

ASSESSING Y-CHROMOSOME VARIATION IN THE SOUTH PACIFIC USING NEWLY DETECTED
NRY MARKERS

A Dissertation
Submitted to
the Temple University Graduate Board

in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

by
Krista Erin Latham
August, 2008

©
By
Krista Erin Latham
2008
All Rights Reserved

ABSTRACT

ASSESSING Y-CHROMOSOME VARIATION IN THE SOUTH PACIFIC USING NEWLY DETECTED NRY MARKERS

Krista Erin Latham

Doctor of Philosophy

Temple University, 2008

Jonathan Friedlaender, PhD

The South Pacific is a region of incredible biological, cultural and linguistic diversity, reflecting its early settlement by human populations. It has been a region of interest to scholars because of this diversity, as well as its unique geography and settlement history. Current evidence suggests there was an initial settlement of Near Oceania during the Pleistocene by Papuan-speaking foragers, followed by a later Holocene settlement of Remote Oceania by Oceanic-speaking agriculturalists. Previous studies of human biological variation have been used to illuminate the migration history of and population relationships within Oceania.

In this study, I analyzed Y-chromosome (NRY) diversity in 842 unrelated males to more fully characterize the phylogeography of paternal genetic lineages in this region, using a large number of regionally informative markers on an intensive sample set from Northern Island Melanesia. This approach facilitated an analysis of NRY haplogroup distributions, an evaluation of the ancestral paternal genetic contribution to the region, and a comparison of regional NRY diversity with that

observed at different genetic loci (e.g., mtDNA). This project is part of a collaborative effort by faculty and graduate students from the Temple University Department of Anthropology that focused on characterizing biological variation and genetic structure in Melanesia, and better resolving the phylogeographic specificity of Northern Island Melanesia.

Overall, this study generated a higher resolution view of NRY haplogroup variation than detected in previous studies through the use of newly defined and very informative SNP markers. It also showed that there is a very small ancestral East Asian paternal contribution to this area, and a rather large proportion of older Melanesian NRY lineages present there. In addition, this study observed extraordinary NRY diversity within Northern Island Melanesia, as well as genetic structure influenced more by geography than linguistic variation. This structure and diversity was essentially equivalent to that noted for mtDNA data for this region. Finally, this study helped to resolve questions about the placement of the 50f2/c deletion within the larger NRY tree. Overall, this work has refined our understanding of the migration and demographic history of Northern Island Melanesia.

ACKNOWLEDGMENTS

The analyses and presentation of this research would not have been possible without the help and support of numerous individuals. I would like to thank the Friedlaender Family Fund and the Temple University Dissertation Completion Grant for funding this study. I would like to thank my advisor, Dr. Jonathan Friedlaender, for his patience and support throughout this journey, as well as my committee members: Dr. Joseph Lorenz, Dr. Charles Weitz, Dr. Leonard Greenfield and Dr. Theodore Schurr. I am also indebted to Françoise Friedlaender and Dr. Christie Rockwell for their intellectual support, and the entire Temple University Anthropology Department for encouragement during my time at Temple University. I would also like to thank the graduate students in the anthropology program, especially Danielle James and Elizabeth Rowe, for their friendship and support. This study could not have been completed without the foundation of work provided by Dr. Laura Scheinfeldt. I would especially like to thank my family and friends, who are too numerous to list, for the love and support they give unconditionally, especially Theresa Luckino, Audra Simpson, and Lukas and Jeremy Beach. There are numerous other individuals that need to be acknowledged for contributing in some way to this project. Know that I am forever grateful.

DEDICATION

To Lukas Darwin Beach,
for making me a better person.

TABLE OF CONTENTS

	Page
ABSTRACT.....	iii
ACKNOWLEDGMENTS.....	v
DEDICATION.....	vi
LIST OF TABLES.....	x
LIST OF FIGURES.....	xi
	CHAPTER
1. INTRODUCTION.....	1
Geographical Terminology.....	3
The Y-Chromosome.....	4
The Use of NRY and mtDNA in Population Genetic Studies.....	5
Worldwide Distribution of Genetic Diversity.....	7
The Y-Chromosome.....	8
The mtDNA.....	9
A Synthesis of Worldwide Genetic Diversity in the Y-Chromosome and mtDNA.....	10
Research Goals.....	11
2. ARCHEOLOGICAL AND LINGUISTIC VARIATION IN THE SOUTH PACIFIC.....	15
Archeological Patterns in Oceania.....	15
Archeological Data from the Pleistocene Epoch.....	15
Archeological Data from the Pleistocene-Holocene Transition.....	18
Archeological Data from the Early Holocene Epoch.....	19
The Lapita Dispersal.....	20
Fossil and Morphological Patterns in Oceania.....	22
Fossil Data.....	22

Linguistic Patterns in Oceania.....	23
Methods of Linguistic Analysis.....	23
The Austronesian Language Group.....	26
The Papuan Language Group.....	28
Summary of Linguistic Patterns in Oceania.....	31
3. BIOLOGICAL VARIATION IN THE SOUTH PACIFIC.....	32
Autosomal DNA Variation in Oceania.....	32
Mitochondrial (mtDNA) Variation in Oceania.....	33
Melanesia.....	33
Polynesia.....	36
Y-Chromosome (NRY) Variation in Oceania.....	37
Melanesia.....	37
Polynesia.....	40
Summary of Genetic Patterns in Oceania.....	41
Skeletal and Dental Patterns in Oceania.....	42
Common Models for the Origins of Remote Pacific Islanders.....	43
The Express Train from Taiwan Models.....	43
The Indigenous Melanesian Origin Model.....	45
The Slow Boat to the Bismarcks Model.....	46
The Triple I Model.....	48
The Entangled Bank Model.....	50
Evaluating the Models for the Origins of Remote Pacific Islanders.....	51
4. MATERIALS AND METHODS.....	53
The Biological Samples.....	53
Isolation of Genomic DNA.....	57
Genetic Analysis.....	57
Restriction Fragment Length Polymorphism Analysis.....	58
Insertion/Deletion Analysis.....	60
Sequence Analysis.....	60

Statistical Analysis.....	60
Measures of Genetic Diversity.....	61
Phylogenetic Tree Construction.....	62
5. RESULTS.....	63
Haplogroup Frequencies.....	63
Geographic and Linguistic Distributions of the Haplogroups.....	69
Macrohaplogroup C.....	69
Macrohaplogroup K.....	70
Haplogroup M.....	71
Haplogroup O.....	71
General Geographic Distributions of the Haplogroups.....	72
Network Analysis.....	73
Population Structure.....	75
AMOVA.....	75
NRY Diversity.....	77
6. DISCUSSION.....	80
Research Objectives.....	84
Haplogroup Distributions.....	84
Measures of Genetic Diversity Within and Between Populations.....	95
Distribution of the 50f2/c Deletion.....	98
Comparison of NRY Diversity with mtDNA and Autosomal Diversity.....	99
The Settlement of Polynesia.....	102
7. FUTURE STUDIES.....	107
LITERATURE CITED.....	109

LIST OF TABLES

Table	Page
4.1 Summary of Samples for NRY Analysis by Island and Language.....	56
4.2 NRY Markers and Parameters for Amplification and Analysis.....	59
5.1 Y-Chromosome Haplogroups in 842 Unrelated Males from Oceania.....	65
5.2 Y-Chromosome Haplogroup Frequencies in 842 Unrelated Males from Oceania.....	66
5.3 Distribution of the 50f2/c Deletion in 842 Unrelated Males from Oceania.....	67
5.4 AMOVA Based on Y-Chromosome Haplogroups (F_{ST}).....	75
5.5 Measures of NRY Diversity.....	77

LIST OF FIGURES

Figure	Page
1.1 Major Geographical Divisions of the Pacific Islands.....	4
1.2 Y-Chromosome SNP Phylogenetic Tree.....	14
4.1 Language Families in Northern Island Melanesia.....	54
5.1 Detailed NRY Tree.....	68
5.2 Network of for the Major NRY Haplogroups in this Study.....	74
6.1 MDS Plot Taken from Scheinfeldt et al. (2006, 2007).....	98
6.2 a. YCC Tree 2005, b. YCC Tree 2008.....	106

CHAPTER 1 INTRODUCTION

The South Pacific is a region of incredible biological, cultural and linguistic diversity reflecting the antiquity of its original settlement by humans. It has been a region of interest to scholars because of its diversity and extraordinary settlement history. This study will investigate a particular type of biological variation in the South Pacific, Y-chromosome diversity. This project is part of a collaborative effort by Temple University anthropology department faculty and graduate students focused on understanding the full extent of biological variation and genetic structure in Melanesia, and will expand upon previous genetic analyses in this region in order to better resolve the phylogeographic specificity of Northern Island Melanesia.

Scheinfeldt (2004) conducted an intensive study of Y-chromosome diversity in the Bismarck Archipelago. She found that the Y-chromosome variation was structured and correlated with both linguistic variation and geographic distances, which contrasted with some of the previously published Y-chromosome data (Capelli et al., 2001; Kayser et al., 2000). These discrepancies were attributed to differential sample sizes and demonstrated the importance of increasing sample sizes, not only in number but also geographic breadth, in order to clarify phylogeographic specificity. Additionally, Scheinfeldt (2004) found that the majority of the lineages in her samples belonged to older, indigenous Melanesian patrilineages rather than the younger Asian lineages present in this region. This study will build upon her work by increasing the number of samples utilized and expanding the number of polymorphic markers used to characterize Y-chromosome variation in order to better understand the genetic structure and distribution of paternal lineages in Northern Island Melanesia.

Friedlaender and colleagues (2007a, b) have extensively investigated mitochondrial DNA (mtDNA) diversity throughout Melanesia. They found that the mtDNA diversity was structured by island, island size and language affiliation. They also detected incredible population structure and observed that the majority of the matrilineages in this region were of younger, Asian origin rather than belonging to the older, indigenous lineages that first settled the region. The mtDNA data are often used to compliment the Y-chromosome data in that one reveals the female-mediated genetic history and the other reveals the male-mediated genetic history. This study will attempt to address this and other discrepancies uncovered by these previous studies by increasing sampling coverage to include more individuals utilized in the mtDNA studies. It will build upon and expand these previous studies by expanding the geographic range of sampling and the number of polymorphic markers utilized to better resolve the structure and distribution of Y-chromosome variation in Northern Island Melanesia and further elucidate the phylogeography of paternal lineages in the region.

This study will investigate patterns of Y chromosome diversity in Northern Island Melanesia using a sample of 842 unrelated males. More specifically, it will: 1) use a larger battery of single nucleotide polymorphisms (SNPs) on a previously tested sub-set of samples (Scheinfeldt, 2004), as well as newly collected samples; 2) expand the geographic coverage of the sample; and 3) compare the NRY diversity with the mtDNA data from the region (Friedlaender et al., 2007a).

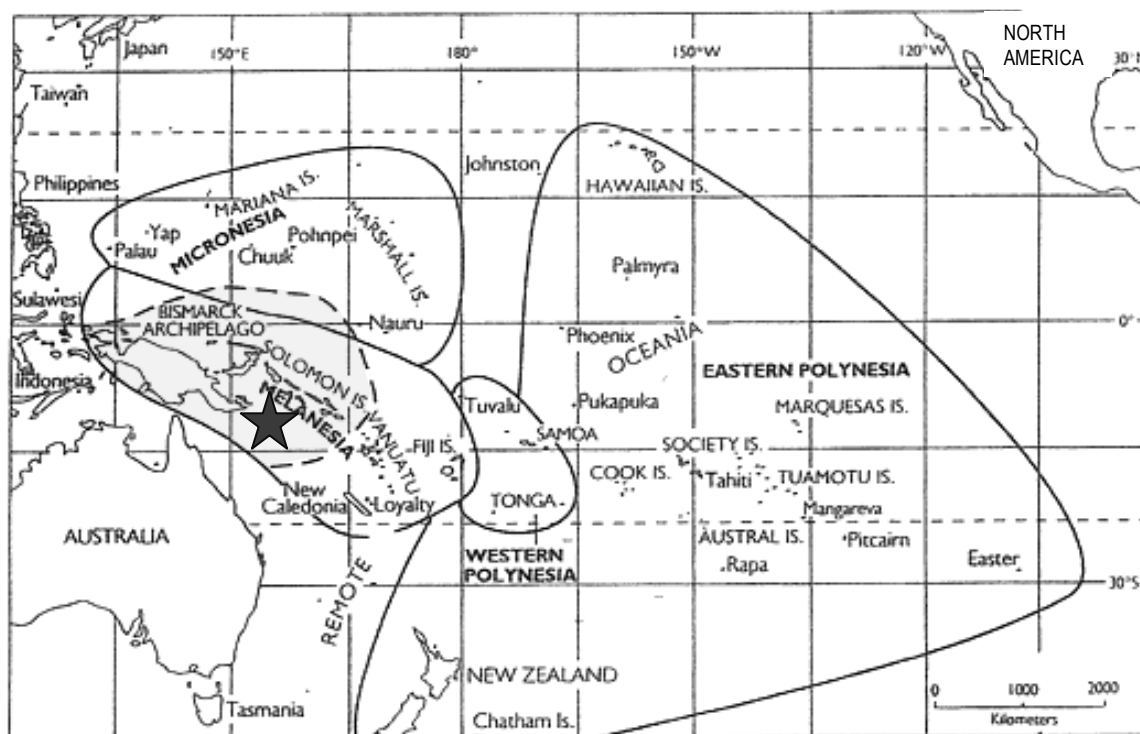
This chapter will introduce the basic terms and concepts that form the background of this study. It will begin by introducing some of the important geographic terms and regions that will be used throughout this manuscript. It will then discuss some general characteristics of the Y-chromosome that make it especially useful for population-based genetic studies, and briefly summarize worldwide

genetic diversity in both the Y-chromosome and the mtDNA. Finally, it will outline the specific research goals of this study.

Geographical Terminology

The Pacific Islands were divided into three broad geographical categories (Fig. 1.1) in 1832 by the French naval commander Durmont d'Urville. Melanesia consists of the islands of New Guinea, the Bismarck Archipelago, the Solomons, Vanuatu, New Caledonia and Fiji. Polynesia encompasses the islands of the central and eastern Pacific as well as New Zealand, and Micronesia consists of the islands of the Pacific to the north of the equator, excluding Hawaii. It is equally important to define several additional geographical terms because this tripartite classification provides only geographic reference points and does not accurately describe cultural histories or biological similarities of the people of the Pacific Islands. To best reflect the anthropological divisions of the region, we can divide it into Near Oceania and Remote Oceania. The Near and Remote Oceania concepts were originally introduced as a means of differentiating the occupational histories of these two areas of the South Pacific (Green, 1991). Near Oceania refers to New Guinea, the Bismarck Archipelago, and the Solomon Islands as far east as San Cristobal and Santa Ana. Remote Oceania encompasses all of the islands to north, east, and southeast of Near Oceania. Northern Island Melanesia consists of all of Near Oceania except New Guinea. (Kirch, 2000; Pawley, 2007). Figure 1.1 shows the major geographical divisions of the South Pacific Islands. Melanesia, Micronesia and Polynesia (Western and Eastern) are distinguished by the solid line boundaries. Near Oceania is bounded by a dashed line and lightly shaded, which distinguishes it from Remote Oceania. Finally, the star indicates the region of interest for this study, the islands of Northern Island Melanesia.

Figure 1.1: Major Geographical Divisions of the Pacific Islands (adapted from Kirch, 2000)



The Y-Chromosome

The Y-chromosome is the smallest human chromosome, and averages 50-60 million base-pairs (bp) in length. It has the largest non-recombining sequence block in the human genome, which consists of 95% of its entire length. This non-recombining portion of the Y-chromosome (NRY) has been shown to be informative in that it is regionally polymorphic, although much less so than the remaining nuclear and mitochondrial genomes, and therefore a useful tool in population based studies. Furthermore, it contains few examples of parallel and back mutations due to its relatively slow rate of mutation, and few functional genes. These properties, along with its haploid, uniparental inheritance, lack of recombination and lack of selection, have made the NRY an extremely attractive area of the

genome for population and lineage-based studies (Cummins, 2001; Hammer and Zegura, 2002; Jobling and Tyler-Smith, 1995; Jobling and Tyler-Smith, 2003b).

The terminology used in this paper to describe Y-chromosome markers will follow those proposed by the Y-Chromosome Consortium (YCC) (Consortium, 2002). The term haplogroup will refer to NRY lineages defined by binary polymorphisms, and the term haplotype will refer to specific short tandem repeat (STR) profiles (combinations of alleles identified at multiple different loci). The specific haplogroups described in this paper will follow YCC (2002) terminology as well. However, updates are continually being added to the NRY tree and any changes will be indicated by the appropriate references.

The Use of the NRY and mtDNA in Population Genetic Studies

The human mtDNA and NRY have been used extensively to investigate evolutionary questions such as population structure (Destro-Bisol et al., 2004a; Destro-Bisol et al., 2004b; Hammer et al., 2003; Seielstad et al., 1998a), phylogenetic relationships (Cann et al., 1987; Underhill et al., 2001a), and divergence times (Friedlaender et al., 2002; Lum et al., 1994). Both genetic systems are inherited in a haploid manner, with the Y-chromosome passing strictly through the male lineage and the mtDNA passing through the maternal line. This mode of inheritance assumes no recombination, such that mutation is the only force that adds sequence variation to them (Jobling and Tyler-Smith, 2003b). Together, these areas of the genome are analyzed and used to present a more complete picture of human history, as the mtDNA describes female-mediated genetic histories and the NRY describes male-mediated genetic histories.

Genetic analyses of SNPs have proven especially useful for reconstructing prehistoric events because the ancestral and derived (mutated) sequences have had enough time to evolve independently (Roewer et al., 2005a). This means the ancestral and mutated sequences are easily distinguishable from each other, and can be used to determine population relationships and past population movements. STRs are another type of genetic marker utilized in population studies. They show a higher degree of genetic variation and a faster mutation rate than do SNPs. Due to these properties, STRs have been used to examine relatively recent population history because they can be shared as identical by state (recurrent mutations) rather than identical by descent (inherited from a common ancestor) (Kayser et al., 2001). Therefore, population relationships may be confused when examining more ancient relationships because of the possibility of shared lineages due to recurrent mutations.

Another assumption about mtDNA and NRY genetic polymorphisms is that sequence variation is selectively neutral at the majority of the loci under investigation. This assumption means that selection has not acted to increase or decrease the frequencies of the markers under investigation, and subsequently influence genetic interpretations by artificially creating closer or more distant genetic relationships. Natural selection will either increase or decrease genetic variation between populations based on the fitness of the individuals possessing a specific genetic variant in a particular environment. Natural selection should act to decrease the genetic variation in populations in similar environments, possibly suggesting a close historic relationship where none exists. A highly successful genetic variant could rapidly replace the other genetic variants in a population, resulting in what is called a selective sweep. On the other hand, populations in different environments become genetically more divergent due to differential selective pressures, which may act to erase close genetic ancestry (Jobling et al.,

2004). The mtDNA and NRY markers of anthropological interest are selectively neutral in that the mutations have little effect on fitness because the DNA base changes are not translated into a different proteins (synonymous mutation), or that variation in the amino acid sequence has little effect on physiology. Therefore, the mutant sequence might be so similar to the ancestral sequence in its effect on survival and reproduction that changes in its frequency would not be governed by natural selection.

As a consequence, the major cause of changes in haplogroup and haplotype frequencies between populations is genetic drift or gene flow. Random genetic drift accelerates differentiation between haplogroup representation in different populations, which makes it a useful property for investigating prehistoric events (Jobling and Tyler-Smith, 2003a). It occurs because a haplogroup frequency may increase by chance in some populations, while decreasing by chance in others. Examples of phenomena that contribute to genetic drift include population bottlenecks, founder effects and non-random mating. Thus, random genetic drift reduces diversity within groups and increases diversity between groups. However, the rate of fixation or loss of genetic polymorphisms is highly dependent on population size. Statistically, the variance in haplogroup frequencies after one generation is likely to be larger in smaller populations. As a result, the rate of random genetic drift is faster in smaller populations (Futuyma, 1998; Wright, 1931). Furthermore, the trail of haplogroup frequencies may reflect the movement of people, and ,therefore, their genes across the landscape.

Worldwide Distribution of Genetic Diversity

Overall, genetic diversity in humans is low relative to other primate species. This observation has been interpreted as reflecting the recent origin of modern humans and a small effective population size of the ancestors of modern human populations. Furthermore, the majority of human autosomal

genetic diversity is found within, rather than between, human populations. Most population-based studies have also found that Africans have the greatest levels of diversity, which has been interpreted as an African origin for modern humans (Jorde et al., 2000; Richards and Macaulay, 2001; Underhill et al., 2001c; Underhill et al., 2000).

The Y-Chromosome

Extant Y-chromosome variation shows an early diversification and dispersal of human males within Africa, and a widespread distribution of them across the African continent. The age of the most recent common ancestor has been estimated at various dates ranging from 70,000 to 130,000 BP (Underhill et al., 2001c; Underhill et al., 2000). This event was followed by a split in the lineage characterized by mutations at three sites: M42, M94 and M139 (Underhill et al., 2001c). All non-African individuals and a majority of African males carry these derived alleles. It is hypothesized that these individuals traveled from East Africa along a southern route to Southeast Asia sometime between 35,000 and 89,000 BP, and colonized both Sahul and the Asian mainland. These populations spread into Eurasia and eventually the Americas. Polynesia was the last part of the world to be colonized by modern humans at ~3,000 BP (Underhill et al., 2001c; Underhill et al., 2000).

On a worldwide level, Y-chromosome diversity tends to be geographically structured (Hammer et al., 2001; Underhill et al., 2001c). This is evident in the higher between group diversity displayed by the NRY versus other areas of the genome. The high degree of geographic differentiation has been explained by sex-specific demographic processes such as the practice of patrilocality, and the small effective population size of the NRY. For example, Hammer and coworkers (2001) examined 43 NRY binary markers in 50 populations and attributed 36% of the total variance to differences between

populations. Additionally, Kayser and colleagues (2001) analyzed 7 NRY STR loci in 986 males from 20 globally dispersed human populations, and found that the majority of the total genetic variance was found between populations, with only ~23% of the total genetic variance found within the populations studied. This extremely high value for population structure is likely due to the very fast mutation of NRY STRs versus the binary markers used by Hammer et al (2001).

The mtDNA

The maternal lineages of all living humans can similarly be traced back to Africa using phylogeographic and statistical analyses. The deepest branches of the mtDNA tree occur within African populations, and African populations tend to be the most genetically diverse in regards to mtDNA variation, suggesting an African origin for all extant mtDNA lineages (Richards and Macaulay, 2001). There is an early expansion of mtDNA lineages within Africa ~100,000-200,000 BP, followed by the movement of modern humans out of Africa by ~60,000 to 75,000 BP detected by a split in the mitochondrial phylogeny (Macaulay et al., 2005; Mishmar et al., 2003). These derived lineages are found in all non-Africans and some Africans today. Some of these individuals took a southern route to populate Sahul and Southeast Asia, and others moved north into northern Eurasia. Modern humans moved into the Americas by ~20,000-15,000 BP, and finally moved into the islands of Remote Oceania ~3,000 BP (Forster, 2004).

MtDNA variation also tends to display some regional variation, although it is not as geographically structured as the NRY. This observation has been interpreted as reflecting a higher female migration rate as compared to males as the result of the practice of patrilocality (Seielstad et al., 1998b). In patrilocal societies, females tend to move to their husbands natal home, and thus

experience greater population movement. Additionally, patterns of mtDNA diversity resemble the NRY in that there tends to be more genetic differentiation between groups than seen with autosomal loci (Hammer et al., 2001; Jorde et al., 2000). This is likely due to the small effective population size of these loci, which is $\frac{1}{4}$ of that in the autosomal genome and more subject to genetic drift.

A Synthesis of Worldwide Genetic Diversity in the Y-Chromosome & mtDNA

There are, therefore, several similarities between the uniparentally inherited mtDNA and NRY. Both areas of the genome tend to show higher levels of diversity in African populations than in non-African populations. These data have been used to support the hypothesis that modern humans first arose in Africa and then spread throughout the rest of the world. The times for the emergence of males versus females from the African continent is slightly different, but this could be accounted for by different demographic events in each sex's history, such as differential effective population size or mating behaviors (Forster, 2004; Ingman et al., 2000; Jorde et al., 2000; Richards and Macaulay, 2001; Seielstad et al., 1998b; Underhill et al., 2001c; Underhill et al., 2000).

On the other hand, global comparisons of human genetic variation in the mtDNA and NRY tend to show some distinct differences. NRY variation has been interpreted as displaying a pattern of regional specificity, a reduction of diversity within groups, and large differences between groups, as compared to the mtDNA pattern. As previously mentioned, it has been argued these global patterns are a result of the long-term human practice of patrilocality or patriphilia, in which females move to their husbands' natal household (Chaix et al., 2004; Hammer et al., 2001; Seielstad et al., 1998b).

However this interpretation has been highly debated, and other suggestions for these global patterns include sex-biased demographic processes, such as a lower male effective population size,

which can be caused by a considerably higher variance in male fertility, followed by subsequent genetic drift (Fay and Wu, 1999; Wilder et al., 2004b). In addition, there is some suspicion that the apparent distinction in male and female mediated variation is simply the result of a bias in the ascertainment of genetic variation in the non-recombining region of the notoriously invariant-Y chromosome versus the mtDNA, which is highly variable, especially in the control region (Wilder et al., 2004a). In other words, this bias is a result of the way in which polymorphic data are ascertained from the Y-chromosome, which relies on previously defined polymorphic markers that may or may not be polymorphic in the investigator's region of interest, versus the mtDNA, which has been relatively uniformly ascertained in all world populations (Jorde et al., 2000).

Research Goals

The purpose of this study is to determine the extent of genetic variation in the NRY in Northern Island Melanesia by more fully characterizing the phylogeography of paternal genetic lineages in this region. This approach is especially useful in providing a better resolution to the high degree of genetic structure that has already been detected in this region (Friedlaender et al., 2007a; Friedlaender et al., 2007b; Friedlaender et al., 2008; Scheinfeldt, 2004). In particular, greater phylogeographic resolution may allow us to address questions surrounding the settlement history of the South Pacific by identifying potential donor populations to the Polynesian gene pool, as well as the extent of island isolation in the region. Furthermore, it can elucidate the contradictory findings of the previously published NRY and mtDNA data, and tackle issues of sex-specific demographic processes in this region.

This study provides an opportunity to expand upon previous research conducted in this region (Scheinfeldt, 2004) by not only increasing the sample size, geographic coverage and number of SNPs

employed in order to overcome the ascertainment of bias, but also by investigating questions raised by the initial study. Ascertainment of bias arises from the way in which we collect and analyze NRY data based on the large size of the Y-chromosome. Because it is too time consuming and expensive to sequence the entire Y-chromosome for each sample in population analyses, population based studies of paternal genetic variation have relied on previously defined polymorphic regions of the Y-chromosome to analyze their samples. However, the majority of NRY polymorphic markers were discovered in European populations (Jorde et al., 2000; Kayser et al., 2001; Roewer et al., 2005b). This creates a situation where these markers may not be polymorphic in other populations, and thus will underestimate the genetic diversity present in them.

Recently, the Y-chromosome has been sequenced in more and more samples from different regions and new regionally-specific SNPs have inevitably been discovered (Cox et al., 2007; Kayser et al., 2008; Scheinfeldt et al., 2006). This study will employ several new markers that will attempt to differentiate some of the larger, general NRY haplogroups in Melanesia, namely the large K-M9* group. Additionally, this study will address one of the questions raised by Scheinfeldt (2004), namely whether the 50f2/c deletion defines its own separate branch of M9, belongs to a sub-branch of another M9 variant, or instead occurs on the background of several haplogroups (Figure 1.2).

This study is related to recently published manuscripts that include the author as a contributor (Scheinfeldt et al., 2006; Scheinfeldt et al., 2007). A large number of the data collected in this study is included in these publications and, therefore, some overlap in the description of this information is unavoidable. However, there are additional populations included in this study that were not included in the previous analyses, namely, the Easter Island and Manus Island groups. This study also addresses

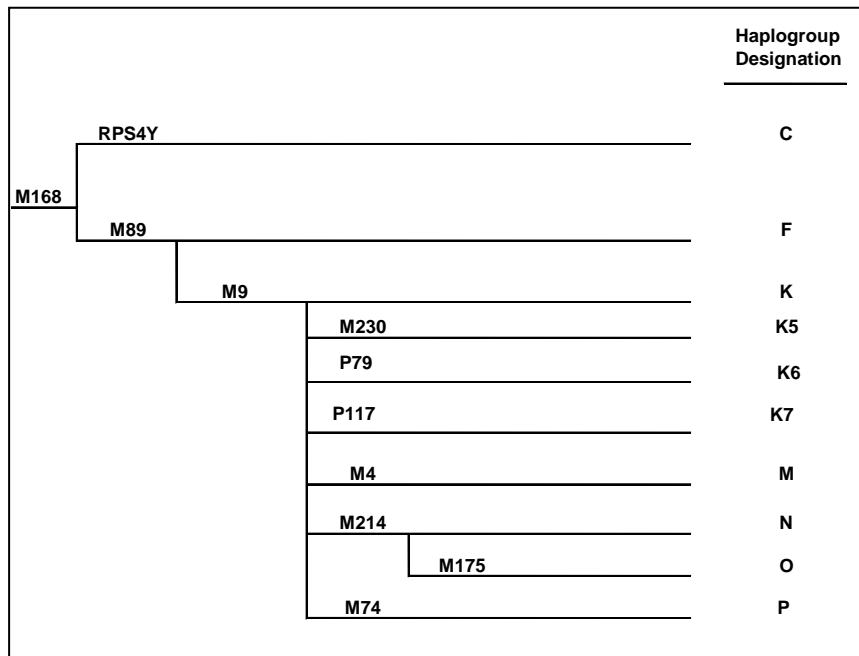
different hypotheses regarding the resolution of genetic structure in Northern Island Melanesia and the relationship of the NRY variation to that of other genetic loci investigated in this region.

More specifically, this study has a number of research objectives which include: 1) to determine the distribution of the newly defined SNPs (K6 and K7), as well as other newly defined SNPs that have not been extensively analyzed yet (M1, M2 and M2a1) across Northern Island Melanesia; 2) to investigate whether these newly defined SNPs will be able to provide evidence of new levels of heterozygosity and regionalization that may aid in overcoming the ascertainment of bias in Northern Island Melanesia; 3) to determine the extent of the paternal contribution of Southeast Asian males to the Northern Island Melanesian populations as recognized by the younger O lineages versus older C lineages that may have been carried into the area by the original settlers versus those lineages indigenous to Island Melanesia (lineages M and K); 4) to investigate the distribution of the 50f2/c deletion; and 5) to compare the regional NRY diversity with that of the mtDNA diversity in Northern Island Melanesia.

This study will show whether the levels of NRY diversity in Northern Island Melanesia are roughly equivalent to those for the mtDNA by using the data collected in support of objectives 1 through 4. This is an important objective (#5) because previous studies have detected different patterns of mtDNA and NRY variation, and have interpreted these patterns as representing different demographic histories between the sexes (Kayser et al, 2008; Kayser et al, 2006; Kayser et al, 2003; Capelli et al, 2001; Kayser et al, 2000; Su et al, 2000). The inclusion of new, regionally specific SNPs will potentially overcome some of the ascertainment of bias in NRY variation in Northern Island Melanesia by showing that levels of male and female mediated genetic variation are essentially equivalent. This goal will be accomplished by comparing the data collected in this study to the large

amount of mtDNA data collected in Northern Island Melanesia by Friedlaender and colleagues (2007a, b). This study will also include an analysis of genetic diversity on Manus and Easter Islands, regions whose populations have not been extensively studied at the molecular level.

Figure 1.2: Simplified Y-Chromosome SNP Phylogenetic Tree (YCC 2005)



CHAPTER 2 ARCHEOLOGICAL AND LINGUISTIC VARIATION IN THE SOUTH PACIFIC

Archeological Patterns in Oceania

Archeological remains provide only indirect evidence of past human presence. However, these items are often more numerous than human skeletal remains in many parts of the world. Therefore, archeological data are important for two main reasons: (1) they provide evidence for the presence of modern humans at sites or times where no human remains have yet been discovered; and (2) they provide evidence for different forms of behavior at those sites (Jobling et al., 2004). However, only a small number of archeological sites of great antiquity have been found in Oceania. The lack of sites is attributable to poor preservation of archeological material due to the tropical climate and high rainfall levels, and cultural practices like forest clearance and gardening that occur in this region. These activities can cause erosion and ground surface alterations that decrease the likelihood of archeological discoveries (Pavrides and Gosden, 1994).

Archeological Data from the Pleistocene Epoch

Geographically, Oceania was different during the Pleistocene when sea levels were lower than at present, due to a large amount of the earth's water being trapped as ice at the earth's poles. Today's Australia, New Guinea, and Tasmania were part of a large landmass known as Sahul, and the modern islands of Java, Sumatra, Bali, and Borneo were part of a combined landmass known as Sunda. Sahul has always been separated from Sunda by a considerable water barrier, which is significant in scientific studies of Oceania because of the biological differences in flora and fauna on either side of the barrier. Marsupials such as kangaroos and koalas, and the platypus are found on the

Sahul side, and apes and tigers are found on the Sunda side of the barrier (Kirch, 2000). These differences were first described by Alfred Russell Wallace, thus, the area located between Sunda and Sahul is called Wallacea (Kirch, 2000).

Analyzing the geography of Wallacea indicates two possible routes of entry into Sahul by Pleistocene human populations. The first route would have involved the islands of Sulawesi, Peleng and Sula as stepping stones to Sahul via a northern route involving Halmahera or a southern route involving Buru and Seram. The second route would have used the islands of Lombok, Sumbawa, Flores, Alor, and Timor as stepping stones to Sahul. While the exact path into Sahul is still under debate, it is obvious that the initial movement into Sahul required a succession of voyages between 10 and 100 km in distance between inter-visible islands (Kirch, 1997; Kirch, 2000; Summerhayes, 2007).

Over 150 archeological sites dating to the Pleistocene have been discovered in the Sahul region, which range from Tasmania, across Australia, to the Bismarck Archipelago, and out into the Solomon Islands (Allen, 1994). The earliest occupation dates for Australia come from the Malakunanja II and Malangangern sites in the northern part of the continent. These sites contain a wide range of stone artifacts in their primary depositional settings. Thermoluminescence dates obtained from quartz grains at the occupation sites yield an age of around 50,000 BP, and suggests the arrival of people in this area between 50,000 and 60,000 BP (Roberts et al., 1990). Other early archeological sites in Australia include hearth features dated at around 42,000 BP at Devil's Lair in Western Australia, lithics and features dated at around 40,000 – 43,000 BP at Lake Mungo in Southern Australia, and artifactual data from the rockshelters of Nauwalabila and Malakunanja in Northern Australia dated at around 50,000 BP. However, the older thermoluminescence dates have been questioned by some scholars

(Kirch, 1997; O'Connell and Allen, 2004). Thus, to date, the most accepted time of entrance into Sahul is around 40,000 years ago.

Other regions of Sahul have also produced relatively early dates for a human presence there. Pleistocene occupation sites discovered at the Peninsula on the north coast of New Guinea have been dated by both radiocarbon and thermoluminescence methods at around 40,000 BP. Hundreds of waisted axes were recovered from this site, both *in situ* and loose, and there is evidence for the hafting of these axes onto wooden stakes (Groube et al., 1986). Similar ages have been reported for the site of Lachitu Cave on the North New Guinea coast, while the earliest dates for the New Guinea interior are around 30,000 BP (Summerhayes, 2007).

Pleistocene humans apparently moved relatively quickly from the Sahul mainland to the islands east of New Guinea. The island of New Britain would have been visible from the northeastern coast of New Guinea, and the earliest site in New Britain, Yombon, has been dated to around 35,000 BP. This site is found in the lowland tropical rainforest of west New Britain, and is important because it demonstrates the early use of inland forest environments and resources (Pavrides and Gosden, 1994). New Ireland would also have been visible from northern New Britain. Rockshelters at Buang Merabak and Matenkupkum on the east coast of New Ireland have been radiocarbon dated to around 35,000 – 40,000 BP (Summerhayes, 2007). The Kilu site on Buka Island has been dated to around 28,000 – 32,000 BP, and indicates human occupation of the Solomon Islands by this time (Kirch, 1997; Summerhayes, 2007).

Archeological Data from the Pleistocene-Holocene Transition

There is no archeological evidence for the transfer of people and goods between the geographically separated areas of New Guinea and the Bismarck Archipelago until 20,000 years after the initial settlement of Sahul (Summerhayes, 2007). At this time, the first evidence for the movement of obsidian from west New Britain to New Ireland is documented in the archeological record. This would require transporting the obsidian over 350 km, and with at least one water crossing (Kirch, 1997). The Gray Cuscus (*Phalanger orientalis*) appears at sites on New Ireland around 20,000 years ago as well. This is significant because the Gray Cuscus is an indigenous New Guinea marsupial (Kirch, 2000). This type of exchange suggests that, instead of moving people to resources, the resources were now being moved over long distances to the people. It also suggests an increased frequency of water voyages and increases in population sizes. This may have been the first time in Oceanic history that chains of exchange networks had developed between mobile foraging groups (Summerhayes, 2007).

Pamwak Cave on Manus shows evidence of human occupation around 21,000 BP. This settlement is significant because it would have required a minimum water crossing of 200 km, which is believed to be one of the earliest water crossings of this magnitude (Kirch, 2000; Summerhayes, 2007). The archeological evidence suggests that Manus was relatively isolated until around 13,000 BP, when evidence of the translocation of indigenous New Guinea plants and animals appears at several sites (Summerhayes, 2007). A different cuscus species from that found on New Ireland, the *Spiloguscus kraemeri*, as well as the bandicoot (*Echymipera kalubu*) and *canarium indicum* nuts were found at Pamwak Cave. However, there seems to have been no exchange between Manus and New Britain, or

Manus and New Ireland, because these flora and fauna are absent from sites on New Britain and New Ireland from this same time period (Summerhayes, 2007).

The Pleistocene-Holocene transition was also a time of global warming at the end of the Last Global Maximum (LGM). This process would have caused a rise in sea levels and a change in the coastlines. Such a geographical change may have led to a change in settlement patterns as people moved inland. The environmental changes may also have triggered a change in subsistence practices because there is evidence of the beginnings of agriculture at this time period (Kirch, 1997).

Archeological Data from the Early Holocene Epoch

Agriculture first appears in the Near Oceanic archeological record during this time period. At the Kuk Plantation site in the Wahgi Valley, six distinct phases of artificial drainage and geomorphological alteration have been defined and radiocarbon dated. Phase 1, dated to 9,000 BP, contains features related to simple horticultural practices. Phase 2 (6,000-5,500 BP) contains raised beds for the cultivation of taro (*Colocasia esculenta*). Phase 3, dated to 4,000-2,500 BP, contains reticulate drainage systems. Phases 4-6 (after 2000 BP) yield extensive drainage systems indicative of sweet potato (*Ipomoea batatas*) cultivation. The Kuk Plantation site is important in defining the history of the South Pacific because it shows that many of the plants found in early Oceanic sites, once thought to be of Southeast Asian origin, were actually domesticated in New Guinea (Golson and Gardner, 1990; Kirch, 2000). Examples of plants domesticated in Near Oceania include taro (*Colocasia esculenta*), yams (*Dioscorea* species), kava (*Piper methysticum*), sweet almond (*Canarium* species), coconut (*Cocos nucifera*), banana (*Musa* species), and sugar cane (*Saccharum officinarum*), among others (Allaby, 2007).

This time period may also mark the beginning of regular exchange between the Bismarck Archipelago and New Guinea. Obsidian, dated to 4,500 BP, was recovered at the highland site of Kafiavana. There are no natural sources of obsidian on New Guinea, and the geographic source of the Kafiavana obsidian was found to be west New Britain. In addition to increased exchange networks within Near Oceania, there appears to have been increased interactions with Asia as well. This is revealed by the introduction of Asian domesticates and cultigens to Near Oceania, and the discovery of indigenous New Guinea animals in Asia (Summerhayes, 2007).

The Lapita Dispersal

A little over 3,000 years ago, there was a major movement of people, language, and culture into Remote Oceania. The most prominent archeological signature of this movement is an ornate style of pottery called Lapita. Lapita pottery is created by a paddle-and-anvil technique, and is low fired at temperatures below 800° C. It is characterized by dentate-stamping and incision with a finely pointed tool. Lapita pottery is found in a variety of forms and shapes, and there is mounting evidence that the decorated pots were nonutilitarian (Kirch, 2000).

Along with this distinctive pottery style, the associated portable artifacts and structural elements constitute the Lapita cultural complex, which is considered the ancestral cultural complex from which the Polynesian culture was derived. This cultural complex is almost always associated with the following elements: (1) dentate-stamped pottery; (2) incorporation of Admiralty Islands obsidian sources; (3) adz/axe kits; (4) fishing gear; (5) shell ornaments; (6) tattooing artifacts; and (7) stilt-house dwellings. Other elements commonly associated with the Lapita cultural complex include: (1) domesticated Oceanic food plants; (2) domesticated commensals, such as pigs, dogs, chickens, and

root and tree crops; (3) outrigger sailing canoes; (4) bark-cloth production; (5) a heteroarchical, house society type of social organization; and (6) aspects of a Proto Oceanic lunar calendar (Green, 2003).

Lapita pottery fragments have been found in a wide arc of the southwestern Pacific, from the Sepik coast of New Guinea and eastward to Fiji, Tonga, and Samoa (Green, 2003; Summerhayes, 2007; Terrell and Welsch, 1997). The Lapita movement was not instantaneous, and likely precursors to it can be found in Wallacea (Kirch, 1997). Around 4,000 BP, human groups using “ornate” pottery, shell tools, and an agriculturally based economy were living in this region. The pottery style was similar to Lapita pottery with low fired paddle-and-anvil construction, and incision and stamped decoration, among other things. These pottery assemblages only lacked the specific decorative motifs that define Lapita style pottery (Kirch, 1997).

Lapita sites show a temporal pattern with a west to east gradient. The earliest known Lapita sites are found in the Bismarck Archipelago. Radiocarbon dates from the Mussau Island group indicate a Lapita presence there by around 3,300 BP. These sites contain pottery, obsidian, the remains of stilt-houses, and other artifacts and ecofacts (Kirch, 1987). Lapita pottery, wood, and animal remains from the Arawe Islands date to around 3,000 BP (Gosden and Webb, 1994). Lapita was established in the Solomon Islands and Vanuatu by around 3,000 BP. Evidence of the Lapita cultural complex appeared on Fiji by around 2,900 BP, and Tonga and Samoa by around 2,800 BP (Kirch, 1997; Kirch, 2000; Summerhayes, 2007).

In summary, Lapita sites are different from anything prior to them in Oceania. They contain large coastal settlements, with stilt-houses, large quantities of ornate pottery, and an agriculturally based economy (Kirch, 1997; Summerhayes, 2007). The Lapita cultural complex has been associated with Austronesian-speakers by the reconstruction language terminology (Green, 2003; Terrell et al.,

2001). For example, terms have been reconstructed for shell ornaments, house terminology, fishing implements, seafaring and navigation terminology, and a variety of cultivation-related activities and crops, among others (Lindstrom et al., 2007). These individuals represent the first group of people moving into Remote Oceania. There is no abrupt end to Lapita, but rather continuity into local archeological sequences as populations adjusted to new environments (Kirch, 2000).

Fossil and Morphological Patterns in Oceania

Fossil Data

The most direct evidence of the time in which modern humans reached Oceania would be accurately dated fossil remains. However, fossil remains from this area and time period are rare, and archeological dating techniques are often disputed. To date, some of the earliest modern human fossils in Sahul come from Lake Mungo, one of the now dry Willandra Lakes in southeastern Australia. This area has produced substantial quantities of archeological material, including skeletal remains, from various stratigraphic layers. These earliest burials include a fully articulated individual that was covered with powdered red ochre before burial (LM III), and a fragmentary female cremation (LM I) from the same stratigraphic unit. The skeletal morphology of LM III is gracile and falls within the range of modern aboriginal Australians. The morphological sex of this individual has been declared ambiguous, although the position of the skeleton for burial (both hands clasping and protecting the penis) indicates male (Bowler et al., 2003; Thorne et al., 1999). Initial radiocarbon dating placed these individuals in the 26,000 to 17,000 BP range; however, later study by Bowler and colleagues (2003), which improved the resolution for the time interval beyond the range of radiocarbon dating by using optically stimulated luminescence, placed the burials in the 40,000 ± 2,000 BP range. While the exact

date of the Lake Mungo remains are still in debate, it is clear that they provide direct evidence of the earliest modern human presence in Sahul by at least 40,000 BP.

A determination of the precise geographic origin of these Pleistocene inhabitants is also difficult due to the paucity of archeological sites in Southeast Asia. There is no evidence of a human presence in Southeast Asia until around 43,000 BP. A poorly preserved juvenile skull has been dated to around this time at Niah Cave on Borneo. Stone tools from the Lang Rongrien Cave site in Southern Thailand were dated at around 38,000 BP (Summerhayes, 2007). Additional sites need to be discovered and analyzed from this area for a true understanding of the route into Sahul.

Linguistic Patterns in Oceania

Linguistic evidence can be used as a tool for analyzing the relationships and migratory patterns of the various groups of individuals in the South Pacific. Currently, all of the languages in this region can be classified into two major clusters: the Austronesian language family and the Papuan language family. The Austronesian languages belong to a single language family and are widely distributed across Oceania, whereas the Papuan languages belong to many different language families and are fairly restricted to New Guinea and a few surrounding islands.

Methods of Linguistic Analysis

The major approach used by linguists to determine the relationships among languages is the Comparative Method. It investigates two or more languages for recurrent sound correspondences in arbitrary form-meaning pairings, such as basic vocabulary and sets of grammatical elements. These correspondences then provide a basis for the reconstruction of the sound system, lexicon and

morphology of the common ancestor. The Comparative Method classifies languages according to the distribution and weighing of shared changes to the reconstructed common ancestral language with a focus on the morphemes (smallest meaningful units) that the related languages have in common (Lindstrom et al., 2007; Pawley, 2000; Pawley and Ross, 1993).

The comparative method employs three steps in its linguistic examination. These include: (1) diagnosis of evidence to propose which groups of languages form genealogical groupings; (2) collecting cognate sets including words; and (3) working out the sound correspondences of the cognate sets. This process allows for subgrouping because it shows that present-day languages belonging to the family are descended through regular sound changes from a common ancestral lineage. Subgroups are sets of languages that form part of a family. Members of a subgroup are more closely related to each other than to other members of the family because they share a common ancestor not shared by other members of the family (Ross, 2001; Ross, 2004).

Comparative data can be analyzed by the family tree model (emphasizes continuity of 'genetic' transmission), the substrate or language mixing model (emphasizes breaks in the continuity of 'genetic' transmission), and the diffusionist model (emphasizes the movement of elements across established language boundaries). The construction of a family tree plays a central role in the comparative method. The strength of family tree building rests on several assumptions. The first is that language change occurs at a steady, constant rate. This is a crucial feature because differential rates of change would obscure the branching pattern. Construction of a family tree also assumes that languages are learned by successive generations of native speakers. Continuity occurs as long as the line of native speaker transmission is unbroken. Linguistic splitting occurs when geographic or social barriers separate a population of native speakers that once spoke the same language, and successive splits

would form a family of related languages. Thus, a family tree diagram schematically represents the sequence of linguistic splitting (Pawley, 2000; Pawley and Ross, 1993; Renfrew, 1994). However, these diagrams tend to oversimplify real events and can be confounded by several phenomena. Especially problematic to the family tree model is linguistic borrowing, which can be difficult to discern in older language contact situations (Lindstrom et al., 2007).

The Comparative Method provides a way for demonstrating 'genetic' relationships between languages, for determining the subgrouping of a language family, and for reconstructing the historical phonology and lexicon of early languages. It is a set of procedures for organizing and evaluating linguistic data in light of a theory concerning the process by which certain kinds of continuity and change occur in languages. While comparative analysis is a strong tool for linguistic analysis, it requires good descriptive data and takes a long time to complete. Despite all of the advantages of the comparative method, it is restricted by a temporal range of 8000 ± 2000 years because linguistic material erodes rather quickly over time by a process of sound change (Pawley, 2000).

Other methods have been employed in historical linguistic reconstruction, but have not been as successful as the Comparative Method. One such example, called the Mass Comparison Method, was used by Greenberg (1971) to place all of the languages of Oceania except the Austronesian languages and the mainland Australian languages into a large group called the Indo-Pacific phylum (Greenberg, 1971; Lindstrom et al., 2007). The Mass Comparison Method compares shared words and meanings across many languages and uses these similarities as evidence of a linguistic relationship. A major weakness of the Mass Comparison Method is that it does not recognize reasons for linguistic similarity outside of a genealogical relationship, namely, linguistic borrowing and chance similarities (Lindstrom et al., 2007).

A typological approach has also been employed in historical linguistic analyses, and classifies languages based on their shared similarities in structure. This method compares structural systems in order to determine universal principles of associations between language types. Statistical analyses of the frequency of particular associations can then be used to form the basis of historical inferences about earlier linguistic systems and directions of change. The major weakness of this method is that it recognizes frequencies, but not historic particulars and should thus be used in conjunction with other methods (Pawley and Ross, 1993). For example, a typological approach has been used to investigate some of the linguistic features of the Proto-Oceanic languages by Lindstrom and colleagues (2007), as a compliment to the stronger Comparative Method approach to linguistic prehistory.

The Austronesian Language Group

Austronesian languages comprise 1/6 of the world's languages, and are widely distributed from Madagascar in the west to Easter Island in the east, and from Taiwan in the north to New Zealand in the south (Pawley and Ross, 1993). All of the Austronesian languages belong to a single language family and, therefore, have a recent common origin. This language family can be broken into subfamilies that are used to create a phylogenetic tree, which represents the sequence of linguistic splitting (Lindstrom et al., 2007). Analysis of its language tree shows that Taiwan contains the most diverse set of Austronesian languages, with nine out of ten of the major subfamilies confined to this area. This is strong evidence that the Austronesian language family originated in Taiwan. All of the remaining Austronesian languages belong to the Malayo-Polynesian subfamily, with a most likely dispersal location in the Philippines (Bellwood, 1991; Lindstrom et al., 2007; Pawley, 2003; Pawley and Ross, 1993).

The Malayo-Polynesian subfamily can be further divided in smaller subfamilies, one of which is the Oceanic subgroup. All of the 480 Austronesian languages spoken in the Pacific Islands fall into the Oceanic subgroup, with an origin in and most likely dispersal location from north New Britain in the Bismarck Archipelago (Lindstrom et al., 2007; Pawley, 2003). The spread of the Oceanic language subfamily is associated with the Lapita cultural complex, and the spread of Lapita across Melanesia into Polynesia is consistent with the pattern of Oceanic linguistic subgrouping (Pawley and Ross, 1993). The Proto-Austronesian languages were most likely brought by Neolithic farmers from Taiwan through the Philippines via Borneo and Sulawesi into Melanesia and western Indonesia. These languages were then spread eastward through east Nusantara, skirting New Guinea to occupy the Bismarck Archipelago around 3500 BP. The Oceanic subfamily then rapidly occupied parts of Melanesia, and radiated outward to Micronesia and Polynesia, which were either empty or sparsely inhabited (Bellwood, 1991; Ross, 2004).

The point of dispersal for the Austronesian languages can be estimated by reconstructing a Proto-Austronesian language, which is the ancestor to modern Oceanic languages. The current consensus is that the speakers of this Proto-Austronesian language migrated from Taiwan ~6,000 BP to other parts of Southeast Asia and Oceania (Lindstrom et al., 2007). The individuals who spoke this language were Neolithic agriculturalists, who generally occupied coastal dwellings, and who possessed sophisticated navigational skills. This information can be surmised from reconstructed words such as rice, garden, landward, seaward, boat, and voyage (Pawley and Green, 1973). Thus, the original settlers of Near Oceania, arriving at least 30,000 BP, spoke a non-Austronesian language.

The Papuan Language Group

The Papuan languages belong to several different language families that may or may not share a common proto-language. Papuan languages are those which are not Austronesian in the region stretching from Timor and Alor in the west, through the island of New Guinea and to the Solomon Islands in the east (Pawley, 2007). The eastern limit of the Papuan languages fits closely with the boundary of Near Oceania. The highest concentration of Papuan languages occurs on the island of New Guinea, while a lower concentration occurs to the west, a spotty distribution appears on the islands to the east, and a sparse distribution of the Papuan languages is found across the Solomons (Ross, 2001). These languages predominate in New Guinea, where the Austronesian languages are restricted to coastal areas, making New Guinea one of the most linguistically diverse areas in the world (Pawley, 2003). New Guinea itself has 750 Papuan languages divided into 18 language families, as well as some isolates that cannot be clustered into families (Pawley, 2007).

The East Papuan language family includes all but one of the non-Austronesian languages spoken on the islands located off the east of mainland New Guinea (Terrill, 2002). These languages are believed to be those spoken by the original inhabitants of Island Melanesia, who were most likely Pleistocene hunter-gatherers that entered Melanesia from Island Southeast Asia at least 30,000 years BP (Dunn et al., 2002). Eight distinct East Papuan language groups have been found through a comparative analysis of pronoun patterning (Ross, 2001). Analysis of East Papuan languages suggests a varied nature with great differences in linguistic type among them as a whole. However, several features were found to be shared by the heterogeneous East Papuan languages, making for some unity among them. These shared characteristics can be interpreted as either inherited from one

or more ancestral languages or the result of contact among these languages in the past (Dunn et al., 2002).

Nominal classification systems have also been investigated in the East Papuan language. These systems preserve older features that tend to be lost elsewhere in a language, are not easily borrowed into languages, and are useful tools for exploring ancient historical relationships between languages. However, little evidence of relationships were found between the nominal classification systems of the East Papuan languages as a whole. This finding can be interpreted as meaning that either the systems were inherited so long ago as to obscure the genetic evidence of relatedness or the borrowing of grammatical systems from other language groups has blurred their historical relationships (Terrill, 2002).

The Trans New Guinea family of languages was first proposed by Wurm (1975) when he suggested a common origin for 500 of the 750 Papuan languages of New Guinea based on typological and lexical resemblances. Wurm presented a complex theory claiming that: (1) three sets of Papuan pronouns could be reconstructed; (2) these forms were reflected in numerous languages of the various Papuan phyla; and (3) the distribution of the forms reflected contact between various waves of migrations (Ross, 1995; Wurm, 1975). This hypothesis was originally regarded as weak due to typological analysis of a linguistic area where languages have been borrowing from one another for thousands of years (Pawley, 2001).

However, recent analysis of the lexical and morphological structure of the Papuan languages supports a modified form of this hypothesis. These studies showed that there is a valid group including 400 of the original 500 Papuan languages assigned by Wurm to the Trans New Guinea language group (Pawley, 1997). Descriptive data on Papuan languages was also analyzed using both a top-

down (comparing distantly related languages to determine rough approximations of ancestral language groups) and bottom-up (beginning with lower order language groups to determine current linguistic relationships) reconstructions. A modified version of the Trans New Guinea language group hypothesis, called Trans New Guinea (TNG) IV, was then proposed and indicated a likely dispersal center for this language phylum in the central highlands of New Guinea. This dispersal may have been powered by the development of agricultural crops such as taro, yam, bananas, and sugar cane (Pawley, 2000).

As mentioned earlier, despite all of the advantages of the comparative method, it is restricted by a temporal range of 8000 ± 2000 years because linguistic features comparatively quickly erode over time by a process of sound change. Applying this method to the Papuan languages of Melanesia has been controversial because it is believed that these languages are much older than the limits of the comparative method, and that those Papuan languages would have survived to the present are not representative of the wide variation of languages spoken in the area prior to the introduction of the Austronesian languages. Dunn and colleagues (2005) believe that applying cladistic methods to language structure (sound systems and grammar) instead of vocabulary may extend the time depth of linguistic analyses. They applied this technique to 15 Papuan languages and concluded that they are able to show an archipelago-based phylogenetic signal that is consistent with the geographical distribution of the languages, and that these languages most likely diverged from a common ancestral language during the late Pleistocene (Dunn et al., 2005).

Summary of Linguistic Patterns in Oceania

Several broad conclusions have been assembled from the linguistic data for the South Pacific. These include: (1) the non-Austronesian or Papuan languages have been in the area much longer than the Austronesian languages; (2) non-Austronesian languages never reached Remote Oceania; (3) Austronesian languages most likely entered Near Oceania from Wallacea; and (4) Austronesian languages dominate Remote Oceania, while Near Oceania is linguistically diverse.

CHAPTER 3 BIOLOGICAL VARIATION IN THE SOUTH PACIFIC

Autosomal DNA Variation in Oceania

Analyses of autosomal variation in Oceania suggest different scenarios depending on the loci being investigated. A study of 14 neutral STRs showed a marked decrease in heterozygosity from island Southeast Asia across Near Oceania and into Remote Oceania in an analysis of 27 Pacific Island and Asian populations. These data suggested the occurrence of a series of population bottlenecks during the settlement of the Pacific. Furthermore, this study indicated that variation followed geographic rather than linguistic boundaries in Near Oceania due to significant gene flow among the populations (Lum, 2007; Lum et al., 2002).

Human leukocyte antigens (HLAs) are highly polymorphic genetic loci that have been used in various population genetic analyses. HLAs are proteins located on the surface of white blood cells that participate in the immune response. Analyses of these loci show that East Asian, Australian, Melanesian, Highland Papua New Guinea, and Polynesian populations can be distinguished from each other (Mack et al., 2000). The HLA loci also revealed a close genetic relationship between aboriginal Taiwanese and Polynesian populations (Serjeantson et al., 1982; Zimdahl et al., 1999). However, these loci have also been shown to be under selective pressure, meaning the modern distributions of HLA diversity may reflect a combination of founder effects, migration and selection (Serjeantson et al., 1982).

Genetic variations in globin genes have also been used to investigate population relationships in Oceania. The limited geographic distribution of the $-\alpha^{3.7}$ III mutation suggests that it arose in one place at one time, with the most likely place of origin in Island Melanesia. This deletion is found in both

Melanesians and Polynesians, but has not been detected in other populations (Hill et al., 1985; Hill et al., 1987; O'Shaughnessy et al., 1990). Globin gene mutations are known to be under positive natural selection in malarial regions, and the high frequency of the $-\alpha^{3.7}$ III mutation in Melanesia may reflect the malaria endemicity in that region. However, this does not explain its presence in parts of Polynesia, which are free of malaria. This observation has led to the conclusion that the $-\alpha^{3.7}$ III mutation was carried into Remote Oceania by migrants from Melanesia (Flint et al., 1986; Hill et al., 1985).

Immunoglobulin (GM) allotype haplogroups have also been used to investigate biological variation in Oceania because these markers show different haplotype frequencies among populations (Schanfield et al., 2007). Schanfield and colleagues (2007) analyzed GM frequencies in several thousand individuals from East Asia and the Pacific. They found the New Guinea populations are closely related to the central Australian populations, and that significant differences in haplotype frequencies are noted between the Papuan and Austronesian speaking populations. Furthermore, there seem to be a Southeast Asian/Austronesian signal by looking at the distribution of the GM haplotypes GM* A, F B indicating a close relationship between the Taiwanese and Polynesian populations (Schanfield et al., 2007).

Mitochondrial DNA (mtDNA) Variation in Oceania

Melanesia

The mtDNA lineages detected in Melanesia have been shown to be extremely diverse and informative for understanding population history in the area. This internal genetic variation reflects linguistic divisions, island-by-island isolations, and marital migration distinctions between coastal and

inland populations (Friedlaender et al., 2005a). A recent study by Friedlaender and coworkers (2007) employed an especially large and intensive sample set in Northern island Melanesia to investigate mtDNA diversity in Near Oceania. They found remarkable population structure in this area as indicated by a large among-group variance, and showed that the variation among islands was almost as great as the variation among groups within islands. The authors concluded that the mtDNA data indicate an initial ancient settlement of Northern Island Melanesia at ~40,000BP, followed by isolation of many of the populations, subsequent internal population expansions, and the introduction of later peoples and haplogroups during the Holocene accompanied by high levels of admixture with the previously established groups.

A study by Ingman and Gyllensten (2003) analyzed the complete mitochondrial genomes in 101 individuals from Australia, New Guinea, Melanesia, Polynesia, Africa, India, Europe and Asia in order to form hypotheses concerning the evolutionary history of Australia and New Guinea. The genetic diversity seen in New Guinea was extremely high, and the authors found that the New Guinea and Australian lineages were more closely related to each other than either was to Asian sequences. These data suggested either a joint colonization of the region or admixture between the two regions soon after colonization between 40,000 – 70,000 BP, which would make sense because both regions were part of the large landmass of Sahul (Ingman and Gyllensten, 2003).

A more recent study by Hudjashov and colleagues (2007) analyzed 172 Australian and Melanesian mtDNA haplotypes and 522 NRY haplotypes to address similar historical questions surrounding Australia and New Guinea. Their work suggested similar dates of settlement and population relationships as suggested by Ingman and Gyllensten (2003) (Hudjashov et al., 2007). Kayser and colleagues (2003) also found high levels of mtDNA diversity in populations of West New

Guinea. These results are in stark contrast to the reduced levels of Y chromosome diversity detected by the same study, which will be discussed below.

There are several major mtDNA haplogroups present in Melanesian populations:

Haplogroup P: Haplogroup P is one of the oldest lineages found in Near Oceania, and the initial branching of P occurred prior to the settlement of Sahul. It is extremely rare west of Wallacea, and some of its branches show very old connections between Near Oceania and Australia. Within Melanesia, haplogroup P is most common in New Guinea, less common in New Britain, and increasingly rare in New Ireland, Bougainville, and islands to the southeast (Forster et al., 2001; Friedlaender et al., 2007a; Friedlaender et al., 2005b).

Haplogroup M: Haplogroup M is also a very ancient macrocluster in Melanesia, and its coalescent date is very similar to that of Haplogroup P. Current data suggest that the Oceanic branches developed around the time of initial settlement at ~ 40,000 BP. The most common branch of haplogroup M in Near Guinea is Q, which appears more frequently than haplogroup P, but is mainly restricted to Melanesia (Forster et al., 2001; Friedlaender et al., 2005b). However, a recent study has detected Q mtDNAs in a single Australian aborigine (Hudjashov et al., 2007). The oldest branch of Haplogroup M in Northern Island Melanesia is M27 (Merriwether et al., 2005). Its origin is most likely in Bougainville, and it has not yet been detected in New Guinea. Different branches of M27 have very distinct distributions within Northern Island Melanesia. The M28 and M29 lineages are also extremely old, and appear to have their origins in Northern Island Melanesia, specifically in New Britain and New Ireland respectively (Friedlaender et al., 2007a; Merriwether et al., 2005).

Haplogroup B: Haplogroup B represents a more recently introduced lineage in Melanesia. It is defined by a COII/tRNA^{LYS} intergenic 9 base-pair deletion, among other SNPs. This haplogroup is

found at moderate frequencies in Southeast Asia, as well as throughout Oceania. A unique pattern of three mutations (16217, 16347 and 16261) in the control region linked to the 9 base-pair deletion has been termed the “Polynesian Motif” (now called branch B4a1a1) because of its high frequencies in Polynesian populations (Lum and Cann, 2000; Melton et al., 1998; Melton et al., 1995; Merriwether et al., 1999). The precursor (B4a1a) to the Polynesian Motif has been identified in Aboriginal Taiwanese groups, and has been dated to $12,200 \pm 4,700$ BP based on control region sequences and $13,169 \pm 3,800$ BP based on coding region data (Friedlaender et al., 2007a). Haplotypes with the full Polynesian Motif have not yet been detected in Taiwan, but occur sporadically in central and eastern Indonesia and Madagascar (Hill et al., 2007). This branch is also found at higher frequencies in Near Oceania and is almost fixed in a majority of populations in Remote Oceania (Friedlaender et al., 2007a; Friedlaender et al., 2007b; Melton et al., 1998; Melton et al., 1995). Interestingly, the Polynesian Motif has not been detected in prehistoric human bones from Melanesia that are believed to represent some of the Holocene voyagers from Taiwan (Hagelberg and Clegg, 1993).

Polynesia

There is a noticeable decrease in mtDNA diversity in Polynesia. A clinal increase of the frequency of haplogroup B from west to east accompanied by a decrease in control region diversity has been interpreted as evidence of founder effects across the Pacific. Studies of mtDNA lineages in Polynesia show that the majority of mtDNA haplogroups found in Polynesian populations are of Southeast Asian origin. This is because the majority of the mtDNA lineages in Polynesia belong to haplogroup B, specifically the B4a1a1 branch. For example, Kayser and colleagues (2006) compared mtDNA diversity in 400 Polynesians from eight island groups to over 900 individuals from Melanesia,

Southeast Asia, East Asia and Australia. They classified 93.8% of the haplogroups in their Polynesian sample as originating in Asia.

The high frequency of the “Polynesian Motif” in Polynesia associated with Austronesian speakers has led some investigators to interpret this as a genetic signature of the people associated with the Lapita cultural complex. It is believed that these individuals originated in Taiwan, spread through Melanesia, and finally settled in remote Oceania. However, Friedlaender and colleagues (2007a, b) do not believe that these data have been correctly interpreted. The full “Polynesian Motif” (B4a1a1 lineage) has been dated to $\sim 8,700 \pm 2,100$ BP, which considerably predates the appearance of the Lapita intrusion at $\sim 3,300$ BP. Thus, the distribution of these haplotypes in Island Melanesia does not fit a neat association with Oceanic-speaking populations. Its frequency is extremely high in some Papuan-speaking groups on Bougainville, New Ireland, New Britain, and New Guinea. Additionally, this haplogroup is not as common in New Britain, the putative home of the Lapita Cultural Complex, as it is in New Ireland and Bougainville (Friedlaender et al., 2007a; Friedlaender et al., 2005a). Furthermore, Friedlaender et al (2007a, b) have distinguished several local variants of the B4a1a1 sublineage in Melanesia. This finding indicates that haplogroup B has been in Northern Island Melanesia long enough to develop these variants, which have reached high frequencies in some populations.

Y-Chromosome (NRY) Variation in Oceania

Melanesia

Y chromosome diversity in Melanesia is relatively high, reflecting the long and complex history of the area. However, previous studies have not detected the range of variation and extreme

population structure as that found in the mtDNA data. Scheinfeldt (2004) found a correlation between patterns of NRY variation and patterns of linguistic variation in that Austronesian-speaking and Papuan-speaking populations showed different haplogroup frequencies. She also showed that the largest proportion of Y chromosome variation occurred within populations, and that populations in the Bismarck Archipelago were extremely diverse when compared to other populations in Near and Remote Oceania, as well as to those in Southeast Asia (Scheinfeldt, 2004; Scheinfeldt et al., 2006; Scheinfeldt et al., 2007).

However, other studies have demonstrated that Y chromosome diversity is greatly reduced in West New Guinea (Irian Jaya) when compared to diversity levels in Papua New Guinea, as well as other populations in Melanesia and Asia. The low levels of NRY diversity contradicted the high levels of mtDNA diversity in the same region. It was suggested that these patterns reflect the long term practice of patrilocality in West New Guinea or biased male reproductive success due to polygyny (Kayser et al., 2003).

There are several major NRY haplogroups present in Melanesian populations, including C, K, M and O.

Macrohaplogroup C: Macrohaplogroup C is believed to be the oldest lineage in the Pacific. It was most likely introduced with the original settlers of Sunda and Sahul, some 59,000 years ago (Underhill et al., 2001c; Underhill et al., 2000). Haplogroup C also shows considerable substructure. For example, the C2-M38 branch shows the highest frequency in eastern Indonesia and coastal New Guinea, while the C2-M208 branch occurs at high frequencies in the highlands of West Papua and the Cook islands (Scheinfeldt et al., 2006).

Macrohaplogroup K: Macrohaplogroup K is the most common lineage in Northern Island Melanesia. The K-M9 branch is found in Wallacea, the Trobriand Islands, New Britain, New Ireland, and Fiji. It is defined by the presence of the M9 SNP and the absence of any other haplogroup-defining mutations (Underhill et al., 2001c). It is also a common lineage that is found widely across Eurasia, the Americas and Australia (Underhill et al., 2001c). The remaining K branches are indigenous to Melanesia and originated ~32,000-50,000 BP (Scheinfeldt et al., 2006; Scheinfeldt et al., 2007). The K5-M230 lineage is found at highest frequencies in highland Papua New Guinea, and decreases in frequency across Northern Island Melanesia, suggesting its origin in New Guinea. The K6-P79 branch is found frequently on New Britain, New Ireland and Mussau, and therefore has a likely origin in New Britain. The K7-P117 lineage is found almost exclusively on New Britain, where it likely originated (Scheinfeldt et al., 2006). K6-P79 and K7-P117 are newly defined haplogroups, thus other studies have not tested their samples for the variants defining them. Due to this restricted testing, the exact distributions of these haplogroups across Oceania cannot be assured, nor can an origin place be inferred.

Haplogroup M: Haplogroup M is a major branch of the K lineage, and is heavily Melanesian in its distribution. The M-M4 lineage is widely distributed from New Guinea to New Ireland, and has a likely origin on New Guinea. The M2-P87 branch is found almost exclusively on New Ireland and east New Britain. This distribution led Scheinfeldt et al. (2006) to propose a likely origin on New Ireland. The M2a-P22 lineage occurs in high frequencies on Bougainville and New Hanover, but at much lower frequencies on New Ireland and New Britain, and appears to have originated on New Britain (Scheinfeldt et al., 2006).

Haplogroup O: Haplogroup O is ubiquitous in East Asia and is found in low frequencies through Northern Island Melanesia. These lineages decrease in frequency across Southeast Asia, Aboriginal Taiwan, Indonesia and Melanesia (Capelli et al., 2001; Kayser et al., 2000).

Polynesia

There is an east to west gradient in the frequency distribution of NRY haplogroups, and overall NRY diversity is greatly reduced in Polynesian as compared to Melanesian populations. This pattern has been interpreted as reflecting a west to east settlement of Remote Oceania as well as a recent population bottleneck during the colonization event (Kayser et al., 2006; Kayser et al., 2000; Kayser et al., 2001). However, the NRY lineage diversity is greater than the mtDNA diversity in Remote Oceania, especially given the high frequency of haplogroup B4a1a1, suggesting a greater Melanesian component in the founding populations than that detected by mtDNA analyses (Kayser et al., 2006; Kayser et al., 2000; Su et al., 2000; Underhill et al., 2001b).

Studies of NRY diversity in Polynesia suggest a mostly Near Oceania origin of the lineages detected in Remote Oceania (Hurles et al., 2002). Furthermore, the majority of the NRY lineages detected in Polynesia are also found in Near Oceania groups. Kayser and colleagues (2006) compared NRY diversity in 400 Polynesians from eight island groups to that found in over 900 individuals from Melanesia, Southeast Asia, East Asia and Australia. They classified 65.8% of the lineages in their Polynesian sample as being Melanesian in origin. Another study investigated NRY variation in 1,209 males from Southeast Asia, Oceania, southern China, and Taiwan. It found that more than half of the NRY lineages present in Remote Oceania could be traced to Melanesia, and were, therefore, associated with the earliest settlers of the region (Capelli et al., 2001). An earlier

investigation of 551 males from Southeast Asia, Taiwan, Micronesia, Melanesia and Polynesia detected no Taiwanese NRY lineages in Micronesia and Polynesia meaning that all NRY lineages detected could be traced back to Melanesia (Su et al., 2000).

Summary of Genetic Patterns in Oceania

The patterns of genetic variation in Oceania are somewhat contradictory. The mtDNA evidence seems to support a strong Southeast Asian influence throughout Oceania, with the “Polynesian Motif”, or haplogroup B4a1a1, being found in high frequencies in many Melanesian populations and reaching almost complete fixation in many Polynesian populations. In contrast, the NRY data suggest a very weak link between Oceanic and Southeast Asian populations. The majority of lineages found in Melanesia are indigenous to the area, and the majority of the lineages in Polynesia can be traced to Near Oceanic populations.

The discrepancy between the extent of Southeast Asian and Melanesian genetic influence in the maternally and paternally inherited portions of the genome may reflect matrilocal residence patterns in ancestral Polynesian societies (Hage and Marck, 2003). It is also possible that the B4a1a1 lineage, whose precursor originated in Taiwan, may have accumulated its distinctive nested mutations in Indonesia or Melanesia. However, a lack of detailed genetic studies on Indonesian populations prevents a thorough investigation of this hypothesis (Trejaut et al., 2005). The patterns also suggest that Polynesian ancestors originated in Southeast Asia, and moved into Near Oceania where they mixed extensively with local populations before moving into Remote Oceania (Kayser et al., 2006).

Furthermore, the HLA data provides a similar picture to that of the mtDNA data, i.e., a strong East Asian influence in Oceania. On the other hand, globin gene variations suggest little Asian

influence in Oceania and detect a strong Melanesian influence in Polynesian populations. However, these loci may be under selective pressure, hence, their distributions may reflect selective pressures rather than migration patterns per se.

Skeletal and Dental Patterns in Oceania

Analyses of skeletal variation in Oceania also show patterns in the biological variation located there. A stepwise discriminant function analysis of 2,531 male crania from Eastern Asia and Oceania resulted in two major divisions (Pietrusewsky, 1994). The first was formed by the crania from Australia, Tasmania, New Guinea and Melanesia, and the second by populations from East Asia, Southeast Asia and Polynesia. The second cluster includes both the Polynesian samples and an aboriginal Taiwanese sample, which led the authors to conclude that Taiwan's aboriginal inhabitants may have been the ancestral source of the populations in Polynesia (Pietrusewsky and Chang, 2003). Similar studies have found distinctions between an Australo-Melanesian group and all other Asian groups, including the Polynesians, with Polynesians being closest to island Southeast Asians (Pietrusewsky, 1994).

Dental variation in mainland Asia and the Pacific can be divided into two major complexes: Sinodonty and Sundadonty (Scott and Turner, 1988). The Sundadont pattern is described as displaying trait retention and simplification, and is found in populations from Southeast Asia, Polynesia, Micronesia and the Jomon in Japan (Scott and Turner, 1988). The Sinodont pattern is considered a specialized derivative of the Sundadont pattern, and includes trait intensification, such as incisor shoveling. This pattern is found among Chinese, Japanese, Siberian, and New World populations (Scott and Turner, 1988). Australian and Melanesian dental patterns do not fit into either category, and are believed to have evolved locally in these regions soon after their initial settlement. Several studies

have noted differences in the dentition of Australian and Melanesian groups versus Asian and Polynesian groups (Brace and Hinton, 1981; Brace and Hunt, 1990; Hanihara, 1992; Scott and Turner, 1988).

Common Models for the Origins of Remote Pacific Islanders

There is little doubt that the ultimate origin of all modern humans is in Africa. However, the more recent origin of Near and Remote Oceanians is still a topic of debate. After leaving Africa, the ancestors of Pacific Islanders would have had to pass through South or East Asia en route to the Pacific, and all of the major models for the origins of Remote Oceanians propose a movement from west to east into the Pacific. It is generally agreed that there were two major migrations into the Pacific, the initial settlement of the area by modern humans and a later spread associated with agriculture. Lapita pottery and Austronesian languages are associated with the movement of individuals into Remote Oceania. However, despite the wealth of archeological discoveries relating to Lapita, the origins of this culture are still highly debated. The various models for the origins of Remote Pacific Islanders differ mostly in their views of ancestral contributions of Island Southeast Asians and Near Oceanians to Remote Oceanians. This section will present a brief overview of the different models for the origins of the Pacific Islanders.

The Express Train from Taiwan Models

The Express Train to Polynesia model proposes an intrusion of Austronesian-speakers into Near Oceania, followed by their rapid dispersal into Remote Oceania with negligible admixture with indigenous Near Oceanians. Proponents of the Express Train from Taiwan model believe the Lapita

cultural complex originated in Southeast Asia, specifically Taiwan, and that highly mobile groups of Lapita explorers expanded quickly through Melanesia around 3500 BP, carrying with them domesticates such as pigs and dogs, an advanced navigational knowledge and sea-faring skills, red-slipped and dentate stamped pottery, Austronesian languages, and Southeast Asian genetic markers into Polynesia. Thus, the proximate origins for Polynesian lineages should be able to be traced to Southeast Asia using an Express Train to Polynesia model (Diamond, 1988; Diamond, 2001; Jobling et al., 2004; Kirch, 1997).

Several investigations of genetic, morphological, archeological, and linguistic data support the Express Train to Polynesia model. Early genetic studies supported this model by revealing allele frequency differences between Polynesians and Melanesians, reduced genetic variability in Polynesian populations, a lack of Papuan genetic markers in Polynesia, and a detection of migration events using population bottlenecks (Cann and Lum, 2004). Craniometric analyses of Pacific and Asian populations also identified two major biological groups, an Australo-Melanesian and an Asian group. Polynesians fell within the Asian cluster and were found to be most closely related to Island Southeast Asians (Pietrusewsky, 1994).

Archeological evidence of the Express Train to Polynesia model includes the discovery of red-slipped pottery in Southeast Asia, followed by the appearance of red-slipped Lapita pottery in Northeastern Melanesia around 3500 BP, and the rapid spread of Lapita pottery into Polynesia by 3200 BP (Bellwood et al., 1995; Summerhayes, 2007). Recent excavations on the Muassau islands have uncovered large Lapita sites dating to 3600 BP, which corresponds to some of the earliest Lapita sites in Polynesia. This site also shows no evidence of gradual indigenous origins for the Lapita culture and has been used as evidence for the Express Train to Polynesia models (Diamond, 1988).

A phylogenetic tree constructed of 77 Austronesian languages, which reflects historical relationships as well as geographical proximity, has been used as linguistic evidence of the Express Train to Polynesia model. This language tree has been interpreted as being most parsimonious with a rapid dispersal from Southeast Asia to Polynesia, with little Melanesian influence (Gray and Jordan, 2000). However, the Express Train to Polynesia model does not compellingly correspond with all of the available biological, linguistic, and archeological evidence, and it tends to be overly simplified in proposing few or no stops in Melanesia along the journey to Polynesia (Kirch, 1997; Green, 2003).

The Indigenous Melanesian Origin Model

The launch of the Lapita Homeland Project in the 1980s stimulated the proposal of the Indigenous Melanesian Origin model to be tested during the fieldwork portion of the project (Allen, 1984). This model proposes that the Lapita cultural complex originated as an indigenous development of the Bismarck Archipelago, which has been settled for around 33,000 years. Proponents of the Indigenous Melanesian Origin model believe that the sudden appearance of Lapita was not due to the influence of new people moving into the area, but was a complex social and subsistence response to a complex environment over a long period of time (Green, 2003; Kirch, 1997). Variations of this model allow for an integration of internal and external technologies, although the weights of the individual contributions of these internal and external forces are not specified. Other variations of this model stress long-term continuity, and claim that there is no need for the migration of people or ideas into this area at all to account for the appearance of the Lapita cultural complex, which is seen as a wholly indigenous development. Thus, the Indigenous Melanesian Origin model would predict that the proximate origins for Polynesian lineages should be traceable to Melanesia (Kirch, 1997; Green, 2003).

The majority of the archeological, linguistic, and biological data from the South Pacific do not support the Indigenous Melanesian Origin model. While archeologists see continuity in some forms of material culture and subsistence, the stimulus for change is not found within Island Melanesia itself. Furthermore, there is no archeological evidence that Lapita social forms, settlement patterns, or domesticated economy developed from the indigenous culture (Summerhayes, 2004; Summerhayes, 2007). However, a strong local component is recognized, and many of the elements that make up the Lapita cultural complex are present at earlier dates in sites to the west (Spriggs, 1997). In addition, the Indigenous Melanesian Origin model fails to explain current linguistic distributions and patterns of biological variation in Oceania (Kirch, 1997; Green, 2003).

The Slow Boat to the Bismarcks Models

The Slow Boat to the Bismarcks model of Polynesian origins proposes that Remote Oceanians trace their ancestry to Wallacea, with subsequent substantial Melanesian influence occurring afterwards. This model is based heavily upon NRY evidence and were first proposed by Oppenheimer in response to the Express Train group of models (Oppenheimer and Richards, 2001a; Oppenheimer and Richards, 2001b) This model is based on the drowning of the Sunda continental shelf around 15,000-7,000 BP (Cann and Lum, 2004). Proponents of the Slow Boat to the Bismarcks model believe that Austronesian-speakers were forced to leave Island Southeast Asia when sea levels began to rise. Different versions of these models suggest that these individuals voyaged toward northern Melanesia, bringing the Lapita culture with them, or even disseminated from Wallacea in all directions, including a migration back to Taiwan (Oppenheimer and Richards, 2001a; Oppenheimer and Richards, 2001b) . This group of models also states that maritime trading and expansion started before the spread of

Lapita, and that Lapita developed out of technological progression among ancient Austronesian-speaking population in eastern Indonesia. Thus, according to the Slow Boat to the Bismarcks models, the proximate origins for the Lapita culture and Polynesian genetic lineages should be Island Southeast Asia (Green, 2003; Cann and Lum, 2004; Jobling et al., 2004; Oppenheimer and Richards 2001a, Oppenheimer and Richards, 2001b).

Several lines of archeological, linguistic, and genetic evidence have been used to support the Slow Boat to the Bismarck models. Obsidian trade in Melanesia pre-dates Lapita, and early obsidian trade distance and dispersal indicate that navigational skills were fairly advanced before the spread of Lapita, suggesting that these individuals were making regular voyages within this region (Green, 2003). Linguistic evidence includes the lack of deep-branch diversity of Austronesian languages in parts of Southeast Asia, and the proposed 'homeland' of the Austronesian language family as the broad triangle-shaped area formed by Taiwan, Sumatra, and Timor. This is an area where some of the oldest Malayo-Polynesian languages are found, and where no other languages are spoken by contemporary populations (Oppenheimer and Richards, 2001b).

Haplogroup B4a1a1, the 'Polynesian motif', predominates in Polynesia, coastal Melanesia, and Wallacea. However, this lineage is virtually absent to the west of Wallace's Line, indicating an ancestral Polynesian origin in Wallacea (Ballinger et al., 1992; Harihara et al., 1992; Horai and Matsunaga, 1986). However, the argument supporting an Island Southeast Asian model of Remote Oceanians tends to disregard a substantial body of archeological evidence documenting the slow spread of domesticated animals, farming, and Neolithic tools and pottery from Taiwan toward Island Southeast Asia, indicating a pre-Island Southeast Asia origin (Green, 2003). Genetic sampling in Indonesia is spotty to nonexistent, which means that more research is required to determine if B4a1a1

is found west of Wallace's Line. Furthermore, if genetic evidence is to be used in support of a Slow Boat to the Bismarcks model of Remote Oceanic origins, it must be based on several loci and integrated with other forms of evidence (Cann and Lum, 2004).

The Triple I Model

A model blending aspects of the Slow Boat models and the Indigenous Melanesian Origin model was created by Green (1991) to attempt to explain the archeological, linguistic, and biological records of the South Pacific. This model suggests significant contributions from both Austronesian-speaking populations from Southeast Asia and indigenous Papuan-speaking Melanesians. The Triple I model of Polynesian origins stands for recognition of the processes of intrusion, innovation, and integration in the emergence of the Lapita cultural complex. Proponents of this model believe that some aspects of Lapita must be accounted for by the intrusion of Austronesian-speakers into Melanesia from Southeast Asia, who brought aspects of their material culture with them. Indigenous populations in the Bismarcks were then integrated into a new culture, and elements of their indigenous culture were also integrated into this new cultural pattern. Finally, innovations and new developments within the Bismarck Archipelago occurred as Lapita remained in the Bismarcks for several hundred years before spreading into Remote Oceania. Thus, according to the Triple I model, genetic lineages in Polynesia should have a significant Southeast Asia and Melanesian ancestry (Kirch, 1997; Jobling et al., 2004; Kirch, 2000; Green, 2003).

The archeological evidence used in favor of the Triple I model indicates a movement of people, ideas, and culture from Southeast Asia into the Bismarcks. The archeological data also suggest that this cultural complex then resided in the Bismarck Archipelago for around 300 years before spreading

into Polynesia, during which time it acquired elements of the local culture and created new elements (Summerhayes, 2004). Further evidence in favor of this model comes from genetic analysis of commensal animals. The Pacific rat (*Rattus exulans*) was transported as a food resource with the colonizing Lapita people, and its skeletal remains first appear in sites associated with the Lapita cultural complex (Matisoo-Smith and Robins, 2004). The Pacific rat originated in Southeast Asia and has not been identified in Near Oceania before the Holocene. Analysis of mtDNA HVSI sequences from both archeological and contemporary rats from Island Southeast Asia and Polynesia show three major haplogroups (I, II and III). The haplogroup phylogenies display geographic patterning, with no haplogroup III lineages being detected in Near Oceania. This rat species is not found in Taiwan, and the distribution of haplogroup II indicates an interaction sphere in the region encompassing the region between the Philippines, Indonesia, and the Solomon Islands. The haplogroup III lineages, therefore, appear to be an intrusive element of the Lapita cultural complex (Matisoo-Smith and Robins, 2004).

However, attempts must be made to distinguish the intrusive elements of the Lapita culture from Southeast Asia from those elements that were acquired through contact with local Melanesian communities, and from those elements that are completely new to each culture. While most agree on the traits that form the Lapita cultural complex, it is still debated as to whether these traits were unique to the earliest Lapita sites in the Bismarcks. This is a daunting task in South Pacific archeology because of the small number of Southeast Asian and Melanesian sites that are as old as or older than Lapita and that lack evidence of the Lapita cultural complex. Without knowing the traits that were present at pre-Lapita sites, it is impossible to distinguish between intrusive, innovative, and integrated features (Terrell and Welsch, 1997).

The Entangled Bank Model

The Entangled Bank model of Polynesian origins was first proposed by Terrell (1988) as a response to what he believed to be overly simplified models for the origins of the Polynesians. He viewed the two peoples (Papuan and Austronesians) and two periods (Pleistocene and Holocene) models of movement into Remote Oceania too simple and based on unrealistic assumptions (Terrell, 1988; Terrell et al., 2001). Terrell instead believes that the current models of Polynesian origins are based on dated assumptions that: 1) populations are historically enduring collective entities; 2) different human populations do have separate origins; and 3) some biological traits are stable enough over time that they can tell us about the beginnings of identifiable corporate human groups rather than about their later history (Terrell et al., 2001).

The Entangled Bank model claims that the settlement of Remote Oceania was the culmination of at least 40,000 years of interactions among peoples inhabiting Near Oceania, and that the history of this area was so complex that it may be difficult or impossible to disentangle past migrations from contemporary data. This model allows for introductions and influences from Asia, but believes that no discernable biological signal could be identified due to the long term complexities of human interactions in this area (Terrell et al., 2001).

Archeological evidence for the Entangled Bank model include the discovery of non-Lapita pottery traditions in Melanesia that are as old or older than Lapita, indicating that there were several related, but not identical, early pottery styles in this region. Obsidian found at archeological sites in this region also points to long range trading or interaction spheres pre-dating Lapita, and a change in obsidian importing and utilization over time (Terrell and Welsch, 1997). However, this model assumes that diverse groups of individuals from different language groups interacted with each other long

enough to produce a new cultural complex, yet were still able to maintain separate pathways of local adaptation and culture change. In addition, evaluating the fit of this model to the genetic data may be problematic because of the difficulty in detecting low-frequency lineages in the populations, which usually require sampling of around 50 unrelated individuals. Thus, small sample sets from groups of individuals in Polynesia may not detect Melanesian lineages that occur in only 1-5% of the population (Cann and Lum, 2004).

Evaluating the Models for the Origins of Remote Pacific Islanders

When placing the different models for the origins of Remote Oceanians into perspective, it is essential to keep in mind that the colonization of an area as big as Oceania is a large-scale phenomenon. A common characteristic of models created to describe the past is that they tend to lack sufficient complexity to describe complicated events. One model may not fit all of the evidence recovered from Oceania; likewise, one model may fit the linguistic data better while another describes the biological patterns best. The spread of Austronesian-speakers and the Lapita cultural complex was most likely a complicated event, with different elements in this complex having their own unique histories.

Furthermore, many of these models use contemporary evidence to attempt to reconstruct the past. However, they only describe the contributions of the first settlers of Remote Oceania and do not take into account post-settlement contact, which would have played a significant role in shaping patterns of modern diversity (Terrell and Welsch, 1997). Nonetheless, direct observation of the past is impossible and thus, models must be created to attempt to reconstruct the past.

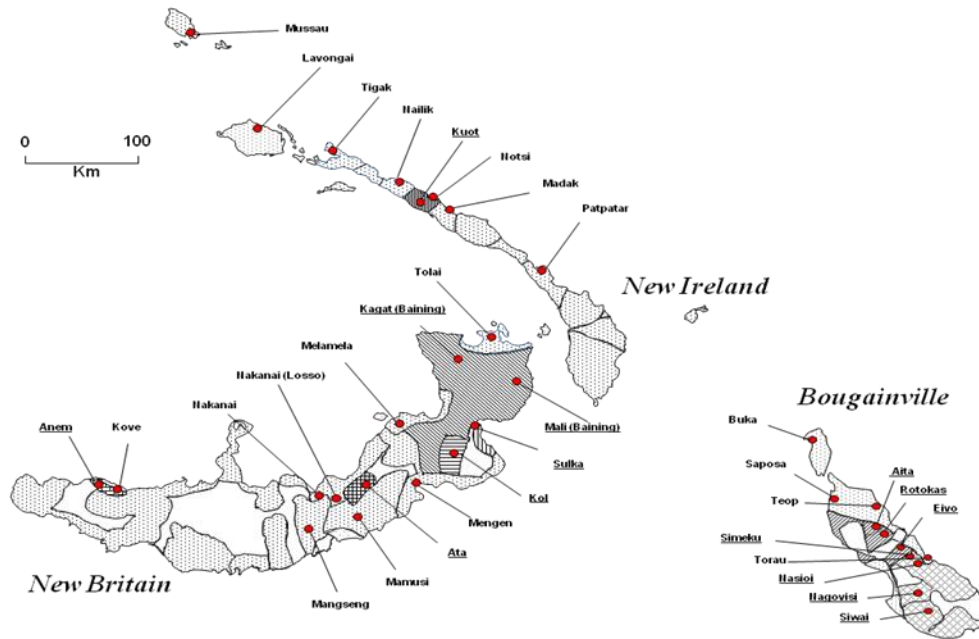
This research will focus on the genetic evidence pertaining to the peopling of Oceania, specifically the NRY data. Currently, the genetic data has produced conflicting results. Many investigators believe the mtDNA data supports the Express Train to Polynesia Model, although recent concerns raised by Friedlaender et al (2007a, 2007b) may change this opinion. The NRY data suggest a greater Melanesian contribution to Polynesian populations, which is more supportive of the Slow Boat to the Bismarcks or Triple I Model. Furthermore, the addition of a large number of samples and new markers analyzed in this study may aid in clarifying the genetic landscape of Near Oceania, as well as possibly detecting these new markers and their contribution to the genetic diversity on the Remote Oceanic island of Easter Island.

CHAPTER 4 MATERIALS AND METHODS

The Biological Samples

Investigating NRY variation for Northern Island Melanesia was accomplished by analyzing the DNA in the samples outlined in Table 4.1 with the appropriate IRB approval from Papua New Guinea (Government of Papua New Guinea – Medical Research Advisory Committee) and Temple University. A total of 842 samples of unrelated males were utilized by this study, and represent a geographically and linguistically diverse panel. The individuals originate from nine islands, and include representatives from Melanesia, Micronesia and Polynesia (Figure 4.1). At least 34 languages are spoken by the individuals for whom we have detailed linguistic information, and an even greater number of dialects are spoken by them. The languages belong to both the Austronesian and Papuan language families (Table 4.1).

Figure 4.1: Language Families in Northern Island Melanesia



The samples from New Guinea were collected by J. Friedlaender of Temple University (38) and by J. and K. Kidd of Yale University (14). The samples from New Britain (427), New Ireland (77), Mussau (20), Lavongai (75) and Bougainville (91) were collected by J. Friedlaender. The samples from Manus Island (60) were collected by G. Koki of the PNG Institute for Medical Research. The samples from Easter Island (28) were collected by M. Shanfield of George Washington University, and the Pohnpei (Micronesian) samples (12) were collected by J. and K. Kidd.

The actual number of samples collected during each field season was much larger than that utilized in this study. The current analysis does not include any females or any males patrilineally related within two generations. Related individuals are more likely to be genetically similar and have alleles that are identical by descent, hence would skew the statistical analyses of their genetic data.

For several reasons, this is a valuable sample set. First, it represents an intensive sampling of Northern Island populations and includes linguistic and genealogical information for the majority of the individuals. Second, it provides an opportunity to conduct an extensive study of NRY diversity across a wide geographic expanse and compare/contrast this diversity with previously collected mtDNA and autosomal diversity data. Finally, it includes regions that have been previously neglected in NRY genetic analyses, such as Manus Island.

Table 4.1: Summary of Samples for NRY Analysis by Island and Language

Island/Region	Language (Dialect)	Language Family	Number of Samples
PNG	PNG Coast	Austronesian and Papuan	28
	PNG Highlands	Papuan	21
	PNG	Austronesian and Papuan	3
New Britain	Anem (Various), West New Britain	Papuan	34
	Ata (Various), West New Britain	Austronesian	45
	Baining (Malasait), East New Britain	Papuan	19
	Baining (Marabu), East New Britain	Papuan	24
	Baining (Rangulit), East New Britain	Papuan	20
	Kol, West New Britain	Papuan	37
	Kove, West New Britain	Austronesian	24
	Mamusi (Various), West New Britain	Austronesian	43
	Mangseng, West New Britain	Austronesian	11
	Melamela, West New Britain	Austronesian	14
	Mengen, West New Britain	Austronesian	23
New Ireland	Nakanai (Various), West New Britain	Austronesian	51
	Sulka (Various), East New Britain	Papuan	33
	Tolai (Various), East New Britain	Austronesian	49
	Kuot (Various), New Ireland	Papuan	32
	Madak, North New Ireland	Austronesian	25
Mussau	Nailik, North New Ireland	Austronesian	17
	Notsi, North New Ireland	Austronesian	14
	Tigak, North New Ireland	Austronesian	21
	Mussau Island	Austronesian	20
Lavongai (New Hanover)	Lavongai (North), Lavongai	Austronesian	28
	Lavongai (South/West), Lavongai	Austronesian	15
Bougainville	Teop, North Bougainville	Austronesian	18
	Saposa, North Bougainville	Austronesian	26
	Buka, North Bougainville	Austronesian	10
	Aita, Central Bougainville	Papuan	18
	Nasioi, South Bougainville	Papuan	11
	Nagovisi/Siwai, South Bougainville	Papuan	8
Manus Island	Manus Island	Austronesian	60
Micronesian	Micronesians		12
Easter Island	Easter Island	Austronesian	28
	Total Austronesian		542
	Total Papuan		257
	Total Mixed, Undesignated or Unknown		43
	TOTAL		842

Isolation of Genomic DNA

DNA was isolated using the QIAamp DNA Mini Kit (Qiagen, Valencia, CA) for the samples consisting of plasma and buccal swabs, and the Puregene Genomic DNA Purification Kit (Gentra Systems, Minneapolis, MN) for the samples consisting of blood and buffy coats as per manufacturer's instructions. DNA isolation was conducted at the Coriell Institute for Medical Research (Camden, NJ).

Genetic Analysis

A panel of 25 polymorphic NRY SNPs previously shown to be informative in the region was used in this study (Table 4.2)¹. These include: M9, RPS4Y711, M38, M208, M4, M230, P117, SRY9138, M16, M74, M122, 50f2/c, M175, M89, P34, P79, P87, P22, M119, and P55 (Bergen et al., 1999; Hammer et al., 2001; Jobling et al., 1996; Karafet et al., 2005; Kayser et al., 2003; Scheinfeldt et al., 2006; Underhill et al., 2001c; Underhill et al., 1997; Underhill et al., 2000; Whitfield et al., 1995). This study utilizes the current haplogroup definitions as defined by the YCC (YCC, 2002) with new markers or revisions as referenced. The primers were purchased from Invitrogen (Carlsbad, CA).

Specifically, this study analyzed a total of 842 samples. Of these, 552 samples represent a previously tested sub-set of samples (Scheinfeldt, 2004) in which the new lineages (K6, K7, M1, M2 and M2a1) were identified during this study. The current study also expanded the investigation of the 50f2/c deletion to all samples in this sub-set. The remaining 290 newly collected samples were analyzed for the entire battery of NRY markers.

¹ The references are for the actual NRY markers. This study used different primers for some of the markers in order to target a smaller amplicon size.

Polymerase chain reaction (PCR) amplification was utilized to amplify the NRY loci. Amplification reactions were carried out in a total volume of 25 μ l and contained 2.5 μ l of buffer, 2.5 μ l of $MgCl_2$, 2.0 μ l of dNTP mix, 0.3 μ l of each primer, 0.3 μ l of Taq polymerase, 15.1 μ l of sterile water, and 2 μ l of sample. The thermal cycling profile followed standard amplification protocols. Amplification products were verified on a 6% acrylamide gel stained with 1% ethidium bromide.

Restriction Fragment Length Polymorphism Analysis

Restriction fragment polymorphism analysis was used for 11 biallelic markers: M9, RPS4Y711, M38, M208, M4, M230, P117, SRY9138, M16, M74 and M122. The restriction enzymes were purchased from New England Biolabs (Beverly, MA) and Promega (Madison, WI). Digest reactions consisted of 1 μ l of enzyme, 1 μ l of sterile water, 3 μ l of buffer and 22 μ l of the PCR product, which was incubated for 6 hours at the temperatures specified by the manufacturer. The results were verified on a 6% acrylamide gel stained with 1% ethidium bromide.

Table 4.2: NRY Markers and Parameters for Amplification Analysis

NRY Marker	Forward Primer	Reverse Primer	Size (bp)	Site (bp)	Mutation	Annealing Temperature	Restriction Enzyme	References
RPS4Y711	tatcctctctattgcag	ccacaagggggaataaaca	208	40	C → T	58	Bsl I	Bergen et al., 1999
M4	tcctaggattatgatacagcgg	ggcacaagctgttcagtaacata	163	88	A → G	60	Nde I	Underhill et al., 1997
M9	gcagcatalaataaaccttcagg	caaacctaaccttgcctcaagc	341	68	C → G	60	Hinf I	Underhill et al., 1997
M16	atattgttattgctcattgaacccagg	ctatttaaaagctgacaaatccaa	105	45	C → A	60	Mnl I	Underhill et al., 1997
M38	cagtttttagagaataatgtoct	ttaagaaaagaaaagcagatg	337	146	T → G	60	HpyCH4 III	Underhill et al., 2000
M74	aactagaaaagctgaaaaataatcaga	gctgctgttcttttaagtaacttact	151	103	G → A	56	Rsa I	Underhill et al., 2001c
M89	agaagcagattgattgccact	tcagttaggagatcccctcatgacc	527	347	C → T	60		Underhill et al., 2001c
M119	gagcttggactttaggacgg	ttcacacaatatacaagatgattctt		224	A → C	60		Underhill et al., 2000
M122	gocctttggaaatgaataatacaag	gagtcactgctctgttagaaaagat: 110	110	54	T → C	58	Hsp92 II	Underhill et al., 2000
M175	ttgagcaagaaaatagtagccca	tgataccttttctgttcattc	114	84	5 bp del	60		Underhill et al., 2001c
M208	ataaatacaaaatcaccctgatggat	ttaaacagaaaatagtaacaaaa	507	352	C → T	60	Taq I	Underhill et al., 2001c
M230	gattttacaatatatacatggcca	acatttattagttaaatctcattgc	164	81	T → A	56	Tsp509 I	Kayser et al., 2003
P22	gaactgtoggaggcaat	gatacacttccctttagtgg		163	G → A	56		Hammer et al., 2001
P34	agggagatgagaagacac	ctggcaattgtcatgtct	661	512	G → A	60		Karafet et al., 2005
P55	tcatacctattggatttctc	tgctctgatacagggtgg	367	142	C → A	60		Hammer (unpublished)
P79	ttgttgcctgacctg	ggaacactattcagcc	301	258	T → C	56		Scheinfeldt et al., 2006
P87	cacigttaacctatgtctgc	ttgagtaaggctctctgag	184	64	A → C	60		Scheinfeldt et al., 2006
P117	ctgattattctttctaccttg	taacatgtaaaaaccctgtc	253	68	G → C	58	Mnl I	Scheinfeldt et al., 2006
SRY9138	gttgatattatattatagaggc	acctacatgcaaaaataaaggatg	72	40	C → T	58	Hae III	Whitfield et al., 1995
50F2/c	ctcaagctaggacaagaaggaaagg	gaggtatgctgaaagcggatag	196	entire	absent	58		Jobling et al., 1996

Insertion/Deletion Analysis

Two insertion/deletion markers were analyzed in this study: 50f2/c and M175. PCR amplification was conducted as previously described, and results were verified on a 6% acrylamide gel stained with 1% ethidium bromide. For each 50f2/c sample, a control band of 1.7-2.7 kb was present. The absence of a 196 bp band indicated that the individual had the deletion, and the presence of the 196 bp band indicated the individual did not possess it. For the M175 samples, the presence of a 114 bp band indicated that the individual lacked the deletion, and the presence of a 109 bp band indicated that the individual had it.

Sequence Analysis

Sequence analysis was used for 8 biallelic markers: M89, P34, P79, P87, M16, P22, M119, and P55 (Table 4.2). PCR amplification was conducted as previously described. The samples were then purified for sequencing using the QIAquick PCR Purification Kit as per manufacturer's instructions (Qiagen). Sequencing was conducted using the BigDye 1.1 Reaction Kit as per manufacturer's instructions (Applied Biosystems, Foster City, CA). Sequences were then purified using the Terminator Dye Removal Kit as per manufacturer's instructions (Qiagen). Sequence data was collected at Coriell Institute for Medical Research on a 3730 DNA Analyzer (Applied Biosystems). Sequence alignment was conducted using Sequencher version 4.1 software (Gene Codes, Inc.).

Statistical Analysis

The statistical analyses were conducted using algorithms available in Arlequin 3.1 software (Excoffier and Schneider, 2005). Genetic structure of populations was analyzed using Analysis of

Molecular Variance (AMOVA). More specifically, AMOVA analyses were conducted to determine if the patterns of variation were subdivided by island or language group. Only populations with 10 or more individuals were included in the AMOVA analysis. This excluded the following populations: PNG and the Nagovisi/Siwai. The Easter Island samples were also removed from the AMOVA analysis because of poor amplification, and therefore unreliable haplogroup affiliation. Furthermore, for the language analysis, the PNG Highlands samples were removed because these individuals included a mixed Papuan-Austronesian sample, and the Micronesian samples were removed because of having an unknown language affiliation.

Measures of Genetic Diversity

Three different tests were used to analyze the amount of NRY variation within the populations. These include gene diversity (Excoffier and Schneider, 2005; Nei, 1987), mean number of pairwise differences (Excoffier and Schneider, 2005; Tajima, 1983; Tajima, 1993), and average gene diversity over loci (Excoffier and Schneider, 2005; Nei, 1987; Tajima, 1983). Gene diversity is the probability that two randomly chosen haplogroups are different in the sample. It is essentially equivalent to the expected heterozygosity for diploid data. A gene diversity value of zero indicates that all haplogroup profiles are identical, while a value of one indicates no shared haplogroup profiles. The mean number of pairwise differences is the average of observed differences between haplogroups of pairs of individuals. Average gene diversity over loci is the probability that two randomly chosen haplogroups are different.

Phylogenetic Tree Construction

A phylogenetic tree of the major haplogroups detected in this study was constructed using Network 4.5.0.0 software (<http://www.fluxus-technology.com/>). The tree was created using the median joining algorithm, with no character weighting or permutations (Bandelt et al., 1999; Bandelt et al., 1995).

CHAPTER 5 RESULTS

This chapter presents the results of the NRY haplogroup analyses of 842 unrelated male samples from Oceania. The geographic and linguistic information for the samples is summarized in Table 4.1. This study detected a total of 17 haplogroups in the current sample. In addition, four samples were tentatively assigned to haplogroup F, although more detailed assignments were not determined in this study. The first section of this chapter describes the NRY haplogroups (Figure 5.1), and presents haplogroup distributions (Tables 5.1 and 5.3) and frequencies (Table 5.2). This is followed by a description of the linguistic and geographic distribution of the haplotypes, and a phylogenetic tree of the major NRY haplogroups in Northern Island Melanesia. Finally, this chapter presents the statistical evaluations of population structure based on NRY diversity.

Haplogroup Frequencies

The number of individuals assigned to each haplogroup can be found in Table 5.1 and their frequencies are presented in Table 5.2. A total of 17 haplogroups were found in this sample. Haplogroup K was the most common lineage, being detected on every island, and in every population that was tested. Haplogroups M2a* and K6 were the second and third most frequent haplogroups, respectively. Haplogroups C*, C2*, C2b, C4, K5, K7, M*, M2a1, O*, O1a, O3 and P were all present in varying frequencies. Haplogroups C4, K1 and M1 were only found in one individual each.

The 50f2/c deletion (Table 5.3) was detected in a total of 71 individuals, or 8% of the samples. However, this deletion was restricted to the K-M9*, K6 and K7 sub-lineages of macrohaplogroup K and

was absent from all other haplogroup backgrounds. Of the 71 individuals with the deletion, 48% belonged to K-M9*, 48% to K6 and 4% to K7.

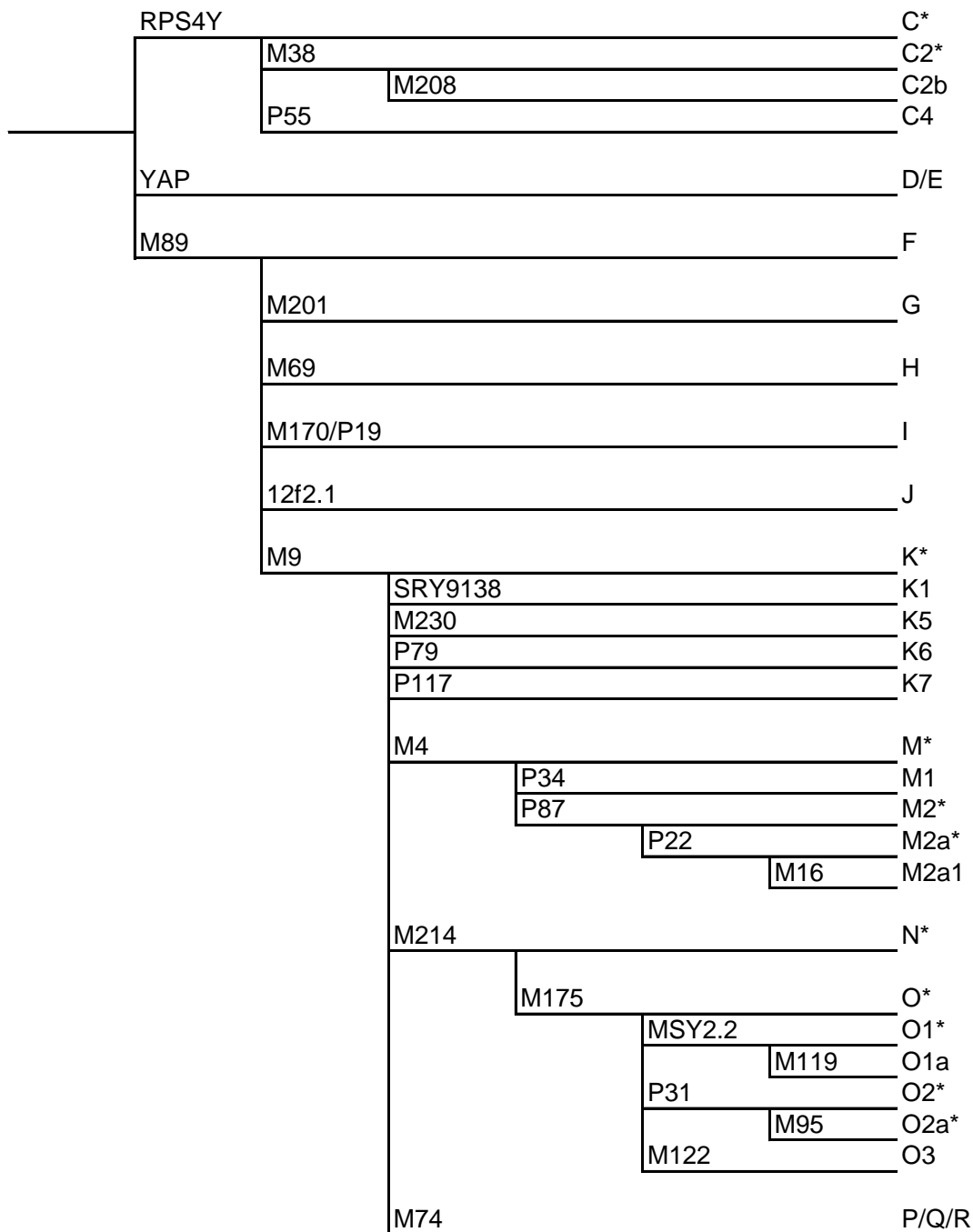
Resolving the position of the 50f2/c deletion within the NRY tree has proven problematic in the past. Scheinfeldt (2004) viewed the 50f2/c deletion as defining a distinct sub-lineage of K-M9*. However, it is now apparent that this deletion appears on the background of several different haplogroups within K (K-M9*, K6 and K7). A study by Jobling and colleagues (1996) observed the 50f2/c deletion in 6% of the males they sampled from a world-wide population sample set. They found that the deletion was restricted to Asia, Australia, southern Europe and northern Europe, and concluded that it was likely the result of multiple mutations and, therefore, not very informative of ancestral relationships. Scheinfeldt (2004) argued that the appearance of the deletion within a region, such as Melanesia, is likely to delineate a particular haplogroup within a larger macrohaplogroup. However, the inclusion of the new K6 and K7 SNPs reveals that this statement is not true.

Scheinfeldt (2004) also concluded that the 50f2/c deletion was geographically restricted to New Britain. However, the current study detected the 50f2/c deletion outside of the island of New Britain on the islands of PNG (1%), Bougainville (4%) and Manus (7%), and in both Papuan and Austronesian populations. The majority of the 50f2/c positive individuals resided on New Britain (87%), suggesting that the increased sample size between Scheinfeldt (2004) and this study allowed the detection of this mutation in a greater geographic area. On the other hand, the deletion was only detected on the backgrounds of K-M9*, K6 and K7, which agrees with previous studies detecting it only within haplogroup K in Melanesia.

Table 5.2: Y-Chromosome Haplogroup Frequencies in 842 Unrelated Males from Oceania

NRY Subhaplogroup	NRY Macrohaplogroup	Island	New Britain													New Ireland					Bougainville																																																																																																																																			
			New Britain													New Ireland					Bougainville																																																																																																																																			
C2*	C	PNG	PNG Coast	0.07	0.24	0.67	0.15	0.29	0.11	0.04	0.03	0.1	0.03	0.08	0.06	0.08	0.12	0.14	0.05	0.06	0.33	0.54	0.17	0.53	0.42	0.14	0.01																																																																																																																													
C2b																												PNG Highlands	0.28	0.48	0.09	0.04	0.05	0.09	0.04	0.05	0.09	0.04	0.04	0.06	0.16	0.24	0.16	0.07	0.1	0.25	0.14	0.07	0.1	0.08	0.02	0.07																																																																																																				
C4																																																					PNG	0.33	0.07	0.26	0.24	0.51	0.2	0.04	0.04	0.04	0.06	0.16	0.24	0.16	0.07	0.1	0.25	0.14	0.07	0.1	0.08	0.02	0.07																																																																													
Totals																																																																												0.07	0.24	0.67	0.15	0.29	0.11	0.04	0.03	0.1	0.03	0.08	0.06	0.08	0.12	0.14	0.05	0.06	0.33	0.54	0.17	0.53	0.42	0.14	0.01																																																					
K-M9*	K	Island	Language (Dialect)	0.11	0.24	0.05	0.15	0.04	0.27	0.29	0.02	0.02	0.02	0.02	0.31	0.24	0.16	0.41	0.16	0.08	0.09	0.05	0.09	0.18	0.27	0.04	0.07																																																																																																																													
K1																												PNG	0.32	0.48	0.09	0.04	0.05	0.09	0.04	0.05	0.09	0.04	0.04	0.06	0.16	0.24	0.16	0.07	0.1	0.25	0.14	0.07	0.1	0.08	0.02	0.07																																																																																																				
K5																																																					PNG Coast	0.07	0.07	0.06	0.27	0.42	0.04	0.4	0.4	0.27	0.36	0.39	0.25	0.18	0.16	0.19	0.16	0.29	0.09	0.5	0.14	0.04	0.04	0.06	0.08	0.15																																																																										
K6																																																																															PNG Highlands	0.47	0.32	0.33	0.15	0.03	0.04	0.09	0.36	0.17	0.16	0.12	0.02	0.08	0.08	0.09	0.05	0.09	0.05	0.09	0.18	0.27	0.13	0.04	0.07																																																	
K7																																																																																																								PNG	0.11	0.24	0.05	0.02	0.02	0.02	0.31	0.24	0.16	0.41	0.16	0.08	0.09	0.05	0.09	0.05	0.09	0.05	0.09	0.18	0.27	0.13	0.04	0.07																								
M*																																																																																																																																	PNG Coast	0.04	0.04	0.03	0.11	0.33	0.33	0.11	0.33	0.08	0.26	0.09	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08
M1																																																																																																																																																								
M2*	K	Island	Language (Dialect)	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04																																																																																																																													
M2a*																												PNG	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04																																																																																																				
M2a1	K	Island	Language (Dialect)	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04																																																																																																																														
O*	K	Island	Language (Dialect)	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04																																																																																																																													
O1*																												K	Island	Language (Dialect)	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04																																																																																																				
O1a																																																					PNG	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04																																																																												
O3	K	Island	Language (Dialect)	0.17	0.21	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1																																																																																																																													
P																												PNG	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04																																																																																																				
F	K	Island	Language (Dialect)	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04																																																																																																																													

Figure 5.1: Detailed NRY Tree



Geographic and Linguistic Distributions of the Haplogroups

Macrohaplogroup C

Haplogroup C is believed to be the oldest lineage in the Pacific, with an origin of ~50,000 BP (Underhill et al., 2001c; Underhill et al., 2000). It was most likely introduced with the original settlers of Sunda and Sahul. Several lineages within macrohaplogroup C appeared in this sample, including C*, C2*, C2b and C4. However, none of them were found on the island of New Hanover or in the Micronesians.

Haplogroup C*: This lineage is defined by the RPS4Y711 (M130) SNP, which is also present in all C*-derived haplotypes. In this study, haplogroup C* was restricted to the Austronesian-speakers of Easter Island.

Haplogroup C2*: This lineage is defined by the M38 SNP. Lineage C2* was found at varying frequencies on all of the islands studied, except for New Hanover and Mussau, and was absent in the Micronesians. It occurred at the highest frequency of all the macrohaplogroup C lineages, and appeared at a frequency of 3.9% in the 842 individuals analyzed. Interestingly, it occurred in both Austronesian and Papuan groups, although being found more frequently in the Austronesian-speakers.

Haplogroup C2b: This lineage is defined by the M208 SNP. Lineage C2b was found on all of the islands studied, except for New Hanover, and it was not detected in the Micronesians. It too appeared at varying degrees in both Austronesian and Papuan groups.

Haplogroup C4: This lineage is defined by the P55 SNP. In this study, C4 was restricted to one Papuan-speaking individual from PNG.

Macrohaplogroup K

Macrohaplogroup K is the most common macrohaplogroup in Northern Island Melanesia, as it appears on every island in this study. Its major branch, haplogroup K-M9* is found in Wallacea, the Trobriand Islands, New Britain, New Ireland, and Fiji, and is commonly observed in Eurasia, the Americas and Australia. The remaining K lineages are indigenous to Melanesia and appear to have originated ~32,000-50,000 BP. Among these are K1, K5*, K6 and K7.

Haplogroup K-M9*: This lineage is defined by the M9 SNP. It occurred at the highest frequency of any K lineage identified in this study. It was found on every island and in every population except for those speaking the Rangulit dialect of the Baining, the Buka languages, and the Nasioi. It occurred in both Austronesian and Papuan populations.

Haplogroup K1: This lineage is defined by the SRY9138 SNP. It was restricted to one Nasioi (Papuan) individual from Bougainville.

Haplogroup K5*: This lineage is defined by the M230 SNP. It was found on every island in this study except for Mussau and Easter Islands. It was also found in both Austronesian and Papuan groups.

Haplogroup K6: This lineage is defined by the P79 SNP. It was found on every island in this study except New Hanover and Easter Island. It was also found in both Austronesian and Papuan populations, and the frequency of haplogroup K6 decreased moving east across Northern Island Melanesia.

Haplogroup K7: This lineage was defined by the P117 SNP. It was restricted to the islands of New Britain and Bougainville, and was found in both Austronesian and Papuan groups. The frequency of haplogroup K7 decreased moving east from New Britain to Bougainville.

Haplogroup M

Haplogroup M is a major sub-branch of macrohaplogroup K, and is heavily Melanesian in its distribution. It was found on every island in this study, but does not occur in the Micronesians. There were several lineages of haplogroup M found in this sample, including M*, M1, M2*, M2a* and M2a1.

Haplogroup M*: This lineage is defined by the M4 SNP. It was found on every island, but not in the Micronesians. It was also found in both Austronesian and Papuan groups.

Haplogroup M1: This lineage is defined by the P34 SNP. It was restricted to one Papuan-speaker of the Papua New Guinea highlands.

Haplogroup M2*: This lineage is defined by the P87 SNP. It was found on all of the islands except Mussau, Manus and Easter Islands, and did not occur in the Micronesians. It was found in both Austronesian- and Papuan-speakers.

Haplogroup M2a*: This lineage is defined by the P22 SNP. It was found on every island except Easter Island, and was absent from the Micronesians. This lineage was also found in both Austronesian and Papuan populations.

Haplogroup M2a1: This lineage is defined by the M16 SNP. It was restricted to the Nasioi and Nagovisi (Papuan) of Bougainville.

Haplogroup O

Haplogroup O is ubiquitous in East Asia and found at low frequencies throughout Northern Island Melanesia. It was found at especially low frequencies in this study, and was absent from the islands of Papua New Guinea, Mussau Island, Manus Island and Easter Island. Several lineages of O were observed in this sample, including O*, O1a and O3.

Haplogroup O*: This lineage is defined by the M175 SNP. It was found on the islands of New Britain, New Ireland, New Hanover and Bougainville, as well as in the Micronesians. This lineage occurred in both Austronesian and Papuan groups, but occurred more frequently in the Austronesian-speakers.

Haplogroup O1a: This lineage is defined by the M119 SNP. It was found on the islands of New Britain, New Ireland, New Hanover, Bougainville and Manus. It occurred in both Austronesian and Papuan groups, but was more common in the Austronesian-speakers.

Haplogroup O3: This lineage is defined by the M122 SNP. It was found on the islands of New Britain and Bougainville, and in the Micronesians. It was found in all Austronesian-speakers in Melanesia, and in one Micronesian of unknown language affiliation.

General Geographic Distributions of the Haplogroups

There were distinct differences in haplogroup frequencies between the populations sampled in Melanesia, Micronesia and Polynesia. In general, the Micronesian and Polynesian samples were less diverse. The lineages within macrohaplogroup C predominated the Easter Island sample, while the K and O lineages were most commonly found in the Micronesians. However, the sample sizes from Polynesia and Micronesia were small and may not reflect the true NRY variation within these regions. The largest proportion of the samples collected came from Melanesia, and the haplogroup frequencies were distinct between the islands of New Guinea, New Britain, New Ireland, Bougainville, Mussau Island and Manus Island. New Guinea was the least diverse of these islands followed by Mussau and Manus Islands, while New Britain was the most diverse, followed by Bougainville and New Ireland.

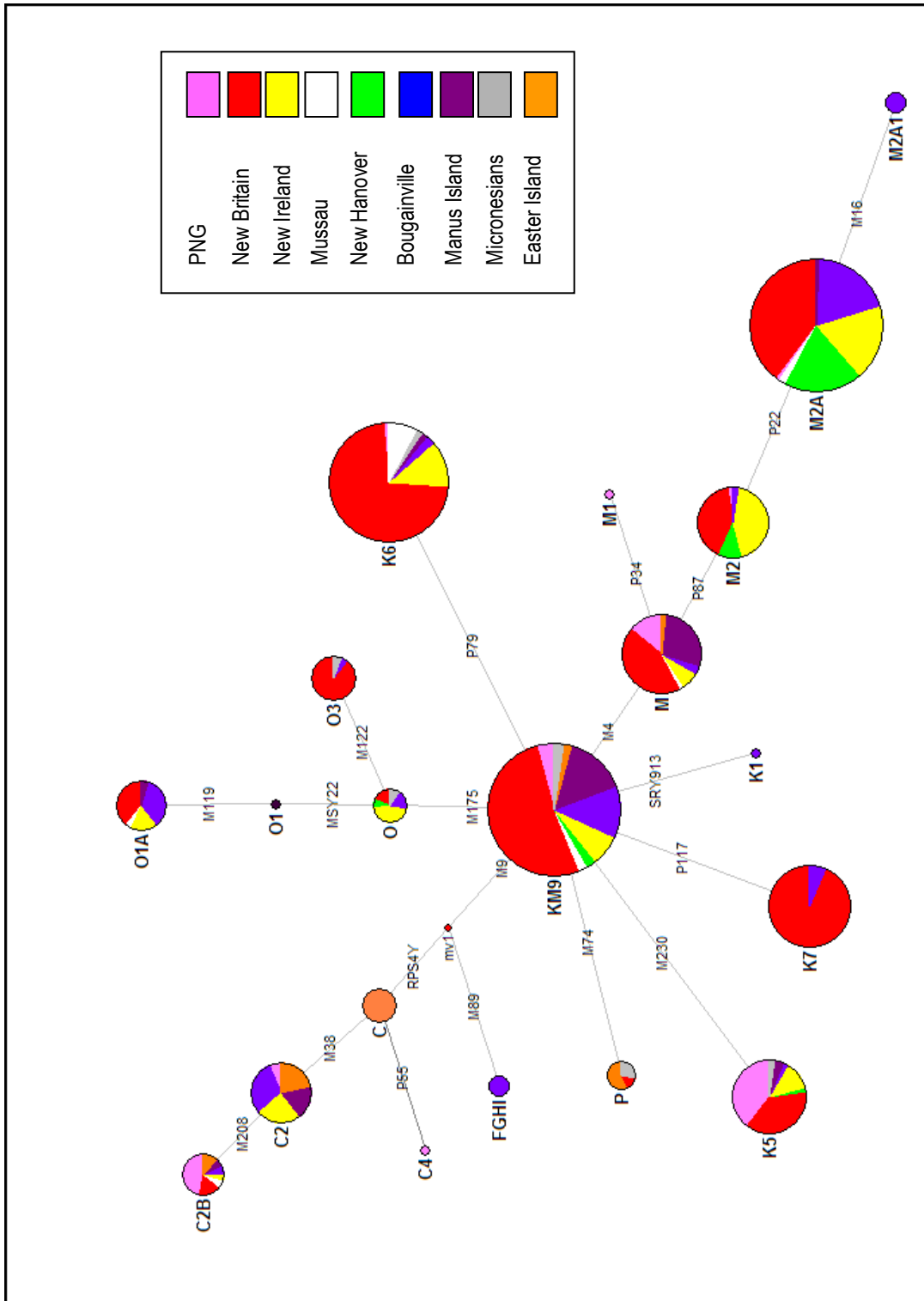
The differences in the distributions of the various lineages likely reflects the long history of isolation and subsequent genetic drift experienced by the island populations.

Network Analysis

The phylogenetic tree (Figure 5.2) is a visual representation of the SNP data presented in Table 5.1. It shows the relationships of the various NRY lineages within Northern Island Melanesia to one another, the frequencies of the lineages based on node size, and the distribution of the lineages by island. In general, the network was star-like and indicates the accumulation of mutations and expansion of these lineages once the populations entered Northern Island Melanesia. Haplogroup K-M9* consisted of the largest group of individuals and was found on every island sampled. Haplogroups M2a* and K6 also contained large numbers of individuals, but their geographic range was more restricted.

The network clearly emphasized that there are still large numbers of individuals clustered into some of the haplogroup categories, namely, K-M9*, K5, K6, K7, M2a*, and M* . This illustrates the need for further efforts to genotype these samples and place them into additional sub-lineages for the large haplogroups, as this will provide a better understanding of the depth and distribution of NRY variation in Northern Island Melanesia.

Figure 5.2: Network of the Major NRY Haplogroups in this Study



Population Structure

AMOVA

An AMOVA was conducted to assess the distribution of genetic variation within and between the defined groups. The groupings tested in this study were categorized in the following ways: (1) geography, (2) language, and (3) population (Table 5.4).

Table 5.4: AMOVA Based on Y-Chromosomal Haplogroups (F_{ST})

Group	Sample size	Number of populations	Number of groups	% Variation between groups	% Variation between populations within groups	% Variation within populations
Geography	792	31	9 islands	6.62	9.58	83.80
Language	752	29	2 languages	0.01	13.99	86.00
Population	792	31	1 population	-	14.52	85.48

AMOVA was used to analyze the structure of NRY variation in the sample set (Table 5.4), and to determine if the patterns of variation clustered by island or language group. Only populations with 10 or more individuals were included in the AMOVA analysis, excluding the following populations: PNG and the Nagovisi/Siwai. The Easter Island samples were also removed from the AMOVA analysis because of poor amplification, and therefore unreliable haplogroup affiliation. Furthermore, for the language analysis the PNG Highlands samples were removed because these individuals included

a mixed Papuan-Austronesian sample, and the Micronesian samples were removed because of their unknown language affiliation.

Over 80% of the variance occurred within the various populations in this analysis, suggesting that all the populations analyzed have high internal SNP diversity. An analysis of genetic variation according to geography showed that 83.8% of the variation was found within populations, 9.6% of the variation was found between populations within groups, and 6.6% was found between the groups at the $p = 0.01$ level of significance. Grouping the samples by language showed that 86% of the variation was found within populations and 14% of variation was found between populations within groups at the $p = 0.01$ level of significance, and 0.01% of the variation was found between groups at a non-significant level ($p = 0.261$). An analysis of variation at the population level showed that 85.48% of the variation was found within populations, and 14.52% of the variation was found between groups at the $p = 0.01$ level of significance. The between group variance was very high, as compared to other worldwide populations, and likely reflects the history of isolation and subsequent drift experienced by these island populations.

These patterns were consistent with earlier studies in this region. Scheinfeldt (2004) recognized 86.52% of the genetic variation in her study occurred within populations and 13.48% was found between populations. While the difference may be small, there was still an increase in the amount of variation found between populations. The increased between population variance between her study and my own may be the result of my using an increased sample size and an increased battery of SNPs in this study. The partitioning of variance within and among human populations in Northern Island Melanesia is consistent with worldwide F_{ST} values, suggesting that there is little evidence of a clear subdivision among major continental groups (Romualdi et al., 2002).

NRY Diversity

Table 5.5: Measures of NRY Diversity

Population	Gene diversity	Mean number of pairwise differences	Average gene diversity over loci
PNG Coast	0.81 +/- 0.05	2.95 +/- 1.59	0.12 +/- 0.07
PNG Highlands	0.70 +/- 0.07	1.05 +/- 0.72	0.04 +/- 0.03
Anem (Various)	0.74 +/- 0.06	1.59 +/- 0.97	0.07 +/- 0.04
Ata (Various)	0.75 +/- 0.03	1.90 +/- 1.10	0.08 +/- 0.05
Baining (Malasait)	0.74 +/- 0.07	1.73 +/- 1.05	0.07 +/- 0.05
Baining (Marabu)	0.74 +/- 0.04	2.03 +/- 1.18	0.08 +/- 0.05
Baining (Rangulit)	0.70 +/- 0.06	1.88 +/- 1.12	0.08 +/- 0.05
Kol	0.61 +/- 0.07	1.19 +/- 0.78	0.05 +/- 0.04
Kove	0.80 +/- 0.05	1.61 +/- 0.99	0.07 +/- 0.05
Mamusi (Various)	0.73 +/- 0.03	2.22 +/- 1.25	0.09 +/- 0.06
Melamela	0.75 +/- 0.07	1.71 +/- 1.06	0.07 +/- 0.05
Mengen	0.77 +/- 0.05	1.79 +/- 1.08	0.07 +/- 0.05
Nakanai (Various)	0.84 +/- 0.02	2.21 +/- 1.24	0.09 +/- 0.06
Sulka (Various)	0.70 +/- 0.07	1.44 +/- 0.90	0.06 +/- 0.04
Tolai (Various)	0.85 +/- 0.02	2.69 +/- 1.46	0.11 +/- 0.07
Mangseng	0.76 +/- 0.11	1.35 +/- 0.90	0.06 +/- 0.04
Kuot (Various)	0.81 +/- 0.03	2.18 +/- 1.24	0.09 +/- 0.06
Madak	0.88 +/- 0.03	2.48 +/- 1.39	0.10 +/- 0.06
Nailik	0.79 +/- 0.08	2.87 +/- 1.59	0.12 +/- 0.07
Notsi	0.86 +/- 0.06	2.93 +/- 1.63	0.12 +/- 0.08
Tigak	0.83 +/- 0.07	2.59 +/- 1.44	0.11 +/- 0.07
Mussau Island	0.69 +/- 0.08	1.80 +/- 1.08	0.08 +/- 0.05
Lavongai (North)	0.48 +/- 0.11	1.26 +/- 0.82	0.05 +/- 0.04
Lavongai (South/West)	0.54 +/- 0.13	1.37 +/- 0.89	0.05 +/- 0.04
Teop	0.84 +/- 0.06	2.71 +/- 1.51	0.11 +/- 0.07
Saposa	0.69 +/- 0.09	1.90 +/- 1.12	0.08 +/- 0.05
Aita	0.29 +/- 0.12	0.88 +/- 0.65	0.04 +/- 0.03
Nasioi	0.75 +/- 0.10	1.67 +/- 1.06	0.07 +/- 0.05
Buka	0.84 +/- 0.10	3.44 +/- 1.92	0.14 +/- 0.09
Manus Island	0.65 +/- 0.05	1.29 +/- 0.82	0.05 +/- 0.04
Micronesians	0.82 +/- 0.10	1.24 +/- 0.84	0.05 +/- 0.04
Easter Island	0.79 +/- 0.04	1.82 +/- 1.08	0.08 +/- 0.05

Overall, haplogroup diversity was high in Northern Island Melanesia. The number of different lineages detected was similar to that found in Southeast Asian populations, and less than that found in Polynesian populations. Within population variation was estimated by calculating the gene diversity, mean number of pairwise differences and average gene diversity (Table 5.5). Gene diversity values can range from 0 to 1, with 0 indicating that all the haplogroups within the population are identical and 1 indicating no shared haplogroups. Population gene diversities ranged from 0.88 ± 0.03 in the Madak (Austronesian) of New Ireland to 0.29 ± 0.12 in the Aita (Papuan) of Bougainville. These values supported the findings of other studies that the coastal populations (like the Madak) tend to be more heterogeneous than the inland populations (like the Aita) (Figure 4.1). These values also differed from those found by Scheinfeldt (2004), illustrating the fact that the use of an increased battery of SNPs and increased sample sizes can affect the statistical results. Scheinfeldt (2004) found the highest gene diversity values (0.90 ± 0.10) in her Manus population, which may be a result of a sample size of only seven. However, her second highest value occurs in her New Ireland sample (0.88 ± 0.06), which was consistent with the values detected in this study. Scheinfeldt's (2004) lowest gene diversity values (0.41 ± 0.15) were in the Sulka (Watwat) population, which is a coastal Papuan population. This result suggested that the location of the population (coastal versus inland) has a greater influence on genetic diversity than its language group.

The mean number of pairwise differences reflects the average number of pairwise differences among any two individuals within a population. The range of mean pairwise differences is dependent upon the number of SNPs analyzed, which are 19 in this study. The mean number of pairwise differences ranged from 3.44 ± 1.92 in the Buka (Austronesian) of Bougainville to 0.88 ± 0.65 in the Aita (Papuan) population of Bougainville. Again, these values suggested that coastal populations were

more genetically heterogeneous than inland populations (Figure 4.1). Scheinfeldt (2004) found the highest mean number of pairwise differences (3.32 ± 1.80) in the Sepik (Papuan) of PNG, and the lowest value (0.50 ± 0.52) in the Kol (Papuan) of New Britain. Both of these populations are coastal dwellers and both speak Papuan languages, suggesting that either location and/or language have little influence on the mean number of pairwise differences. While not enormous, the estimates from the two studies were somewhat different. The likely reason for this distinction is that the increased number of SNPs and increased sample size was producing a better resolution to the genetic picture of Northern Island Melanesia (Scheinfeldt 2004 analyzed 14 SNPs and this study used 19 SNPs).

The average gene diversity reflects the probability that any two randomly chosen individuals within a population will have a different haplogroup. Average gene diversity over loci ranged from 0.14 ± 0.09 in the Buka (Austronesian) of Bougainville to 0.04 ± 0.03 in the Aita (Papuan) of Bougainville and the PNG Highlands population (Papuan). The Buka are coastal Austronesian-speakers, and the Aita and PNG Highlanders are both more isolated inland Papuan-speakers (Figure 4.1). Scheinfeldt (2004) found average gene diversity values that ranged from 0.24 ± 0.14 in the PNG Sepik (Papuan) population to 0.04 ± 0.04 in the Kol (Papuan) of New Britain. Again, both of these populations are coastal dwellers that speak Papuan languages.

CHAPTER 6 DISCUSSION

This study analyzed NRY haplogroup data in order to better resolve the phylogeographic distribution of NRY lineages in Northern Island Melanesia. Within this general objective, a number of specific aims were pursued: The first was to determine the distribution of the newly defined SNPs (K6 and K7), as well as other recently characterized SNPs that have not been extensively analyzed yet (M1, M2 and M2a1) in Northern Island Melanesia. The second was to determine the extent of the paternal contribution of Southeast Asian males to the Northern Island Melanesian populations as indicated by the younger O lineages versus older C lineages. A related goal was to delineate the original lineages brought into the area by the initial settlers from those which arose in and remained indigenous to Island Melanesia (lineages M and K). The third was to determine whether these newly defined SNPs provided evidence of new levels of heterozygosity that might aid in overcoming the ascertainment of bias resulting from surveying Northern Island Melanesian populations with SNPs defined in non-Melanesian populations. A fourth aim was to investigate the distribution of the 50f2/c deletion in this region. The final aim was to compare the regional NRY diversity with the mtDNA diversity in Northern Island Melanesia because of the discrepancies in the male- and female-mediated genetic histories in this region.

Improving the understanding of NRY phylogeography specificity in Northern Island Melanesia is important for addressing questions surrounding the settlement of the South Pacific by identifying potential donor populations to the Polynesian gene pool. This is especially important for the NRY since the majority of Polynesian NRY lineages are Near Oceanic in origin. Better phylogeographic resolution can also provide insights into the effects of island isolation by allowing a comparison of genetic

similarities and differences among populations in the region. This approach also facilitated an examination of the contradictory findings of previously published NRY and mtDNA data as well as issues of sex-specific demographic processes in the region.

This study analyzed a larger battery of SNPs, and increased the number of samples and the coverage of populations within Northern Island Melanesia. Increasing the number of SNPs has successfully subdivided the large K-M9* lineage into smaller sub-lineages, which, in turn, has provided better resolution of NRY variation in this region. Defining new NRY SNPs is an important aspect of understanding the extent of Y-chromosome variation in any population because of the general bias in the analysis of the Y-chromosome sequence in previous studies. As more regional SNPs are detected and analyzed, the ascertainment of bias will slowly be overcome. Because new SNP data are constantly being generated, it is difficult to make comparisons between research papers, as many are almost outdated at the time of their publication. It will be seen in the following discussion that new SNPs have been defined since the laboratory phase of this project began, and thus cannot be used to analyze the sample populations here.

Another effort to overcome the ascertainment of bias was increasing the sample size and geographic coverage in Northern Island Melanesia. Increasing the coverage of Northern Island Melanesian populations has revealed the presence of SNPs and sub-lineages in areas where they were previously unrecognized, and provides a better understanding of the regionalization of NRY variation in the area. Since this is a continual process, future researchers will continue to expand coverage in Northern Island Melanesia, as this is an area of interest to many anthropologists. In fact, several studies published very recently are directly relevant to this study in that they serve to further

increase the number of individuals sampled in Northern Island Melanesia (Cox et al., 2007; Kayser et al., 2008).

This study included an analysis of Manus Island, a location where populations have not been extensively studied at the molecular level. Manus was settled relatively early in the expansion into Melanesia with archeological evidence of a human presence at 21,000 BP (Summerhayes, 2007). The archeological record suggests that Manus was relatively isolated until around 13,000 BP, when there is evidence of the translocation of the cuscus, bandicoot and canarium nut from New Guinea. The island also lacks any flora and fauna from New Britain and New Ireland, suggesting that these large islands were not in contact with Manus populations (Summerhayes, 2007). The early occupation of Manus and later translocations suggests a Pleistocene link with New Guinea, although the predominantly Austronesian languages spoken on Manus also suggest a strong Holocene influence from incoming Lapita peoples.

This study investigated whether there was a genetic link to New Guinea or to other early settled islands in Northern Island Melanesia, or to the intruding Asian populations of the Holocene. Only 1.6% of the Manus samples were assigned to an East Asian haplogroup (O1a), suggesting they had a predominantly Melanesian genetic constitution. All of the haplogroups detected in the Manus sample, including C2*, C2b, K-M9*, K5, K6, M*, M2a*, and O1a, (except the one O1a individual), were also observed on New Guinea. This finding along with the archeological evidence, suggests that there was ongoing contact between the two islands. Two other possible source populations for Manus Island populations are New Ireland and New Hanover. Although all of the haplogroups detected on Manus were also found on New Ireland, the archeological data does not support a close prehistoric

relationship between the two islands. By contrast, the haplogroup frequencies on Manus and New Hanover are very different, with C2*, C2b and M* occurring on Manus but not New Hanover.

A recent study by Kayser and coworkers (2008) investigated NRY variation in the Admiralty Islands, which includes Manus. They detected O-M110 at a frequency of 17.7% in their samples. This marker defines a sub-branch of O1a for which the Northern Island Melanesian samples were not screened. However, we can assume that those individuals are also positive for the mutation defining O1a, which is the highest frequency O lineage in this study sample, because O-M110 is a SNP nested within the O1a background. Kayser et al. (2008) also observed O-M110 to be present at high frequencies in Taiwan (34.1%), Philippines (12.8%) and the Trobriand Islands (17.3%), but at lower frequencies in Island Southeast Asia (2.5-9.7%), New Guinea, Fiji and Tuvalu. O1a has previously been associated with the Austronesian expansion into Remote Oceania, and is believed to have originated in Taiwan (Kayser et al., 2008).

Kayser et al. (2008) also surveyed their samples for the O-M324 SNP, which defines a sub-group of O3. They found this sub-lineage in just one individual from Manus Island, and stated that it is East Asian in origin. While the current study did not test samples for M324, it did screen samples for the O3 marker. All O-M324 individuals should also be O3 positive, as well. O3 was detected in 2% of the entire sample set in this study, and appeared on the islands of New Britain and Bougainville. Believed to have originated in East Asia, O3 is also associated with the Austronesian expansion, and is the major East Asian haplogroup detected in Polynesian samples tested to date (Kayser et al., 2008).

The NRY data presented here allow a comparison between the high levels of variation and structured organization of the NRY with those aspects of the mtDNA data. Both types of genetic data are structured heavily by island and island size, and both show regional distinctions. The larger, more

rugged islands, such as New Britain, Bougainville and New Ireland, are the most diverse, followed by the smaller island populations. This pattern of variation is the result of a coastal versus inland distinction in diversity levels. The more isolated inland groups tend to have less overall diversity than the coastal groups. The main difference in the apportionment of variance between the two loci is related to the lower within population variation of the mtDNA due to the high frequencies of haplogroup B4a1a1. This chapter will return to the discussion of the way in which this study contributes to our understanding of the peopling of Polynesia.

Research Objectives

Haplogroup Distributions

A major goal of this study was to determine the distribution of the lineages defined by the K6 and K7 SNPs, as well as those delineated by the M1, M2* and M2a1 SNPs across Northern Island Melanesia. The second was to determine the extent of the paternal contribution of Southeast Asian males to the Northern Island Melanesian populations as indicated by the younger O lineages versus older C lineages. A related goal was to delineate the original lineages brought into the area by the initial settlers from those which arose in and remained indigenous to Island Melanesia (lineages M and K).

This study has expanded the number of NRY SNPs used to analyze Northern Island Melanesian populations, and has revealed their distribution with the objective of better characterizing the remarkable NRY diversity in that region. These SNPs have provided a better resolution of the NRY variation in Melanesia, and the larger sample size facilitated the detection of SNPs (hence, lineages) in populations where they had not previously been detected. The haplogroups discussed in this section

will be grouped together according to their proposed time and place of origin. Haplogroups K6, K7, M1, M2* and M2a1 will be emphasized by bold print in the following discussion.

Early Melanesian Haplogroups: There are several old mtDNA and NRY lineages in Northern Island Melanesia. For the mtDNA, several branches of macrohaplogroup M arose in Northern Island Melanesia around 40,000 years ago (Forster et al., 2001; Friedlaender et al., 2005b). In addition to these older lineages, the younger B and E lineages are also present. In general, there is a higher frequency of the older lineages in Papua New Guinea, with these decreasing in frequency moving east towards Polynesia. There is also a significant presence of the younger lineages, especially haplogroup B, in Northern Island Melanesia, with the younger lineages reaching very high frequencies moving into Polynesia (Friedlaender et al., 2007b).

Several old Y chromosome lineages also arose in Northern Island Melanesia, as well. These branches of macrohaplogroup K arose around 35,000 years ago in this area (Scheinfeldt et al., 2006). The younger haplogroups are represented by the O lineages. However, there are generally very few young NRY lineages detected in Northern Island Melanesia.

It is important to note that age estimates will be mentioned in the following discussion. Many geneticists agree that such calculations should be evaluated relative to one another, rather than taken as absolute dates. The various publications, from which the dates were taken used different mutation rates in their calculations, thus are not directly comparable.

As generally established in earlier studies, certain haplogroups in the region seem to have originated in Melanesia and spread from there to Near and Remote Oceania. Haplogroup C* is one such lineage, and likely one of the first to be brought into this region as modern humans expanded out

of Africa. Haplogroups C2* and M* appears to have arisen in Melanesia, and haplogroups C2b, K5, and **M1** appear to have originated in New Guinea.

Macrohaplogroup C has several branches that represent very ancient NRY lineages in Melanesia. Haplogroup C* is considered to be one of the oldest lineages in the Southwest Pacific. Interestingly the only place that it was detected in this study was Easter Island. However, the Easter Island samples in this study were collected in the 1950s, and many of them failed amplification for the sub-lineages of C*, which is why they are classified as C*. Thus, re-evaluation of these samples with additional SNPs may place them in a different sub-haplogroup.

C* is believed to be one of the first lineages introduced onto the ancient continent of Sahul, meaning that it traveled across Oceania to Easter Island at the most eastern extent of Remote Oceania. Easter Islanders are Austronesian-speakers, suggesting that haplogroup C* was introduced into the population as people moved east with the Lapita expansion. Cox (2003) also found C* in 18% of the Vanuatu samples. This finding is interesting because Vanuatu is geographically closer to Northern Island Melanesia (where C* was not observed) than to Easter Island (where it was located exclusively in this study). However, Cox (2003) did not screen these samples for the SNPs defining sub-lineages of C* (C2* or C2b), and these individuals may be reclassified if re-evaluated for additional SNPs. On the other hand, Kayser and coworkers (2006) did detect one C* individual in their PNG Coast samples.

Haplogroup C2* is one of the older branches of C*. Its occurs at high frequencies on Easter Island (25%), and at lower frequencies on PNG (4%), New Britain (2%), New Ireland (8%), New Hanover (3%), Bougainville (1%) and Manus Island (8%). Mona and coworkers (2007) also found C2* at high frequencies in New Guinea, and Kayser and colleagues (2006) found it at lower frequencies

(4%) in their New Guinea sample. Since this lineage is believed to have originated in Melanesia around 12,375 YBP, its presence in Polynesia indicates that it is a Melanesian haplogroup that was carried to Remote Oceania with Austronesian farmers.

A sub-branch of C2, haplogroup C2b is detected on every island except for New Hanover, and did not appear in the Micronesians. It occurs at the highest frequency on PNG (15%), and at lower frequencies on the islands of New Britain (1%), New Ireland (1%), Mussau (5%), Bougainville (1%), Easter Island (7%) and Manus Island (2%). It is believed to have originated in northwestern New Guinea ~46,200 YBP (Mona et al., 2007; Scheinfeldt et al., 2007). Interestingly, C2b was detected at much lower frequencies on New Guinea by Mona et al. (2007) and Kayser et al. (2006). However, Kayser et al. (2006) did detect haplogroup C2b at high frequency (15%) in the Admiralty Islands. A recent study by Cox and colleagues (2007) also found 0.38% of the men from the Indo-Pacific belonged to the C2b lineage, although 40 more individuals were identified as C-P33, which has a SNP that is downstream that which defines C2b. C2b is found in both Papuan and Austronesian speakers in this study, and has been detected in high frequencies in many Polynesian populations such as the Cook Islanders. Thus C2b has been viewed as supporting heavily Melanesian origin of Y-chromosomes in Polynesia (Kayser et al., 2008; Mona et al., 2007).

Macrohaplogroup K also has several ancient lineages in Melanesia. Haplogroup K-M9* is the most predominant lineage in this study, constituting 25% of all individuals. It is found at high frequencies on all the islands investigated: PNG (17%), New Britain (26%), New Ireland (13%), Mussau (25%), New Hanover (13%), Bougainville (30%), Manus Island (53%), Easter Island (14%) and in the Micronesians (41%). Given its antiquity K-M9* was likely carried into Melanesia with the original settlers of Sahul based on its wide distribution across Eurasia and the Americas.

K-M9* is presently defined as having the derived state for the M9 mutation, but lacking other lineage-defining SNPs. Currently, there is a lack of markers defining any sub-lineages of K-M9*, but future analyses may be able to subdivide this group into smaller groupings. Other studies have also had difficulty in dividing K-M9* into sub-lineages (Cox, 2003; Kayser et al., 2006).

This study reanalyzed samples employed by Scheinfeldt (2004) using the newly defined K6 and K7 SNPs, which divide K-M9* into two different sub-lineages. This study successfully reduced the frequency of K-M9* individuals from 41% to 25% with this marker analysis. Other studies conducted before K6 and K7 were characterized have similarly large numbers of K-M9* individuals. For example, Cox and Lahr (2006) noted that 20.3% of their samples belonged to K-M9*, while 47% of the sample in Cox (2003), and 25% of the Trobriand Islanders in Kayser and coworkers (2006) were K-M9* haplogroups.

Haplogroup K5 is detected on every island except Mussau and Easter Island, and is found within both Austronesian and Papuan speaking groups. It is found in very high frequencies on PNG (37%), and at smaller frequencies on New Britain (9%), New Ireland (6%), New Hanover (3%), Bougainville (1%), Manus Island (3%) and the Micronesians (8%). However, there were no individuals classified as K5 in the larger study of New Guinea by Mona et al (2007), although, they did detect the M254 SNP in 16 (9.8%) individuals, with M254 being a downstream marker of K5. The high frequencies of K5 in our sample reflect the likely origin of K5 in New Guinea (Mona et al., 2007). This lineage is also found at high frequencies in parts of Polynesia (Cox, 2003).

Haplogroup M* was detected on all islands except for New Hanover and those in Micronesia. It is found at its highest frequencies on Manus (27%) and PNG (15%), and at lower frequencies on New Britain (6%), New Ireland (5%), Mussau (5%), Bougainville (2%) and Easter Island (4%). This lineage

is restricted to Indonesia, Melanesia and Polynesia, and is believed to have appeared ~8200 YBP in Melanesia (Kayser et al., 2003). It is interesting that the larger sample from New Guinea analyzed by Mona et al. (2007) did not detect M* in their samples while Kayser et al. (2006) detected it at a low frequency in their samples, and it occurs at a high frequency in our dataset. This lineage is also found at a high frequency (30%) in Vanuatu (Cox, 2003). However, Cox did not test his samples for M2 or M2a, so it is possible that a reevaluation of these samples will decrease the prevalence of the M* lineage.

Haplogroup **M1** is a sub-lineage of M*, and is restricted to one Papuan-speaking individual of the Papua New Guinea highlands in this study. Other studies have found **M1** at high frequencies (47%) on the island of New Guinea, and suggest that this lineage is a marker of the expansion of Trans-New Guinea speakers from the central highlands (Mona et al., 2007). Kayser and colleagues (2006) found **M1** at extremely high frequencies in their Melanesian data set (53.25%); however, 82% of their samples were collected on New Guinea, further supporting the TNG expansion hypothesis. It is believed to have arisen in PNG ~7,356 YBP, and then subsequently carried from the central highlands to the rest of New Guinea and nearby islands fueled by agricultural expansion (Mona et al., 2007).

M1 was detected at a high frequency (10.2%) in the Admiralty Islands by Kayser and coworkers (2008). However, this lineage was not detected in the Manus Island samples in this study. Furthermore, it was detected in one Tolai (Austronesian) individual from New Britain (5%) and 15 Trobriand Islanders (28.3%) (Kayser et al., 2006). Unfortunately, haplogroup **M1** has not been extensively studied in Northern Island Melanesia, and it is possible that **M1** will be detected at higher frequencies across this region once this marker is utilized in future NRY analyses.

Indigenous Northern Island Melanesian Haplogroups: Remarkably, certain other haplogroups seem to have originated in specific parts of Northern Island Melanesia, particularly in New Britain (K6 and K7) and Bougainville (M2* and M2a1). For the most part, K6 and K7 have not spread to the west of New Britain, but have, in some cases, been carried to the southeast. The low frequency to absence of these lineages of considerable age in the western-most populations suggests that the populations of New Britain were relatively isolated for the majority of Melanesian habitation. However, the movement of these lineages in the other direction suggests that the Holocene farmers came into contact with these populations and picked them up while moving into Remote Oceania.

The large macrohaplogroup K has several lineages that appear to be indigenous to Northern Island Melanesia. Haplogroup **K6** is widely distributed within this sample set (15%), is found on every island in this study except for New Hanover and Easter Island, and appears in both Austronesian and Papuan speaking groups. While **K6** may be widespread across Northern Island Melanesia, its frequencies differ per island. It is found at very high frequencies on Mussau (50%), New Britain (22%) and in the Micronesian samples (17%). It is found at moderate frequencies on New Ireland (10%), and low frequencies on New Hanover (8%), PNG (4%), Bougainville (3%) and Manus Island (3%), but is absent from Easter Island. Haplogroup **K6** therefore likely originated on the island of New Britain. The TMRCA for K6 has been estimated at 37,100 YBP, using STR data (Scheinfeldt et al., 2007).

K6 is a recently discovered haplogroup and, therefore, has not been extensively analyzed in other studies. A recent study by Kayser and colleagues (2008) found **K6** at low frequencies (4.1%) in the Admiralty Islands, which is similar to its frequency on Manus Island. Kayser et al. (2006) also detected **K6** at high frequencies in western Polynesia and at lower frequencies in central Polynesia. These data suggest a decrease in the frequency of **K6** as one moves east across Island Melanesia

and into Remote Oceania, as **K6** was not detected in our eastern Polynesia sample (Easter Island). Thus, **K6** is an ancient lineage that arose in New Britain and has its highest frequencies in Papuan-speaking groups. The distribution of **K6** across Polynesia is consistent with the distribution of Lapita sites across Oceania, suggesting that the individuals moving through New Britain in the Holocene picked up this lineage and carried it into Remote Oceania during their subsequent colonization of Polynesia.

Haplogroup **K7** is restricted to the islands of New Britain and Bougainville in this study, and is found in both Austronesian and Papuan speaking groups (Tables 5.1 and 5.2). Actually, **K7** is found at higher frequencies on New Britain (13%) than on Bougainville (4%) (Figure 5.2). Thus, it likely originated in populations on the island of New Britain, with an estimated time of origin at ~35,400 YBP (Scheinfeldt et al., 2007). This time frame is consistent with the earliest settlement dates for this area (Summerhayes, 2007).

Haplogroup **M2***, a branch of **M***, is another newly defined lineage that has not been extensively employed in NRY studies of Oceania. Haplogroup **M2*** is found at its highest frequency on the islands of New Hanover (20%) and New Ireland (13%), and at lesser frequencies on the islands of PNG (2%), New Britain (4%), and Bougainville (1%). It, too, is found in both Papuan and Austronesian speakers. Scheinfeldt et al. (2007) suggested that **M2*** developed on New Ireland ~ 32,700 YBP. However, based on additional SNP markers, it appears that **M2*** likely developed in Bougainville, rather than New Ireland. Haplogroup **M2a***, which occurs on the **M2*** background, has the highest frequency (34%) in Bougainville populations. In addition, **M2a1** is only found in Bougainville in this study.

The **M2a*** haplogroup is a sub-lineage of **M2**, and was detected on all the islands in this study except on Easter Island and the Micronesians. This lineage is prevalent at an overall frequency of 19%

or 158 individuals analyzed. It is found in very high frequencies on the islands of Bougainville (34%), New Hanover (49%), Mussau (15%), New Ireland (29%) and New Britain (15%), and at lower frequencies at PNG (2%) and Manus Island (2%).

Sub-lineage **M2a1** is restricted to the Nagovisi and Nasioi (Papuan-speakers) of South Bougainville. **M2a1** most likely originated in South Bougainville, but it is such a rare lineage that the exact time and location of origin have not been extensively investigated. Nevertheless, this is a most interesting distribution for this rare haplogroup, suggesting the isolation of Bougainville populations.

Younger Asian Haplogroups: Another set of haplogroups is clearly identified with Austronesian-speaker origins (O and its sub-branches). They are distributed from Island Southeast Asia through Island Melanesia, and was carried out into Remote Oceania, but only as a minor component. There is a very small East Asian male contribution detected in this sample, as judged by the haplogroup O data. There are only 53 O haplogroups observed in the 842 males tested, comprising only 6.3% of the population. Specifically, the individuals that fall within macrohaplogroup O are found on the islands of New Britain, New Hanover, New Ireland, Bougainville and Manus, and in the Micronesians. Interestingly, these lineages are found in all Austronesian-speaking populations except for O* in one Papuan-speaking Kuot from New Ireland, and O* and O3 in two Micronesians of unknown linguistic affiliation. This is expected if the O lineages represent the expansion of Austronesian-speaking Asian horticulturalists through Melanesia during the Holocene.

In this study, haplogroup O* is found in New Britain (1%), New Ireland (5%), New Hanover (3%) and Bougainville (2%), and in the Micronesians (8%). However, it is possible that these individuals may be placed in the sub-haplogroups O2* or O2a* upon reevaluation, because they were not tested for the P31 and M95 SNPs. The O1a haplogroup is found on the islands of New Britain

(2%), New Ireland (5%), New Hanover (1%), Bougainville (9%) and Manus (2%). Haplogroup O3 is detected on New Britain (4%) and Bougainville (1%), and in the Micronesians (8%).

Other studies of NRY variation in Melanesia similarly detected a small East Asian component in Melanesian groups, with none of their samples being typed as O*, 4% as O1a and 4.75% as O3 (Kayser et al., 2006). A more recent study by Kayser and colleagues (2008) also detected a small East Asian contribution to the Admiralty Islands. However, they detected O haplotypes at higher frequencies than what was detected here by just one O1a individual from Manus. Again, this likely reflects different sample sizes and sampling strategies.

It is important to mention here that Cox et al (2007) has recently proposed a sub-lineage of C* as a “Polynesian motif” equivalent to the B4a1a1 mtDNA haplogroup. This newly discovered marker (P33) occurs downstream from that defining C2b (M208) and has only been found in Polynesian samples. It is believed that this marker arose just before or during the initial settlement of Remote Oceania, and has been dated at 7500 to 1500 YBP. The Austronesian expansion likely assimilated C2b individuals as they moved through Melanesia into Remote Oceania, and the P33 mutation arose soon after that (Cox et al., 2007). While this is not an East Asian marker, but instead one apparently restricted to Polynesians, it may serve to address many of the same issues that the O lineages are currently employed to do so. Cox et al (2007) found 100% of their Easter Island samples (n=9) to belong to the P33 lineage. In this study, the C2b haplogroup only represents 7% of the Easter Island samples, but they have not been tested for P33 and may actually belong to this newly defined Polynesian-specific group. Additionally, the Easter Island samples in this study were very old, and many of them failed amplification for the sub-lineages of C*, which is why they are classified as C*.

The distribution of East Asian haplogroups contrasts sharply with the mtDNA data from this region. The frequency of the haplogroup B4a1a1, known as the Polynesian motif, is very high in Northern Island Melanesia, and reaches near fixation in some groups, most particularly in south Bougainville and New Ireland (Friedlaender et al., 2007b). The discrepancy between the distribution of the two uni-parentally inherited haplogroups may be explained by differential demographic histories or drift. Since the mtDNA and Y chromosome have $\frac{1}{4}$ the effective population size of the autosomal DNA, they are extremely sensitive to the influence of genetic drift. Additionally, the practice of Oceanic matrilocality may produce this pattern (Hage and Marck, 2003; Kayser et al., 2008). This finding agrees with other Y chromosome studies in Northern Island Melanesia, which also found a very low East Asian male contribution to Melanesians (Kayser et al., 2008; Scheinfeldt, 2004; Scheinfeldt et al., 2006; Scheinfeldt et al., 2007).

Studies of NRY variation in Polynesia similarly suggest a low East Asian male contribution to Remote Oceania (Kayser et al., 2006). This finding has been attributed to the Slow Boat hypothesis of the settlement of Polynesia, where significant admixture occurred between the incoming Austronesian speaking Lapita people and the indigenous Papuan speakers. This process was followed by sex-specific demographic processes such as admixture bias in Polynesia towards more Melanesian men than women (Kayser et al., 2006). The only Polynesian sample in this study, Easter Island, supports the conclusions drawn from previous studies in that there are no O lineages detected there. Rather, the majority of these individuals belong to macrohaplogroup C, which suggests either a Melanesian origin or possibly a Polynesian-specific genetic signature.

Other Haplogroups: Finally, there are some rare SNPs whose distribution we cannot make much sense of at this point given their limited distributions. Haplogroups C4 and K1 were only

detected in one individual each in this study, with C4 being found in one Papuan-speaker from PNG, and K1 being found in one Papuan-speaking Nasioi from Bougainville. Additionally, Cox and Lahr (2006) identified a single K1 individual from the Malaita Province of the Solomon Islands. Until that time, K1 had been considered a private polymorphism and, thus, its defining SNP not tested regularly in studies. The authors suggested that this lineage represents ancient contacts between the two islands (Cox and Mirazon Lahr, 2006). Incorporating this haplogroup into population analyses of Near Oceania would provide a better understanding of the extent of the spread of this lineage.

The discrepancies between the frequencies detected in this study and other recent studies analyzing the same SNPs is likely due to differences in sample size and sampling strategies. This is especially true of the non-Northern Island Melanesia samples (PNG, Easter Island and the Micronesians), which were included as comparisons to the other samples from Northern Island Melanesia. Other discrepancies are likely due to the inclusion of different sets of markers for analysis. This is the result of the relatively rapid identification of new regionally informative markers for Melanesia.

Measures of Genetic Diversity Within and Between Populations

The third research objective was to investigate whether these newly defined SNPs will be able to provide evidence of new levels of heterozygosity and regionalization that may aid in overcoming the ascertainment of bias in Northern Island Melanesia. Ascertainment bias is the systematic distortion in a data set as a result of the way in which markers or sampled are collected (Jobling and Tyler-Smith, 2003b). It is a huge problem for NRY studies because of the large size of the Y chromosome, and the smaller numbers of populations sampled for NRY variation as those surveyed for mtDNA diversity. In

addition, it is not cost effective or timely to sequence entire Y-chromosomes. As a result, many studies type markers that were previously utilized in other studies. The problem here is that the same markers may not be polymorphic in the subsequent populations being analyzed or that the original analyses may have employed small sample sets that are not representative of certain groups. Therefore, the ascertainment of bias on the Y-chromosome essentially results in an underrepresentation of the genetic variation present in a particular region. For this reason, the discovery of new regionally specific markers can aid in an understanding of the biological variation in a region.

One of the major goals of this study was to investigate whether typing samples for the markers defining the new sub-lineages of K-M9*, K6 and K7, would reduce the large number of individuals classified as K-M9* in previous studies. This study built upon the work of Scheinfeldt (2004), which identified 41% of her samples as belonging to K-M9*. As a result, this study re-evaluated her samples for the newly defined SNPs, as well as increased the sample size and coverage of Northern Island Melanesia. This study was successful in overcoming the ascertainment of bias and specifically reduced the number of individuals assigned to K-M9* from 41% to 25%. By defining new subgroups within a much larger category, (e.g. K6 and K7), this study was able to detect a level of genetic variation not present in previous analyses of Melanesian populations. The individuals classified as K6 or K7 in this study constitutes 22.3% of the entire sample set, emphasizing the importance of discovering new polymorphic markers within Melanesian Y-chromosomes.

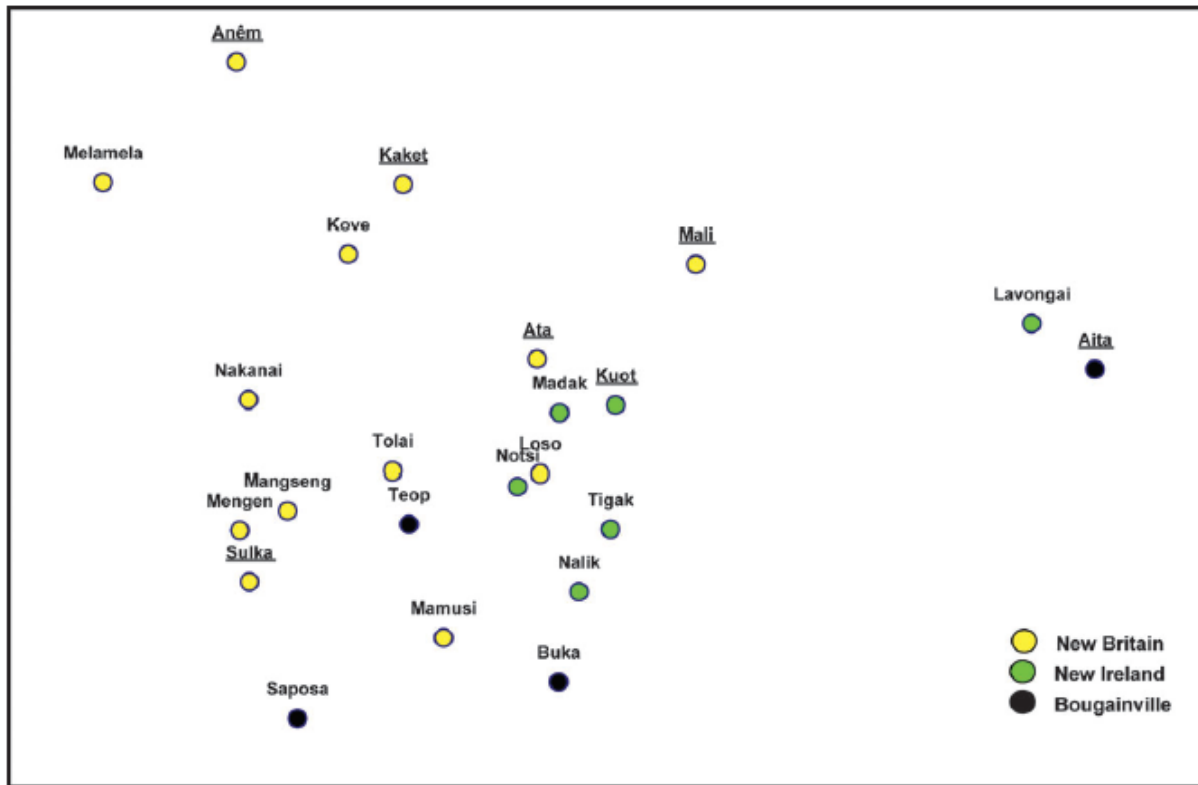
AMOVA Analysis: AMOVA was used to analyze the structure of NRY variation in the sample set (Table 5.4), and to determine if the patterns of variation clustered by island or language group. The patterns detected in this analysis were consistent with those of other studies (Scheinfeldt, 2004),

showing that the Melanesian groups are structured by geography and have a high internal SNP diversity.

The newly defined SNPs have provided evidence of highly structured NRY variation in Northern Island Melanesia, which is consistent with a history of relative isolation among the islands. As more samples are tested or reanalyzed for these regionally informative markers, a greater understanding of biological variation in Melanesia as well as a refined picture of Melanesian population history can be obtained.

Non-parametric multidimensional scaling (MDS) can be used to complement AMOVA analyses. MDS provides a two-dimensional representation of population relationships based on pairwise distances. Such an analysis was constructed using the SNP data in this study for the islands of New Britain, New Ireland and Bougainville (Figure 6.1), and published in Scheinfeldt et al. (2006 & 2007). This level of phylogeographic resolution can address issues of island isolation and show that the New Britain populations generally fall to one side of the plot, the New Ireland samples fall to the other side of the plot, and the Bougainville populations are mixed throughout. This suggests that New Britain and New Ireland populations have less within population diversity and have been isolated more than the Bougainville populations.

Figure 6.1: MDS Plot taken from Scheinfeldt et al. (2006, 2007)



Measures of Genetic Diversity: Within population variation was estimated by calculating the gene diversity, mean number of pairwise differences and average gene diversity (Table 5.5). The measures of genetic diversity all suggest that there is high haplogroup diversity in Northern Island Melanesia. Furthermore, the amount of genetic variability is due more to geographic than linguistic designations, as the coastal populations are more diverse than the isolated inland populations.

Distribution of the 50f2/c Deletion

The fourth research objective was to investigate the distribution of the 50f2/c deletion in this region. The appearance of the 50f2/c deletion on multiple haplogroup backgrounds suggests that it is

the result of multiple mutations (Jobling et al., 1996). If the deletion were nested within the greater K-M9* haplogroup, then we would expect to see the deletion in more individuals. If it occurred once in the K-M9* lineage, then we would expect to see it in all the individuals identified as belonging to haplogroups branching from the K-M9* lineage. In addition, the structure of the NRY tree has recently been modified (Figure 6.2) and the K7 branch is now separated from the K-M9* and K6 branches by the inclusion of new markers. Furthermore, seven STRs were previously analyzed for these individuals (Scheinfeldt et al., 2006; Scheinfeldt et al., 2007), and the alleles are different enough to suggest multiple mutations of the 50f2/c deletion. There are three main clusters of STR haplotypes on New Britain in the individuals carrying the deletion. Each of these clusters is separated from the others by at least four allelic differences across the seven loci. The STR profiles are even more divergent on Bougainville. Therefore, all the evidence suggests that the deletion has arisen several times within Melanesia, within a specific NRY macrohaplogroup.

Comparison of NRY Diversity with mtDNA and Autosomal Diversity

The final research objective was to compare the regional NRY diversity with that of the mtDNA diversity and autosomal diversity in Northern Island Melanesia by comparing the data collected in this study to the large amount of data collected on mtDNA variability in Northern Island Melanesia by Friedlaender et al. (2008, 2007a, b). They have thoroughly investigated mtDNA diversity throughout Melanesia. Through extensive sampling of the area and large sample sizes, they have deciphered the patterns of mtDNA diversity in this region of the world. This comparison is especially significant because their mtDNA study includes the same individuals utilized in this study.

The mtDNA diversity on Northern Island Melanesia is structured by island, island size and language affiliation. The greatest amount of mtDNA diversity is found within the more isolated Papuan-speaking groups that live on the larger island interiors versus the shore dwelling Austronesian-speakers. This is similar to the structure detected by this study and by Scheinfeldt (2004) for NRY diversity. Both the mtDNA and NRY variation show an island-by-island distinction of haplogroup occurrence and frequencies, with New Britain being the most diverse, followed by Bougainville and New Ireland.

The mtDNA data shows incredible population structure, with 28% of the population variance occurring among populations based on AMOVA results. The NRY among group variance is smaller, with 14.5% occurring among the populations, but still suggests significant population structure with regards to NRY haplotypes. These high values may be reflective of genetic drift among the small island populations. The mtDNA and the NRY variance figures indicate that variation among the islands was significant, with values of 12% and 6.6%, respectively. The amount of variation between populations within groups was larger for each locus as well, with the mtDNA at 17.8% and the NRY at 9.6%. The distinction between the NRY and mtDNA within population variation is likely due to the very common occurrence of haplogroup B4a1a1 in Melanesian groups, which acts to decrease the within population variation relative to the Y-chromosome. Additionally, these values are reflective of the size and topographical complexity of the islands in the study. Partitioning the variation by language produced non-significant results for both the mtDNA and Y chromosome data sets.

The apparent contradictions provided by the mitochondrial and Y chromosome data can be put in a larger context through comparisons with the results of analyses of neutral, bi-parentally inherited autosomal DNA analyses. Friedlaender and coworkers (2008) conducted an extensive autosomal

analysis of 952 individuals from 41 Pacific populations using 687 microsatellites and 203 insertions/deletions, with many of the same populations as covered by the NRY study being analyzed. The patterns of genetic diversity are generally similar to those detected in the mtDNA and NRY data. The Pacific populations had greater within population (95.4%) than among population (4.6%) genetic variation. However, in comparison to global populations, the Pacific groups had reduced diversity within populations.

The AMOVA results suggest that the apportionment of diversity is related to island size, with the larger and more rugged islands showing greater differentiation among populations, although New Guinea presents some exceptions to this trend. The PNG pattern is apparently due to the older Papuan groups inhabiting the island interiors and in part to marital migration patterns in the region. Similarly to the mtDNA and NRY data, they did not find a significant association between language affiliation and genetic variation (Friedlaender et al., 2008).

Friedlaender and coworkers (2008) used the program STRUCTURE (Pritchard et al., 2000) to investigate individual and population similarities. Their analysis of East Asian and Pacific populations provided even more detail about the relationships between Melanesians, Polynesians, Taiwan Aborigines and East Asians. They identified distinctions between Melanesian populations within the islands of Bougainville, New Guinea, New Britain and New Ireland. The results also showed a clear East Asian-Polynesian signature that the authors refer to as the “Austronesian” genetic signature. The Polynesian and Micronesian samples showed a primary affinity to Taiwan Aborigines and a secondary affinity to East Asians. The Polynesian samples showed very few connections to the Melanesian populations and a small genetic contribution from New Ireland and New Britain. Additionally, there were several Austronesian-speaking populations in Melanesia that showed the “Austronesian” genetic

signature. The Kove (New Britain) and Saposa (Bougainville) had the highest “Austronesian” contribution at just under 20% (Friedlaender et al., 2008).

A comparison of the NRY and autosomal data also illuminated Melanesian population history. The major difference between the two studies is the low contribution of East Asian males to Melanesian and Polynesian populations. This is in complete contrast to the clear Asian signal produced by the autosomal data. Instead, both the Melanesian and Polynesian populations show a preponderance of indigenous Melanesian NRY lineages. However, the genetic distinctions between the larger Melanesian islands of New Guinea, New Britain, New Ireland and Bougainville are apparent in both the autosomal and NRY data. Interestingly, the Austronesian-speaking Melanesian populations that show the greatest “Austronesian” contribution also have a high East Asian Y-chromosome frequency. For example, the Kove have a 20.8% frequency of O lineages and the Saposa have an 11.5% frequency of O lineages.

The Settlement of Polynesia

The greater phylogeographic resolution provided by this study can also contribute to a discussion of the settlement of Polynesia. The three areas of the genome that have been analyzed, NRY, mtDNA and autosomal loci, all contribute to a better understanding of the movement of people into Polynesia. The different types of genetic loci agree and contradict each other on certain aspects of Oceanic demography, although they all help play a role in deciphering the peopling of Remote Oceania.

This study has confirmed the generally low contribution of East Asian Y-chromosomes to Oceania, as detected in earlier studies. Only 6% of the NRY lineages in this sample belonged to

haplogroup O, which has been associated with the relatively recent expansion of East Asians into the region. The remaining NRY lineages are much older, and many are of indigenous Melanesian origin. This observation suggests that the genetic impact of the Austronesian-speakers on the Melanesian landscape was quite small. This pattern has also been detected in an analysis of NRY variation in Polynesia, with a greater contribution of Melanesian NRY lineages versus Asian NRY lineages (Hurles et al., 2002).

This weak signal can be interpreted as supporting the Slow Boat hypothesis of Polynesian origins in which Austronesian-speakers spread from Taiwan through Melanesia, intermixed with the Melanesians, and then moved out into Remote Oceania. Therefore, the Polynesians would show a strong Melanesian origin. This model of movement into Remote Oceania is supported by archeological evidence showing the slow movement of domesticates, farming and tools from Taiwan towards Island Southeast Asia and Melanesia over the past 6,000 years.

The greater phylogeographic specificity detected in this study has allowed for some hypotheses regarding the best islands for NRY contributions to the Polynesian gene pool. The earliest Lapita sites appear in the Bismarck archipelago (Summerhayes, 2007) making this an important area to investigate for genetic expansions as well. The major NRY lineages detected in Polynesia belong to macrohaplogroup C and its sub-branches. For example, Kayser et al. (2006) found 34.5% of their samples to be haplogroup C2b and 17.9% to be K-M9*, Hurles et al. (2002) found 63.3% of their samples to be K-M9*, and Capelli et al. (2001) found the predominant lineage in their Polynesian samples to be C*. New Britain, Mussau, New Ireland and Bougainville all have the haplogroups C2b and K-M9*, making them possible candidates. However; Bougainville and New Ireland have high frequencies of the M lineages, which occur at a lower frequency in Polynesia. It may be possible to

further resolve these relationships by looking for K6 and K7 haplogroups in Polynesia because the presence or absence of these SNPs may suggest one particular island versus another.

A very different picture arises from the mtDNA data. There is a large contribution of East Asian mtDNA lineages in Melanesia as evidenced by the haplogroup B4a1a1 and branches of haplogroup E. The frequencies of these haplogroups vary among the populations, and reach almost fixation in some of them. Additionally, there are the older P and Q lineages detected among the Melanesian groups. The amount of genetic diversity decreases moving east into Polynesia, as the frequency of the haplogroup B4a1a1 increases. This haplogroup pattern has been interpreted as a rapid movement of Austronesian-speaking East Asians through Melanesia and out into Remote Oceania with minimal admixture with the Melanesians. Thus, the earlier genetic analyses of Oceania, which involved mainly mtDNA data, supported an Express Train model of movement into Polynesia, while the NRY data show that this hypothesis may be overly simplistic

The discrepancies between the two uniparentally inherited loci may be the result of sex-biased admixture between the Austronesians and indigenous Melanesians followed by drift. As mentioned, the mtDNA and NRY loci have an effective population size $\frac{1}{4}$ that of the autosomal genome and are more susceptible to drift. The large Asian maternal contribution and lower male paternal contribution can possibly be explained by the practice of matrilocality, in which the Austronesian women were taking Melanesian men as husbands and thus moving them out into Polynesia (Hage, 1999). This pattern may also be the result of a matrilineal social system in which individuals identify with their mother's social group.

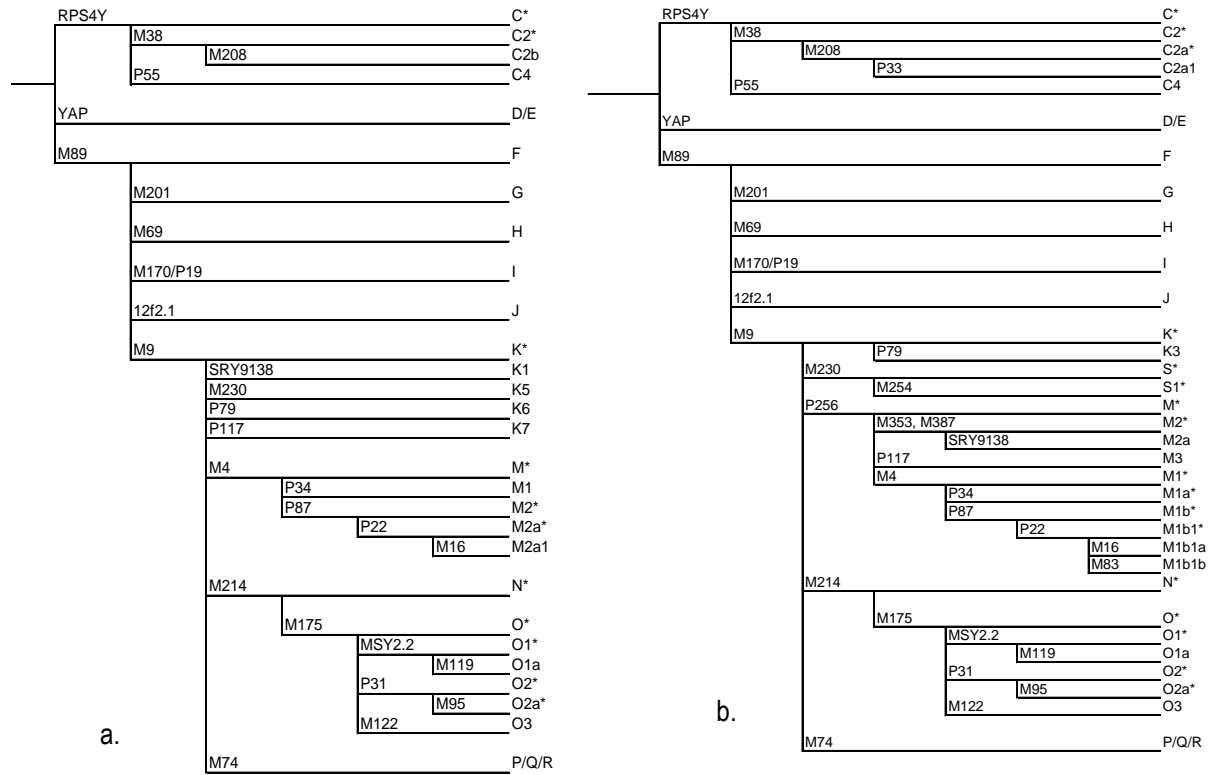
Under the prediction of matrilocality we would expect a high frequency of East Asian mtDNA lineages and a high frequency of Near Oceanic NRY lineages in Polynesia. As the Austronesians

moved through Melanesia the females would have picked up males to carry with them to their new homelands. This is clearly the picture that we see in Polynesia. However; under this same model, we would expect a low frequency of East Asian mtDNAs in Melanesia as few women would be expected to remain behind, and this is not what we see in Melanesia with the high frequencies of haplogroup B4a1a1. Thus, it is possible that the high frequencies of B4a1a1 in Melanesia are the result of drift and represent the descendants of a few very successful East Asian women.

A better understanding of the genetic constitution of the Oceanic people recently came from an analysis of the autosomal genome, which is inherited from both parents. Overall, the autosomal data suggests that genetic diversity within the Melanesian populations tends to be low, while differentiation between them is high. These data also identified a unique allele frequency profile detectable in Polynesian populations, called the “Austronesian” signature. This profile was detected at low frequencies only in some Austronesian speaking populations in Melanesia. This suggests that linguistic replacement of Papuan by Austronesian languages had occurred but not the genetic replacement of the older Melanesian autosomal lineages by new East Asian ones. In fact, the Austronesian-speaking Melanesian populations that did not contain the “Austronesian” signature were almost genetically indistinguishable from the Papuan-speaking groups. This finding confirms the Taiwanese origin of the Polynesians, as suggested by the earlier mtDNA studies. The autosomal data also suggest that the East Asian individuals contributed the majority of the autosomal genes to the Polynesians, a scenario suggested by the mtDNA data. However, they did leave a small genetic signature in Melanesia while picking up paternal lineages on their way to Remote Oceania. While the NRY data has not had the ability to completely resolve many of the issues surrounding the population

history of the South Pacific, it has been useful in investigating patterns of population diversity in the area.

Figure 6.2: a. YCC Tree 2005, b. YCC Tree 2008



CHAPTER 7 FUTURE STUDIES

The new markers utilized in this study, namely K6, K7, M2* and M2a1, were added to the YCC tree in 2005. They were instrumental in subdividing the K macrohaplogroup into smaller branches. Additional newly defined SNPs have recently been incorporated into the YCC tree for 2008 (Figure 6.2). These markers have the potential to further differentiate haplogroups within the large categories within macrohaplogroups C and K. Also note in Figure 6.1 that the lineage names have changed since the writing of this dissertation. The newly defined C2a1 (P33) lineage has already been detected in high frequencies in Polynesia, and would prove very useful if tested on additional Polynesian groups, as well as Melanesian groups for comparison.

The P256 SNP occurs between M9 and M4, and redefines M*. This marker might also serve to further break down the K-M9* category. For example, the combination of P256 and the newly defined M353, M387 may shed light on the origin and spread of the rare lineage M2a (formerly K1), since its defining SNP (SRY9138) is upstream of this marker. The newly defined SNP M254 (S1*), occurring downstream of M230 (S*), might allow the subdivision of the large group of M230 (formerly called K5) individuals in this study. Furthermore, it would be useful to compare the P117 data to the available P256 positive individuals to shed light on the origin of K7 (now M3). Finally, the new M83 marker (M1b1b) might be helpful in subdividing the samples that have been designated as M-P22 (xM16).

The availability of information about these newly defined markers provides an opportunity to reevaluate this data set with the new markers. This is a valuable data set due to its size, geographic coverage, and available linguistic and genealogical information. These samples should continue to be screened for new markers as they are discovered. Additionally, the geographic area screened for

these markers (as well as the markers introduced in this study) should be increased. It is very possible that ascertainment bias is hiding recognizable patterns of genetic variation. This study has made strides in uncovering detectable patterns of NRY variation in Melanesia. It was able to subdivide some of the larger haplogroups, and these newly defined markers will continue to make progress in our attempts to detect and understand human biological variation in Oceania.

Additionally, analyses of Pacific populations have focused on comparisons between individuals of Near Oceanic and Remote Oceanic descent with East Asians. Recent studies of mtDNA and autosomal DNA have suggested that there is a strong link between Taiwanese and Polynesian populations. However, populations in Indonesia and Southeast Asia have not been extensively studied using the same battery of markers as subsequent analyses. Future work may show an even stronger link between the Pacific populations and those of Island Southeast Asia.

LITERATURE CITED

- Allaby R (2007) Origins of Plant Exploitation in Near Oceania: A Review. In J Friedlaender (ed.): Genes, Language, and Culture History in the Southwest Pacific. New York: Oxford University Press, pp. 181-198.
- Allen J (1984) In search of the Lapita homeland. *Journal of Pacific History* 19:186-201.
- Allen J (1994) Radiocarbon determinations, luminescence dating and Australian archaeology. *Antiquity* 68.
- Ballinger SW, Schurr TG, Torroni A, Gan YY, Hodge JA, Hassan K, Chen KH, and Wallace DC (1992) Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient mongoloid migrations. *Genetics* 130:139-52.
- Bandelt HJ, Forster P, and Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16:37-48.
- Bandelt HJ, Forster P, Sykes BC, and Richards MB (1995) Mitochondrial portraits of human populations using median networks. *Genetics* 141:743-53.
- Bellwood P, Fox JJ, and Tryon D, eds. (1995) *The Austronesians*. Canberra: Department of Anthropology, The Australian National University.
- Bellwood P (1991) The Austronesian dispersal and the origin of languages. *Scientific American July*:70-75.
- Bergen A, Wang C, Tsai J, Jefferson K, Dey C, Smith K, Park S, Tsai S, and Goldman D (1999) An Asian-native American paternal lineage identified by RPS4Y resequencing and by microsatellite haplotyping. *Annals of Human Genetics* 63:63-80.
- Bowler JM, Johnston H, Olley JM, Prescott JR, Roberts RG, Shawcross W, and Spooner NA (2003) New ages for human occupation and climatic change at Lake Mungo, Australia. *Nature* 421:837-840.
- Brace CL, and Hinton RJ (1981) Oceanic tooth-size variation as a reflection of biological and cultural mixing. *Current Anthropology* 22:549-569.
- Brace CL, and Hunt KD (1990) A nonracial craniofacial perspective on human variation: A(ustralia) to Z(uni). *Am J Phys Anthropol* 82:341-60.
- Cann R, Stoneking M, and Wilson A (1987) Mitochondrial DNA and human evolution. *Nature* 325:31-36.
- Cann RL, and Lum JK (2004) Dispersal ghosts in Oceania. *American Journal of Human Biology* 16:440-451.

- Capelli C, Wilson JF, Richards M, Stumpf MPH, Gratix F, Oppenheimer S, Underhill PA, Pascali V, Ko TM, and Goldstein DB (2001) A predominantly indigenous paternal heritage for the Austronesian-speaking peoples of insular Southeast Asia and Oceania. *American Journal of Human Genetics* 68:432-443.
- Chaix R, Austerlitz F, Morar B, Kalaydjieva L, and Heyer E (2004) Vlach Roma history: what do coalescent-based methods tell us? *Eur J Hum Genet* 12:285-92.
- Consortium Y-C (2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Research* 12:339-348.
- Cox M (2003) Genetic patterning at Austronesian contact zones. PhD Thesis, Otago University, Dunedin, New Zealand.
- Cox MP, and Mirazon Lahr M (2006) Y-chromosome diversity is inversely associated with language affiliation in paired Austronesian- and Papuan-speaking communities from Solomon Islands. *Am J Hum Biol* 18:35-50.
- Cox MP, Redd AJ, Karafet TM, Ponder CA, Lansing S, Sudoyo H, and Hammer MF (2007) A polynesian motif on the Y chromosome: population structure in remote Oceania. *Hum Biol* 79:525-35.
- Cummins J (2001) Mitochondrial DNA and the Y chromosome: parallels and paradoxes. *Reprod. Fertil. Dev.* 13:533-542.
- Destro-Bisol G, Donati F, Coia V, Boschi H, Verginelli F, Caglia A, Tofanelli S, Spedini G, and Capelli C (2004a) Variation of female and male lineages in Sub-Saharan populations: the importance of sociocultural factors. *Molecular Biology and Evolution* 21:1673-1682.
- Destro-Bisol G, Donati F, Coia V, Boschi I, Verginelli F, Caglia A, Tofanelli S, Spedini G, and Capelli C (2004b) Variation of female and male lineages in sub-saharan populations: the importance of sociocultural factors. *Molecular Biology and Evolution* 21:1673-1682.
- Diamond JM (1988) Express train to Polynesia. *Nature* 336:307-308.
- Diamond JM (2001) Slow boat to Melanesia: a reply. *Nature* 410:167.
- Dunn M, Reesink G, and Terrill A (2002) The East Papuan languages: a typological appraisal. *Oceanic Linguistics* 41:28-62.
- Dunn M, Terrill A, Reesink G, Foley R, and Levinson S (2005) Structural phylogenetics and the reconstruction of ancient language history. *Science* 309:2072-2075.
- Excoffier L, and Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1:47-50.

- Fay JC, and Wu CI (1999) A human population bottleneck can account for the discordance between patterns of mitochondrial versus nuclear DNA variation. *Molecular Biology and Evolution* 16:1003-1005.
- Flint J, Hill AV, Bowden DK, Oppenheimer SJ, Sill PR, Serjeantson SW, Bana-Koiri J, Bhatia K, Alpers MP, Boyce AJ, and et al. (1986) High frequencies of alpha-thalassaemia are the result of natural selection by malaria. *Nature* 321:744-50.
- Forster P (2004) Ice ages and the mitochondrial DNA chronology of human dispersals: a review. *Philosophical Transactions of the Royal Society of London B* 359:255-264.
- Forster P, Torroni A, Renfrew C, and Rohl A (2001) Phylogenetic star contraction applied to Asian and Papuan mtDNA. *Molecular Biology and Evolution* 18:1864-1881.
- Friedlaender J, Friedlaender F, Hodgson J, McGrath S, Stoltz M, Koki G, Schurr T, and Merriwether DA (2007a) Mitochondrial DNA Variation in Northern Island Melanesia. In J Friedlaender (ed.): *Genes, Language, and Culture History in the Southeast Pacific*. New York: Oxford University Press, pp. 61-80.
- Friedlaender J, Friedlaender F, Hodgson J, Stoltz M, Koki G, Horvat G, Zhadanov S, Schurr T, and Merriwether DA (2007b) Melanesian mtDNA complexity. *PLoS ONE* 2:e248.
- Friedlaender J, Friedlaender F, Reed F, Kidd K, Kidd J, Chambers G, Lea R, Loo J, Koki G, Hodgson J, Merriwether DA, and Weber J (2008) The genetic structure of Pacific Islanders. *PLoS Genetics* 4:e19.
- Friedlaender J, Gentz F, Friedlaender F, Kaestle F, Schurr T, Koki G, Schanfield M, McDonough J, Smith L, Cerchio S, Mgone C, and Merriwether DA (2005a) Mitochondrial Genetic Diversity and its Determinants in Island Melanesia. In A Pawley, R Attenborough, J Golson and R Hide (eds.): *Papuan Pasts: Studies in the Cultural, Linguistic, and Biological History of the Papuan Speaking Peoples*. Canberra: Pacific Linguistics, pp. 693-716.
- Friedlaender J, Gentz F, Green K, and Merriwether D (2002) A cautionary tale on ancient migration detection: mitochondrial DNA variation in Santa Cruz Islands, Solomon Islands. *Human Biology* 74:453-471.
- Friedlaender J, Schurr T, Gentz F, Koki G, Friedlaender F, Horvat G, Babb P, Cerchio S, Kaestle F, Schanfield M, Deka R, Yanagihara R, and Merriwether DA (2005b) Expanding Southwest Pacific mitochondrial haplogroups P and Q. *Molecular Biology and Evolution* *Epub ahead of print*.
- Futuyma D (1998) *Evolutionary Biology*. Sunderland, Massachusetts: Sinauer Associates, Inc.
- Golson J, and Gardner DS (1990) Agriculture and sociopolitical organization in New Guinea highlands prehistory. *Annual Review of Anthropology* 19:395-417.

- Gosden C, and Webb J (1994) The creation of a Papua New Guinea landscape: archaeological and geomorphological evidence. *Journal of Field Archaeology* 21:29-51.
- Gray RD, and Jordan FM (2000) Language trees support the express-train sequence of Austronesian expansion. *Nature* 405:1052-1055.
- Green R (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*. Auckland: The Polynesian Society.
- Green RC (2003) The Lapita horizon and traditions - signature for one set of Oceanic migrations. In C Sand (ed.): *Pacific Archaeology: Assessments and Anniversary of the First Lapita Excavation (July 1952)*. Koné, Nouméa, 2002. *Le Cahiers de l'Archéologie en Nouvelle-Calédonie*, Nouméa, New Caledonia., pp. 1-26.
- Greenberg J (1971) The Indo-Pacific hypothesis. In T Sebeok (ed.): *Current Trends in Linguistics 8: Linguistics in Oceania*. Mouton: The Hague, pp. 807-871.
- Groube L, Chappell J, Muke J, and Price D (1986) A 40,000 year-old human occupation site at Huon Peninsula, Papua New Guinea. *Nature* 324:453-455.
- Hage P (1999) Reconstructing ancestral Oceanic society. *Asian Perspectives* 38:200-228.
- Hage P, and Marck J (2003) Matrilineality and the Melanesian origin of Polynesian Y chromosomes. *Current Anthropology* 44:S121-127.
- Hagelberg E, and Clegg J (1993) Genetic polymorphisms in prehistoric Pacific Islanders determined by analysis of ancient bone DNA. *Proceedings of the Royal Society of London, Biological Sciences* 252:163-170.
- Hammer M, Blackmen F, Garrigan D, Nachman M, and Wilder J (2003) Human population structure and its effects on sampling y chromosome sequence variation. *Genetics* 164:1495-1509.
- Hammer MF, Karafet TM, Redd AJ, Jarjanazi H, Santachiara-Benercetti S, Soodyall H, and Zegura SL (2001) Hierarchical patterns of global human Y-chromosome diversity. *Molecular Biology and Evolution* 18:1189-1203.
- Hammer MF, and Zegura SL (2002) The human Y chromosome haplogroup tree: nomenclature and phylogeography of its major divisions. *Annual Review of Anthropology* 31:303-321.
- Hanihara T (1992) Dental and cranial affinities among populations of east Asia and the Pacific: the basic populations in east Asia, IV. *Am J Phys Anthropol* 88:163-82.
- Harihara S, Hirai M, Suutou Y, Shimizu K, and Omoto K (1992) Frequency of a 9-bp deletion in the mitochondrial DNA among Asian populations. *Hum Biol* 64:161-6.

- Hill AV, Bowden DK, Trent RJ, Higgs D, Oppenheimer S, Thein S, Mickleson KNP, Weatherall DJ, and Clegg JB (1985) Melanesians and Polynesians share a unique alpha-thalassemia mutation. *American Journal of Human Genetics* 37:571-580.
- Hill AV, Gentile B, Bonnardot J, Roux J, Weatherall DJ, and Clegg JB (1987) Polynesian origins and affinities: Globin gene variants in Eastern Polynesia. *American Journal of Human Genetics* 40:453-463.
- Hill C, Soares P, Mormina M, Macaulay V, Clarke D, Blumbach PB, Vizuete-Forster M, Forster P, Bulbeck D, Oppenheimer S, and Richards M (2007) A mitochondrial stratigraphy for island southeast Asia. *Am J Hum Genet* 80:29-43.
- Horai S, and Matsunaga E (1986) Mitochondrial DNA polymorphism in Japanese. II. Analysis with restriction enzymes of four or five base pair recognition. *Hum Genet* 72:105-17.
- Hudjashov G, Kivisild T, Underhill PA, Endicott P, Sanchez J, Lin AA, Shen P, Oefner PJ, Renfrew C, Villems R, and Forster P (2007) Revealing the prehistoric settlement of Australia by Y chromosome and mtDNA analysis. *Proceedings of the National Academy of Sciences of the USA* 104:8726-8730.
- Hurles ME, Nicholson J, Bosch E, Renfrew C, Sykes BC, and Jobling MA (2002) Y chromosomal evidence for the origins of Oceanic-speaking peoples. *Genetics* 160:289-303.
- Ingman M, and Gyllensten U (2003) Mitochondrial genome variation and evolutionary history of Australian and New Guinean aborigines. *Genome Research* 13:1600-1606.
- Ingman M, Kaessmann H, Paabo S, and Gyllensten U (2000) Mitochondrial variation and the origin of modern humans. *Nature* 408:708-713.
- Jobling M, and Tyler-Smith C (2003a) The human Y chromosome: an evolutionary marker comes of age. *Nature Reviews Genetics* 4:598-612.
- Jobling MA, Hurles M, and Tyler-Smith C (2004) *Human Evolutionary Genetics: Origins, Peoples, and Disease*. New York: Garland Science.
- Jobling MA, Samara V, Pandya A, Fretwell N, Bernasconi B, Mitchell RJ, Gerelsaikhan T, Dashnyam B, Sajantila A, Salo PJ, Nakahori Y, Disteche CM, Thangaraj K, Singh L, Crawford MH, and Tyler-Smith C (1996) Recurrent duplication and deletion polymorphisms on the long arm of the Y chromosome in normal males. *Hum Mol Genet* 5:1767-75.
- Jobling MA, and Tyler-Smith C (1995) Fathers and sons - the Y chromosome and human evolution. *Trends in Genetics* 11:449-456.
- Jobling MA, and Tyler-Smith C (2003b) The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet* 4:598-612.

- Jorde LB, Watkins WS, Bamshad MJ, Dixon ME, Ricker CE, Seiselstad MT, and Batzer MA (2000) The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y chromosome data. *American Journal of Human Genetics* 66:113-155.
- Karafet TM, Lansing J, Redd AJ, Reznikova S, Watkins J, Surata S, Arthawiguna W, Mayer L, Bamshad MJ, Jorde LB, and Hammer MF (2005) Balinese Y-chromosome perspective on the peopling of Indonesia: genetic contributions from pre-Neolithic hunter-gatherers, Austronesian farmers, and Indian traders. *Human Biology* 77:93-114.
- Kayser M, Brauer S, Cordaux R, Casto A, Lao O, Zhivotovsky L, Moyses-Faurie C, Rutledge R, Schiefenhover W, Gil D, Lin AA, Underhill PA, Oefner PJ, Trent RJ, and Stoneking M (2006) Melanesian and Asian origins of Polynesians: mtDNA and Y-chromosome gradients across the Pacific. *Molecular Biology and Evolution* 23:2234-2244.
- Kayser M, Brauer S, Weiss G, Schiefenhover W, Underhill PA, Shen P, Oefner PJ, Tommaseo-Ponzetta M, and Stoneking M (2003) Reduced Y-chromosome, but not mitochondrial DNA, diversity in human populations from west New Guinea. *American Journal of Human Genetics* 72:281-302.
- Kayser M, Brauer S, Weiss G, Underhill PA, Roewer L, Schiefenhover W, and Stoneking M (2000) Melanesian origin of Polynesian chromosomes. *Current Biology* 10:1237-1246.
- Kayser M, Choi Y, van Oven M, Mona S, Brauer S, Suarkia D, Schiefenhover W, and Stoneking M (2008) The impact of Austronesian expansion: evidence from mtDNA and Y-chromosome diversity in the Admiralty Islands of Melanesia. *Molecular Biology and Evolution* *Advance Access*.
- Kayser M, Krawczak M, Excoffier L, Deltjes P, Corach D, Pascali V, Gehrig C, Bernini LF, Jespersen J, Bakker E, Roewer L, and de Knijff P (2001) An extensive analysis of Y-chromosomal microsatellite haplotypes in globally dispersed human populations. *American Journal of Human Genetics* 68:990-1018.
- Kirch PV (1987) Lapita and Oceanic cultural origins: excavations in the Mussau Islands, Bismarck Archipelago. *Journal of Field Archaeology* 14:163-180.
- Kirch PV (1997) *The Lapita Peoples: Ancestors of the Oceanic World*. Malden, Massachusetts: Blackwell Publishers, Inc.
- Kirch PV (2000) *On the Road of the Winds: an Archaeological History of the Pacific Islands before European Contact*. Berkeley: University of California Press.
- Lindstrom E, Terrill A, Reesink G, and Dunn M (2007) The Languages of Island Melanesia. In J Friedlaender (ed.): *Genes, Language, and Culture History in the Southwest Pacific*. New York: Oxford University Press, pp. 118-140.

- Lum JK (2007) Contributions of population origins and gene flow to the diversity of neutral and malaria selected autosomal genetic loci of Pacific Island populations. In J Friedlaender (ed.): *Genes, Language, and Culture History in the Southwest Pacific*. New York: Oxford University Press, pp. 219-230.
- Lum JK, and Cann R (2000) mtDNA lineage analyses: origins and migrations of Micronesians and Polynesians. *American Journal of Physical Anthropology* 113:151-168.
- Lum JK, Jorde LB, and Schiefenhover W (2002) Affinities among Melanesians, Micronesians, and Polynesians: a neutral, biparental genetic perspective. *Human Biology* 74:413-430.
- Lum K, Rickards O, Ching C, and Cann R (1994) Polynesian mitochondrial DNAs reveal three deep maternal lineage clusters. *Human Biology* 66:567-590.
- Macaulay V, Hill C, Achilli A, Rengo C, Clarke D, Meehan W, Blackburn J, Semino O, Scozzari R, Cruciani F, Taha A, Shaari NK, Raja JM, Ismail P, Zainuddin Z, Goodwin W, Bulbeck D, Bandelt HJ, Oppenheimer S, Torroni A, and Richards M (2005) Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* 308:1034-6.
- Mack SJ, Bugawan TL, Moonsamy PV, Erlich JA, Trachtenberg EA, Paik YK, Begovich AB, Saha N, Beck HP, Stoneking M, and Erlich HA (2000) Evolution of Pacific/Asian populations inferred from HLA class II allele frequency distributions. *Tissue Antigens* 55:383-400.
- Matisoo-Smith E, and Robins JH (2004) Origins and dispersals of Pacific peoples: evidence from mtDNA phylogenies of the Pacific rat. *Proceedings of the National Academy of Sciences of the USA* 101:9167-9172.
- Melton T, Clifford S, Martinson J, Batzer MA, and Stoneking M (1998) Genetic evidence for the Proto-Austronesian homeland in Asia: mtDNA and nuclear DNA variation in Taiwanese aboriginal tribes. *American Journal of Human Genetics* 63:1807-1823.
- Melton T, Peterson R, Redd AJ, Saha N, Sofro ASM, Martinson J, and Stoneking M (1995) Polynesian genetic affinities with Southeast Asian populations as identified by mtDNA analysis. *American Journal of Human Genetics* 57:403-414.
- Merriwether DA, Friedlaender F, Mediavilla J, Mgone C, Gentz F, and Ferrell R (1999) Mitochondrial DNA variation as an indicator of Austronesian influence in Island Melanesia. *American Journal of Physical Anthropology* 110:243-270.
- Merriwether DA, Hodgson J, Friedlaender F, Allaby R, Cerchio S, Koki G, and Friedlaender J (2005) Ancient mitochondrial M haplogroups identified in the Southwest Pacific. *Proceedings of the National Academy of Sciences of the USA* 102:13034-13039.

- Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, Hosseini S, Brandon M, Easley K, Chen E, Brown MD, Sukernik RI, Olckers A, and Wallace DC (2003) Natural selection shaped regional mtDNA variation in humans. *Proceedings of the National Academy of Sciences of the USA* 100:171-176.
- Mona S, Tommaseo-Ponzetta M, Brauer S, Sudoyo H, Marzuki S, and Kayser M (2007) Patterns of Y-chromosome diversity intersect with the Trans-New Guinea hypothesis. *Mol Biol Evol* 24:2546-55.
- Nei M (1987) *Molecular Evolutionary Genetics*. New York: Columbia University Press.
- O'Connell, and Allen J (2004) Dating the colonization of Sahul (Pleistocene Australia-New Guinea): a review of recent research. *Journal of Archaeological Science* 31:835-853.
- O'Shaughnessy DF, Hill AV, Bowden DK, Weatherall DJ, and Clegg JB (1990) Globin genes in Micronesia: origins and affinities of Pacific Island peoples. *Am J Hum Genet* 46:144-55.
- Oppenheimer S, and Richards M (2001a) Fast trains, slow boats, and the ancestry of the Polynesian islanders. *Science Progress* 84:157-181.
- Oppenheimer S, and Richards M (2001b) Slow boat to Melanesia. *Nature* 410:166-167.
- Pavildes C, and Gosden C (1994) 35,000-year-old sites in the rainforests of West New Britain, Papua New Guinea. *Antiquity* 68:604-610.
- Pawley A (1997) The Trans New Guinea phylum hypothesis: a reassessment.
- Pawley A (2000) The chequered career of the Trans New Guinea phylum: recent historical research and its implications. In A Pawley, Attenborough, Hide and Golsen (eds.): *Papuan Pasts: Investigations into the Cultural, Linguistic, and Biological History of the Papuan Speaking Peoples*. Adelaide: Crawford House Australia.
- Pawley A (2001) The Proto Trans New Guinea obstruents: arguments from top-down reconstruction: *Studies in Melanesian Linguistics in Honour of Tom Dutton*. Canberra: Pacific Linguistics, pp. 261-300.
- Pawley A (2003) Locating Proto Oceanic. In M Ross, A Pawley and M Osmond (eds.): *The Lexicon of Proto Oceanic: The Physical Environment*: Pacific Linguistics.
- Pawley A (2007) Recent Research on the Historical Relationships of the Papuan Languages, or, What Does Linguistics Say about the Prehistory of Melanesia? In J Friedlaender (ed.): *Genes, Language, and Culture History in the Southwest Pacific*. New York: Oxford University Press, pp. 36-60.
- Pawley A, and Green R (1973) Dating the dispersal of the Oceanic languages. *Oceanic Linguistics* 12:1-67.

- Pawley A, and Ross M (1993) Austronesian historical linguistics and culture history. *Annual Review of Anthropology* 22:425-459.
- Pietrusewsky M (1994) Pacific-Asian relationships: A physical anthropological perspective. *Oceanic Linguistics* 33:407-429.
- Pietrusewsky M, and Chang C (2003) Taiwan aboriginals and peoples of the Pacific-Asia region: multivariate craniometric comparisons. *Anthropological Science* 111:293-332.
- Pritchard JK, Stephens M, and Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945-59.
- Renfrew C (1994) World linguistic diversity. *Scientific American January*:116-123.
- Richards M, and Macaulay V (2001) The mitochondrial gene tree comes of age. *Am J Hum Genet* 68:1315-1320.
- Roberts RG, Jones R, and Smith MA (1990) Thermoluminescence dating of a 50,000-year-old human occupation site in northern Australia. *Nature* 345:153-156.
- Roewer L, Croucher P, Willuweit S, Lu T, Kayser M, Lessig R, de Knijff P, Jobling M, Tyler-Smith C, and Krawczak M (2005a) Signature of recent historical events in the European Y-chromosomal STR haplotype distribution. *Human Genetics* 116:279-291.
- Roewer L, Croucher PJ, Willuweit S, Lu TT, Kayser M, Lessig R, de Knijff P, Jobling MA, Tyler-Smith C, and Krawczak M (2005b) Signature of recent historical events in the European Y-chromosomal STR haplotype distribution. *Hum Genet* 116:279-91.
- Romualdi C, Balding D, Nasidze I, Risch G, Robichaux M, Sherry S, Stoneking M, Batzer M, and Barbujani G (2002) Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms. *Genome Research* 12:602-612.
- Ross M (1995) The great Papuan hunt: recalibrating our sites.
- Ross M, ed. (2001) *Is there a Papuan phylum? Evidence from pronouns*. Canberra: Pacific Linguistics.
- Ross M (2004) Pronouns as a preliminary diagnostic for grouping Papuan Languages. In A Pawley, Attenborough, Hide and Golsen (eds.): *Papuan Pasts: Investigations into the Cultural, Linguistic, and Biological History of the Papuan Speaking Peoples*. Adelaide: Crawford House Australia.
- Schanfield M, Austin F, Booth P, Gajdusek D, Hornabrook R, McAdams K, Saave J, Serjeantson SW, and Woodfield G (2007) Immunoglobulin Allotypes as a Marker of Population History in the Southwest Pacific. In J Friedlaender (ed.): *Genes, Language, and Culture History in the Southeast Pacific*. New York: Oxford University Press.

- Scheinfeldt L (2004) Y chromosome variation in the Bismarck Archipelago: disentangling genetic relationships among Austronesian and Papuan speaking populations. PhD Thesis, Temple University, Philadelphia, PA.
- Scheinfeldt L, Friedlaender F, Friedlaender J, Latham K, Koki G, Karafet TM, Hammer MF, and Lorenz J (2006) Unexpected NRY chromosome variation in Northern Island Melanesia. *Molecular Biology and Evolution* 23:1628-1641.
- Scheinfeldt L, Friedlaender F, Friedlaender J, Latham K, Koki G, Karafet TM, Hammer MF, and Lorenz J (2007) Y Chromosome Variation in Northern Island Melanesia. In J Friedlaender (ed.): *Genes, Language, and Culture History in the Southwest Pacific*. Oxford: Oxford University Press, pp. 81-95.
- Scott G, and Turner C (1988) Dental Anthropology. *Annual Review of Anthropology* 17:99-126.
- Seielstad M, Minch E, and Cavalli-Sforza L (1998a) Genetic evidence for a higher female migration rate in humans. *Nature Genetics* 20:278-280.
- Seielstad MT, Minch E, and Cavalli-Sforza LL (1998b) Genetic evidence for a higher female migration rate in humans. *Nature Genetics* 20:278-280.
- Serjeantson SW, Ryan D, and Thomson R (1982) The colonization of the Pacific: The story according to human leukocyte antigens. *American Journal of Human Genetics* 34:904-918.
- Spriggs M (1997) *The Island Melanesians*. Cambridge: Blackwell Publishers Inc.
- Su B, Jin L, Underhill PA, Martinson J, Saha N, McGarvey S, Shriver M, Chu J, Oefner PJ, Chakraborty R, and Deka R (2000) Polynesian origins: insights from the Y chromosome. *Proceedings of the National Academy of Sciences of the USA* 97:8225-8228.
- Summerhayes GR (2004) Island Melanesian Pasts - A View from Archaeology: The Second Garden of Eden - Island Melanesian Genetic Diversity, pp. 1-31.
- Summerhayes GR (2007) Island Melanesian pasts - a view from archaeology. In J Friedlaender (ed.): *Genes, Language, and Culture in the Southwest Pacific*. New York: Oxford University Press, pp. 10-35.
- Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105:437-460.
- Tajima F (1993) Measurement of DNA polymorphism. In N Takahata and AG Clark (eds.): *Mechanisms of Molecular Evolution. Introduction to Molecular Paleopopulation Biology*. Tokyo: Japan Scientific Societies Press, Sinauer Associates, Inc, pp. 269-285.
- Terrell JE (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, and Rainbird P (2001) Foregone conclusions? In search of "Papuan" and "Austronesians". *Current Anthropology* 42:97-124.
- Terrell JE, and Welsch RL (1997) Lapita and the temporal geography of prehistory. *Antiquity* 71:548-572.
- Terrill A (2002) Systems of nominal classification in East Papuan languages. *Oceanic Linguistics* 41:63-88.
- Thorne A, Grun R, Mortimer G, Spooner NA, Simpson JJ, McCulloch M, Taylor L, and Curnoe D (1999) Australia's oldest human remains: age of the Lake Mungo 3 skeleton. *Journal of Human Evolution* 36:591-612.
- Trejaut J, Kivisild T, Loo J, Lee C, He C, Hsu C, Li Z, and Lin M (2005) Traces of archaic mitochondrial lineages persist in Austronesian-speaking Formosan populations. *PLoS Biology* 3:e247.
- Underhill P, Passarino G, Lin A, Shen P, Lahr M, Foley R, Oefner P, and Cavalli-Sforza L (2001a) The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Annals of Human Genetics* 65.
- Underhill PA, Passarino G, Lin AA, Marzuki S, Oefner PJ, Cavalli-Sforza LL, and Chambers G (2001b) Maori origins, Y-chromosome haplotypes, and implications for human history in the Pacific. *Human Mutation* 17:271-280.
- Underhill PA, Passarino G, Lin AA, Shen P, Lahr MM, Foley RA, Oefner PJ, and Cavalli-Sforza LL (2001c) The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Annals of Human Genetics* 65:43-62.
- Underhill PA, Shen P, Jin L, Lin AA, Mehdi S, Jenkins T, Vollrath D, Davis R, Cavalli-Sforza LL, and Oefner PJ (1997) Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Research* 7:996-1005.
- Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang W, Kauffman E, Bonne-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd J, Mehdi S, Seielstad MT, Wells R, Piazza A, Davis R, Feldman MW, Cavalli-Sforza LL, and Oefner PJ (2000) Y chromosome sequence variation and the history of human populations. *Nature Genetics* 26:358-361.
- Whitfield L, Sulston J, and Goodfellow P (1995) Sequence variation of the human Y chromosome. *Nature* 378:379-380.
- Wilder JA, Kingan SB, Mobasher Z, Pilkington MM, and Hammer MF (2004a) Global patterns of human mitochondrial DNA and Y-chromosome structure are not influenced by higher migration of females versus males. *Nature Genetics* 36:1122-1125.
- Wilder JA, Mobasher Z, and Hammer MF (2004b) Genetic evidence for unequal effective population size of human females and males. *Molecular Biology and Evolution* 21:2047-2057.

Wright S (1931) Evolution in mendelian populations. *Genetics* 16:97-159.

Wurm SA (1975) New Guinea Area Languages. In SA Wurm (ed.): *Papuan Languages and the New Guinea Linguistic Scene*. Canberra: Pacific Linguistics.

YCC (2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Research* 12:339-348.

Zimdahl H, Schiefenhovel W, Kayser M, Roewer L, and Nagy M (1999) Towards understanding the origin and dispersal of Austronesians in the Solomon Sea: HLA class II polymorphism in eight distinct populations of Asia-Oceania. *European Journal of Immunogenetics* 26:405-416.