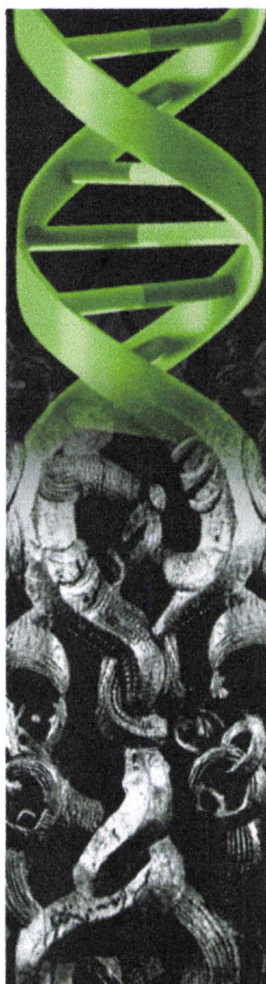


Human Evolution in Polynesia: A Molecular Biological Study



Adele Lorene Hineao Whyte

Submitted in partial fulfilment of the requirements of Master of Science in Genetics and Molecular Biology



Victoria University
of Wellington, New Zealand
*Te Whare Wananga o te
Upoko o te Ika a Maui
Aotearoa*

2002

Abstract

Human evolution is an extremely interesting and contentious topic that incorporates data from a wide variety of disciplines. Molecular studies are becoming increasingly important for reconstructing human history, as new techniques allow faster recovery of results, and genetic tests provide an independent test for colonisation theories that are usually based on archaeological and linguistic evidence. Mitochondrial DNA (mtDNA) is an extremely useful genetic marker that is widely used in molecular biology for establishing phylogenetic links between individuals and populations.

In the current study the 3' Hypervariable Region (3' HVI) of mtDNA from human populations living in New Zealand (Aotearoa) was analysed. The HVI mtDNA diversity in the New Zealand Maori population was significantly reduced compared with Polynesian and Melanesian populations, with an extremely high frequency (~87%) of the 'Polynesian CGT motif' haplotype. Additionally, the 9-bp deletion, common in Polynesian populations, has reached fixation (100%) in the Maori samples. These results support a settlement scenario with repeated population bottlenecks.

The mtDNA HVI haplotype frequencies in this study, combined with those from previous studies, were used in computer simulations to estimate the number of females required to found the current Maori population. Approximately 56 women, and an equal or greater number of men, were estimated to have been present in the founding waka. This estimate is too large to support any settlement models with a small number of founders and effectively rules out the possibility of 'accidental discovery', instead supporting a planned settlement of Aotearoa in agreement with traditional knowledge.

Analysis of interdisciplinary data has allowed current theories for the origins of the Polynesians and proto-Austronesians to be consolidated by introducing a ‘Synthetic Total Evidence Theory’. It is likely that once published this theory will stimulate intense discussion and debate and will continue to grow as further evidence is obtained and incorporated into this model.

Acknowledgements

I would like to thank the following people who have provided support, encouragement, and expertise throughout the years:

My supervisor Dr Geoff Chambers for his advice, editing, and encouragement throughout the past 2 years.

Dr Elizabeth MacAvoy, for her support, expertise in the laboratory techniques and helpful advice.

My fiancée Jason Allen, for his support, both emotional and financial. Thank you so much for believing in me, pushing me, bullying me and dragging me in the end to get this done. I couldn't have done it without you.

Shirley Pledger for helping me with some of the Stats stuff when no one else would.

Stephen Marshall for doing all the computer simulations (how do people work with computers?) and offering encouragement and advice. Also Professor David Penny (Massey University) for providing the original computer programme and advice.

George Andrews (George Andrews Productions) for his support of this project and the use of the graphic shown on the cover page of this thesis.

The financial support provided by the Foundation for Research, Science and Technology (FRST), in the form of a Tuapapa Putaiao Maori Fellowship (Contract #VUWX9902). This allowed me to concentrate on my studies without having to work part-time, keeping me off the breadline, and giving me free time (and money) outside of my research to keep sane.

The Victoria University of Wellington (VUW) Science Faculty (SciFac) and School of Biological Sciences (SBS) for providing financial assistance to purchase consumables required for this research. Also the associated staff for encouragement!

The students associated with the Institute for Molecular Sciences (IMS) facility within SBS, including but not limited to, James Sainsbury, Lorna McGibbon, Kathy Hill and Matthew Chan. Thanks for all the good times hanging out in the computer lab, when we should've been doing work elsewhere ☺

Special thanks to my office mate and good friend Jess who has been a constant source of yummy sugar snacks (is she trying to fatten me up?), and a comforting person to turn to for advice throughout all this madness. Jess you are a great friend and I really appreciate you! Thanks also to Nick Wong (former office mate) for all his support and journal articles sent from Melbourne. We sure are the 3 “Asses” – and it's been great getting to know you both.

Special thanks to my other non-science friends who remind me that there is 'occasionally' life outside of Science!

Thanks to Liz Richardson and my family on campus *Awhina*, I've had a wonderful time being a mentor, and look forward to being involved further in the future.

Finally a great big thanks to all my whanau back in Hawke's Bay who have been instrumental in my decision to do my Masters, and my decision to continue further study with my PhD. Thanks particularly to my Mum and Dad who have instilled in me the will to succeed and do well. Also thanks to my younger brothers, Daniel, James and Jason, as while I strive to inspire them and be a good role model, they have provided inspiration for me too! They also make me work harder with the thought of them nipping at my heels to outdo me! I will continue to work hard, for the benefit of all and to make our ancestors proud. Whaia Te Iti Kahurangi - To the stars through hard work!

Table of Contents

Abstract.....	ii
Acknowledgements.....	iv
Table of Contents.....	vi
List of Figures.....	xi
List of Tables.....	xiii
Abbreviations.....	xiv
 1. Introduction.....	 1
1.1 Why study Human Evolution?	1
1.2 Introducing Oceania	4
1.2.1 Historical definition of Oceania	6
1.2.2 Oral History of Oceania	7
1.3 Small population effects in Pacific People	9
1.3.1 Founder Effect	9
1.3.2 Bottleneck	11
1.4 Interdisciplinary view of Polynesians	12
1.4.1 Archaeology: Lapita Culture	12
1.4.2 Linguistics: Austronesian Language	16
1.4.3 Anthropology	21
1.4.4 Human Population Genetics	21
1.4.4.1 Blood groups	22
1.4.4.2 HLA	22
1.4.4.3 Globin genes	23
1.4.5 Minisatellites	23
1.4.6 Mitochondrial DNA	24
1.4.6.1 Mitochondrial genome structure	26
1.4.6.2 Variable sites	27
1.4.6.3 Restriction analysis	28

1.4.6.4	Hypervariable I Region (HVI)	29
1.4.6.4.1	CGT motif	29
1.4.6.4.2	Homopolymeric cytosine tract	30
1.4.6.5	9 base-pair (9-bp) deletion (Region V)	31
1.4.6.6	Hypervariable Region II (HVII)	32
1.4.6.7	Heteroplasmy	33
1.4.6.8	Ancient DNA	35
1.4.6.9	Use of Mitochondrial DNA in Forensic Science	37
1.4.6.9.1	Disaster Victim Identification (DVI)	38
	- An Australasian example	
1.4.6.9.2	Missing Person Identification	39
	(Australia)	
1.4.7	Y-Chromosome studies	39
1.5	Models of the Oceanic Expansion of people across the Pacific	40
1.5.1	Express Train Model (Bellwood, 1991)	41
1.5.2	Entangled Bank Model (Terrell, 1988)	42
1.5.3	The Slow Boat Model (Oppenheimer and Richards, 2001)	43
1.6	Pacific Colonisation	44
1.7	Research Objectives	46
2.	Materials and Methods.....	47
2.1	Source of DNA	47
2.2	DNA Extraction Protocol	48
2.2.1	Recovery of white blood cells from whole blood	48
2.2.2	Proteinase K Digestion	49
2.2.3	Phenol/Chloroform Extraction	49
2.2.4	Ethanol Precipitation of Extracted DNA	50
2.3	Mitochondrial DNA Target Amplification	50
2.3.1	Primers	50
2.3.2	PCR Protocol	51
2.4	Submarine Agarose Gel Electrophoresis	52
2.5	Detection of the 9-bp deletion using a 4% NuSieve® gel	53
2.6	Purification of PCR reactions using Low Melting point (LMP) gel	53

2.7	Purification of excised amplified dsDNA products	54
2.7.1	Prep-a-Gene	54
2.7.2	Concert™ Rapid Gel Extraction	55
2.7.3	MinElute™ spin columns	55
2.8	Molecular weight and concentration of purified DNA	56
2.9	DNA Cycle Sequencing Protocol	57
2.10	Ethanol Precipitation of DNA	58
2.11	Automated DNA Sequencing Gel	58
2.12	Data Analysis	60
2.12.1	Determining suitable parameters for analysis of the data used in the founding population simulations	60
2.12.2	Haplotype Identification	62
2.12.3	Estimate of the number of Maori females required to found the New Zealand's (Aotearoa) Maori population	64
3.	Results.....	66
3.1	DNA Extraction	66
3.2	Mitochondrial DNA 9-bp deletion Target Amplification	66
3.3	Mitochondrial DNA HVI Target Amplification	68
3.4	Mitochondrial DNA HVII Target Amplification	71
3.5	Mitochondrial DNA Sequencing	72
3.5.1	Presence or Absence of the 9-bp Deletion	74
3.5.2	Mitochondrial DNA HVI Sequence	75
3.5.3	Mitochondrial DNA HVII Sequence	77
3.6	Population Simulation Data Analysis	85
3.6.1	Capture-Recapture Estimates	85
3.6.2	Population growth simulation to estimate the size of the Maori founding population	86
4.	Discussion.....	90
4.1	DNA Extraction	90
4.2	Mitochondrial 9-bp, HVI and HVII Amplification	91
4.3	Sequence Analysis	93

4.3.1	9-bp deletion assignment	93
4.3.2	Mitochondrial HVI sequence	94
4.3.3	Mitochondrial HVII sequence	97
4.4	Number of Maori female founders to settle New Zealand (Aotearoa)	98
4.4.1	Determining suitable parameters for analysis of the data used in the founding population simulations	99
A	What region(s) of the mitochondrial sequence should be included in the simulation dataset?	99
B	Which dataset should be used for the Maori founder analysis?	100
C	Which Eastern Polynesian (EP) Haplotypes should be used for the frequency comparison?	100
D	Is the dataset based on decisions A-C suitable for analysis (see Table 3.3)?	101
	▪ New Mutations	102
	▪ The Discovery Curve Assumption	103
E	What Population Growth Model should be used?	107
4.4.2	The Simulation	108
4.5	Source of Maori Female Founders to settle Aotearoa	111
4.5.1	Theories for the colonisation of Aotearoa	111
4.5.2	Source of the Maori founders	113
4.6	Origin of Eastern Polynesian Populations: Introducing the ‘Synthetic Total Evidence Theory’	115
4.6.1	The ‘Synthetic Total Evidence Theory’	117
4.6.2	New mtDNA data support for the ‘Synthetic Total Evidence Theory’	125
4.6.3	Archaeology	128
	4.6.3.1 Agricultural evidence	128
	4.6.3.2 Lapita Culture	130
4.6.4	Linguistics	131
4.6.5	Y Chromosome markers	132
4.6.6	ADH and ALDH data	135

5.	Summary and Conclusions.....	136
5.1	Synopsis of Major Findings	136
5.2	Conclusions	138
	Future Investigations.....	139
	Appendices.....	142
Appendix One :	Solutions	142
Appendix Two:	Population Growth Models used for estimating the number of Female founders for the current Maori population	144
Appendix Three:	Complete 9-BP/HVI (16 024-16, 400) haplotype for each sample	148
	HVII (00 073-00 340) haplotype for selected samples	155
Appendix Four:	Electrophoresis of HVII agarose gel and HVII Calibration Curve	156
Appendix Five:	Combined Haplotype data for Maori, Eastern and Western Polynesian samples (i.e. EP+)	158
Appendix Six:	Total number of Haplotypes found in combined MM98 and AW02 data combining haplotypes that differ before position 16 189	161
Appendix Seven:	Discovery curves for the Maori and Eastern Polynesian data	162
Appendix Eight:	Comparison of total data with 'unequal' distribution of haplotypes seen in the modern data	162
Appendix Nine:	Map to illustrate the 'Synthetic Total Evidence Theory'	164
	References.....	166

List of Figures

1.1	Map of Oceania	5
1.2	Theories for the number of Waka to journey to Aotearoa	9
1.3	Uni-linear maternal inheritance of mtDNA	11
1.4	Founder Effect and Bottleneck	12
1.5	Distribution of Lapita sites within Oceania	14
1.6	Lapita vessel forms	16
1.7	Distribution of Austronesian and Non-Austronesian languages	17
1.8	Sub-grouping within the Austronesian language trees	19
1.9	Major Austronesian language relationships	20
1.10	Dendrogram of Polynesian biological populations	21
1.11	Map of the Human mitochondrial genome structure	26
1.12	Phylogeny tree	27
1.13	Haplogroup specific restriction sites	28
1.14	Map of continent specific haplogroups	29
1.15	Cytosine tract (C-tract)	30
1.16	Electropherogram for the identification of the remains of Tsar Nicholas	34
1.17	Map showing alternative views for Austronesian origins	43
1.18	Hiroa's theories for Polynesian origins via Micronesia	45
3.1	Presence or Absence of the 9-bp deletion as visualised using a 4% NuSieve® Agarose Gel	67
3.2	Calibration Curve for 9-bp deletion	68

3.3	HVI target region PCR products as visualised using a 1% SeaKem® Agarose Gel	70
3.4	HVI Calibration Curve	71
3.5	Electropherogram showing samples with and without the homopolymeric C-tract	74
3.6	Simulation results using sigmoid early population expansion model	87
3.7	Simulation results using sigmoid middle population expansion model	88
3.8	Simulation results using sigmoid late population expansion model	89
4.1	Discovery curve for the Caucasian population data	103
4.2	Discovery curves comparing Eastern Polynesian and Maori populations with Caucasian populations	105
4.3	Probable population growth curve for Maori	108
4.4	Synthetic Total Evidence Theory	123
4.5	Most likely order for the occurrence of the transitions required to produce the Polynesian motif haplotype	126

List of Tables

3.1	The HVI and 9-bp deletion mtDNA haplotype summary data for the current study (AW02)	79
3.2	Variable sites in the HVI region of combined MM98 and AW02 mtDNA data	81
3.3	Total number of haplotypes found in full combined dataset (i.e.MM98+) and AW02 data from the 3' HVI region	84

Abbreviations

9-bp	-	9 base pair
A	-	Adenine
AW	-	Adele Whyte (current study)
bp	-	base pairs
BP	-	before present
C	-	Cytosine
CRS	-	Cambridge Reference Sequence
ddH ₂ O	-	double distilled water
EP	-	Eastern Polynesian
G	-	Guanine
ht	-	Haplotype
HVI	-	Hypervariable One
HVII	-	Hypervariable Two
Ma	-	New Zealand Maori
MM	-	Murray McIntosh et al. (1998)
MN	-	Melanesian
nt	-	Nucleotide
PCR	-	Polymerase Chain Reaction
PN	-	Pacific Nation (Pacific Island)
sdd H ₂ O	-	Sterile double distilled water
SDS	-	sodium dodecyl sulphate
T	-	Thymine
UV	-	Ultra-Violet
WP	-	Western Polynesian

1. Introduction

1.1 Why study Human Evolution?

Historically, questions surrounding human origins have fascinated people for centuries, but over the last 20 years new technologies have been developed which allow researchers from many different disciplines to explore the evolution and migration of anatomically modern *Homo sapiens* in remarkable detail. There are two main competing hypotheses for the evolution of modern humans. Although both agree that *Homo erectus* spread from Africa around 2 million years ago, the ‘recent African origin’ hypothesis (alternatively known as the ‘out of Africa’ and the ‘replacement model’) holds that anatomically modern humans (*Homo sapiens sapiens*) originated in Africa 100,000-200,000 years ago and subsequently colonised the rest of the world, replacing archaic human forms with little or no genetic mixing (McKie and Stringer, 1996). The alternative, ‘multi-regional’ hypothesis proposes that the transformation to anatomically modern humans occurred in different parts of the world, and claims support for this model via fossil evidence of cultural and morphological continuity between archaic and modern humans outside Africa (Caspari and Wolpoff, 1997). Pioneering work by the late Allan Wilson and colleagues (e.g. Rebecca Cann) using mitochondrial DNA (mtDNA), found that all extant human populations trace their mitochondrial ancestry to an ‘African Eve’ within the last 100,000-200,000 years (Cann et al., 1987), which is too recent for the multi-regional hypothesis. Although this research has been criticised and continues to be controversial (For a review: see Templeton, 2002): there is general agreement for a recent African origin(s), a view which has been supported using revised analytical techniques, additional mtDNA studies (e.g. Chen et al., 1995; Chu et al., 1998), and other markers (i.e. Y-chromosomes and autosomal regions). Ingman et al.

(2000) used the results from analysis of complete mitochondrial DNA sequences from 53 humans to suggest that anatomically modern humans left their postulated homeland of Africa approximately 50,000 years ago, and from this staging post they started great journeys of discovery and colonisation across the rest of the planet. Since the departure from Africa ~50,000 years ago, the small human population (~50,000 individuals in Paleolithic distributed in small 'bands': Cavalli-Sforza, 2000) spread to every habitable place in the world and expanded to a worldwide population today of over 6 billion. Although it is now widely accepted that Africa is the homeland for all humankind, questions remain such as: what happened during the years after our exit from Africa? how did humans develop the diverse range of physical attributes, cultures and languages that are present today around the globe? and which groups are most closely related?

These long-standing questions about the origins and evolution of humans are being explored by the combined research effort of diverse disciplines such as Archaeology, Anthropology, Linguistics, Forensic Science, Medical Science and Molecular Biology/Genetics. The field of Molecular Biology/Genetics has advanced quickly due to the development of new diagnostic techniques, which allow the collection of entire or partial mitochondrial DNA sequences for phylogenetic comparisons. In the 1960s the first molecular data on populations was from serological and biochemical polymorphisms (e.g. by allozyme analysis) and restriction fragment length polymorphisms (RFLPs: Papiha and Mastana, 1999). In the 1980s DNA sequence data was produced using mitochondrial DNA (mtDNA) control region (D-loop) sequences (Di Rienzo and Wilson, 1991; Ward et al., 1991). Finally, during the late 1990s, two broad categories of markers (Biallelic and Multiallelic) were developed to test for diversity in the non-recombining region of the Y-chromosome. Biallelic markers have low mutation rates, so each allele therefore represents unique, or a near unique event in

human evolution (e.g. including single nucleotide polymorphisms or SNPs, and *Alu* element insertions). Multiallelic markers have much higher mutation rates. Such markers include microsatellites (average mutation rate ~0.2% per generation) and minisatellites (mutation rate 6-11% per generation). These biallelic and multiallelic Y-chromosome markers are scored by sequencing, denaturing high performance liquid chromatography (DHPLC), or single-strand conformation polymorphism (SSCP) techniques (Underhill et al., 1997). Mitochondrial DNA and Y-chromosome sequences are ideal for elucidating population relationships as they are uni-parentally inherited (maternally and paternally respectively) and thus do not recombine allowing the production of uncomplicated phylogenetic trees.

Although all humankind can trace their ultimate origins to Africa, recurrent 'waves' of migration into nearby Europe and Asia, extensive interbreeding between resident and incoming groups and ~100, 000 years of genetic drift and local allelic selection has meant that underlying relationships between current populations can be difficult to interpret. Settlement of Remote Oceania was the concluding chapter of human migration and colonisation, and included some of the last places on earth to be found and settled. This final migration occurred relatively recently in evolutionary time, within the last 6,000 years, so it is still possible to reconstruct relationships among resident populations. For this reason Oceania is considered to be a particularly good region to study human evolution and migration. Therefore, the central topic of this thesis is the settlement of Polynesia. Results are presented from sampling the genetic variability in the Maori population living in Aotearoa (New Zealand). The following sections of this introduction include an overview of Pacific Island history and culture, theories for the settlement: origins and interaction with surrounding Pacific Nation

populations, and a review of scientific advances: focussing on examples using mitochondrial DNA and Y-chromosome marker data from the Oceania region.

1.2 Introducing Oceania

The Pacific Ocean is a vast area encompassing one-third of the globe and containing over 20,000 islands (see Figure 1.1). However, when these islands were ‘discovered’ and mapped during the voyages of Captain James Cook between 1768-80, he found many islands already inhabited by indigenous people, who had reached them using Neolithic stone-age technology. Other European explorers reaching these isolated islands in their ships for scientific, commercial, imperial or ‘enlightenment’ voyages from the end of the 1500’s were also surprised to find that every sizable island in the vast Pacific Ocean was either currently or had been occupied. Cook in particular was struck by the similarities of language and culture he perceived among the peoples we now call Polynesians and outlined a theory of origins:

“From what continent they originally emigrated, and by what steps they have spread through so vast a space, those who are curious in disquisitions of this nature, may perhaps not find it very difficult to conjecture. It has been already observed, that they bear strong marks of affinity to some of the Indian tribes, that inhabit the Ladrões (Marianas) and Caroline Islands; and the same affinity may again be traced amongst the Battaks of northern Sumatra and the Malays. When these events happened, it is not so easy to ascertain; it was probably not very lately, as they are extremely populous, and have no tradition of their own origin, but what is perfectly fabulous; whilst, on the other hand, the unadulterated state of their general language, and the simplicity which still prevails in their customs and manners, seem to indicate that it could not have been at any very distant period” (Reproduced from Kirch, 2000).

We now know that the Pacific region was colonised within the last 6,000 years, which is very recent in evolutionary terms. Because of the isolation of Pacific populations and limited amount of gene flow between European populations, the patterns and processes involved in human evolution can be explored. The pattern element of human evolution in the Pacific refers to the source of the 'proto-Polynesian' people, i.e. where the people(s) that later became modern Polynesians came from. The process element refers to the settlement dynamic of the proto-Polynesians, e.g. presence/absence of intermarriage along the migration route. The human colonisation of the Pacific has left a 'genetic trail', –an exact record of these voyages preserved in the genes of the modern day descendants of the original voyagers and it is therefore possible to evaluate pattern and process elements in the Pacific region (Hill and Serjeantson, 1989).

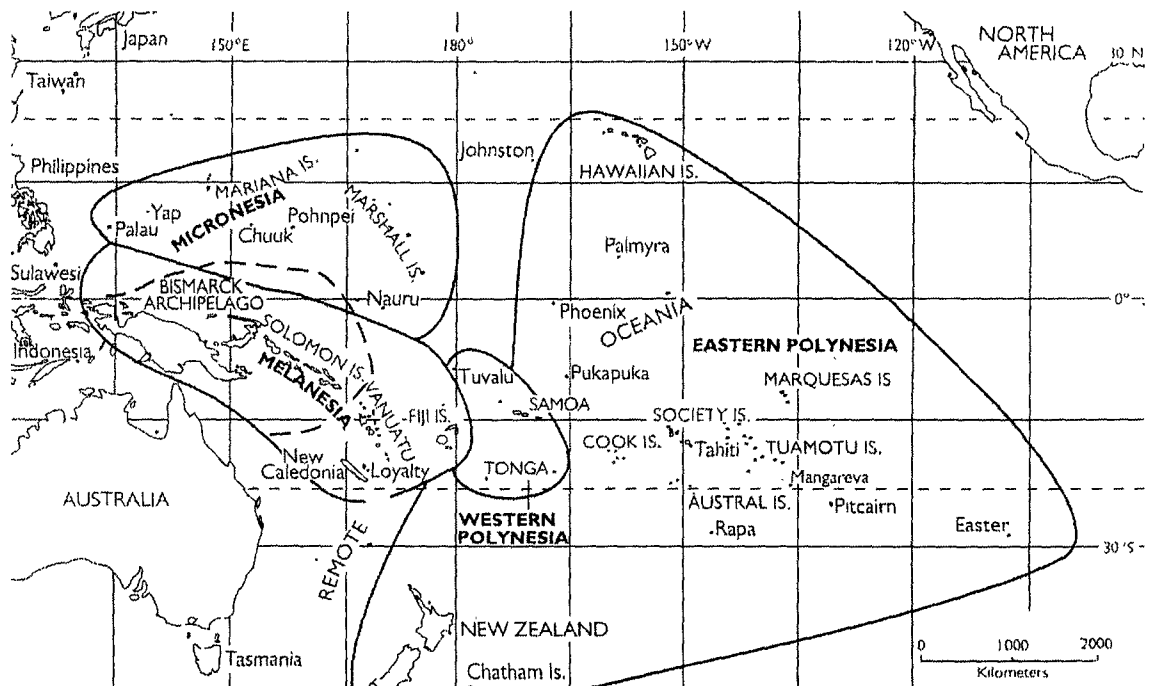


Figure 1.1 Oceania showing the traditional cultural regions of Polynesia, Melanesia, and Micronesia. The heavy dashed line indicates the boundary between Near Oceania (to the west) and Remote Oceania (to the east). The Polynesian Triangle is the area bounded by New Zealand, Hawaii and Rapanui (Easter Island). This Figure has been reproduced from Kirch (2000).

1.2.1 Historical definition of Oceania

Kirch (2000) notes that Dumont d'Urville (1832) classified the people of the Pacific Islands into three groups (see Figure 1.1). The first of these were the Polynesian (of many islands), a generally light-skinned people spread over the islands of the eastern Pacific, including Tahiti, Hawaii, Easter Island and New Zealand. However, for the purpose of the current study the Maori people of New Zealand are considered separately to Polynesians (East and West) unless otherwise specified. In the western Pacific north of the equator, on many small atolls lived the Micronesian (of little islands). Finally, Melanesian (of dark islands), consisting of the generally darker-skinned peoples inhabiting the large islands of New Guinea, the Solomons, Vanuatu (then the New Hebrides), New Caledonia and Fiji. This tripartite classification is still used today, yet maybe misleading if equated to meaningful segments of cultural history (Kirch, 2000). Only Polynesia has really stood the test of time and increased knowledge, to be recognised as a category with historical significance. Therefore, Polynesians may be said to form a meaningful unit for cultural-historical analysis, whereas the labels Micronesia and most particularly, Melanesia imply no true underlying cultural-historical unity. In fact, whether looking at language, human biological variation, or culture the peoples of Melanesia defy categorization, and they are among the most diverse and heterogenous to be found in any comparably sized geographic space on earth. Dumont d'Urville's three groups, taken together, are generally understood to make up Oceania and usually exclude the islands of Southeast Asia (the Indonesian and Philippine archipelagos in particular) and Australia.

Near Oceania and Far Oceania are terms that were proposed by Green (1991) in reaction against the inadequacy of the 'Melanesia' concept (see Figure 1.1). Near Oceania includes the large island of New Guinea, along with the Bismarck Archipelago and as far east as San Cristobal and Santa Ana. This is not only the region of greatest biogeographical diversity within Oceania, but also had human occupation beginning in the late Pleistocene (ca. 40,000 years ago or longer). Within Near Oceania, we find peoples who speak both Austronesian and Non-Austronesian (Papuan) languages. Remote Oceania includes all the Pacific islands to the north, east and southeast of Near Oceania, yet its inhabitants speak exclusively Austronesian languages (see Section 1.4.2). Archaeology confirms that the Remote Oceanic islands were not discovered or settled by humans until after about 1500B.C. Thus, the distinction between Near Oceania and Remote Oceania is not merely a geographic division, but one that encapsulates two major epochs in the history of the Pacific islanders.

1.2.2 Oral History of Oceania

Polynesian people have no written history before European contact and oral tradition was their only means of passing on knowledge. Consequently, much knowledge about historical relationships between populations has been lost. However, oral tradition does link many island groups together e.g. Tupaia, a Tahitian priest navigator interviewed by Cook, named 130 islands for which he claimed to know sailing directions and relative distances (Kirch, 2000).

The discovery and colonisation of Aotearoa, the last sizable island in the entire Oceanic region, and indeed the world, is credited by oral history of Polynesia to Kupe who voyaged deep into the southern ocean in his waka Matahourua. Two accounts known as

‘the legend of Kupe’ and ‘the tradition of Kupe’ are retold in *‘The Discovery of Aotearoa’* (Evans, 1998) and the subtle differences in one account compared with the other were thought to function as a navigators guide to Aotearoa. These subtle ‘clues’ present in the retelling of Kupe’s voyages were actually obvious sailing directions to Aotearoa, hidden for detection only by trained and skilled navigators, although it is thought that current versions have lost much of this information. Traditional knowledge dictates that after the discovery of Aotearoa by Kupe, he returned to his homeland Hawaiki (Tahiti: according to the accounts in Evans, 1998; or Tonga-Samoa archipelagos: Kirch, 2000) to pass on the navigational directions. Over time, an unspecified number of waka made the long journey to Aotearoa, carrying Polynesians from their homeland. Aotearoa was found using advice obtained from Kupe pertaining to directional cues present in the sky (stars and the setting sun) oceans currents and migrating birds (for a description of these traditional navigational methods see Evans, 1998). The exact number of waka to attempt and successfully complete this journey is unknown, but the single 7-waka migration fleet theory (agued by Smith, Hale and Thomson: see Sorrenson 1979) appears to fall short of the true number of canoes. Nga Waka O Nehera –the first voyaging canoes (Evans, 1997) lists the names and in many cases the crews of almost 200 waka. However, Sorrenson (1979) states that many of the ‘canoe traditions’ including the ‘great fleet’, ‘Kupe’ and other ‘traditional’ discoverers of Aotearoa were not based on fact and instead “*were essentially the constructs of European collectors and editors of Maori traditions, not of the Maoris themselves*” and therefore may not reflect the true migration history of the Maori people. Additionally, *Hawaiki* is thought to represent a mythical rather than a geographic location, as *Hawaiki* is regarded as the source of all mankind (Peter Adds, Te Atiawa, Senior Lecturer in Maori Studies, Victoria University: pers. comm., 2001). Although it seems unlikely we will ever know for sure the exact number of canoes to come to Aotearoa, it is probably

safe to discount any models that assume only one canoe or accidental discovery (see Figure 1.2). However, it may be possible to pinpoint an island(s) of origin using DNA technology.

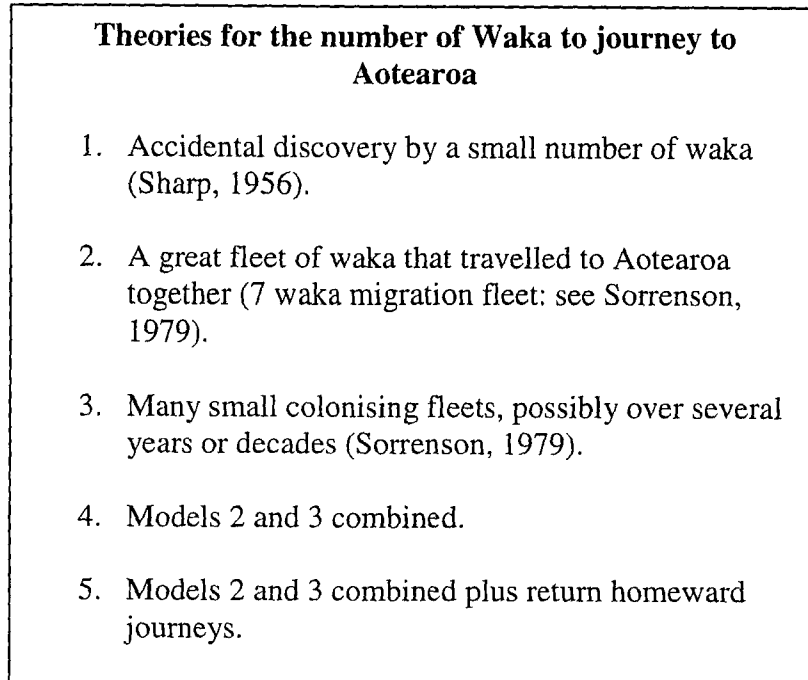


Figure 1.2 Theories for the number of waka to journey to Aotearoa, carrying colonising Polynesians.

1.3 Small population effects in Pacific People

1.3.1 Founder Effect

A Founder effect is the result of starting a new population with a low number of individuals (founders), so that their gene pool may not contain the same proportions of alleles at a particular locus as in the original population. In Pacific populations it is easy to imagine how such situations may arise. Islands have limited space and resources and eventually a population will reach the carrying capacity of the land. Therefore, people may choose to leave to find more resources and space elsewhere, and/or escape possible

tribal conflict and warfare over scarce resources. The effects of a devastating storm could collapse delicate tropical ecosystems causing starvation if another suitable island is not found. However, many islands were not heavily populated at the time of European contact so scarce resources may not have been a factor. Alternatively, due to the chieftdom system operating on many islands, ambitious younger sons who will not inherit the status of their father (this traditionally goes to the eldest son) may choose to find and rule over another island, taking followers along with them (Evans, 1998). Regardless of the reason why an island is left, not everyone will leave and consequently not every allele (or in this study mitochondrial haplotype) will get into the boat to travel on to the next island. Therefore, the diversity present in the founders on the boat is only a sub-set of the diversity present in the original population. Some of the mitochondrial haplotypes have been lost to this new population. Further mitochondrial haplotypes may be lost due to the process of genetic drift whereby chance fluctuations in the reproductive output of individuals can lead to the loss or fixation of haplotypes. Due to the maternal inheritance of mitochondrial DNA, the production of only sons would result in the extinction of that maternal lineage, as although sons and daughters each receive their mitochondrial haplotype from their mother only daughters can pass this genetic material on to the next generation (see Figure 1.3). The loss of rare haplotypes is most likely with the resulting population characterised by a decrease in the number of haplotypes. The process of genetic drift and the founder effect combined can alter the frequencies of the remaining haplotypes so genetically the resulting population may be very different to the original population.

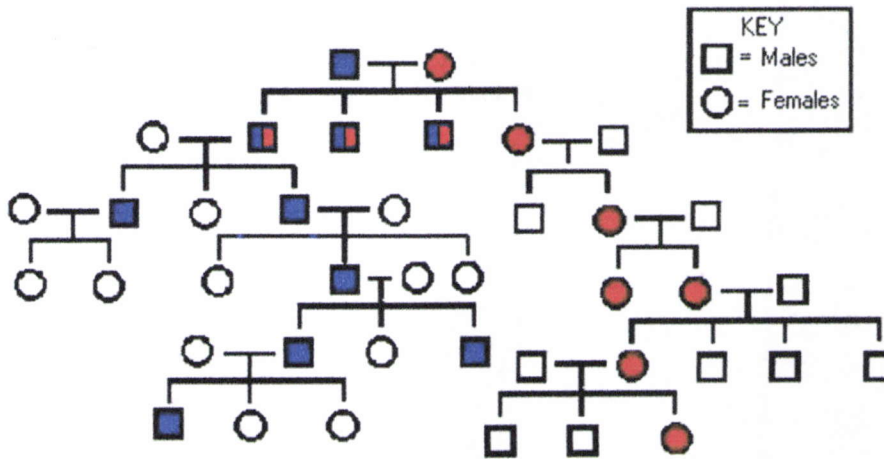


Figure 1.3 Diagram illustrating the uni-linear maternal inheritance of mitochondrial DNA (red) and paternal inheritance of Y-chromosomes (blue). Although both males and females receive a copy of mitochondrial DNA from their mother, only daughters can pass this marker on to the next generation. Only males receive a copy of the Y-chromosome, as this is the basis for sex determination.

1.3.2 Population Bottleneck

A population bottleneck (Flint, 1989) is a period in the history of a population when the number of individuals is reduced to a low number (founder effect) for a number of generations (the bottleneck: see Figure 1.4). During the colonisation of the Pacific it is theorised that as each voyaging group reached a new island, the initial population would have been relatively small for a few generations and hence represent a founder effect and genetic bottleneck. In NZ Maori the effects of a bottleneck and small founder population are reflected in the near fixation (98-100%) of the 9-bp deletion, a marker present in mtDNA (see Section 1.4.6.5). This increase in the 9-bp deletion is accompanied by an increase in the frequency of the CGT motif (another mtDNA marker) from east to west Polynesia (see Section 1.4.6.4.1).

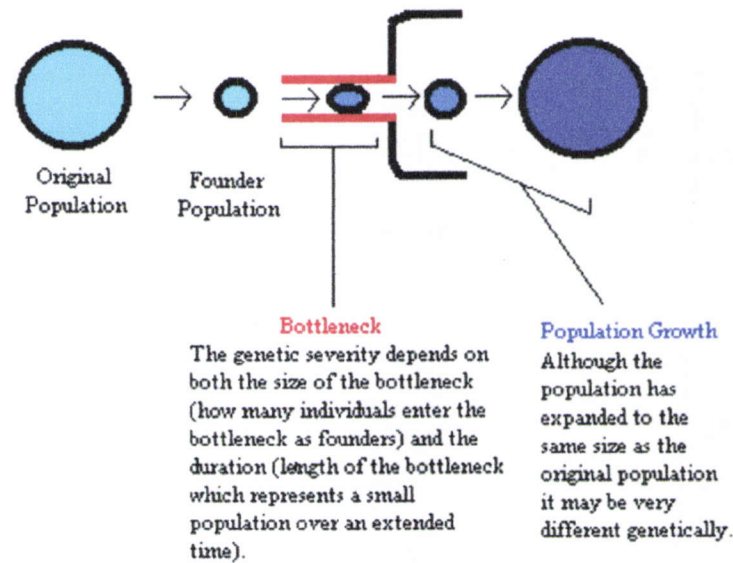


Figure 1.4 A diagram illustrating the concept and consequences of a Founder Effect and Bottleneck.

1.4 Interdisciplinary view of Polynesians

1.4.1 Archaeology: Lapita Culture

Archaeology is the study of ancient cultures from their physical remains, a discipline that enables insights into how our ancient ancestors lived. Recovery and examination of artefacts at sites associated with ancient human remains can be used to reconstruct how people lived. Archaeologists can identify ash fall sites around Oceania that are associated with the Mount Witori (W-K2) eruption in Papua New Guinea 3,600 years ago, and have found differences in site distribution patterns and artefact types below and above this stratification marker. Below the tephra obsidian tools are found, with a shift to lithic technology after the W-K2 eruption. This is accompanied by a change in land use and settlement patterns with a shift to more localised settlement and substantial

obsidian exchange. Ceramics, which were absent before the W-K2 event appear suddenly, and have become known as Lapita pottery (Kirch, 2000).

Since the 1950s and 60s 'Lapita' sites have been discovered across Oceania (see Figure 1.5). These Lapita sites have been intensely studied, with several characteristics found that render these sites wholly different from anything preceding them in Near Oceania. First, they are good-sized settlements, situated on coastal beach terraces or built out over the shallow lagoons as clusters of stilt-houses. Second, their occupants made, traded, and used large quantities of earthenware ceramics, of both plain and decorated varieties. The plain ware consisted largely of red-slipped globular jars with out-turned rims, while the decorated pottery varied in size and shape and was covered in finely executed motifs many representing human faces, made by pressing small toothed ('dentate') stamps into the leather-hard clay before firing (see Figure 1.6). Third, the economic base of Lapita people had expanded from that of the preceding phase in Near Oceania, utilizing all of the tree crops that had been domesticated in this region, but also including pigs, dogs and chickens. Fishing strategies were sophisticated, and they employed a variety of fishhooks, including trolling lures for taking tuna and other pelagic fish on the open ocean. Lapita people were seafarers, venturing beyond coastal waters to trade substantial quantities of pottery, chert, obsidian, oven stones, and other materials between communities. Their material culture exhibits a greater range of tools, implements, and ornaments than any earlier sites in Near Oceania (Kirch, 2000).

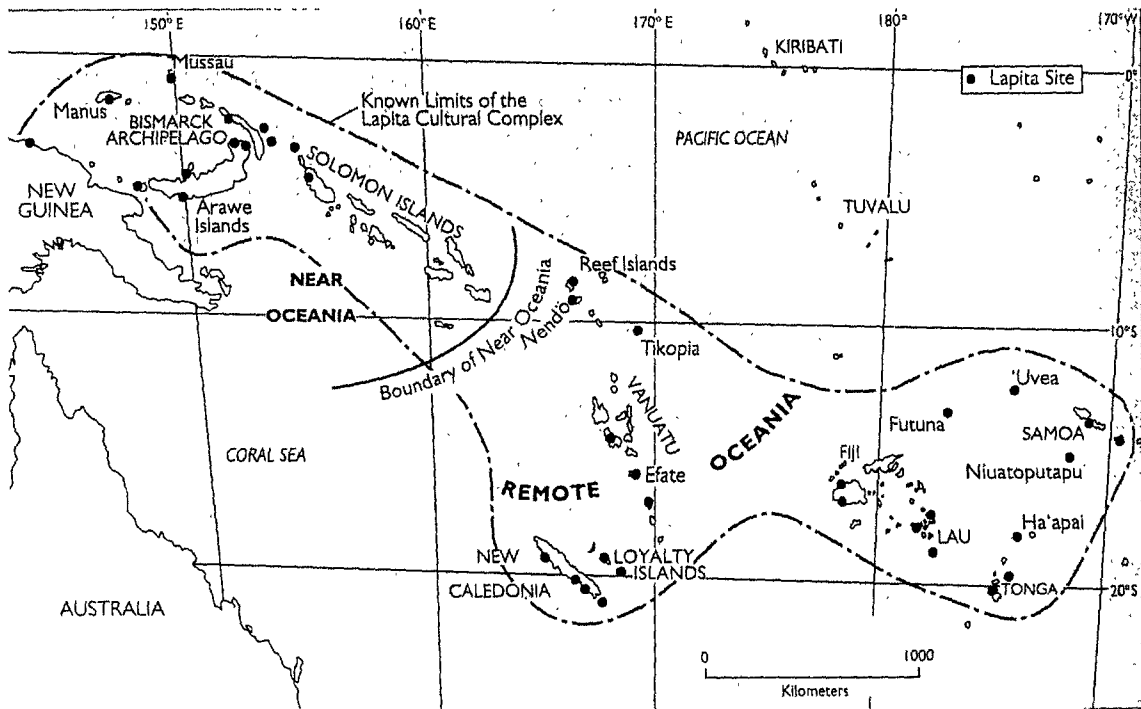


Figure 1.5 The distribution of Lapita sites in Near and Remote Oceania. This Figure has been reproduced from Kirch (2000).

Despite some proposals that the Lapita people represent an indigenous, *in situ* development, many archaeologists conclude that the Lapita culture was entirely new and intrusive and invaded along with the potters, from outside Near Oceania. A popular view shared by many archaeologists is that of Bellwood (1997), who argues that the archaeological record of the expansion of pottery-using, agricultural peoples out of Southeast Asia and into the islands of the Pacific, correlates with the spread of the Austronesian language (see Section 1.4.2). The first ceramic-making peoples are in evidence on the large island of Taiwan, at around 3000 B.C., and over the course of the next 1,500 years, their descendants or closely related peoples expanded south and southeast, toward Near Oceania. More than 200 radiocarbon dates have been obtained from Lapita sites ranging geographically from the Bismarck Archipelago to Samoa and Tonga, providing a firm basis for tracking the chronology and rate of population spread. The earliest dates for sites with the distinctive Lapita style of dentate-stamped

decoration range between 1500 and 1400 B.C. For the next two to three centuries, there was no expansion of Lapita populations beyond the immediate Near Oceania region. Then, around 1200B.C., a phase of long-distance voyaging and colonisation commenced with Lapita groups breaking through the invisible boundary of Near Oceania, which for 30,000 years or more had marked the limit of human existence in the Pacific. Lapita groups moved in every direction southeast to Santa Cruz group and south through the Banks Islands and into the main Vanuatu archipelago, in quick succession to the upraised Makatea islands of the Loyalty group and on to discover New Caledonia (1200-1100 B.C.). Lapita people reached Fiji by 1100-900 B.C., and then went on to Samoa and Tonga. The radiocarbon dates conclude that the Lapita expansion out of Near Oceania into Remote Oceania, as far south as New Caledonia and as far east as Tonga and Samoa was accomplished in only two to three centuries (Kirch, 2000).

Debate regarding Eastern Polynesian settlement revolves around what is called the 'long pause', whereby settlement beyond Tonga/Samoa occurred as much as 1,600 years after the first Polynesians reached these areas. This pause was followed by a sudden burst of exploration and discovery that would take people into virtually every other island and archipelago within a mere four centuries (i.e. by A.D. 1000). Historical linguists such as Fawley (1996) argue for a period of common development in the Western Polynesian homeland prior to later dispersals in order to explain the many innovations in Proto Polynesian language, and to allow time for the differentiation of Proto Tongic and Proto Nuclear Polynesian languages. However, the time required for such language changes is unknown, and if the pause was very long, as some maintain, then this would leave an insufficient period of time for similar innovations to arise in Proto Eastern Polynesian language and culture (see Section 1.4.2).

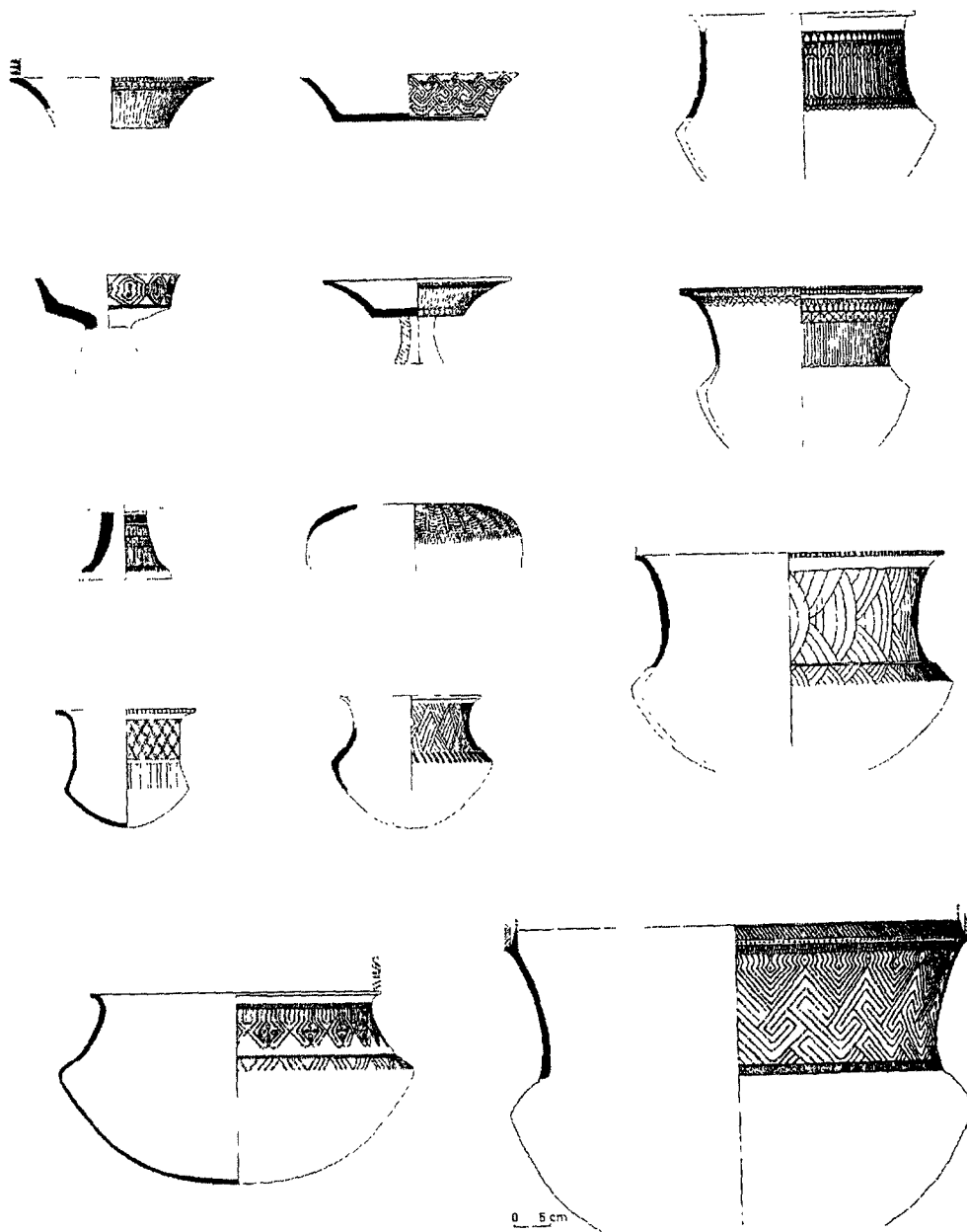


Figure 1.6 The varieties of Lapita vessel forms occurring in early Lapita sites in New Caledonia. This Figure has been reproduced from Kirch (2000). Please refer to this book for an extensive guide to the Archaeological and Linguistic History of the Pacific.

1.4.2 Linguistics: Austronesian Language

The indigenous people of Oceania all speak related languages, which belong to the well-defined and geographically widespread Austronesian family. Austronesian is spoken as

far west as Madagascar, and as far east as Easter Island. It is spoken by most of the populations of island Southeast Asia (including the Taiwanese Aborigines), and makes up the majority of languages spoken in Melanesia outside New Guinea, and all the languages spoken within Micronesia and Polynesia. Whereas in New Guinea, and in a few scattered locales elsewhere in Near Oceania, such as on New Britain and the Bougainville Islands, the indigenous languages are Non-Austronesian or Papuan (see Figure 1.7).

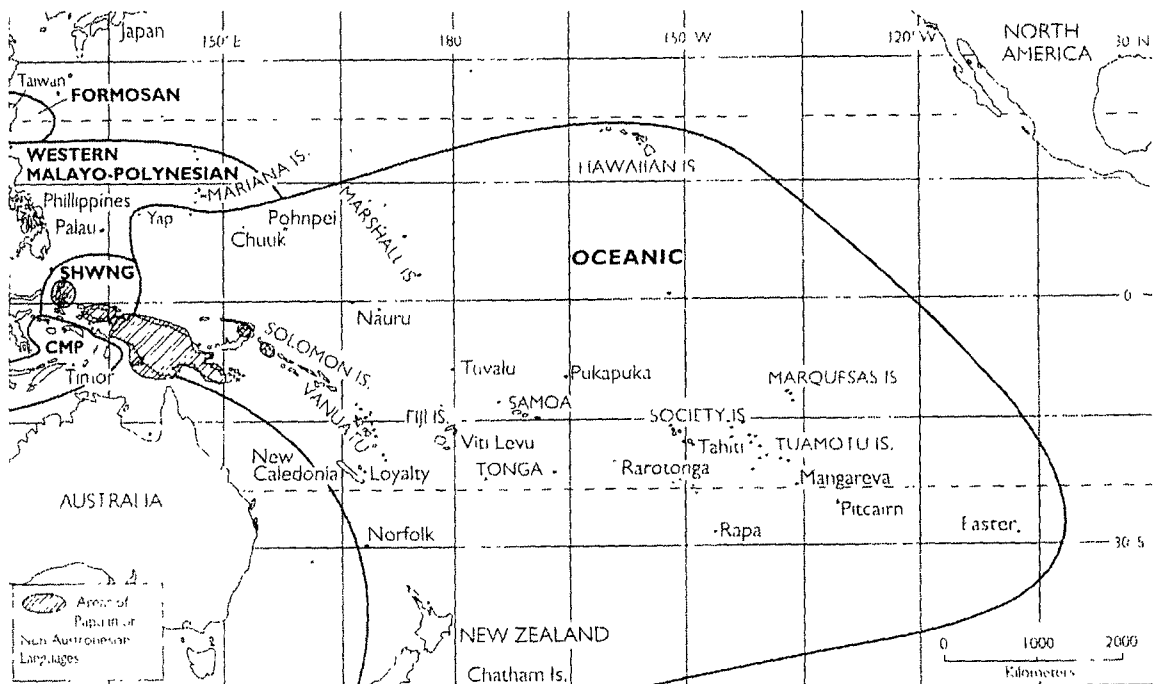


Figure 1.7 The distribution of Austronesian and Non-Austronesian Languages in Oceania. The Non-Austronesian languages are situated in the shaded areas, while the heavy lines delineate several major subgroups of Austronesian languages. SHWNG, South Halmaher-West New Guinea; CMP, Central Malayo-Polynesian. This Figure has been reproduced from Kirch (2000).

There are 1,200 or so languages of the Austronesian language family, possibly the largest among the 6,000 languages of the modern world (Diamond, 2000). This language has 10 subgroups of which 9 (containing only 26 languages) are spoken only by the non-Chinese aborigines of the island of Taiwan (Diamond, 2000). The tenth subgroup encompasses all Austronesian languages outside Taiwan, from Madagascar to east Polynesia and proves to be of great culture-historical significance due to the supposed association of this Austronesian language expansion with the Lapita culture.

The archaeological models and dating are remarkably consistent with the internal relationships or sub-grouping of the Austronesian languages as worked out through careful linguistic comparisons (see Figure 1.8 i.). The Oceanic branch has no deep history of successive language ‘splits’ such as would be expected had speakers of these languages been present for a long period in Near Oceania (as were the speakers of the highly diverse Papuan languages). Rather, the first-order branches or sub-groups of Oceanic form a chainlike *linkage* (Figure 1.8 ii.). To linguists, this implies that the Proto Oceanic speech community moved rapidly into the area east of Halmahera and as far as the Solomon Islands, quickly forming a dialect chain, along which communication continued for some time. Differentiation eventually occurred as links in this chain were severed, leading to the four or five major first-order subgroups of Oceanic language observed in the area today. This model correlates well with archaeological evidence for the rapid intrusion of a new population, coming out of island Southeast Asia around 1500B.C., with trade and exchange as evidenced by archaeology, to be expected among peoples who spoke related dialects. Links between Lapita and Oceanic language in Remote Oceania have support, as prior to Lapita there is no evidence for any human occupation in these regions, unlike in Near Oceania which had been occupied for ~30,000 years prior to contact with Lapita peoples. Sub-

grouping within the model for the Oceanic languages spoken in Remote Oceania closely mirrors the sequence of island settlement as given by archaeology. Polynesian languages are shown in Figure 1.8 iii., and the branching structure of this tree may also correlate with successive stages of migration.

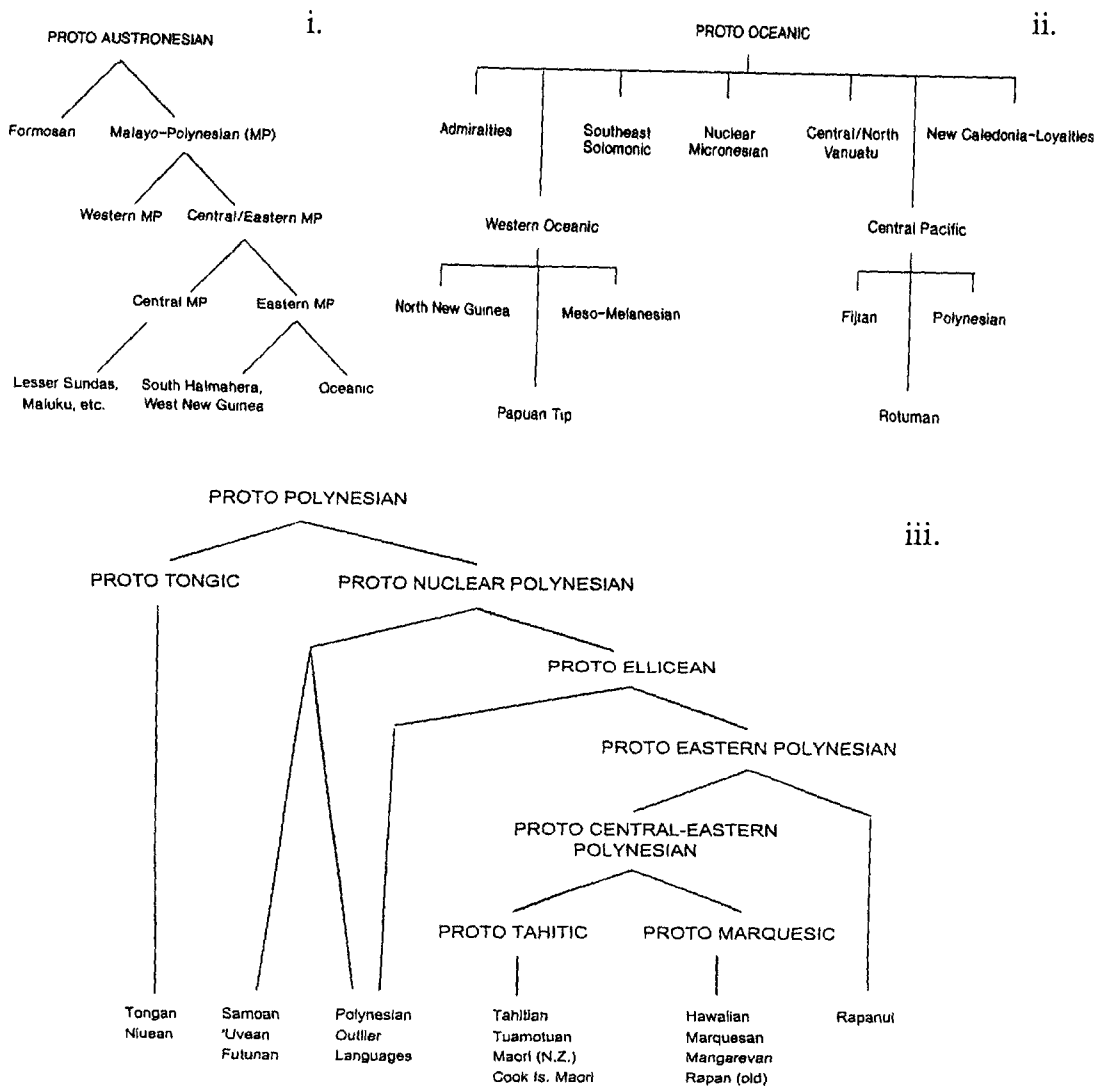


Figure 1.8 Diagram i. shows the subgrouping or 'family tree' for the Austronesian languages (After Kirch, 1997), Diagram ii. shows the sub-grouping of the Oceanic languages (After Kirch, 1997) and Diagram iii. shows the subgrouping of the Polynesian languages (After Marck, 1999). These Figures are all reproduced from Kirch (2000): the original references are shown in brackets.

A recent paper by Gray and Jordan (2000) shows support for the Austronesian expansion via language trees (see Figure 1.9). This study used a parsimony analysis of a large language data set, to test competing theories (see Sections 1.5.1 and 1.5.2) for the colonisation of the Pacific by Austronesian-speaking peoples.

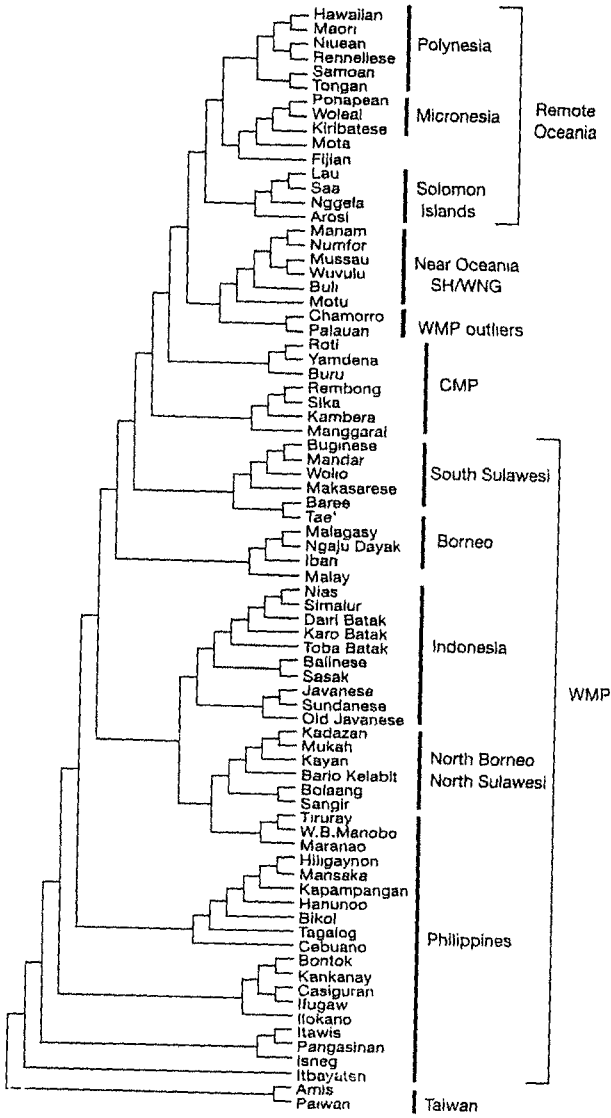


Figure 1.9 Diagram showing the relationships between the major Austronesian Language groups of the Pacific. This Figure is reproduced from Gray and Jordon (2000). Abbreviations are as follows: WMP, Western Malayo-Polynesian; CMP, Central Malayo-Polynesian; Sh/wng, South Halmaher/West New Guinea.

1.4.3 Anthropology

Anthropology is the study of the variety, distribution, characteristics and culture of mankind. In the Pacific, studies have often focussed on variability in skull shape, facial features, skin colour, stature and body form. In short, biological evidence reinforces the linguistic model of a robust phyletic grouping for the Polynesian populations.

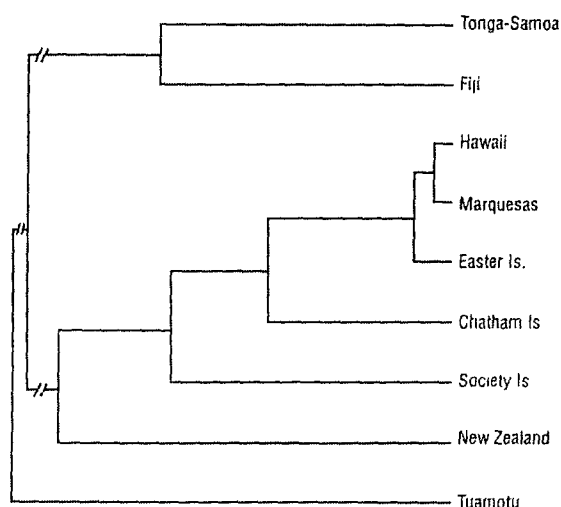


Figure 1.10 A dendrogram of Polynesian biological populations, based on anthropometric analysis, mirrors the Western-Eastern Polynesian division suggested by linguistic and cultural comparison. This Figure is reproduced from Pietrusewsky (1996).

1.4.4 Human Population Genetics

Early work by population geneticists was done to discern patterns of historical population movements, using comparisons from blood groups, serum proteins and enzymes of a few hundred donors. Although techniques have moved on to incorporate direct DNA sequence comparisons from tens of thousands of individuals, some of the early findings by molecular population genetic studies will be reviewed (see Hill and Serjeantson, 1989).

1.4.4.1 Blood groups

The ABO blood groups were discovered at the beginning of the 20th century, and have been subject to numerous studies due to their importance for successful blood transfusions. During World War I, Ludwik and Hanka Hirschfeld, two Polish immunologists, examined several different ethnic groups among the soldiers in the English and French colonial armies and the World War I prisoners, including Vietnamese, Senegalese, and Indians. The Hirschfelds' discovered that the proportions of individuals belonging to the different blood groups were different in every population (see Cavalli-Sforza, 2000). This 'uneven' blood group distribution pattern is now known to be universal, and each population is different for most other polymorphisms as well. In the Pacific region although the frequency of the ABO blood groups varies widely, some generalizations can be made. The blood group B is absent entirely or met with only sporadically, and it is possible that the blood group A gene was absent in the central Pacific before European contact (see Hill and Serjeantson, 1989). With the discovery and combined data of other blood group systems (e.g. RH and MN) it is possible to identify populations from the 5 continents. However blood groups on their own are insufficient to provide satisfactory resolution for population relationships (Cavalli-Sforza, 2000).

1.4.4.2 HLA

The class I and II genes of the Major Histocompatibility Complex (MHC) are among the most variable loci known in many vertebrates, including humans. Variation in the genes of the human MHC, more commonly known as Human Leukocyte Antigens (HLA) genes have been subject to intense study for many years due to their importance

in transplants, autoimmune disease and recognition of pathogens. In the Pacific region a study using HLA markers was conducted by Hagelberg et al. (1999b), who found reduced overall variability in island Melanesia. An 'Asian' DPB1 0501 HLA allele was found at a high percentage in Chinese (47%), Taiwanese (70%), the Roro (71%) and Samoans (70%), which was thought to indicate an ultimate Asian origin of these people. The reduced levels of HLA diversity could be explained by a genetic bottleneck in the founding of these populations.

1.4.3 Globin genes

Flint et al. (1986) studied the frequency of α -thalassemia in Melanesia and found an altitude and latitude dependent correlation with malaria endemicity throughout Melanesia, supporting the hypothesis that protection against this parasitic disease is a major factor responsible for the high frequencies of various haemoglobinopathies in many parts of the world. However, evidence of α -thalassemia at gene frequencies of ~7% in parts of Polynesia that are historically free of malaria suggest a link between Melanesia and Polynesia, whereby the α -thalassemia genotype was carried out into Polynesia from Melanesia.

1.4.5 Minisatellites

A study by Flint et al. (1989) of minisatellite diversity in Polynesians showed that in comparison to Melanesians the number of alleles at each loci is smaller. However, the allele distribution and the mutation rates at Polynesian minisatellite loci do not deviate from those predicted by the neutral mutation/infinite allele model. The low gene diversity is therefore likely to be a result of the maintenance of small population sizes

and bottleneck effects during the colonisation of the Pacific (Flint et al., 1989; Hamilton PhD thesis, 1994). The interesting point about the study by Flint et al. (1989) is that minisatellites identified a significant loss of heterozygosity (or gene diversity) that was not noted using more traditional markers, such as the degree of polymorphism in protein and antigen loci.

1.4.6 Mitochondrial DNA

Mitochondria are small cellular organelles present in cell cytoplasm and responsible for energy production. In contrast to nuclear DNA (nucDNA), which has only two copies per diploid cell, with one copy originating from each parent, mitochondria are inherited maternally and contain mtDNA at a much higher copy number. Mitochondrial DNA has a high copy number because the cell cytoplasm can have many mitochondria, each one containing multiple copies of mtDNA, with the overall number of mtDNA copies ranging from ~200-1700 in somatic cells (depending on the tissue type), to more than 100,000 in oocytes. The maternal mode of transmission of mtDNA is thought to be due to the small number of mitochondria contained in sperm, in relation to the many thousands of mitochondria contained in each ovum. Additional experiments have also shown that paternal mitochondria may be eliminated from the ovum after fertilisation (Ankel-Simons, 1996). The large number of mtDNA copies in each cell allows a high likelihood of recovery of mtDNA for genetic testing (see Section 1.4.6.9). The special features of mitochondrial DNA including high copy number and uni-linear maternal transmission as mentioned, and absence of recombination and high mutational rate, make mtDNA ideal for phylogenetic studies. Consequently, for many years, mitochondrial DNA has been the marker of choice for evolutionary studies and many laboratory groups have taken advantage of these features to infer relationships between

diverse populations, and draw phylogenies illustrating population interactions (e.g. Horai et al., 1991; Horai et al., 1993; Torroni et al., 1993a; Torroni et al., 1993b; Bonatto et al., 1997; Torroni et al., 1998; Helgason et al., 2001).

While there is no general dispute regarding the high mutation rate of mtDNA (many studies illustrate this property: Jazin et al., 1998; Seielstad et al., 1998), the mode of inheritance and lack of recombination in humans have been recently challenged (Ashkenan et al., 1997; Howell et al., 1997; Awadalla et al., 1999; Eyre-Walker et al., 1999; Eyre-Walker et al., 2000). However, these assertions have been emphatically refuted (Macaulay et al., 1999a; Jorde et al., 2000; Kivisild and Villems, 2000; Kumar et al., 2000; Parsons and Irwin, 2000). Thus, despite some studies that may show recombination as a factor in mitochondrial evolution, these views are in the minority and have yet to achieve widespread support. A recent report (Stoneking, 2000) indicates that the hypervariable regions (sites that evolve at a much faster than average rate) present in the non-coding regions of human mtDNA are mutational hotspots, rather than the result of ancient mutations that have been redistributed among mtDNA lineages via recombination (Eyre-Walker et al., 1999; Hagelberg et al., 1999a). The impact of solid evidence proving recombination in human mitochondrial DNA would be widespread as current mtDNA phylogenetic studies assume that mutations accumulate at a steady rate on independent lineages. Recombination would effectively 'shuffle' this variation making phylogenetic inferences difficult.

1.4.6.1 Mitochondrial genome structure

Figure 1.11 shows a diagram of the human mitochondrial genome with an enlarged view of control region (CR), also called the D-loop (due to the visible displacement structure formed during mtDNA replication). The CR is about 1,000 bp long subdivided in two hypervariable regions: HVI and HVII of about 400 bp each, the length and position of these regions vary slightly depending on the definition used in various studies. The two hypervariable regions are separated by a more conservative region (16 365-00 073), and a longer coding region of about 15,000 bp, that contains 37 genes (Anderson et al., 1981). The mtDNA evolves very quickly, up to 5-10 fold faster than nuclear DNA and the control region of mtDNA has a mutation rate about 10 times higher than the coding region of mtDNA (Greenberg et al., 1983).

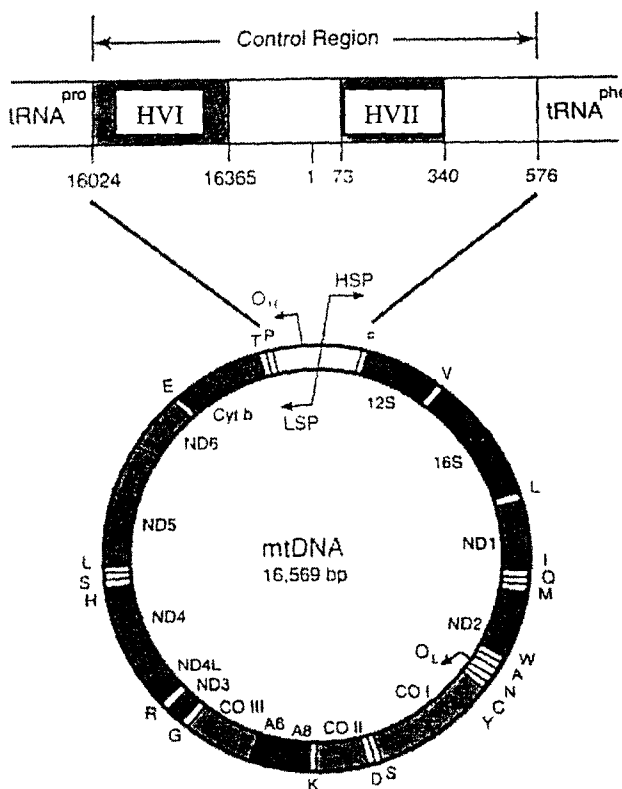


Figure 1.11 Map of the human mitochondrial genome and expanded diagram of the non-coding control region.

Listed are the genes for 12S and 16S ribosomal RNAs, subunits of the NADH coenzyme Q oxidoreductase complex (ND), cytochrome c oxidase complex (CO), cytochrome b (cyt b), ATP synthase (A), and 22tRNAs (labelled with single letter amino acid code). The light strand (O_L) and heavy strand (O_H) origins of replication, and the light strand (LSP) and heavy strand (HSP) transcriptional promoters are shown by arrows. Control region diagram shows flanking tRNAs and location of hypervariable region 1 (HVI) and 2 (HVII); numbering follows that of the standard reference sequence of Anderson et al. (1981).

Figure reproduced from Holland and Parsons (1999).

1.4.6.2 Variable sites

Within the control region of human mtDNA some sites are prone to changes in nucleotides, but these variable (polymorphic) sites are not evenly distributed. The number of polymorphic sites in HVI is about double that of HVII, and within both of these two segments the mutation rate is site dependent. There are some extremely variable positions and some more conserved sites, with a large bias towards transitional ($T \leftrightarrow C$ and $A \leftrightarrow G$) versus transversional ($G \leftrightarrow T$, $G \leftrightarrow C$, $A \leftrightarrow T$ and $A \leftrightarrow C$) changes (Piercy et al., 1993).

The number and distribution of polymorphic sites in human mitochondria have some important implications for mitochondrial DNA studies as a nucleotide change may be frequent in human populations for two very different reasons: either they represent an ancient mutational event and it is present in all the individuals derived from any particular ancestor, or it is a site with an extremely high mutation rate, thus representing independent mutational events in different lineages, with no phylogenetic signal (see Figure 1.12). Analyses, which do not take into account this effect, giving the same weight to all the nucleotide changes detected may give poor resolution in the study of populations at the micro evolutionary level (Francalacci et al., 1999).

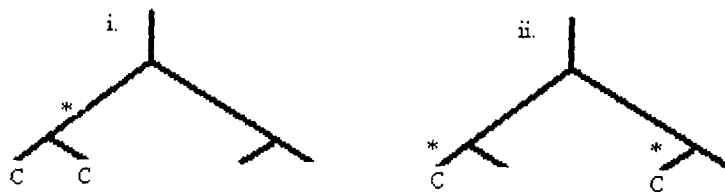


Figure 1.12 High rates of nucleotide mutation (indicated by *) may give parallel changes in nucleotides, which can complicate phylogenetic inferences. Diagram i. shows two lineages with a common ancestor. In Diagram ii. the same mutation has occurred independently in different lineages and does not indicate a phylogenetic signal.

1.4.6.3 Restriction analysis

Restriction enzyme digestion and analysis allows mtDNA lineages to be grouped into clusters defined by specific endonuclease restriction sites. These clusters are called haplogroups and are based on high-definition restriction analysis using a defined set of 14 endonucleases. The positions of haplogroup specific restriction sites are shown in Figure 1.13 following the alphabetic nomenclature initiated by Torroni et al. (see Francalacci et al., 1999). Figure 1.14 illustrates the resolving power of phylogenetic reconstructs behind restriction analysis revealing that haplogroups are largely continent specific. Recent studies combining both RFLP and control region sequence (Macaulay et al., 1999b) allow for finer resolution of the haplotypes present in each haplogroup, and provide distinction between clusters. For example, in the European lineages, samples may have identical haplotypes using control region sequence, but when tested with RFLP may be found to belong to different haplogroups. This is the case for some haplotypes in the H and U haplogroups.

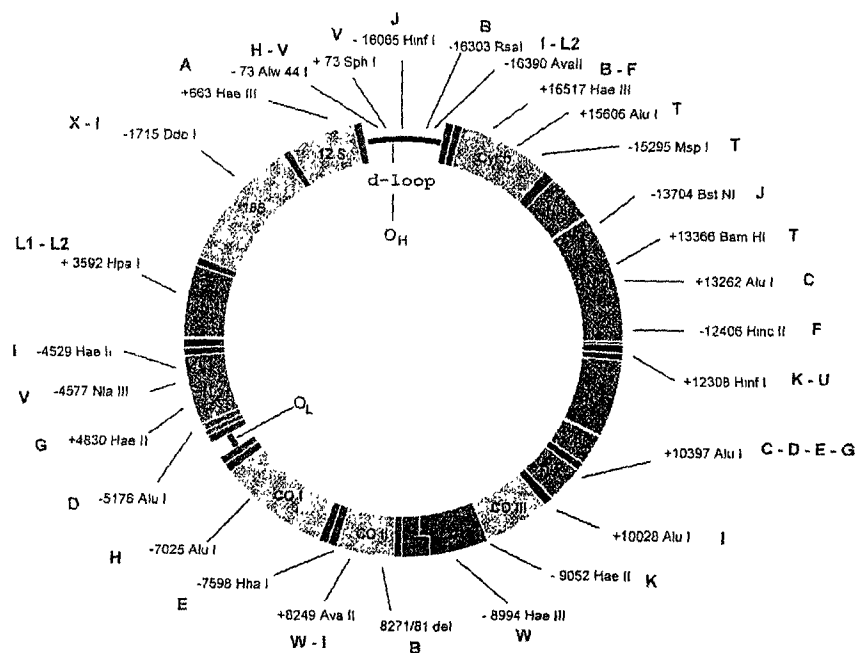


Figure 1.13 The position of the haplogroup specific restriction sites in the human mitochondrial genome. This Figure has been reproduced from Francalacci et al. (1999).

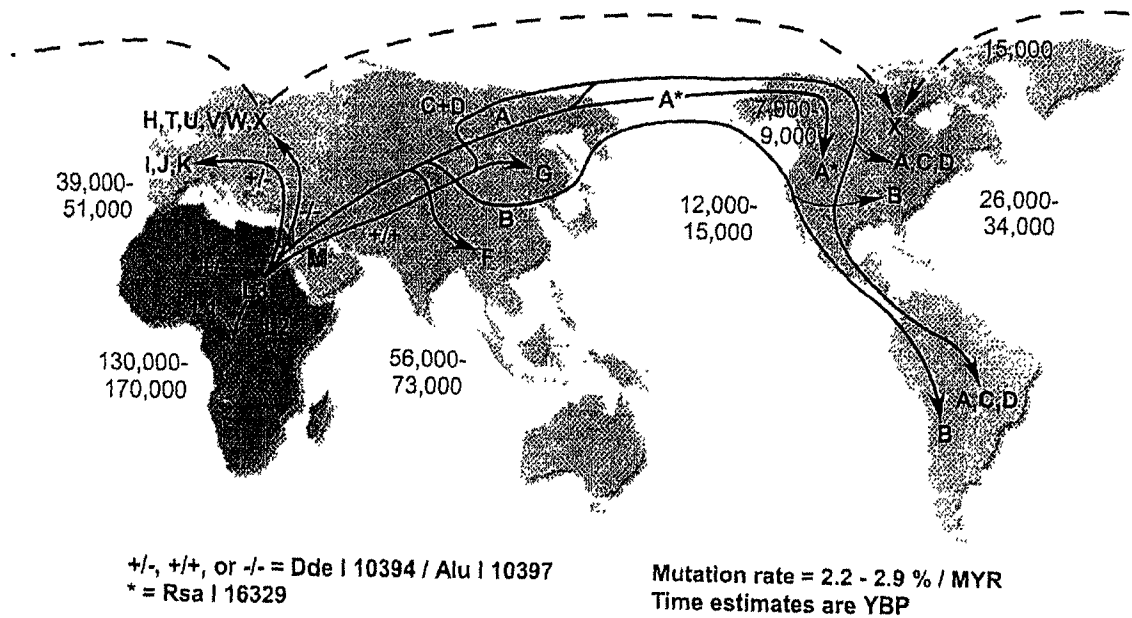


Figure 1.14 Map of continent specific haplogroups (Taken from MITOMAP website: <http://www.gen.emory.edu/MITOMAP/WorldMigrations.pdf>)

Analyses of mtDNA RFLP results identify haplogroup lineages L1, L2 and L3 as specific to the African continent. Asian-specific haplogroups are A, B, F and M, with the M supergroup further split into haplogroups C, D, E and G. Native Americans have mainly A, B, C and D haplogroup lineages. The European-specific mtDNA haplogroups are H, I, J, K, T, U, V, W and X; with V the sister group to H.

1.4.6.4 Hypervariable I Region (HVI)

1.4.6.4.1 CGT motif

As mentioned previously mtDNA sequencing efforts have largely been concentrated on the HVI (~16 024 to ~16 383) and HVII (~16 057 to ~16 372) regions (numbering after Anderson et al., 1981). Within the HVI region there is a set of specific nucleotide transition substitutions that are frequently found in Polynesian populations (~94% of

mtDNA lineages: Sykes et al., 1995) and have become known as the Polynesian ‘CGT motif’ (Hagelberg and Clegg, 1993; Redd et al., 1995; Hagelberg, 1997). The Polynesian ‘CGT motif’ is located at nucleotide (nt) positions 16 217, 16 247, and 16 261, numbering according to the Cambridge reference sequence (CRS: Anderson et al., 1981) with T→C, A→G and C→T transitions at each site respectively (see Figure 1.15).

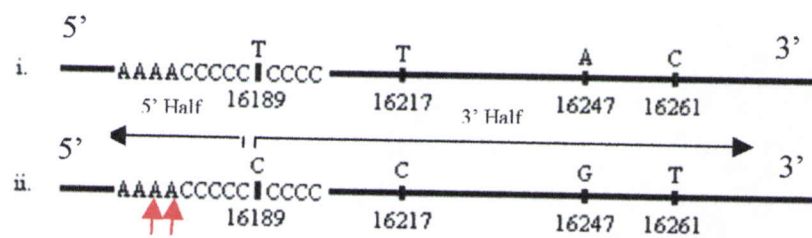


Figure 1.15 Enlarged view of the mtDNA HVI region showing the CRS (i.) and the transitions required to form the CGT motif (ii.). The transition at position 16 189 creates a homopolymeric cytosine tract (see Section 1.4.6.4.2 and Figure 3.5). Often the two adjacent adenines (A) at positions 16 182 and 16 183 (indicated by red arrows) acquire A→C transversions producing the characteristic 12 C-tract (see Figure 3.5). Black arrows indicate the 3' and 5' halves of the HVI region. This study focuses on the 3' half, as this is the region with greatest mtDNA sequence diversity (see Section 3.5).

1.4.6.4.2 Homopolymeric cytosine tract

In addition to the CGT motif, the control region of Pacific Island people often contains a homopolymeric cytosine tract between nts 16 184 and 16 193. This tract is present at a high frequency in many populations in the Pacific (~100%, current study) but is also found in ~15% of Europeans (Bendall and Sykes, 1995), and is thought to occur in

~20% of the general population (Holland and Parsons, 1999). This tract is caused by a T→C transition at nt position 16 189 compared with the CRS, resulting in an uninterrupted homopolymeric tract (see Figure 1.15). Due to the homopolymeric cytosine tract, direct nucleotide sequencing of individuals with this trait produces a characteristically blurred sequence of nucleotides 3' beyond the tract. This difficulty in characterising the sequence after the C-tract is caused by the low fidelity of the mitochondrial DNA polymerase system (Holland and Parsons, 1999), as the population of mtDNA molecules within an individual will differ for the length of the C-tract (heteroplasmy: see Section 1.4.6.7; e.g. refer to G204 in Figure 3.5). Therefore, during sequencing the template molecules are out of register with one another. Sequencing clones from individuals with this tract by Bendall and Sykes (1995) revealed that ambiguous sequences are caused by high levels of length heteroplasmy in the homopolymeric tract coupled with lower levels of length heteroplasmy in the four adenines proceeding the tract. It is also possible that these blurred sequences are further complicated by errors during PCR and/or sequencing.

In the HVII region C tract-length heteroplasmy also occurs between nt positions 00 303-00 315, but in this case there is often an identifiable predominant length variant, which allows sequencing to proceed normally (Holland and Parsons, 1999).

1.4.6.5 9 base pair (9-bp) deletion (Region V)

Another mtDNA characteristic common in Polynesian populations is the deletion of one of the two adjacent copies of a 9 base pair (9-bp) direct repeat sequence (CCCCCTCTA) in the COII/tRNA^{Lys} intergenic region (also known as region V) of mitochondrial DNA (Wrischnik et al., 1987). In the ancestral state there are two copies

of this 9-bp direct repeat, but some populations have lost one copy and therefore have a 7-bp deletion. This marker has previously been used as a polymorphic anthropological marker for people of east Asian origin (Wrischink et al., 1987; Melton et al., 1995), and links Asiatic populations with Oceania (Hertzberg et al., 1989; Hagelberg and Clegg, 1993; Lum et al., 1994; Redd et al., 1995) and the Americas (Ward et al., 1991; Ward et al., 1993). This deletion was originally thought to have arisen only once in the evolution of east Asian populations. It was thus thought of as an important 'Asian specific' mutation that could be used to trace the genetic trail out of Asia into Oceania and the New World. However, this mutation has also been found at low frequencies in African people (Soodyall et al., 1996), and in a single Spanish Caucasian person (Barrientos et al., 1995) indicating that the 9-bp deletion can originate independently and could be considered as a recurrent mutation. Despite the fact that this marker has been found at low frequencies in other populations, it is still very important for characterising Asian and Polynesian populations, and can be used as an additional character for assigning haplotypes (i.e. distinguishing between samples with and without the 9-bp deletion).

1.4.6.6 Hypervariable II Region (HVII)

Although many studies have found substantial variation in both the HVI and HVII regions of mtDNA in worldwide populations, a previous study in the Pacific region found almost no variation in HVII rendering it unsuitable for determining phylogenetic relationships among Oceanic people (Lum et al., 1994). Despite this, a small number of samples were sequenced for this study using the HVII region, but results did indeed show very little variation when compared with the diversity of sequences obtained using samples from other populations. A study by Lutz et al. (1996) identified a third

hypervariable region, which they named HVIII between positions 00 440 and 00 560. This region has not been mentioned in any previous studies conducted on Pacific or worldwide populations to the best of my knowledge. The proposed new hypervariable region was published in a forensic paper and may be well characterised on the forensic database. However, a review paper by Holland and Parsons (1999) indicated that this forensic database was highly regulated and contained HVI and HVII regions only.

1.4.6.7 Heteroplasmy

There is usually only one mitochondrial DNA population within an individual, but in rare cases (2-8% of the human population) heteroplasmy has been observed (Holland and Parsons, 1999). Heteroplasmy is the term given to an individual with two or more mitochondrial DNA populations. Many cases of heteroplasmy occur as point mutations in individuals affected with mtDNA diseases, whereby there will be population of 'normal' and 'abnormal': disease causing molecules within the patient's mitochondrial DNA population. The proportion of mutant mtDNA transmitted from mother to offspring is variable, and the 'dose' of the mutant abnormal mtDNA received from the mother may influence the severity of the phenotype in the offspring.

There are a few reports of silent (i.e. non-disease causing) heteroplasmic point mutations in mtDNA in human populations. A famous example is that of a single heteroplasmic point mutation found in the remains of Tsar Nicholas II (see Figure 1.16), which was valuable in subsequent identification (Comas et al., 1995). When the mtDNA sequence from the putative remains of the Tsar (see Figure 1.16 a) was compared with mtDNA sequence from a known living maternal relative, there was one difference in the sequence (see Figure 1.16 c). This mtDNA sequence difference was at

nt position 16 169, correlating to the heteroplasmic position found in the skeletal remains of the Tsar. Automated sequence analysis of the Tsar's remains assigned a C to position 16 169, as although another mtDNA population was present with a T at this position, C was predominant. A sample taken from the Tsar's brother, Georgij Romanov also showed heteroplasmy at 16 169 but with a predominance of T (see Figure 1.16 b). Using sequence databases and considering only a sequence match between the Tsar and his deceased brother Georgij, a likelihood ratio of 150 was calculated implying the observed data was 150 times more likely if the remains were those of the Tsar as opposed to a random individual. When a sequence match at the same heteroplasmic position, i.e., 16 169 was taken into consideration the likelihood ratio jumped to more than 300,000, and when the sequence match between the Tsarina and her living maternal relative HRH Prince Philip was taken into consideration, the total mtDNA data gave a likelihood ratio of more than 100,000,000.

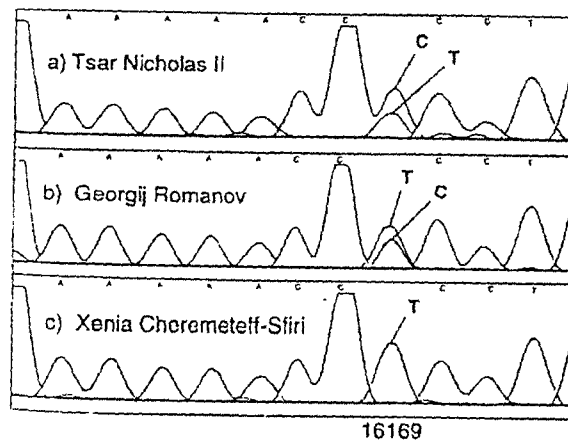


Figure 1.16 Automated sequence electropherograms comparing mtDNA sequences at position 16 169: (a) sequence from Tsar Nicholas II, showing heteroplasmy with cytosine predominating thymine; (b) sequence form Gran Duke Georgij Romanov, showing heteroplasmy with thymine predominating over cytosine; (c) sequence for living maternal relative of Tsar Nicholas, separated by five generational events, apparently homoplasmic for thymine. Diagram is reproduced from Holland and Parsons (1999), using information from the study by Ivanov et al. (1996).

A study by Bendall et al. (1996) suggests that heteroplasmic point mutation variants in the control region are more widespread than has been reported. The lack of heteroplasmy observed in earlier studies is attributed to difficulties of detection by direct sequencing. A study by Holland and Parsons (1999) using denaturing gradient gel electrophoresis polymorphism (DGGE), has shown that this method has high sensitivity for detecting mixed DNA sequences compared with direct sequencing (e.g. using 253 samples, DGGE analysis identified 14% heteroplasmy compared to 1% using direct sequencing).

As mentioned previously (see Section 1.4.6.4.2), tract-length heteroplasmy occurs frequently in the HVI and HVII regions of worldwide populations (up to ~20%), and is particularly common in Polynesian populations (100%, present study). In these instances the point mutation occurs within a region of mtDNA that is rich in C's (see Figure 1.15 ii). A single point mutation creates a tract of C's which may lead to a compromise in the mitochondrial copying fidelity, so that as the mitochondrion makes a copy of its DNA, slip-strand mispairing is more common, and incorporation of additional C's can easily occur. This copying error can result in a copied mtDNA that is longer (more C's) than the template, and effectively results in two (or more) populations of mtDNA 'types' within a single individual, leading to sequencing difficulties.

1.4.6.8 Ancient DNA

With the advent of PCR technology and improved methods for the prevention of contamination, it is now possible to gather authentic mtDNA sequence information from trace and ancient samples. The majority of mtDNA sequence information is obtained from bone samples. Recently, mtDNA has even been recovered from the remains of

two Neanderthal specimens (Krings et al., 1997; Krings et al., 1999; Ovchinnikov et al., 2000). Using HVI and HVII comparisons with contemporary human mtDNA have shown that the Neanderthal mtDNA variation falls well outside the range present in modern human populations. These findings suggest that modern humans and Neanderthals are different species, thereby supporting the 'recent African origin' hypothesis. However, a study by Wall (2000) determined that sequence data from ~50–100 unlinked nuclear loci is necessary to distinguish between the 'recent African origin' and the 'multiregional' model of modern human evolution. Sufficient evidence to enable thorough evaluation of these two models is expected to be available within the next few years (Wall, 2000).

Another study on ancient mtDNA (Adcock et al., 2001) using ancient (~60,000 years old) anatomically modern Australian skeletons, places the deepest mtDNA lineage from an anatomically modern human in Australia, whereas analysis using contemporary human mtDNA genomes places the deepest branches in east Africa. This study challenges current concepts of modern human origins, but is only based on one set of skeletal remains. Further ancient skeletons from Australia and other regions would be needed to confirm this assertion, as although these data imply that the deepest branches occurred in Australia, the study is highly contentious and is yet to receive widespread support as other alternative theories could equally well explain this result.

Work done using mtDNA extracted from ancient bone samples of prehistoric Pacific Islanders (Hagelberg and Clegg, 1993) has shown that patterns present in the mtDNA of the population presently residing in a region are not necessarily reflected in the ancient DNA. Although there is a high frequency of the 9-bp deletion (indicating an east Asian origin) in the present Pacific populations, the 9-bp deletion was missing in the oldest

samples from Fiji, Tonga and Samoa. In contemporary DNA the deletion is present at frequencies of 82% (Hertzberg et al., 1989), 93% and 98% (Sykes et al., 1995) respectively. The findings by Hagelberg and Clegg (1993) were therefore unexpected as many of these samples were taken from 'Lapita sites' (see Section 1.4.1), and the remains were assumed to be ancestral to modern Polynesians. These results were taken to suggest that the Lapita culture was originally carried from its Melanesian homeland into the central Pacific by indigenous inhabitants of island Melanesia, rather than by Austronesian-speaking migrants from Southeast Asia who settled the region en-route to the eastern Pacific. Furthermore, these results would suggest that Polynesians may have descended from both Melanesian and Southeast Asian ancestors, and that the combined effects of population bottlenecks and genetic drift (see Section 1.3) have led to the extinction of all but a very few mtDNA lineages during the colonization of the eastern Pacific (Hagelberg and Clegg, 1993). However, once again these conclusions are based on a limited number of specimens and even the authors concede that their results should be taken with a note of caution as discrepancies between their expected and actual results could have alternative explanations.

1.4.6.9 Use of Mitochondrial DNA in Forensic Science

Human mtDNA has proved to be a very powerful tool for forensic identification of individuals using DNA from very old (see Section 1.4.6.8) or degraded tissue (Lutz et al., 1996), to provide genetic identification criteria that would not often be possible using nuclear markers. This is due to the fact that mtDNA is present in many copies in each cell and is more likely than nuclear DNA to be found intact in compromised samples. Since the advent of PCR only a very small number of mtDNA need to be recovered, and once amplified the mitochondrial DNA sequence can be used for

comparison with living maternal relatives of the deceased. This has led to the identification of many high profile skeletal remains such as those of Tsar Nicholas II (Gill et al., 1994) and of Martin Bormann, the second most powerful man of the Third Reich (Anslinger et al., 2001).

Currently in New Zealand a number of cases that have physical matches using microscopical hair comparisons are supported using mtDNA sequences, usually for Caucasian victims only. It is now proposed that mtDNA analysis will be introduced for hair work, skeletal remains and other decomposing/compromised samples (Sally-Ann Harbison, ESR Forensic Team Leader: pers. comm., 2001). Typical examples of decomposed highly degraded samples include corpses on battlefields, disaster victims, and skeletal remains of missing people degraded by long-time exposure to weather.

1.4.6.9.1 Disaster Victim Identification (DVI)- An Australasian example

At the New South Wales Institute of Forensic Medicine (Australia), Dr Sarah Robinson is involved in the set-up of processes to identify disaster victims using mitochondrial DNA analysis (pers. comm., 2001). INTERPOL regulations currently require at least 2 forms of primary information to successfully and accurately identify a victim after a disaster. The most commonly used identification methods are medical, dental and fingerprinting records, but DNA has become increasingly popular over the last 4 years. Although nuclear DNA markers provide greatest resolution, the circumstances of many disasters lead to severe degradation of nuclear DNA, making high copy number mtDNA a viable alternative target for victim identification. However, strict controls must be applied to the matching and exclusion standards when using mtDNA, as false positives could be returned if the mtDNA haplotype was shared among more than one of the

unidentified victims. This is a particular concern for the Polynesian populations resident in both Australia and New Zealand as very few mtDNA haplotypes have yet been found in these populations. However mtDNA results taken together with another primary identification (id) method could provide a suitable and satisfactory back up, if none of the other primary id methods were available.

1.4.8.9.2 Missing Person Identification (Australia)

To assist in the identification of skeletal remains found by police, Dr Robinson is also contracted to set-up a Missing Persons Database. The aim of this database is to reduce stress on the relatives of missing people. Relatives can now donate a blood sample to determine their mtDNA sequence with their results stored in the Missing Person database. These mtDNA sequences will be compared to each set of unidentified remains found. Further contact with the police is only necessary when a match has been found, whereas presently relatives have blood taken each time a new set of unidentified remains are found. Precautions identical to those taken with DVI work will need to be implemented, as there is still the possibility that remains could be returned to the wrong relatives. In both Missing Person Identification and DVI work, additional information on different mtDNA haplotype frequencies for people of different cultural backgrounds is critical so statistical probabilities of a mismatch can be worked out (refer to section D of the review by Holland and Parsons, 1999).

1.4.7 Y-Chromosome studies

There are many recent publications using the non-recombining region of the Y-chromosome (Hammer and Horai, 1995; Underhill et al., 1996; Underhill et al., 1997;

Karafet et al., 1999; Hill et al., 2000; Thomson et al., 2000; Kayser et al., 2001) for phylogenetic analysis. This region may be considered the male equivalent of mitochondria, due to its inability to form a homologous pair during meiosis and mitosis. Therefore, it does not undergo recombination and is passed through the paternal lineage, like a family name. Unfortunately Y-chromosomes are not present in high copy numbers in cells, and therefore PCR amplification and sequencing is not as straightforward as it is for mtDNA. Early work on Y-chromosomes using DNA sequencing found almost no variation among individuals with different population backgrounds (Dorit et al., 1995). However perseverance has been rewarded as new markers were found and phylogenies can now be drawn (Underhill et al., 2001b).

Studies using Y-chromosomal polymorphisms to trace paternal lineages in Polynesia have found various levels (~2-27%) of European admixture (Hurles et al., 1998; Su et al., 2000). Initial work done on Maori Y-Chromosomes indicate a substantial (~40%) European ancestry (Hurles et al., 1998; Su et al., 2000; Underhill et al., 2001a), with lineages representative of East Asia and Melanesia also present. These data, together with the mitochondrial information, indicate a gender-modulated pattern of differential gene flow in the history of Polynesia as has been found elsewhere e.g. in Asian (Perez-Lezaun et al., 1999) and Colombian populations (Seielstad et al., 2000: for a review see Pennisi 2001).

1.5. Models of the Oceanic Expansion of people across the Pacific

With the considerable wealth of combined interdisciplinary knowledge available about the populations present in the Pacific region, it is obvious that some attempts will be

made to assimilate these ideas into a model that will explain how the population expansions took place. Although there is some disagreement as to the relative genetic contributions of males and females, and where the expansion originated, most experts in all disciplines do not support (Heyerdahl, 1950) idea that American Indians played a major role in the settlement of the Pacific, and instead favour settlement from somewhere in the west, around island southeast Asia. Although many archaeologists agree that there was a rapid and recent migration of Austronesian speakers into Remote Oceania, what is open to debate is precisely where the Lapita peoples came from and how much they interbred along the way with the indigenous people whose ancestors had been living in Near Oceania for at least 33,000 years. The two leading, and supposedly most strongly contrasting, models are explained in detail below. A recent news article (Gibbons, 2001) notes that the polarised views may not necessarily coincide with those of the author most closely associated with the model, and geneticists may have made the situation worse by testing what were seen as ‘competing’ theories. A third model is also explained in detail below, as this is the latest attempt at an interdisciplinary theory. For a brief description of other models see Section 1.6.

1.5.1 The Express Train Model (Bellwood, 1991)

The most recent human migration is postulated to have started from a homeland in China ca. 6,000 years ago and spreading outwards to Taiwan, Philippines, eastern Indonesia and Melanesia reaching Fiji by ca. 3,500 years ago. The colonisation of the Pacific concluded with a radiation filling the Polynesian Triangle ca. 1000 years ago and Aotearoa (New Zealand) within the last 800-1000 years, or maybe even more recently (McFadgen et al., 1994). This leading historical description of human expansion across the Pacific was made by archaeologist Peter Bellwood (1991), and has

become known as the ‘Express Train to Polynesia’ hypothesis (Diamond, 1988). This assumes that the Lapita culture (see Section 1.4.1) arrived full-blown in the Bismarcks, having arisen far to the west in island Southeast Asia and spread rapidly eastwards (Diamond, 1988). This Express Train model was thought to suggest a very fast migration out of Taiwan with no integration of indigenous people already along the colonisation route. However, although the Express Train terminology was quick to spread, Bellwood himself has expressed concern over the false dichotomy between his model and Terrell’s (see Section 1.5.2). In fact Gibbons (2001) suggests that geneticists looking to test opposing views have made assumptions about Bellwood’s model, suggestions backed up by quotes “*I don’t believe in express trains*” says Bellwood. “*It was a fast movement out of Taiwan, but it wasn’t totally closed off. Of course there has been intermarriage. The express train was a kind of journalistic statement*”.

1.5.2 Entangled Bank Model (Terrell, 1988)

The Entangled Bank Model (Terrell, 1988) states that the various Pacific people are thought of as a geographic set of local and island populations, more or less in touch with each other, who have followed separate but often connected historical pathways of local adaptation and cultural change. This view argues that the Lapita culture and peoples arose from a diverse Melanesian population that was linked in a complex social and trade network with people from island Southeast Asia over thousands of years. Terrell claims that describing the events of Polynesian colonisation in terms of a fusion between two peoples and two cultures is too simple. Terrell uses two metaphors to describe the interlocking interdependent nature of these populations: a playing field and an Entangled Bank; the latter has become a well-known metaphor.

1.5.3 The Slow Boat Model (Oppenheimer and Richards, 2001a,b)

The Slow Boat Model (Oppenheimer and Richards, 2001a,b) suggests that Polynesians originated in island Southeast Asia, specifically eastern Indonesia (Oppenheimer and Richards, 2001b). While the idea of an offshore origin for Polynesians is not new (Solheim, 1996; Meacham, 1985), reinterpretation of mtDNA and linguistic evidence may provide renewed support for these theories.

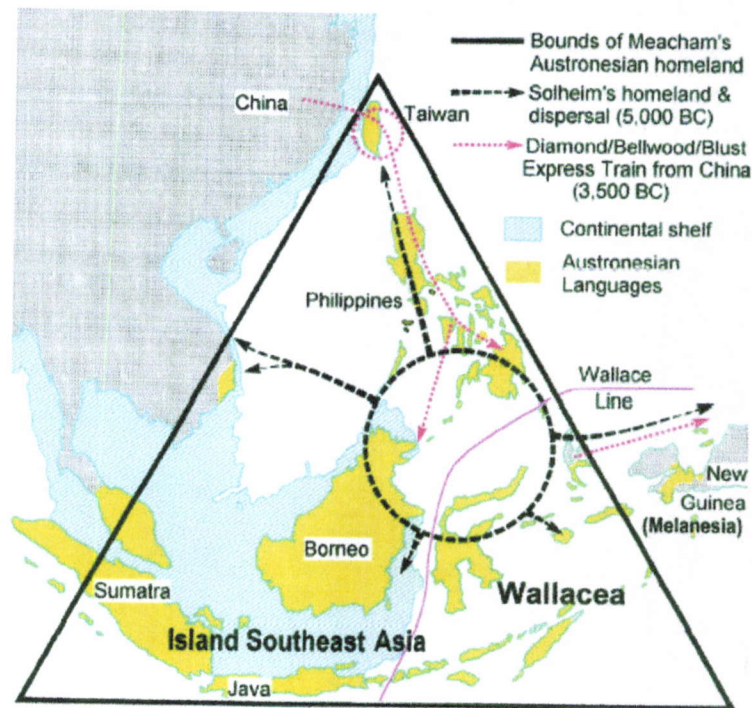


Figure 1.17 Map showing the alternative views for Austronesian origins. The oldest view represented by Meacham (solid triangle), Terrell and Solheim (interrupted solid black line and circle) argues for an island Southeast Asian homeland (>5,000 B.C.). The ‘Express Train’ view of a recent rapid migration from China via Taiwan (3,000-4,000 B.C.), spreading to replace the older populations of Indonesia after 2,000 B.C. is shown by the red dotted line.

1.6 Pacific Colonisation

Two levels of complexity:

1. Source of original population (Pattern)

- Express Train (Mainland China/South China) – Bellwood, 1997: see Figure 1.17
- Entangled Bank (Melanesia) – Terrell, 1988
- Taiwan – Redd et al., 1995; Sykes et al., 1995 and Lum et al., 1994
- Island South East Asia (Indonesia) – Richards et al., 1998; Oppenheimer and Richards, 2000a, b; Solheim, 1996; Meacham, 1985: see Figure 1.17
- South America – Heyerdahl, 1950
- Micronesia – Hiroa, 1938: see Figure 1.18

2. Settlement Dynamic (Process)

i.e. How did these people settle? Was there intermarriage? Does it really depend on whether you look at male (Y-chromosome) or female (mtDNA) markers?

- Express Train (Bellwood, 1991; Diamond, 1988) – Settlement population spread rapidly through the Pacific without admixture with populations already present.
- Entangled Bank (Terrell, 1988) – Contribution from both males and females from populations already present along the migration route via intermarriage.
- Geneflow (Devlin, 2001) – Contribution of males only along the colonisation route.

- Triple-I Model: Intrusion, Innovation and Integration (Green, 1991) – Intermarriage between the intrusive Austronesian speakers and the long-resident populations of Near Oceania.
- Slow Boat (Oppenheimer and Richards, 2001 a, b) – Melanesian admixture through intermarriage (same as Triple-I?)

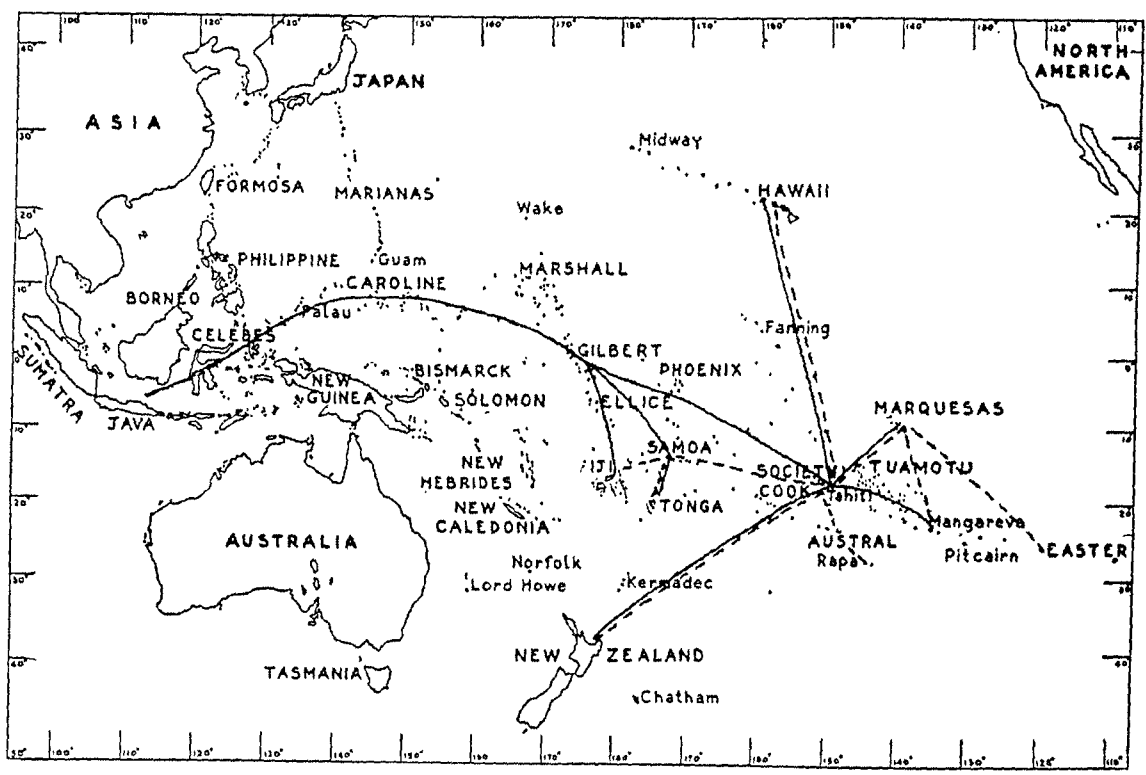


Figure 1.18 Diagram illustrating Hiroa’s theories for Polynesian origins via Micronesia.

This Figure is reproduced from Kirch (2000).

1.7 Research objectives

The objectives of my thesis are as follows:

- To collect a molecular database of at least 50 partial mtDNA sequences from Maori and Pacific Nation volunteers.
- To search this database for characteristic elements including the 9-bp deletion and the Polynesian CGT motif.
- To use this database to refine the estimates of Murray-McIntosh et al. (1998) for the number of Maori women founders for New Zealand's Maori population.
- To evaluate the 'competing' models of Oceanic population expansion, and attempt to consolidate these theories into the 'Synthetic Total Evidence Theory'.

2. Materials and Methods

2.1 Source of DNA

The samples used for mitochondrial DNA sequence analysis were obtained from blood, frozen in plastic vials at -80°C. These samples were obtained from the human DNA bank, held at the Institute for Molecular Systematics (IMS) in the School of Biological Sciences, Victoria University of Wellington. Permission to use the samples was obtained from Dr Geoffrey K. Chambers, the curator of this collection, and from the Wellington Ethics Committee (WEC). Support from WEC was obtained after consultation with various Maori and Pacific Nation Health Organisations: such as the Wellington Maori Health Committee, Maori Health Unit, Pacific Health Research Centre, Te Putahi-a-Toi and the Eru Pomare Centre. Approval and support for this project was obtained from these organisations by late December 2000, and WEC granted ethics approval on the proviso that additional support from a Katiaki group would be sought, as was advised by the consulted Maori and Pacific Nation Health groups. A Katiaki group was established in 2001 using members from Toi huarewa, a group with high Maori representation, established at Victoria University to deal with Maori academic issues.

The blood samples were collected anonymously with informed consent from volunteers at the Wellington blood transfusion service. Volunteers were asked to provide a small blood sample for analysis, state their gender and identify their ethnic origin. The self-declared ethnic origin was used to randomly select participants for this study. In this study 61 participants who identified themselves as Maori were used as a representative sample to survey the genetic variability of the NZ Maori population. A further 24

individuals who identified their ethnic origin as Pacific Island (Hawaii, Fijian, Rarotonga, Cook Island, Samoan, Niue and Tokelau) had their DNA sequences analysed as a comparison for the Maori data.

2.2 DNA Extraction Protocol

The DNA samples used in this study were a sub-set of the samples available from the IMS human DNA bank. The DNA had already been extracted from the blood samples which were collected by a previous student and have been stored in the IMS facility (Hamilton PhD thesis, 1994). However, it was necessary to extract and analyse a sample of DNA from my own blood to ensure I was not contaminating my work.

2.2.1 Recovery of white blood cells from whole blood

Fresh blood was collected in a 4.5 ml Vacutainer tube. The whole blood was diluted with 10-20 volumes of cold (4°C) red blood cell lysis solution (RBS: see Appendix One), in two 30 ml polythene centrifuge tubes, and left on ice until the solution cleared (~10 minutes). The white blood cells were pelleted using a Heraeus Sepratech centrifuge at 4,000 rpm for 10 minutes. The pellets were resuspended in 5 ml cold TBS (see Appendix One) and combined in one graduated 10 ml polypropylene centrifuge tube. The washed white blood cells were recovered by centrifugation at 3,000 rpm for 15 minutes at 4°C, using a Sorvall RC5 centrifuge, with a SS34 rotor. Cell debris was removed from surface of pellet by gentle inversion prior to pouring off the supernatant. The pellets were stored frozen in the centrifuge tubes at -20°C until required.

2.2.2 Proteinase K Digestion

The pellets were taken from the freezer and resuspended in 4.5 ml of TE buffer (see Appendix One) by vortexing and shaking. The following reagents were then added in the following order: 250 µl of 0.5M EDTA (see Appendix One), 250 µl 10% sodium dodecyl sulphate (sds) reagent (see Appendix One) and 25 µl of 20 mg/ml Proteinase K (Life Technologies). The mixture was divided among five 1.5 ml tubes and incubated at 55°C for 2-3 hours, with mixing by inversion every hour.

2.2.3 Phenol/Chloroform Extraction

The contents of the five 1.5 ml tubes were recombined in a single 15 ml polypropylene tube and an equal volume of neutral phenol (Life Technologies, Gibco BRL[®]) was added, mixing by inversion. The tube was spun for 5 minutes at 10,000 rpm (4°C) using a SS34 rotor in a Sorvall RC5 centrifuge. The top aqueous layer was transferred to a clean 1.5 ml plastic microcentrifuge tube. An equal volume of phenol:chloroform:isomyl (25:24:1) was added to the aqueous solution, mixed gently by inversion and spun at 10,000 rpm for 5 minutes using a Sorvall MC12V microcentrifuge. The top aqueous layer containing the DNA was removed, placed in a clean 1.5 ml tube and an equal volume of chloroform:isomyl (24:1) was added. The final extraction mixture was combined by inversion and spun in a Sorvall MC12V microcentrifuge for 2 minutes and the top aqueous layer was transferred to a clean 30 ml tube.

2.2.4 Ethanol Precipitation of Extracted DNA

To precipitate the extracted DNA 100µl of 3M NaCl and 15 ml of cold ethanol was added to the 30 ml tube, and spun at 10,000 rpm, 4°C for 30 minutes. The supernatant was removed and replaced with 70% ethanol to resuspend and wash the pellet, which was then and spun for 10 minutes at 10,000 rpm, 4°C. The ethanol was removed and the newly precipitated DNA was allowed to air dry. The dry pellet was then redissolved in 200 µl of TE buffer, with a 20 µl aliquot taken for immediate use and stored at 4°C and the rest stored at -20°C in a clean 1.5 ml tube.

2.3 Mitochondrial DNA Target Amplification

Three areas of the mitochondrial genome: HVI, HVII and Region V (see Sections 1.4.6.4, 1.4.6.5 and 1.4.6.6) were amplified for each sample using the Polymerase Chain Reaction (PCR). The expected product size in the HVI target region was ~442 bp. In the HVII region sequences were expected to be ~400 bp long. In Region V sequences were either 112 or 121 bp long depending on the presence or absence of the deletion respectively.

2.3.1 Primers

Each mtDNA target region amplified using the polymerase chain reaction required primers complementary to its flanking regions. The primer pair sets used for HVI were those used by Murray-McIntosh et al. (1998), H16401 (5'-TGA TTT CAC GGA GGA TGG TG-3') and L15997 (5'-CAC CAT TAG CAC CCA AAG CT-3'). Due to the homopolymeric cytosine tract present in many samples, sequencing was difficult (see

Section 1.4.6.4.2). Therefore, internal primers as used by Dr Sarah Robinson of the New South Wales Institute of Forensic Medicine (pers. comm., 2001), were used to identify ambiguous nucleotides and provide sequence results in both the forward and reverse orientations (L16209 5'-CCC CAT GCT TAC AAG CAA GT-3' and H16164 5'-GAT GTG GAT TGG GTT T-3').

The HVII region was amplified using primers communicated by Dr Robinson (pers. comm., 2001: L048 5'-CTC ACG GGA GCT CTC CAT GC-3' and H408 5'-CTG TTA AAA GTG CAT ACC GCC A-3').

Region V (9-bp deletion) was amplified using the primers as described by Wrischnik et al. (1986: A8316 5'-ATG CTA AGT TAG CTT TAC AG-3' and B8196 5' ACA GTT TCA TGC CCA TCG TC-3').

2.3.2 PCR Protocol

The standard 25 µl double stranded PCR reaction contained:

18.35 µl	sterile double distilled water (sddH ₂ O)
2.5 µl	10X Qiagen [®] PCR Buffer (containing 15mM MgCl ₂)
1.0 µl	2.5 mM Pharmacia [®] dNTPs (deoxynucleotide triphosphate mix with dATP, dTTP, dCTP, dGTP at 2.5 mM each)
1.0 µl	Forward (F) Primer (10 µM, Life Technologies)
1.0 µl	Reverse (R) Primer (10 µM, Life Technologies)
0.15 µl	Qiagen [®] Taq Polymerase (5units/µl)
1.0 µl	Template DNA (~20 ng/µl)

Before setting up PCR reactions the surfaces and equipment within the PCR set-up lamina flow hood were irradiated by Ultra-Violet (UV) light and cleaned with ethanol. This process ensures the degradation of any DNA present in the air and on the surfaces that could contaminate samples. Due to the small amount of *Taq* required for each reaction, precision measurements using pipettes are difficult, so a Master Mix was made. Each Master Mix was made by multiplying the volume of each reagent (listed above) by the number of samples used for the experiment. A 24 µl aliquot of the Master Mix was distributed into each 0.5 ml sterile tube. These steps were followed under sterile conditions, within a dedicated DNA-free laboratory, so the template DNA was added to the reaction in the main laboratory area. A negative control containing 1 µl ddH₂O instead of template DNA was also PCR amplified to check for contamination. The reaction mixes were overlayed with 20 µl of mineral oil (Sigma) to prevent evaporation. The target sequences were amplified in a Perkin Elmer DNA Model 480 Thermal Cycler set to the following conditions: 30 cycles of denaturation at 94°C for 30 seconds, primer annealing at 55°C for 30 seconds, extension at 72°C for 60 seconds, followed by 4°C soak.

2.4 Submarine agarose gel electrophoresis

The purity and quantity of the HVI and HVII PCR products were analysed by electrophoresis on 2% SeaKem[®] LE agarose gels (FMC Bioproducts: see Appendix One). A 5 µl aliquot of the PCR product from each sample plus 0.5 µl of gel loading buffer (Blue Juice, Life Technologies) was loaded into each well. One or more lanes were loaded with 3 µl of 123 bp ladder (Life Technologies) size standard to determine the relative size of the DNA products. The gel was run in Tris Borate Buffer (TBE:

Appendix One) containing $0.4 \mu\text{g ml}^{-1}$ ethidium bromide at 90 volts for 90 minutes. Polaroid® photographs of the gels were taken under UV illumination with a Polaroid® MP 4 Land camera using f5.6 aperture and 1 second exposure settings.

2.5 Detection of the 9-bp deletion using a 4% NuSieve® gel

A 4% NuSieve® agarose gel (see Appendix One) was used to distinguish between samples with and without the 9-bp deletion using the Region V PCR products. Gels containing $0.6 \mu\text{g ml}^{-1}$ ethidium bromide were run in Tris Borate (TBE) Buffer containing $0.4 \mu\text{g ml}^{-1}$ ethidium bromide. The gels were run at 80 Volts for 210 minutes. Under UV illumination it was possible to detect the presence and absence of the 9-bp deletion by the position of the band, in relation to a known size standard. To test the reliability of this method, 5 samples were sequenced. The sequence results matched the results obtained using the gel, indicating this method is suitable for detecting the 9-bp deletion.

2.6 Purification of PCR reactions using Low Melting point (LMP) gel

If the results from the HVI and HVII PCR reactions were satisfactory, a low melting point agarose gel was used to purify the PCR products. The remaining PCR product (~20 μl volume) was loaded on to a 1% NuSieve® agarose gel (see Appendix One) with 2 μl of loading buffer. The gel contained $0.6 \mu\text{g ml}^{-1}$ ethidium bromide and was run in Tris Acetate (TAE) Buffer (see Appendix One) containing $0.4 \mu\text{g ml}^{-1}$ ethidium bromide. The gel was run at 80 Volts for 45 minutes and the resultant bands of HVI or HVII DNA were excised under UV illumination.

2.7 Purification of excised amplified dsDNA products

DNA Purification was performed using various methods as outlined below. The most successful purifications were done using MinElute™ spin columns, which were used for all subsequent samples.

2.7.1 Prep-a-Gene

The manufacturer's (Biorad) instructions were followed, using reagents contained in the Prep-A-Gene kit. Briefly, the excised gel band containing the target DNA was weighed and 3 times its weight in Binding Buffer (Biorad) was added. The gel was melted at 50°C for 2-3 minutes. When completely melted 5 µl of vortex mixed matrix (Biorad) was added, and left to rotate at room temperature for 10 minutes on a slowly turning wheel. The wheel allows the matrix to bind to the DNA, by repeated inversion. The DNA and matrix were spun in a Sorval MC12V microcentrifuge at 10,000 rpm for 30 seconds to produce a pellet. The supernatant was removed and the pellet washed with 125 µl of Binding Buffer. This step was repeated twice using Wash Buffer (Biorad) instead of the Binding Buffer. A final spin with no added Wash Buffer was done to ensure that all Wash Buffer had been removed. Finally 10 µl of Elution Buffer (Biorad) was added to release the DNA from the matrix. The tube containing the matrix, DNA and Elution Buffer was incubated for 5 minutes at 50°C and then spun at 5,000 rpm for 30 seconds to pellet the matrix. The eluted supernatant containing the DNA was carefully removed and put in a clean 1.5 ml plastic microcentrifuge tube, with care taken to avoid contamination from the pelleted matrix. This elution step was repeated with another 10 µl of Elution Buffer and combined for a total yield of ~20 µl of DNA.

2.7.2 Concert™ Rapid Gel Extraction

The manufacturer's (Life Technologies, Gibco BRL®) instructions were followed using reagents contained in the Rapid Gel Extraction Kit. Briefly, 30 µl of Gel Solubilization Buffer (L1: Life Technologies, Gibco BRL®) was added for every 10 mg of excised gel and incubated at 50°C for up to 15 minutes, mixing every 3 minutes to ensure complete gel dissolution. After the gel appeared to be dissolved the solution was incubated for a further 5 minutes. Next a spin cartridge (Life Technologies, Gibco BRL®) was placed into a 2 ml wash tube (Life Technologies, Gibco BRL®) and the dissolved gel solution was placed into the spin cartridge. This was spun at $\geq 12,000g$ in a Sorval MC12V for 1 minute and the flow through discarded. A further 500 µl of Gel Solubilization Buffer was added and the wash tube was incubated at room temperature for 1 minute, followed by centrifugation at $\geq 12,000g$ using a Sorvall MC12V microcentrifuge, with the flow through discarded. A further wash followed, using 700 µl of Wash Buffer (L2: Life Technologies, Gibco BRL®), incubated for 5 minutes at room temperature, centrifuged at $\geq 12,000g$ for 1 minute and discarding the flow through. The spin columns were then spun for a further minute to remove any residual Wash Buffer. Finally, the columns were placed in a 1.5 ml recovery tube, and the DNA was eluted with 20 µl of warm (50°C) TE buffer (see Appendix One) applied to the centre of the spin cartridge. After incubation at room temperature for 1 minute, the column was spun for 2 minutes eluting 20 µl of DNA.

2.7.3 MinElute™ spin columns

Manufacturer's (Qiagen®) instructions were followed using reagents contained in the MinElute kit. Briefly, the excised DNA band was weighed and 3 µl/mg of QG Buffer

(Qiagen[®]) was added (e.g. gel slice = 50 mg then 150 μ l of QG Buffer was added). The samples were incubated for 10 minutes at 55°C until the slice was completely dissolved. A check was made to ensure that the solution was yellow, if not then 10 μ l of 3M NaAc was added. Isopropanol was then added (1 μ l/mg of gel) and mixed gently by inversion (e.g. gel slice = 50 mg then 50 μ l of isopropanol was added). The contents of the 1.5 ml plastic microcentrifuge tube were transferred carefully to the MinElute™ columns and then spun using a Sorvall MC12V microcentrifuge for 1 minute at 10,000 rpm. The flow through was discarded and 500 μ l of QG Buffer was added. The tubes were centrifuged for 1 minute at 10,000 rpm and the flow through again discarded. The DNA bound in the tubes was then washed using 750 μ l of PE Buffer (Qiagen[®]), and spun at 10,000 rpm for 1 minute. The flow through was discarded and the tubes were spun for a further minute to ensure that all traces of the wash buffer were removed. Finally, 10 μ l EB Buffer (Qiagen[®]) was used to elute the purified DNA. After incubation at room temperature for 1 minute the DNA was spun at 10,000 rpm into clean plastic 1.5 ml microcentrifuge tubes.

2.8 Molecular weight and concentration of purified DNA

The concentration and molecular weight of the purified DNA was determined by electrophoresis on a 1% SeaKem[®] LE Agarose gel. Gels containing 0.6 μ g ml⁻¹ ethidium bromide were run in TBE Buffer containing 0.4 μ g ml⁻¹ ethidium bromide. Samples were prepared using 2 μ l of purified dsDNA product, 3 μ l of ddH₂O and 0.5 μ l of loading buffer. To provide a size and concentration standard 4 μ l of Low mass ladder (Life Technologies) was loaded in one or more lanes, alongside the DNA samples and run at 90V for 45 minutes.

For MinElute™ purified dsDNA products, samples were prepared for electrophoresis using 1 µl DNA, 4 µl of ddH₂O and 0.5 µl of loading buffer. Samples were run on 1% SeaKem® LE Agarose gels as described above.

After electrophoresis, gels were photographed and visually compared to the Low Mass Ladder. From this result, it was possible to estimate the concentration of the purified DNA and calculate suitable quantities to use for cycle sequencing reactions.

2.9 DNA Cycle Sequencing Protocol

The standard 20 µl Cycle Sequencing Reaction contained:

- 8.0 µl Big Dye Terminator Mix™ (Applied Biosystems).
- 3.2 µl Primer (diluted to 1mM).
- ? Purified DNA Template (30-90 ng, see Section 2.7).
- ? ddH₂O (depends on volume of DNA Template added).
- 20 µl Total (* for half reactions reduce all volumes by half).

The reactions were overlayed with 20 µl of mineral oil to prevent evaporation. The cycling parameters were: 25 cycles of 94°C for 30 seconds, 50°C for 15 seconds and 60°C for 4 minutes.

2.10 Ethanol Precipitation of DNA

The cycle sequencing products were removed from under the mineral oil and $\frac{1}{10}$ volume of 3M NaOAc was added and mixed gently on ice for a few minutes. Twice the total volume of cold ethanol was added and mixed gently by inversion. The samples were kept at 4°C overnight.

Some later samples were processed using half reactions (see Section 2.9*) and Big Dye Terminator™ Mix version III (as opposed to version II) in the cycle sequencing reaction. This required a slightly revised ethanol precipitation procedure, using 1.5 µl of 3M sodium acetate (NaOAc), 31.25 µl of 95% Ethanol and 7.25 µl of deionised water, together with the 10 µl cycle sequence product. This was left to stand for no more than an hour as any longer allowed unwanted primer to precipitate out of the solution, producing ambiguous sequences.

2.11 Automated DNA Sequencing Gel

Sequencing plates were prepared by soaking in 3M HCl for 10 minutes. The acid was thoroughly washed off using cold water and the plates were scrubbed vigorously by hand using Alconox (Alconox Inc.®). The detergent was washed off the plates using hot water, rinsed with ddH₂O and left to dry. Once dry, the plates were polished with ethanol and then assembled.

Once the plate assembly was set-up the 4% Polyacrylamide gel was made using a Long Ranger sequencing gel kit (BioWhittaker Applications: BWA). The manufacturer's (BioWhittaker Applications: BWA) instructions were followed, using the reagents

supplied in the self-contained pouch. Briefly, the black clip was removed and the contents of each half of the pouch were mixed thoroughly by hand for 1 minute. This pouch was shaken at medium speed (Series 25 Incubator, New Brunswick Scientific) for 5 minutes, and then the pouches were again mixed by hand for 1 minute. After further shaking for 5 minutes the red clip was removed and the contents of this part of the pouch was mixed in with the rest for 1 minute. Finally, the white clip was removed and the contents of the pouch were squeezed towards the filter embedded in the pouch, and out through this filter into a beaker. The liquid polyacrylamide was drawn out of the beaker into a syringe and dispensed in between the assembled sequencing plates. The gel was then left to set for at least 2 hours.

The DNA samples were precipitated out of solution (see Section 2.9), and prepared for loading onto the polyacrylamide gel, by spinning at 10,000 rpm using a Jouann MR1812 centrifuge (4°C) for 30 minutes. The supernatant was removed and 100 µl of 70% ethanol was added. After a further spin for 10-15 minutes the supernatant was again removed and the samples were allowed to air dry for 30-60 minutes. Once the DNA pellet was dry, loading buffer was prepared (see Appendix One) and 6 µl of the loading buffer was added to the pellet (3 µl for a half reaction). The samples were vortexed vigorously to ensure the DNA pellet was fully dissolved in the loading buffer, and then spun briefly to ensure enough sample was available to load onto the polyacrylamide gel.

The ABI Prism Sequencer 377 was prepared for the sequencing run by loading the gel, filling the appropriate tanks with 1X TBE (see Appendix One), and starting the computer to run a plate check and pre-run. When the gel had reach 47-48°C the samples were prepared for loading. The dsDNA samples were heated to 95°C for 2 minutes, to

denature the DNA to single strands, and then placed on ice to keep their denatured state. The samples were loaded into alternate wells using a duck billed tipped pipette. While these samples were run into the gel, the second batch of samples were heated to 95°C for 2 minutes and kept on ice. These samples were loaded into the remaining alternate wells, and run into the gel for 2 minutes. Once all the samples were loaded, the pre-run was stopped and the sequencing run was commenced. The run lasted for 6 hours at 2.4 kV.

2.12 Data analysis

2.12.1 Determining suitable parameters for analysis of the data used in the founding population simulations

Before simulations to estimate the number of female founders for the current Maori population could occur, a number of decisions had to be made regarding the treatment of the data. Below is a summary of the parameters that were used to select the most appropriate datasets for this simulation. The parameters selected for use in the simulations are in bold and the justifications for each selection are mentioned briefly in italics, these decisions will be discussed more fully later (see Section 4.4.1).

A. What region(s) of the mitochondrial sequence should be included in the simulation dataset?

- Long DNA Sequences (Whole HVI and HVII if available)
- **Short Sequences** (3' section of HVI, from nt position 16 189 onwards)

Maximises sample numbers in dataset

B. Which dataset should be used for the Maori founder analysis?

- Dataset from current study (AW02) only: to be used as a comparison with the Murray-McIntosh et al. 1998 (MM98) dataset
- AW02 and MM98 datasets combined
- **All available data combined (i.e. including those omitted from MM98 dataset)**

Maximises size of dataset

C. Which Eastern Polynesian (EP) Haplotypes should be used for the frequency comparison?

- Only EP (as used in MM98)
- **EP+ (all haplotypes found across all studies i.e. including those omitted from MM98)**

Best dataset, includes all possible comparable data

D. Is a dataset constructed to decisions A-C above currently suitable for analysis?
Answer: No it isn't as some haplotypes that are found in the Maori samples are not found in the EP samples, therefore what are the options?

- do nothing (this option is rejected as no simulation is possible)
- Assume new Maori Haplotypes have arisen by mutation post settlement of Aotearoa
- **add missing EP haplotypes (Discovery Curve assumption)**

Minimal assumption with respect to structure of EP dataset and requires no speculative estimates of mutation rates

E. What Population Growth Model should be used?

- Exponential (as used in MM98)
- **Sigmoid (3 types of population growth):**
 - Early expansion
 - Middle expansion
 - Late expansion

Sigmoid model is more plausible population growth model

2.12.2 Haplotype identification

The mitochondrial haplotype for each of the 85 samples used in this study was determined by examining the nucleotide sequence in the HVI region and presence or absence of the 9-bp deletion in Region V of the mtDNA. The haplotype for each sample was compiled into a database, and these results (see Table 3.1) were used for further analyses. A table containing the combined haplotype data from this study and the study by Murray-McIntosh et al. (1998) was also compiled (see Table 3.2) as a resource for future research. However, for the population growth modelling simulations this combined table is not appropriate, as the different studies summarised in the MM98 table focus on different regions of the mtDNA (i.e. some studies obtained haplotypes by sequencing the entire HVI and/or HVII region(s), while others only focused on the 3' HVI region from nt position 16 189 onwards). The MM98 table had also omitted some of the Group III haplotypes found in previous studies. Consequently, a third table was compiled (see Appendix Five), using the sequence information from all of the studies that have been conducted on Oceanic populations. From this table, a new summary table (see Table 3.3) was compiled making sure that the same region of the mtDNA was used in each study (i.e. the 3' HVI region from nt position 16 189 onwards, see Figure 1.15), to ensure that the data were comparable across studies. This revised combined data shown in Table 3.3, is used in the computer simulations. Mitochondrial HVII sequences were obtained from a total of 29 samples and were not used in the simulations.

Mitochondrial DNA sequence data was obtained for the HVI region using an ABI Model 377 Prism automated DNA sequencer, in conjunction with Perkin-Elmer Big Dye™ Technology. The results of the sequencing run were automatically compiled by

the ABI software (DNASar Inc.), and were available as a file of nucleotide sequences for each sample. An electropherogram (see Figure 3.5) was also generated for each sample and was used to assist with manual editing of DNA sequence files, specifically, to identify ambiguous or erroneous machine assignment of nucleotides and exclude areas of sequence at the start and end of each run that were particularly subject to reading/assignment errors. The sequences were edited manually using the electropherogram and the EditSeq™ facility in the Lasergene 99 DNASar Inc. package. Each edited sequence was aligned with the CRS using the SeqMan™ facility within the DNASar Inc. Lasergene 99 package. Aligned sequences were examined to identify nucleotide positions that differed from the reference sequence (Anderson et al., 1981). Each sequence was assigned a haplotype # based on its nucleotide sequence, compiled into a table (see Table 3.1) and used for further analysis in the computer simulations (see Section 2.12.3). Although mtDNA haplogroups are defined (see Figure 1.13) there is no universal nomenclature for the naming of mtDNA haplotypes. This means in most cases that haplotype numbering is not equivalent across studies. Haplotypes found in the present study have the prefix AW (see Table 3.1), whereas those found by Murray-McIntosh et al. (1998) have had the prefix MM added. Adding a prefix ensures that haplotypes are identified by the haplotype number and the study i.e. AW6 is haplotype 6 in the current study, this is distinguished from MM6, haplotype 6 in the study by MM98, these mtDNA sequences are not equivalent, instead AW6≈MM12 (see Table 3.2). The haplotypes in the combined AW02 and MM98 dataset have the prefix AWMM (see Table 3.2), whereas the complete combined dataset haplotypes have the prefix CD (see Appendix Five).

Some sequences from the current study were excluded from the computer simulation as the samples were of Western Polynesian, Melanesian, or unspecified origin. Other

samples were excluded if they had additional sequence variation in the 5' HVI region, to ensure that same region of mtDNA was being examined across all studies.

2.12.3 Estimate of the number of Maori females required to found New Zealand's (Aotearoa) Maori population.

The number of Maori females required to found the current Maori population was estimated based on the approach developed by Murray-McIntosh et al. (1998). Briefly, a three-step simulation was done based on the observed (or estimated) haplotypes frequencies in Eastern Polynesia (see Table 3.3) and in New Zealand. On Table 3.3, the combined data show there are a total of 21 haplotypes in the Maori and Eastern Polynesian samples. Some of these haplotypes are found in both Ma and EP samples, others are found only in the Ma or only in the EP samples. However, for the purpose of this simulation it is assumed that all haplotypes found in the Ma population are also to be found in the EP population, but that they are not present on the table as they are rare and haven't been discovered in the EP populations yet (Discovery Curve Assumption). All such undiscovered EP haplotypes are given an estimated occurrence of 1. The first step of the simulation selected a maternal founding population of between 2 to 1000 people randomly with replacement, from the frequencies of the 21 haplotypes observed in eastern Polynesia (see Table 3.3). Therefore the initial frequencies of the 21 haplotypes based on Table 3.3 is: 72, 13, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1 (see Appendix Three). The second step of the simulation allowed this maternal founding population to expand using one of three different sigmoid (logistic) growth models (see Appendix Two) over 30 generations to reach 50,000 females (corresponding to an overall population of ~100,000) and thus to give the haplotypes expected to be present in New Zealand Maori at the time of European settlement. The

third step of the simulation randomly selected 107 people from this expanded pre-European population. This enabled a comparison between the number of haplotypes observed in the simulations versus the number of haplotypes seen in the current Maori population (107 people, 8 Haplotypes: see Table 3.3).

The number and frequency of haplotypes was recorded at all three stages, the founding population, the estimated present population, and the final sample of 107 individuals. The simulations were repeated 20,000 times and performed by Dr Stephen Marshall (University Teaching Development Centre, Victoria University). For additional information or to request a copy of the program code please refer to the 'Founder' website: <http://www.vuw.ac.nz/~marshall/Programming/Founder/Founder.html>

3. Results

3.1 DNA extraction

Total genomic DNA was successfully extracted from whole blood provided by this investigator, and was of sufficient concentration and quality to allow PCR amplification and sequencing of the mtDNA HVI region. The sequence obtained was of European origin and differed from the Cambridge Reference sequence at 5 positions, of which two nucleotide transitions at positions 16 051 (G→A) and 16 189 (T→C), were also found in the Maori (Ma) and Pacific Nation (PN) samples (see Appendix Three, Sample 87). The other transitions (16 092, T→C and 16 362 A→G) and a single transversion substitution (16 129, G→C) were only found in the investigator's sample so contamination of results with sequences sourced from this DNA is considered unlikely.

3.2 Mitochondrial DNA 9-bp deletion Target Amplification

The 9-bp deletion target region (see Section 1.4.6.5) was PCR amplified from a total of 85 banked DNA samples. The products of each PCR reaction were assessed by 4% NuSieve[®] agarose gel electrophoresis (see Section 2.5). Figure 3.1 shows results obtained from analysis of 9-bp deletion PCR products. The expected sizes of the PCR products are either 112 or 121 base pairs (bp) depending on the presence or absence of the 9-bp deletion respectively. The migration distances of the PCR products correspond to fragment sizes of either ~115 or ~126 bp, as calculated using the 123 bp DNA standard ladder calibration curve (see Figure 3.2), and were considered to

represent the expected product sizes of 112 or 121 bp. Direct DNA sequencing of 5 amplified 9-bp deletion PCR products confirmed that the sequences were in fact either 112 or 121 bp long and these were in agreement with published results (see Section 4.2). Therefore, for all subsequent samples, assignment of the 9-bp deletion was based only on direct comparison results of 4% agarose gel electrophoresis.

Presence or Absence of the 9-bp deletion as visualised using a 4% NuSieve® Agarose Gel

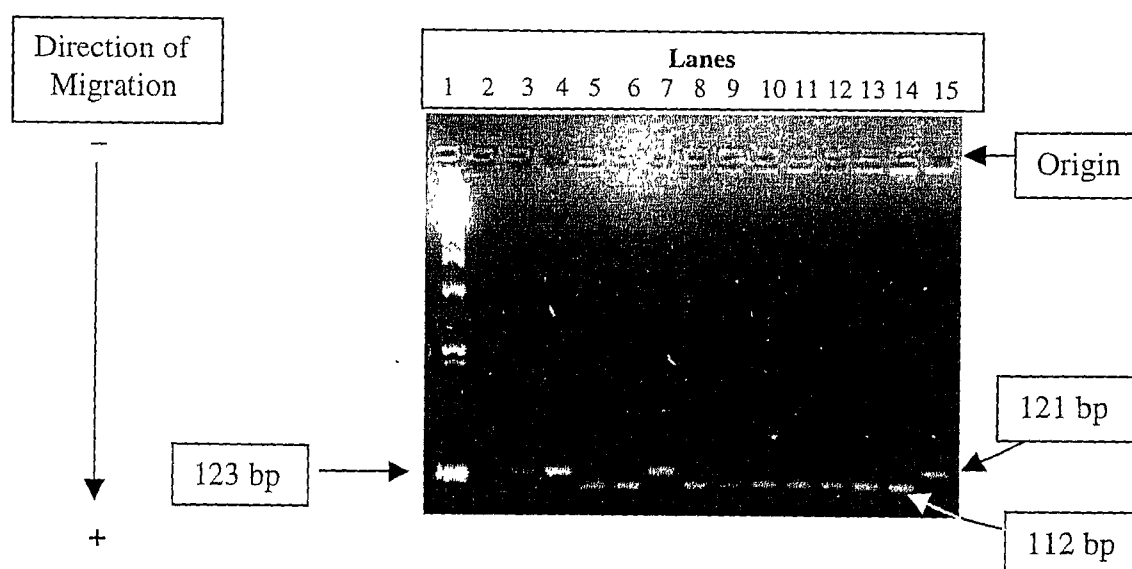


Figure 3.1 Results show 5 μ l aliquots of representative 9-bp deletion PCR products amplified from 13 samples and visualised on a 4% NuSieve® agarose gel.

- Lane: 1. 123 bp DNA standard ladder
 2. Negative reagent control (no template DNA)
 3. Sample #273, positive control for 121 bp (i.e. without 9-bp deletion)
 4, 7 and 15. Samples without the 9-bp deletion; #421, 471, 604
 5, 6, 8 and 9. Samples with the 9-bp deletion #457, 469, 608, 609
 10. Sample #610, positive control for 112 bp (i.e. with 9-bp deletion)
 11-14. Samples with the 9-bp deletion #611, 612, 616, 617

Calibration Curve for 9-bp deletion

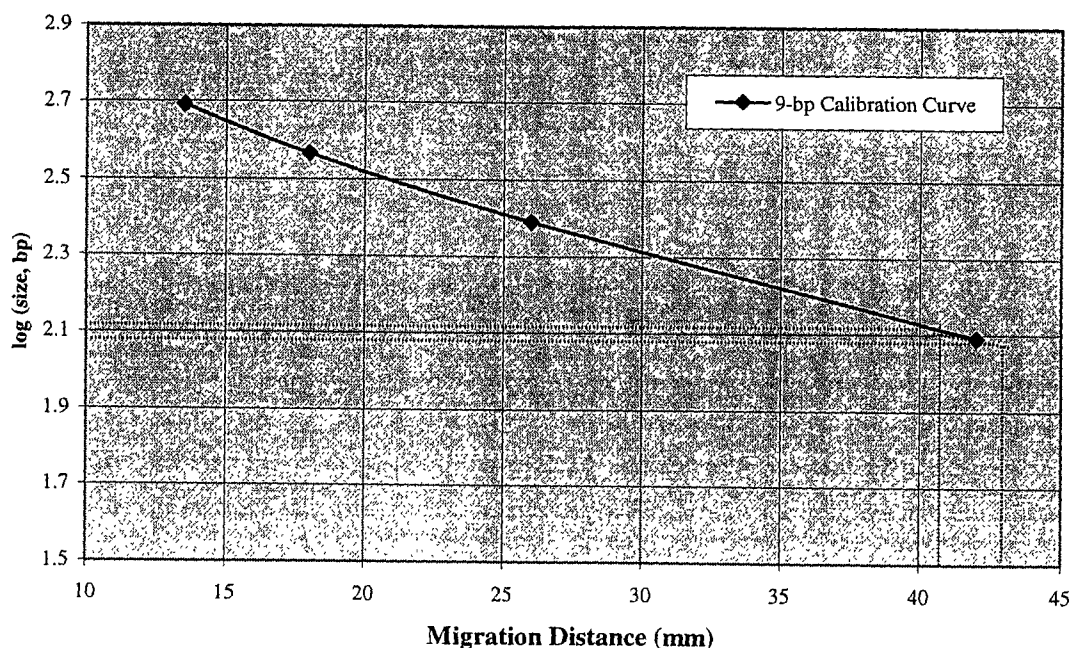


Figure 3.2 Calibration curve of log (size, base pairs) versus migration distance (mm) of the fragments that make up the 123 bp DNA standard ladder, shown in Figure 3.1. The migration distance of PCR products was measured and their corresponding sizes estimated from calibration curves such as this one. The actual examples marked on this curve are the products in lanes 4 and 5 of Figure 3.1. From their migration distances of 41.5 mm and 44 mm respectively product sizes of ~126 bp and ~115 bp were estimated. These sizes approximate their expected sizes of 121 and 112 bp (see text). The calibration curve was constructed using Microsoft Excel 2000.

3.3 Mitochondrial DNA HVI Target Amplification

The mitochondrial HVI target region was PCR amplified, from a total of 85 samples (61 Ma and 24 PN) using the primers and PCR protocol as outlined in methods (see Section 2.3). The products from each PCR reaction were assessed by agarose gel electrophoresis, which revealed varying degrees of success. A successful PCR product

appeared as a single, sharp band, of the expected size (~440 bp), on the agarose gel. Figure 3.3 shows an agarose gel electrophoresis of eleven HVI region PCR products. In this image there is a single bright band representing the major products and slight smearing below these bands, representing a mixture of minor lower molecular weight contaminants.

The migration distances of the PCR products correspond to a maximum DNA fragment size of ~492 bp (major band) and a minimum of ~407 bp (including the 'smear') estimated from the 123 bp DNA standard ladder calibration curve (see Figure 3.4). These maximum and minimum values bracket the expected size of ~442 bp and were considered to indicate that the PCR reactions were successful in targeting the HVI region of interest. When sequenced all of the 11 PCR products seen in Figure 3.3 were ~442 bp long, and were successfully aligned with the Anderson reference sequence (data not shown). The alignments were used to produce a table of results that will be presented later (see Section 3.5.2).

The high intensities of the major bands in Figure 3.3 reflect high DNA yields of the PCR product (>50 ng/μl, estimated). The major band in lane 6 is thinner and not as bright as the other bands, which may represent slightly lower DNA product yield. The negative control (Lane 3) contained no template DNA and the absence of a band in this lane reflects the absence of a PCR product, indicating there was no DNA contamination of the reagents used in these reactions.

HVI target region PCR products as visualised using a 1% SeaKem[®] Agarose Gel

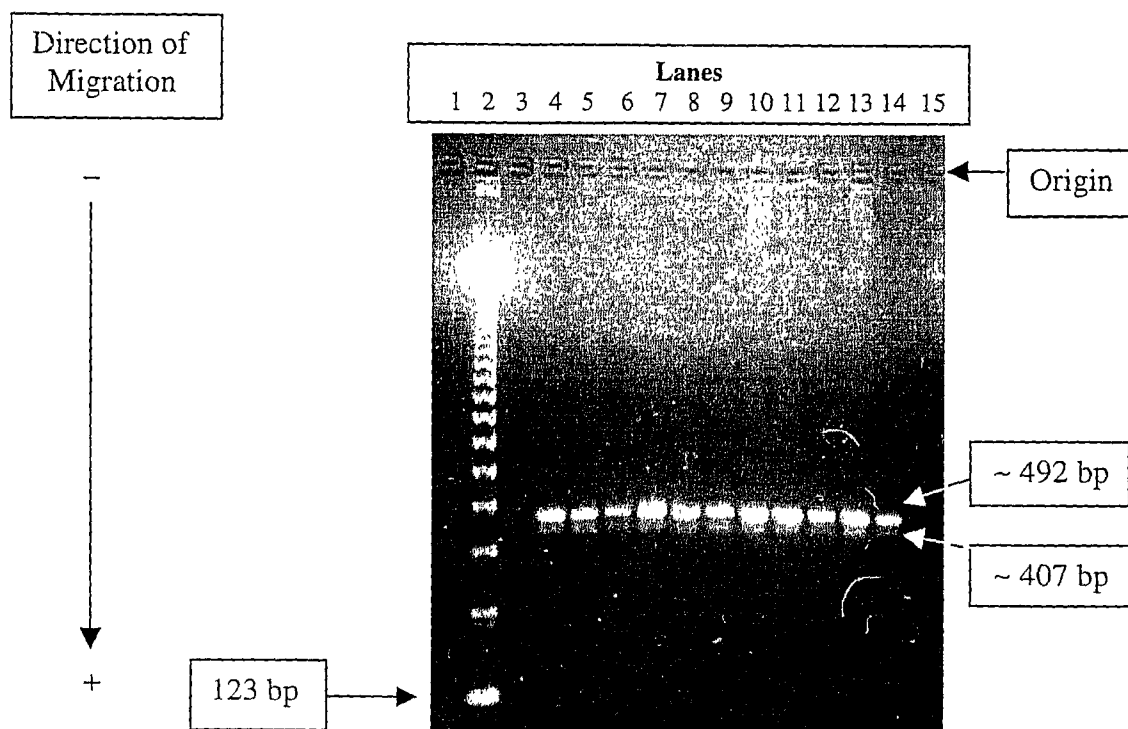


Figure 3.3 Results above show 5 μ l of PCR products from the HVI region of human mtDNA, amplified from 11 samples, visualised on a 1% agarose gel.

Lane: 1. No Sample

2. 123 bp DNA standard ladder

3. Negative reagent control (no template DNA)

4-14. Samples #418, 420, 422, 426, 433, 436, 479, 480, 484, 500, 501

15. No sample

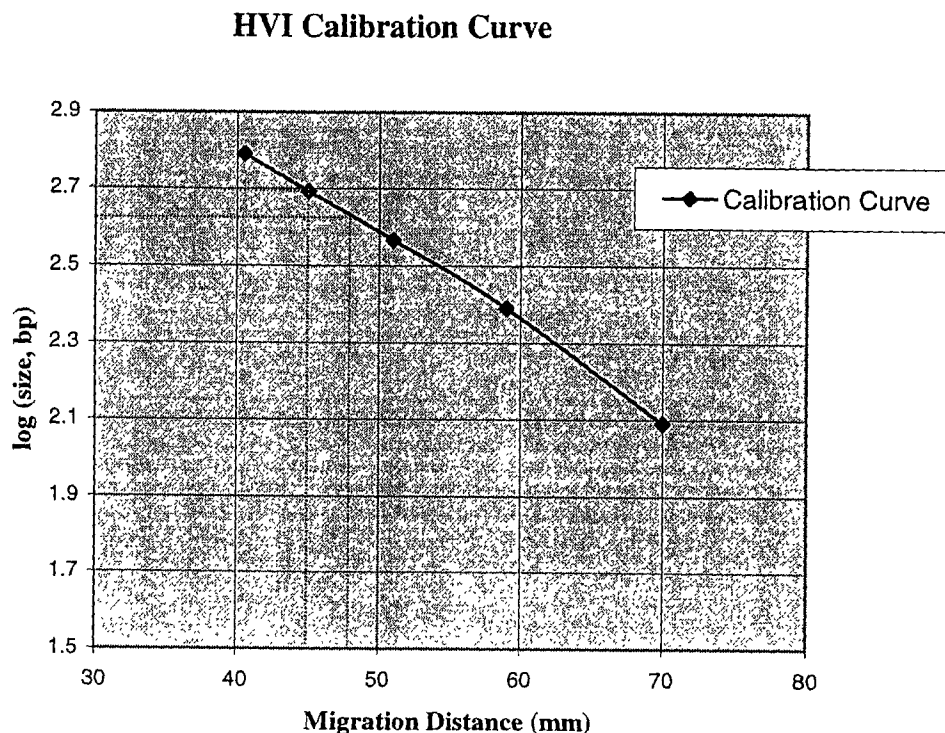


Figure 3.4 Calibration curve of log (size, base pairs) versus migration distance (mm) of the fragments that make up the 123 bp DNA standard ladder, seen in Figure 3.3, Lane 2. The examples seen on this curve are for the maximum (major band) and minimum ('smear') size products seen in lane 14 of Figure 3.3. The migration distances of 45 mm and 48 mm respectively infers a maximum size of ~492 bp and a minimum of ~407 bp respectively. These values bracket the expected size of 442 bp. The calibration curve was constructed on Microsoft Excel 2000.

3.4 Mitochondrial DNA HVII Target Amplification

The mitochondrial HVII region was PCR amplified, from a total of 38 samples (26 Ma and 12 PN), and the products were assessed for quality and quantity using 1% agarose gel electrophoresis (results not shown: see Appendix Four). Each sample produced a

single PCR product approximate to the expected size for HVII (~400 bp) as assessed using calibration curves (results not shown: see Appendix Four). Therefore, these PCR products were considered appropriate for sequencing (see Section 3.5.3).

3.5 Mitochondrial DNA Sequencing

Uncontaminated, appropriately sized PCR products were obtained from the HVI and 9-bp deletion target regions for all 85 samples. The exact sizes and sequences of all 85 PCR amplified samples were determined for the HVI region using an ABI™ Prism 377 Automatic DNA Sequencer (Perkin-Elmer), and the presence or absence status of the 9-bp deletion was inferred from the results of a 4% gel electrophoresis (see Section 3.2). The HVII target region was amplified and sequenced from a total of thirty-eight individuals to examine nucleotide transitions and transversions in this less well studied region of the mitochondrial genome. The sequence results from successful experiments will be added to GenBank, for access by the scientific community, after work has been submitted for publication in a peer-reviewed international journal.

The sequencing protocol used in this experiment produced an electropherogram (see Figure 3.5), which revealed the nucleotide sequence for each sample. All samples with the 'Polynesian CGT motif', also had a T→C transition at nucleotide position 16 189. This transition results in a homopolymeric C-tract (see Section 1.4.6.4.2) followed by a characteristically blurred sequence from this point onwards (see Figure 3.5, B). Consequently, a decision was made for this project to focus on acquiring HVI sequences from the 3' HVI region (nucleotide position 16 189 onwards, up until 16 360). This decision was made in order to encompass the region of greatest sequences diversity

within the HVI region including the 'Polynesian CTG motif', and to bypass the need to figure out complicated ambiguous sequence data produced by DNA polymerase stutters on the C-tract (see Section 4.3.2). Therefore, for most samples, only primer H16401 was used in the DNA Cycle Sequencing Protocol (see Section 2.9). However, 17 samples were also sequenced in the opposite direction using primer L15996, to see if there was any extra variation in the 5' HVI target region (see Figure 1.15).

Although the primary reason for the unreadable sequences before nucleotide position 16 189 was due to stutters caused by the homopolymeric C-tract, additional tract-length variation was created by transversions (A→C) at nt positions 16 182 and 16183 (i.e. directly preceding the tract). After careful examination of numerous electropherograms and comparing the results of samples that had been sequenced in both directions, it was concluded that most Polynesian subjects have two A→C transversions 5' before the tract (at nt positions 16 182 and 16183), meaning that in most cases the tract was estimated to consist of 12 C's in succession (see Table 3.1). Sample 601 appeared to have only 11 C's after H strand sequencing (i.e. does not have the A→C transversion at nt 16 182) and additional sequence information using the L strand supported this idea, although it is impossible without cloning to know for sure whether this is an accurate result or an artefact of the sequencing process. Other samples that revealed rare and/or novel haplotypes, were re-sequenced using either internal primer L16209 or H16164 to provide forward and reverse sequence for the unique HVI target area. Thus, novel sequences were examined at least once in forward and reverse orientations.

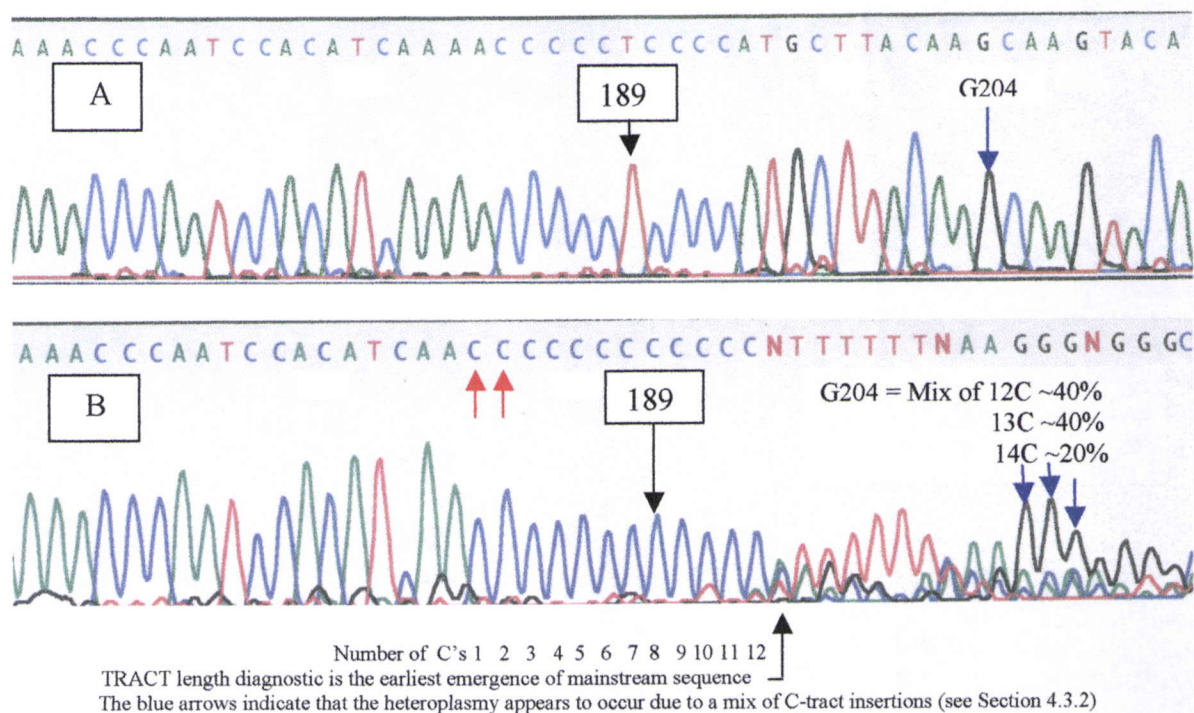


Figure 3.5 Electropherograms showing the results from samples with (A) and without (B) the homopolymeric C-tract. In B, the presence of the T → C transition at position 16 189 and A → C transitions at nt positions 16 182 and 16183 (indicated by red arrows) leads to heteroplasmy and results in difficult nucleotide assignment from this region onwards. Nucleotide G204 (blue arrows) illustrates the stuttering effect of heteroplasmy whereby a mixture of sequences with a varying number of Cytosines in the C-tract causes 'unreadable' sequences.

3.5.1 Presence or Absence of the 9-bp Deletion

Out of 85 banked DNA samples, 76 have the 9-bp deletion (see Appendix Three), indicative of Polynesian ancestry. Samples lacking the 9-bp deletion were inferred to have been sourced from volunteers with part-European ancestry. This conclusion was tested by HVI region sequencing, which revealed that these individuals also lack the characteristic Polynesian CGT motif at positions 16 217 (C→T), 16 247 (A→G) and 16

261 (C→T) and instead have haplotypes corresponding with those that are found in European populations (Macaulay et al., 1999b).

3.5.2 Mitochondrial DNA HVI Sequence

As with previous studies using DNA samples obtained from individuals living in Polynesia (Murray-McIntosh et al., 1998; Sykes et al., 1995; Lum et al., 1994; Redd et al., 1995), very low haplotype diversity was found in both the Maori (Ma) and Pacific Nation (PN) samples. In total, only 13 haplotypes containing the ‘CGT’ motif were identified in this set of Maori and Pacific Nation samples (see Table 3.1). Out of the 61 Maori samples 8 CGT motif haplotypes were found, with an additional 7 haplotypes ruled out due to probable European ancestry (see above and Table 3.1), leaving a dataset of 54 Maori sequences for further analysis. Of the remaining 54 Maori sequences, 44 were identical (see Table 3.1, ht #AW1), and have been found in previous studies of Maori populations (Murray-McIntosh et al., 1998, ht MM1; Sykes et al., 1995, ht Sykes11; see Appendix Five). An additional 6 samples contained previously reported haplotypes: ht #AW2, AW3 and AW6 found 4, 1 and 1 time(s) respectively (see Table 3.1). This leaves 4 Maori samples that exhibit haplotypes that have not been found in previous studies (Haplotype #AW4, Sample 447; ht #AW5, Sample 468; ht #AW7, Sample 484; ht #AW8, Sample 510). Each of these unique haplotypes is rare, occurring only once each. Of these four unique haplotypes only 3 are used for further analysis, the 4th (ht #AW8, Sample 510) has been included as a subclass of haplotype CD2 (see Appendix Five), and was not used in the computer simulations. Leaving out this haplotype ensures that all samples included in the simulation have been scored across comparable nucleotides from position 16 189 onwards (3’ HVI region). Thus

although haplotype #AW8 is similar to haplotype #AW2 from position 16189 onwards, L strand sequencing shows haplotype #AW8 has additional variation in the 5'HVI region before 16 189 (at nucleotide position 16 051). It might bias the results to lump these sequences together into the single #AW2 category. This is an important consideration and illustrates the need for care in the treatment of data. In the MM98 study, mtDNA haplotype diversity may have been underestimated as some samples that had nucleotide transitions or transversions (compared with the reference sequence) in the 5'HVI region were grouped together with sequences without these changes simply because they have the same nucleotide sequences in the 3' HVI region (i.e. identical sequences from position 16 189 onwards). The consequence is a table with more people represented by fewer haplotypes.

Among the 24 Pacific Nation (PN) samples, a total of 9 haplotypes were found (see Table 3.1), of which 2 can be excluded due to probable European ancestry (ht #AW16 and AW21). Of the remaining 7 haplotypes (22 samples), one type is common and occurs in 13 samples (ht #AW1), this is the same common type as found in the Maori samples, and has also been reported in previous studies using PN subjects (Sykes et al., 1995; Redd et al., 1995; Lum et al., 1994). The other common haplotype (ht #AW 2) occurs 3 times, and was also found in previous studies. The remaining 5 haplotypes are all unique to the present study and are all rare, with one haplotype occurring twice (ht #AW9, Samples 602 and 616) and the other four only once each (ht #AW10, 613; ht #AW11, 624; ht #AW12, 611; and ht #AW13, 601). Only one of these samples (616) is of Eastern Polynesian origin and hence able to be included in later analysis. The other haplotypes are from samples scored from Western Polynesia (613, 611), Melanesia (601, 602), or of unspecified Polynesian origins (624). Although these have been excluded from the dataset used in later computer simulations run in the current study,

they have been included on the combined DNA sequence table (see Table 3.2) using the geographic categories of Murray-McIntosh et al. (1998), or excluded altogether due to lack of information (624).

3.5.3 Mitochondrial DNA HVII Sequence

Of the 38 DNA samples that were considered to have yielded adequate HVII PCR products suitable for sequencing, only 29 of them could be sequenced successfully. Of these the full sequence data for the HVII region of interest (~00 060-00 280) were obtained from just 13 samples (7 Ma and 6 PN), while the other 16 samples (11 Ma and 5 PN) yielded partial sequences. Previous studies have indicated that the HVII region is very homogeneous for Polynesian populations (Lum et al., 1994). A previous study by Redd et al. (1995) using the HVII region found a set of specific nucleotide changes that were very common in Maori and Pacific Nation populations. These were transitions at nucleotides 00 073 A→G, 00 146 T→C and 00 263 A→G. Out of the 13 full HVII sequences available (see Appendix Three) only 4 differ from this expected haplotype, and 3 out of the 4 samples had suspected European ancestry implied using HVI data. The remaining sample (507) had the CGT motif in the HVI target region, but did not have the expected HVII motif, as it was missing the transition at nucleotide position 00 146. Due to the small amount of HVII data available in this study and previous studies, no further analysis was done using this sequence information. The 13 full HVII sequences (see Appendix Three) will be deposited in GenBank for access by the scientific community.

Table 3.1 A summary of HVI haplotypes found in this study.

This table shows all the haplotypes found in the current study (AW02: for the haplotype of each individual sample please see Appendix Three). The numbering and reference sequence follow the CRS (Anderson et al., 1981). Dots indicate that the DNA sequence found in a haplotype is identical to the reference sequence at that nucleotide position, transition and transversion substitutions are shown as a changed nucleotide. Each haplotype has a number (#) assigned to allow easy reference e.g. ht #AW1 has transversions at nucleotide positions 16 182 (A→C) and 16 183 (A→C), and transitions at positions 16 189 (T→C), 16 217 (T→C), 16 247 (A→G) and 16 261 (C→T), it is also the most common haplotype occurring in 44/54 Maori and 13/22 Pacific Nation samples. The TRACT column refers to the presence of the homopolymeric tract, with the number referring to the length of the tract (either 11 or 12 C's). The presence or absence of the 9-bp deletion is indicated by a tick or a cross respectively.

Table 3.2 Haplotype data summary incorporating information contained in Table 1 (Murray-McIntosh et al., 1998), and Table 3.1 (AW02).

This table combines information from Table One in the study by Murray-McIntosh et al. 1998 (MM98), with the information from Table 3.1 of the current study (excluding sequences with suspected European Ancestry and unknown PN island of origin). The numbering and the reference sequence follow Anderson et al. (1981). The numbers in the AW02 REF and the MM98 REF columns refer to the reference number of these haplotypes in the original AW02 or MM98 tables for easy reference to the original data. Table 1 from MM98 includes data from; Murray-McIntosh et al. 1998 and *selected* data from Sykes et al., 1995; Lum et al., 1994; Redd et al., 1995.

Variable sites in Hypervariable Region I (HVI) of combined AW02 and MM98 mtDNA data sets.

[illegible]

Anderson Reference Sequence				Ma	EP	WP	MN	A	A	A	A	T	A	G	C	A	C	A	C	A	C	A	C	T	A	T	A	T	A	C	C	G	T	C	MM REF			
Group II																																						
AWMM31	0	1	0	0																																MM23		
AWMM32	0	0	1	0																																MM24		
AWMM33	1	16	1	0																																MM25		
AWMM34	0	0	2	4																																MM26		
AWMM35	0	3	0	0																																MM27		
AWMM36	0	0	2	0																																MM28		
AWMM37	0	0	0	1																																MM29		
AWMM38	0	0	0	2																																MM30		
AWMM39	0	0	0	8																																MM31		
AWMM40	0	0	0	1																																MM32		
AWMM41	0	0	0	1																																MM33		
AWMM42	0	0	0	2																																MM34		
AWMM43	0	0	0	1																																MM35		
Group III																																						
AWMM44	0	0	1	0																																MM36		
AWMM45	0	2	0	0																																MM37		
AWMM46	0	1	0	0																																MM38		
AWMM47	0	0	1	0																																MM39		
Totals	108	113	81	88																																		

The MN total was taken from the MM98 paper and includes additional haplotypes from Sykes et al. (1995)
† Haplotype Groups are as defined in Lum et al. (1994)

Re-examination of the source data used in the MM study revealed that some sequences were combined in the MM table even though sequencing of the HVII region and/or before position 16 189 in the 5' HVI region had meant that these sequences were probably distinct haplotypes. For example, see Appendix Five: haplotypes CD1, CD1a, CD1b, CD1c, CD1d were combined into a single haplotype category (MM1) in Table One of the study by MM98, even though these haplotypes have clear sequence differences in either the HVII region, or in the 5' HVI region. Other sequences (mainly Group III) appear to have been left off the MM table entirely. The full combined data set for all studies conducted using samples from Maori, Eastern Polynesian and Western Polynesian samples with references to the original sources are shown in Appendix Five and these data have been summarised in Figure 3.3 (below) and were used for further analysis in the computer simulations.

Table 3.3 Total number of Haplotypes found in full combined dataset (i.e. MM98+) and AW02 data from the 3' HVI region (16 189 onwards).

Haplotype Number (see Appendix 5)	MM98+		AW02	
Group I	Ma	EP	Ma	EP
CD1	47	69	44	3
CD2	5	12	4	1
CD3	1	1	1	
CD7		1	1	
CD8		1		
CD12		1		
CD17			1	
CD18			1	
CD19			1	
CD20				1
Group II				
CD22		1		
Group III				
CD30		1		
CD31		1		
CD32		1		
CD33		1		
CD34		1		
CD36	1			
CD37		1		
CD41		1		
CD43		1		
CD44		1		
Totals (samples)	54	95	53	5
(haplotypes)	4	16	7	3

Table 3.3 combines all the sequence data from Polynesia (see Appendix Five), but omits sub-haplotype data e.g. CD1a, b etc. The omission of sub-haplotype data ensures that the data is homogenous, with each haplotype covering the same nucleotide positions (i.e. 3' HVI region). This table adds in some data that was left off the MM98 table (all the Group III sequences), and leaves out some data that was put on the MM98 table (i.e. haplotypes CD22a, 23a, 27 and 28, see Appendix Five).

Appendix Six contains a further summary of all the mtDNA haplotype data from Polynesia. This table is based on the table in the study by MM98, whereby some

sequences have been combined into a single category and other sequences (Group III) have been excluded from analysis. Comparison of Appendix Six with the Table 3.3, give different total numbers of haplotypes, and illustrates the importance of the data classification system. Table 3.3 has been used for the computer simulations as the 3' HVI region has been consistently used in all studies, and is therefore comparable data across studies.

3.6 Population Simulation Data Analysis

3.6.1 Capture-Recapture Estimates

The data summarised on Table 3.3 were used by Shirley Pledger (Senior Lecturer, School of Mathematical and Computing Science, Victoria University) to estimate the true haplotype abundance within Maori (Ma) and Eastern Polynesian (EP) populations, using capture-recapture estimates of species abundance. Using the information summarised on Table 3.3, i.e. EP: 15 haplotypes observed once, 1 ht observed 13 times and 1 ht observed 72 times (17 different haplotypes observed in total); Ma: 5 haplotypes observed once, 1 ht observed 2 times, 1 ht observed 9 times, 1 ht observed 91 times (8 different haplotypes observed in total); using Burnham and Overton's (1979) first-order jack-knife it was estimated that there are ~32 haplotypes in the EP population and ~13 in the Maori population. This indicates that there are further rare haplotypes still to be found within these populations and suggests maximum haplotype values which might be used in a population growth simulation to estimate the size of the Maori founding population.

3.6.2 Population growth simulation to estimate the size of the Maori founding population

The data summarised in Table 3.3 were used to estimate the number of female founders required to establish the current Maori population, using the computer simulation methods outlined in section 2.12.3. The exponential growth pattern used in the original MM98 study was considered inappropriate for the current study, as Brewis (1990) suggests that sigmoid curves are a more appropriate model for population growth. Therefore, three different sigmoid (logistic) population growth models were tested to see which model if any would be best to apply to this dataset. The first two sigmoid population growth models with early and middle expansion curves (see Appendix Two, Model A and B) gave good founder estimates, whereas the third (Model C) gave an estimate that was very high and considered inappropriate. For comparison a simulation using an exponential growth curve was also run (see Appendix Two, Model D).

The results shown in Figure 3.6 were produced using a sigmoid early population expansion model (Appendix Two, Model A). Inspection of the plot predicts that approximately 47 women (between 25 and 115) would have been necessary for the initial founding population to provide sufficient genetic input to account for the observed genetic diversity among mitochondrial haplotypes seen in the current Maori population (a combined total of 8 haplotypes in 107 people: see Table 3.3).

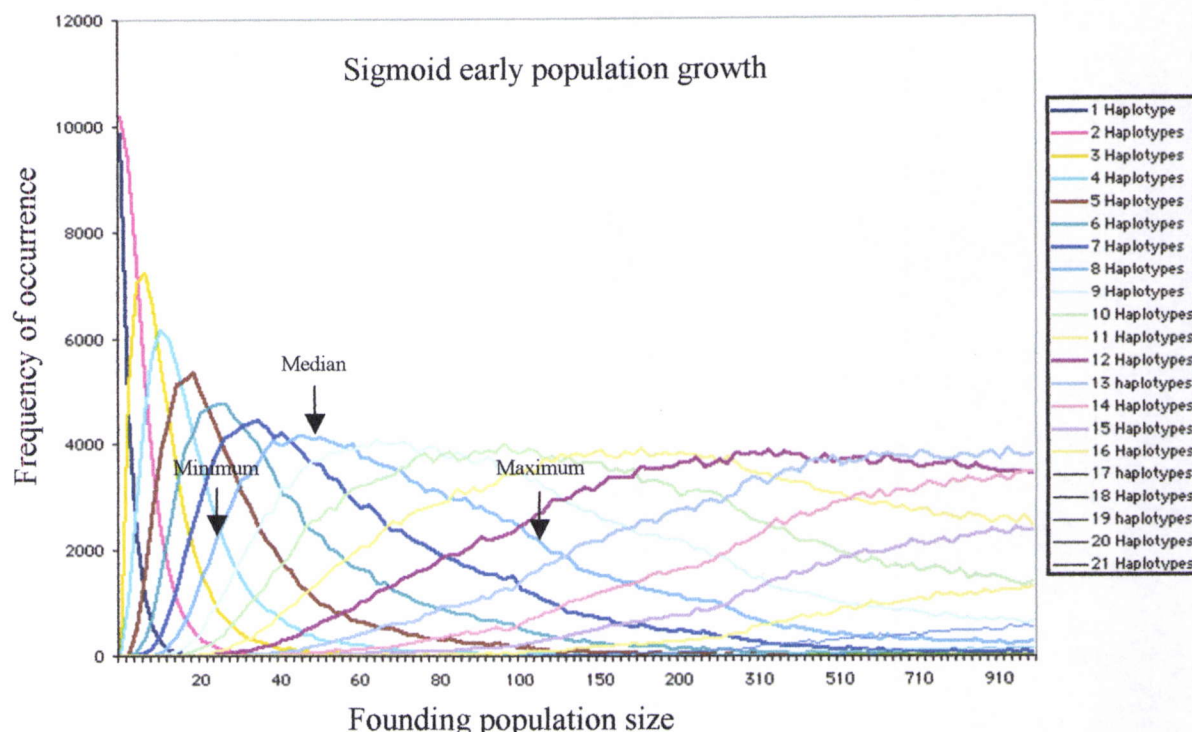


Figure 3.6 The results of a computer simulation using a sigmoid early population growth model (see Appendix Two, Model A) to predict that a median number of approximately 47 Maori female founders would have been required to found the current Maori population. The minimum, median, and maximum values are marked with black arrows.

The results shown in Figure 3.7 were produced using a sigmoid middle population expansion model (see Appendix Two, Model B). This model predicts that approximately 260 women (between 115 and 670) were required for the initial founding population to provide sufficient genetic input to account for the observed genetic diversity among the mitochondrial haplotypes in the current Maori population (a total of 8 haplotypes in 107 people, see Table 3.3). This number is larger than the estimate obtained using the early sigmoid population growth model, but is consistent

with some previous studies suggesting a maximum founding population of ~250 women (McGlone et al., 1994).

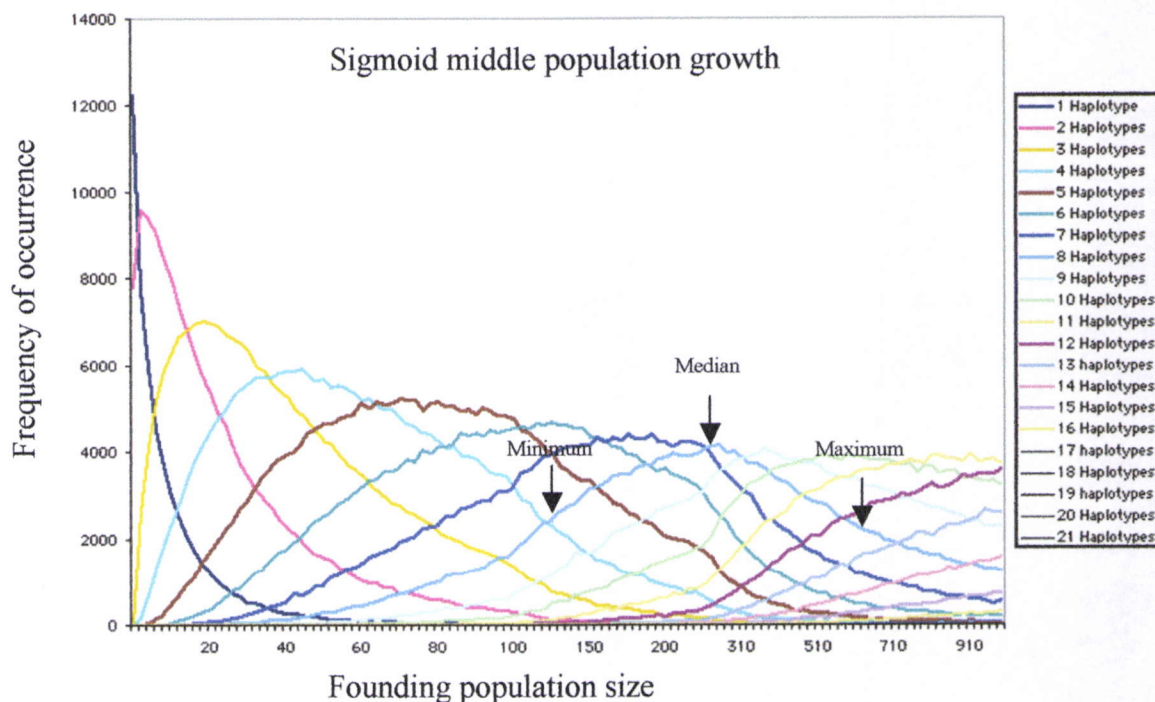


Figure 3.7 The results of the simulations using the sigmoid middle population growth model (see Appendix Two, Model B) predict that a median number of approximately 260 Maori female founders would have been required to found the current Maori population.

Finally, the results shown in Figure 3.8 were produced using the sigmoid late population expansion model (Appendix Two, Model C). This model predicts that approximately 510 women (between 230 and >1000) were required for the initial founding population to provide sufficient genetic input to account for the observed genetic diversity among mitochondrial haplotypes in the current Maori population (a total of 8 haplotypes in 107 people). This number is significantly larger than either estimate obtained using the

early or middle sigmoid population growth model. It is also outside the maximum number of settlers considered likely for this population. Therefore, this model is considered inappropriate for modelling the Maori population expansion.

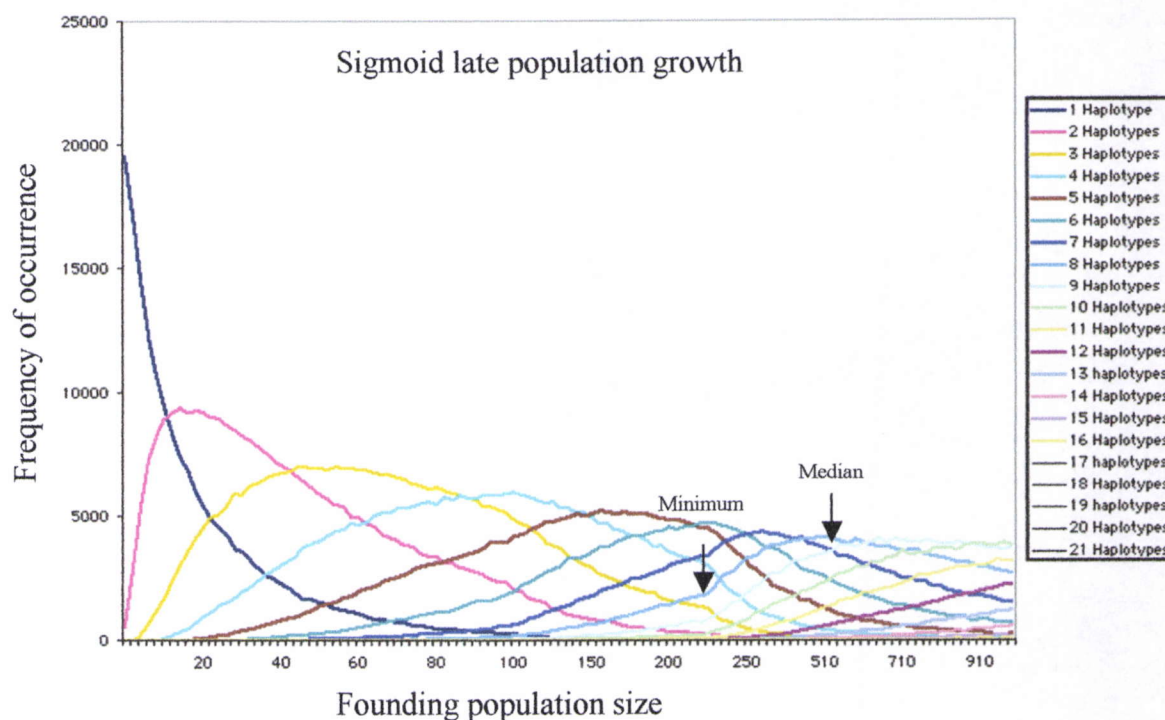


Figure 3.8 The results of the simulations using the late sigmoid population growth model (see Appendix Two, Model C) to predict that a median number of approximately 510 Maori female founders would have been required to found the current Maori population.

4. Discussion

4.1 DNA Extraction

The SDS/Proteinase K and phenol/chloroform extraction was successful in isolating DNA from the blood provided by this investigator. The sequence analysis results of this DNA extraction were used as a comparison against each of the other samples, to ensure that DNA from the investigator was not contaminating these results. The mitochondrial DNA sequence of the investigator was of Caucasian origin and differed at 5 nucleotide positions relative to the Anderson reference sequence (16 051, 16 092, 16 129, 16 189 and 16 362: see Appendix Two). The transition at nucleotide position 16 051 was only found in one other sample (510), but in conjunction with the CGT motif. It is unlikely for contamination to influence only a single nucleotide position i.e. 16 051, and if this result was due to contamination and was indeed derived investigator sequence or a mixture, then it is expected that all the other transitions (i.e. at positions 16 092, 16 189 and 16 362) and the transversion (16 129) would have also been present. This was not the case and implies that the observed transition at position 16 051 in sample 510 was not the result of contamination, and instead was an authentic result.

Another transition at position 16 189 was present in the mtDNA sequence of the investigator and was very common in this study, found in ~90% of samples, or 100% excluding those with suspected European ancestry. However, the high incidence of this transition has already been noted in previous studies of mtDNA HVI regions (Murray-McIntosh et al., 1998; Sykes et al., 1995; Lum et al., 1994; Redd et al., 1995) and was therefore expected. In all cases the Maori and Pacific Nation samples used in this study

which have the 16 189 transition also had the CGT motif and the 9-bp deletion indicating authentic Polynesian ancestry.

Additionally, contamination was rendered unlikely by the use of a negative control for each new PCR Master Mix and consequent set of PCR reactions. Each negative control contained the same Master Mix as the samples but with no DNA added, so a negative control would only produce a band if DNA contaminated the sample to provide a template. All negative controls used in this study did not show any DNA amplification and were therefore considered to be free of contamination.

4.2 Mitochondrial 9-bp, HVI and HVII Amplification

Successful PCR amplification of the target HVI and 9-bp deletion regions were achieved in all 85 samples. The HVII target region was successfully amplified in 29 samples. The ~20 ng/μl concentration of template DNA added to tubes for PCR reactions proved sufficient to allow for high yield amplification of the target mtDNA regions. The selected primer pairs (see Section 2.3.1) were effective in amplifying the DNA targets, and were specific, generating fragment sizes within the expected ranges.

The size of the PCR products was estimated using the 123 bp DNA standard ladder (Life Technologies) run in Lane 1 of the agarose gel during electrophoresis. Using the migration distances of the PCR products and a calibration curve constructed using the 123 bp ladder, it was possible to estimate the approximate sizes of the PCR products, providing a useful tool for assessing whether the correct region of DNA had been amplified.

In some cases, PCR product bands show different migration distances across lanes, producing the well known 'smile effect' (data not shown). This result might be taken to suggest that the products in the higher and lower numbered lanes are of a lower molecular weight than those products run in the middle of the gel, when in fact sequencing analysis shows them to be of identical size. This effect may have occurred because the direction of product migration in the gel may not have been parallel to the wells, probably due to the gel not being placed perpendicular to the electric current during electrophoresis. Uneven migration can also occur when a gel is not of homogenous density. However, the most likely explanation for the 'smile effect' is an electrophoresis voltage that is too high causing the centre of the agarose to melt slightly. A more accurate assessment of size during agarose gel electrophoresis could be achieved by running a 123 bp DNA standard ladder in both the centre and the extreme left and right lanes. The complication of migration direction in establishing the size of the PCR products was not a significant problem in this project as the purpose of the 123 bp DNA standard ladder was to provide an estimate, not an exact, assessment of fragment size. However, for the assignment of the 9-bp deletion haplotype it was more important to have the samples running evenly across the gel, as if the samples were not running straight, a single sample without the deletion may be difficult to identify. However, samples without the deletion (positive controls) were run in the middle and in the extreme left and right lanes of gels so scoring and 9-bp haplotype assignment was unambiguous.

The PCR amplification products from the HVI target region were visualised on a 1% agarose gel as distinct bright bands, with slight smearing after this band that probably represents lower weight molecular contaminants (see Figure 3.3). These lower weight products are ~407 bp and may have been caused by unspecific primer annealing, or they

may represent an overabundance of HVI target PCR products. Reducing the number of amplifications cycles, reducing the concentration of the template or a combination of both, may be ways to avoid this problem. However, corrective measures were not required as the small amount of smearing did not affect the sequencing reactions and all amplification products were successfully sequenced.

4.3 Sequence Analysis

The ABI™ Prism 377 Automatic DNA Sequencer (Perkin-Elmer) provided a reliable and efficient procedure for obtaining the exact sequences of the HVI, HVII and 9-bp deletion target sequences. Sequences obtained from the automated sequencer were edited and aligned with the Anderson reference sequence. For the HVI and HVII target regions complications in sequence assignment arose due to the presence of homopolymeric tracts.

4.3.1 9-bp deletion assignment

The assignment of the 9-bp deletion was successfully determined using 4% NuSieve® agarose electrophoresis. In all cases electrophoresis gave fragments that were only one of two possible sizes. The smaller fragments were taken to represent the deleted 112 bp fragment and the larger the undeleted 121 bp fragment. In each run a positive control known to have the 9-bp deletion and another without the 9-bp deletion were run with the unknown samples to allow comparison. Product bands were consistently separated well enough to allow 9-bp deletion assignment.

In all cases, samples without the 9-bp deletion also lacked the CTG motif, and had control region haplotypes that were common in European populations. This observation provides further evidence for the reliability of the electrophoresis results, as Europeans are unlikely to have the 9-bp deletion. Additionally, the assignment of the presence or absence of the 9-bp deletion was done before sequencing of the HVI region was complete, so interpretation of the electrophoresis results were not biased by expectation. Although the electrophoresis results were adequate to distinguish between the larger and smaller sized fragments, intermediate fragments, or slightly larger fragments may have been missed. A triplication of the 9-bp repeat has been observed in some populations (Chuckchi from Siberia: Shields et al., 1992 and Tharu from Nepal: Passarino et al., 1993 referenced in Redd et al., 1995). However, it is easy to identify the difference between samples with and without the deletion, even though the difference is only 9-bp, therefore, it is expected that samples with a triplication would be equally easy to identify. There was no evidence of triplication in any samples used for this study.

4.3.2 Mitochondrial HVI sequence

In the HVI region a homopolymeric C-tract occurs between nucleotide positions 16 182 and 16 193. In all cases, the presence of the homopolymeric tract within the HVI region results in a characteristically 'blurred' sequence that was unable to be interpreted by the automated sequencer. Manual sequence assignment based on the results of the electropherogram also proved difficult and all attempts to use these ambiguous data were abandoned. Instead, the results include the HVI region from nt ~16 182 onwards, as this 3' HVI region (see Figure 1.15) is more variable and the sequence can be easily interpreted.

However, concentrating on only the most variable 3' HVI region means that potentially valuable information from the 5' HVI has not been extensively studied or evaluated in Polynesian populations (including Maori). Although some studies have suggested that this region is very homogeneous in Ma and PN populations (Lum et al., 1994) a small number of samples were surveyed using primer L15996, to see whether any variation could be found. A total of 17 samples were sequenced using both primers H16401 and L15996 (see Section 2.3.1), together providing the entire HVI sequence for these samples (see Appendix Three). Samples with inferred European ancestry did not have the homopolymeric C-tract and therefore amplification and sequencing using primer H16401 alone was sufficient to read through this region and provide sequence information for both 5' and 3' HVI regions. Therefore, there are a total of 24 samples with sequence information for the entire HVI target region (5' and 3' HVI: see Figure 1.15), of which 15 were from Ma and PN samples and 9 are from samples with inferred European ancestry. Out of the 9 European samples, 3 had variation in the 5' HVI region before nucleotide position 16 189, compared with only 1 sample out of 15 for the Ma and PN samples. Despite the very small sample sizes, the comparative results above do indeed indicate that Maori and Pacific Nation populations may have a small amount of variation in the 5' HVI region when compared with the variation seen in Caucasian populations.

However, the same reduction in haplotype diversity between Polynesian and Caucasian samples also holds for the 3' HVI region. In the 3' HVI region there are only 11 different haplotypes observed out of a total of 74 samples (see Table 3.1). Whereas out of the 9 samples with inferred European ancestry there are 8 different haplotypes. Therefore, the question becomes "Would sequencing the 5' HVI region reveal further variation sufficient to warrant the extra time and money that would need to be spent on

this endeavour?” Ideally sequence information from the entire HVI target region would be required to ensure that the database is complete and more easily comparable with data from other populations worldwide that report sequence data from the entire HVI (i.e. both 5' and 3') region.

The rare Maori haplotypes found in the current study (Haplotype #AW4, AW5, AW7 and AW8: see Table 3.1) were compared with other populations worldwide, using the mtDNA haplotype database <http://db.eva.mpg.de/hvrbase/>. Although no exact matches were made, some similarities could be found between Haplotypes #AW4 and #AW8 and South American haplotypes (HISP_98 and CHI 32, 48 and 53: see Horai et al., 1993). A previous study by Sykes et al. (1995) found a very small number (<0.02%) of mtDNA haplotypes among Polynesians that *exactly* matched haplotypes described for South American populations (Haplotypes #45 and 47: see Sykes et al., 1995). Therefore, although Heyerdahl's (1950) theory for a substantial input from South America is not supported by most researchers, evidence from the study by Sykes et al. (1995) provides slight support for a tiny amount of gene flow. Some contact between Polynesian and South American populations is necessary to explain the presence of *kumara* within Polynesia, but it is unknown when this 'contact' occurred, or how substantially these populations interacted. Present genetic evidence is ambiguous because both populations share some features that are probably attributable to their common Asian ancestry, e.g. the 9-bp deletion is relatively common in South American populations, as is the first transition required to create the Polynesian motif (T→C 16217: see Figure 4.4). The rare haplotypes found in the current study have either the CGT motif, or the intermediate ancestor (CAT: see Figure 4.4) *plus* the unique transition that distinguishes the haplotype i.e. haplotype #AW4 has the CTG motif, *plus*

a unique transition at nt 16207 whereas haplotype #AW8 has a CAT motif, *plus* a unique transition at nt 16051. The four rare haplotypes were searched for within the database using their unique transition and the T→C transition at nt 16217, resulting in partial matches reported above. Closer examination of these partial matches revealed that none of the South American haplotypes had any additional CGT motif transitions. Therefore, although the rare Maori haplotypes in the current study do have potential links to South America, due to the 9-bp deletion and T→C transition at nt 16217, it is likely that these similarities are due to directly shared ancestry between the South American and Polynesian populations, rather than more recent contact. It is probable that the shared unique transitions are simply parallel changes that have occurred independently within the South American and Maori lineages since diverging from the parental Asian population (see Figure 1.12).

4.3.3 Mitochondrial HVII sequence

In the HVII region a homopolymeric C-tract occurs between nucleotide positions 00 303 and 00 315 which leads to ambiguous nucleotide assignment and incomplete sequences. The incomplete sequences were unsuitable for further analysis and are not reported in this thesis, whereas the 13 full HVII sequences that were obtained (see Appendix Two) will be deposited in GenBank for access by the scientific community.

Of these 13 complete HVII sequences 10 were from Maori and PN samples, whereas 3 were from samples with suspected European ancestry. Out of the Maori and PN samples 9 had identical HVII region sequences which conformed to the expected haplotype (00 073 A→G, 00 146 T→G, 00 263 A→G, compared with the Cambridge

Reference Sequence: Anderson et al., 1981), whereas one sample was missing the expected T→G transition at nt 00 146 and had the same sequence as 2 of the samples with suspected Caucasian ancestry. The third Caucasian sample had the same haplotype as the other Caucasians (00 073 A→G, 00 263 A→G), and an additional transition at nucleotide position 00 279 T→C. In the study by Redd et al. (1995), out of a total of 24 Samoan samples only 3 had HVII region sequences that differed from the expected sequence. So the question once again is “Would sequencing the HVII region reveal further significant variation to warrant the extra time and money that would need to be spent?”

4.4 Number of Maori female founders to settle New Zealand (Aotearoa)

Before commencing simulations to estimate the number of female founders to settle New Zealand, decisions were required to identify the appropriate treatment of sequence information that would provide the best set of data for this analysis and also ensure accurate estimates for the number of female founders to settle Aotearoa. A summary of the parameters used to identify the appropriate dataset, the decision process outcomes, and brief explanations were summarised in section 2.12.1. The following section explores in depth the reason behind each decision.

4.4.1 Determining suitable parameters for analysis of the data used in the founding population simulations

A. What region(s) of the mitochondrial sequence should be included in the simulation dataset?

- Long sequence (Whole of HVI and HVII if available)
- **Short Sequence** (3' HVI, from nt position 16 189 onwards only)

It was decided that for the purpose of these simulations it would be best to focus on short 3' HVI sequences (see Figure 1.15). Focusing on the 3' HVI region avoids the sequencing complications arising due to artefacts in the polymeric cytosine tract that starts around nt position 16 189. The 3' HVI region includes the CGT motif, and has been used in many previous studies as it has much more variation than the 5' HVI region. Therefore, using the short 3' HVI sequence allows comparison of the largest possible set of data from previous studies. It is clear that this treatment of data does underestimate haplotype diversity slightly e.g. for the Maori samples there is a decrease from 10 to 8 haplotypes (see Appendix Five versus Table 3.3). However, it is better to exclude data that are not comparable rather than to try to combine different haplotypes into a single category as was done in the MM98 study (i.e. see Appendix Five, CD1a, CD1b etc were combined into MM1). Not only does the MM treatment of data lead to an underestimate of haplotype diversity, but there is also an apparent increase in the frequencies of the remaining haplotype categories, because more samples are represented by fewer haplotypes.

B. Which dataset should be used for the Maori founder analysis?

- AW02 only (to be used as a comparison to the MM98 data)
- AW02 and MM98 data combined
- **All available data combined (i.e. including data omitted from MM98)**

To use the data from the current study alone in comparison against the MM data, would require the use of some Pacific Nation (PN) haplotype frequencies from other studies, as in the current study only 5 samples from EP were sequenced. The MM study did not sequence any PN samples, but instead relied on PN data compiled from other studies. Therefore, it was impossible to keep the simulations for the current study totally separate from the MM98 study and other studies, as the PN data would have been taken from the same sources. So if a combination of data must be used for the EP data it is sensible to use all the available Ma 3'HVI sequence data for analysis too.

C. Which Eastern Polynesian (EP) Haplotypes should be used for the frequency comparison?

- EP (as used by MM)
- **EP+ (all haplotypes found in all studies)**

When MM98 compiled the EP data from other previous studies, a total of 11 EP haplotypes were identified. Reanalysis of these data found many more EP haplotypes, as many Group III sequences seem to have been left out of the MM table entirely and some sub-haplotypes may have been combined into a single haplotype category. Consequently, a new table was compiled (see Appendix Five) and it was decided that these revised total haplotype data (see Table 3.3) should be used for analysis.

D. Is the dataset based on decisions A-C suitable for analysis (see Table 3.3)?

No it isn't as some of the hts that are found in the Ma samples are not found in the EP samples. What are the options?

- do nothing (this option is rejected as no simulation is possible)
- Mutation accounts for the extra Maori haplotypes
- **add missing EP haplotypes (Discovery Curve Assumption)**

In this study and the previous study conducted by Murray-McIntosh et al. (1998), the expectation is that the settlement of Aotearoa occurred as one or a series of founder events and bottlenecks, probably modified by immigration and emigration over several generations leading to the low mtDNA haplotype diversity encountered in the present New Zealand Maori population (see Section 1.3). Under this model, Maori are expected to contain a subset of the genetic variability to be found in central Eastern Polynesia, which in turn is expected to be a subset of the diversity found across the whole of Remote Oceania. The 3 new rare Maori haplotypes found in the present study and the one found by Sykes et al. (1995) are included in the simulations, and have not, to the best of my knowledge, been found and reported in any other Polynesian or Oceanic population. This means that these haplotypes are unique to Maori and therefore do not fit directly into the expected founder scenario outlined above. There are two explanations that could explain the presence of these haplotypes in the Maori population and/or their unexpected absence in Eastern Polynesian. Firstly, it is possible that these unique haplotypes are present in the Eastern Polynesian populations but have not yet been observed (Discovery Curve Assumption). This was the first assumption tested during the computer simulations. Secondly, it is possible that these unique haplotypes are the result of mutations that have occurred in the Maori population since the original founders left their parent population. To do the computer simulations using the simulation format as used in the original MM98 study, requires an estimate to be

made of the probable frequency of the unobserved rare haplotypes in the Eastern Polynesian population. For the purpose of this simulation the rare haplotypes were each given a frequency of 1 in the total Eastern Polynesian population dataset. However, as will be shown below, a more accurate estimate of the true frequency is difficult, as there may be conflict between the apparent and actual frequencies due to the discovery curve phenomenon.

▪ *New Mutations*

It is possible that the 4 new unique haplotypes found in Maori are not present in the Eastern Polynesian population at all as they could have occurred as new mutations in the Maori population since the original founders left their parent population. The reported mutation rate of mtDNA has an extremely wide variation. The highest rate ever suggested is $\sim 1.5\text{--}2.5/\text{site/Myr}$ ($\sim 1/33$ generations: Parsons et al., 1997). This mutation rate was found by studying intergenerational substitution rates and is highly contentious, drawing criticism (e.g. Jazin et al., 1998) as it is ~ 20 -fold higher than estimates from phylogenetic analysis. The mutation rate is higher for the hypervariable regions, $\text{HVI} = 10.3 \times 10^{-8} / \text{site per year}$, $\text{HVII} = 7.39 \times 10^{-8} / \text{site per year}$ (Horai 1995). A range of mutation rates could be applied to see which, if any, could give the haplotype distribution seen in the present study. However, the possible effect of mutations on the simulations was considered beyond the scope of the present study due to the above uncertainties and has not been attempted. This could be an important aspect for analysis in future investigations.

▪ *The Discovery Curve Assumption*

All population databases of mtDNA HVI sequences studied previously show a similar frequency distribution of haplotypes, with many rare haplotypes and fewer common haplotypes (Holland and Parsons, 1999). As the number of samples surveyed increases, there will be an increase in the total number of mtDNA haplotypes found, and a simultaneous increase in the number of unique haplotypes found. When any new sequences are added to an established database, the great majority of sequences that were unique in the smaller database remain unique in the expanded database, with a concomitant decrease in their apparent frequency (see example using Caucasian population data: Figure 4.1). Therefore, since the total overall number of haplotypes in the Caucasian population in Figure 4.1 has not yet been closely approached (judging by the shape of the graph), then the number of rare haplotypes in the population can only be an estimate or the 'apparent frequency' ($1/n$, where n is the database size) and is likely to be an overestimate of the true frequency.

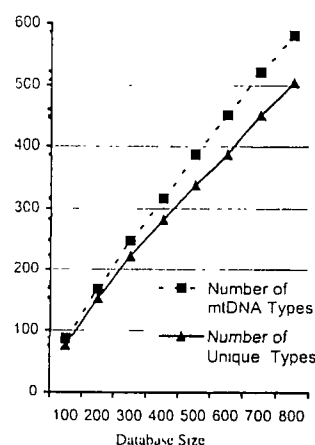


Figure 4.1 Discovery curve for Caucasian population data

Graph showing the effect of database size on the total number of mtDNA types that are unique in the database. Databases increasing in size by increments of 100 were obtained by random sub-sampling of a large database of Caucasian sequences. This Figure has been reproduced from Holland and Parsons (1999).

The discovery curves for the Maori and Eastern Polynesian data are very flat by comparison to Caucasian populations (see Figure 4.2). This indicates a much lower overall number of haplotypes in these Pacific populations, and may indicate that most, if not all, mtDNA haplotypes present in these populations have already been sampled. These discovery curves provide additional evidence for the impact of recurrent founder effects and bottlenecks; the severe reduction of the number of haplotypes as shown by comparison of the decreased haplotype diversity in Eastern Polynesia versus Europeans and the further reduction in haplotype numbers in Maori versus Eastern Polynesian populations. However, this decrease in diversity from East Polynesia to New Zealand is expected under the model of settlement outlined in the first paragraph of this section. Although it may be more appropriate to compare the Polynesian population discovery curves to similar data from Melanesian and Asian populations, this information is not readily available. However, the Asian population is expected to have a similar haplotype diversity to Caucasian populations, and Asians may even have slightly higher diversity (Holland and Parsons, 1999). Therefore this Caucasian comparison is adequate to illustrate the much lower overall diversity in Polynesians and to show the decreased slope of the discovery curve.

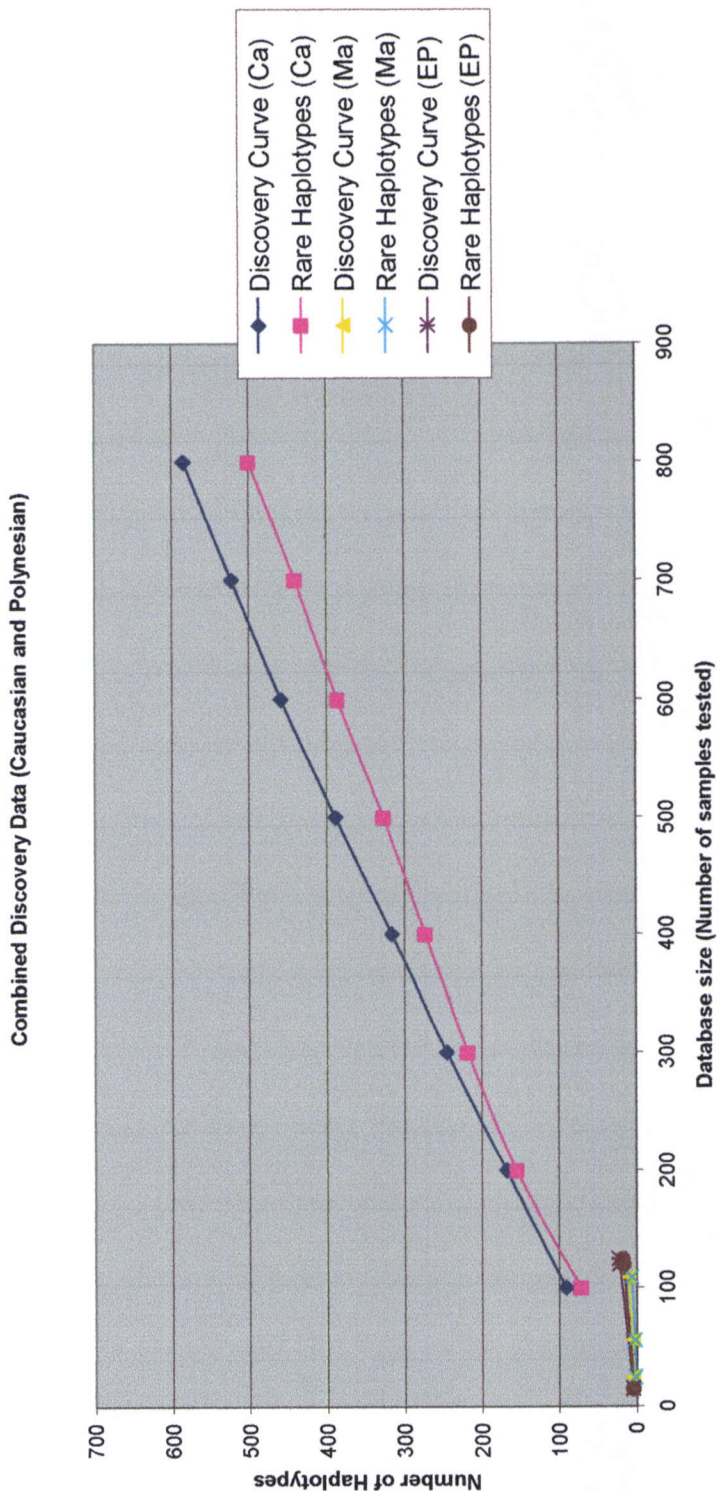


Figure 4.2 Discovery Curves comparing Eastern Polynesian and Maori populations with Caucasian populations. The graph above plots the discovery of haplotypes overall versus the number of rare haplotypes in Eastern Polynesian (EP) and Maori (Ma) populations (see Appendix Seven), compared with Caucasian populations. The Caucasian curve has been reproduced using data from Holland and Parsons (1999).

The power of the simulation for this project is limited by the size of the current database, as the total number of haplotypes and their true frequencies in Maori and Eastern Polynesian populations is still unknown. However, for the purpose of this simulation the rare Maori haplotypes are assumed also to be rare in the Eastern Polynesian population too (hence the reason why they haven't been discovered yet), and have each been given frequencies of 1. Further analysis would need to be done to determine how changing these assumed frequencies (probably a slight underestimate) might influence the results of the simulation.

In the current study, various capture-recapture calculations were performed on the data to estimate the most likely number of haplotypes present overall in the total Maori and Eastern Polynesian populations. Using the Burnham and Overton (1979), first order jack knife, the total number of haplotypes expected was found to be ~32 for Eastern Polynesians and ~13 for the Maori population. In future these estimates could be used in computer simulations to calculate the maximum number of founders for the Maori population. These capture-recapture estimates may fit in with the maximum number of haplotypes estimated by extrapolating the Ma and EP discovery curves (see Appendix Seven).

Currently, there is no way to determine whether these rare haplotypes are actually present in Eastern Polynesia without doing more sampling. However, if these newly discovered rare haplotypes are also present in Eastern Polynesia or elsewhere in Oceania it may even be possible to trace them to a single island or island group, thus providing evidence for a Maori homeland.

E. What Population Growth Model should be used?

- Exponential (as used by MM98)
- **Sigmoid (3 types of population growth):**
 - i. **Early expansion**
 - ii. **Middle expansion**
 - iii. **Late expansion**

In the study by Murray-McIntosh et al. (1998), an exponential growth curve was used to model the population expansion. In the current study 8 haplotypes were identified, and preliminary trials using exponential growth curve with the current data predicted an exceptionally large founding population of ~185 females (range between 75 and 475: see Appendix Two, Exponential population growth). These results suggest that an exponential population growth model may be inappropriate. A study by Brewis (1990) modelling the Maori prehistoric population expansion suggests that a sigmoid expansion would be a more appropriate model as it fits in well with the archaeological record, whereby pa fortifications were built for defence as competition for scarce resources increased. Maori colonised Aotearoa ~1000 years ago, with the first evidence of pa building ~700 years ago. However, Kirch (2000) suggests that the majority of pa were built approximately 200-350 years ago (A.D. 1650-1800), during the time the Maori population may have reached the carrying capacity of the land (see Figure 4.3). Three different sigmoid population growth curves were used in the current study to model the prehistoric expansion of the Maori population. A range of early, middle and late expansion sigmoid curves were used to model the population growth scenarios as it is currently unknown which growth pattern is most appropriate.

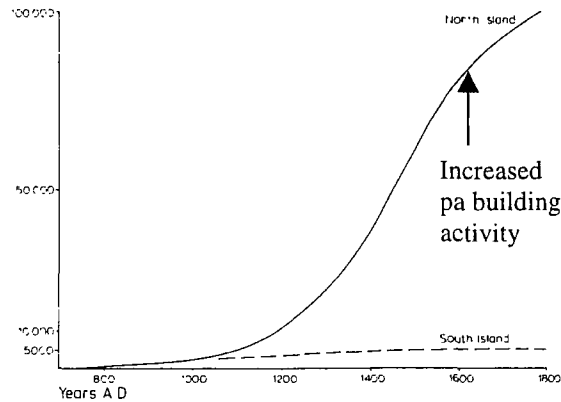


Figure 4.3 A probable population growth curve for the Maori, as reconstructed by Janet Davidson (1984). Diagram is reproduced from Kirch (2000).

4.4.2 The simulation

In the Murray-McIntosh et al. (1998) study, founder simulations indicated that ~70 woman (between 50 and 100) were required to found the haplotype diversity present in the current Maori population. However, this study was based on an exponential growth curve, which may not be the best fit for the population expansion (see above). The current study has used more realistic sigmoid growth curves. Success of these models varied. The late sigmoid population expansion model for example was considered inappropriate because it predicts an extremely large number (~510) of female founders (between 230 and >1000), a model that is not supported by other evidence (see above). The early and middle expansions give female founder estimates of ~47 (between 25 and 115) or ~260 (between 115 and 670) respectively, both of which are considered to better match the maximum number of female founders (250) suggested (McGlone et al., 1994). However, a sigmoid expansion curve intermediate between the early and middle curves, may give the best answer. It is currently unknown exactly how the Maori population in Aotearoa expanded, and for this reason a range of values (early, middle

and late curves) was examined to see if any would give values consistent with the population estimates defined by archaeological evidence.

Although the exact demography of the prehistoric Maori population is unknown, some records are available from around the 18th century, at the time of first European settlement (Anderson, 1990), a time that corresponds with the end of the simulations (NB: The simulations follow MM98 and run for 30 generations at 20 years per generation over 600 years from 1200 A.D. to 1800 A.D., i.e. to the time of first European settlement). It would be possible to use these records to establish an estimate of birth and death rates at the time of settlement and to use these derivative figures to check whether the answers given using sigmoid growth curves are appropriate. Given that there may have been relatively few infectious diseases present in Aotearoa prior to increased outside contact and an abundance of resources for the original founders, an extremely rapid population growth scenario is reasonable (Professor David Penny, Institute of Molecular BioSciences, Massey University: pers. comm., 2002).

The frequency distribution of the haplotypes found in the Ma and EP populations is 'unequal': i.e. a few common and many rare haplotypes. This 'unequal' distribution is taken in to consideration by matching simulation data (roughly) to the modern distribution. This was done by looking for sets of results where one haplotype was present 50 or more times in the sample of 107, and 5 other haplotypes were represented by less than 5 individuals, leaving 2 haplotypes of between 5 and 50 individuals. Using these criteria it was found that the overall shape of the curves did not change but the founder estimates were affected, slightly increasing the number of founders required to give the current haplotype frequency distribution (see Appendix Eight). For the early population growth pattern the number of founders increases from ~47 (between 25 and

115) to ~56 (between 34 and 120), whereas for the middle population growth the number of founders increases from ~260 (between 115 and 670) to ~320 (between 155 and 695). The late growth pattern was considered inappropriate (see Appendix Eight).

In the previous study by Murray-McIntosh et al. (1998) various reasons were suggested for the very 'unequal' haplotype frequencies found, including one or more factors such as chance, continuing migration, and genetic and/or cultural selection (for a complete description see MM98). The current study supports these as possible explanations and offers an additional idea. The original study (MM) allowed the maternal founding population to expand to 50,000 females (100,000 total population) over a period of ~600 years, from 1200 A.D. to 1800 A.D. The graphs shown in Figures 3.6, 3.7 and 3.8 illustrate the effect of the founding maternal population size on the expected number of haplotypes for the New Zealand Maori sample. The size of the founding population from Eastern Polynesia is indicated on the *x* axis; the *y* axis shows the frequency of 20,000 simulations that the indicated number of haplotypes (between 1-21) are present in a sample of 107 sequences after the founding population has expanded to 50,000 females. Using the early and middle sigmoid expansion, the estimated founding population size (using the mid-point values) is ~47-260 Maori females to give the 8 haplotypes observed in the present day population. But this simulation stopped in the 1800's (i.e. at the time of first European settlement), there was an additional ~10 generations for the female population to expand to its present size (268, 797: 2001 NZ census). Would the present-day number of haplotypes and their frequencies be expected to be the same as they were back in the 1800's? Additional time has been available for the frequencies and the number of haplotypes to change since European settlement. It might also be that the impact of European settlement itself has led to the current 'unequal' haplotype distribution in the Maori population. Obviously, it would be

difficult to extend the simulations to the present day (i.e. allowing for admixture etc.), but it is also hard to interpolate the current gene pool back to what was present 200 years ago. Despite this problem the estimates given in the current study for the number of Maori female founders is considered to be a very close approximation to those suggested by archaeology and oral history.

4.5 Source of Maori female founders to settle Aotearoa

4.5.1 Theories for the colonisation of Aotearoa

There are ~5 major competing theories for the settlement of Aotearoa (see Figure 1.2), but these can be amalgamated into 2 major theories representing opposite ends of a continuum. The most widely accepted theory is a planned settlement, with multiple voyages over time bringing people, animals, plants, and cultural artefacts. This fits in with the current understanding of Maori oral traditions (elaborated by 19th century ethnologists) describing epic voyages requiring great navigational skills with sailing directions obtained from Kupe the legendary discoverer of Aotearoa. Long sea voyages by Polynesians to Aotearoa are supported by experience of canoe voyages using only traditional navigation knowledge; these voyages have supported the oral tradition on such matters as seasonal timing, the setting of courses, and pohutukawa trees in bloom on arrival. Traditional stories of return voyages from New Zealand to eastern Polynesia are supported by the discovery of New Zealand obsidian in the Kermadec Islands. McClone et al. (1994) propose that a maximum of 500 settlers arriving over several generations around 800 years before present would have provided both the necessary numbers to occupy the early coastal sites found throughout the country and the

necessary skills, traditions, and knowledge for successful colonisation. This initial colonising population would have provided the necessary basis for a high population growth expected in an environment previously uninhabited and extremely rich in seals, large birds (including moa), fish, and shellfish.

At the other end of the spectrum is the theory originally suggested by Sharp (1956), stating that Aotearoa was settled accidentally as opposed to planned voyages (see above). These random voyages were thought to have been made by one canoe with perhaps 10-20 people (or at the most a few canoes), drifting before the wind or currents, or by people forced to leave their home island, blindly searching for a new island on which to settle. However, throughout the colonisation of Oceania the voyaging people have always brought a large variety of domesticated plants and animals with them to every new island. It is this portable man-made environment, a virtual ark full of biotic resources, that is the foundation for the conquest of the Pacific, as the number of naturally occurring edible species declines rapidly outside of Near Oceania. It is for this reason that accidental voyage and discovery by, for example, a fishing boat blown off course, is thought unlikely. The presence of a few food crops that were successfully transported to New Zealand including: taro, yam (*Dioscorea batatas*), bottle gourds, paper mulberry, Ti Tree (*Cordyline fruticosa*) and the staple crop sweet potato or Kumara (*Ipomoea batatas*) imply that the journey was planned. To ensure successful transportation of these crops to a new island they would have to be carefully packed, and it would be unusual for food prepared for a day fishing trip to include important sprouting planting shoots, or for that matter enough women to found a new population. The previous section (4.4) estimates that absolute minimum number of women founders for the Maori population would be ~34 which is already too high to support models that stipulate a small number of waka and/or surviving founders. The waka would also

contain at least an equal or greater number of males. However, the number of original canoes suggested by oral tradition may actually include later subsidiary voyages within Aotearoa, a scenario that has not been accounted for in the founder simulations.

4.5.2 Source of the Maori founders

The 3' HVI mtDNA haplotype data obtained in the current study has been used to successfully estimate how many women were in the waka that journeyed to colonise Aotearoa and effectively rules out all models assuming a small number of founders. However the same data may also be able to answer some fundamental questions about Polynesian origins; where did these canoes originate from and is it possible to pinpoint proximal and distal 'homelands' for New Zealand Maori? Throughout the 19th century and the first half of the 20th century researchers sought to resolve the "problem of Polynesian origins" by searching for some external homeland, usually in Asia. Te Rangi Hiroa (1932) believed the Polynesian "Vikings of the Sunrise" had migrated in substantial fleets out from an Asiatic homeland passing through Micronesia to enter the Polynesian triangle via Samoa and on to the Society Islands. Most scholars followed some version of Hiroa's theories. Around the 1950's and 60's, Jack Golson, Roger Green and students excavated in Samoa and Tonga where they uncovered evidence that distinctly Polynesian cultural traits had emerged within these archipelagos more than 1000 years ago. Green (1991) summarised this new understanding of Polynesian origins:

"Thus there never was a Polynesian migration from elsewhere, becoming Polynesian took place in Polynesia itself as the archaeology of Tonga and Samoa over the last 3000 years readily attests. One begins with Eastern Lapita and ends with Polynesian".

Therefore, archaeology positions the primary geographic homeland of the ancestral Polynesian population in the Tonga-Samoa archipelagos and their smaller neighbours including Futuna, Uvea and Niuatoputapa (Kirch, 2000). Bellwood (1989) suggests the Cook Islands whereas an account by Evans (1998), places the homeland in Tahiti. It is thought that it is one or a combination of these regions that later Polynesian people would call in their myths and traditions, Havaiki (or Hawaiki), the Polynesian 'homeland'. The archaeological data not only specify a probable homeland, it also estimates the emergence time for the ancestral culture out of Early Eastern Lapita as the middle of the first millennium B.C. (Kirch, 2000).

If the present populations in these 'homeland' regions should prove to have any of the rare haplotypes found in the Maori population in this study, it would be compelling evidence for their particular role as participants in the settlement scenario outlined above, whereby current Eastern Polynesian and Maori populations are derived from a parental population that resided on one or a few closely associated islands in the Tonga-Samoa archipelagos. All Polynesian populations are very similar genetically and general links can already be found amongst many islands (i.e. shown by the high frequency of the CGT motif and the 9-bp deletion haplotypes). Currently, we do not have data from enough samples collected in any of these homeland regions to be able to conclude whether or not a specific genetic link could be established between the Maori people of Aotearoa and a single parental Eastern Polynesian population. It is perhaps rare mtDNA haplotypes that might lead to identification of a specific parent population. It is not certain that further sampling would actually discover these new rare haplotypes in the Samoa-Tonga archipelagos, as they may have occurred once only by mutation in Aotearoa and would therefore not ever be found elsewhere, even with extensive sampling. Comparisons of mtDNA haplotypes with South American populations have

indicated some ‘contact’ with Maori and/or proto-Polynesians may have occurred. However, additional evidence would be required to fully support this scenario, as the similarities could well be due to common ancestry (see Section 4.3.2).

But even if one or more of these rare Maori population haplotypes were to be found in a supposed ‘homeland’ region or island, then on its own this ‘genetic trail’ would not prove that this is the actual parental population, or that one had found the mythical Havaiki. Since the population split, chance changes in haplotype diversity could lead to parallel changes in nucleotides, with no phylogenetic signal for reconstruction of population relationships (see Figure 1.12). Reciprocally, even if these rare haplotypes cannot be found in the Tonga-Samoa region today, it does not mean that settlement of New Zealand and other Eastern Polynesian Islands did not occur from these archipelagos. Just as these haplotypes are currently rare in the Maori population, they may also have been rare in the original parental population, time has passed and chance events may have eliminated them from the parent population since leaving the island of origin. However, positive genetic associations considered alongside the Archaeological and Linguistic evidence may ultimately prove capable of providing once and for all a definitive homeland for Maori.

4.6 Origin of Austronesian-speaking Oceanic Populations: Introducing the ‘Synthetic Total Evidence Theory’

The previous section traced the origins of the Maori people back to a probable homeland within Eastern Polynesia, but where did these eastern Polynesian people

come from? Looking back further in the past, is it possible to identify the source of all the speakers of Austronesian languages?

In the introduction to this thesis (see Section 1.5), three ‘competing’ colonisation models were outlined, the ‘Express Train Model’ (Bellwood, 1991, 1997, 1998), the ‘Entangled Bank Model’ (Terrell, 1988), and the ‘Slow Boat Model’ (Oppenheimer and Richards, 2001). Briefly, the ‘Express Train’ argues for a Taiwanese origin, and links the expansion of Austronesian-speaking people with the colonisation of Polynesia, whereas the ‘Entangled Bank’ and ‘Slow boat’ decouple Austronesian language and Polynesian origins stating that:

“the origins of the Polynesian dispersals may not follow on directly from the earlier history of the bulk of the Austronesian speakers in island Southeast Asia” (Oppenheimer and Richards, 2001a: emphasis added).

Furthermore, the ‘decoupled’ theories place the Austronesian language expansion somewhere in the region where these languages are currently spoken, within island Southeast Asia and/or Melanesia. The ‘Entangled Bank’ is more closely associated with an indigenous Melanesian origin and the ‘Slow Boat’ is associated with an island Southeast Asian origin for Polynesians. This thesis contends that these ‘competing’ hypotheses can in fact be incorporated into an inclusive ‘Synthetic Total Evidence Theory’ that encapsulates aspects of each model.

Currently, most theories of Pacific settlement concentrate on only one of the two aspects that are important in understanding Pacific colonisation, i.e. either a pattern or process element. Additionally, models may only focus on expansions occurring over one part of

the genetic trail e.g. concentrating on island Southeast Asia, without explaining the patterns and processes occurring over the whole migration route i.e. including information before and after the island Southeast Asia expansion. Therefore, present theories are incomplete and thus unsatisfactory for explaining the complicated migration history of the entire Oceania region, from the proposed proto-Austronesian homeland, right through to the final migration to Aotearoa. Additionally, results from interdisciplinary studies are often misinterpreted by non-specialists, and there are some who contend that when archaeological and linguistic lines of evidence are combined, they become mutually self-supporting to the extent of circularity (Oppenheimer and Richards, 2001). Now, perhaps for the first time, it is possible to look at the interdisciplinary evidence in its entirety, combining data from archaeology, linguistics, anthropology and genetic studies to explore both pattern and process elements to create a more complete picture of Pacific settlement, and from consideration present a total evidence account of this amazing human expansion. The timing of this thesis fits in to a period where many pieces of this complicated puzzle have been found, and researchers can now fit them all together to make sense of these voyages. The ‘Synthetic Total Evidence Theory’ (plus diagrammatic representation) introduced in the following section will show how ‘competing’ theories have a lot in common and present arguments showing that a consensus view using information from different disciplines may finally be in sight.

4.6.1 The Synthetic Total Evidence Theory

The colonisation of the Pacific region can be broken down into three phases each with associated pattern and process elements. The first phase is the initial expansion of people into Near Oceania, which occurred at least 33,000 years ago. These populations

would have probably originated in Asia (pattern). However, consequent differentiation of genes, languages and cultures (processes) within Asia and Near Oceania since the original expansion, would have created unique markers for each population. The distinction between *ancient* Near Oceanic populations and *migrant* populations is important for the second phase of the expansion, as both populations have a shared history with Asia, which must be taken into consideration. The second phase is the expansion of a proto-Austronesian population (*migrant*) into Near Oceania. This phase of population expansion is contentious as it involves locating an origin for the Austronesian expansion (pattern element). Proposed centres of expansion include specific regions, e.g. Taiwan, Indonesia, and South America, or more general areas e.g. island Southeast Asia, Melanesia or Micronesia. Further complications arise due to the unknown interactions (process elements) between the ancient Papuan-speaking residents of Near Oceania (Phase One) and the incoming migrant population(s). The third phase involves the expansion of the Austronesian-speaking populations (Phase Two) out into the previously unoccupied Remote Oceanic region. The third phase is less complicated, as the origin (pattern) of the expansion can be more reliably traced using archaeological dates and there were no previous populations to obscure relationships with the parental population. Debate in this third phase centres on the 'long pause' (process), whereby there was a pause of up to 1,600 years before the settlement of islands beyond Tonga/Samoa (see Section 1.4.1). Phases Two and Three are discussed in detail, as they are most relevant to the Synthetic Total Evidence Theory, and because it is these phases which are causing problems for consensus models, Phase One appears in summary form only.

▪ **Phase One: Initial settlement of Near Oceania**

Hunter-gatherer Papuan-speaking populations migrated over small water gaps and land bridges to colonise Sahul (the combined Australia-New Guinea landmass) $\geq 50,000$ -60,000 years ago, and Near Oceania (see Figure 1.1) between 30,000-40,000 years ago. There is good multi-disciplinary evidence to support the antiquity of these populations, using archaeology to date the earliest occupation sites, and linguistics and genetics to measure the diversity of languages and genes respectively. Changes in sea levels meant that much of the continental shelf was above water for long periods (see Figure 1.17), making water crossings easier as only deep-water channels persisted. These ancient populations would most likely have originated from somewhere on the Asian mainland, and reached Australia, PNG and the Bismarck Archipelago (today ancient populations persist in the PNG Highlands and Australia). As there is general agreement on events occurring in this region, no further discussion of phase one is necessary as part of the Synthetic Total Evidence Theory.

▪ **Phase Two: Neolithic expansion into Near Oceania**

Neolithic proto-Austronesian speakers expanded from central and south China to colonise Taiwan between 8000-6000 years ago (Melton et al., 1998). This expansion may have been driven by the development of agriculture, the consequent spread of rice cultivation and the increase in population associated with increased food production (Bellwood, 1998). The contrast between the big differences to be found among Taiwanese Austronesian languages and much more modest differences among extra-Taiwanese Austronesian languages suggests there was a 'long pause' between the

colonisation of Taiwan and the expansion of Austronesian speakers out of Taiwan (Diamond, 2000). Archaeology supports the long pause scenario suggesting there was a 1,000-year gap (from about 4300 to 3300 B.C.) between farmers' colonisation of Taiwan and their subsequent colonisation of the Philippines. Blust suggests that the pause was due to the time required to develop sufficiently good boat technology (perhaps outrigger canoes?) to enable them to cross the 375-km sea gap separating Taiwan from the Philippines (Diamond, 2000). Therefore, ~6,000 years ago the proto-Austronesian population began to spread outwards from Taiwan to colonise the Philippines, eastern Indonesia and Melanesia (Bellwood, 1997). Lapita culture (see Section 1.4.1) arose ~3,500 years ago around the Bismarck archipelago and provides a firm basis for tracking the chronology and rate of population spread (see later, Section 4.6.3.2). Linguistic evidence points to Taiwan as the source for the Austronesian languages, and phylogenetic analysis of the language branching structure supports the 'Taiwanese Homeland' component of the 'Express Train' theory (Gray and Jordan, 2000). Additionally, mtDNA data (Melton et al., 1995) and some Y-chromosome evidence (Capelli, 2001) connects the Taiwanese population to Polynesians. As previously mentioned, Phase One evidence indicates that Near Oceania (Philippines, eastern Indonesia and Melanesia) had human occupants beginning in the late Pleistocene (ca. 40,000 years). Consequently, the Austronesian-speaking Proto-Polynesian migrants would have had to compete and interact with human populations already present in these areas, and intermarriage, mediated by force or otherwise, is likely to have occurred. However, comparisons between mtDNA and Y chromosome data, suggest differential geneflow, with greater Melanesian male than Melanesian female contribution to the proto-Polynesian gene pool (e.g. Kayser et al., 2000).

▪ Phase Three: Remote Oceania

Based on archaeological evidence, the proto-Polynesian population reached Fiji by 3,500 years ago, spread into the Polynesian triangle by ca. 1000 years ago, and concluded the migration with the colonisation of Aotearoa within the last 800-1000 years. Remote Oceania was not colonised prior to the arrival of the Polynesians, therefore there are no underlying population interactions with older residents to take into consideration (as is the case for Near Oceania). Consequently, descriptions of the final phase of this population expansion are in general agreement across most disciplines. However, there is still some debate regarding the Eastern Polynesian settlement possibly arising from a second 'long pause'. Further archaeological work is required to determine whether this pause is real, or an artefact due to insufficient excavations in these areas. Pawley (1996) argues that a period of common development in the Western Polynesian region is required prior to later dispersals in order to explain the many innovations in proto-Polynesian languages. Currently, archaeology confirms the linguistically deduced long pause, from about 1200 to 200 B.C. (Diamond, 2000). Therefore, it is likely that the pause is real, and may have been caused by the increasing inter-island distances past the Samoan-Tongan region, coinciding with the development of double-hulled platform sailing canoes (Diamond, 2000).

Micronesia is often ignored when prehistory and human diversity in the Pacific are discussed (Intoh, 2001). However, any 'Synthetic Total Evidence Theory' would not be complete without some discussion of the colonisation of this area. Archaeological evidence suggests relatively long habitation of western Micronesia, with sites dating to >3,500 years in the Mariana's archipelago and Palau, and dates ~3,300 years old from Yap, suggesting that western Micronesia was colonised about the same time or even

before the Lapita making Austronesian-speaking people reached the Admiralty Islands (Intoh, 2001). Despite the long habitation history of Western Micronesia the rest of Micronesia was not settled until ~2,000 years ago. All Micronesian peoples speak Austronesian languages, but not all of these belong to the Oceanic subgroup, a clue to their diverse origins (Kirch, 2000). Studies of Oceanic languages suggest that the late Micronesian colonisation (eastern Micronesia) originated in the Solomon-Vanuatu region, which is close to the date of dispersal from Samoa to East Polynesia (Intoh, 1997). However, some western Micronesian islanders speak western Malayo-Polynesian languages and linguistics links have been made with the Philippines, Indonesia and possibly Taiwan (Kirch, 2000). In contrast to Polynesia, Micronesia does not form a 'monophyletic' cultural group, although its history is shallower and less complex than that of Near Oceania (Kirch, 2000). Despite biological similarities between Polynesian and Micronesian populations, according to Intoh (2001) the influence of Polynesian-speaking populations in the history of Micronesia is late and minimal compared with that of the Nuclear-Micronesian-speaking population coming from Melanesia. Genetic studies have shown some evidence of mixing between eastern Micronesia and Melanesia (Lum and Cann, 1998; Lum et al., 1998).

Figure 4.4 Synthetic Total Evidence Theory

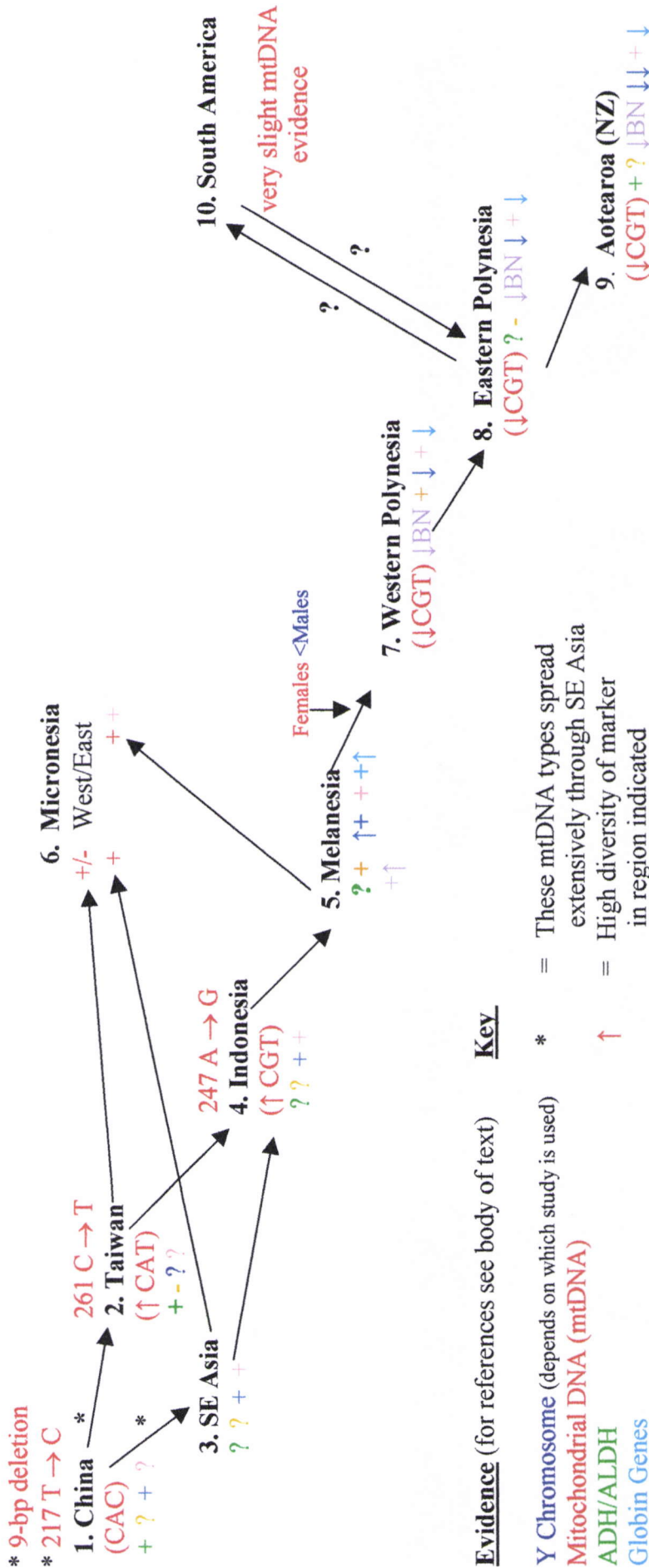


Figure 4.4 Synthetic Total Evidence Theory contd...

Linguistics: Spread of Austronesian (AN) Languages

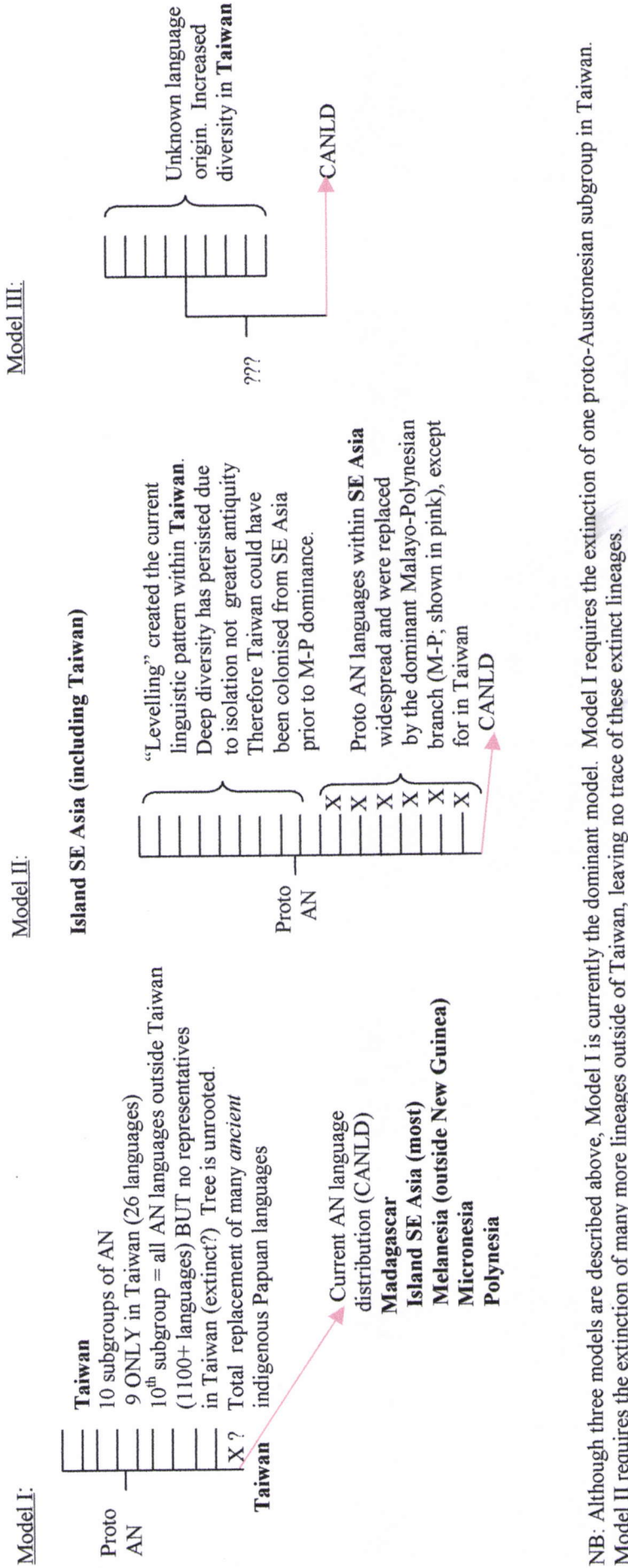


Figure 4.4 A diagrammatic summary of the ‘Synthetic Total Evidence Theory’ as outlined in section 4.6.1. By combining research from diverse disciplines, for the first time the complete picture of the final major migration of human history is becoming clear. Support for this model is discussed in the following sections, which outline current developments and opinions in each discipline highlighting their relevance to the ‘Synthetic Total Evidence Theory’. Appendix Nine shows the ‘Synthetic Total Evidence Model’ outlined on a map of the Oceanic region.

4.6.2 New mtDNA data support for the Synthetic Total Evidence Theory

Researchers have found that 90-95% of Polynesians have inherited the 9-bp deletion seen in Southeast Asians including the Taiwanese, but it is largely absent in Melanesians (Melton et al., 1995). The mtDNA CGT motif pattern is also found at a high frequency in Polynesian people and the ultimate origin for this motif is likely to be in Taiwan (Melton, 1995). These patterns fit in with Bellwood's idea of a Taiwanese homeland and rapid expansion with limited intermixing (Gibbons, 2001). This idea does not support Terrell's idea for an exclusive Melanesian origin of these populations, but could support the 'Slow Boat' model for an island Southeast Asian origin proposed by Oppenheimer and Richards (2001a). The conclusions drawn from mtDNA data depend on the way the mtDNA tree is rooted.

According to Redd et al. (1995) the 9-bp deletion first arose approximately 58,000 years ago somewhere in Asia (probably in China or Southeast Asia). The deletion was quickly followed by the transition at nt position 16 217, and mtDNAs with these markers spread extensively throughout Southeast Asia (Melton, 1995: see Figure 4.4). The change at nt 16 261 occurred next, probably in Taiwan, and people bearing these haplotypes spread outwards beginning ~8,000 years ago throughout the Philippines and Indonesia (Melton, 1995). Finally, between 900 and 23,000 years ago the substitution at position 16 247 occurred creating the Polynesian motif. This final substitution probably occurred in Indonesia and there was a subsequent expansion ~5,500 years ago of early proto-Polynesians from Indonesia eastward (Redd, 1995). Figure 4.4 shows the most likely order for the occurrence of the transitions required to produce the Polynesian motif, in association with 9-bp deletion based on the mitochondrial evidence acquired from previous studies e.g. Redd et al., 1995.

Figure 4.5 Most likely order for the occurrence of the transitions required to produce the Polynesian Motif haplotype

Order of occurrence	Nucleotide Position	Transition /Deletion	Haplotype (16 217, 247, 261)
First	Intergenic	9-bp del	TAC (CRS <i>plus</i> 9-bp deletion)
Second	16 217	T → C	CAC (ancestral)
Third	16 261	C → T	CAT (intermediate)
Fourth	16 247	A → G	CGT (Poly Motif)
** ??	16 189	T → C	

** This substitution also occurs at high frequencies in control region sequences with and without the 9-bp deletion and very likely predates the 9-bp deletion

Richards et al. (1998) and Oppenheimer and Richards (2001a,b) argue for an east Indonesian origin (somewhere between Wallace's line and the island of New Guinea) for Polynesian populations, as the control region diversity among Polynesian motif mtDNAs is highest in Indonesia, and they suggest that it is therefore likely that the Polynesian motif originated there (Redd et al., 1995). However, as illustrated in Figure 4.5 there are 3 patterns of nucleotide substitution CAC → CAT → CGT. The intermediate motif 'CAT' is useful for population origins, as it predates that of the CGT pattern. The 'CAT' pattern occurs at high frequency in 3 aboriginal groups of Taiwan. Furthermore, the mtDNA diversity of the 'CAT motif' is highest in Taiwan and together these results suggest that the intermediate CAT motif arose in Taiwan and spread south through the Philippines and Indonesia (Melton, 1995). Therefore, although it is likely that the final element of the Polynesian motif arose in Indonesia, closer examination of the intermediate motif suggests an earlier link between these populations and Taiwan.

Using a molecular clock to estimate divergence times based on observed haplotype diversity accumulated around the Polynesian motif, Richards et al. (1998) were able to estimate the age of the CGT motif as roughly 17,000 years before present (95% Confidence Interval: 5,500-34,000 years). Because archaeology argues for a tightly constrained arrival and departure of the express train from Wallacea ~4,000 years ago (Bellwood, 1997), Oppenheimer and Richards claim that the molecular clock evidence suggests the CGT motif originated before an express train carrying Taiwanese farmers could have arrived in Wallacea. However genetic evidence should be based on many loci, integrated with other evidence and the extremely wide confidence limits for the calculations of marker ages should be viewed with caution (Diamond, 2001). Therefore overall, the mtDNA diversity could be taken to suggest either an Indonesian/island Southeast Asian origin, or a Taiwanese origin, depending on whether focusing on the diversity of the CAT or CGT motifs. An exclusive Melanesian origin (Terrell, 1988) is not supported as the mtDNA evidence suggests a relatively minor genetic (<5%) contribution from Melanesian females to the current Polynesian gene pool (Sykes et al., 1995). However minor, this mtDNA input hints at intermarriage between *migrant* and *ancient* populations, a link that is strongly supported by some other markers, suggesting that Terrell's Melanesian origin theory contains some elements that are correct. Additionally, there is some evidence of South American input into Polynesian populations (<0.02%), suggesting that the largely discounted claim of Thor Hyerdahl (1950) is also at least partially supported by mtDNA evidence (Sykes et al., 1995).

4.6.3 Archaeology

4.6.3.1 Agricultural evidence

Peter Bellwood is the archaeologist most often associated with the predominant view for Polynesian ancestry, popularly known as the ‘Express Train to Polynesia’ (Diamond, 1988). Bellwood (1998) came up with many reasons to explain the underlying causes and motivation behind Austronesian dispersals including those outlined below:

- (a) dependence on agriculture with its potential for population growth (especially in the early stages in China and Taiwan)
- (b) dependence on maritime resources with consequent development of a voyaging technology
- (c) social processes which have been termed founder-focused rank enhancement

Agricultural motivation alone (a) has long been associated with the Austronesian expansion out from the China/Taiwan region, but Oceanic populations show no evidence, past or present, of growing rice (except in the Marianas) and instead cultivate the same root crops as Melanesians (Oppenheimer and Richards, 2001a). Also until about 4,000 years ago island Southeast Asia was entirely inhabited by non-Austronesian speaking ‘Australoid’ foragers and now almost every language in island Southeast Asia is Austronesian implying there was a near complete linguistic and ethnic replacement (Oppenheimer and Richards, 2001a). The cultural and technological superiority of the incoming migrant farming population is the reason usually given to account for this dramatic replacement. But this idea is modelled on the spread of the more thoroughly documented European Neolithic, and there are now some serious doubts about the total

replacement of the indigenous foraging population by the advancing Indo-European farmers (Cavalli-Sforza, 2000). Additionally, some archaeologists are beginning to think that a Neolithic-type revolution was not even required to get people moving around the Pacific at the end of the Pleistocene. By 6,000 years ago the world's sea levels had finally stabilised to within a few metres of their modern position. Coastal lagoon systems would therefore have reached a higher natural resource productivity that would have led to higher human population growth. There are two ways to view the spread of agriculture in the Oceanic region; the agriculturalists spread through their own demographic impetus, in most cases intermarrying, but mostly dominating the indigenous hunter-gatherers; or early agriculture was the result of enterprising adoptions by pre-existing foraging people. Debate continues, therefore agriculture alone does not appear to provide a sufficient positive link between the Chinese/Taiwanese expansion and the spread of Austronesian-speaking people.

Maritime foraging and boat skills (b) are motivations that appear to have been important throughout island Southeast Asia, and Polynesia, so an origin for the voyaging Neolithic is very hard to pinpoint. The possible social reasons (c) as motivation for colonisation journeys were outlined briefly in section 1.3.1, but the desire to be a high ranking chief and to rule one's own land could be recognised as a strong feature of any culture. Therefore, overall, agriculture currently provides weak evidence for the Taiwanese origin, as the technology could equally plausibly have arisen within island Southeast Asia or Melanesia.

4.6.3.2 Lapita Culture

According to Bellwood (1997) and other archaeologists (e.g. Kirch, 1997) Lapita sites within the Bismarck Archipelago and Solomon Islands are correlated with the spread of the Austronesian language (see Section 1.4.1). The archaeologically well documented spread of Lapita culture, throughout Near Oceania and into Remote Oceania provide the basis for tracking the chronology and rate of population spread (see Figure 1.5). However, Oppenheimer and Richards (2001a) state that archaeology can reconstruct material culture to date it, but the absence of written text or skeletal remains means less success in identifying the origins of the *manufacturers*. Therefore, they ‘decouple’ the spread of Lapita culture, from the spread of Austronesian languages and the colonisation of Polynesia.

Although agreeing that the Austronesian languages and Lapita culture are linked, Patrick Kirch has some problems with the archaeological record in Taiwan stating:

“I don’t think there’s any question that the Austronesian expansion comes out of island Southeast Asia”, “The danger is getting too specific about Taiwan when we don’t know enough archaeologically about the coastal China area, Taiwan or the Philippines” (Gibbons, 2001).

Therefore, although Lapita culture is a reliable marker within Melanesia and Western Polynesia, links before and after these regions are more contentious and many sites require further excavation.

4.6.4 Linguistics

Large language families (e.g. Austronesian, Indo-European, Afro-Asiatic, Bantu, Sino-Tibetan, Uto-Aztecan) owe their very existences to the geographical expansions of speakers of proto-languages within these families, consequent initially upon the regional beginnings of agriculture with its stimulus to rapid demographic growth (Bellwood, 1998). Bellwood further concludes that major language families on this scale did not spread by simple cultural contact alone, but that population movement would have been required.

Blust concludes that Malayo-Polynesian languages all descend from the same ancestor, a proto-Austronesian language spoken “in or near Taiwan” (Gibbons, 2001: see Figure 4.4, Model I). However, Oppenheimer and Richards (2001a) disagree, suggesting that the linguistic phenomenon of ‘levelling’ may have occurred. Levelling refers to the phenomenon whereby the expansion of one or several closely related languages erases the previously existing diversity (see Figure 4.4, Model II). Therefore, if Taiwan had simply been an Austronesian backwater (Meacham, 1985), earlier levels of diversity may well have survived. Deeper diversity would have persisted more as a result of isolation than of greater antiquity; whilst in island Southeast Asia the language groups may have been levelled as a result of contact and the demographic and cultural expansions of Neolithic peoples (Oppenheimer and Richards, 2001). Model I requires the extinction of one proto-Austronesian subgroup in Taiwan, whereas model II requires the extinction of an unspecified number of proto-Austronesian languages outside of Taiwan. Which scenario is most likely? In addition to the extinction of the proto-Malayo-Polynesian sub-group in Taiwan, Model I also requires the extinction of the languages that were spoken originally by the *ancient* populations that entered Near

Oceania during Phase One of Pacific colonisation. Currently most languages in Near Oceania are Austronesian, implying a complete linguistic sweep, usually explained by the demographic advantage of the *migrant* population. Alternatively Model II softens the sweep, by inferring that the populations in these regions already spoke related languages turning complete replacement into an emergence of a dominant sub-type. It is unknown how many proto-Austronesian language subtypes were present in the Near Oceanic region, but they must all have become extinct when the Malayo-Polynesian language became dominant (except in Taiwan).

4.6.5 Y Chromosome markers

Polynesian men have exceptionally low Y chromosome haplotype diversity, indicating a small number of male founders (Capelli et al., 2001, Kayser et al., 2000), in agreement with bottlenecks inferred from using mtDNA and other genetic evidence. However, in contrast to the mtDNA studies, there is strong evidence for a substantial input of Melanesian Y chromosomes (Kayser et al., 2000) and in some cases extensive European admixture (Hurles et al., 1998) in Polynesian populations. Interpretations of Y chromosome data seem to vary depending on which markers are chosen for study, with some studies supporting, and others rejecting possible Taiwanese origins.

First, a study by Su et al. (2000) using biallelic markers, found very few Taiwanese haplotypes in Micronesia and Polynesia, and a complete absence of a Melanesia-specific haplotype among Polynesians. Therefore, this study concluded that Southeast Asia provided the genetic source for two independent migrations: one toward Taiwan and the other toward Polynesia through island Southeast Asia. It also concluded that

Melanesian Y chromosomal contribution to the Polynesian expansion was very low or negligible!

A later study by Kayser et al. (2000) using single nucleotide polymorphisms (SNPs) and short tandem repeat (STR) loci, found that all Polynesians belonged to just three Y chromosome haplotypes, of which the major type (82%) was restricted to eastern Indonesia and Melanesia, where it likely arose. The other two Polynesian Y haplotypes were widespread in Asia, but are also found in Melanesia, thereby not providing a diagnostic feature for the influence of Asian populations on Polynesia. Additionally, this paper claims to introduce the 'slow-boat model', although Oppenheimer and Richards are more generally credited with being the first to propose this type of model (Gibbons, 2001).

A more recent study by Capelli et al. (2001) found 2 major haplogroups (C and L), which together account for over 95% of the Y chromosome diversity in Polynesian populations. Haplogroup C chromosomes are probably indigenous to Southeast Asia and Melanesia as this is where the highest diversity occurs. Back migration may explain the small frequency of Haplogroup C in Taiwan (2%) and Southern China (4%), as presumably it did not originate in these areas. Polynesians are most strongly associated with Melanesian haplogroup C clusters, indicating that most (~64%) Polynesian Y chromosomes are derived from Melanesian populations. This evidence supports Terrell's Melanesian theory, at least regarding male input. Conversely, the frequency and diversity of haplotype L is highest in one of the Taiwanese aboriginal populations (Ami), where it probably arose. Therefore, Haplogroup L links most of the other Polynesian Y chromosomes (32%) with southern China and Taiwan, supporting at least partially Bellwood's Express Train theory. However, despite admitting that there is a

clear increase of southern Chinese and Taiwanese L haplogroups in Polynesia relative to Melanesia, Capelli et al. (2001) dismiss this link stating that “*the extensive genetic drift in Polynesia makes it difficult to draw any strong inferences based on this difference*”.

A study focusing specifically on the New Zealand Maori population (Underhill et al., 2001), using Y chromosome binary marker analysis, found that approximately 40% of the samples had haplotypes of unequivocal European origin. Another ~40% of samples had single binary haplotypes that are also observed in Indonesia and New Guinea, indicative of common indigenous Melanesian ancestry. The remaining haplotypes had either typical East Asian signatures or alternative compositions consistent with their affinity to Melanesia or New Guinea. Therefore this study does not provide any direct links with Taiwan and consequently does not support the ‘Express Train’ scenario for male lineages.

The most recent study by Hurles et al. (2002) does not support the ‘Express Train’ theory either and instead suggests:

“most Micronesian and Polynesian Y chromosomes appear to originate from different source populations within Melanesia and Eastern Indonesia”.

This evidence is based on MSYI, binary markers and microsatellites.

Overall the Y chromosome evidence appears to vary depending on what markers are used for analysis. Although most studies agree on some Melanesian input into Polynesian populations, disagreements centre on the relative contributions of Melanesian versus Southeast Asian populations. Additionally, there is disagreement over whether Taiwan was involved in the Polynesia expansion.

4.6.6 ADH and ALDH data

A study on Alcohol and Aldehyde Dehydrogenase metabolism (Chambers et al., 2002) supports a link between Polynesians (including Maori) and Taiwanese hill tribes. The frequency of *ALDH 2*2* in Polynesians was found to be 0.00, in contrast with the high frequency (0.24) commonly reported for Orientals. However, the frequency of *ALDH 2*2* was found to be particularly low in the Taiwanese aborigine tribes specifically, Ami (0.02) and Paiwan (0.05), suggesting that Taiwan may have been an early staging post during migration in line with the ideas discussed by Diamond (1988).

5. Summary and Conclusions

5.1 Synopsis of Major findings

The major findings of this thesis are summarised below in relation to the research aims (italicised) from section 1.7:

- *To collect a molecular database of at least 50 partial mtDNA sequences from Maori and Pacific Nation volunteers.*

In this study, approximately 98 partial mitochondrial DNA sequences were obtained from 85 Maori and Polynesian banked DNA samples using PCR amplification and automated sequencing. The resulting molecular database consists of 61 Maori and 24 Pacific Nation 3' HVI mtDNA sequences, and 7 Maori and 6 Pacific Nation HVII mtDNA sequences. Additionally, each of the 85 DNA samples was typed for the presence or absence of the 9-bp deletion. The mtDNA sequence and associated 9-bp deletion status of each sample will be deposited in GenBank, for access by the scientific community, after this thesis is submitted.

- *To search this database for characteristic elements including the 9-bp deletion and the Polynesian CGT motif.*

The 9-bp deletion and the Polynesian CGT motif were present at extremely high frequencies (~87% and 100% respectively) in the mtDNA haplotype database compiled for the current study. This pattern strongly supports the theory of repeated bottlenecks during the journey across Far Oceania to colonise Aotearoa.

- *To use this database to refine the estimates of Murray-McIntosh et al. (1998) for the number of Maori women founders for New Zealand's Maori population.*

After taking into account the number of haplotypes found AND their frequencies, the current study suggests that ~56 women (between 34 and 120) were necessary to found the current New Zealand Maori population, whereas the previous study by Murray-McIntosh et al (1998) estimated 70 woman (between 50 and 100) were required. Both studies reject models that support a small number of founders, or accidental settlement by just a single waka. Instead, the large number of female founders needed to account for the present Maori diversity, support oral history whereby an unknown number of waka deliberately made the journey from Eastern Polynesia to Aotearoa (New Zealand). The current study is superior to the MM98 study in the following 3 ways:

1. uses larger Ma and EP mtDNA haplotype database (AW02 new data + all old data)
2. more accurate collection of data for comparison (3' regions compared across all studies)
3. computer simulations use more appropriate sigmoid population growth curves

- *To evaluate the ‘competing’ models of Oceanic population expansion, and attempt to consolidate these theories into the Synthetic Total Evidence Theory.*

All current Oceanic colonisation theories have elements in common that allow them to be incorporated into the ‘Synthetic Total Evidence Theory’. The Synthetic Total Evidence Theory introduced in this thesis is a first attempt at consolidating ‘competing’ theories using interdisciplinary evidence.

5.2 Conclusions

The research presented in this thesis clearly demonstrates the ability of mtDNA to elucidate the patterns and processes involved in the evolution and migration of anatomically modern *Homo sapiens* within the Oceania region. The mtDNA HVI and 9-bp deletion evidence strongly links together the populations resident in the islands of Remote Oceania, and provides evidence for recurring bottlenecks in the settlement history of this region. The haplotype frequencies identified in this study, were used in conjunction with frequencies described in previous studies, to estimate the number of Maori women required to found the current Maori population.

In Near Oceania, the relationships between populations are more difficult to trace due to admixture between ancient and migrant populations. However, combining interdisciplinary evidence from archaeology, anthropology and linguistics, has allowed some conclusions regarding the settlement of near Oceania to be made (see Synthetic Total Evidence Theory, 4.6.1).

Future Investigations

There is still a considerable amount of future work that could be done using human molecular population genetics to elucidate the history of Oceanic people, and to provide additional answers to the long-standing questions about human evolution. Possible future directions will be outlined briefly below:

Mitochondrial DNA

As mentioned previously, the homopolymeric cytosine tract may be a good model substrate for looking at DNA polymerase error rates. Using a variety of primer templates incorporating a variety of C-tracts lengths, e.g. CCCCTCCCC; CCCCCCTCCC; CCCCCCTCC it should be possible to determine when the sequencing becomes difficult to read, and how these errors occur. Once the underlying mechanisms behind the production of the 'blurred' sequence are understood, a system for correcting this error could be devised thereby creating a more effective system for sequencing the entire HVI region.

Increased Sampling

It might even be possible for a Maori scholar to trace the ancestry of the rare mtDNA haplotypes back to a specific waka. On the molecular side this would be straight forward, but in reality this endeavour would require a very full knowledge of oral history and to cope with an enormous amount of tea drinking! Additionally (as mentioned previously), it may be possible to trace these rare haplotypes back to a

specific island or group of islands, providing evidence for one or possibly more Maori 'homeland(s)'. To achieve this objective, more extensive sampling in the suspected homeland regions of Tonga-Samoa and Tahiti would need to occur. To this end collaboration may be possible with Professor David Penny at the Massey Institute of Molecular BioSciences, who is planning further research into the characterisation of Hepatitis B virus from Melanesian and Polynesian individuals (pers. comm., 2002). This project requires the cooperation of the Fiji School of Medicine, and samples may be shared with the IMS facility at Victoria University.

Additional Computer simulations

There are numerous additional population growth simulations that could be carried out using the new enhanced dataset. First, more work could be done to ensure that the underlying population growth model is appropriate for the population expansion of the prehistoric Maori people. Preliminary work done during this study, suggests that an exponential growth curve might not be appropriate, and that sigmoid growth curves give improved founder estimates. However, the most appropriate sigmoid curve to describe the population growth of the Maori people is currently unknown. The early and middle sigmoid curves used in the current study, both give founding population values that are within the accepted range, but an intermediate value for where the curve starts expanding from (parameter b: see Appendix Two) may give the best results. However, one must be careful in this situation, as there could be a tendency to change the underlying model to give results within the 'expected' range thereby manipulating the results to unjustly favour an 'intermediate' model over the others. Before using an intermediate model additional research would be required to substantiate the claim that this model 'best' fits the archaeological data. Second, simulations could be done using

mutations to account for the extra haplotypes that have been found in the Maori population. Some additional simulations using the mutation rates discussed in Section 4.4 (see New Mutations) are planned, before submitting the results of this thesis to a peer-reviewed journal. Thirdly, the haplotype frequencies in the New Zealand Maori population had an unusual skewed distribution. It was found using additional computer analyses that to produce the 'unequal' haplotype frequencies present in the current population requires an increase the founding population. Additional simulations could be done to refine these calculations. Finally, as Y chromosome information becomes available an identical simulation could be done to estimate the number of Maori male founders required for the present Maori population.

The 'Synthetic Total Evidence Theory'

It is anticipated that the introduction of this theory in a peer-reviewed journal will stimulate further discussion and debate. Throughout the years attempts have been made by noted authors such as Peter Bellwood to make sense and consolidate the wide-ranging interdisciplinary data that has been collected in the Oceanic region over the past 20-30 years. The 'Synthetic Total Evidence Theory' may provide the necessary framework to allow experts from interdisciplinary fields to cooperate together to refine aspects of a single model. By focusing only on their area of expertise, collaboration with other experts in other disciplines would be encouraged, which would remove some of the opportunities for misinterpreted data and lead to the first truly interdisciplinary model.

Appendices

Appendix One: Solutions

0.5M EDTA (pH 8.0)

93.05 g disodium ethylenediaminetetraacetate.2H₂O (EDTA, GibcoBRL[®]) in 400 ml ddH₂O.

Stir vigorously on magnetic stirrer to dissolve.

Adjust pH to 8.0 using NaOH pellets (approximately 10 g).

Make up volume to 500 ml with ddH₂O and autoclave before use.

5x Tris Borate (TBE) Buffer

108g of Tris Base.

55g of Boric Acid.

40ml of 0.5 EDTA pH 8.0.

Make up to 2 L with ddH₂O.

1x TBE Buffer

1:5 ddH₂O dilution of stock solution made up as needed.

50x Tris Acetate (TAE) Buffer

484g Tris base.

114.2 ml Glacial acetic acid.

200 ml of 0.5 EDTA (pH 8.0).

Make up volume to 2 L with ddH₂O.

1x TAE Buffer

1:50 ddH₂O dilution of stock solution made up as needed.

Red Blood Cell Lysis Solution (10x)

16.04g of NH₄Cl (1.5M).

2g of KHCO₃.

0.4 ml of EDTA (0.5M).

Dissolve in 180 ml of ddH₂O.

Make up volume to 200 ml with ddH₂O.

Autoclave and store at 4°C.

Dilute solution to 1x before using.

TBS

16g of NaCl
0.76g of KCl
6.06g Tris-HCl base
Dissolve in 180 ml of ddH₂O.
Adjust to pH 7.4 using concentrated HCl
Make up volume to 200 ml with ddH₂O.
Autoclave and store at 4°C.

Dilute solution to 1x before using.

10% SDS

10 g Sodium Dodecyl Sulphate.
100 ml ddH₂O.
Dissolve by inversion

Sequencing Loading Buffer

200 µl formamide
40 µl blue dextran

Agarose Gel Protocols

1% Agarose Gel

0.6 g SeaKem LE[®] agarose (FMC BioProducts).
60 ml 1X TBE buffer.

Microwave for 1 minute, mix by swirling. Microwave for another 15 seconds until all the agarose has dissolved. Add 2 µl ethidium bromide (10 mg/ml). Allow solution to cool to approximately 50°C. Pour into taped minigel plate/comb apparatus. Leave to set for 30 minutes at room temperature. Gel contains 0.6 µgml⁻¹ ethidium bromide, used for DNA visualisation

2% Agarose Gel

1.2 g SeaKem LE[®] agarose (FMC BioProducts).
60 ml 1X TBE buffer.

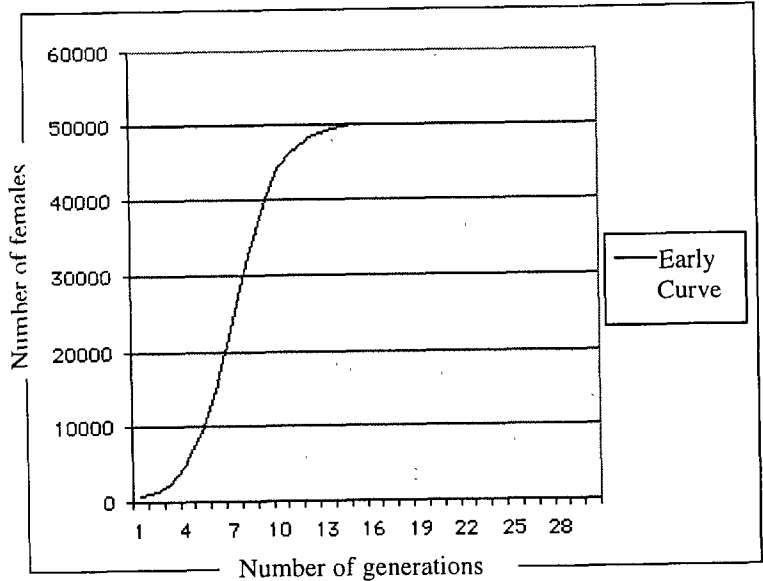
Follow directions as shown for 1% Agarose Gel.

Appendix Two: Population Growth Models used for estimating the number of Female founders for the current Maori population

Model A : Early population expansion

Parameter	Value
number of initial haplotypes	21
initial frequencies	72,13,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1
model	sigmoid (logistic)
a	5
b	-0.7
initial population	2-1000
final population	50000
generations	30
final sample	107
number of runs	20000

Population growth followed the following curve:

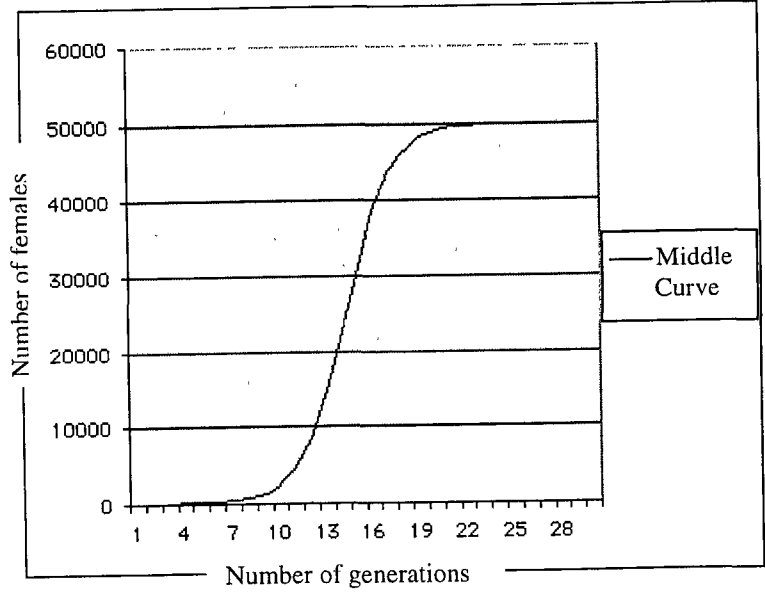


NB: Parameter 'a' controls where in the x axis the curve starts growing from
Parameter 'b' controls the slope of the midpoint of the curve (how steep or shallow it is)

Model B: Middle population expansion

Parameter	Value
number of initial haplotypes	21
initial frequencies	72,13,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1
model	sigmoid (logistic)
a	10
b	-0.7
initial population	2-1000
final population	50000
generations	30
final sample	107
number of runs	20000

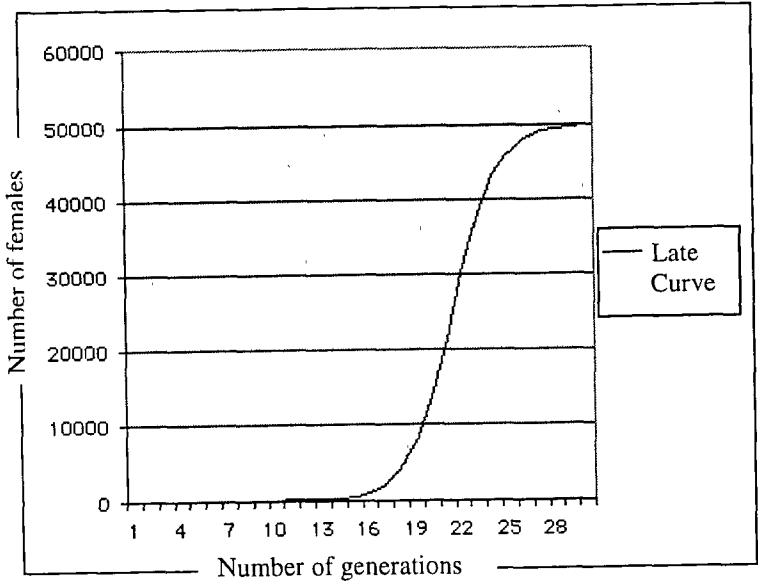
Population growth followed the following curve:



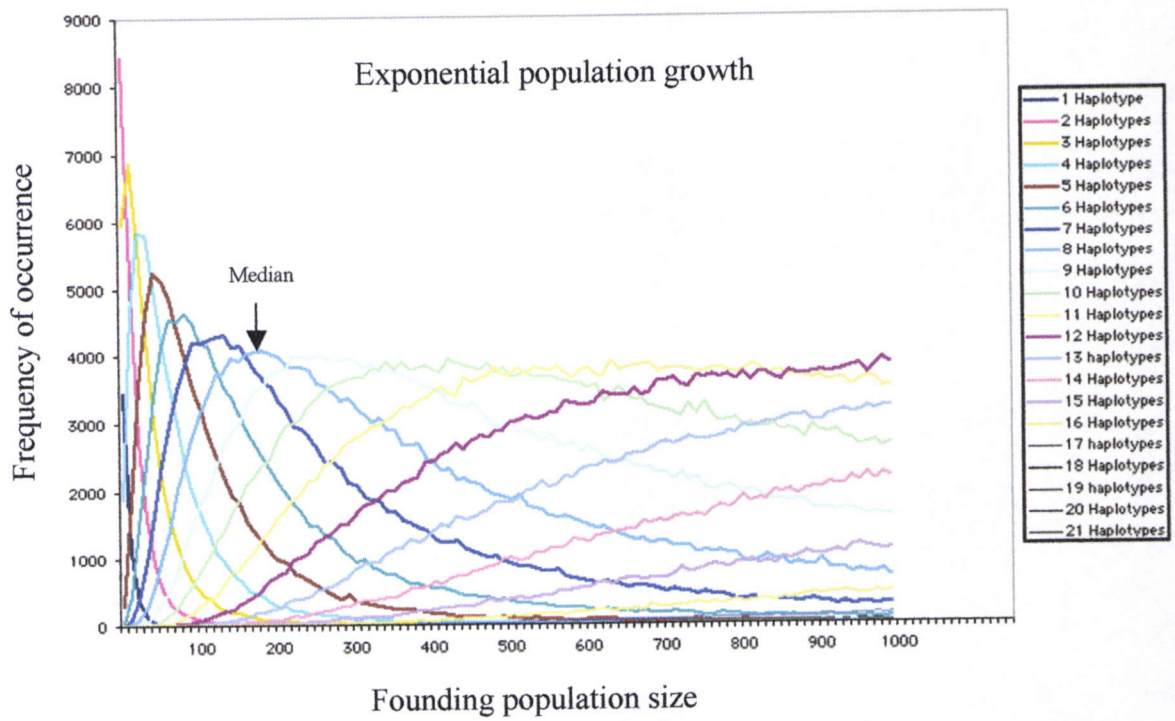
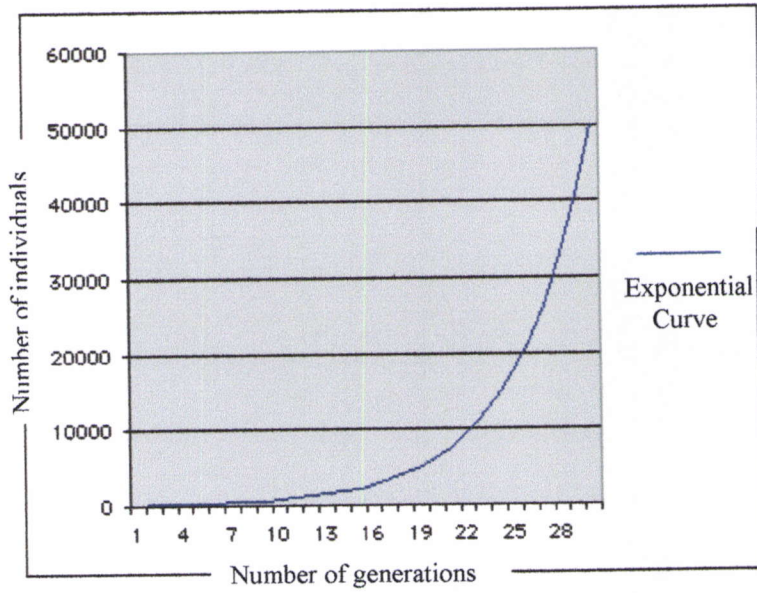
Model C: Late population expansion

Parameter	Value
number of initial haplotypes	21
initial frequencies	72,13,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1
model	sigmoid (logistic)
a	15
b	-0.7
initial population	2-1000
final population	50000
generations	30
final sample	107
number of runs	20000

Population growth followed the following curve:



Model D: Exponential population expansion



Appendix Three: Complete 9-BP/HVI (16 024-16 400) haplotype description for each sample

NB Identifying transitions are indicated in bold. This identifies the rare haplotypes, found on Table 3.1.

1. 273- No 9p Deletion, Caucasian (Not on shown on Table 3.1)
Control for 9-bp deletion
16 270 C→T 16304 T→C
2. 418- Deletion, Maori (Ht #AW1 on Table 3.1)
16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
Tract =12C's, Confirmed by L strand ht, therefore 16 182 A⇒C 16 183 A⇒C
3. 420- No 9bp Deletion, European Haplotype H, Maori (Ht #AW14 on Table 3.1)
16 147 C→T 16304 T→C No Tract
4. 421- No 9bp Deletion, European Haplotype T, Maori (Ht #AW15 on Table 3.1)
16 292 C→T 16 294 C→T 16 126 G→C No Tract
5. 422- Deletion, Maori 100% (Ht #AW1 on Table 3.1)
16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
Tract =12C's (est) so 16 182 A→C 16 183 A→C
6. 426- Deletion, Maori 100% (Ht #AW1 on Table 3.1)
16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
Tract =12C's, Confirmed by L strand ht, therefore 16 182 A→C 16 183 A→C
7. 427- Deletion, Maori (Ht #AW1 on Table 3.1)
16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
Tract =12C's, Confirmed by L strand ht, therefore 16 182 A→C 16 183 A→C
8. 429- Deletion, Maori (Ht #AW1 on Table 3.1)
16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
Tract =12C's (est) so 16 182 A→C 16 183 A→C
9. 432- Deletion, Maori (Ht #AW1 on Table 3.1)
16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
Tract =12C's, Confirmed by L strand ht, therefore 16 182 A→C 16 183 A→C
10. 433- Deletion, Maori (Ht #AW1 on Table 3.1)
16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
Tract =12C's (est) so 16 182 A→C 16 183 A→C
11. 436- Deletion, Maori 100% (Ht #AW1 on Table 3.1)
16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
Tract =12C's (est) so 16 182 A→C 16 183 A→C

12. 447- Deletion, Maori (Ht #AW4 on Table 3.1)
16 189 T→C **16 207 A→G** 16 217 T→C 16 247 A→G 16 261 C→T
Tract =12C's (est) so 16 182 A→C 16 183 A→C
13. 449- Deletion, Maori (Ht #AW1 on Table 3.1)
16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
Tract =12C's (est) so 16 182 A→C 16 183 A→C
14. 451- Deletion, Maori (Ht #AW1 on Table 3.1)
16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
Tract =12C's (est) so 16 182 A→C 16 183 A→C
15. 452- Deletion, Maori (Ht #AW2 on Table 3.1)
16 189 T→C 16 217 T→C **16 247 same as ref seq** 16 261 C→T
Tract =12C's (est) so 16 182 A→C 16 183 A→C
16. 453- Deletion, Maori (Ht #AW1 on Table 3.1)
16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
Tract =12C's (est) so 16 182 A→C 16 183 A→C
17. 454- Deletion, Maori (Ht #AW1 on Table 3.1)
16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
Tract =12C's (est) so 16 182 A→C 16 183 A→C
18. 455- European Haplotype U, Maori (Ht #AW16 on Table 3.1)
16 256 C→T 16 270 C→T
19. 456- Deletion, Maori 100% (Ht #AW1 on Table 3.1)
16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
Tract =12C's (est) so 16 182 A→C 16 183 A→C
20. 457- Deletion, Maori 100% (Ht #AW1 on Table 3.1)
16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
Tract =12C's (est) so 16 182 A→C 16 183 A→C
21. 458- No 9bp Deletion, European Haplotype H, Maori/Euro 50% (Ht #AW17 on Table 3.1) 16 162 A→G 16 209 T→C
22. 459- Deletion, Maori/Euro 50% (Ht #AW1 on Table 3.1)
16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
Tract =12C's (est) so 16 182 A→C 16 183 A→C
23. 460- Deletion, Maori 50% (Ht #AW1 on Table 3.1)
16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
Tract =12C's, Confirmed by L strand ht, therefore 16 182 A→C 16 183 A→C
24. 462- Deletion, Maori 100% (Ht #AW1 on Table 3.1)
16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
Tract =12C's (est) so 16 182 A→C 16 183 A→C

25. 464- Deletion, Maori 100% (Ht #AW1 on Table 3.1)
16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
Tract =12C's (est) so 16 182 A→C 16 183 A→C

26. 466- Deletion, Maori 100% (Ht #AW1 on Table 3.1)
16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
Tract =12C's (est) so 16 182 A→C 16 183 A→C

27. 468- Deletion, Maori 100% (Ht #AW5 on Table 3.1)
16 189 T→C 16 217 T→C **16 142 C→T** 16 247 A→G
16 261 C→T Tract =12C's, Confirmed by L strand ht, therefore so 16 182
A→C 16 183 A→C

28. 469- Deletion, Maori 100% (Ht #AW1 on Table 3.1)
16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
Tract =12C's (est) so 16 182 A→C 16 183 A→C

29. 470- Deletion, Maori 100% (Ht #AW1 on Table 3.1)
16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
Tract =12C's (est) so 16 182 A→C 16 183 A→C

30. 471- No 9bp Deletion, European Haplotype H, Maori/Euro 50% (Ht #AW18 on
Table 3.1) Same as reference sequence

31. 472- No 9bp Deletion, European Haplotype U, Maori/Euro 50% (Ht #AW20 on
Table 3.1) 16 224 T→C 16 311 T→C

32. 474- Deletion, Maori 100% (Ht #AW1 on Table 3.1)
16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
Tract =12C's (est) so 16 182 A→C 16 183 A→C

33. 476- Deletion, Maori 100% (Ht #AW6 on Table 3.1)
16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
Tract =12C's (est) so 16 182 A→C 16 183 A→C **16291 C→T**

34. 477- Deletion, Maori (Ht #AW1 on Table 3.1)
16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
Tract =12C's (est) so 16 182 A→C 16 183 A→C

35. 479- Deletion, Maori 100% (Ht #AW1 on Table 3.1)
16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
Tract =12C's (est) so 16 182 A→C 16 183 A→C

36. 480- Deletion, Maori 100% (Ht #AW1 on Table 3.1)
16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
Tract =12C's (est) so 16 182 A→C 16 183 A→C

37. 484- Deletion, Maori 100% (Ht #AW7 on Table 3.1)
16 271 T→C 16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
Tract =12C's (est) so 16 182 A→C 16 183 A→C

38. 488- No 9bp Deletion, European Haplotype U, Maori (Ht #AW19 on Table 3.1)
16 356 T→C
39. 500- Deletion, Maori 100% (Ht #AW1 on Table 3.1)
16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
Tract =12C's (est) so 16 182 A→C 16 183 A→C
40. 501- Deletion, Maori 100% (Ht #AW2 on Table 3.1)
16 189 T→C 16 217 T→C **16 247 same as ref seq** 16 261 C→T
Tract =12C's (est) so 16 182 A→C 16 183 A→C
41. 502- Deletion, Maori 100% (Ht #AW1 on Table 3.1)
16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
Tract =12C's (est) so 16 182 A→C 16 183 A→C
42. 503- Deletion, Maori 100% (Ht #AW1 on Table 3.1)
16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
Tract =12C's (est) so 16 182 A→C 16 183 A→C
43. 504- Deletion, Maori 100% (Ht #AW2 on Table 3.1)
16 189 T→C 16 217 T→C **16 247 same as ref seq** 16 261 C→T
Tract =12C's (est) so 16 182 A→C 16 183 A→C
44. 505- Deletion, Maori 100% (Ht #AW1 on Table 3.1)
16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
Tract =12C's (est) so 16 182 A→C 16 183 A→C
45. 506- Deletion, Maori 100% (Ht #AW1 on Table 3.1)
16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
Tract =12C's (est) so 16 182 A→C 16 183 A→C
46. 507- Deletion, Maori 100% (Ht #AW1 on Table 3.1)
16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
Tract =12C's (est) so 16 182 A→C 16 183 A→C
47. 508- Deletion, Maori 75% (Ht #AW1 on Table 3.1)
16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
Tract =12C's (est) so 16 182 A→C 16 183 A→C
48. 509- Deletion, Maori 100% (Ht #AW2 on Table 3.1)
16 189 T→C 16 217 T→C **16 247 same as ref seq** 16 261 C→T
Tract =12C's (est) so 16 182 A→C 16 183 A→C
49. 510- Deletion, Maori 100% (Ht #AW8 on Table 3.1)
16 189 T→C 16 217 T→C **16 247 same as ref seq** 16 261 C→T
Tract =12C's Confirmed by L strand ht, therefore 16 182 A→C 16 183 A→C
16 051 A→G
50. 511- Deletion, Maori 100% (Ht #AW1 on Table 3.1)
16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
Tract =12C's Confirmed by L strand ht, therefore 16 182 A→C 16 183 A→C

51. 512- Deletion, Maori 100% (Ht #AW3 on Table 3.1)
 16 189 T→C 16 217 T→C 16 247 A→G **16 261 same as ref seq**
 Tract =12C's (est) so 16 182 A→C 16 183 A→C
52. 514- Deletion, Maori 100% (Ht #AW1 on Table 3.1)
 16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
 Tract =12C's (est) so 16 182 A→C 16 183 A→C
53. 516- Deletion, Maori 100% (Ht #AW1 on Table 3.1)
 16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
 Tract =12C's (est) so 16 182 A→C 16 183 A→C
54. 517- Deletion, Maori 100% (Ht #AW1 on Table 3.1)
 16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
 Tract =12C's (est) so 16 182 A→C 16 183 A→C
55. 518- Deletion, Maori 75% (Ht #AW1 on Table 3.1)
 16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
 Tract =12C's (est) so 16 182 A→C 16 183 A→C
56. 519- Deletion, Maori 100% (Ht #AW1 on Table 3.1)
 16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
 Tract =12C's (est) so 16 182 A→C 16 183 A→C
57. 520- Deletion, Maori 100% (Ht #AW1 on Table 3.1)
 16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
 Tract =12C's (est) so 16 182 A→C 16 183 A→C
58. 521- Deletion, Maori 100% (Ht #AW1 on Table 3.1)
 16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
 Tract =12C's (est) so 16 182 A→C 16 183 A→C
59. 523- Deletion, Maori 100% (Ht #AW1 on Table 3.1)
 16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
 Tract =12C's (est) so 16 182 A→C 16 183 A→C
60. 524- Deletion, Maori 100% (Ht #AW1 on Table 3.1)
 16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
 Tract =12C's (est) so 16 182 A→C 16 183 A→C
61. 525- Deletion, Maori 100% (Ht #AW1 on Table 3.1)
 16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
 Tract =12C's (est) so 16 182 A→C 16 183 A→C
62. 534- Deletion, Maori 100% (Ht #AW1 on Table 3.1)
 16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
 Tract =12C's (est) so 16 182 A→C 16 183 A→C
63. 600- European Haplotype H, Maori/Hawaiian/Euro 50% (Ht #AW21 on Table 3.1) 16 231 T→C 16 304 T→C 16 311 T→C

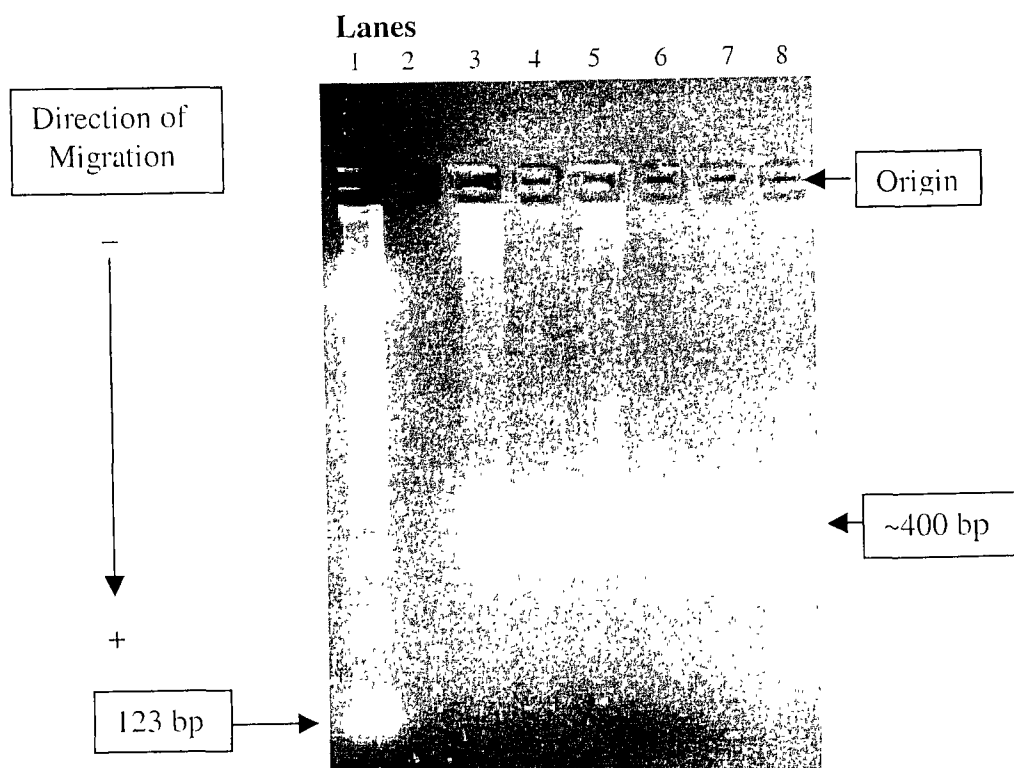
64. 601- Deletion, Polynesian Fijian/Euro 50% (Ht #AW13 on Table 3.1)
 16 189 T→C 16 217 T→C **16 247 same as ref seq 16 261 same as ref seq**
Tract =11C's, confirmed by L strand? so **16 182 same as ref seq 16 183**
 A→C
65. 602- Deletion, Polynesian Fijian 100% (Ht #AW9 on Table 3.1)
 16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
 Tract =12C's (est) so 16 182 A→C 16 183 A→C **16360 C→T**
66. 603- Deletion, Polynesian (Ht #AW2 on Table 3.1)
 16 189 T→C 16 217 T→C **16 247 same as ref seq 16 261 C→T**
 Tract =12C's (est) so 16 182 A→C 16 183 A→C
67. 604- No Deletion, European Haplotype U, Polynesian Rarotonga (Ht #AW16 on Table 3.1) 16 256 C→T 16 270 C→T
68. 605- Deletion, Polynesian Cook Island 50% (Ht #AW1 on Table 3.1)
 16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
 Tract =12C's (est) so 16 182 A→C 16 183 A→C
69. 606- Deletion, Polynesian Samoan (Ht #AW1 on Table 3.1)
 16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
 Tract =12C's, Confirmed by L strand ht, therefore 16 182 A→C 16 183 A→C
70. 607- Deletion, Polynesian Samoan (Ht #AW2 on Table 3.1)
 16 189 T→C 16 217 T→C **16 247 same as ref seq 16 261 C→T**
 Tract =12C's (est) so 16 182 A→C 16 183 A→C
71. 608- Deletion, Polynesian W. Samoan (Ht #AW1 on Table 3.1)
 16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
 Tract =12C's (est) so 16 182 A→C 16 183 A→C
72. 609- Deletion, Polynesian W. Samoan (Ht #AW1 on Table 3.1)
 16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
 Tract =12C's, Confirmed by L strand ht, therefore 16 182 A→C 16 183 A→C
73. 610- Deletion, Polynesian Tonga (Ht #AW1 on Table 3.1)
 16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
 Tract =12C's, Confirmed by L strand ht, therefore 16 182 A→C 16 183 A→C
74. 611- Deletion, Polynesian Tonga (Ht #AW12 on Table 3.1)
 16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T **16 311 T→C**
 Tract =12C's, Confirmed by L strand ht, therefore 16 182 A⇒C 16 183 A⇒C
75. 612- Deletion, Polynesian Rarotonga/Scot/Germ/Eng/Maori 50% (Ht #AW1 on Table 3.1) 16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
 Tract =12C's (est) so 16 182 A→C 16 183 A→C
76. 613- Deletion, Polynesian Niue (Ht #AW10 on Table 3.1)
 16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T **16 181 A→G**
 Tract =12C's, Confirmed by L strand ht, therefore 16 182 A→C 16 183 A→C

77. 614- Deletion, Polynesian Samoan/Germ 50% (Ht #AW1 on Table 3.1)
 16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
 Tract =12C's (est) so 16 182 A→C 16 183 A→C
78. 615- Deletion, Polynesian (Ht #AW1 on Table 3.1)
 16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
 Tract =12C's (est) so 16 182 A→C 16 183 A→C
79. 616- Deletion, Polynesian Tokelau (Ht #AW9 on Table 3.1)
 16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
 Tract =12C's (est) so 16 182 A→C 16 183 A→C **16 360 C→T**
80. 617- Deletion, Polynesian Cook/ Maori/Scot 50% (Ht #AW1 on Table 3.1)
 16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
 Tract =12C's (est) so 16 182 A→C 16 183 A→C
81. 618- Deletion, Polynesian Cook Island (Ht #AW2 on Table 3.1)
 16 189 T→C 16 217 T→C **16 247 same as ref seq** 16 261 C→T
 Tract =12C's (est) so 16 182 A→C 16 183 A→C
82. 621- Deletion, Polynesian Tonga/Nuie 50% (Ht #AW1 on Table 3.1)
 16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
 Tract =12C's (est) so 16 182 A→C 16 183 A→C
83. 622- Deletion, Polynesian (Ht #AW1 on Table 3.1)
 16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
 Tract =12C's (est) so 16 182 A→C 16 183 A→C
84. 623- Deletion, Polynesian (Ht #AW1 on Table 3.1)
 16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
 Tract =12C's (est) so 16 182 A→C 16 183 A→C
85. 624- Deletion, Polynesian (Ht #AW11 on Table 3.1)
 16 189 T→C 16 217 T→C **16 234 C→T** 16 247 A→G 16 261 C→T
 Tract =12C's (est) so 16 182 A→C 16 183 A→C
86. 625- Deletion, Polynesian (Ht #AW1 on Table 3.1)
 16 189 T→C 16 217 T→C 16 247 A→G 16 261 C→T
 Tract =12C's, Confirmed by L strand ht, therefore 16 182 A→C 16 183 A→C
87. A- **Author, Maori on fathers side, European on Mothers side, European Haplotype U** (Not on shown on Table 3.1)
 16 051 G→A 16 092 T→C 16 129 G→C 16 189 T→C 16 362 A→G
88. N- **Philippine sample, used as a control make sure DNA extraction protocol was working** (Not on shown on Table 3.1)
 16 223 C→T 16 295 C→T

HVII (16 024-16 400) haplotype for selected samples

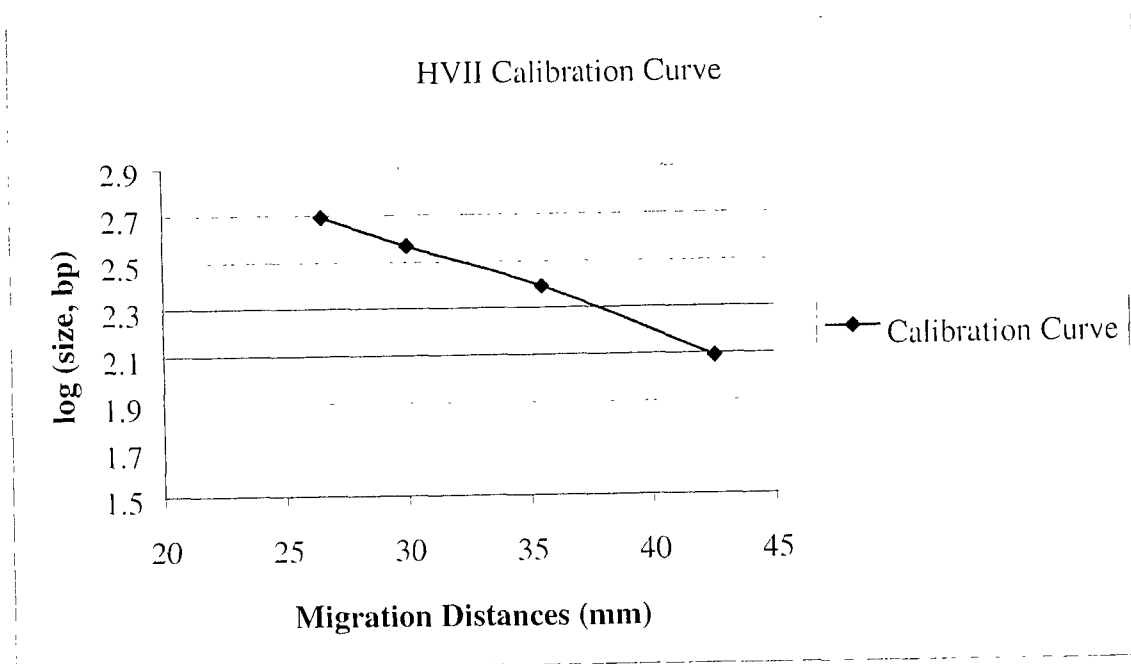
1. 449 – Maori, 00 073 A→G, 00 146 T→G, 00 263 A→G
2. 455 – Maori, **Euro U**, 00 073 A→G, **00 146 same as ref seq**, 00 263 A→G
3. 458 – Maori, **Euro H**, 00 073 A→G, **00 146 same as ref seq**, 00 263 A→G
4. 460 – Maori 50%, 00 073 A→G, 00 146 T→G, 00 263 A→G
5. 466 – Maori 100%, 00 073 A→G, 00 146 T→G, 00 263 A→G
6. 472 – Maori/Euro 50%, **Euro U**, 00 073 A→G, **00 146 same as ref seq**, 00 263 A→G, **00 279 T→C**
7. 507 – Maori 100%, **CGT Motif**, 00 073 A→G, **00 146 same as ref seq**, 00 263 A→G
8. 612 – Polynesian Rarotonga/Scot/Germ/English/Maori 50%, 00 073 A→G, 00 146 T→G, 00 263 A→G
9. 614 – Polynesian Samoan/Germ 50%, 00 073 A→G, 00 146 T→G, 00 263 A→G
10. 615 – Polynesian, 00 073 A→G, 00 146 T→G, 00 263 A→G
11. 617 – Polynesian Cook/Maori/Scot 50%, 00 073 A→G, 00 146 T→G, 00 263 A→G
12. 618 – Polynesian Cook Island, 00 073 A→G, 00 146 T→G, 00 263 A→G
13. 623 – Polynesian, 00 073 A→G, 00 146 T→G, 00 263 A→G

Appendix Four: Electrophoresis of HVII agarose gel and HVII Calibration Curve



Appendix Four: Results show 5 μ l aliquots of representative HVII PCR products amplified from 6 samples and visualised on a 2% SeaKem[®] agarose gel.

- Lane: 1. 123 bp DNA standard ladder
2. Negative reagent control (no template DNA)
3-8. Samples # 449, 455, 458, 460, 466, 472



Appendix Four: Calibration curve of log (size, base pairs) versus migration distance (mm) of the fragments that make up the 123 bp DNA standard ladder, shown above. The migration distance of PCR products was measured and their corresponding sizes estimated from calibration curves such as this one. From the migration distance of 28 mm a fragment size of ~400 bp was estimated. This size is approximate the expected size. The calibration curve was constructed using Microsoft Excel 2000.

Appendix Five: Combined Haplotype data (CD) for Maori, Eastern and Western Polynesian samples (i.e. EP+).

Group I	Haplotype (numbering according to reference sequence +16000 unless HVII*)	Number and location of individuals with each haplotype.	Ma	EP	WP	h# in original study
CD1	189, 217, 247, 261 *Redd sequences in HV1 as above (CD1), plus HVII 00073, 00146, 00263	11 (Haw), 5 (Sa), 2 (Ton) 12 (Marq), 20 (Ma), 2 (Ton), 14 (Sam), 46 (Cook) 14 (Sam) 27 (Ma) 44 (Ma), 2 (Cook), 1 (Raro), 4 (Sam), 2 (Ton)	20 27 44	11 58 3	7 16 14 6	Lum2 Sykes11 Redd1* MM1 AW1
CD1a	HV1 as above (CD1), HVII as above (*) plus 00385	1 (Sam)			1	Redd4*
CD1b	HV1 as above (CD1), HVII as above (*) plus 00151	1 (Sam)			1	Redd8*
CD1c	HV1 as above (CD1), HVII as above (*) plus 00196	1 (Sam)			1	Redd14*
CD1d	Rest of HV1 as above (CD1) plus mutation in HV1 region before 189	1 (Sam)			1	Redd10*
CD1e	163 (plus HVII*) 181	1 (Nu)			1	AW10
CD2	189, 217, __, 261	1 (Haw-ad), 5 (Sam) 3 (Sam), 10 (Cook), 1 (Marq), 2 (Ma) 2 (Sam) 3 (Ma) 1 (Cook), 4 (Ma), 1 (Sam)	2 3 4	1 11 1	5 3 2 1	Lum5 Sykes6 Redd 22* MM15 AW2
CD2a	Rest of HV1 as above (CD2) plus mutation in HV1 region before 189	1 (Ma)	1			AW8
CD3	051 189, 217, 247, __	1 (Cook) 1 (Ma) 1 (Ma)	1 1 1	1		Sykes19 MM11 AW3
CD4	189, 217, 247, 261, 361 (HVII*)	1 (Sam)			1	Redd2*
CD5	189, 217, __, 261, 311 (HVII*)	2 (Sam)			2	Redd18*
CD6	189, 213, 217, __, 261 (HVII*)	1 (Sam)			1	Redd24*

Group	Haplotype (numbering according to reference sequence +16000 unless HVII)	Number and location of individuals with each haplotype.	Ma	EP	WP	Study (plus ref# in that study)
Group I (confd)						
CD7	189, 217, 247, 261, 291	2 (Sam) 1 (Cook) 1 (Ma)	1	1	2	Lum1 Sykes13 AW6
CD8	189, 247, 261	1 (Hav)		1		Lum4
CD9	189, 213, 217, 261	1 (Sam)			1	Lum7
CD10	189, 217, 261, 362	1 (Sam)			1	Lum8
CD11	189, 217	1 (Sam; also 1 Amerindian-indic Sth Am infl?)			1	Lum11
CD12	189, 217, 261, 312	1 (Cook)		1		Sykes10
CD13	189, 217, 247, 261, 354 C-A	1 (Sam)			1	Sykes14
CD14	189, 217, 247, 261, 274 G-C	1 (Sam)			1	Sykes15
CD15	189, 217, 247, 261, 342	1 (Sam)			1	Sykes16
CD16	189, 217, 247, 261, 362	2 (Sam)			2	Sykes18
CD17	189, 217, 242, 247, 261	1 (Ma)	1			AW5
CD18	189, 207, 217, 247, 261	1 (Ma)	1			AW4
CD19	189, 217, 247, 261, 271	1 (Ma)	1			AW7
CD20	189, 217, 247, 261, 360	1 (Tok)		1		AW9
CD21	189, 217, 247, 261, 311	1 (Ton)			1	AW12
Group II						
	Haplotype (numbering according to reference sequence +16000 unless HVII)	Number and location of individuals with each haplotype.	Ma	EP	WP	Study (plus ref# in that study)
CD22	223, 241, 265, 274, 311	1 (Hav)		1		Lum39
CD22a	as above (CD22) plus 343 and mutations before 189 (129, 144, 148)	3 (Cook)		3		Sykes22
CD23	223, 241, 265, 293, 311	1 (Sam)			1	Lum40
CD23a	as above (23) plus 343 and mutations before 189 (129, 144, 148)	1 (Sam), 11 (Cook), 2 (Tah), 3 (Aus), 1 (Ma)	1	16	1	Sykes20
CD24	223, 241, 265, 311, 343 and mutations before 189 (129, 144, 148)	1 (Ton), 1 (Sam)			2	Sykes21
CD25	223, 241, 265, 311 and mutations before 189 (129, 144, 148)	2 (Ton)			2	Sykes23

Group III	Haplotype (numbering according to reference sequence +16000 unless HVII) Ht marked ** have mutation(s) before 189 shown in brackets	Number and location of individuals with each haplotype.	Ma	EP	WP	Study (plus ref # in that study)
CD26	294, 304, 362	1 (Sam)			1	Lum20
CD27	(172), 304 **	2 (Marq)		2		Sykes34
CD28	(172), 189, 304 **	1 (Tah)		1		Sykes35
CD29	(172), 304, 311 **	1 (Ton)			1	Sykes36
CD30	263	1 (Cook)		1		Sykes40
CD31	312	1 (Cook)		1		Sykes41
CD32	356	1 (Cook)		1		Sykes42
CD33	221, 264	1 (Cook)		1		Sykes43
CD34	210, 256, 269	1 (Cook)		1		Sykes44
CD35	(093), 223, 298, 327 **	1 (Cook)		1		Sykes45
CD36	192, 218, 270, 291	1 (Ma)	1			Sykes46
CD37	223, 290, 319, 362	1 (Tah)		1		Sykes47
CD38	235, 291, 293	1 (Ton)			1	Sykes48
CD39	189, 223, 291	1 (Ton)			1	Sykes49
CD40	(172), 223 **	1 (Marq)		1		Sykes50
CD41	298, 311	1 (Marq)		1		Sykes51
CD42	(126), 153, 294 **	1 (Marq)		1		Sykes52
CD43	223, 224, 270, 274, 311	1 (Marq)		1		Sykes53
CD44	223, 320	1 (Haw-ad)		1		Lum28

Key:

Ma	= Maori	= Ma
Haw	= Hawaiian	= EP
Marq	= Marquesan	= EP
Raro	= Rarotongan	= EP
Cook	= Cook Islander	= EP
Tok	= Tokelauan	= EP
Tah	= Tahitian	= EP
Aus	= Australes	= EP
Sam	= Samoan	= WP

Nu = Nuie = WP
Ton = Tongan = WP

Haw-ad, indicates that the individual acknowledges admixed ancestry.

Nucleotides which characterise a haplotype (ht) are underlined (e.g. for ht CD4 nt 361 is the key nt change that differentiates CD4 from ht CD1)

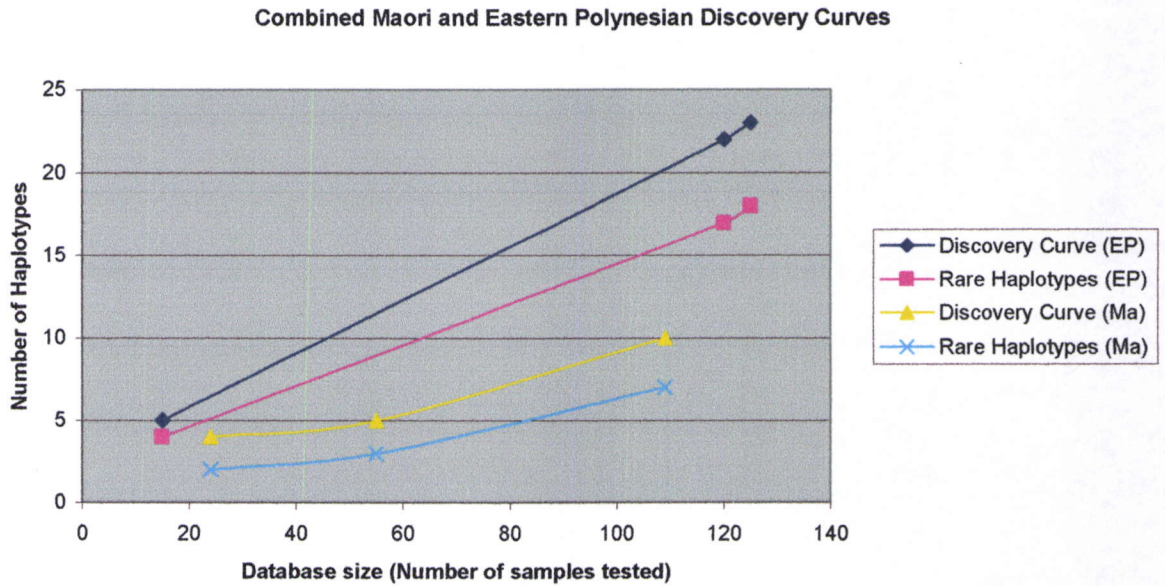
Nucleotides which characterise a ht because they are missing are shown by a dash: i.e. ____ (e.g. for ht CD3 nt 261 is missing, and it is this nt change that differentiates CD3 from CD1)

Appendix Six: Total number of Haplotypes found in combined MM98 and A W02 data combining haplotypes that differ before position 16 189.

Group I	MM (ref)		AW (ref)	
	Ma	EP	Ma	EP
CD1	47 (1)	69 (1)	45 (1,10)	3 (1)
CD2	5 (15)	12 (15)	4 (2)	1 (2)
CD3	1 (11)	1 (11)	1 (3)	
CD7		1 (12)	1 (6)	
CD8		1 (13)		
CD12		1 (19)		
CD17			1 (5)	
CD18			1 (4)	
CD19			1 (7)	
CD20				1 (9)
Group II				
CD22		1 (23)		
CD22a		3 (24)		
CD23a	1 (25)	16 (25)		
Group III				
CD27		2 (37)		
CD28		1 (38)		
Totals (people)	54	108	54	5
(haplotypes)	4	11	8	3

The haplotype numbering on this table follows that used in Appendix Five. This table is compiled treating the data in the same manner as employed in the study by Murray-McIntosh et al 1998. In the MM98 column there are 11 EP haplotypes as was found in Table One of their study. The main problem with the MM98 study is the failure to use the Group III EP sequences on their table, as this underestimates the number of haplotypes in the EP population. Also, the MM treatment of data includes sequences that have differences in their sequence in the 5' HVI region (before nucleotide position 16 189) into the same haplotype category. Instead of acknowledging that some of these sequences are different they were lumped together in MM98 as the sequences were identical from 16 189 onwards. This concern greater for the WP and MN data.

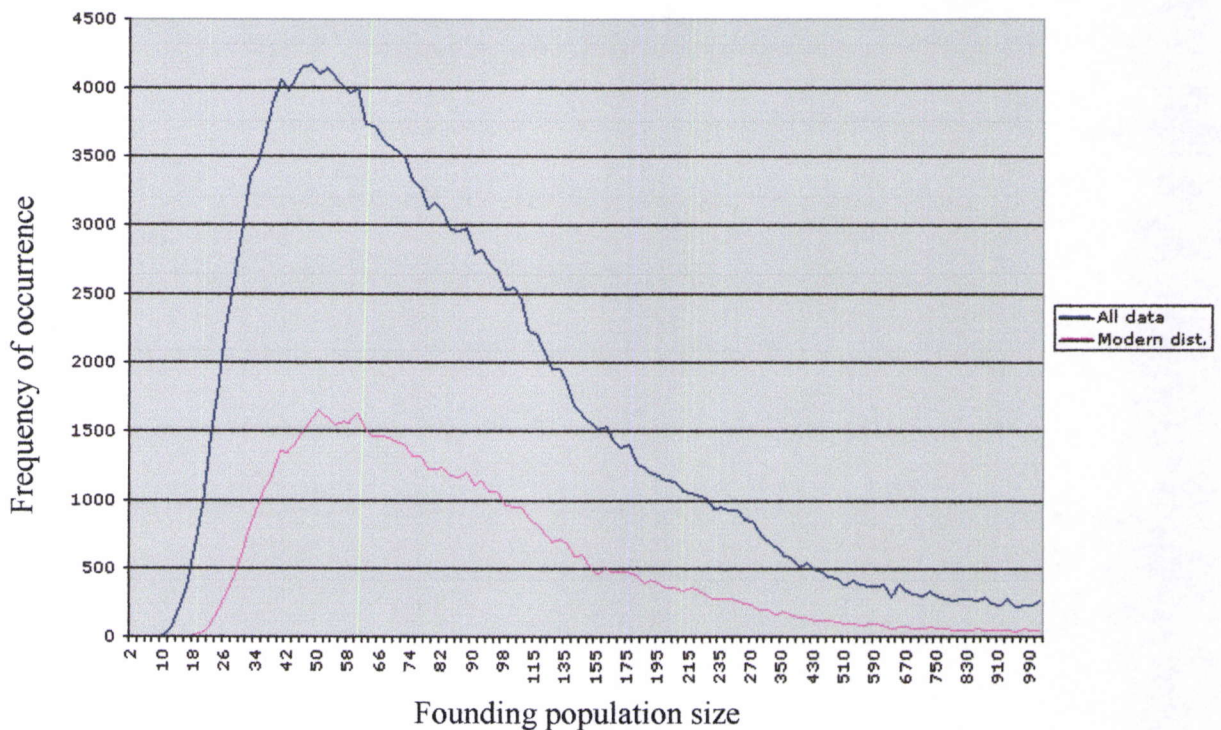
Appendix Seven: Discovery curves for the Maori and Eastern Polynesian data



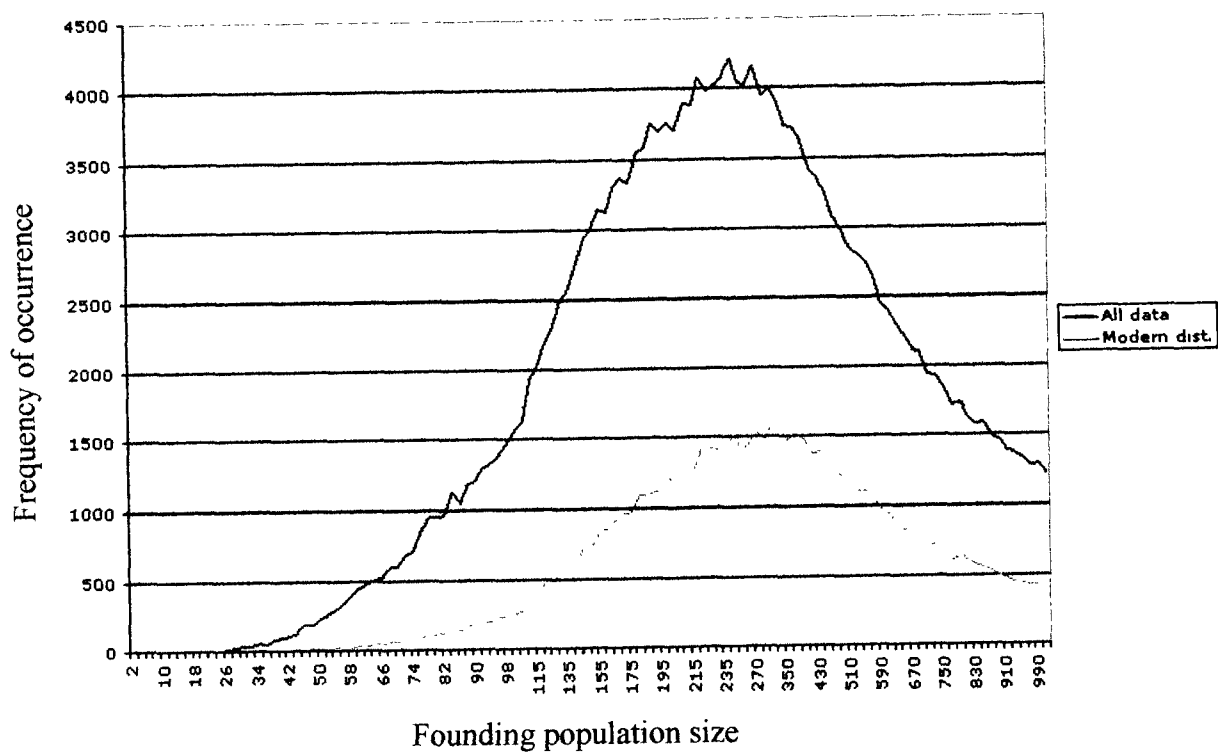
Appendix Eight: Comparison of total data (blue) with ‘unequal’ distribution of haplotypes seen in the modern data (pink).

NB The peaks stay in the same place, but the number of founders required increases slightly.

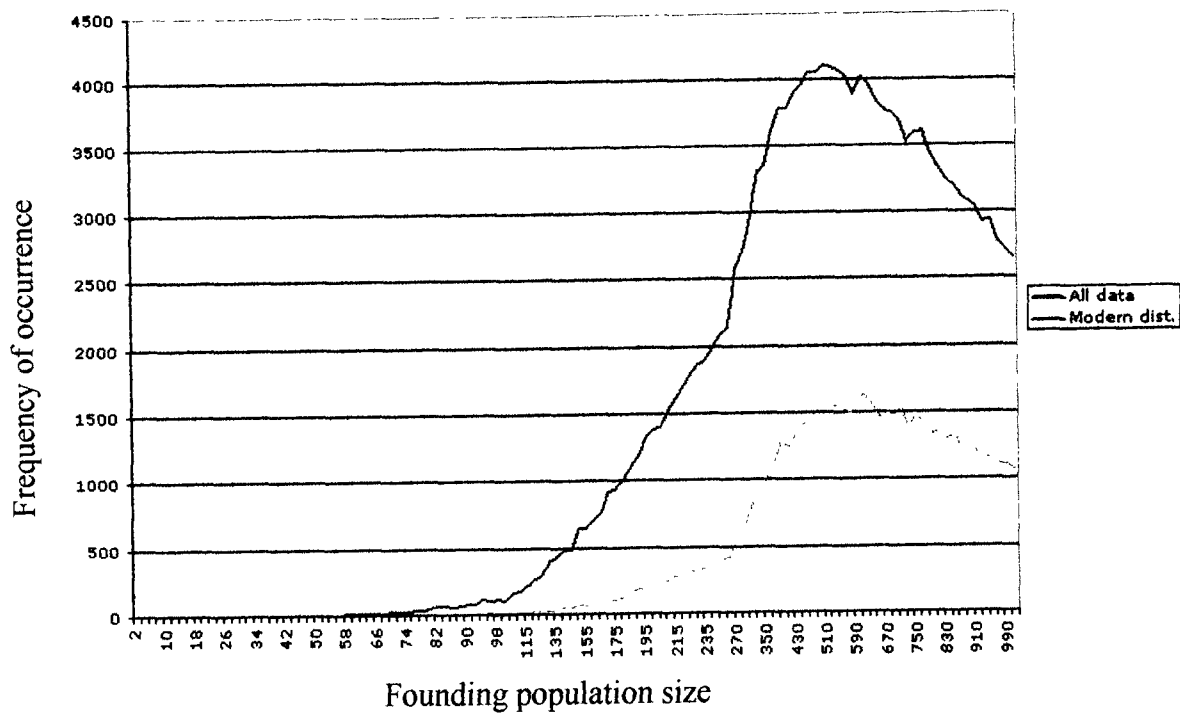
i. Early population expansion



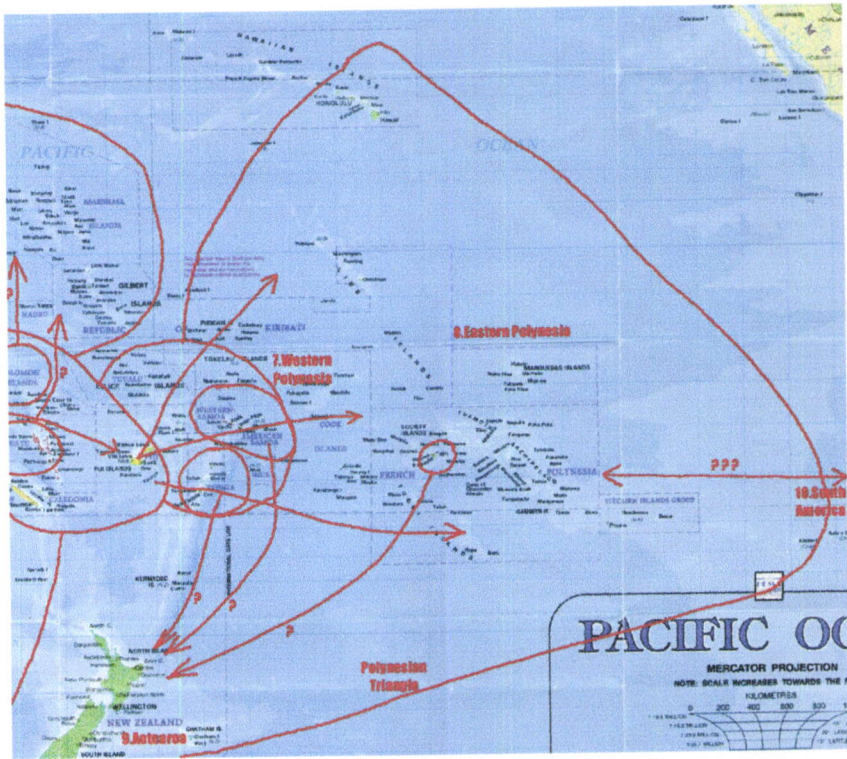
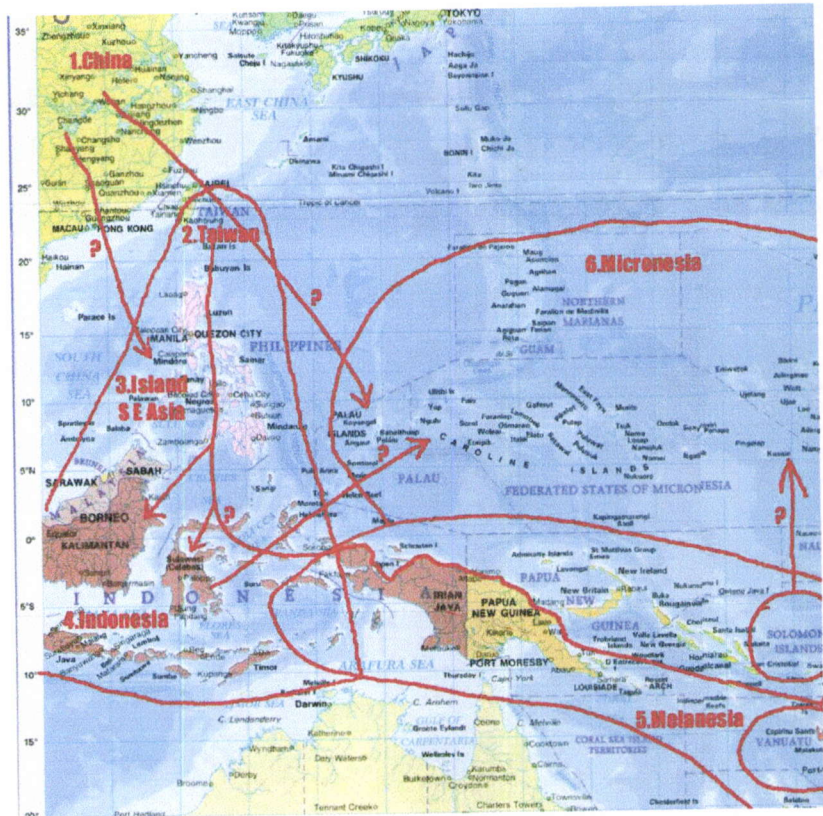
ii. Middle population expansion

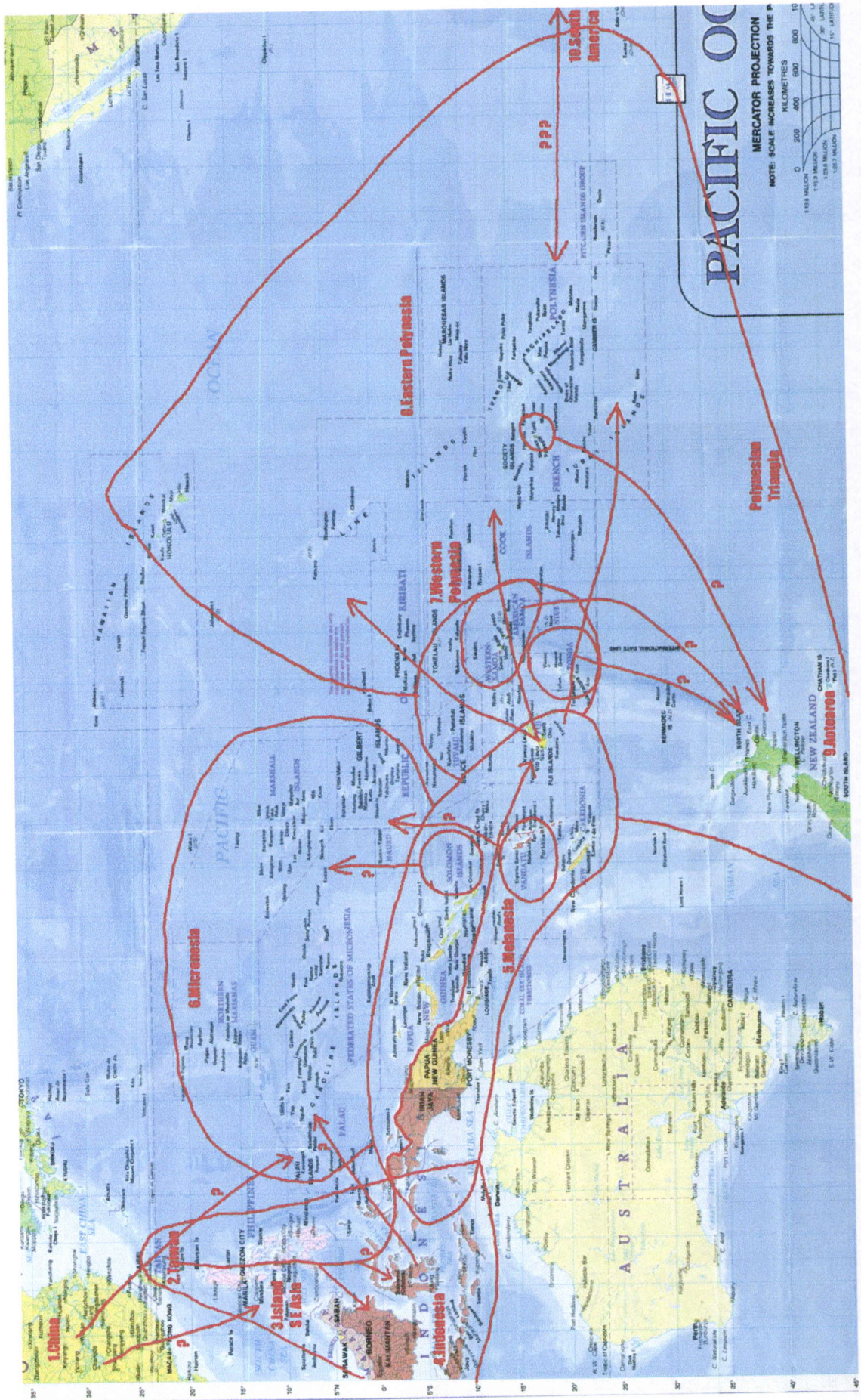


iii. Late population expansion



Appendix Nine: Maps to illustrate the ‘Synthetic Total Evidence Theory’





References

- Adcock GJ, Dennis ES, Eastaie S, Huttley GA, Jermin LS, Peacock WJ, Thorne A (2001) Mitochondrial DNA sequences in ancient Australians: Implications for modern human origins. *Proc. Natl. Acad. Sci. USA* 98:537-542
- Anderson A (1991) Race against time: The early Maori-Pakeha families and the development of the mixed-race population in southern New Zealand. University of Otago, Dunedin
- Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon I C, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young I G (1981) Sequence and organization of the human mitochondrial genome. *Nature* 290:457-464
- Ankel-Simons F, Cummins JM (1996) Misconceptions about mitochondria and mammalian fertilization: Implications for theories on human evolution. *Proc. Natl. Acad. Sci. USA* 93:13859-13863
- Anslinger K, Weichhold G, Keil W, Bayer B, Eisenmenger W (2001) Identification of the skeletal remains of Martin Bormann by mtDNA analysis. *Int. J. Legal Med.* 114:194-196
- Ashkenan J (1997) Homologous Recombination in Human Mitochondria? *Am. J. Hum. Genet.* 61:18
- Awadalla P, Eyre-Walker A, Smith J M (1999) Linkage Disequilibrium and Recombination in Hominid Mitochondrial DNA. *Science* 286:2524-2525
- Barrientos A, Casademont J, Solans A, Moral P, Cardellach F, Urbano-Marquez A, Estivill X, Nunes V (1995) The 9-bp deletion in region V of mitochondrial DNA: evidence of mutation recurrence. *Hum. Genet.* 96:225-228
- Bellwood P (1991) The Austronesian Dispersal and the origin of languages. *Sci. Am.* 265:70-75
- Bellwood P (1997) The Austronesian Dispersal. *Newsletter of Chinese Ethnology*:1-25
- Bellwood P (1998) Human Dispersals and colonizations in prehistory- the Southeast Asian data and their implications. In: Omoto K, Todias P V (ed) *The Origins and Past of Modern Humans-Towards Reconciliation*. World Scientific, Singapore, pp 199-205
- Bendall KE, Sykes BC (1995) Length Heteroplasmy in the First Hypervariable Segment of the Human mtDNA Control Region. *Am. J. Hum. Genet.* 57:248-256
- Bendall KE, Macaulay VA, Baker JR, Sykes BC (1996) Heteroplasmic Point Mutations in the Human mtDNA Control Region. *Am. J. Hum. Genet.* 59:1276-1287

- Bonatto S, Salzano F (1997) Diversity and Age of the Four Major mtDNA Haplogroups, and Their Implications for the Peopling of the New World. *Am. J. Hum. Genet.* 61:1413-1423
- Brewis AA, Molloy MA, Sutton DG (1990) Modeling the Prehistoric Maori Population. *Am. J. Phys. Anthro.* 81:343-356
- Burnham KP, Overton WS (1979) Robust estimation of population size when capture probabilities vary among animals. *Ecology* 60:927-936
- Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. *Nature* 325:31-36
- Capelli C, Wilson JF, Richards M, Stumpf MP, Gratrix F, Oppenheimer S, Underhill P, Pascali VL, Ko TM Goldstein DB (2001) A Predominantly Indigenous Paternal Heritage for the Austronesian-Speaking Peoples of Insular Southeast Asia and Oceania. *Am. J. Hum. Genet.* 68: 432-443
- Caspari R, Wolpoff M (1997) *Race and Human Evolution*. Simon and Schuster, New York.
- Cavalli-Sforza LL (2000) *Genes People and Languages*. North Point Press, New York
- Chambers GK, Marshall SJ, Robinson GM, Maguire S, Newton-Howes J, Chong N (2002) The Genetics of Alcoholism in Polynesians: Alcohol and Aldehyde Dehydrogenase genotypes in young men. In Press.
- Chen YS, Torroni A, Excoffier L, Santachiara-Benerecetti AS, Wallace, DC (1995) Analysis of mtDNA Variation in African Populations Reveals the Most Ancient of all Human Continent-Specific Haplogroups. *Am. J. Hum. Genet.* 57:133-149
- Chu JY, Huang W, Kuang SQ, Wang JM, Xu JJ, Chu ZT, Yang ZQ, Lin KQ, Li P, Wu M, Geng ZC, Tan CC, Du RF, Jin L (1998) Genetic relationship of populations in China. *Proc. Natl. Acad. Sci. U S A* 95: 11763-8.
- Comas D, Paabo S, Bertranpetit J (1995) Heteroplasmy in the Control Region of Human Mitochondrial DNA. *Genome Research* 5:89-90
- Davidson JM (1984) *The Prehistory of New Zealand*. Auckland: Longman Paul.
- Devlin B, Roeder, K, Otto C, Tiobech S Byerley W (2001) Genome-wide distribution of linkage disequilibrium in the population of Palau and its implications for gene flow in Remote Oceania. *Hum. Genet.* 108:521-528
- Di Rienzo A, Wilson AC (1991) Branching in the evolutionary tree for human mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* 88:1597-1601
- Diamond J (1988) Express Train to Polynesia. *Nature* 336:307-308
- Diamond J (2000) Taiwan's gift to the world. *Nature* 403:709-710

- Dorit RL, Akashi H, Gilbert W (1995) Absence of polymorphism at the ZFY locus on the human Y chromosome. *Science* 268: 1183-1185
- Dumont d'Urville MJ (1832) Notice sure les Tles du Grand Océan et sur l'orgine des peuples qui les habitent. *Société de Géographie Bulletin*. 17:1-21
- Evans J (1997) *Nga Waka O Nehera; the first voyaging canoes*. Reed Books; a division of Reed Publishing, Birkenhead, Auckland
- Evans J (1998) *The Discovery of Aotearoa*. Reed Books; Division of Reed Publishing (NZ), Birkenhead, Auckland
- Eyre-Walker A (2000) Do mitochondria recombine in humans? *Phil. Trans. R. Soc. Lond. B* 355:1573-1580
- Eyre-Walker A, Smith NH, Smith JM (1999) How clonal are human mitochondria? *Proc. R. Soc. Lond. B*. 266:477-483
- Flint J, Hill AVS, Bowden DK, Oppenheimer S J, Sill PI R., Sergeantson, S. W., Bana-Koiri, J., Bhatia, K., Alpers, M. P., Boyce, A. J., Weatherall, D. J., and Clegg, J. B. (1986) High frequencies of alpha-thalassaemia are the result of natural selection by malaria. *Nature* 321:744-750
- Flint J, Boyce AJ, Martinson JJ Clegg JB (1989) Population bottlenecks in Polynesia revealed by minisatellites. *Hum. Genet.* 83:257-263
- Francalacci P, Montiel R, Malgosa A (1999) A mitochondrial DNA database. In: Papiha S, Deka, R. and Chakraborty, R. (ed) *Genomic Diversity :Applications in Human Population Genetics*. Vol Chapter 9. Kluwer Academic/Plenum Publishing, New York, pp 103-119
- Gibbons A (2001) The Peopling of the Pacific. *Science* 291:1735-1737
- Gill P, Ivanov PL, Kimption C, Piercy R, Benson N, Tully G, Evett I, Hagelberg E, Sullivan K (1994) Identification of the remains of the Romanov family by DNA analysis. *Nature Genetics* 6:130-135
- Gray R Jordon F (2000) Language trees support the express-train sequence of Austronesian expansion. *Nature* 405:1052-1055
- Green R C (1991) Near and Remote Oceania: Disestablishing "Melanesia" in culture history. In A. Pawley, ed., *Man and a Half: Essays in Pacific Anthropology and Ethnobiology in Honour of Ralph Bulmer* 491-502
- Greenberg BD, Newbold JE, Sugino A (1983) Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. *Gene* 21:33-49
- Hagelberg E (1997) Ancient and modern mitochondrial DNA sequences and the colonization of the Pacific. *Electrophoresis* 18:1529-1533

- Hagelberg E, Clegg JB (1993) Genetic polymorphisms in prehistoric Pacific islanders determined by analysis of ancient bone DNA. *Proc. R. Soc. Lond. B.* 252:163-170
- Hagelberg E, Goldman N, Lio P, Whelan S, Schiefenhover W, Clegg JB, Bowden DK (1999a) Evidence for mitochondrial DNA recombination in a human population of island Melanesia. *Proc. R. Soc. Lond. B.* 266:485-492
- Hagelberg E, Kayser M, Nagy M, Roewer L, Zimdahl H, Krawczak M, Lio P Schiefenhover W (1999b) Molecular genetic evidence for the human settlement of the Pacific: analysis of mitochondrial DNA, Y chromosome and HLA markers. *Phil. Trans. R. Soc. Lond. B* 354:141-152
- Hamilton F (1994) Multi-Locus and Single-Locus DNA Profiling in New Zealand. Unpublished Ph.D. thesis. Victoria University, Wellington
- Hammer M, Horai S (1995) Y chromosomal DNA Variation and the peopling of Japan. *Am. J. Hum. Genet.* 56:951-962
- Helgason A, Hickey E, Goodacre S, Bosnes V, Stefansson K, Ward R, Sykes B (2001) mtDNA and the Islands of the North Atlantic: Estimating the Proportions of Norse and Gaelic Ancestry. *Am. J. Hum. Genet.* 68:723-737
- Hertzberg M, Mickleson KN, Serjeantson SW, Prior JF, Trent RJ (1989) An Asian-specific 9-bp deletion of mitochondrial DNA is frequently found in polynesians. *Am. J. Hum. Genet.* 44:504-510
- Heyerdahl T (1950) *The Kon-Tiki Expedition: By raft across the south seas.* Allen and Unwin, London
- Hill AVS, Serjeantson SW (1989) *The Colonization of the Pacific: A Genetic Trail.* Clarendon Press, Oxford
- Hill E, Jobling MA, Bradley DG (2000) Y-chromosome variation and Irish origins. *Nature* 404:351-352
- Hiroa TR (1938) *Vikings of the Sunrise.* New York: Frederick Stokes.
- Holland MM, Parsons TJ (1999) Mitochondrial DNA Sequence Analysis-Validation and use for forensic casework. *Forensic Sci Rev* 11:21-50
- Horai S, Kondo R, Murayama K, Hayashi S, Koike H, Nakai N (1991) Phylogenetic affiliation of ancient and contemporary humans inferred from mitochondrial DNA. *Phil. Trans. R. Soc. Lond. B* 333:409-417
- Horai S, Kondo R, Nakagawa-Hattori Y, Hayashi S, Sonoda S, Tajima K (1993) Peopling of the Americas, Founded by Four Major Lineages of Mitochondrial DNA. *Mol. Bio. Evol.* 10:23-47

- Horai S, Hayasaka K, Kondo R, Tsugane K, Takahata N (1995) Recent African Origin of Modern Humans Revealed by Complete Sequences of Hominoid Mitochondrial DNAs. *Proc. Natl. Acad. Sci. USA* 92: 532 - 536.
- Howell N (1997) mtDNA Recombination: What Do In Vitro Data Mean? *Am. J. Hum. Genet.* 61:19-22
- Hurles M, Irven C, Nicholson J, Taylor PG, Santos FR, Loughlin J, Jobling MA, Sykes BC (1998) European Y-Chromosomal Lineages in Polynesians: A Contrast to the Population Structure Revealed by mtDNA. *Am. J. Hum. Genet.* 63:1793-1806
- Hurles ME, Nicholson J, Bosh E, Renfrew C, Sykes B, Jobling MA (2002) Y Chromosomal Evidence for the origins of Oceanic-Speaking peoples. *Genetics* 160: 289-303
- Ingman M, Kaessmann H, Paabo S, Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. *Nature* 408:708-713
- Intoh M (2001) In: Foregone Conclusions? In search of "Papuan" and "Austronesians" by Terrell JE, Kelly KM, Rainbird P. *Current Anthropology* 42: 97-124
- Ivanov PL, Wadhams MJ, Roby RK, Holland MM, Weedn VW, Parsons TJ (1996) Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II. *Nat Genet.* 12(4): 417-20.
- Jazin E, Soodyall H, Jalonon P, Lindholm E, Stoneking M, Gyllensten U. (1998) Mitochondrial mutation rate revisited: hot spots and polymorphism. *Nature Genetics* 18:109-110
- Jorde L, Bamshad M (2000) Questioning evidence for Recombination in Human mtDNA. *Science* 288:1931
- Karafet TM, Zegura SL, Posukh O, Bergen A, Long J, Goldman D, Klitz W, Harihara S, de Knijff P, Wiebe V, Griffiths RC, Templeton AR, Hammer MF (1999) Ancestral Asian source(s) of new world Y-chromosome founder haplotypes. *Am. J. Hum. Genet.* 64:817-831
- Kayser M, Brauer S, Weiss G, Underhill PA, Roewer L, Schiefenhover W, Stoneking M (2000) Melanesian origin of Polynesian Y chromosomes. *Current Biology* 10: 1237-1246
- Kayser M, Brauer S, Weiss G, Schiefenhover W, Underhill PA, Stoneking M (2001) Independent Histories of Human Y Chromosomes from Melanesia and Australia. *Am. J. Hum. Genet.* 68:173-190
- Kirch PV (1997) *The Lapita Peoples: Ancestors of the Oceanic World*. Oxford: Blackworld

- Kirch PV (2000) On the road of the winds: and archaeological history of the Pacific Islands before European contact. University of California Press, Berkley and Los Angeles, California
- Kivisild T, Villems R (2000) Questioning evidence for Recombination in Human mtDNA. *Science* 288:1931
- Krings M, Geisert H, Schmitz RW, Krainitzki H, Paabo S (1999) DNA sequence of the mitochondrial hypervariable region II from the Neandertal type specimen. *Proc. Natl. Acad. Sci. USA* 96:5581-5585
- Krings M, Stone A, Schmitz RW, Krainitzki H, Stoneking M, Paabo S (1997) Neandertal DNA Sequences and the Origin of Modern Humans. *Cell* 90:19-30
- Kumar S, Hedrick P, Dowling T, Stoneking M (2000) Questioning evidence for Recombination in Human mtDNA. *Science* 288:1931
- Lum JK, Rickards O, Ching C, Cann R (1994) Polynesian Mitochondrial DNAs Reveal Three Deep Maternal Lineage Clusters. *Hum. Biol.* 66:567-590
- Lum JK, Cann RL, Martinson JJ, Jorde LB (1998) Mitochondrial and Nuclear Genetic Relationships among Pacific Island and Asian Populations. *Am. J. Hum. Genet.* 63: 613-624
- Lutz S, Weisser HJ, Heizmann J, Pollak S (1996) mtDNA as a tool for identification of human remains: Identification using mtDNA. *Int. J. Legal Med.* 109:205-209
- Macaulay V, Richards M Sykes B (1999a) Mitochondria DNA recombination- no need to panic. *Proc. R. Soc. Lond. B.* 266:2037-2039
- Macaulay V, Richards M, Hickey E, Vega E, Cruciani F, Guida V, Scozzari R, Bonne-Tamir B, Torroni A (1999b) The emerging tree of West Eurasian mtDNAs: A synthesis of control region sequences and RFLPs. *Am. J. Hum. Genet.* 64:232-249
- Marck J (1999) Polynesian Language and Culture History. Unpublished Ph.D. thesis. Australian National University (Canberra)
- McFadgen BG, Knox FB, Cole TR (1994) Radiocarbon calibration curve variations and their implications for the interpretation of New Zealand prehistory. *Radiocarbon* 36:221-236
- McGlone MS, Anderson AJ, Holdaway RN (1994) The Origins of the First New Zealanders, ed. Sutton DG. Auckland Univ. Press, Auckland.
- McKie R, Stringer C (1996) African exodus: The Origins of Modern Humanity. Henry Hold and Company: New York.
- Meacham W (1984-85) On the improbability of Austronesian origins in South China. *Asian Perspectives* XXVI: 89-106

- Melton T, Peterson R, Redd AJ, Saha N, Sofro ASM, Martinson J, Stoneking M (1995) Polynesian genetic affinities with Southeast Asian populations as identified by mtDNA analysis. *Am. J. Hum. Genet.* 57:403-414
- Melton T, Clifford S, Martinson J, Batzer M, Stoneking M (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
- Murray-McIntosh RP, Scrimshaw BJ, Hatfield PJ, Penny D (1998) Testing migration patterns and estimating founding population size in Polynesia using human mtDNA sequences. *Proc. Natl. Acad. Sci. USA* 95: 9047-9052
- Oppenheimer S, Richards M (2001a) Fast trains, slow boats, and the ancestry of the Polynesian islanders. *Science Progress* 84:157-181
- Oppenheimer S, Richards M (2001b) Slow Boat to Melanesia? *Nature* 410:166-167
- Ovchinnikov IV, Gotherstrom A, Romanova GP, Kharitonov VM, Liden K, Goodwin W (2000) Molecular analysis of Neanderthal DNA from the northern Caucasus. *Nature* 404:490-493
- Papiha S, Mastana S (1999) Classical to Molecular Polymorphisms: Population Genetic Studies from the Indian Sub-Continent. In: Papiha S, Deka R, Chakraborty R (ed) *Genetic Diversity: Applications in Human Population Genetics*. Vol Chapter 1. Kluwer Academic/ Plenum Publishers, New York
- Parsons T, Muniec D, Sullivan K, Woodyatt N, Alliston-Greiner R, Wilson M, Berry D, Holland K, Weedn V, Gill P, Holland M (1997) A high observed substitution rate in the human mitochondrial DNA control region. *Nature Genetics* 15: 363-368
- Parsons T, Irwin J (2000) Questioning evidence for recombination in human mitochondrial DNA. *Science* 288:1931
- Passarino G, Semino O, Modiano G, Santa-Chiara-Benerecetti AS (1993) COII/tRNA^{Lys} intergenic 9-bp deletion and other mtDNA markers clearly reveal that the Tharus (southern Nepal) have Oriental affinities. *Am. J. Hum. Genet.* 53: 609-618
- Pawley A (1996) On the Polynesian subgroup as a problem for Irwin's continuous settlement hypothesis. In Davidson J, Irwin F, Leach F, Pawley A, Brown D eds, *Oceanic Culture History: Essays in Honour of Roger Green*. New Zealand Journal of Archaeology Special Publication. Dunedin.
- Pennisi E (2001) Tracking the Sexes by their Genes. *Science* 291:1733-1734
- Perez-Lezaun A, Calafell F, Comas D, Mateu E, Bosch E, Martinez-Arias R, Clarimon J, Fiori G, Luiselli D, Facchini F, Pettener D, Beruranpetit J (1999) Sex-specific migration patterns in central Asian populations, revealed by analysis of Y-chromosome short tandem repeats and mtDNA. *Am. J. Hum. Genet.* 65:208-219

- Piercy R, Sullivan KM, Benson N, Gill P (1993) The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. *Int. J. Legal Med.* 106:85-90
- Pietrusewsky M (1996) The physical anthropology of Polynesia: A review of some cranial and skeletal studies. In Davidson, J., Irwin, F., F. Leach, Pawley, A. and Brown, D. eds, *Oceanic Culture History: Essays in Honour of Roger Green*. New Zealand Journal of Archaeology Special Publication. Dunedin.
- Redd AJ, Takezake N, Sherry ST, McGarvey ST, Sofro ASM, Stoneking M (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
- Richards M, Oppenheimer S, Sykes B (1998) mtDNA suggests Polynesian origins in Eastern Indonesia. *Am. J. Hum. Genet.* 63:1234-1236
- Seielstad MT (2000) Asymmetries in the Maternal and Paternal Genetic Histories of Colombian Populations. *Am. J. Hum. Genet.* 67:1062-1066
- Seielstad MT, Minch E, Cavalli-Sforza LL (1998) Genetic evidence for a higher female migration rate in humans. *Nature Genetics* 20:278-280
- Sharp A (1956) *Ancient Voyagers in the Pacific*. Wellington Polynesian Society
- Shields GF, Hecker K, Voevoda MI, Reed JK (1992) Absence of the Asian-specific region V mitochondrial marker in Native Beringians. *Am. J. Hum. Genet.* 50:758-765
- Solheim WG, (1996) The Nusantara and north-south dispersals. *Bulletin of the Indo-Pacific Prehistory Association* 15. The Chiang Mai Papers 2: 101-109
- Soodyall H, Vigilant L, Hill A, Stoneking M, Jenkins T (1996) mtDNA Control-Region Sequence Variation Suggests Multiple Independent Origins of an "Asian-Specific" 9-bp Deletion in Sub-Saharan Africans. *Am. J. Hum. Genet.* 58:595-608
- Sorrenson MPK (1979) *Maori origins and migrations: the genesis of some pakeha myths and legends*. Auckland University Press
- Stoneking M (2000) Hypervariable Sites in the mtDNA Control Region Are Mutation Hotspots. *Am. J. Hum. Genet.* 67:1029-1032
- Su B, Jin L, Underhill P, Martinson J, Saha N, McGarvey ST, Shriver MD, Chu J, Oefner P, Chakraborty R, Deka R (2000) Polynesian origins: insights from the Y chromosome. *Proc. Natl. Acad. Sci. USA* 97:8225-8228
- Sykes B, Leiboff A, Low-Beer J, Tetzner S, Richards M (1995) The origins of the Polynesians: An interpretation from mitochondrial lineage analysis. *Am. J. Hum. Genet.* 57:1463-1475

- Templeton A (2002) Out of Africa again and again. *Nature* 416:45-51
- Terrell J (1988) History as a family tree, history as an entangled bank: constructing images and interpretations of prehistory in the South Pacific. *Antiquity* 62:642-657
- Terrell JE, Kelly KM, Rainbird P (2001) Foregone Conclusions? In search of "Papuan" and "Austronesians". *Current Anthropology* 42: 97-124
- Thomson R, Pritchard JK, Shen P, Oefner PJ, Feldman MW (2000) Recent common ancestry of human Y chromosomes: Evidence from DNA sequence data. *Proc. Natl. Acad. Sci. USA* 97:7360-7365
- Torroni A, Bandelt HJ, D'Urbano L, Lahermo P, Moral P, Sellitto D, Rengo C, Forster P, Savontaus ML, Bonne-Tamir B, Scozzari R (1998) mtDNA Analysis Reveals a Major Late Paleolithic Population Expansion from Southwestern to Northeastern Europe. *Am. J. Hum. Genet.* 62:1137-1152
- Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, Vullo C M, Wallace DC (1993a) Asian Affinities and Continental Radiation of the Four Founding Native American mtDNAs. *Am. J. Hum. Genet.* 53:563-590
- Torroni A, Sukernik RI, Schurr TG, Starikovskaya YB, Cabell MF, Crawford MH, Comuzzie AG, Wallace D (1993b) mtDNA variation of Aboriginal Siberians reveals distinct genetic affinities with Native Americans. *Am. J. Hum. Genet.* 53:591-608
- Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T, Vollrath D, Davis RW, Cavalli-Sforza LL, Oefner PJ (1997) Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Research* 7: 996-1005
- Underhill PA, Jin L, Zemans R, Oefner PJ, Cavalli-Sforza LL (1996) A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. *Proc. Natl. Acad. Sci. USA* 93:196-200
- Underhill PA, Passrino G, Lin AA, Marzuki S, Cavalli-Sforza LL and Chambers, GK (2001a) Maori Origins, Y-Chromosome Haplotypes and implications for human history in the Pacific. *Human Mutation* 17:271-280
- Underhill PA, Passrino G, Lin A, Shen P, Lahr MM, Foley R, Oefner P, Cavalli-Sforza LL (2001b) The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann. Hum. Genet.* 65:43-62
- Wall JD (2000) Detecting Ancient Admixture in Humans Using Sequence Polymorphism Data. *Genetics* 154: 1271-1279
- Ward R, Frazier B, Dew-Jager K, Paabo S (1991) Extensive mitochondrial diversity within a single Ameridian tribe. *Proc. Natl. Acad. Sci. USA* 88:8720-8724

- Ward R, Redd A, Velencia D, Frazier B, Paabo S (1993) Genetic and linguistic differentiation in the Americas. *Proc. Natl. Acad. Sci. USA* 90:10663-105667
- Wrischinik LA, Higuchi, RG, Stoneking M, Erlich HA, Arnheim N, Wilson AC (1987) Length mutations in human mitochondrial DNA: direct sequencing of enzymatically amplified DNA. *Nucleic Acid Res.* 15:529-541