

Mitochondrial Genome Variation and Metabolic Traits in a Maori Community

Miles Clifford Benton

A thesis submitted in partial fulfilment of the requirements for
the degree of Master of Biomedical Science

December 2009

School of Biological Sciences
Victoria University of Wellington

Abstract

The mitochondrion is the energy producing factory of cells and it has long been thought that disruption to mitochondrial systems is linked to energy metabolism dysfunction. Sequence variants in the mitochondrial genome are plausible candidate risk factors for numerous human diseases, and research has identified specific mitochondrial DNA (mtDNA) variants associated with metabolic disorders such as obesity and type-2 diabetes. As part of the Rakaipaaka Health and Ancestry Study (RHAS) it has been observed that the Maori community of Nuhaka (Ngati Rakaipaaka) have a high incidence of certain metabolic diseases, namely obesity and diabetes. The reason for this is not well understood, but is likely to be a combination of both current lifestyle (e.g. dietary) and ancestral genetic factors. This study set out to sequence the entire mitochondrial genome in a sample of RHAS Maori participants. The aim was to discover genetic variation that might be specific to this Maori community and test whether such variants are associated with diabetes and other metabolic traits.

This study used a novel RFLP assay to screen the mtDNA control region for Polynesian mtDNA ancestry. This established an initial group (n=30) with high levels of Maori mtDNA. Hypervariable (HVRI) sequencing was then used to generate a large dataset of sequences (n=94). This dataset was representative of individuals showing high Maori ancestry and aided the selection of 20 mtDNA's for Mitochip analysis. Combining the RHAS Maori HVRI sequences with those from previous studies indicated elevated variation in Maori mtDNA. Haplotype analysis identified 17 unique Maori haplotypes, 10 more than previously recorded. Mitochip resequencing has provided the first

complete Maori mtDNA sequences to date. When compared to other mtDNA sequences it was identified that RHAS Maori share similar haplotype markers with Polynesians.

Seven novel undocumented variants were found, as well as four variants that had previously been associated with various metabolic disorders. Mitochip analysis of mtDNA sequences revealed three variants which created a RHAS Maori specific signature; C1185T, G4769A, and T16126C. These variants also defined 3 unique mtDNA haplotypes within RHAS Maori, which are the first Maori specific haplotypes reported. Variant genotyping and correlation with metabolic traits identified significant associations within the wider RHAS Maori community. It was identified that RHAS Maori with T16189C showed elevation in both vitamin B12 levels and mean diastolic blood pressure. Individuals with the G4769A variant were shown to have significant increases in specific metabolic risk factors for cardiovascular disease. Conversely individuals with the more common A4769G variant were 2 times less likely to be diagnosed with diabetes.

The findings from this study have identified a series of potential markers of metabolic disease within the RHAS Maori community. The goal now is to understand how these markers interact with environmental variables to increase the risk of metabolic syndrome. Such an outcome may open the way to designing personalised intervention strategies (e.g. dietary) to increase the health and well-being of at risk individuals.

Acknowledgements

I would like to thank the following people who have provided support, encouragement, wisdom, and the odd poke of a stick for motivation over the past two years:

Firstly the iwi of Ngati Rakaipaaka and all individuals involved in the RHAS, without whom all this would not have been possible.

My supervisors; Dr Geoff Chambers, Dr Rod Lea, and Dr Donia Macartney.

Geoff, I remember coming to see you about a final grade for genetics, and walking away an hour later effectively enrolled in a Masters programme! I would like to thank you for your wisdom, editing, and all the little chats over the past years.

Rod, I don't know how you managed to turn around my complete draft twice in such a short amount of time, but I am eternally grateful. I thank you for the opportunity and the trust that you vested in me. It's been a hard slog, but it's done. Bring on the next one!

Donia, what can I say? I don't know how you put up with me when I first started out at ESR. I thank you for your patience, constant tutelage, and all the random talks about everything from building to ethics. You always seemed to be able to judge my sanity levels and find the right things to say to clam me down.

ESR for all the funding (both for the project, and myself), the amazing lab resources and support.

A big thanks to Liz Richardson, for all the support and opportunities. For inviting me to be part of Awhina, and Equity, I learnt and received so much, it was a privilege.

The staff of the Centre for Genomics and Proteomics, University of Auckland, for their hard work and patience 'tweaking' the Mitochip assay.

The Allan Wilson Centre for Molecular Ecology and Evolution for quickly turning around all my sequencing needs.

David Hall, for the use of several of his PhD figures.

All the great people that have come and gone from the office KK417, and the ‘others’ around SBS. Bridget and Pete, for the proofing and coffee. Emma and Surrey for putting up with my writing induced ranting.

Kirsty, for the years of friendship and through the trials and tribulations of Team Retardo, thank you. I know I made the joke about the stick poking at the start, but that came from a real place Kirsty!

My flatmates past and present, Matty, Raybag, Mark, Will, and Toookie for the constant distractions and late nights of ‘work’ (read Xbox).

My family, Mum and Dad for the constant support and love, and also making sure I keep my feet on the ground. It’s nice to be able to head home to Taranaki and feel truly at peace, if only for a weekend. My sisters Delle and Thea, for being there for me. I get great pride and inspiration from you both, even if I don’t tell you that enough. You are both amazing people, and I’m lucky to have you as sisters.

And last, but defiantly not least, Ellie. There is nothing I can write here that you don’t already know. You are my constant support, you’re belief in me is what has got me this far. You make me push myself to be better. Even in my worst moments you were there, as I know you will always be there, and for that I thank you with all my heart.

Table of Contents

ABSTRACT.....	II
ACKNOWLEDGEMENTS.....	III
TABLE OF CONTENTS.....	V
LIST OF TABLES.....	X
LIST OF FIGURES.....	XI
ABBREVIATIONS.....	XIII
 CHAPTER 1 INTRODUCTION.....	 1
1.1 Rationale.....	1
1.2 Hypothesis and Objectives.....	2
1.3 The Mitochondrion.....	3
1.3.1 Mitochondrial Structure.....	3
1.3.2 Mitochondrial Function.....	5
1.3.3 Oxidative Phosphorylation.....	5
1.3.4 The Mitochondrial Genome.....	7
1.3.5 Mitochondrial OXPHOS Genes.....	9
1.3.6 mtDNA Markers of Ancestry.....	10
1.3.6.1 Restriction Fragment Length Polymorphisms.....	12
1.3.6.2 mtDNA Haplogroup B.....	14
1.3.6.3 The 9-bp Deletion.....	15
1.3.6.4 Hypervariable Region I and the Polynesian ‘CCGT’ Motif.....	16
1.3.6.5 Hypervariable Region II.....	17
1.3.7 A New Era of mtDNA Sequencing.....	18
1.3.7.1 Affymetrix Mitochip.....	18

1.3.7.2	Applied Biosystems mitoSEQr.....	20
1.3.7.3	High-Resolution Melt Profiling.....	20
1.4	The Role of Mitochondria in Disease.....	21
1.4.1	Heteroplasmy and Mutational Threshold.....	23
1.4.2	ROS and Mitochondrial Disease.....	24
1.4.3	mtDNA Haplotypes Associated with Disease.....	27
1.4.4	Coupling and Thrifty Genes.....	28
1.4.5	mtDNA Disease Detection and Future Therapies.....	30
1.5	Genomics and Polynesian Ancestry.....	32
1.5.1	The Settlement of Remote Oceania by Polynesians.....	32
1.6	The Rakaipaaka Health and Ancestry Study.....	37
1.7	Specific Research Objectives.....	41
CHAPTER 2	MATERIALS AND METHODS.....	42
2.1	Study Population Characteristics.....	42
2.2	Overview of Experimental Design.....	43
2.3	Source of DNA and Ethics Approval.....	46
2.3.1	Blood Extraction.....	46
2.3.2	Buccal Swabs Extraction.....	47
2.4	Mitochondrial Control Region (HVRI) Analysis.....	47
2.4.1	Primers.....	48
2.4.2	HVRI PCR Protocol.....	48
2.4.3	Agarose Gel Electrophoresis.....	49
2.4.4	Purification of PCR products.....	49
2.4.5	DNA Quantitation: Pico Green assay.....	50
2.4.6	Sequencing of HVRI Products.....	50
2.5	Restriction Fragment Length Polymorphism Analysis.....	51
2.6	Mitochip Analysis.....	52

2.7	Software Packages.....	53
2.7.1	DNA Sequence Alignment and Phylogenetic Tree Creation.....	53
2.7.2	Mitochip Data Analysis in GSEQ 4.1.....	53
2.7.3	rRNA Secondary Structure Modelling.....	54
2.7.4	Protein Modelling with Sirius.....	55
2.7.5	mtPhyl: software for mtDNA analysis and phylogeny reconstruction....	55
2.8	Genotyping and Statistical Analysis.....	56
CHAPTER 3	RESULTS.....	57
3.1	DNA Extraction.....	57
3.2	Mitochondrial DNA HVRI Target Amplification.....	57
3.3	Identification of <i>Pst</i> I Restriction in RHAS Maori mtDNA.....	61
3.4	Mitochondrial DNA Sequencing.....	64
3.4.1	30 RHAS HVRI Sequences.....	71
3.4.2	94 RHAS HVRI Sequences.....	75
3.5	Complete mtDNA sequence data for 20 RHAS Maori individuals.....	78
3.5.1	Mitochip Analysis.....	79
3.5.2	Variation in 20 Maori Mitochondrial Genomes.....	79
3.5.3	Variants within Protein Coding Genes.....	82
3.6	Population analysis.....	84
3.6.1	Variant Frequency Tables.....	85
3.6.2	Haplotype Analysis.....	88
3.6.3	Phylogenetic Analysis.....	91
3.7	Secondary Structure Analysis of 12S and 16S rRNA.....	95
3.8	Protein Modelling Analysis.....	102
3.9	Genotype Results and Metabolic Trait Correlation.....	107

3.9.1	16189 and Diabetes.....	109
3.9.2	4769 and Diabetes.....	110
3.9.3	Hypothesis Generating.....	110
3.9.3.1	Association Testing 16189 and Metabolic Traits.....	111
3.9.3.1	Association Testing 4769 and Metabolic Traits.....	113
3.10	Summary of Results.....	115
 CHAPTER 4 DISCUSSION.....		116
4.1	HVRI Variation of 94 RHAS mtDNA.....	116
4.1.1	RFLP Assay.....	117
4.1.2	HVRI Sequencing and 16189.....	118
4.2	Variation within 20 complete RHAS Maori mt Genomes.....	119
4.2.1	Identification of a RHAS Maori Specific mtDNA Signature.....	120
4.2.2	Identification of Seven Novel Mitochondrial Variants.....	120
4.2.3	Strengths and Limitations of the Affymetrix Mitochip.....	121
4.3	Phylogenetic Analysis.....	122
4.3.1	Using complete mtDNA to explore RHAS Maori ancestry.....	124
4.3.2	Haplotype Analysis.....	125
4.3.2.1	RHAS Maori mtDNA Exhibits Increased Variation.....	125
4.3.2.2	Identification of Maori Specific mtDNA haplotypes.....	127
4.4	rRNA Secondary Structure Predictions.....	131
4.5	Protein Modelling Analysis.....	133
4.6	mtDNA Variants and Disease Association.....	138
4.6.1	Disease Association in the RHAS Maori Community.....	140
4.6.1.1	16189 possible disease associations in RHAS Maori.....	140
4.6.1.2	4768 possible disease associations in RHAS Maori.....	141

CHAPTER 5	CONCLUSION.....	143
5.1	Summary.....	143
5.2	Conclusion.....	146
	FUTURE DIRECTIONS.....	147
	APPENDICES.....	149
Appendix 1	RHAS community fact sheet.....	149
Appendix 2	mtDNA HVRI PCR primers and product.....	151
Appendix 3	Invitrogen 1Kb Plus DNA Ladder reference.....	151
Appendix 4	Guide to Pico-Green DNA standards.....	152
Appendix 5	Long range Mitochip PCR primers.....	152
Appendix 6	Pico-Green output of HVRI mtDNA concentrations.....	153
Appendix 7	Location and Identification for 20 RHAS Maori DNA specimens.....	154
Appendix 8	Maori HVRI GENBANK Accession numbers.....	155
Appendix 9	Frequency tables of RHAS variants.....	156
	REFERENCES.....	158

List of Tables

Table 3.1	Comparison alignment of four HVRI sequences.....	66
Table 3.2	30 RHAS Maori mtDNA HVRI sequences.....	72
Table 3.3	94 RHAS Maori mtDNA HVRI sequences.....	76
Table 3.4	20 complete RHAS Maori mitochondrial genomes.....	81
Table 3.5	Documentation of 7 novel mtDNA variants.....	83
Table 3.6	Population frequency analysis of mtDNA variants.....	86
Table 3.7	HVRI haplotype data combining AW, MM98, and MB datasets.....	88
Table 3.8	HVRI haplotype variants using extended AW and MB datasets.....	90
Table 3.9	Genotyping frequency data for 3 RHAS Maori variants.....	108
Table 3.10	Chi squared diabetes association testing 16189.....	109
Table 3.11	Chi squared diabetes association testing 4769.....	110
Table 3.12	Correlating 16189 and metabolic traits using t-test statistic.....	112
Table 3.13	Correlating 4769 and metabolic traits using t-test statistic.....	114
Table 4.1	RHAS Maori specific mtDNA haplotype markers.....	128

List of Figures

Figure 1.1	Mitochondrial structure and electron transport.....	4
Figure 1.2	The human mitochondrial genome.....	8
Figure 1.3	Haplogroup specific restriction sites in the mitochondrial genome.....	13
Figure 1.4	Mitochondrial markers of haplogroup B.....	15
Figure 1.5	Overview of the Affymetrix Mitochip™.....	19
Figure 1.6	The mitochondrial genome and a disease association.....	22
Figure 1.7	Mitochondrial heteroplasmy and disease.....	23
Figure 1.8	Possible routes to metabolic disease in the pancreatic β cell.....	25
Figure 1.9	Migration of Austronesian speaking peoples.....	34
Figure 1.10	Map of continent specific mitochondrial haplogroups.....	36
Figure 2.1	RHAS Maori autosomal ancestry evaluation.....	42
Figure 2.2	Workflow diagram.....	44
Figure 3.1	Extended RHAS Maori autosomal ancestry score graph.....	58
Figure 3.2	Gel electrophoresis of RHAS Maori HVRI PCR products.....	59
Figure 3.3	PCR product migration calibration curve.....	60
Figure 3.4	<i>Pst</i> I restriction site within HVRI.....	61
Figure 3.5	Gel Picture of restriction enzyme digest of 32 HVRI regions.....	63
Figure 3.6	HVRI sequence electropherograms.....	68
Figure 3.7	mtDNA variations within RHAS Maori HVRI.....	74
Figure 3.8	Phylogenetic tree of geographic mtDNA haplotypes.....	92

Figure 3.9	Phylogenetic tree of haplogroup B.....	94
Figure 3.10	Predicted 12S rRNA secondary structures.....	96
Figure 3.11	Predicted 16S rRNA secondary structures.....	100
Figure 3.12	Variant models of mtDNA encoded OXPHOS proteins.....	104
Figure 3.13	Hypothesis generating graph of 16189 variant.....	111
Figure 3.14	Hypothesis generating graph of 4769 variant.....	113
Figure 4.1	HVRI sequence showing the homopolymeric C-tract.....	118
Figure 4.2	Phylogenetic tree portraying haplotype B4.....	130
Figure 4.3	Protein model of COI showing redox centres.....	135

Abbreviations

Abbreviations used in text are described below. Unit abbreviations are described in S.I. [System Internationale (d'unites)] form, and standard notations are used for chemical formulae.

A	adenine
AGRF	Australian Genome Research Facility
ANT	adenine nucleotide translocator
ADP	adenine diphosphate
ATP	adenine triphosphate
AWC	Allan Wilson Centre
BMI	body mass index
Bp	base pairs
C	cytosine
del	deletion
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
EDTA	ethylenediaminetetra-acetic acid
ETC	electron transport chain
G	guanine
H ⁺	proton
HSCR	high sensitivity C-reactive protein
HVR	hyper variable region
IDDM	insulin dependent diabetes mellitus
ins	insertion
KSS	Kearns-Sayre Syndrome
LHON	Leber's hereditary optic neuropathy
M	molar
MELAS	mitochondrial myopathy
MERRF	myoclonic epilepsy and ragged-red fiber disease
MIDD	maternally inherited diabetes and deafness
μg	microgram

μl	microlitre
mg	milligram
ml	millilitre
mmol	millimole
mPTP	mitochondrial permeability transition pore
mtDNA	mitochondrial DNA
NADH	nicotinamide adenine dinucleotide
OXPHOS	oxidative phosphorylation
PD	Parkinson's Disease
PDH	pyruvate dehydrogenase complex
P _i	inorganic phosphate
RFLP	restriction fragment length polymorphism
RHAS	Rakaipaaka Health and Ancestry Study
ROS	reactive oxygen species
RNA	ribonucleic acid
rRNA	ribosomal RNA
SIDS	sudden infant death syndrome
SNP	single nucleotide polymorphism
T	thymine
TNF-α	tumour necrosis factor-α
tRNA	transfer RNA
T2DM	type 2 diabetes mellitus
UCP	uncoupling protein
VDAC	voltage dependent anion channel
WT	wild-type

Chapter 1 *Introduction*

1.1 Rationale

As part of the Rakaipaaka Health and Ancestry Study (RHAS) it has been observed that the Maori community of Nuhaka, New Zealand (i.e. Ngati Rakaipaaka) have quite a high prevalence of certain metabolic traits such as obesity and type II diabetes. In a recent community health survey it was shown that the rate of self-reported diabetes in the Rakaipaaka community was 12%, which is approximately 3 times greater than the New Zealand (NZ) average (NZ Health Survey 02/03). The reason for this is not well understood, but will no doubt be explained by both current lifestyle (e.g. dietary), and ancestral genetic factors, including perhaps underlying dysfunction in energy metabolism systems.

The mitochondrion is the energy producing factory of cells and it has long been thought that disruption to mitochondrial systems is linked to metabolic disorders. The objective of this study was to screen the entire mitochondrial genome in a sample of RHAS Maori using novel high-throughput sequencing technologies to identify mitochondrial genome (mtDNA) variants that might be specific to this Maori community and that might be correlated with metabolic-related traits such as diabetes.

The goal of this project was to identify genetic markers for metabolic syndrome in this Maori community and to understand how these markers interact with environmental variables to increase risk of metabolic disorders. Such an outcome should allow for more personalised intervention strategies (e.g. dietary) to be designed to increase the health and wellbeing of people with those genetic markers.

1.2 Hypothesis and Objectives

It is hypothesised that Maori have distinctive DNA sequence patterns (or signatures) in the mitochondrial genome, which have been shaped by a unique population history. It is suspected that these specific mtDNA signatures modify the risk of developing metabolic disorders such as diabetes and obesity in this population.

This project aims to utilise a new high-throughput microarray technology (Mitochip™ resequencing assay 2.0) to sequence the entire mitochondrial genome in subjects from the RHAS, and identify novel genetic markers of metabolic disease traits such as diabetes and obesity.

The specific objectives of this study are as follows:

1. Design and optimisation of an RFLP assay and HVRI PCR to aid with subject selection for Mitochip array analysis.
2. Obtain entire mtDNA sequence information from a subsample of Rakaipaaka individuals with Maori ancestry (n=20) using the Affymetrix Mitochip.
3. Align and compare this sequence information with independent worldwide sequences to identify group differences, i.e. a mtDNA signature that is specific to Rakaipaaka Maori.
4. Validate the signature in a larger community sample of the RHAS Maori population (n=400).
5. Correlate the validated mtDNA signature with metabolic traits (eg. BMI and diabetes).

1.3 The Mitochondrion

Type 2 diabetes and obesity are prevalent in the Maori community of Ngati Rakaipaaka (Lea 2009, Paper in prep). Metabolic traits such as these are known to be influenced by inherited genetic factors. As the mitochondrion is the metabolic powerhouse of the cell and contains its own genetic information, it was deemed that this research would concentrate on identifying unique variants within mtDNA that could possibly be contributing to certain metabolic disorders. This section provides an overview of the mitochondrion.

1.3.1 Mitochondrial Structure

Mitochondria are double membrane organelles are $\sim 1 \mu\text{m}$ in size and located in the cytoplasm of eukaryotic cells (Kakkar *et al.*, 2007). The basic structure of a mitochondrion consists of; outer membrane, intermembrane space, inner membrane, cristae, and matrix (Figure 1.1). The outer membrane is very similar in structure and composition to that of a eukaryotic cell membrane (Voet *et al.*, 2004). This outer membrane contains large amounts of the protein porin, which creates transport pores for the diffusion of small molecules (Ryan 2005). Any large proteins that enter the mitochondrion must be labelled at the N-terminus and actively transported by the protein translocase of the outer membrane (Ryan 2005). As small molecules are free to diffuse through pores, the intermembrane space is similar in content to the cytoplasm. Larger molecules however, are actively transported and reside in the intermembrane space, an example being cytochrome c (Lister *et al.*, 2005). The inner membrane contains all the proteins required for electron transport and production of ATP, as well as specific transport proteins and others that are needed for fusion and fission (Hansford 2002; Voet *et al.*, 2004). The cristae are the invaginating folds of the inner membrane

that extend into the matrix of a mitochondrion. These increase the surface area of the inner membrane, which effectively increases the amount of oxidative phosphorylation (OXPHOS) proteins (Voet *et al.*, 2004). Mitochondria that reside in tissue with higher energy requirements have many cristae, while those in less energy dependent areas have fewer cristae. The mitochondrial matrix contains the bulk of all mitochondrial proteins, and is the major site of aerobic metabolism (Hansford 2002). It also contains several copies of mtDNA and the ribosomes and associated proteins required for mtDNA transcription and translation (Voet *et al.*, 2004).

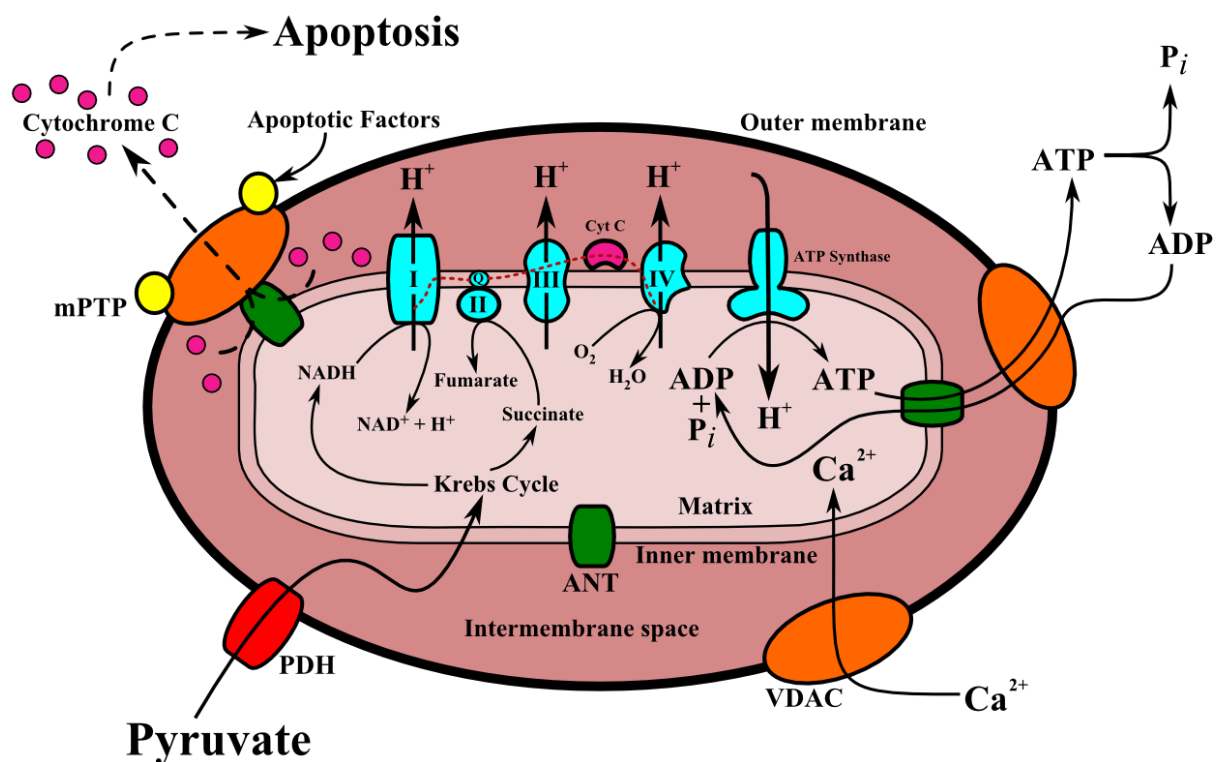


Figure 1.1: Overview of the mitochondrial electron transport chain (ETC) and other complexes. Flow of electrons through ETC complexes is indicated by the red dotted line. ANT, adenine nucleotide translocator; Cyt C, cytochrome C; mPTP, mitochondrial permeability transition pore; PDH, pyruvate dehydrogenase complex; VDAC, voltage-dependent transition pore. Adapted from Wallace (2006).

1.3.2 Mitochondrial Function

The mitochondrion is the energy producing factory of the cell (Wallace 2006). It is inside mitochondria that the conversion of metabolic energy into adenosine triphosphate (ATP) takes place (Figure 1.1). ATP powers all cellular processes, and it is the mitochondrion's job to ensure that oxygen consumption is correctly coupled to ATP production. The system put in place is known as oxidative phosphorylation (Hansford 2002). This requires the passing of electrons (via electron transport chain, ETC) through specific protein complexes, the OXPHOS proteins (Hansford 2002). Electron transport through this system enables the pumping of protons (H^+) from the mitochondrial matrix into the intermembrane space, generating a proton motive gradient (Schultz *et al.*, 2001). It is this gradient that drives the production of ATP. The mitochondrion has roles other than energy production, including cell signalling, programmed cell death (apoptosis), and control of the cell cycle (Wallace 2006; Kakkar *et al.*, 2007).

1.3.3 Oxidative Phosphorylation

Oxidative phosphorylation (OXPHOS) is the metabolic pathway that uses energy released from the oxidation of nutrition resources to produce ATP (Voet *et al.*, 2004). This process occurs inside mitochondria, with the initial energy being supplied in the form of the nutrient pyruvate. Glycolysis breaks down glucose producing the metabolic product pyruvate, which is an essential component of aerobic respiration (Hansford 2002). Pyruvate is transported into the mitochondria (Figure 1.1) via pyruvate dehydrogenase (PDH), where it is then converted via a series of reactions to acetyl CoA. Acetyl CoA is the first substrate essential for Krebs cycle (Hansford 2002). The Krebs cycle (citric acid cycle) provides several electron donors required for the ETC. The ETC pertains to a series of 5 enzymes (Complex I-V) located within the mitochondrial

inner membrane (Figure 1.1). These enzymes mediated the passage of electrons along the ETC via redox centers (Schultz *et al.*, 2001). The passage of electrons in this manner induces various charges which allow the pumping of protons from the matrix to the intermembrane space.

The first enzyme in the ETC is NADH dehydrogenase (complex I). This is a very large protein (~1000 kDa), in humans consisting of 46 subunits (Voet *et al.*, 2004). Complex I receives electrons from NADH supplied via the Krebs cycle, which it then transfers to ubiquinone (coenzyme Q10). During this transfer 4 protons (H^+) are pumped across the membrane (Hansford 2002). The second enzyme succinate dehydrogenase (complex II) also transfers electrons to Q10. Complex II is unique as it is the only ETC complex which has no mtDNA encoded subunits. It is also the only complex enzyme that has a role in the Krebs cycle (Horsefield *et al.*, 2004). In the Krebs cycle, complex II oxidises succinate to fumarate, transferring these electrons to Q10. Coenzyme Q10 when fully reduced becomes ubiquinol (Voet *et al.*, 2004). Ubiquinol is oxidised by cytochrome bc1 complex (complex III), resulting in the transfer of electrons and pumping of more protons. Complex III contains several cytochrome subunits; proteins that contain hemegroups and transfer electrons (Kakkar *et al.*, 2007). Oxidisation of ubiquinol allows complex III to transfer electrons to a mitochondrial associated protein, cytochrome c (Schultz *et al.*, 2001), which in turn transfers electrons to the final ETC complex, cytochrome c oxidase (complex IV). It is in this complex that the terminal electron receptor resides in the form of oxygen (Voet *et al.*, 2004). A series of hemegroups and metal co-factors in complex IV are utilised to transfer electrons to oxygen, ultimately reducing it to H_2O . Again potential energy created at this step allows for the pumping of protons (Hansford 2002). The net pumping of protons by ETC creates a chemiosmotic gradient, which allows the passage of protons, via proton-

motive force, through the enzyme ATP synthase (Schultz *et al.*, 2001). As these protons pass back down into the matrix, a conformational change in the 'head' of ATP synthase forces ADP and P_i (inorganic phosphate) to bind, resulting in the production of ATP (Dimroth *et al.*, 2000). This ATP is transported out of the mitochondrion and used to power cellular processes.

1.3.4 The Mitochondrial Genome

The human genome stretches over a total of 3 billion base pairs, and codes for approximately 20,000-25,000 genes (Venter *et al.*, 2001). In contrast, the human mitochondrial genome is ~16,600 base pairs in size with only 37 genes in total (Anderson *et al.*, 1981). However, while smaller in size there are far more protein coding genes per base in the mtDNA (~69%) than in the human nuclear genome (~1.5%). Essentially the entire human mtDNA encodes products; there is no 'junk' DNA. Unlike nuclear DNA which is only present in two copies per autosome, there can be in excess of thousands of mtDNA copies per cell (Chen *et al.*, 2005). Mitochondrial DNA (mtDNA) has a mutation rate that is ~15 times that of the nuclear genome (Wallace 2006). It is believed that this could be due to the apparent lack of mtDNA repair mechanisms and error checking, the lack of histones, as well as the constant production of reactive oxygen species (ROS), a by-product of ATP production (Chinnery *et al.*, 2000; Hoffman *et al.*, 2006).

The mtDNA is located within the matrix of mitochondria, typically exhibiting usually 2-10 copies per organelle (Kakkar *et al.*, 2007). Nuclear DNA and mtDNA are commonly thought to be of separate evolutionary origins, with mtDNA believed to have originated from the genomes of bacteria which were engulfed by an early eukaryotic cell ancestor (Davidov *et al.*, 2006). Together these organisms formed a symbiotic relationship,

allowing the flow of genes from the bacterial DNA into the host nuclear DNA. Human mtDNA is a double stranded, circular molecule (Anderson *et al.*, 1981). Within this are coded 37 genes; 13 dedicated to OXPHOS subunits, 22 to tRNAs, and 2 for rRNAs (Figure. 1.2).

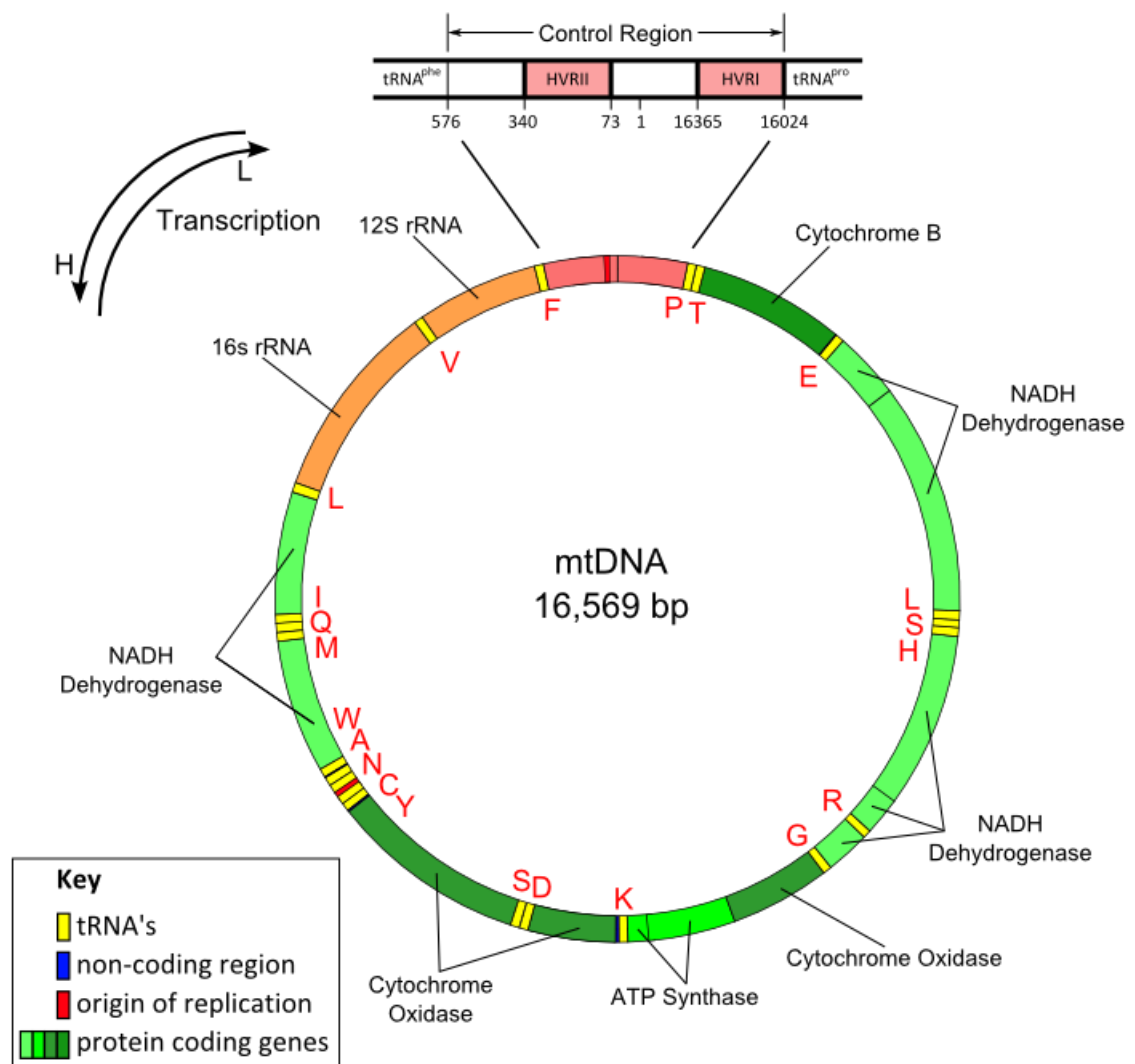


Figure 1.2: The human mitochondrial genome is a circular molecule that is 16.569bp in size. Human mtDNA contains 37 genes; 13 protein coding (green), 22 transfer RNA's (yellow; single letter code indicates coded amino acid), and 2 rRNA (orange). In the control region there are two hypervariable regions, HVRI and HVRII (pink). Origins of replication for each strand (red) and directions of transcription are detailed.

Being solely maternally inherited (Giles *et al.*, 1980; Chen *et al.*, 2005), mitochondrial genetics do not follow simple Mendelian laws (DiMauro *et al.*, 2007). In each cell there are potentially hundreds to thousands of mitochondrial genomes (Hoffman *et al.*, 2006). This means that instead of dealing with two alleles of every gene, as in Mendelian genetics, there are hundreds to thousands of gene copies within each cell, all of which may have variable mutations (Chinnery *et al.*, 2000). During cell proliferation, mitochondria are randomly segregated to the two new daughter cells at cytokinesis. Due to this random distribution of mtDNA segregating to daughter cells, the frequency of mutant mtDNA within a daughter cell can increase or decrease in frequency (Chinnery *et al.*, 2000).

1.3.5 Mitochondrial OXPHOS Genes

The majority of the OXPHOS proteins are encoded by the nuclear genome, with only 13 OXPHOS subunit proteins stemming from the mtDNA (Figure 1.4). These 13 subunits form integral components within four of the protein complexes essential for electron transport and OXPHOS (Kakkar *et al.*, 2007). A brief overview of the complexes containing mtDNA encoded subunits and an overview of their function will be given for reference:

- NADH Dehydrogenase (Complex I)

(mitochondrial genes: *NAD1*, *NAD2*, *NAD3*, *NAD4*, *NAD4L*, *NAD5*, *NAD6*)

NADH dehydrogenase catalyzes the transfer of electrons from NADH to coenzyme Q. Four protons are translocated per NADH, helping build the electrochemical gradient.

- Cytochrome b (subunit of Complex III)

(mitochondrial gene: *cyt b* – solely encoded by mtDNA)

The mitochondrial gene that codes for cytb subunit of the cytochrome bc₁ complex (complex III). This complex non-covalently binds two heme groups. Four conserved histidine residues are postulated to be the ligands of the iron atoms of these two heme groups. Complex III passes electrons on to cytochrome c and pumps protons out of the matrix.

- Cytochrome c Oxidase (Complex IV)

(mitochondrial genes: *COI*, *COII*, *COIII*)

Cytochrome c Oxidase accepts electrons from cytochrome c, and uses them to convert molecular oxygen into water and in the process translocates protons across the innermembrane helping establish the electrochemical gradient.

- ATP Synthase (Complex V)

(mitochondrial genes: *ATP6*, *ATP8*)

ATP Synthase is an enzyme that can create ATP from ADP and an inorganic phosphate (P_i) and energy. In mitochondria this energy is supplied via the protons moving against the electrochemical gradient set up via electron transport.

1.3.6 mtDNA Markers of Ancestry

Over the last 30 years mtDNA has shown considerable potential in the revealing of evolutionary secrets. Mitochondrial DNA is uniparently inherited; it is only passed down the maternal line (Giles *et al.*, 1980). Male mtDNA is either lost during the fusion of the sperm and ovum, or targeted for destruction within the oocyte (Giles *et al.*,

1980). While there is scarce evidence of male mtDNA inheritance (Schwartz *et al.*, 2002), it appears to be at a very diminutive level. The fact that mtDNA is haploid and does not exhibit the recombination seen in nuclear genes, means that markers (mutations) are conserved in blocks and passed on to future generations. These unique blocks are termed haplogroups. Haplogroups are defined by specific variants that create major branching points in the mitochondrial phylogenetic tree (van Oven *et al.*, 2009). They can be broken down into more specific haplotypes, which refer to an individual group of variants found on top of the core haplogroup defining variants (a collection of potential haplotypes belong to a broader haplogroup). Due to the nature of mtDNA inheritance and the path of early human evolution specific haplogroups have become established in varying parts of the world (for a review see Forster [2004]). These properties make mtDNA an ideal candidate for investigating the evolutionary relationship between inherited variants and among human sub populations.

Founder effects and bottlenecks have shaped the Polynesian mitochondrial genome in a unique way that allows it to be used as a powerful tool for identification. It was discovered that a 9-bp deletion existed in moderate frequencies in the mitochondrial haplotypes of Southeast Asia and Papa New Guinea, and later in Taiwan and the Pacific Island (Polynesia) populations (Hertzberg 1989). Alongside this was discovered a transition (T16189C), and two transversions, which create a homopolymeric tract (Poulton 1998). A few years later a unique set of transition mutations termed the ‘Polynesian Motif’ were identified in high frequencies in the mitochondrial control regions of Polynesians (Redd *et al.*, 1995). This motif includes three separate transition mutations; T16217C, A16247G, and C16261T. An individual with maternal Polynesian descent will most likely show the motif ‘CGT’. It is also important to note that the Polynesian Motif is only found in haplogroup B, in the presence of the 9-bp deletion.

1.3.6.1 *Restriction Fragment Length Polymorphism's (RFLP)*

It was in 1968 that Meselson and Yucan discovered a group of enzymes that recognised and cut specific sequences of DNA (Meselson *et al.*, 1968). These were termed restriction endonucleases', now more commonly known as restriction enzymes. Owing to the nature of these enzymes specificity a simple mutation, such as a substitution of a single nucleotide (e.g. single nucleotide polymorphisms or SNPs), can create or destroy a restriction site, thus leaving larger or smaller fragments of digested DNA. This recognised change in fragment size was first noticed in human β globin with the restriction enzyme *Hpa* I (Waikan *et al.*, 1978). The presence of the specific polymorphism was coined a restriction fragment length polymorphism (RFLP). It was later realised that restriction sites could be used to investigate genetic affinities within and between specific human ethnicities. Different mitochondrial haplogroups were found to contain unique restriction sites (Figure 1.3) which are seen to be continent specific (Ruiz-Pesini 2007).

The high mutation rate and population specific properties of mtDNA lend itself well to the nature of RFLP investigation. Early studies investigated mitochondrial polymorphisms with various restriction enzymes (Cann *et al.*, 1987), ultimately generating a group of RFLP based haplogroups (Francalacci *et al.*, 1999). Figure 1.3 shows the restriction sites and corresponding enzymes distributed around the mtDNA, haplogroups are indicated using alphabetic nomenclature.

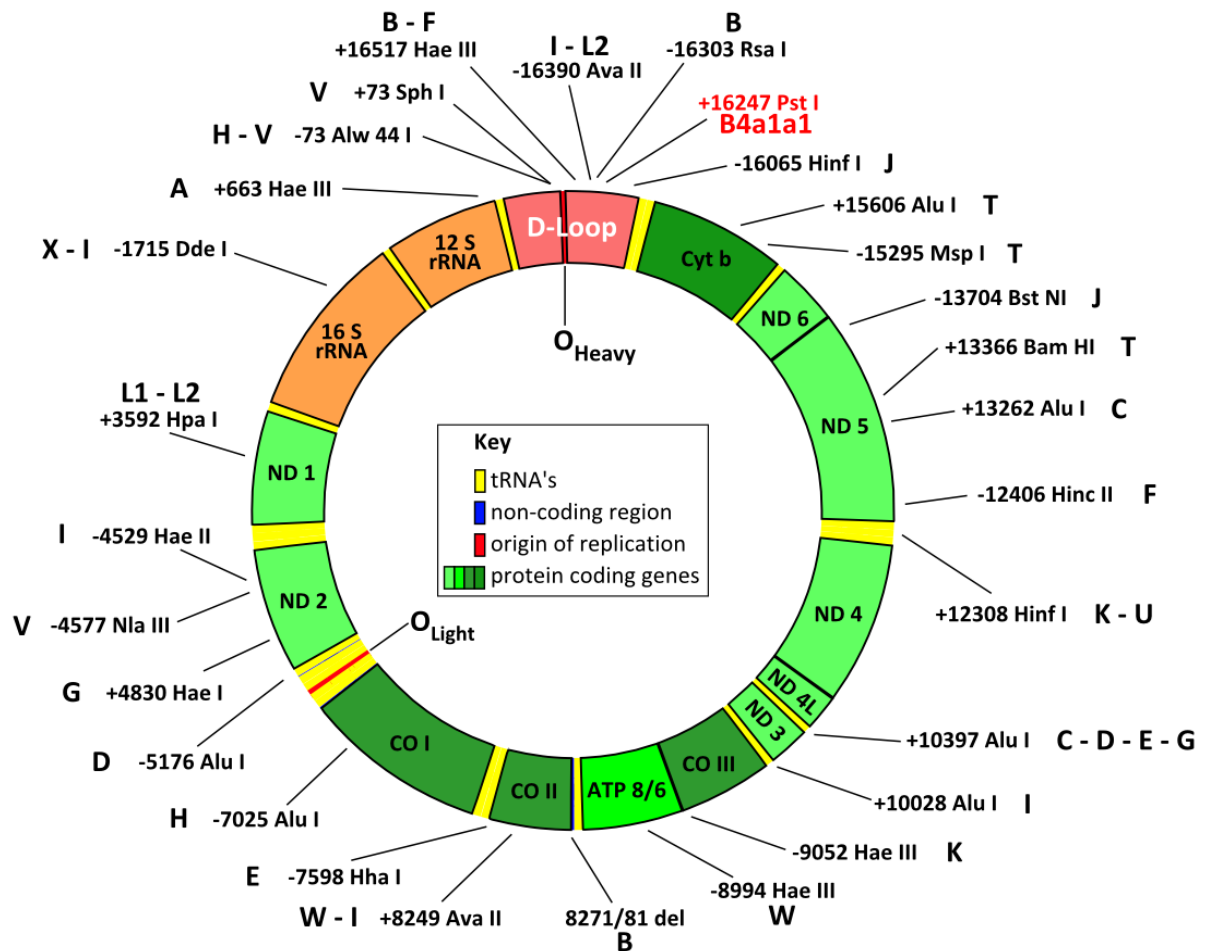


Figure 1.3: The position of haplogroup specific restriction sites in the human mtDNA. (+) indicates a restriction site is created by respective variant, (-) indicates a restriction site is lost. Restriction site labelled in red was discovered in this study and will be explored in more detail later. This figure has been modified from Francalacci et al. (1999).

1.3.6.2 *mtDNA Haplogroup B*

Haplogroup B is believed to have arisen in Asia some 50,000 years before present (YBP), stemming from the ancestral haplogroup R (van Oven *et al.*, 2009). Present day haplogroup B is spread over all of Asia, the Americas, Melanesia and Polynesia (Hertzberg 1989; Melton *et al.*, 1995). An important characteristic that defines mtDNA of this group is a 9-base deletion at 8281-8289 (Hertzberg 1989). Haplogroup B is broadly distributed in Asian populations. Recent population expansions associated with the spread of Austronesian languages appear to have brought these haplotypes from Southeast Asia into Polynesia around 5000-1000 YBP (Sykes *et al.*, 1995). The vast majority of Polynesian haplogroup B mtDNA's exhibit a set of polymorphisms that distinguishes them from similar types in other Asian populations (Lum *et al.*, 1994; Melton *et al.*, 1995; Redd *et al.*, 1995; Melton 1998). This set of mutations, which includes the T16217C, A16247G, and C16261T transitions (or "CGT"), has been called the "Polynesian motif" because of its uniqueness to Polynesian and related populations (Hagelberg and Clegg 1993; Hagelberg *et al.* 1994; Lum *et al.* 1994). The Polynesian motif is believed to have descended from mtDNAs bearing an ancestral origin in Taiwan, and continued as populations spread south into the Philippines, Indonesia, and Melanesia (Melton *et al.* 1995, 1998; Redd *et al.* 1995), although some favour an Indonesian source for this lineage (Richards *et al.* 1998). Melton *et al.* (1995, 1998) observed the Polynesian Motif at the highest frequency in Polynesians and coastal Papua New Guineans, and at modest frequencies in East Indonesians and Malays, with the intermediate form (CAT) occurring at the highest frequencies in Taiwanese aborigines and moderate frequencies in Filipinos and east Indonesians (Trejaut *et al.*, 2005).

1.3.6.3 9bp-deletion

Haplogroup B is found in several places around the world; Southeast Asia, South America, and the Pacific region (van Oven *et al.*, 2009). The major B haplogroup defining mutation is the intergenic 9-bp deletion between cytochrome oxidase II and tRNA lysine (Figure 1.4). Population frequencies of the 9-bp deletion have nearly reached 100% in much of the island Oceania region (Hertzberg 1989; Sykes *et al.*, 1995) and appear to have reached fixation Maori mtDNA of Polynesian ancestry (Whyte *et al.*, 2005). A complete phylogenetic tree of mitochondrial haplogroups and their individual haplotypes can be found on the PhyloTree database (van Oven *et al.*, 2009).

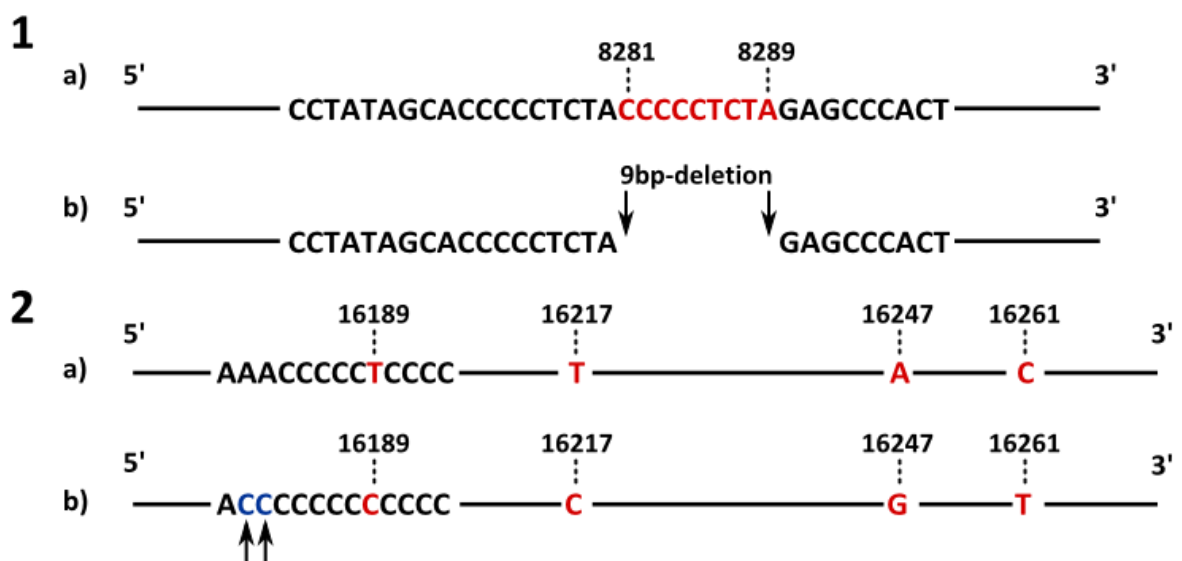


Figure 1.4: Mitochondrial markers of haplogroup B. 1. (a) the Cambridge Reference

Sequence mtDNA showing the position of the 9-bp deletion (indicated in red text). (b) the deleted region in mtDNA of haplogroup B. 2. (a) the mtDNA HVRI of the Cambridge Reference Sequence. (b) the required transitions for the formation of the Polynesian motif. The 16189 transition creates a poly-cytosine tract. The two adenines often undergo A to C transversions in many Polynesian and Maori (indicated by blue text and arrows). This figure has been adapted from Whyte *et al.* (2004).

1.3.6.4 *Hypervariable Region I and the Polynesian 'CCGT' motif*

The hypervariable region of the human mitochondrial genome is divided in two parts: hypervariable region I (HVRI), and hypervariable region II (HVRII). Both HVRI and HVRII are located within the D-loop (Anderson *et al.*, 1981; Andrews *et al.*, 1999). These regions are named such because they exhibit a higher mutation rate than any other genes within mtDNA (Stoneking 2000). Both regions are believed to contain sections of DNA that are involved in the initiation and regulation of mtDNA replication and transcription. There are mutational hotspots within the hypervariable regions, and the HVRI mutates faster than HVRII (Stoneking 2000).

Haplogroup B is of predominantly higher frequency within the Polynesian peoples of Oceania (Hertzberg 1989). The majority of Polynesian mitochondrial haplotypes are specifically B4a1a1 (Ingman 2001). This haplotype is represented by a specific set of variants located within the HVRI of the mitochondrial D-loop (Hagelberg *et al.*, 1993; Lum *et al.*, 1994; Redd *et al.*, 1995). This has been colloquially termed the Polynesian motif (Redd *et al.*, 1995), for the reason that frequencies are extremely high in all Polynesian populations. The Polynesian motif consists of three transition mutations which are all found within the HVR of the mtDNA. The transitions are as follows; T16217C, A16247G, and C16261T (Figure 1.4). As mentioned previously, haplogroup B is also typified by a T→C transition at position 16189 which creates a homopolymeric C tract. This is often referred to as the forth transition within the Polynesian motif (Whyte *et al.*, 2005). For the purpose of this thesis from this point on the Polynesian motif will be represented as 'CCGT' for the four transitions. This 'CCGT' motif defines the haplotype B4a1a1a.

In 2005 it was identified that there existed a possible pre-Polynesian motif in the coastal regions of Taiwan, the aboriginal Taiwanese Amis tribe (Trejaut *et al.*, 2005). This pre-Polynesian motif is again represented by the same four HVRI positions, yet it only exhibits three of the transitions (T16189C, T16217C, and C16261T). This then is ‘CCAT’, missing the A→G transition at 16247. Interestingly the researchers found that when they ran several phylogenetic tests, Polynesian mitochondrial sequences grouped within the Amis sequences (Trejaut *et al.*, 2005). Not only this, other sequences from areas in Micronesia and Melanesia (including the PNG coast) also grouped within these aboriginal Taiwanese samples. This supports the theory that present day Polynesians are decedents of Austronesian speaking migrants from coastal Taiwan, who moved down through coastal regions of PNG and surrounding islands, finally branching out into the Pacific region. It appears that it was during this migration into the Pacific region that the 16247 transition occurred, as it is at its highest frequencies in and around the Pacific countries (Redd *et al.*, 1995; Cann 2004).

1.3.6.5 *Hypervariable Region II*

Previous research has shown that the HVRII shows little to no variation within Polynesian and NZ Maori populations (Ingman 2001; Whyte *et al.*, 2005). The only informative markers within this region are shared by numerous different haplogroups, thus analysing the HVRII region is not as informative as the HVRI. Studies have also discovered that there is a high mutation rate within HVRI, explaining the increased variation (Stoneking 2000). It has been postulated that this could be due to the fact that the HVII contains vital DNA segments that are related to the functional aspects of replication and translation of mtDNA (Chang *et al.*, 1985).

1.3.7 A New Era of mtDNA Sequencing

Historically the mitochondrial genome has been investigated using RFLP and targeted polymerase chain reaction (PCR) and sequencing, mainly of the HVR's (Cann *et al.*, 1984). The last few years have seen a dramatic increase in number and availability of complete mitochondrial genomes, see databases such as Mitomap, mtDBA, and Phylotree (the latter having a curated database of 6729 as of October 2009). Bar the recent rise there were a limited number of complete sequences, due to the fact that using traditional methods (numerous overlapping PCR fragments and traditional Sanger sequencing) was expensive and time consuming. Recognising the need for more cost effective and less labour intensive mtDNA sequencing analyses there have been the recent release of several new technologies.

1.3.7.1 Affymetrix Mitochip™

Affymetrix have designed an alternative method based on their GeneChip technology. The Mitochip is a resequencing array that binds mtDNA fragments and interrogates each position of the mitochondrial genome (Figure 1.5). The design and first documented use of Mitochip technology was in detection of head and neck cancer (Maitra *et al.*, 2004). Information of positional variation is determined and a complete mtDNA sequence generated. This method has several benefits, the first being that it is very quick and requires little preparation in the way of PCR's and sequencing (the human contact time is reduced). One of the major technical developments was the introduction of the long-range PCR method. Using a set of specially designed primers and Takara LA Taq polymerase, the entire mitochondrial genome can be amplified in only three overlapping fragments, far less than traditional PCR methods. The Mitochip process takes no more than 48 hours and 8 chips can be processed at the same time.

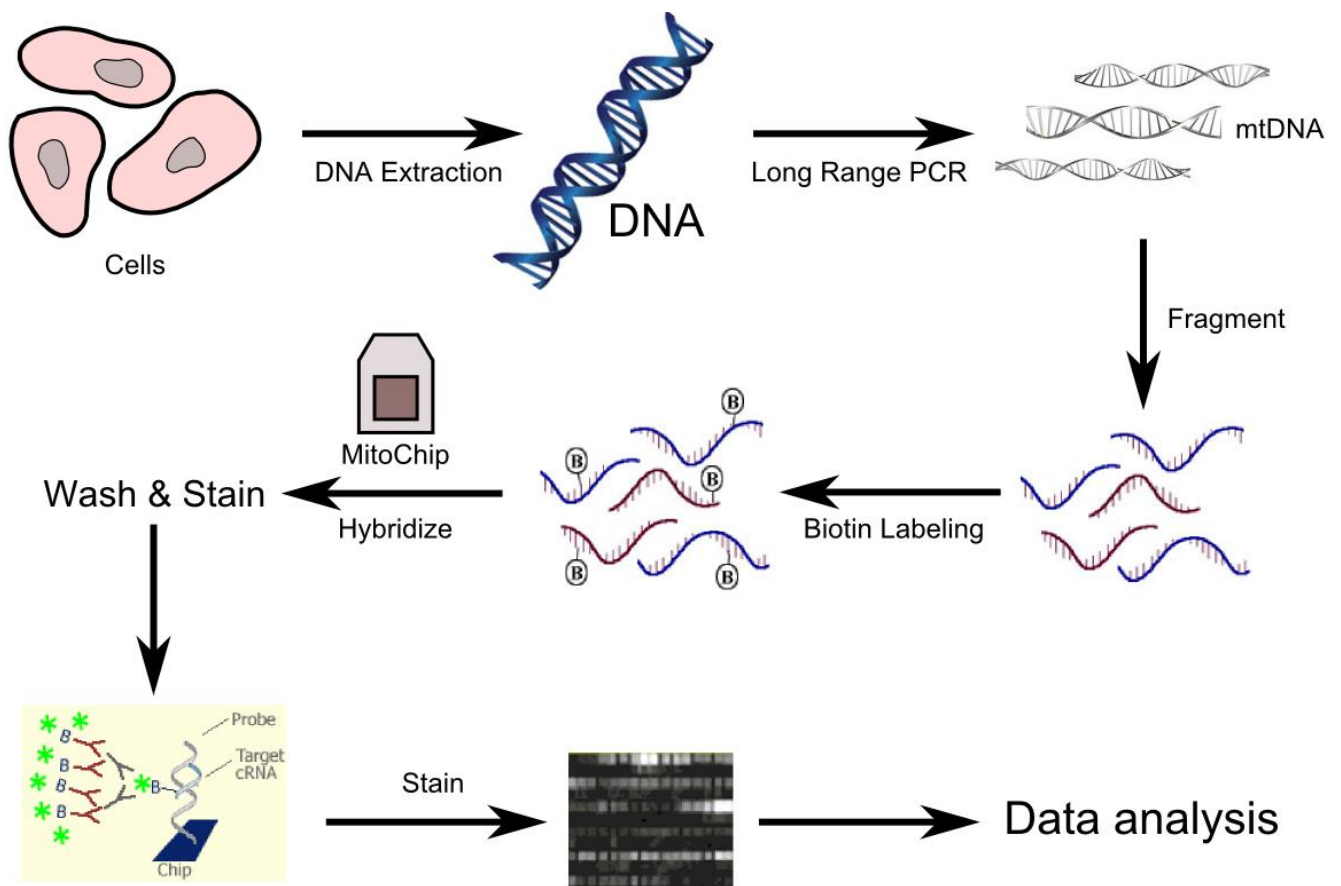


Figure 1.5: Overview of the Affymetrix Mitochip™. Extracted DNA is used as template for long range PCR. Three sets of primers are used to amplify the mtDNA. The PCR products are then fragmented, labelled with biotin, and hybridised to the Mitochip. The chip is then washed with streptavidin-phycoerythrin which binds to the biotin. A signal amplification step is carried out by washing the chip with anti-streptavidin antibody (goat), followed with biotinylated goat IgG antibody, and a final wash with streptavidin. This procedure ensures an ample signal is present when the chip is scanned with a confocal laser. When scanned the chip creates a distribution pattern, which is read by Affymetrix GeneChip Operating Software (GCOS). Data can then be analysed using GeneChip Sequence Analysis Software (GSEQ).

1.3.7.2 *mitoSEQr™*

Applied BioSystems have recently developed a resequencing assay which is able to identify sequence variants in entire mitochondrial genomes (including the control region). This new plate PCR method uses a set of specific universal primer pairs, which generate 46 overlapping mtDNA fragments covering the complete mitochondrial genome. These fragments are used as templates for sequencing using universal M13 primers. The assay design allows all primer pairs to anneal and operate at the same temperature, making it faster and less labour intensive. Being a direct sequencing assay makes the mitoSEQr protocol very accurate at detecting areas of heteroplasmy, as well as in-dels. Further details on mitoSEQr can be found on the Applied BioSystems website (<https://products.appliedbiosystems.com/>).

1.3.7.3 *High-Resolution Melt Profiling*

High-Resolution Melt (HRM) profiling has recently been proven as a method for accurately detecting specific mtDNA mutations (Dobrowolski *et al.*, 2009). It has the benefits of being a fast, simple, and low cost screen that looks to become widely adopted as a mutational diagnostic tool. It is also less labour intensive as the PCR amplification and melting curve are able to be run from the same plate. Other advantages include high levels of accuracy for detection of heteroplasmy, indels, and SNPs. High resolution melt profiling is however not an ideal method for complete reconstruction of the mtDNA, as this would end up being expensive and time consuming.

1.4 The Role of Mitochondrial DNA in Disease

Faults in the mtDNA can cause around 50 known diseases, some of which lead to disability and death (Hoffman *et al.*, 2006). About one in every 6500 people is affected by such conditions, which include fatal liver failure, stroke-like episodes, blindness, muscular dystrophy, diabetes and deafness. Metabolic syndrome comprises of numerous 'component' disorders, such as; insulin resistance, hyperglycemia, high blood pressure, visceral obesity, elevated triglycerides, and decreased HDL. Together these increase the risk of developing heart disease (cardiovascular disease) and diabetes. Several diseases are in fact directly related to specific mtDNA mutations, such as the large scale deletion Kearns-Sayre Syndrome (Puoti *et al.*, 2003).

The effects of mitochondrial disorders vary greatly. Due to the nature of mtDNA it is possible that the expression of mutant mtDNA can differ from tissue to tissue. It is evident that many mitochondrial diseases are tissue dependent. Muscle and nervous tissue are generally accepted as being most vulnerable to deleterious mutations due to the high energy requirement of these tissue types. Many mtDNA mutations have been associated with myopathy and neurological disorders such as Parkinson's disease (van der Walt *et al.*, 2003; Gaweda-Walerych *et al.*, 2008). It is now clear that mtDNA plays a role in many other diseases, with metabolic disorders such as insulin resistance (Poulton *et al.*, 1998), type-II-diabetes mellitus (Kim *et al.*, 2002), deafness (Ballana *et al.*, 2008), and obesity being some of the most important (Figure 1.6). It is however, not always easy to attribute disease phenotype to mitochondrial dysfunction. The major reason for this is that mtDNA can also vary inside the mitochondrion itself, a state known as heteroplasmy.

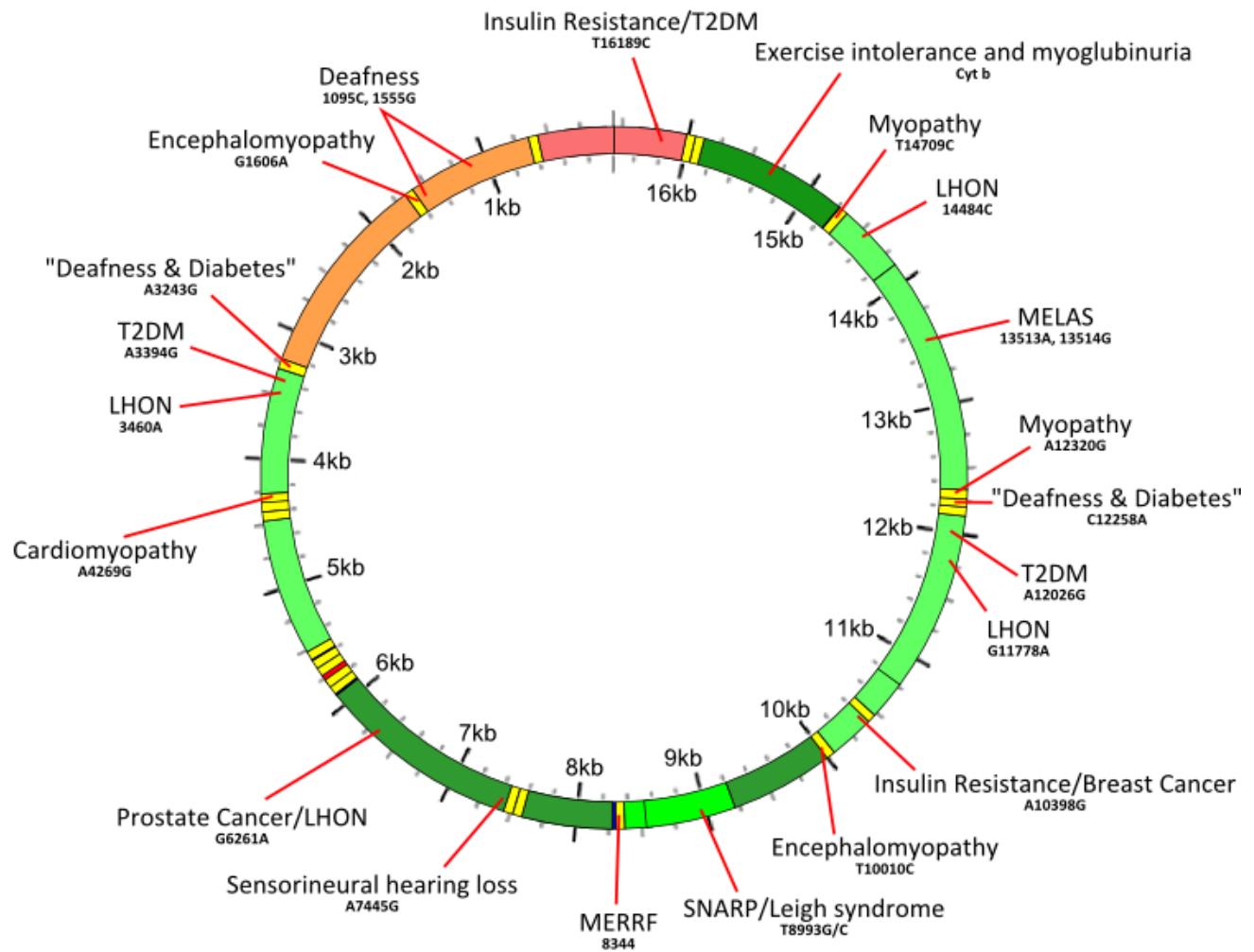


Figure 1.6: The mitochondrial genome and disease association. The position in mtDNA of the disorders is indicated along with variant type (i.e. transition/transversion, or insertion/deletion). LHON, leber's hereditary optic neuropathy; MELAS, mitochondrial encephalopathy lactic acidosis with stroke-like episodes; MERRF, myoclonic epilepsy and ragged-red fiber; T2DM, type II diabetes mellitus. Adapted from Wallace (2006).

1.4.1 Heteroplasmy and Mutational Threshold Effect

Heteroplasmy is the state of having two variant mtDNA populations within a cell (Acton *et al.*, 2007). This means that mitochondria can contain varying levels of two different types of mtDNA, one set being ‘normal’ the other being ‘mutant’. Different levels of heteroplasmy, also the tissue type, can mask the disease phenotype (Figure 1.7). The relative frequency of mutant to wild-type mtDNA within a cell, a tissue, or even an organism, determines whether an altered phenotype will be observed clinically (Ballana *et al.*, 2008). Thus, a specific threshold of mutations are required within each cell to alter metabolism, and a threshold of mutant cells within a tissue are required for the phenotype to be observed clinically (DiMauro *et al.*, 2007; Ballana *et al.*, 2008).

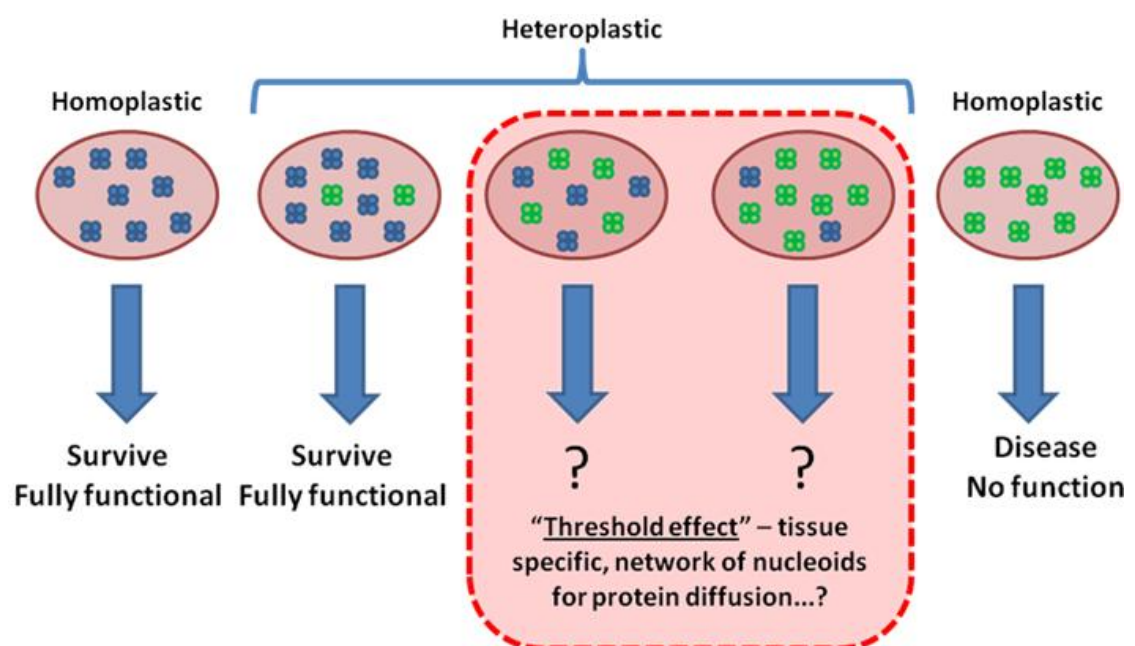


Figure 1.7: The role of mitochondrial heteroplasmy and disease. Varying degrees of heteroplasmy are shown, along with possible outcomes to cell survival. Wild-type mtDNA is represented in blue, mutant mtDNA in green.

Heteroplasmy can also be useful when tracing mtDNA ancestry. The most famous documented example of mtDNA heteroplasmy is the historical case of the Romanov's, the Russian royal family that were brutally murdered during the Russian revolution. Independent lab testing of mtDNA extracted from exhumed remains identified a position of heteroplasmy shared between close maternal relatives (Ivanov *et al.*, 1996). Using this positional variation it was determined that over several generations the heteroplasmic variant had segregated to homoplasmy in living relatives that had supplied DNA. This was most likely attributed to a mtDNA bottleneck effect over the several generations of mtDNA transmission.

1.4.2 Reactive Oxygen Species and Mitochondrial Disease

Reactive oxygen species (ROS) are free radicals that contain the oxygen atom (Voet *et al.*, 2004). They are very small molecules that include oxygen ions and peroxides and can be either inorganic or organic (Brand 2007). They are highly reactive due to the presence of unpaired valence shell electrons. ROS form as a natural by-product of the normal metabolism of oxygen and have important roles in cell signalling (Figure 1.8). Many of mitochondrial functions actually rely on specific forms of ROS, such as apoptosis (Cadenas 2004). Raquel Moreno-Loshuertos and colleagues reported that ROS could also act homeostatically (Moreno-Loshuertos *et al.*, 2006). Their research shows that ROS can control processes such as increasing mtDNA replication, the activation of uncoupling proteins, and growth factor signalling. These properties of ROS may be a contributing factor towards the masking of phenotypic consequences of mtDNA variation (Baughman *et al.*, 2006). However, during times of environmental stress (e.g. UV or heat exposure) ROS levels can increase dramatically, which can result in significant damage to cell structures. This cumulates into a situation known as oxidative stress.

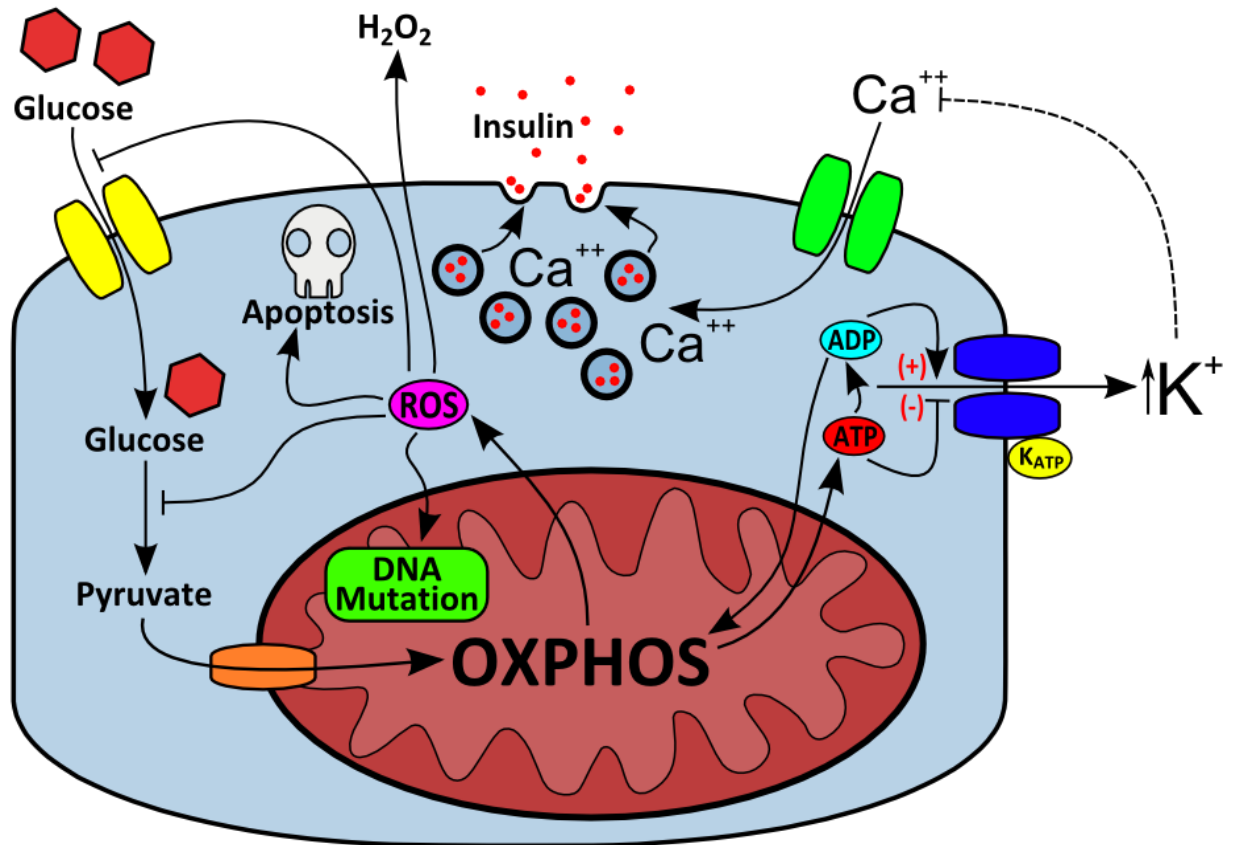


Figure 1.8: Possible routes to metabolic disease in the pancreatic β cell. Effecting the production of ATP can ultimately alter the amount of insulin released by the β cells. Decreasing ATP production favours ADP, which opens the membrane potassium channels. Extracellular potassium can feed back, inhibiting calcium uptake, which is directly related to insulin release. Other outcomes to errors in mitochondrial function include the increased production and build up of reactive oxygen species (ROS), which can be harmful to the cell. Adapted from Wallace (2006).

Recent studies have been investigating the relationship between certain mtDNA variants and increased ROS levels (Brand 2007; Gaster 2007). One of the major correlations seems to be that increasing ROS levels lead to increased instances of insulin resistance (Lin *et al.*, 2005; Houstis *et al.*, 2006). Insulin resistance occurs when the normal action of insulin is impaired, leading to muscle, fat and liver cells unable to illicit an insulin response (Stannard *et al.*, 2004). Production of insulin increases to try and combat sensitisation of cells, this leads to increased insulin in blood plasma. Ultimately the action of insulin decreases leading to increased blood glucose (Gaster 2007). High

plasma insulin levels and high blood glucose owing to insulin resistance is believed to be the precursor and primary cause of type II diabetes and many metabolic disorders.

Researchers are still debating as to how insulin resistance actually occurs (Taubes 2009). Numerous reports are putting forth specific mtDNA variants as contributing or possibly having a causal role in insulin resistance (Poulton *et al.*, 1998; Evans *et al.*, 2003; Petersen *et al.*, 2003; Bhat *et al.*, 2007). Significance of the mutation T16189C having a role in insulin resistance has been discussed in many studies (Poulton *et al.*, 1998; Kim *et al.*, 2002; Lin *et al.*, 2005; Weng *et al.*, 2005; Bhat *et al.*, 2007; Liou *et al.*, 2007). This mutation is found within the non-coding control region (HVRI), it is also a marker for haplogroup B alongside the intergenic 9-bp deletion. It is hypothesised mitochondria with this mutation become less competent in dealing with oxidative stresses (Bhat *et al.*, 2007). Research has associated this variant with increased levels of oxidative damage, not only in diabetic patients but also in non-diabetic patients harbouring the mutation (Lin *et al.*, 2005; Bhat *et al.*, 2007). One possible pathway is via mitochondria containing this mutation being prone to producing higher than normal levels of ROS. A build up of ROS in the mitochondria can have many pathologic effects, including DNA damage, inhibition of glucose metabolism, and promotion of apoptosis (Figure 1.8).

ROS levels appear to be a trigger for insulin resistance in numerous settings (Houstis *et al.*, 2006). Houstis *et al.*, (2006) found they were able to induce ROS levels with either cytokine TNF- α or glucocorticoid dexamethasone. They noticed that as ROS levels increased so did insulin resistance. They then tested whether ROS has a causal effect in insulin resistance by applying pharmacological and genetic interventions. While these treatments did not fully prevent the development of insulin resistance, they did ease it.

1.4.3 *mtDNA Haplogroups Associated with Metabolic Diseases*

Phylogenetic studies have shown that human populations can be divided into mtDNA haplogroups which are defined by specific single nucleotide polymorphisms (SNP's) (van Oven *et al.*, 2009). These haplogroups are representative of accumulated mutations which have been inherited along discrete lineages. Some of these specific inherited variants might be predisposing factors in numerous diseases; insulin resistance (Poulton *et al.*, 1998) and diabetes (Chinnery *et al.*, 2005; Lin *et al.*, 2005; Bhat *et al.*, 2007), cancer (Bai *et al.*, 2007; Ray *et al.*, 2009), Alzheimer's disease (van der Walt *et al.*, 2004; Santoro *et al.*, 2006), and Parkinson's disease (Van Der Walt *et al.*, 2002; Ghezzi *et al.*, 2005; Gaweda-Walerych *et al.*, 2008).

The variant T16189C, common to haplogroup B (Polynesian motif, this variant is heavily conserved in Polynesian and Maori populations) has been associated in numerous studies with insulin resistance and a possible role in type II diabetes (Poulton *et al.*, 1998; Kim *et al.*, 2002; Livesey *et al.*, 2004; Bhat *et al.*, 2007; Liao *et al.*, 2008). Other markers from haplogroup B have also been associated with metabolic traits. The previously mentioned 9-bp deletion has a role in disease. It has been reported that there is a high prevalence of 9-bp in Taiwanese populations with MELAS or MERRF (Liu *et al.*, 2005).

Some haplogroups have been shown to provide a certain degree of protection from certain diseases. The Asian haplogroups N9a and N9b appears to offer protection from specific metabolic disorders. N9a has been associated with protection against type 2 diabetes in Asian females (Fuku *et al.*, 2007), while N9b has been associated with protection against myocardial infarction in Asian males (Nishigaki *et al.*, 2007). Protection is also seen in arctic populations. Several non-synonymous mitochondrial

mutations appear to be advantageous in Arctic populations (Mishmar *et al.*, 2003; Ruiz-Pesini *et al.*, 2004). These two studies suggest that non-synonymous mutations in mtDNA encoded OXPHOS proteins might disturb ATP production in a way that leads to the production of extra heat. These mutations could then be viewed as adaptive features.

The A10398G has been found to be both protective and harmful. In Irish populations (mutation is associated with haplotype J) this mutation has been associated with longevity, found to be well represented in haplotypes of centenarians (Ross *et al.*, 2001). The same variant in Northern Indian populations has been correlated with increased ROS levels and insulin resistance, possibly also predisposing individuals to type 2 diabetes (Bhat *et al.*, 2007).

1.4.4 Coupling and Thrifty Genes

As humans migrated out of Africa they came in contact with varying environmental conditions and challenges (Forster 2004). Mitochondrial DNA is theorised to have played a large role in the survival and success of migrating human populations (Wallace 2006). This advantage revolves around the process of coupling. Coupling explains the efficiency of the joint effort of OXPHOS complexes and electron transport to produce ATP (Schultz *et al.*, 2001). This is an indirect process, being maintained via the generated proton electrochemical gradient. Loosely coupled mitochondria are able to disengage the processes of OXPHOS ETC and ATP production (Malmgren *et al.*, 2009). Using either uncoupling proteins (UCPs) or nonspecific leak pathways, they have the ability to usurp potential redox energy and dissipate it as heat rather than synthesising ATP. This property therefore allowed certain populations the ability to

persevere in colder climates, and forced those with tightly coupled mitochondria to the warmer more temperate climates (Wallace 2006).

Tightly coupled mitochondria have increased efficiency for using available calories to produce ATP. Fewer calories are needed in a tightly coupled system as the mitochondria are able to generate a larger electrochemical potential, thus produce more ATP. This would have been an advantage in warmer climates (Wallace 2006). Not being able to uncouple however, can have drawbacks. Over eating can lead to outcomes other than fat storage. An excess of calories can overload OXPHOS, effectively stalling electron transport. Overloaded electron transport causes the electron carriers to become highly reduced, which in turn leads to the increased production of ROS in an attempt to utilise some of the surplus electrons (Wallace 2006). ATP production decreases due to the stalling of electron transport, thus the ATP:ADP ratio changes to favour ADP (Maechler *et al.*, 2006). An increase of ADP within the cell activates the potassium ATP (KATP) channel which inhibits the cellular influx of Ca^{2+} (Section 1.4.2, Figure 1.8). This is important to consider in the context of the β -cells of the pancreas (Maassen *et al.*, 2006). Decreasing the amount of Ca^{2+} taken up by the pancreatic islet cells, directly affects insulin release. Less insulin release means higher blood glucose. This is one hypothesised route to insulin resistance (Maassen *et al.*, 2006; Maechler *et al.*, 2006).

1.4.5 *mtDNA Disease Detection and Future Therapies*

In the past sequencing mtDNA used to be relatively difficult and very expensive for complete mitochondrial genomes. The majority of the earlier mtDNA research focused on RFLP (Restriction Fragment Length Polymorphism) analysis and sequencing of the hypervariable regions (Francalacci *et al.*, 1999). Slowly it has been recognised that complete mtDNA sequences offer the ability to analyse more fully the role and function of mitochondria (Zhou *et al.*, 2006). Complete mitochondrial sequencing has provided an advanced resolution for phylogenetic studies (Torroni *et al.*, 2006), and is going to be of great benefit for further investigation into how metabolic disorders effect not only individuals but also specific populations (so effectively improve how we understand the mitochondrion).

One such example of technology that has evolved for analysing the mtDNA is the Mitochip, designed by Affymetrix (Maitra *et al.*, 2004). This is a gene array chip that is able to resequence the entire mtDNA, provide additional information for the control and hypervariable regions (HVR), as well as a host of common known variants. A simple PCR recoups the mtDNA from a genomic preparation, then it is fragmented and hybridised to the Mitochip. Not only does the Mitochip lend itself to provide complete mtDNA sequence information, it has several other novel uses. Several groups have found a potential use for the Mitochip in the early detection of cancers (Maitra *et al.*, 2004; Zhou *et al.*, 2006). In fact one group have managed to successfully identify breast cancer patients by just using mtDNA from nipple aspirate (Jakupciak *et al.*, 2008). This process offers much faster results than traditional sequencing, and is very non-intrusive as nipple aspirate fluid is easy and painless to collect. This technique could then be utilised for other cancers such as prostate or urinary tract cancers by analysing mtDNA from cells in urine.

Another Mitochip application is forensic identification (Vallone *et al.*, 2007). Previously HVR data has been used for forensics, but now the ability to increase the resolution of mtDNA means that it becomes an even more potent and accurate tool (Vallone *et al.*, 2007). Therefore the next few years should see the amount of available complete mtDNA sequences increase dramatically (~6729 online, PhyloTree). The offshoot of this in terms of metabolic disorders means that more mtDNA sequences from an extended variety of population backgrounds will become available online, and will allow for increased interrogation.

Due to the maternal inheritance of mtDNA, mitochondrially inherited disorders are nearly impossible to treat (Dimauro *et al.*, 2001). Recently the first viable treatment for mitochondrial disorders was explored in monkeys (Tachibana *et al.*, 2009). This procedure has been termed the three-parent embryos. The principle is to effectively transfer a healthy nucleus from a mitochondrial defected cell into an embryonic cell that has had its nucleus removed (Tachibana *et al.*, 2009). This new cell has a healthy mitochondrial background. The procedure has been tested recently in monkeys, and some have given birth to healthy offspring (Tachibana *et al.*, 2009). The offspring show nuclear DNA from the donor female and male, and mtDNA from the cytoplasm donor. If the three-parent embryo procedure became a viable option for humans, it would be the ideal treatment for mothers with metabolic syndrome who wished to one day have children.

1.5 Genomics and Polynesian Ancestry

When thinking about RHAS and the study being conducted it is important to consider the effect that ancestry and migration have on the subsequent origin of specific variants and genes (in this case the mtDNA). Ancestry has shaped the mitochondrial genome in all human populations and this holds true for RHAS. Recounting ancestral history is valuable not only from an evolutionary perspective, but also for identifying associations between mtDNA and disease, with the goal of being able to deliver the best possible diagnosis and treatment.

1.5.1 *The Settlement of Remote Oceania by Polynesians*

Now that the ‘out of Africa’ model is generally accepted (Vigilant *et al.*, 1991; Forster 2004), the question becomes re-orientated on the process of how we got to where we are now. The vast majority of populations around the world are able to be traced back through their ancestral intermediates to Africa, but one region has perplexed evolutionary scientists for nearly 50 years. This area is Oceania.

Oceania encompasses the Pacific Islands and New Zealand, and is thought to have been the last place in the world to be populated. As such there is great interest in discovering where pre-Polynesian populations originated from (Cann 2004). This interest has sparked major debate within the scientific community. It is generally accepted that Oceania was populated by Austronesian speaking peoples (Bellwood 1991). The debate centres on where these people originally came from, and how they reached their final destinations. Early evidence arose in the form of linguistic (Bellwood 1991; Gray *et al.*, 2000) and archaeological information (Kirch *et al.*, 1988), and recently genetic analysis has been used to reinforce the picture. Genetic markers have been invaluable tools in

the attempt to elucidate the facts about the settlement of Oceania. Analyses of the Y-chromosome are pointing towards a Melanesian origin, or a stopover in this area for a longer period of time (Hagelberg 1999; Kayser *et al.*, 2000). More recent research on ALDH points towards a South East Asian origin, with NZ Maori sharing the many of the same allele's as some Asian populations (Chambers *et al.*, 2002). The plausible theories incorporate a combination of these pieces of evidence, suggesting models that provide a variety of answers.

There are three main theories that have previously emerged to try and explain the origins of Polynesians: the Slow Boat model (Oppenheimer 2001), the Express Train model (Bellwood 1991), and the Entangled Bank model (Terrell 1988). The Slow Boat model proposes Island Southeast Asia as the 'melting pot' for early Polynesian culture. It assumes that migrants from southern China travelled down into island Southeast Asia, where they resided for many years (~15000). Gradual, complex interactions with neighbouring islands additionally shaped the Polynesian culture and genome, before populations began travelling further abroad, finally reaching Polynesia (Oppenheimer 2001). The Express Train model suggests that the migration into Oceania occurred rapidly, in the space of ~6000 years. This theory assumes Taiwan as the original birth place of Polynesian migration. Early ancestors migrated from mainland China and Southeast Asia to colonize Taiwan, and from there rapidly spread out into the Pacific (Diamond 1988). The Entangled Bank model supposes a long history of cultural and genetic interactions amongst south-east Asians, Melanesians, and already-established Polynesians. This model suggests that these Melanesian and Polynesian peoples were in constant contact with each other, making a complex social and trade network (Terrell 1988).

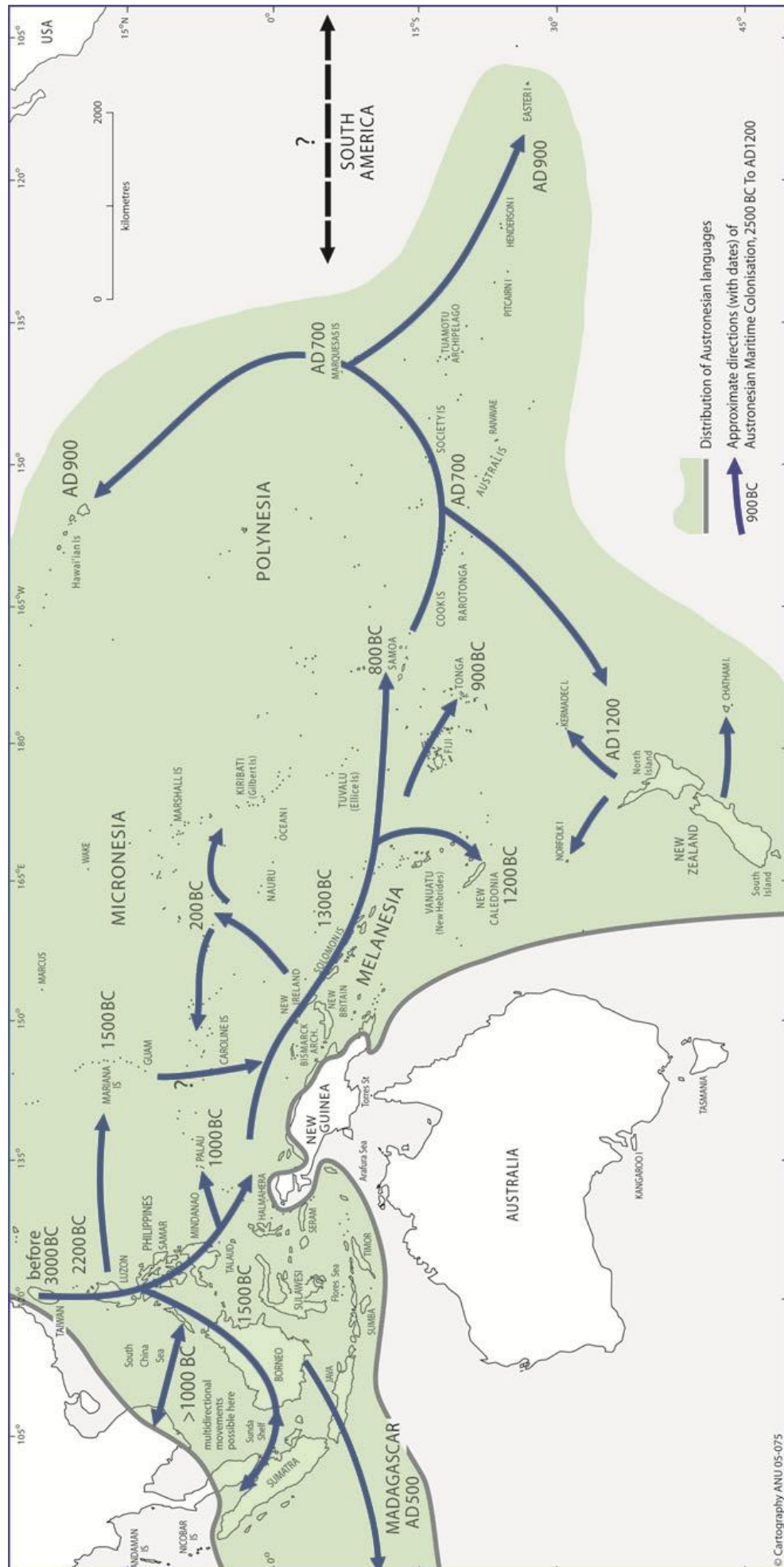


Figure 1.9: Map outlining migratory paths of Austronesian speaking populations. Adapted from Bellwood *et al.*, (unpublished).

This is a brief overview as it stands today from all the available evidence (Figure 1.9). Between 3000-1000 BC speakers of Austronesian languages migrated through island Southeast Asia. Current linguistic (Bellwood 1991; Lum 1998; Gray *et al.*, 2009) and archaeological evidence (Kirch *et al.*, 1988) suggest that these people originated from Taiwan (Figure 1.9). Carbon dating has placed the appearance of Lapita culture in the Bismarck Archipelago, northwest Melanesia, ~3000 years BC (Kirch *et al.*, 1988). This culture is likely to have spread from China/Taiwan. The Lapita culture has been found as far as 6000 km east of the Bismarck Archipelago (Kirch *et al.*, 1988). According to dating its spread occurred rapidly, between 1300 and 900 BC, ultimately reaching as far as Fiji, Tonga and Samoa (Kirch *et al.*, 1988). It is calculated that these islands were populated about 2000 years ago. It was during this time that the Polynesian culture began to develop and, around 300 BC, spread further out into remote Oceania populating the Cook Islands, Tahiti, the Marguesas Islands, and the Tuamotus. Easter Island was settled by the Polynesians around 300 AD. Around AD 1200 a grouping of Polynesian peoples migrated to and settled New Zealand (Aotearoa), these people are known as the New Zealand Maori (Cann 2004).

One major genetic clue that has proven to be important to the elucidation of this migration is the variation that exists within the genome of the human mitochondrion (Cann *et al.*, 1984; Cann *et al.*, 1987; Lum *et al.*, 1994). As mentioned previously, mitochondria are found within cells and are unique in the fact that they have their own genomic material. Mitochondrial DNA is only passed down via the maternal line, in contrast to nuclear DNA which is obtained from both parents (Giles *et al.*, 1980). This property of mtDNA allows it to be traced back to a single common ancestor. As such, all human mtDNA should be an exact copy of this original, but this is not the case. Mutation plays a major role in the mtDNA, which exhibits a mutational rate ~15 times

higher than that of the nuclear genome (Stoneking 2000). As human populations have migrated from Africa different mutations have arisen within the mtDNA. Those that persist for long enough will be passed down the maternal line to descendants (Forster 2004). Over time as human populations have moved apart, the number of mutations has grown, creating very different mtDNA (haplogroups) when compared to the original population. These haplogroups can be compared, and relationships established between populations. Mitochondrial haplogroups have been used to trace migration and settlements of populations (Figure 1.10).

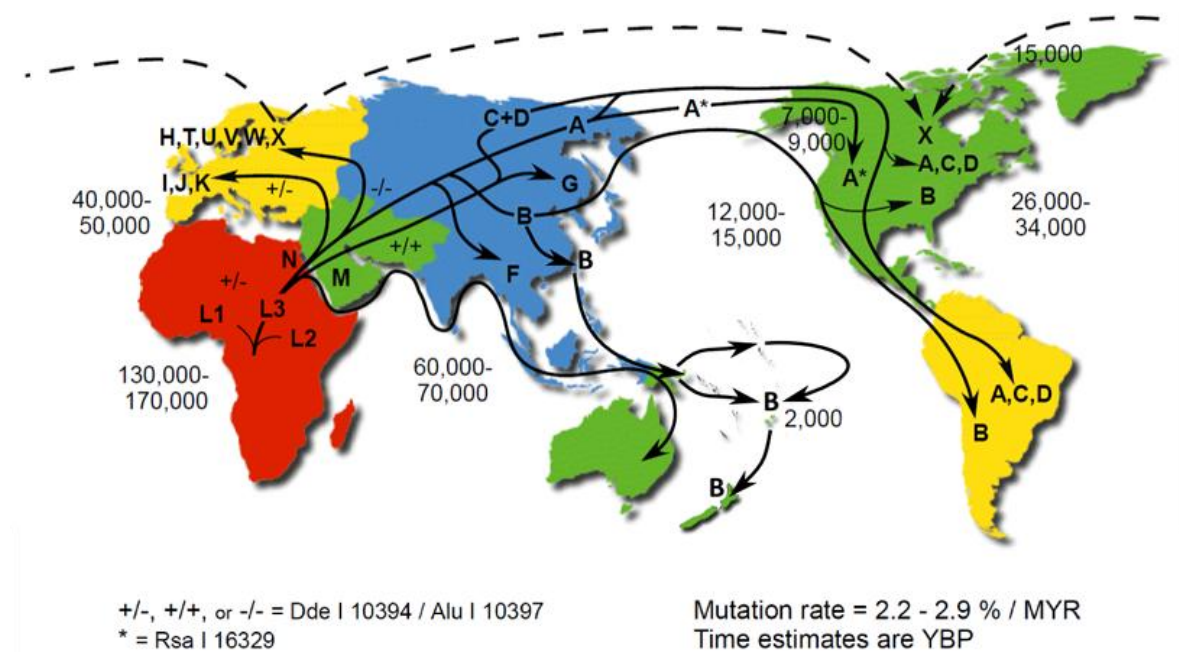


Figure 1.10: Map of continent specific mitochondrial haplogroups. Haplogroups are indicated with a letter, +/- indicates the presence or absence of specific restriction sites. Estimates of migration time are in years before present. This figure has been modified from the MITOMAP website: <http://www.mitomap.org/WorldMigrations.pdf>

1.6 The Rakaipaaka Health and Ancestry Study

Ngati Rakaipaaka is a Maori iwi (tribe) located predominantly on the east coast (Hawke's Bay) of the North Island of Aotearoa (New Zealand). This iwi is rich in history, with the living descendents of Rakaipaaka now extending into the 15th generation. The iwi have extremely well documented whakapapa (family) and a comprehensive family tree. It has been observed that within the iwi there are high incidences of serious diseases. This has been one of the driving forces behind the development of the Rakaipaaka Health and Ancestry Study (RHAS). The RHAS has been designed to utilise whakapapa to understand patterns of genetic and disease variation within the Rakaipaaka community (Appendix One – Community Factsheet). The RHAS is a large-scale, long-term epidemiological project revolving around the genetic variation observed within the iwi. There are several main aims to the RHAS:

- i) utilise genetics to understand the serious diseases that affect the community.
- ii) understand the heritability of these diseases through the use of whakapapa.
- iii) identify genetic and environmental factors that influence these diseases.

The ultimate goal of the RHAS is to generate an understanding of these diseases at the genetic level, and to use this to develop treatment strategies to benefit Rakaipaaka. All individuals enrolled in the RHAS are informed consenting participants. Each has provided detailed health information in the form of a questionnaire, undergone physical assessment, and given blood samples for analysis. Detailed whakapapa information has also been given where available. All this will be used to successfully track patterns of disease with particular genetic markers. Already genetic research is providing a better understanding of disease pathways within the Maori community, with the goal of offering enhanced treatment strategies. Recent research conducted in Rakaipaaka Maori

explored the association of DNA variants within several genes, including CYP2A6 (Lea *et al.*, 2005; Lea *et al.*, 2008), and ADH2/ALDH2 (Chambers *et al.*, 2002). These studies investigated variation within drug metabolising genes between Maori and Caucasian genomic backgrounds. It was found that Maori have a very different genomic background when compared to Caucasians (Hall *et al.*, 2007). One region of common Maori genetic variation that contains several single nucleotide polymorphisms (SNPs) is found within cytochrome P450 (specifically CYP2A6). This variation has been associated with slower nicotine metabolism, and appears to be more prevalent in Maori (Lea *et al.*, 2005). Further research identified important differences in drug metabolism between Maori and Caucasians (Lea *et al.*, 2008). These differences could possibly influence the manner by which Maori respond to specific treatments, i.e. slower nicotine metabolism could require an alteration of the cessation therapy. Identifying such differences between groups with varying ancestry is the first step to offering personalised treatment based on an individual's genetic background.

Studies suggest that this Maori iwi has a distinguishable genomic background when compared to Caucasians (Hall *et al.*, 2007; Lea *et al.*, 2007) and preliminary work on mtDNA variation in New Zealand Maori has shown different frequencies compared to other groups worldwide (Murray-McIntosh *et al.*, 1998; Whyte *et al.*, 2005). This apparent uniqueness of genomic patterns in Maori is probably the result of genetic bottlenecks and founder effects that occurred during the Polynesian migrations (Lea *et al.*, 2007). Because of these unusual population genetic forces, it cannot be assumed that the genetic factors for metabolic traits in other worldwide populations are the same as in groups with Maori heritage. Identifying the specific genetic factors that influence risk of metabolic disorder in Rakaipaaka community, and understanding how these

interact with environmental (eg. dietary) factors, will contribute to better treatment strategies, reducing this problem.

Mitochondria are the energy producing factories of cells. It has long been thought that disruption to mitochondrial systems is linked with metabolic disorders. The mitochondrion has its own distinct genome. Mitochondrial DNA sequence variants have been linked to diabetes risk in several epidemiological studies (Poulton *et al.*, 2002; Tanaka *et al.*, 2007). As part of the RHAS (Hudson *et al.*, 2007) it has been observed that the Maori community of Nuhaka (Ngati Rakaipaaka) have a high incidence of certain metabolic diseases such as diabetes and obesity. These disorders are thought to be influenced by a combination of ancestry specific genetic predisposition and environmental exposures (envirogenomics).

It is important to begin to understand the ancestry of a population when we come to the creation of treatments, whether pharmaceutical or interventional. It is this ancestry that has shaped population genomics, meaning that individualised population treatment should be explored, what works for one group does not mean that it will work for another. As such, ancestry is starting to be considered in the development of personalised medicine (pharmacogenomics). Pharmacogenomics attempts to explain the influence of genetic variation on the response and efficacy of a drug treatment in patients. The aim is to develop a means of optimising the manufacture and delivery of drugs based on the patient's genotype. So this is replacing the 'one size fits all' approach with a more tailored treatment that should be effective with minimal risk of adverse effects.

It is hypothesised that Maori from Ngati Rakaipaaka have distinctive DNA sequence patterns (or signatures) in the mitochondrial genome, which have been shaped by unique population history, and that specific mtDNA signatures modify the risk of developing metabolic disorders such as diabetes in this population. This study aimed to screen the mitochondrial genome using the Mitochip™ in selected RHAS participants to highlight genetic variants that might be specific to this community and subsequently correlate these particular variants with metabolic traits such as diabetes, cardiovascular disease, and obesity. These metabolic conditions are a significant health issue in New Zealand, with the rates being especially high in the Rakaipaaka community. A successful outcome of this research may allow for more personalised intervention strategies (e.g. dietary) to be designed, with the ultimate goal of reducing metabolic disease and increasing health and wellbeing within the Rakaipaaka iwi, and perhaps Maori and Polynesians more generally.

1.7 Specific Research Objectives

The objectives of this thesis are as follows:

1. Design and optimisation of an RFLP assay and HVRI PCR to aid with subject selection for Mitochip array analysis.
2. Obtain entire mtDNA sequence information from a subsample of Rakaipaaka individuals with Maori ancestry (n=20) using the Affymetrix Mitochip.
3. Align and compare this sequence information with independent worldwide sequences to identify group differences, i.e. a mtDNA signature that is specific to Rakaipaaka Maori.
4. Validate the signature in a larger community sample of the RHAS Maori population (n=400).
5. Correlate the validated mtDNA signature with metabolic traits (eg. BMI and diabetes).

Chapter 2 *Materials and Methods*

2.1 Study Population Characteristics

DNA was collected either from blood or buccal samples for ~400 adults enrolled in the RHAS. Extensive clinical data was also obtained, including BMI, lipid profiles (LDL, HDL, etc.), cardiovascular disease makers, and disease history. Autosomal ancestry has been used to investigate population structure within RHAS Maori (Hall 2009) and is shown in Figure 2.1. This autosomal ancestry score will be used to identify those individuals most likely to contain the mitochondrial Polynesian motif, hence Maori mtDNA. This eliminates any admixture of European mtDNA present and provides a Maori specific mtDNA sample group with which to screen for sequence variation.

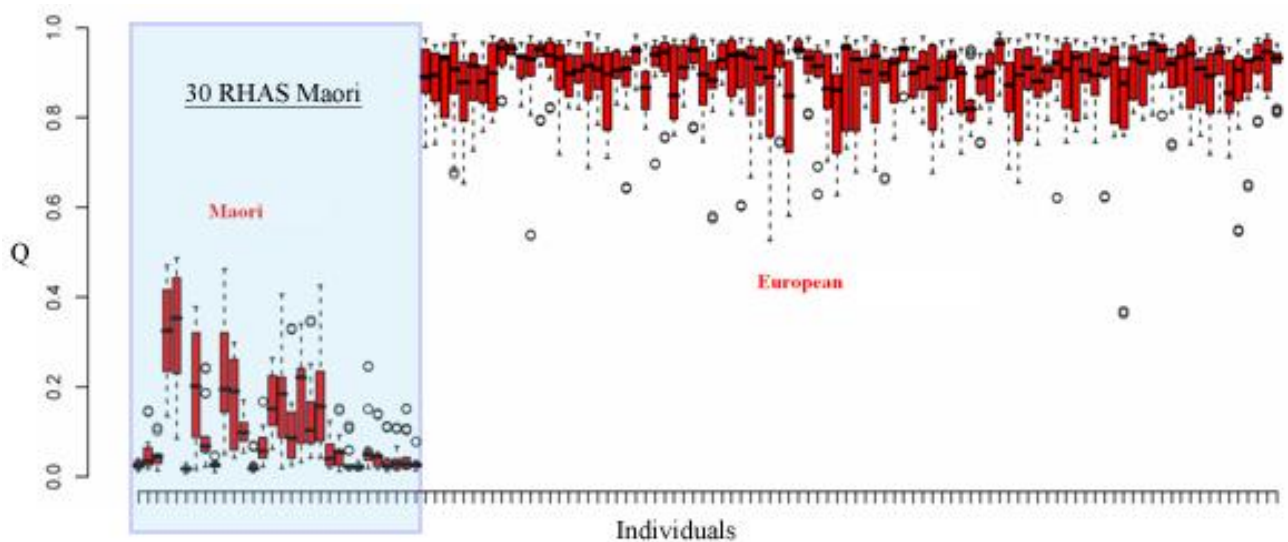


Figure 2.1: The figure shows genomic clustering of individuals with Maori and European ancestry. The X axis shows individuals for whom DNA has been genotyped using 300K SNP chips. Y axis shows estimates for % of European ancestry (1.0=100%). The 30 RHAS Maori individuals highlighted have been selected for this study because little, if any, of their genome is of European origin. Figure reproduced from Hall (2009).

2.2 Overview of Experimental Design

This study was designed as pioneer research, an exploration of complete mitochondrial genome sequences in New Zealand Maori community. This next section aims to give a general summary of the methodology and experimental design of the research conducted in this thesis. A workflow diagram has been included to give an overview of experimental structure (Figure 2.2).

The research was conducted with DNA samples kindly supplied to ESR as part of the Rakaipaaka Health and Ancestry Study (RHAS). As far as is known this is the first research of its kind in New Zealand Maori. Twenty Mitochip arrays were purchased due to budget limitations and the expense of the technology. This then became a proof of principle study, with the goals of testing the Mitochip technology and gaining an indication from a smaller sample of mtDNA sequences if there was anything worth following up. This meant that much care had to be shown in the selection of individual mtDNA sequences for Mitochip analysis; the final 20 had to be a good representation of the RHAS mtDNA pool. Using autosomal ancestry data supplied by David Hall (Figure 2.1), 30 RHAS individuals were identified as having very high Maori ancestry.

In testing the mtDNA for Polynesian ancestry two approaches have been designed and optimised (Figure 2.2). The first is the more traditional PCR and sequencing of the HVRI region. This protocol was used to generate the first 30 HVRI sequences. The second analysis was the designing of an RFLP assay, which was very useful as a quick screen for a specific marker of Polynesian mtDNA. This was a novel RFLP assay, the restriction site being first documented in this research.

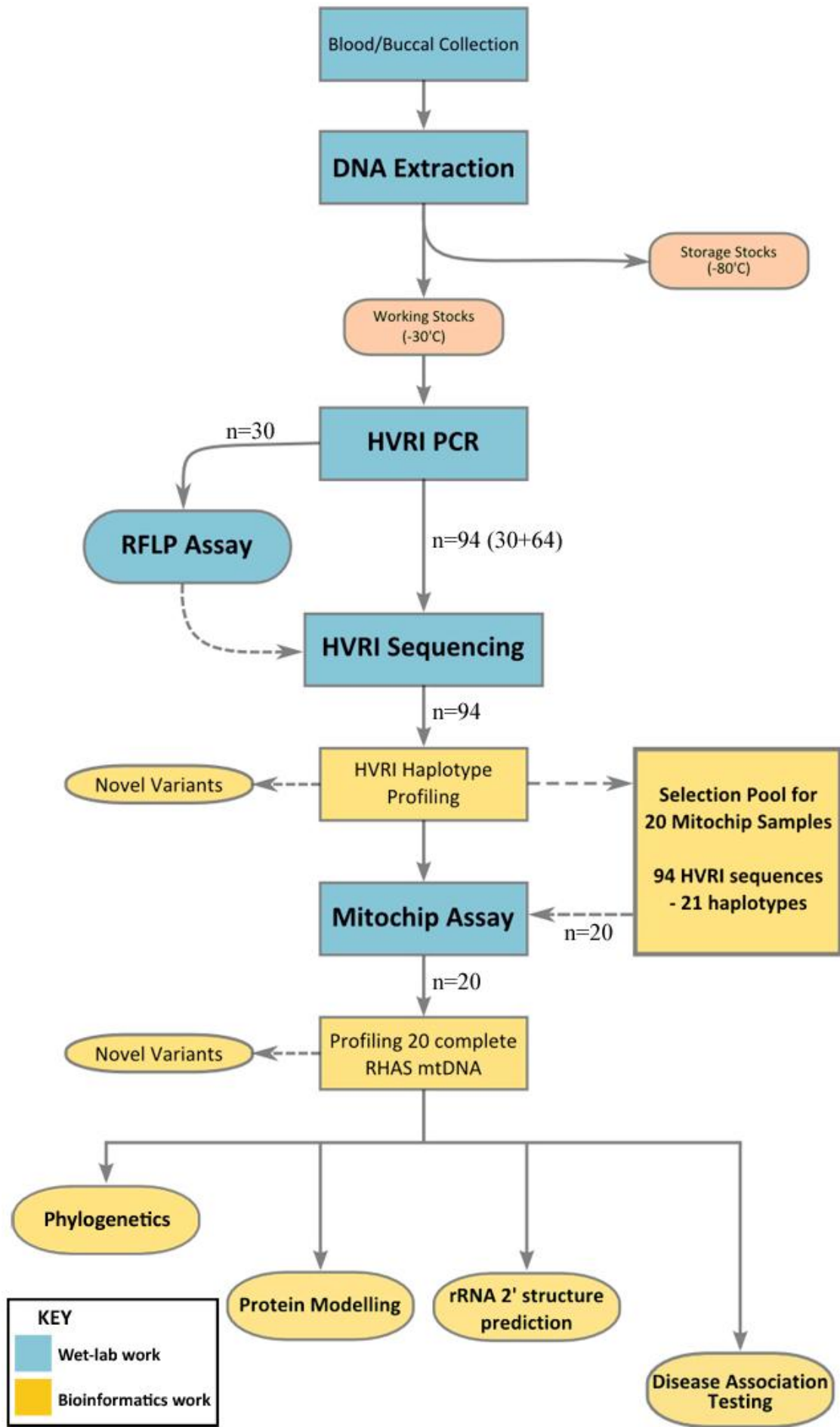


Figure 2.2: Workflow schematic detailing experimental layout.

After confirmation that the PCR protocol was working for the HVRI region, the RFLP assay was used as a quick survey to determine if the majority of RHAS Maori PCR products were of Polynesian ancestry. The RFLP assay interrogated a Polynesian mtDNA marker, A16247G. The presence of this marker indicated mtDNA as being Polynesian in ancestry. This eliminated the chance of wasting resources on sequencing unwanted products. This analysis proved useful, with some samples showing digestion and others not.

When all 30 HVRI PCR products were sequenced and analysed, it was observed that all sequences were in fact Polynesian, regardless of the RFLP results. To make completely sure the best representative group of mtDNAs was selected a further 64 HVRI sequences were added, making the final cohort 94. This created a larger pool of candidate sequences for final Mitochip resequencing. Based on the frequency of reported haplotypes and specific variants of interest the 20 RHAS Maori products were chosen.

Once the complete mitochondrial sequence data had been acquired, it was aligned against the Cambridge Reference Sequence (CRS). This is the human mitochondrial reference sequence which all variants are identified against. This data was then collated in Microsoft Excel and MEGA 4.1, providing the major dataset of this project. Obtaining this dataset was one of the primary objectives, which lead to further analysis experiments. These were; phylogenetic testing, protein modelling, ribosomal RNA secondary structure prediction, and genotyping. Genotyping several signature SNPs in the larger RHAS Maori cohort group (n=400) allowed for disease association statistics to be calculated.

2.3 Source of DNA and Ethics Approval

Being a genetic study involving Maori participants, the RHAS has taken several years to develop in terms of ethical and cultural clearance. The study is directed and hosted by Te Iwi o Rakaipaaka (TIORI) in Nuhaka and has received full ethical approval from the Multi-regional ethics committee (MEC/05/12/174). All individuals have signed a consent form acknowledging they understand the genetic nature of this health research and wish to participate. All blood samples were sent to Aotea pathology, where plasma was removed the plasma for testing. The remaining blood samples and buccal swabs were sent to ESR Ltd. Te Iwi o Rakaipaaka are consulted and kept updated in respect to major findings, manuscripts and presentations.

2.3.1 Blood Extraction

Genomic DNA was extracted from blood using QIAGEN FlexiGene DNA extraction kit and a modified extraction procedure. Briefly, 4ml of blood was collected in an EDTA vacutainer. The sample was sent to Aotea pathology who removed the plasma for testing and sent the remaining sample to ESR where it was stored at -20 before extraction. The 4ml tube was topped up with sterile phosphate buffered saline. Blood suspension was decanted into a 50ml falcon containing 20ml FG1 Buffer, and then mixed by inversion. Tubes were centrifuged at 2500 x g in a swing out rotor for 10 minutes. The supernatant was discarded and the tubes inverted on absorbent paper for 2 minutes, care was taken to ensure the pellet remained in the tube. 4 ml of FG2 Protease was added and the tubes immediately vortexed until the pellet was completely homogenised. The tubes were then inverted 3 times and placed in a heat block at 65°C for 10 minutes. 4 ml of isopropanol (100%) was added each tube and mixed by inversion, causing the DNA to precipitate and become visible. At this stage the

clumped DNA was removed using a pipette tip and transferred to clean labelled 1.5 ml microcentrifuge tubes. 1 ml of 70% ethanol was added and tubes were vortexed. Tubes were then centrifuged at 10000 rpm for 2 minutes then vortexed. Tubes underwent a final centrifugation for 2 minutes, the supernatant was discarded and tubes inverted and pellet allowed to air dry. The DNA pellet was resuspended in 1 ml dH₂O. Where necessary the tube was incubated 65°C to aid in resuspension. DNA yield was quantitated by pico-green (Section 2.4.5) and samples stored at -20°C.

2.3.2 Buccal Swabs Extraction

DNA extraction from these was performed using BuccalAMP™ DNA Extraction Kit (Epicentre®). Briefly, the subject's mouth was rinsed thoroughly with water and tissue was collected by rolling the Catch-All sample collection swab firmly on the inside of both cheeks. Swabs were sent to ESR and stored at -20 until extraction. The swabs were placed into tubes containing QuickExtract DNA extraction solution and then rotated a minimum of 5 times. Swabs were removed and screw caps applied to the tubes, which were then vortexed for 10 seconds. They were then incubated at 65°C for 1 minute before a further vortex of 15 seconds before incubation at 98°C for 2 minutes. Tubes were then vortexed for 15 seconds then incubated at 98°C for 2 minutes. Tubes were then vortexed for 15 seconds, centrifuged for 5min at 14,000rpm and the extract removed to a fresh tube with care not to carry over any pelleted material. DNA yield was quantitated by pico-green (Section 2.4.5) and samples stored at -20°C.

2.4 Mitochondrial Control Region (HVRI) Analysis

The HVRI area of the mitochondrial genome was amplified for each sample (94 in total) using the Polymerase Chain Reaction (PCR). The expected product size of the HVRI

target region was ~543bp. The presence of the 16189 T→C transition in the majority of samples made sequencing the complete product using one primer impossible. Therefore products were sequenced in both the forward and reverse directions, generating a consensus HVRI sequence for each. Additionally any nucleotide variants from the Cambridge Reference Sequence (CRS) were confirmed by generating independent PCR amplicons of PCR product and resequencing.

2.4.1 Primers

The HVRI target region was amplified by polymerase chain reaction (PCR), using primers that complemented the flanking regions. The primer pair used for this reaction were L15975 (5'-CTC CAC CAT TAG CAC CCA AAG C-3') and H16517 (5'-CCT GAA GTA GGA ACC AGA TG-3'). See Appendix Two for primer information.

2.4.2 HVRI PCR Protocol

The standard 30 µl double stranded PCR reaction contained:

16.1 µl	sterile double distilled water (sddH ₂ O)
3.0 µl	10X PCR Buffer with MgCl ₂
6.0 µl	GC-rich Buffer (10x)
1.2 µl	2.5 mM dNTPs (deoxynucleotide triphosphate mix with dATP, dTTP, dCTP, dGTP at 2.5 mM each)
1.2 µl	Forward (F) Primer (10 µM, Invitrogen)
1.2 µl	Reverse (R) Primer (10 µM, Invitrogen)
0.3 µl	FastStart <i>Taq</i> Polymerase
1.0 µl	Template DNA (~20 ng/µl)

Due to the small amount of *Taq* required for each reaction a master mix was prepared beforehand. Each master mix was made by multiplying the PCR reagents above by the number of samples to be amplified (*Taq* was added to the master mix just before use).

The master mix was made up in a 1.5 ml microcentrifuge tube and kept on ice until required. 1 µl Template DNA was pipette into sterile strip tubes, and then 29 µl of master mix was added to each tube. Negative controls contained 1 µl ddH₂O instead of template DNA, and were also PCR amplified to check for contamination. The target region was amplified in an Eppendorf Mastercycler ep. Thermal Cycler under the following profile: initial denaturation at 95°C for 5 minutes, 35 cycles of denaturation at 94°C for 30 seconds, primer annealing at 60°C for 30 seconds, extension at 72°C for 60 seconds, a final extension at 72°C for 60 seconds, then held at 4°C.

2.4.3 Agarose Gel Electrophoresis

Successful amplification of PCR products was assessed using agarose gel electrophoresis. A 5 µl aliquot of loading dye was added to each PCR product containing strip tube, and then a 5 µl aliquot of each PCR product was loaded onto a 1% agarose, 1 x TBE gel containing SYBR safe (1µl of 1/10 dilution [Invitrogen]) for DNA visualisation (Appendix Three). 5 µl of 1 kbp ladder (Invitrogen) to gauge the size of PCR products was used as reference. The gel was run in Tris Borate Buffer (TBE) at 110V for 60 minutes. Gel photographs were taken under UV illumination.

2.4.4 Purification of PCR Products

Manufacturer's (Qiagen®) instructions were followed using the QIAquick PCR Purification kit. Briefly, 5 volumes of Buffer PB (kit) were added for every 1 volume of PCR sample and mixed. The DNA was bound by applying each sample to a QIAquick column, which was then centrifuged for 60 seconds at ~13000 rpm. Flow-through was then discarded and 750 µl of wash buffer (buffer PE) added to each column. Columns were then centrifuged for 60 seconds. Flow-through was again discarded and columns underwent an additional centrifugation for 60 seconds. In the

final step columns were placed into labelled 1.5 ml plastic microcentrifuge tubes. To elute DNA product 30 µl of dH₂O was applied to each column which was then centrifuged for 60 seconds. The columns were then discarded and the 30 µl DNA samples were stored at -20°C.

2.4.5 DNA Quantitation: Pico Green assay

A Quant-iT PicoGreen dsDNA quantitation kit (Invitrogen) was used to quantify PCR products. Briefly, 96-well flat-bottomed (Nunc) plates were used. Standard solutions were prepared in duplicate in the first two rows of the plate wells using 1 µl DNA standard stock (Appendix Four) and 99 µl TE buffer to a total of 100 µl. The standard preparations used were 200 ng, 100 ng, 50 ng, 25 ng, 12.5 ng, 1 ng, 0.5 ng total DNA. The blank wells contained 1 µl dH₂O and 99 µl TE buffer. 1 µl of DNA samples were added to wells in duplicate, then 99 µl TE buffer was added to each well. Pico Green quantitation reagent was diluted 1:200 in dH₂O, 100 µl was added to each well used, the plate was then covered and mixed gently, before reading with a FLUOstar OPTIMA microplate reader after 1-5 minutes. The excitation wavelength used was 484 nm and the emission was 538 nm. A curve was created from the standard solutions. The standard curve and the sample values corrected for the blank were used to calculate the amount of DNA present in each sample.

2.4.6 Sequencing of HVRI Products

Sequencing of the HVRI product was difficult due to the homopolymeric cytosine tract present in all but one of the 94 RHAS samples. Therefore both the forward and reverse strands were sequenced to generate a consensus HVRI sequence. Sequencing was conducted at the Allan Wilson Centre of Molecular Biology. Sequencing reactions were prepared in accordance to specifications detailed on the Allan Wilson Centre

(AWC) website (<http://awcmee.massey.ac.nz/genome-primer.htm>). Sequencing reactions were set up with 13 ng template PCR product, 3.2 pmol primer (forward or reverse) and then made up to 15 μ L total volume with PCR grade dH₂O. DNA sequencing was primed with the following sequencing primers: forward strand sequence 5'-CTCCACCATTAGCACCCAAAGC-3' and reverse strand sequence 5'-CCTGAAGTAGGAACCAGATG-3'. The BigDyeTM Terminator Version 3.1 Ready Reaction Cycle Sequencing kit was used and samples were run on an ABI3730 DNA Analyzer. Sequences were made available from AWC via web drop off using an AWC user profile. Sequence data was interpreted using FinchTV software v1.4.0 (Geospiza, Inc.; Seattle, WA, USA).

2.5 Restriction Fragment Length Polymorphism Analysis

The 50 μ L restriction digest reaction contained:

38 μ L	dH ₂ O
4.0 μ L	10x Buffer (NEB)
0.5 μ L	BSA
0.5 μ L	<i>Pst</i> I
7.0 μ L	DNA (HVRI PCR product at 15 ng/ μ L)

A master mix was made by multiplying the reagents by the number of samples. 7 μ L of each DNA sample was pipette into individual 0.2 ml strip tubes, and then 43 μ L of master mix was added to each tube. The samples were incubated in an Eppendorf Mastercycler ep. Thermal Cycler under the following parameters: 2 hours at 37°C, 80°C inactivation for 20 minutes, then held at 4°C. Restriction products were run on a 2% agarose, 1 X TBE gel (containing SYBR safe) at 100V for 2 hours, then viewed under UV illumination.

2.6 Mitochip Analysis

The Mitochip Resequencing Array was conducted offsite by the Centre for Genomics and Proteomics at the University of Auckland. The Mitochip Resequencing protocol (Affymetrix, Santa Clara, CA) laid out by Affymetrix was followed (Affymetrix GeneChip CustomSeq Resequencing Array Protocol version 2.1.), and the chips were run on Affymetrix GeneChip equipment (GeneChip Hybridization Oven, GeneChip Fluidics Station, and GeneChip Scanner 3000). Briefly, the Mitochip has array tiles for both strands of the complete human mtDNA (~16,550 bp), as well as an additional 12,935 bp arrayed in duplicate. The assay required 100ng of DNA from 20 selected RHAS individuals as template for mtDNA specific long range PCR. Two sets of long range primers were used to amplify the mtDNA: Mito1-2 forward primer 5'-ACATAGCACATTACAGTCAAATCCCTTCTCGTCCC-3' and Mito1-2 reverse primer 5'-ATTGCTAGGGTGGCGCTTCCAATTAGGTGC-3', Mito3 forward primer 5'-TCATTTTTTATTGCCACAACCTCCTCGGACTC-3' and Mito3 reverse primer 5'-CGTGATGTCTTATTTAAGGGGAACGTGTGGGCTAT-3' (Appendix Five). The PCR products were then fragmented, labelled with biotin, and hybridised to the Mitochip. The chip with bound mtDNA was washed with streptavidin-phycoerythrin, which binds to the biotin. A signal amplification step was conducted by washing the chip with anti-Streptavidin antibody, followed with biotinylated goat IgG antibody, and a final wash with streptavidin. This procedure ensures an ample signal is present when the chips are scanned with a confocal laser. The raw data from the scanned Mitochips was uploaded and made available for collection via the University of Auckland web-dropoff system. These files were then analysed using the GeneChip Sequence Analysis Software 4.1 (GSEQ 4.1).

2.7 Software Packages

2.7.1 DNA Sequence Alignment and Phylogenetic Tree Creation

MEGA4 was used to align both partial HVRI sequences and complete mtDNA sequences. All sequences were aligned against the human mitochondrial reference sequence (CRS) to preserve historic positional information. The default ClustalW alignment was implemented, with final alignment checked manually for error or discrepancies. A phylogenetic reconstruction was created in MEGA4 using Tamura-Nei neighbour-joining method (Tamura 2007). All 20 complete RHAS mtDNA sequences were added to a data set containing complete sequences from various B haplogroups (Trejaut *et al.*, 2005).

2.7.2 Mitochip Data Analysis in GSEQ 4.1

The Mitochip data received from the University of Auckland was in .CEL format. These files were analysed using GeneChip® Sequence Analysis Software 4.1 (Affymetrix). The CEL files store the probe intensity data for each nucleotide position tiled on the Mitochip array. There are 8 probes for each position (4 each strand) which, depending on probe binding, will give various fluorescent levels. These levels are interpreted by the GSEQ 4.1 software by running a resequencing algorithm. This algorithm consists of specific parameters which when adjusted can alter the overall accuracy and base calling ability of the array. The final algorithm determined to give the highest call rate in this study is as follows:

Probe Array Type - Mitochip_2

Number of Bases per Array = 37756

Number of Samples = 20

Parameters Used By Algorithm:

No Signal Threshold (probe signal/noise ratio) = 1

Weak Signal Fold Threshold (mean/probe ratio) = 10

Large SNR Threshold (probe signal/noise ratio) = 20

Perfect Call Quality Threshold (quality score) = 3

Model Type (0=Heterozygote, 1=Homozygote) = 0

Min Fraction of Calls of Samples (0-1) = 0.5

Trace Threshold = 1

Sequence Profile Threshold = -0.175

2.7.3 rRNA secondary structure modelling

Structures for the wild type and mutated human mitochondrial 12S rRNAs were generated using the RNAfold webserver software (<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>) from the Vienna RNA package (Hofacker 2003). RNAfold predicts RNA secondary structure based on minimum energy requirements and pair probabilities. Parameters used were adapted from models found in Ballana *et al.* (2006). Parameters used are as follows: minimum free energy (MFE) only = selected, avoid isolated base pairs = selected, unpaired bases can participate in at most one dangling end = selected, RNA parameters (Turner Model) = selected, rescale energy to given temperature (C) = 37. The same parameters were used to predict secondary structure of 16SrRNA models. Amino acid conservation at positions of interest was examined by aligning 12SrRNA sequences of five different species (idea adapted from Ballana *et al.* (2006)).

2.7.4 Protein Modelling with Sirius

Models were found for those proteins which contained amino acid changes: (model numbers listed, all models can be found at <http://swissmodel.expasy.org/repository/>)

Human Cytochrome b: P00156

Human ATP Synthase subunit 6: P18859

Human Cytochrome Oxidase subunit I: P00395

Human Cytochrome Oxidase subunit III: P00414

Sirius v1.2 was used to view and render the protein models. It also allowed for the substitution of specific amino acids. Using this feature wild-type amino acids could be substituted for their respective variant, which allowed the examination of what effects amino acid change had on the possible structure and interactions within the proteins. (Sirius can be found at <http://sirius.sdsc.edu/index.php>)

2.7.5 mtPhyl: software for mtDNA analysis and phylogeny reconstruction

Haplotype identification and amino acid conservation indexes were generated using the human mitochondrial DNA software suite mtPhyl (Eltsov 2009). Sequences can be imported into mtPhyl in either text or fasta format. Once the 20 complete RHAS Maori mtDNA's were loaded into the program a variation report with conservation analysis was conducted. This tests all sequences against the Cambridge Reference Sequence (CRS) and generates a table that can be exported and edited in Microsoft Excel.

2.8 Genotyping and Statistical Analysis

Samples were sent for genotyping to a service provider, Australian Genome Research Facility Ltd. (AGRF, www.agrf.org.au). Three mitochondrial SNPs were included in a Genotyping assay alongside selected RHAS nuclear SNPs (in collaboration with ESR). These were positions 4769, 16189, and 16519. AGRF's Sequenom custom SNP genotyping service is run on the Sequenom MassARRAY system and utilises a homogenous MassExtend (hME – single base extension) reaction termed iPLEX GOLD, for analysis of up to 36 SNP's in a single multiplex. SNP assays are designed by the AGRF based on provided SNP sequences (RS Numbers). Oligonucleotides are then obtained by the AGRF and used to process samples in multiplex format. Following the MassExtend reaction, oligonucleotides are printed onto specialised Spectro CHiPs and analysed by the Sequenom mass spectrometer. This platform has the ability to replex SNP panels so SNPs can be dropped or added at a later date.

Results from genotyping were entered into an SPSS database which contained all recorded information for participants enrolled in the RHAS project. Statistical analysis was therefore performed using the SPSS statistical package (www.spss.com). Possible correlations between metabolic traits and mitochondrial SNP's were tested using Spearman's two-tailed Bivariate test for correlation. Traits that showed a $P=0.05$ (95%) level of significance were then investigated using appropriate statistical tests, either an Independent-Samples T test or Chi squared test. As data for the diabetes trait was nominal in nature (either diabetic or non-diabetic) risk association was explored using the Chi-squared test.

Chapter 3 Results

3.1 DNA Extraction

Genomic DNA was extracted from ~400 Maori individuals who had given consent to be part of the RHAS research project. These individuals donated DNA in one of two forms, either as a buccal swab or as a small blood sample (4 ml). Total genomic DNA (includes mtDNA) was extracted using the respective protocol; blood extraction (QIAGEN®; Section 2.3.1), or buccal swab extraction (Epicentre®; Section 2.3.2). Rakaipaaka Maori DNA was then PCR amplified, purified (Chapter 2, Section 2.4), and quantitated to approx. 15 ng/μl (see Appendix Six for an example Pico-Green output).

OBJECTIVE 1: Design and optimisation of an RFLP assay and HVRI PCR to aid with subject selection for Mitochip array analysis.

3.2 Mitochondrial DNA HVRI Target Amplification

Due to budget constraints there was only enough funding for 20 Mitochip assays. As such it was determined that preliminary work would involve screening HVRI sequences of an initial RHAS Maori group (n=30) identified to be of high Maori autosomal ancestry (Methods, Figure 2.1). The goal of this initial screen was to identify mtDNA sequences that were of Polynesian, and thus Maori, ancestry. The initial 30 RHAS Maori selected revealed high autosomal ancestry which was a good predictor of mtDNA ancestry. As such, an extended group of HVRI sequences was generated (n=94) consisting of another 64 RHAS Maori based on ancestry score (Figure 3.1).

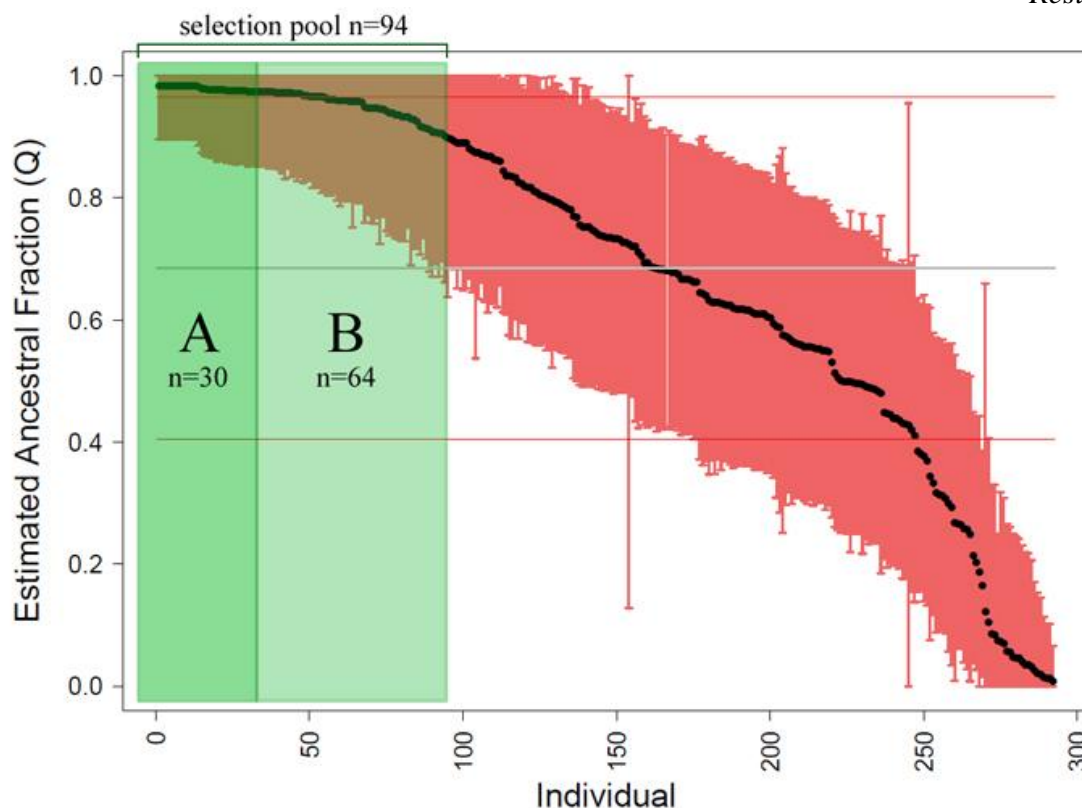


Figure 3.1: shows the 2 HVRI subgroups and the overall Mitochip sample selection pool (n=94). The Y axis has ancestry fraction (Q) based on autosomal data (Hall 2009), while the X axis shows RHAS Maori individuals.

RHAS Maori were selected for HVRI analysis based on autosomal ancestry (Hall 2009). Figure 3.1 details the selection pool; subgroup A consisted of the initial 30 RHAS Maori, and subgroup B (n=64) extended this to a larger pool (n=94) of HVRI sequences to allow for Mitochip selection. Having this pool of HVRI sequences meant that the 20 sequences chosen were a good representation of RHAS Maori mtDNA. The mitochondrial HVRI control region was successfully PCR amplified from a total of 94 RHAS Maori DNA solutions, using the primers and protocol outlined in Methods Section 2.4. All PCR products were validated for size and quality using agarose gel electrophoresis (Figure 3.2). An agarose calibration curve was calculated to verify that products of the correct size were being amplified (Figure 3.3). The expected PCR product size was 543 bp based on primer design software (Primer Blast; <http://www.ncbi.nlm.nih.gov/>). A successful PCR amplification should appear as a

single band of this size when run through an agarose gel. Figure 3.2 shows an agarose gel electrophoresis of four HVRI PCR products (specimen ID's; mt3, mt6, mt12, and mt20). In each of the lanes loaded with PCR product in this figure there is a single bright band of approximately ~550 bp (when compared the 1 Kb Plus DNA Ladder; Invitrogen®). The lanes containing negative reaction controls show no bands, indicating there is no PCR product present.

Figure 3.2 also shows the reaction used to find the optimal primer annealing temperature for further HVR1 PCR reactions. A thermal cycle was designed to set the annealing temperature across the heat block in three sections as follows: 55°C, 60°C, and 65°C. As all three temperatures yielded acceptable products of the correct size, the central temperature (60°C) was selected as the final annealing temperature.

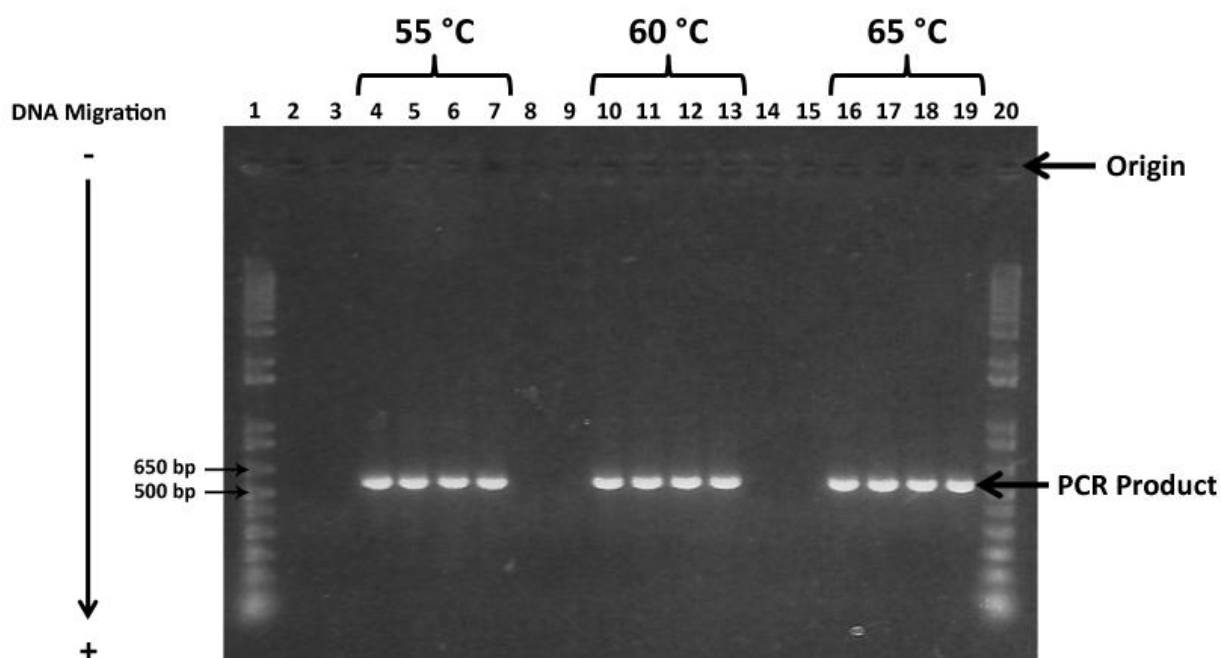


Figure 3.2: Gel showing electrophoresis of amplified mtDNA HVRI target region. PCR amplicons of the same four DNA specimens were used in replicate for each temperature gradient, as loaded from left to right; mt3, mt6, mt12, and mt20. Lanes 1 and 20 contain a 1 Kb Plus DNA Ladder (Invitrogen®; Appendix Three). Lanes 2, 3, 8, 9, 14, and 15 were negative reaction controls used to identify if contamination was present.

The size of the amplified products was inferred to be ~550 bp, as indicated by the DNA ladder fragments. To confirm the fragment size of the PCR product bands from Figure 3.2 an agarose gel calibration curve was calculated (Figure 3.3). This curve was created by plotting the \log_{10} of the ladder fragment sizes against the distance the fragment bands migrated. A trend line established a linear relationship allowing the calculation of the 'unknown' PCR product size. The PCR product band was measured to have migrated 45.5 mm from the origin. When this distance was input into the equation generated by the trend line, a product size of 550 bp was inferred (Figure 3.3). This correlates well with both the size estimated from the DNA ladder and also the size given by *in silico* analysis (543 bp).

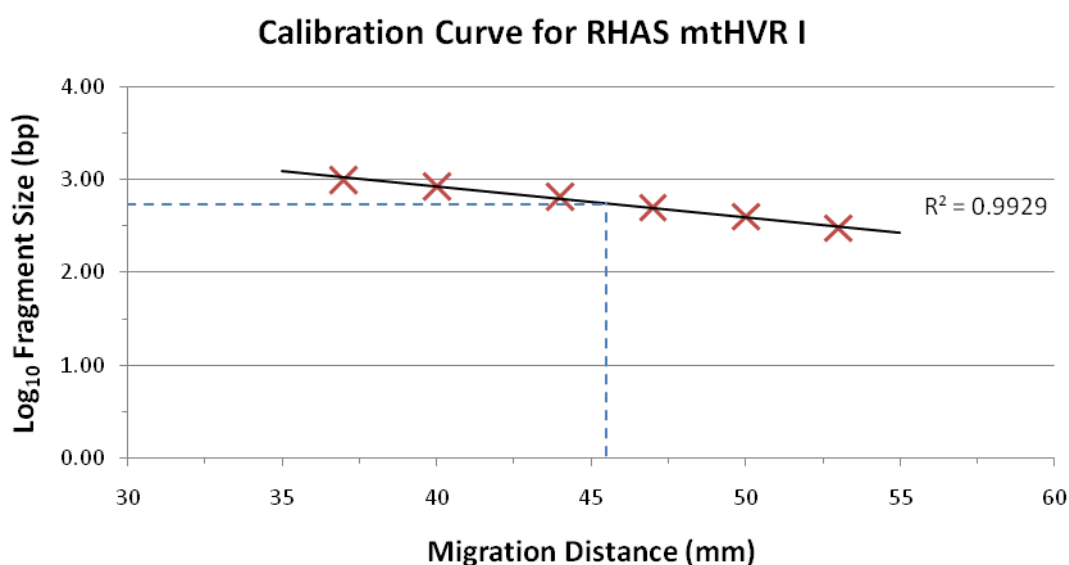


Figure 3.3: Calibration curve of Log (fragment size, bp) versus the migration distance (mm) of 1 Kb Plus DNA Ladder (Invitrogen®) fragments, as seen in Figure 3.2. The blue dashed line on the graph represents the migratory distance of PCR products in Figure 4.1. These products had migrated 45.5 mm through the gel giving an estimated product size of 550 bp. This size is an approximation of the PCR product size expected (543 bp).

3.3 Identification of *Pst* I Restriction in RHAS Maori mtDNA

An online restriction digest program (http://insilico.ehu.es/restriction/two_seq/) allowed the comparison of ~300 bp of Maori HVRI sequence (Whyte *et al*, 2005, GENBANK No. AY604073.1) and the same section from the Cambridge Reference Sequence (CRS: NC_012920). The online web server analysed both sequences and identified several restriction sites. The only enzyme found to digest specifically within Maori HVRI sequences was *Pst* I. The restriction site was found to be formed by the transition at 16247, A→G, (Figure 3.4).

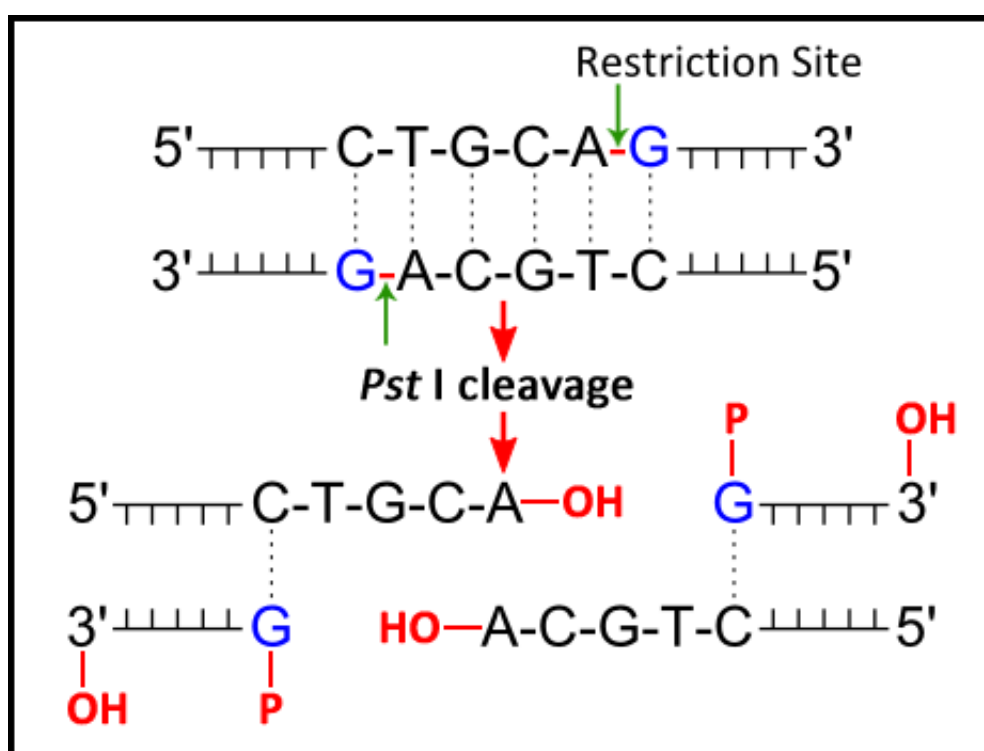


Figure 3.4: Restriction enzyme site created by the A16247G transition within the HVRI of haplogroup B4a1a1 individuals. When a G (blue) is present at position 16247 it creates a restriction site that is specific for *Pst* I. When treated with *Pst* I digestion occurs creating DNA fragments which can be identified via agarose gel electrophoresis.

For the restriction analysis 30 RHAS PCR products were investigated. Two Amis Taiwanese Aboriginal samples were HVRI PCR amplified and used as positive controls in the digest. These Amis products were used as they were known lack the 16247 A→G transition. The negative controls contained dH₂O instead of PCR product. Figure 3.5 shows the agarose gel electrophoresis of the restriction digest, a total of 32 individual PCR products were tested for the 16247 A→G transition (2 Amis and 30 RHAS). The gel was set up so each HVRI product had 2 lanes; the first lane lacked *Pst* I restriction enzyme, the second lane contained *Pst* I. If the A→G transition was present the product in the second lane would show differing band size to that in the first lane. The positive controls (lanes 4-8) do not exhibit any sign of restriction, confirming that Amis Taiwanese lack the transition that creates the restriction site. The negative controls (lanes 2, 3 and 72, 73) do not contain any bands, indicating that there was no contamination. Some of the restriction lanes contain faint bands around the 500-600 bp mark, this is possibly due to partial digestion of the original ~540 bp product. As this restriction analysis was used to quickly identify mitochondrial sequences which belong to haplogroup B4a1a1a it was not further optimised, all that was needed was proof of principle. Further optimisations could include; adjusting the amounts of PCR product, the concentration of the enzyme, and the length of restriction reaction.

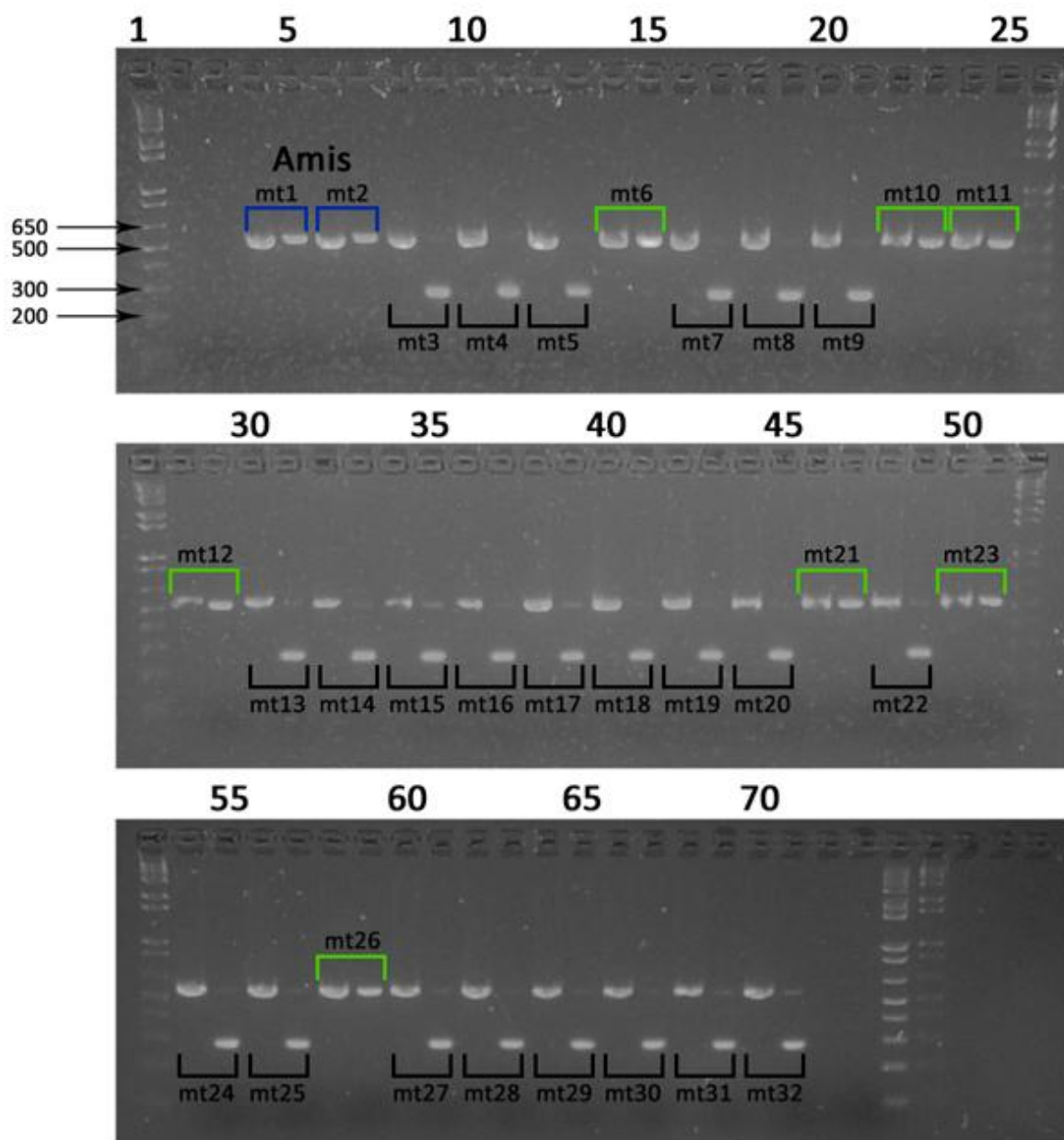


Figure 3.5: Gel Picture of restriction enzyme digest of 32 HVRI regions. Shown above, DNA ladder, negative controls, 2 Taiwanese HVRI samples, and 30 RHAS HVRI samples. Uncut product is ~540 bp, cut product is ~270 bp. Each sample has two lanes, the first lane contains PCR product only, while the second lane contains PCR product and *Pst* I restriction enzyme. Shown in red are the sample numbers of those PCR products that did not undergo restriction.

It can be seen from Figure 3.5 that 23 (77%) of the 30 RHAS HVRI PCR products underwent restriction indicating the presence of the Polynesian motif. This is identified by the smaller digested fragments migrating further in the second lane than the original product in the first lane. The original product was ~550 bp; the fragments formed by restriction are ~270 bp, exactly half of the original size. This explains why there is only one visible band present in the lanes that have undergone restriction; the fragments are the same size. There were 7 RHAS HVRI PCR products that did not digest. This indicates that these samples did not contain the 16247 A→G transition, or that there was another mutation within the restriction site that stops the enzyme from digesting. This became clear once the results from sequencing were returned (Section 3.5). It was assumed that the samples that did not digest were of different mitochondrial haplotype, most likely European introduced through genetic admixture.

3.4 Mitochondrial DNA Sequencing

As this is a discovery project, and the Mitochip technology is still very novel, only 20 chips were available for complete mtDNA sequence analysis. This meant that from the final pool of HVRI sequences a candidate group of 20 RHAS mtDNA had to be selected as a representative. It was an important requirement that these mtDNA sequences be of the haplotype B4a1a1a, as this indicates mtDNA of Maori origin. A sampling pool of 94 individual RHAS mitochondrial HVRI sequences was generated from PCR products. This pool was used, along with a selection criterion based on individual variants of particular interest, to select the 20 mtDNA candidate group. It was hoped that this candidate group would provide information regarding a Maori specific signature within the coding region of the mitochondrial genome. This signature could possibly prove valuable not only in disease association studies, but also in evolutionary genetics. At

the time of writing these are the only documented complete Maori mitochondrial genomes.

The exact sequences of all 94 PCR amplified HVRI samples were determined using an ABI3730 DNA Analyzer (Methods, Section 2.4.6). All HVRI sequences were viewed and edited in the FinchTV software (primer cleanup and base read corrections). They were then aligned with the Cambridge Reference Sequence (CRS) in MEGA4, which verified that all products were human mitochondrial HVRI (see Table 3.1). After primer cleanup and alignment the DNA sequences were ~470 bp long. This size is due to the presence of the C variant at 16189, which creates a homopolymeric C-tract. Automated sequencing is unable to proceed past this point. Therefore complete HVRI sequence was obtained by sequencing in both 5' and 3' directions. Table 3.1 shows four HVRI sequences (mt6, mt3, mt13, and mt20) have been aligned in MEGA and exported to indicate comparison to the CRS. The incredibly high degree of similarity with the CRS indicates that the correct PCR product has been amplified, and is homo sapien in nature. Over the 450 bp sequence shown in Table 3.1, there are 8 variants (highlighted in yellow). An insert of A at position 8 in the table is represented as dashes in the other sequences.

[illegible]

Initially sequencing of the RHAS HVRI samples was conducted using the 5' L15975 primer (Forward). Figure 3.6 shows example electropherograms produced by the sequencing protocol employed in this experiment. These revealed the nucleotide sequence for each sample. The first sequencing reaction used to test the primers was conducted on several Taiwanese Aboriginal HVRI samples. Sequence returned from the initial forward primer reactions was of exceptional quality (Figure 3.6 A). The sequence shown in this figure is the HVRI of Amis Taiwanese specimen AM011. This electropherogram shows little to no background 'noise', and very clear consistent nucleotide peaks. Coverage of the HVRI is also shown to be good with sequence covering the majority of this region; once edited HVRI sequences were 498 bp in length. These results confirmed that the forward primer was in fact selecting for the correct region of the mtDNA, and was also giving quality sequence data.



Figure 3.6 A: Sequence electropherograms for two mitochondrial HVRI products. Sequence electropherogram showing forward primer analysis of a Taiwanese Amis H16182 and 16183) that are acquired transversions in RHAS mtDNA.

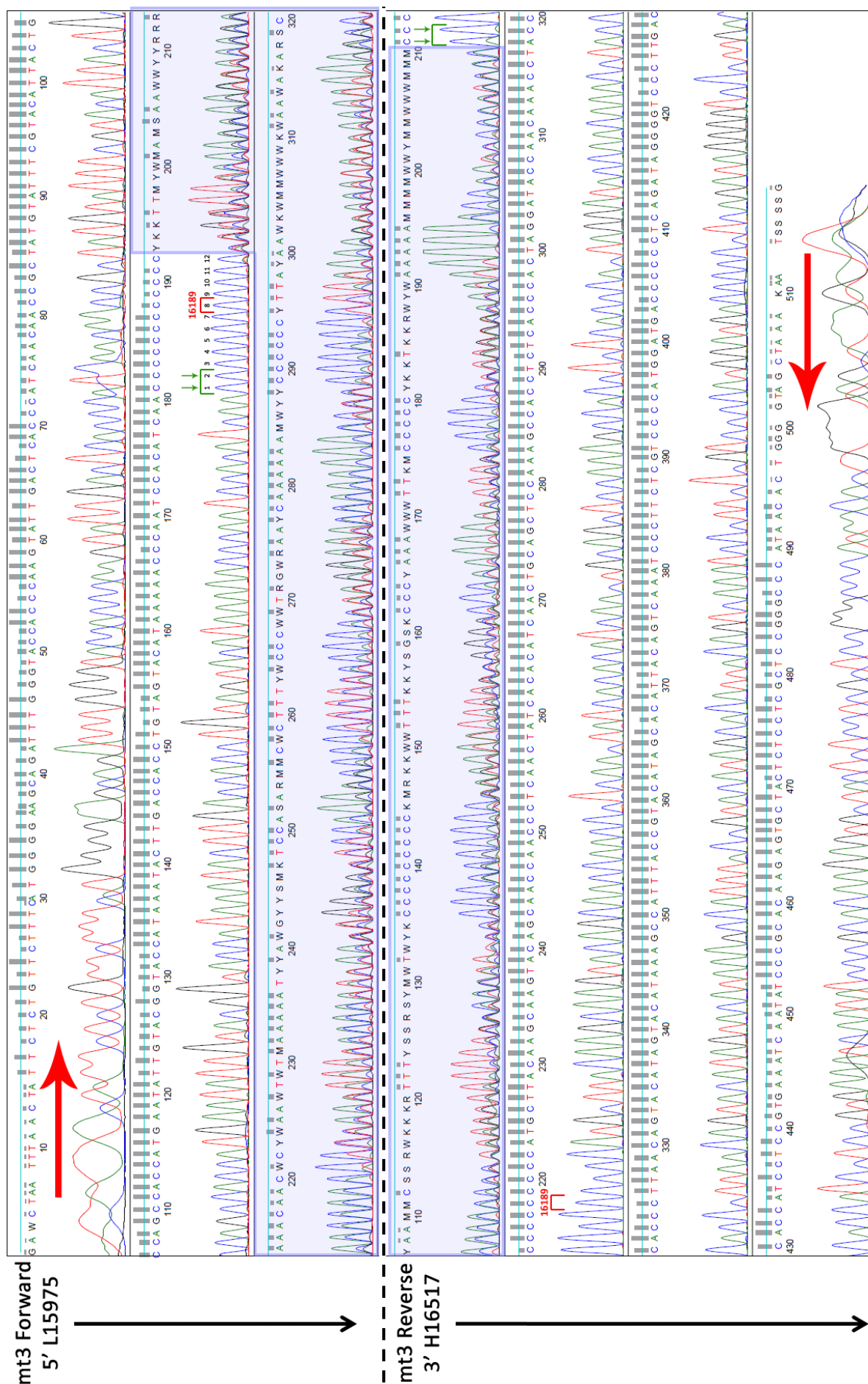


Figure 3.6 B: Sequence electropherograms for two mitochondrial HVRI products.

Upon confirmation of the sequencing protocol, 30 RHAS HVRI specimens were sent to the AWC for analysis. When the data was returned the HVRI sequences were viewed in FinchTV, a sample RHAS HVRI sequence is shown in Figure 3.6 B. All 30 RHAS sequences were of good quality for the first 190 bp, after which they became ‘noisy’ and could not be reliably interpreted. Several attempts were made to sequence through this area. First by using the reverse, 3’ H16517 primer (Figure 3.6 B shows the reverse complement of this data for mt3). The reverse primer had the same problem, giving readable sequence to a point. A modified sequencing protocol (dGTP BigDye™ Terminator Version 3.0) was then trialled to make sequencing of GC rich regions easier. None of the employed methods allowed for sequencing through this region. As all but 1 of the 94 RHAS HVRI samples contained this T16189C transition (Section 3.4.2) the decision was made to use the forward primer (L15975) to generate the 5’ region of the sample, up to the C-tract, and the reverse primer (H16517) was used to provide the 3’ region after the C-tract (Figure 3.6 B). Sequences were then joined and aligned using MEGA4. Therefore both primers were used to acquire the final HVRI sequences. The HVRI sequences generated in this study will be entered in to GenBank, for access by the scientific community, after this thesis has been submitted.

As noted in previous work (Lum *et al.*, 1994; Redd *et al.*, 1995; Whyte *et al.*, 2005) most Polynesian subjects have additional C-tract length variation introduced by two transversions (A→C) at positions 16182 and 16183 (Figure 3.6 B). This combined with the 16189 transition inevitably made it impossible for the complete sequencing in one direction. After close examination of both forward and reverse sequences, as well as other available Polynesian sequence data, the majority of the RHAS HVRI samples (again all but one) were found to have a C-tract length of 12 C’s. This was later confirmed by Mitochip analysis (Section 3.5)

3.4.1 30 RHAS HVRI Sequences

These were the first 30 RHAS individuals that had been identified as prime candidates for Mitochip analysis. These individuals were chosen on the basis of their genetic Maori ancestry (Hall 2009). Accordingly they all exhibited very high autosomal Maori ancestry (Figure 3.1). It was suggested that by selecting individuals with high nuclear ancestry individuals with Maori mtDNA would also be selected, which was crucial for this project. The variability in the HVRI sequences was evaluated in 94 RHAS mtDNA by analysing the difference between RHAS sequences and the CRS. The sequence data was compiled and variants were documented in Table 3.2.

Of the 30 sequences, 24 exhibited the classic Polynesian CCGT Motif (T16189C, T16217C, A16247G, and C16261T; CCGT), while the other 6 lack the transition, A16247G (Motif being CCAT). From these sequencing results we are able to explain the restriction digest pattern seen in Figure 3.5. As seen in this figure 7 of the 30 samples did not undergo restriction digest (mt6, mt10, mt11, mt12, mt21, mt23, and mt26). Now using the sequence information in Table 3.2 we can see that mt6, 10, 11, 12, 21, and 23 all lack the A→G transition required to create a *Pst* I restriction site, explaining why they did not digest. As for the seventh undigested sample, mt26, the sequence for this individual shows a T→C transition at position 16243. This mutation falls within the recognition site of the restriction enzyme, and alters it so that *Pst* I will no longer cut. Individual mt26 therefore has the classic Polynesian Motif, but another mutation interferes with the restriction by *Pst* I. This is then a novel restriction assay for identifying mtDNA that contains the 16247 transition, making the assay specific for haplotype B4a1a1 as shown by Figure 3.5.

Table 3.2: Table showing HVRI sequence data for 30 RHAS individuals

Position	16051	16674	16686	16626	16619	16672	16682	16683	16689	16690	16692	16627	16623	16647	16660	16662	16661	16627	16622	16629	16629	Ins@162295	16637	16630	16633	16631	16633	16635	16636	16638
	A	A	T	T	G	C	T	A	A	T	C	T	T	C	C	A	C	C	T	G	A		C	—	T	C	T	T	A	C
mt1	A	.	C	T	—	.	.	C	.	.	.	C	.
mt2	.	.	.	C	A	C	.	T	—	C	T	.	C
mt3	C	C	C	.	C	.	G	T	—
mt4	C	C	C	.	C	.	G	T	—	Y	.	.
mt5	C	C	C	.	C	.	G	T	.	A	.	.	.	—
mt6	.	.	C	C	C	C	.	C	.	.	T	—
mt7	C	C	C	.	C	.	G	T	—	A	.
mt8	C	C	C	.	C	.	G	T	—
mt9	.	.	C	C	C	C	.	C	.	G	T	—
mt10	C	C	C	.	C	.	.	T	—
mt11	.	.	C	C	C	C	.	C	.	.	T	—
mt12	C	C	C	.	C	.	.	T	—
mt13	C	C	C	.	C	.	G	T	A
mt14	C	C	C	.	C	.	G	T	—
mt15	C	C	C	.	C	.	G	T	—
mt16	C	C	C	.	C	.	G	T	T	—
mt17	C	C	C	.	C	.	G	T	—
mt18	G	.	C	C	C	C	.	C	.	G	T	—
mt19	.	.	C	C	C	C	.	C	.	G	T	—	Y	.	.
mt20	C	C	C	.	C	.	G	T	—	Y	.	.
mt21	.	.	.	C	.	.	.	C	C	C	.	C	.	.	T	—
mt22	C	C	C	.	C	.	G	T	—
mt23	.	.	C	C	C	C	.	C	.	.	T	—
mt24	C	C	C	.	C	.	G	T	—
mt25	C	C	C	.	C	.	G	T	—
mt26	C	C	C	.	C	C	G	T	—
mt27	C	C	C	.	C	.	G	T	.	.	G	.	.	—
mt28	C	C	C	.	C	.	G	T	—
mt29	C	C	C	.	C	.	G	T	—
mt30	C	C	C	.	C	.	G	T	—
mt31	C	C	C	.	C	.	G	T	—
mt32	C	C	C	.	C	.	G	T	—

The above table shows the mitochondrial HVRI sequence data for 32 HVRI products. Shown in blue are 2 Taiwanese Amis HVRI sequences, 30 RHAS Maori HVRI sequences are shown in green. All sequences are aligned to the CRS with positions shown across the top. The four positions of the Polynesian Motif are highlighted; 16189, 16217, 16247, and 16261. Nucleotides that differ to the CRS are represented by their single letter code. All HVRI sequence ID's are listed in numerical order with the prefix mt, these ID's are maintained through-out this study. Sequences that did not undergo restriction are underlined. A dash at 16295 represents no nucleotide insertion.

Nucleotide heteroplasmy is represented as Y in the table (in this case C/T heteroplasmy).

All RHAS mtDNA sequences lacking the A16247G transition were independently PCR amplified a second time, with this PCR product being sequenced allowing the validation of this variant (Figure 3.7 A). This procedure was replicated for all HVRI sequences containing a variant to the CRS. The 6 sequences lacking the final Polynesian motif marker (underlined in Table 3.2), were the same sequences which did not undergo restriction digest detailed in Section 3.3. This again proved that the restriction assay was working correctly. However HVRI sequencing had identified that the 6 sequences were definitely not European as was originally suspected from restriction results.

There are several other mutations of note among these sequences, the first being an insertion of A at position 16295 in individual mt13. Insertions are not uncommon in the D-loop of the mitochondria, but none have been documented in previous work with Polynesian and Maori HVRI samples. As such this mutation was identified as genuine (Figure 3.7 B), and a database search suggested that this is a novel undocumented mutation.

Another point of interest is that 3 individuals demonstrated nucleotide heteroplasmy at the same position, 16354 (Figure 3.7 C). The sequence chromatograms were examined to ensure the position had been called correctly, and then validated by independent PCR of a new HVRI sequence from the same sample, followed by sequencing. All areas of heteroplasmy were identified and replicated in sequences this way, and were deemed to be genuine.

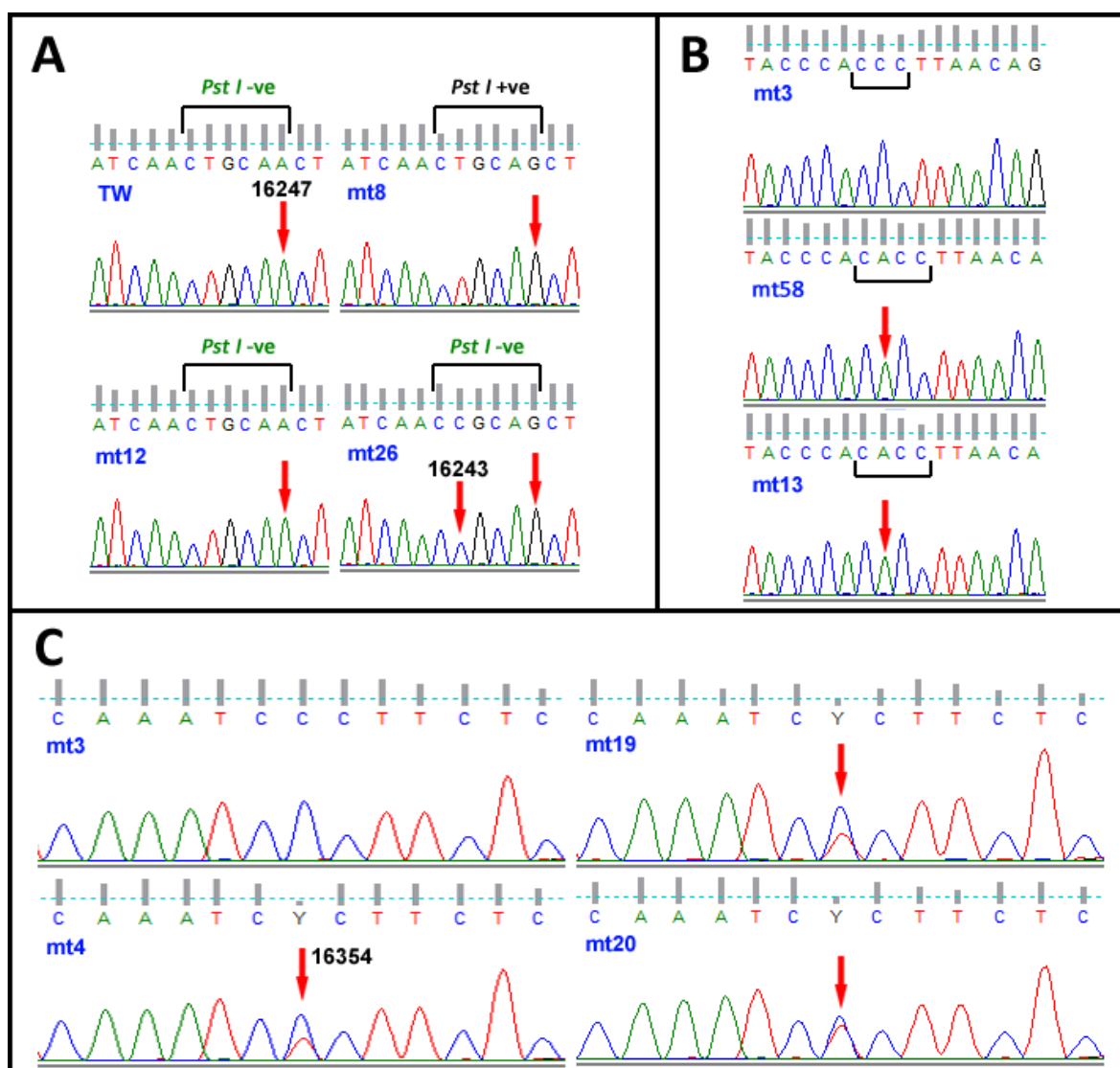


Figure 3.7: Positional variants of the mitochondrial HVRI in RHAS individuals.

A) Shows the presence (*Pst* I +ve) or absence (*Pst* I -ve) of the restriction site created by A16247G. B) Demonstrates the insertion of A at 16295 in two RHAS HVRI. C) Shows C/T heteroplasmy in 3 RHAS HVRI sequences at position 16354.

3.4.2 94 RHAS HVRI Sequences

Following on from the sequencing of the first 30 RHAS HVRI samples it was decided that we would validate the frequency of these haplotypes in a larger cohort before making the final selection for the Mitochip analysis. The reasoning for this is that we wanted to select 20 individual mitochondrial haplotypes as a wide representative of the RHAS mtDNA pool as a whole. So another 64 individual RHAS DNA samples were PCR amplified for the HVRI region, and then added to the original 30 sequences, giving us a total of 94. Again any novel variants were confirmed by preparing separate PCR amplicons and resequencing these.

All sequences were collated into a haplotype table (Table 3.3). From this table it can be seen that there are three major haplotypes; those that contain the full Polynesian ‘CCGT’ Motif, those that have the ‘CCAT’ motif lacking the 16247 A→G transition, and one individual (haplotype MB21) that shares the European haplotype H (Helena) with the CRS. This in itself is interesting as all samples were selected on autosomal DNA ancestry, which makes finding a European mitochondrial haplotype on a predominantly Maori genome peculiar.

All RHAS mtDNA sequences lacking the A16247G transition were independently PCR amplified a second time, with this PCR product being sequenced allowing the validation of this variant (Figure 3.7 A). There were 18 individuals in the RHAS Maori cohort lacking this transition (20%). This proportion seems to correlate well with those found in other HVRI studies that have been conducted in Polynesian and Maori (see position 16247 in Table 3.6).

Table 3.3: Table showing HVRI sequence data for 94 RHAS individuals

CRS	f	16051	16608	16611	16667	16668	16669	16670	16671	16672	16673	16674	16675	16676	16677	16678	16679	16680	16681	16682	16683	16684	16685	16686	16687	16688	16689	16690	16691	16692	16693	16694	16695	Total
CRS		A	T	T	C	T	T	T	C	T	T	A	C	C	T	G	C	A	-	T	A	C	G											
MB 01	45	C	C	C	.	C	.	G	.	T	-											
MB 02	4	C	C	C	.	C	.	G	.	T	.	A	.	.	-											
MB 03	1	G	.	.	.	C	C	C	.	C	.	G	.	T	-											
MB 04	2	G	C	.	.	C	C	C	.	C	.	G	.	T	-											
MB 05	4	C	C	C	.	C	.	G	T	T	-											
MB 06	1	.	.	.	Y	C	C	C	.	C	.	G	T	T	-											
MB 07	5	.	.	C	.	C	C	C	.	C	.	G	.	T	-											
MB 08	2	.	.	C	.	C	C	C	.	C	.	G	.	T	-	.	.	.	Y											
MB 09	2	C	C	C	.	C	.	G	.	T	-	.	.	.	Y											
MB 10	2	C	C	C	.	C	.	G	.	T	A											
MB 11	2	C	C	C	.	C	.	G	.	T	.	.	.	G	-											
MB 12	1	C	C	C	T	C	.	G	.	T	.	.	.	G	-											
MB 13	1	C	C	C	.	C	C	G	.	T	-											
MB 14	1	C	C	C	.	C	.	G	.	T	-	C	.	.	.											
MB 15	1	C	C	C	.	C	.	G	.	T	C	.	.	.	-											
MB 16	1	C	C	C	.	C	.	G	.	T	-	.	.	.	A											
MB 17	7	C	C	C	.	C	.	.	.	T	-											
MB 18	9	.	.	C	.	C	C	C	.	C	.	.	.	T	-											
MB 19	1	.	.	C	.	C	C	C	.	C	.	.	.	T	-	Y	.	.	.											
MB 20	1	C	C	C	.	C	.	.	.	T	-	.	G	.	.											
MB 21	1	T	.	-											
To tal	94																																	1

Table 3.3 shows the mitochondrial HVRI sequence data for all 94 RHAS HVRI products. All sequences are aligned to the CRS with the positions shown across the top. Nucleotides that differ to the CRS are shown. Three major haplogroups are shown; CCGT (green, 80%), CCAT (violet, 19%), and TTAC (orange, 1%). Positions shaded in grey indicated Polynesian motif transitions. A dash at 16295 represents no nucleotide insertion. Nucleotide heteroplasmy is represented as Y in the table (in this case C/T heteroplasmy). These three major haplotypes have been broken down into 21 sub-haplotypes in Table 3.3 (MB01-MB21). This is based on individual variants contained within the sequenced region. Being the Hyper Variable Region of the mitochondrial D-loop it comes as no surprise that such a high rate of mutation was

detected. As mentioned in Section 3.4.1 there are several positions that show heteroplasmy; one more individual that shows the previously mentioned position (16354), and two other individuals showing heteroplasmy at positions 16147 and 16311 respectively. These positions were validated as being genuine heteroplasmy, and tend to show varying degrees between individuals. Another individual sequence contains the same insertion seen in individual mt13 (Table 3.2), insertion of A at 16295. This could indicate that these two individuals are related. The insertion was validated by generating separate amplicons and resequencing, and was confirmed in both individuals (Figure 3.7 B). Searching online databases yielded no result for this polymorphism, leading us to believe it is undocumented.

The variant at position 16126 (T→C) is present in 18% HVRI sequences (17/94). It is interesting to note that this C transition at 16126 is present in 7 sequences exhibiting the traditional Polynesian ‘CCGT’ motif, as well as in 10 sequences which have the ‘CCAT’ motif. This could indicate two separate mutational events involving both variant 16126 as well as the variant 16247. Variant 16126 could well be a RHAS Maori specific signature marker worth validating in the larger Maori community.

OBJECTIVE 2: Obtain entire sequence information from a subsample of Rakaipaaka individuals with Maori ancestry (n=20) using the Mitochip.

3.5 Complete mtDNA sequence data for 20 RHAS Maori individuals

One of this project's primary objectives was to ascertain whether a RHAS Maori specific mtDNA signature existed. This was determined by utilising the Affymetrix Mitochip assay and resequencing 20 entire RHAS Maori mitochondrial genomes. The previous section demonstrates the design and optimisation of a specific RFLP assay and HVRI PCR protocol. These were used to generate a selection pool of HVRI sequences, from which 20 Maori mtDNA could be selected for Mitochip resequencing. After the successful PCR amplification and sequencing of RHAS Maori HVRI the selection pool was n=94. These sequences were interrogated for variation and the results collated into haplotypes (Table 3.3). These 94 sequences became an informative representation of the mitochondrial variation within the Rakaipaaka community. From this cohort 20 individuals were selected to undergo Mitochip analysis. This chip analysis is a resequencing assay that interrogates each nucleotide position of a selected mitochondrial genome (Introduction, Section 1.3.7.1), with the final result being a reconstructed complete genome sequence (~16540 bp). It was hypothesised that by investigating and comparing the complete mtDNA sequences of RHAS Maori individuals against available databases a specific 'signature' motif common to RHAS Maori would be identified. This signature could potentially be useful as a genetic marker for the early diagnosis of specific metabolic disorders, and could provide an invaluable insight into the early maternal ancestry of not only Rakaipaaka, but also other New Zealand Maori. The selected 20 individuals are shown in Appendix Seven.

3.5.1 Mitochip Analysis

Once results from the Mitochip analysis were made available from the University of Auckland Centre for Genomics and Proteomics it was analysed using Affymetrix GSEQ4.1 software. The GSEQ4.1 software package allows the user to run a series of algorithms to gain ‘base calls’ for each position of the mitochondrial genome (see Methods, Section 2.7.2 for algorithm parameters). The algorithms restrict or relax decisions to make a call based on the intensity (fluorescence) of four separate probes at each nucleotide position. Therefore the highest intensity indicates which nucleotide is present at each position. This way the entire mitochondrial genome can be resequenced base by base, and interpreted by the software GSEQ4.1. Due to the nature of the mitochondrial genome and the chemistry involved in the Mitochip probe design, areas that contain runs of 4 or more C’s sometimes result in miss calls. These miss calls are indicated by n’s in the sequence. Adjustment of the GSEQ4.1 algorithm parameters allowed the number of miss calls to be limited to a minimum. It was also discovered that by setting the ‘model type’ parameter to ‘heterozygote’ the chip was better able to detect areas of nucleotide heteroplasmy. Once the software was calling accurately, consensus sequence was obtained for 20 ‘complete’ Maori mitochondrial genomes. These sequences were exported as FASTA files and then imported and aligned in MEGA4.

3.5.2 Variation in 20 Maori mitochondrial genomes

Once the sequence data was exported from GSEQ4.1 it was aligned in MEGA4. FASTA files were then imported into the program mtPhyl (Eltsov 2009). MtPhyl is a new piece of software that can be used to analyse and compare mtDNA, and can be used to create tables and phylogenetic trees. Table 3.4 was created by mtPhyl and edited in Microsoft Excel. This table summarises all the information gained from the

Mitochip analysis of 20 mitochondrial genomes RHAS Maori individuals. The individual sequences are listed down the left of the table, with positional information listed across the top. All sequences are aligned against the CRS; mutations to the CRS are shown as the respective nucleotide at the given position. Blank squares indicate the nucleotide at that position is identical to that within the CRS. Position numbers bolded indicate rare polymorphisms found within the CRS (Andrews *et al.*, 1999). Map locus refers to the region of the mitochondrial genome a given variant is found in (i.e. the HVR, or a specific protein coding gene). For the protein coding regions there is extra information included in the table; Amino acid change indicates whether there has been a substitution and if so what amino acid, the Conservation Index represents how many (in percentage) of 39 species have the same amino acid in that position as human wild type does, and the Protein Position indicates where in the protein the specific amino acid is located. Areas of common variation within the RHAS Maori sequences can clearly be seen highlighted green in Table 3.4. These positions were found in all 20 RHAS Maori sequences. Positions that are highlighted in purple are variants that appear to form unique Maori haplotypes. The rare CRS polymorphisms (positions bolded) were excluded from the suggested RHAS Maori signature. This is due to the fact that they are present in high frequency in other mtDNA's. The rest of the variant positions became our candidate mitochondrial signature for RHAS Maori mtDNA. It is interesting to note that the 7 individuals that exhibit the C1185T transition, lack the A4769G transition which all other individuals contain, indicating some form of genetic linkage.

Table 3.4: Correlation table showing mitochondrial sequence data for 20 RHAS Maori individuals.

	HVRII	12SrRNA	16SrRNA	ND1	ND2	COI	ATP6	COIII	ND3	ND4	S(AGY)	ND5	Cyt b	NC	D-loop & HVRI	Position
	263	1438	2706	3909	5465	7028	9145	9722	10238	11719	12239	14022	15746	16051	16519	CRS
	151	1185	2416		4769	6905	9123	9255					15326	16086	16295	mt3
	146	750	1806			6782	8865						15014	16051	16261	mt6
	73		1692			6261	8860						14766		16260	mt8
															16247	mt9
															16243	mt10
															16217	mt11
															16189	mt12
															16126	mt13
																mt14
																mt16
																mt18
																mt19
																mt21
																mt23
																mt24
																mt25
																mt26
																mt28
																mt29
																mt30
n (variant)	20 20 2	20 20 2	20 20 2	13 20 3	20 20 2	20 20 2	20 20 2	20 20 2	20 20 2	20 20 2	20 20 2	20 20 2	20 20 2	20 20 2	20 20 2	20 20 2
Amino-acid Change																
Conservation Index																
Protein Position																

3.5.3 Variants within Protein Coding Genes

Amino acid information is included in the data generated by the mtPhyl software (see Table 3.4). Information is provided regarding any variant within a protein coding gene, as well as respective conservation indexes where applicable. These values represent how many (in percentage) of 39 species have the same amino acid in that position relative to the human CRS (Ruiz-Pesini *et al.*, 2004; Eltsov 2009). A confidence index of 100 indicates that the site is highly evolutionarily conserved. This provides information about the predicted mutation rate of specific amino acid positions. The positions of the amino acid in the protein, as well as the type of mutation (synonymous, non-synonymous), are listed in the table.

There are 13 synonymous mutations and 8 amino acid substitutions. Of the 8 substitutions it is interesting to note that 4 occur within the gene for cytochrome b; C14766T, T15014C, A15326G, and A15746G. Of these A15326G can be accounted for as one of the 7 rare polymorphisms of the CRS. Extended variant analysis of protein coding and rRNA genes was conducted and explored in Sections 3.7 and 3.8.

This study identified six novel variants within the coding region, and one within the control region (Table 3.5). An extended database search of Mitomap (Ruiz-Pesini 2007) and mtDB (Ruiz-Pesini 2007) returned no hits, thus these positions are deemed to be unidentified in previous literature. Upon the submission of this thesis these variant positions will be entered into the respective databases for the general use of the scientific community.

Table 3.5: Novel mtDNA variants documented in 20 RHAS individuals*

Gene	Nucleotide Change	Protein Change	No. Individuals	Percentage
16SrRNA	m.1806T>C	NA	1 (20)	5
ND1	m.3909C>T	syn	3 (20)	15
COXI	m.6782T>C	syn	1 (20)	5
COXIII	m.9255C>T	p.MT-COXIII:Pro17Ser	1 (20)	5
COXIII	m.9722T>C	syn	1 (20)	5
Cyt b	m.15014T>C	p.MT-Cyt b:Phe90Leu	1 (20)	5
HVRI**	m.16295C insA	NA	2 (98)	2

*Reference sequence is the revised Cambridge Sequence (www.mitomap.org/mitoseq.html)

**HVRI is non-coding and the variant was investigated in a group of 94 individuals, the 6 novel coding variants have been investigated in a group of 20.

NA, not applicable; 16SrRNA, 16S ribosomal RNA; ND1, NADH dehydrogenase subunit 1; COXI, cytochrome c oxidase I; COXIII, cytochrome c oxidase III; Cyt b, cytochrome b; HVRI, mitochondrial hyper variable region I.

The main reason behind obtaining 20 complete mtDNA sequences was to identify if a RHAS Maori signature was present and, if so, to validate this in a larger global dataset. The common variants are highlighted green in Table 3.4. There are 22 variants to the CRS which are shared by all 20 RHAS Maori over their entire mtDNA sequence. It is these variants that will form the basis of the RHAS Maori mtDNA signature. It must be noted that these 22 variants are those ancestral to, and therefore shared by, Polynesian mtDNA. Highlighted in purple (Table 3.4) however, are 3 other variants (C1185T, A4769G, and T16126C) that form unique haplotypes, which appear to be RHAS Maori specific. Together all these variants become the specific RHAS mtDNA signature. It is the 3 ‘unique’ variants that will be of interest for validation in the larger RHAS Maori population. Another point of interest is the fact that position 16247 (indicative of Polynesian ancestry) appears to be variable among these sequences, this will be reflected on in discussion. This newly defined Maori mtDNA signature is explored and validated within a global dataset in the following section (Section 3.6).

OBJECTIVE 3: Align and compare the sequence information with independent worldwide data to identify a RHAS Maori specific signature.

3.6 Population analysis

These 20 RHAS Maori sequences are the only entire Maori mtDNA's currently available (at time of writing). As such this project provides the first opportunity to investigate the evolutionary similarities between Polynesian and Maori mtDNA. It has been speculated that they are identical, with little to no diversity, based on HVR I & II data (Hertzberg 1989; Lum *et al.*, 1994; Sykes *et al.*, 1995). It is currently suggested that there is very limited mtDNA variation in Polynesians in general, and even less in Maori. As such, no Maori informative genetic mtDNA markers have yet been identified, and excluding novel variants, most Maori are expected to share the same mtDNA lineage.

By utilising RHAS HVRI and entire mtDNA sequence data this study aims to answer several important questions:

- How does Maori mtDNA compare with other global ethnicities mtDNA?
- Can this be used to establish a specific RHAS Maori mtDNA profile signature?
- What does this tell us about Maori mtDNA ancestry?

3.6.1 Variant Frequency Tables - Identification of a RHAS signature?

Online databases (Mitomap, mtDB, and PhyloTree) were used to gather global mtDNA sequence data from various ethnic groups; European, Chinese, Japanese, Taiwanese, Melanesian, and Polynesian. Additional HVRI sequence data was also obtained for 100 New Zealand Maori sequences (see Appendix Eight for accession numbers). All

sequences were combined with the 20 RHAS Maori mtDNA sequences. Frequency tables were generated to validate the possible RHAS specific mtDNA signature (see Appendix Nine for raw data). Frequency data was collated and compared to RHAS Maori variant data gained from both HVRI sequencing and Mitochip analysis (see Table 3.6). This table shows all RHAS Maori variants that exhibit a high frequency (nearly 100% in all cases) when compared to differing ethnic groups. These groups are; European, Chinese, Japanese, Taiwanese, Melanesian, Polynesian, and Maori (RHAS for complete mtDNA, and includes other NZ Maori HVRI sequences where applicable). Variant position and the mitochondrial region containing it is shown, with wild-type and respective variant nucleotide given for each position. Each nucleotide is represented by a percentage, with the number of sequences investigated in brackets next to it. In some cases only partial mtDNA sequence information was available. Also some positions show low sequence count due to the lack of complete mitochondrial genomes available for certain ethnic groups (e.g. Taiwanese and Polynesian).

Table 3.6 shows the comparison and validation of a proposed Maori mtDNA signature. It is not yet RHAS Maori specific as, with the exception of 1185 and 4769, all the other included variants are common in Polynesian mtDNA sequences. This is the signature that can be used to identify Polynesian and Maori mtDNA, an 'extended' Polynesian motif that covers the entire mtDNA. This correctly suggests that Maori mtDNA has come from Polynesia. Yet the RHAS sequences are not identical. Positions 1185 and 4769 have been included as they show high frequency in RHAS Maori relative to other included ethnic groups, including Polynesians, and are suspected to be RHAS Maori specific. The frequencies seen in other populations support this theory, with both 1185>T and 4769>A being present in 35% of the RHAS Maori sequences.

Table 3.6: Variant Frequencies of Mitochondrial Polymorphisms between different ethnic groups

Gene	HVRII		12S rRNA				ND2				COI	
	146		1185				4769				6719	
	T	C	T	C	T	C	A	G	T	C	T	C
European, % (n)	88 (22)	12 (3)	...	100 (101)	1 (1)	99 (100)	100 (101)	...	100 (101)	...
Chinese, % (n)	84 (41)	16 (8)	...	100 (49)	100 (49)	100 (49)	...	100 (49)	...
Japanese, % (n)	95 (51)	5 (3)	...	100 (62)	100 (62)	98 (61)	2 (1)	100 (62)	...
Taiwanese, % (n)	42 (5)	58 (7)	...	100 (12)	8 (1)	92 (11)	33 (4)	67 (8)	50 (6)	50 (6)
Melanesian, % (n)	63 (31)	37 (18)	2 (1)	98 (55)	100 (56)	84 (47)	16 (9)	84 (47)	16 (9)
Polynesian, % (n)	...	100 (9)	...	100 (9)	100 (9)	22 (2)	78 (7)	22 (2)	78 (7)
RHAS Maori, % (n)	...	100 (20)	35 (7)	65 (13)	35 (7)	65 (13)	...	100 (20)	...	100 (20)
Gene	ATP6		ND3				TS2				Cytb	
	9123		10238				12239				15746	
	G	A	T	C	T	C	C	T	A	G	A	G
European, % (n)	100 (101)	...	97 (98)	3 (3)	100 (101)	...	100 (101)	...	100 (101)	...
Chinese, % (n)	100 (49)	...	100 (49)	100 (49)	...	100 (49)	...	100 (49)	...
Japanese, % (n)	98 (61)	2 (1)	100 (62)	100 (62)	...	100 (62)	...	100 (62)	...
Taiwanese, % (n)	33 (4)	67 (8)	50 (6)	50 (6)	50 (6)	50 (6)	100 (12)	...	50 (6)	50 (6)
Melanesian, % (n)	84 (47)	16 (9)	84 (47)	16 (9)	84 (47)	16 (9)	86 (48)	14 (8)	84 (47)	16 (9)
Polynesian, % (n)	22 (2)	78 (7)	22 (2)	78 (7)	22 (2)	78 (7)	22 (2)	78 (7)	22 (2)	78 (7)
RHAS Maori, % (n)	...	100 (20)	...	100 (20)	100 (20)	...	100 (20)	...	100 (20)
Gene	HVR1		16217				16247				16519	
	16189		16217				16247				16519	
	T	C	T	C	T	C	A	G	C	T	T	C
European, % (n)	97 (105)	3 (3)	100 (108)	100 (108)	...	100 (108)	...	26 (28)	74 (80)
Chinese, % (n)	65 (33)	35 (18)	88 (55)	12 (6)	100 (51)	...	94 (58)	6 (3)	55 (28)	45 (23)
Japanese, % (n)	63 (75)	37 (45)	93 (111)	7 (9)	100 (120)	...	93 (111)	7 (9)	53 (63)	47 (57)
Taiwanese, % (n)	63 (230)	37 (133)	73 (264)	27 (99)	100 (363)	...	84 (304)	16 (59)	41 (78)	59 (110)
Melanesian, % (n)	59 (95)	41 (65)	71 (114)	29 (46)	74 (119)	26 (41)	76 (121)	24 (39)	46 (73)	54 (87)
Polynesian, % (n)	5 (6)	95 (114)	5 (6)	95 (114)	19 (23)	81 (97)	7 (9)	93 (111)	21 (25)	79 (95)
NZ Maori, % (n)	11 (8)	89 (67)	11 (8)	89 (67)	20 (15)	80 (60)	12 (9)	88 (66)	no data	no data
RHAS Maori, % (n)	1 (1)	99 (93)	1 (1)	99 (93)	20 (19)	80 (75)	1 (1)	99 (93)	...	100 (20)

Again from Table 3.6 it was noted that the C→T transition at position 1185 is quite interesting. This variant demonstrates a 35% frequency in the RHAS Maori cohort, while the other ethnic groups investigated do not contain this mutation at all (with the exception of 2; an Indian [AY71401] and a Melanesian [DQ137401] individual). As can be seen in Table 3.6 this is also the case for the variant A4769G. These two variants appear to exhibit a form of genetic linkage and are present together in 7 RHAS Maori.

As this study is the first to sequence and document entire Maori mtDNA sequences, there was no available Maori coding region data for comparison. There is also unfortunately little coding region information available for other Polynesian mtDNA sequences, with only 9 complete Polynesian sequences listed in mtDB (Ingman 2006). As such the possible RHAS Maori specific variants identified in Section 3.5 were unable to be incorporated into this analysis. A database screen of Mitomap (Ruiz-Pesini 2007) and mtDB (Ingman 2006) however, provided information about the global frequencies of these variants. As mentioned, 1185>T exhibits a frequency of 0.07% (2/2800). The other two RHAS signature variants show marginally higher global frequencies; 4769>A = 1% (30/2800) and 16126>C = 6% (166/2800). All comparisons were made using mtDB (Ingman 2006). It is obvious that all three variants exist at very low frequencies globally, and an extended search returned no catalogued sequence containing all three. This further confirmed that these markers would be useful in defining a RHAS Maori mtDNA specific signature.

3.6.2 Haplotype analysis: New Zealand Maori mtDNA HVRI

Previous research using Maori HVRI sequences has attempted to shed light on the possible size of the Maori female founding population. Earlier work in 1998 used 3' HVRI haplotypes to test migration patterns, and estimated the founding population size of NZ Maori to be ~70 (Murray-McIntosh *et al.*, 1998). This was later expanded when Whyte *et al.* (2005) identified that additional unique HVRI haplotypes were present in Maori. These findings increased the total number of haplotypes from 3 to 7. When simulations of the combined data sets were conducted the estimated number of founding females was greatly expanded to ~190 (Whyte *et al.*, 2005). With the discovery of numerous unique HVRI haplotypes in Rakaipaaka HVRI sequences, it was decided to investigate how these affect the total amount of Maori haplotypes. In Section 3.4.2, 21 different HVRI haplotypes were reported from 94 Rakaipaaka Maori HVRI sequences. These were combined with the two existing data sets, AW (Whyte *et al.*, 2005) and MM98+ (Murray-McIntosh *et al.*, 1998), and collated in Table 3.7.

Table 3.7: 3' HVRI Control Region mtDNA Haplotypes: New Zealand Maori

Haplotype	Number Found in Each Data Set			
	RHAS Maori	AW	MM98+	Combined Data Set
MB1/AW1†	54	44	47	145
AW3	0	1	1	2
MB15/AW7	1	1	0	2
MB02	4	0	0	4
MB05†	5	0	0	5
MB08†	4	0	0	4
MB10	2	0	0	2
MB11	2	0	0	2
MB17/AW2†	16	5	5	26
Others‡	5	3	0	8
<i>N</i> (samples)	93	54	53	200
<i>N</i> (haplotypes)	13	7	3	17

† samples within these categories are known to contain variation outside np 16189-16365, the region of interest. ‡ observations in these categories/groups are all singleton haplotypes.

NB data exclude all mtDNA haplotypes that are European-like.

Table 3.7 details all three data sets, and the number of sequences within each haplotype. It can be clearly seen that the MB1/AW1 haplotype is the most common found in NZ Maori, showing the highest numbers in all three data sets and 73% of total sequences (145 out of 200). All sequences within MB1/AW1 contain the ‘CCGT’ Polynesian motif. The second most common haplotype is MB17/AW2, which is present in approximately 8% (16 out of 200) of the surveyed sequences. All sequences in this haplogroup have the pre cursor ‘CCAT’ motif. The addition of the RHAS HVRI data has greatly expanded the previous number of recorded haplotypes, from 7 to 17. Re-running population simulations is outside the scope of this study, however implications of the expanded haplotype diversity will be discussed.

The previous research used 3’ HVRI sequences between nucleotide position 16189-16365. However in this study there were several unique variants identified outside of this range. This was also the case with the AW data set. To expand on this the RHAS haplotypes have been combined with the AW data set (Table 3.8), which increases the number of detected HVRI haplotypes in Maori.

In Table 3.8 the length of HVRI sequence included is extended past the range set by Murray-McIntosh (1998). It can be seen that variation outside of this range is relatively high, when included in the analysis it increases the number of haplotypes from 17 to 25. This is nearly three times the number of previous calculated Maori mtDNA haplotypes (Whyte *et al.*, 2005).

Table 3.8: HVRI Control Region mtDNA Haplotypes: 2 New Zealand Maori Data Sets

<i>Haplotype</i>	<i>Number Found in Each Data Set</i>		
	<i>RHAS Data Set</i>	<i>AW Data Set</i>	<i>Combined Data Set</i>
MB1/AW1	45	44	89
MB18	9	0	9
MB17/AW2	7	4	11
MB07	5	0	5
MB02	4	0	4
MB05	4	0	4
MB04	2	0	2
MB08	2	0	2
MB09	2	0	2
MB10	2	0	2
MB11	2	0	2
MB15/AW7	1	1	0
Others‡	8	5	13
<i>N</i> (samples)	93	54	147
<i>N</i> (haplotypes)	20	8	25

‡ observations in these categories/groups are all singleton haplotypes.

NB data exclude all mtDNA haplotypes that are European-like.

So with this haplotype analysis the number of distinct Maori mtDNA haplotypes had been greatly expanded upon. It must be taken into account that this number is actually still an underestimate of the total haplotype diversity likely to be observed. This is due to the fact that the RHAS dataset has come from one Maori community, the Rakaipaaka iwi. As such there are family members enrolled in the study who would share the same mtDNA, thus underestimating the total diversity within the community.

3.6.3 Phylogenetic Analysis

To further investigate the phylogenetic structure of the RHAS Maori sequences, a neighbour-joining tree was created in MEGA4. The tree with the highest parsimony is shown in Figure 3.8. A selection of complete mtDNA sequences that represent various haplogroup were downloaded from mtDB (Ingman 2006). These included sequences from Africa, Aboriginal Australia, Highland and Coastal PNG, Europe (CRS), Aboriginal Taiwan, Japan, Korea, Polynesia, and NZ Maori. Figure 3.8 shows the neighbour-joining tree, with bootstrap values. The tree was rooted on the oldest mtDNA lineage, deemed to be L from Africa. It can be clearly seen that the two RHAS Maori sequences are grouped well within the major haplogroup B, showing close association with other Polynesian mtDNA sequences. These in turn group within the broader B haplogroup, grouping within Coastal PNG and Aboriginal Taiwanese sequences. This is supported by previous work (Trejaut *et al.*, 2005).

Maori and Pacific Islanders mitochondrial haplotypes belong to the haplogroup B. Haplogroup B arises from haplogroup R on the base of two mutations, 8271-9bp deletion and 16189 T→C. B4 is created by 16217 T→C, B4a is typified by 16261 C→T. Three transitions, 146 T→C, 6719 T→C, and 10238 T→C, create B4a1a, this is the sub-haplotype that the Taiwanese Amis Tribe belong to (Trejaut *et al.*, 2005). Another transition mutation at 14022 A→G forms B4a1a1, then another at 16247 A→G forms B4a1a1a. It is these haplotypes, B4a1a1 and B4a1a1a, which form the bulk of Rakaipaaka mitochondrial genomes. It is also these haplotypes which are directly attributed to by variants of the Polynesian motif; the ‘CCGT’ motif is B4a1a1a, while the ‘CCAT’ motif is B4a1a1 (van Oven *et al.*, 2009).

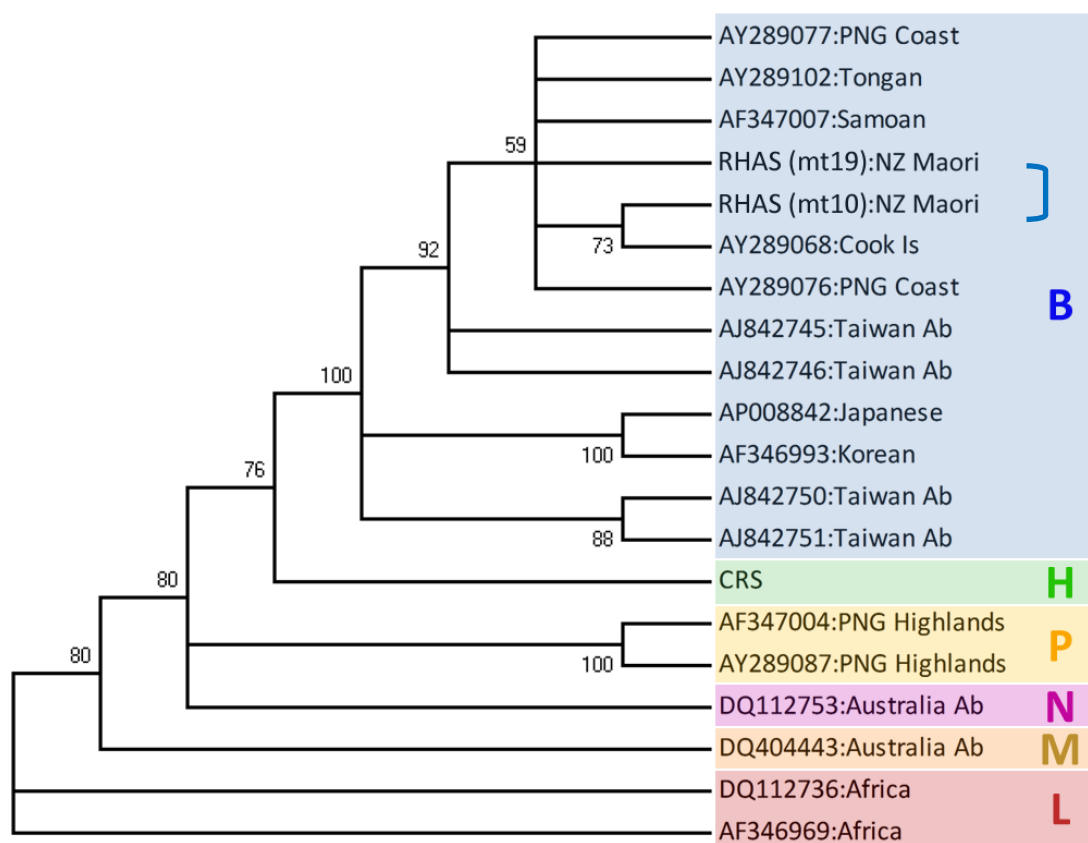


Figure 3.8: A neighbour-joining tree showing haplotype mapping across a selection of complete mitochondrial genome sequences. The tree was created in MEGA4 using 1000 bootstrap replicas and the Tamura-Nei model of base pair substitution. All GENBANK Accession numbers are listed for each sequence. Sequences are grouped and coloured according to mitochondrial haplogroup. The tree was rooted on the African mtDNA sequence AF346969.

It was established within the software suite mtPhyl (Eltsov 2009) that all 20 RHAS Maori sequences belonged to haplogroup B. This was the expected outcome as previous studies have detailed that the majority of Polynesian mtDNA belong to haplogroup B (Hertzberg 1989; Lum *et al.*, 1994; Redd *et al.*, 1995; Sykes *et al.*, 1995), as well as work previously conducted on Maori HVRI sequences (Murray-McIntosh *et al.*, 1998; Whyte *et al.*, 2005). This was further validated in MEGA4 by generating a phylogenetic tree to demonstrate haplotype structuring (Figure 3.8). It was decided to investigate the extent of haplotypes within the RHAS mtDNA pool. To further explore the phylogenetic relationship Figure 3.9 shows a reconstruction based on the sequences from Trejaut *et al.*, (2005). The 20 complete RHAS mitochondrial genomes were

combined with Trejaut's dataset. This analysis was conducted to investigate where the complete RHAS Maori mtDNAs would fit in the overall structure of haplogroup B. Figure 3.9 clearly demonstrates RHAS sequences grouping within the subhaplogroup B4a1a1, this was determined by using the complete mitochondrial haplogroup tree available from PhyloTree (van Oven *et al.*, 2009). The individual sequence mt18 was found to be of subhaplogroup B4a1a1a1, owing to the presence of the variant A6905G.

This tree clearly shows the presence of 2 clades which contain a large portion of the RHAS sequences; one with 7 sequences, the other with 8 (15 out of the total 20). In the previous section it was eluded to the fact that within the 20 RHAS Maori sequences there markers that defined distinct haplotypes within Maori. This tree confirms a new haplotype within the B4a1 grouping, defined by the presence of the C1185T variant (and the absence of A4769G). Going from the naming system implemented in the Phylotree database this would be termed B4a1a1a2. This is a group of 7 sequences clearly seen separated and underlined on the tree (Figure 3.9). The clade of 8 RHAS sequences are grouping due to the presence of the T16126C variant found within the HVRI. This variant could also indicate the presence of another Maori specific haplotype.

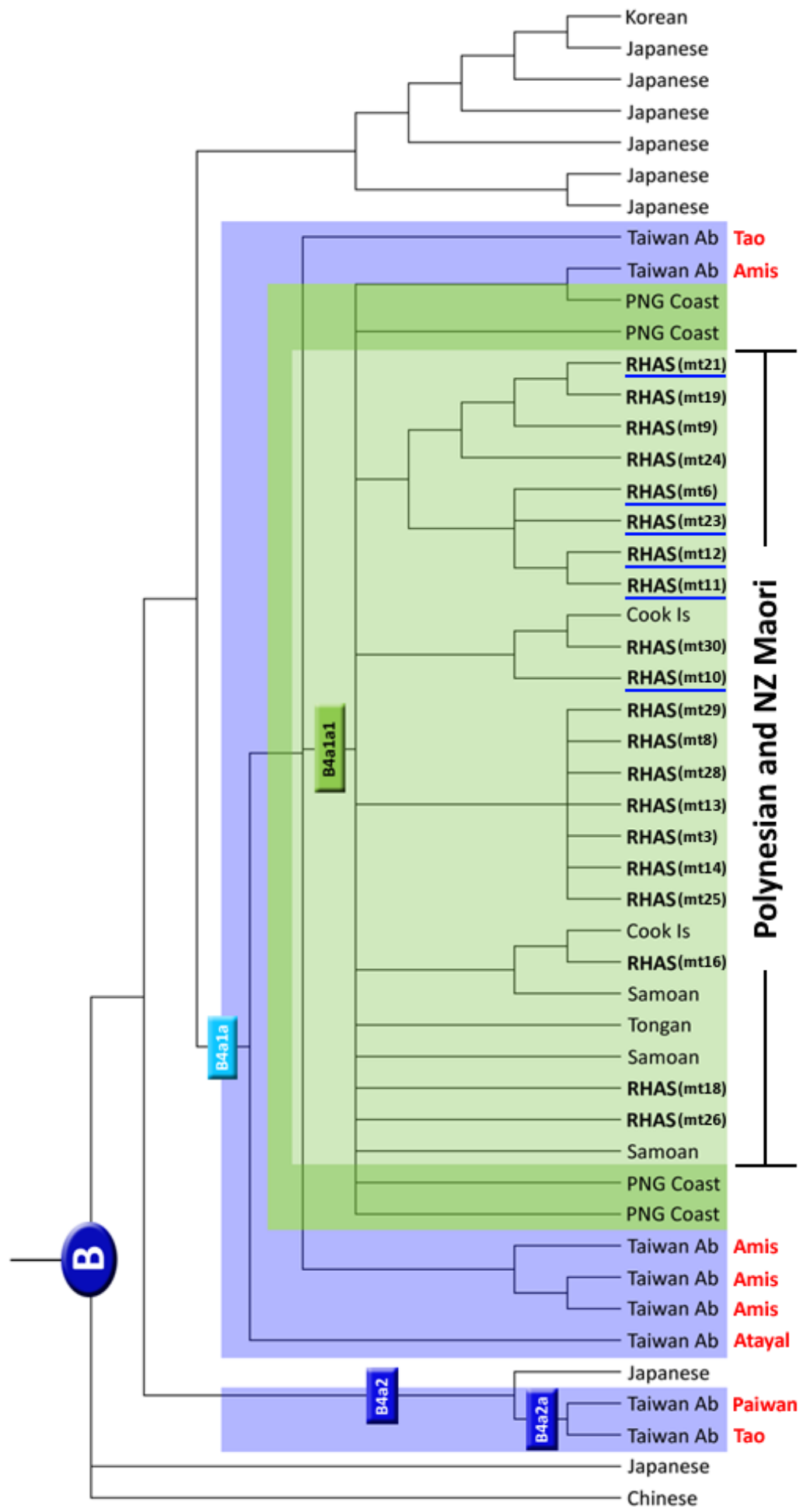


Figure 3.9: Phylogenetic reconstruction detailing haplogroup B4a1 in complete mtDNA sequences.

This neighbour-joining tree was created in MEGA4, using the Tamura-Nei substitution model. The sequences used were sourced from Trejaut *et al.* (2005), and include the 20 complete RHAS Maori mtDNA's. Shown in red are the respective tribes of the Taiwanese Aboriginal sequences. The 6 RHAS sequences underlined in blue have the 'CCAT' motif, while the other 14 RHAS sequences have the 'CCGT' motif.

3.7 Secondary structure analysis of 12S and 16S rRNA

With information gathered from the Mitochip data, we have noted 3 12S rRNA variants occurring within the 20 individuals; A750G, C1185T, and A1438G. Two of these variants (A750G and A1438G) have been previously identified as rare European polymorphisms within the CRS (see Andrews *et al.*, 1999 for a complete list of rare polymorphisms). This means that variants at these positions are actually the most globally common base, but appear as variation when compared to the CRS. All 20 individuals had these two variants, creating the 12S rRNA motif 'GCG'. The third variant that is present within the Mitochip group is C1185T. This variant is of great interest as it was identified in 7 of the 20 complete mtDNA sequences (35%), and is not a known B haplogroup mutation. It also seems to share linkage association with the variant at position 4769 (previously reported in Section 3.5.4, Table 3.4). The presence of this variant creates a sub-haplotype within the 12S rRNA sequences; with 13 individuals having the 12S motif 'GCG', while the 7 containing C1185T are 'GTG'. To further explore this variation at 1185 two tests were conducted; prediction of 12S rRNA secondary structure models, and a conservation analysis.

To determine if any of these 12S rRNA variants had an effect on the possible secondary structure, a series of modelling predictions were generated. These predictions were made using RNAfold (Vienna RNA web servers; <http://rna.tbi.univie.ac.at/website>) by utilising minimum free energy (MFE) parameters adapted from those in Ballana *et al.* (2006). Three separate 12S rRNA sequences were used to predict the final models (Figure 3.10); A) CRS to predict a wild-type model, B) mt6 (represents 'GCG' motif), and C) mt3 (represents 'GTG' motif).

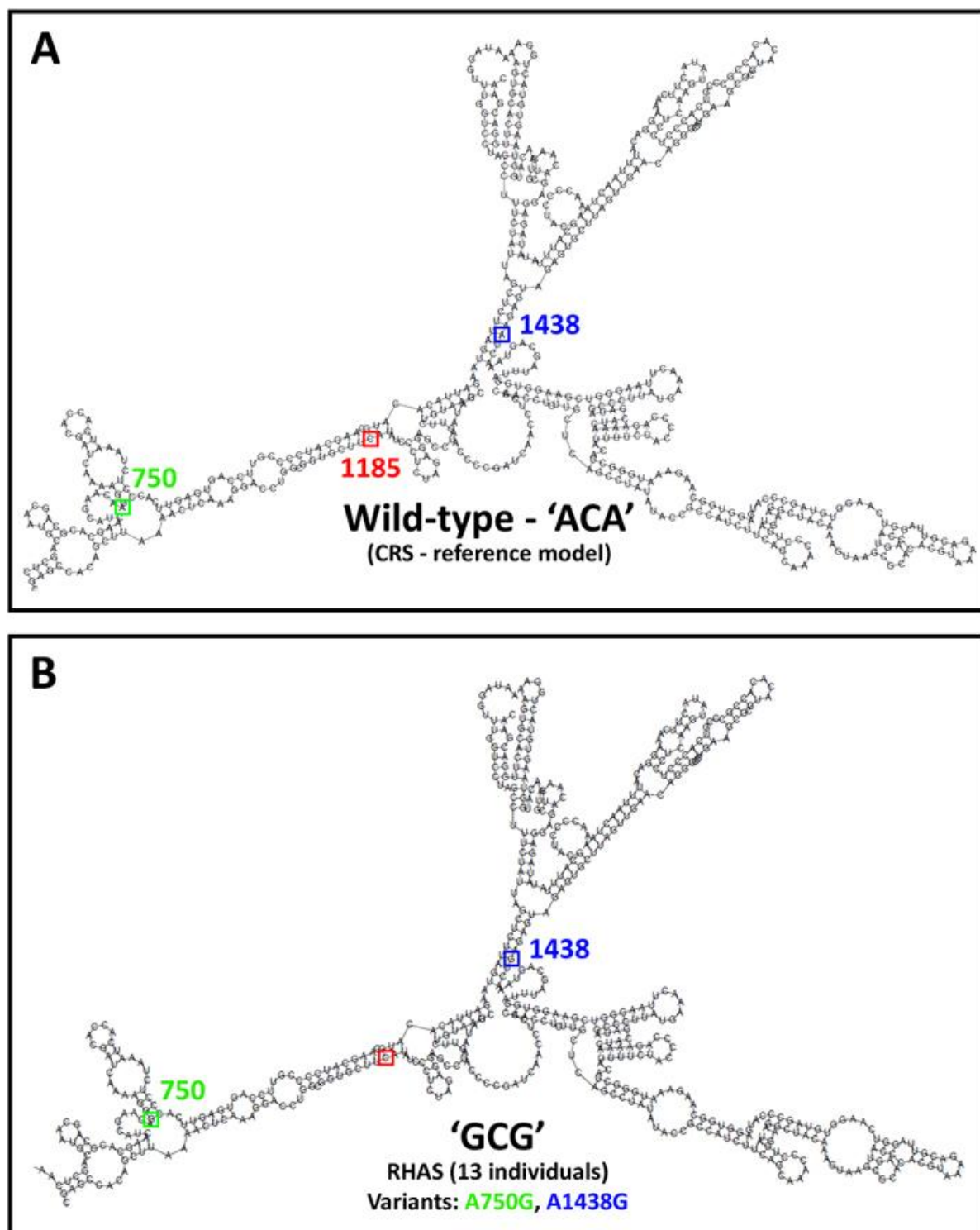


Figure 3.10: Predicted secondary structure models of human mitochondrial 12S rRNA.

A) wild-type model predicted from CRS, showing the location of all 3 RHAS variants (750, 1185, and 1438). B) 'GCG' model was predicted from 12S rRNA sequence from mt6, this represents 13 of the 20 RHAS 12S rRNA sequences. Two variants are present in this sequence, A750G and A1438G. 12S rRNA secondary structures were predicted using RNAfold webserver.

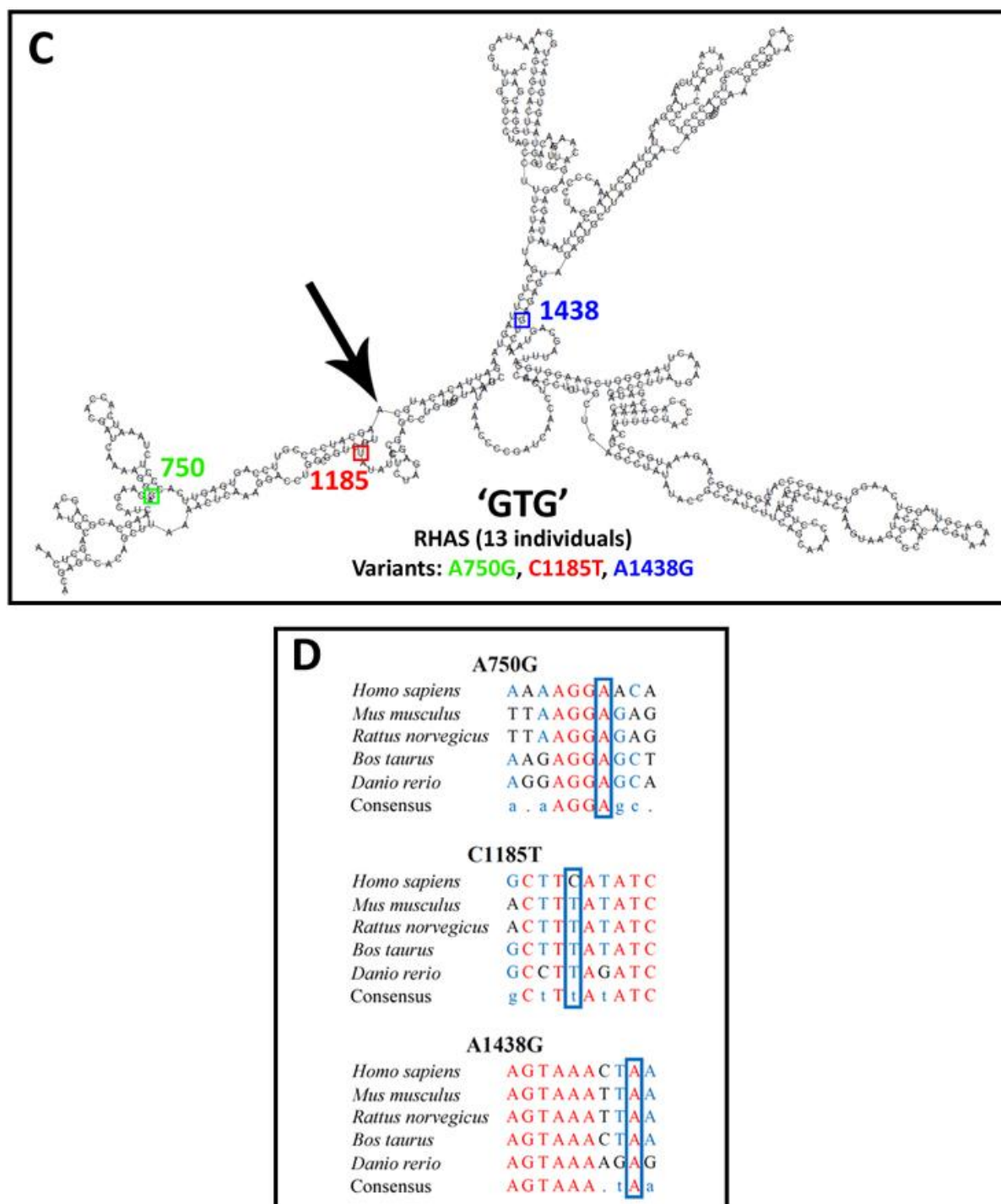


Figure 3.10: Predicted secondary structure models of human mitochondrial 12S rRNA.

C) ‘GTG’ model was predicted from 12S rRNA sequence from mt3, this represents the 7 of the 20 RHAS 12S rRNA sequences. All three variants are present in this model. The arrow indicates a significant conformational change induced by the presence of the C1185T variant. D) a conservation analysis comparing 12S sequences between 5 different organisms; *Homo sapiens*, *Mus musculus*, *Rattus norvegicus*, *Bos Taurus*, and *Danio rerio*. A consensus sequence surrounding each variant position is shown. 12S rRNA secondary structures were predicted using RNAfold webserver.

The wild-type 12S rRNA model generated by CRS sequence (Figure 3.10 A) shows the MFE predicted model. This model was used to identify any structural differences due to 12S rRNA variants. As variants at A750G and A1438G are actually rare polymorphisms found in the CRS, it was expected that there would be little influence of these variants on secondary structure. This was proven to be the case as seen in both models B and C from Figure 3.10. However a noticeable structural change is seen in the 12S rRNA sequence containing the C1185T variant (Figure 3.10 C). In this model it appears that C1185T causes a shortening of an RNA ladder and a disassociation of a small hairpin from a larger loop (indicated by arrow in Figure 3.10 C). A database search for other C1185T variants returned only 2 other sequences (DQ137401, AY714047). MEGA4 analysis revealed that while these 2 sequences contained C1185T, they did not have the same association with position 4769 as observed in these 7 RHAS sequences. This suggests that C1185T is a rare mutation, even more so when in combination with 4769A, which could possibly define a unique haplotype in RHAS Maori.

The C1185T variant was found in 7 of the 20 mtDNA sequences. A conservation analysis was used to determine the evolutionary significance of each of the 3 variants (adapted from Ballana *et al.* (2006)). Five different organism's 12S rRNA sequences were aligned and a consensus sequence was generated for the areas directly surrounding each of the 3 variants (Figure 3.10 D). The conservation analysis showed that all 3 positions appear to be highly conserved through evolution, with the exception of position 1185. All other sequences show a T at 1185, but the human sequence has a C.

Four variants were also identified within the 16S rRNA of the 20 complete mtDNA's; A1692G, T1806C, T2416C, and A2706G. Again RNAfold webserver was used to predict secondary structure models for the 4 separate 16S motifs. The same MFE parameters utilised in the 12S analysis were used as a starting point (see Methods Section 2.7.3 for detailed parameters). These parameters were adjusted until a consensus model was generated for all 4 sequences. It is important to note that these are predicted models based on MFE, as yet the secondary structure model of human mitochondrial 16S rRNA is yet to be elucidated. The predicted secondary structure models are shown in Figure 3.11. In this figure A is the wild-type model, based on CRS 16S sequence, predicted by the RNAfold webserver. This figure indicates the 4 separate positions identified as variable within RHAS mtDNA from the complete sequence data; A1692G, T1806C, T2416C, and A2706G. So the wild-type 16S motif is 'ATTA' (Figure 3.10 A). The other 3 16S motifs are; B) 'ATTG' the most common (85%), C) 'ACTG' only 1 individual (5%), and D) 'GTCG' (10%). Comparing these 3 models to the wild-type shows no significant change to the secondary structure due to the variation within the 16S rRNA gene. There is a slight change in the due to the A2706G, the loss of a small loop as seen in the wild-type model. By looking at the position of each variant on the wild-type model, they all seem to be in structural regions that are 'flexible' in terms of base conservation – a transition mutation won't affect the structure so conservation appears to be relaxed.

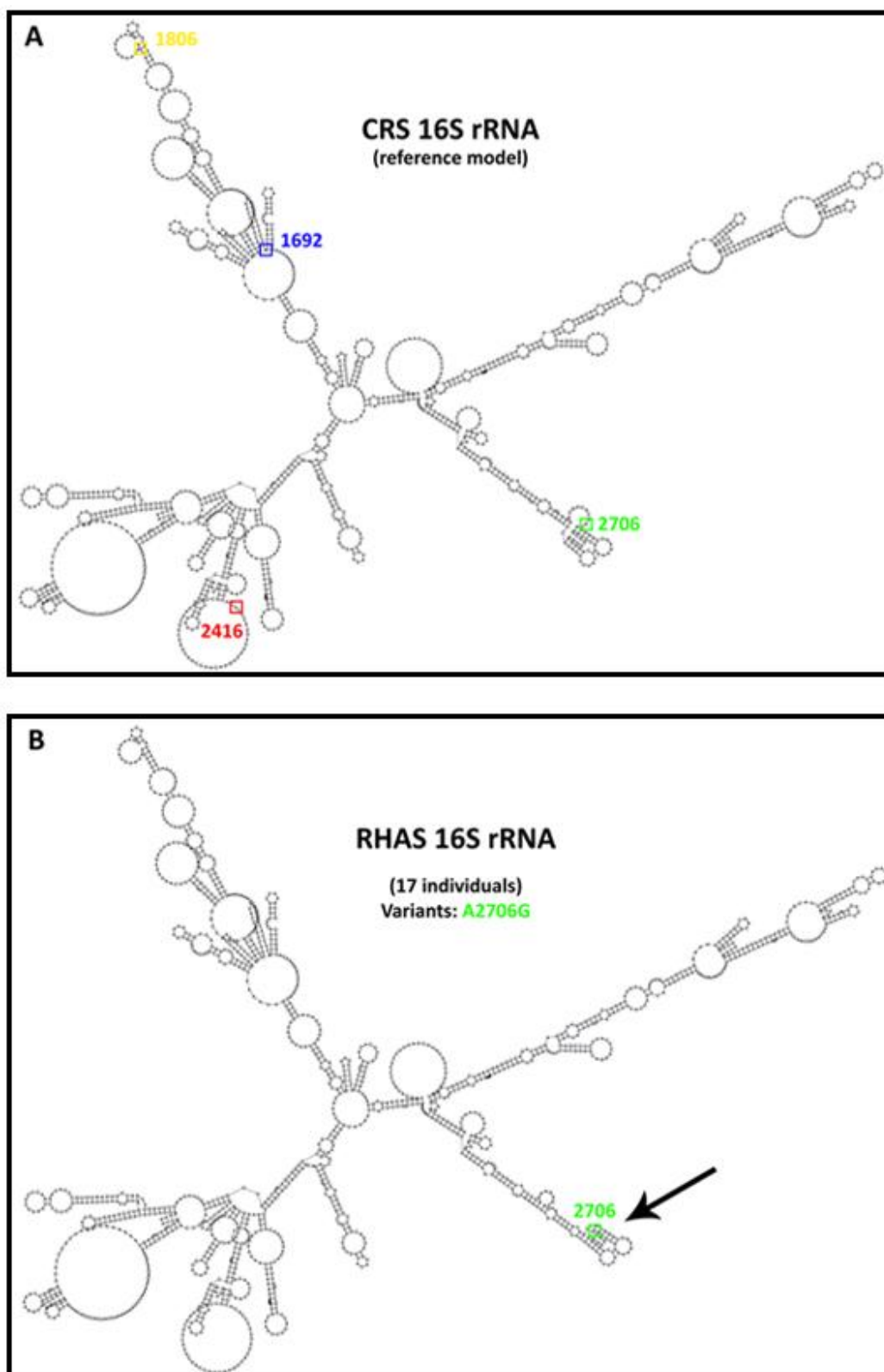


Figure 3.11: Predicted secondary structure of human mitochondrial 16S rRNA.

16S rRNA secondary structures were predicted using RNAfold webserver. A) wild-type model predicted from CRS, showing the location of all 4 RHAS variants (1692, 1806, 2416, and 2706). B) was predicted from 16S rRNA sequence from mt3, this represents 17 of the 20 RHAS 16S rRNA sequences. Figure generated using the RNAfold webserver software (<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>).

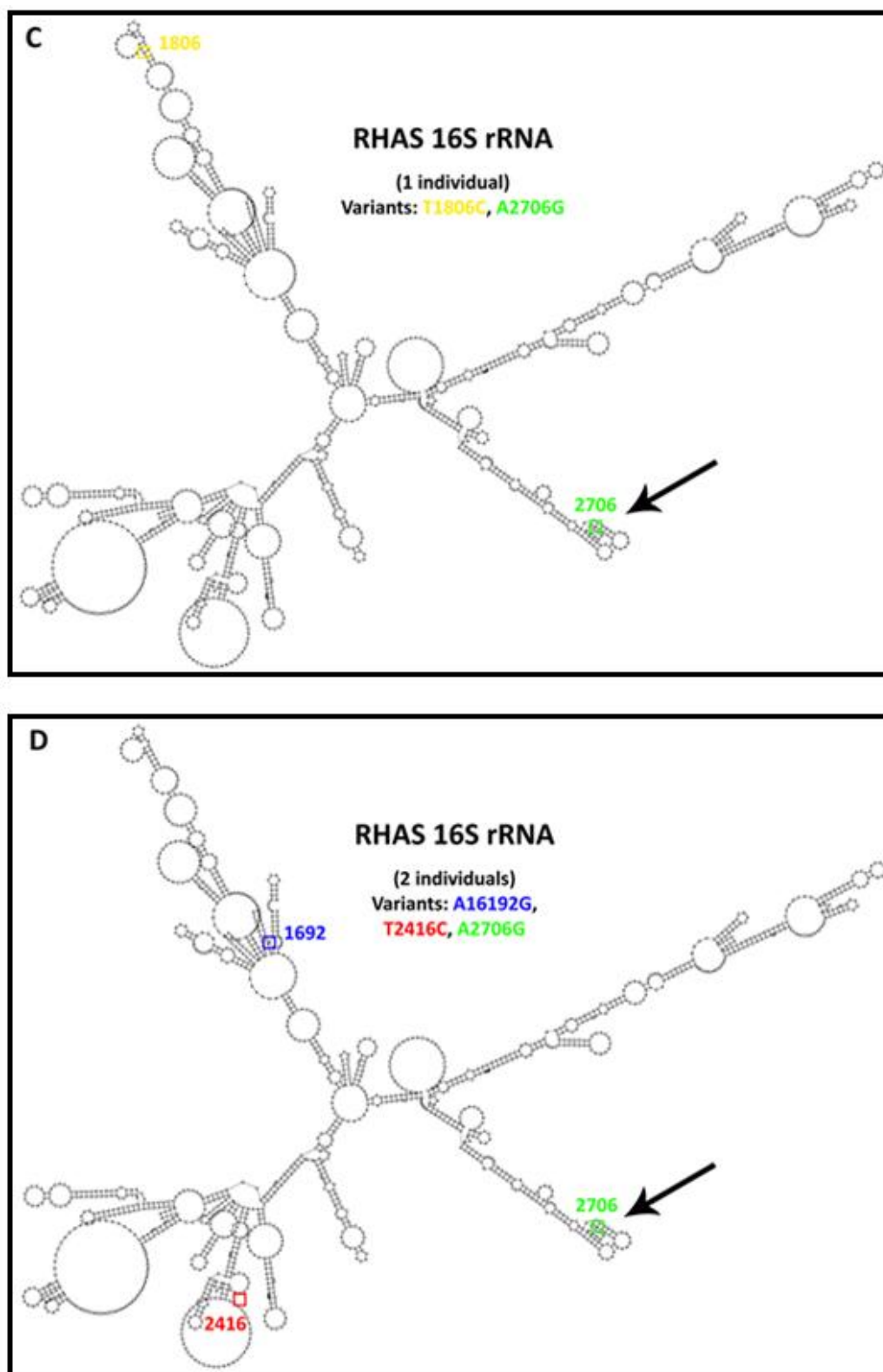


Figure 3.11: Predicted secondary structure of human mitochondrial 16S rRNA.

C) was predicted from 16S rRNA sequence from mt29, this represents the 1 of the 20 RHAS 16S rRNA sequences. D) was predicted from 16S rRNA sequence from mt10, this represents 2 of the 20 RHAS 16S rRNA sequences. The arrows indicate the loss of a small loop due to the A2706G transition in all 3 variant models. Figure generated using the RNAfold webserver software (<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>).

3.8 Protein modelling

While the mitochondrial genome only encodes 13 proteins, they are all crucial subunits of the OXPHOS system. Hence any mutations that detrimentally alter proteins can disrupt production of ATP, increase ROS levels, and be associated with diseases such as exercise intolerance, cancer, neurological disorders, and deafness. These proteins are vital to the correct functioning of OXPHOS, and much has been discovered about their functions, molecular structure, and interactions. Elucidated structural models for mitochondrial encoded OXPHOS proteins have been created and can be freely download from databases such as SwissProt (Kopp *et al.*, 2004; Kiefer *et al.*, 2009).

The aim of this modelling analysis was to identify if any of the substituting variants within the mtDNA encoded OXPHOS proteins of the 20 RHAS sequences affected structure or interactions. In Section 3.5, the variation of the 20 RHAS mtDNA's was described. There were 21 mutations within 9 protein coding genes. Of these 8 mutations resulted in a change in the respective amino acid, the other 13 mutations were synonymous. The protein coding genes containing amino acid changes were; COI, ATP6, COIII, and Cyt b. It should be noted that two of the eight amino acid substitutions are in fact variants that have been found to be rare polymorphisms to the CRS. As such these positional variants are not indicative of a conserved wild-type position, but have been included for historical purposes. Another two of the eight amino acid substitution variants are haplogroup B markers that result in favoured substitutions. The last four variants are of the most interest as they are found in very low frequency in the 20 complete mtDNA's. It is these variants that may benefit from the structural analysis conducted within the modelling software.

The Complex III subunit cytochrome b (cyt b) exhibited the most amino acid variation across all RHAS sequences interrogated. All four variants identified within the cyt b gene were substitutions; C14766T Threonine to Isoleucine, T15014C Phenylalanine to Leucine, A15326G Threonine to Alanine, and A15746G Isoleucine to Valine. Figure 3.12 A shows the 3D structural model for cyt b, amino acids implicated in substitutions are displayed. The variant A15326G is one of the 7 rare polymorphisms found in the CRS, as such the G allele at this position is common to most mtDNA's. As is expected all 20 RHAS individuals had this variant in reference to the CRS. All sequences also shared C14766T and A15746G which have been identified as haplogroup B specific markers. Also highlighted in this figure are two regions which anchor redox prosthetic groups found within cyt b. These are two *b* heme groups; heme *b_L*, and heme *b_H* (Figure 3.12 A). This area is of interest as the forth variant found within the cyt b gene, T15014C, resides between both heme groups. The variant (C) allele at this position was only identified in one individual sequence (mt16). The wild-type amino acid at this position, phenylalanine (UUC), is substituted in the T15014C variant by leucine (CUC). From interaction models within Sirius, this substitution appears to leave a somewhat more open pocket where the aromatic group of the phenylalanine would have been. Implications of this will be covered in discussion.

Within ATP6 there were 4 variants, two of these resulting in amino acid substitutions; A8860G Threonine to Alanine, and G9145A Alanine to Threonine. Again it is noted that variant A8860G is one of the 7 rare CRS polymorphisms identified by Andrews *et al.* (1999). Hence all 20 mtDNA's contained this 'variant' as would be expected. The other variant causing amino acid substitution, G9145A, was present only in one individual sequence (mt18). Alanine (GCC) to threonine (ACC). Figure 3.12 B outlines the structure of ATP6 showing positional information of variants identified.

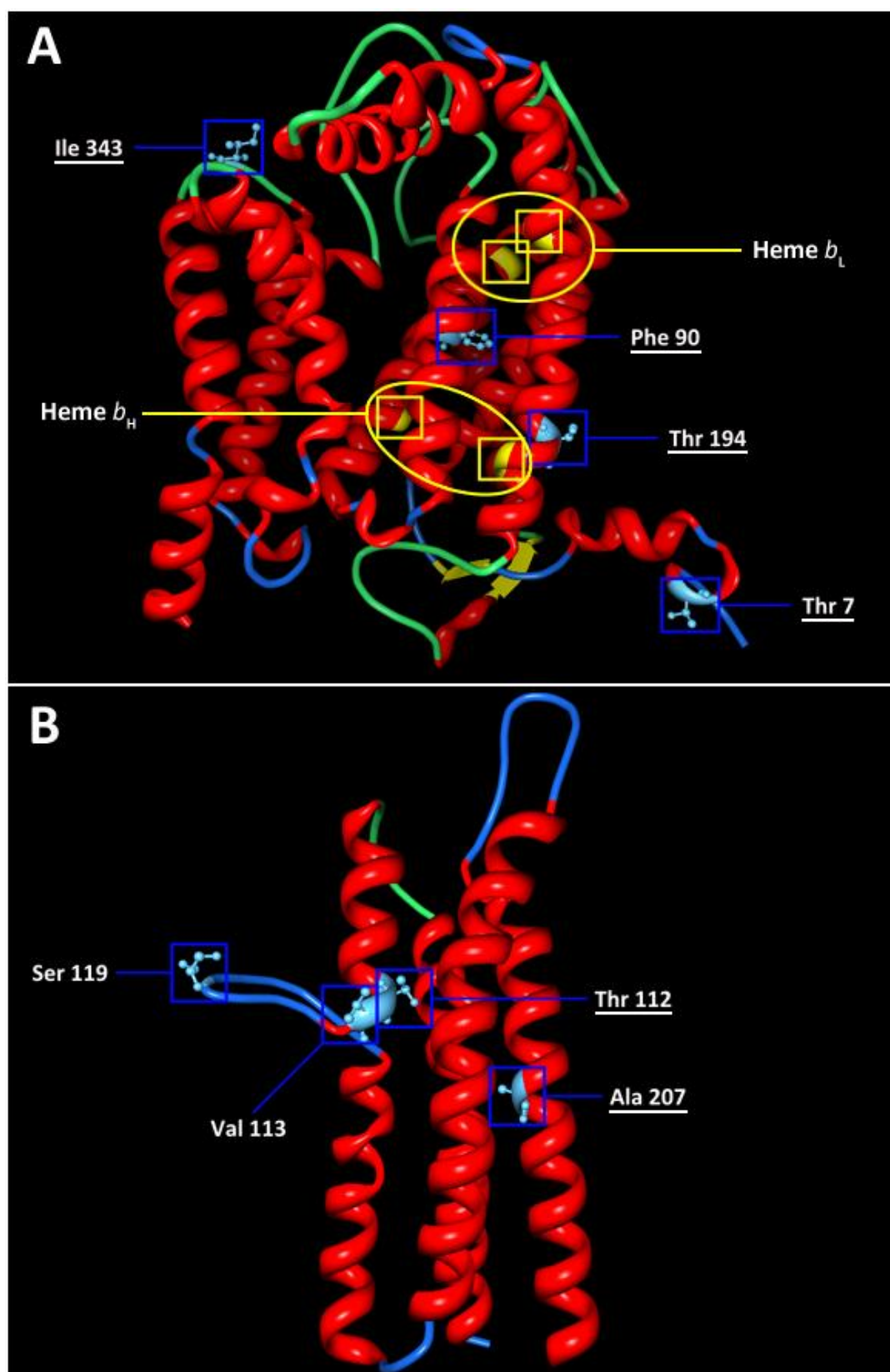


Figure 3.12: Protein modelling of human mitochondrial respiratory chain proteins.

A) Human Cytochrome b B) Human ATPase subunit 6. All positions rendered in blue are wild-type amino acids that have been modified in some or all of the 20 RHAS individuals. Amino acids are given their three letter code and the corresponding position within the protein. Underlined AA positions indicate an AA change. Protein models generated in Sirius (<http://sirius.sdsc.edu/index.php>).

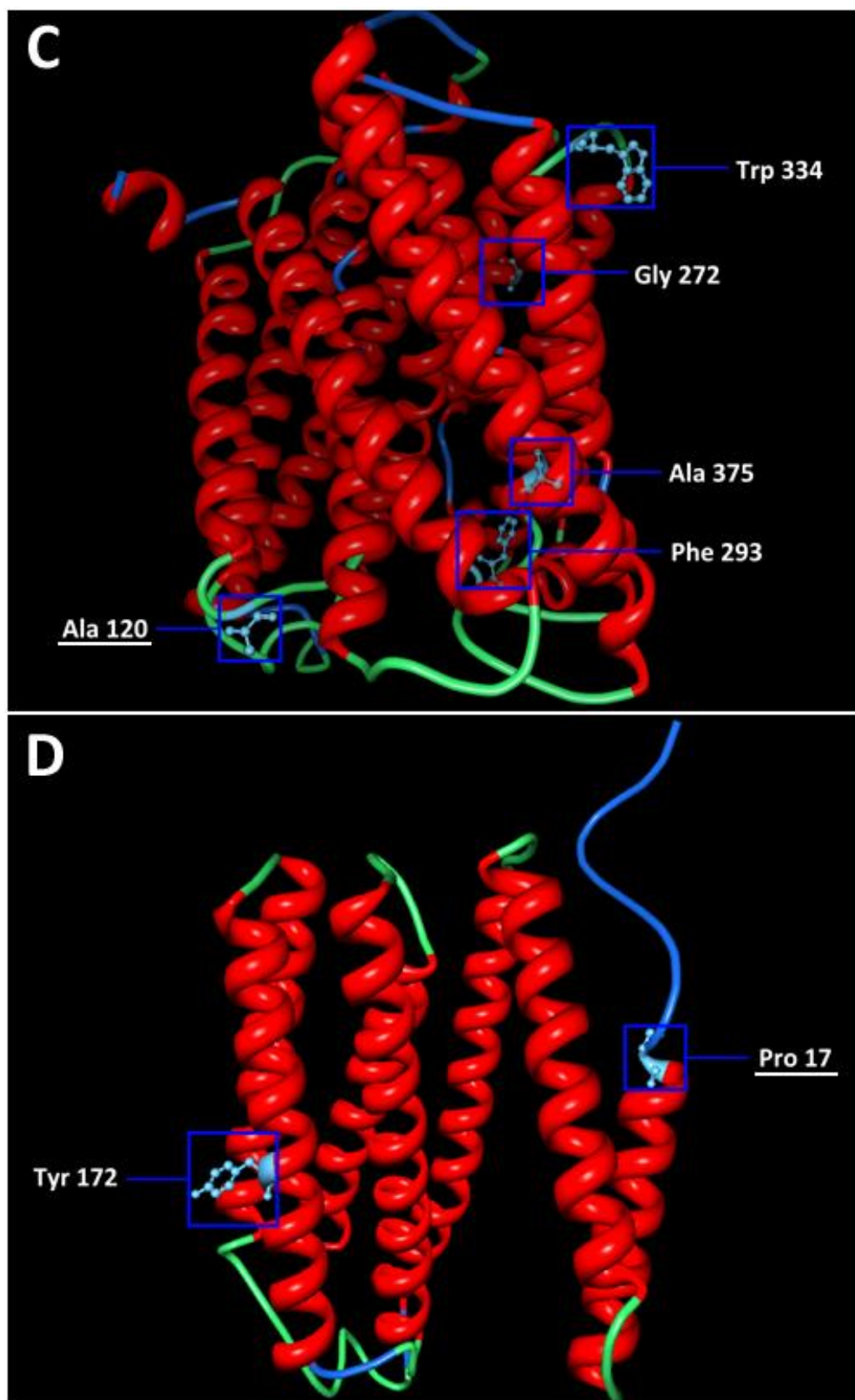


Figure 3.12: Protein modelling of human mitochondrial respiratory chain proteins.

C) Human Cytochrome c Oxidase I D) Human Cytochrome c Oxidase III. All positions rendered in blue are wild-type amino acids that have been modified in some or all of the 20 RHAS individuals. Amino acids are given their three letter code and the corresponding position within the protein. Underlined AA positions indicate an AA change. Protein models generated in Sirius (<http://sirius.sdsc.edu/index.php>).

Within COI there were 5 variants, however only one of these resulted in an amino acid substitution; G6261A Alanine to Threonine. The wild-type alanine is (GCC) is substituted for threonine (ACC). Figure 3.12 C shows the 3D structure for the modelled COI protein subunit, and highlights the areas of variation. A database search was also conducted to identify if this variant had previously been reported, and if so, to investigate its frequency in recorded sequences. This variant had been reported in 20 individual sequences, out of 6729 (as of time of writing) that were available in the database. It had also been associated with prostate cancer and LHON.

Within COIII there were two variants, with one resulting in an amino acid substitution; C9255T Proline to Serine. This was identified as an unfavoured amino acid substitution (Betts *et al.*, 2003). In the C9255T variant the wild-type proline (CCC) is substituted for serine (UCC). This variant was again only identified within one of the 20 RHAS mtDNA sequences (mt26), and as mentioned is also novel. Figure 3.12 D shows the 3D structure for the modelled COIII protein subunit, and highlights the areas of variation.

From the structural modelling analysis of these four OXPHOS proteins several variants of interest have been discovered. Predicted models have been used to identify substitutions of interest. The A15746G (Phe→Leu) variant appeared to impact on predicted protein structure the most. Also the properties of serine make the variant C9255T (Ser→Pro) interesting to discuss. As mentioned in previous sections these two variants are also novel, making them prime candidates for further study. A database search for the Ala→Thr (G6261A) substitution identified it being associated in other research with prostate cancer and LHON. Further research would need to be conducted into these variants to elucidate whether they have any implication on final protein structure or function.

OBJECTIVE 4: Validate the signature in a larger community sample of the RHAS Maori population (n=400 adults).

3.9 Genotype Results and Metabolic Trait Correlation

Selected mitochondrial SNPs were included in a Genotyping assay alongside numerous RHAS nuclear SNPs (in collaboration with ESR Ltd.). These were variant positions 1185, 16189, and 16519. Position 1185 was an obvious selection for screening in the larger RHAS group due to the noted high frequency (Section 3.5.2, Table 3.4). The variant at 16189 was selected as this position has been associated with disease in several prior studies (Poulton *et al.*, 1998; Lin *et al.*, 2005; Bhat *et al.*, 2007). It is also a marker of Polynesian mtDNA and would therefore provide an indication of the total frequency of Polynesian/Maori mtDNA within the RHAS population. As this variant forms the C-tract which causes sequencing problems it was unsure if genotyping this position would be possible. In anticipation of this position 16519 was selected, also another marker of interest already highlighted in previous work. This assay was conducted offsite by AGRF, and results were made available by email. Unfortunately position 1185 was excluded from the assay as it was found to be flanked by regions that showed high copy number within genomic DNA. As noted in Section 3.5, position 1185 appeared to be linked to another variant at position 4769. It was decided that another genotyping assay would use this variant as a possible surrogate marker for the variant at 1185.

Results from both genotyping assays were inputted into an SPSS database which contained all recorded information for participants enrolled in the RHAS project. The information was gathered in a questionnaire, and includes extensive clinical data. Table

3.9 shows all mitochondrial SNPs screened and the resulting genotypes. Final totals exclude samples that failed the genotype assay at the given position.

Table 3.9: Frequencies of three mitochondrial SNP's in large RHAS cohort.

SNP	Allele	RHAS Maori		European	
		Frequency	Percentage	Frequency	Percentage
4769	A (CRS)	92	25	1	1
	G (variant)	274	75	100	99
	<i>Total</i>	366	100	101	100
16189	T (CRS)	42	12	105	97
	C (variant)	314	88	3	3
	<i>Total</i>	356	100	107	100
16519	T (CRS)	15	4	28	26
	C (variant)	345	96	80	74
	<i>Total</i>	360	100	108	100

Note: The total number of participants enrolled in the RHAS study was 448. The totals reported above only included those that were; a) sent for genotyping, and b) passed the genotyping experiment. It should also be noted that samples were sent only from consensual participants. European frequency data was that generated from the variant frequency table in section 3.6.1 (Table 3.6).

It is interesting to note the frequencies of the 3 variants in the larger RHAS cohort group. Both T16189C and T16519C were found to be present at very high levels, 88% and 96% respectively. In this group the variants greatly outnumber the globally common mtDNA alleles, suggesting that there is minimal European mtDNA admixture within the RHAS Maori. It is also interesting to note the frequency of variant A4769G, the A allele appears to reside in 25% of RHAS Maori which in contrast to non-Maori sequences is magnitudes higher. Due to the 16519T frequency being so low (4.2%) it was not possible to use it in further statistical analysis. This left two variants, T16189C and A4769G, for disease association and hypothesis generating analysis.

OBJECTIVE 5: Correlate the validated mtDNA signature with metabolic traits (e.g.BMI and Diabetes).

3.9.1 16189 and Diabetes

As the information entered into the SPSS database for several traits was categorical, a t-test of means was not always the appropriate analysis. This was the case for the diabetes trait. In the database each individual had an entry under the variable diabetes, either; 0 for non-diabetic, or 1 for diabetic. To explore variant association a chi squared test with risk analysis was conducted using the mtDNA SNP 16189 (Table 3.10).

Table 3.10: Association statistics of mitochondrial variant 16189 and diabetes.

	16189		Total
	T	C	
non-diabetic	32 (15%)	183 (85%)	215 (100%)
diabetic	1 (4%)	24 (96%)	25 (100%)
Total	33 (14%)	207 (86%)	240 (100%)

Note: Total number of cases investigated has been adjusted to; a) adults, those aged 18 years or older, and b) adult cases that have available data for both diabetes and genotyping values.

The chi squared test statistic was $\chi^2 = 2.20$, d.f. = 1, with a p-value of 0.138. As one of the cells in the chi squared analysis fell below 5% (Table 3.10), an alternative test statistic was adopted. This was the Fisher's Exact test (2 sided) which gave a p-value of 0.217. This test again was non-significant so no further follow up statistics were conducted.

3.9.2 4769 and diabetes

As the information entered into the SPSS database for several traits was categorical, a t-test of means was not always the appropriate analysis. This was the case for the diabetes trait. In the database each individual had an entry under the variable diabetes, either; 0 for non-diabetic, or 1 for diabetic. To explore variant association a chi squared test with risk analysis was conducted using the mtDNA SNP 4769 (Table 3.11).

Table 3.11: Association statistics of mitochondrial variant 4769 and diabetes.

	4769		Total
	G	A	
non-diabetic	167 (77%)	51 (23%)	218 (100%)
diabetic	16 (62%)	10 (38%)	26 (100%)
Total	183 (75%)	61 (25%)	244 (100%)

Note: Total number of cases investigated has been adjusted to; a) adults, those aged 18 years or older, and b) adult cases that have available data for both diabetes and genotyping values.

The chi squared test statistic was $\chi^2 = 2.813$, d.f. = 1, with a p-value of 0.094. This test was non-significant; however there appeared to be a trend which was explored with risk analysis. This gave an odds ratio of 2.05, which suggests that individuals with the G allele are 2 times less likely to be diabetic than those who have the A allele.

3.9.3 Hypothesis Generating

In addition to the specific hypotheses tested above a set of exploratory (or hypothesis generating) tests were run to further analyse mtDNA markers and trait correlation in the larger RHAS Maori database. Both genotyped SNPs (16189 and 4769) were tested for significance against an extensive list of metabolic markers. These included; BMI, WHR, age, disease associations, blood serum markers (vitamin, lipids, glucose, etc.), to name a few.

3.9.3.1 Association Testing 16189 and Metabolic Traits

A list of metabolic traits was generated within the SPSS database. Possible correlations between metabolic traits and mitochondrial variant T16189C were tested using a Spearman's two-tailed bivariate test for correlation. Results from this test were shown as a hypothesis generating graph by plotting each traits p-value against its correlation statistic (Figure 3.13). A cut-off for significance was set at the 0.05 a conservative level was chosen to ensure that no relationships of interest were overlooked.

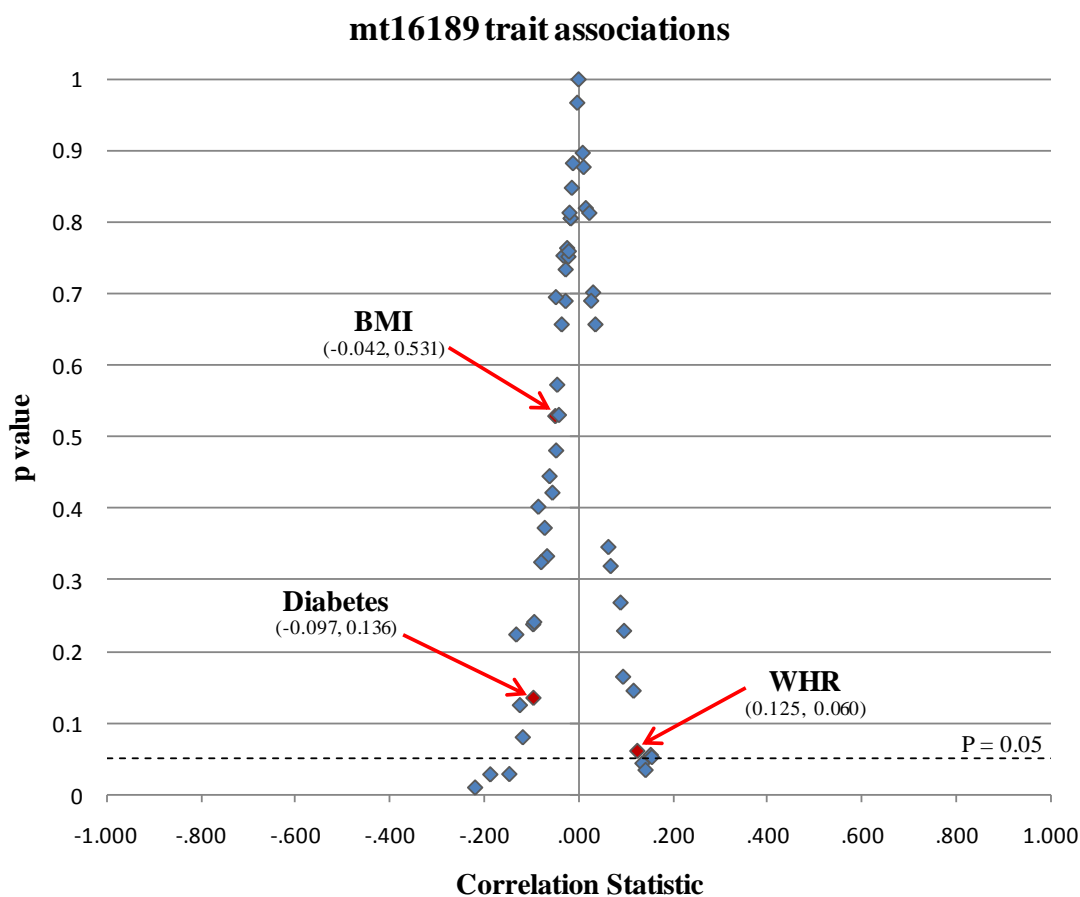


Figure 3.13: Hypothesis generating graph showing metabolic associated traits and their statistical correlation to mitochondrial variant T16189C. Traits below the 0.05 level of significance were selected for further statistical investigation.

An independent samples student t-test was conducted on metabolic traits which had a significant correlation ($p=0.05$) with the T16189C variant (Table 3.12). It should be noted that several of the traits showing significant correlation were nominal in nature and were unable to be analysed using the student t-test. These traits were analysed separately using appropriate tests. BMI and WHR were also included in the t-test as traits of interest to investigate whether any associated trends may be worth further investigation.

Table 3.12: t-test statistics for specific clinical traits

Clinical Trait	wt (T)	16189 variant (C)	t score	df	p value
B12 (pmol/L)	367.774	442.370	3.016	58.167	0.004**
CREA ($\mu\text{mol/L}$)	84.613	80.141	-1.310	50.779	0.196
ICIN	13.097	12.047	-1.443	54.704	0.155
AUDIT	5.333	7.596	1.744	33.250	0.090
BMI (kg/m^2)	30.858	32.841	1.455	80.298	0.150
mean_dias (mmHg)	77.214	81.927	2.151	40.485	0.037*
WHR	0.930	0.929	-0.037	123.335	0.970

For all traits equal variances are not assumed. ** indicates significance at the 0.005 level. * indicates significance at the 0.05 level. B12 (vitamin B12), CREA (creatinine), ICIN (bilirubin levels), AUDIT (dangerous consumption of alcohol), BMI (body mass index), mean_dias (mean diastolic blood pressure), WHR (waist hip ratio).

The t-test results (Table 3.12) indicated two of the metabolic traits were deemed to have statistically significant difference of means; B12 ($p=0.004$) and mean_dia ($p=0.037$). Blood serum vitamin B12 levels appear to be significantly higher in those RHAS individuals who carry the T16189C variant. A similar trend was noticed in mean diastolic blood pressure, again significantly higher in variant groups. Another interesting trend identified was that of the decreased creatinine levels in T16189C

variant individuals. While this trend was not significant, the noted difference in means and creatinine's possible role in early identification of T2DM make it a prime candidate for follow up research.

3.9.3.2 Association Testing 4769 and Metabolic Traits

Possible correlations between metabolic traits and mitochondrial variant A4769 were tested using a Spearman's two-tailed bivariate test for correlation. Results from this test were shown as a hypothesis generating graph by plotting each traits p-value against its correlation statistic (Figure 3.14). A cut-off for significance was set at the 0.05 level a conservative level was chosen to ensure that no relationships of interest were overlooked.

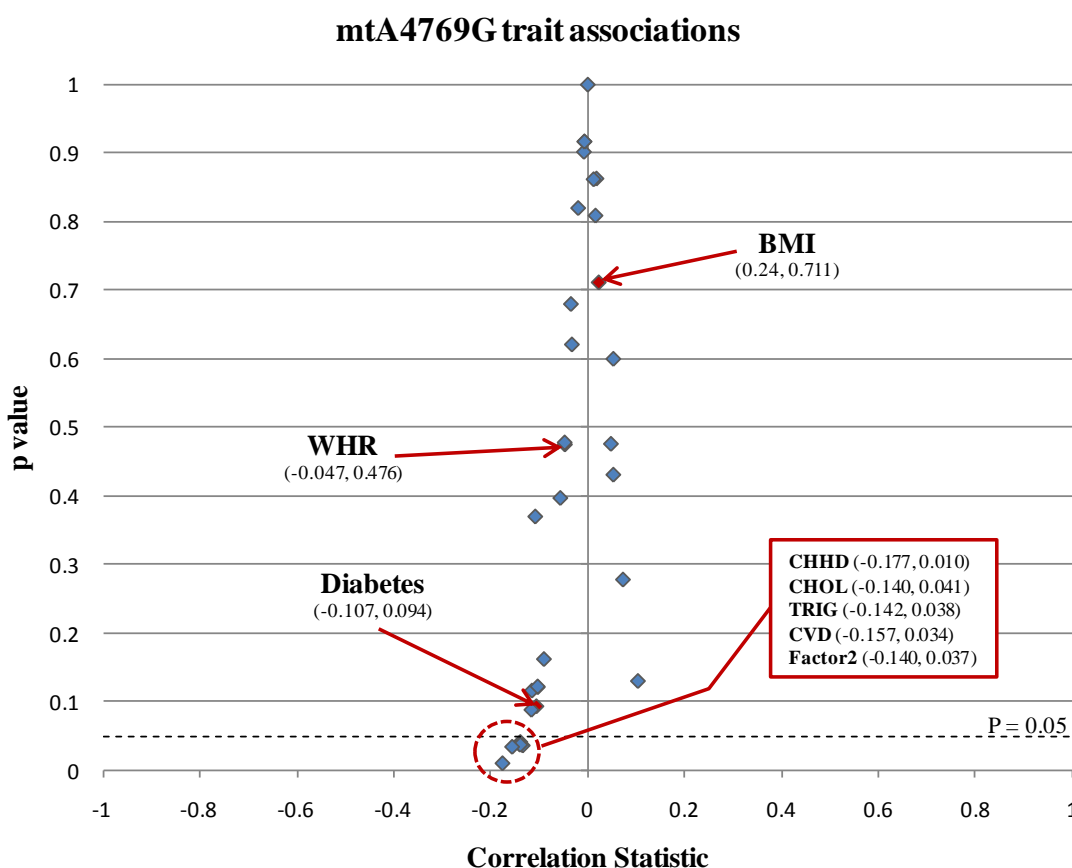


Figure 3.14: Hypothesis generating graph showing metabolic associated traits and their statistical correlation to mitochondrial variant A4769G. Traits below the 0.05 level of significance were selected for further statistical investigation.

It was noted that all significant correlated metabolic traits were related to blood lipids and cardiovascular disease. An independent samples student t-test was conducted on metabolic traits which had a significant correlation ($p=0.05$) with the A4769G variant (Table 3.13).

Table 3.13: t-test statistics for specific clinical traits

Clinical Trait	wt (A)	⁴⁷⁶⁹ variant (G)	t score	df	p value
CHHD	4.339	3.947	2.311	109.497	0.023*
CHOL (mmol/L)	5.219	4.944	1.826	105.939	0.071
TRIG (mmol/L)	2.280	1.837	2.113	80.851	0.038*
CVD	6.895	5.191	1.514	71.801	0.134
Factor2	0.507	0.218	2.076	101.406	0.040*

For all traits equal variances are not assumed. * indicates significance at the 0.05 level. CHHD (composite lipid profile statistic), CHOL (cholesterol), TRIG (triglycerides), CVD (risk of cardio vascular disease), Factor2 (risk factor for CVD).

The t-test results (Table 3.13) indicated three of the metabolic traits were deemed have statistically significant difference of means; CHHD ($p=0.023$), TRIG ($p=0.038$) and Factor2 ($p=0.040$). All three traits are related to blood lipid and triglyceride levels. Factor2 is a composition of several traits which creates risk factor statistic for cardiovascular disease. It is therefore interesting to note that the Factor2 mean for the A allele is more than double that of the G allele. These trends pave the way for future investigation and form a group of prime candidate markers for follow up research.

3.10 Summary of Results

RFLP and HVRI sequencing was used to generate a Mitochip selection group RHAS Maori (n=94). Analysis of these sequences with other Maori HVRI data identified increased mtDNA variation. This increased variation has led to the discovery of larger Maori haplotype diversity than was once thought, including 3 unique to RHAS Maori.

Using a novel high-throughput resequencing assay this study has sequenced the first complete Maori mitochondrial genomes (n=20). It has also documented 7 novel variants, which will be added to the Mitomap database. When compared to independent worldwide sequence datasets it was determined that RHAS Maori share the same haplogroup markers as Polynesian mtDNA; with both belonging to haplogroup B. It has been identified that RHAS Maori exhibit high frequencies of three variants; 1185>T, 4769>A, and 16126>C. These variants reside in very low frequencies worldwide (>1%), and form an RHAS Maori specific mtDNA signature.

Four variants discovered within RHAS Maori mtDNA sequences have previously been reported to have association with disease. These were; 4769, 6261, 16189, and 16519. Two of these variants (4769 and 16189) were genotyped in the wider RHAS Maori community (n=400). Frequencies of these variants in the community were; T16189C (88%) and G4769A (25%). When correlated against metabolic traits it was found that RHAS Maori with T16189C showed elevated vitamin B12 levels, as well as increased mean diastolic blood pressure. The variant G4769A was found to be associated with a group of cardiovascular disease risk factors (CHHD, TRIG, and Factor2). An association between G4769A and diabetes was also explored, with indications that individuals with the G variant were 4 times less likely to develop diabetes.

Chapter 4 Discussion

4.1 HVRI Analysis of 94 RHAS Maori

The control region of the human mitochondrial genome (mostly HVRI) is a focus in many scientific fields (Poulton *et al.*, 1998; Stoneking 2000; Lambert *et al.*, 2002). It has proved a great success in tracing mtDNA migration and structuring populations (Vigilant *et al.*, 1991; Lum *et al.*, 1994; Hagelberg 1997), as well as being a powerful forensic identifier (Vallone *et al.*, 2007). More recently specific mtDNA variants within the control region have been highlighted as having associations with a variety of disorders (Kim *et al.*, 2002; Navaglia *et al.*, 2006; Bhat *et al.*, 2007). Several of these disease studies involve variants within Maori HVRI sequences (Poulton *et al.*, 1998; Poulton *et al.*, 2002).

In this study the Hyper Variable Segment I (HVRI) of the mtDNA control region was analysed in two ways. The first involved a novel restriction assay, which analysed a specific Polynesian variant marker within PCR amplified HVRI products. The second method employed was sequencing of 94 HVRI from New Zealand Maori individuals from the Rakaipaaka iwi. Twenty one different haplotypes were identified, defined by 22 variable positions, or SNPs. The majority of these variants were transitions (minimal transversions were detected), however, there were several areas of visible heteroplasmy, and two sequences contained an insertion at 16295.

4.1.1 RFLP assay

Restriction Fragment Length Polymorphism (RFLP) assays have been proven to be useful for quick identification of specific sites within segments of DNA (Introduction, Section 1.3.6.1). These assays have been well established in mtDNA research as a means of quickly identifying specific haplotype variants (Francalacci *et al.*, 1999). This study identified a novel restriction site that resides within a large proportion of Polynesian mtDNA types. The assay used the restriction enzyme *Pst* I to screen for the Polynesian specific haplotype B4a1a1a (Results, Section 3.3). This haplotype is defined by the transition variant 16247>G, which is also the position that determines the *Pst* I restriction site. New Zealand Maori have been identified to share this Polynesian mtDNA haplotype (Sykes *et al.*, 1995; Murray-McIntosh *et al.*, 1998; Whyte *et al.*, 2005). As such it was hypothesised that this test would provide a means to quickly and cheaply screen for Maori mtDNA sequences and would detect the presence of European mtDNA admixture.

The RFLP assay conducted showed 77% of the RHAS Maori HVRI PCR products were digested. It was originally suspected that the 33% undigested HVRI products were European in nature. This was proven wrong when sequencing determined that one of these sequences had a variant within the restriction site, meaning it did not undergo restriction. While the other sequences were found to be Polynesian in origin, they lacked the 16247>G variant, having the previously identified 'CCAT' motif (Trejaut *et al.*, 2005). This was the first documented use of this RFLP assay. It has been proven to be specific for haplogroup B4a1a1a. It was useful as a quick screen of ancestry in RHAS Maori, and identified the B4a1a1a haplotype in 77% of the HVRI products tested. Once this thesis has been submitted this RFLP assay will be entered into the consortium located in the Mitomap database (Ruiz-Pesini 2007).

4.1.2 HVRI sequencing and 16189

Sequencing of PCR products was used to generate a database of 94 HVRI RHAS Maori sequences. All but one sequence was identified as belonging to haplogroup B (Results, Section 3.4.2). One of the variants that define haplogroup B is a T to C transition at position 16189 in the HVRI of the D-loop. This transition creates a poly-C tract, which has been shown to exhibit length variation (Poulton *et al.*, 1998). This variant is within the control region of the mitochondrial genome, while it is non-coding, it does contain many of the controlling elements of transcription and translation. Previous research had demonstrated that due to the nature of variants within the Polynesian (B4a1a1a) mtDNA haplotype, sequencing through the HVRI region becomes troublesome (Cann *et al.*, 1987; Redd *et al.*, 1995; Whyte *et al.*, 2005). The problem involves the C→T transition at position 16189 (Figure 4.1), and two A→C transversions at 16182 and 16183. The transition, along with both transversions, generates a homopolymeric C-tract which causes polymerase instability during sequencing, leading to blurred results after this point (Figure 4.1 B).

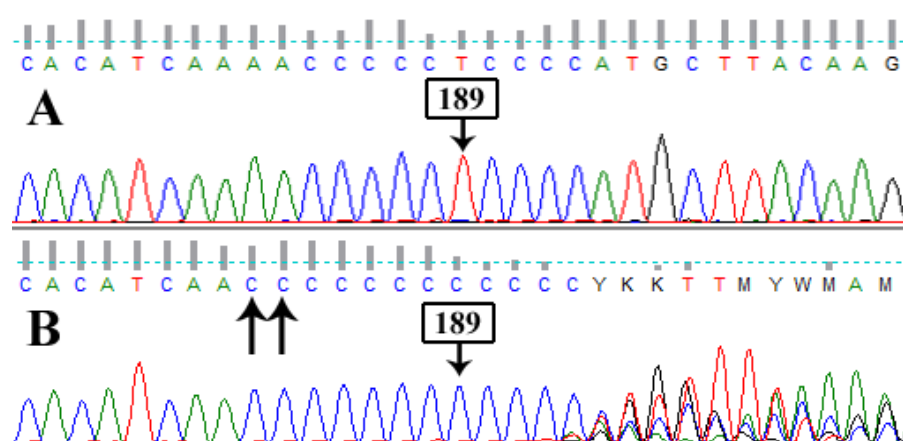


Figure 4.1: Chromatographs showing; A) a Taiwanese Aboriginal HVRI sequence (AMI011) with a T at position 16189 (indicated above), B) a Maori HVRI sequence (mt3) showing the T to C transition at position 16189, as well as the two A to C transversions (indicated by arrows). These two events create a homopolymeric cytosine tract in the mitochondrial HVRI, about 200 bp into the primer selected sequence.

As noted in previous work (Lum *et al.*, 1994; Redd *et al.*, 1995; Whyte *et al.*, 2005) the additional C-tract length variation caused technical difficulty for one directional sequencing of the HVRI. To overcome this, both forward and reverse primers were used to obtain the full HVRI sequence. After close examination of obtained forward and reverse sequences, as well as other available Polynesian sequence data, the majority of the RHAS HVRI samples (again all but one) were found to have a C-tract length of 12 C's. This was later confirmed by Mitochip analysis (Section 3.5). Mitochip analysis also revealed the presence of the 9-bp deletion in the 20 RHAS Maori sequences.

4.2 Variation within 20 complete RHAS Maori mt genomes

Mitochip resequencing analysis of 20 NZ Rakaipaaka Maori mitochondrial genomes resulted in 14 different haplotypes, defined by 44 variants (SNPs) of the revised Cambridge Reference Sequence. The majority of these variants were transitions, only two transversions were identified. As an internal control several of the HVRI mtDNA's were chosen as they contained novel variants that could be used to test the Mitochip array. One of the control sequences contained an insertion at 16295. The Mitochip resequencing assay confirmed this insertion of A at base 16295, as was noted in HVRI sequencing results (Results, Section 3.4.2). This indicated that the Mitochip array was accurate and working correctly. As all 20 sequences belonged to haplogroup B, they were hypothesised to contain the 9 bp deletion (Introduction, Section 1.3.6.3). The array detected the 9 bp deletion in all 20 RHAS Maori. This detection is traditionally analysed using separate tests (either RFLP, directed PCR and sequencing, or a combination of both), yet the Mitochip was able to detect its presence. This could be validated by following the procedure laid out by Whyte *et al.*, (2005).

4.2.1 Identification of a RHAS Maori specific mtDNA signature

One of the specific aims of this project was to identify whether RHAS Maori exhibit a specific mtDNA signature. It was hypothesised that if a RHAS Maori signature was present it could possibly be used to help explain the increased frequencies of metabolic disorders such as diabetes and obesity within the community. If any RHAS Maori specific variants were found to associate with metabolic traits they could ultimately become part of an early diagnostic screen. This would allow for the development of early intervention strategies.

Three variants were identified as belonging to RHAS Maori, and were found on the background haplotype of Polynesian mtDNA. The three RHAS Maori specific signature variants are; 1185, 4769, and 16126. With the finding of this signature this project has identified that RHAS Maori do indeed exhibit a unique mtDNA background when compared to Polynesians. Implications of this will be discussed further through this section.

4.2.2 Identification of seven novel mitochondrial variants

Of the 44 variants identified from the 20 RHAS Maori mtDNA sequences, 16% (n=7) were found to be undocumented. As such they were termed novel (Results, Section 3.5.5, Table 3.5). As they were only present in very low frequencies, none of these variants were considered to be part of the RHAS Maori mtDNA signature. They are however still worth documenting, as future studies could possibly screen for these in other Maori mtDNA. These novel variants will be entered into the Mitomap database upon submission of this thesis.

4.2.3 *Strengths and Limitations of the Affymetrix Mitochip*

The Affymetrix Mitochip array was a relatively novel technique when it was decided upon for the mtDNA screening in this project, and had not previously been employed within NZ. However over the last 2-3 years it has seen use in many different research projects, including forensics (Vallone *et al.*, 2007), early breast cancer detection (Jakupciak *et al.*, 2008), and disease correlation studies (Rollins *et al.*, 2009). The Mitochip assay offers complete resequencing of mtDNA with additional probes for variant identification and confirmation. It also requires a very small amount of DNA to be effective, having been shown to work with DNA amounts ~10-50 ng.

The assay does however have limitations; it's relatively expensive (~NZ\$700 per chip), chips have to be analysed on Affymetrix GeneChip machinery. Additionally due to an oversight in the design there are 12bps missing from each end of the mitochondrial genome, and due the GC rich nature of mtDNA there are many areas that consist of solid cytosine blocks (runs 4 or more) which interfere slightly with probe hybridisation and ultimately base call data surrounding the area.

There were also several quirks to the Mitochip protocol that became clear during the analysis phase. The first of these was that 12bp of sequence was noted to be missing from the beginning and end of the mtDNA sequences. Communications with Affymetrix confirmed this as a known 'oversight' in the design of the Mitochip, specifically the design of the probes. This means that all mtDNA sequences generated using the Mitochip protocol were missing 24bp. However this region of the D-loop contains very little variation and no key markers so for the purpose of this work these sequences will be referred to as complete.

4.3 Phylogenetic Analysis

Genotyping data for three SNP's was generated primarily for disease association testing within the wider RHAS Maori community (Results, Section 3.9.) and is discussed further in this chapter. However this frequency information also gives a good indication about the mtDNA haplotype composition of RHAS Maori mtDNA. The three SNP's were found in varying frequencies in RHAS Maori; G4769A = 25%, T16189C = 88%, and T16519C = 96%. The presence of the T16189C variant is a good indicator of Polynesian, and subsequently, Maori mtDNA ancestry as it is part of the Polynesian 'CCGT' motif. It was seen in the genotype screen for this SNP that 88% of the enrolled RHAS individuals had the C variant. This left 12% with mtDNA that is suspected to be European-like. The variant G4769A has been identified as a possible RHAS Maori specific variant, and was found to be present in the wider RHAS community (n~400) in proportions similar to what was identified in the Mitochip analysis. Originally discovered in 7 of the 20 mtDNA's that were resequenced via Mitochip, larger scale genotyping has revealed this position to be present in 25% of the RHAS Maori. It should be noted that the G4769A SNP is only present on a Polynesian/Maori mtDNA background, meaning that markers such as the 9-bp deletion and the Polynesian motif have to be present for this SNP to be found in Maori. This indicates that it is most likely to have originated in RHAS Maori rather than somewhere in Polynesia. It should also be noted that at this stage the G4769A SNP has only be validated in RHAS Maori. Further investigation in wider Maori communities would determine if this variant is RHAS Maori specific or indeed a marker found in other Maori. As yet it has not been observed in other Polynesian populations; a database search discovered no sequences containing both the Polynesian 'CCGT' motif and G4769A. This test was also run for variants C1185T and T16126C (other RHAS

specific signature markers) with the same results, no sequences in the available databases contained the specific variants. This leads towards the possibility of the discovery of the first specific Maori haplotype defining variants.

The C variant at 16519 was identified in all 20 RHAS Maori mtDNA sequences. It was originally thought to be another ancestral Polynesian mtDNA marker. As such it was genotyped and was to be employed as a secondary screen in case genotyping of the T16189C variant was not possible. It was thought that the difficulties encountered when sequencing mtDNA with the T16189C variant might have made genotyping the SNP unlikely. This however was not the case.

When examined in the RHAS Maori 16519>C was present in 96% of the community. This was considered odd as it was expected that the variant at T16519C would be present at the same frequency as T16189C, if they were both markers of Polynesian mtDNA ancestry. This appears to be the result of a unique mutational relationship with the 16189 C variant. When the homopolymeric C-tract generating transition (16189>C) is present it is hypothesised that position 16519 becomes hyper-mutable, that is it is more likely to undergo a transition mutation (Navaglia *et al.*, 2006). The T16189C variant does appear randomly in various geographical mitochondrial haplotypes, but has become a mainstay in haplogroup B (van Oven *et al.*, 2009). So it has been noticed that sequences exhibiting the T16189C variant are more likely to have the T16519C variant. This was also demonstrated in this study when the 16519 SNP was genotyped in the wider RHAS Maori community. The T16519C variant was found to be present in 96% of RHAS Maori (Results, Section 3.9). Therefore it looks to be moving towards fixation in Maori mtDNA.

4.3.1 Using complete mtDNA to explore RHAS Maori ancestry

This is the first research to completely sequence the mitochondrial genome of New Zealand Maori individuals. Three variants have been identified that appear to have arisen in Amis Taiwanese Aborigines: T6719C, C12239T, and A15746G. These 3 variants are present in mitochondrial sequences from Micronesia [Kapingamarangi Atoll, Majuro Atoll (Pierson 2006)], Melanesia [Espiritu Santo, Maewo, Trobriand Islands (Pierson 2006), Bougainville (Macaulay V. 2005), PNG coast (Ingman 2003)], Polynesia [Tonga (Pierson 2006), Cook Islands, Samoa (Ingman 2000; Ingman 2003)], and now NZ Maori (this study). Another variant has been identified that is believed to have occurred either in or around coastal Melanesia (PNG) or in Oceania. This is A14022G, which has been identified in peoples from Micronesia, Melanesia (coastal PNG), and Polynesia (Trejaut *et al.*, 2005). The A16247G variant of the Polynesian motif is not found in Amis or Taiwan, but seen in high frequencies in Melanesia and Polynesia (Trejaut *et al.*, 2005).

Early work has suggested the region around Taiwan as the staging post for Polynesian migration, this has been based on linguistics (Bellwood 1991; Gray *et al.*, 2009), archaeological (Kirch *et al.*, 1988), and DNA in the form of mitochondrial control region sequences (Lum *et al.*, 1994; Sykes *et al.*, 1995; Cann 2004). More recently this evidence has been supported by complete mitochondrial genome sequence data (Ingman 2001; Trejaut *et al.*, 2005). The RHAS Maori mtDNA from this study seem to provide support to this theory as well. There are specific markers inside of the mitochondrial coding region that look to have arisen in Taiwan (Results, Section 3.6.1 Table 3.6). It can be seen that the variants T6719C, T10238C, C12239T, and A15746G, appear first in Taiwan and grow in frequency through Melanesia and into Polynesia. These Maori mtDNA therefore add weight to the ongoing argument of Polynesian migration.

4.3.2 Haplotype Analysis

Of the 94 HVRI sequences 93 were reported as belonging to haplogroup B; 75 were identified as having the ‘CCGT’ motif (haplotype B4a1a1a), while 18 were found to have a previously reported pre-Polynesian ‘CCAT’ motif (haplotype B4a1a1). This ratio of ‘CCGT’/‘CCAT’ within the Rakaipaaka Maori is an interesting one, suggesting possible parallel migration of separate mtDNA lineages. It could also indicate a reversion mutation at position 16247. Both motifs have been identified previously within Maori populations (Murray-McIntosh *et al.*, 1998; Whyte *et al.*, 2005).

One individual sequence lacked any Polynesian mtDNA markers and was determined to belong to the European haplogroup H (Helena). Haplogroup H is the haplogroup that the CRS belongs to. This was an intriguing discovery as all RHAS Maori individuals were chosen based upon an autosomal ancestry score (Hall 2009). This meant that all 94 individuals exhibited high percentage Maori ancestry, from which it was hypothesised that they would also therefore show Polynesian mtDNA ancestry. Mitochondrial HVRI sequencing confirmed this hypothesis, with the exception of this one individual. This phenomenon could be explained by admixture, all it would take is one European female in a lineage to pass on her mtDNA and then it would move down through all her female descendants.

4.3.2.1 RHAS Maori mtDNA Exhibits Increased Variation

This study has built on the previous research of Murray-McIntosh *et al.*, (1998) and Whyte *et al.*, (2005). By combining HVRI datasets from these studies with the RHAS Maori HVRI dataset it has been discovered that the Maori mitochondrial genome exhibits more variation than once believed (Results, Section 3.6.2). The previous estimate of Maori mtDNA haplotypes was 7 (Whyte *et al.*, 2005) based on 109 HVRI

sequences. In the same study analysis of East Polynesian HVRI sequences revealed 24 haplotypes. This suggested that mtDNA variation decreased as Austronesian speaking Polynesians migrated through Oceania (Hertzberg 1989). As Aotearoa was the last place to be inhabited, it was concluded that there is little variation within Maori mtDNA. It is thought that this was due to genetic bottlenecks created by founder effects (Lum 2000; Whyte *et al.*, 2005). Due to the then apparent low haplotype diversity it was speculated that Maori would show no specific mtDNA haplotype, one that would differentiate Maori mtDNA from the Polynesian mtDNA haplotype B4a1a1a (Trejaut *et al.*, 2005).

Combining the RHAS Maori HVRI dataset with that of Murray-McIntosh *et al.*, (1998) and Whyte *et al.*, (2005) greatly expanded the number of identified Maori HVRI haplotypes from 7 to 17 (Results, Section 3.6.2). It was noted that several Maori sequences in Whyte *et al.*, (2005) contained HVRI variation outside of the region explored by Murray-McIntosh *et al.*, (1998). When the wider HVRI sequences of Whyte *et al.*, (2005) was combined with RHAS Maori HVRI data the extra variation increased the number of mtDNA haplotypes from 17 to 25. This analysis reveals that there is indeed more variation within the Maori mtDNA HVRI than was once thought.

There are several possible explanations for the increased mtDNA diversity identified. The first explanation is that there were more Maori founders that arrived in Aotearoa than have been previously estimated. Whyte *et al.*, (2005) went on to use Maori haplotype number to estimate, via population simulation, the number of founding females that would have been required to have arrived in Aotearoa to form the present day Maori population. The final number of founding females was estimated to be ~190

(Whyte *et al.*, 2005). Based on the extended haplotype diversity discovered in this study, it is postulated that this number could now be considered an underestimate, and in fact the number of founding females is actually much higher. Future study could explore re-running the computer simulations to estimate the number of female founders based on the new HVRI diversity.

Another explanation for the noticed variation in RHAS Maori HVRI sequences is that it is possible that the human mtDNA is acquiring mutations at a much faster rate than previously expected (Howell *et al.*, 2007). When calculating mitochondrial mutational rates and applying them to current molecular clocks, it has been identified that the mtDNA of penguins actually mutates at 2-7 times the rate that was currently accepted (Subramanian *et al.*, 2009). It is possible that this is similar in human mtDNA mutation rates, and the evolutionary genetics field can only benefit from accruing more complete mtDNA sequences to examine this trait further.

4.3.2.2 Identification of Maori Specific mtDNA haplotypes

In Results Section 3.5.2, it was identified that within the 20 RHAS mtDNA sequences there were 44 variants when compared to the CRS. Of these it was determined that the bulk of the common RHAS variants were the same as those found in Polynesian mtDNA. However it was identified that several variants appeared in higher frequency in some RHAS Maori sequences, suggesting the presence of extended mtDNA haplotypes. Further analysis of the 20 complete Maori mtDNA sequences has revealed that at least 3 unique haplotypes reside within the RHAS Maori community. These novel haplotypes are defined by the RHAS specific variants C1185T, G4769A, T16126C, and A162476G and are documented in Table 4.1.

Table 4.1: RHAS Maori mtDNA haplotype markers

Haplotype	<i>f</i>	<u>1185</u>	<u>4769</u>	<u>16126</u>	16189	16217	16247	16261
RM1	7	T	A	T	C	C	G	T
RM2	4	C	G	C	C	C	G	T
RM3	6	C	G	C	C	C	A	T
RM4*	3	C	G	T	C	C	G	T

*Previously reported Polynesian mtDNA haplotype B4a1a1a (Trejaut *et al.*, 2005; van Oven *et al.*, 2009). Positions underlined are RHAS specific haplotype markers. Bolded positions represent Polynesian mtDNA motif ‘CCGT’. Shaded positions indicate variants to the CRS, usually positions in common to CRS are represented as dots, however it was chosen to represent these as the respective base to portray haplotype motif’s.

Table 4.1 details the variant positions that define 4 RHAS Maori haplotypes within the well document B4a1a1 haplotype (Trejaut *et al.*, 2005). While there are singleton variants (noticed in only one sequence) that further expand these haplotypes in many of the sequences, it is these RHAS signature variant’s that define the major haplotype motifs. The 4 haplotype motifs are: RM1 (TATCCGT), RM2 (CGCCCGT), RM3 (CGCCCAT), and RM4 (CGTCCGT). RM1, RM2, and RM3 are novel haplotypes that have not been previously documented. This was ascertained by searching for motif positions in several databases (mtDB and Mitomap) as well as Phylotree (van Oven *et al.*, 2009). RM4 is the haplotype more traditionally characterised by the Polynesian ‘CCGT’ motif, as can be seen in Table 4.1. It is interesting to note that out of the 20 sequences, only 15% show the traditional Polynesian motif. Haplotype RM1 displays the highest frequency at 35%. This haplotype is formed by two of the three markers defined as RHAS Maori specific (C1185T and G4769A). It appears that these variants have most likely occurred either in, or on the way to Aotearoa, or more likely, relatively recently within the Rakaipaaka iwi. Evidence for this is the fact that these two variants exist on top of the ‘CCGT’ Polynesian motif. This ‘CCGT’ motif is derived from the

HVRI sequence only, and has been used to classify Maori haplotypes until now (Murray-McIntosh *et al.*, 1998; Whyte *et al.*, 2005). It is highly possible that sequences generated in these studies show the same or similar variant positions that are important for haplotype analysis. Further research into Maori mtDNA haplotypes would benefit from typing these specific variant sequences, as it is obvious that there are important areas of haplotype variation outside of the HVRI.

The other two RHAS Maori specific haplotypes (RM2 and RM3) both have the RHAS specific marker T16126C. These haplotypes are interesting as they are actually defined by the variant A16247G, which is the final evolutionary marker of the Polynesian motif. Haplotype RM2 has the 16247>G transition present, where as RM3 has the more common A allele. It has been generally thought that the Polynesian motif was the dominant motif in Oceania (Polynesians and Maori), but from this study it appears that the ‘CCAT’ motif is present in high frequencies than documented (Whyte *et al.*, 2005). There are several possibilities for this. The first possibility being, two mtDNA populations have co-existed during the migration and population of Oceania. The ‘CCAT’ motif is found in very high frequencies around Taiwan/South East Asia (Trejaut *et al.*, 2005; Soares *et al.*, 2008). As the Austronesian speaking peoples migrated down and out into the Pacific some of their mtDNA acquired the 16247>G transition thus the ‘CCGT’ motif, while others retained the ‘CCAT’ motif. They have since resided within Polynesian populations, with the ‘CCGT’ motif reaching much higher frequencies. Another possibility is that position 16247 is hypermutable, and it’s possible that there has been a reversion mutation at 16247. After the G allele has been acquired along the Polynesian migration, it has reverted back to A at some point. It’s hard to say whether the T16126C variant occurred twice, once on each motif, or whether it was present first and then there was a reversion at 16247 bringing about the

A allele again. Further research trying to identify these variants in other Polynesian and Maori mtDNA is required.

The substructure of haplogroup B is extrapolated out in Figure 4.2 to indicate where on the phylogenetic tree these mtDNA sequences would be placed. It is yet to be seen if these RHAS Maori haplotypes exist in Polynesia, and only research generating more complete mtDNA data will provide the answers to this. If these haplotype motifs were found to be in the wider Oceania, it would provide vital information towards finding Hawaiki, the original launching place for New Zealand Maori.

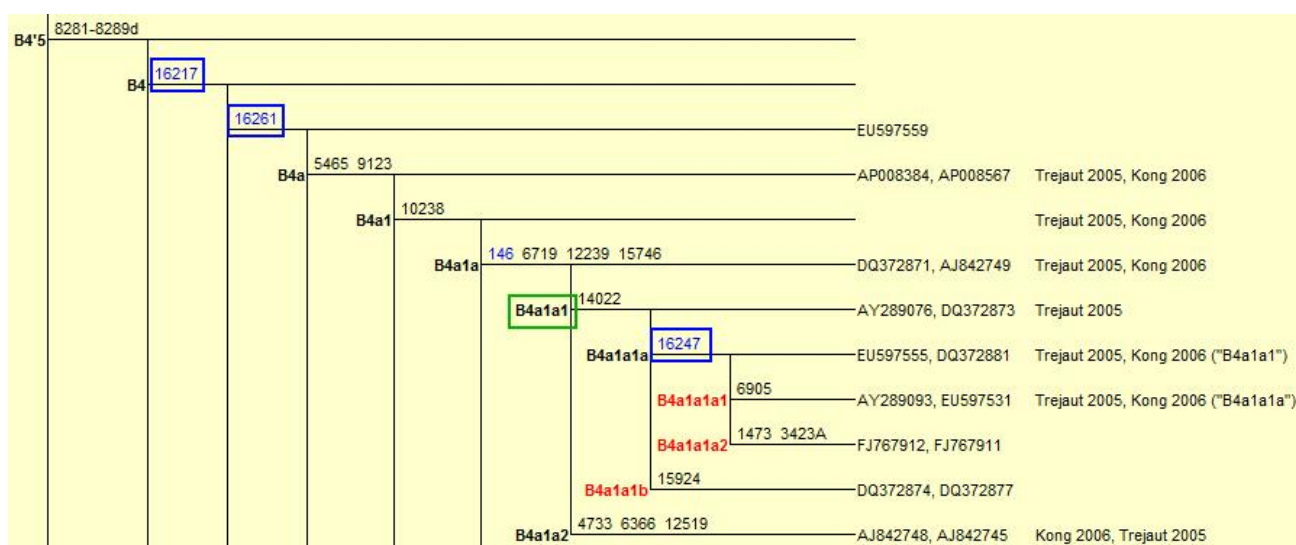


Figure 4.2: extended haplotype structure for B4, data obtained from Phylotree (van Oven *et al.*, 2009). Three of the HVRI Polynesian motif markers are indicated in blue boxes, while haplotype B4a1a1 is shown in green.

4.4 rRNA Secondary Structure Predictions

As mentioned previously (Introduction, Section 1.3), mitochondrial proteins stem from two separate locations. The majority of proteins incorporated into a mitochondrion are encoded by the nuclear genome and are translated in the cytoplasm, then transported via chaperone proteins to the mitochondrion. There are however 13 protein coding genes in the mtDNA that are essential for correct mitochondrial function that are translated within the matrix (Wallace 2006). This translation occurs on ribosomes encoded by the 12S (small subunit) and 16S (large subunit) rRNA transcripts within the mtDNA. It is therefore important that these ribosomes assemble and function correctly to avoid a decrease or mis-translation of these proteins crucial to OXPHOS (Bai *et al.*, 2005). Mutations to the 12S rRNA have been found to play a significant role in several diseases. The majority of the research conducted into mutations within the 12S rRNA gene has looked at sensorineural hearing loss and mitochondrially inherited deafness (Prezant *et al.*, 1993; Ballana *et al.*, 2008; Hobbie *et al.*, 2008).

The characteristics of ribosomal RNA are important from both a medical and an evolutionary point of view. Ribosomal RNA is the target of numerous clinically available antibiotics; including erythromycin, streptomycin (Ballana *et al.*, 2008). Many antibiotics target the ribosomal small subunit (SSU) of bacteria, exhibiting a bactericidal effect by interrupting protein synthesis. Extensive research has been conducted into the role of 12S rRNA mutations in syndromic and nonsyndromic hearing impairment (Prezant *et al.*, 1993; Ballana *et al.*, 2008; Hobbie *et al.*, 2008). It is suggested that specific mutations in human 12S rRNA alter the secondary structure so that it more closely resembles a bacterial counterpart (Ballana *et al.*, 2008). This has lead to the discovery of several mutations (A1555G, C1494T, T1291C) that seem to be

susceptible to aminoglycoside antibiotics (Prezant *et al.*, 1993; Ballana *et al.*, 2008; Lu *et al.*, 2009). It has been shown that these mutations decrease the accuracy of translation, and render the ribosomal decoding site hypersusceptible to aminoglycosides (Hobbie *et al.*, 2008). These aminoglycosides appear to have their mode of action on mitochondrial ribosomes located in the inner ear hair cells of affected individuals. Identification of variants that pose health risks, such as susceptibility to specific antibiotics, could potentially be developed into a genetic screen which would allow health providers to decide on alternative treatments.

Several areas of variation have been identified with the 20 RHAS Maori 12S rRNA sequences (Results, Section 3.5.2). One of particular interest is 1185>T, which was identified in 35% of the complete mtDNA sequences analysed. This variant is within the 12S rRNA gene. Secondary structure predictions were made using RNAwebserver to explore possible ribosomal structural changes (Results, Section 3.7). This prediction analysis was based on the same parameters utilised by Ballana *et al.*, (2006). When compared to the predicted consensus (wild-type) model, it was discovered that 12S rRNA which contained the 1185>T transition exhibited a small conformational change. It was seen that in these 12S models, a small hair pin seems to disassociate from a larger loop. The significance of this is unknown and would no doubt benefit from in-depth research in the future. No real structural changes were evident when the same analysis was conducted on 16S rRNA containing variants.

4.5 Protein Modelling Analysis

While it is important to identify correlations between identified variants and disease traits, it is perhaps more imperative to understand the biological mechanisms that contribute to disease. It is only when these mechanisms are elucidated that putative treatments and pharmaceuticals can be developed. Therefore it is essential to consider these modelling analyses in addition to the correlation work that has been conducted in this study (Section 4.6).

Much of the variation found within the RHAS Maori mtDNA's protein coding genes resulted in synonymous SNPs (Results, Section 3.5.2). Synonymous SNPs do not produce altered coding sequences, thus the same amino acid is coded for. These are termed silent mutations; as such they are not expected to have an effect on protein function (Kimchi-Sarfaty *et al.*, 2007). Because silent mutations do not alter protein function they are often treated as though they are evolutionarily neutral. However, many organisms are known to exhibit codon usage biases, suggesting that there is selection for the use of particular codons due to translational stability (Kimchi-Sarfaty *et al.*, 2007). Silent mutations may also affect splicing, or transcriptional control (Komar 2007). Recent results suggest that silent mutations can have an effect on subsequent protein structure and activity (Kimchi-Sarfaty *et al.*, 2007; Komar 2007) It has been postulated that silent mutations could slow the translation machinery, which in turn could lead to misfolded, or differently folded proteins (Komar 2007).

Cytochrome bc_1 complex (Complex III) is the third enzyme in the electron transport chain (Hansford 2002). It accepts electrons from reduced Coenzyme Q and passes them on to cytochrome c. Complex III contains four redox centers; two *b*-type hemes, a *c*-

type heme, and an iron-sulphur cluster. Of the eleven protein subunits that make up Complex III, with cytochrome *b* (cyt *b*) is the only one that is encoded by the mitochondrial genome (Hansford 2002). Cytochrome *b* is located in the center of the transmembrane region of the *bc*₁ complex. It contains 2 of the total 4 redox centers; heme *b*_L and heme *b*_H. These heme groups are vital to the transport of electrons (Voet *et al.*, 2004). As such the amino acid substitution (T15014C, Phe to Leu) could possibly interfere with this process. Phenylalanine prefers to be buried in protein hydrophobic cores, is fairly non-reactive, and is rarely directly involved in protein function (Betts *et al.*, 2003). Leucine has very similar properties; as such this substitution is most likely neutral. However there is a slight difference in structure between the aromatic side chain of phenylalanine and that of the aliphatic leucine (Betts *et al.*, 2003). Due to its location between the two *b* heme groups, in close proximity to both, this substitution could influence the surrounding structures eliciting and have some functional effect. Mutations in cytochrome *b* primarily result in exercise intolerance in human patients; though more rare, severe multi-system pathologies have also been reported (Blakely *et al.*, 2005).

Cytochrome *c* oxidase (Complex IV) is the terminal enzyme of the electron transport chain (Voet *et al.*, 2004). It receives an electron from each of 4 cytochrome *c* molecules, and uses these to create two water molecules (H₂O) from one oxygen molecule (O₂) (Voet *et al.*, 2004). One of the variants discovered in this study, A6261G, has been associated with prostate cancer (Petros *et al.*, 2005; M. Esther Gallardo 2006; Ray *et al.*, 2009). This variant results in a non-conservative amino acid substitution of threonine (encoded by the G allele) from alanine (encoded by the A allele) within the cytochrome *c* oxidase I (COI) subunit of Complex IV. It appears to be a fairly harmless amino acid substitution. In membrane proteins Ala and Thr are

usually the favoured AA substitutions (Betts *et al.*, 2003). However there have been cases where Ala/Thr substitutions have been associated with various disease states (Canter *et al.*, 2005; Bai *et al.*, 2007; Bhat *et al.*, 2007; Liao *et al.*, 2008). This variant could also impact on the interactions of COXI and the other subunits that bind to form the complete cytochrome c oxidase complex.

Protein modelling of specific RHAS Maori variants identified two common variants in close proximity to crucial COXI redox centres (Figure 4.3). These variants are transitions that result in synonymous mutations; there is no change to the coded amino acid.

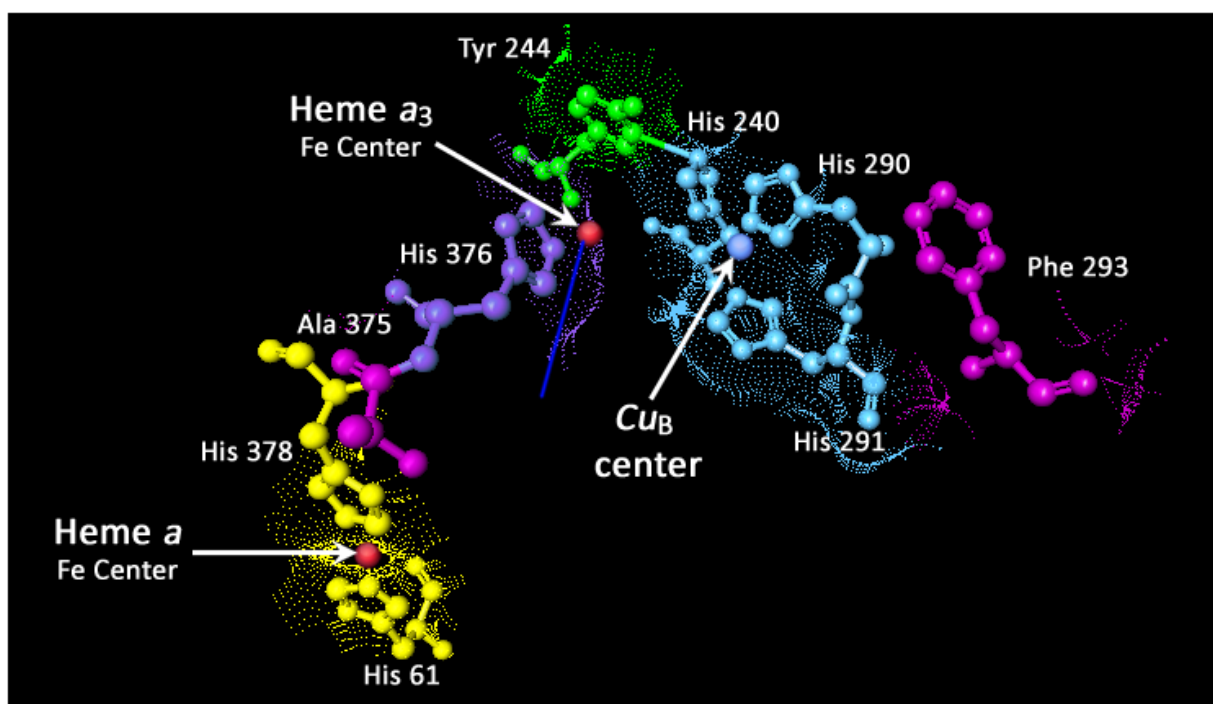


Figure 4.3: Redox prosthetic groups within Cytochrome c oxidase subunit I (COI). Structures shown represent those amino acids involved in the binding of the redox centers of COI. Amino acids are labelled and colour coded; yellow binding heme *a*, purple binding heme *a*₃, light blue binding Cu_B, interacting Tyr shown in green, and identified variants shown in pink. The Fe and Cu ions are represented by red and cyan spheres. Note that the side chains of Tyr 244 and His 240 are joined by a covalent bond. Protein model generated in Sirius (<http://sirius.sdsc.edu/index.php>).

As previously mentioned, synonymous mutations have been found to possibly cause misfolding due to codon usage bias (Kimchi-Sarfaty *et al.*, 2007). It is possible that these variants could affect the surrounding structures, which in turn could affect transfer of electrons and ATP production. It has been hypothesised that variants such as these have persisted in mtDNA as they provide functional advantages, such as uncoupling (Wallace 2006).

The protein COIII is a subunit of cytochrome oxidase, which is involved in the binding of COI and COII; COII brings the final complex enzyme together (Voet *et al.*, 2004). Two variants were discovered in two separate COIII RHAS Maori sequences. Interestingly neither of the variants have been reported previously, and so are novel to these mtDNA's at time of writing. One of the variants found was synonymous, T9722C. The other variant however, is of more interest; a transition C9255T. This transition results in a codon change, from CCU (proline) to UCU (serine). Of the 20 protein forming amino acids proline is the only one that has its α -amino group secondary (Betts *et al.*, 2003). Due to its unique structural properties proline can introduce kinks into α -helices in the protein. Therefore proline does not favour substitution with other amino acids easily (Betts *et al.*, 2003), this is also confirmed by the conservation index (100) seen in Results (Section 3.5.2, Table 3.4). This high index value indicates that the variant position has extremely high evolutionary conservation across 37 different organisms. With Pro17 being in close proximity to the N-terminus, this variant could possibly affect the position/structure this strand. This could in turn affect how this binds other subunits within Complex IV. Proline plays important roles in molecular recognition, particularly in intracellular signalling. Domains such as WW and SH3 bind to specific proline containing peptides that are key parts of many signalling cascades (Betts *et al.*, 2003). So if this proline was within a domain such as

these it could have adversely affect related pathways (i.e. second messenger, ROS signalling).

The protein modelling work conducted in this study is currently hypothetical and would greatly benefit from extensive protein interaction simulations. Such work could possibly begin to identify the effects of these variants within the protein coding genes. Regardless, this study has progressed significantly in providing the first examples of variation within the coding region of Maori mtDNA. Several variants of interest have been identified, and are plausible starting points for future analysis.

4.6 mtDNA variants and disease association

As outlined in Chapter 1 (Introduction, Section 1.4) mtDNA variation can contribute towards disease susceptibility and phenotype. Due to the recent general acceptance of this fact, mtDNA is starting to be implemented in numerous disorders, ranging from insulin resistance (Poulton *et al.*, 1998; Wiederkehr *et al.*, 2006; Gaster 2007) to Alzheimer's disease (Pericak-Vance *et al.*, 2002; Santoro *et al.*, 2006). As such, Mitomap, a database of recorded disease associated variants is growing substantially (Ruiz-Pesini 2007). With the aid of this database, specific mtDNA variants can be searched and any previously reported disease associations detected. The Mitochip assay of 20 complete Maori mtDNA sequences revealed a profile of 44 variants compared to the CRS. When queried against variants in the Mitomap database, four of these identified variants were found to have been previously associated with several disorders. These four variants were: A4769G, G6261A, T16189C, and T16519C.

The variant A4769G is somewhat of a misnomer due to the fact that is actually a rare polymorphism found within the CRS (Andrews *et al.*, 1999). The A allele at 4769 is actually the rarer variant and not a good representation of this position in other mtDNA's. Searching this variant in mtDB (Ingman 2006) returned 30 sequences out of ~3000 (~1%). This becomes interesting as 35% of the 20 RHAS Maori mtDNA's were found to exhibit the A allele at 4679, the other 75% having the more common G allele. This variant resides within the *ND2* gene, a subunit of OXPHOS Complex I. The G to A transition results in a synonymous mutation of the ATG(A) anticodon of methionine. A literature search identified that 4769>A had been recently associated with a group of schizophrenia patients (Rollins *et al.*, 2009). There were other possible associated variants identified within the same study, many of which were apparently associated

with various mental disorders including; schizophrenia, bipolar disorder, and depression. It has been hypothesised that the higher energy dependency of neural tissue makes it more susceptible to adverse mtDNA mutations than tissue that is less energy dependent (Santoro *et al.*, 2006; Moreira *et al.*, 2007; Rollins *et al.*, 2009). It was recently identified that Maori appear to show elevated levels of schizophrenia when compared to non-Maori (Kake *et al.*, 2008). Unfortunately schizophrenia or other mental disorders were not screened in the RHAS, meaning that at this time no association testing can be conducted. This is however an excellent avenue for a future research project.

Another variant identified in RHAS Maori that has previously been associated with disease is G6261A. This variant has been associated with prostate cancer in several studies (Jeronimo *et al.*, 2001; Petros *et al.*, 2005). The 6261 variant was found to be present in low frequency (1%) in the Mitochip RHAS Maori sequences. It must be noted that this is a very small sample size (n=20). As such this is another variant that could form part of a possible future screen in the wider Maori community.

Specific HVRI variants have been found to have associations with metabolic disorders. The most documented of these is the T16189C transition, which also happens to be a major haplotype defining variant of haplogroup B, mentioned extensively in this thesis. Interestingly the 16189 variant has been associated with a wide range of metabolic disorders: insulin resistance and diabetes (Poulton *et al.*, 1998; Lin *et al.*, 2005; Bhat *et al.*, 2007; Liao *et al.*, 2008), cancer (Liu *et al.*, 2003), metabolic syndrome (Weng *et al.*, 2005), and cardiomyopathy (Khogali *et al.*, 2001). A study by Liao *et al.* (2008) identified several mtSNPs which showed an increased association with type-2-diabetes mellitus (T2DM), including the two variants T16189C and T16519C. These two

variants were found in significantly higher frequencies in patients suffering T2DM than in individuals that didn't (Liao *et al.*, 2008). Caution is advised however when considering all of these associations. As the variants in question are major markers of mtDNA ancestry, it is possible that they are picking up diseases associated with other underlying genetic factors. For example, the presence of T16189C is a very good indicator of mtDNA haplogroup B. Hence any associations involving frequencies of this variant with disease could well be identifying the contribution of a multitude of ancestral genetic markers.

4.6.1 Disease Association in the RHAS Maori Community

As part of the RHAS individuals were asked to supply blood and questionnaire information for clinical trait profiling. As such the final database of RHAS Maori adults contained ~400 individuals, with extensively documented clinical data. Added to this was also genotyping screens of three mtDNA variants identified within RHAS Maori. This database allowed the statistical analysis and association testing of specific metabolic traits such as; diabetes, BMI, lipid profiles (LDL, HDL), reactive peptide levels, waist hip ratio, and many more. Of the three SNPs, only two were identified in frequencies that allowed for statistical correlations. These two SNPs were 4769>A and 16189>C. The next section discusses possible disease association with these markers.

4.6.1.1 T16189C possible disease associations in RHAS Maori

The variant 16189>C was identified in 88% of the wider RHAS Maori community. Association testing with diabetes returned non-significant results, so hypothesis testing was used to identify any significant correlations with metabolic traits (Results, Section 3.9.3). Significant correlations identified included increased means between the variant and common allele in: vitamin B12 and mean diastolic blood pressure.

Vitamin B12 is directly related to energy metabolism. Generally levels can fluctuate broadly, with higher levels of vitamin B not having any adverse effects. In fact higher levels of vitamin B12 are possibly protective against ROS, and aging. Lower levels of B vitamin however have been found to cause disease states, usually neurological in nature (Herrmann *et al.*, 2008). This could possibly be due to an interaction between vitamin B12 and energy metabolism of the mitochondria in neural tissue. Lower levels of vitamin B12 would lower metabolism levels decreasing energy production. Decreased energy production in neural tissue has been thought to lead to disorder such as Parkinson's, Alzheimer's, depression, and others (Ghezzi *et al.*, 2005). If this is the case then having higher vitamin B12 levels is beneficial, and could possibly be somewhat protective against neurological disorders in RHAS Maori.

A trend towards decreased CREA (creatinine) levels was also noticed in RHAS Maori with the 16189>C variant. Decreased creatinine levels have been shown to correlate with the presence of type-2-diabetes in Asian men (Harita *et al.*, 2009). It has been suggested that lower creatinine levels could be a good risk factor for early diabetes detection. As levels of diabetes are elevated in RHAS Maori, it could be worth following this analysis up in a larger group to further investigate the trend of creatinine. A weak correlation with a general increase in mean diastolic blood pressure was also noticed with the variant T16189C. This could be a possible link to cardiovascular disease risk, and is worth following up.

4.6.1.2 A4769G possible disease associations in RHAS Maori

In the correlation against metabolic traits, G4769A was found to associate significantly with five specific traits; CHHD, TRIG, Factor2, CHOL, and CVD. It is interesting that all five traits are showed significant correlations with similar metabolic traits. These

traits were; LDL (low-density lipoprotein), CHOL (cholesterol), CHHD (blood lipid ratio), TRIG (triglycerides), and Factor2 (cardiovascular risk statistic), all of which can be considered risk factors for cardiovascular disease. Further testing of means found that three of these traits with have statistically higher means than the common allele in RHAS Maori. These were CHHD, TRIG, and Factor2. This variant has been identified as a RHAS specific variant, and could well be associated with metabolic disorder, specifically cardiovascular disease, in the Rakaipaaka community. It has been observed that RHAS Maori exhibit higher than national levels of metabolic disorders. It's therefore possible that variant G4769A could help explain this and become the basis of early diagnostic screens.

In a chi squared analysis with risk association of 4769 and diabetes, an interesting trend was observed. While not quite at the $p=0.05$ level of significance, it is still interesting to document. The risk analysis odds ratio suggests that individuals with the 4769>G variant are more than 2 times less likely to develop diabetes. Again this was not significant, but warrants further genotyping in the wider Maori community.

A recent study considered mtDNA haplogroups and individual variants and their possible role in susceptibility of psychiatric disorders (Rollins *et al.*, 2009). In this work the 4769A variant showed a significant allelic association with Schizophrenia. It has been identified that Maori exhibit elevated incidences of Schizophrenia than non-Maori (Kake *et al.*, 2008). This provides strong evidence that a possible relationship between mtDNA variants and mental disorders such as Schizophrenia in Maori may exist, which would benefit from screens in the wider Maori community.

Chapter 5 Summary and Conclusion

5.1 Summary

The major findings of this thesis are summarised below in relation to the research aims from Introduction, Section 1.8:

- 1) *Design and optimisation of an RFLP assay and HVRI PCR to aid with subject selection for Mitochip array analysis.*

A total of 94 HVRI sequences were obtained from Maori individuals enrolled in the RHAS. Sequences were generated using PCR and automated sequencing. All HVRI sequences will be entered in to GenBank for access by the scientific community after this thesis has been submitted. Of the 94 HVRI sequences 93 were reported as belonging to haplogroup B; 75 were identified as having the ‘CCGT’ motif, while 18 were found to have a previously reported pre-Polynesian ‘CCAT’ motif. One individual sequence lacked any Polynesian mtDNA markers and was determined to belong to the European haplogroup H (Helena). Several novel variants were found within these HVRI sequences, of note was an insertion at 16295. From the HVRI database 20 RHAS mtDNA specimens were selected for complete re-sequencing using the Affymetrix Mitochip assay. These 20 were selected to be a representation of mtDNA diversity within the Maori of the Rakaipaaka iwi. As such 14 sequences containing the ‘CCGT’ motif, and 6 sequences with ‘CCAT’ motif, were selected. To further refine selection, HVRI sequences showing unique individual variants were chosen as internal controls for the Mitochip assay.

- 2) *Obtain entire mtDNA sequence information from a subsample of Rakaipaaka individuals with Maori ancestry (n=20) using the Affymetrix Mitochip.*

The 20 complete mtDNA from Rakaipaaka Maori revealed 9 novel variants previously undiscovered. Of the 20 mtDNA sequences 7 had variants 1185>T and 4769>A that define a unique mtDNA haplotype. Another variant specific to RHAS Maori, 16216>C, also defined a unique haplotype. These are the first Maori specific haplotypes to be recorded. It is yet to be determined if these are Rakaipaaka specific haplotypes, or if they exist in the wider Maori community.

- 3) *Align and compare this sequence information with independent worldwide sequences to identify group differences, i.e. a mtDNA signature that is specific to Rakaipaaka Maori.*

These complete mtDNA sequences are the first documented for Maori. Comparison with available Polynesian mtDNA sequences revealed extremely high similarity. It was determined that Maori mtDNA shares all the same haplotype B4a1a1 markers across mtDNA as Polynesians, which was previously assumed but lacked proof from any complete Maori mtDNA analysis. Validation of the 3 variants, C1185T, G4768A, and T16126C, in global databases identified these as being RHAS Maori specific. These variants then define a unique mtDNA signature within the RHAS Maori community.

- 4) *Validate the signature in a larger community sample of the RHAS Maori population (n=400).*

Several of the variants identified have previously been associated with possible disease traits and susceptibility; 4769, 6261, 16189, 16519. The previously associated diseases included; mental disorders (schizophrenia), prostate cancer, insulin resistance, diabetes,

and cardiomyopathy. Of the four variants, the three exhibiting highest frequencies were genotyped in the larger RHAS Maori community.

- 5) *Correlate the validated mtDNA signature with metabolic traits (eg. BMI and diabetes).*

Three SNP's of interest were genotyped in the wider RHAS cohort group; 4769, 16189, and 16519. These SNP's were present in large numbers in the extended RHAS database, C variant at 16189 was present in 88%, C variant at 16519 was present in 96%, and the A variant at 4769 was present in 25%. Hypothesis testing was conducted using variants 4769 and 16189, exploring possible correlations and associations with metabolic traits. The A variant at 4769 was observed to associate with three cardiovascular disease traits (CHHD, TRIG, and Factor2), while the C variant at 16189 appears to be associated with increased mean vitamin B12 levels as well as increased mean diastolic blood pressure. In terms of disease association, neither SNP proved to have significant associations with metabolic traits such as waist hip ratio (WHR), BMI, or diabetes. However a chi squared test indicated a possible trend between the variant 4769>G and diabetes in the RHAS Maori community. It appears that if the variant (G) is present at 4769 then it is more than 2 times less likely that an individual will develop diabetes.

5.2 Conclusion

This study has answered all the specific objectives which were where laid out. Complete mitochondrial genome sequencing of 20 RHAS Maori has provided pioneer research into the mtDNA variation within New Zealand Maori. It has been discovered that a unique mtDNA signature exists within RHAS Maori. The three variants in this signature create the first unique Maori mtDNA haplotypes where there were originally thought to be none.

Previously the only mitochondrial genetic information that had been extensively analysed in New Zealand Maori was HVRI sequences. This study therefore documents the first ‘complete’ Maori mitochondrial genomes, those of 20 RHAS Maori individuals. The mitochondrial variants discovered could ultimately lead to the development of early diagnostic and prevention of metabolic disorders, such as diabetes and cardiovascular disease, for New Zealand Maori.

The goal now is to understand how these markers interact with environmental variables to increase the risk of metabolic syndrome. Such an outcome may open the way to designing personalised intervention strategies (e.g. dietary) to increase the health and well-being of at risk individuals. As part of RHAS program the final results from this research add great knowledge to both the health and ancestry aspects of the studies main aims.

Future Directions

This project set out to screen the entire mtDNA of selected RHAS Maori individuals for unique genetic signature markers. The screen revealed that alongside the expected haplogroup B4a1a1a (Polynesian) markers, RHAS Maori mtDNA contained a combination of signature variants. When compared to global frequency data, the amount of variation within Maori mtDNA was found to be much higher than currently thought. Genotyping analysis identified high frequencies of two of these variants within the RHAS Maori community. Statistical correlation studies suggested that relationships exist between these variants and metabolic traits, such as diabetes, vitamin B12, and cardiovascular disease. The identification of specific mtDNA markers and their possible associations with metabolic traits opens up many avenues of future research.

Increase Sample Size

There is now technology available which allows for the cheap and efficient resequencing of mitochondrial genomes. By generating more sequence data the larger mtDNA variation noticed in this study can be testing in outside Maori communities. This could give information about whether mtDNA variation is different between Maori iwis. This would also establish if the signature markers discovered in RHAS Maori, extend out in to the wider Maori community.

Haplotype Simulations and Founding Population

The finding of unique Maori mitochondrial haplotypes raises important evolutionary questions. Identifying increased numbers of Maori specific haplotypes allows for simulations, such as that presented by Whyte *et al.*, (2005), to be conducted. This data would extrapolate on previous work, and it is likely the estimate of founding population

size would increase. Applying complete Maori mtDNA in phylogenetic applications could also prove beneficial in population and evolutionary genetics.

Variant Genotyping and Correlation

This project investigated and identified metabolic associated mtDNA variants in a single Maori iwi. Further research extending this into the wider NZ Maori community would be very informative. Variants previously associated with disease were identified within the RHAS Maori community. One identified variant, G4769A, has recently been implicated in schizophrenia. It has also been previously noted that Maori have elevated incidences of schizophrenia compared to non-Maori. This therefore presents the perfect opportunity to screen for mtDNA variants in Maori mental health patients.

In RHAS Maori two variants, G4769A and T16189C, were found to associate with metabolic traits including; increased vitamin B12, increased mean diastolic blood pressure, and cardiovascular risk factors. These two were the only variants genotyped in the wider RHAS Maori community. There were however, several other variants of note that would be informative if screened. These are two other RHAS Maori specific variants; 16126 and A16247G. They appear to be in a similar linkage relationship as 1185 and 4769. These four SNP's together create 3 unique mtDNA haplotype that have not been identified previously. From genotyping G4769A it was identified that it resides at a frequency of about 25% in RHAS Maori. Having similar information for other variants, such as T16126C and A16247G, would prove highly beneficial. It would also allow for further disease association studies, which could ultimately lead to the development of early diagnostic and prevention of metabolic disorders, such as diabetes and cardiovascular disease, for New Zealand Maori.

APPENDICES

Appendix One

Rakaipaaka Health and Ancestry Study fact sheet.



Rakaipaaka Health & Ancestry Study

Norma Hinepahauariki Bardell,
23.11.1940 – 19.09.1994

Letter to the Whanau,

My mother Norma Hinepahauariki Bardell (nee Nepia) lost her battle with cancer in 1994, aged 53. Sadly, she never had the chance to meet her two mokopuna (my son's). I am aware that many of the whanau (some very young) have also been affected by cancer and other horrible illnesses such as diabetes. Rod, my husband of 12 years, is a medical researcher interested in applying his knowledge to studying diseases that affect the Rakaipaaka community. We are both passionate about the future health of the Iwi and through research aim to determine the major health issues affecting the Rakaipaaka community. We ultimately hope to determine ways of preventing these illnesses from occurring in our future generations.

Marino Lea,



How will information be made available to Rakaipaaka?

- The monthly Rakaipaaka Panui
- The Rakaipaaka Website (www.rakaipaaka.co.nz)
- Executive management committee hui
- General community hui
- Events such as "Rakaipaaka Day"

The research will be conducted with full respect for the culture and beliefs of the Rakaipaaka people. The health and genetic information gained from this research will be strictly confidential and will only be made accessible if authorised (or requested) by the consenting individual participant. All biological material (ie. blood and DNA) will remain the property of the individual (person) from whom it originated and belongs.

Information will be dealt with in accordance with the rules and guidelines specified by the Maori Health Committee and the Health Research Council of New Zealand (www.hrc.govt.nz/).

All ethics approval will be obtained through the Central Regional Ethics Committee.

For more information please contact;

Marino Lea (Senior Researcher)
Phone: 04 914 0717 (work),
0210649954 (cell)
Email: Marino.Lea@esr.cri.nz

Dr Rod A Lea
Phone: 04 914 0664 (work),
0211887876 (cell)
Email: Rod.Lea@esr.cri.nz



The photo above shows Melvin Nepia (Uncle Bub) donating blood for the research. My young son (watching his Koro) was brave enough to donate a blood sample.

Who will conduct the study? And where will it be conducted?

A small team of medical researchers and public health/social workers and volunteers will conduct the research. Where possible the staff and volunteers will be members of Rakaipaaka. The project will be led by Dr Rod Lea and assisted by his wife Marino Lea (nee Bardell). The research will involve collection from participants in their homes and events in the Nuhaka, Wairoa & Wellington regions. The data and medical analyses will be conducted at the Environmental Science and Research Centre (ESR).

How long will the proposed research take?

This research will be a long-term multi-year endeavour. It is envisaged that the community consultation, health surveying and data collection will take up to 2 years to complete. Identifying the genetic, medical and lifestyle factors will take at least 3 years to complete. After 5 years we will also reassess the Rakaipaaka health status to monitor the trends that have occurred and to assess the success of any public health education and intervention initiatives. Long term follow up research (10 years +) may also be conducted.

The Rakaipaaka Health & Ancestry Study will investigate the serious diseases and conditions that are common in Rakaipaaka and determine whether there are underlying ancestral reasons for these conditions. The knowledge gained from the research may help improve the overall wellbeing of the Rakaipaaka community.

- To determine the general health status of Rakaipaaka and identify the major diseases affecting the Rakaipaaka community
- To develop a computerised whakapapa health registry for Rakaipaaka
- Identify the genes and lifestyle reasons that might cause diseases in Rakaipaaka
- Utilise the new knowledge gained to inform future health planning and action for Rakaipaaka

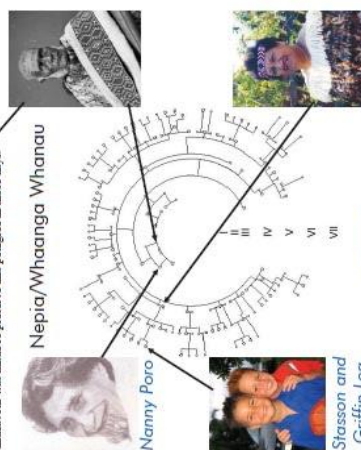
There is some evidence to suggest that the Rakaipaaka community has high rates of diseases such as diabetes, gout, heart disease and cancer. Some of these tend to be passed down through generations suggesting an inherited component. However, there is no research-based information about which conditions form serious problems in the Rakaipaaka community as a whole. A study such as

this will help to identify the major health issues of Rakaipaaka as well as identify those diseases, which might have a genetic basis in the Iwi.

The most successful strategy researchers have for uncovering genes that cause disease is to carry out genetical inheritance studies of very large, extended families. The living descendants of Rakaipaaka have now extended into the 15th generation and are all part of a very impressive family tree with many living relatives across several generations. Rakaipaaka has excellent historical knowledge and records of their whakapapa. This information combined with medical science and technology may help reveal the reason why some of the Rakaipaaka people suffer from these diseases.

Year (AD)	Chiefs	Wives
1350	Tamisa Atikuni Ma Tawabi (Commander in Chief of Takitimu)	Toto (Resident of Aotearoa)
1500	Koukako Tamata Ureakea Manungu (Paramount Chief) Rahakuruni Rakapoko (Paramount Chief) Kaitiaki Tutakamo	Muriwhenu Wharou Rongomawhene Tunelongo Tunamanga Hawete Huahehema Huahehema Te Rangihounga Makangia I Te Tahi Te Wharehaka Te Wharehaka Te Waiora Rangia Ahuranga Te Haka Rakano Epeka Tahera Te Paea Tamaru
1600		
1700	Te Haki (Paramount Chief) Pumaua Manatua Munioewangi Te Puman Whangau Isaka Tem Whangau (Paramount Chief)	
1800		

All the chief's wives were the daughters or granddaughters of other chiefs and every endeavour was made to preserve this pedigree in those days.



- Establishment of a health information system detailing genealogical, and medical records of Rakaipaaka will be useful for future healthcare planning and general Rakaipaaka education.
- Employment and training of Rakaipaaka researchers, community workers and administrators.
- Generation of commercial knowledge (intellectual property) for developing new medicines and diagnostics for disease.

Participants of this research you will receive a general health check eg. Blood pressure, cholesterol, blood sugar. You will also contribute valuable health and ancestral information about the Rakaipaaka community as a whole. This information will be useful for future health care planning and action.

Many of the younger children of Rakaipaaka do not suffer from illnesses today however; this may change, as they grow older. The health knowledge we gain now may allow us to protect our children from these illnesses, enhancing their lives and the future of Rakaipaaka.

- Informed consent acknowledging that you understand the nature of the research and agree to participate
- Details about each person's specific family medical history (if known)
- Completion of a Health Questionnaire
- Measurement of each person's height, weight, blood pressure, heart rate etc
- A small sample of blood (50mls) to conduct biochemical testing and obtain DNA

Appendix Two

HVRI PCR Particulars

HVRI PCR Primers:

Forward	L15975	5'-CTCCACCATTAGCACCCAAAGC-3'
Reverse	H16517	5'-CCTGAAGTAGGAACCAGATG-3'

HVRI PCR Product:

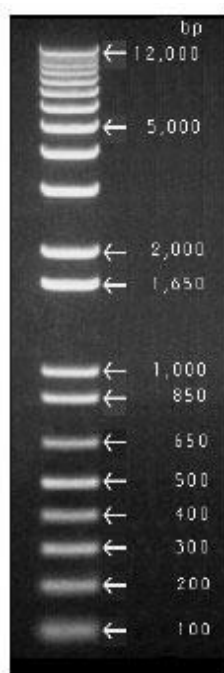
PCR product 543bps

CTCCACCATTAGCACCCAAAGCTAAGATTCTAATTTAAACTATTCTCTGTTCTTTCATGGGGA
AGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTTCGTAC
ATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACAT
AAAAACCCAATCCACATCAAAACCCCTCCCCATGCTTACAAGCAAGTACAGCAATCAACCC
TCAACTATCACACATCAACTGCA^GCTCCAAAGCCACCCCTCACCCACTAGGATACCAACAA
ACCTACCCACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAG
TCAAATCCCTTCTCGTCCCCATGGATGACCCCTCAGATAGGGGTCCCTTGACCACCATCCT
CCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCTCGTCCGGGCCATAACACTTGG
GGGTAGCTAAAGTGAAGTGTATCCGACATCTGGTTCCTACTTCAGG

NB: seq is from CRS, changed A to G to highlight *Pst* I restriction site

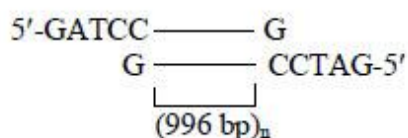
Appendix Three

1Kb Plus DNA Ladder from Invitrogen



1 Kb Plus DNA Ladder
0.7 µg/lane
0.9% agarose gel
stained with ethidium bromide

Structure of Fragments in 1-Kb Increments:



Notes:

During 1% agarose gel electrophoresis with Tris-acetate (pH 7.5) as the running buffer, bromophenol blue migrates together with the 500 bp band.

The 1650 bp band is generated from pUC. The bands smaller than 1000 bp are derived from lambda DNA.

Appendix Four

Pico Green Standards

To make 1 ml of each standard:

Add 100 µl DNA from Invitrogen kit to 4900 µl TE buffer = 2 µg/ml

Add 100 µl of the above 2 µg/ml solution to 900 µl TE buffer = 0.2 µg/ml

200 ng = 1000 µl of 2 µg/ml solution

100 ng = 500 µl of 2 µg/ml solution + 500 µl TE buffer

50 ng = 250 µl of 2 µg/ml solution + 750 µl TE buffer

25 ng = 125 µl of 2 µg/ml solution + 875 µl TE buffer

12.5 ng = 62.5 µl of 2 µg/ml solution + 937.5 µl TE buffer

2.5 ng = 125 µl of 0.2 µg/ml solution + 875 µl TE buffer

1 ng = 50 µl of 0.2 µg/ml solution + 950 µl TE buffer

0.5 ng = 25 µl of 0.2 µg/ml solution + 975 µl TE buffer

Appendix Five

Long range PCR primers for Mitochip analysis

Mito1-2	Forward Primer	Total Bp
	ACATAGCACATTACAGTCAAATCCCTTCTCGTCCC	
	Reverse Primer	
	ATTGCTAGGGTGGCGCTTCCAATTAGGTGC	
Mito3	Forward Primer	Total Bp
	TCATTTTATTGCCACAACCTCCTCGGACTC	
	Reverse Primer	
	CGTGATGTCTTATTTAAGGGGAACGTGTGGGCTAT	

Appendix Six

Example Pico-Green assay output showing average the concentrations for 15 RHAS Maori HVRI PCR products.

Testname: ENVGPICO

2008/12/19 14:54:18

ID 1,2,3: Miles 15 RHAS samples

Content	Sample ID	Well	Raw data	Avg of replicates	SD of replicates	%CV	Calculated conc. (ng/μl)	
Blank		H01	226	217	10	4.4	-0.230	
Blank		H02	207				-0.263	
S1	Std	A01	58534	55165	3369	6.1	100.538	
S1	Std	A02	51796				88.893	
S2	Std	B01	49098	50670	1572	3.1	84.230	
S2	Std	B02	52242				89.664	
S3	Std	C01	34199	35066	867	2.5	58.482	
S3	Std	C02	35933				61.479	
S4	Std	D01	16934	17341	407	2.3	28.644	
S4	Std	D02	17748				30.051	
S5	Std	E01	5947	5911	37	0.6	9.657	
S5	Std	E02	5874				9.530	
S6	Std	F01	2729	2977	248	8.3	4.095	
S6	Std	F02	3225				4.952	
S7	Std	G01	802	876	74	8.4	0.765	Avg Conc (ng/μl)
S7	Std	G02	949				1.019	
X1	mt41	A03	8067	8324	257	3.1	13.320	13.8
X1	mt41	A04	8581				14.209	
X2	mt42	B03	9222	8894	328	3.7	15.317	14.7
X2	mt42	B04	8566				14.183	
X3	mt43	C03	8703	8970	267	3.0	14.420	14.9
X3	mt43	C04	9236				15.341	
X4	mt44	D03	8970	8948	23	0.3	14.881	14.8
X4	mt44	D04	8925				14.803	
X5	mt45	E03	8990	9344	354	3.8	14.916	15.5
X5	mt45	E04	9697				16.137	
X6	mt46	F03	7960	8167	207	2.5	13.136	13.5
X6	mt46	F04	8373				13.849	
X7	mt47	G03	8166	8582	416	4.8	13.492	14.2
X7	mt47	G04	8998				14.929	
X8	mt50	H03	7481	7483	2	0.0	12.308	12.3
X8	mt50	H04	7485				12.315	
X9	mt51	A05	3638	4097	459	11.2	5.666	6.5
X9	mt51	A06	4556				7.253	
X10	mt52	B05	7850	7800	51	0.6	12.945	12.9
X10	mt52	B06	7749				12.771	
X11	mt53	C05	6124	7396	1272	17.2	9.963	12.2
X11	mt53	C06	8667				14.357	
X12	mt54	D05	5165	5520	355	6.4	8.305	8.9
X12	mt54	D06	5874				9.530	
X13	mt55	E05	4421	4564	143	3.1	7.019	7.3
X13	mt55	E06	4706				7.512	
X14	mt56	F05	7128	7433	305	4.1	11.698	12.2
X14	mt56	F06	7737				12.750	
X15	mt57	G05	8021	7651	370	4.8	13.241	12.6
X15	mt57	G06	7281				11.962	

Appendix Seven

Sample ID's, freezer position, and working stock concentrations for the 20 RHAS Maori DNA specimens selected for Mitochip array analysis.

Mitochip Code	Alt. sample no.	Box	Position	ng/μl
DM001	mt3	30LT00017	A11	34.2
DM002	mt6	30LT00018	B6	32.3
DM003	mt8	30LT00018	A7	25.2
DM004	mt9	30LT00018	E7	31.9
DM005	mt10	30LT00018	H7	24.6
DM006	mt11	30LT00018	A8	33.8
DM007	mt12	30LT00018	F8	32.7
DM008	mt13	30LT00018	G8	42.2
DM009	mt14	30LT00018	B9	37.9
DM010	mt16	30LT00018	A10	23.6
DM011	mt18	30LT00017	G11	30.9
DM012	mt19	30LT00017	H11	22.4
DM013	mt21	30LT00017	C12	17.9
DM014	mt23	30LT00018	C1	22.1
DM015	mt24	30LT00018	E1	38.1
DM016	mt25	30LT00018	C2	47.3
DM017	mt26	30LT00018	D2	24.0
DM018	mt28	30LT00018	F2	28.3
DM019	mt29	30LT00018	A3	35.4
DM020	mt30	30LT00018	B3	32.8

Appendix Eight

Catalogue of GENBANK Accession number's Maori HVRI mtDNA sequences used in variant frequency and haplotype analyses.

1. mtDNA HVRI data set from *Whyte et al.*, (2005).

Adele Whyte Data Set					
A Y604156	A Y604141	A Y604126	A Y604111	A Y604096	A Y604081
A Y604155	A Y604140	A Y604125	A Y604110	A Y604095	A Y604080
A Y604154	A Y604139	A Y604124	A Y604109	A Y604094	A Y604079
A Y604153	A Y604138	A Y604123	A Y604108	A Y604093	A Y604078
A Y604152	A Y604137	A Y604122	A Y604107	A Y604092	A Y604077
A Y604151	A Y604136	A Y604121	A Y604106	A Y604091	A Y604076
A Y604150	A Y604135	A Y604120	A Y604105	A Y604090	A Y604075
A Y604149	A Y604134	A Y604119	A Y604104	A Y604089	A Y604074
A Y604148	A Y604133	A Y604118	A Y604103	A Y604088	A Y604073
A Y604147	A Y604132	A Y604117	A Y604102	A Y604087	A Y604072
A Y604146	A Y604131	A Y604116	A Y604101	A Y604086	A Y604071
A Y604145	A Y604130	A Y604115	A Y604100	A Y604085	
A Y604144	A Y604129	A Y604114	A Y604099	A Y604084	
A Y604143	A Y604128	A Y604113	A Y604098	A Y604083	
A Y604142	A Y604127	A Y604112	A Y604097	A Y604082	

2. mtDNA HVRI data set from Pierson (Unpublished Data).

Maori	Tongan	Cook Is	Samoa	Solomon Is	Niuean	Tokelauan	Fijian
EF077399	EF077401	EF077402	EF077400	EF077389	EF077386	EF077371	EF077363
EF077393	EF077398	EF077394	EF077396				
EF077385	EF077397	EF077388	EF077395				
EF077378	EF077392	EF077387	EF077391				
EF077369	EF077384	EF077381	EF077390				
EF077368	EF077375	EF077377	EF077383				
EF077367	EF077374	EF077376	EF077382				
EF077365	EF077370	EF077362	EF077380				
EF077361			EF077379				
EF077360			EF077373				
EF077359			EF077372				
			EF077366				
			EF077364				
			EF077358				
			EF077357				

Appendix Nine

Global mtDNA variant frequency data for all informative positions within RHAS Maori sequences.

	European		Chinese		Japanese		Taiwanese		Melanesian		Polynesian		NZ Maori		RHAS Maori	
	SEQS	Percent	SEQS	Percent	SEQS	Percent	SEQS	Percent	SEQS	Percent	SEQS	Percent	SEQS	Percent	SEQS	Percent
HVRI	A16051G	1	0	0.0	1	0.8	10	2.8	0	0.0	2	1.7	2	8.0	3	3.2
	T16086C	0	1	2.0	1	0.8	54	14.9	4	2.5	4	3.3	1	4.0	2	2.1
	T16126C	5	1	2.0	1	0.8	9	2.5	1	0.6	1	0.8	2	8.0	17	18.1
	A16182C	1	5	9.8	13	10.8	91	25.1	49	30.6	93	77.5	67	89.3	93	98.9
	A16183C	2	13	25.5	32	26.7	124	34.2	50	31.3	114	95.0	67	89.3	93	98.9
	T16189C	3	18	35.3	45	37.5	133	36.6	65	40.6	114	95.0	67	89.3	93	98.9
	C16190T	0	0	0.0	1	0.8	1	0.3	0	0.0	0	0.0	0	0.0	1	1.1
	T16217C	0	6	11.8	9	7.5	99	27.3	46	28.8	114	95.0	67	89.3	93	98.9
	T16243C	0	2	3.9	7	5.8		0.0	0	0.0	0	0.0	0	0.0	1	1.1
	A16247G	0	0	0.0	0	0.0	0	0.0	41	25.6	97	80.8	60	80.0	75	79.8
	C16260T	0	1	2.0	3	2.5	1	0.3	0	0.0	0	0.0	0	0.0	5	5.3
	C16261T	0	3	5.9	9	7.5	59	16.3	39	24.4	111	92.5	66	88.0	93	98.9
	T16271C	0	0	0.0	1	0.8	1	0.3	1	0.6	0	0.0	1	1.3	1	1.1
	G16274A	0	1	2.0	6	5.0	17	4.7	5	3.1	0	0.0	0	0.0	4	4.3
	C16290C	0	3	5.9	11	9.2	2	0.6	0	0.0	0	0.0	0	0.0	1	1.1
	A16293G	1	0	0.0	0	0.0	0	0.0	3	1.9	0	0.0	0	0.0	3	3.2
	T16311C	3	11	21.6	11	9.2	78	21.5	63	39.4	8	6.7	2	2.7	2	2.1
	A16317G	0	0	0.0	0	0.0	3	0.8	0	0.0	1	0.8	0	0.0	1	1.1
	T16519C	80	74.1	23	45.1	57	47.5	110	58.5	87	54.4	95	79.2	No Data	20	100.0

Global mtDNA variant frequency data for all informative positions within RHAS Maori sequences.

Gene	Variant	European		Chinese		Japanese		Taiwanese		Melanesian		Polynesian		RHAS	
		Seqs	Percentage	Seqs	Percentage	Seqs	Percentage	Seqs	Percentage	Seqs	Percentage	Seqs	Percentage	Seqs	Percentage
HVRII	A73G	15	60.0	49	100.0	62	100.0	12	100.0	49	100.0	9	100.0	20	100.0
	T146C	3	12.0	8	16.3	3	4.8	7	58.3	18	36.7	9	100.0	20	100.0
	C151T	0	0.0	1	2.0	1	1.6	0	0.0	4	8.2	1	11.1	2	10.0
	A263G	25	100.0	48	98.0	62	100.0	12	100.0	49	100.0	9	100.0	20	100.0
12SrRNA	A750G	101	100.0	49	100.0	62	100.0	12	100.0	56	100.0	9	100.0	20	100.0
	C1185T	0	0.0	0	0.0	0	0.0	0	0.0	1	1.8	0	0.0	7	35.0
	A1438G	100	99.0	47	95.9	61	98.4	12	100.0	51	91.1	9	100.0	20	100.0
16SrRNA	A1692G	0	0.0	0	0.0	0	0.0	0	0.0	1	1.8	1	11.1	2	10.0
	T2416C	0	0.0	0	0.0	1	1.6	0	0.0	0	0.0	1	11.1	2	10.0
	A2706G	64	63.4	49	100.0	60	96.8	12	100.0	56	100.0	9	100.0	20	100.0
	C3909T	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	3	15.0
ND2	A4769G	100	99.0	49	100.0	62	100.0	11	91.7	56	100.0	9	100.0	13	65.0
	T5465C	0	0.0	0	0.0	1	1.6	8	66.7	9	16.1	7	77.8	20	100.0
COI	G6261A	1	1.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	2	10.0
	T6719C	0	0.0	0	0.0	0	0.0	6	50.0	9	16.1	7	77.8	20	100.0
	C7028T	64	63.4	49	100.0	62	100.0	12	100.0	55	98.2	9	100.0	20	100.0
	A8860G	101	100.0	49	100.0	62	100.0	12	100.0	56	100.0	9	100.0	20	100.0
ATP6	G8865A	0	0.0	0	0.0	0	0.0	0	0.0	1	1.8	0	0.0	2	10.0
	G9123A	0	0.0	0	0.0	1	1.6	8	66.7	9	16.1	7	77.8	20	100.0
ND3	T10238C	3	3.0	0	0.0	0	0.0	6	50.0	9	16.1	7	77.8	20	100.0
ND4	G11719A	57	56.4	49	100.0	62	100.0	12	100.0	56	100.0	9	100.0	20	100.0
TS2	C12239T	0	0.0	0	0.0	0	0.0	6	50.0	9	16.1	7	77.8	20	100.0
ND5	A14022G	0	0.0	0	0.0	0	0.0	0	0.0	8	14.3	7	77.8	20	100.0
Cyt b	C14766T	58	57.4	49	100.0	62	100.0	12	100.0	56	100.0	9	100.0	20	100.0
	A15326G	101	100.0	49	100.0	62	100.0	12	100.0	56	100.0	9	100.0	20	100.0
	A15746G	0	0.0	0	0.0	0	0.0	6	50.0	9	16.1	7	77.8	20	100.0

References

- Acton, B. M., I. Lai, X. Shang, A. Jurisicova and R. F. Casper (2007). "Neutral mitochondrial heteroplasmy alters physiological function in mice." *Biology of Reproduction* 77: 569-576.
- Anderson, S., A. T. Bankier, B. G. Barrell, M. H. L. de Bruijn, A. R. Coulson, J. Drouin, I. C. Eperon, D. P. Nierlich, B. A. Roe, F. Sanger, P. H. Schreier, A. J. H. Smith, R. Staden and I. G. Young (1981). "Sequence and organization of the human mitochondrial genome." *Nature* 290(5806): 457-465.
- Andrews, R. M., I. Kubacka, P. F. Chinnery, R. N. Lightowlers, D. M. Turnbull and N. Howell (1999). "Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA." *Nat Genet* 23(2): 147-147.
- Bai, R. K., S. M. Leal, D. Covarrubias, A. Liu and L. J. C. Wong (2007). "Mitochondrial genetic background modifies breast cancer risk." *Cancer Research* 67(10): 4687-4694.
- Bai, Y. D., J. S. Park, J. H. Deng, Y. F. Li and P. Q. Hu (2005). "Restoration of mitochondrial function in cells with complex I deficiency." *Role of the Mitochondria in Human Aging and Disease: From Genes to Cell Signaling* 1042: 25-35.
- Ballana, E., N. Govea, R. De Cid, C. Garcia, C. Arribas, J. Rosell and X. Estivill (2008). "Detection of unrecognized low-level mtDNA heteroplasmy may explain the variable phenotypic expressivity of apparently homoplasmic mtDNA mutations." *Human Mutation* 29(2): 248-257.
- Baughman, J. M. and V. K. Mootha (2006). "Buffering mitochondrial DNA variation." *Nature Genetics* 38(11): 1232-1233.
- Bellwood, P. (1991). "The Austronesian dispersal and the origin of languages." *Scientific American* 265(1): 88-93.
- Betts, M. J. and R. B. Russell (2003). Amino Acid Properties and Consequences of Substitutions. *Bioinformatics for Geneticists*. I. C. G. Michael R. Barnes: 289-316.
- Bhat, A., A. Koul, S. Sharma, E. Rai, S. I. A. Bukhari, M. K. Dhar and R. N. K. Bamezai (2007). "The possible role of 10398A and 16189C mtDNA variants in providing susceptibility to T2DM in two North Indian populations: a replicative study." *Human Genetics* 120: 821-826.
- Blakely, E. L., Mitchell, A. L., Fisher, N., Meunier, B., Nijtmans, L. G., Schaefer, A. M., Jackson, M. J., Turnbull, D. M. and Taylor, R. W. (2005). "A mitochondrial cytochrome *b* mutation causing severe respiratory chain enzyme deficiency in humans and yeast." *FEBS Journal* 272(14): 3583-3592.

- Brand, M. (2007). "Mitochondrial ROS production." *Comparative Biochemistry and Physiology a-Molecular & Integrative Physiology* 146(4): DOI 10.1016/j.cbpa.2007.01.044|12.
- Cadenas, E. (2004). "Mitochondrial free radical production and cell signaling." *Molecular Aspects of Medicine* 25(1-2): 17-26.
- Cann, R. L., and Lum, J.K. (2004). "Dispersal Ghosts in Oceania." *AMERICAN JOURNAL OF HUMAN BIOLOGY* 16: 440-451.
- Cann, R. L., W. M. Brown and A. C. Wilson (1984). "Polymorphic sites and the mechanism of evolution in human mitochondrial DNA." *Genetics* 106(3): 479-499.
- Cann, R. L., M. Stoneking and A. C. Wilson (1987). "Mitochondrial DNA and human evolution." *Nature* 325(6099): 31-36.
- Canter, J. A., A. R. Kallianpur, F. F. Parl and R. C. Millikan (2005). "Mitochondrial DNA G10398A polymorphism and invasive breast cancer in African-American women." *Cancer Research* 65(17): 8028-8033.
- Chambers, G. K., S. J. Marshall, G. M. Robinson, S. Maguire, J. Newton-Howes and N. L. Chong (2002). "The genetics of alcoholism in Polynesians: Alcohol and aldehyde dehydrogenase genotypes in young men." *Alcohol Clin. Exp. Res.* 26: 949-955.
- Chang, D. D. and D. A. Clayton (1985). "Priming of human mitochondrial-DNA replication occurs at the light-strand promoter." *Proceedings of the National Academy of Sciences of the United States of America* 82(2): 351-355.
- Chen, X. J. and R. A. Butow (2005). "The organization and inheritance of the mitochondrial genome." *Nature Reviews Genetics* 6(11): 815-825.
- Chinnery, P. F., H. R. Elliott, S. Patel, C. Lambert, S. M. Keers, S. E. Durham, M. I. McCarthy, G. A. Hitman, A. T. Hattersley and M. Walker (2005). "Role of the mitochondrial DNA 16184-16193 poly-C tract in type 2 diabetes." *Lancet* 366(9497): 1650-1651.
- Chinnery, P. F. and D. M. Turnbull (2000). "Mitochondrial DNA mutations in the pathogenesis of human disease." *Molecular Medicine Today* 6(11): 425-432.
- Davidov, Y., D. Huchon, S. F. Koval and E. Jurkevitch (2006). "A new alpha-proteobacterial clade of Bdellovibrio-like predators: implications for the mitochondrial endosymbiotic theory." *Environmental Microbiology* 8(12): 2179-2188.
- Diamond, J. (1988). "Express train to Polynesia." *Nature* 336: 307-308.
- DiMauro, S. and M. Mancuso (2007). "Mitochondrial diseases: Therapeutic approaches." *Bioscience Reports* 27(1-3): 125-137.

- Dimauro, S. and E. A. Schon (2001). "Mitochondrial DNA mutations in human disease." *American Journal of Medical Genetics* 106(1): 18-26.
- Dimroth, P., G. Kaim and U. Matthey (2000). "Crucial role of the membrane potential for ATP synthesis by F(1)F(o) ATP synthases." *J Exp Biol* 203(1): 51-59.
- Dobrowolski, S. F., J. Gray, T. Miller and M. Sears (2009). "Identifying sequence variants in the human mitochondrial genome using high-resolution melt (HRM) profiling." *Human Mutation* 30(6): 891-898.
- Eltsov, N., and Volodko, N. (2009). "mtPhyl - software tool for human mtDNA analysis and phylogeny reconstruction. ." from <http://eltsov.org/mtphyl.aspx>.
- Evans, J. L., I. D. Goldfine, B. A. Maddux and G. M. Grodsky (2003). "Are oxidative stress-activated signaling pathways mediators of insulin resistance and beta-cell dysfunction?" *Diabetes* 52(1): 1-8.
- Forster (2004). "Ice Ages and the mitochondrial DNA chronology of human dispersals: a review." *Phil. Trans. R. Soc. Lond. B* 359: 255-264.
- Francalacci, P., R. Montiel and A. Malgosa (1999). "A mitochondrial DNA database: applications to problems of nomenclature and to population genetics." In: *Genomic diversity: applications in human population genetics* (S.S. Papiha, R. Deka & R. Chakraborty Eds), Kluwer Academic / Plenum Publishing Corp, New York: pp 103-119.
- Fuku, N., K. S. Park, Y. Yamada, Y. Nishigaki, Y. M. Cho, H. Matsuo, T. Segawa, S. Watanabe, K. Kato, K. Yokoi, Y. Nozawa, H. K. Lee and M. Tanaka (2007). "Mitochondrial haplogroup N9a confers resistance against type 2 diabetes in Asians." *American Journal of Human Genetics* 80(3): 407-415.
- Gaster, M. (2007). "Insulin resistance and the mitochondrial link. Lessons from cultured human myotubes." *Biochimica Et Biophysica Acta-Molecular Basis of Disease* 1772(7): 755-765.
- Gaweda-Walerych, K., A. Maruszak, K. Safranow, M. Bialecka, G. Klodowska-Duda, K. Czyzewski, J. Slawek, M. Rudzinska, M. Styczynska, G. Opala, M. Drozdziak, J. A. Canter, M. Barcikowska and C. Zekanowski (2008). "Mitochondrial DNA haplogroups and subhaplogroups are associated with Parkinson's disease risk in a Polish PD cohort." *Journal of Neural Transmission* 115(11): 1521-1526.
- Ghezzi, D., C. Marelli, A. Achilli, S. Goldwurm, G. Pezzoli, P. Barone, M. T. Pellecchia, P. Stanzione, L. Brusa, A. R. Bentivoglio, U. Bonuccelli, L. Petrozzi, G. Abbruzzese, R. Marchese, P. Cortelli, D. Grimaldi, P. Martinelli, C. Ferrarese, B. Garavaglia, S. Sangiorgi, V. Carelli, A. Torroni, A. Albanese and M. Zeviani (2005). "Mitochondrial DNA haplogroup K is associated with a lower risk of Parkinson's disease in Italians." *European Journal of Human Genetics* 13(6): 748-752.

- Giles, R. E., H. Blanc, H. M. Cann and D. C. Wallace (1980). "Maternal inheritance of human mitochondrial-DNA." *Proceedings of the National Academy of Sciences* 77(11): 6715-6719.
- Gray, R. D., A. J. Drummond and S. J. Greenhill (2009). "Language Phylogenies Reveal Expansion Pulses and Pauses in Pacific Settlement." *Science* 323(5913): 479-483.
- Gray, R. D. and F. M. Jordan (2000). "Language trees support the express-train sequence of Austronesian expansion." *Nature* 405(6790): 1052-1055.
- Hagelberg, E. (1997). "Ancient and modern mitochondrial DNA sequences and the colonization of the pacific." *Electrophoresis* 18(9): 1529-1533.
- Hagelberg, E. and J. B. Clegg (1993). "Genetic Polymorphisms in Prehistoric Pacific Islanders Determined by Analysis of Ancient Bone DNA." *Proceedings of the Royal Society of London. Series B: Biological Sciences* 252(1334): 163-170.
- Hagelberg, E., Kayser, M., Nagy, M., Roewer, L., Zimdahl, H., Krawczak, M., Lio, P., and Schiefenhove, W. (1999). "Molecular genetic evidence for the human settlement of the Pacific: analysis of mitochondrial DNA, Y chromosome and HLA markers." *Phil.Trans. R. Soc. Lond.* 354: 141-152.
- Hall, D. (2009). "Genomic analysis of human populations." (David Hall PhD thesis, in preparation).
- Hall, D. A., G. K. Chambers and R. A. Lea (2007). "Haplotype analysis at the alcohol dehydrogenase gene region in New Zealand Maori." *Journal of Human Genetics* 52(2): 191-194.
- Hansford, R. (2002) "Oxidative Phosphorylation." *Encyclopedia of Life Sciences* DOI: 10.1038/npg.els.0001371.
- Harita, N., T. Hayashi, K. K. Sato, Y. Nakamura, T. Yoneda, G. Endo and H. Kambe (2009). "Lower Serum Creatinine Is a New Risk Factor of Type 2 Diabetes." *Diabetes Care* 32(3): 424-426.
- Herrmann, W. and R. Obeid (2008). "Causes and Early Diagnosis of Vitamin B-12 Deficiency." *Deutsches Arzteblatt International* 105(40): 680-U22.
- Hertzberg, M., K. N. Mickleson, S. W. Serjeantson et al. (1989). "An Asian-specific 9-bp deletion of mitochondrial DNA is frequently found in Polynesians." *Am. J. Hum. Genet.*(44): 504–510.
- Hobbie, S. N., C. M. Bruell, S. Akshay, S. K. Kalapala, D. Shcherbakov and E. C. Bottger (2008). "Mitochondrial deafness alleles confer misreading of the genetic code." *Proceedings of the National Academy of Sciences of the United States of America* 105(9): 3244-3249.
- Hofacker, I. L. (2003). "Vienna RNA secondary structure server." *Nucl. Acids Res.* 31(13): 3429-3431.

- Hoffman, S. and M. Bauer (2006) "Mitochondrial Disorders." Encyclopedia of Life Sciences DOI: 10.1038/npg.els.0005539.
- Horsefield, R., S. Iwata and B. Byrne (2004). "Complex II from a structural perspective." *Current Protein & Peptide Science* 5(2): 107-118.
- Houstis, N., E. D. Rosen and E. S. Lander (2006). "Reactive oxygen species have a causal role in multiple forms of insulin resistance." *Nature* 440(7086): 944-948.
- Howell, N., J. L. Elson, C. Howell and D. M. Turnbull (2007). "Relative Rates of Evolution in the Coding and Control Regions of African mtDNAs." *Mol Biol Evol* 24(10): 2213-2221.
- Hudson, M. L., A. L. M. Ahuriri-Driscoll, M. G. Lea and R. A. Lea (2007). "Whakapapa - A foundation for genetic research?" *Journal of Bioethical Inquiry* 4(1): 43-49.
- Ingman, M., and Gyllensten, U. (2001). "Analysis of the Complete Human mtDNA Genome: Methodology and Inferences for Human Evolution." *The American Genetic Association* 92: 454–461.
- Ingman, M., and Gyllensten, U. (2003). "Mitochondrial Genome Variation and Evolutionary History of Australian and New Guinean Aborigines." *Genome Research* 13: 1600-1606.
- Ingman, M., and Gyllensten, U. (2006). "mtDB: Human Mitochondrial Genome Database, a resource for population genetics and medical sciences." *Nucleic Acids Research* 34: D749–D751.
- Ingman, M., Kaessmann, H, Paabo, S., and Ulf Gyllensten, U. (2000). "Mitochondrial genome variation and the origin of modern humans." *Nature* 408: 708-713.
- Ivanov, P. L., M. J. Wadhams, R. K. Roby, M. M. Holland, V. W. Weedn and T. J. Parsons (1996). "Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of TsarNicholas II." *Nature Genetics* 12(4): 417-420.
- Jakupciak, J. P., A. Maggrah, S. Maragh, J. Maki, B. Reguly, K. Maki, R. Wittcock, K. Robinson, P. D. Wagner, R. E. Thayer, K. Gehman, T. Gehman, S. Srivastava, A. Ngom, G. D. Dakubo and R. L. Parr (2008). "Facile whole mitochondrial genome resequencing from nipple aspirate fluid using MitoChip v2.0." *Bmc Cancer* 8.
- Jeronimo, C., S. Nomoto, O. L. Caballero, H. Usadel, R. Henrique, G. Varzim, J. Oliveira, C. Lopes, M. S. Fliss and D. Sidransky (2001). "Mitochondrial mutations in early stage prostate cancer and bodily fluids." *Oncogene* 20(37): 5195-5198.

- Kake, T. R., R. Arnold and P. Ellis (2008). "Estimating the prevalence of schizophrenia among New Zealand Maori: a capture-recapture approach." *Australian and New Zealand Journal of Psychiatry* 42(11): 941 - 949.
- Kakkar, P. and B. K. Singh (2007). "Mitochondria: a hub of redox activities and cellular distress control." *Molecular and Cellular Biochemistry* 305(1-2): 235-253.
- Kayser, M., S. Brauer, G. Weiss, P. A. Underhill, L. Roewer, W. Schiefenhover and M. Stoneking (2000). "Melanesian origin of Polynesian Y chromosomes." *Current Biology* 10(20): 1237-1246.
- Khogali, S. S., B. M. Mayosi, J. M. Beattie, W. J. McKenna, H. Watkins and J. Poulton (2001). "A common mitochondrial DNA variant associated with susceptibility to dilated cardiomyopathy in two different populations." *The Lancet* 357(9264): 1265-1267.
- Kiefer, F., K. Arnold, M. Kunzli, L. Bordoli and T. Schwede (2009). "The SWISS-MODEL Repository and associated resources." *Nucl. Acids Res.* 37(suppl_1): D387-392.
- Kim, J. H., K. S. Park, Y. M. Cho, B. S. Kang, S. K. Kim, H. J. Jeon, S. Y. Kim and H. K. Lee (2002). "The prevalence of the mitochondrial DNA 16189 variant in non-diabetic Korean adults and its association with higher fasting glucose and body mass index." *Diabetic Medicine* 19(8): 681-684.
- Kimchi-Sarfaty, C., J. M. Oh, I.-W. Kim, Z. E. Sauna, A. M. Calcagno, S. V. Ambudkar and M. M. Gottesman (2007). "A "Silent" Polymorphism in the MDR1 Gene Changes Substrate Specificity." *Science* 315(5811): 525-528.
- Kirch, P. V. and T. L. Hunt (1988). "Radiocarbon-dates from the Mussau Islands and the Lapita colonization of the Southwestern Pacific." *Radiocarbon* 30(2): 161-169.
- Komar, A. A. (2007). "GENETICS: SNPs, Silent But Not Invisible." *Science* 315(5811): 466-467.
- Kopp, J. and T. Schwede (2004). "The SWISS-MODEL Repository of annotated three-dimensional protein structure homology models." *Nucl. Acids Res.* 32(suppl_1): D230-234.
- Lambert, D. M., P. A. Ritchie, C. D. Millar, B. Holland, A. J. Drummond and C. Baroni (2002). "Rates of evolution in ancient DNA from Adelie penguins." *Science* 295(5563): 2270-2273.
- Lea, R., N. Benowitz, M. Green, J. Fowles, A. Vishvanath, S. Dickson, M. Lea, A. Woodward, G. Chambers and D. Phillips (2005). "Ethnic differences in nicotine metabolic rate among New Zealanders." *N Z Med J* 118(1227): U1773.
- Lea, R. A. and G. K. Chambers (2007). "Pharmacogenetics in Admixed Polynesian Populations." Chapter in *Pharmacogenetics and Admixed populations*. Edited

- by: Guilherme Suarez-Kurtz. Eureka Bioscience Database (ISBN: 1-58706-130-9).
- Lea, R. A., R. L. Roberts, M. R. Green, M. A. Kennedy and G. K. Chambers (2008). "Allele frequency differences of cytochrome P450 polymorphisms in a sample of New Zealand Maori." *N Z Med J* 121(1272): 33-7.
- Liao, W. Q., Y. Pang, C. A. Yu, J. Y. Wen, Y. G. Zhang and X. H. Li (2008). "Novel mutations of mitochondrial DNA associated with type 2 diabetes in Chinese Han population." *Tohoku Journal of Experimental Medicine* 215(4): 377-384.
- Lin, T. K., S. D. Chen, P. W. Wang, Y. H. Wei, C. F. Lee, T. L. Chen, Y. C. Chuang, T. Y. Tan, K. C. Chang and C. W. Liou (2005). "Increased oxidative damage with altered antioxidative status in type 2 diabetic patients harboring the 16189 T to C variant of mitochondrial DNA." *Role of the Mitochondria in Human Aging and Disease: From Genes to Cell Signaling* 1042: 64-69.
- Lin, Y., A. H. Berg, P. Iyengar, T. K. T. Lam, A. Giacca, T. P. Combs, M. W. Rajala, X. L. Du, B. Rollman, W. J. Li, M. Hawkins, N. Barzilai, C. J. Rhodes, I. G. Fantus, M. Brownlee and P. E. Scherer (2005). "The hyperglycemia-induced inflammatory response in adipocytes - The role of reactive oxygen species." *Journal of Biological Chemistry* 280(6): 4617-4626.
- Liou, C. W., T. K. Lin, H. H. Weng, C. F. Lee, T. L. Chen, Y. H. Wei, S. D. Chen, Y. C. Chuang, S. W. Weng and P. W. Wang (2007). "A common mitochondrial DNA variant and increased body mass index as associated factors for development of type 2 diabetes: Additive effects of genetic and environmental factors." *Journal of Clinical Endocrinology and Metabolism* 92(1): 235-239.
- Lister, R., J. M. Hulett, T. Lithgow and J. Whelan (2005). "Protein import into mitochondria: origins and functions today (Review)." *Molecular Membrane Biology* 22(1-2): 87-100.
- Liu, C. S., W. L. Cheng, Y. Y. Chen, Y. S. Ma, C. Y. Pang and Y. H. Wei (2005). "High prevalence of the COII/tRNA(Lys) intergenic 9-bp deletion in mitochondrial DNA of Taiwanese patients with MELAS or MERRF syndrome." *Role of the Mitochondria in Human Aging and Disease: From Genes to Cell Signaling* 1042: 82-87.
- Liu, V. W. S., Y. Wang, H.-J. Yang, P. C. K. Tsang, T.-Y. Ng, L.-C. Wong, P. Nagley and H. Y. S. Ngan (2003). "Mitochondrial DNA variant 16189T→C is associated with susceptibility to endometrial cancer." *Human Mutation* 22(2): 173-174.
- Livesey, K. J., V. L. C. Wilmhurst, K. Carter, M. Worwood, E. Cadet, J. Rochette, A. G. Roberts, J. J. Pointon, A. T. Merryweather-Clarke, M. L. Bassett, A. M. Jouanolle, A. Mosser, V. David, J. Poulton and K. J. H. Robson (2004). "The 16189 variant of mitochondrial DNA occurs more frequently in C282Y homozygotes with haemochromatosis than those without iron loading.(Original Article)." *Journal of Medical Genetics* 41(1): 6(5).

- Lu, S. Y., S. Nishio, K. Tsukada, T. Oguchi, K. Kobayashi, S. Abe and S. Usami (2009). "Factors that affect hearing level in individuals with the mitochondrial 1555A > G mutation." *Clinical Genetics* 75(5): 480-484.
- Lum, J. K., and Cann, R.L. (1998). "mtDNA and Language Support a Common Origin of Micronesians and Polynesians in Island Southeast Asia." *AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY* 105: 109–119.
- Lum, J. K., and Cann, R.L. (2000). "mtDNA Lineage Analyses: Origins and Migrations of Micronesians and Polynesians." *AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY* 113: 151-168.
- Lum, J. K., O. Rickards, C. Ching and R. L. Cann (1994). "Polynesian mitochondrial DNAs reveal three deep maternal lineage clusters." *Human Biology* v66(n4): p567(24).
- M. Esther Gallardo, R. M.-L., Celia López, Mercedes Casqueiro, Javier Silva, Felix Bonilla, Santiago Rodríguez de Córdoba, José Antonio Enríquez, (2006). "m.6267G>A: a recurrent mutation in the human mitochondrial DNA that reduces cytochrome c oxidase activity and is associated with tumors." *Human Mutation* 27(6): 575-582.
- Maassen, J. A., L. M. Hart, G. M. C. Janssen, E. Reiling, J. A. Romijn and H. H. Lemkes (2006). "Mitochondrial diabetes and its lessons for common Type 2 diabetes." *Biochemical Society Transactions* 34: 819-823.
- Macaulay V., H. C., Achilli A., Rengo C., Clarke D., et al. (2005). "Single, Rapid Coastal Settlement of Asia Revealed by Analysis of Complete Mitochondrial Genomes." *Science* 308: 1034-1036.
- Maechler, P. and P. B. M. de Andrade (2006). "Mitochondrial damages and the regulation of insulin secretion." *Biochemical Society Transactions* 34: 824-827.
- Maitra, A., Y. Cohen, S. E. D. Gillespie, E. Mambo, N. Fukushima, M. O. Hoque, N. Shah, M. Goggins, J. Califano, D. Sidransky and A. Chakravarti (2004). "The human MitoChip: A high-throughput sequencing microarray for mitochondrial mutation detection." *Genome Research* 14(5): 812-819.
- Malmgren, S., D. G. Nicholls, J. Taneera, K. Bacos, T. Koeck, A. Tamaddon, R. Wibom, L. Groop, C. Ling, H. Mulder and V. V. Sharoyko (2009). "Tight Coupling between Glucose and Mitochondrial Metabolism in Clonal beta-Cells Is Required for Robust Insulin Secretion." *Journal of Biological Chemistry* 284(47): 32395-32404.
- Melton, T., R. Peterson, A. J. Redd, N. Saha, A. S. M. Sofro, J. Martinson and M. Stoneking (1995). "Polynesian Genetic Affinities with Southeast Asian Populations as Identified by mtDNA Analysis." *The American Society of Human Genetics* 57: 403 -414.

- Melton, T., S. Clifford, J. Martinson et al. (1998). "Genetic evidence for the Proto-Austronesian homeland in Asia: mtDNA and nuclear DNA variation in Taiwanese aboriginal tribes." *Am. J. Hum. Genet.* 63: 1807–1823.
- Meselson, M. and R. Yuan (1968). "DNA restriction enzyme from E coli." *Nature* 217(5134): 1110-&.
- Mishmar, D., E. Ruiz-Pesini, P. Golik, V. Macaulay, A. G. Clark, S. Hosseini, M. Brandon, K. Easley, E. Chen, M. D. Brown, R. I. Sukernik, A. Olckers and D. C. Wallace (2003). "Natural selection shaped regional mtDNA variation in humans." *Proceedings of the National Academy of Sciences of the United States of America* 100(1): 171-176.
- Moreira, P. I., M. S. Santos, R. Seica and C. R. Oliveira (2007). "Brain mitochondrial dysfunction as a link between Alzheimer's disease and diabetes." *Journal of the Neurological Sciences* 257(1-2): 206-214.
- Moreno-Loshuertos, R., R. Acin-Perez, P. Fernandez-Silva, N. Movilla, A. Perez-Martos, S. R. de Cordoba, M. E. Gallardo and J. A. Enriquez (2006). "Differences in reactive oxygen species production explain the phenotypes associated with common mouse mitochondrial DNA variants." *Nat Genet* 38(11): 1261-1268.
- Murray-McIntosh, R. P., B. J. Scrimshaw, P. J. Hatfield and D. Penny (1998). "Testing migration patterns and estimating founding population size in Polynesia by using human mtDNA sequences." *Proc. Natl. Acad. Sci. USA* 95: 9047-9052.
- Navaglia, F., D. Basso, P. Fogar, C. Sperti, E. Greco, C. F. Zambon, A. Stranges, A. Falda, S. Pizzi, A. Parenti, S. Pedrazzoli and M. Plebani (2006). "Mitochondrial DNA D-loop in pancreatic cancer: Somatic mutations are epiphenomena while the germline 16519 T variant worsens metabolism and outcome." *American Journal of Clinical Pathology* 126(4): 593-601.
- Nishigaki, Y., Y. Yamada, N. Fuku, H. Matsuo, T. Segawa, S. Watanabe, K. Kato, K. Yokoi, S. Yamaguchi, Y. Nozawa and M. Tanaka (2007). "Mitochondrial haplogroup N9b is protective against myocardial infarction in Japanese males." *Human Genetics* 120(6): 827-836.
- Oppenheimer, S. J., and Richards, M. (2001). "Slow boat to Melanesia?" *Nature* 410: 166-167.
- Pericak-Vance, M. A., J. M. van der Walt, K. K. Nicodemus, E. R. Martin, C. C. Kroner, W. K. Scott, D. E. Schmechel, G. W. Small, P. M. Conneally, A. M. Saunders, J. Haines, J. R. Gilbert and J. M. Vance (2002). "Mitochondrial effect on risk of developing late-onset Alzheimer disease (AD) is modified by gender." *American Journal of Human Genetics* 71(4): 501-501.
- Petersen, K. F., D. Befroy, S. Dufour, J. Dziura, C. Ariyan, D. L. Rothman, L. DiPietro, G. W. Cline and G. I. Shulman (2003). "Mitochondrial dysfunction in the elderly: Possible role in insulin resistance." *Science* 300(5622): 1140-1142.

- Petros, J. A., A. K. Baumann, E. Ruiz-Pesini, M. B. Amin, C. Q. Sun, J. Hall, S. Lim, M. M. Issa, W. D. Flanders, S. H. Hosseini, F. F. Marshall and D. C. Wallace (2005). "mtDNA mutations increase tumorigenicity in prostate cancer." *Proceedings of the National Academy of Sciences of the United States of America* 102(3): 719-724.
- Pierson, M. J., Martinez-Arias, R., Holland, B.R., Gemmell, N.J., Hurles, M.E, and Penny, D. (2006). "Deciphering Past Human Population Movements in Oceania: Provably Optimal Trees of 127 mtDNA Genomes." *Society for Molecular Biology and Evolution*.
- Poulton, J. (1998). "Does a common mitochondrial DNA polymorphism underlie susceptibility to diabetes and the thrifty genotype?" *Trends in Genetics* 14(10): 387-389.
- Poulton, J., A. L. Bednarz, M. Scott-Brown, C. Thompson, V. A. Macaulay and D. Simmons (2002). "The presence of a common mitochondrial DNA variant is associated with fasting insulin levels in Europeans in Auckland." *Diabetic Medicine* 19(11): 969-971.
- Poulton, J., M. S. Brown, A. Cooper, D. R. Marchington and D. I. W. Phillips (1998). "A common mitochondrial DNA variant is associated with insulin resistance in adult life." *Diabetologia* 41(1): 54-58.
- Poulton, J., J. Luan, V. Macaulay, S. Hennings, J. Mitchell and N. J. Wareham (2002). "Type 2 diabetes is associated with a common mitochondrial variant: evidence from a population-based case-control study." *Human Molecular Genetics* 11(13): 1581-1583.
- Prezant, T. R., J. V. Agapian, M. C. Bohlman, X. D. Bu, S. Oztas, W. Q. Qiu, K. S. Arnos, G. A. Cortopassi, L. Jaber, J. I. Rotter, M. Shohat and N. Fischelghodsian (1993). "Mitochondrial ribosomal-RNA mutation associated with both antibiotic-induced and non-syndromic deafness." *Nature Genetics* 4(3): 289-294.
- Puoti, G., F. Carrara, S. Sampaolo, M. De Caro, C. M. Vincitorio, F. Invernizzi and M. Zeviani (2003). "Identical large scale rearrangement of mitochondrial DNA causes Kearns-Sayre syndrome in a mother and her son." *Journal of Medical Genetics* 40(11): 858-863.
- Ray, A. M., K. A. Zuhlke, A. M. Levin, J. A. Douglas, K. A. Cooney and J. A. Petros (2009). "Sequence variation in the mitochondrial gene cytochrome c oxidase subunit I and prostate cancer in African American men." *The Prostate* 69(9): 956-960.
- Redd, A., J. N. Takezake and S. T. Sherry (1995). "Evolutionary history of the COII/tRNA lys intergenic 9 base pair deletion in human mitochondrial DNAs from the Pacific." *Mol. Biol. Evol.* 12: 604–615.
- Rollins, B., M. V. Martin, P. A. Sequeira, E. A. Moon, L. Z. Morgan, S. J. Watson, A. Schatzberg, H. Akil, R. M. Myers, E. G. Jones, D. C. Wallace, W. E. Bunney

- and M. P. Vawter (2009). "Mitochondrial variants in schizophrenia, bipolar disorder, and major depressive disorder." *PLoS ONE* 4(3): e4913.
- Ross, O. A., R. McCormack, M. D. Curran, R. Alistair Duguid, Y. A. Barnett, I. Maeve Rea and D. Middleton (2001). "Mitochondrial DNA polymorphism: its role in longevity of the Irish population." *Experimental Gerontology* 36(7): 1161-1178.
- Ruiz-Pesini, E., Lott, M.T., Procaccio, V., Poole, J., Brandon, M.C., Mishmar, D., Yi, C., Kreuziger, J., Baldi, P., and Wallace, D.C. (2007). "An enhanced MITOMAP with a global mtDNA mutational phylogeny." *Nucleic Acids Research* 35 (Database issue):D823-D828.
- Ruiz-Pesini, E., D. Mishmar, M. Brandon, V. Procaccio and D. C. Wallace (2004). "Effects of purifying and adaptive selection on regional variation in human mtDNA." *Science* 303(5655): 223-226.
- Ryan, M. T. (2005). "Mitochondria Protein Import: Methods." *Encyclopedia of Life Sciences*: 1-3.
- Santoro, A., E. Balducci, F. Rosini, A. Montesanto, S. Salvioli, G. Rose, E. Bellavista, M. Mishto, G. Gasparre, E. Bonora, A. Chiamenti, G. Romeo, C. Gabelli, G. Crepaldi, T. Grune, G. De Benedictis and C. Franceschi (2006). "mtDNA inherited and somatic variability in Alzheimer's Disease." *Free Radical Research* 40: S46-S46.
- Schultz, B. E. and S. I. Chan (2001). "Structures and proton-pumping strategies of mitochondrial respiratory enzymes." *Annual Review of Biophysics and Biomolecular Structure* 30(1): 23-65.
- Schwartz, M. and J. Vissing (2002). "Paternal inheritance of mitochondrial DNA." *New England Journal of Medicine* 347(8): 576-580.
- Soares, P., J. A. Trejaut, J. H. Loo, C. Hill, M. Mormina, C. L. Lee, Y. M. Chen, G. Hudjashov, P. Forster, V. Macaulay, D. Bulbeck, S. Oppenheimer, M. Lin and M. B. Richards (2008). "Climate change and postglacial human dispersals in Southeast Asia." *Molecular Biology and Evolution* 25(6): 1209-1218.
- Stannard, S. R. and N. A. Johnson (2004). "Insulin resistance and elevated triglyceride in muscle: more important for survival than 'thrifty' genes?" *Journal of Physiology-London* 554(3): 595-607.
- Stoneking, M. (2000). "Hypervariable Sites in the mtDNA Control Region Are Mutational Hotspots." 67(4): 1029-1032.
- Subramanian, S., D. R. Denver, C. D. Millar, T. Heupink, A. Aschrafi, S. D. Emslie, C. Baroni and D. M. Lambert (2009). "High mitogenomic evolutionary rates and time dependency." *Trends in Genetics* 25(11): 482-486.
- Sykes, B., A. Leiboff, J. Lowbeer, S. Tetzner and M. Richards (1995). "The origins of the Polynesians - An interpretation from mitochondrial lineage analysis." *American Journal of Human Genetics* 57(6): 1463-1475.

- Tachibana, M., M. Sparman, H. Sritanaudomchai, H. Ma, L. Clepper, J. Woodward, Y. Li, C. Ramsey, O. Kolotushkina and S. Mitalipov (2009). "Mitochondrial gene replacement in primate offspring and embryonic stem cells." *Nature* 461(7262): 367-372.
- Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). "MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0." *Molecular Biology and Evolution* 10.1093/molbev/msm092.
- Tanaka, M., N. Fuku, Y. Nishigaki, H. Matsuo, T. Segawa, S. Watanabe, K. Kato, K. Yoko, M. Ito, Y. Nozawa and Y. Yamada (2007). "Women with mitochondrial haplogroup N9a are protected against metabolic syndrome." *Diabetes* 56(2): 518-521.
- Taubes, G. (2009). "Prosperity's Plague." *Science* 325(5938): 256-260.
- Terrell, J. (1988). "History as a family tree, history as an entangled bank - constructing images and interpretations of prehistory in the South-Pacific." *Antiquity* 62(237): 642-657.
- Torroni, A., A. Achilli, V. Macaulay, M. Richards and H.-J. Bandelt (2006). "Harvesting the fruit of the human mtDNA tree." *Trends in Genetics* 22(6): 339-345.
- Trejaut, J. A., T. Kivisild, J. H. Loo, C. L. Lee, C. L. He, C. J. Hsu, Z. Y. Lee and M. Lin (2005). "Traces of archaic mitochondrial lineages persist in Austronesian-speaking Formosan populations." *PLoS Biol* 3(8): 1-11.
- Vallone, P. M., J. P. Jakupciak and M. D. Coble (2007). "Forensic application of the affymetrix human mitochondrial resequencing array." *Forensic Science International-Genetics* 1(2): 196-198.
- Van der Walt, J., K. K. Nicodemus, E. R. Martin, W. K. Scott, M. A. Nance, J. P. Hubble, J. L. Haines, W. C. Koller, R. Pahwa, M. Stern, B. C. Hiner, J. Jankovic, C. G. Goetz, G. W. Small, F. Mastaglia, R. Gibson, L. T. Middleton, A. C. McLaurin, M. A. Pericak-Vance and J. M. Vance (2002). "Specific mitochondrial haplogroups affect risk for late-onset Parkinson disease." *American Journal of Human Genetics* 71(4): 364-364.
- Van der Walt, J. M., Y. A. Dementieva, E. R. Martin, W. K. Scott, K. K. Nicodemus, C. C. Kroner, K. A. Welsh-Bohmer, A. M. Saunders, A. D. Roses, G. W. Small, D. E. Schmechel, P. M. Doraiswamy, J. R. Gilbert, J. L. Haines, J. M. Vance and M. A. Pericak-Vance (2004). "Analysis of European mitochondrial haplogroups with Alzheimer disease risk." *Neuroscience Letters* 365(1): 28-32.
- Van der Walt, J. M., K. K. Nicodemus, E. R. Martin, W. K. Scott, M. A. Nance, R. L. Watts, J. P. Hubble, J. L. Haines, W. C. Koller, K. Lyons, R. Pahwa, M. B. Stern, A. Colcher, B. C. Hiner, J. Jankovic, W. G. Ondo, F. H. Allen, C. G. Goetz, G. W. Small, F. Mastaglia, J. M. Stajich, A. C. McLaurin, L. T. Middleton, B. L. Scott, D. E. Schmechel, M. A. Pericak-Vance and J. M. Vance

- (2003). "Mitochondrial polymorphisms significantly reduce the risk of Parkinson disease." *American Journal of Human Genetics* 72(4): 804-811.
- van Oven, M. and M. Kayser (2009). "Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation." *Human Mutation* 30(2): E386-E394.
- Venter, J. C., M. D. Adams, E. W. Myers, *et al.*, (2001). "The sequence of the human genome." *Science* 291(5507): 1304-+.
- Vigilant, L., M. Stoneking, H. Harpending, K. Hawkes and A. Wilson (1991). "African populations and the evolution of human mitochondrial DNA." *Science* 253(5027): 1503-1507.
- Voet, D. and J. G. Voet (2004). *Biochemistry*, Third Edition. U.S., Wiley International Edition.
- Waikan, Y. and A. M. Dozy (1978). "Polymorphism of DNA-sequence adjacent to human beta-globin structural gene - relationship to sickle mutation." *Proceedings of the National Academy of Sciences of the United States of America* 75(11): 5631-5635.
- Wallace, D. C. (2006). "A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: A dawn for evolutionary medicine." *Faseb Journal* 20(5): A1474-A1474.
- Weng, S. W., C. W. Liou, T. K. Lin, Y. H. Wei, C. F. Lee, H. L. Eng, S. D. Chen, R. T. Liu, J. F. Chen, I. Y. Chen, M. H. Chen and P. W. Wang (2005). "Association of mitochondrial deoxyribonucleic acid 16189 variant (T -> C transition) with metabolic syndrome in Chinese adults." *Journal of Clinical Endocrinology and Metabolism* 90(9): 5037-5040.
- Whyte, A. L. H., S. J. Marshall and G. K. Chambers (2005). "Human evolution in Polynesia." *Human Biology* 77(2): 157-177.
- Wiederkehr, A. and C. B. Wollheim (2006). "Minireview: Implication of mitochondria in insulin secretion and action." *Endocrinology* 147(6): 2643-2649.
- Zhou, S. Y., K. Kassauei, D. J. Cutler, G. C. Kennedy, D. Sidransky, A. Maitra and J. Califano (2006). "An oligonucleotide microarray for high-throughput sequencing of the mitochondrial genome." *Journal of Molecular Diagnostics* 8(4): 476-482.