.A.; and Deininger, P.L. (1991). Evolution of the Master *Alu* Gene. *Journal of Molecular Evolution* **33**: 311-320.

Sherry, S.T.; Rogers, A.R.; Harpending, H.; Soodyall, H.; Jenkins, T. and Stoneking, M. (1994). Mismatch Distributions of mtDNA Reveal Recent Human Population Expansions. *Human Biology* **66**: 761-755.

Singer, R., Budtz-Olsen, O.E., Brain, P. and Saugrain, J. (1957). Physical features, sickling and serology of the Malagasy of Madagascar. *American Journal of Physical Anthropology*, **15**: 91-124.

Soodyall, H.; Stoneking, M. and Jenkins, T. (1995). Polynesian mtDNA in the Malagasy. *Nature Genetics* **10**: 377-378.

Soodyall, H.; Vigilant, L.; Hill, A.; Stoneking, M. and Jenkins, T. (1996). mtDNA Control-Region Sequence Variation Suggests Multiple Independent Origins of an "Asian-Specific" 9-bp Deletion in Sub-Saharan Africans. *American Journal of Human Genetics* **58**: 595-608.

Soodyall, H.; Jenkins, T.; Hewitt, R.; Krause, A. and Stoneking, M. (1997). The Peopling of Madagascar. Published in Boyce, A.J. and Mascie-Taylor, C.G.M. (1997) Molecular Biology of Human Diversity. Cambridge University Press, Cambridge.

Spuhler, J.N. (1988). Evolution of mitochondrial DNA in monkey, apes and humans. Yearbook of Physical Anthropology 31: 15-48.

Spurdle, A.; Hammer, M.F.; and Jenkins, T. (1994). The Y *Alu* Polymorphism in Southern African Populations, and its Relationship to other Y-Specific Polymorphisms. American Journal of Human Genetics (in press).

Strachan, T. and Read, A.P. (1996). Human molecular Genetics. BIOS Scientific Publishers. Stoneking, 1986

Stoneking, M. (1993). DNA and recent human evolution. Evolutionary Anthropology, 2: 60-73.

Stoneking, M. and Soodyall, H. (1996). Human evolution and the mitochondrial genome. Current Opinions in Genetics and Development, 6: 731 – 736.

Stoneking, M. and Wilson, A.C. (1989). Mitochondrial DNA. In Hill, A. V. S. and Serjeanston, S. W. (editors). Colonization of the Pacific: a genetic trail, 215-245. Oxford University Press, New York.

Summerfield, J.A.; Sumiya, M.; Levin, M. and Turner, M.W. (1997). Association of mutations in mannose binding protein gene with childhood infection in consecutive hospital series. British Medical Journal 314: 1229-1232.

Thaler, E. and Lechner, K. (1981). Antithrombin III deficiency and thromboembolism. Clinical Haematology. 10: 369-390.

Torrioni, A.; Schurr, T.G.; Yang, C.-C.; Szathmary, E.J.E.; Williams, R.C.; Schanield, M.S. and Troup, G.A. (1992). Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. Genetics 130: 153-162.

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

Underhill, P.A.; Jin, L.; Lin, A.A.; Mehdi, S.Q.; Jenkins, T.; Vollrath, D.; Davis, R.W.; Cavalli Sforza, L.L. and Oefner, P.J. (1997). Detection of Numerous Y Chromosome Biallelic Polymorphisms by Denaturing High-Performance Liquid Chromatography. *Genome Research* 7: 996-1005.

Uzoigwe, G.N. (1995). Madagascar. The 1996 Grolier electronic encyclopaedia, 8th edition. Grolier Electronic Publishing, Incorporated.

Verin, P.; Koyyak, C.P. and Gorlin, P. (1970). The glottochronology of Malagasy speech communities. *Oceanic Linguistics*, 8: 26-81.

Vigilant, L. (1990). Control region sequences from African populations and the evolution of human mitochondrial DNA. Ph.D. Thesis. University of California, Berkeley.

Vigilant, L.; Stoneking, M.; Harpending, H.; Hawkes, K. and Wilson, A.C. (1991). African populations and the evolution of mitochondrial DNA. *Science* 253: 1503-1507.

Walberg, M.W. and Clayton, D.A. (1981). Sequence and properties of the human KB cell and mouse L cell D-loop regions of mitochondrial DNA. *Nucleic Acids Research* 9: 5411-5421.

Wallace, D.C. (1995). Mitochondrial DNA variation in human evolution, degenerative disease and aging. 1994 William Allan Award address, *Nucleic Acids Research* 9: 5411-5421

Watson, E.; Forster, P.; Richards, M. and Bandelt, H.J. (1997). Mitochondrial footprints of human expansions in Africa. *American Journal of Human Genetics* 61: 691-704.

Wilson, A.C.; Cann, R.L.; Carr, S.M.; George, M.; Gyllenstein, U.B.; Helm-Bychowski, K.M.; Higuchi, R.G.; Palumbi, S.R.; Prager, E.M.; Sage, R.D. and Stoneking, M. (1985). Mitochondrial DNA and two perspectives on evolutionary genetics. *Biol J Linn Soc* 26: 375-400.

Wilson, M.R.; Stoneking, M.; Holland, M.M.; DiZinno, J.A. and Budowle, B. (1993). Guidelines for the use of mitochondrial DNA sequencing in forensic science. *Crime Laboratory Digest*, 20: 68-77.

Wrischnik, L.A.; Higuichi, R.G.; Stoneking, M.; Erlich, H.A.; Arnheim, N. and Wilson, A.C. (1987). Length mutations in human mitochondrial DNA: direct sequencing of enzymatically amplified DNA. Nucleic Acids Research 15: 529-542.

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

Division of the Deputy Registrar (Research)

COMMITTEE FOR RESEARCH ON HUMAN SUBJECTS (MEDICAL)

Ref: R14/49 Dorf

CLEARANCE CERTIFICATE

PROTOCOL NUMBER M980606

PROJECT

Molecular Genetic Variation In The Malagasy

INVESTIGATORS

Miss C Dorf

DEPARTMENT

Dept of Human Genetics, SAIMR

DATE CONSIDERED

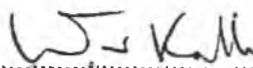
980626

DECISION OF THE COMMITTEE *

Approved unconditionally

DATE 980629

CHAIRMAN



(Professor P E Cleaton-Jones)

* Guidelines for written "informed consent" attached where applicable.

c c Supervisor: Prof T Jenkins

Dept of Dept of Human Genetics, SAIMR

Works2\ain0015\HumEth97.wdb\M 980606

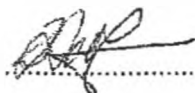
DECLARATION OF INVESTIGATOR(S)

To be completed in duplicate and **ONE COPY** returned to the Secretary at Room 10001, 10th Floor, Senate House, University.

I/we fully understand the conditions under which I am/we are authorized to carry out the abovementioned research and I/we guarantee to ensure compliance with these conditions. Should any departure to be contemplated from the research procedure as approved I/we undertake to resubmit the protocol to the Committee.

DATE 6/6/98

SIGNATURE



PROTOCOL NO.: M 980606

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES