Genetic history and pre-Columbian Diaspora of Chibchan speaking populations: Molecular genetic evidence

by

Copyright 2008
Phillip Edward Melton
M.A., University of Kansas, 2005

Submitted to the Department of Anthropology and the Faculty of the Graduate School of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy

____
Chairperson
____

____

____

____

Date Defended:____________________________

The Dissertation Committee for Phillip E. Melton certifies
that this is the approved version of the following dissertation:

Genetic history and pre-Columbian Diaspora of Chibchan speaking populations: Molecular genetic evidence

by

Copyright 2008
Phillip Edward Melton
M.A., University of Kansas, 2005

Submitted to the Department of Anthropology and the Faculty of the Graduate School of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy

____
Chairperson

____

____

____

____

Date Defended:____________________________

Abstract

This dissertation examined Y-Chromosome and mitochondrial DNA (mtDNA) genetic variation in 230 individuals from five (Rama, Chorotega, Huetar, Maléku, and Guaymi) indigenous populations inhabiting lower Central America in order to determine the evolutionary history and biological relationship among Chibchan-speakers and neighboring groups. Mitochondrial genetic diversity observed in Chibchan populations indicates a biological relationship with two Mesoamerican groups (Chorotega and K'iche Maya). However, Chibchan populations are biologically differentiated from eastern and Andean South American indigenous groups. Y-chromosome variation demonstrates a shared paternal biological relationship between Mesoamerican and northern Chibchan populations, whereas southern Chibchan and South American groups demonstrate a closer genetic association. Genetic diversity values in all five study populations were higher for Y-chromosome and lower for mtDNA haplotypes. Low mtDNA diversity and positive neutrality tests statistic values indicate that genetic drift has operated on this locus in Chibchan populations. Coalescent dates based on two haplogroup A2 (16360 and 16187) nodes indicate the divergence of Chibchan groups from earlier Paleoindian groups between 8,000 and 10,000 years ago. In addition, a genetic discontinuity was detected in Chibchan populations and is associated with the region around Lake Nicaragua. This discontinuity is connected with rising sea levels that occurred 8,000 years before present, separated lower Central from North and South America, and isolated the predecessors of modern Chibchan populations. Archaeological and linguistic evidence is used to support the endogenous development of Chibchan populations in lower Central America. Difference between low maternal and high paternal genetic diversities are interpreted as the result of two male dominated migrations. The first migration occurred with the arrival of Nahua and Oto-Manguean populations on the Pacific coast of Costa Rica and Nicaragua between 800 and 1000 A.D. The second migration occurred with European contact in the early sixteenth century. This study is concluded by supporting mtDNA evidence for the endogenous development of Chibchan-speaking populations in lower Central America. However, recent demographic events impacted the paternal genetic structure of these groups.

This dissertation is dedicated to Jennifer, Spencer, Xavier, and Cordelia for their encouragement, understanding and patience

Acknowledgements

I would like to thank the indigenous people who participated in this study and allowed for this study to be possible. I would especially like to thank Cleveland Macrea and Jerry Macrea for their hospitality and courtesy during the fieldwork with the Rama.

I would also like to thank my committee members: Dr. Michael Crawford; Dr. John Hoopes; Dr. Bart Dean; Dr. Jim Mielke; and Dr. William Woods; my fellow collaborators on this project, Dr. Ramiro Barrantes, and Norberto Baldi from the University of Costa Rica; my fellow students from the Laboratory of Biological Anthropology specifically Kristin Young; Anne Justice, Geetha Chitoor, Kritie Beaty, Chris Krawzack, Dr. Mark Zlojutro, Dr. Rohina Rubicz, and Orion Graf. The Department of Anthropology office staff: Judy Ross, Carol Archinal, and Kathleen Womack; and of course my family for providing me encouragement throughout the entire process.

This research was funded by University of Kansas a General Research Fund and an International studies grants awarded to Dr. Michael Crawford plus a Carroll D. Clark award from the department of Anthropology awarded to the author.

Table of Contents

CHAPTER ONE: INTRODUCTION 1

CHAPTER TWO: LITERATURE REVIEW 7
- Molecular Genetic Overview 7
  - Mitochondrial DNA 10
    - Native American mtDNA 16
    - mtDNA and the origins of Native Americans 19
    - mtDNA and the peopling of South America 21
  - Y-Chromosome 23
    - Native American Y-chromosomes 26
    - Y-chromosomes and the peopling of the Americas 29
    - Y-chromosomes and peopling of South America 32
- Language and the Peopling of the Americas 33
- Genes and Language 37
- The origins of Agriculture in the Americas 44
- Summary 51

CHAPTER THREE: CHIBCHAN POPULATION BACKGROUND 53
- Chibchan populations 53
- Chibchan cultural history 55
  - Hunter/Gatherer Period 57
  - Specialized Domestication 57
  - Early Tribal Social Agriculture 58
  - Specialized Agriculture 59
  - Chiefdoms and Increased Specialization 60
  - Post European Contact 61
- Genetic of Chibchan Populations 62
  - Classical Genetic Markers 63
  - mtDNA 70
  - Y-chromosome 73
  - Genetics summary 77
- Chibchan populations Divergence 78
- Mesoamerican-Chibchan Boundary 80
  - Rama 81
  - Maléku 82
  - Huetar 83
  - Guaymi 84
  - Chorotega 84
- Summary 85

CHAPTER FOUR: MATERIALS AND METHODS 87
Field Methods 87
Laboratory Methods 89
DNA Extraction 89
mtDNA analysis 90
Y-chromosome analysis 95
Analytical Techniques 98
Diversity Measures 98
Measures of Selective Neutrality 99
Multidimensional Scaling Plots 101
Median-Joining Networks 103
Heterozygosity vs. r_{ii} 104
Population structure 105
Divergence times 107
Phylogeographic analysis 108
Admixture estimates 110
Summary 111

CHAPTER FIVE: RESULTS 113
MtDNA 113
Restriction Length Polymorphisms 113
HVS-I sequences 118
Diversity and Neutrality Measures 120
Median Joining Networks 122
Multidimensional Scaling Plot 127
Gene diversity versus r_{ii} 128
Phylogeographic Analysis 129
SAMOVA 130
Monmonier's Algorithm 131
Interpolated Genetic Landscape 133
Chibchan Divergence Times 133
Y-Chromosome 134
Single Nucleotide Polymorphisms 135
Short Tandem Repeats 137
Diversity Measures 140
Y-Chromosome STR Median Joining Networks 140
Multidimensional Scaling Plot 144
Gene diversity versus r_{ii} 144
Classical Genetic Polymorphims 147
Diversity Measures 148
Multidimensional Scaling Plot 148
Heterozygosity versus r_{ii} 150
Population Structure 152

Wright's F_{ST} 152
Analysis of Molecular Variance (AMOVA) 154
Admixture Estimates 159
Summary 160

CHAPTER SIX: DISCUSSSION: 162
Genetic Diversity 162
Mesoamerican-Chibchan Population Relationships 173
Maternal vs. Paternal Genetic histories 178
Genetic drift vs. Gene Flow 184
Phylogeography, Paleoecology and Chibchan Origins 188
Summary 193

CHAPTER SEVEN: CONCULUSION 195

LITERATURE CITED 199

APPENDIX A 230

List of Figures

Figure 1: Central America geographic areas 2

Figure 2: mtDNA genome with positions of major mtDNA haplogroups 13

Figure 3: Geographic distribution of mtDNA haplogroups 14

Figure 4: Schematic diagram of the Y-chromosome 24

Figure 5: Geographic location of Y-chromosome haplotypes 26

Figure 6: Mesoamerican linguistic area 41

Figure 7: Isthmo-Colombian area 42

Figure 8: Chibchan language family 43

Figure 9: Agricultural regions in the Americas 45

Figure 10: Central America cultural areas 53

Figure 11: Lower Central America culture area and archaeological sites 56

Figure 12: Geographic location of the five study populations 88

Figure 13: Comparative populations for mtDNA 116

Figure 14:Median Joining Network for Chibchan populations 123

Figure 15: Median Joining Network for mtDNA haplogroup A2 124

Figure 16: MDS Plot for mtDNA HVS-I sequences 128

Figure 17: Gene diversity vs. r_{ii} for mtDNA HVS-I sequences 129

Figure 18: MDS – SAMOVA plot for mtDNA HVS-I sequences 131

Figure 19: Delaunay triangulation using Monmonier's Algorithm 132

Figure 20: Interpolated genetic landscape for mtDNA HVS-I sequences 133

Figure 21: Comparative populations for Y-chromosome 136

Figure 22: Median joining network for Y-chromosome haplogroup Q3 143

Figure 23: Median joining network for Y-chromosome haplogroup R1b 143

Figure 24: MDS plot of Y-chromosome STR haplotype data 145

Figure 25: Gene diversity vs. r_{ii} for Y-chromosome STR haplotypes 146

Figure 26: Comparative populations for classical genetic markers 147
Figure 27: MDS plot for classical genetic markers 150
Figure 28: Heterozygosity vs. r_{ii} for classical genetic markers 151

List of Tables

Table 1: Native American mtDNA coding region RFLP and HVS-I sites 17

Table 2: Number of participants from the five study populations 89

Table 3: Primer pairs for mtDNA haplogroups 91

Table 4: Y-chromosome STRs used in this study 95

Table 5: Primers for Y-SNP analysis 97

Table 6: Native American mtDNA haplogroups for study populations 114

Table 7: Native American mtDNA haplogroups – comparative populations 115

Table 8: mtDNA HVS-I sequences for five study populations 119

Table 9: Diversity values for mtDNA HVS-I sequences 122

Table 10: mtDNA divergence times for Chibchan populations 134

Table 11: Y-Chromosome SNPs in study and comparative populations 137

Table 12: Y-chromosome STR variation in five study populations 138

Table 13: Y-Chromosome STR Diversity values 141

Table 14: Gene diversity for classical genetic markers 149

Table 15: Wright's F_{ST} Values for Cultural Area and Language Affiliation 153

Table 16: mtDNA AMOVA between geographic regions 156

Table 17: mtDNA AMOVA between Chibchan linguistic stocks 157

Table 18: Y-chromosome AMOVA between geographic regions 158

Table 19: Admixture estimates for five study populations 160

I: INTRODUCTION:

The indigenous inhabitants of Central America have played an important role in understanding the population dynamics and cultural development in the Americas, due to their unique geographic location spanning the two American continents. Geographically, the region is divided into two areas: 1) Mesoamerica; and 2) lower Central America (figure 1). Until recently, the majority of research focused on complex cultures and social stratification in Mesoamerica, while lower Central America was considered to be an intermediate region dominated by autonomous chiefdoms heavily influenced by state level societies from Mesoamerica and the central Andes. However, new evidence contests this idea and supports the existence of a major cultural area dominated by Chibchan-speakers (Hoopes and Fonseca 2003). Kirchoff (1943) was the earliest scholar who attempted to identify a "Chibchan" cultural area in lower Central America. This concept of a major Pre-Columbian cultural complex dominated by Chibchan-speaking populations was neglected by other researchers, who considered the area as peripheral and intermediate between Mesoamerica and the Andes (Hoopes and Fonseca 2003; Sheets 1992). On the other hand, recent research in lower Central America has refuted the notion of the region as a "cultural intermediary". Current archaeological (Cooke 2005; Hoopes 2005), biological (Barrantes *et al.* 1990; Kolman and Bermingham 1997; Melton *et al.* 2007), and linguistic (Constenla 1991) evidence support Kirchoff's explanation for the existence of a major Chibchan cultural area. This research also indicates that Chibchan populations have inhabited this region

continuously for the last 14,000 years and has supported the endogenous development of these groups from earlier human inhabitants (Barrantes *et al.* 1990; Constenla 1991; Kolman and Bermingham 1997; Cooke 2005). Recent evidence also supports a biological relationship between northern South and Central American Chibchan populations (Melton et al. 2007). However, research into the biological relationship between Mesoamerican and Chibchan populations and their divergence from earlier human groups in the region remains unresolved. This dissertation examines mitochondrial DNA (mtDNA) and Y-chromosome genetic variation in four Chibchan-speaking populations (Rama, Zapáton Huetar, Guaymi, and Maléku) and one Oto-Manguean speaking group (Chorotega) in order to determine their biological relationship to each other, as well as attempt to reconstruct Chibchan genetic history.

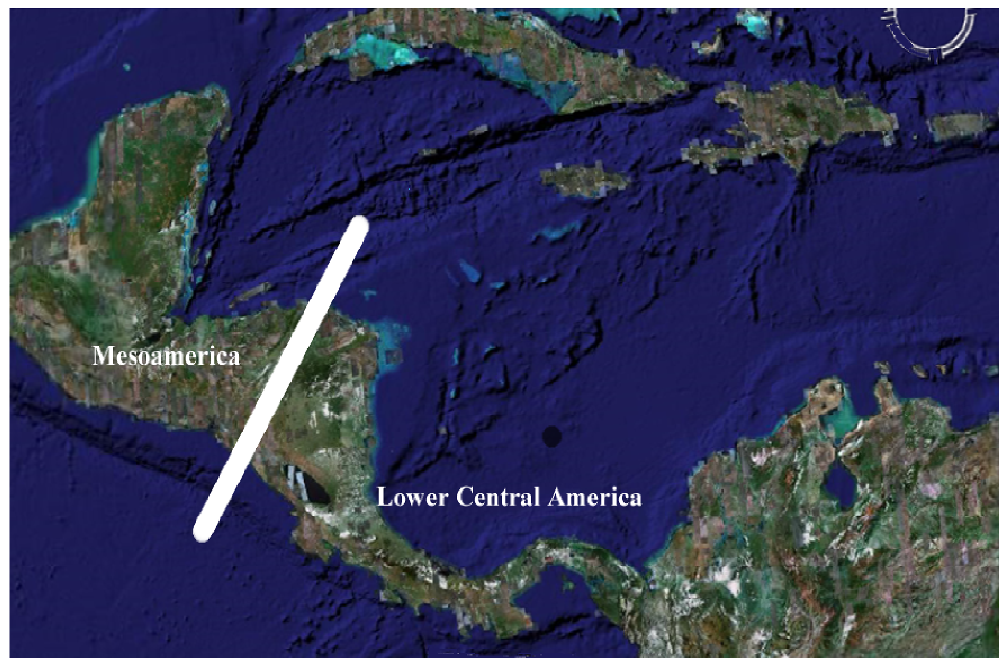

Figure 1: Central America divided by geographic region.

The term "Chibchan" is a useful identifier for both the autochonous populations of lower Central and northern South America is supported by numerous lines of evidence from archaeology (Hoopes and Fonseca 2003; Hoopes 2005), population genetics (Barrantes *et al.* 1990; Kolman and Bermingham 1997; Melton *et al.* 2007) and historical linguistics (Constenla 1991; 1995). Chibchan refers to a family of Native American languages spoken from eastern Honduras to Venezuela. Similar to the terms "Maya" and "Mayan", which refer to another heterogeneous group of approximately 30 indigenous Central American languages that are amalgamated under one label, but in the past were never subsumed under a distinct political entity, "Chibchan" refers to a wide array of ethnic identities. Modern Chibchan-speakers range from the Pech and Rama (who inhabit the tropical Mosquito coast of Nicaragua and Honduras) to the present day Kogi, Ijka, and Arsario (Wiwa) (who reside in the mountainous Sierra Nevada de Santa Marta in Colombia). Based on archaeological evidence, these populations hint at ancient relationships that indicate a rich shared cultural history throughout the region (Hoopes and Fonseca 2004). However, the impact and severity of European contact dramatically altered the histories of these indigenous populations. Recent advances in molecular genetics may provide insight into the genetic history of these Chibchan populations.

Over the past twenty years, methodological and theoretical progress in molecular genetics has enhanced the ability to test scientific hypotheses regarding the origins and interrelationships of human populations. Two of the most widely used molecular applications for studying genetic history are mitochondrial DNA (mtDNA)

and the Y-chromosome. Both these genetic regions are useful because they can be used to test the maternal (mtDNA) or paternal (Y-chromosome) evolutionary history of a single group, or a number of historically linked populations. Recent research into genetic variation within these molecular regions has allowed for considerable insight into the origins of Native American populations. Initial evidence from mtDNA revealed that Amerindian populations are characterized by a series of mutations that cluster into five haplogroups termed A2, B2, C1, D1, and X2a (Schurr *et al.* 1990; Torroni *et al.* 1993; Schurr and Sherry 2004). This current evidence from complete mtDNA genome sequencing has further enhanced our understanding of Native American origins (Tamm *et al.* 2007; Achilli *et al.* 2008). Research on the Y-chromosome has also identified three haplogroups termed Q, Q3, and C that can be used to determine Native American genetic ancestry (Karafet *et al*. 1999; Lell *et al*; 2002; Zegura *et al.* 2004). This molecular genetic evidence has increased our understanding of both the initial peopling and the post-colonization population dynamics of Native American populations.

Chibchan-speaking populations of lower Central and northern South America provide one group that has important implications for understanding the subsequent genetic history of Native American populations. Chibchan populations share linguistic and cultural characteristics that indicate these groups may share an evolutionary history. These shared cultural characteristics include iconography, settlement patterns, and social structure (Hoopes and Fonseca 2003). In addition, lower Central America has a well defined archaeological record that demonstrates

continuous occupation of the region from the Paleoindian era through European contact (Cooke 2005). This long tern human occupation allows for the comparison between biological and cultural history in order to understand the evolution of Chibchan populations.

The goals of this dissertation are: 1) to characterize Y-chromosome and mtDNA genetic variation from five indigenous Central American populations; 2) compare these data to molecular and classical genetic polymorphisms obtained from the literature; and 3) investigate the genetic history and the biological relationships of Chibchan-speaking populations to surrounding Native American groups using molecular and classical genetic data. The molecular markers are used to characterize maternal and paternal lineages, while classical genetic polymorphisms are investigated as proxies for bi-parentally inherited markers. The four specific questions addressed by this research are:

1) What evolutionary forces have previously operated on Chibchan populations? How have they impacted these groups? and Do they differ between mtDNA and the Y-chromosome?

2) What is the maternal/paternal genetic relationships of Chibchan populations with neighboring Mesoamerican and South American indigenous groups?

3) What can mtDNA and Y-chromosome variation reveal regarding the genetic history of Chibchan populations?

4) Are discontinuities observable in the genetic variation of Chibchan and neighboring populations and if they are present how can they be related to the cultural and genetic history of the region.

This dissertation is divided into six additional chapters. Chapter two summarizes previous research on the anthropological genetic studies of Native American populations using Y-chromosome and mitochondrial DNA (mtDNA) data, linguistic diversity in the Americas, and the origins of pre-Columbian agriculture in the Americas. Chapter three reviews background information on Chibchan populations from lower Central America along with previous archaeological and genetic research in the region in order to better understand Chibchan evolutionary history. The materials and methods chapter discusses the fieldwork, laboratory and analytical methods used for data analyses. The results are presented in chapter four for molecular and classical genetic data. The relevance of these results to understanding the genetic and cultural of history of Chibchan populations are discussed in chapter five. The last chapter provides a summary of these results and their implications for future research.

II. LITERATURE REVIEW

This chapter summarizes previous research on the anthropological genetics of Native American populations using Y-chromosome and mitochondrial DNA (mtDNA) data, linguistic diversity, and the origins of Pre-Columbian agriculture. In addition, this chapter provides a theoretical overview of current research pertaining to these summaries and how they relate to the initial peopling of the Americas in order to better understand the evolutionary history of the Chibchan populations in Central and South America.

MOLECULAR GENETIC OVERVIEW:

Advances in molecular biology and computer technology have led to a number of useful genomic markers for investigating human genetic population structure and history. This genetic information is inherited from the mother (mtDNA), the father (Y-chromosome) and from both parents (autosomal markers). When alleles or DNA sequence variants occur in the human genome in a population with a frequency above one percent they are termed genetic polymorphisms (Jobling *et al.* 2004). The presence of these genetic polymorphisms in a group of individuals requires an evolutionary explanation that excludes mutation. These alternate evolutionary explanations include natural selection (directional and balanced), gene flow (admixture), or genetic drift (bottleneck and founder effect). Directional selection and genetic drift reduce the amount of genetic variation within a population and increase genetic differences between groups. Whereas, balancing selection and

gene flow increase genetic variation within populations and decreases variation among groups. When genetic polymorphisms are studied at the population level, they provide insight into the genetic history of a population. Subsequent correlation of genetic history with archaeological, environmental, and cultural data provides greater understanding of evolutionary, ecological, or historical processes. Recent molecular research in anthropological genetics has focused on broad-based theoretical questions of major demographic expansions. These studies include: the origins of modern humans (Cann *et al.* 1987; Vigilant *et al.* 1991; Sherry *et al.* 1994; Garrigan and Hammer, 2006); the spread of farming in Europe (Sokal *et al.* 1991; Rosser *et al.* 2000; Currat and Excoffier, 2005); the peopling of the Americas (Schurr *et al.* 1990; Torroni *et al.* 1993; ; Karafet *et al.* 1999; Schurr and Sherry 2004; Zegura *et al.* 2004; Tamm *et al.* 2007; Wang *et al.* 2007) and; the evolutionary relationship between modern *H. sapiens* and Neanderthals (Krings *et al.* 1997; Green *et al.* 2006; Noonan *et al.* 2006). Increasingly, greater importance has been placed on examinations of microevolutionary events that precipitated large scale population expansion or genetic bottlenecks on a smaller geographic scale (Barrantes *et al.* 1990; Redd and Stoneking 1999; Crawford 2007). Two of the most widely used human genetic regions for understanding the Pleistocene peopling of the Americas are mtDNA and the Y-chromosome.

The human genome (DNA) consists of 22 pairs of autosomal chromosomes, two sex chromosomes (X or Y), and mtDNA. This genome is subdivided into exons (executing regions), introns (intervening segments), and other non-coding regions.

Exons are DNA sequences or series of sequences that maintain the instructions required for synthesizing proteins. There are between 20,000 to 25,000 exons in the human genome (International Human Genome Sequencing Consortium 2004). The majority of these exons are found in the autosomes, with only 1,529 exons being located on the X-chromosome and 344 on the Y-chromosome (NCBI Human Genome Map 2007). The remaining 98.5% of the genome consists of introns, repeat elements, transposons, and pseudogenes. Introns are genetic segments that are transcribed but are spliced from messenger RNA (mRNA) prior to protein translation. Repeat elements are genetic sequences characterized by a repetitive sequence of nucleotide base pairs (i.e, GACGACGAC). Transposons are sequences of DNA that move around to different positions in the genome of a cell. Pseudogenes are nonfunctional relatives of known exons that have lost their protein coding abilities.

The current biological function of these genomic regions is unknown and genetic polymorphisms in this region are considered selectively neutral. Therefore, the absence of natural selection as an evolutionary explanation for genetic variation in human noncoding regions allows only for two potential evolutionary outcomes, genetic drift or gene flow. This does not mean that mutation does not occur in these noncoding regions, but that it does not exceed a frequency of 1% within a population. The dichotomous relationship between gene flow (more genetic variation within populations, less genetic differences between populations) and genetic drift (less genetic variation, more genetic differences) permits the testing of hypotheses

regarding the evolutionary history of a population or group of populations using human genomic regions.

Advances in biotechnology have recently accelerated the ability and understanding of the human genome. The advent of automated sequencing technology permits the determination of individual DNA base pairs (Sanger *et al.* 1977). The subsequent development of polymerase chain reaction (PCR) methodology in the mid-1980s allows for large numbers of the same sequence to be replicated (Saiki *et al.* 1985). Both these techniques permit greater characterization of molecular polymorphisms in the human genome. A number of molecular polymorphisms are useful for understanding population structure and dynamics in the human genome including: single nucleotide polymorphims (SNPs); insertion/deletions (indels); microsatellites (STRs), minisatellites (VNTRs), retroelements (*Alus*, LI, endogenous retroviruses) (Rubicz *et al.* 2007).

MITOCHONDRIAL DNA

Mitochondrial DNA is one of the most informative and widely used genomic regions for understanding Native American genetic variation. This wide use is due to its elevated mutation rate (Brown *et al.* 1979), small number of base pairs, high cellular copy number, maternal transmission (Giles *et al.* 1980), lack of recombination (Olivo *et al.* 1983), and small effective population size (1/4 of autosomal DNA). The mitochondrial genome is located within the energy producing mitochondria in the cytoplasm of the cell. This genome consists of 16,569 base pairs

(bp) divided into a coding region with 37 exons (22 transfer RNAs, 13 proteins, and two other RNAs) and an intron control region separated into three hypervariable segments (HVS-I, HVS-II, and HVS-III ~400 bp each). The mutation rate of mtDNA is high and occurs at five to ten times the rate of nuclear DNA, with the mtDNA control region higher than the rate of the coding section. The mutation rate in the coding region is estimated as 3.2% per million years (Francalaci *et al.*1999) and the rate increases to 8.4% in the control region (Vigilant *et al.* 1989). Variation in the mtDNA control region is not evenly distributed, with a higher number of polymorphic sites being present in HVS-I then HVS-II. Mitochondrial DNA is maternally inherited, meaning that it is passed from mother to all of her offspring. However, only daughters subsequently pass it onto their progeny (Giles *et al.* 1980) which implies that mtDNA does not undergo recombination and is passed unaltered from generation to generation. Therefore, any observed variation in mtDNA can be attributed to mutation and the timing of the event can be estimated using the aforementioned mutation rates.

Early human mtDNA genetic history studies used high resolution restriction fragment length polymorphism (RFLP) analyses and focused on the entire mtDNA genome (Cann *et al.*1987; Schurr *et al.*1990; Torroni *et al.*1992). This research focused on the conservative coding region and suggested that some selectively neutral mutations occurred only once in human evolutionary history. These unique mutational events allowed for the grouping of certain shared mtDNA polymorphisms into lineages called haplogroups that were defined by their RFLPs. Two different

nomenclatures were originally proposed: 1) Roman numerals used by Horai *et al.* (1993) and; 2) the English alphabet employed by Torroni *et al.* (1992). The English alphabet nomenclature achieved greater acceptance and is currently used by geneticists. Figure 2 displays the position of these haplogroups in the mtDNA genome, along with their corresponding RFLPs and associated English alphabet notation. Several of these haplogroups have been found to be continent specific and are used to trace the migration patterns of human populations (Francalacci *et al.*1999). Figure 3 illustrates the major geographic locations of these haplogroups along with their hypothetical migration routes (Wallace and Lott, 2004).

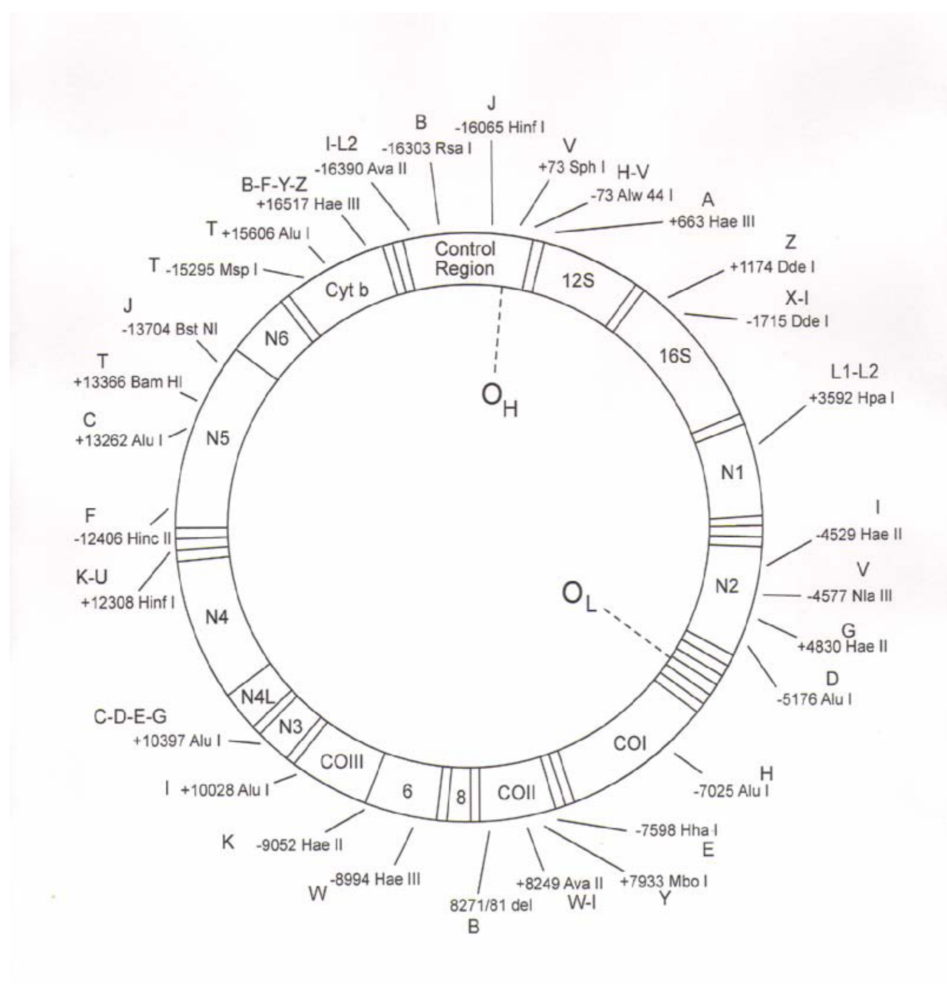

Figure 2: mtDNA genome with positions of major mtDNA haplogroups and their associated RFLP cut sites (Rubicz *et al.* 2007)**.**

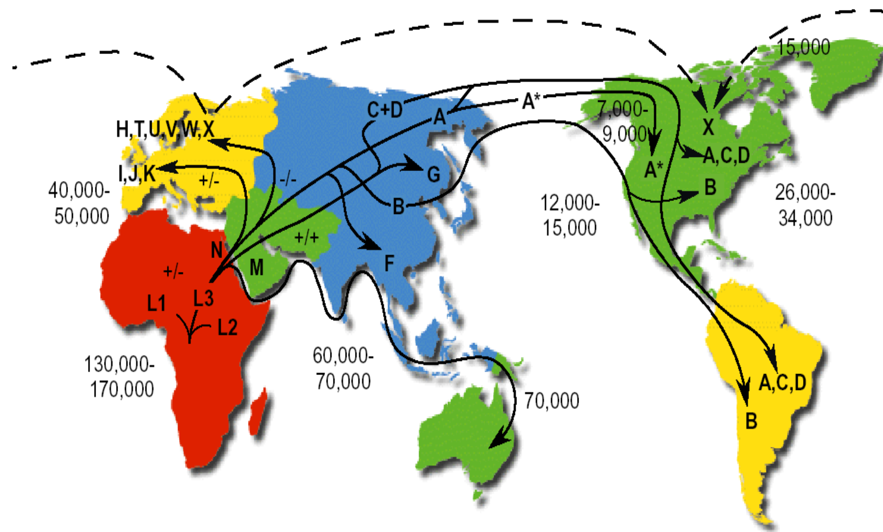

Figure 3: Geographic distribution of human mtDNA haplogroups. Time estimates are thousands years ago (ya) and based on a mutation rate of 2.2-2.9%/million years. +/+/+-, and -/- refers to *Dde*I 10394/*Alu*I 10397, * refers to +*Rsa*I (Wallace and Lott 2004).

Subsequent human mtDNA research focused on the high variability within the control region and applied direct DNA sequencing to the HVS regions. These analyses allowed for deeper understanding of population differentiation at the micorevolutionary level (Vigilant *et al.*1989; Vigilant 1991). However, there is a certain risk in only focusing on the mtDNA control region. The high mutation rate found in this region increases the potential for homoplasy occurring at the same nucleotide site. These redundancies result in a reduction of observed genetic variability between diverged human populations and complicate further phylogenetic analysis. In early studies that applied only this method, sequences detected in different populations often clustered together, with no apparent phylogenetic relationship (Piercy *et al.* 1993). There are two potential alternative explanations for

reduced mtDNA control region genetic variation: 1) it represents an ancient mutation derived from a most recent common ancestor (MRCA) and is present in all individuals within a population; or 2) a change occurs at a nucleotide site with a higher than expected mutation rate in different mtDNA lineages reducing value for establishing phylogenetic relationships. Analyses that do not account for these explanations and assign all mtDNA coding region mutations the same weight provide poor resolution in the understanding of microevolutionary history at the population level (Francalacci *et al.* 1999).

Current mtDNA research into human genetic history consists of utilizing both RFLPs and HVS sequencing because both regions are inherited as a unit (Torroni *et al.* 1996). This approach enhanced mtDNA research because HVS sequencing information is kept intact but arranged according to its phylogenetic relationship through association with specific RFLPs. This relationship demonstrated that some HVS site specific polymorphisms always coincide with certain RFLP sites from the coding region. In addition, these associations allow for an evolutionary balance between the two mtDNA regions and permits researchers to verify their results. Recently, complete sequencing of the mtDNA genome has been applied to the understanding of human evolutionary history (Rieder *et al.* 1998; Tamm *et al.* 2007; Torroni *et al.* 2006, Kitchen *et al.* 2008, Achilli *et al.*2008). Complete sequencing allows for the highest possible phylogenetic resolution because each of the 16,569 bases pairs of the mtDNA genome is identified in all individuals. However, the cost of this methodology often precludes it from being widely used at present.

Native American mtDNA

There are currently four major (A2, B2, C1, and D1) and five potential (C4c, D2a, D3, D4c, and X2a) founding mtDNA haplogroups identified in Native American populations (Tamm *et al.* 2007, Achilli *et al.* 2008). These haplogroups are characterized by RFLPs, sequence variation, and a 9 base pair deletion in the COII gene of the mitochondrial genome (Table 1). These haplogroups account for between 95-100% of known mtDNA polymorphisms from indigenous Amerind populations (Schurr and Sherry, 2004). Variable frequencies of these haplogroups are found in North, Central, and South American populations. Six of these mtDNA haplogroups (A2, B2, C1, D1, D2a and X2a) have been identified in pre-Columbian skeletal materials found in North America (Fox 1996; Stone and Stoneking, 1998; O'Rourke *et al.* 2000; Gilbert *et al.* 2008). The remaining three mtDNA haplogroups (C4c, D3, and D4c) have not, as yet, been detected in skeletal populations because they have never been investigated. Haplogroup X2a is restricted to North American populations and has never been identified in any ancient or living Central or South American groups (Salzano 2002; Dornelles *et al.* 2005). These mtDNA haplogroups have also been found in low frequencies in indigenous populations from central and northeast Asia, suggesting this area as a potential source for maternal lineages in Native American populations (Schurr and Sherry 2004). Other researchers propose that these haplogroups do not represent all haplogroups present in the Americas prior to European contact. These researchers suggest that other founding haplogroups were lost due to severe depopulation that followed European contact and/or through

genetic drift (Bailliet *et al.*1994; Easton *et al.* 1996; Lorenz and Smith, 1996; Rickards *et al.*1999). Subsequent research on these indigenous populations documented that most of these unknown haplogroups were derivatives of haplogroups A-D, belonged to haplogroup X2a, or were the result of non-native admixture (Schurr and Sherry, 2004).

Table 1: Correlation between coding region RFLP sites and control region variants in human mtDNA.

Native American Haplogroup	Geographic distribution	16111	16189	16217	16223	16278	16290	16298	16319	16327	16362	Coding Region RFLP sites
CRS*		**C**	**T**	**T**	**C**	**C**	**C**	**T**	**G**	**C**	**T**	
A2	Asia-America	T			T		T		A		C	+663 *Hae*III
B2	Asia-America		C	C								9bp deletion
C1	Asia-America				T			C		T		-13259/+13262 *Hinc*II/*Alu*I
D1	Asia-America				T						C	-5176 *Dde*I
X2A	Asia, Europe, America		C		T	T						-1715*Dde*I

*Revised Cambridge Reference Sequence (Anderson *et al.* 1981; Anderson *et al.* 1999)

Mitochondrial DNA haplogroups occur in Native American populations in varying frequencies. Haplogroup A2 displays the most diversity and is divided into 13 subclades (A2c, A2d, A2d1, A2d2, A2e, A2f, A2g, A2h, A2i, A2j, A2j1, A2k, and A2k1) (Achilli *et al.* 2008). A2 occurs at differing frequencies throughout the Americas but has also been detected in Siberian populations (Pakendorf *et al.* 2006; Tamm *et al.* 2007). Haplogroup B2 contains at least four sub-haplogroups (B2a-B2d). B2a is found throughout North America and is common among the Pima, Ojibwa, and Navajo. B2d is frequent in lower Central and northern South American populations and has been detected in both the Wayuú from Colombia and the Ngöbé

from Panama (Achilli *et al.* 2008). Haplogroup C1 is divided into three subclades (C1b, C1c, and C1d) located in populations throughout the Americas (Achilli *et al.* 2008). Haplogroup C4c is only known from a single population (Ijka) in Colombia (Tamm *et al.* 2007; Melton *et al.* 2007). Haplogroup D2 consists of two sister subclades (D2a, D2b) with D2b being found only in Siberia and D2a only in Arctic and subarctic populations (Tamm *et al.*2007). Haplogroup X2a has only been found in North American populations but has yet to be detected in South American populations (Dornelles *et al.* 2005).

A clinal distribution of mtDNA haplogroup data is evident in South America involving haplogroups A2 and D1. Haplogroup A2 occurs at high frequencies (>50%) in lower Central and northern South America, but is absent in native populations from the southern cone (Kolman *et al.* 1995; Moraga *et al.* 2002). Haplogroup D1 mirrors A2, with high frequencies in the south and almost a complete absence in northern populations (Fox 1996; Keyeux *et al.* 2002). One Central American population, the Huetar, show moderate frequencies of haplogroup D1 (26%), while populations from Argentina and Chile exhibit the highest frequencies (Santos *et al.* 1994; Fox 1996). Haplogroup B2 is detected in high frequencies among Andean populations, some populations from the Brazilian highlands and the Gran Chaco, but is absent from the southern cone (Ward *et al.* 1996; Moraga *et al.* 2000; Schmidt *et al.* 2004). Haplogroup C1 is absent in Central America but is detected in groups throughout South America, while it is at its highest frequency in the Caribbean among the extinct Tainos (75%) and Ciboney (60%) from Cuba (Lalueza-Fox *et al.* 2001; Lalueza-Fox

et al. 2003). Several South American populations lack one or more haplogroups reflecting the extent that genetic drift may have played in forming the mtDNA distribution in modern Native Americas (Schurr and Sherry, 2004).

mtDNA and the origins of Native Americans

There is a consensus that ancestral Native American populations originated in Asia and migrated to the Americas over the Bering Land Bridge during the Pleistocene. However, the Asian source(s), timing, and the number of migrations remain a matter of considerable scientific debate. Mitochondrial DNA has been used extensively to examine the initial peopling of the Americas in order to determine potential entry dates, number of migrations, and Asian sources for Native Americans. Early mtDNA RFLP research provided molecular dates between 35,000 – 20,000 years ago (ya) for haplogroups A2, C1, and D1 (Schurr, 2004) and between 17,000-13,000 ya for haplogroup B2 (Brown *et al.* 1998) suggesting two migrations. Recent studies based on full mtDNA genome sequencing have resulted in more refined molecular dates between 24,000 to 17,000 ya for haplogroups A2, B2, C1, D1 (Achilli *et al.* 2008). These molecular coalescent dates provide further support for a pre-Clovis occupation of the Americas. According to the pre-Clovis model, humans entered the Americas from Siberia between 20,000-14,000 years before present (YBP, refers to radiocarbon dates and prior to 1950) and migrated along the Pacific coast (Dixon 1999; Dillehay 2000). However, these dates contrast with the "Clovis First" peopling model, that argues humans arrived later to the Americas around

11,000 YBP and migrated into the North American interior through an interglacial passageway (Bonnichsen and Turnmire 1992).

Mitochondrial DNA has been used to develop models for the number of migrations into the Americas. The earliest mtDNA RFLP research suggested two major migrations to the New World. The first migration brought haplogroups A2, C1, and D1 found in Native American populations throughout the Americas and a later migration brought B2 along the Pacific coast from Siberia. However, B2 is absent among Eskimo-Aleuts and Na-Dene speaking populations (Torroni *et al.* 1992; 1993; Schurr and Wallace 1999) indicating that these new B2 populations either bypassed the Arctic and arrived by boat from Polynesia (Cann 1994), or that Eskimo-Aleuts and Na-Dene speakers represent a separate migration (Rubicz *et al.* 2003; Zlojutro *et al.* 2006). Subsequent research on the mtDNA HVS-I region argues for a single human migration into the Americas (Merriwether *et al.* 1995; Kolman *et al.* 1996; Bonatto and Salzano, 1997 Stone and Stoneking, 1998). The current favored explanation, based on full mtDNA genome sequencing, is the Beringian Incubation Model (BIM). This model maintains a small human population (~1,000 to 5,400) (Kitchen *et al.* 2008) migrated from Siberia to Beringia where they were isolated for a significant period of time and that these founding Amerindian mtDNA haplogroups developed *in situ* in Beringia (Tamm *et al.* 2007). Then after the Last Glacial Maximum (LGM) (~20,000 YBP), human populations migrated rapidly down the Pacific coast, and this is reflected in current Native American mtDNA diversity (Tamm *et al.* 2007; Kitchen *et al.* 2008). The BIM demonstrates a strong relationship

with a pre-Clovis archaeological explanation for the peopling of the Americas. This mtDNA/archaeological association is based on overlapping coalescent mtDNA (13,900 ± 2,700 ya) (Tamm *et al.* 2007) and early archaeological radiocarbon dates from Monte Verde (14,500 YBP), which is now widely accepted among archaeologists (Lavalleé 2000). However, more recent mtDNA coalescent dates for all Native American haplogroups place the origin of these groups between 20,200 ya (± 1,600 ya) and 19,000 ya (± 1,400 ya) (Achilli *et al.* 2008), which is earlier than the more controversial sites in the Americas, including those in South America.

The majority of these migratory explanations argue for the peopling of the Americas as a combined occurrence and little research has focused on the peopling of South America as a separate evolutionary event. The South American continent was the last major inhabitable land mass populated by humans during the late Pleistocene and presents a number of ecological challenges (high altitude; tropical forest) that may have required significant cultural and biological adaptation.

mtDNA and the Peopling of South America

Archaeological and genetic research indicates that South America was initially settled by humans by 14,000 YBP (Dillehay 1999; Dixon 2001; Fuselli *et al.* 2003). However, there is no consensus on the number or timing of migrations into South America (Rothhammer and Silva 1989, 1992; Fox 1996; Lalueza *et al.* 1997; Luiselli *et al.* 2000; Moraga *et al.* 2000; Rothhammer *et al.* 2001; Tarazona-Santos *et al.* 2001). The peopling of South America remains largely unresolved because of the

unique distribution of genetic diversity and a number of controversial radiocarbon dates from archaeological sites (i.e, Taima-Taima 14,000 YBP, El Jobo 12,400 YBP, Pedra Furada 45,000 YBP) found throughout the continent (Lavalleé 2000). Indigenous South American populations inhabiting the eastern lowland regions demonstrate lower mtDNA genetic diversity values than those groups residing in or near the Andes (Lewis *et al.* 2007). Native eastern South American populations often lack one or more of the major mtDNA haplogroups, and exhibit low amounts of mtDNA haplotype diversity compared to western Andean groups. This difference in diversity values has led some researchers to suggest two migrations, one group moving along the Atlantic coast and inhabiting the East and a second group travelling down the Pacific and occupying the Andes (Fox 1996; Tarazona-Santos *et al.* 2001). An alternative view is that there was a single human migration during the late Pleistocene into South America with different evolutionary forces operating on eastern and western populations, accompanied with low gene flow between regions (Fuselli *et al.* 2003; Lewis *et al.* 2005; Lewis *et al.* 2007; Lewis and Long 2008). In this scenario, genetic drift (loss of diversity) is seen as the primary evolutionary force operating on eastern groups and gene flow (gain of diversity) is functioning on Andean populations (Lewis *et al.* 2007). This is based on a low amount of mtDNA haplotype diversity observed in eastern South American populations and being differentiated from each other based on mtDNA HVS-I sequences. Theoretically, this hypothesis is based on the idea that movement among populations is more feasible in the Andes than in the tropical forests of eastern South America. A third explanation is

that there was an initial migration of populations into the southern portion of the continent and a subsequent migration of Chibchan and Arawak speaking populations into northern South America and the Caribbean (Keyeux *et al.* 2002; Melton *et al.* 2007). This view suggests that Chibchan groups spread into northern South America from Central America and that Arawak populations moved into the region from the Caribbean. An alternative these explanations, is that natural selection is operating on coding regions in some indigenous South America populations. However, this idea requires further evidence from genetic variation within the mtDNA coding region.

Y-CHROMOSOME

The Y-chromosome is the male equivalent to mtDNA and is therefore a useful tool in evolutionary genetics studies. This chromosome includes the largest nonrecombining (95%) portion (NRY) of the human genome (Hammer and Zegura, 2002), determines male sex, spans approximately 60 million bps, represents 1% of the nuclear genomic content, and contains 307 exons (Jobling and Smith, 2003). The Y-chromosome is comprised of a heterochromatic centomere that separates the long (Yq) from the short arm (Yp) (figure 4). The NRY is flanked by two pseudoautosmal regions (PAR1 and PAR2) that undergo recombination with the X-chromosome during meiosis (Vogt *et al.* 1997). The genetically active part of the Y-chromosome, or euchromatin, contains the sex-determining region of the chromosome (SRY gene) which produces a transcription factor that turns on other genes involved in the development of the male testes from unspecified gonads (Graves 2002), and the AZFa, AZFb, and AZFc regions that have genes vital for sperm development. The

NRY (also referred to as Male-specific region or MRY) provides a depiction of the evolutionary pressures that influence population structure and may be used as a tool for reconstructing coalescent and demographic events affecting a population from a paternal perspective (Jobling and Tyler-Smith, 2003).

The earliest molecular characterization of the Y-chromosome was done with RFLP analysis, but this method identified 60 polymorphisms, only eleven of which could be determined through traditional PCR methods (Hammer 1994; Seilstad *et al.* 1994; Hammer and Horai 1995; Santos *et al.* 1995; Whitfield et al. 1995; Jobling *et al.* 1996; Underhill *et al.* 1996). One reason for this reduced RFLP number is that the

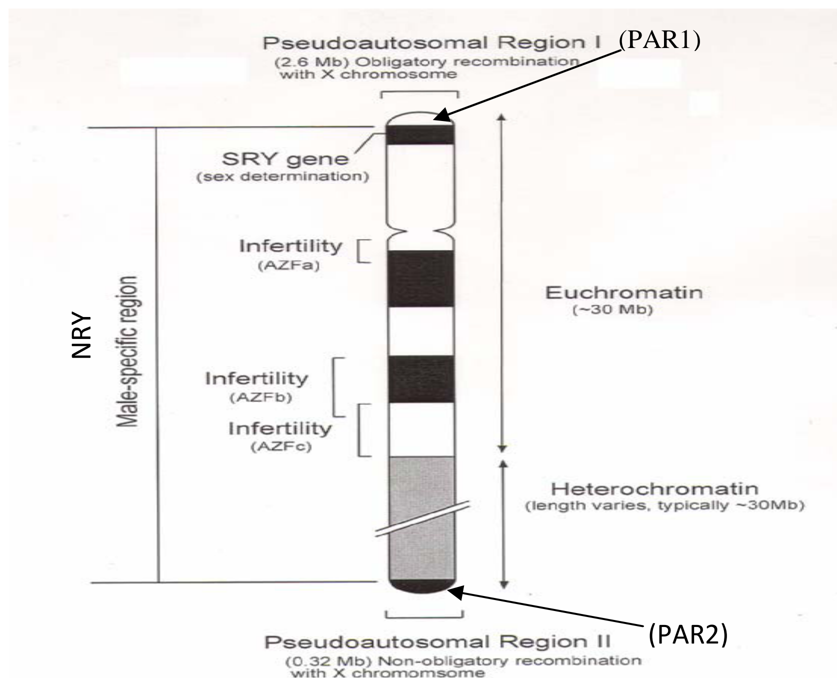

Figure 4: Schematic diagram of the Y-chromosome (Jobling *et al.* 2004). PAR1 and PAR2 are labeled on the top and bottom of the chromosome. Boxed area in the center contains the NRY (MRY) region.

NRY has fewer polymorphisms than any other portion of the human genome (International SNP Map Working Group 2001). With the application of denaturing high performance liquid chromatography (DHPLC) in the late 1990s, the number of identified polymorphisms in the region increased to over 200 SNPs and insertion/deletions (indels) (Underhill *et al.* 1997, 2000; Shen *et al.* 2000). These polymorphisms are especially useful due to their low levels of reticulations, which make them ideal for detecting stable paternal lineages that can be traced back thousands of years (Karafet *et al.* 1999; Thompson *et al.* 2000). This new application led to an increased number of investigations into Y-chromosome variation and its role in understanding human genetic history. However, often these researchers did not consistently use the same markers and applied different nomenclature making it difficult to interpret their results (Hammer and Zegura, 2002). The resulting differing nomenclatures necessitated the formation of the Y-Chromosome Consortium (YCC) (2002) in order to develop a consistent naming system for this genetic region. This YCC nomenclature was then updated again in 2008 (Karafet *et al.* 2008).

Currently, paternal lineages associated with Y-chromosome variation are characterized by derived mutations at a given biallelic locus (SNPs or indels). The current NRY haplogroup nomenclature (figure 5) consists of haplogroups identified by 21 capital English letters (A-T, and Y) and the primary SNP or indel that defines it (e.g, Q-M3) (Y-Chromosome Consortium 2002). The M refers to a mutation and the P refers to a polymorphism (Hammer and Zegura, 2002). Further categorization of these haplogroups can be made through the utilization of STRs. When compared to

the biallelic markers, these STRs aid in the estimation of population divergence, admixture and ancestry in determining the origins and microevolutionary forces affecting Native American populations.

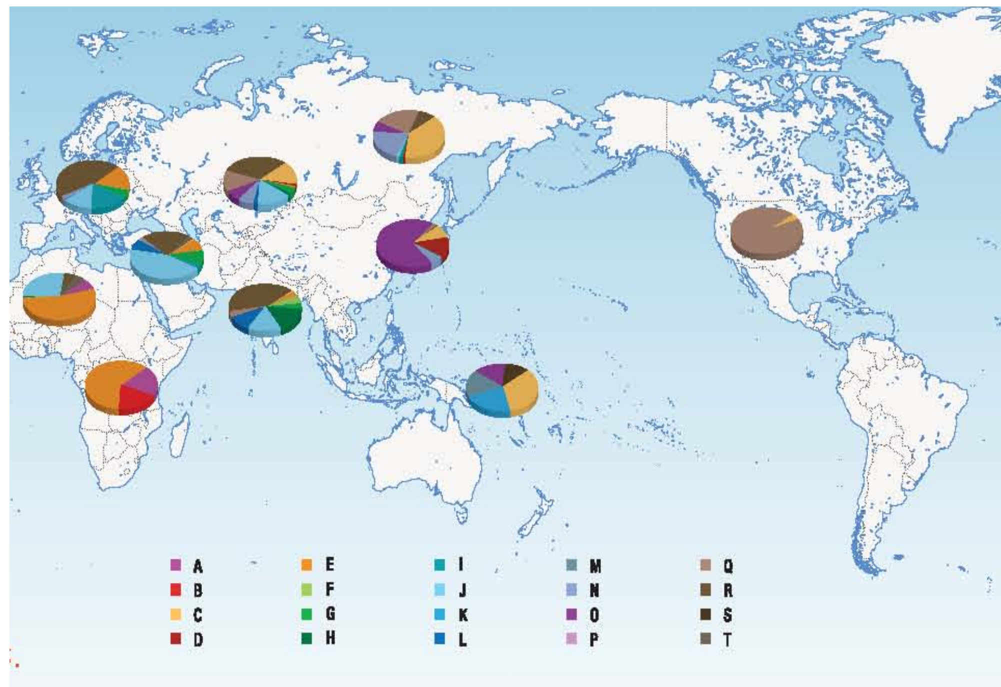

Figure 5: Geographic location of major Y-chromosome haplotypes in the World (Karafet *et al.* 2008) Native South Americans are considered entirely haplogroup Q-M3 and Native Australians and New Guineans belong entirely to haplogroup K.

Native American Y-Chromosomes

While the YCC nomenclature (Karafet *et al.* 2008) has been successfully used to identify Y-chromosome mutations, there are still discrepancies when applied to Native American populations. A number of recent studies investigated Y-chromosome variation in the Americas (Karafet *et al.* 1999; Lell *et al.* 2002; Bortolini *et al.* 2003; Zegura *et al.* 2004). A problem with these investigations is that they often

use different numbers of biallelic markers and STRs, as well as a separate terminology when applying the Y-chromosome nomenclature to Native American populations. When compounded these differences make it difficult to compare data or to relate it new research findings. As an example, Schurr (2004) suggests five founding Native American Y-chromosome haplogroups (Q-M3, R1a1-M17, P-M45, F-M89, and C-M130) that are also present in Siberia. Whereas, Zegura and colleagues (2004) claim that Native American populations are characterized by the presence of only three Y-chromosome haplogroups: (Q (M3 and P36) and C-P39) along with a third haplogroup R that is believed to represent Eurasian admixture. Schurr (2004) relied heavily on results from Lell *et al.* (2002) who investigated 12 biallelic markers and four STRs in 549 individuals from Siberia and the Americas. Wheras Zegura *et al.* (2004) examined 63 binary markers and 10 STRs from 2,344 individuals. The higher number of binary markers and STRs applied in Zegura *et al.*'s (2004) article has been adopted by the most recent YCC nomenclature (Karafet *et al.* 2008) and will be applied in this dissertation. Both these researchers demonstrate that haplotypes belonging to haplogroup Q represent the majority of all Native American Y-chromosome diversity (Schurr 2004; Zegura *et al.* 2004). Haplogroup Q-M3 (52.6%) shows the widest spatial distribution and is generally at the highest frequency in all Native American populations (Lell *et al.* 1997; Karafet *et al.*1999; Lell et al. 2002; Zegura *et al.* 2004). Underlying STR haplotypes associated with the Q-M3 haplogroup also reveal significant differences in distributions among North, Central, and South American populations (Schurr 2004). A second major Q haplogroup (P36)

occurs in approximately 24% of Native American populations and is found in indigenous North American groups (Zegura *et al.* 2004). However, its distribution in South American groups is unknown as Zegura *et al.* (2004) only included a single population (Wayuú) from this continent and all other investigations of Y-chromosome diversity in the region only investigated Q3 (Tarazona-Santos *et al.* 2001; Demarchi and Mitchell 2004).

The two remaining major Y haplogroups (R and C) found in the Americans constitute 20% of Amerindian variation and are largely restricted to North and Central American populations (Zegura *et al.* 2004). Haplogroup R (13.4%) is widely distributed throughout North and Central America but is thought to be the result of European admixture over the last 500 years (Zegura *et al.* 2004). This Y-chromosome haplogroup is believed to have arisen in Southwest Asia between 30,000 and 35,000 YBP and spread into Europe with migrating farmers (Kivisild *et al.* 2003).The most common derivative Y-chromosome haplogroup detected in Native Americans is R1b. This haplogroup is in its highest frequencies (>70%) in Iberian (Spain, Portugal) and northwestern European populations (British Isles, France, Germany, Scandinavia). Haplogroup C (5.4%) is found primarily in Na-Dene speaking groups from the southwest of North America but has also been detected in the Wayuú of Colombia (Zegura *et al.* 2004). This haplogroup appears to be a founding haplogroup as those Native American populations with the P39 SNP mutation cluster separately from other European and Asian populations (Zegura *et al.* 2004). However, Amerindian populations belonging to R haplogroups show a random pattern and are often

associated with European populations indicating male admixture in these groups (Zegura *et al.* 2004).

Other Y-chromosome haplogroups have also been found in Native American populations but are thought to be the result of recent admixture with African, European or Asian populations (Zegura *et al.* 2004; Schurr 2004). Schurr (2004) defines haplogroup F-M89 as a founding Native American haplogroup, but this has only been detected by Lell *et al.* (2002) who identified low frequencies in two Native American populations (Seminole and Boruca). However, this may be attributed to admixture, as F is a Y-macrohaplogroup (found over a large geographic range) that has been identified throughout Asia and Europe (Karafet *et al.* 2008).

Y-Chromosomes and the peopling of the Americas

As with molecular data from mtDNA, information from the Y-chromosome has been used to investigate the origins and migration patterns of Native American populations. Research on Native American Y-chromosomes has suggested that specific markers may derive from two sources in Siberia. However, once again differing terminologies often lead to confusing interpretations when contrasting the different peopling of the Americas ideas proposed from Y-chromosome variation. Schurr (2004) states that those populations belonging to Y haplogroups Q-M3 and Q-P36 derive from the ancestral state P-M45a and that these populations could be traced to central Siberia. Both these regions contain frequencies of haplogroup Q, which derive from the larger haplogroup P (Karafet *et al.* 2008). Schurr (2004) also suggests that those belonging to Y haplogroups C and R were derivatives of haplogroup P-

M45b that originated in Northeastern Siberia. However, this is based on the mistaken assumption that Y-chromosome haplogroup C is included within the larger P macrohaplogroup, which it is not. Rather, it is included within the older F macrohaplogroup (Karafet *et al.* 2008). Despite these inconsistencies, current research on Y-chromosome genetic variation has suggested some important findings regarding genetic variation in Amerindian populations.

Both mtDNA and Y-chromosome polymorphisms point to a Siberian origin for Native Americans, but the general geographic location is unclear and does not take into account the proposed mtDNA Beringian incubation model (BIM). The majority of Native American populations belong to Y-chromosome haplogroup Q, which provides the most consistent molecular dates for entry into the Americas. Molecular dates for haplogroup Q range from 30,000 ya (Underhill *et al.* 1996) to 17,200 ± 3,200ya (Zegura *et al.* 2004). The latter date is compatible with a Pre-Clovis archaeological model for the peopling of the Americas (Dillehay 2000) and also overlap with recent mtDNA molecular dates (Achilli *et al.* 2008). Early research on Native American Y-chromosome variation supported a single migration of people into the New World. In addition, molecular dates for haplogroup C date to around 27,500 ± 10,200 YBP and haplogroup R to around 16,300 ± 4430 ya (Hammer and Zegura, 2002). These dates would appear to support a single, post last glacial maximum (LGM) entry into the Americas but other migration models have also been proposed (Karafet *et al.* 1999; Tarazona-Santos *et al.* 2001; Lell *et al.* 2002).

Research into the number of migrations based on Y-chromosome data usually favors one or two migrations. The earliest research on Y-chromosomes in the Americas was used to support a single migration (Pena *et al.* 1995; Santos *et al.* 1995, 1996; Karafet *et al.*1997; Lell *et al.* 1997). This idea was based on the high frequency of haplogroup Q-M3 individuals found in Native American populations from Alaska to Argentina. However, this research was often based on a low number of Y-chromosome polymorphisms or number of populations sampled. Karafet *et al.* (1999) investigated potential paternal peopling models and identified two possible sources in Siberia for Native Americans (the Lake Baikal and Altai regions). These researchers argue for two migrations into the Americas, with further support by Lell *et al.* (2002). Both these groups of authors suggest that the presence of Y-haplogroup Q throughout the Americas point to an early migration and that because Y-haplogroups R and C were restricted to North and Central America these latter haplogroups were representative of a later migration originating in Northeast Siberia (Lell *et al.* 2002). The absence of haplogroup C in South American populations is not entirely accurate as it has been found in the Wayuú, indicating that its absence may be due to lack of investigation. In addition, if haplogroup R represents Native American admixture, it is often removed from phylogenetic analysis (Tarazona-Santos *et al.* 2001; Ruiz-Narvaez *et al.* 2005). Zegura *et al.* (2004) challenged the two migration model based on their phylogenetic analyses of STR haplotypes and argue that a single migration best represents the available data. They also argue for a central Siberian homeland for Native Americans. While indigenous South American populations have always been

included in the Y-chromosome studies of the earliest Americans, only a single study has used this marker to investigate the peopling of South America (Tarazona-Santos *et al.* 2001).

Y-Chromosomes and peopling of South America

The reason for this lack of study of Y-chromosome variation in South American indigenous populations is the low amount of diversity seen in these groups. A study by Bortolini *et al.* (2003) investigated 23 South American populations and found 13 of these groups were 100% Q-M3, while the remaining ten were at least 60% haplogroup Q. A similar finding was made by Lell *et al.* (2002) who identified three out of seven South American populations that were 100% Q-M3 and only one population under 50%, the Yanamamo (26.7% Q-M3). The low amount of Y-chromosome diversity is sometimes attributed to a single migration into the South American continent but others have argued for multiple migrations. Tarazona-Santos and colleagues (2001) are the only current researchers to focus on the evolution of paternal lineages in South America. These researchers investigated a single SNP and six STRs in twelve South American groups but only investigated those individuals belonging to haplogroup Q-M3. These authors proposed a differential model of evolution for paternal lineages in South America involving gene flow in the Andean region and genetic drift in the Amazon and Gran Chaco regions. This explanation is supported by similar evidence from mtDNA haplotype diversity (Fuselli et al, 2003; Lewis *et al.* 2004; Lewis and Long 2008) and has been used to support a single migration into South America. Tarazona-Santos *et al.* (2001) also indicate that their

model is consistent with linguistic and paleoecological evidence available from South America, but they do not fully develop this proposed relationship through statistical tests nor do they offer any phylogenetic support for their idea.

LANGUAGE AND THE PEOPLING OF THE AMERICAS

Over the last twenty years, a number of attempts have been made to model the peopling of the Americas using linguistics (Greenberg *et al.* 1986; Greenberg 1987; Campbell 1997; Ruhlen 1992; Nichols 2002). The two models most often used in anthropology are Greenberg *et al.*'s (1986) "three wave" and Nichols' (2002) "Pacific Rim" hypotheses. By far the most well known of these models is, the "three wave" hypothesis, argued for by Greenberg and colleagues (1986). This idea suggests that three major linguistic families gave rise to all existing and extinct Native American languages. These three language families were named Amerind, Na-Dené, and Eskimo-Aleut. The Amerindian group included the majority of Native American populations from North and South America, while Na-Dené groups inhabited portions of the northwest and southwest in North America, and Eskimo-Aleut speakers were isolated to Alaska and the Arctic circle. Greenberg *et al.* applied sound correspondence and a glottochronological method to suggest that Native Americans arrived in three successive migrations from Siberia. This scenario inferred that Amerindian populations arrived around 11,000 ya as big game hunters and were associated with the Clovis culture that had originated in the Lena River basin in Siberia and passed over the Bering Land Bridge. Another subsequent migration occurred when Na-Dené populations arrived between 7000-10,000 ya, and these

groups were associated with the Paleo-arctic archaeology tradition. Populations of the Na-Dené language family were suggested to have originated in the forested region between the Lena and Amur river basins and migrated to Alaska prior to the flooding of the Bering land bridge. A third migration involved final Eskimo-Aleut speakers who arrived between 8500-10,000 ya from the lower Amur region and migrated along the southern coast of the Bering land bridge.

Greenberg's classification of American Indian languages is widely used by anthropologists but has suffered numerous criticisms from fellow linguists (Chafe 1987; Goddard; 1987; Campbell 1988; Nichols 1990). These linguistic criticisms have questioned the legitimacy of lumping the vast majority of Native American languages into a single Amerindian family as well as the validity of glottochronology. One criticism is the use of sound correspondence, that only allows for a language decay rate of between 8000 to 12,000 ya (Bateman *et al.* 1990). This linguistic decay rate is indicative of the weaknesses in applying the study of language to deep time. According to the glottochronological decay rate applied in the "three wave" model, the oldest a Native American language could be is 12,000. This linguistic rate artificially supports a Clovis entry into the Americas, as populations belonging to this archaeological time period are known to have existed within this time frame. However, an earlier entry date would be impossible to calculate as languages that may have pre-dated the current linguistic diversity in the Americas have gone extinct (Schurr 1999). There is considerable evidence of a pre-Clovis migration to the Americas from both archaeology (Dillehay 2000) and molecular genetics (Schurr

2004) that support human occupation between 15,000 and 20,000 ya. This sound correspondence method indicates that language diversity may be useful in understanding recent human history but not for recognizing older demographic events. In addition to these linguistic criticisms, numerous studies from molecular genetics using mtDNA (Merriwether *et al.* 1995; Bonatto and Salzano 1997; Rubicz *et al.* 2002; Tamm *et al.* 2007), Y-chromosome (Karafet *et al.* 1999; Lell *et al.* 2002; Bortolini *et al.* 2003; Zegura *et al.* 2004) and autosomal markers (Schroeder *et al.* 2007; Wang *et al.* 2007) have refuted the idea of three migrations and instead have indicated one or two migrations for the origins of Native American populations.

The second major linguistic model for the origin of Native American populations, Nichols' "Pacific Rim" hypothesis (1992; 2002), also attempts to reconstruct the pattern of linguistic diversity in northern Asia and the Americas. This explanation applied structural features (dominant alignment, word order, and complexity) in order to calculate the frequency of these attributes over broad geographic areas. These resulting frequencies were then tabulated in order to approximate linguistic colonizations and the amount of time for the present linguistic diversity to appear within a given geographic region (Nichols 1992). Using this model, Nichols estimated the spread of linguistic stocks to occur every 2,000-5,000 years, or on average a new linguistic family evolving every 3,500 years. Based on the number of language families in the Americas, the initial colonization of the Americas was determined to be approximately 37,500 ya with multiple migrations. This linguistic model suggests that as language families spread in the Americas, existing

populations were displaced and forced to migrate to interior regions of the continent. Nichols (2002) also inferred that, due to high linguistic density among populations inhabiting the Pacific Rim versus the American continental interior, coastal regions of eastern Asia were the source populations for Native Americans. Other linguists argued that the assumption that structural features evolve in linear fashion with time are not valid and high linguistic diversity in the Americas is more compatible with recent human colonization (Nettle 1999). This "Pacific Rim" linguistic model is also not supported by evidence from molecular genetics. The majority of genetic evidence points to a Siberian (Schurr *et al.* 1990; Torroni *et al.* 1993; Lell *et al.* 2002; ; Zegura *et al.* 2004; Schroeder *et al.* 2007) origin for Native American populations and not to Asian coastal populations.

Both these linguistic explanations have been criticized for making sweeping generalizations regarding the peopling of the Americas because they are not compatible with data from other scientific disciplines. These linguistic models over or under estimate the timing of the peopling of the Americas based on current archaeological and genetic evidence and lump large amounts of linguistic diversity into overly general classification systems. However, this does not imply that genetic research cannot be criticized for making these same generalizations. Modern humans as a species, are defined by the intersection of their biological and cultural histories that has clearly impacted their evolutionary history. An alternate approach to understanding the role of linguistics in human evolutionary history is to understand the relationship of genes and language within a specific language family. This

microevolutionary methodology is useful for understanding a single or small group of languages within a geographic region and correlating this diversity with evidence from genetics, archaeology, and known demographic or historical events. This latter method allows for the understanding of biocultural diversity within a given geographic area and the testing of hypotheses in order to draw inferences upon past events of historical or evolutionary significance to a region's current inhabitants. Native American populations belonging to the Chibchan language family of lower Central and northern South America provide an opportunity to understand this biocultural relationship between genes and languages. These populations have well defined archaeological (Hoopes and Fonseca 2003), linguistic (Constenla 1991; 1995) and genetic (Barrantes *et al.* 1990; Kolman and Bermingham 1997; Melton *et al.* 2007) histories that allow for determining if this gene-language relationship is present in these groups or if the perceived relationship is based on the lack of genetic research on neighboring populations.

GENES AND LANGUAGE

Anthropological geneticists and linguists are interested in the parallels between linguistic and genetic evolution. Generally, people who speak languages within the same linguistic family and people who live within close proximity share similar gene frequencies. This gene-language association has often been used to imply evolutionary relationships among populations. There are two possible explanations for these gene-language connections: 1) processes leading to linguistic division also brought about genetic differentiation (i.e, population expansion); or 2)

linguistic variation among populations operate as a reproductive barrier making geographic neighbors biologically distinct (Barbujani and Sokal, 1990). These gene-language associations have been exhibited using classical and molecular genetic markers for several geographic regions including; 1) Africa (Excoffier *et al.* 1991); 2) Arctic (Crawford and Duggirala 1992); 3) Europe (Menozzi *et al.* 1978); 4) North America (Suarez *et al.* 1985); and 5) Siberia (Crawford *et al.* 1997). While these gene-language associations have been demonstrated in other global areas, research into the gene-language relationship among Native South American populations has often led to dissimilar outcomes.

Traditionally, investigators in South America have been unable to find a clear association between the distribution of genes and languages on the continent (Black 1991). This has led some researchers to advocate genetic drift as the principal evolutionary mechanism shaping the underlying population structure throughout the entire continent (O'Rourke and Suarez 1986; Black 1991; O'Rourke *et al.* 1992). These researchers used classical genetic polymorphims to argue that South American environments (high altitude and tropical forests) are poorly adapted to human subsistence. According to this explanation, marginal ecosystems have led to small effective population size and greater genetic differentiation among groups creating an absence of any genetic-linguistic associations. Other researchers agree that genetic drift is the primary force, but have argued that South American gene-language connections do exist but are only detectable over short geographic distances rather than at the continental level (Luiselli. *et al.* 2000; Fagundes *et al.* 2002; Hunley and

Long 2007). Recent studies using molecular markers (mtDNA and Y-chromosome) have supported the idea of genetic drift as the primary evolutionary force in the Amazon region, but have argued for gene flow along the Andes in western South America (Tarazona-Santos *et al.* 2001; Fuselli *et al.* 2003; Lewis *et al.* 2004; 2007; Lewis and Long 2008).

One geographic region that may have an important effect on this relationship is Central America. Indigenous populations in this area are informative because they bridge the two American continents. Furthermore, the linguist Johanna Nichols (2002) has identified a linguistic hub using structural features in the Gulf of Mexico or eastern Caribbean. This linguistic hub contains connections that reach out to eastern South and eastern North America. In fact, this connection contains strong linguistic bonds among Central America, eastern South America, and the North American southeast. According to this idea, the bond between Central America and the North American southeast is so strong that this latter geographic region should be considered a linguistic offshoot of Central America. Therefore, understanding the linguistic and genetic relationships of Central America may have important consequences for understanding the post colonization population dynamics of the Americas.

Previous research on the relationship between genes and language in Central America has been sparse, with only Chibchan populations exhibiting a clear genetic linguistic association (Barrantes *et al.* 1990; Kolman and Bermingham 1997; Kemp

2006; Melton *et al.* 2007). This lack of study may be due to the common misconception that the majority of indigenous populations in the region are extinct or have been assimilated into the larger Hispanic populations of these countries. There are currently 56 Native American populations in Central America divided into seven linguistic families and three language isolates (Lenca, Xinca, and Tolupan) (Campbell *et al.* 1976; Constenla 1991; 1995; Brignoli 2005). These languages may be further divided into two broad linguistic areas, which include the Mesoamerican and Isthmo-Colombian regions (Campbell *et al.* 1976; Constenla 1991; Hoopes and Fonseca 2003). A linguistic area is defined as a continuous geographic area where linguistic features are shared by a number of different languages or various language families (Campbell *et al.* 1986). The Mesoamerican linguistic area (figure 6) is characterized by relational nouns, a base 20 numerical system, and a verb-final syntax. Mesoamerica includes languages from the Mayan, Uto-Aztecan, Oto-Manguean, Mixe-Zoquean, and Misumplan families (Campbell *et al.* 1986). Mayan languages are spoken in southern Mexico, Guatemala, and Belize and are represented by 28 languages spoken by approximately 6 million people (Ibarra-Rivera *et al.* 2007). Uto-Aztecan is one of the largest linguistic groups in the Americas and spreads from Central America into the Great Basin region of western North America (Campbell 1979). Currently, Oto-Manguean languages are spoken in Mexico but were once spoken as far south as Nicaragua and included the extant populations of Chorotega and Subtiva (Campbell 1997). Mixe-Zoquean languages are spoken along the Tehuantepec isthmus in Mexico (Campbell and Kaufman 1976).

Figure 6: Mesoamerican Linguistic Area and geographic localities of populations found within the region (Wikipedia.org)

The second major linguistic area in Central America is termed the Isthmo-Colombian area and is defined as the area from eastern Honduras to the shores of Lake Maracaibo in Venezuela (figure 7). This linguistic region has also been termed the Intermediate Area, lower Central America, and the Chibchan Historical Area but, numerous problems have arisen with these alternate classification schemes as they are largely based on artificial geographic boundaries or the exclusion of specific populations (Hoopes and Fonseca 2003). This language area contains indigenous populations primarily belonging to the Chibchan linguistic stock but also contains groups from the Misumplan, Paez/Barbocoan and Chocoan language families. The Chibchan language family is divided into four subfamilies and is spoken from eastern Honduras to Venezuela. Chocoan languages are a small group of languages spoken

from Panama to Ecuador along the Pacific coast (Campbell, 1997). Misumplan languages are spoken in eastern Nicaragua. This linguistic family consists of two extant (Miskito, Sumo) and one extinct language (Matagalpa) (Hale and Salamanca, 2001). Paez is a language family from central Colombia. In addition, three language isolates occur in Central American and include Xinca from Guatemala, as well as Lenca and Tolupan from Honduras (Campbell, 1997). All of the languages within this region are thought to have connections to Chibchan languages and therefore the understanding of the gene-language association within this language family has important implications for understanding population structure within Central America.

Figure 7: Isthmo-Colombian region, dark shade areas show locations of Chibchan populations (Hoopes 2005).

The Chibchan language family is divided into four subfamilies (figure 8) and includes: 1) Pech (Paya) spoken in eastern Honduras; 2) Votic, spoken along the Caribbean coast of Nicaragua and northern Costa Rica; 3) Isthmic, spoken in Costa Rica and Panama; and 4) Magdalenic, spoken in northern South America. Chibchan languages are thought to have diverged from related linguistic families in Central America approximately 7,000 ya and are believed to have originated in either Costa Rica or Panama. The Isthmic subfamily displays the highest amount of Chibchan language diversification and, based on the principle that areas with higher levels of diversity are older, would suggest this region is where the language originated. The fragmentation of these Chibchan subdivisions is believed to have begun around 5,000 ya with divisions evident by 4,000 ya. This linguistic differentiation is thought to have occurred due to a subsistence shift to agriculture and an adaptation of a sedentary lifestyle that is associated with this time period (Constenla 1991).

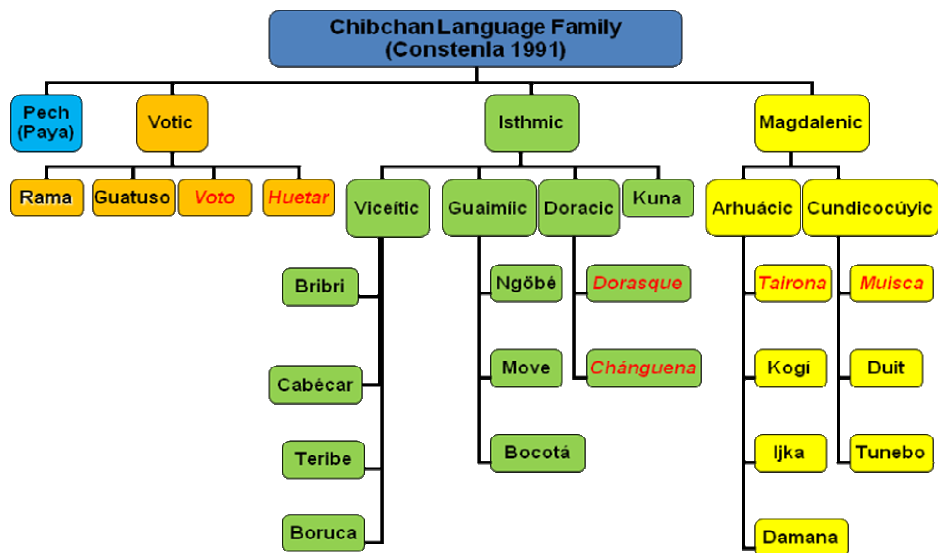

Figure 8: Hierarchical diagram of languages in the Chibchan linguistic family. Existing languages are in black and extinct languages shown in red italics (Constenla 1991).

This view suggests that geographic areas within regions of agricultural development will demonstrate higher linguistic diversity and spread faster along latitudinal axes than in a north-south direction. However, according to Diamond and Bellwood (2005) this basic premise is complicated by five issues that make total acceptance of this idea difficult. These six issues are the: 1) potential of genetic admixture between hunter/gathers and expanding farmers; 2) adoption of agriculture by hunter-gatherers on the periphery of the agriculturalist's territory; 3) reversion of agriculturists to a hunter-gathering subsistence strategy; 4) replacement of the farmer's language in its homeland; and 5) expansion of hunter/gathers groups. While there are clear complexities in accepting this explanation, it does provide for the testing of questions regarding the distribution of modern day Native American language families, genetics and their association with the origins of agriculture in the Americas. A goal of this dissertation is to determine if this explanation can be extended to the understanding of Chibchan population dispersal in Central and South America or if these groups adopted agriculture from neighboring tribes.

THE ORIGINS OF AGRICULTURE IN THE AMERICAS

The traditional view regarding the origins of agriculture in the Americas is that it occurred later in time than either the Middle East or China and, due to geographic boundaries, it was more diffuse in its adoption by hunter-gatherers than in other parts of the world. Evidence for the transition to agriculture is often determined by the appearance of ceramics within a given archaeological site and the appearance

of a widespread use of small number of plant domesticates (Bellwood 2005). Within the Americas this cultural transition occurred with either the presence of maize or manioc and is suggested to have led to the independent development of agriculture in different areas of North and South America rather than radiating out from a central region. There are currently four independent regions of agricultural development identified in the Americas. These four regions include the 1) Andean region of coastal Ecuador, Peru, Bolivia, and northern Chile; 2) Middle America region from northern Colombia to Mexico; 3) Southwest region of the United States; and 4) Eastern Woodlands of North America (figure 9) (Bellwood 2005). Agricultural development

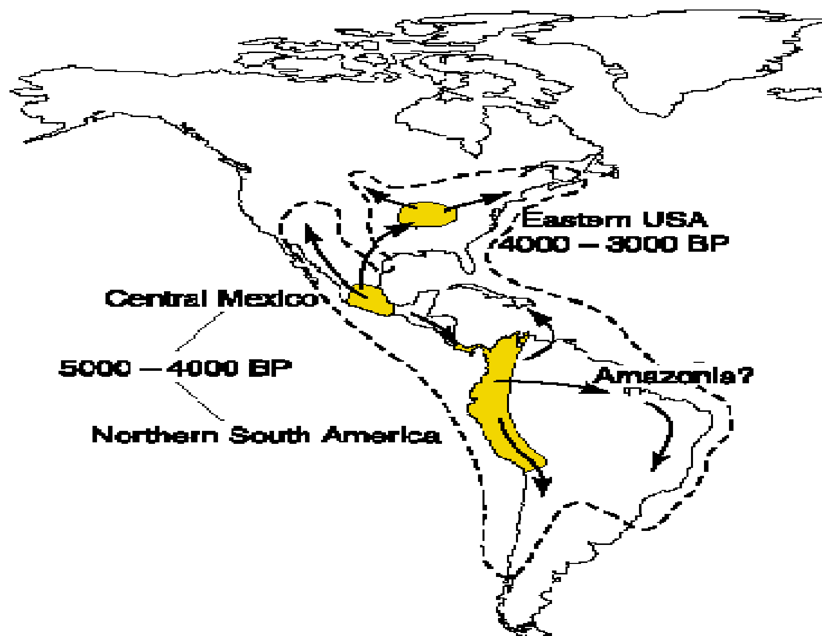

Figure 9: Distribution of agricultural regions in the Americas and their extent (modified from Diamond and Bellwood 2003)

in the Andes is thought to have begun with the Valdiva cultural complex beginning around 6000 YBP along the arid coast of Ecuador. Agriculture in Middle America is suggested to have begun around 4000 YBP and stretched from central Mexico to western Honduras. This date is also congruent with the presence of agriculture in the North American Southwest, which is also thought to have begun about 4000 YBP. The last place of agricultural development in the Americas was the Eastern Woodlands of North America which began about 1000 YBP. Both the North American Southwest and Eastern woodlands may have been heavily influenced by the spread of agriculture from Mesoamerica, and both may have linguistic associations with this region (Nichols 2002). Some researchers have suggested that the spread of plant domesticates to both the American southwest and the Eastern Woodlands originated in Middle America and subsequently spread into these regions (Bellwood 2005). New evidence from paleoecology is challenging the traditional viewpoint of the origins of agriculture in the Americas and suggests that there are only two major areas of independent development, the Andes and Mesoamerica (Piperno and Pearsall 1999). In addition, current archaeological evidence in the Americas suggests that plant domestication may have preceded the development of ceramic technology (Cooke 2005). These new research findings have altered the traditional view regarding New World agricultural development.

The current opinion regarding the earliest origins of agriculture in the Americas is that it began gradually in the lowland deciduous forests of Central and northern South America between 10,000 – 11,000 YBP (Piperno and Pearsall, 1998).

This subsistence shift to an agriculture lifestyle was precipitated by climatic changes at the end of the Pleistocene and ecological changes in the region, as dry habitats became wetter and heavily forested. This subsistence shift occurred slowly with the use of small house gardens amid shifting agricultural fields and the absence of domesticated animals. Then, around 7,000 YBP, large scale food production emerged. The beginnings of slash and burn agriculture are evident in the paleoecological record (Piperno and Pearsall, 1998). This idea is not a new suggestion and is in fact an updated version of theories proposed earlier by both Sauer (1952) and Lathrap (1970). However, this view differs from that of MacNeish (1992), who favored an origin of agriculture in the Americas in the semi-arid central Mexican highlands along with the development of maize based agriculture. However, it is now evident that teosinte, considered the ancestor to maize, was absent in the Mesoamerican highlands until at least 6,000 YBP and was probably first cultivated in the lowlands of Mexico (Pohl *et al.* 1996; Buckler *et al.* 1998; Piperno and Flannery 2001). This absence is more supportive of a lowland tropics origin of maize agriculture than for a Mesoamerican highland source. Agriculture is thought to have developed independently in the three other areas of the America at later dates. Therefore, understanding the distribution of linguistic and genetic diversity in Central America has important implications for understanding the movement of plant domesticates throughout the Americas.

The lowland tropical region of Central America is one of the independent global regions of agriculture in the Americas. This cultural development clearly had a

significant impact on the genetic structure and linguistic diversity of populations in the region. A number of regional Mesoamerican language families are thought to have dispersed along with agriculture and include Oto-Manguean, Chibchan, Mayan, Mixe-Zoquen, and Uto-Aztecan (Bellwood 2005). Early languages associated with the Oto-Manguean family are thought to have originated between the valley of Mexico and Oaxaca in modern day Mexico. Populations within this language family then began to expand but were hemmed in by Mayan and Mixe-Zoquean speakers in the region expanding at the same time (Bellwood 2005). The Mayan language family originated in the highlands of Mexico and Guatemala beginning around 4,000 YBP (Campbell 1997). Other linguistic researchers have suggested that these groups were all interrelated at one time into a larger Proto-Mesoamerican linguistic group that was included with Uto-Aztecan that subsequently splintered into different language families after the intensification of agriculture in Mesoamerica (Witkowski and Brown 1981). By far the best evidence for the simultaneous expansion of a language family and agriculture in the Americas is Uto-Aztecan. This language is found in Native American populations ranging from the Pacific coast of Nicaragua to the Southwest United States and offers a clear potential for understanding the relationship among languages, genes, and agricultural expansion.

Traditionally, the Uto-Aztecan linguistic family is thought to have originated in the Great Basin region of North America and then spread south into Mesoamerica. However, the majority of historic populations in the Great Basin are hunter-gatherers and not agriculturalists. Recently, others have suggested a southern origin of Uto-

Aztecan populations (Hill 2002), and when these groups migrated into agriculturally inhospitable habitats (i.e, Great Basin, Sonora desert), they reverted back to being hunter/gatherers. One genetic study has tested the "First Farmer" idea in the Americas and investigated the relationship between the American southwest and central Mexico using mtDNA diversity (Kemp 2006). This research detected no maternal genetic relationship between southern and northern Uto-Aztecan populations. However, more evidence is needed to support or refute this hypothesis in Central America. Uto-Aztecan groups also migrated along the Pacific coast of Mesoamerica, and ethnohistoric accounts indicate that Aztec traders were also found along the Caribbean as far south as Panama (Cooke 1997). This may indicate some cultural and biological interaction between Mesoamerica and lower Central America, but a lacuna of molecular genetic data in indigenous populations from Honduras and Nicaragua makes biological characterization difficult. Another possibility is that Chibchan populations also spread with the development of agriculture from Central America into South America, but this explanation has only been hinted at through cursory evidence from linguistics (Constenla 1991; 1995), archaeology (Cooke 2005), and classical genetic polymorphims (Barrantes *et al.* 1990).

The present day distribution of Chibchan populations spanning the Central American land bridge and into northern South America demonstrates their importance for understanding the dispersal of both people and cultural traits through the region. There is significant paleoecological and archaeological evidence for the presence of domesticated plant material from the central Pacific site of La Yeguada in Panama

beginning around 9,000 YBP (Piperno and Pearsall 1998; Cooke 2005). Subsequent cultural developments between 7,000-4,500 YBP brought several plants domesticated from outside the region, including maize, manioc, squash, and possibly bean, which appear in the archaeological record at La Yeguada. There are two alternative explanations for how these plants arrived in Panama: 1) they were traded for from outside neighboring populations; or 2) they were brought into this region by expanding agriculturists who replaced existing hunter-gatherer populations (Cooke 2005). The presence of a sedentary lifestyle led to a population increase that is demonstrated in the archaeological record. In lower Central America, ceramics appear in the archaeological record of both Panama and Costa Rica at approximately 4,000 YBP (Cooke 2005). The occurrence of pottery is important because it is demonstrative of long term sedentary occupation and an increase in population size in the region (Cooke 2005). Early agriculturists in the region are believed to have spread from the Pacific foothills into the Caribbean region between 3,000-2,000 YBP, although it is pertinent to point out that due to different microclimates in the region, populations within the Caribbean may have been more reliant on root crops such as manioc than on maize (Baldi 2001; Cooke 2005).

Social stratification in lower Central and northern South America began to take place at approximately 2,500 YBP when jadeite appears in burials and rich residences appear in archaeological sites throughout the region (Hoopes 2005). At European contact in the early 15th century, lower Central and northern South America were fully developed agricultural communities with high population densities.

However, they were rapidly decimated by disease and the severity of Spanish colonization (Cooke 1997). Therefore, the importance of Chibchan populations in understanding the complexity of Pre-Columbian cultural history in the Americas has been neglected by researchers more interested in the state level societies in Mesoamerica and the Andes.

SUMMARY

This chapter reviewed current explanations for the peopling of the Americas based on mtDNA, Y-chromosome, and linguistic evidence as well as briefly summarized the origins of agriculture in the Americas. There appears to be a consensus forming among researchers from multiple scientific disciplines that the Americas were populated by a single human migration that occurred somewhere between 20,000-15,000 ya. However, data from molecular genetics also allows for insight into subsequent demographic events, environmental changes and cultural transitions that affected the genetic diversity of living indigenous populations. One of these cultural transitions is the origin of Native American agriculture. Based on current paleoecological evidence, this cultural transition appears to have first originated in lower Central or northern South America between 11,000 - 10,000 YBP. This new evidence makes understanding the population structure of Chibchan populations intriguing as numerous lines of archaeological, biological, and linguistic evidence have suggested that these populations also originated within this time frame.

III. CHIBCHAN POPULATION BACKGROUND

This chapter provides background information on Chibchan populations from lower Central America along with previous archaeological, linguistic, and genetic research in the region in order to better understand Chibchan evolutionary history.

CHIBCHAN POPULATIONS

Chibchan populations provide an opportunity for understanding population structure and the impact of microevolution within the confines of a small geographic area. Currently, 56 Native American populations affiliated with seven linguistic families (Mayan, Misumplan, Mixe-Zoquean, Oto-Manguean, Uto-Aztecan, Chibchan, Chocoan) inhabit the area from southern Mexico to the Panama/Colombia border. Geographically, Central America is divided into two major regions, Mesoamerica and lower Central America (figure 1), which are distinguished by their underlying geology. Mesoamerica is geologically older and is defined as the region between south-central Mexico and the Pacific coast of Nicaragua. Lower Central America is younger and is identified as the area from the Caribbean lowlands of Nicaragua to the Panama/Colombia border (Coates, 1997). Both of these geographic regions have also been used to define cultural areas by archaeologists (figure 10) (Brignoli 2005). The Mesoamerican culture area (MCA) is characterized by agricultural villages, complex ceremonial centers, and regional polities (Brignoli, 2005). Archaeologists have had a more difficult time delineating a cohesive cultural area for Pre-Columbian populations inhabiting lower Central America. Cultural

material found in this region suggests these populations were primarily chiefdoms, lacked monumental architecture and, until recently, were considered intermediate between state societies found to the north and south (Hoopes and Fonseca, 2003). The heterogeneous nature of cultural material in the region has led the cultural area for the region being termed the "Intermediate Area" (Willey 1971), Lower Central America (Lange and Stone 1984), or more recently the Isthmo-Colombian Area (Hoopes and Fonseca 2003).

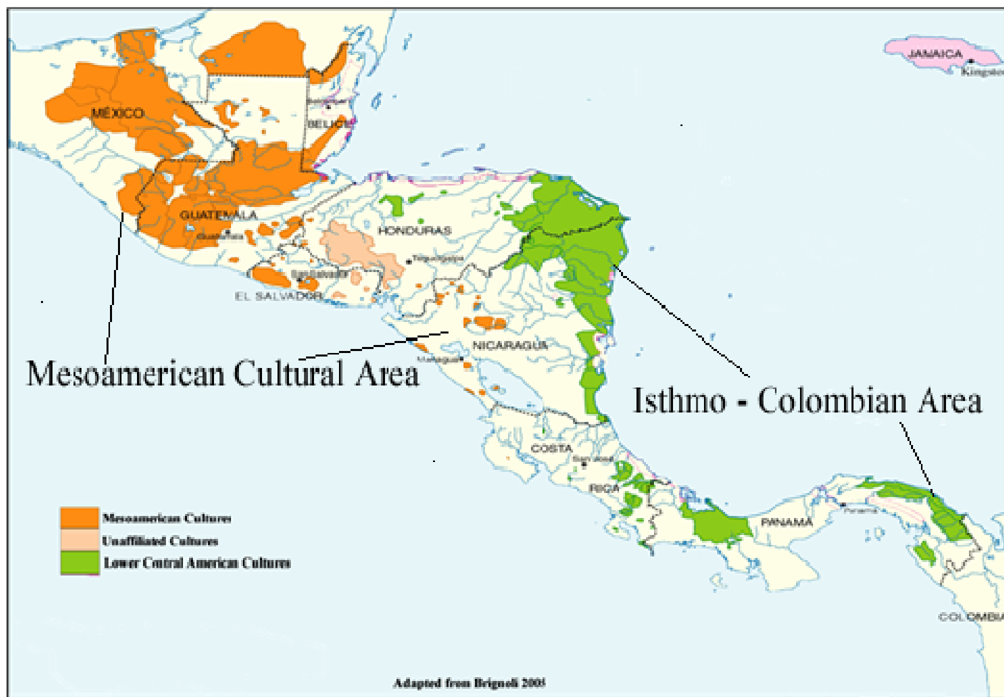

Figure 10: Central America divided by regional cultural affiliation

The first attempt to define a cultural area in lower Central and northern South America was proposed by Kirchoff (1943) when he attempted to define it as "Chibchan". This definition was disregarded by other archaeologists at the time who

considered the region to be "peripheral" or "intermediate" between the Mesoamerican and Andean regions. Currently, there is dissatisfaction with the underlying subtext implied in the terms "Intermediate Area" and "Lower Central America" among researchers. The term "Intermediate Area" implies that populations inhabiting the region are not as culturally complex as those surrounding the region and invokes a hierarchical cultural evolutionary model (Hoopes and Fonseca 2003). Similar problems arise with "Lower Central America" as it implies an arbitrary boundary between Central and South America that simply did not exist in prehistory (Bray 1984). Since the early 1990s, when Fonseca (1992) proposed the Chibcha Historical Region, a concerted effort has been made to return to Kirchoff's idea of a region dominated by Chibchan populations. This nomenclature has gone through several derivations and also been termed Región Histórica Chibcha-Chocó (Cooke 1993), Área de Tradición Chibchoide (Fonseca 1994) and Área Histórica Chibchoide (Fonseca 1997; 1998). However, there are also problems with this classification system. The term "Chibcha" is often used to refer to the extinct Musica culture of highland Colombia, and the terminology ignores the presence of other linguistically distinct indigenous populations in the region. This led Hoopes and Fonseca (2003) to propose the term "Isthmo-Colombian Area". This label acknowledges the presence of other culturally distinct populations in the region, as well as recognizes pre-Columbian archaeological connections between Central and South America. These cultural similarities are detected through a number of shared traits evident in the rich archaeological history of the region.

CHIBCHAN CULTURAL HISTORY

Archaeological evidence indicates continuous human occupation in Central America over the last 14,000 years (Cooke 2005). Therefore, understanding parallels between Chibchan genetic and cultural history has important implications for understanding the initial colonization of the Americas and the complex population dynamics that led to the peopling of South America. The complexity and depth of archaeological diversity in the region is outside the scope of this dissertation (for detailed reviews of the region see Cooke 2005 and Hoopes 2005), but a brief summary is necessary in order to understand cultural and historical events that may have affected the genetic structure of Chibchan populations. This dissertation uses Fonseca's (1992) temporal classification in order to provide a better understanding of Chibchan cultural history. This timeline is organized into six categories: 1) Hunter Gatherer (14,000-6,000 YBP); 2) Specialized Domestication (6,000-4,000 YBP); 3) Early Agriculture (4,000-2,500 YBP); 4) Specialized Agriculture (2,500 1,500 YBP); 5) Chiefdoms and increased cultural specialization (1,500-500 YBP); and 6) Post European contact (500 YBP to present). The geographic location of archaeological sites and regional cultural areas for lower Central America are shown in figure 11. The Isthmo Colombian cultural area is divided into five sub-cultural areas: 1) Gran Nicoya a region of Mesoamerican influence that ranges from the Pacific coast of Nicaragua to the Gran Nicoya peninsula in Costa Rica; 2) Costa Rican Central Valley historically occupied by Votic speaking Chibchan populations; 3) Gran Chirqui an area that extends through the majority of southern Costa Rica; 4) Gran Coclé the

region that includes the Costa Rican/Panamanian border and extends into central Panama; and 5) Gran Darien a dense tropical forest that extends from southern Panama into northern Colombia (Cooke 2005). Other sub-cultural areas not displayed on the map include the Muisca cultural area located around Bogotá, Colombia and the Tairona region located along the northern Caribbean coast of Colombia near Santa Marta. This long term human occupation of the Isthmo-Colombian region has led to rich but heterogeneous archaeological records in the region. There is also the possibility that this cultural history has impacted the genetic structure of indigenous populations in the region.

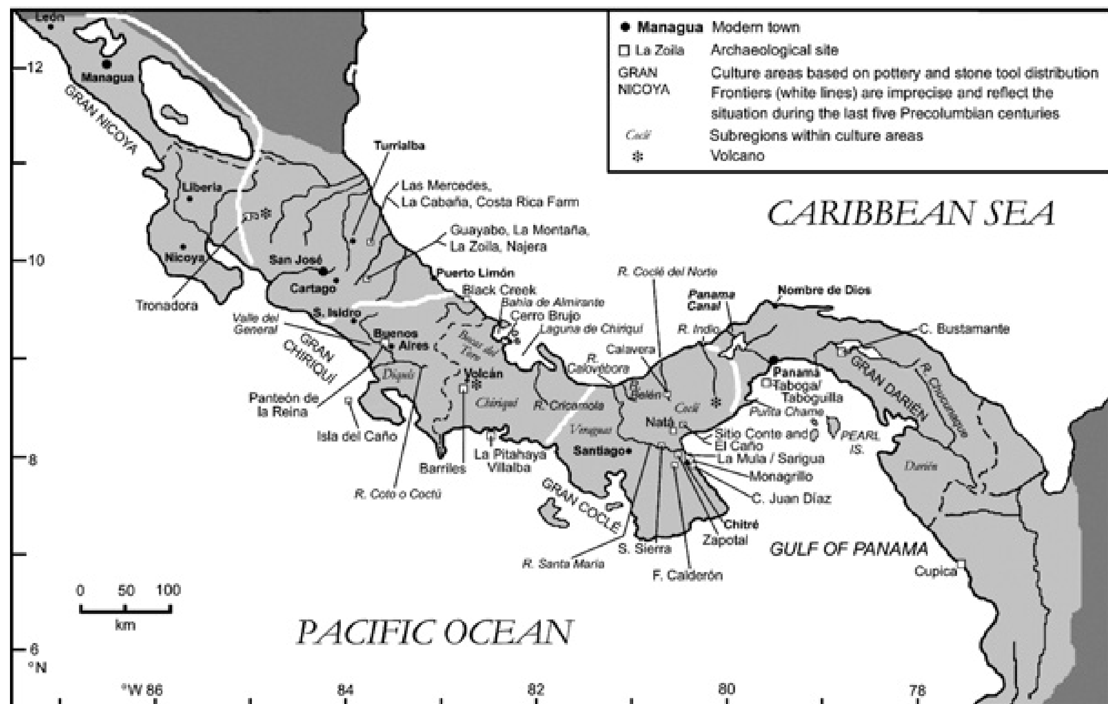

Figure 11: Geographic locations of local cultural areas and archaeological sites in lower Central America (Cooke 2005).

HUNTER-GATHERER PERIOD (14,000-6,000 YBP)

The earliest inhabitants on the Central American isthmus belonged to Paleoindian cultures (Cooke, 1997). The first humans probably entered the region during the Late Glacial Stage (14,000-10,000 YBP) of the Pleistocene epoch. Surface finds of Fishtail (Jobo) projectile points are known from near the Panamanian continental divide (Pearson 2002; Ranere and Cooke 2003). More definitive evidence is available for Clovis populations who entered the region between 10,000 to 9,000 YBP based on radiocarbon dating of the Cueva de los Vampiros, Corona, and Aguadulce archaeological sites (Cooke 2005). These original inhabitants were hunter-gatherers and subsisted on the collection of wild plants and the hunting of megafauna. Climatic changes associated with the shift from the Pleistocene to the Holocene (e.g, Younger Dryas) in Central America began approximately 10,500 YBP (Cooke 1997). This climatic shift led to a change from dry xeric environments to wetter mesic forests. This rise in sea levels also may have shifted human occupation away from the coasts. Shortly after this climatic transition, humans domesticated plant materials and shifted to hunting small mammals.

SPECIALIZED DOMESTICATION (6,000-4,000 YBP)

This subsequent archaeological period known as Specialized Domestication (Archaic) ranged from as early as 9,000 to 4,000 YBP and contains early evidence for the development of agriculture (Demarest 2004; Cooke 2005). During this time period projectile points were constructed with serrated edges and barbs in order to

hunt small animals including deer, peccaries, and other species that had avoided extinction at the end of the Pleistocene. The distribution and density of these projectile points that are found in archaeological sites from Guatemala to central Panama suggest that population size was increasing during this time period. There is also increasing evidence of a shift to a more sedentary lifestyle or at least a seasonally based occupation. Archaeologists in the region believe that people were dedicating more time to the collection and preparation of plant foods. The ecological diversity of lowland tropical forests encouraged experimentation with a wide variety of plant material including tubers, roots, fruits, and trees (Piperno and Pearsall 1999; Baldi 2001). The majority of these plants could not be consumed without adequate processing, which necessitated a shift in the archaeological toolkits. These new toolkits included scrapers, stones for milling, hammers for cracking open nuts, and polished stones that may have been more useful for working wood or other plant materials (Baldi 2001). This time period is also when linguists suggest the dispersal of Chibchan language first began to occur (Constenla 1991; 1995).

EARLY TRIBAL SOCIAL AGRICULTURE (4,000-2,500 YBP)

Domestication led to increased population size and social stratification, which allowed for the appearance of complex polities during the Early Agricultural Period. Archaeologists have identified differences in residential structures and funerary sites that appear to represent basic forms of early social stratification occurring during this time period in lower Central America. These cultural differences are interpreted as

important changes in social structure occurring during this time period. These groups were now required to change their interaction with their environment as well as with other neighboring human populations in order to maintain continuity in their social interactions and subsistence patterns (Cooke 2005)

SPECIALIZED AGRICULTURE (2,500 1,500 YBP)

The defining characteristic of this time period is the amount of cultural variation that appears in the archaeological record. These cultural characteristics include permanent settlements in different ecological settings, a wider array of cultivated plant material, mining of mineral products, and riverine fishing (Baldi 2001). It is also important to note prior to ~2,500 YBP, differences in cultural material permit archaeologists to differentiate local artifact traditions and demonstrate their connotations for prehistoric ethnic and linguistic variation. However, it is not until this time period that clear regional distinctions can be construed from the archaeological record (Cooke 2005). Archaeological research during this time period demonstrates: 1) the importance of random events (i.e, volcanic eruptions) in shifting settlement patterns; 2) sociopolitical fragmentation that resulted from population growth and agricultural intensity; 3) changes in social relationships and their association with raw material sources, and craft centers; 4) increased community specialization and exchange of certain prestige items; and 5) varying regional cultural area boundaries, where greater artifact diversity is seen at the edges of the regions than at the epicenter (Cooke 2005). One regional cultural area that embodies this

research well is the distribution of ceramics in the Gran Chirquí. Ceramic diversity in this region would appear to be local based on stylistic categories but a shared cultural connection is seen through a common ideology, mutual artifact traditions. and trade of prestige items (Baldi 2001; Cooke 2005). The shifting cultural geography seen in ceramic distribution during this time period correlates well with linguistic and genetic data, where Gran Chiriquí Chibchan populations diverged into their own political entities, each with their own languages (Bribri, Cabecar, and Ngöbé) approximately 1,500 to 1,000 YBP (Barrantes *et al.* 1990; Constenla 1991).

CHIEFDOMS AND INCREASED CULTURAL SPECIALIZATION (1,500-500 YBP)

The chiefdoms and cultural specialization time period represents greater interaction between Central and South American Chibchan populations. During this time, the region is characterized by the presence of gold and jade artifacts, interment of elites, and iconography not shared with Mesoamerican populations. The iconography consisted of anthropomorphized birds, crocodiles, bats, monkeys and other animals that may have been linked to religion in the region. This development of complex hierarchical societies is thought to have occurred rapidly and occurred throughout the Chibchan region from Honduras to Colombia (Hoopes 2005). The earliest evidence for this development is seen in Costa Rica, Panama, and Colombia with later development in Honduras and Nicaragua. This later development of the peripheries of the Chibchan world may indicate the migration of these populations from Central American into the Caribbean region.

Socially, this period is dominated by centralized chiefdoms and this may be related to cultural interaction with Mesoamerican and South American populations. A number of reasons have been proposed for this rapid cultural transition. These cultural changes included ecological events, such as volcanic eruptions in El Salvador (Ilopango 429 A.D.) and Panama (Volcán Barú 600 A.D.) (Linares *et al.* 1975; Dull *et al.* 2001) as well as demographic movements occurring in Mesoamerica at both Tikal and Copán (Gunn 2000; Martin and Grube 2001). This time period also sees the immigration of Mesoamerican cultures (Chorotega, Nicarao, and Subtiva) into the Pacific coast of Nicaragua and the Gran Nicoya cultural area (Fowler 1989). These populations are believed to have been displaced from central Mexico by Nahua speaking groups (Kaufman 2001). There is clear evidence that this time period represents interaction between Mesoamerican and Chibchan populations but the nature of this contact is not fully understood at present.

POST EUROPEAN CONTACT (500 YBP TO PRESENT)

At European contact, Central America (~1500 A.D.) was densely populated and culturally heterogeneous. Distinct from Mexico or the Andes, the region was not dominated by state level societies but rather small chiefdoms. The cultural landscape of Central America changed rapidly and within a short time period. Columbus visited the region on his fourth voyage to the New World in 1502 and described a multitude of cultures and languages along the Caribbean coast of Central America (Cooke 1997). The Pacific region became dominated by the Spanish, and indigenous

populations were decimated by disease, slavery, and violence (Cooke 1997). In a forty year period, it is estimated that the indigenous population of Pacific Nicaragua declined from 600,000 in 1520 to 6000 between 1560 and 1570 (Cooke 1997). However, the Caribbean side remained relatively isolated throughout the early period of European contact due to the dense forestation and fierce resistance by indigenous populations. These groups used their knowledge of the environment and political divisions between the English and Spanish to their advantage to remain relatively autonomous until the modern era (Cooke 1997). Extant Chibchan populations represent the effect of this early contact and are currently distributed in small enclaves in throughout the region. Therefore, a better understanding of their genetic history will allow for a better reconstruction of their pre-Columbian distribution.

GENETICS OF CHIBCHAN POPULATIONS

Previous genetic research on Chibchan speaking populations has focused on populations from lower Central America in order to understand their biological relationship to each other as well as their divergence from Paleoindian populations. Due to their geographic and cultural location, bridging the two American continents between complex cultural societies in Mesoamerica and the Andes, it was thought that indigenous populations in the region would demonstrate high genetic diversity as evidence of large amounts of gene flow. However, all previous studies on both classical and molecular genetic markers in the region have demonstrated low amounts of genetic diversity as well as high numbers of private genetic polymorphisms that

are not generally shared with neighboring populations even within the Chibchan language family (Barrantes *et al.*; 1990, Bieber *et al.* 1996; Kolman and Bermingham 1997; Ruiz-Narvaez *et al.* 2005). Based on these results, researchers have advocated an endogenous development of Chibchan populations from earlier hunter/gatherer populations and that genetic drift is the major evolutionary force operating on these populations. These researchers also argue that in order for the presence of these private polymorphisms in several culturally related populations that significant time depth (~10,000 YBP) is necessary for their development. The following section provides information on previous classical and molecular genetics research on Chibchan and neighboring Mesoamerican populations in order to address these potential theoretical deficiencies in the literature.

CLASSICAL GENETIC MARKERS

The earliest investigations of Central American indigenous populations using classical genetic polymorphisms was conducted by Matson and Swanson (1959, 1963a, b, c, d, e, 1964, 1965). These authors systematically investigated blood group polymorphisms (ABO, MNS, P, Kell, Diego, Rhesus, Duffy, and Diego), haptoglobins, transferrins, and hemoglobin types in Amerind groups throughout Central America. Their research was descriptive and consisted of gene frequency tables, Chi-square tests, and brief discussions of admixture from European or African populations. Admixture estimates were based on the presence of blood group alleles that were not considered to be of Native American origin, including the presence of

blood groups A and B, or the presence of the MNSs haplotype considered to be of African origin (Matson & Swanson 1963a). Statistical analyses of their results were conducted using Chi-square (χ^2), and no phylogenetic analyses were attempted to clarify biological relationships among Central American indigenous populations. However, data contained in this research series represent one of the most comprehensive analyses of Central American indigenous populations.

In the late 1970s, researchers focused greater attention on lower Central American and Chibchan speaking populations due to a proposed phylogenetic relationship between the Yanamamo of Venezuela and the Guaymi of Costa Rica (Ward *et al.* 1975). Speilman *et al.* (1979) investigated this relationship using ten classical genetic loci and eleven anthropometric markers. These authors concluded that taxonomic similarity was based on the presence of private polymorphisms that were not shared with other populations. These private markers then clustered groups in a dendogram and were not based on shared genetic histories. Barrantes *et al.* (1982) focused on intrapopulation differences among two Guaymi communities (Abrojo and Limoncito) using blood groups, plasma proteins, and erythrocyte proteins. They demonstrated a closer genetic affinity among the Guaymi and other linguistically related populations than to other South American populations

Further research investigations on Chibchan speaking populations from lower Central America were conducted in order to determine their relationship to other Native American groups. Barrantes *et al.* (1990) investigated 48 classical genetic loci from eight Chibchan speaking groups from Costa Rica and Panama (Boruca, Bribri,

Cabecar, Maléku, Teribe, Kuna, Guaymi, and Bokota). Using standard genetic distances converted to linear form and correlated with divergence time, linguistics, and geographic location, they demonstrated low correspondence between taxonomic affinity and geographic location. They also detected higher than expected frequencies of private genetic polymorphisms that included: 1) absence of the Diego A* allele (DiA*) in six of the eight groups; 2) high frequencies of the transferrin D-Chi allele; 3) the G6PD C allele; and 4) five regionally restricted variants (TPI*3-Bribri, TF*D – Guatuso, ACP*Guataso1, LDHB*Guataso1, and PEPA*2-Kuna). They concluded that the large number of private polymorphisms indicated an endogenous development of Chibchan populations that occurred within the last 10,000 years (Barrantes *et al.* 1990). Thompson *et al.* (1992) tested several statistical models developed for understanding migration, survival and history in these same Chibchan populations. These researchers investigated the previously described private polymorphisms (Barrantes *et al.* 1990) and determined that the LDHB-GUA, ESA2*BOK, DI*A, and 6PGD*C alleles were recent genetic introductions into these Chibchan populations from other central or South American indigenous populations. They also indicated that the GOTS*2, PEPA*F, and TF*D-CHI alleles were ancestral to the first Chibchan populations. However, they suggested significant population fluctuations occurred in these groups since contact and this may have influenced their results. These results differ from those of Barrantes *et al.* (1990) and indicated that further research into the evolutionary history of these populations was necessary.

Two previous classical genetic studies investigated Costa Rican Chibchan populations in order to understand their biological relationship to each other. Bieber *et al.* (1996) evaluated genetic variation in six Costa Rican Chibchan populations (Guaymi, Bribri, Cabecar, Huetar, Teribe, Maléku) using four protein polymorphisms (transferrins (TF), alpha-1-antitrypsin (P1), α2-HS-glycoprotein (ASHG), and human coagulation factor B (F13B)). They found low G_{st} (<0.05) values for all four polymorphisms, which demonstrated an absence of significant population substructure among these groups. Using Nei's standard genetic distance and neighbor-joining trees, they found that these six populations cluster into three groups: 1) Bribri/Cabecar; 2) Huetar/Maléku; and 3) Guaymi/Teribe. This relationship is attributable to geography as all three groups contain adjacent populations. A second study by Azofeifa *et al.* (2001) investigated 39 classic polymorphisms in two Chibchan populations (Cabecar and Huetar) in order to determine the relationship between an isolated population (Cabecar) and a highly admixed group (Huetar). They detected two private polymorphisms (TF*DGUA and PEPA*F) only found in Costa Rican and Panamanian Chibchan groups. They identified a strong genetic relationship between the Huetar and Cabecar despite admixture in the Huetar (Barrantes 1993).

Only a handful of classical genetic studies have been conducted on living non-Chibchan populations in Central America over the past twenty years. Azofeifa *et al.* (1997) examined genetic diversity and admixture using red blood cell enzymes, hemoglobins, and serum proteins in 59 Miskito refugees. The origins of the Miskito are uncertain, with some ethnologists considering them a post-contact population

while others consider them a Native American population with linguistic affinities with Chibchan populations. These authors sought to determine whether the Miskito demonstrated significant amounts of European and African admixture. They concluded that this population contained at a minimum 80% Native American genetic polymorphisms as well as private alleles previously only detected in lower Central American Chibchan populations.

Two studies investigated HLA polymorphisms in extant Olmec and Mayan populations (Arnaiz-Villena *et al.* 2000; Gomez-Casado *et al.* 2003). Arnaiz-Villena and colleagues (2000) studied HLA class I and II markers in 80 Mazateacans and compared them to worldwide indigenous populations. Using correspondence analysis they found a genetic relationship among Amerindian populations that distinguished them from other global indigenous groups. They also postulated indirect evidence for a genetic relationship between Mazateacan and Mayan groups but did not provide statistical support for this relationship. Gomez-Casado *et al.* (2003) studied HLA diversity in 132 Mayan individuals from Guatemala and compared them to eight other Native American populations. They concluded that: 1) Mayan populations were genetically closest to the Ijka (Arhucao), a Chibchan speaking population from northeast Colombia; 2) Mayans did not appear to be related to geographically neighboring populations based on HLA data (Zapotec, Mixe, and Mixtec); and 3) a separate cluster of Mesoamerican and South American indigenous populations was represented in their correspondence analysis. However, a few problems may be noted for these results. First, they recruited participants based on their “Mayan” appearance

and not based on genealogical history, which questions the validity of their sampling method. They also misidentify the Arhuaco (Ijka), a Chibchan speaking population, as the original inhabitants of the Caribbean. In this case, they are misidentifying the Arhuaco with Arawak, a South American linguistic family and Caribbean ethnic populations. The term *Arhuac* was introduced by the Spanish in the 17th century to define the southern slopes of the Sierra Nevada de Santa Marta in Colombia and referred to pacified Indians but is now often used to refer to the Ijka (Uribe 2000). All three of the previous classical genetic polymorphism research on non-Chibchan populations has suggested a biological relationship with Mesoamerican populations, but this relationship has never thoroughly been investigated.

Another research question that has not been fully addressed by investigators is the biological relationship between Central and South American Chibchan speaking populations. Five previous studies have investigated classical polymorphims in South American Chibchan populations, and the majority of this research was concerned with genetic admixture in these populations (Layrisse *et al.* 1963; Salzano and Callegari-Jacques 1988; Yunis *et al.* 1994; Briceño *et al.* 1996). Layrisse *et al.* (1963) compared blood group polymorphisms in four South American "Chibchan-speaking" populations. These authors could not find a clear connection among these populations and concluded that Chibchan populations did not form a homogeneous genetic group, but stipulated that more research was necessary. However, three (Tunebo, Páez, and Warrau) of these populations are Paez-speakers and do not belong to the Chibchan linguistic family. Salzano and Callegari-Jacques (1988) utilized data from Layrisse *et*

al. (1963) on the Ijka to investigate European and African admixture in the Ijka. This study found a small amount of African (7%) and European (7%) admixture in this population. Yunis *et al.* (1994) investigated seven HLA (DQA1, DQB1, DRB1, DRBE, DRB3, DRB4, and DRB5) alleles and eight blood groups polymorphisms (ABO, MNS, Rh, Kell, Duffy, Kidd, and Diego) in three Chibchan populations (Kogi, Arsario, and Ijka). These researchers found no genetic admixture among the Kogi or Arsario but a small amount of European (<1%) and a high amount of African admixture (22%) in the Ijka. Briceño *et al.* 1996 investigated genetic variation in the HLA-DRB1 locus in these same three South American Chibchan populations. They found no evidence for genetic admixture in the Kogi, Arsario, or Ijka. Only a single classical genetic study investigated the biological relationship among Central and South American Chibchan populations, Layrisse *et al.* (1995) investigated 22 Chibchan and Paezan speaking populations and detected a correlation between genetic and linguistic data with higher affinity among Chibchan-speaking populations.

Previous research utilizing classical genetic polymorphisms on Chibchan populations has supported the idea that these groups are closer genetically to each other than with other Central, South, or North American indigenous populations. However, these analyses have not focused on Mayan populations from the north and have not attempted to demonstrate shared biological relationships among Chibchan speaking groups from northern South America.

MTDNA

Indigenous Central American populations are primarily characterized by high frequencies of mtDNA haplogroup A2 (> 50%), moderate frequencies of haplogroup B2 (20% to 40%), and low frequencies of haplogroups C1 and D1 (<10%). Haplogroup X2a has not been identify in Central American populations but has been detected in small frequencies in Mexican indigenous populations (Penaloza-Espinosa *et al.* 2007). Evidence from mtDNA research on Chibchan populations has also been used to support the idea of *in situ* genetic evolution model promoted by Barrantes *et al.* (1990). This model suggests that Chibchan populations originated in lower Central America from local Paleoindian populations approximately 10,000 YBP due to ecological changes in the region and have differentiated into the various subgroups since that time.

A number of studies (Torroni *et al.* 1993, 1994; Santos *et al.* 1994; Batista *et al.* 1995; Kolman *et al.* 1995; Kolman and Bermingham 1997) have investigated mtDNA haplogroup data in Chibchan populations in order to establish their role in the peopling of the Americas. These groups are characterized by high frequencies of mtDNA haplogroup A2 (>65%), moderate frequencies of B1 (20-30%), absence of C, and only two populations (Huetar, Boruca) demonstrated the presence of D1 (< 05%) (Torroni *et al.* 1994b; Santos *et al.* 1994). Kolman *et al.* (1995) argued the absence of haplogroup C1 in Central American populations suggest that it was not present at the onset of Chibchan genetic history. However, haplogroup C1 has been found in

Chibchan speaking groups from Colombia, which would appear to refute this claim (Melton *et al.* 2007). Torroni *et al.* (1994) used high resolution RFLP analysis to study eight Costa Rican and Panamanian tribes (Teribe, Maléku, Boruca, Kuna, Guaymi, Bribri, Cabecar) as a reference point for wider investigation of the origins of Amerind populations and to develop a coalescent based mtDNA molecular clock. They found a total of fifteen haplotypes, eleven of which were unique to Chibchan populations. They also found a *MspI* RFLP cut site only in Central American Chibchan populations. This RFLP would appear to be restricted to Central American groups as it has not been detected in South American Chibchan speakers (Melton *et al.* 2007).

A small number of previous research studies have investigated mtDNA HVS-I sequence variation in three Central American populations (Huetar, Guaymi, and Kuna) (Santos *et al.* 1994; Batista et al 1995; Kolman *et al.* 1995) and three South American Chibchan populations (Kogi, Arsario, and Ijka) (Melton *et al.* 2007). These studies all demonstrated low haplogroup and haplotype diversity, reproductive isolation from outside groups, and long-term independent evolution had occurred in lower Chibchan-speaking populations. In order to determine if neighboring populations speaking different languages demonstrated similar genetic diversity, Kolman and Bermingham (1997) compared two Chibchan populations (Kuna, Guaymi) to two Chocoan (Emberá, Wounan) speaking groups from Panama. They found that Chocoan populations demonstrated larger haplotype and nucleotide diversity than their Chibchan neighbors. This indicated a loss of genetic diversity in

Chibchan population in Panama and indicated greater differentiation among Chibchan populations and their neighbors than one would expect through random chance. In addition, the K'iche Maya population demonstrated high haplogroup diversity (0.93), even though these individuals were considered related as part of a human rights forensic investigation when compared to Chibchan populations. The traditional interpretation of low diversity values for Chibchan populations has been attributed to genetic drift but alternative explanations include depopulation due to European contact or cultural factors associated with marriage practices (Barrantes *et al.* 1990; Thompson *et al.*. 1992; Kolman and Bermingham 1997).

A few mtDNA coalescent dates have been published for Central American populations. These include molecular dates for the Ngöbé, of 6,800 ya (Kolman *et al.* 1995) and Kuna of ~10,000 ya (Batista *et al.*1995) along with a combined haplogroup A2 date for both these groups of 10,900 ya (Kolman & Bermingham 1997). These molecular dates have often been used to support the idea that Chibchan populations diverged from earlier Paleoindian populations in the last 10,000 years. These molecular dates are also consistent when compared to archaeological (Hoopes and Fonseca, 2003) and glottochronological evidence (Constenla 1991) for the divergence of Chibchan populations.

Available evidence from previous mtDNA studies supports the idea of long-term population continuity for Chibchan populations in lower Central and northern America. This long occupation of the region may have slowed the migrations of

populations in lower Central America between north and south over the last 10,000 years. The maternal genetic diversity of Chibchan groups is lower than neighboring populations and would appear to have a biological relationship with Mesoamerican populations. This genetic relationship between Chibchan and Mesoamerican populations has often been neglected by researchers in the region. This oversight may be due to the lack of genetic information on living indigenous Mesoamerican populations with only one previous study of mtDNA HVS-I variation being done on the K'iche Maya (Boles *et al.* 1995) and the majority of other mtDNA Mesoamerican research only consist of mtDNA haplogroup data.

Y-CHROMOSOME

Only a few studies have investigated Y-chromosome diversity in indigenous Central American populations. Two studies included Chibchan and Mesoamerican populations in wider analyses regarding the initial peopling of the Americas (Lell *et al.* 2002; Zegura *et al.* 2004). Lell *et al.* (2002) included four Chibchan (Boruca, Guaymi, Bribri/Cabecar, and Kuna) populations. Two of these populations (Guaymi, Kuna) demonstrated only haplogroup Q, with the Native American specific haplogroup Q-M3 being at high frequencies (<60%) in all four Chibchan populations. The Boruca demonstrated a low frequency of haplogroup F (4.3%) and the Bribri/Cabecar showed a small amount of haplogroup K (7.7%). However, these latter two frequencies may be anomalies as haplogroup Q is a derivative of both Y-chromosome haplogroups F and K. In the hierarchical nomenclature of human Y-

chromosome haplogroups and the limited number of SNPs (11) used by Lell *et al.* (2002) it is possible that these small frequencies are related either to back mutations within haplogroup Q or may be attributed to laboratory error. Zegura *et al.* (2004) also included four Mesoamerican (Mixtec, Zapotec, Mixe, and Maya), two Chibchan (Ngöbé, Kuna) and two Chocoan (Emberá, Wounan) in their genetic analysis of Native American Y-chromosome diversity. The Mesoamerican and Chocoan groups primarily demonstrated haplogroup Q with a small amount of haplogroup R. The two Chibchan populations showed only haplogroup Q. The majority of these populations also demonstrated low SNP diversity and repeat number variance. Two of these populations (Zapotec, Ngöbé) contained high repeat number variance, 0.76 and 0.75 respectively. This may be indicative of higher levels of male population movement in the past or it may be representative of larger population size. The Ngöbé of Panamá is one of the largest indigenous populations (~170,000) in the region and therefore this high repeat number may be the result of a larger effective population size (Brignoli 2005). However, the sample size for all of these populations was low and no attempt in either of these studies was made to understand the population structure in Central America.

Only four studies have examined Y-Chromosome variation in Central American indigenous populations (Ruiz-Narvaez *et al.*, 2005; Campos-Sanchez *et al.*; 2006; Barrot *et al.*, 2007; Lovo-Gomez *et al.*, 2007). Lovo-Gomez and colleagues (2007) investigated 16 Y-chromosome SNPs and eight STRs in four El Salvadoran ethnic groups and compared them with admixed metropolitan populations. As

expected, they found a high frequency of haplogroup Q3 in indigenous groups and a high frequency of European haplogroup R1b in metropolitan populations. Barrot *et al.* (2007) studied eight Y-chromosome STR loci in three Amerindian populations (Huasteco, Otomies de la Sierra, Otomie de Valle, Tepehuas) inhabiting the Mexican state of Hidalgo. They found high gene diversity values for all four groups ranging from 0.95 for Otomies de la Sierra to 1.0 for the Tepehuas. The high gene diversity for the Tepehuas may be explained by the low sample size of 13 individuals from this population (Barrot *et al.* 2007). However, these authors did not include any SNPs and therefore did not categorize any of these populations into Y-chromosome haplogroups.

Campos-Sanchéz *et al.* (2006) investigated Y-chromosome diversity along with mtDNA diversity in admixed Hispanic populations related to schizophrenia from Costa Rica, Mexico, and the Southwest U.S. They detected a strong maternal association based on mtDNA and a weaker relationship between paternal lineages based on Y-chromosome diversity. Similar to the Hidalgo study (Barrot *et al.* 2007) they identified high Y-chromosome gene diversity levels for all three populations (Costa Rica=0.997, Mexico=0.995, U.S=0.998). However, these authors did not attempt to place the sampled populations into a proper phylogenetic context or develop admixture estimates between European and Native American samples. The mtDNA and Y-chromosome phylogenies are well developed and would have helped them better elucidate the relationship between these populations.

One study examined Y-chromosome diversity in five Chibchan-speaking populations (Huetar, Bribri, Cabecar, Guaymi, and Teribe) from Costa Rica (Ruiz-Narvaez *et al.*, 2005). Estimated haplotype diversities for all five of these populations was lower than in either the Campos-Sanchéz *et al.* (2006) or Barrot *et al.* (2007) study. Diversity values ranged from 0.942 in the Huetar to 0.679 in the Guaymi (other values Bribri=0.89, Cabecar=0.717, Teribe=0.724). The authors explained the high diversity in the Huetar as potentially due to increased admixture from European populations. Low Y-chromosome diversity values of the Guaymi may be based on their patrilocal marriage pattern (Kolman *et al.* 1995). Based on their phylogenetic analysis, four Chibchan populations (Cabecar, Huetar, Teribe, Bribi) clustered with the Gavaio from Brazil and Cayapa from Ecuador. The phylogenetic relationship between the Huetar and Cabecar was also demonstrated in classical genetic data (Azofeifa *et al.*, 2001). The only Chibchan group separated from other lower Central American populations are the Guaymi, who are isolated on their own branch. The relationship between the Gavaio and the Bribri is uncertain and may be related to an anomaly similar to the pairing of the Yanamamo and the Guaymi with classical genetic data (Spielman *et al.* 1979).

A single study investigated Y-chromosome variation in three South America Chibchan (Ijka, Arsario, and Kogi) speaking populations (Guarino *et al.* 1999). This study examined Y-chromosome variation in five STRs (DYS19, DYS390, DYS391, DYS392, and DYS393) on a small number of males in the study populations. These researchers found shared haplotypes between all three of these South American

Chibchan , but not with neighboring populations indicating a shared evolutionary history between these three Chibchan groups. Based on unpublished SNP data for these three populations, the majority of these individuals belong to haplogroup Q-M3 (R. Mitchell pers. comm.). As with both classical and mtDNA, the available Y-chromosome literature does not attempt to construct a broader picture of Central American population structure or determine the relationship between Central and South American Chibchan speaking populations.

GENETICS SUMMARY

A number of important observations can be made from previous genetics research on Central American and Chibchan populations. First, there is consistent evidence of some characteristics that appear to distinguish Chibchan-speakers from other Native American populations in the region. Chibchan populations. These groups are characterized by certain classical genetic private polymorphisms that are not shared with surrounding populations. Low mtDNA haplotype diversity in these groups suggest indicates that genetic drift may have impacted these populations in the past. Finally, molecular dates, indicate a long-term occupation of Central America for Chibchan populations. However, there are two potential issues in previous genetic research on Chibchan populations. First, this research has not focused on the biological relationship Chibchan-speakers with other Mesoamerican populations. This omission is unfortunate as it fails to take into account complex population dynamics that occurred in the region during the last millennia before European contact. During

this thousand year period, several populations moved along the Pacific coast into the region from Mexico (Fowler 1989; Cooke 2005). The second deficiency is biological research on Chibchan populations often does not account for the presence of several linguistically related Chibchan populations in northern South America. Modern political boundaries are often used as Pre-Columbian barriers when interpreting Pre-Columbian history in the region. However, these barriers are often arbitrarily defined and often do not reflect the Pre-Columbian demographic of genetic history in the regions. Through addressing these issues it is possible to fill in a gap in our understanding regarding the population dynamics and genetic history of indigenous populations inhabiting Central America.

DIVERGENCE OF CHIBCHAN POPULATIONS

Four models have been proposed for the divergence of Chibchan populations from earlier human groups in Central America. The first of these theories suggests that Chibchan groups migrated into Central America from northern South America and replaced earlier populations. Early scholars in the region proposed that the region was populated by groups migrating north from Colombia around 1,000 YBP (Stone 1966). This idea of external South American cultural influences on the origins of Central American populations was first proposed by Stone (1972; 1977) who advocated that the isthmus was an “archaeological bridge” based on iconography and architectural affinities in cultural material in Central America. Originally, this cultural relationship was attributed to the movement of Arawak speaking groups moving into

the region (Stone 1972), but was later revised to indicate a greater influence of an Ecuadorian religious complex (Stone 1977). The second explanation proposes that Chibchan population are an offshoot of Mesoamerican groups that spread south into the region along with the introduction of maize approximately 1,500 YBP (Snarskis 1978; 1998). Swadesh (1955) also advocated a Mesoamerican origin for Chibchan populations that occurred approximately at 7,000 YBP following the linguistic fission the Macro-Maya, Uto-Aztecan and Chibchan language families. More recently, Snarskis (2003) has advocated a "two-wave" model of population movement with an original movement south by Olmec-like Mesoamerican groups occurring about 2,000 YBP along with the introduction of maize and a second migration north of tropical lowland South Americans occurring around 1,500 YBP with the introduction of gold, tombs, and platform mounds, which are believed to have appeared earlier in northern Colombia (Snarskis 2003). The final model is that Chibchan populations evolved endogenously in Central America from earlier Paleoindian groups, without influence from other regions (Barrantes *et al.* 1990; Constenla 1991; Hoopes and Fonseca 2003). Based on current, archaeological, genetic, and linguistic evidence it would appear that the endogenous development model is the most parsimonious explanation for Chibchan origins. However little research has been conducted on the biological relationship between Mesoamerican and Chibchan populations.

MESOAMERICAN-CHIBCHAN BOUNDARY

One linguistic stock of Chibchan populations that has important implications for understanding the biological and cultural relationship between Mesoamerican-and the Isthmo-Colombian populations is those groups that speak Votic dialects. There are currently three living Votic (Rama, Maléku, Huetar) and two extinct Votic (Voto, Corobici) speaking Chibchan groups from Nicaragua and Costa Rica. These populations form the northern and western boundaries of Chibchan populations (with the exception of the Pech in Honduras) in Central America. There is archaeological evidence that indicates these Chibchan groups interacted with both Oto-Manguean (Chorotega) and Nahua (Nicarao) speaking groups. These latter two linguistic families are believed to have migrated from Mesoamerican between 1,200 and 800 YBP (Fowler 1989). The modern day Rama are bordered to the north by Misumplan (Miskito, Sumo) groups (Riverstone 2004). Votic populations are bordered to the south by Isthmic Chibchan groups. The Isthmic (Guaymi, Cabecar, Bribi, Teribe, and Ngöbé) speaking populations are considered to be the ancestral Chibchan populations based on linguistic (Constenla 1991, 1995), genetic (Barrantes *et al.* 1990), and archaeological diversity (Hoopes and Fonseca 2003). Therefore, Votic Chibchan populations provide an opportunity to better understand the genetic structure and their biological relationship to Mesoamerican populations. This section provides historical and ethnographic information regarding the five Native American populations (Rama, Huetar, Maléku, Guaymi, and Chorotega) investigated in this dissertation.

RAMA

The Rama Indians are a Votic-speaking Chibchan population who inhabit the southeastern Caribbean coast of Nicaragua. Current population estimates for the Rama range from 1,350 (Brignoli 2005) to their own estimates of approximately 3,000 (L. Martinez pers. comm.). Rama individuals live in nine communities spread from the coastal port town of Bluefields, Nicaragua, to San Juan del Norte, Costa Rica. The largest Rama community is Rama Cay, a small island in Bluefields Bay, where approximately 1,000 Rama reside. The modern Rama subsist through hunting, gathering, horticulture, and fishing. This subsistence pattern does not appear to have changed significantly since their first contact with Europeans (Riverstone 2004). The Rama belong to the Moravian Church, which arrived in the Caribbean in the 1850s (Riverstone 2004, Conzemius 1932). However, in recent years, a number of Rama have advocated the retention and revitalization of traditional cultural practices (Riverstone 2004). The Rama view this cultural revitalization as a way to maintain a distinct ethnic identity in an increasingly homogenous global society.

An indigenous group named the Rama does not appear in historical documents until the 18th century. This has led to numerous competing scenarios regarding their origins. These proposals suggest that the modern Rama are a remnant population of the Suerre, Huetar, Corobici, Maléku, Voto, Melchora, Kukra, or the Caribs. Several of these proposed groups are extinct (Suerre, Voto, Corobici, Kukra, Melchora) and their relationship with the Rama is based solely on ethnohistoric

accounts that often contradict each other (Riverstone 2004). The association of the Rama with the Huetar is more difficult to determine as this population was decimated by the Spanish in the 16th century and some of these individuals may have fled north and intermixed with indigenous populations in the Caribbean. The currently most favored explanation is that this population split from the Maléku shortly after contact with the Spanish and has been isolated in the mangrove forest since that time (Riverstone, 2005, Conzemius 1932). However, the Rama themselves believe that they have inhabited the region for centuries and their territory once spread to the shores of Lake Nicaragua (C. McCrea pers comm.). There is very little known about the biological relationships of the Rama to other Native American groups in the area. The Rama were investigated for classical genetic markers in the 1950s (Matson and Swanson 1963c) and for PTC tasting in the 1970s (DeStefano and Molieri 1976).

MALÉKU

The Maléku (Guatuso, Maléku Jaíka) are a Votic-speaking Chibchan population that currently resides in three small localities in northern Costa Rica. Historically, the Maléku occupied the northern highlands of Costa Rica and this may have extended to the shores of Lake Nicaragua (Cooke 1997). Today, the population number of this group is approximately 1,000 individuals (Barrantes *et al.* 1990, Brignoli 2005). The origins of the Maléku are also poorly understood. They have been considered to be related to the Rama, a sub-group of the extinct Corobici (Lothrop 1926), a branch of the Huetar (Gabb 1875), or a remnant group of the Voto,

an extinct Chibchan population that once inhabited the mouth of the Rio San Juan in northern Costa Rica (Riverstone 2004). The Maléku have been previously characterized for both classical genetic polymorphisms and mtDNA RFLP variation (Barrantes *et al.* 1990; Torroni *et al.* 1994b). These studies have suggested that they are biologically related to other Chibchan populations and are distinct from other indigenous groups in the region.

HUETAR

The Huetar (Güetar) are a Votic-speaking Chibchan population who now inhabit two isolated communities (Quitirrisi and Zapáton) west of San Jose, Costa Rica and are estimated to number around 1600 individuals (Brignoli 2005). In the 16th century, the Huetar occupied the central highlands of Costa Rica west to the Gulf of Nicoya. The Huetar population was decimated and acculturated after contact with the Spanish. Other than the two extant highly acculturated communities, they are considered to be culturally extinct. The Quitirrisi Huetar have previously been investigated for classical genetic polymorphisms (Barrantes *et al.* 1990; Bieber *et al.* 1996; Azofeifa *et al.* 2001), mtDNA HVS-I (Santos *et al.* 1994), and Y-chromosomes polymorphims (Ruiz-Narvaez *et al.* 2005). This previous genetic research has supported the idea that the Huetar are biologically related to other Chibchan populations but demonstrate distinct characteristics that differentiate them from other Native American groups.

GUAYMI

The Guaymi (Ngöbé) are an Isthmic-Chibchan speaking group that inhabits several southern localities in Costa Rica. This group immigrated to the region from Panama in the 1940s and is considered a smaller offshoot (n=3,500) of the larger Panamanian Ngöbé (n~169,000) population (Barrantes *et al.* 1982). There is little known regarding the ethnography of these communities but they appear to differ from other Chibchan groups in practicing patrilocal residence patterns (Kolman and Bermingham 1997). The Guaymi have previously been studied for classical genetic polymorphims (Spielman *et al.* 1979; Barrantes *et al.* 1982, Barrantes *et al.* 1990), mtDNA RFLPs (Torroni *et al.* 1994b), and Y-chromosome variation (Ruiz-Narvaez *et al.* 2005). These studies have supported the relationship of Guaymi with other Chibchan population while differentiating them from surround Native American groups.

CHOROTEGA

The Chorotega are an Oto-Manguean speaking population that inhabits the Gran Nicoya peninsula of Costa Rica and the Pacific coastal region of southern Nicaragua. This population is thought to have been forced from central Mexico by the Olmec and subsequently migrated down the Pacific coast of Central America into the region between 800 – 1,200 YBP (Fowler 1989; Constenla 1995). However, their biological relationship with neighboring Chibchan populations is unresolved. There is little previous biological research on the Chorotega, with the only previous study

being conducted on classical genetic polymorphims (Matson and Swanson 1963c). Therefore, understanding the biological relationship that this population has with surrounding Chibchan groups has important implications for understanding the association between Mesoamerican and Chibchan groups.

SUMMARY

This chapter provided background information for Chibchan populations and included previous archaeological and genetic research on population structure and cultural history of Central American indigenous groups in order to better understand their evolutionary history. Archaeological evidence indicates that lower Central America has been continuously occupied for the last 14,000 years (Cooke 2005) and that the region had developed into a culturally complex region by European contact (Hoopes 2005). A number of models have been proposed for the divergence of Chibchan populations from earlier Paleoindian groups. These models include populations migrating south from Mesoamerica, moving north from South America or the endogenous development of these groups within the Panamanian isthmus. Genetically, Chibchan populations demonstrate a number of characteristics that differentiate them from surrounding indigenous populations but their biological relationship to other Mesoamerican and South American populations is not completely understood. These questions of population divergence may be addressed through the use of molecular genetic markers that allows for a greater understanding

of population dynamics as well as the maternal and paternal genetic history of Chibchan populations.

IV: MATERIALS AND METHODS

This chapter describes the field work conducted in Nicaragua, laboratory methods used in the characterization of the DNA, and analytical techniques employed in this research. DNA was extracted from mouth rinses and was characterized for mtDNA and Y-chromosome markers including: mtDNA RFLPs; mtDNA HVS-I sequencing; Y-Chromosome SNPs; and Y-Chromosome STRs. In addition, classic genetic markers were compiled from the literature. Measures of genetic diversity were calculated. Phylogenetic relationships were visualized through multi-dimensional scaling (MDS) and phylogenetic networks. Population structure was characterized through the use of F-statistics (F_{ST}) and analysis of molecular variance (AMOVA). Genetic drift was ascertained through neutrality test statistics and heterozygosity vs. distance from the centroid. Phylogeographic analysis included SAMOVA, genetic landscapes, and Monmonier's algorithm.

FIELD METHODS

During the Fall of 2007, buccal swabs, mouth rinses, and genealogical information were collected for 75 (48 females, 27 males) Rama participants by Phillip Melton and Norberto Baldi, with the assistance of Rama elder Cleveland Macrea and his son Jerry Macrea in the communities of Rama Cay and Sumo Kat, Nicaragua. Permission for this study was granted by the University of Kansas Institutional Review Board (IRB) on human subjects. All participants signed an informed consent statement (Appendix A) and were provided as required contact information for the

researchers in case of future questions. In Sumo Kat a community meeting was organized and the basic goals of the project were discussed. Individuals who wished to participate visited the research site the next day. In Rama Cay, Cleveland Macrea made an announcement during the Sunday church service and individuals visited the research station the same day. Additional biological information was obtained from Dr. Ramiro Barrantes of the University of Costa Rica for 155 individuals from four indigenous Costa Rican populations (Chorotega n=30, Maléku n=35, Guaymi n=50, Zapáton Huetar n=40). The composition of each population by community and sex is shown in table 2 and their geographic location is displayed in figure 12. This resulted in a sample of 230 individuals from five indigenous Central American populations for this project. Comparative data on additional Chibchan populations was obtained from the literature (Santos *et al.* 1994; Kolman *et al.* 1995; Batista *et al.* 1995; Kolman and Bermingham 1997) and Master's thesis research conducted by Phillip Melton on Colombian Chibchan populations (Melton 2005; Melton *et al.*, 2007).

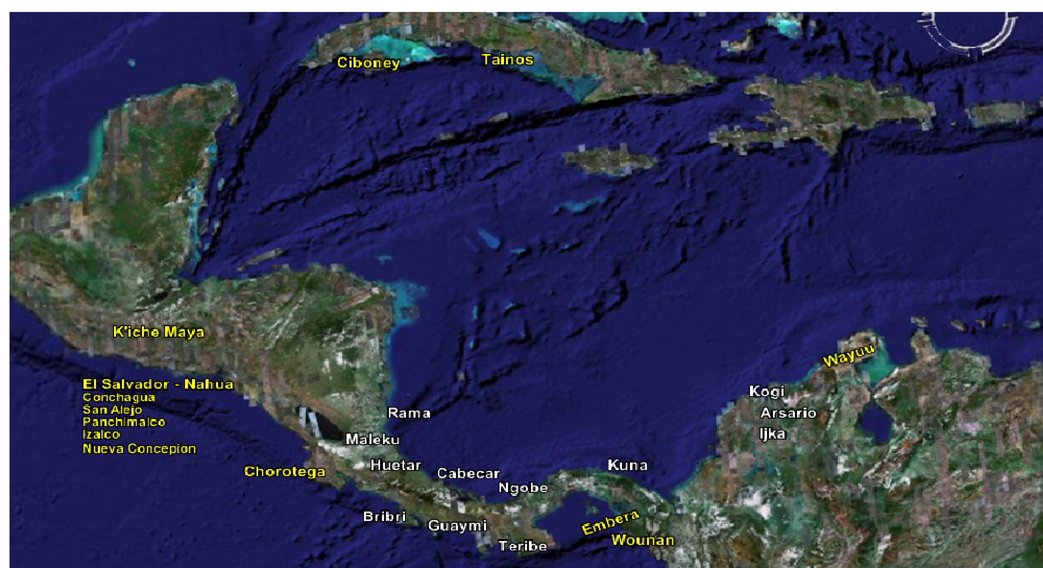

Figure 11: Geographic distribution of the five studied populations and other neighboring Native American indigenous groups. Chibchan speakers in white and neighboring populations in yellow.

Table 2: Study populations, communities where samples were obtained, male and female sample size and total sample size**.**

Population	**Community**	**Males**	**Females**	**Total**
Nicaragua				
Rama	Sumo Kat	14	17	31
	Rama Cay	13	31	44
Rama Total		27	48	75
Costa Rica				
Chorotega	Matambú	25	5	30
Maléku	Tongibe	4	11	15
	Margarita	10	10	20
Maléku Total		14	21	35
Huetar	Zapáton	13	27	40
Guaymi	Dos Brazos	0	16	16
	Abrojo	19	15	
Guaymi Total		19	31	50
TOTAL		98	132	230

LABORATORY METHODS

DNA EXTRACTION

All Rama DNA was extracted from mouth rinses using the Chelex method in the Genetics Laboratory at the University of Costa Rica. In the field, saliva was collected by having individuals swish 10mL of distilled water in their mouths and expectorate into a Dixie cup. The biological samples were then poured into a 15 mL collection tube and cells were allowed to settle to the bottom. Upon the return to the Genetics Laboratory in Costa Rica, a bulb pipette was used to transfer cells into a 2.0 mL microcentrifuge tube. The cells were pelleted in a microcentrifuge at 10,000 RPM for five minutes. The aqueous overflow was discarded and 100 µL of 10% Chelex solution was added to each sample. The suspended cells were placed in a 100°C water

bath for 10 minutes. This step lysed the cells and released DNA and proteins. The next step involved placing the samples on ice for three minutes, allowing for the Chelex to bind to everything except the DNA. Samples were centrifuged for five minutes at 10,000 RPM in order to pellet the Chelex bound material. The aqueous overflow, containing the DNA was transferred to new 0.5 mL tubes and stored at 4°C. This procedure was repeated twice for all Rama individuals. The presence and quantity of DNA was determined through the use of a spectrophotometer in the Molecular Biosciences department at the University of Costa Rica. DNA from the four Native Costa Rican populations was obtained from blood and extracted using phenol chloroform (Barrantes *et al.* 1990). Twenty five µM of DNA from each of the 155 samples was transferred to new 0.5 mL tubes and transported to the Laboratory of Biological Anthropology at the University of Kansas for analysis.

MTDNA ANALYSIS

Mitochondrial DNA analysis consisted of RFLP and sequencing of the HVS-I region of the mitochondrial genome. RFLP analysis consisted of identifying specific regions of mtDNA using polymerase chain reaction (PCR). Components for the RFLP PCR included 2.5 µL of 10X PCR Buffer (provided by Promega), 4.0 µL of $MgCl_2$ (25mM), 0.5 µL of dNTP nucleotide mix, 0.2 µL of GoTaq polymerase (Promega), 7.8 µL of ddH_2O, 2.5 µL of forward primer (10 pmole/µL), 2.5 µL of reverse primer (10 pmole/µL), and 1 to 3 µL of sample DNA (if more or less DNA was used, the ddH_2O was adjusted accordingly). This resulted in a 25 µL volume for

each sample. All reagents were purchased from Promega (Madison, WI) except for olgionucleotide primers, which were synthesized at Integrated DNA Technologies (Coralville, IA). Primers and their associated mtDNA haplogroups used in this research are listed in table 3. PCR reactions were run on a Perkin Elmer 2400 or an Applied Biosystems 9600.

Table 3: Primer pairs for mtDNA haplogroups and HVS-I sequence primers used in this study.

Haplogroup	Primer Pair	Sequence (5´→ 3´)	AT*
A2 (+*Hae*III 663)	535FOR 725REV	CCCATACCCCGAACCAACC GGTGAACYCACYGGAAGGGG	57ºC
B2 (+*Hae*III 8250)	8149FOR 8366REV	ACCGGGGGTATACTAACGGT TTTCACTGTAAAGAGGGTTGTTGG	53ºC
C1 (-*Hinc*II 13259 & +*Alu*I 13262)	13172FOR 13383REV	GCTTAGGCCCTATCACCA GTTGTGGATGATGGACCC	51ºC
D2 (-*Alu*I 5176)	5151FOR 5481REV	CTACTACTATCTTCGCACCTG GTAGGAGTAGCGTGGTAAG	53ºC
HVS-I	15976 FOR 16422 REV	CCACCATTAGCACCCAAAGCTAAG AATGATTTCACGGGAGGATGG	55ºC

*=Annealing temperature

The reactions were run according to the following thermal profile: 1) denaturing at 94°C for 40 seconds; 2) annealing for 30 seconds at the lowest melting temperature for each primer pair (see table 3); 3) extending at 72°C for 45 seconds. These steps were repeated for between 30 to 35 cycles. Reactions were checked for PCR product on a 1.5% agarose amplification gel using electrophoresis at 97 volts for approximately one hour. The reagents for the agarose gel included 100µL of 1X TBE, 1.5g of SeaKem agarose, cooled to 45°C and stained with ethidium bromide. A total of 5 µL of PCR product and 2 µL of 6X loading dye (Promega, Madison WI) were added to each well and checked against a 50 bp DNA ladder (Promega, WI) and a

positive (sample of know haplogroup) and negative control (ddH_2O) DNA was visualized through ultraviolet light.

After the initial PCR amplification, base substitutions or insertion/deletion events were characterized with restriction endonucleases. This research investigated the restriction site variation of samples characterized for each of the four major Native American mtDNA haplogroups (A2, B2, C1, and D1). These haplogroups are characterized by cut sites determined by specific restriction enzymes (*Hae*III, *Hinc*II, *Alu*I, *and DdeI*). The restriction enzymes and associated mtDNA haplogroup used in the study are listed in table 3. Reagents for the restriction digest include 2.0 µL of 10X RFLP buffer (New England Biolabs, Beverly, MA), 1.0 µL of 100 X bovine serum albumin (BSA), 0.5 µL of restriction enzyme (New England Biolabs, Beverly, MA), 9.0 µL of ddH_2O, and 7.5 µL of PCR product DNA. This resulted in a reaction volume of 20 µL per sample. RFLP samples were digested for ten to eighteen hours at 37°C using the appropriate restriction enzyme. The reaction was stopped by the addition of 5 µL of 3X loading dye to each sample. Digested fragments were visualized on a 3% 3:1 NuSieve gel (ISC BioExpress, Kaysville, UT) at 97 volts for two hours. The digested products were visualized under ultraviolet light.

All 230 samples were characterized for the four founding mtDNA Native American haplogroups (A2, B2, C1, and D1) in a hierarchical fashion. Previous research on Chibchan populations had indicated that the majority of these populations belonged to haplogroups A2 and B2. All 230 individuals in this research were tested

for haplogroup A2, which is recognized through the presence of a *Hae*III cut site at mtDNA nucleotide (nt) site 663. Those who do not belong to A2 were tested for haplogroup B2 through the presence of a *Hae*III cut site at nt 8250. Remaining non A2 and B2 individuals were tested for haplogroup D1 through the presence of an *Alu*I cut site at nt 5176. The remaining individuals were tested for haplogroup C1. If the individual did not belong to any of the Native American haplogroups, the HVS-I sequence of that individual was used to determine the mtDNA haplogroup. Samples were scored with a 1 if a restriction site was absent and a 2 if a site was present.

Approximately 400 base pairs (16000-16400 nt) of HVS-I of the mtDNA control region were sequenced on an automated capillary system using the Sanger dideoxy method (Sanger 1977). In preparation for sequencing, DNA templates were created using a 25 µL PCR reaction (same protocol as RFLP PCR). Primers annealing temperatures and primer information used for DNA sequencing are listed in table 3. Amplified PCR products were checked for on a 1.5% agarose amplification gel. These products were then purified using a QIAquick kit (Qiagen Valencia, CA) according to the manufacturer's instructions. A 5:1 ratio of Buffer PB was added to the PCR product in order to bind the DNA and this solution was then placed in a spin column and centrifuged at 13,000 RPM for one minute. Each sample was washed with 750 µL of Buffer PE and centrifuged at 13,000 RPM for one minute. The column was transferred to a new 1.5 mL tube and 30 µL of ddH_2O was added and allowed to stand for one minute. The solution in the 1.5 mL tube was centrifuged for

one minute at 13,000 RPM in order to release purified DNA sample from column and collected in the bottom of the tube.

The DNA templates were sequenced at the University of Kansas Sequencing Laboratory by Dr. Michael Grose. These samples were analyzed using Big Dye Sequencing kits on an ABI 3130 Sequencer (Applied Biosystems, Foster City, CA). The sequencing reaction included: 4.0 µL of Big Dye Ready Reaction Mix; 1.0 µL of either forward or reverse primer; and 4.0 µL of DNA template. This PCR reaction was run according to the following thermal profile: 1) 96°C for 10 seconds; 2) 50 °C for five second; and 3) 55 °C for four minutes and repeated 25 times. Unused primers and dNTPs were removed by passing them through a spin column and dried in a speed vacuum. Following manufacturer's instructions, dried samples were prepared for sequencing by adding 20 µL of ABI template suppression buffer heated to 95 °C for three minutes and then placed on ice. DNA sample were then transferred to ABI tubes and loaded onto the ABI 3130 sequencer. The sequencing gels were run overnight and the resulting chromatogram data were recorded on a computer.

The mtDNA sequencing chromatograms were edited using the BioEdit computer program (Hall 1999) and compared to the revised human Cambridge Reference Sequence (CRS) (Anderson *et al.* 1981; Andrews *et al.* 1999). Variations in nucleotides deviating from the CRS were recorded as DNA sequence variants.

Y-CHROMOSOME ANALYSIS

Male participants in this study were characterized for Y-chromosome STRs and SNPs. The STR analysis was done at the Laboratory of Biological Anthropology at the University of Kansas by the author. Eight STRs (DYS19, DYS389I, DYS389II, DYS 390, DYS391, DYS 392, DYS 393, and DYS439 see table 4) were analyzed in three separate PCR reactions and multiplexed for fragment analysis on an ABI 3130 at the KU Sequencing laboratory by Dr. Michael Grose.

Table 4: Y-Chromosome STRs used in the study

STR	Repeat Sequence*	Ref.
DYS19	(TAGA)3tagg(TAGA)n	4
DYS389I	(TCTG)3 (TCTA)n	2, 3
DYS389I	(TCTG)n(TCTA)nN28(TCTG)3 (TCTA)n	2, 3
DYS390	(tcta)2(TCTG)n (TCTA)n(TCTG)n(TCTA)ntca(tcta)2	2, 3
DYS391	(tctg)3(TCTA)n	2, 3
DYS392	(TAT)n	2, 3
DYS393	(AGAT)n	2, 3
DYS439	(GATA)n	1

* GenBank top strand; References: 1. Ayub *et al.* 2000; 2. De Knijff *et al.* 1997; 3. Kayser *et al.* 1997; 4. Roewer *et al.* 1992;

The following PCR protocol was used for the first Y-Chromosome multiplex reaction (DYS390, DYS391, and DYS393): 4.4 µL 5X Flexi Buffer: 3.8 µL $MgCl_2$; 1 µL dNTP; 0.3 µL GoTaq Polymerase; 0.4 µL BSA; 5.7 µL ddH_2O; 2.0 µL primer mix; 4.4 µL DNA dilution. This solution resulted in a total volume of 22 µL for each sample. Amplification reactions were run on a Perkins Elmer 2400 according to the following parameters: initial incubation at 94 °C for 3 minutes; 35 cycles of denaturing at 94°C for 30 seconds, annealing at 57°C for 30 seconds, and extension at

72 °C for 30 seconds; a final extension at 72 °C for three minutes; and then on hold at 4 °C.

The following PCR protocol was used for the second Y-Chromosome multiplex (DYS19, DYS391, DYS393, DYS439) and for DYS389 I/II: 3.6 µL 5X Flexi Buffer; 3.1 µL $MgCl_2$; 1.0 µL dNTP; 0.3 µL GoTaq polymerase; 0.3 µL BSA; 4.1 µL ddH_2O; 2.0 µL (0.8 µL forward and reverse for DYS389I/II); and 3.6 µL DNA dilution. This solution resulted in an 18 µL total volume for each sample. Amplification reactions were run on a Perkins Elmer 2400 according to the following parameters: initial incubation at 94 °C for 3 minutes; 35 cycles of denaturing at 94°C for 25 seconds, annealing at 56°C for 30 seconds, and extension at 72 °C for 30 seconds; a final extension at 72 °C for three minutes; and then on hold at 4 °C. All reagents for these PCR protocols were purchased from Promega (Madison, WI) except for the fluorescent tagged primers that were purchased from Integrated DNA Technologies (Coralville, IA). A total of 5 µL from each PCR sample was added to 85 µL of ddH_2O for restriction fragment analysis at the KU Sequencing laboratory. An ABI 3130 Sequencer (Applied Biosystems) was used for electrophoresis and detection of amplified product.

Y-Chromosome SNP analysis was completed at the University of Kansas Laboratory of Biological Anthropology by the author. In order to identify specific Y-SNPs, Y-Chromosome STR haplotypes matches were checked at http://ysearch.org and tested for corresponding Y-SNP haplogroups. Y-SNPs were investigated in a

hierarchical fashion. All male samples were investigated for the Y-chromosome SNP Q-M3. Subsequent SNPs investigated included R1b, E3, and Q-P36. The PCR protocol for this procedure was as follows: 5µL of 5X flexi buffer; 4.3µL MgCl2; 0.5µL dNTP mix; 0.2µL GoTaq polymerase; 5.0µL ddH2O; 2.5µ forward primer; 2.5µL reverse primer; and 5.0µL DNA dilution. PCR ingredients were purchased from Promega (Madison, WI), except for the primers which were synthesized by IDT (Coralville, IA). Primers and annealing temperatures are show in table 5. PCR products were run on a Perkin Elmer 2400 according to the following thermal profile: initial denaturation at 94ºC for one minute; and then 35 cycles of denaturing at 94ºC for 40 seconds, annealing for 30 seconds, and extension

Table 5: Primers for Y-SNP Analysis

Y SNP (Haplogroup)	Primer Pairs	Sequence (5´→ 3´)	AT*
M3 (Hap Q3)	M3 FOR M3 REV	TAATCAGTCTCCTCCCAGCA AAAATTGTGAATCTGAAATTTAAGG	60ºC
P36 (Hap Q)	P36 FOR P36 REV	TGAAGGACAGTAAGTACACA TAAGTCCATTGATCTACAGA	62ºC
M269 (Hap R1b)	R1b FOR R1b REV	CTAAAGATCAGAGTATCTCCCTTTG AAATTGTTTTCAATTTACCAG	58ºC
P39 (Hap C)	P39FOR P39REV	AGAAGGACTGCCTCAGAATGC GTTCGAAAGGGGATCCCTGG	60ºC
P2 (Hap E3)	P2 FOR P2 REV	GATGCAAATGAGAAAGAACT CTAAAAACTGGAGGGAGAAA	62ºC
M286 (Hap G2a)	M287 FOR M287 REV	TTATCCTGAGCCGTTGTCCCTG TGTAGAGACACGGTTGTACCCT	60ºC
M170 (Hap I)	M170 FOR M170 REV	TGCTTCACACAAATGCGTT CCAATTACTTTCACCATTTAAGACC	60ºC

*AT=Annealing temperature.

at 72ºC for 45 seconds; a final extension of 5 minutes at 72ºC, and a hold at 4ºC. Resulting DNA templates were cleaned using QiaQuick kits (see page 92), and

sequenced by Dr. Michael Grose at the University of Kansas Sequencing Laboratory. Sequences were aligned in the program BioEdit (Hall 1999) and the presence of SNP DNS sequence variants were recorded and used for Y chromosome haplogroup assignment.

ANALTYICAL TECHNIQUES

Diversity Measures

Haplotype diversity, nucleotide diversity, and neutrality test statistics were calculated for haplotypic data using the computer program Arlequin 3.1 (Excoffier *et al.* 2005). Haplotype diversity (Nei 1987) was calculated for mtDNA HVS-I sequences and Y-chromosome STR data. This measure is equivalent to the expected heterozygosity for diploid systems and is calculated using the equation

$$H = \left(\frac{n}{n-2}\right)\left(1 - \sum_{i=1}^{k} p_i^2\right) \qquad (1)$$

where n represents the sample size, k is the number of haplotypes, and p_i is the frequency of the ith haplotype. This is considered a stable measurement and is believed to be less responsive to genetic drift and recent demographic events (Helgason *et al.* 2003). Nucleotide diversity for mtDNA HVS-I sequence data was calculated as

$$\pi = \sum_{ij}^{a} X_i X_j d_{ij} \qquad (2)$$

where q is the total number of alleles, X_i is the frequency of the i-th allele in the population, and d_{ij} is the number of nucleotide differences between alleles i and j. Nei's (1987) coefficient of gene differentiation (G_{ST}) was used to determine diversity in classical genetic polymorphims. This was calculated in DISPAN (Ota 1993) using the equation

$$G_{ST} = \frac{(HT - HS)}{HT} \qquad (3)$$

where HT is the gene diversity among subpopulations (i.e, average of the allele frequencies for total data set), and HS is the gene diversity within subpopulations (i.e, the average of gene diversities for individual populations).

MEASURES OF SELECTIVE NEUTRALITY

In order to determine if mtDNA HVS-I sequences were statistically significant under the neutral equilibrium model, two measures of selective neutrality: Tajima's D (Tajima 1989) and Fu's F_S (Fu 1997) were investigated. The neutral evolution model is based on the standard Wright-Fisher evolutionary model and makes the following five assumptions: 1) a constant population size of N individuals; 2) random mating; 3) no overlapping generation: 4) no recombination; and 5) an infinite sites constant mutation rate process whereby an offspring differs from its parent allele by a Poisson-distributed number of mutations with mean μ (Simonsen *et al.* 1995). These neutrality test statistics are appropriate for distinguishing population growth from constant size groups. Population growth generates an excess of mutation in the external branches of a genealogy and therefore an excess of substitutions that are present in only one

sampled sequence (Ramos-Onsins and Rozas 2002). These mutations then lead to a star-like phylogeny that includes a large central node with several radiating spokes represented by a single individual.

Tajima's D uses mutation frequency information and is based on an infinite sites model without recombination (Tajima 1989). Therefore, this statistic is more appropriate for short DNA sequences or RFLP haplotypes. Tajima's D compares two estimators of mutation parameter θ, where $\theta=2N_e\mu$ (for haploid data, $4N_e\mu$). The parameter N_e is the effective population size and μ is equal to the mutation rate of the locus under investigation. The neutrality test statistic D is estimated as

$$D = \frac{\theta_\pi - \theta_S}{\sqrt{\mathrm{var}(\theta_\pi - \theta_S)}} \qquad (4)$$

where θ_π is equivalent to the mean number of pairwise differences between sequences (π) and θ_s is based on the number of polymorphic sites. Negative scores are indicative of larger values for θ_s relative to θ_π and signify the potential effects of population expansion. However, significant Tajima's D values may be caused by factors other than population expansion, including population bottlenecks, background selection or mutation rate heterogeneity (Aris-Brosou and Excoffier 1996; Schneider *et al.* 2000).

Fu's Fs is also based on the infinite-site model without recombination but utilizes information from the haplotypes distribution. This neutrality test statistic is defined using the equation

$$F_S = \ln \frac{S'}{1 - S'} \qquad (5)$$

where S' is the probability of observing a random neutral sample and is defined as S' = PR($K \geq k_{obs}|\theta = \theta\pi$), where k equals the number of alleles similar or less than the observed value given $\theta\pi$ and Fs is the logit of S'. Negative Fs values are indicative of an excess of alleles and signify population expansion. This neutrality test statistic is considered less conservative that Tajima's D and more sensitive to large population expansions expressed as large negative numbers whereas positive Fs values indicate groups impacted by genetic drift (Fu 1997; Ramos-Onsins and Rozas 2002).

MULTIDIMENSIONAL SCALING PLOTS (MDS)

In order to visualize biological relationships in reduced dimensional space multi-dimensional scaling plots (MDS) were constructed for mtDNA sequences, Y-chromosome STR data, and classical genetic markers using the NTSYS 2.1 computer program (Applied Biostatistics, Inc. Setanket NY). For mtDNA HVS-I sequences Tamura and Nei (1994) distances were used with a γ-value of 0.26 (Meyer *et al.* 1999). The Tamura and Nei distance corrects the percentage of nucleotides by which two base pairs differ and takes into consideration different mutation rates for transitions and transversions, by making a distinction between purine and pyrimidine transition rates. The γ-value of Meyer et al. (1999) corrects for mutation rate heterogeneity in the mtDNA HVS-I region and allows for a less biased estimate of genetic diversity and population history parameters. For Y-STR data Slatkins linear F_{ST} distance was utilized (Slatkin 1995). Both these genetic distances were calculated in Arlequin 3.1 (Excoffier *et al.* 2005). For classical genetic markers standard genetic distances were used and computed in the computer program DISPAN (Ota 1993).

MDS is a ordination method that is similar to Principal Coordinates and displays the dissimilarity of n objects represented in k-dimensional space (Kruskal 1964). This method is considered to be more precise than Principal Coordinate Analysis (PCA) in preserving small inter-point differences. This is because PCA maximizes variances and thereby gives greater weight to larger distances. Therefore, MDS is more appropriate for geographically proximal populations such as Chibchan-speaking groups. The MDS algorithm begins with a set of points produced by a PCA. The first step is to compute distances (d^*_{ij}) between all pairs of points (ij). These distances are then compared to the original distance (d_{ij}). A monotonic function (d^f_{ij}) is fitted to the variables and differences are computed as a sum of squared deviations. A stress value is then calculated in order to determine the goodness of fit of the projected to the original distances and is calculated with the equation

$$Stress = \sqrt{\frac{\Sigma\left(d^*_{ij} - d^f_{ij}\right)*2}{\Sigma d^{*2}_{ij}}} \qquad (6)$$

The project points are adjusted in order to lower the stress value. The stress value is considered a goodness-of-fit value. A high stress value is indicative that the chosen number of dimensions is not an accurate reflection of relationship between a set of objects. The stress level is dependent on the number of investigated data points. If the stress level is representative of less than 1% probability of obtaining a random distribution then resulting plot may be considered non-random and an accurate reflection of the investigated data. For example, a data set of 38 objects with a stress of 0.168 and displayed in two dimensions has a maximum stress of 0.348 before it

will be considered random. Therefore, a stress value of 0.168 in two-dimensions is considered non-random or not without structure (Sturrock and Roche 2000). MDS plots were constructed in Minitab v12 (Minitab Inc., State College, PA).

MEDIAN JOINING NETWORKS

Phylogenetic network analysis utilizing the median-joining method was used for determining genetic relationships among haplotypes found within the studied populations for mtDNA HVS-I sequence and Y-chromosome STR data. These types of networks offer an advantage over traditional phylogenetic tree building methods that utilize maximum parsimony or likelihood, because networks are able to differentiate between irresolvable and resolvable character conflict errors that may occur due to homoplasy. The resulting network then represents "all most parsimonious trees" by highlighting conflicts in the form of reticulations that can be interpreted as homoplasy, recombination, or sequencing errors (Bandelt *et al.* 1995).

Median joining (MJ) phylogenetic networks are applicable for this research because this analysis deals with multi-state data (nucleotide sequences and STRs). The assumptions of this type of network are that ambiguous states are infrequent and that recombination is absent. These assumptions are met for mtDNA sequences and Y-chromosome STRs when a stepwise mutation model is utilized (Bandelt *et al.* 1999). MJ networks for all four major Native American haplogroups (A-D) were generated individually. These resulting mtDNA networks were then reconstructed in the computer graphics program Fireworks (Adobe-Macromedia, San-Jose CA) and joined together based on a phylogeny of East-Asian HVS-I sequences (Kivisild *et al.*

2002). Y-chromosome STR networks were reconstructed for the major Y-Chromosome networks present in these Chibchan populations. For these analyses the Y389I STR repeat was subtracted from DYS389II repeat in order to treat them as two separate alleles. The output from the reduced median network analysis was used for input in constructing the median joining network in order reduce any large phylogenetic unrealistic reticulations in the network. Both mtDNA HVS-I sequencing and Y-chromosome STR networks were visualized using the computer program Network 4.0 (www.fluxus-engineering.com).

HETEROZYGOZITY Versus r_{ii}

In order to determine the effects of gene flow vs. genetic drift heterozygosity vs. distance from the centroid (r_{ii}) were plotted for classical genetic markers, Y-Chromosome STRs, and mtDNA HVS-I sequence data. This method uses the following equation:

$$r_{ij} = \frac{(p_i - \bar{p})^2}{\bar{p}(1 - \bar{p})} \qquad (7)$$

where r_{ii} is the distance from the centroid for a particular allele in the *i*th population, p_i is the frequency of the allele in the *i*th population, and $\bar{p}$ is mean frequency of the allele for all populations. Mean heterozygosity and r_{ii} values were calculated in the computer program ANTANA (Harpending and Rogers 1984), and regressed in Minitab ver. 12.0 (Minitab, Inc., State College, PA). Gene diversity measures were substituted for heterozygosity data for haplotypic data (Y-chromosome STRs and mtDNA HVS-I). According to Harpending and Ward (1982), there is a linear

relationship between heterozygosity and $r_{ii.}$. Therefore, any deviations from this relationship indicate that a particular population experiences either gene flow or genetic drift depending on its relationship to the theoretical regression line. Those populations that appear above the regression line and close to the X-Y axis are considered to be undergoing gene flow while those below the regression line and furthest from the X-Y axis are considered to be experiencing genetic drift.

POPULATION STRUCTURE

In order to quantify the effect of population substructure between the two major cultural areas (Mesoamerican and Isthmo-Colombian) in Central America, F_{ST} (Wright, 1931) values were calculated for classical genetic polymorphisms, mtDNA haplogroups, and Y-chromosome STRs with frequency data obtained from the literature. F_{ST} is a useful measure for examining the overall level of population subdivision relative to the total population and is calculated using the equation

$$F_{ST} = \frac{HT - HS}{HT} \quad (8)$$

where HT is equivalent to the total heterozygosity in the population and HS is the average heterozygosity across all subpopulations. Traditionally, values from 0 to 0.05 indicate little differentiation between populations, while 0.05 to 0.15 indicate moderate genetic differentiation, values from 0.15 to 0.25 indicate high differentiation, and values above 0.25 indicate great differentiation (Hartl and Clark 1997). Data for classical genetic polymorphisms were obtained for seven (20 alleles) blood group markers (ABO, MN, Rh, Fy, Duffy, Kidd, and Diego) in 18 Native

Central American populations. In order to test if F_{ST} values were statistically different between cultural areas, populations were subdivided into two groups: the first from lower Central America (Miskito, Sumo, Rama, Lenca, Pech, Boruca, Bribri, Cabecar, Teribe, Guaymi, Choco); and seven populations from Mesoamerica (Subtiva, Chorotega, Mam, Tolupan, Kekchi, Cackchi, and Choco). mtDNA data for four Native American haplogroups (A2, B2, C1, and D1) were obtained from study data and the from the literature. Y-chromosome frequency data were obtained for four common STR makers (DYS19, DYS390, DYS391, and DYS393) and divided into five populations from lower Central America (Teribe, Huetar Combined, Guaymi, Bribri, and Cabecar) and a Pipil Nahua Mesoamerican group from five communities (Conchagua, San Alejo, Panchimalco, Izalco, and Nueva Concepción) from El Salvador. All F_{ST} values from these frequency data were obtained using the *popstr* routine in the computer program ANTANA. This routine calculates F_{ST} for each allele and then obtains the average for each allele (Harpending and Rogers 1984).

In order to determine if population structuring could be determined between Mesoamerica and lower Central America using mtDNA HVS-I sequence data an AMOVA was conducted in Arlequin 3.11 (Excoffier *et al.* 2005). The populations were divided into three groups based on linguistic affinity (Mesoamerican, Chibchan, and Chocoan). The first group included two Mesoamerican populations (Maya and Chorotega), the 2nd group included seven populations from the Chibchan language family (Maléku, Ngöbé, Kuna, Quitirrisi Huetar, Zapáton Huetar, Guaymi, and Rama) and the third group included two Chocoan speaking groups (Emberá,

Wounan). AMOVA is analogous to a nested hierarchical analysis of variance (ANOVA) derived from a matrix of squared distances among all pairs of haplotypes. This method produces variance estimates and F_{ST} analogs that reflect the correlation of haplotypic diversity at different levels of population subdivision (Excoffier *et al.* 1992). The following equation is used for the calculation of total sum of squared deviations (SSD):

$$SSD_{(Total)} = \frac{1}{2N} \sum_{j=1}^{N} \sum_{k=1}^{N} \delta^2 jk \,, \tag{9}$$

where N equals the number of haplotypes and δ^2jk is the Euclidean distance between haplotypes *j* and *k*. This partitions haplotypes into three categories: SSD within populations, SSD between populations, and SSD among populations. Mean squared deviation is calculated by dividing the corresponding SSD with the appropriate degrees of freedom (Excoffier *et al.* 1992). Statistical significance is then determined through covariance components with different levels of genetic structure (Excoffier *et al.* 2005). Analysis was conducted in Arlequin 3.11 using pairwise F_{ST} values corrected with a Tamura and Nei (1993) model of substitution and γ-value of 0.26 (Meyer *et al.* 1999). Statistical significance was estimated by bootstrapping 1,000 times (Excoffier *et al.* 2005).

DIVERGENCE TIMES

Divergence times were calculated in Arlequin 3.11 (Excoffier *et al.* 2005) using the model of Reynolds *et al.* (1983) for mtDNA HVS-I data. This model uses

pairwise F_{ST} data and estimates divergence with the equation $D = -\log(1-F_{ST})$, which approximates dates as t/N where t refers to the divergence time and N refers to haploid population size. Dates were approximated using the mtDNA mutation rate of 33% per million years (Ward *et al.* 1991).

PHYLOGEOGRAPHICAL ANALYSIS

Spatial analysis of molecular variance (SAMOVA), Monmonier's maximum difference algorithm, and genetic landscapes were used to characterize patterns of genetic discontinuity in studied groups and neighboring indigenous populations inhabiting lower Central America. First, SAMOVA was applied in order to identify partitions of population groupings based on HVS-I sequence using the computer program SAMOVA 1.0 (Dupanloup *et al.* 2002). This analytical method is based on simulated annealing procedure and maximizes the proportion of total genetic variance due to differences among groups of populations (F_{CT}). The final number of groups (K) is based on the largest amount of variation shown value F_{CT} and it remains statistically significant. In contrast to other measures of genetic structure where groups are defined *a priori* based on geographical, linguistics, or ecological factors, SAMOVA determines groups classification solely on genetic information (Nasidze *et al.* 2004). Analyses were based on 100 annealing steps and examined maximum indicators of differentiation (F_{CT} values) when the program was instructed to categorize $K=2$ to $K=12$ partitions of population groupings. In addition to determine population groups based on genetic data, SAMOVA also putatively identifies genetic

barriers between groups. In this way SAMOVA is comparable to Monmonier's maximum difference algorithm.

Monmonier's maximum difference algorithm (Monmonier 1973) is a geographical regionalization procedure used to detect barriers to gene flow by identifying groups of contiguous distances along connectivity networks (Dupanloup *et al.* 2002; Manni *et al.* 2004; Miller 2005). This algorithm locates the edges ("barriers") associated with the highest rate of change in a given distance measure. This algorithm is then applied to connectivity network generated using Delaunay triangulation between geographical coordinates of studied populations (Brouns *et al.* 2003). After the triangulation is constructed, genetic barriers are identified by superimposing the highest rates of change of the connectivity network, which determines potential geographical barriers reflected by genetic data. Monmonier's algorithm was applied using the computer program Alleles in Space (Miller 2005) for mtDNA HVS-I sequences.

The final phylogeographic technique applied in this research was interpolated genetic landscapes (Miller 2005) for HVS-I sequence data. This procedure also begins with the construction of a Delaunay triangulation connectivity network and assigning genetic distances (Z_i, computed as the average pairwise distances between individuals from different populations) to geographical coordinates (X_i, Y_i) at n edges of the connectivity network. Next, inverse-weighted diverse interpolation (Watson 1992) is used to infer genetic distances between locations assigned on a uniformly

spaced grid of the geographic landscape. For each grid coordinate (x, y), a genetic distance (z) is inferred from each of i=1 to n genetic distances (Z_i) using the equation

$$Z = \frac{\sum_{i=1}^{n} w_i * Z_i}{\sum_{i=1}^{n} w_i} \qquad (10)$$

where w_i is a weighting function assigned to each Z_i and is inversely proportional to geographic distances between grid coordinates (x, y) and the actual geographic coordinates (X_i, Y_i), assigned to each of the n values of Z_i. The value for w_i is calculated using the equation:

$$w_i = \frac{\left\{\left[(X_i - x)^2 + (Y_i - y)^2\right]^{-\frac{a}{2}}\right\}}{1} \qquad (12)$$

and a is a weight value for distance. Larger values of a cause interpolated points to be closer together and lower values cause allow all points to equally influence interpolated values. This procedure results in a three-dimensional representation of geographic and genetic distances with higher regions of genetic differentiation observed above the x-y plane and regions of genetic similarity below the x-y plane. Interpolated genetic landscapes were generated in the program Alleles in Space (Miller 2005).

ADMIXTURE ESTIMATES

Admixture estimates for paternal and maternal lineages for the five study populations were algebraically calculated. This estimation assumed that mtDNA haplogroups (A2, B2, and D1) and Y-chromosome haplogroup Q3 represent the

Native American component and all other haplogroups are the result of European or African gene flow into these populations. Admixture estimates for mtDNA from four of the five study populations was unnecessary as only the Huetar Zapáton demonstrated any non-Native maternal admixture.

SUMMARY:

This chapter reviewed the field work, laboratory methods, and analytical techniques applied in this dissertation research. Fieldwork was conducted with the Rama Indians of the Caribbean coast of Nicaragua and resulted in the collection of biological samples from 75 individuals from two communities (Sumo Kat, Rama Cay) after obtaining informed consent. An additional 155 biological samples were obtained from four Costa Rican Native American groups (Chorotega, Maléku, Zapáton Huetar, and Guaymi) from Dr. Ramiro Barrantes at the University of Costa Rica. These biological samples were characterized for mtDNA RFLP and HVS-I sequences as well as Y-chromosome SNPs and STRs. Comparative classical genetic markers were obtained from the literature and analyzed. A number of analytical techniques were applied to these resulting molecular data in order to determine: 1) the impact of evolutionary forces operating on these populations, biological relationships among Chibchan-speaking groups and other indigenous populations in lower Central and northern South America; 2) if significant differences in populations structure were observable between Mesoamerica and lower Central America; 3) if genetic "barriers" were present in the region. These analytical techniques included: 1) diversity measures; 2) Wright's F_{ST} and Nei's G_{ST}; 3) neutral test statistics; 4)

AMOVA; 5) SAMOVA; 6) MDS; 7) Median –Joining phylogenetic networks; 8) Monmonier's maximum difference algorithm; and 9) interpolated genetic landscapes.

V. RESULTS

This chapter presents the results of analyses using mitochondrial DNA RFLP and HVS-I sequences, Y-chromosome SNPs and STRs, and classical genetic polymorphisms. Analytical methods applied to these data included: diversity measures; neutrality tests; phylogenetic networks; MDS plots; AMOVA, SAMOVA; heterozygosity versus r_{ii}; Monmonier's maximum difference algorithm; and interpolated genetic landscapes.

MITOCHONDRIAL DNA

RESTRICTION FRAGMENT LENGTH POLYMORPHISMS (RFLPs)

The RFLP analysis revealed that the five study populations contained three (A2, B2, D1) of the five major Native American haplogroups as defined by Schurr *et al.* (1990) and Torroni *et al.* (1992). Table 6 presents the mtDNA RFLP haplogroup data for the 230 individuals in the five study populations. Three of the populations (Rama, Maléku, and Guaymi) only contained haplogroups A2 and B2. Haplogroup A2 is the most common haplogroup in all of the study populations except for the Rama where B2 is at a higher frequency. This high frequency of haplogroup B2 distinguishes the Rama from the other three Chibchan populations. Haplogroup D1 was present in both the Zapáton Huetar and the Chorotega. Haplogroup D1 has previously been identified in the Quitirrisi Huetar (Santos *et al.* 1994). In addition, the Zapáton Huetar population exhibits African L2a (n = 2) and one European H1 (n =1) haplogroups. Haplogroup C1 was absent from all five of the study populations. These

RFLP data support earlier analysis that indicated that Central American Chibchan populations were primarily characterized by high frequencies of mtDNA haplogroup A and moderate frequencies of haplogroup B2 (Santos *et al.* 1994; Kolman *et al.* 1995, Batista *et al.* 1995; Kolman and Bermingham 1997). The one exception to this finding is the Rama who demonstrated an exceptionally high frequency (92%) of haplogroup B2.

Table 6: Percent of Native American RFLP haplogroups for five study populations:

Population	N	%A2	%B2	%C1	%D2	%other
Rama	75	8	92	0	0	0
Maléku	35	92	8	0	0	0
Zapáton Huetar	40	60	24	0	8	8
Guaymi	50	78	22	0	0	0
Chorotega	30	73	10	0	17	0

Table 7 compares mtDNA haplogroup data from the five study populations to 23 other Central (9), Caribbean (2) and South American (12) populations divided by geographic location and Chibchan ethnicity. These populations were also used for comparative data for the MDS plot (figure 16) used in this study and the geographic location of these populations is shown in figure 13. The Maléku have the highest frequency (92%) of haplogroup A2 in all the investigated populations. The high frequency of haplogroup A2 is shared with all the other Chibchan populations, with the exception of the Rama. This population is characterized by a high occurrence (92%) of haplogroup B2. The only other two comparative populations who share this high frequency of B2 are the Aché (90%) from Paraguay and the Xavante (84%) from Brazil. The absence of haplogroup C1 in Chibchan speakers from lower Central

Table 7: mtDNA Native American haplogroup frequencies in 28 South, Central, and Caribbean Native American mtDNA haplogroup frequencies.

Population*	N	%A	%B	%C	%D	%Other
CHIBCHAN						
Rama[1]	75	8	92	0	0	0
Maléku[1]	35	92	8	0	0	0
Zapáton Huetar[1]	40	60	24	0	8	8
Quitirrisi Huetar[2]	27	70	4	0	26	0
Guaymi[1]	50	78	22	0	0	0
Ngöbé[3]	46	67	33	0	0	0
Kuna[4]	79	77	23	0	0	0
Kogi[5]	48	65	0	35	0	0
Arsario[5]	50	68	0	32	0	0
Ijka[5]	40	90	3	7	0	0
CENTRAL AMERICAN						
Chorotega[1]	30	73	10	0	17	0
K'iche Maya[6]	27	52	22	15	7	4
Emberá[7]	44	23	52	25	0	0
Wounan[7]	31	29	19	48	3	0
SOUTH AMERICAN						
Wayuú[5]	46	37	26	35		2
Cayapa[8]	120	29	40	9	22	0
Shamatari[9]	151	0	56	32	12	0
Yanamamõ[10]	129	2	9	26	31	5
Aché[11]	64	10	90	0	0	0
Xavante[12]	25	16	84	0	0	0
Gavaio[12]	27	15	15	0	70	0
Zoro[12]	30	20	7	13	60	0
Pehuenche pooled)[13,14]	205	2	8	40	50	0
Mapuche (pooled)[14, 15, 16]	208	5	20	33	39	3
Yahgan[14]	21	0	0	48	52	0
CARIBBEAN						
Ciboney[17]	15	7	0	60	33	0
Tainos[18]	24	0	0	75	25	0

***References**: (1) This Study; (2) Santos *et al.* 1994; (3) Kolman *et al.* 1995; (4) Batista *et al.* 1995; (5) Melton *et al.* 2007; (6) Boles *et al.* 1995; (7) Kolman and Bermingham 1997; (8) Rickards *et al.* 1999; (9) Williams *et al.* 2002; (10) Meriwether *et al.* 2000; (11) Schmidt *et al.* 2005; (12) Ward *et al.* 1996; (13) Meriwether *et al.* 1995; (14) Moraga *et al.* 2000; (15) Ginther *et al.* 1993; (16) Baillet *et al.* 1994; (17) Lalueza-Fox *et al.* 2003; (18) Lalueza-Fox *et al.* 2001.

Figure 13: Geographic location of 28 Native Central and South American populations used to investigate mtDNA HVS-I sequence variation (16050-16383) and used to create MDS plot shown in figure 16. *K'iche Maya (Guatemala), Tainos, and Ciboney (Cuba) geographic location shown in figure 12*.

America is also confirmed by these analyses (Melton *et al.* 2007). The Chorotega, an Oto-Manguean speaking population from Costa Rica and Nicaragua are also characterized by high frequencies of haplogroup A2, low frequencies of B2, and moderate frequencies of D1. The presence of haplogroup D1, which is largely absent from northeastern South American populations is found among the Chibchan speaking Huetar, the Chocoan Wounan, and the K'iche Maya. In addition, this haplogroup occur at high frequencies in skeletal populations (Tainos, Ciboney) from Cuba. The presence of this haplogroup in populations neighboring Chibchan groups suggests that the majority of Chibchan speakers may have lost haplogroup D1 through genetic drift or may have been introduced into the Huetar by neighboring Mesoamerican populations.

Based on the four hypotheses for the divergence of Chibchan populations the resulting mtDNA haplogroup data may be used to interpret the geographic location where this event occurred. The first hypothesis suggests that these populations diverged in South America and then migrated into Central America from either the Amazon or Caribbean region. Given the absence of haplogroup C1 and low frequency of D1 in Central American Chibchan populations, this scenario seems unlikely. The two Caribbean populations (Ciboney and Tainos) are characterized by high frequencies of C1 and D1, and the eastern South American groups also contain varying frequencies of these two haplogroups. The absences of these two haplogroups in Central American Chibchan populations indicates a reduced genetic diversity in these groups compared to other populations. Mesoamerican groups also contain three

or four of the Native American mtDNA haplogroups, once again indicating that genetic drift may be operating on Chibchan populations. The two-wave model suggests a combination of Mesoamerican and South American influence in Chibchan populations, but if either of these two earlier models do not address the geographic location question, this explanation may be rejected as well. This leaves the endogenous development of Chibchan populations as the only available alternative. However, the presence of two haplogroups (A2 and B2) does not directly answer the question if Chibchan groups diverged from an earlier hunter/gatherer population in the region as little ancient DNA data are available for the region but evidence from mtDNA haplotype data may provide more evidence regarding this hypothesis.

HVS-I SEQUENCES

The HVS-I sequencing results for a subset 131 (30 Rama; 14 Maléku; 24 Zapáton Huetar; 39 Guaymi; and 24 Chorotega) individual sequences are shown in table 8. A total of 18 different haplotypes characterized by 22 different nucleotide variant sites were observed in these populations. All 22 of the observed DNA sequence variants were transitions and none were transversions. The Rama and Zapáton Huetar demonstrated the most haplotypes (7), followed by the Chorotega and Guaymi (6). The Maléku contained the least number of haplotypes (3) of any of the study populations. The haplotype CA8 is the most common mtDNA lineage, present in 33 individuals and all five study populations. This haplotype corresponds to the founding Native American mtDNA haplogroup B2 and is common throughout the

Americas. The second most frequent haplotypes CA2 individuals is shared by 26 individuals and four (Rama, Guaymi, Zapáton Huetar, and Chorotega) of the five study populations. Haplotypes CA1 (19 individuals, CA4 (14 individuals), and CA5 (13 individuals) are also common in these populations. Five (CA1, CA2, CA4, CA6, CA8) of the 18 observed haplotypes are shared among populations. Haplotype CA5 is characterized by a T-C transition at 16189 and is found only in the Chorotega. However, this haplotype is known to occur in other South American Chibchan speaking populations (Melton *et al.* 2007). The remaining 12 haplotypes occur in one and three individuals. The Chorotega share three haplotypes (CA2, CA6, and CA8) with the four Chibchan speaking populations.

Table 8: mtDNA HVS-I sequences for five study population samples. Only variable sites are shown

Haplogroup*	Haplotype	16111	16124	16150	16187	16189	16209	16217	16223	16242	16231	16256	16264	16270	16274	16278	16290	16291	16319	16325	16327	16354	16360	16362	Rama	Malekú	Zapáton Huetar	Guaymi	Chorotega	Total #
	CRS**	**C**	**T**	**C**	**C**	**T**	**T**	**T**	**C**	**C**	**T**	**C**	**C**	**C**	**G**	**C**	**C**	**C**	**G**	**T**	**C**	**C**	**C**	**T**	-	-	-	-	-	
	CA1	T	-	-	-	-	-	-	T	-	-	-	-	-	-	-	T	-	A	-	-	-	-	C	1	12	-	6	-	**19**
A2	**CA2**	T	-	-	-	-	-	-	T	-	-	-	-	-	-	-	T	-	A	-	-	-	T	C	1	-	9	11	5	**26**
	CA3	T	C	-	-	-	-	-	T	-	-	-	-	-	-	-	T	-	A	-	-	-	-	C	-	-	-	1	-	**1**
	CA4	T	-	-	T	-	-	-	T	-	-	-	-	-	-	-	T	-	A	-	-	-	-	C	3	-	-	11	-	**14**
	CA5	T	-	-	-	C	-	-	T	-	-	-	-	-	-	-	T	-	A	-	-	-	-	C	-	-	-	-	13	**13**
	CA6	T	-	-	-	-	C	-	T	-	-	-	-	-	-	-	T	-	A	-	-	-	-	C	-	1		-	1	**2**
	CA7	T	-	T	-	-	-	-	T	-	-	-	-	-	-	-	T	-	A	-	-	-	-	C	-	-	4	-	-	**4**
B4	**CA8**	-	-	-	-	C	-	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	19	1	5	8	1	**33**
	CA9	-	-	-	-	C	-	C	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	2	-	-	-	-	**2**
	CA10	-	-	-	-	C	-	C	-	T	-	-	-	-	-	-	-	T	-	-	-	-	-	-	1	-	-	-	-	**1**
	CA11	-	-	-	-	C	-	C	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3	-	-	-	-	**3**
	CA12	-	-	-	-	C	-	C	-	-	-	-	-	-	-	-	-	-	-	C	-	-	-	-	-	-	-	2	-	**2**
	CA13	-	-	-	-	C	-	C	-	-	-	-	-	-	-	-	-	-	-	C	-	-	-	C	-	-	2	-	-	**2**
	CA14	-	-		-	C	-	-	T	-	-	-	-	-	-	-	-	-	-	-	T	-	-	C	-	-	1	-	-	**1**
D1	**CA15**	-	-		-	-	-	-	T	-	-	-	-	-	A	-	-	-	-	C	-	-	-	C	-	-	-	-	3	**3**
	CA16	-	-		-	-	-	-	-	-	C	-		-	A	-	-	-	-	C	-	-	-	C-	-	-		-	1	**1**
L2a	**CA17**	-	-	-	-	-	-	-	T	-	-	-	T	-	-	T	-	-	-	-	-	-	-	-	-	-	2	-	-	**2**
H1	**CA18**	-	-	-	-	-	-	-	-	-	-	T	-	T	-	-	-	-	-	-	-	T	-	-	-	-	1	-	-	**1**
Total																									**30**	**14**	**24**	**39**	**24**	**131**

*= Native American haplogroup nomenclature (Achilli *et al.* 2008), European haplogroup nomenclature (Loogvali *et al.* 2004), African haplogroup nomenclature (Salas *et al.* 2002). ** Revised Cambridge Reference Sequence (Andrews *et al.* 1999, Anderson *et al.*, 1981).

HVS-I sequences that belonged to mtDNA haplogroups A2 and B2 demonstrate the most variability and each contain seven haplotypes. All seven

lineages that belonged to A contain the 16111 C→T transition, that along with 16223T, 16290T, 16319A, and 16362C, which places them in the A2 haplogroup, found primarily in Native American populations. All seven B haplogroups contain the 16217 T→C transition, which places them in the B2 subclade that is common in Native American and East Asian population (Achilli *et al.* 2008). Four of the five study populations were characterized exclusively by Native American haplotypes. The only study population that demonstrated non-European maternal lineages was the Zapáton Huetar, which contain two African L2a (Salas *et al.* 2002) and one European H1 haplogroup (Loogvali *et al.* 2004). It is unlikely that the presence of HI is due to laboratory contamination because the author also belongs to European haplogroup H but his mtDNA HVS-I sequence is three base pairs different then the observed Zapáton Huetar sequence. This indicates that this HVS-I sequence was not the result of contamination during the PCR process.

DIVERSITY AND NEUTRALITY MEASURES

Haplotype and nucleotide diversity values were calculated for all five study populations as well as for an additional nine Central and South American groups (table 9). Low haplotype diversity values are indicative of low genetic diversity as they represent polymorphisms shared by several individuals, whereas high haplotype diversity values suggest genetic differentiation among individuals within a population. Of the five study populations, the Maléku demonstrate the lowest haplotype diversity (0.274) whereas the Guaymi (0.812) have the highest haplotype

diversity. This is also the highest haplotype diversity values for all the investigated Chibchan speaking groups. Among comparative populations, the Ijka have the lowest haplotype diversity (*h*) value (0.185), which is the lowest known mtDNA diversity value for any South American population (Melton *et al.* 2007). The Emberá hace the highest *h* value (0.942) of any of the comparative populations. All groups demonstrate low nucleotide diversity values ranging from 0.001 in the Guaymi to 0.02 in the Wounan.

Table 9: Diversity values in Central American populations based on mtDNA HVS-I sequence data.

Population*	N	# ht	# variant sites	*h*	Π	D	Fu's Fs
CHIBCHAN							
Ngöbe[2]	46	7	12	0.763	0.012	1.68	3.39
Kuna[3]	63	7	10	0.592	0.009	1.52	2.77
Quitirrisi Huetar[4]	29	7	12	0.709	0.010	0.41	1.18
Zapáton Huetar[7]	25	7	17	0.806	0.017	0.35	2.3
Rama[7]	31	7	11	0.591	0.008	-0.49	0.06
Guaymi[7]	39	7	12	0.812	0.001	1.01	2.33
Maléku[7]	14	3	9	0.274	0.004	-1.93*	1.63
Ijka[8]	31	3	12	0.185	0.004	-1.58*	2.96
Kogi[8]	21	3	10	0.523	0.009	0.58	5.30
Arsario[8]	28	4	10	0.725	0.012	1.98	5.74
MESOAMERICAN							
Chorotega[7]	24	6	14	0.670	0.01	-0.585	1.42
Maya[5,6]	34	18	27	0.930	0.017	-0.578	-4.90*
CHOCOAN							
Emberá[1]	44	20	23	0.942	0.018	0.46	-4.38
Wounan[1]	31	14	29	0.912	0.020	-0.27	-1.01

#ht=number of haplotypes; *h*=haplotype diversity; ***π***=nucleotide diversity; Tj's D=Tajima's D. 1=Kolman and Bermingham 1997; 2=Kolman *et al.* 1995; 3=Batista *et al.* 1995; 4=Santos *et al.* 1994; 5=Boles *et al.* 1995; 6=Torroni *et al.* 1993; 7=This Study; 8=Melton *et al.* 2007), * = $p < 0.05$, **= $p < 0.01$.

Two measures of selective neutrality (Tajima's D and Fu's Fs) were calculated in order to detect any deviations from the neutral model of evolution (table 9). Both the Maléku (-1.93) and the Ijka (-1.58) exhibit statistically significant Tajima's D values at the $p< 0.05$ level. However, these significant values are the product of low haplotypes diversity values, which may be inflating the overall Tajima's D value to make it appear as a population expansion. The only population to possess a significant Fu's Fs value is the K'iche Maya (-4.90). A large negative Fs value is indicative of population expansion, whereas positive Fs values suggest population differentiation. All Chibchan populations have positive Fs values; the Rama exhibit the lowest (0.06) and the Arsario the highest, (5.74). The Chorotega were the only non-Chibchan population to exhibit a positive Fs value (1.42).

MEDIAN JOINING NETWORKS

Two different median joining networks were constructed from mtDNA HVS-I data. The first median network (figure 14) combines four Chibchan (Rama, Zapáton Huetar, Guaymi, Maléku) study populations with an additional six Chibchan (Arsario, Kogi, Ijka, Ngöbé, Quitirrisi Huetar, Kuna) groups from Central and South America into a single network of all four major mtDNA haplogroups (A2, B2, C1, D1). This network was rooted on the African L3 haplogroup and connected through RFLP

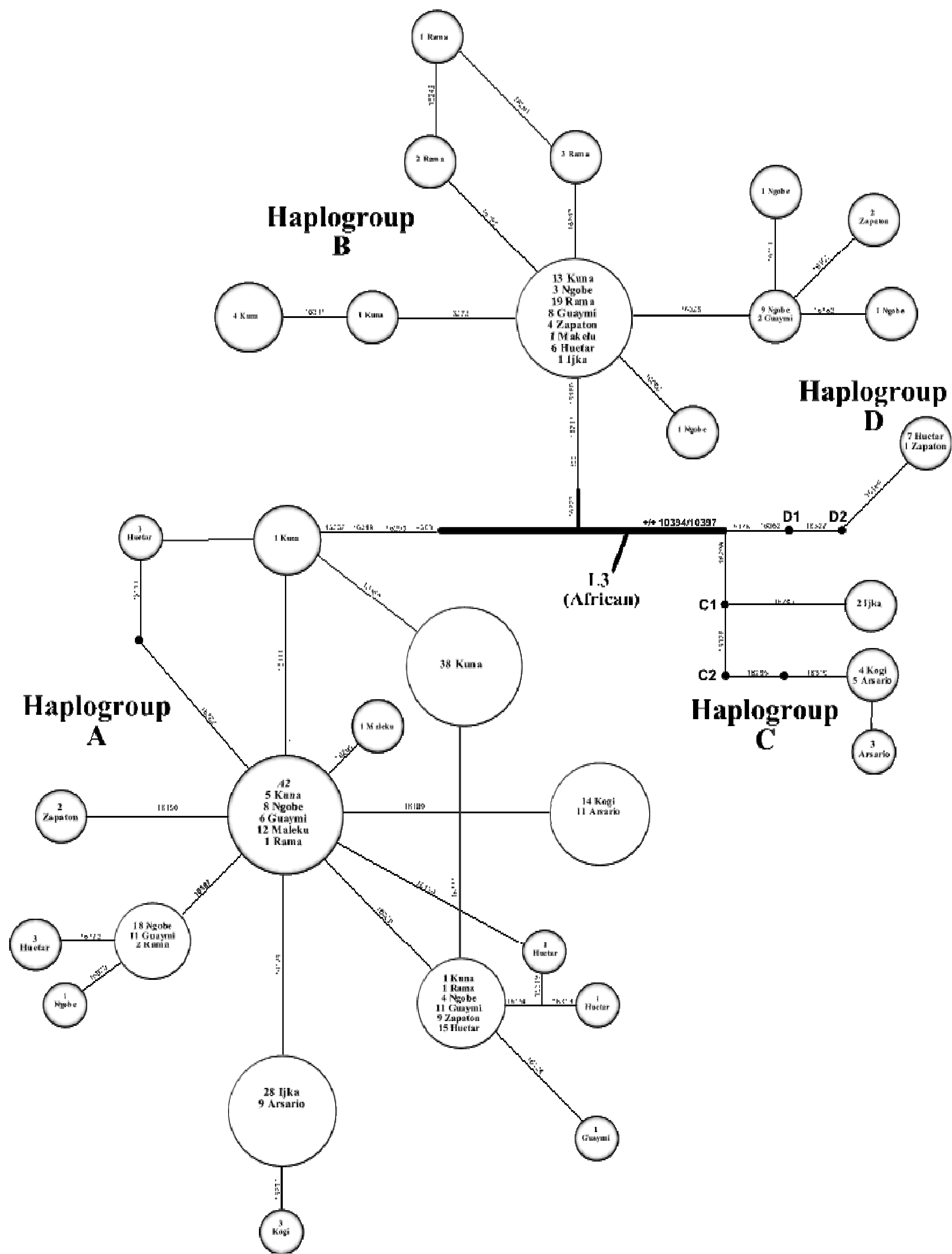

Figure 14: Median Joining phylogenetic network (Bandelt *et al.* 1999) of mtDNA HVS-I sequence date (16050-16383) for 10 Chibchan populations (Arsario, Ngöbé, Ijka, Kuna, Kogi, Huetar (Zapáton and Quitirrisi), Guaymi, Maléku, and Rama). Rooted on African L3 haplogroup.

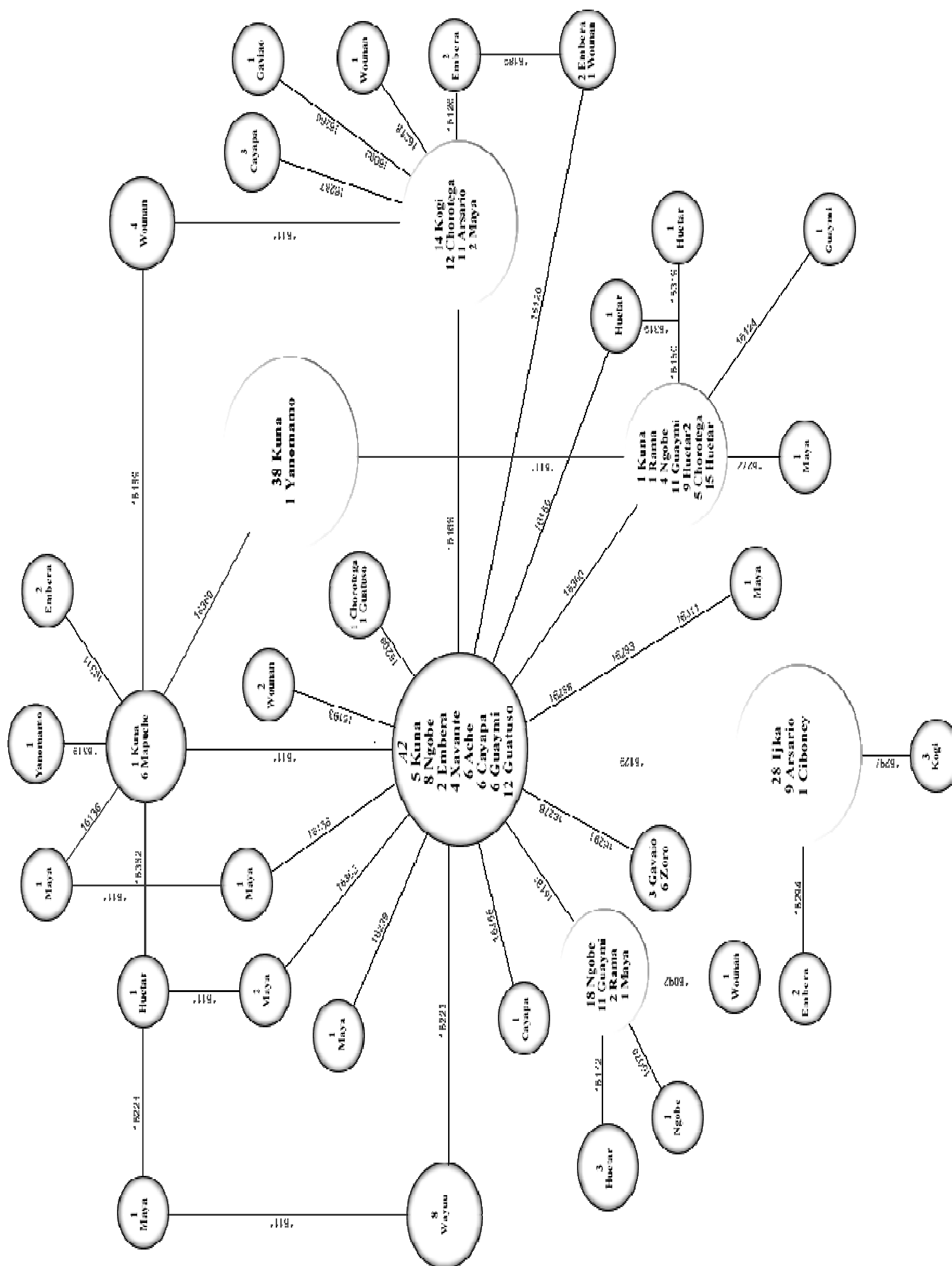

Figure 14: Median Joining phylogenetic network (Bandelt *et al.* 1999) of mtDNA HVS-I sequence date (16050-16383) for mtDNA haplogroup A. reticulations are highlighted in red. Centered on A2 node (includes diagnostic nucleotide sites 16111, 16223, 16290, 16319, 16362).

defining DNA sequence variants. The second network consists of the ten Chibchan populations and 12 comparative Central and South American populations (Emberá, K'iche Maya, Wounan, Chorotega, Wayuú, Aché, Xavante, Cayapa, Gavaio, Zoro, Mapuche, Yanamamo) for haplogroup A2. This network is rooted on the Native American A2 haplogroup that is defined by the mtDNA HVS-I sequence variants (16111, 16223, 16290, 16319, 16362).

The median joining network pictured in figure 14 provides a graphical representation of the 30 mtDNA haplotypes found in 13 Chibchan populations. The majority of Chibchan mtDNA haplotypes (15) are found in haplogroup A2, followed by haplogroups B2 (11), C1 (3), and D1 (1). Only the two Huetar populations exhibit haplogroup D1 and only the South American Chibchan groups (Kogi, Arsario, Ijka) contain haplogroup C1. Haplogroup A2 exhibits four satellite clusters (16189, 16129, 16360, and 16187). The satellite node, 16360, is shared among six of the seven Central American Chibchan speaking populations. Two of these nodes (16189, 16129) are located only in South American Chibchan speakers and are not shared with Central American Chibchan speaking populations. These results suggest long-term separation of Chibchan-speaking population inhabiting lower Central and northern South America.

The results for the haplogroup A2 median joining network are shown in figure 15. This network contains six satellite clusters (16189, 16129, 16360, 16187, -16223, and -16111) and was coalescent dated to 23,582 ± 7,769 ya, using the ρ-statistic (ρ=

1.16 ± 0.38) (Forster *et al.*, 1996; Saillard *et al.*, 2000). This date is consistent with other estimates for the origins of haplogroup A2 in the Americas (Schurr and Sherry 2004; Tamm *et al.* 2007; Achilli *et al.* 2008). Four of the six satellite nodes are composed of Chibchan speaking groups but are regionally differentiated between South and Central American Chibchan speaking groups. The Chorotega are shown in both the major South (16189C) and Central American Chibchan (16360C) clades as are the K'iche Maya, that are present in the 16189C and the 16187 Chibchan mtDNA lineages. These shared haplotypes indicate a shared maternal genetic history in these groups. This shared maternal genetic history suggests that at some point in the past, these populations shared a common female ancestor(s). This is because mtDNA is passed from the mother to all of her offspring but only daughters pass that on to their offspring, making mtDNA a useful tool for investigating the maternal genetic history of a population.

The satellite node 16189 demonstrates that genetic variation coalesced at 6,985 ± 3,055 ya (ρ = 0.34 ± 0.15) in these populations. The other three satellite nodes containing Chibchan populations coalesce at dates ranging from 10,967 ± 8,657 ya (16360; ρ = 0.54 ± 0.42), 8,221 ± 6,829 ya (16187; ρ = 0.40 ± 0.33), and 2,346 ±1,692 ya (16129, ρ = 0.11 ± 0.08). All these coalescent dates are consistent with previous estimates regarding the origins of Chibchan maternal genetic history (Kolman *et al.* 1995; Batista *et al.* 1995; Kolman and Bermingham 1997; Melton *et al.* 2007).

MULTIDIMENSIONAL SCALING PLOT (MDS)

In order to ascertain the relationship of the five study populations to 25 other groups inhabiting Central and South America a multidimensional scaling plot (MDS) was designed with mtDNA HVS-I sequence data. The results of this analysis are shown in figure 16. The stress value for the plot is moderate (0.16) and the goodness of fit (0.97) is high. The upper bound for a stress level for 30 data points displayed in two dimensions is 0.38 indicating that this represents a non-random visualization of these data (Sturrock and Roche 2000). There are three distinct clusters of populations that can be determined from the plot. The first cluster contains the majority of Chibchan populations that are grouped together in the lower right quadrant of the plot area. This cluster also demonstrates a genetic relationship among two Central American populations (K'iche Maya and Chorotega) and the Cayapa, an indigenous population from Ecuador. This cluster is composed of populations characterized by high frequencies of haplogroup A2 (data not shown). The second cluster contains the only Chibchan group, Rama, not found in the first group that exhibits a relationship with the Xavante and Aché, from Brazil and Paraguay. This second grouping is characterized by populations that demonstrate high frequencies of mtDNA haplogroup B2. A final group is found in the lower left quadrant of the graph and contains the Tainos and Ciboney, two skeletal populations from Cuba. This group is exhibits the absence of mtDNA haplogroups A2 and B2 and high frequencies of C1 and D1. The center of the graph consists of a mixture of populations from throughout South American and also includes two Central American populations, the Emberá and

Wounan. These populations are known to have migrated to Panama from Colombia after European contact.

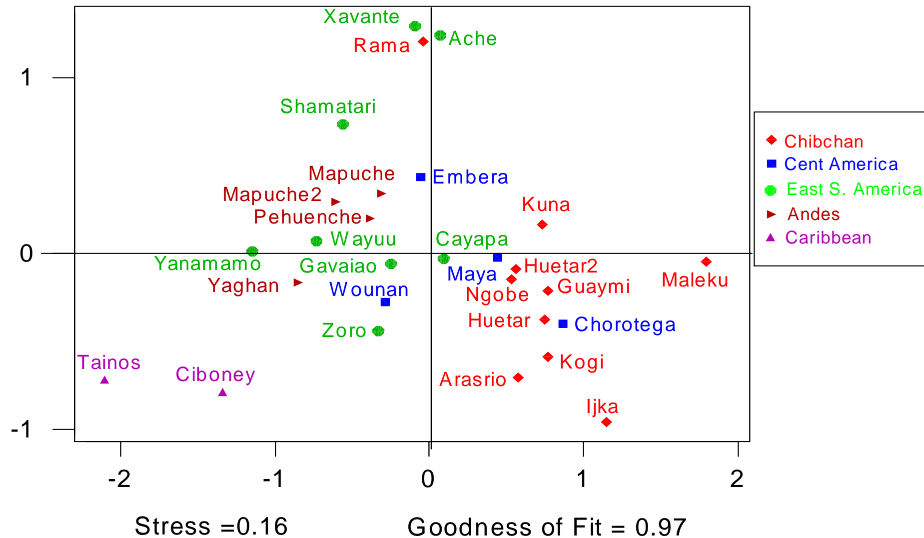

Figure 16: MDS plot of mtDNA HVS-I sequence data (bp 16050-16383) for 28 Native Central and South American populations (Figure 13). Plot was created using DA genetic distances (Nei, 1987) using the assumption of Tamura and Nei (1993) model of substitution and a γ level of 0.26 (Meyer *et al.* 1999).

GENE DIVERSITY Versus. r_{ii}

The plot for gene diversity versus r_{ii} using mtDNA HVS-I sequence data is shown in figure 17. There are seven populations (Ngöbé, Quitirrisi Huetar, Chorotega, Rama, Kuna, Arsario, and Kogi) below the theoretical regression indicating greater genetic differentiation in these populations. Three of these populations (Rama, Kuna, and Kogi) are also distant along the r_{ii} axis indicating

maternal genetic isolation than the other four groups who are close to the centroid (gene diversity/r_{ii} axis). A total of five (Guaymi, Zapáton Huetar, K'iche Maya, Wounan, and Emberá) groups are found above the axis. The two Chocoan populations (Wounan and Emberá) cluster with the K'iche Maya. All three of these groups demonstrate higher haplotype (gene) diversity values (average = 0.87) than Chibchan populations (average = 0.59).

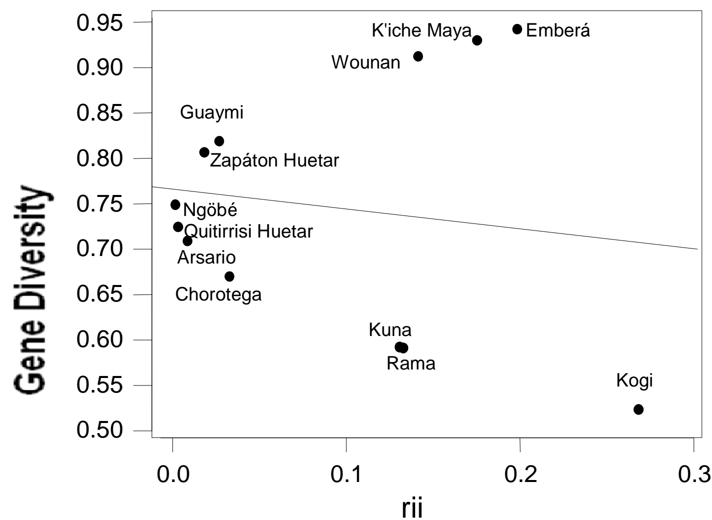

Figure 17: Regression plot of gene diversity values as a proxy for heterozygosity and distances from the centroid (r_{ii}) for 12 Native Central American populations using mtDNA HVS-I sequence data (16050-16383). Outliers Maléku and Ijka are not shown.

PHYLOGEORAPHIC ANALYSIS

Phylogeographic analysis with three different methods were conducted using mtDNA HVS-I sequence data in order to investigate the presence or absence of

genetic discontinuities among Central American populations. These analyses included spatial analysis of molecular variance (SAMOVA), interpolated genetic landscape analysis, and Monmonier's maximum likelihood difference algorithm. Five study populations were included these analyses along with three Central American Chibchan (Ngöbé, Kuna, and Quitirrisi Huetar), three South American Chibchan (Kogi, Ijka, and Arsario), two Chocoan (Emberá, Wounan), one Arawak (Wayuú), and one Mayan (K'iche) group. The geographic locations of these populations are shown in figure 13

SAMOVA

SAMOVA analysis was run for *k* (number of groups) from 2 to 13 to establish if populations in the region could be grouped based on their geographic location. The most parsimonious SAMOVA was found for $k = 4$ ($F_{CT} = 0.148$, p-value < 0.00001), which demonstrates the highest percentage of variation explained among groups (21.76%) and remained statistically significant. The results for this analysis are shown in a MDS plot (figure 18). This plot has a high stress value 0.21 (upper bound .228) and a moderate goodness of fit (0.92). This analysis separates Native Central Americans from non-Chibchan South American populations and splits two Votic Chibchan-speakers (Rama and Maléku) from other Central American groups. The Ijka appear closest to non-Chibchan South American groups but are grouped consistently with Central American populations until $k = 5$, where they are included

in their own group. These results confirm a strong maternal genetic relationship among Mesoamerican and Chibchan groups based on mtDNA HVS-I sequence data.

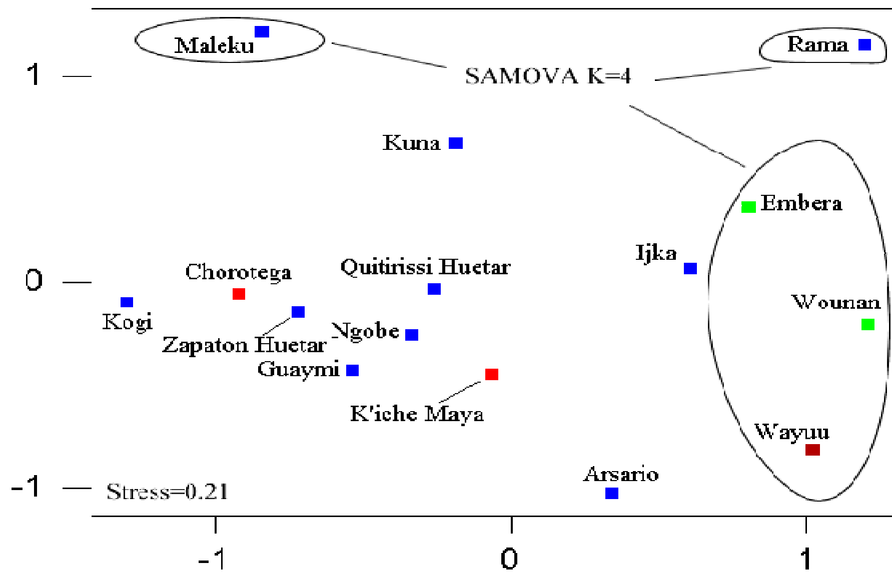

Figure 18: Combined MDS and SAMOVA Plot of genetic discontinuity for Central and northern South American populations. Chibchan population = blue square; Central American = red square; Chocoan = green squares; Wayuú = dark red square.

Monmonier's Algorithm

The results of the Monmonier's Algorithm are presented in figure 19. This analysis demonstrates a single genetic discontinuity that separates the K'iche Maya, and three Votic-speaking (Zapáton Huetar, Rama, and Maléku) Chibchan populations from the other groups investigated. In addition, the Chorotega and Quitirrisi Huetar are border the genetic discontinuity. Based on geography, the most likely candidate for a discontinuity in the region is Lake Nicaragua and this discontinuity may be attributed to rising sea occurred in the region approximately 8,000 YBP (Urquhart

1997). This date also overlaps with previously discussed coalescent genetic dates for Chibchan populations in the region.

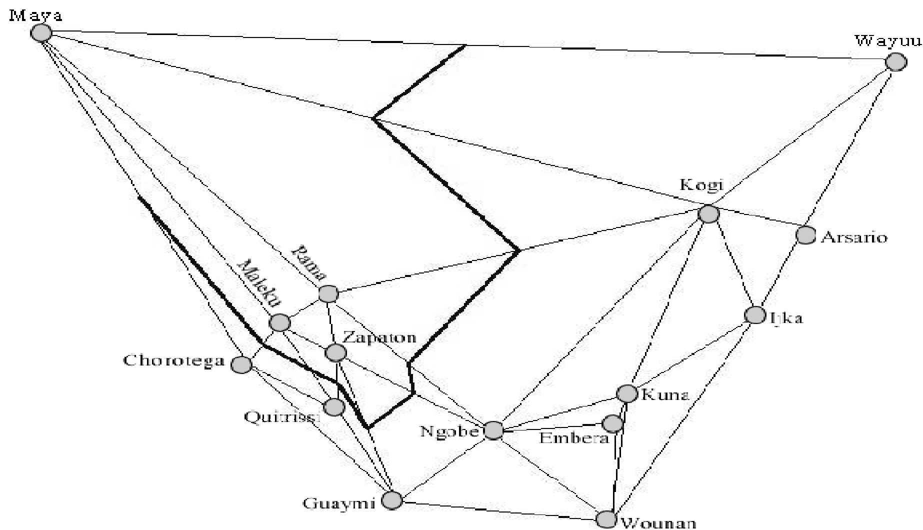

Figure 19: Delaunay triangulation using Monmonier's Algorithm for 15 Central and South American populations. Genetic barrier is shown as a bold line

Interpolated Genetic Landscape

The interpolated genetic landscape among 14 Central and South American populations is presented in figure 20. The XY axis represents geographic coordinates and the Z-axis represent pairwise genetic distances between populations. Peaks and depressions below the XY plane signify more genetic similarities. Greater genetic differentiation are seen among populations inhabiting the northern regions of lower Central America. The highest peaks are between the Rama and the Maléku populations. Based on this figure, more genetic similarities are observed among indigenous populations inhabiting Colombia and Panamá. This landscape confirms

the presence of a significant genetic discontinuity in northern Costa Rica or southern Nicaragua that may correspond to Lake Nicaragua.

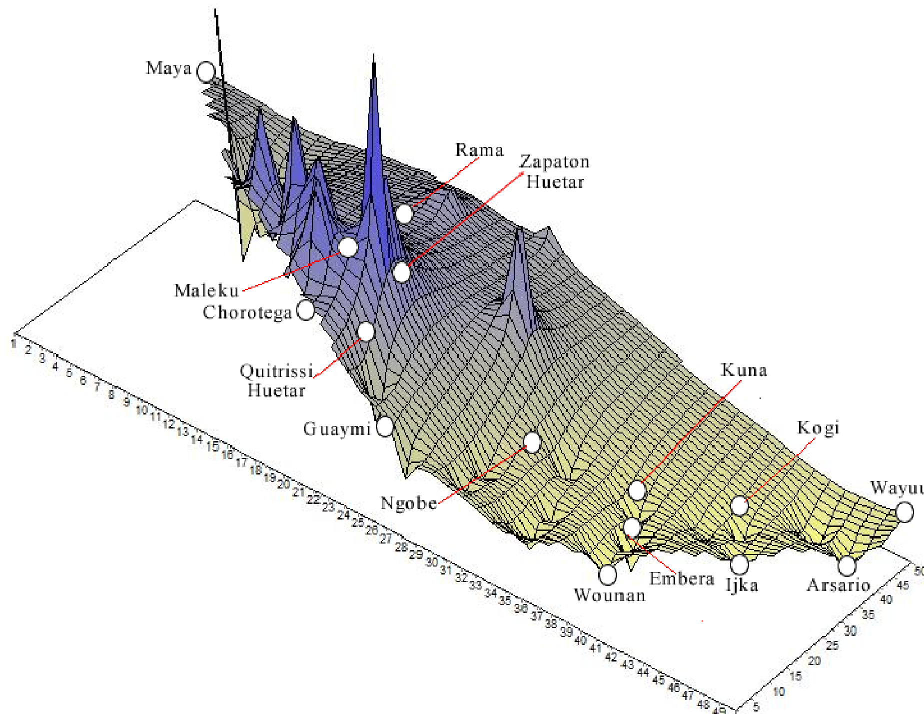

Figure 20: Interpolated genetic landscape for 14 indigenous Central and South American populations. The XY axis represents geographic coordinates and the Z-axis represents pairwise genetic distances among populations. Peaks above the XY plane represent greater genetic differentiation among populations and peaks below the XY plane represent more genetic similarities.

CHIBCHAN DIVERGENCE TIMES

Divergence times in ten Chibchan populations based on mtDNA HVS-I sequence variation are shown in table 10. The divergence estimate between the Guaymi and the Ngöbé can be used as a control. The Guaymi are a recent splinter group of the Ngöbé that migrated to Costa Rica in the 1940s (Barrantes et al. 1990; Herlihy 1997). The divergence estimate of approximately 30 years would suggest that this is an accurate measure of divergence time among populations. The average

divergence time between the Rama and all other Chibchan populations is 7,178 ya, which places them at the earlier boundary of when Chibchan populations are estimated to have diverged. If the three Colombian Chibchan populations and the Maléku are removed, the average divergence time of the Rama with other Chibchan populations is 5,422 ya, which is approximately the time frame that linguists believe that Chibchan languages differentiated into distinct sublanguage families (Constenla, 1991). The divergence time between the Maléku and the Rama is estimated at 10,531 ya, suggesting that they did not originate from the same source population as indicated by historians and linguists.

Table 10: mtDNA divergence times for 10 Chibchan populations using Reynolds *et al.* (1983) method.

	NG	KU	GY	HQ	HZ	ML	RA	AR	IJ	KG
NG	0	1582	30	883	395	2964	4310	1286	3110	1110
KU	0.174	0	1331	1266	589	5135	5326	2799	4904	2920
GY	0.003	0.1467	0	492	200.	3029	5690	1251	2809	1003
HQ	0.0974	0.1396	0.0542	0	410.	4931	7553	1510.	3751	1347
HZ	0.0435	0.0649	0.0220	0.0452	0	3102	4249	1453	3292	1366
ML	0.3267	0.5660	0.3339	0.5435	0.3419	0	10,531.	4544	7994	5226
RA	0.4751	0.5871	0.6272	0.8325	0.4684	1.16075	0	7240	11,927	7795
AR	0.1417	0.3085	0.1379	0.1664	0.1602	0.50087	0.79808	0	2011	101
IJ	0.3428	0.5406	0.3096	0.4134	0.3629	0.8812	1.31462	0.22168	0	3096
KG	0.1223	0.3218	0.1105	0.1484	0.1506	0.57606	0.8592	0.01121	0.3412	0

Divergence estimates are shown in the lower matrix. Time estimates are shown in the upper half of the matrix. Abbreviations are: NG= Ngöbé; KU=Kuna; GY=Guaymi; HQ= Quitirrisi Huetar; HZ= Zapáton Huetar; ML=Maléku; RA=Rama; AR=Arsario; IJ= Ijka; KG=Kogi.

Y-CHROMOSOME

Y-chromosome SNP and STR data were analyzed for the five study populations and comparative data from an additional 18 South and Central American groups compiled from the literature. These comparative data included: five Pipil

Nahua communities from El Salvador (Conchagua, San Alejo, Panchimalco, Izalco, and Nueva Concepción) (Lovo-Gomez *et al.* 2006), five Chibchan (Bribri, Huetar, Cabecar, Osa-Guaymi, and Teribe) (Ruiz-Navarez *et al.* 2005), and eight South American (Xavante, WaiWai, Ticuna, Gavaio, Karitiana, Cayapa, Tayacaja, and Arequipa) populations (Tarazona-Santos *et al.* 2001) (figure 21).

SINGLE NUCLEOTIDE POLYMORPHISMS (SNPs)

The results of the Y-chromosome SNP analysis for the five study populations and ten other Central American groups are presented in table 11. The five study populations are characterized by Native American haplogroups Q and Q3, Eurasian haplogroups R1b, G2a, and I1b, Euro-Afro Asian haplogroup E1b, and "other" undetermined haplogroups. Among the study populations, the Native American haplogroup Q3 occurs in all five groups with the highest frequency in the Zapáton Huetar (84.6%) and the lowest in the Chorotega (34.7%). In addition, the major western European haplogroup, R1b, is found in four (Rama, Abrojo Guaymi, Zapáton Huetar, and Chorotega) of the five study populations. The Rama are characterized by three Y-chromosome haplogroups Q3 (50%), R1b (40%) and G2a (10%). The Zapáton Huetar are characterized by two haplogroups Q3 (84.5%) and R1b (15.5%). The Abrojo Guaymi are characterized by haplogroups Q3 (68.4%), R1b (10.5), as well as unknown haplogroups (21%). The Maléku are characterized by haplogroups Q3 (46.1%), E3 (46.1%), and I1b (7.7%). The Chorotega demonstrate

Figure 21: Comparative Populations used for Y-chromosome analysis. Five Mesoamerican populations from El Salvador: Conchagua; San Alejo; Panchimalco; Izalco; and Nueva Concepción shown in figure 12.

the most Y-chromosome SNP diversity and are characterized by haplogroups Q (15%), Q3 (34.7%), R1b (43.4%), E3 (4%) and "other" (13%). Among comparative populations, the Bribri display the highest frequency of Native American haplogroup Q3 (92.5%) and the Mesoamerican Izalco of El Salvador have the lowest (27.2%). The Izalco also contain the highest frequency of European haplogroup R1b (45.4%).

Table 11: Y-Chromosome SNPs in study and comparative populations

Population	N	Q	Q3	R1b	E3	G2a	I1b	other
CHIBCHAN								
Rama[1]	20	0.000	0.500	0.400	0.000	0.100	0.000	0.000
Huetar - Zapáton[1]	13	0.000	0.845	0.155	0.000	0.000	0.000	0.000
Huetar - Comb.[2]	28	0.000	0.642	0.000	0.000	0.000	0.000	0.358
Guaymi - Abrojo[1]	19	0.000	0.684	0.105	0.000	0.000	0.000	0.210
Guaymi – Oso[2]	8	0.000	0.625	0.000	0.000	0.000	0.000	0.375
Maléku[1]	13	0.000	0.461	0.000	0.461	0.000	0.078	0.000
Bribri[2]	14	0.000	0.928	0.000	0.000	0.000	0.000	0.072
Cabecar[2]	28	0.000	0.821	0.000	0.000	0.000	0.000	0.179
Teribe[2]	15	0.000	0.800	0.000	0.000	0.000	0.000	0.200
MESOAMERICAN								
Chorotega[1]	23	0.150	0.347	0.434	0.04	0.000	0.000	0.130
Conchagua[3]	23	0.000	0.600	0.170	0.000	0.000	0.000	0.210
San Alejo[3]	9	0.000	0.444	0.222	0.000	0.000	0.000	0.333
Panchimalco[3]	11	0.000	0.545	0.090	0.000	0.000	0.000	0.363
Izalco[3]	11	0.000	0.272	0.454	0.000	0.000	0.000	0.272
Nueva Con.[3]	9	0.000	0.333	0.362	0.000	0.000	0.111	0.222

1=This study; *2*=Ruiz-Navarez *et al.* 2005 (these populations were only characterized for the Q-M3 Y-chromosome haplogroup); *3*= Lovo-Gomez *et al.* 2006;

SHORT TANDEM REPEATS (STRs)

The results of the Y-STR analysis for the five study populations are presented in Table 12. These groups were analyzed for eight Y-chromosome STR loci: DYS19; DYS389I; DYS389II; DYS390; DYS391; DYS392; DYS393; and DYS439. The Rama (n = 20) are characterized by 15 haplotypes that are not shared with other study

populations. The Zapáton Huetar (n = 13) have 12 haplotypes and share a single haplotype with the Abrojo Guaymi (n = 19) and have 15 haplotypes. This was the only Y-chromosome haplotype shared among study populations. The Chorotega contain the highest number of Y-chromosome STR haplotypes (21) and the Maléku have the lowest (8). The Rama contain 15 Y-STR haplotypes, the Abrojo Guaymi have 14, and the Zapáton Huetar have 12 Y-chromosome haplotypes.

Table 12: Y-chromosome STR variation in five study populations

			DYS STRs							
Ht[A]	**Hg[B]**	**N**	**19**	**389I**	**389II**	**390**	**391**	**392**	**393**	**439**
RAMA (N=20)										
R1	R1b	2	13	13	29	24	9	13	13	13
R2	Q3	2	13	13	29	24	12	16	13	11
R3	Q3	1	13	13	30	24	10	12	13	12
R4	Q3	1	13	13	30	24	10	14	13	11
R5	Q3	1	13	13	30	24	10	13	13	12
R6	Q3	2	13	13	30	24	10	14	14	13
R7	Q3	1	13	13	30	24	12	14	14	12
R8	Q3	2	13	13	30	24	12	16	13	11
R9	R1b	1	14	13	30	23	10	14	13	12
R10	R1b	1	14	13	30	24	10	14	13	12
R11	R1b	1	14	13	30	24	10	16	13	12
R12	R1b	1	14	15	32	24	10	13	13	11
R13	G2a	2	15	12	29	23	11	11	13	11
R14	R1b	1	15	13	29	23	10	13	13	11
R15	R1b	1	16	13	30	24	10	12	13	12
HUETAR ZAPÁTON (N=13)										
Z1	Q3	1	12	14	32	24	10	13	13	13
Z2	Q3	1	13	13	29	24	10	13	13	12
Z3	R1b	1	13	13	29	24	10	14	14	13
Z4	Q3	1	13	13	30	24	10	15	13	12
Z5	Q3	1	13	13	30	25	10	15	13	12
Z6	R1b	1	13	14	29	24	10	14	13	12
Z7	Q3	1	13	14	30	24	11	12	13	12
Z8	Q3	1	13	14	30	25	10	14	13	12
Z9	Q3	1	13	14	30	25	10	16	14	12
Z10	Q3	1	13	15	31	25	12	13	13	12

Z11	Q3	2	13	15	32	25	10	16	13	12
Z12	Q3	1	14	14	30	24	12	13	13	11
ABROJO GUAYMI (N=19)										
G1	Q3	1	9	13	32	24	10	15	13	12
G2	?	1	12	15	33	24	9	15	12	11
G3	Q3	1	13	10	27	24	10	15	13	12
G4	Q3	3	13	10	27	25	10	16	13	12
G5	?	1	13	10	28	24	10	16	13	12
G6	?	1	13	10	30	24	10	13	14	11
G7	Q3	1	13	12	29	24	10	14	12	11
G8	Q3	1	13	12	30	25	11	15	12	12
G9	R1b	1	13	13	28	24	10	14	12	11
G10	Q3	1	13	13	30	24	9	15	13	12
G11	Q3	3	13	13	30	24	10	15	13	12
G12	R1b	1	13	14	30	24	10	13	13	11
G13	Q3	2	13	14	30	24	10	15	14	12
G14	?	1	18	11	26	24	10	16	12	10
MALÉKU (N=13)										
M1	E3	1	12	12	28	24	9	11	13	12
M2	Q3	1	12	13	29	24	10	15	12	13
M3	Q3	2	13	12	29	24	10	14	13	11
M4	E3	4	13	12	30	24	9	11	12	10
M5	Q3	1	13	12	30	24	10	15	12	12
M6	Q3	2	13	12	31	24	10	15	12	12
M7	E3	1	13	14	30	24	9	11	12	10
M8	I2b	1	16	12	30	24	10	12	15	11
CHOROTEGA (N=23)										
C1	R1b	1	11	14	27	24	11	14	13	12
C2	?	1	13	11	25	21	10	14	13	14
C3	Q3	2	13	12	25	21	10	14	13	12
C4	Q	1	13	12	27	22	11	15	13	12
C5	Q3	1	13	12	29	21	10	14	13	13
C6	Q	1	13	12	30	21	10	14	13	12
C7	Q3	1	13	12	30	21	10	14	13	13
C8	?	1	13	13	26	19	10	14	14	12
C9	R1b	1	13	13	26	24	10	13	13	11
C10	R1b	1	13	13	26	24	10	14	14	14
C11	Q3	2	13	14	27	23	10	14	13	11
C12	Q	1	13	14	30	23	10	13	13	12
C13	R1b	1	14	13	26	24	10	13	13	11
C14	R1b	1	14	13	26	24	10	13	14	13
C15	R1b	1	14	13	26	24	10	14	13	12

C16	Q3	1	14	14	31	23	10	11	13	14
C17	R1b	1	14	15	28	24	10	14	13	11
C18	R1b	1	15	13	25	24	9	14	13	12
C19	E1b	1	15	13	30	19	10	11	14	13
C20	R1b	1	16	12	29	24	10	14	14	13
C21	R1b	1	16	13	26	24	10	13	14	13

DIVERSITY MEASURES

Gene diversity measures, by population for each Y-STR locus and entire haplotypes are presented in table 13 in both study and comparative populations. The highest gene diversity value (0.599) is found in the 389II locus, while the lowest (0.337) is found in the DYS19 locus. The highest amount of average gene diversity (0.974) is found among Mesoamerican populations with two populations having a value of 1.00, meaning that each individual belongs to a different haplotype. The lowest amount (0.777) is detected in South American populations. However, both of these values are likely skewed due to the small sample size of some of the populations in each group. Among the study populations the highest gene diversity value was found among the Zapáton Huetar (0.987) and the lowest value was among the Maléku (0.897). All five study populations demonstrate higher Y-chromosome gene diversity values than mtDNA gene diversity values.

Y-Chromosome STR Median Joining Networks

Two median-joining networks based on Y-STR variation present in the five study populations are presented in this section in order to provide a visual representation of this diversity. These networks are each comprised of a single

haplogroup and represent the two most common haplogroups (Q3 and R1b) present in the sample groups. Native American haplogroup Q3 (figure 22) is found in all five of the study populations. These populations are characterized by a number of nodes

Table 13: Y-Chromosome STR Diversity values in five study and 19 comparative populations

Population	N	DYS19	389I	389II	390	391	393	GD[A]
CHIBCHAN								
Rama[1]	20	0.605	0.278	0.542	0.336	0.647	0.268	0.952
Huetar - Zapáton[1]	13	0.294	0.692	0.730	0.538	0.410	0.282	0.987
Huetar - Comb.[2]	28	0.507	0.611	0.769	0.568	0.455	0.473	0.928
Guaymi - Abrojo[1]	19	0.298	0.801	0.608	0.350	0.292	0.409	0.953
Guaymi – Oso[2]	8	0.000	0.571	0.535	0.571	0.678	0.571	0.678
Maléku[1]	13	0.410	0.294	0.679	0.000	0.538	0.564	0.897
Bribri[2]	14	0.00	0.439	0.692	0.626	0.142	0.142	0.758
Cabecar[2]	28	0.428	0.560	0.563	0.584	0.203	0.267	0.716
Teribe[2]	15	0.133	0.000	0.619	0.247	0.457	0.133	0.723
Subtotal	*158*	*0.297*	*0.471*	*0.637*	*0.424*	*0.425*	*0.354*	*0.834*
MESOAMERICAN								
Chorotega[1]	23	0.644	0.735	0.822	0.735	0.245	0.403	0.976
Conchagua[3]	23	0.644	0.608	0.818	0.498	0.446	0.486	0.956
San Alejo[3]	9	0.750	0.388	0.833	0.416	0.555	0.666	1.000
Panchimalco[3]	11	0.618	0.345	0.745	0.618	0.654	0.472	0.945
Izalco[3]	11	0.690	0.727	0.890	0.781	0.600	0.490	1.000
Nueva Con.[3]	9	0.694	0.388	0.750	0.805	0.555	0.416	0.972
Subtotal	*86*	*0.673*	*0.531*	*0.809*	*0.642*	*0.509*	*0.488*	*0.974*
SOUTH AMERICAN								
Xavante[4]	5	0.000	0.600	0.000	0.000	0.600	0.000	0.800
Ticuna[4]	32	0.000	0.364	0.467	0.526	0.000	0.578	0.697
WaiWai[4]	5	0.000	0.700	0.000	0.400	0.000	0.400	0.700
Gavaio[4]	34	0.000	0.370	0.661	0.470	0.443	0.000	0.882
Karitiana[4]	8	0.000	0.000	0.000	0.000	0.250	0.000	0.250
Cayapa[4]	26	0.212	0.680	0.621	0.643	0.516	0.396	0.963
Tayacaja[4]	44	0.368	0.582	0.731	0.645	0.311	0.459	0.979
Arequipa[4]	15	0.466	0.590	0.704	0.780	0.133	0.647	0.952
Subtotal	*169*	*0.130*	*0.485*	*0.398*	*0.433*	*0.281*	*0.310*	*0.777*
Total	**413**	**0.337**	**0.492**	**0.599**	**0.484**	**0.396**	**0.370**	**0.854**

A= Gene Diversity values; *1*=This study; 2=Ruiz-Navarez *et al.* 2005; *3*= Lovo-Gomez *et al.* 2006; *4*=Tarazona-Santos *et al.* 2001. Geographic location of population shown in figure 21.

representing 1-2 individuals and are not shared between populations. There are also a large number of reticulations in this graph, indicative of high male genetic diversity found in these populations. There is one star like structure located in the lower right quadrant of the graph that is based on a Zapáton Huetar haplotype. This node contains branches that connect to all four of the other studied Chibchan populations and may indicate a paternal Chibchan haplotype. The Zapáton Huetar also demonstrate the highest amount of haplotype diversity and characterized by 10 nodes, followed by the Rama (7), Guaymi (6), Chorotega (6), and Maléku (4).

The median joining network for haplogroup for R1b STR variation is shown in figure 23 for the five study populations. As with the Q3 network (figure 22) these populations do not share haplotypes. However, this network does show better segregation among male lineages, with the Rama and Chorotega being characterized by separate pathways. The majority of haplotypes (10) present in this network are found among the Chorotega. The second highest number of haplotypes (8) is found among the Rama. Both these population inhabit coastal regions, with the Rama living along the Caribbean coast and the Chorotega residing along the Pacific on the Nicoya peninsula of Costa Rica. Therefore, given over 500 years of European contact in the region this is indicative of high amounts of European male admixture into these communities. The other two represented populations, Abrojo Guaymi and Zapáton Huetar, are each characterized by two haplotypes. The only study population that did not contain haplogroup R1b was the Maléku.

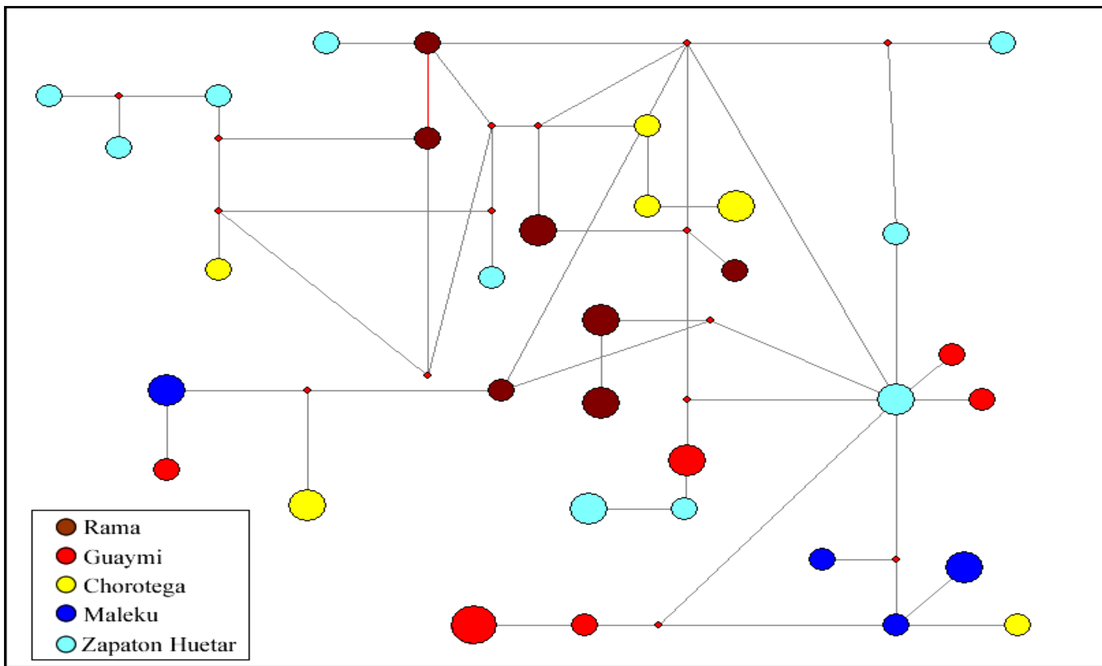

Figure 22: Median Joining Network for Y-Chromosome haplogroup Q3 haplotypes in the five studied populations. Based on eight Y-chromosome STRs.

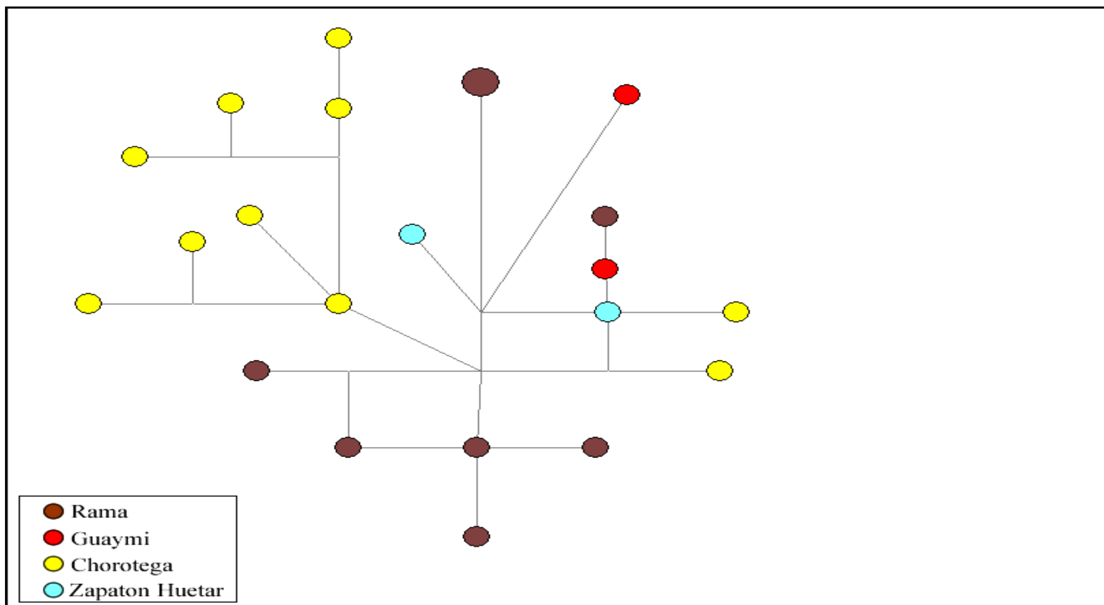

Figure 23: Median Joining Network for Y-Chromosome haplogroup R1b STR haplotypes in four of the five studied populations:

MULTIDIMENSIONAL SCALING PLOT (MDS)

A MDS plot for Y-chromosome STR data in 23 Native Central and South American populations is presented in figure 24. The stress value (0.174, upper bound 0.301) and goodness of fit (0.92) of these data are moderate indicating that the resulting distribution of these populations in two dimensional space is non-random. The Mesoamerican (Conchagua, San Alejo, Panchimalco, Izalco, and Nueva Concepción) Pipil communities from El Salvador form a major cluster that includes the Rama, Zapáton Huetar, and Chorotega. Five Chibchan (Combined Huetar, Bribri, Abrojo Guaymi, Cabecar, and Teribe) populations are interspersed with other South American indigenous populations. A total of five populations are located on the upper and far right periphery of the graph and include the Maléku, Osa Guaymi, Xavante, Ticuna, and Karitiana showing little biological relationships with other study or comparative populations.

Gene Diversity Versus r_{ii}

The results for gene diversity versus r_{ii} using Y-chromosome STR data for 22 Central and South Native American populations are presented in figure 25. A mixture of South American (Tayacaja, Cayapa, Arequipa, and Gavaio) Chibchan (Huetar, Zapáton, Abrojo-Guaymi, and Rama) and Mesoamerican (Chorotega, Izalco, Nuevo Concepcion, San Alejo, and Panchimalco) groups are above the theoretical regression line in the upper left corner of the plot, indicating a high magnitude of male gene flow in these groups. Three of the four Chibchan populations belong to the

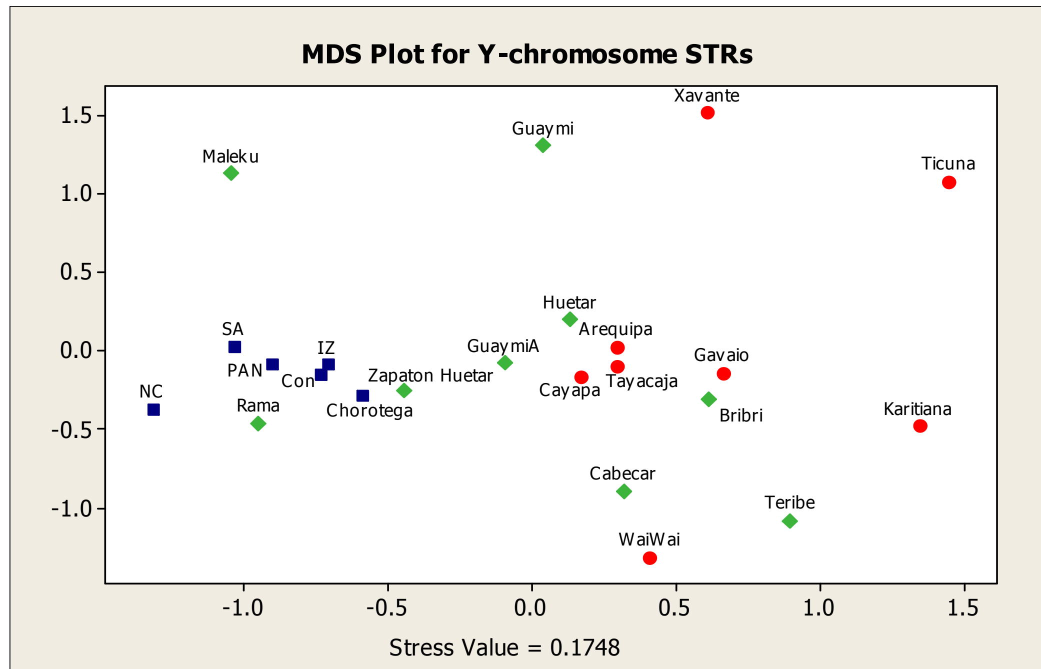

Figure 24: MDS plot of Y-chromosome STR data with 23 populations. Chibchan populations are green triangles; Mesoamerican populations are blue squares; South American populations are red triangles. Geographic locations of these populations are shown in figure 21. Black squares=Mesoamerican, light gray triangles=Chibchan, dark gray triangles=South American.

Votic-speaking linguistic Chibchan sub-family and are geographically located along the northern border of the Chibchan linguistic range indicating the potential admixture with other Mesoamerican populations as well as with European or African populations. All of the Mesoamerican populations are found above the regression line and are geographically located along the Pacific coast of Central America. In addition, four populations (Huetar-Combined, Zapáton Huetar, Abrojo Guaymi, and Gavaio) are not located near either the Pacific of Caribbean coast. Both Huetar populations are located in the Central Valley of Costa Rica, a region with a long term European presence and the town of Abrojo is located along a major road near the towns of Neilly, Villa Neilly, and Carmen. Therefore, this high amount of gene flow

in these groups may be indicative of more than 500 years of contact with European, Asian, and African populations.

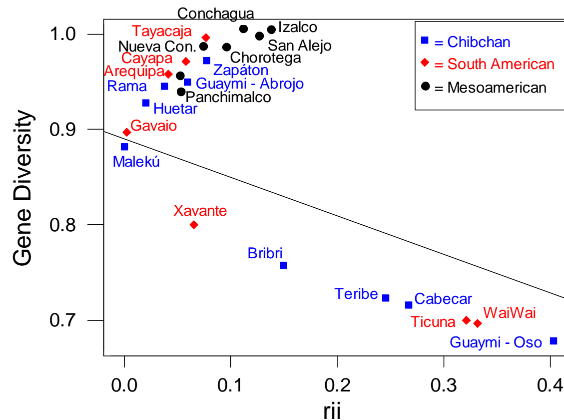

Figure 25: Regression plot of gene diversity values as a proxy for heterozygosity and distances from the centroid (r_{ii}) for 22 Native Central and South American populations using six Y-chromosome STR data.

Eight of the populations in the Y-chromosome gene diversity versus r_{ii} plot are below the theoretical regression line and five (Maléku, Bribri, Teribe, Cabecar, and Guaymi Osa) of these belong to the Chibchan linguistic family, indicating the presence of paternal genetic drift. This paternal genetic drift indicates that males are not moving between populations in these groups. This differs from the mtDNA gene diversity versus r_{ii} plot where all Chibchan populations are characterized by maternal genetic drift. The Maléku are found closest to the centroid and are found just below the regression line indicating a slightly higher amount of gene flow in this population than the other groups below the regression line. The Guaymi-Osa is furthest from the centroid and has the highest r_{ii} values. However, this may be due to the small sample

(n = 8) size of this group, which may underestimate Y-chromosome genetic diversity in this population.

CLASSICAL GENETIC POLYMORPHISMS

Classic genetic polymorphisms for Chibchan and other Central American indigenous populations were obtained from the literature (Matson and Swanson 1963a; 1963b; 1963c; 1965a; 1965b). Comparative populations used in these analyses included: Mayan populations (Mam, Cakchi, Kekchi); Oto-Manguan (Chorotega, Subtiva); Misumplan (Sumo, Miskito); Chibchan (Pech, Rama, Boruca, Bribri, Cabecar, Teribe, Guaymi, Kuna) and Chocoan (figure 26).

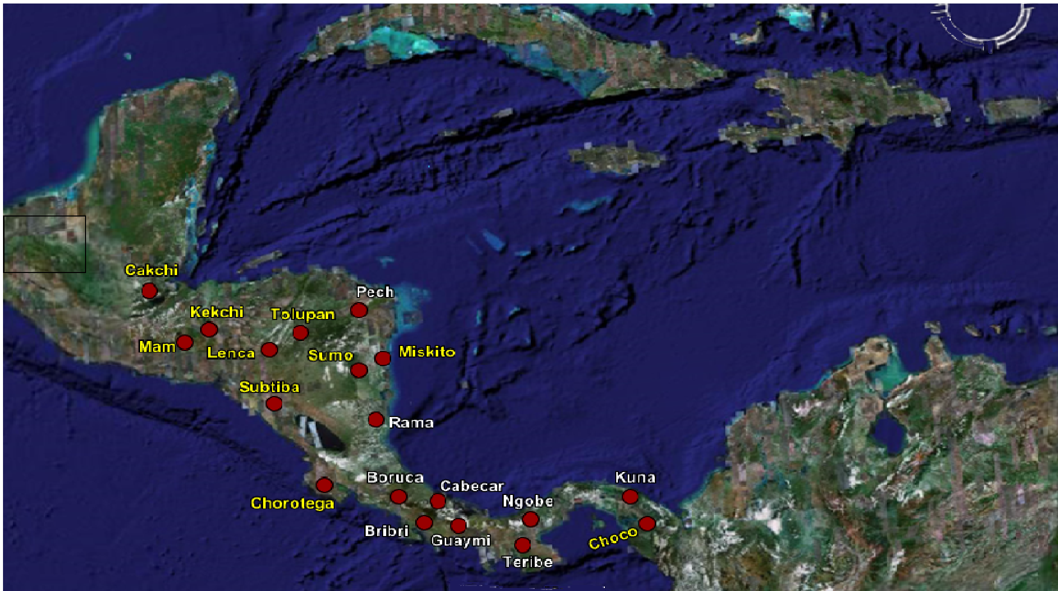

Figure 26: Comparative populations used for 18 Native American populations used in analysis of classical genetic polymorphims.

DIVERSITY MEASURES

In order to determine diversity in classical genetic polymorphisms G_{ST} was calculated by means of the computer program, DISPAN (Ota 1993). These analyses included 18 Central American populations and eight blood groups (ABO, MN/S, P, Rh, Duffy, Kidd, and Diego) loci (23 alleles), with the result shown in table 14. All loci analyzed demonstrate high gene diversity (G_{ST}) values above 0.05. The overall G_{ST} was 0.07, and values were above 0.05 for all other loci except the Kidd blood group (G_{ST} = 0.03). These estimates indicate population subdivision in these groups based on classical polymorphisms. The highest value G_{ST} was found in the Diego blood group (G_{st} = 0.23). The highest gene diversity among populations (H_T) and within subpopulations (H_S) values (0.686 and 0.648 respectively) are found in the MN/S loci. Two blood group loci (ABO and Diego) demonstrated fixed alleles in some populations. The ABO locus was fixed (100%) for the O allele in seven (Rama, Bribri, Cabecar, Teribe, Guaymi, Kuna, and Chocoan) populations. The Di^a allele (100%) was found to be fixed in four (Bribri, Cabecar, Teribe, and Sumo) populations. The highest diversity value (0.501) for all alleles was found in the Tolupan from Honduras and the lowest (0.283) was found in the Sumo from Nicaragua.

MULTIDIMENSIONAL SCALING PLOT (MDS)

A MDS plot for classical genetic polymorphims is displayed in figure 27. This diagram is divided by linguistic affiliation into six groups (Mayan, Chibchan,

Misumplan, Uto-Aztecan, Chocoan, and Unaffiliated). Stress for this plot is high (0.27 upper bound 0.263) and the goodness of fit value is moderate (0.92). The high

Table 14: Gene Diversity of 8 Blood group Loci for 18 Central American indigenous populations

Population	ABO	MNS	P	Rh	Duffy	Kidd	Diego	All Loc
Chibchan								
Pech[1]	0.110	0.678	0.500	0.583	0.493	0.448	0.093	0.415
Rama[2]	*0.000*	0.421	0.478	0.381	0.093	0.504	0.104	0.356
Boruca[3]	0.104	0.658	0.423	0.544	0.506	0.502	0.139	0.411
Bribri[3]	*0.000*	0.717	0.489	0.572	0.354	0.434	*0.000*	0.366
Cabecar[3]	*0.000*	0.499	0.225	0.581	0.435	0.507	*0.000*	0.321
Teribe[3]	*0.000*	0.639	0.356	0.655	0.390	0.393	*0.000*	0.347
Guaymi[4]	*0.000*	0.674	0.490	0.240	0.501	0.362	0.004	0.324
Kuna[4]	*0.000*	0.669	0.478	0.437	0.450	0.487	0.092	0.373
Misumplan								
Sumo[2]	0.052	0.743	0.494	0.496	0.198	0.511	*0.000*	0.283
Miskito[2]	0.184	0.632	0.475	0.616	0.422	0.470	0.013	0.402
Oto-Manguean								
Chorotega[2]	0.148	0.681	0.506	0.577	0.075	0.505	0.013	0.358
Subtiaba[2]	0.256	0.688	0.516	0.655	0.068	0.502	0.502	0.455
Unaffiliated								
Tolupan[1]	0.537	0.711	0.498	0.672	0.502	0.502	0.087	0.501
Lenca[1]	0.173	0.682	0.482	0.601	0.355	0.452	0.093	0.405
Chocoan								
Chocoan[4]	*0.000*	0.706	0.503	0.516	0.451	0.493	0.493	0.451
Mayan								
Cakchi[5]	0.097	0.673	0.502	0.580	0.435	0.494	0.085	0.409
Kekchi[5]	0.096	0.677	0.480	0.553	0.476	0.477	0.056	0.402
Mam[5]	0.083	0.676	0.503	0.529	0.354	0.386	0.101	0.376
$\mathbf{H_S}^a$	0.101	0.648	0.460	0.536	0.360	0.462	0.102	0.381
$\mathbf{H_T}^b$	0.110	0.686	0.486	0.571	0.424	0.480	0.133	0.413
$\mathbf{G_{ST}}^c$	0.086	0.056	0.052	0.061	0.149	0.037	0.230	0.076

a=Gene diversity within subpopulations, b=Gene diversity among subpopulations **c**=Coefficient of gene differentiation: 1= Matson and Swanson (1963a); 2= Matson and Swanson (1963c); 3= Matson and Swanson (1965a); 4= Matson and Swanson (1965b); 5=Matson and Swanson (1963b). *Null values are show in italics:* 0.000 equals fixation for O allele in ABO locus and Dia for Diego locus.

stress value suggests that this may be a random distribution of the classical genetic variation displayed by these data. However, a cluster near the middle consists of

Chibchan (Pech, Boruca, Bribri, and Kuna) and Mayan (Mam, Cakchi, and Kekchi) populations. Four populations (Lenca, Chorotega, Lenca, and Cabecar) appear as satellites around this cluster. Six populations (Rama, Chocoan, Teribe, Guaymi, Tolupan, and Subtiva) are positioned on the periphery of this plot. Three of these populations (Rama, Teribe, Guaymi) belong to the Chibchan linguistic family, one (Tolupan) is a linguistic isolate and one (Subtiva) belongs to the Oto-Manguean linguistic family.

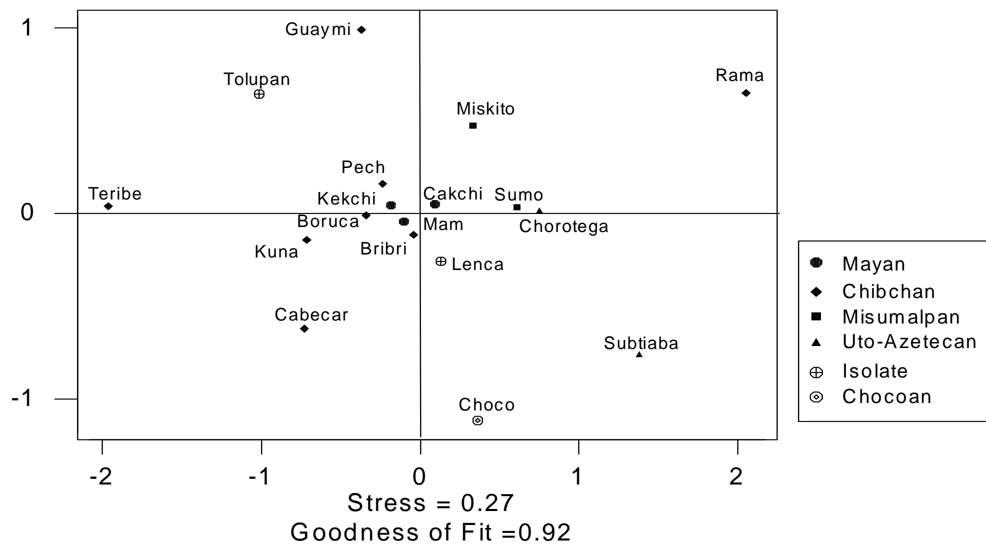

Figure 27: Multi-Dimensional Scaling (MDS) using standard genetic distances for 18 Central American populations using 23 alleles from 7 classical genetic polymorphisms (ABO, MNS, P, Duffy, Rh, Kidd, and Diego). Geographic locations of populations are shown in figure 25.

HETEROZYGOSITY VERSUS r_{ii}

Figure 28 displays the plot of heterozygosity versus r_{ii} for classical genetic polymorphisms. Nine populations (Kekchi, Kuna, Mam, Bribri, Sumo, Chorotega,

Guaymi, Cabecar, and Rama) demonstrate a lower than expected heterozygosity, relative to the theoretical regression line, indicating a greater degree of isolation than other populations investigated. The remaining nine populations (Tolupan, Choco, Subtiva, Pech, Boruca, Cakchi, Lenca, Miskito, and Teribe) demonstrate higher than expected diversity, indicating gene flow in these populations. However, six of these populations (Tolupan, Choco, Subtiva, Pech, Teribe, and Boruca) are to the far right of this plot indicating that they are farthest from the centroid (r_{ii}).

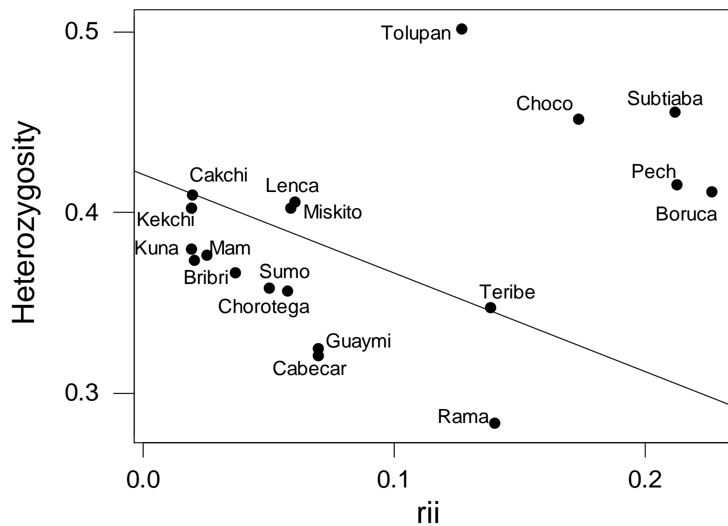

Figure 28: Regression plot of heterozygosity values and distances from the centroid (rii) for 18 Native Central American populations using 7 classical genetic markers (ABO, MNS, P, Duffy, Rh, Kidd, and Diego).

POPULATION STRUCTURE

WRIGHT'S F_{ST}

In order to determine if there are significant differences in population structure between Mesoamerican and lower Central American populations a series of Wright's F_{ST} statistics were calculated based on geographic locality and linguistic affiliation for classical genetic polymorphisms, mtDNA haplotypes diversity, and Y-chromosome STR variability. The results of these analyses are presented in table 15. The overall F_{ST} value for all indigenous Central American classical polymorphisms is 0.09. After populations were divided by cultural region, lower Central American groups exhibit a higher F_{ST} value (0.12) than Mesoamerican populations (F_{ST} = 0.06) based on classical genetic data. These values indicate a moderate amount of genetic differentiation occurring in the region based on classical genetic polymorphisms. A similar pattern emerges with both the mtDNA and Y-chromosome frequency data. Mitochondrial DNA haplogroup frequencies present the highest F_{ST} values of the three genetic systems investigated (Central America = 0.184, Mesoamerica = 0.161, and lower Central America = 0.229). These values all indicate that the highest level of population subdivision in the region occur at the maternal level. The Y-chromosome demonstrate F_{ST} values comparable to classical polymorphisms (Central America = 0.096, Mesoamerica = 0.069, lower Central America = 0.141) and indicate moderate population differentiation at the paternal level. None of these F_{ST} values differentiate the region based on cultural area.

Table 15: Wright's F_{ST} Values for Cultural Area and Language Affiliation

Culture Area	Group Size	F_{ST}
Mesoamerica		
Classical	10	0.069
mtDNA haplogroup	12	0.161
Y-chromosome	9	0.069
Lower Central America		
Classical	11	0.124
mtDNA haplogroup	10	0.229
Y-chromosome	9	0.141
Language Affiliation		
Maya		
Classical	6	0.046
mtDNA haplogroup	5	0.205
Uto-Aztecan		
mtDNA haplogroup	7	0.053
Misumplan		
Classical	2	0.047
Oto-Manguean		
Classical	2	0.029
Isolate		
Classical	2	0.056
Chibchan		
Classical (Matson & Swanson)	8	0.092
Classical (Barrantes *et al.*, 1990)	8	0.080
mtDNA haplogroup	8	0.175
Y-chromosome	8	0.096
Chocoan		
mtDNA haplogroup	2	0.048
All Populations		
Classical	20	0.096
mtDNA haplogroup	26	0.184
Y-chromosome	18	0.096

Based on linguistic classification, mtDNA haplogroup data demonstrate the highest F_{ST} values. The Mayan groups contain the highest linguistic F_{ST} (0.205), followed by Chibchan (0.175), Uto-Aztecan (0.053), and Chocoan (0.048). These results indicate the highest population subdivision in Mayan and Chibchan groups

and borderline F_{ST} value (~0.05), indicating some population subdivision in Uto-Aztecans. Among classical polymorphisms, the only groups with F_{ST} values above 0.05 are the Chibchan speakers (0.092) and the Mesoamerican linguistic isolates (0.056), which indicates genetic differentiation in these populations. These results indicate that there are genetic subdivisions in indigenous Central American populations based on linguistic affiliations and that mtDNA exhibits the highest genetic differentiation.

Both Y-chromosome and classical genetic polymorphisms indicate moderate genetic differences while mtDNA indicates high genetic differentiation. Based on these results there appears to be a slight difference between population subdivision in lower Central and Mesoamerican groups. On average lower Central America F_{ST} values are twice those in Mesoamerican groups but this may be due to small sample size, geographic isolation, sampling error, and lower effective population size that are difficult to estimate using frequency data obtained from the literature.

ANALYSIS OF MOLECULAR VARIANCE (AMOVA)

Three analyses of molecular variance (AMOVAs) using mtDNA HVS-I and Y-chromosome STR data were conducted in order to determine if population substructure was observable among indigenous Central American populations. AMOVA results in three F_{ST} analogs that partition the observed genetic variation in among groups (F_{CT}), within populations in groups (F_{SC}), and within individual populations (F_{ST}). The first AMOVA separated three groups: 1) Mesoamerican

(Chorotega and K'iche Maya), 2) Chibchan (Rama, Maléku, Quitirrisi Huetar, Zapáton Huetar, Guaymi, Kuna, Ngöbé, Kogi, Ijka, and Arsario); and 3) northern South America (Emberá, Wounan, and Wayuú) and was based on mtDNA HVS-I data. The second AMOVA also used mtDNA HVS-I data and separated the ten Chibchan populations into three groups based on their linguistic subfamilies: 1) Votic (Rama, Maléku, Quitirrisi Huetar, and Zapáton Huetar); 2) Isthmic (Kuna, Guaymi, and Ngöbé) and 3) Magdalenic (Kogi, Ijka, and Arsario). The third AMOVA applied Y-STR data into three groups 1) Mesoamerican (Conchagua, San Alejo, Panchimalco, Izalco, Chorotega, and Nueva Concepción); 2) Chibchan (Bribri, Huetar-Combined, Zapáton Huetar, Abrojo-Guaymi, Rama, Maléku, Cabecar, Osa-Guaymi, and Teribe); and 3) South American (WaiWai, Ticuna, Gavaio, Karitiana, Cayapa, Tayacaja, and Arequipa).

The results of the AMOVA among Mesoamerican, Chibchan and northern South American populations are shown in table 16. The amount of variation observed among groups is 7.74% ($F_{CT} = 0.07$, $p = 0.04$) and is statistically significant at the $p < 0.05$ level. This F_{CT} value indicates that there may be maternal population structure among these groups but this is borderline statistically significant. This maternal population structure indicates that there is differentiation in mtDNA genetic variation among the three groups. There is also statistically significant variation ($F_{SC} = 0.219$ and p-value <0.0001) among populations within these three groups and this accounted for 20.25% of the observed mtDNA genetic variation. This high F_{SC} value indicates that there is population subdivision within these three groups. This

subdivision indicates that there is also differences in mtDNA among population within the three groups. However, this may be an artifact of the small groups size of two of the groups (N = 2 and N = 3). Unfortunately, there is limited mtDNA HVS-I sequence data available for Mesoamerican populations (Boles *et al.* 1995 and this study), which limits the ability to successfully discriminate among these populations. The greatest amount of variation for this AMOVA was observed within populations and accounted for 72.01% of all variation and was statistically significant ($p < 0.001$). The high F_{ST} is indicative of high population structure within these studied populations. However, this magnitude of F_{ST} value from mtDNA HVS-I sequences is not different from similar values provided for other South American populations including the Yanamamo $F_{ST} = 0.25$ (Hunley *et al.* 2008) and for South American mtDNA haplogroup data F_{ST}.= 0.25 (Fagundes *et al.* 2002).

Table 16: AMOVA among three groups (Mesoamerican; Chibchan; and northern South America) using mtDNA HVS-I sequence data.

Source of Variation	D.F.	Sum of Squares	Variance Components	Percentage Variation	F Statistic
Among groups	2	94.57	0.219	7.74	$F_{CT} = 0.07^{*}$
Within groups	12	244.37	0.573	20.25	$F_{SC} = 0.219^{**}$
Within populations	472	962.96	2.040	72.01	$F_{ST} = 0.279^{**}$
Total	486	1303.04	2.83		

The results of the mtDNA HVS-I AMOVA among populations belonging to three different Chibchan linguistic stocks (Votic, Isthmic, and Magdalenic) are shown in table 17. The amount of variation among groups accounted for 6.17% of the

variation and was considered statistically significant ($F_{CT} = 0.061$ and $p = 0.02$). The amount of variation ($F_{SC} = 0.248$ and $p < 0.0001$) within groups described 23.35% of the variation and the amount within populations ($F_{ST} = 0.295$ and $p < 0.0001$) explained 70.48% of the observed variation. These high F_{SC} and F_{ST} values indicate significant mtDNA genetic differences among Chibchan populations even within linguistic subfamilies and may support the idea that these groups are partitioned along maternal lineages. These maternal genetic differences indicate that Chibchan population demonstrate high genetic variability among the three linguistic subdivisions and suggest that there are biological differences among these groups. This genetic differentiation is also seen in the haplogroup A2 median-joining network (figure 14), where Chibchan populations do not share specific mtDNA HVS-I polymorphisms.

Table 17: AMOVA among three Chibchan language stocks (Isthmic; Votic; and Magdalenic) using mtDNA HVS-I sequence data.

Source of Variation	D.F.	Sum Squares	Variance Components	Percentage Variation	F Statistic
Among groups	2	71.199	0.147	6.17	$F_{CT} = 0.061$*
Within groups	7	134.539	0.558	23.35	$F_{SC} = 0.248$**
Within populations	314	529.364	1.685	70.48	$F_{ST} = 0.295$**
Total	323	735.102	2.391		

* p-value <0.05 level, ** p-value < 0.0001

The results of the AMOVA using Y-Chromosome STR data separated into three groups are shown in table 18. The lowest amount of observed variation, 5.54%, was among groups ($F_{CT} = 0.055$), while within groups ($F_{SC} = 0.158$) explained 14.93%,

and within populations (F_{ST} = 0.204) described 79.53%. These high F_{SC} and F_{ST} values are statistically significant and indicate genetic differentiation based on paternal lineages within these populations. In addition, these results differ significantly for those F_{ST} (0.09) provided by Y-chromosome frequency data for Central American populations (table 15). This difference may be due to greater diversity seen among paternal lineages based on STR data than based solely on frequency information. This high Y-chromosome F_{ST} value is also not unseen in other Native South American populations. Hunley *et al.* (2008) reported a F_{ST} value of 0.27 for the Yanamamo and indicates potentially higher genetic differentiation in Native South American male lineages. These data also demonstrate no clear differences among groups indicating that there may not be significant differences among Mesoamerica, Chibchan, and South American groups based on Y-chromosome STR data. This suggests that higher gene flow among males occurs in these indigenous Central American groups than is observed in mtDNA data.

Table 18: AMOVA among three groups (Mesoamerican; Chibchan; and South American) using Y-chromosome STR data.

Source of Variation	D.F.	Sum Squares	Variance Components	Percentage Variation	F Statistic
Among groups	2	39.734	0.096	5.54	F_{CT} =0.055**
Within groups	20	117.60	0.260	14.93	F_{SC} = 0.158**
Within populations	390	540.26	1.385	79.53	F_{ST} = 0.204**
Total	412	697.60	1.741		

*p-value <0.01, ** p-value < 0.0001

ADMIXTURE ESTIMATES

Admixture estimates based on mtDNA data are unnecessary because all individuals in four (Rama, Maléku, Guaymi, Chorotega) of the five study populations belong to Native American mtDNA haplogroups A2, B2, and D1. The only population to demonstrate non-Native maternal gene flow was the Zapáton Huetar, where 92% of the lineages belonged to A2, B2, or D1, 5% belonged to the African L2a haplogroup, and 2.5% were European haplogroup H1.

Estimates for paternal gene flow are presented in table 19. The Zapáton Huetar demonstrated the least amount of non-Native admixture (15%), which differs from expectations as they are the only study population to have experienced detectable maternal gene flow. The Rama ware characterized by 40% European gene flow and 10% Asian and/or African gene flow. This non-Native, non European gene flow is based on the presence of Y-chromosome haplogroup G2a, which is frequent in southwestern Asia but also occurs at low frequencies in Europe. The Maléku are characterized by the lowest amount of European admixture based on the presence of haplogroup E3, which is common in Africa but has also been detected in southern European and Middle Eastern populations. The Guaymi are characterized by 11% European admixture, and 20% other non-Native American admixture. The Chorotega contain the lowest amount of Native American male genetic material and 34% was considered to be of European origin.

Table 19: Admixture Estimates based on Y-haplogroup data.

Population	N	Native American	European	Other
Rama	20	50	40	10
Maléku	13	46	8	46
Zapáton Huetar	13	85	15	0
Guaymi	19	69	11	20
Chorotega	23	34	34	32

SUMMARY

This chapter presented the results for mtDNA RFLP and HVS-I sequences, Y-chromosome SNPs and STRs as well as classical genetic polymorphisms. Based on maternal genetic lineages Chibchan populations are characterized by high frequencies of mtDNA haplogroup A2 and moderate frequencies of haplogroup B2. Mitochondrial DNA HVS-I variation also demonstrates that Chibchan-speakers cluster with other Mesoamerican populations and are separated from indigenous South American groups. The one exception to this pattern are the Chibchan-speaking Rama who are characterized by a high frequency of mtDNA haplogroup B2 (92%) and cluster with two eastern South American populations (Xavante and Ache). However, this relationship is based on a shared high frequency of haplogroup B2 among these groups and HVS-I sequences indicate that the Rama are a Chibchan population that has undergone genetic drift. Gene diversity versus r_{ii} analysis indicates that the maternal genetic history of Chibchan populations has been affected by genetic drift. These results differ from Y-chromosome SNP analysis, which suggests that Chibchan populations contain high paternal genetic diversity, indicative of male gene flow. However, gene diversity versus r_{ii} plots for Y-chromosome STRs

indicate a differentiation among Chibchan populations in regards to genetic drift vs. gene flow where those groups belonging to the northern Votic-Chibchan speaking populations are being affected by gene flow, whereas the majority of groups belonging to the southern Isthmic-Chibchan linguistic family are being impacted by genetic drift. This is indicative of higher gene flow from European and Mesoamerican males into the northern and central Votic Chibchan groups than in the southern and eastern Isthmic Chibchan populations. Phylogeographic analysis of mtDNA HVS-I variation detected a genetic between Votic and other Chibchan-speaking populations in the region. This discontinuity corresponds to Lake Nicaragua, which may have provided an impediment to gene flow at the same time that mtDNA coalescent dates indicate the genetic divergence of Chibchan populations from earlier groups. Population structure analysis indicates that Mesoamerican and lower Central American populations are characterized by high F_{ST} values for mtDNA, and moderate amounts of Y-chromosome and classical genetic F_{ST} values, which indicates population subdivision in the region within cultural areas and based on linguistic affiliation.

VI: DISCUSSION

This chapter places the results of this study (summarized with the previous chapter) within the broader context of previous molecular genetic, archaeological, and anthropological research in indigenous lower Central America populations and their relationship to surrounding Native American groups. The specific topics addressed in this chapter are: 1) Genetic diversity; 2) Maternal versus paternal genetic histories; 3) Biological relationships among the K'iche Maya, Chorotega, and Chibchan populations; 4) Selectively neutral evolutionary forces operating on Chibchan populations (genetic drift versus gene flow); and 5) Phylogeographic analysis, paleoecology and the origin of Chibchan populations;

GENETIC DIVERSITY

Mitochondrial DNA markers examined in five indigenous Central American populations belong to Native American haplogroups A2, B2, and D2 (Schurr *et al.* 1990; Torroni *et al.* 1993). The Chibchan speaking Rama, Maléku, and Guaymi are characterized by only by two (A2 and B2) of these haplogroups. These results are consistent with other previously reported data for the Maléku and the Guaymi (Torroni *et al.* 1994b, Kolman and Bermingham 1997). The previously uninvestigated Rama differed by having a much higher frequency of haplogroup B2 (92%) than either the Maléku (8%) or the Guaymi (24%). The Chibchan Zapáton Huetar were characterized by three Native American (A2, B2, and D1), one African (L2a), and one European derived (H1) haplogroup. One previous study of the Quitirrisi Huetar had identified the presence of haplogroup D1 in this population (Santos *et al.* 1994).

The Oto-Manguean speaking Chorotega exhibited three (A2, B2, and D1) of the Native American mtDNA haplogroups. Only one (Zapáton Huetar) of the five study populations exhibited non-Native American mtDNA haplogroups (L2a and H1). The Zapáton Huetar contained both European and African haplogroups. This is consistent with previously published data using classical markers that indicated genetic admixture in this population (Barrantes *et al.* 1990; Azofeifa *et al.* 2001). The absence of maternal admixture in the other four populations would indicate that gene flow into these groups was generated from male lineages.

In contrast to other neighboring Native American populations, Chibchan-speaking populations of lower Central America demonstrate a distinctive pattern of mtDNA haplogroups and haplotypes. Neighboring Central American populations have previously exhibited three or four of the Native American founding mtDNA haplogroups (A2, B2, D1). The K'iche Maya (Bolas *et al.* 1995) contain four (A2, B2, C1, D1) of the five (A2, B2, C1, D1, X2a) mtDNA haplogroups, as do the Chocoan-speaking Wounan (Kolman and Bermingham 1997). In addition, the Chorotega (this study) and the Emberá (Kolman and Bermingham 1997) contain three (A2, B2, D1) of the five mtDNA haplogroups. According to some researchers, this is indicative of a Chibchan population bottleneck that did not include individuals with haplogroups C1 or D1 (Kolman and Bermingham 1997). The majority of individuals in Chibchan populations belong to haplogroup A2. This is shared with Chibchan-speaking populations from northern South America (Melton *et al.* 2007). Kolman and Bermingham (1997) suggested that haplogroup C1 was not present in Chibchan

groups throughout their genetic history. However, Melton (2005) and Melton *et al.* (2007) suggested that all four Native American haplogroups were present at the beginning of their genetic history and subsequently lost through genetic drift. Melton *et al.* (2007) also offered the alternative explanation that individuals belonging to haplogroup C1 were part of an earlier migration and these individuals adopted the Chibchan language as these populations expanded from Central America in to northern South America (Keyeux *et al.* 2002; Fox 1996).

The Huetar and the Boruca are the only two Chibchan populations that have previously demonstrated low frequencies of haplogroup D1 (Santos *et al.* 1994; Torroni *et al.* 1994b). They also share this with the Chorotega, who are believed to have migrated into the Gran Nicoya region approximately 1,200 YBP (800 A.D.) (Fowler 1989). Haplogroup D1 is detected in low frequencies in other Mesoamerican populations including the K'iche Maya (Boles *et al.* 1995) and ancient Maya from Copán (Merriwether *et al.* 1997). Therefore, there is the possibility that haplogroup D1 was introduced into Chibchan populations from Mesoamerica. This would explain the presence of D1 in the Huetar and Boruca as indicative of pre-Columbian maternal gene flow. Similar to post Chibchan populations, the marriage residence pattern of these groups is matrilocal. However, the Chorotega also share a high frequency of haplogroup A2 (73%) with neighboring Chibchan populations. An alternate possibility is that the Chorotega represent the remnants of a Chibchan population that adopted a language and certain Mesoamerican cultural traits from Nahua and Oto-Manguean-speaking populations that migrated from central Mexico into the region

(Fowler 1989). One Chibchan population that may have inhabited the Gran Nicoya region prior to the arrival of the Chorotega are the Corobici. The Corobici are an extinct Votic-speaking Chibchan population, with close linguistic affinities to the Rama (Constenla 1995). The Chorotega population share HVS-I polymorphisms with other Chibchan populations. The shared relationship of mtDNA among Chorotega and Chibchan groups and of Y-chromosome among the Chorotega and Mesoamerican groups is suggestive of migrating Mesoamerican males intermarrying with Chibchan females.

Recent archaeological evidence offers contradictory evidence to the assumption that there was complete replacement of earlier Chibchan cultures with Mesoamerican motifs that are identified in ceramics, burial patterns, and social organizations during the Sapoa period (650 to 1,200 YBP) (González and Leiva 2006). The Chorotega are believed to have migrated from the central Mexican highlands in the 8th century and migrated to the Rivas region of Pacific Nicaragua. Subsequent migrations of Nahua speakers forced this group further south into the Gran Nicoya region of Costa Rica where they resided when contacted by the Spanish (Fowler 1989). Recent archaeological evidence from the site of San Isabel (800 to 1200 A.D.), on the western shore of Lake Nicaragua, suggest only partial replacement with the presence of superficial Mesoamerican traits related to iconography and ideology but this group lacked shared domestic characteristics such as maize and dog food production, absence of *comales* (large cooking plates), and maintained different housing structures then groups to the north (McCafferty 2008). This interpretation

differs from that of González and Leiva (2006) who argue that total population replacement occurred with the arrival of Nahua and Oto-Manguean groups, and this is seen in the archaeological record of the Gran Nicoya region.. This archaeological evidence includes Mesoamerican ceramic motifs and social organization (González and Leiva 2006) In addition, unpublished data from ancient DNA indicates a high frequency of haplogroup B2 in Gran Nicoya skeletal material suggesting a closer biological relationship to the Rama than to present day populations (MH Crawford pers. comm.).

The Rama is the most divergent Chibchan population at the mtDNA locus, due to their high frequency of haplogroup B2 (92%). In the MDS plot (figure 16) they cluster with the Aché and Xavante, two eastern South American populations. This genetic relationship between the Rama and the eastern South American groups is largely due to their shared haplogroup frequency for B2 and probably not due to recent common ancestry. Haplogroup B2 is common throughout several Native American populations. The Aché (90%) (Schmidt *et al.* 2004) and Xavante (84%) (Ward *et al.* 1996) both exhibit high frequencies of haplogroup B2. In addition, the Chibchan- speaking Bari of Colombia had previously been reported to contain 100% haplogroup B2 (Keyeux *et al.* 2002). The Rama also share a number of haplotypes with other Chibchan populations particularly the 16360 sequence variant that is shared by the majority (6 out 7) of Central American Chibchan populations and the 16187 sequence variant that is shared with Isthmic-speaking groups (figure 14 and 15). Analysis from ancient DNA from highland Colombia also demonstrated a higher

frequency of haplogroup B2 in northern South America during the Pre-Columbian era. Jara *et al.* (2008) investigated mtDNA haplogroups in 14 skeletons from the Muisca era (400 to 1,200 YBP) from Colombia and found all of these specimens exhibit haplogroup A2. However, in the preceding Herrera period (1,200 to 2,600 YBP) all skeletal material contain haplogroup B2 (Briceño *et al.* 2008), signifying population replacement. Archaeological evidence also indicate changes between the Herrera and Early Muisca periods with increased population growth and an increase in pottery styles that may be related to feasting (Langeback 2003). Given the time frame (~1,200 YBP), this replacement may be reflective of Mesoamerican migrations pushing Chibchan populations into northern South America and towards the Caribbean coast. However, little archaeological research has been done regarding the movement of Chibchan populations into South America.

Based on the aforementioned high frequency of haplogroup B2 in Gran Nicoya ancient DNA analysis, the Rama potentially represents a single or group of Chibchan populations that migrated to the Caribbean after contact with Mesoamerican groups. An alternative interpretation is that the Rama are undergoing genetic drift due to their isolation in the tropical forests of eastern Nicaragua. Until recently there was little archaeological research conducted on the Atlantic coast of Nicaragua (Magnus 1978; Hoopes 2005). The earliest archaeological site in the region is a shell mound excavated near Monkey Point in 1969 that was radiocarbon dated to 7,600 and 5,500 YBP and contained stone artifacts, animal material, and shark bone indicative of a seafaring population with advanced fishing techniques

(Riverstone 2004). Further research by Spanish archaeologists since the mid 1990s has identified 25 sites and 85 shell mounds along the coast. The earliest of these sites, "El Cascal de Flor de Pino" near the Nicaraguan town of Kukra Hill dates to approximately 2,700 YBP with the major occupation occurring around 1,500 YBP (Gassiot and Clemente 2003). During this time period, the site appears to be a settlement with evidence for cooperative labor and the presence of basalt columns indicative of monumental architecture. Other archaeological sites from eastern Honduras and Costa Rica are contemporary with this site and indicate long term human occupation along the Caribbean coast (Hoopes 2005). This region may represent an expansion of Chibchan populations north along the Caribbean as the site is located between the current location of the Pech and the Rama, two Chibchan groups that are separated by the Miskito.

Central American historians have long debated where the Rama originated and have attributed their presence to numerous indigenous populations mentioned at European contact. These groups include the extinct populations, Suerre, Voto, Melchora, Corobici and Caribes, as well as the extant groups, Maléku and Huetar (Riverstone 2004). Coalescent times for population divergence presented in this dissertation (table 10) rule out a recent shared origin with either the Huetar (4,249 ya Zapáton and 7,553 ya Quitirrisi) or the Maléku (10,531 ya) and indicate no recent shared biological relationships with these groups. The Suerre are an indigenous population that lived along the Caribbean coast of Costa Rica during the 1500s but were heavily decimated by disease and violence and the survivors were believed to

have been integrated with the Voto. The Voto were an important indigenous population that inhabited the region around the Rio San Juan. This population offered fierce resistance against the Spanish until 1666, when the they were forcibly relocated inland to prevent them from assisting British pirates. The last A 1745 account of the Voto mentions 200 Voto-Rama along the shores of Lake Nicaragua (Riverstone 2004). The location of the Voto along the Rio San Juan, which today forms the southern boundary of Rama territory may indicate a possible relationship between these two populations. There is little information regarding the Melchora, an indigenous population that inhabited river systems southeast of Lake Nicragaua, and the only possible relationship with the Rama is geography. The argument for a shared relationship with the extinct Corobici is more difficult to discount. The Corobici are believed to have inhabited the southern region of Lake Nicaragua at European contact and may have once occupied the Gran Nicoya peninsula (Johnson 1948). Then, after European contact, the remnants of this population were forced toward the Caribbean where they were absorbed into the Maléku or became the Rama. The term "Caribes" also appears in Spanish ethnohistoric accounts of indigenous peoples inhabiting the regions around Lake Nicaragua but this is believed to differentiate the native Caribbean groups from the Mesoamerican populations along the Pacific (Riverstone 2004). An alternative to the several populations that may be the predecessors of the Rama, is that these groups could be representative of a single indigenous populations. All of these populations are only known from ethnohistoric references and it is

possible that the same population was being described using different terminology by the early Spanish chroniclers in the region.

Rama oral history, on the other hand, indicates they occupied their current location prior to European contact and this also included a larger region that extended from the Caribbean coast to the eastern shores of Lake Nicaragua. Subsequent conflict with the Miskito, armed by the British between the 16th and 19th centuries, forced the Rama into isolated geographic regions. These isolated regions are where the majority of Rama communities are located today. It was not until the introduction of the Moravian church in 1859 on Rama Cay, does there appear to be resurgence in population numbers. According to Conzemius (1930), there were only 270 Rama speakers at that time. Current estimates place population size around 1,300 (Brignoli 2005) to 3,000 (L. Martinez pers. comm.). This small population size and current rebound over the last 50 year period would indicate that genetic drift is the most likely interpretation for the high frequency of haplogroup B2 in this population. Evidence is also seen for genetic drift in the heterozygosity versus r_{ii} plot for classical markers (figure 28) and mtDNA (figure 17). The higher Y-chromosome diversity values for the Rama may be the result of British political influence in the region and the introduction of the Moravian church (Riverstone 2004).

Y-Chromosome lineages present among the five study populations include the Native American haplogroups Q and Q3, European derived haplogroups R1b, G2a, and I1b, and the African haplogroup E3. The frequency of Native American specific haplogroup Q3 vary from 85.4% in the Huetar Zapáton to 34.7% in the Chorotega.

Native American Y-chromosome haplotypes are characterized by two haplogroups Q3 and C (Zegura *et al.* 2004). Haplogroup Q3 is found in Native American populations throughout both North and South America. Y-chromosome haplogroup C is highest in northwestern North America, found in lower frequencies in Athabaskan populations of the southwest, and only been detected in the Wayuú of South America (Zegura *et al.* 2004; Mahli *et al.* 2008). Haplogroup C is absent in all of the Chibchan populations indentified in this study and has not been previously characterized in other Chibchan populations studied for Y-chromosome variation (Lell *et al.* 2002; Zegura *et al.* 2004; Ruiz-Narvaez *et al.* 2005). Therefore, the Native American paternal component in these populations is represented by Q (15%) and Q3 (66%). Y-chromosome haplogroup Q is also found in Siberian populations and in low frequencies in Europe, East Asia, and the Middle East indicating the potential for admixture with other populations (Karafet *et al.* 2008).

The European haplogroup with the highest frequency in the study and comparative populations is R1b, with values ranging from 43.7% in the Chorotega to 10.5% in the Abrojo Guaymi. Haplogroup R1b is found in high frequencies in the British Isles and Iberian peninsula (Wilson *et al.* 2001). The presence of this haplotype in high frequencies in these populations is indicative of paternal gene flow from western Europe after contact. The presence of high frequencies of R1b in these populations is consistent with historical information regarding European contact. Columbus visited the region on his fourth voyage to the Americas in 1502. He termed the region “Costa Rica” (Rich Coast) due to the large number of gold objects worn by

the indigenous populace (Morrison 1942). By 1520, the Spanish conquistador Gil Gonzalez Dávila headed from the Spanish settlement of Panama along the Pacific coast, where he first contacted the Chorotega chief Nicoya. Nicoya informed him of a larger Nicarao settlement headed by chief Nicaragua. By 1560, the indigenous population of Pacific Nicaragua is believed to have collapsed from an estimated 600,000 to 6,000 "tributaries" due to disease, violence and the forced removal of slaves to Peru (Cooke 1997). A caveat to this is that "tributaries" may not be equivalent to population size and may be more reflective of the dwindling amount of easily obtainable gold in the region as the indigenous population was decimated or that it refers to male members of a population. This is in contrast with indigenous populations on the Caribbean coast, that were influenced and armed by the British in order to attack Spanish interests in the Caribbean region until 1860 (Riverstone 2004). Therefore, the high frequencies of Y-chromosome R1b are likely the result of admixture from Spanish and British males.

The analysis of classical genetic polymorphisms indicates population subdivision in these populations based on G_{ST} values above 0.05. In addition, the Classical genetic MDS plot (figure 27) demonstrates a shared biological relationship between Mayan and Chibchan speaking populations. This is similar to the mtDNA MDS plot (figure 16) where the K'iche Maya, Chorotega, and the majority of Chibchan population cluster based on HVS-I sequence data. Melton et al. (2007) argued that this shared biological relationship among the K'iche Maya and Chibchan groups was based on a shared common biological origin for these two groups, but

each of these groups had been affected by different evolutionary forces. In the case of the K'iche Maya, that demonstrate higher haplotype diversity and positive Fu's Fs scores, they had been impacted by gene flow. Whereas Chibchan populations contain low haplotype diversity and negative neutrality test statistics, indicative of genetic drift. This relationship is further supported by recent unpublished data from both archaeology and linguistics (J. Hoopes pers. comm.). However, the K'iche Maya are different than the other surrounding Mayan populations and are considered to have migrated to Central Mexico around 800 YBP (1200 A.D.) and may be Nahua and Chontal Maya speakers influenced by the Toltec empire that migrated into the Guatemala highlands during this time period (Carmack 1968). This time frame is similar to the movement of other indigenous populations from central Mexico, including the Chorotega and the Nicarao. However, further genetic and archaeological work is necessitated in order to elucidate this relationship.

MESOMAERICAN – CHIBCHAN POPULATION RELATIONSHIPS

Contrary to previous results that have argued for the distinction of Chibchan populations from other indigenous groups in Central America (Barrantes *et al.*, 1990; Torroni *et al.*, 1994b; Batista *et al.*, 1995; Kolman *et al.*, 1995; Bieber *et al.*, 1996; Kolman and Bermingham, 1997; Ruiz- Narvaez *et al.*, 2005), this dissertation demonstrated further support for a shared biological affinity between Mesoamerican and Chibchan populations (Melton *et al.* 2007). This biological relationship is shown in the MDS plots for mtDNA (figure 17) and classical genetic markers (figure 27) as

well as the SAMOVA (figure 18) for mtDNA. The MDS plot for mtDNA presents a cluster of nine of the ten Chibchan populations, with the K'iche Maya and the Chorotega. Melton *et al.* (2007) investigated intermatch distances between Chibchan and the K'iche Maya Mayan and found that they shared an overlapping peak that was not shared with other South American groups. This was used to suggest a shared biological origin for Chibchan and the K'iche Maya.

While these indigenous Central American populations may have originated within the same timeframe, different evolutionary forces are operating on these populations. The K'iche Maya demonstrate genetic characteristics consistent with gene flow and included high mtDNA haplotype diversity values and significantly negative neutrality test statistics. In contrast, based on mtDNA data, all Chibchan populations demonstrate genetic characteristics consistent with genetic drift. The separation of Central from South American populations is also displayed in the SAMOVA, which identified four groups. Two of these groups appear to be Chibchan outliers, the Rama and Maléku, that are characterized by high frequencies of mtDNA haplogroups, B2 and A2, respectively. Whereas, the Chocoan-speakers (Emberá and Wounan) and the Wayuú cluster together in their own group. The two Chocoan groups are known to be recent immigrants to Central America and migrated along the Pacific coast from Colombia into the Darien region of Panama since European contact (Kolman and Bermingham 1997; Herlihy 1997). Previous research on mtDNA in the Wayuú has also differentiated them from Chibchan populations and suggests they share closer biological affinities to other South American populations

(Melton 2005; Melton *et al.* 2007). One argument against a shared relationship between Chibchan and Mesoamerican populations is that this relationship is based only on two Mesoamerican groups (K'iche Maya and Chorotega), both of which have distinct cultural histories.

Today, the Maya are a heterogeneous group of populations that speak 28 different languages and represent 6 million different people (Ibarra-Rivera *et al.* 2008) making the biological relationship of one Mayan group (K'iche Maya) from Guatemala (Boles *et al.* 1995) with Chibchan populations somewhat tenuous. There is also the possibility that the Chorotega mtDNA sequences represent the maternal genetic remnants of an earlier Chibchan population that was subsumed by migrating Mesoamerican groups after 1,200 YBP. In order for further confirmation of this proposed Central American relationship, further mtDNA sequence data are needed from other populations in the region, especially those that inhabit Honduras and Nicaragua where there is a lacunae of biological information regarding Native Americans.

In addition, the Y-chromosome MDS plot (figure 24) demonstrates a closer biological relationship between Votic-speaking Chibchan and Mesoamerican populations whereas the Isthmic-speakers cluster closer to other South American groups. This latter pattern was also observed in previous Y-chromosome research where some South American populations clustered together in a neighbor-joining tree with Chibchan populations (Ruiz-Narvaez *et al.* 2005). The paternal relationship between Votic and Mesoamerican groups is not unexpected. The Votic-speaking

Chibchan groups would have bordered the incoming Mesoamerican groups and would have had more opportunity for interaction with them. The Votic-speaking groups may have used this position to act as middlemen between the Mesoamerican groups to the north and other Chibchan populations further south. The Huetar were the largest Central American Chibchan group at European contact (Hoopes and Fonseca 2003). The Huetar may have increased in population size during the period between 300-600 AD when the widespread use of prestige items manufactured from jadeite and gold begin to appear in abundance in the region's archaeological record (Hoopes 2005). This role would have expanded after the migration of Mesoamerican populations along the Pacific coast around 1,200 YBP. In addition, the presence of Votic-speakers may have impeded the further migration of these Mesoamerican tribes further south. The evidence presented here suggests that male and smaller amounts of maternal gene flow did occur between Mesoamerican and Chibchan populations, but it was primarily restricted to the Votic-speaking groups.

The Rama and Chorotega also grouped closely with other Mesoamerican populations based on Y-chromosome STR data. This relationship is not surprising since the Chorotega are a Mesoamerican population. The close paternal biological relationship between the Chibchan Rama with the Pipil Nahua-speaking groups from El Salvador (Lovo-Gomez *et al.* 2005) could be based on the influence of Mesoamerican groups in the Caribbean coastal region. Lothrop (1942) used ethnohistoric evidence to suggest the presence of an Aztec trading colony, Sigua, at the mouth of the Rio Sixaola on the Caribbean between Costa Rica and Panama. This

outpost may have represented the southernmost expansion of the Aztecs along the Caribbean coast. The trading colony was destroyed in 1722 by Miskito pirates and most of the 2,000 inhabitants were enslaved and shipped to Jamaica. The survivors fled to the Bocas del Tora region of Panama and are believed to have intermixed with the Chibchan groups of the region (Cooke et al. 2003). Beekman and Christensen (2003) also locate a Uto-Aztecan linguistic population on the Caribbean coast at the mouth of the Rio San Juan. The location of this latter proposed trading outpost is considered the southern border of Rama territory, and the Aztec and Nicarao are both Nahua-speaking populations. Therefore, this paternal genetic relationship may be representative of male gene flow into the Rama from these groups. However, physical evidence for these outposts remains unsubstantiated.

Another reasonable interpretation for this paternal relationship is that it may represent an influx of Mesoamerican paternal lineages from Miskito Indians. Starting in 1687 the British crowned a "Miskito King", which allowed them to wield political influence over the Caribbean coast of Central America from Honduras to Panama until the late 19th century. Beginning in the early 1700s, the Miskito expanded their territory into Rama territory and enslave other indigenous groups along the coast. In these cases, the males within a village were often killed and the females and children sold into slavery. The majority of the enslaved populace was shipped to Jamaica or taken as personal slaves by the Miskito. After 1740, when large numbers of Africans were brought to Jamaica, the demand for Central American slaves lessened but still continued (Riverstone 2004). Given the violence towards males during this period,

coupled with the increasing isolation of the Rama, it is reasonable to assume that male gene flow from other populations occurred during this time period. The biological relationship between the Miskito and Rama is tenuous at best. There is little known regarding the genetics of populations along the Miskito coast. Aside from previous research on classical genetic markers by Matson and Swanson (1963) and Azofeifa *et al.* (1998), neither of which assessed population relationships, there is no comparative molecular marker data for this population. Based on certain classical genetic private polymorphisms, the Miskito appear to be closely related to Chibchan populations (Azofeifa *et al.* 1998). Therefore, further evidence is required for this proposed biological relationship between the Rama and Miskito to be verified.

DIFFERENCES IN MATERNAL AND PATERNAL GENETIC HISTORIES

Chibchan genetic history varies considerably between mtDNA and Y-chromosome markers. The maternal genetic profile is that of a Native American population with little admixture from outside groups. Only one (Zapáton Huetar) of the five study populations demonstrated maternal gene flow from European and African populations. This population is also characterized by the highest frequency of male Native American haplogroup Q3. The introduction of female genetic material may be due to their location in the central valley of Costa Rica, which is the most populated area in the country. The Huetar are the indigenous population that inhabited the Central Valley prior to European contact and were also the largest indigenous population in the region at the time (Hoopes and Fonseca 2003). The other four populations are characterized exclusively by the presence of Native

America haplogroups (A2, B2, and D1). This, along with reduced mtDNA gene diversity values and positive neutrality test statistics, are the result of genetic drift and the isolation of these communities. The paternal markers, however, indicate a substantial amount of male mediated gene flow from both Native Americans and Europeans. The Zapáton Huetar have one of the highest Y-chromosome diversity values (0.987) and are characterized by the second highest frequency of Native American haplogroup Q3 (84.5%). However, based on the Y-chromosome Q3 median joining network (figure 22) there appears to be an absence of a paternal phylogenetic relationship between populations.

The diversity measures for Y-chromosome variation in the five study populations are high, which is likely the product of two different demographic events: 1) the movement of Mesoamerican populations from central Mexico with the rise of the Toltec empire around 1200 YBP (Fowler 1989); and 2) the arrival of European populations in the early sixteenth century. The first of these demographic events is demonstrated in the Y-chromosome MDS plot (figure 24) where Votic-speaking Chibchan populations cluster with the five El Salvador Pipil Nahua-speaking groups and not with Isthmic-speaking populations from southern Costa Rica and Panama. This is indicative of a shared biological relationship between these populations and suggests gene flow between these groups prior to European contact. Contact with European groups impacted the structure of these population through the introduction of male genetic material. This is evidenced through the presence of male European Y-chromosome haplogroups in all of the five study populations. In addition, there is

abundant documentation of the severity of contact from disease, forced population relocation, and violence that decimated and altered the genetic structure of Native American populations (Crawford 1998). However, the expected outcome of this would be reduced genetic variation and not the high genetic diversity observed in the Y-chromosome. This higher diversity may be representative of higher male population movement in the region.

A number of studies have indicated high paternal European and African admixture estimates for other South American indigenous populations (Santos *et al.* 1999; Mesa *et al.* 2000; Carvajal-Carmona *et al.* 2000). Mesa *et al.* (2000) investigated genetic admixture in five indigenous populations (Emberá, Ingano, Wayuú, Zenu, and Ticuna) from Colombia and found between 95-100% of the mtDNA variation was of Native American origin in all five populations. This study also demonstrated that the male Native American Y-chromosome component varied dramatically between populations. The Amazonian Ticuna had the highest Native American Y-chromosome component (97%) whereas the more urban Zenu (near Medelin) had the lowest (35%). Estimates of European admixture ranged from 62% in the Zenu to 3% in the Ticuna. The Emberá had the highest African paternal component (11%) which was absent in the Ingano and the Ticuna. Carvajal-Carmona *et al.* (2000) found similar results for an admixed population in the Colombian department (state) of Antioquia where 90% of the sampled population had Native American mtDNA but 94% of the Y-chromosomes were of European origin. These authors traced the genetic origins of males back to southern Spain and attributed a

small proportion (~14%) to a Sephardic genetic contribution from northern Iberia. Paternal genetic admixture from Central American populations is unknown, but a recent study of autosomal STR variation in Latin American Mestizos indicated a higher than 70% European admixture rate for the central valley of Costa Rica (Wang *et al.* 2008). Similar studies among Arctic populations have also found differences between Y-chromosome and mtDNA markers indicating unidirectional mating between European males and Native American women (Bosch *et al.* 2003; Rubicz 2007).

The migration of the Spanish to the Americas was primarily done by males. According to the 15,000 names listed in the "Catálogo de Pasajeros a Indias," a list of Spanish passengers migrating to the Americas between 1500 and 1559, only 10% were female (Sanchez-Albornoz 1977). The interactions between the new European immigrants and indigenous peoples were often violent. Early historical accounts of slaughter and violence against Native Americans including mass executions, torture, and enslavement was documented by Spanish chroniclers (Anglería 1976; Oviedo 1976; 1977; Casas 2007). The early Spanish conquistadors in the region considered Native Americans to be the agents of Satan, and under the system of *econmiendia* they were granted the right to enslave a maximum of 300 indigenous inhabitants each (Thomas 2004). Females were often taken as slaves and the males were often killed. Casas (2007) describes one event in Nicaragua in the 1520s where a small group of conquistadors raided a village during the night, killed the males and took the females hostage. On their return, the Spanish were attacked by another group of Native

Americans and rather than return the female captives, they killed them. The enslavement and slaughter of indigenous peoples continued until the papal bull of *Sublimus Dei* was decreed in 1537, forbidding the enslavement of the indigenous populations of the Americas. This decree was widely ignored by the Spanish colonizers but led to the Valladolid debate between 1550 and 1551. This debate between two Dominican friars; the bishop of Chiapas, Bartolomé de las Casas, who argued that Native Americans were free men and deserved the same treatment as others. His opponent, Juan Ginés de Sepúlveda, argued that Native Americans were natural slaves, and that their enslavement was consistent with Catholic theology (Crow 1992). While both of these individuals claimed victory in the debate, the outcome was moot as the decimation of Native American populations had already occurred. The remaining Native Americans were either subsumed into the larger Spanish culture or fled into isolated regions in Central America.

Violence and slavery were not unknown to the indigenous inhabitants of Central America prior to the arrival of Europeans. Based on archaeological evidence, it would appear that violence was endemic to the region prior to contact and intertribal warfare and human sacrifice were common (Fowler 1989). However, the level of violence that occurred with the arrival of the Spanish, along with new diseases, rapidly depopulated the region. The well documented smallpox epidemic of 1520-1527 spread from Mexico to the Andes and is estimated to have killed more than a million people (Crosby 1972). The historic events of Nahua migrations from central Mexico around 1,200 YBP and the arrival of Europeans help explain the

introduction of Mesoamerican haplogroups and European Y-chromosome lineages into northern Votic-speaking Chibchan populations. However, the Isthmic speaking groups have lower Y-chromosome diversity values and cluster closer to South American populations in the MDS plot (figure 26). These diversity values are still higher than their maternal counterparts.

Another potential explanation for this observed dichotomy between low mtDNA and high Y-chromosome diversity is that ethnographic information indicates that Chibchan populations historically practiced matrilocal residence patterns (Reichel-Dolmatoff 1950; Kolman and Bermingham 1997). Almost all of the available comparative data demonstrates low mtDNA haplotype diversity values when compared to Y-chromosome values for Chibchan populations. The only population to show a lower Y-chromosome (0.678) to mtDNA diversity (0.81) value is the Guaymi-Oso, and this may be attributed to the small sample size (n = 8) or that this population practice a patrilocal residence pattern. In the other This matrilocal resident cultural practice is demonstrated by these genetic data which demonstrates higher gene flow in males than in females in the majority of Chibchan populations. An alternative explanation is that the Guaymi/Ngöbé practiced patrilocality and this mtDNA/Y-chromosome relationship is due to this cultural practice (Kolman and Bermingham 1997). However, the other Guaymi-Abrojo population demonstrates high Y-chromosome diversity values, which suggests more population movement by males. An interesting observation from these data is that it does not appear that males move between Chibchan populations. Whereas there are shared mtDNA matrilineages

between Chibchan populations, there do not appear to be shared patrilineages among groups. This indicates limited male gene flow between groups. This is also consistent with previous research on Chibchan population Y-chromosomes, where an exact test of differentiation demonstrated that Y-chromosome haplotypes were not randomly distributed among Chibchan populations (Ruiz-Narvaez *et al.* 2005). This difference in male gene flow versus genetic drift between the Votic and Isthmic Chibchan groups may be explained by the isolation of these southern Isthmic speaking communities. The Cabecar and Bribri inhabit the Talamanca mountain range in southern Costa Rica and the Teribe are an isolated group located along the Caribbean coast of Panama. These groups are known to have revolted against the Spanish several times after contact and maintained some degree of isolation from European populations in the region into the 20th century (Herlihy 1997). However, all Chibchan populations appear to be highly impacted by maternal genetic drift and this demonstrated through shared matrilocal residence pattern.

GENETIC DRIFT VS. GENE FLOW

As mentioned previously there appears to be a significant difference in the evolutionary forces operating on mtDNA and the Y-chromosome in Chibchan populations. MtDNA analysis supports the idea that genetic drift is the primary evolutionary force operating on Chibchan populations. Whereas the Y-chromosome analysis supports the idea that the northern speaking Votic-speaking groups have

experienced higher levels of admixture with both Mesoamerican and European populations than southern Isthmic groups.

The heterozygosity (gene diversity) versus r_{ii} plots (figures 17, 25 and 27) were used to examine the interactions between genetic drift and gene flow and their effect on genetic markers. At the mtDNA locus (figure 17) all but two (Guaymi, Zapáton Huetar) of the Chibchan populations fall below the theoretical regression line, indicating that genetic drift is operating on the majority of Chibchan populations. A similar pattern is also observed for classical genetic markers (figure 25) with only two Chibchan (Pech, Boruca) populations above the regression line. For the Y-chromosome, four Chibchan populations (Zapáton Huetar, Huetar, Rama, and Guaymi-Abrojo) appear above the regression line whereas the other Chibchan populations appear below the regression line. Three of these four populations belong to the Votic-speaking linguistic stock and given their geographic location this is indicative of great gene flow due to the two previously discussed demographic events.

Further evidence for genetic drift in Chibchan population maternal lineages is observed in positive neutrality test scores for Fu's FS (average = 2.76 ± 1.73) and low haplotype diversity (average = 0.598 ± 0.217) values. The only other Central American population to contain a positive Fu's FS value is the Chorotega, and this may be representative of an earlier female Chibchan population in the region. The Chorotega share a number of mtDNA haplotypes (figure 14) with other Chibchan

populations including, 16189 (12) and 16360 (5) both of which are frequent in other Central and South American Chibchan populations. Genetic drift is most pronounced in the Rama, Maléku, and Ijka Chibchan populations. Both the Ijka (90%) and Maléku (92%) are almost fixed for mtDNA haplogroup A2 and the Rama (92%) for B2. The most applicable explanation for the observed type of genetic drift is that these populations underwent a genetic bottleneck within the last 10,000 years followed by subsequent fission along maternal lineages.

Evidence for this Chibchan population bottleneck is seen in the median joining network for mtDNA haplogroup A2 (figure 14). The majority of Chibchan populations are characterized by large nodes that are shared between a large number of individuals within Chibchan groups, whereas, other Central American populations (K'iche Maya, Emberá) are characterized by a large number of singletons indicative of single groups. The four major Chibchan nodes (16189, 16360, 16187, and 16129) are all one mutational unit from the founding Native American haplogroup. Based on coalescent dates these four nodes range from 10,967 ya(16360) to 2,346 ya (16129). All of these dates are consistent with other reported dates for Chibchan populations that place the genetic origins of Chibchan populations within the last 10,000 years (Barrantes *et al.* 1990; Santos *et al.* 1994; Batista *et al.* 1995; Kolman *et al.* 1995; Kolman and Bermingham 1997).

An argument against this genetic drift interpretation is that, due to the severe depopulation of the region that occurred after European contact, genetic variation

within these populations was lost and therefore does not show up in the observed mtDNA diversity. However, all Native American populations experienced contact with Europeans and not all of these populations demonstrate this reduced haplotype diversity. In fact the populations neighboring these groups to the north (K'iche Maya) and the south (Emberá and Wounan) also demonstrate signatures associated with expanding populations and include statistically significant negative neutrality test statistics and high mtDNA haplotype diversity values. There is the possibility that the reduced genetic diversity in these populations is the result of natural selection operating on the mtDNA genome in Chibchan populations. However, this argument may be ruled out as the HVS-I region is considered to be non-coding and therefore it is unlikely that it would be impacted by selection. If natural selection was operating on Chibchan groups it would have to be demonstrated in the coding region, which may be determined from full mtDNA genome sequencing. The surrounding populations (K'iche Maya, Emberá, and Wounan) inhabit similar geographic regions and ecosystems, indicating that if selective pressure were acting on Chibchan populations it should also be reflected in K'iche Maya and Chocoan mtDNA diversity. A final argument that can be made regarding the idea of genetic drift is that neutrality test statistics, such as Tajima's D, may not be sufficiently robust enough to detect evolutionary events that occurred within the last 10,000 years because they are based on rare variation and number of segregating sites and do not account for mutation rate heterogeneity (Aris-Brousou and Excoffier 1996). However, the less conservative Fu's Fs statistics presented here does take into account mutation rate

heterogeneity and provides similar results, further supporting the idea of a genetic bottleneck occurring in Chibchan populations. Therefore, the question becomes, if a population bottleneck did occur in Chibchan population and not in neighboring populations, what may have precipitated it?

PHYLOGEOGRAPHY, PALEOECOLOGY, AND CHIBCHAN ORIGINS

Recent research regarding paleoenvironmental reconstructions has demonstrated that Central and northern South America were cooler and drier during the Pleistocene than during the Holocene (Cooke 2005). This Pleistocene environment was characterized by open grasslands and dry xeric (sagebrush) habitats that were favorable to megafauna and hunter-gatherer populations (Lynch 1983). Climatic oscillations during the Younger Dryas (10,000 to 8,500 YBP) created a warmer, moister environment and a rise in sea level which may have submerged coastal areas (Cooke 2005). Elevated ocean levels in the Caribbean are estimated to have occurred twice (12,300 and 8,800 YBP) and resulted a rise in sea water of 30 meters each time (Fairbanks 1989). This rise in sea level also would have covered earlier human occupation in the coastal regions of Central America (Cooke 2005). This time period also corresponds to the formation of swamps in Caribbean Nicaragua. Urquhart (1997) obtained pollen cores from eastern Nicaragua and determined the earliest formation of swamps occurred between 8,000 and 7,800 YBP. These swamps were composed of the Paurotis palm (*Acoelorraphe wrightii*) and fern species *Blechnum serrulatum*. Both of these plants are typical of deep water and

indicate Caribbean around Lake Nicaragua was largely a swampy area. These species were then replaced in the pollen record between 7,800 and 5,000 YBP by *Myrica Mexicana*, a plant that grows on floating vegetation hammocks in swamps more than a meter in depth (Phillips 1995). After 5,000 YBP, the sea level again decreased and then settled at its current level around 3,800 YBP. Other paleoecological evidence suggests that lower Central America was isolated from Mesoamerica as Lake Nicaragua was flooded around 7,500 YBP, creating an interoceanic corridor between the Atlantic and Pacific (J. Hoopes pers. comm.). This Lake Nicaragua region corresponds to the genetic barrier (figure 18) identified from mtDNA HVS-I sequences and is apparent in the interpolated genetic landscape (figure 19). The proposed paleoecology pollen core dates also overlap with previous genetic dates suggesting Chibchan populations originated between 14,000 and 8,000 ya (Barrantes *et al.* 1990; Kolman and Bermingham 1997; Melton *et al.* 2007).

There is also archaeological evidence that human environmental disturbance intensified and a subsistence shift to horticulture took place during this time period (Cooke 2005). Three archaeological sites from central Panama (Carabalí, Cueva de los Vampiros, and Aguadulce) demonstrate the presence of four domesticated plants: bottle gourd (*Lagenaria siceraria*); arrowroot (*Maranta arundinacea*); lerén (*Calathe allouia*); and squash (*Cucurbita moschata*) date to between 9,000–7,000 YBP (Cooke 2005). Domesticated arrowroot has also been found in highland Colombia and dates to 10,000 to 9,000 YBP (Bray 2000). In addition maize (*Zea mays*) and manioc (*Manihot esculenta*) have been found in found in archaeological sites from Pacific

and Caribbean Panama that date to between 7,800 and 5,400 YBP (Dickau et al. 2007). Cultural complexity also began to occur in the region around 6,000 YBP, and greater interaction among groups occurred (Cooke 2005). This time period (~5,000 ya) overlaps when glottochronological evidence indicates that Chibchan linguistic stocks diverged from one another (Constenla 1991). However, there are significant differences for coalescent dates for population divergence within Votic-speaking groups, indicating greater temporal differentiation between these groups.

The congruence of biological, paleoenvironmental, and archaeological data indicate that Chibchan populations first diverged from earlier Paleoindian populations in response to ecological changes that began during the Pleistocene/Holocene boundaries. These climatic changes resulted in a shift in subsistence strategies that transitioned these groups from hunter-gatherers to plant domesticators. These Pleistocene/Holocene climate changes led to a number of microenvironments. With subsequent geographic isolation, people began to alter their environment, which led to reduced genetic diversity and movement among populations in the region until 5,000 YBP.

The resulting subsistence shift to horticulture is also observable in the mtDNA evidence presented here for Chibchan populations. The majority of Chibchan populations are characterized by low mtDNA haplotype diversity values when compared to neighboring Central American and South American populations. Chibchan populations are also characterized by very few nodes in the median joining networks (figures 14 and 15) where Chibchan groups are composed of one or two

major haplotypes per population along with a few singletons. These biological characteristics are indicative of a population that has undergone a genetic bottleneck within the last 10,000 years.

The timing of this genetic bottleneck corresponds to climatic changes (occurring in Central America) at the time and a shift in subsistence strategies from hunter-gathering to plant domestication is observed in the archaeological record. This would suggest that Chibchan-speaking populations diverged during this transition to agriculture, which occurred within Central America. The earliest evidence for this subsistence transition is found in the highland regions of Central Panama. While there is also evidence of population continuity in the central highlands of Costa Rica, this region lacks vegetative support for an agricultural transition (Cooke 2005). During the time period from 7,000 to 5,000 YBP, archaeological evidence indicate human activity along the Pacific and Caribbean coasts intensified during the time period between 5,000 and 2,500 YBP. After 2,500 YBP, greater regional specialization occurs in the archaeological record, but a number of shared characteristics in ceramic motifs are apparent across regions (Cooke 2005, Hoopes 2005). There is clear archaeological evidence that humans expanded lower Central America into South America after the adoption of plant domestication and this expansion corresponds well with both glottochronological (Constenla 1991; 1995) and coalescent divergence time presented in this study between Chibchan linguistic stocks. These temporal estimates provide further evidence for the diaspora of Chibchan populations and indicate that the “First Farmer” hypothesis (Bellwood 2005) is applicable to these

data. The genetic data appears to fit the assumptions of this hypothesis. There appears to be central area where genetic diversity is higher (in this case Panama or Costa Rica) and the lowest values on the periphery of the Chibchan linguistic area, the earliest presence of domestication in the region also appears to have occurred in Panama or Costa Rica. However, the specific cultigens may have been domesticated outside the isthmus so how they arrived either through trade or population movement needs further research. In addition, climatic changes occurred during this time period which may have necessitated this cultural change. As human populations were forced into smaller geographic ranges and as areas became more forested, changes in subsistence patterns occurred. An effect of agriculture is population growth. As the sea level decreased these populations expanded from lower Central America into new territories along the Caribbean coast. However, further evidence from linguistics, genetics, and archaeology is necessary to frame this idea in a broader anthropological/historical context that includes northern South America as well as uninvestigated Chibchan areas in Nicaragua and Honduras.

Of the four models proposed for the divergences of Chibchan populations, the genetic data presented in this study offers further biological evidence for endogenous development in lower Central America. The genetic evidence from mtDNA provides the best evidence for a Central American divergence from earlier human inhabitants in the region. The highest haplotype diversity values are observed in Isthmic populations (with the exception of the Huetar) and then decrease in both the Votic and Magdalenic speaking groups. In order for higher diversity values to be observed

in these Isthmic groups, greater time depth is needed. The lower diversity values on the periphery of the Votic and Magdalenic groups can then be interpreted as the expansion and diversification of Chibchan-speakers. A similar pattern is also observed in the linguistic structure of Chibchan populations. More languages are spoken among the Isthmic-stock than in both the Votic and Magdalenic groups (Constenla 1991; 1995). Diversity from the Y-chromosome presents a different picture, with higher diversity levels being present in Votic-speakers groups than in the Isthmic groups. However, given the higher amount of European admixture and the closer relationship of Votic groups to Mesoamerican populations, the Y-chromosome data may be more reflective of recent historical events that have impacted the male genetic structure of Chibchan populations.

SUMMARY

This research cannot determine the number of pre-Columbian migrations into South America. However, these molecular genetic data are suggestive of a diaspora of Chibchan-speaking groups into the northern South American prior to European contact. The northern South American Chibchan populations cluster with lower Central American Chibchan and Mesoamerican groups and not with other South American populations. This is demonstrated in both the HVS-I MDS Plot (figure 16) and the SAMOVA (figure 18) that show a close genetic affinity for these populations to one another are suggestive of shared biological history. The Y-chromosome results presented here appear to differ and suggest a intermediate paternal relationship between Chibchan, Mesoamerican and South American groups. There appears to be a

greater biological relationship between Votic-speaking Chibchan groups and Mesoamerican populations whereas Isthmic groups cluster with indigenous South American populations. There is a little data regarding the relationship of northern South American Chibchan groups to Central American populations but it appears in these communities males move more frequently than females.

VII: CONCLUSION

A number of recent studies have investigated the role of Chibchan populations in the peopling of the Americas due to their contemporary geographic location bridging the two American continents. This previous research has rejected the idea of lower Central America as intermediate between the complex cultures of Mesoamerica and the Andes and instead has concluded that the region is its own major cultural area. However, recent research has suggested a biological connection among the K'iche Maya and Chibchan populations, suggesting further evidence was needed in order to evaluate this association.

This dissertation characterized mitochondrial and Y-chromosome genetic diversity in five (Rama, Maléku, Zapáton Huetar, Chorotega, and Guaymi) indigenous populations from Costa Rica and Nicaragua in order to address the origins of the Chibchan language family and the biological relationship of these groups to each other. The specific questions asked were:

1) What evolutionary forces have previously operated on Chibchan populations? How have they impacted these groups? and Do they differ between mtDNA and the Y-chromosome?

2) What is the maternal/paternal genetic relationships of Chibchan populations with neighboring Central and South American indigenous groups?

3) What can mtDNA and Y-chromosome variation reveal regarding the genetic history of Chibchan populations?

4) Are discontinuities observable in the genetic variation of Chibchan and neighboring populations and if they are present how can they be related to the cultural and genetic history of the region.

This study demonstrated genetic similarities between Chibchan populations presented in this dissertation and elsewhere (Barrantes *et al.* 1990; Bieber *et al.* 1994; Torroni *et al.* 1994b; Batista *et al.* 1995; Kolman *et al.* 1995; Kolman and Bermingham 1997; Melton *et al.* 2007) suggesting an endogenous development of Chibchan populations. This model proposed Chibchan populations diverged from earlier hunter-gathering populations in lower Central America. This divergence occurred due to climatic changes that occurred between 10,000 and 8,500 YBP, which resulted in higher sea levels. This rise in sea level created an inter-oceanic corridor between Atlantic and the Pacific around modern Lake Nicaragua that isolated human populations in the region. This geographic isolation resulted in a genetic bottleneck in Chibchan populations. After population size increased, a subsequent expansion of these groups occurred along with the spread of agriculture along the Pacific and Caribbean coastline.

Relationships among Mesoamerican and Chibchan populations are also apparent based on both mtDNA and Y-chromosome data. Two demographic events impacted populations from Mesoamerica and the Isthmo-Colombian area. The first event occurred with the arrival of Mesoamerican populations from Central Mexico that migrated along the Pacific coast into Nicaragua and Costa Rica. The second

event occurred with the arrival of Europeans in the region early in the sixteenth century. This latter contact dramatically altered the cultural landscape in the region as the indigenous population was decimated due to violence, disease, and slavery with an intense effect on the male populations. The impact of European and Mesoamerican contact appears to have been greater in northern Chibchan groups than in southern populations. There is also a difference in evolutionary forces operating on the K'iche Maya and Chibchan populations. Genetic drift has had a greater impact on matrilineages in Chibchan populations and the Chorotega whereas the K'iche Maya are characterized by high diversity values representative of gene flow. Y-chromosome data indicated gene flow is more apparent in Votic-speaking Chibchan and Mesoamerican groups while the Isthmic-speaking populations are characterized by genetic drift. However, there is a general lacunae of available biological evidence for Mesoamerican and Chibchan populations inhabiting Nicaragua and Honduras, making their relationship to neighboring groups difficult to assess. This is unfortunate, as indigenous groups from these regions represented an area of increasing Pre-Columbian cultural complexity and interactions that would provide greater understanding of both pre- and post- European contact in the Americas.

Indigenous populations belonging to the Chibchan linguistic family demonstrate a relationship between genetics and language and are useful for determining biological relationships in Central America. This biological relationship is augmented by numerous shared Chibchan cultural characteristics However, this interpretation of Chibchan gene language relationship is not without reservations. The

first and most important of these is that the measures anthropological geneticists use are not always compatible with archaeological assemblages and linguistic classifications. Neither culture or language are determined by genes. The second caution is that European contact has obviously altered the genetic structure of the extant populations in the region. As globalization continues, this trend is likely to continue. Gene flow from African, Asian, and European populations decreases observable biological relationships among living indigenous populations and decreases genetic variation between these populations. Finally, it is difficult to determine the effects that the extinctions of people and languages in the region had on the resulting phylogenetic relationships as we know little about the biological makeup of extinct populations, like the Cueva, in the area. However, if these cautions are taken into account and the application of the term "Chibchan" is made heuristically, it is a useful function for investigating shared cultural, historical and biological characteristics between lower Central and northern South America.

LITERATURE CITED

Achilli A, Perego UA, Bravi CM, Coble MD, Kong Q-P, Woodward SR, Salas A, Torroni A, and Bandelt H-Jr. 2008. The Phylogeny of the Four Pan-American MtDNA Haplogroups: Implications for Evolutionary and Disease Studies. PLoS ONE 3(3):e1764.

Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, and Young IG (1981) Sequence and organization of the human mitochondrial genome. Nature *290:*457-65.

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, and Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet *23:*147.

Anglería, PM (1976) *Nicaragua en los cronistas de Indias*. Serie Cronistas 1. Banco de América, Managua.

Aris-Brosou S, and Excoffier L. (1996). The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism. Mol Biol Evol 13(3):494-504.

Arnaiz-Villena A, Vargas-Alarcon G, Granados J, Gomez-Casado E, Longas J, Gonzales-Hevilla M, Zuniga J, Salgado N, Hernandez-Pacheco G, Guillen J, and Martinez-Laso J (2000) HLA genes in Mexican Mazatecans, the peopling of the Americas and the uniqueness of Amerindians. Tissue Antigens *56:*405-16.

Ayub Q, Mohyuddin A, Qamar R, Mazhar K, Zerjal T, and Tyler-Smith C (2000) Identification and characterization of novel human Y-chromosomal microsatellites from sequence database information. Nucl. Acids Res. 28:2-8

Azofeifa J, Ruiz E, and Barrantes R (1998) Genetic variation and racial admixture in the Miskito of the southern Mosquito Shore, Nicaragua. Revista De Biologia Tropical *46:*157-165.

Azofeifa J, Ruiz E, and Barrantes R (2001) Blood group, red cell, and serum protein variation in the Cabecar and Huetar, two Chibchan Amerindian tribes of Costa Rica. American Journal of Human Biology *13:*57-64.

Bailliet G, Rothhammer F, Carnese FR, Bravi CM, and Bianchi NO. 1994. Founder mitochondrial haplotypes in Amerindian populations. Am J Hum Genet 55(1):27-33.

Baldi, N. (2001). Black Creek (Cat. UCR 467): primeras interpretaciones de un modo de vida costero en el Caribe sur de Costa Rica para el segundo milenio antes de Cristo. Unpublished Licenciatura thesis, Universidad de Costa Rica, San Jose.

Bandelt HJ, Forster P, Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37–48.

Bandelt HJ, Forster P, Sykes BC, Richards MB. (1995). Mitochondrial portraits of human populations using median networks. Genetics 141:743–753.

Barbujani G, and Sokal RR. (1990). Zones of sharp genetic change in Europe are also linguistic boundaries. Proc Natl Acad Sci U S A 87(5):1816-1819.

Barrantes R (1993) Genetic Diversity and Racial Mixture in Amerindians from Costa-Rica and Panama. Revista De Biologia Tropical *41:*379-384.

Barrantes R, Smouse PE, Mohrenweiser HW, Gershowitz H, Azofeifa J, Arias TD, and Neel JV (1990) Microevolution in lower Central America: genetic characterization of the Chibcha-speaking groups of Costa Rica and Panama, and a consensus taxonomy based on genetic and linguistic affinity. Am J Hum Genet *46:*63-84.

Barrantes R, Smouse PE, Neel JV, Mohrenweiser HW, and Gershowitz H (1982) Migration and Genetic Infrastructure of the Central American Guaymi and Their Affinities with Other Tribal Groups. American Journal of Physical Anthropology *58:*201-214.

Barrot C, Simili C, Sanchez C, Brandt-Casadevall C, Gonzalez-Martin A, Xifro A, Ortega M, Huguet E, Corbella J, and Gene M (2007) Haplotype frequencies of eight Y-chromosome short tandem repeats loci in four Amerindian populations (state of Hidalgo, Mexico). J Forensic Sci *52:*504-6.

Bateman, R., I. Goddard, R. O'Grady, V. Funk, R. Mooi,W. Kress, and P. Cannell. (1990). Speaking of forked tongues: The feasibility of reconciling human phylogeny and the history of language. Curr. Anthropol. 31:1–24.

Batista O, Kolman CJ, and Bermingham E (1995) Mitochondrial DNA diversity in the Kuna Amerinds of Panama. Hum Mol Genet *4:*921-9.

Beekman CS, and Christensen AF (2003) Controlling for doubt and uncertainty through multiple lines of evidence: A new look at the Mesoamerican Nahua migrations. Journal of Archaeological Method and Theory *10:*111-164.

Begley CT (1999) Elite Power Strategies and External Connections in Eastern Honduras, University of Chicago, Chicago.

Bellwood PS (2005) The first farmers : the origins of agricultural societies. Malden, MA: Blackwell Pub.

Bieber H, Bieber SW, Rodewald A, and Barrantes R (1996) Microevolution and genetic affinities among six Amerindian tribes of lower Central America: Comparative genetic study of serum proteins. Human Biology *68:*929-953.

Black FL. (1991). Reasons for failure of genetic classifications of South Amerind populations. Hum Biol 63(6):763-774

Boles TC, Snow CC, and Stover E (1995) Forensic DNA Testing on Skeletal Remains from Mass Graves - a Pilot Project in Guatemala. Journal of Forensic Sciences *40:*349-355.

Bonatto SL, and Salzano FM. (1997). A single and early migration for the peopling of the Americas supported by mitochondrial DNA sequence data. Proc Natl Acad Sci U S A 94(5):1866-1871.

Bonnichsen R and Turnmire K (1992) Clovis: Origins and Adaptations. Center for the Study of the First Americans, Corvallis, Oregon.

Bortolini MC, Salzano FM, Thomas MG, Stuart S, Nasanen SP, *et al.* (2003) Y-chromosome evidence for differing ancient demographic histories in the Americas. Am J Hum Genet *73:*524-39.

Bosch E, Lee AC, Calafell F, Arroyo E, Henneman P, de Knijff P, and Jobling MA (2002) High resolution Y chromosome typing: 19 STRs amplified in three multiplex reactions. Forensic Sci Int *125:*42-51.

Bray W. (2000). Archaeology. Ancient food for thought. Nature 408:145–146.

Bray W. (1984). Across the Darien Gap: A Colombian view of Isthmian Archaeology. In: Lange FW, and Stone DZ, editors. The Archaeology of Lower Central America. Albuqueque: University of New Mexico Press. p 305-340.

Briceño, I, Sánchez, MC, Silva, A., Jara, P., Torres D., Villega V, Burgos J., Rodríguez, J., Bernal, JE., Gómez, A (2008) Mitochondrial DNA analysis in pre-Columbian archaic bones from two periods; belonging to the region of Bogotá, Colombia. II Congreso Latinoamericano de Antropología, San Jose, Costa Rica

Briceño I, Bernal J, Duran S, Ortega J, Umana A, Gomez A, Papiha S. (1996). HLA antigens in Amerindian groups of two different linguistic families from Colombia. Eur J Immunogenet 23:21–28.

Brignoli HP (2005) La dinámica demográfica de las poblaciones indígenas del tropic húmedo en América Central (censos del 2000). IUSSP XXVth International Population Conference.

Brouns G, De Wulf A, Constales D, (2003). Delaunay triangulation algorithms useful for multibeam echosounding. J Surv Eng 129:79–84.

Brown M, Hosseini S, Torroni A, Bandelt H, Allen J, Schurr T, Scozzari R, Cruciani F and Wallace D (1998) MtDNA haplogroup X: An ancient link between Europe/West Asia and North America? American Journal of Human Genetics 63:1852-1861.

Brown WM, George M, Wilson AC (1979) Rapid Evolution of Animal Mitochondrial-DNA. Proceedings of the National Academy of Sciences of the United States of America *76:*1967-1971.

Buckler, E., D. Pearsall, and T. Holtsford. (1998). Climate, plant ecology, and Central Mexican Archaic subsistence. Curr. Anthrop. 39:152-164.

Campbell L (1997) American Indian Languages: The historical languages of the Americas. New York: Oxford University Press.

Campbell L. (1988). [Review of *Language in the Americas*, Greenberg 1987]. Language, *64*, 591-615.

Campbell L (1979) Middle American languages. In L Campbell and M Mithun (eds.): The Languages of Native America: Historical and Comparative Assessment. Austin: University of Texas Press, pp. 902-100.

Campbell L, and Kaufman T (1976) A Linguistic Look at the Olmec. American Antiquity *41:*80-89.

Campell L, Kaufman T, and Stark TS (1976) Meso-America as a linguistic area. Language *62:*530-558.

Campos-Sanchez R, Barrantes R, Silva S, Escamilla M, Ontiveros A, Nicolini H, Mendoza R, Munoz R, and Raventos H (2006) Genetic structure analysis of three Hispanic populations from Costa Rica, Mexico, and the southwestern United States using Y-chromosome STR markers and mtDNA sequences. Hum Biol *78:*551-63.

Cann RL. (1994). mtDNA and Native Americans: a Southern perspective. Am J Hum Genet 55(1):7-11.

Cann RL, Stoneking M, and Wilson AC (1987) Mitochondrial DNA and human evolution. Nature *325:*31-6.

Carmack, R.M. (1981) The Quiché Mayas of Utatlán: The Evolution of a Highland Guatemala Kingdom, Civilization of the American Indian series, no. 155. Norman: University of Oklahoma Press.

Carvajal-Carmona L, Soto I, Pineda N, *et al.* (2000) Strong Amerind/white sex bias and a possible sephardic contribution among the founders of a population in northwest Columbia. American Journal of Human Genetics 67:281-286.

Casas, Bartolomé de las (2007) A Brief Account of the Destruction of the Indies Or, a faithful NARRATIVE OF THE Horrid and Unexampled Massacres, Butcheries, and all manner of Cruelties, that Hell and Malice could invent, committed by the Popish Spanish Party on the inhabitants of West-India, TOGETHER With the Devastations of several Kingdoms in America by Fire and Sword, for the space of Forty and Two Years, from the time of its first Discovery by them. Translated from Spanish and accessed through Project Gutenburg (http://www.gutenberg.org/etext/20321)

Chafe, Wallace. (1987). [Review of Greenberg 1987]. Current Anthropology, *28*, 652-653

Coates AG (1997) Central America: A Natural and Cultural History. New Haven: Yale University Press.

Coe MD (1996) Mexico: from the Olmecs to the Aztecs. New York: Thomas and Hudson.

Constenla, A. (1995). Sobre el estudio diacronico de las lenguas chibchenses y su contribucion al estudio del pasado de sus hablantes. Boletın del Museo del Oro 38–39: 13–56.

Constenla-Umaña A (1991) Las lenguas del Area intermedia: introducción a su estudio areal. San José, Costa Rica: Editorial de la Universidad de Costa Rica.

Conzemius E, and Smithsonian Institution. Bureau of American Ethnology. (1932). Ethnographical survey of the Miskito and Sumu Indians of Honduras and Nicaragua [microform]. Washington: U.S. G.P.O.,.

Conzemius E. (1930). Une tribe inconnue du Costa-Rica: les Indiens du Rio Zapote. L'Anthropologie 40:93-108.

Cooke R (2005) Prehistory of native Americans on the Central American land bridge: Colonization, dispersal, and divergence. Journal of Archaeological Research *13:*129-187.

Cooke R (1997) The Native Peoples of Central America during Precolumbian and Colonial Times. In AG Coates (ed.): Central America: A Natural and Cultural History. New Haven: Yale University Press, pp. 137-176.

Cook R. (1993) La utilización y domesticación de alimentos autóctonos en la Región Histórica Chibcha-Chocó. *Revista de Arqueología Americana* 6 (July–December 1992).

Cooke R, Isaza I, Griggs J, Desjardins B, and Sánchez LA. (2003) Who Crafted, Exchanged, and Displayed Gold in Pre-Columbian Panama? In: Quilter J, and Hoopes JW, editors. Gold and Power in Ancient Costa Rica, Panama, and Colombia. Washington, D.C.: Dumbarton Oaks. p 91-158

Crawford MH (2007) Genetic structure of circumpolar populations: a synthesis. Am J Hum Biol *19:*203-17.

Crawford MH (1998) The Origins of Native Americans: Evidence from anthropological Genetics. Cambridge, UK: Cambridge University Press.

Crawford MH, Williams JT, and Duggirala R. (1997). Genetic structure of the indigenous populations of Siberia. Am J Phys Anthropol 104(2):177-192.

Crawford MH, and Duggirala R. (1992). Digital dermatoglyphic patterns of Eskimo and Amerindian populations: relationships between geographic, dermatoglyphic, genetic, and linguistic distances. Hum Biol 64(5):683-704

Crosby, A. (1972) The Columbian Exchange: Biological and Cultural Consequences of 1492. Greenwood Press 1972, Praeger Publishers 2003.

Crow, John A. (1992) The Epic of Latin America, 4th ed. University of California Press, Berkeley: 1992

Currat M, and Excoffier L (2005) The effect of the Neolithic expansion on European molecular diversity. Proc Biol Sci *272:*679-88.

De Knijff P, Kayser M, Caglia A, et al. (1997) Chromosome Y microsatellites: population genetics and evolutionary aspects. International Journal of Legal Medicine 110:134-149.

Deka R, Majumder PP, Shriver MD, Stivers DN, Zhong YX, Yu LM, Barrantes R, Yin SJ, Miki T, Hundrieser J, Bunker CH, McGarvey ST, Sakallah S, Ferrell RE, and Chakraborty R (1996) Distribution and evolution of CTG repeats at the myotonin protein kinase gene in human populations. Genome Research *6:*142-154.

Demarchi DA, and Mitchell RJ. (2004). Genetic structure and gene flow in Gran Chaco populations of Argentina: evidence from Y-chromosome markers. Hum Biol 76(3):413-429.

Demarest AA (2004) Ancient Maya : the rise and fall of a rainforest civilization. Cambridge ; New York: Cambridge University Press.

De Stefano G, and Molieri JJ. (1976). P.T.C. tasting among three Indian groups of Nicaragua. Am J Phys Anthropol 44(2):371-373.

Diamond J, and Bellwood P (2003) Farmers and their languages: the first expansions. Science *300:*597-603.

Dickau R, Ranere AJ, and Cooke RG. (2007) Starch grain evidence for the preceramic dispersals of maize and root crops into tropical dry and humid forests of Panama. P Natl Acad Sci USA 104(9):3651-3656.

Dillehay TD (2000) The settlement of the Americas: a new prehistory. New York: Basic Books.

Dixon EJ (1999) Bones, Boats, and Bison. Albuquerque: University of New Mexico Press.

Dornelles CL, Bonatto SL, De Freitas LB, and Salzano FM (2005) Is haplogroup X present in extant South American Indians? Am J Phys Anthropol *127:*439-48.

Dull R A, Southon J R, Sheets P, (2001). Volcanism, ecology and culture: a reassessment of the Volcan Ilopango TBJ eruption in the southern Maya realm. Latin Amer Antiquity, 12: 25-44

Dupanloup I, Schneider S, and Excoffier L. (2002). A simulated annealing approach to define the genetic structure of populations. Mol Ecol 11(12):2571-2581.

Easton RD, Merriwether DA, Crews DE, and Ferrell RE. (1996). mtDNA variation in the Yanomami: evidence for additional New World founding lineages. Am J Hum Genet 59(1):213-225.

Excoffier L, Laval G, Schneider S. (2005). Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1.

Excoffier L (2004) Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. Mol Ecol *13:*853-64.

Excoffier, L., Smouse, P.E., and Quattro, J.M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131: 479-491.

Excoffier L, Harding RM, Sokal RR, Pellegrini B, and Sanchez-Mazas A. (1991). Spatial differentiation of RH and GM haplotype frequencies in Sub-Saharan Africa and its relation to linguistic affinities. Hum Biol 63(3):273-307.

Fagundes NJ, Bonatto SL, Callegari-Jacques SM, Salzano FM. (2002). Genetic, geographics and linguistic variation among South American Indians: possible sex influence. Am. J. Phys. Anthropol. 117:68-78

Fairbanks RG. (1989). A 17,000-year glacio-eustatic sea level record: influence of glacial melting rates on the Younger Dryas event and deep-ocean circulation. Nature 342(6250):637-642.

Fonseca O. (1998) El Espacio histórico de los amerindios de filiación chibcha: El Área Histórica Chibchoide. In Congreso Científico sobre pueblos indígenas de Costa Rica y sus fronteras (M. E. Bozzoli *et al.*): 36–60. EUNED, San José.

Fonseca O. (1997) La Confirmación de los espacios históricos, el Caso de América Central y el noroccidente colombiano. In III Simposio Panamericano de Historia: 21–357. Instituto Panamericano de Geografía e Historia, México

Fonseca O. (1994) El concepto de area de tradición chibchoide y su pertinencia para entender Gran Nicoya. Vínculos 18: 209–228.

Fonseca O. (1992) Historia antigua de Costa Rica, Colección Historia de Costa Rica.Universidad de Costa Rica. 260 p.

Forster P, Harding R, Torroni A, and Bandelt HJ. (1996). Origin and evolution of Native American mtDNA variation: a reappraisal. Am J Hum Genet 59(4):935-945

Fowler, W. (1989) The Cultural Evolution of Ancient Nahua Civilizations: The Pipil-Nicarao of Central America, University of Oklahoma Press, Norman.

Fox C (1996) Mitochondrial DNA Haplogroups in Four Tribes from Tierra del Fuego-Patagonia: Inferences about the Peopling of the Americas. Hum Biol *68:*855-871.

Francalacci, P., Montiel, R. & Malgosa, A. (1999) A mitochondrial DNA database. In *Genomic Diversity*, ed. S. Papiha, R. Deka & R. Chakraborty, pp.103-19. New York: Kluwer Academic/Plenum Publishers.

Fu YX. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147:915–925.

Fuselli S, Tarazona-Santos E, Dupanloup I, Soto A, Luiselli D, and Pettener D. (2003). Mitochondrial DNA Diversity in South America and the Genetic History of Andean Highlanders. Mol Biol Evol 20(10):1682-1691

Gabb WM (1875). "On the Indian tribes and languages of Costa Rica". *American Philosophical Society Proceedings* 14: 483–602.

Garrigan D, and Hammer MF (2006) Reconstructing human origins in the genomic era. Nat Rev Genet *7:*669-80.

Gassiot, E.; Clemente, I.; Briz, I. López, JA. (2003). El Cascal de Flor de Pino, centro neurálgico de una civilización desconocida en la Costa Atlántica de Nicaragua. Revista de Arqueología*,* .21. 268, pp. 32-37

Giffford JC (1976) Prehistoric Pottery Analysis and the Ceramics of Barton Ramie in the Belize Valley. Cambridge, MA: Harvard University Press.

Gilbert MT, Kivisild T, Gronnow B, Andersen PK, Metspalu E, Reidla M, Tamm E, Axelsson E, Gotherstrom A, Campos PF and others. (2008). Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. Science 320(5884):1787-1789.

Giles RE, Blanc H, Cann HM, and Wallace DC (1980) Maternal Inheritance of Human Mitochondrial-DNA. Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences *77:*6715-6719.

Ginther C, Corach D, Penacino G, Rey J, Carnese F, Hutz M, Anderson A, Just J, Salzano FM, and King M (1993) Genetic variation among the Mapuche Indians from the Patogonian region of Argentina: Mitochondrial DNA sequence variation and allele frequencies of several nuclear genes. In S Pena, R Chakraborty, J Epplen and A Jeffreys (eds.): DNA Fingerprinting: State of the Science. Basel: Birkhäuser-Verlag, pp. 211-219.

Goddard I. (1987). [Review of Joseph Greenberg, *Language in the Americas*]. Current Anthropology, *28*, 656-657.

Gomez-Casado E, Martinez-Laso J, Moscoso J, Zamora J, Martin-Villa M, Perez-Blas M, Lopez-Santalla M, Gramajo PL, Silvera C, Lowy E, and Arnaiz-Villena A (2003) Origin of Mayans according to HLA genes and the uniqueness of Amerindians. Tissue Antigens *61:*425-436.

González SS, and Leiva RV. (2006) Was there a Greater Nicoya subarea during the postclassic. Vínculos 29(1-2):1-16.

Graves J (2002) The rise and fall of SRY. TRENDS in Genetics 18(5):259-264.

Green RE, Krause J, Ptak SE, Briggs AW, Ronan MT, Simons JF, Du L, Egholm M, Rothberg JM, Paunovic M, and Paabo S (2006) Analysis of one million base pairs of Neanderthal DNA. Nature *444:*330-6.

Greenberg JH. 1987. Language in the Americas. Stanford, Calif.: Stanford University Press.

Greenberg JH, Turner CG, and Zegura SL. (1986). The Settlement of the Americas - a Comparison of the Linguistic, Dental, and Genetic-Evidence. Curr Anthropol 27(5):477-497.

Gunn, J. D. (ed.). (2000). *The Years Without Summer: Tracing A.D. 536 and Its Aftermath*, British Archaeological Reports International Series 872, Archaeopress, Oxford.

Hale K, and Salamanca D (2001) Theoretical and Universal Implications of Certain Verbal Entries in Dictionaries of the Misumalpan Languages. In Frawley, EW Hill and Munro (eds.): Making Dictionaries: Preserving indigenous Languages of the Americas. Berkeley: University of California Press.

Hall T. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 41:95–98.

Hammer M and Zegura S (2002) The human Y chromosome haplogroup tree: nomenclature and phylogeny of its major divisions. Annual Review of Anthropology 31:303-321.

Hammer MF, Karafet TM, Redd AJ, Jarjanazi H, Santachiara-Benerecetti S, Soodyall H, and Zegura SL (2001) Hierarchical patterns of global human Y-chromosome diversity. Mol Biol Evol *18:*1189-203.

Hammer MF, and Horai S. (1995). Y chromosomal DNA variation and the peopling of Japan. American Journal of Human Genetics 56: 951-962

Hammer MF. (1994). A recent insertion of an Alu element on the Y chromosome is a useful marker for human population studies. Molecular Biology and Evolution 11: 749-761

Harpending, H., and A. Rogers. (1984). ANTANA: A Package for Multivariate Data Analysis V.1.1. Distributed by the authors.

Harpending, H. C., and R. H.Ward. (1982). Chemical systematics and human populations. In Biochemical Aspects of Evolutionary Biology, M. H. Nitechi, ed. Chicago: University of Chicago Press.

Hartl, DL Clark AG.(1997). Principles of population genetics. 3rd edition. Sunderland (MA): Sinauer Associates

Helgason A, Nicholson G, Stefansson K, and Donnely P (2003) A reassessment of genetic diversity in Icelanders: Strong evidence from multiple loci for relative homogeneity caused by genetic drift. Annals of Human Genetics 67:281-197.

Helms MW (1969) The cultural ecology of a colonial tribe. Ethnology *8:*76-84.

Herlihy PH. 1997. Central American Indian Peoples and Land Today. In: Coates AG, editor. Central America: A cultural and natural history. New Haven, CT: Yale University Press. p 215-240.

Hill JH (2002) Proto-Uto-Aztecan Cultivation and the Northern Devolution. In P Bellwood and C Renfrew (eds.): Examining the Farming/Language Dispersal Hypothesis. Cambridge: McDonald Institute for Archaeological Research, pp. 331-340.

Hoopes J (2005) The emergence of Social Complexity in the Chibchan World of Southern Central America and Northern Colombia, AD 300-600. Journal of World Prehistory *13:*1-32.

Hoopes J, and Fonseca O (2003) Goldwork and Chibchan Identity: Endogenous Change and Diffuse Unity in the Isthmo-Colombian Area. In J Quilter and J Hoopes (eds.): Gold and Power in Ancient Costa Rica, Panama, and Colombia. Washington DC: Dumbarton Oaks, pp. 49-89.

Horai S, Konodo R, Nakagawa-Hottori Y, Hayashi S, Sonoda S, and Tajima K (1993) Peopling of the Americas, founded by four major lineages of mitochondrial DNA. Molecular Biological Evolution 10:23-47.

http://www.fluxus-engineering.com accessed in May 2008

http://www.ysearch.org accessed June 2008

Hunley KL, Spence JE, Merriwether DA. (2008) The Impact of Group Fissions on Genetic Structure of Native South America and Implications for Human Evolution. Am. J. Phys. Anth. 135: 195-205

Hunley KL, Cabana GS, Merriwether DA, and Long JC. (2007). A formal test of linguistic and genetic coevolution in native Central and South America. Am J Phys Anthropol 132(4):622-631.

Ibarra-Rivera L, Mirabal S, Regueiro MM, and Herrera RJ (2007) Delineating genetic relationships among the Maya. Am J Phys Anthropol.

International Human Genome Sequencing Consortium (2001). "Initial sequencing and analysis of the human genome." Nature 409 (6822): 860–921.

International SNP Map Working Group. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409: 928-933

Jara P, Villegas V, C.L. dM, Torres D, Bautista E, Bernal JE, A. G, and Briceño I. (2008). Análisis de polimorfismos del ADN mitocondrial (ADNmt) en un grupo precolombino Muisca. Cuadernos de medicina en investigación y salud 1-2:103-111.

Jobling M, Hurles M, and Tyler-Smith C (2004) Human Evolutionary Genetics: Origins, Peoples & Disease. New York: Garland Publishing.

Jobling MA, and Tyler-Smith C (2003) The human Y chromosome: an evolutionary marker comes of age. Nat Rev Genet *4:*598-612.

Jobling MA, Samara V, Pandya A, Fretwell N, Bernasconi B, Mitchell RJ, Gerelsaikhan T, Dashnyam B, Sanjantila A, Salo PJ, *et al.* (1996) Recurrent duplication and deletion polymorphism on the long arm of the Y chromosome in normal males. Human Molecular Genetics 5: 1767-1775

Johnson F. (1948). Central American Cultures: An introduction. In: Steward J, editor. Handbook of South American Indians: The Circum-Caribbean Tribes. New York: Cooper Aquare. p 43-69

Jones PN (2004) American Indian mtDNA and Y Chromosome Genetic Data: A Comprehensive Report of their Use in Migration and Other Anthropological Studies.

Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, and Hammer MF. (2008). New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. Genome Res 18(5):830-838.

Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, Goldman D, Klitz W, Harihara S, de Knijff P, Wiebe V, Griffiths RC, Templeton AR, and Hammer MF (1999) Ancestral Asian source(s) of new world Y-chromosome founder haplotypes. Am J Hum Genet *64:*817-31.

Karafet T, Zegura SL, Vuturo-Brady J, Posukh O, Osipova L, Wiebe V, Romero F, Long JC, Harihara S, Jin F and others. (1997). Y chromosome markers and Trans-Bering Strait dispersals. Am J Phys Anthropol 102(3):301-314.

Kaufman T. (2001) Nawa linguistic prehistory. published at website of the Mesoamerican Language Documentation Project http://www.albany.edu/anthro/maldp/Nawa.pdf

Kemp BM (2006) Mesoamerica and Southwest prehistory, and the entrance of humans into the Americas: Mitochondrial DNA evidence, Ph.D. dissertation University of California, Davis.

Kemp BM, Malhi RS, McDonough J, Bolnick DA, Eshleman JA, Rickards O, Martinez-Labarga C, Johnson JR, Lorenz JG, Dixon EJ and others. (2007). Genetic analysis of early Holocene skeletal remains from Alaska and its implications for the settlement of the Americas. American Journal of Physical Anthropology 132(4):605-621.

Kemp BM, Reséndez A, Berrelleza JAR, Mahli RS, and Smith DG (2005) An analysis of Ancient Aztec mtDNA from Tlatelocloco: Pre-Columbian Relations and the Spread of Uto-Aztecan. In DM Reed (ed.): Biomolecular Archaeology. Carbondale: Southern Illinois University Press, pp. 22-46.

Keyeux G, Rodas C, Gelvez N, and Carter D. (2002). Possible migration routes into South America deduced from mitochondrial DNA studies in Colombian Amerindian populations. Hum Biol 74(2):211-233.

Kayser M, Caglia A, Corach D, *et al*. (1997) Evaluation of Y-chromosomal STRs: a multicenter study. International Journal of Legal Medicine 110:125-133 (Appendix 141-149).

Kitchen A, Miyamoto MM, and Mulligan CJ. (2008). A Three-Stage Colonization Model for the Peopling of the Americas. PLoS ONE 3(2):e1596.

Kirchoff P. (1943). Mesoamérica. Acta Americana 1(1):92-107

Kivisild T, Rootsi S, Metspalu M, Mastana S, Kaldma K, Parik J, Metspalu E, Adojaan M, Tolk HV, Stepanov V and others. (2003). The Genetic Heritage of the Earliest Settlers Persists Both in Indian Tribal and Caste Populations. The American Journal of Human Genetics 72(2):313-332.

Kivisild T, Tolk HV, Parik J, Wang Y, Papiha SS, Bandelt HJ, and Villems R. (2002). The emerging limbs and twigs of the East Asian mtDNA tree. Mol Biol Evol 19(10):1737-1751.

Kolman CJ, and Bermingham E (1997) Mitochondrial and nuclear DNA diversity in the Choco and Chibcha Amerinds of Panama. Genetics *147:*1289-302.

Kolman C, Sambuughin N, Bermingham E (1996) Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders. Genetics 142:1321-1334.

Kolman CJ, Bermingham E, Cooke R, Ward RH, Arias TD, and Guionneau-Sinclair F (1995) Reduced mtDNA diversity in the Ngöbé Amerinds of Panama. Genetics *140:*275-83.

Krings M, Stone A, Schmitz RW, Krainitzki H, Stoneking M, and Paabo S (1997) Neanderthal DNA sequences and the origin of modern humans. Cell *90:*19-30.

Kruskal J. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika 29: 1–27.

Lalueza C, Perez-Perez A, Prats E, Cornudella L, and Turbon D. (1997). Lack of founding Amerindian mitochondrial DNA lineages in extinct aborigines from Tierra del Fuego-Patagonia. Hum Mol Genet 6(1):41-46.

Lalueza-Fox C, Calderon FL, Calafell F, Morera B, and Bertranpetit J (2001) MtDNA from extinct Tainos and the peopling of the Caribbean. Ann Hum Genet *65:*137-51.

Lalueza-Fox C, Gilbert MT, Martinez-Fuentes AJ, Calafell F, and Bertranpetit J (2003) Mitochondrial DNA from pre-Columbian Ciboneys from Cuba and the prehistoric colonization of the Caribbean. Am J Phys Anthropol *121:*97-108.

Lange FW, and Stone DZ. (1984). The Archaeology of Lower Central America. Albuquerque: University of New Mexico Press

Langeback CH. (2003). The Political Economy of Pre-Colombian Goldwork: Four Examples from Northern South America. In: Quilter J, and Hoopes JW, editors. Gold and Power in Ancient Costa Rica, Panama, and Colombia. Washington, D.C.: Dumbarton Oaks. p 245-278.

Lampl M, and Blumberg BS (1979) Blood Polymorphisms and the Origin of New World Populations. In AB Harper (ed.): The First Americans: Origins, Affinities and Adaptations. New York: Gustav Fischer.

Lathrap DW. 1970. The Upper Amazon. London,: Thames & Hudson.

Lavalleé.D. 2000. The First South Americans: The Peopling of a continent from the earliest evidence to High Culture. Salt Lake City: University of Utah Press.

Layrisse M, and Wilbert J (1961) Absence of the Diego antigen, a genetic characteristic of early immigrants to South America. Science *134:*1077-8.

Layrisse Z., Layrisse M,, Rodríguez-Larraide A. (1995) Estudios genéticos en poblaciones de lengua chibcha. Boletín del Museo del Oro 38–39: 71–86.

Lell JT, Sukernik RI, Starikovskaya YB, Su B, Jin L, Schurr TG, Underhill PA, and Wallace DC. (2002). The dual origin and Siberian affinities of Native American Y chromosomes. American Journal of Human Genetics 70: 192-206

Lell JT, Brown MD, Schurr TG, Sukernik RI, Starikovskaya YB, Torroni A, Moore LG, Troup GM, and Wallace DC. (1997). Y chromosome polymorphisms in native American and Siberian populations: identification of native American Y chromosome haplotypes. Hum Genet 100(5-6):536-543.

Lewis CM, Jr., and Long JC. (2008). Native South American genetic structure and prehistory inferred from hierarchical modeling of mtDNA. Mol Biol Evol 25(3):478-486.

Lewis CM, Jr., Lizarraga B, Tito RY, Lopez PW, Iannacone GC, Medina A, Martinez R, Polo SI, De La Cruz AF, Caceres AM and others. (2007). Mitochondrial DNA and the peopling of South America. Hum Biol 79(2):159-178.

Lewis CM, Jr., Tito RY, Lizarraga B, and Stone AC. (2005). Land, language, and loci: mtDNA in Native Americans and the genetic history of Peru. Am J Phys Anthropol 127(3):351-360.

Linares, O. F., Sheets, P. D., and Rosenthal, E. J. (1975). Prehistoric agriculture in tropical highlands. Science 187: 137–145.

Loogvali EL, Roostalu U, Malyarchuk BA, Derenko MV, Kivisild T, Metspalu E, Tambets K, Reidla M, Tolk HV, Parik J and others. (2004). Disuniting

uniformity: a pied cladistic canvas of mtDNA haplogroup H in Eurasia. Mol Biol Evol 21(11):2012-2021.

Lorenz JG, and Smith DG (1996) Distribution of four founding mtDNA haplogroups among Native North Americans. Am J Phys Anthropol *101:*307-23.

Lothrop SK (1942) The Sigua: Southernmost Aztec Outpost. Proceedings of the 8th American Scientific Congress, pp. 109-116. Washington, D.C.

Lothrop SK. (1926). Pottery of Costa Rica and Nicaragua. New York,: Museum of the American Indian, Heye Foundation. 2 v. p.

Lovo-Gomez J, Blanco-Verea A, Lareu MV, Brion A, and Carracedo A (2007a) The genetic male legacy from El Salvador. Forensic Science International *171:*198-203.

Luiselli D, Simoni L, Tarazona-Santos E, Pastor S, and Pettener D. (2000) Genetic structure of Quechua-speakers of the Central Andes and geographic patterns of gene frequencies in South Amerindian populations. Am J Phys Anthropol 113(1):5-17

Lynch TF. (1983). The Paleo-Indians. In: Jennings JD, editor. Ancient South Americans. South America: Freeman. p 87–137.

MacNeish RS. (1992). The origins of agriculture and settled life. Norman: University of Oklahoma Press.

Magnus RW. (1978). The prehistoric and modern subsistence patterns of the Atlantic coast of Nicaragua: a comparison. In: Stark BL, and Voorhies B, editors. Prehistoric coastal adaptations: the economy and ecology of maritime middle America. New York: Academic Press. p 61-80.

Malhi R, Gonzalez-Oliver A, Schroeder KB, Kemp BM, Greenberg JA, Dobrowski S, Smith D, Resendez A, Karafet TM, Hammer MF and others. (2008). Distribution of Y chromosomes among native North Americans: A study of Athapaskan population history. American Journal of Physical Anthropology . early view

Manni F, Guerard E, and Heyer E. (2004). Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by using Monmonier's algorithm. Hum Biol 76(2):173-190.

Martin S, Grube N (2000) Chronicle of the Maya Kings and Queens: Deciphering the Dynasties of the Ancient Maya. Thames and Hudson, London.

Matson GA, and Swanson J (1959) Distribution of Hereditary Blood Antigens among the Maya and non-Maya Indians in Mexico and Guatemala. Am J Phys Anthropol *17:*49-74.

Matson GA, and Swanson J (1963a) Distribution of Blood Groups among Indians in Middle America: Iv. Am J Phys Anthropol *21:*319-33.

Matson GA, and Swanson J (1963b) Distribution of hereditary blood antigens among Indians in Middle America. II. Tzotzil and other Maya. Am J Phys Anthropol *21:*1-14.

Matson GA, and Swanson J (1963c) Distribution of Hereditary Blood Antigens among Indians in Middle America. V. In Nicaragua. Am J Phys Anthropol *21:*545-59.

Matson GA, and Swanson J (1964) Distribution of Hereditary Blood Antigens among Indians in Middle America. Vi. in British Honduras. Am J Phys Anthropol *22:*271-84.

Matson GA, and Swanson J (1965) Distribution of hereditary blood antigens among Indians in Middle America. 8. In Panama. Am J Phys Anthropol *23:*413-26.

Matson GA, Sutton HE, Swanson J, and Robinson AR (1963) Distribution of Haptoglobin, Transferrin, and Hemoglobin Types among Indians of Middle America: Southern Mexico, Guatemala, Honduras and Nicaragua. Hum Biol *35:*474-83.

McCafferty GC. (2008). Domestic practice in Postclassic Santa Isabel, Nicaragua. Latin American Antiquity 19(1):64-82.

Melton PE, Briceño I, Gomez A, Devor EJ, Bernal JE, and Crawford MH (2007) Biological relationship between Central and South American Chibchan speaking populations: evidence from mtDNA. Am J Phys Anthropol *133:*753-70.

Melton P (2005) Molecular Perspectives on the Origins of Chibchan Populations from the Sierra Nevada de Santa Marta, Colombia. Master's thesis, University of Kansas, Lawrence, KS.

Menozzi P, Piazza A, and Cavalli-Sforza L. (1978). Synthetic maps of human gene frequencies in Europeans. Science 201(4358):786-792.

Merriwether DA, Kemp BM, Crews DE, and Neel JV (2000) Gene flow and genetic variation in the Yanomama as revealed by mitochondrial DNA. In C Renfrew (ed.): America Past, America Present: Genes and Languages in the Americas and Beyond. Cambridge, UK: McDonald Institute for Archaeological Research, pp. 89-124.

Merriwether DA, Reed DM, and Ferrell RE (1997) Ancient and contemporary mitochondrial variation in the Maya. In SL Whitington and DM Reed (eds.): Bones of the Maya: studies of ancient skeletons. Washington D.C.: Smithsonian Institute, pp. 208-217.

Merriwether DA, Rothhammer F, and Ferrell RE (1995) Distribution of the four founding lineage haplotypes in Native Americans suggests a single wave of migration for the New World. Am J Phys Anthropol *98:*411-30.

Mesa N, Mondragon M, Soto I, *et al.* (2000) Autosomal, mtDNA, and Ychromosome diversity in Amerinds: Pre- and post-Columbian patterns of gene flow in South America. American Journal of Human Genetics 67:1277-1286.

Meyer S, Weiss G, von Haeseler A. (1999). Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. Genetics 152:1103–1110.

Miller MP (2005). Alleles in space (AIS): computer software for the joint analysis of interindividual spatial and genetic information. J Hered 96(6):722-724.

Monmonier, M. (1973). Maximum-difference barriers: An alternative numerical regionalization method. Geogr. Anal. 3:245–261.

Moraga ML, Rocco P, Miquel JF, Nervi F, Llop E, Chakraborty R, Rothhammer F, and Carvallo P. 2000. Mitochondrial DNA polymorphisms in Chilean aboriginal populations: implications for the peopling of the southern cone of the continent. Am J Phys Anthropol 113(1):19-29.

Morison, S.E (1942) Admiral of the Ocean Sea: A Life of Christopher Columbus, Boston, Little, Brown and Company,

NCBI Human Genome map website accessed July, 2008 http://www.ncbi.nlm.nih.gov/projects/genome/guide/human/

Nei M. (1987). Molecular evolutionary genetics. New York: Columbia University Press.

Nettle, D. (1999). Linguistic diversity of the Americas can be reconciled with a recent colonization. Proc. Natl. Acad. Sci. USA 96**:** 3325-3329.

Nichols J. (2002). The First American Languages. In: Jablonski N, editor. The First Americans. San Francisco, CA: California Academy of Sciences. p 273-293

Nichols J (1992), Linguistic diversity in space and time, Chicago: University of Chicago Press

Noonan JP, Coop G, Kudaravalli S, Smith D, Krause J, Alessi J, Chen F, Platt D, Paabo S, Pritchard JK, and Rubin EM (2006) Sequencing and analysis of Neanderthal genomic DNA. Science *314:*1113-8.

Novick GE, Novick CC, Yunis J, Yunis E, Antunez de Mayolo P, Scheer WD, Deininger PL, Stoneking M, York DS, Batzer MA, and Herrera RJ (1998) Polymorphic Alu insertions and the Asian origin of Native American populations. Hum Biol *70:*23-39.

Olivo PD, Vandewalle MJ, Laipis PJ, and Hauswirth WW (1983) Nucleotide-Sequence Evidence for Rapid Genotypic Shifts in the Bovine Mitochondrial-DNA D-Loop. Nature *306:*400-402.

O'Rourke DH, Hayes MG, and Carlyle SW. (2000). Spatial and temporal stability of mtDNA haplogroup frequencies in native North America. Hum Biol 72(1):15-34.

O'Rourke DH, Mobarry A, and Suarez BK. (1992). Patterns of genetic variation in Native America. Hum Biol 64(3):417-434.

O'Rourke DH, and Suarez BK. (1986). Patterns and correlates of genetic variation in South Amerindians. Ann Hum Biol 13(1):13-31.

Ota T (1993) DISPAN: Genetic Distance and Phylogenetic Analysis. University Park, PA, : Pennsylvania State University.

Oviedo GF (1976) Nicaragua en los cronistas de Indias. Serie Cronistas 3. Banco de América, Managua.

Oviedo GF (1977) Centroamérica en los cronistas de Indias: Oviedo. Serie Cronistas 4. Banco de América, Managua.

Pakendorf B, Novgorodov IN, Osakovskij VL, Danilova AP, Protod'jakonov AP, and Stoneking M. (2006). Investigating the effects of prehistoric migrations in Siberia: genetic variation and the origins of Yakuts. Hum Genet 120(3):334-353

Pearson, G. A. (2002). Pan-Continental Paleoindian Expansions and Interactions as Viewed from the Earliest Lithic Industries of Lower Central America, Ph.D. Dissertation, Department of Anthropology, University of Kansas, Lawrence.

Pena SD, Santos FR, Bianchi NO, Bravi CM, Carnese FR, Rothhammer F, Gerelsaikhan T, Munkhtuja B, and Oyunsuren T. (1995). A major founder Y-chromosome haplotype in Amerindians. Nat Genet 11(1):15-16.

Penaloza-Espinosa RI, Arenas-Aranda D, Cerda-Flores RM, *et al.* (2007) Characterization of mtDNA haplogroups in 14 Mexican indigenous populations. Human Biology *79:*313-320.

Perez-Miranda AM, Alfonso-Sanchez MA, Kalantar A, Garcia-Obregon S, de Pancorbo MM, Pena JA, and Herrera RJ (2005) Microsatellite data support subpopulation structuring among Basques. J Hum Genet *50:*403-14.

Piercy R, Sullivan K, Benson N, and Gill P. (1993). The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. Int J Legal Med 106(2):85-90.

Piperno DR, and Flannery KV. (2001). The earliest archaeological maize (*Zea mays* L.) from highland Mexico: new accelerator mass spectrometry dates and their implications. Proc Natl Acad Sci U S A 98(4):2101-2103.

Piperno DR, Pearsall DM (1998) The Origins of Agriculture in the Lowland Neotropics. Academic Press, San Diego

Phillips S. (1995). Holocene evolution of the Changuinola peat deposit, Panama: Sedimentology of a marine-influenced tropical peat deposit on a tectonically active coast: The University of British Columbia (Canada). Ph.D. Dissertation.

Pohl, M. D., Pope, K. O., Jones, J. G., Jacob, J. S., Piperno, D. R., deFrance, S. D., Lentz, D. L., Gifford, J. A., Danforth, M. E., and Josserand, K. J. (1996) Early agriculture in the Maya lowlands. Latin American Antiquity 7(4):355-372.

Pring DC (1977) The Preclassic Ceramics of Northern Belize, University of London, London.

Ramos-Onsins SE, Rozas J. 2002. Statistical properties of new neutrality tests against population growth. Mol Biol Evol 19:2092–2100.

Ranere, A. J., and Cooke, R. G. (2003). Late glacial and early Holocene occupation of Central American tropical forests. In Mercader, J. (ed.), Under the Canopy. The Archaeology of Tropical Rain Forests, Rutgers University Press, New Brunswick, pp. 219–248.

Redd AJ, and Stoneking M (1999) Peopling of Sahul: mtDNA variation in aboriginal Australian and Papua New Guinean populations. Am J Hum Genet *65:*808-28.

Reichel-Dolmatoff G. (1950). The Kogi: a tribe of the Sierra Nevada de Santa Marta Colombia. Bogota´ :El Instituto.

Reynolds, J., Weir, B.S., and Cockerham, C.C. (1983) Estimation for the coancestry coefficient: basis for a short-term genetic distance. Genetics 105:767-779.

Rickards O, Martinez-Labarga C, Lum JK, De Stefano GF, and Cann RL (1999) mtDNA history of the Cayapa Amerinds of Ecuador: detection of additional founding lineages for the Native American populations. Am J Hum Genet *65:*519-30.

Rieder MJ, Taylor SL, Tobe VO, and Nickerson DA (1998) Automating the identification of DNA variations using quality-based fluorescence re-sequencing: analysis of the human mitochondrial genome. Nucleic Acids Research *26:*967-973.

Riverstone G. (2004) Living in the Land of Our Ancestors: Rama Indian and Creole Territory in Caribbean Nicaragua. ASDI Managua, Nicaragua

Roewer L., Arnemann J., Spurr N.K., Grzeschik K.H. and Epplen J.T. (1992) Simple repeat sequences on the human Y-chromosome are equally polymorphic as their autosomal counterparts. Hum. Genet. 89:389-394.

Rogers AR, and Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. Mol Biol Evol *9:*552-69.

Rohlf, F. J. (2002) NTSYSpc: Numerical Taxonomy System, ver. 2.1. Setauket, NY: Exeter Publishing.

Rosser ZH, Zerjal T, Hurles ME, Adojaan M, Alavantic D, Amorim A, *et al.* (2000) Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. Am J Hum Genet *67:*1526-43.

Rothhammer F, Llop E, Carvallo P, and Moraga M. (2001). Origin and evolutionary relationships of native Andean populations. High Alt Med Biol 2(2):227-233.

Rothhammer F, and Silva C. (1992). Gene geography of South America: testing models of population displacement based on archeological evidence. Am J Phys Anthropol 89(4):441-446.

Rothhammer F, and Silva C. (1989). Peopling of Andean South America. Am J Phys Anthropol 78(3):403-410.

Rowold DJ, and Herrera RJ (2003) Inferring recent human phylogenies using forensic STR technology. Forensic Sci Int *133:*260-5.

Roy-Engel AM, Carroll ML, Vogel E, Garber RK, Nguyen SV, Salem AH, Batzer MA, and Deininger PL (2001) Alu insertion polymorphisms for the study of human genomic diversity. Genetics *159:*279-90.

Rubicz R. (2007). Evolutionary Consequences of Recently Founded Aleut Communities in the Commander and Pribilof Islands. Lawrence: University of Kansas. Ph.D. Dissertation

Rubicz R, Melton PE, and Crawford MH (2007) Molecular Markers in Anthropological Studies. In MH Crawford (ed.): Anthropological Genetics: Theory, Method and Applications. Cambridge: Cambridge University Press.

Rubicz R, Schurr TG, Babb PL, and Crawford MH (2003) Mitochondrial DNA variation and the origins of the Aleuts. Hum Biol *75:*809-35.

Ruhlen M (1991) A Guide to the World’s Languages. Stanford University Press: Stanford, CA.

Ruiz-Narvaez EA, Santos FR, Carvalho-Silva DR, Azofeifa J, Barrantes R, and Pena SD (2005) Genetic variation of the Y chromosome in Chibcha-speaking Amerindians of Costa Rica and Panama. Hum Biol *77:*71-91.

Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, and Arnheim N (1985) Enzymatic amplification of beta-globin genomic sequences and

restriction site analysis for diagnosis of sickle cell anemia. Science *230:*1350-4.

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors Proc Natl Acad Sci U S A. 74(12): 5463–5467.

Saillard J, Forster P, Lynnerup N, Bandelt HJ, Norby S (2000) mtDNA variation among Greenland Eskimos. The edge of the Beringian expansion. American Journal of Human Genetics 67:718-726.

Sauer CO. (1952). Agricultural origins and dispersals. New York,: American Geographical Society.

Salas A, Richards M, De la Fe T, Lareu MV, Sobrino B, Sanchez-Diz P, Macaulay V, and Carracedo A. (2002). The making of the African mtDNA landscape. Am J Hum Genet 71(5):1082-1111

Salzano FM (2002) Molecular variability in Amerindians: widespread but uneven information. An Acad Bras Cienc *74:*223-63.

Salzano FM, and Callegari-Jacques SM. (1988). South American Indians : a case study in evolution. Oxford New York: Clarendon Press - Oxford University Press.

Sánchez-Albornoz N (1977) La población de America Latina: desde los tiempos precolombinos al año 2000. 2nd ed. Alianza Editorial, Madrid.

Sanger F, Nicklen S, and Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A *74:*5463-7.

Santos FR, Rodriguez-Delfin L, Pena SD, Moore J, and Weiss KM. (1996). North and South Amerindians may have the same major founder Y chromosome haplotype. Am J Hum Genet 58(6):1369-1370.

Santos FR, Pena SDJ, and Tyler-Smith C. (1995). PCR haplotypes for the human Y chromosome based on alphoid satellite variants and heteroduplex analysis. *Gene* 165: 191-198

Santos M, Ward RH, and Barrantes R (1994) mtDNA variation in the Chibcha Amerindian Huetar from Costa Rica. Hum Biol *66:*963-77.

Santos S, Rodrigues J, Santos A, and Zago M (1999) Differential contribution of indigenous men and women to the formation of an Urban population in the

Amazon region as revealed by mtDNA and Y-DNA. American Journal of Physical Anthropology 109:175-180.

Schneider S, Roessli D, and Excoffier L (2000) Arlequin ver. 2.00: A Software for Population Genetic Analysis. Switzerland: Genetics and Biometry Laboratory, University of Geneva.

Schmitt R, Bonatto SL, Freitas LB, Muschner VC, Hill K, Hurtado AM, and Salzano FM (2004) Extremely limited mitochondrial DNA variability among the Ache Natives of Paraguay. Ann Hum Biol *31:*87-94.

Schroeder KB, Schurr TG, Long JC, Rosenberg NA, Crawford MH, Tarskaia LA, Osipova LP, Zhadanov SI, and Smith DG. (200)7. A private allele ubiquitous in the Americas. Biol Lett 3(2):218-223.

Schull W, and MacCluer J (1968) Human genetics: structures of populations. Annual Review of Human Genetics *2:*279-304.

Schurr T (2004) The peopling of the New World: Perspectives from molecular anthropology. Annual Review of Anthropology 33:551-583.

Schurr TG, and Sherry ST (2004) Mitochondrial DNA and Y chromosome diversity and the peopling of the Americas: evolutionary and demographic evidence. Am J Hum Biol *16:*420-39.

Schurr T and Wallace D (1999) Mitochondrial DNA variation in Native Americans and Siberians and its implications for the peopling of the New World. In: Who Were the First Americans: Proceedings of the 58th Annual Biological Colloquium, Oregon State University. Ed. by Bonnichsen R. Center for the Study of the First Americans: Corvallis, OR. 41-77.

Schurr TG, Ballinger SW, Gan YY, Hodge JA, Merriwether DA, Lawrence DN, Knowler WC, Weiss KM, and Wallace DC (1990) Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. Am J Hum Genet *46:*613-23.

Seielstad MT, Hebert JM, Lin AA, Underhill PA, Ibrahim M, Vollrath D, and Cavalli-Sforza LL. (1994) Construction of human Y-chromosomal haplotypes using a new polymorphic A to G transition. Human Molecular Genetics 3: 2159-2161.

Sharer RJ, and Grove DC, eds. (1989) Regional Perspectives on the Olmec. Cambridge: Cambridge University Press.

Sheets P. 1992. The Pervasive Pejorative in Intermediate Area Studies. In: Lange F, editor. Wealth and Hierarchy in the Intermediate Area. Washington, D.C.: Dumbarton Oaks. p 15-42.

Shen P, Wang F, Underhill PA, Franco C, Yang WH, Roxas A, Sung R, Lin AA, Hyman RW, Vollrath D, *et al.* (2000). Population genetic implications from sequence variation in four Y chromosome genes. *Proceedings National Academy of Sciences USA* 97: 7354-7359.

Sherry ST, Rogers AR, Harpending H, Soodyall H, Jenkins T, and Stoneking M (1994) Mismatch distributions of mtDNA reveal recent human population expansions. Hum Biol *66:*761-75.

Simonsen KL, Churchill GA, Aquadro CF. (1995). Properties of statistical tests of neutrality for DNA polymorphism data. Genetics 141(1):413-429.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457-462.

Snarskis MJ. (2003). From jade to gold in Costa Rica: How, why, and when. In: Quilter J, and Hoopes JW, editors. Gold and Power in Ancient Costa Rica, Panama, and Colombia. Washington, D.C.: Dumbarton Oaks. p 159-204.

Snarskis,M. J. (1998). Stone sculpture and Pre-Columbian cultural evolution in the Central Highlands- Atlantic Watershed of Costa Rica/La sculpture en perre et l'evolution culturelle pr´ecolombienne sur Altiplano Central et sur le versant atlantique du Costa Rica. *Precolombart* 1: 19–41.

Snarskis MJ. 1978. The Archaeology of the Central Atlantic Watershed of Costa Rica: Columbia University. Ph.D. Dissertation.

Sokal RR, Oden NL, and Wilson C (1991) Genetic evidence for the spread of agriculture in Europe by demic diffusion. Nature *351:*143-5.

Spielman RS, Migliazza EC, Neel JV, Gershowitz H, and Arauz RTD (1979) Evolutionary Relationships of 2 Populations - Study of the Guaymi and the Yanomama. Current Anthropology *20:*377-388.

Stone AC, and Stoneking M. (1998). mtDNA analysis of a prehistoric Oneota population: implications for the peopling of the New World. Am J Hum Genet 62(5):1153-1170.

Stone DZ. (1977). Pre-Columbian man in Costa Rica. Cambridge, MA: Peabody Museum Press.

Stone DZ. 1972. Pre-Columbian man finds Central America; the archaeological bridge. Cambridge, Mass.,: Peabody Museum Press..

Stone DZ. (1966). Synthesis of Lower Central American Ethnohistory. In: Willey GR, and Eckholm GF, editors. Handbook of Middle American Indians. Austin, TX: University of Texas Press. p 209-233.

Sturrock K, and Roche J. (2000). A Multidimensional Scaling Stress Evaluation Table Field Methods 12(1):49-60.

Suarez BK, Crouse JD, and O'Rourke DH. (1985). Genetic variation in North Amerindian populations: the geography of gene frequencies. Am J Phys Anthropol 67(3):217-232.

Swadesh. M (1955). Time Depths of American Linguistic Groupings. American Anthropologist 56 121-137.

Tajima F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595.

Tamm E, Kivisild T, Reidla M, Metspalu M, Smith DG, Mulligan CJ, Bravi CM, Rickards O, Martinez-Labarga C, Khusnutdinova EK, *et al.* (2007) Beringian standstill and spread of Native American founders. PLoS ONE *2:*e829.

Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10:512–526.

Tarazona-Santos E, Carvalho-Silva DR, Pettener D, Luiselli D, *et al.* (2001). Genetic differentiation in South Amerindians is related to environmental and cultural diversity: evidence from the Y chromosome. Am J Hum Genet 68(6):1485-1496.

Thomas, H (2004) Rivers of Gold New York: Random House

Thompson EA, Neel JV, Smouse PE, and Barrantes R (1992) Microevolution of the Chibcha-Speaking Peoples of Lower Central-America - Rare Genes in an Amerindian Complex. American Journal of Human Genetics *51:*609-626.

Thomson R, Pritchard JK, Shen P, Oefner PJ, and Feldman MW. (2000). Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. *Proceedings of the National Academy of Sciences USA* 97: 7360-7365.

Torroni A, Achilli A, Macaulay V, Richards M, and Bandelt HJ (2006) Harvesting the fruit of the human mtDNA tree. Trends in Genetics *22:*339-345.

Torroni A, Huoponen K, Francalacci P, Petrozzi M, Morelli L, Scozzari R, Obinu D, Savontaus ML, and Wallace DC. 1996. Classification of European mtDNAs From an Analysis of Three European Populations. Genetics 144(4):1835-1850.

Torroni A, Chen YS, Semino O, Santachiarabeneceretti AS, Scott CR, Lott MT, Winter M, and Wallace DC (1994a) Mtdna and Y-Chromosome Polymorphisms in 4 Native-American Populations from Southern Mexico. American Journal of Human Genetics *54:*303-318.

Torroni A, Neel JV, Barrantes R, Schurr TG, and Wallace DC (1994b) Mitochondrial DNA "clock" for the Amerinds and its implications for timing their entry into North America. Proc Natl Acad Sci U S A *91:*1158-62.

Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, Vullo CM, and Wallace DC (1993) Asian affinities and continental radiation of the four founding Native American mtDNAs. Am J Hum Genet *53:*563-90.

Torroni A, Schurr TG, Yang CC, Szathmary EJE, Williams RC, Schanfield MS, Troup GA, Knowler WC, Lawrence DN, Weiss KM and others. 1992. Native American Mitochondrial DNA Analysis Indicates That the Amerind and the Nadene Populations Were Founded by Two Independent Migrations. Genetics 130(1):153-162.

Underhill PA, Passarino G, Lin AA, Shen P, Mirazon Lahr M, Foley RA, Oefner PJ, and Cavalli-Sforza LL (2001) The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. Ann Hum Genet *65:*43-62.

Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T, Vollrath D, Davis RW, 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, Jin L, Zemans R, Oefner PJ, and Cavalli-Sforza LL (1996) A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. Proc Natl Acad Sci U S A *93:*196-200.

Uribe C (2000) La Ethnografia de las Sierra Nevada de Santa Marta y las Tierras Bajas, Adyacentes: Geografia Humana de Colombia Nordeste Indígena (Tomo II). Bogatá: Bank of the Republic.

Vigilant L, Stoneking M, Harpending H, Hawkes K, and Wilson AC (1991) African populations and the evolution of human mitochondrial DNA. Science *253:*1503-7.

Vigilant L, Pennington R, Harpending H, Kocher TD, Wilson AC (1989) Mitochondrial DNA sequences in single hairs from Southern African populations. Proceedings of the National Academy of Sciences, USA 86:9350-9354.

Vogt PH, Affara N, Davey P, Hammer M, Jobling MA, Lau YF, Mitchell M, Schempp W, Tyler-Smith C, Williams G, Yen P, and Rappold GA (1997) Report of the Third International Workshop on Y Chromosome Mapping 1997. Heidelberg, Germany, April 13-16, 1997. Cytogenet Cell Genet *79:*1-20.

Urquhart G. (1997). Disturbance and Regeneration of swamp forests in Nicaragua: Evidence from Ecology and Paleoecology: University of Michigan. Ph.D. Dissertation.

Wallace DC, and Lott MT. 2004. MITOMAP: A Human Mitochondrial Database. http://www.mitomap.org – Accessed June 2008

Wallace DR (1997) Central American Landscapes. In AG Coates (ed.): Central America: A Natural and Cultural History. New Haven: Yale University Press, pp. 72-96.

Wang S, Ray N, Rojas W, Parra MV, Bedoya G, Gallo C, Poletti G, Mazzotti G, Hill K, Hurtado AM and others. (2008). Geographic Patterns of Genome Admixture in Latin American Mestizos. PLoS Genetics 4(3):e1000037

Wang S, Lewis CM, Jakobsson M, Ramachandran S, Ray N, Bedoya G, *et al.* (2007) Genetic Variation and Population Structure in Native Americans. PLoS Genetics *3:*e185.

Ward RH, Salzano FM, Bonatto SL, Hutz MH, Coimbra CEA, and Santos RV (1996) Mitochondrial DNA polymorphism in three Brazilian Indian tribes. American Journal of Human Biology *8:*317-323.

Ward RH, Frazier BL, Dew-Jager K, and Paabo S. (1991). Extensive mitochondrial diversity within a single Amerindian tribe. Proc Natl Acad Sci USA 88:8720-4.

Ward RH, Gershowitz H, Layrisse M, and Neel JV (1975) Genetic Structure of a Tribal Population, Yanomama Indians .11. Gene-Frequencies for 10 Blood-Groups and Abh-Le Secretor Traits in Yanomama and Their Neighbors - Uniqueness of Tribe. American Journal of Human Genetics *27:*1-30.

Watson DF, 1992. Contouring: a guide to the analysis and display of spatial data. New York: Pergamon Press.

Whitfield LS, Sulston JE, and Goodfellow PN (1995).Sequence variation of the human Y chromosome. Nature 378: 379-380

Ward RH, Gershowitz H, Layrisse M, and Neel JV (1975) Genetic Structure of a Tribal Population, Yanomama Indians .11. Gene-Frequencies for 10 Blood-Groups and Abh-Le Secretor Traits in Yanomama and Their Neighbors - Uniqueness of Tribe. American Journal of Human Genetics *27:*1-30.

Willey GR (1974) The Classic Maya Hiatus. In RF Heizer and JA Graham (eds.): Observations on the Emergence of Civilization in Mesoamerica. Berkeley: University of California Press, pp. 97-111.

Willey GR. 1971. An Introduction of American Archaeology, Volume II: South America. Englewood Cliffs: Prentice Hall, Inc.

Williams SR, Chagnon NA, and Spielman RS (2002) Nuclear and mitochondrial genetic variation in the Yanomamo: a test case for ancient DNA studies of prehistoric populations. Am J Phys Anthropol *117:*246-59.

Wilson JF, Weiss DA, Richards M, Thomas MG, Bradman N, and Goldstein DB. (2001). Genetic evidence for different male and female roles during cultural transitions in the British Isles. Proc Natl Acad Sci U S A 98(9):5078-5083.

Witkowski S, Brown C (l98l), "Mesoamerican historical linguistics and distant genetic relationship. American Anthropologist. 83:9O5-9ll.

Wright S (1931) Evolution in Mendelian populations. Genetics 16:96-159.

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

Yunis JJ, Ossa H, Salazar M, Delgado MB, Deulofeut R, de la Hoz A, Bing DH, Ramos O, Yunis EJ, and Yunis EJ (1994) Major histocompatibility complex class II alleles and haplotypes and blood groups of four Amerindian tribes of northern Colombia. Hum Immunol *41:*248-58.

Zegura SL, Karafet TM, Zhivotovsky LA, and Hammer MF (2004) High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas. Mol Biol Evol *21:*164-75.

Zlojutro M, Rubicz R, Devor EJ, Spitsyn VA, Makarov SV, Wilson K, and Crawford MH. 2006. Genetic structure of the Aleuts and Circumpolar populations based on mitochondrial DNA sequences: a synthesis. Am J Phys Anthropol 129(3):446-464.

APPENDIX A

The Co-Evolution of Genes and Language in Chibchan populations from Middle America

INFORMED CONSENT STATEMENT

The Department of Anthropology at the University of Kansas supports the practice of protection for human subjects participating in research. The following information is provided for you to decide whether you wish to participate in the present study. You may refuse to sign this form and not participate in this study. You should be aware that even if you agree to participate, you are free to withdraw at any time. If you do withdraw from this study, it will not affect your relationship with this unit, the services it may provide to you, or the University of Kansas.

We are interested in using molecular genetic techniques to reconstruct the origins and migrations of indigenous human populations inhabiting Central America belonging to the same language family, Chibchan. You will be participating in one session that should require a few minutes of your time. During that time you will be asked about the geographic origins of your parents, your age, places of birth and two buccal smears, one on each side of your mouth, will be taken. The buccal smear technique consists of a sterile wooden applicator being gently stroked across the cheeks and gums, followed by rinsing the mouth with distilled water.

The DNA extracted from the buccal smears will be used solely to reconstruct the genetic history of Chibchan speaking populations. Although participation will not directly benefit you, we believe that the information will be useful in revealing the origins of Central American indigenous people and their relationship to other Native American groups. There are no risks associated with this study. All DNA will be used up in this analysis. Only personnel working directly on the Chibchan-speakers project will have access to the DNA.

Your participation is strictly voluntary. We assure you that your name will not be associated in any way with the research findings. There is a minor risk that personal information may be identifiable, however this information will be identified only be a code number and all study data will be maintained in a locked file cabinet in a separate room from all DNA. You are not required to sign this Consent and Authorization form and you may refuse to do so without affecting your right to any services you are receiving or may receive from the University of Kansas or to

participate in any programs or events of the University of Kansas. However, if you refuse to sign, you cannot participate in this study.

If you would like additional information concerning this study before or after it is complete, please feel free to contact either of the individuals listed below by phone or e-mail.

Researcher Contact Information

Phillip E. Melton	Dr. Michael H. Crawford Ph.
Principal Investigator	Faculty Supervisor
Dept. of Anthropology	Dept. of Anthropology
1415 Jayhawk Blvd	1415 Jayhawk Blvd
University of Kansas	University of Kansas Lawrence, KS 66045
785 864 _4172___	785 864 _4170_____

PARTICIPANT CERTIFICATION:

I have read this Consent and Authorization form. I have had the opportunity to ask, and I have received answers to, any questions I had regarding the study. I understand that if I have any additional questions about my rights as a research participant, I may call (785) 864-7429 or (785) 864-7385 or write the Human Subjects Committee Lawrence Campus (HSCL), University of Kansas, 2385 Irving Hill Road, Lawrence, Kansas 66045-7563, email dhann@ku.edu or mdenning@ku.edu.

I agree to take part in this study as a research participant.

_________________________________ ______________________

Type/Print Participant's Name Date

____

Participant's Signature

By my signature I affirm that I am at least 18 years old and that I have received a copy of this Consent and Authorization form.