Molecular Perspectives on the Origins of Chibchan Populations from the Sierra Nevada de Santa Marta, Colombia

by Phillip Edward Melton

B. A., University of Montana, 1996

Professor in Charge
Michael H. Crawford

Committee Members
James H. Mielke
Deborah R. Smith

MOLECULAR PERSPECTIVES ON THE ORIGINS OF CHIBCHAN POPULATIONS FROM THE SIERRA NEVADA DE SANTA MARTA, COLOMBIA

BY P. MELTON

Molecular perspectives on the origins of Chibchan populations from the Sierra Nevada de Santa Marta, Colombia

by

Phillip Edward Melton
B.A., University of Montana 1996

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 Master of Arts

M. H. Crawford
Dr. Michael H. Crawford, Chair

James H. Mielke
Dr. James H. Mielke

Deborah R. Smith
Dr. Deborah R. Smith

Date Defended: 11/30/04

Abstract

Current archaeological, biological and linguistic evidence points to a lower Central American origin for Chibchan speaking populations who are thought to have continuously occupied the region for the last 10,000 years. However, the biological relationship of these groups to Chibchan speakers from Northern South America remains largely unresolved. This thesis examines mitochondrial DNA (mtDNA) haplogroup and haplotype diversity in three Sierra Nevada de Santa Marta Chibchan (Kogí, Arsario, Ijka) speaking populations and one neighboring Arawakan (Wayuú) group from Northeast Colombia in order to determine: (1) the nature of the biological relationship between the four study populations, (2) whether or not a relationship between Central and Northern South American Chibchan groups exists, (3) a potential timeframe for a Chibchan diaspora, (4) test hypothetical models regarding the initial peopling of the Santa Marta region and, (5) the role of Chibchan populations in the peopling of the Americas.

Amerindian mtDNA haplogroups were characterized for 190 individuals using RFLP analysis and 61 HVS-I sequences were obtained. Three (A, B, and C) of the five founding Amerindian mtDNA haplogroups (A, B, C, D, and X) were found in these populations. The Kogí and Arsario exhibited only haplogroups A and C (Kogí 65% A, 35% C, Arsario 68% A, 32% C). The Ijka primarily exhibited haplogroup A (90%) with a single B (2.5%) and two C (7.5%) individuals. The Wayuú contained haplogroups A (34%), B (24%), C (32%), and undetermined (10%). Haplogroup D was not found in any of the groups examined. R-matrix analysis demonstrates that the three Santa Marta Chibchan populations are related to each other but not to the neighboring Wayuú. Analysis of these three South America Chibchan populations at the sequence level shows that they share low mtDNA haplotype diversity, low negative or positive values for Fu's Fs and Tajima's D and a peak between zero and one unit of mutational time with linguistically related populations from lower Central America and not with other indigenous South American groups. Phylogenetic reconstruction of these populations using median-joining networks indicates that all sampled Chibchan speaking populations had undergone a bottleneck and were highly influenced by a founder effect within the last 10,000 years. Using the ρ-statistic of Saillard *et al.* (2000) on two clusters of Santa Marta Chibchan haplotypes gives mtDNA coalescence dates of 8,072 (±4943) and 6,985 (±3557) both of which are consistent with other temporal estimates of Chibchan genetic history. This time depth points to a long term occupation of Chibchan populations within Northern South America suggesting an *in situ* development for the Santa Marta groups.

This study concludes that while there are biological similarities between the Chibchan speakers from the Sierra Nevada de Santa Marta and the Panamanian isthmus they diverged in the distant past. If a Chibchan diaspora did occur it may have been geographically widespread and would have occurred early during the peopling of the Americas. This diaspora may have blocked gene flow from the north and south possibly leaving genetic drift as the primary evolutionary force on the South American continent.

This thesis is dedicated to my family past, present, and future for their encouragement understanding, and patience.

Acknowledgements

I would like to thank my committee members: Dr. Michael Crawford, Dr. James Mielke, and Dr. Deborah Smith; my fellow collaborators Dr. Ignacio Briceño, Dr. Eric Devor, Dr. S.S. Papiha; my fellow students, particularly Kristin Melvin, Rohina Rubicz, and Mark Zlojutro; and the office staff: Judy Ross and Carol Archinal.

I would also like to thank the individuals who participated in the original study; Dr. John Hoopes for several useful discussions and encouragement regarding Chibchan populations and the archaeology of the region. Dr. Carlos Langeback and Dr. Augusto Oyuela-Caycedo for sharing their archaeological research with me; and of course my wife Jennifer Rack and two sons, Spencer and Xavier, who put up with me throughout the whole process.

This research was funded in part by a grant from the National Science foundation and a Carroll D. Clark award from the Department of Anthropology.

Table of Contents

List of Figures

List of Tables

Chapter I: Introduction

In 1599, Don Juan Guiral Velón, Spanish governor of the South American colonies of Santa Marta and Riohacha, raised an army of settlers to quell another rebellion by the indigenous inhabitants of the nearby Santa Marta mountain range. After a year of fierce fighting the Spanish prevailed and ruthlessly punished the indigenous inhabitants, killing their leaders, burning their villages and destroying one of the great pre-Columbian cultures of South America, the Tairona (Uribe 2000). However, a few native survivors fled to the interior of the Santa Martas where their descendents branched into the three indigenous communities found there today; the Kogí, Ijka, and Arsario. These three groups all speak languages within the Chibchan linguistic stock that extends from Honduras into northern South America. Populations belonging to this linguistic family are thought to have played important roles in peopling the Americas due to their geographic location bridging the two American continents.

The extant indigenous populations of the Americas exhibit a remarkable result of human evolution. They represent the descendents of populations who entered a vast unoccupied land mass and remained relatively isolated for a significant period of time. There is general consensus that ancestral Native American populations originated in Asia but there is disagreement among scholars regarding the number and timing of the earliest human entry into the Americas (Salzano 2002, Crawford 1998). The timing of these events range from an "early" entry approximately 30,000 years before present (YBP) to a "late" entry less than 13,000 YBP (Lavalleé 2000). One

group of Native American populations that has been used to develop an evolutionary clock for this peopling event are Chibchan-speakers from lower Central and northern South America. Current archaeological (Hoopes & Fonseca 2003), linguistic (Constenla 1981, 1991) and genetic (Barrantes *et al.* 1990, Kolman & Bermingham 1997, Torroni *et al.* 1994) evidence points to a Central American origin of Chibchan-speakers and suggests that these groups separated from other Native American populations 8,000-10,000 YBP. However, the biological relationship of these Central American groups to the Chibchan-speakers of South America has gone uninvestigated.

Over the last decade, research into mitochondrial DNA (mtDNA) variation within Native American populations has revealed that these indigenous groups are characterized by a series of specific mutations that cluster into five haplogroups termed A, B, C, D, and X (Schurr & Sherry 2004, Torroni *et al.* 1993b, Schurr *et al.* 1990). These findings and the knowledge that each of the five haplogroups are found in low frequencies in northeast Asian and Siberian populations, indicates a shared maternal ancestry between the Asian and American continents. This shared maternal ancestry permits quantification of mtDNA variation accumulated within each haplogroup from the time of the first human arrival in the Americas assuming that each of the five haplogroups was founded by a single haplotype present in Asia during the Pleistocene. The resulting quantification can then be used to develop a molecular clock based on the mtDNA mutation rate (Torroni *et al.* 1993a, 1993b, 1994).

This thesis investigates mitochondrial DNA (mtDNA) inter- and intra- group diversity in the three Chibchan populations from the Santa Marta region (Kogí, Ijka, Arsario) and one neighboring Arawak (Wayuú) speaking group from the Guajiro peninsula in northeast Colombia in order to determine; (1) the relationship of the three Santa Marta Chibchan populations to each other and to the Wayuú; (2) the relationship of the Santa Marta Chibchan populations to Central American Chibchan and South American populations; (3) a potential time frame for Chibchan genetic history; (4) and test hypothetical models for the peopling of the Santa Marta region using archaeological, linguistic, and biological data; and (5) the role of Chibchan-speaking populations in peopling the Americas.

The chapters that follow present a literature review, materials, methods, results, and a discussion pertinent to the research presented in this thesis. Chapter two summarizes the known mtDNA research on Native American populations, provides archaeological, ethnographic, and linguistic information on the Santa Marta groups and Wayuú, and reviews classic and molecular genetic data regarding South and Central American Chibchan-speakers. Chapter three presents the laboratory and analytical procedures used during this research. The results are presented in chapter four and discussed in chapter five. Finally, chapter six concludes through a brief review of findings presented in this thesis and offers suggestions for future research.

Chapter II: Literature Review

Mitochondrial DNA Review

In recent years, advances in molecular biology have led to a number of useful genomic markers for investigating human genetic population structure and history. One of the most informative and widely used of these new markers has been mitochondrial DNA (mtDNA). This is due to its elevated mutation rate, maternal transmission, and lack of recombination. This locus is located within the energy producing mitochondria in the cytoplasm of the cell. The human mtDNA genome is circular and has approximately 16,569 base pairs (bp). The mitochondrial genome (Figure 1) is divided into two regions: a coding region of approximately 15,000 bp which codes for 37 tightly packed genes (22 transfer RNAs, 13 proteins, and two other RNAs), and a control region of approximately 1,000 bp that is subdivided into two hypervariable segments (HVS-I and HVS-II, ~ 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 mutation rate ten times higher than the rate of the mtDNA coding region (Francalacci *et al.* 1999). This mtDNA coding region mutation rate is estimated as 3.2% per million years (Francalacci *et al.* 1999) and the rate increases to approximately 8.4% per million years in the mtDNA control region (Vigilant *et al.* 1989). This mtDNA control region variability is not evenly distributed, with twice the number of polymorphic sites being present in HVS-I than in HVS-II. MtDNA is maternally inherited, meaning that it is passed from the mother to all of her offspring. However, only the daughters pass it onto their offspring (Giles *et al.* 1980). This

suggests that mtDNA does not undergo recombination and is passed unaltered generation to generation. Therefore, any observed variations in mtDNA can be attributed to mutation and the timing of the mutational event can be estimated using the mtDNA mutation rate.

The earliest studies of population history in mtDNA variation used restriction fragment length polymorphism (RFLP) analysis and focused on the entire mtDNA genome (Cann *et al.* 1987, Schurr *et al.* 1990, Torroni *et al.* 1990). This early RFLP research focused on the conservative coding region and suggested that some neutral mutations occurred only once in evolutionary history, which allowed for the grouping of different lineages in clusters, called haplogroups, defined by their restriction sites. Two different nomenclatures were originally proposed: Roman numerals being used by Horai *et al.* (1993) and the English alphabet being used by Torroni *et al.* (1992). The alphabetic nomenclature gained wider acceptance and is currently used by researchers. The high resolution RFLP involves copying the entire mitochondrial genome in nine overlapping segments using the polymerase chain reaction (PCR) method. These nine segments are digested with 14 restriction enzymes that cleave mtDNA at diagnostic restriction sites. Figure 1 shows the position of these sites in the mtDNA, the corresponding restriction enzyme and their associated English alphabet haplogroup. Several of these haplogroups were found to be continent specific and could be used to trace the migration patterns of human populations (Francalacci *et al.* 1999). Figure 2 shows the geographic location of these haplogroups and hypothetical

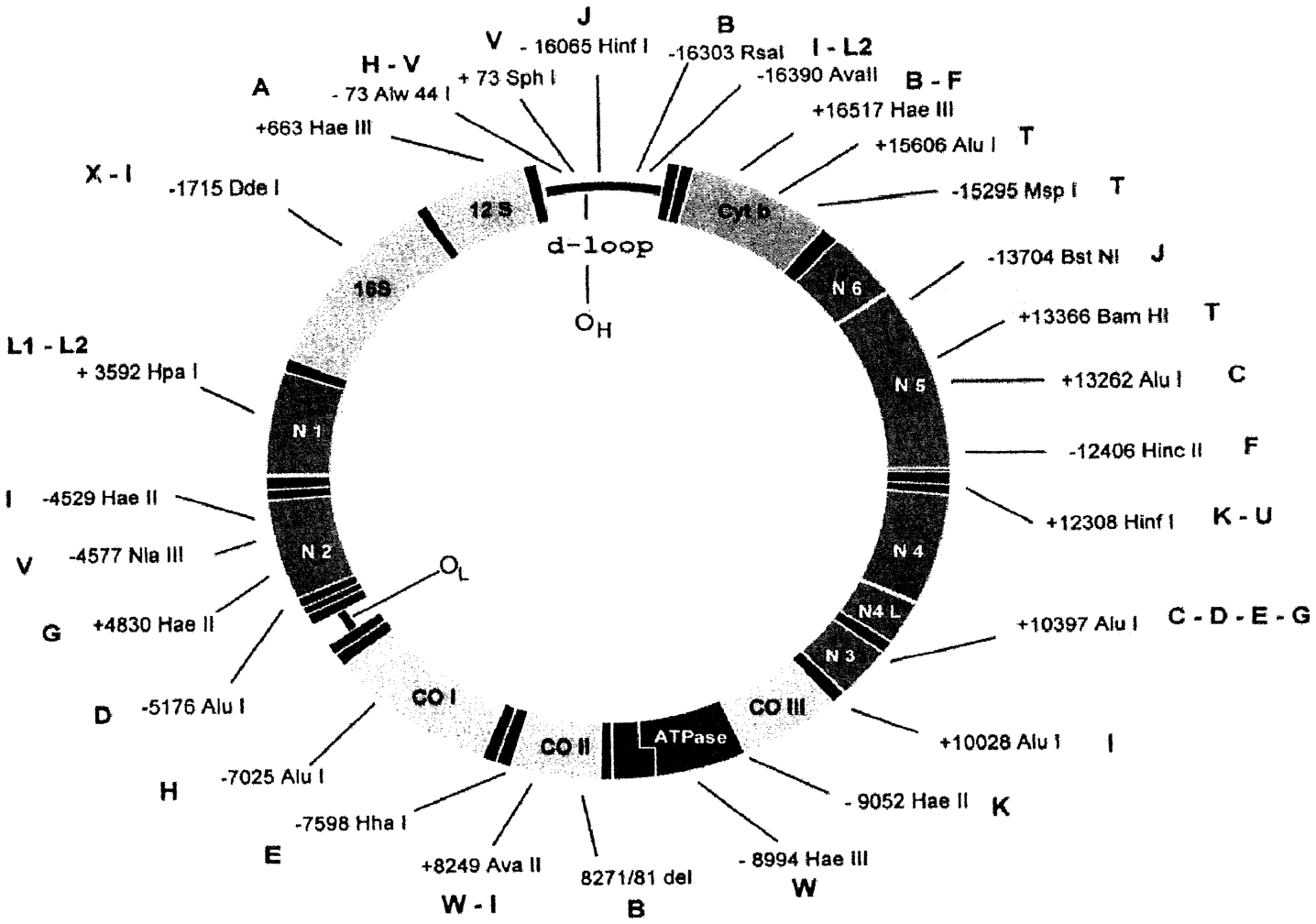


Figure 1: Location of haplogroup specific restriction sites in human mtDNA. Capital bold letters outside circle indicate English nomenclature RFLP site, + or – indicates presence or absence of restriction enzyme cut site, numbers indicate nucleotide position of mtDNA RFLP site, Text within circle represents mtDNA coding region genes, d-loop refers to mtDNA control region (from Francalacci *et al.* 1999)

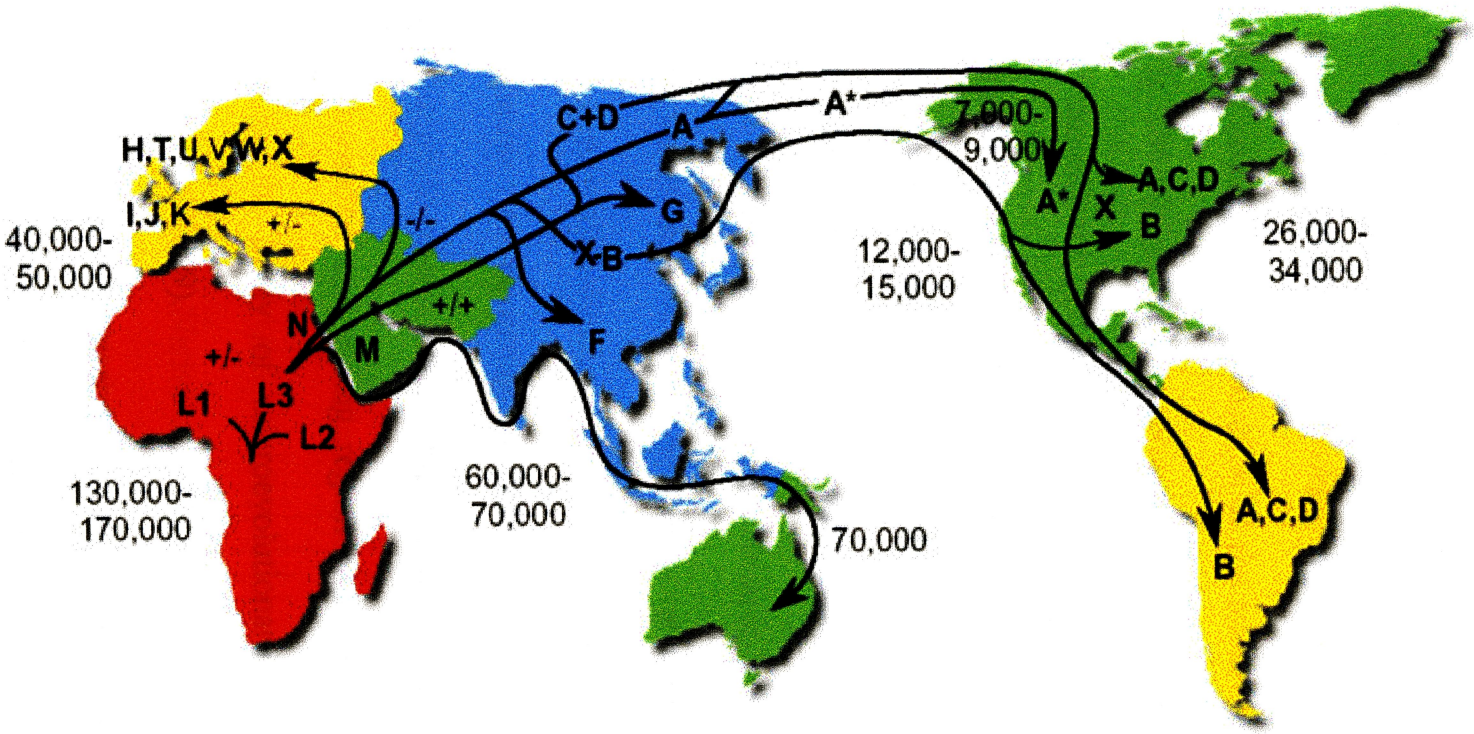


Figure 2: Geographic distribution of human mtDNA haplogroups. Time estimates are YBP and based on mutation rate of 2.2-2.9%/million years +/+, +/-, -/- refers to *Dde*I 10394/ *Alu*I 10397, * refer to +*RsaI* 16329 (Wallace & Lott 2004)

migration patterns that may have occurred during human evolutionary history (Wallace & Lott 2004).

Subsequent human mtDNA research focused on the high variability within the control region and used direct DNA sequencing of HVS-I, which allowed for a deeper analysis of population differentiation at the microevolutionary level (Vigilant *et al.* 1989, Vigilant 1990). There is certain risk, however, in just focusing on the control region of the mtDNA genome. The high mutation rate found in the human mtDNA control region increases the potential for homoplasy occurring at the same nucleotide site. This may result in reduced genetic variability between known evolutionarily divergent populations, complicating phylogenetic analysis. There are two plausible explanations for this reduced genetic variation in human populations, it: (1) represents an ancient mutational event derived from a most recent common ancestor (MRCA) and is present in all individuals within a population, (2) is a site with an extremely high mutation rate occurring in different mtDNA lineages reducing value for establishing phylogenetic relationships. Analyses that do not account for this and give all mtDNA control and coding region mutations the same weight provide poor resolution in the study of microevolutionary history at the population level (Francalacci *et al.* 1999).

Current mtDNA research has combined mtDNA control region sequencing and coding region RFLP analysis because both regions are inherited as a unit (Torroni *et al.* 1996, Graven *et al.* 1995, Torroni *et al.* 1993 a, b). This combined mtDNA region approach has further elucidated the evolution of mtDNA because sequencing

information is kept intact. However, it is arranged according to its phylogenetic relationship value through association with monophyletic restriction enzyme cut sites (Francalacci *et al.* 1999). This has shown that some HVS-I site-specific sequence mutations always coincide with certain restriction sites from the coding region. This mtDNA control-coding region relationship allows for a balance between the two regions and allows researchers to check their results.

Amerindian mtDNA haplogroups

There are currently five major founding mtDNA haplogroups (A, B, C, D, & X) for Native Americans, characterized by RFLPs and a 9-base pair deletion located in the cytochrome c oxidase subunit II (CO II) gene of the mitochondrial genome (figure 1). Table 1 shows the relationship between coding region RFLPs and control region HVS-I sequences that define the five Native American haplogroups. An example of this relationship is that individuals belonging to haplogroup A are defined through the presence of a *Hae* III cut site at mtDNA nucleotide (nt) 663 in the coding region and control region transitional mutations at nucleotides 16223(T), 16290(T), 16319(A), and 13362(C). Combined, these five haplogroups account for 95-100% of all mtDNAs from autochthonous populations present in the Americas (Schurr & Sherry 2004). All five Native American haplogroups have been detected in pre-Columbian skeletal materials found in North America (O'Rourke *et al.* 2000, Stone & Stoneking 1998, Fox 1996). However, haplogroup X has never been detected in any ancient or living South American population (Salzano 2002). These five haplogroups have also been found in low frequencies in Central and Northeast Asia, indicating that

they were the founding mtDNA lineages in Native American populations (Schurr & Sherry 2004, Schurr *et al.* 1990). Some researchers have suggested that these five haplogroups do not represent all of the initial haplogroups present in pre-Columbian America. These researchers propose that other founding haplogroups may have been lost due to the severe depopulation that occurred after contact or through genetic drift (Rickards *et al.* 1999, Easton *et al.* 1996, Lorenz & Smith 1996, Bailliet *et al.* 1994). Subsequent research conducted on these populations has documented that most of these unknown haplotypes were derivatives of haplogroups A-D, belonged to haplogroup X, were the result of non-native admixture, or not sufficiently analyzed to determine their haplogroup status (Schurr & Sherry 2004).

Native American Haplogroup	**Geographic Distribution**	**16189**	**16223**	**16278**	**16290**	**16298**	**16319**	**16327**	**16362**	**Coding region RFLP sites**
CRS*		**T**	**C**	**C**	**C**	**T**	**G**	**C**	**T**	
A	Asia-America		T		T		A		C	+ 663 *Hae*III
B	Asia-America	C								8271-8281 9 bp del,
C	Asia-America		T			C		T		-13259/+13262 *Hinc*II/*Alu*I
D	Asia-America		T						C	-5176 *Alu*I
X	Asia, Europe, America		T	T						-1715 *Dde*I

Table 1: Correlation between coding region restriction sites and control region variants in human mtDNA. *Cambridge Reference Sequence (Anderson *et al.* 1981)

Haplogroups A, B, C, and D are found throughout North, Central and South American indigenous populations in differing frequencies. Haplogroup A occurs at higher frequencies in North America, and is frequent in Artic and Subartic populations but is also common among Chibchan-speaking populations from lower

Central America (Rubicz *et al.* 2003, Lorenze & Smith 1996, Merriwether *et al.* 1995, Torroni *et al.* 1994, Schurr *et al.* 1990). Haplogroup B, almost absent among Artic populations, is found in moderate frequencies in the Southwest United States, is high among Central American populations and in the Andes (Merriwether *et al.* 1995, Fox 1996). Haplogroups C and D are generally rare in North America but are found in high frequencies among South American populations from the Amazon basin and the southern cone of the continent (Schurr & Sherry 2004, Merriwether *et al.* 1995, Fox 1996, Ginther *et al.* 1993). Haplogroup X has only been found in North American populations (Schurr & Sherry 2004) but has not been detected in any South American populations (Salzano 2002).

There does not appear to be a distinct clinal distribution within North America based on mtDNA haplogroup data but there does appear to be one in South America involving haplogroups A and D. Haplogroup A occurs at high frequencies (>50%) in the northern parts of the continent and is completely absent from the southern cone. Haplogroup D mirrors the A, with high frequencies in the southern cone, and almost a complete absence from the northern parts of South America (Keyeux *et al.* 2002, Fox 1996). Haplogroup A occurs at high frequencies in lower Central America and northern South America but is completely absent from groups in the southern areas of the continent (Moraga *et al.* 2000, Kolman *et al.* 1995). Haplogroup B is detected in high frequencies among Andean populations and some populations from the Brazilian highlands but is absent from groups in the southern cone (Moraga *et al.* 2000, Ward *et al.* 1996). Haplogroup C is absent in populations from Central America but is located

in groups throughout South America and reaches its highest frequency among the Yanamamõ (68%) from the Amazon basin (Kolman & Bermingham 1997, Easton *et al.* 1996). Haplogroup D is observed at moderate frequencies among one Central American population, the Huetar (26%), and is at high frequencies among groups from the southern cone of the continent (Santos *et al.* 1994, Fox 1996). All four haplogroups may appear in different populations. Several populations lack one or more of the haplogroups reflecting the extent to which evolutionary forces such as genetic drift may have played in forming the mtDNA distribution in modern Native American indigenous populations (Schurr & Sherry 2004).

Antiquity of mtDNA haplogroups in the Americas

There is general consensus that ancestral Native American populations originated in Asia and migrated to the New World over the Bering Land Bridge. However, the Asian starting place or places, the timing, and number of migrations remain a matter of discussion. MtDNA provides researchers with a tool to document the possible timing of this event through mtDNA coalescence dates and comparing human populations to determine when they diverged. The early mtDNA coding region RFLP dates gave a range between 35,000-20,000 years before present (YBP) for haplogroups A, C, and D, and a date of between 17,000-13,000 YBP for haplogroups B and X (Figure 2). These RFLP results suggested two migrations into the Americas, each bringing different founding haplogroups from Asia (Schurr & Sherry 2004). However, data from the mtDNA control HVS-I region indicated that all five haplogroups arrived in the New World at roughly the same time between, 35,000-

20,000 YBP, and indicated a single migratory wave into the Americas (Stone & Stoneking 1998, Bonatto & Salzano 1997). More recent research using data from both the mtDNA coding and control region have suggested that the New World was colonized through at least two migrations. The first migration arrived between 30,000-25,000 YBP and contained ancestral Native American populations found throughout North and South America, while a second migration arrived approximately 14,000-12,000 YBP and contained Beringian populations (Aleuts, Eskimos, and Athapaskans). This latter migration is distinguished by the presence of several distinct mtDNA Native American haplotypes found in Beringian groups but not in other Amerindian populations (Rubicz *et al.* 2003). All of these molecular dates point to an earlier entry into the New World than associated with the Clovis culture, once considered the best archaeological evidence for the earliest inhabitants of the Americas (Schurr & Sherry 2004).

Chibchan Language Family

In addition to the peopling of the Americas, another inquiry that has intrigued anthropologists is the strong parallelism between linguistic and genetic evolution. These correlations between genes and languages have long been used to imply biological relationships between populations. There are two potential explanations for this association: (1) those processes leading to linguistic differentiation also brought about genetic differentiation; or (2) linguistic differences between populations act as a reproductive barrier causing geographically close populations to be genetically distinct (Barbujani & Sokal 1990). These linguistic-genetic associations have been

demonstrated for several geographic areas of the world including, Africa (Excoffier *et al.* 1991), the Arctic (Crawford & Duggirala 1992), Europe (Menozzi *et al.* 1978), North America (Suarez *et al.* 1985), and Siberia (Crawford *et al.* 1997).

However, research into this relationship in South America has often led to discordant results. This has led some researchers to advocate genetic drift as the primary evolutionary force determining population structure on the continent (O'Rourke & Suarez 1986, Black 1991, O'Rourke *et al.* 1992). They argue that the majority of environments in South America are highly marginal to human subsistence patterns, which has led to small population size and greater genetic differentiation between groups. Recently, other researchers have argued that a genetic-linguistic relationship does exist over short distances but occurs rarely over further distances (Luiselli *et al.* 2000, Goichoechea *et al* 2001, Fagundes *et al.* 2002). One Latin American language family that exhibits this latter associative pattern of a linguistic-genetic relationship is Chibchan (Barrantes *et al.* 1990). Populations belonging to the Chibchan language family are informative because they bridge the two American continents. Archaeological evidence indicates that these groups have inhabited the region for the last 10,000 years and therefore understanding their genetic history has important implications for understanding the initial colonization of the Americas by humans.

Populations of the Chibchan language family are distributed from eastern Honduras to the eastern shores of Lake Maricaibo in Venezuela (Hoopes & Fonseca 2003). This family contains fourteen extant and six extinct languages in Honduras,

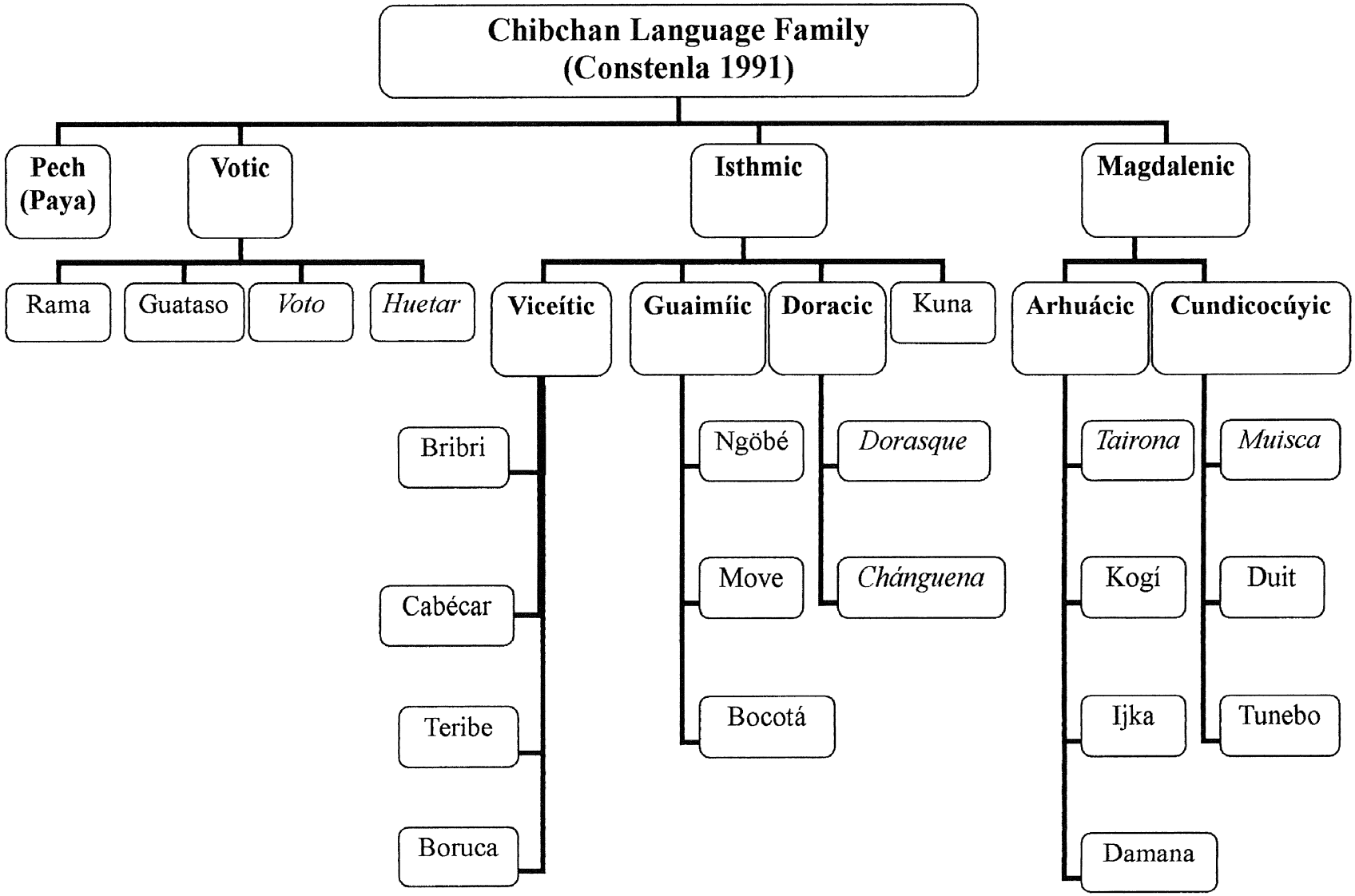


Figure 3: Hierarchical diagram of languages in Chibcha. Subgroups are in bold and extinct languages in *italics*

Nicaragua, Costa Rica, Panama, Colombia, and Venezuela. Figure 3 illustrates the hierarchical structure of the Chibchan language family and shows subgroups, extinct and extant languages. The Chibchan family is divided into four subfamilies and includes: (1) Pech (Paya), spoken in eastern Honduras; (2) Votic, spoken in Nicaragua and northern Costa Rica, containing two extant and two extinct languages; (3) Isthmic, spoken in Costa Rica and Panama, including three subgroups, eight extant and two extinct languages; and (4) Magdalenic, spoken in northern SouthAmerica, comprising two subgroups, five extant and two extinct languages. This language stock is thought to have diverged from related languages in the region approximately 7000 YBP and is thought to have occurred in either Costa Rica or Panama as these regions have higher linguistic diversity then those Chibchan areas from the north or the south (Constenla 1991). The fragmentation of these Chibchan subdivisions is thought to have begun around 5,000 YBP with divisions clearly evident by 4,000 YBP. This division is thought to have occurred due to a shift to agriculture and an adaptation to a sedentary lifestyle that coincided with this time period (Constenla 1991).

Population Background

The majority of archaeological work conducted in lower Central and northern South America has focused on the formation of Chibchan cultures from the Panamanian isthmus and until recently has been less focused on the relationship between populations from lower Central America to those from Northern South America (Quilter & Hoopes 2003, Lange & Stone 1984). Recently, several lines of

evidence have suggested that Chibchan populations initially arose in the lower isthmus of Central America and subsequently migrated out into adjoining regions forming a large, culturally homogenous region that stretched from eastern Honduras into northern South America and blocked migrations from either the north or the south (Hoopes & Fonseca 2003, Kolman & Bermingham 1997, Constenla 1991, Barrantes *et al.* 1990). While linguistic (Constenla 1991) and archaeological (Hoopes 2005a, 2005b, Bray 2003, Hoopes & Fonseca 2003) evidence points to a clear relationship between lower Central America and northern South America the biological relationship between the two regions has yet to be investigated. One Chibchan region in northern South America containing a wealth of comparative archaeological, biological, linguistic, ethnographic, and ethnohistoric information is the Santa Marta mountain range in northeast Colombia. The extant indigenous populations (Kogí, Arsario, Ijka) of the Santa Martas may represent the remnants of the ancient Tairona, one of the largest Chibchan populations present at European contact (Bray 2003, Langeback 2003, Wilson 1999). While all three of these populations demonstrate a well developed archaeological and linguistic relationship with populations from lower Central America (Hoopes & Fonseca 2003, Constenla 1981, 1991) no research has been conducted into their biological relationship with Chibchan populations from the Panamanian isthmus.

Sierra Nevada de Santa Marta ecology

The Sierra Nevada de Santa Marta is an isolated geologic massif that rises to 5,780 meters in a little over 35 kilometers and occurs within the Caribbean littoral

ecological environment of South America. This ecological zone includes isolated mountain ranges of the Andes and runs along the Caribbean coast from the Orinoco river delta in Venezuela to the Isthmus of Panama. The Santa Marta range is pyramidal in shape with a gradually sloped southeastern side, a steep narrow northern side that drops abruptly to Caribbean Sea, and a western side that drops less abruptly into a large swamp; the Ciénaga Grande of Santa Marta. Several rivers originate from lakes high in the range and run down all three sides (Wilson 1999). Reichel-Dolmatoff (1951) separated the northern portion of the range in five climatic zones. These included the: (1) tropical zone, including the coastal plain and lower slopes from sea level to ~1,300 meters; (2) subtropical zone, occurring from ~1,300 meters to ~2,500 meter; (3) temperate zone, from ~2,500 meters to 3,500 meters; (4) *páramo*, from 3,500 meters to 4,500 meters; and (5) snowline, occurring above 4,500 meters. In the coastal regions the average temperature is 24°C and rainfall is from 1,000 to 2,100 millimeters per year, in the subtropical region the temperature is 17°C and annual rainfall is between 2,000 and 4,000 millimeters a years. The average rainfall for the temperate zone is the same as the subtropical but the average temperature is lower, between 11°C and 15°C. The average temperature in the *páramo* is between 3°C and 11°C, so no vegetation beyond short grasses survives in the area. The native vertebrate fauna of the region is varied but not abundant and includes jaguar, puma, tigrillo, tapir, paca, peccary, capybara, armadillo, howler monkey, toucan, heron, pelican, marine turtles, and fish (Wilson 1999).

Archaeology of the Sierra Nevada de Santa Marta

The archeological record of the Santa Marta region indicates that it has been continuously occupied by humans from 3,000 BP to the present. Indications of earlier inhabitants come from two isolated projectile points corresponding to the hunter-gather period (~8,000 BP) (Langeback 2003). However, there are no definitive stratified sites that can verify this date, but archaeological evidence for the 10,000-7,000 YBP era is rare throughout the Americas. Other evidence for a human presence in the area comes from shell middens along the western peripheries of the range and along rivers and the Caribbean coast. These sites have radio-carbon dates between ~5,000 YBP and ~3,000 YBP (Oyuela-Caycedo 1996). Several archaeological sequences have been proposed over the last fifty years for the region and the one presented here is largely an amalgamation of those proposed by archaeologists for the area. The Malambo period represents the first evidence for sedentism in the Santa Marta region. This sedentism indicates stable subsistence patterns and suggests continuous occupation of the region by humans. The earliest evidence for this period comes from a site adjacent to the swamp on the western side of the range near Papare. Ceramics found at this site are identical to those found at the Malambo site (Angulo 1962) and date to between 3,000 and 2,000 YBP (Langeback 2003). Populations during this time period used a wide variety of resources from diverse environments and are characterized by having larger amounts of ceramics relative to lithic material. This time period is also thought to be the start of the manioc-to-maize transition. However, this inference is based on artifact function and no macro-or microbotanical

remains or biochemical analysis of human skeletons has been conducted (Oyuela-Caycedo 1996).

The Neguanje period represents the next unit in the archaeological sequence and occurs from ~2,000 and 1,800 YBP. This period is identified by painted ceramics and the presence of personal adornments in the archaeological record of the region.. Some researchers have suggested the archeological complexity of some sites during this time period point to a rise in tribal societies within the region (Oyuela-Caycedo 1986). These tribal societies are defined through the presence of prestige good manufactured through gold and jade, special cemeteries for the elite, and a rich iconography (Hoopes 2005a). A third period, known as Buritaca, overlaps the later part of the Neguanje and is dated between 1,800 and 1,300 YBP. This period is described from a site from the lower terrace of the Buritaca River excavated by Jack Wynn (1975). The final prehistoric archaeological sequence is known as Tairona and overlaps the end of the Buritaca and continues to after European contact (Bray 2003).

The Ancient Tairona

The term "Tairona" is a Spanish word that refers to "smelting of gold" and is used to refer to the contact period indigenous population of the Santa Marta range and adjacent areas of the Caribbean coast. It also refers to the archaeological context created by their ancestors from 800 A.D. to European contact (Bray 2003). Archaeologically, this period is known for its dense concentration of settlements that occurred from the sea level into the *páramo*. These settlements included complex stone road ways, irrigation drainages, platforms for houses, and terraces for

agriculture (Langeback 2003). Prior to the 1970s the Tairona were largely known for their goldwork and other artifacts, but little was known about their settlement patterns. One of the better known Tairona sites today is Ciudad Perdida (Lost City) that occurs at 1,130 meters in elevation and is located 50 kilometers to the southeast of the modern port city of Santa Marta. This site is situated along a ridge overlooking the Buritaca River and extends over a 20 hectare area. It included a series of man made terraces that range from 50 centimeters to over 30 meters in height and contained dwellings that ranged in size from 5 meters in diameter to over 30 meters. Artifacts found at the site included gold and *tumbaga* (an alloy of copper and gold). The oldest C-14 date for the site is 1,000 YBP. While 20 other nearby sites have not been dated, they have similar artifacts and pottery indicating that they date to the same time period (Wilson 1999). Other archaeological evidence demonstrates the Tairona were involved in a wide trade network that included the extinct Muisca cultural group from near present day Bogotá and westward into Costa Rica and Panama. The trade items included stone, gold, and shell jewelry as well as other perishable items that have not been maintained in the archaeological records (Bray 2003).

Ethnohistoric information collected by Spanish conquistadores and priests from the 1500s indicates that the Tairona were a series of large loosely organized chiefdoms found throughout the region (Bray 2003, Wilson 1999, Reichel-Dolmatoff 1950). Spanish contact first occurred around 1500 A.D. and then intensified after the port of Santa Marta was founded in 1524. The Tairona resisted Spanish colonization

and began a rebellion that was not ended until 1600. The Spanish punished those Tairona communities that had participated in the resistance by burning settlements and prohibited the repopulation of the areas with punishment by death (Uribe 2000). During this time several Spanish soldiers and government officials recorded their visits to the area and described a densely populated area with towns of 20 to 1,000 structures that included ceremonial houses and temples and that major towns were governed by chiefs (Bray 2003). These Spanish chroniclers did not estimate overall Tairona population size but did describe different regions as densely occupied with hundreds of settlements and thousands of individuals. The Spanish recorder, Pedro Simón, noted that within the Caldera valley in the western Santa Marta range contained 250 settlements with an estimated ten thousand dwellings (Wilson 1999). The modern ecological anthropologist David Wilson (1999) used this ethnohistoric information along with archaeological evidence regarding settlement patterns from Ciudad Perdida to infer a Tairona population of 390,000 at European contact. This number was based on an estimated population density of 260 individuals per square-kilometer and then applied to the entire 1,500 square kilometer range of Tairona occupation. Other demographic estimates for the Santa Marta region range from a high of 468,000 (312 individuals per km^2) to approximately 110,000 (75 individuals per km^2) (Langeback 2003). These demographic calculations are arbitrary because they are based on ethnohistoric information that details only contact with Europeans and do not reconstruct population growth or decline over time. The archaeological evidence that is used is primarily based on evidence of settlement patterns from a

single, large-scale site (Ciudad Perdida) and then applied to the whole region of Tairona occupation, which does not account for the potential of differential population density between the inhabited ecological zones of the Santa Marta range.

Extant indigenous populations of the Santa Martas

The current indigenous population size of the Santa Martas is 17,032 (Uribe 2000). This catastrophic decline in population size can be traced to the Tairona rebellion that the Spanish ended in 1600. The Spanish forcibly removed or killed the majority of inhabitants and burned their villages to the ground. However, after the rebellion a few individuals fled to the interior of the Santa Marta Mountains and their descendents are presumed to be the living indigenous populations of the region. All three extant indigenous populations (Kogí, Ijka, Arsario) claim ancestry to the Tairona but this assertion is not without controversy. All three groups speak different languages within the Arhuácic subgroup of Chibcha. These groups also present cultural differences in the way they dress, perform certain rituals, and display subtle phenotypic differences (Uribe 2000). Some anthropologists have argued that these differences necessitate different interpretations regarding the relationships of these populations to one other (Uribe 2000). The discussion that follows presents pertinent cultural and demographic information on each of the Santa Marta populations as well as the neighboring Wayuú.

Kogí

The Kogí (Kagaba, Cogui) are Chibchan speaking seasonal slash and burn horticulturalists who inhabit the northern slopes of the Santa Martas. The Kogí are

divided into a tribal moiety of *Túxe* composed of all men and *Dáke* composed of all women. All *Túxe* and *Dáke* recognize a "totem" and declare themselves related to a certain animal, plant, object, or natural phenomenon. These "totems" are divided into one series which is male (i.e. jaguar-men, fox-men) and another that is female (i.e. armadillo-women, deer-women). A *Túxe* and a *Dáke* can never have the same "totem". Each of these groups has specific cultural rules regarding members of other groups that are eligible for marriage. For example a jaguar-man will marry a pig-woman, because the pig is considered to be the preferred food of the jaguar (Reichel-Dolmatoff 1950). Marriage occurs between the ages of fourteen and eighteen for both men and women. However, men are said to prefer to marry women between the ages of fifty and sixty because they believe that they could learn more from such a marriage partner. This marriage preference was rarely achieved because there are few women of this age, and those who attained that age did not wish to marry naïve young men (Wilson 1999). The men live matrilocally in order to serve the bride's parent for at least one year or until the religious leaders (*Mámas*) decide the time for the couple to live independently (Reichel-Dolmatoff 1950).

There is little demographic information regarding the Kogí from the 1600s to the 1900s. In 1938 and 1941 Park (1965) estimated the number of Kogí at two thousand individuals. This figure coincides well with Reichel-Dolmatoff's (1950) ethnographic research in the late 1940s that estimated two thousand individuals inhabiting the northern slope of the Sierra de Santa Marta. The most accurate current estimate of total population size is 6,138 spread throughout twenty-five communities

(Uribe 2000) but others have suggested a number of about half this size 2,700 (Yunis *et al.* 1994). There are a number of potential reasons for these discrepancies including the isolation of these communities, the xenophobia of the indigenous inhabitants, and that the area is one of the most contested zones of Colombia's forty year civil war, with every major armed group in the country fighting for control of the land (Muse 2004).

Ijka

The Ijka (Ica, Bíntukua, Busínka, Busintana, Arhuaco) are a Chibchan speaking agricultural population who inhabit the San Sebastián valley, the upper Río Fundación, and the Río Templado on the southern side of Sierra de Nevada de Santa Marta. They are the largest Chibchan speaking group in the area and number either 8,700 (Yunis *et al.* 1994) or 9,394 in eleven communities (Uribe 2000). The Ijka have had the greatest amount of contact with the outside world through the work of missionaries and encroachment from Spanish settlers who were interested in the area for agricultural purposes. Beginning in the 1700s Capuchin Catholic missionaries began to work in the area and by the 1940s Catholic priests performed marriage and burial rites with the indigenous religious leaders (*Mámas*) (Layrisse *et al.* 1963). The social organization of the Ijka is based on clans and lineages similar to the one used by the Kogí. However, the Ijka do not follow the "totem" model associated with masculine and feminine lineages (Uribe 2000).

Park (1965) estimated that the number of Ijka in the Santa Marta region was 500 in the 1930s and early 1940s. This differs from present Ijka population estimates

that range between 8,500 and 9,300 (Uribe 2000, Yunis *et al.* 1994). If Park's 1940s estimates are correct this population would need a growth rate of between 0.75 – 0.80 every ten years to account for this rapid increase in size. Based on population estimates given by Spanish chroniclers, the environment could sustain a population several times this size. However, other factors may account for this dramatic population increase including inaccurate estimates, admixture, or ethnic misidentification. It is possible that Park's (1965) 1940s population estimates are wrong because he never actually visited the area and relied primarily on secondary sources. Another potential factor for the rapid increase in size is admixture. The Ijka are the most acculturated of the Santa Marta populations and have been known to intermarry with both Europeans, Africans, and other neighboring Native American groups (Uribe 2000). A third potential reason for this size discrepancy is that the Ijka are often identified in the literature as *Arhuac*, which is phonetically indistinguishable from the term Arawak that refers to a South American and Caribbean language family. The term *Arhuac* was introduced by the Spanish in the 17th century to define the southern slopes of the Santa Marta range and referred to pacified Indians but is now often used to refer to the Ijka (Uribe 2000). This similarity between the two words may have caused significant enumeration problems with regards to ethnic identity.

Arsario

The Arsario (Sanká, Sanhá, Guamaca, Marocaseros, Wiwa) are a Chibchan speaking population who inhabit the southeastern slope of the Sierra Nevada de Santa

Marta. This population is closely related to the Kogí and at times will refer to themselves using the name and occasionally inhabit the same communities. However, the Kogí do not consider the Arsario members of the same population and refer to them as the "some people" (Reichel-Dolmatoff 1950). In the Kogí hamlets of San Francisco, Mamaróngo, Surlibáka, and Cherrúa both groups live side by side with the only difference being that the Kogí live in circular houses and the Arsario in rectangular houses (Reichel-Dolmatoff 1950). The Arsario are the least well known of all the Santa Marta population but we can assume that due to their close relationship with the Kogí, they maintain a similar subsistence pattern, and ideology. Their social organization is identical to the Ijka, based on clans, lineages, and the absence of "totems" (Uribe 2000). There is also little demographic information regarding this population. Park (1965) estimated the total population to be about 500 individuals in the early 1940s. The current population of the Arsario is estimated to be between 1,500 (Uribe 2000) and 2,700 individuals (Yunis *et al.* 1994). This increase reflects a population growth rate between 0.25 and 0.30 every ten years. One potential factor for this growth rate is encroachment from other Santa Marta indigenous populations. According to Uribe (2000), both the Kogí and the Ijka have began moving into to traditional Arsario territory in order to find more arable lands for agricultural purposes and that these two groups have begun to subsume the Arsario.

Wayuú

The Wayuú (Goajira) are a semi-nomadic pastoralist Arawakan speaking population who inhabit the arid La Goajira (11°N, 73°W) peninsula that extends into

the Caribbean Sea (Saler 1992). This population was believed to be either a hunter-gatherer or a horticulturist society until 1550, when they adopted a series of European domesticates including cattle, goats, donkeys, and sheep and became pastoralists. It is unlikely they ever achieved the cultural complexity observed in Arawakan speaking groups located in Venezuela or the Guineas (Layrisse *et al.* 1961). Today, Wayuú society is divided into 30 matrilineal sibships with each of these having a separate chief. Social structure follows a matrilineal avunculocal residence pattern and is typified by extended families with polygynous family households. The biggest division in Wayuú society is between "blood relatives" and "flesh-relatives." "Blood-relatives" are related through the paternal line and "flesh-relatives" are related maternally with stronger bonds being between "flesh-relatives". Varying degrees of consanguineous marriages are allowed except ego's mother's sisters children. Third degree marriages are allowed except for ego's sister's daughters children, and all fourth degree marriages are allowed (Layrisse *et al.* 1961). They represent one of the largest current extant indigenous populations in South America and number approximately 80,000 (Yunis *et al.* 1994).

Genetics of Chibchan populations

Central American populations

The bulk of genetic research on Chibchan speaking populations has focused on populations from lower Central America in order to determine the relationship of these inhabitants to other Native American groups to the north and south. Due to their geographic location, bridging the two continents, between the complex cultural

societies in Mesoamerica and the Andes, it was thought that these groups would demonstrate high genetic diversity as evidence of large amounts of gene flow through the region. However, all of the 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 shared with neighboring populations. Barrantes *et al.* (1990) investigated 48 genetic loci from eight Chibchan speaking groups from Costa Rica and Panama (Boruca, Bribri, Cabecar, Guataso, Teribe, Kuna, Guaymi, Bokota). Using a standard genetic distance matrix converted to linear form correlated with time since divergence and compared with linguistic and geographic location data for the same groups, they found that geographic proximity accounts for some of the genetic and linguistic divergence but the correspondence between phylogenetic affinity and geographic location was not high. They also found a high number of private genetic polymorphisms including an absence of the Diego blood group A* allele (DiA*) in all except one group and high frequencies of transferrin D-Chi allele, the G6PD C allele and five regionally restricted variants. Using this evidence they argued that the large number of private polymorphisms indicated a long term *in situ* development of Chibchan populations over the last 10,000 years and that "waves of migration are not compatible with either the genetic and linguistic data or with the archaeological history of the region" (Barrantes *et al.* 1990: 80). The absence of the DiA* allele had earlier been suggested as an important genetic marker differentiating North and South American Amerind populations (Layrisse & Wilbert 1961). Bieber *et al.* (1996) looked at the six Central American

Chibchan (Guyami, Bribri, Cabecar, Huetar, Teribe, Guataso) populations and four protein polymorphisms (transferrin (TF), alpha-1-antitryrpsin (P1), α2-HS-glycoprotein (ASHG), human coagulation factor (F13B)) that distinguished continental human populations and tested whether they could be used to differentiate Amerindian groups. They found that low genetic diversity (G_{st}<0.05) for ASHG and F13B were the most useful for separating Amerindian populations. Using cluster analysis they found that Chibchan-speakers grouped together but not with neighboring populations, and indicated reproductive isolation of these populations. Layrisse *et al.* (1995) investigated twenty-two Chibchan and Páezan speaking populations and found a general correspondence between genetic and linguistic data with higher affinity between Chibchan speaking populations. They estimated the antiquity of these Chibchan-speakers at a minimum of 7,000 YBP. All of these studies suggested that these Chibchan populations have greater affinity with neighboring linguistically similar populations than they have with other neighboring Central, South, or North American indigenous populations.

Evidence from mtDNA research on Central American Chibchan populations belonging to the Votic and Isthmic linguistic subfamilies also support the theory of *in situ* genetic evolution that occurred between 10,000-7,000 YBP. Several studies (Torroni *et al.* 1993b, 1994, Santos *et al.* 1994, Batista *et al.* 1995, Kolman *et al.* 1995, Kolman & Bermingham 1997) have investigated mtDNA haplogroup data in these populations and discovered that they demonstrate a distinct pattern of founding Amerindian haplogroup frequencies. These groups are characterized by high

frequencies (>65%) of haplogroup A and moderate frequencies (20% - 30%) of haplogroup B and absence of haplogroup C; only two populations (Huetar, Boruca) demonstrated presence of haplogroup D. All of the Huetar individuals with haplogroup D were from the same island (Santos *et al.* 1994) and the haplogroup was present in only one Boruca individual (Torroni *et al.* 1994). Kolman *et al.* (1995) argued that the absence of haplogroup C in any of the lower Central American populations indicated that it was not present during Chibchan genetic history. Torroni *et al.* (1994) used high resolution RFLP analysis to study seven Chibchan (Teribe, Guataso, Boruca, Kuna, Guyami, and Bribri/Cabecar) populations as a reference point for a wider study of eighteen Amerindian populations in order to develop a mtDNA "clock" for the peopling of the Americas. They found a total of fifteen haplotypes of which eleven were unique to Chibchan speaking populations. They also found a unique loss of a *Msp* I restriction site at mtDNA nucleotide 104 that occurred only in Chibchan populations. Three studies have investigated mtDNA HVS-1 sequence variation in three (Huetar, Ngöbé, Kuna) of these populations (Santos *et al.* 1994, Batista *et al.* 1995, Kolman *et al.* 1995). These studies all obtained similar results and showed low haplogroup and nucleotide diversity. The authors argued that these results indicate that these populations had been reproductively isolated from outside groups and that independent genetic evolution had occurred in these lower Central America populations. Kolman and Bermingham (1997) compared two of these populations (Kuna, Ngöbé) to two Chocoan (Emberá, Wounan) speaking populations of the Chocoan linguistic family who inhabit the Pacific coast of Panama and

Colombia. They found that the Chocoan populations demonstrated larger haplogroup and nucleotide diversity similar to other Native American populations and concluded that "ideas have flowed across the isthmian region more successfully than genes" (Kolman & Bermingham 1997: 1301). A small number of mtDNA coalescent dates for Central American Chibchan populations have been published. These include the Ngöbé, 6800 YBP (Kolman *et al.* 1995), the Kuna approximately 10,000 YBP (Batista *et al.* 1995), and a combined haplogroup A date for both group of 10,900 BP (Kolman & Bermingham 1997). These mtDNA coalescent dates have often been used to support the idea that Chibchan populations arose early during the peopling of the Americas and then compared to archaeological and glottochronological evidence to imply long-term population continuity of Chibchan-speakers within the region.

Santa Marta populations

Seven recent studies have investigated genetic diversity in the three Santa Marta Chibchan populations and the neighboring Wayuú (Salzano & Callegari-Jacques 1988, Yunis *et al.* 1994, Briceño *et al.* 1996, Briceño *et al.* n.d, Guarino *et al.* 1999, Mitchell *et al.* 2000, Keyeaux *et al.* 2002). Five of these studies focused on admixture between African, European, and Native American populations and the function that it has played in the development of these populations over the last 500 years (Salzano & Callegari-Jacques 1988, Yunis *et al.* 1994, Briceño *et al.* 1996, Guarino *et al.* 1999, Mitchell *et al.* 2000). Two of these studies investigated mtDNA haplogroup diversity in these four populations and other indigenous populations from

Colombia in order to elucidate their role in the peopling of South America (Briceño *et al.* n.d , Keyeaux *et al.* 2002).

Three of these studies used "classical genetic" markers (Salzano & Callegari-Jacques 1988, Yunis *et al.* 1994, Briceño *et al.* 1996) to investigate admixture in the four study populations (Kogí, Ijka, Arsario, Wayuú). Salzano and Callegari-Jacques (1988) utilized data from two previous studies to investigate European and African admixture in the Wayuú (Layrisse *et al.* 1961) and the Ijka (Layrisse *et al.* 1963). They investigated a number of genetic markers including blood groups, red cell enzymes and red cell proteins. Table 2 shows the results of this study that found a small amount of admixture from African (7%) and European (7%) populations in the Ijka and a high degree of admixture (43% European, 39% African) from both groups in the Wayuú. Yunis *et al.* (1994) investigated seven HLA (DQA1, DQB1, DRB1, DRBR, DRB3, DRB4, DRB5) alleles and eight blood group systems (ABO, MNSs, Rhesus, Kell, Duffy, Kidd, Diego) in the three Santa Marta Chibchan populations and the Wayuú. They found no evidence of admixture among the Kogí or the Arsario, a miniscule amount of European (<1%) and a high amount of African (22%) admixture in the Ijka. High levels of admixture were found from both Europeans (36%) and Africans (24%) in the Wayuú. Briceno *et al* (1996a) investigated variation in the HLA-DPB1 locus in the same four populations. They found no evidence of admixture in the Kogí, Arsario, or Ijka but found an exotic allele (*0101) in the Wayuú, indicating admixture with African populations. Two of these studies used molecular mircosatellite markers to investigate admixture in these populations (Guarino *et al.*

1999, Mitchell *et al.* 2000). Guarino *et al.* (1999) used nine microsatellite markers (D2S1358, vWa, FGA, D8S1179, D21S11, D18S51, D5S818, D13S317, D7S820) and found no evidence of admixture in either the Kogí or Arsario. The Ijka were genetically close to these two groups but contained some levels of admixture and the Wayuú demonstrated the highest amount of admixture. However, they were unable to determine whether the admixture came from European or African populations. Mitchell *et al.* (2000) investigated the D12S67 allele and compared these four groups to other worldwide populations (Minihasans, Torajans, Chinese Han, and Europeans). They found no levels of admixture in the Kogí or the Arsario, high levels of admixture in the Wayuú and moderate levels in the Ijka. The general consensus of these studies is that no admixture is present in either the Kogí or the Arsario, there is some admixture in the Ijka and the Wayuú demonstrated considerable amounts of gene flow from both European and African populations.

Population	European	African	Amerindian	Source[1]
Arsario	0	0	1.00	2
Kogí	0	0	1.00	2
Ijka	.058	.070	.872	1
	.0025	.2166	.7809	2
Wayuú	.425	.39	.185	1
	.3633	.2423	.3934	2

Table 2: Admixture data from three Santa Marta Chibchan populations and the Wayuú. Sources: 1= Salzano & Callegari-Jacques (1988), 2= Yunis *et al.* 1994

Two studies (Yunis *et al.* 1994, Guarino *et al.* 1999) contained data that are important in understanding Chibchan genetics but were not fully addressed in the original articles. Yunis *et al.* (1994) found an absence of the DiA* allele in both the Kogí and the Arsario but the found a relative high frequency of the allele (0.1852) in

the Ijka, and a very small frequency (0.0082) in the Wayuú, a neighboring Arawak speaking group. The loss of this allele has previously been used as an important diagnostic marker for Chibchan speaking populations (Barrantes *et al.* 1990), and it is not found in two of the Santa Marta groups (Kogí, Arsario). Previous studies have demonstrated admixture in the Ijka, and this may explain the moderate frequency of the allele found in this population. However, it does not explain the low frequency in the Wayuú, an Arawak speaking group and presents the question of how accurate a diagnostic marker this is for Chibchan populations. Guarino *et al.* (1999) investigated five Y-chromosome (DYS19, DYS390, DYS391, DYS392, and DYS393) markers on a small number of males from the four study populations. They found the largest number of haplotypes in the Wayuú (8) and the lowest number in the Ijka (3). They found sharing of haplotypes among the Arsario, Ijka, and Kogí but not with the Wayuú indicating a shared evolutionary history with the three Chibchan groups but not with the neighboring Arawakan group.

Two previous studies (Keyaux *et al.* 2002, Briceño *et al.* n.d) have investigated mtDNA haplogroup diversity in the four study populations (Arsario, Ijka, Kogí, Wayuú). Table 3 presents the mtDNA RFLP results of these two investigations. Keyaux *et al.* (2002) investigated all four groups as part of a larger examination of mtDNA haplogroup diversity in 25 indigenous Colombian populations. The results of this study found that the three Santa Marta Chibchan populations only contained haplogroups A and C. The neighboring Wayuú possessed three of the four haplogroups (A, B, C). Briceño *et al.* (nd) also investigated mtDNA haplogroup

diversity in these same four groups. Their results showed that two of the Santa Marta Chibchan populations (Kogí and Arsario) contained only haplogroups A and C. The Ijka and the Wayuú contained three of the four haplogroups. A small number of Wayuú (4%) possessed a haplogroup of unknown origin that was thought to represent a previously unknown mtDNA haplogroup, designated haplogroup F. All of the Santa Marta Chibchan speaking populations shared high frequencies of haplogroup A with groups from lower Central America however they differ in the presence of haplogroup C, which is not found in Panamanian and Costa Rican Chibchan speaking populations.

		% Native American mtDNA haplogroups				
Population	**N**	**A**	**B**	**C**	**D**	**Source***
Arsario	8	25	0	75	0	1
	50	64	0	36	0	2
Kogí	30	34	0	64	0	1
	50	58	0	42	0	2
Ijka	40	82.5	0	17.5	0	1
	40	92	3	5	0	2
Wayuú	20	25	15	60	0	1
	55	31	23	42	0	2**

Table 3: mtDNA RFLP data from three Santa Marta Chibchan populations and Wayuú.
***Sources: 1= Keyaux *et al.* 2002, 2= Briceño *et al*, **= Wayuú 4% unknown haplogroup F**

A number of important observations can be made from these earlier studies. Firstly, there is consistent evidence of distinctive genetic characteristics that distinguish Chibchan-speakers from other Native American populations in the region. The Central American Chibchan-speakers possess a number of private polymorphisms, low amounts of mtDNA haplotype diversity and mtDNA coalescent dates indicating long-term occupation of the Panamanian isthmus. Several studies

among Santa Marta Chibchan-speakers have shown that the Kogí and Arsario are non-admixed and that the Ijka show some evidence for gene flow. Research into mtDNA diversity among the Santa Marta groups demonstrates a reduced amount of Native American haplogroups compared to neighboring indigenous populations.

Models for the peopling of the Santa Martas

There are several proposed hypotheses for the peopling of the Santa Marta region. These include the hypothesis that modern and ancient cultures in the region do not fit the South American pattern and their cosmological and philosophical patterns indicate a migration from Mesoamerica prior to European contact (Reichel-Dolmatoff 1965). A second hypothesis is that close contact and mutual influence linked the populations of the Santa Martas with populations from lower Central America and along the Caribbean coast to form a Chibchan "interactions sphere" that extended from eastern Honduras into Colombia. However, the closet archaeological parallels of the both Neguanje and Tairona periods occur along the Atlantic coast of Costa Rica and not with those from Panama (Hoopes 2005b, Bray 2003). A third model suggests that these cultures evolved *in situ* without influence from other regions (Langeback 2003, Bray 1984). A final model suggests that these populations represent Chibchan populations that migrated into the region between 300-600 A.D., due to drastic environmental events that changed the ecology of the region and led to the differential access to resources and a social hierarchy primarily controlled by religious elite (Hoopes 2005a, Bray 2003, Oyuela-Caycedo 1987).

Chapter III: Materials and Methods

Population Background

This research investigated mtDNA inter- and intra- group diversity in three Chibchan speaking populations (Arsario, Ijka, Kogí) from the Sierra Nevada de Santa Marta and one Arawak speaking group (Wayuú) from the Goajira peninsula in northeast Colombia. Figure 4 shows the geographic distribution of the four study populations as well as other lower Central and northern South American indigenous populations. The molecular techniques included polymerase chain reaction (PCR), restriction fragment length polymorphisms (RFLP), and sequencing of mtDNA HVS-I. The analytical techniques included measures of selective neutrality (Fu 1997, Tajima 1989), R-matrix (Harpending & Jenkins 1973), median joining network phylogenetic construction (Bandelt *et al* 1995, 1999), analysis of molecular variance (AMOVA) (Excoffier *et al.* 1992), and coalescent time measures (Rogers & Harpending 1992, Saillard *et al.* 2000). The total sample size for the study was 190 with the following population sample sizes: Kogí (n=50), Ijka (n=40), Arsario (n=50), and Wayuú (n=50).

Population Samples:

The biological samples utilized during this research were collected by researchers from Javeriana University, Bogotá, Colombia, during the multidisciplinary "Expedicíon Humana" project. After informed consent was obtained, blood samples were collected from unrelated individuals (Uribe 2000). Table 4 presents information regarding the study populations, the community where

biological samples were collected, the Colombian department (province) of the community, and year when samples collected. This thesis project was approved by the University of Kansas Human Research Committee (Hann, personal communication)

Figure 4: Geographic distribution of the four populations used in this study as well as other extant and extinct lower Central and northern South American populations. Chibchan-speakers in dark red, Wayuú in red, Chocoan speakers in black. Tairona and Muisca refer to extinct Chibchan-speaking populations.

Population	Community	Department	Year
Arsario	Umandita	La Guajira	1992
Ijka	Las Cuevas	Cesar	1992
Kogí	Nibiyaca	La Guajira	1992
Wayuú	Kareme	La Guajira	1990
	Manuare	La Guajira	1990
	Macú	La Guajira	1990
	Uribia	La Guajira	1990
	Carizial	La Guajira	1990
	Portete	La Guajira	1990

Table 4: Study Population, the community where biological samples were collected, Colombian department of collection community, and year when samples were collected.

DNA Extraction:

Researchers at the University of Newcastle upon Tyne extracted DNA from whole blood using the phenol chloroform method (Briceño *et al* 1996). DNA samples were then sent to the Dr. Michael H. Crawford at the University of Kansas by Dr. Surindar S. Paphia. The original buffer used to suspend the DNA in the original extraction had evaporated. Due to this desiccation the DNA samples were rehydrated in 50-75μL of nuclease free water (ddH_2O).

Laboratory Methods:

Polymerase Chain Reaction (PCR)

Since polymerase chain reaction (PCR) was first conceptualized and described by Kerry Mullis and others at Cetus Corporation in the mid-1980s (Saiki *et al.* 1985), it has become the most widespread technique used in molecular biology to date (Devor 2004). The components of a common PCR reaction are water, a PCR reaction buffer, forward and reverse primers, magnesium chloride ($MgCl_2$), a DNA polymerase enzyme (e.g., *Taq*), a mix of deoxynucleotide triphosphates (dNTPs) and target DNA. The water is primarily used as the medium in which the reaction will occur. However, due to the high potential for contamination, it is now common practice to use nuclease free water (ddH_2O). Hence ddH_2O was used for all PCRs conducted during this research. The reaction buffer is used to provide an optimal pH environment for the reaction to occur in. The purpose of the $MgCl_2$ is to supply the Mg^{++} cations necessary for Type II enzymes, which include restriction endonucleases and the DNA polymerases used in PCR. dNTPs provide the raw materials necessary

for the synthesis of new strands of DNA. The forward and reverse primers are used to specify a location of the region to be amplified within the target DNA. In order to denature the original double stranded DNA molecule at a high temperature, a stable thermophillic DNA polymerase enzyme is required. The most common DNA polymerase enzyme used in PCR is from the aquatic species *Thermus aquaticus* (*Taq*), which inhabits heated pools in and around Yellowstone Park. The final and most important component for PCR is the target DNA. In order to prevent contamination, control samples must be utilized because the PCR method does not discriminate between the target DNA and any other nucleic acid (Devor 2004). In this study, positive and negative controls were used in all analysis. Positive controls included DNA samples with known Native American haplogroups and negative controls were run with ddH_2O.

Components for the PCR reaction included 2.5μL of 10X PCR Buffer (provided by manufacturer), 4.0μL of $MgCl_2$ (25mM), 0.5μL of dNTP nucleotide mix, 0.2μL of *Taq* Polymerase, 9.8μL of ddH_2O, 2.5μL of forward Primer (10 pmoles/μL), 2.5μL of reverse primer (10 pmoles/μL), and 3.0 μL of diluted DNA. This resulted in 25μL reaction volume for each sample. All reagents were purchased from Promega (Madison, WI) except for the oliglionucleotide primers, which were synthesized at Integrated DNA Technologies (Coralville, IA). Primers and their associated mtDNA haplogroups used in this analysis are listed in Table 5. Reactions were run on a Perkin Elmer 2400, an Applied Biosystems 9600, or a Biometra PC48 thermocycler.

Primer	Sequence	A.T.*
	Primer Pair for Haplogroup A (+*Hae*III 663)	
535 FOR	5' CCC ATA CCC CGA ACC AAC C 3'	55°C
725 REV	5'GGT GAA CTC ACT GGA AGG GG 3'	
	Primer Pair for Haplogroup C (-*Hinc*II 13259 & +*Alu*I 13262)	
13172 FOR	5'GCT TAG GCC CTA TCA CCA 3'	51°C
13383 REV	5'GTT GTG GAT GAT GGA CCC 3'	
	Primer Pair for Haplogroup B (+*Hae*III 8250)	
8149 FOR	5'ACC GGG GGT ATA CTA ACG GT 3'	51°C
8366 REV	5' TTT CAC TGT AAA GAG GGT TGT TGG 3'	
	Primer Pair for Haplogroup D (-*Alu*I 5176)	
5151 FOR	5' CTA CTA CTA TCT TCG CAC CTG 3'	53°C
5481 REV	5' GTA GGA GTA GCG TGG TAA G 3'	
	Primer Pair for HVS-1	
15976 FOR	5' CCA CCA TTA GCA CCC AAA GCT AAG 3'	55°C
16422 REV	5' ATT GAT TTC ACG GGA GGA TGG 3'	
	Primer Pair for Chibchan specific RFLP (-*Msp*I 104)	
1 FOR	5'GAT CAC AGG TCT ATC ACC CT 3'	47°C
240 REV	5'TAT TAT TAT GTC CTA CAA GC 3'	

Table 5: Primer pairs for four Native American founding mtDNA haplogroups (A-D), one Chibchan specific haplotype (*MspI*) and HVSI sequence used in this analysis. *A.T.=Annealing temperature.

PCR reactions were run according to the following three step thermal profile: (1) denaturing at 94°C for forty seconds; (2) annealing for thirty seconds at the lowest melting temperature for each primer pair (see Table 5); (3) extending at 72°C for forty-five seconds. These steps were repeated for thirty five cycles. Reactions were checked for PCR product on a 1.5% agarose amplification gel using electrophoresis at

ninety-seven volts for approximately one hour. Depending on the size of the gel box, the reagents for the gel included 100μL or 150μL of 1X TBE, 1.5g or 2.25g or SeaKem agarose, cooled to 45°C and stained with ethidium bromide. A total of 5μL of PCR product and 2μL of 6X loading dye (provided by manufacturer) was added to each well and checked against a 25bp or 100bp DNA ladder (Promega, Madison, WI). DNA was visualized though ultraviolet light and photodocumented.

Restriction Fragment Length Polymorphisms (RFLP)

After amplification, base substitutions or insertion/deletion events were characterized. These were detected with restriction endoculeases, which are bacterial enzymes that cut the DNA at a specific recognition sequence that is typically 4-6 base pairs long. These enzymes cleaved the DNA at this specific location and left a 5', 3'or no overhang. Size variation in these restriction fragments was then detected by gel electrophoresis (Dowling *et al.* 1996). This research investigated the restriction site variation of samples characterized for each of the four major Amerindian mtDNA haplogroups (A, B, C, & D). These four haplogroups are characterized by a restriction cut site determined by a specific restriction enzyme (*Hae*III, *Hinc*II, *Alu*I). An additional RFLP was examined for the loss of an *Msp* I cut site at mtDNA nt104 that previously had been detected only in Chibchan speaking populations from lower Central America (Torroni *et al.* 1994). The restriction enzymes and associated mtDNA haplogroup used in this study are listed in Table 5.

The reagents for the restriction digest included 2.0μL of 10X RFLP buffer (provided by manufactuer), 1.0μL of 100X bovine serum albumin (BSA), 0.5μL of

restriction enzyme (New England Biolabs Beverely, MA), 9.0μL of ddH_2O, and 7.5μL of PCR product DNA. This resulted in a reaction volume of 20μL per sample. All reagents except the ddH_2O water were purchased from New England Biolabs (Beverly, MA). The samples were then digested for ten to eighteen hours at 37°C using one of the three restriction enzymes (*Hae*III, *Alu*I, *Hinc*II). The reaction was stopped by the addition of 5μL of 3X loading dye to each sample. Digested fragments were electrophoresed on an ethidium bromide stained 3% 3:1 NuSieve gel (ISC BioExpress, Kaysville, UT) at 97 volts for two hours. The digested products were visualized on under ultraviolet light and photo documented.

All 190 samples were characterized for the four founding Native American mtDNA haplogroups (A, B, C, D) from South America. Haplogroup A was recognized by the presence of a *Hae*III cut site at mtDNA nucleotide (nt) 663. The haplogroup B 9bp-deletion was characterized by cutting the amplified DNA with *Hae*III at nt8250. Although this particular cut site is present in all humans, verification of the deletion is easier to detect when smaller DNA fragments are compared. Haplogroup C was characterized by the absence of a *Hinc*II site at nt13259 and the presence of an *Alu*I cut site at nt13262. Haplogroup D was characterized by the absence of an *Alu*I cut site at nt5176. A subset of 20 samples each of the three Santa Marta Chibchan groups (Arsario, Ijka, Kogí) that belonged to haplogroup A were further characterized for the absence of the *Msp*I nt104 cut site. Samples were scored with a 1 if a restriction site was absent and a 2 if a site was present.

DNA sequencing:

Four hundred base pairs (nt16001-nt16400) of HVS-I of the mtDNA control region were sequenced using the Sanger dideoxy sequencing method. Twenty-five samples from each of the four study groups were randomly chosen for sequencing. DNA stands were synthesized in a single direction with the forward HVS-I primer (Table 5). Due to a T-C transition at nt16189 that caused the automated sequencing reaction to fail, all samples that contained this mutation were sequenced with the HVS-I reverse primer (Table 5). All DNA sequencing was conducted at Integrated DNA Technologies in Coralville, Iowa, under the direction of Dr. Eric J. Devor. The author was present for the initial sequencing of samples but subsequent samples were conducted by personnel at IDT.

In preparation for the sequencing reaction, DNA templates were created using either a 25 μL (same as haplogroup PCR) or a 50 μL PCR reaction: 37.5μL ddH_2O, 5.0μL of 10X buffer (provided by manufacturer, Sigma-Aldrich, St. Louis, MO), 3.0μL $MgCl_2$,, 1.0μL 10mM dNTPs, 1.0μL forward and reverse primer, 1.0μL DNA sample, and 0.5μL of Taq. This resulted in 50.0μL total reaction volume. Primers and annealing temperature used for DNA sequencing are listed in table 5. Amplified PCR products were then checked on a 1.5% agarose gel

Amplified PCR products were then purified using a QIAquick kit (Qiagen, Valencia, CA) kit. All reagents used in this portion of the sequencing analysis were provided by Qiagen (Valencia, CA). Following manufacturer's instructions, 5:1 ratio of Buffer PB was added to PCR product in order to bind the DNA. Solution was

transferred to a new 1.5mL tube and centrifuged for one minute at 13,000 rpm in order to bind the DNA to a positively charged filter. Contents of tube were discarded and column was placed back in tube. 750μL of wash buffer was placed in column and 1.5mL tube was centrifuged for one minute at 13,000 rpm. Column was transferred to new 1.5mL tube. 30μL of EB buffer was added to column and allowed to stand for one minute. The 1.5mL 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 t tube.

Automated fluorescence method by an Applied Biosystems 310 capillary system sequencer was used to sequence the mtDNA HVS-I region. This technique involves tagging rhodamine derivatives on each of the dideoxynucleotides (ddNTP) terminators (ddG, ddA, ddC, ddT). When a ddNTP is added to the end of a growing strand, polymerization is terminated because it lacks the 3'OH needed for the attachment of the next nucleotide. This resulted in several DNA fragments of varying lengths, which were fluorescently-labeled. These were visualized by exciting the terminators with an argon laser at 488nm producing a peak spectral emission that differs for each specific nucleotide based on differing chemical structures (Devor 2004). An Applied Biosystems (ABI, Foster City, CA) protocol was used for this reaction. All reagents used in this portion of the sequencing analysis were provided by Applied Biosystems (Foster City, CA). The following components were added to a 0.2 mL tube: 4.0μL of terminator ready mix, 2.0μL 5x sequencing buffer, 1.0μL of forward or reverse primer, 11.0μL ddH_2O, and 2.0μL of target DNA. This PCR reaction was run according to the following three step thermal profile: (1) 96°C for

ten seconds (2) 50°C for five seconds (3) 55°C for four minutes, and repeated for twenty five cycles. Unused primers and ddNTPs were removed by passing it through a spin column and dried in a speed-vacuum. Following manufacturer's instructions, dried sample were prepared for sequencing by adding 20μL of ABI template suppression buffer heated to 95°C for three minutes and placed on ice. DNA samples were transferred to ABI tubes and loaded onto an ABI 310 sequencer. The sequencing gels were run overnight and resulting chromatogram data were recorded by a computer.

The mtDNA sequence chromatograms were aligned with Sequencher software (Gene Codes Corporation, Ann Arbor, MI). Resulting sequences from the four study populations were compared against the Cambridge reference sequence (published mtDNA sequence, Anderson *et al.* 1980) and differences between the nucleotide based were recorded as mutations.

Analytical Procedure and Methods:

Measures of Selective Neutrality

In order to determine whether or not resulting sequence data were statistically significant under the neutral equilibrium model of sequence evolution two measures of selective neutrality, Tajima's D (Tajima 1989) and Fu's Fs (Fu 1997), were investigated. This model is based on the standard Wright-Fisher model and makes the following assumptions: (1) a large constant population size of N individuals; (2) random mating; (3) no overlapping generations; (4) no recombination; and (5) an infinite sites constant rates mutation process whereby an offspring differs from its

parent allele by a Poisson-distributed number of mutations with mean μ (Simonsen *et al.* 1995). These statistics are appropriate in distinguishing population growth from constant size populations. Population growth generates an excess of mutations in the external branches of the genealogy and therefore an excess of substitutions that are present in only one sampled sequence (Ramos-Onsins& Rozas 2002). This leads to a star like phylogeny that includes a large central node with several radiating spokes represented by a single individual.

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

$$D = \frac{\theta\pi - \theta s}{\sqrt{Var(\theta\pi - \theta s)}} \qquad (1)$$

where $\theta\pi$ 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 $\theta\pi$ signifying the potential effects of population expansion. However, significant D values may be caused by factors other than population expansion, including population bottlenecks, background selection, or mutation rate heterogeneity (Schneider *et al.* 2000, Aris-Brosou & Excoffier 1996, Tajima 1993)

Fu's Fs (1997) is also based on the infinite-site model without recombination but utilizes information from the haplotype distribution. This test statistic is defined using the equation.

$$Fs = \ln(\frac{S'}{1-S'}) \qquad (2)$$

where S' is the probability of observing a random neutral sample and defined as S'=PR(K≥$k_{obs}|\theta=\theta\pi$), where (k) is equal to the number of alleles similar or smaller than the observed value given θπ and Fs is logit of S'. Negative scores indicate an excess of allele, a signature of population expansion. This test is less conservative than Tajima's D and is more sensitive to large population expansions expressed as large negative numbers whereas positive numbers indicate populations impacted by genetic drift (Schneider *et al.* 2000, Fu 1997). Tests of selective neutrality were generated using the statistical program Arlequin 2.0 (Schneider *et al.* 2000).

R-Matrix

R-matrix analysis was used to determine the relationship between the four study population and neighboring South American indigenous groups using RFLP haplogroup data. R-matrix is a relational statistical technique that allows for the visualization of population structure and history in two or three dimensional space (Harpending & Jenkins 1973). The first step is to calculate a variance-covariance matrix of genetic similarity or dissimilarity between populations, using the equation

$$R_{ij} = (p_i - \overline{p})(p_j - \overline{p}) / \overline{p}(1-\overline{p}), \qquad (3)$$

where R_{ij} is the kinship coefficient for every allele, p_i and p_j are the allele frequencies in populations i and j and $\bar{p}$ is the weighted gene mean frequency of allele p in the matrix R. The matrix R is defined by the equation

$$R=E\Lambda E^T, \qquad (4)$$

where Λ is a diagonal matrix with elements λ_1, λ_2, λ_3 and E are the eigenvalues of R. The first set of eigenvalues is arranged in descending order along the axis and the second set is placed at a right angle to the first maximizing dispersion. A covariance matrix is constructed by multiplying the eigenvectors by the variable for each object. In this way, the R-matrix is similar to a Principal Component Analysis (PCA) and the first two or three eigenvalues explain the majority of variation within the sampled populations. In addition to the R-matrix an S-matrix of covariances was calculated from the eigenvectors of the R-matrix in order to determine the distribution of alleles and the effect that they have on population dispersion. The S-matrix is calculated as the Z-matrix premultiplied by its transpose (Harpending & Jenkins 1973).

Two separate R-matrices were run: (1) comparing these four study populations to one another using mtDNA haplogroup diversity along with data obtained from the literature for four blood groups systems (MNS, RH, Diego, Duffy) (Yunis *et al.* 1994); (2) comparing these four studied populations with mtDNA haplogroup variation in 36 existing Central and South American populations (Table 7). Originally, the R-matrix technique was designed for situations in which numbers of alleles exceed the number of populations studied. However, in this latter analysis a single locus (mtDNA) with multiple alleles was used. The following methodology was

applied because of the violation of the underlying assumptions. If the dimensions of the hyperspace occupied by k populations defined by p allele frequencies is the minimum of k-1 and p so that both the S and R matrices contain all the information regarding kinship and genetic distance. Indicating that if the number of populations (k) are larger than number of alleles (p) then the R-matrix is more informative, however, if alleles (p) are larger than populations (k) the S-matrix is more useful (Harpending & Jenkins 1973:188). Both the R-matrix and S-matrix were calculated with ANTANA (Harpending & Rogers 1984) and PCA with Minitab 12.0 (Minitab, Inc. State College, PA).

Median Joining Network analysis

Phylogenetic network analysis utilizing the median-joining methods was used for determining genetic relationships between haplotypes found within the studied populations. Networks offer an advantage over traditional phylogenetic tree building methods that utilize maximum parsimony or maximum likelihood, because networks can distinguish between irresolvable and resolvable character conflict errors that may occur due to homoplasy. These networks then represent "all most parsimonious trees" by highlighting conflicts in the form of reticulations that can then be interpreted as homoplasy, recombination, or sequence errors (Bandelt *et al.* 1995). Through the addition of consensus points (median vectors) to three mutually close sequences at a time the network is sequentially constructed. These median vectors can then be interpreted as either extinct sequences or extant unsampled sequences within the population. There are four different types of networks, Minimum Spanning Networks

(MSN), Reduced Median (RM), Median-Joining (MJ), and Quasimedian (QM) networks. The minimum spanning network is the simplest to compute but does not generally represent the most parsimonious tree. However, MSNs are the starting point for all higher order networks. QM networks represent all possible combinations and are generally too complex to be visualized in three-dimensional space. The only difference between RM and MJ networks is that RM networks only work with binary data. MJ networks permit the use of multi-state data MJ networks and add an additional tolerance parameter ε that can be adjusted to the level of homoplasy within a population (Bandelt *et al.* 1999). Applicable network analyses for this research are either RM or MJ, but because this analysis deals with multi-state data (nucleotide sequences) and with the addition of the ε parameter, MJ networks were determined the most appropriate.

The assumptions of the MJ method are that ambiguous states are infrequent and that recombination is absent. These assumptions are met for mtDNA RFLP and control region sequence data as well Y-chromosome STRs when a single repeat mutation model is utilized (Bandelt *et al.* 1999). The process of constructing a MJ network is completed through a five-step algorithm after specifying that the tolerance parameter $\varepsilon \geq 0$. The first step involves determining the distance matrix D for difference between sequences. The second step identifies the links between sequence types that are feasible. The third step removes obsolete sequences that were not among the sampled sequences but can be feasibly linked. If obsolete sequences are detected, the second step is repeated. The fourth step involves determining the

feasible triplets where at least two pairs are feasibly linked and at least one median vector is not yet a current sequence type. If no feasible triplets are present, then the algorithm continues to step five. Otherwise, the minimum connection cost λ is computed: and if it does not exceed $\lambda + \varepsilon$, these new median vectors are added to the current sequence types and the process is looped back to step one. The last step is the construction of the final network (Bandelt *et al.* 1999). MJ networks for this research were visualized using the computer program Network 4.0 (www.fluxus-engineering.com) and then reconstructed by hand.

Analysis of Molecular Variance (AMOVA)

Analysis of Molecular Variance (AMOVA) was used to determine whether there was population subdivision among the four study populations or among neighboring populations based on their linguistic affiliation. Several AMOVAs were run that consisted of HVS-I haplotypes of individuals, traced to their linguistic affiliation and geographic place of origin in lower Central or northern South America. AMOVA is analogous to a nested analysis of variance (ANOVA) derived from a matrix of squared distances among all pairs of haplotypes. This in turn produces variance estimates and F-statistic analogs designated as Φ-statistics that reflect the correlation of haplotypic diversity at different levels of hierarchical subdivision (Excoffier *et al.* 1992). This is accomplished though the following equation for the total sum of squared deviations (SSD).

$$SSD_{(Total)} = \frac{1}{2N}\sum_{j=1}^{N}\sum_{k=1}^{N}\delta_{jk}^{2} \qquad (5)$$

where N equals the number of haplotypes, δ^2_{jk} is the Euclidean distance between haplotypes j and k. This allows for the hierarchical partition of the haplotypes into SSD within populations, SSD within regional groups, and SSD among populations within regional groups. The mean squared deviation (MSD) is obtained by dividing the corresponding SSD by the appropriate degrees of freedom. (Excoffier *et al.* 1992). Due to the fact that mtDNA sequence data contain a greater number of transitions, distance matrices were constructed using the Kimura-2 parameter model (Kimura 1980). This model accounts for unequal rates of transitions (T↔C, A↔G) and transversions (T or C ↔ A or G).

Two different AMOVAs were run: (1) using linguistic affiliation and geographic location of the four study populations; (2) using linguistic affiliation and geographic location from ten neighboring lower Central and South American populations (Table 9). This hierarchical analysis partitioned the variance into: (1) variance among population grouped by their geographic location or linguistic affiliation as defined by Ruhlen (1991) or Constenla (1991); (2) variance among populations within these groups; and (3) variance within these populations. Significance values were obtained by bootstrapping data 1000 times. AMOVAs were calculated with Arlequin 2.0 (Schneider *et al.* 2000)

Time Estimate Measures

HVS-I sequences were used to generate two different time estimates based on the mtDNA coalescent (Hudson 1990). The first of these was mismatch analysis, a useful statistical approach for estimating time of expansion for recently established

populations. This method produces histograms based on the number of pairwise differences within a population sample, which theoretically preserves a record of population history including expansion and genetic drift (Harpending *et al.* 1993). Rogers and Harpending (1992) demonstrated that pairwise differences between nucleotide sequences increased by a rate of 2μ (where μ equals the mutation rate) for each generation during population growth. By taking this estimated substitution rate, it is possible to estimate N (population size) of a sample prior to population expansion. It is also possible by using the mtDNA coalescent (Hudson 1990) to estimate the initial timing of population growth in mutational units using the equation

$$\tau = 2\mu t \qquad (6)$$

where t is time in generations and μ is the mutation rate. Then taking the parameters θ_0 and θ_1 as the population estimates before and after expansion, respectively, and fitting these and τ through the least squares method to the observed mtDNA mismatch distribution permits an estimate of expansion time in mutational units over divisions of time (Rogers & Harpending 1992). This analysis was conducted in order to estimate the date of a potential Chibchan diaspora. Mismatch analysis was conducted using the statistical program Arlequin 2.0 (Schneider *et al.*2000) and frequency distributions were calculated by hand and plotted in Microsoft Excel 2003. Time estimates were generated using the mutation rate of 1 event every 9,300 years from Ward *et al.* (1991).

A second time estimator based on the mtDNA coalescent was used to account for the potential problem of mutation rate heterogeneity. A total of 29 sites within the

human mtDNA hypervariable region I mutate at 12 times the rate of other sites in the area (Schneider & Excoffier 1999). Therefore, an alternative time estimate ($\rho \pm \sigma$) that is less susceptible to this mutation rate heterogeneity and based on median joining trees was used to account for clusters within the phylogeny (Saillard *et al.* 2000) and to determine if these time estimates provide a different chronology than obtained from mismatch analysis. These time estimates were generated using Network 4.0 (www.fluxus-engineering.com).

Chapter IV: Results

Restriction Fragment Length Polymorphisms

The RFLP analysis revealed that the four study populations contain three of the five major Native American haplogroups as defined by Schurr *et al* (1990) and Torroni *et al.* (1992). Table 6 six presents the mtDNA RFLP haplogroup data for the four study populations. Two of the populations (Kogí and Arsario) contain only haplogroups A and C. Two Kogí samples did not contain usable DNA reducing the sample size of that population from 50 to 48. The Ijka contain three of the five major haplogroups (A, B, C). The Wayuú also contain three haplogroups but B occurs in higher frequencies and haplogroup A in lower frequencies than in the other three study groups. Five samples from this population could not be assigned to any of the five major Native American haplogroups and may represent either African or European admixture that is known to be present in this population (Yunis *et al.* 1994, Guarino *et al.* 1999). However, these five were not analyzed for characteristic African or European RFLP markers. Haplogroup D is absent from all four of the populations in this study.

Population	N	% A	%B	%C	% D	%Other
Arsario	50	68	0	32	0	0
Ijka	40	90	2.5	7.5	0	0
Kogí*	48	65	0	35	0	0
Wayuú**	50	34	24	32	0	10

Table 6: Percent Native America mtDNA RFLP haplogroups for four study populations. *=2 Kogí samples contained no usable DNA, **=5 Wayuú samples were undetermined haplogroups

Table 7 compares the four study populations to mtDNA haplogroup DNA variation in 36 other South American and lower Central American populations divided by linguistic affiliation as defined by Ruhlen (1991). These data are also used for R-matrix analyses in this study. These populations include eleven additional Chibchan-Paezan speaking groups including seven from lower Central America (Teribe, Guataso, Kuna, Bribri-Cabecar, Huetar, Ngöbé, and Emberá) and four from South American (Zenu, Atacemeño, Cayapa, Yanamamõ), nine Andean speaking populations, eight Equatorial-Tucanonan groups, and seven Ge-Pano-Carib populations. Geographic locations of these populations are shown in Figure 5 with the corresponding population number being present in the first column of Table 7. The three studied Chibchan populations share a high frequency of haplogroup A, especially with those groups inhabiting lower Central America. The Ijka have the highest frequency of haplogroup A (90%) of any of the populations. The Santa Marta Chibchan-speakers differ from Central American Chibchan populations through the presence of haplogroup C. This haplogroup is absent in populations from lower Central America but is found throughout populations in South America. The Santa Marta populations also differ in the absence of haplogroup B from the majority of individuals present in the Santa Marta area. With the exception of one Ijka individual, this haplogroup is virtually absent from these groups. The presence of one haplogroup B individual in the Ijka population may be due to admixture with other Native American populations, but other factors cannot be ruled out. The Wayuú, who inhabit the nearby Guajiro peninsula, demonstrate a moderate frequency of haplogroup B

(24%) and may have introduced it into the population. An additional Chibchan specific RFLP was examined in a subset of 20 A individuals from each of the Santa Marta populations. The absence of an *Msp*I cut site at mtDNA nucleotide 104 had been found exclusively in Chibchan speakers from Central America (Torroni *et al.* 1994). All sixty individuals from the three Chibchan populations here contained the *Msp*I cut site, indicating that this mutation may have occurred after Chibchan ethnogenesis.

DNA Sequencing

The HVS-1 sequencing results for 61 (18 Arsario, 8 Kogí, 19 Ijka, and 17 Wayuú) individuals are shown in Table 8. Due to the degraded nature of the DNA it was not possible to obtain sequences for all 104 individuals. A total of 18 different haplotypes, characterized by 25 different variable sites were observed despite this degradation. The haplotype SMA7 is the most common mtDNA lineage, present in 20 individuals, followed by SMC3 (8 individuals), SMA3 (7 individuals), SMA1 (5 individuals), SMC1 (4 individuals), and SMB1 (3 individuals). The remaining twelve haplotypes are unique and account for only one or two individuals each. Three haplotypes are shared between the Chibchan-speaking populations SMA3 and SMC3 are shared by the Kogí and the Arsario and SMA7 is shared by the Ijka and the Arsario. The Wayuú do not share haplotypes with any of the three Chibchan speaking groups.

Seventeen of the mutations are transitions and eight are transversions. Control region sequences belonging to haplogroup A demonstrate the most variability and

Population	LAF*	N	% A	%B	% C	% D	%OTH	Ref.**
(1)[#] Teribe	CP	20	80	20	0	0	0	1
(2) Guataso	CP	20	85	15	0	0	0	1
(3) Kuna[pool]	CP	79	77	23	0	0	0	2
(4) Bribri-Cabecar	CP	24	54	46	0	0	0	1
(5) Huetar	CP	27	70	4	0	26	0	3
(6) Ngöbé	CP	46	67	33	0	0	0	4
(7) Cayapa	CP	120	29	40	9	0	22	2
(8) Atacemeño	CP	50	12	72	10	6	0	6
(9) Yanamamõ	CP	107	0	8	68	15	9	1
(19) Emberá	CP	22	73	22	0	0	5	7
(22) Zenu	CP	37	19	41	3	5	5	7
Kogí	**CP**	**48**	**65**	**0**	**35**	**0**	**0**	**This Study**
Ijka	**CP**	**40**	**90**	**2.5**	**7.5**	**0**	**0**	**This Study**
Arsario	**CP**	**50**	**68**	**0**	**32**	**0**	**0**	**This Study**
Chibchan Total		**690**	**46.7**	**24**	**19.7**	**4**	**5.6**	
(28) Quechua[pool]	AD	51	20	61	8	12	0	6, 8
(10) Aymara[pool]	AD	205	5	72	11	12	0	6, 8
(11) Mapuche[pool]	AD	208	5	20	33	39	3	9, 10, 11
(12) Huilliche[pool]	AD	118	4	29	19	48	0	2
(13) Pehunche[pool]	AD	205	2	8	40	50	0	6, 11
(18) Yahgan	AD	21	0	0	48	52	0	11
(25) Fuegian	AD	45	0	0	42	56	0	12
(22) Ingano	AD	27	15	44	37	0	4	7
Pueños	AD	65	12	65	8	15	0	13
Andean Total		**945**	**5.5**	**34.3**	**25.7**	**33.6**	**.8**	
(23) Ticuna[pool]	ET	82	15	10	36	39	0	1, 7
(14) Zoro	ET	30	20	7	13	60	0	14
(15) Gavião	ET	27	15	15	0	70	0	14
(20) Wayuú	ET	40	25	35	38	0	2	7
Wayuú	**ET**	**50**	**34**	**24**	**32**	**0**	**10**	**This Study**
(32) Ignaciano	ET	22	18	36	41	0	5	8
(33) Trinitarion	ET	35	14	40	37	3	6	8
(34) Movima	ET	22	9	9	64	18	0	8
(35) Yuracare	ET	28	39	32	21	4	4	8
Equatorial Total		**336**	**36.7**	**21.8**	**31.7**	**22.3**	**2.9**	
(16) Mataco[pool]	GPC	129	8	54	9	27	2	1, 15, 16
(27) Toba[pool]	GPC	56	20	41	5	30	4	15
(24) Chorote[pool]	GPC	34	15	44	23	18	0	11, 16
(17) Xavante	GPC	25	16	84	0	0	0	14
(29) Pilaga	GPC	26	27	35	4	35	0	15
(30) Chimane	GPC	41	39	54	5	0	2	8
(31) Moseten	GPC	20	40	55	0	0	5	8
Ge Total		**331**	**18.6**	**51.5**	**7.6**	**20.2**	**2.1**	
Total		**2302**	**24.3**	**31.9**	**22.2**	**21.2**	**2.7**	

Table 7: South American Indigenous population mitochondrial DNA haplogroup frequencies

= Refers to geographic location shown in Figure 5. * = Refers to Linguistic affiliation as defined by Ruhlen (1991), CP=Chibchan Paezan, AD=Andean, ET=Equatorial-Tucanoan, GPC=Ge-Pano-Carib **= References 1= Torroni *et al.*1993b, 2=Rickards *et al.*1999, 3=Santos *et al.*1993, 4=Kolman *et al.*1995, 6=Merriwether *et al.*1995, 7=Mesa *et al.*2000, 8=Bert *et al.*2001, 9=Ginther *et al.*1993, 10=Bailliet *et al.*1994, 11=Moraga *et al.*2000, 12=Lalueza *et al.*1997, 13=Dipierri *et al.*1998, 14=Ward *et al.*1996, 15=Demarchi *et al.*2001, 16=Bianchi *et al.*1995

Figure 5: Geographic location of 36 South American study populations. Numbers correspond to populations from Table 7.

contain twelve haplotypes. All lineages belonging to this haplogroup contain the 16111C→T, which place them all in the A2 subgroup of haplogroup A, found only in Native American populations. These haplogroup A lineages also contain the following three transitions: (1)16290 C →T; (2)16319 G → A; and (3)16362 T → C. Ten of the haplotypes have the 16223 C → T transition, six haplotypes have the 16129 G →A transition, and four haplotypes have the 16189 T → C transition. All of the other haplotypes are defined by only one or two other mutations. None of the A haplotypes found in this study are known to be shared by other indigenous populations from South or Central America.

Haplogroup C accounts for four of the eighteen haplotypes found in this study. All four haplotypes contain the 16223 C→ T transition, a 16298T → C transition, a 6325 T → C transition, and a 16327 C → T transition. The two Chibchan speaking populations (Kogí and Arsario) that represent this haplogroup also have an additional 16265 A → G transition that the Arawakan speaking Wayuú lack. The Wayuú share haplotype SMC1 with two other South American populations, the Yanamamõ and the Mapuche, and this corresponds to haplotype 24 shown in Table 9. All other haplotypes for this group are represented by a single mutation. Haplogroup B accounts for two of the eighteen haplotypes. Both these haplotypes have a 16189 T → C transition and a 16217 T → C transition. The only population to demonstrate B haplotypes is the Wayuú. The haplotype SMB1 is shared with four Central American (Huetar, Kuna, Ngöbé, and Emberá) and two South American populations (Yanamamõ, Xavante) and corresponds to the haplotype 01 shown in Table 9.

Haplotype	16054	16059	16066	16078	16111	16129	16184	16189	16202	16214	16217	16223	16233	16235	16250	16265	16290	16297	16298	16311	16319	16325	16327	16330	16362	Arsario	Ijka	Kogí	Wayuú	Total #
CRS*	A	A	A	A	C	G	C	T	A	C	T	C	A	A	C	A	C	T	T	T	G	T	C	T	T					
SM A1	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	A	-	-	-	C	-	-	-	5	5
SM A2	-	-	G	-	T	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	A	-	-	-	C	-	-	-	1	1
SM A3	-	-	-	-	T	-	-	C	-	-	-	T	-	-	-	-	T	-	-	-	A	-	-	-	C	4	-	3	-	7
SM A4	-	-	-	-	T	-	-	C	-	G	-	T	-	-	-	-	T	-	-	-	A	-	-	G	C	1	-	-	-	1
SM A5	-	-	-	-	T	-	-	C	-	-	-	T	-	-	-	-	T	-	-	-	A	-	-	G	C	1	-	-	-	1
SM A6	T	-	-	-	T	-	-	C	-	-	-	T	-	-	-	-	T	-	-	-	A	-	-	-	C	-	-	1	-	1
SM A7	-	-	-	-	T	A	-	-	-	-	-	T	-	-	-	-	T	-	-	-	A	-	-	-	C	6	14	-	-	20
SM A8	-	T	-	-	T	A	-	-	-	-	-	T	-	-	-	-	T	-	-	-	A	-	-	-	C	-	2	-	-	2
SM A9	-	T	-	-	T	A	-	-	-	-	-	T	-	T	-	-	T	-	-	-	A	-	-	-	C	-	1	-	-	1
SM A10	-	T	-	-	T	A	-	-	-	-	-	T	T	-	-	-	T	-	-	-	A	-	-	-	C	-	1	-	-	1
SM A11	-	-	-	-	T	A	-	-	-	-	-	T	-	-	-	-	T	C	-	-	A	-	-	-	C	-	-	1	-	1
SM A12	-	-	-	-	T	A	-	-	T	-	-	T	-	-	-	-	T	-	-	-	A	-	-	-	C	-	1	-	-	1
SM B1	-	-	-	-	-	-	-	C	-	-	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3	3
SM B2	-	-	-	G	-	-	-	C	-	-	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1
SM C1	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	C	-	-	C	T	-	-	-	-	-	4	4
SM C2	-	-	G	-	-	-	-	-	-	-	-	T	-	-	G	-	-	-	C	-	-	C	T	-	-	-	-	-	2	2
SM C3	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	G	-	-	C	-	-	C	T	-	-	5	-	3	-	8
SM C4	-	-	-	-	-	-	T	-	-	-	-	T	-	-	-	G	-	-	C	-	-	C	T	-	-	1	-	-	-	1
Total																										18	19	8	16	61

Table 8: mtDNA Hypervariable Region I Sequences for Sierra Nevada de Santa Marta samples, only variable sites are shown
* = Cambridge Reference Sequence, Anderson *et al.* 1981

	Haplogroup		A						B			C					
			Shared mtDNA South American haplotypes														
Population	**n**	**Ns***	**02**	**10**	**64**	**65**	**86**	**107**	**01**	**07**	**13**	**24**	**66**	**89**	**91**	**92**	**References**
Huetar	27	21 (77%)	7	-	-	-	-	-	1	13	-	-	-	-	-	-	1
Kuna	63	19 (30%)	-	5	-	-	-	-	13	1	-	-	-	-	-	-	2
Ngöbe	46	16 (13%)	-	8	-	-	-	-	3	4	1	-	-	-	-	-	3
Zoro	29	1 (3%)	1	-	-	-	-	-		-	-	-	-	-	-	-	4
Yanamamõ	53	22 (42%)	-	-	-	-	-	-	1	-	-	21	-	-	-	-	5, 6
Xavante	25	15 (60%)	-	4	-	-	-	-	11	-	-	-	-	-	-	-	4
Gavião	28	0 (0%)	-	-	-	-	-	-		-	-	-	-	-	-	-	4
Mapuche	39	6 (15%)	-	-	-	-	-	-		-	-	6	-	-	-	-	7
Arsario	**18**	**15 (83%)**	-	-	**6**	**4**	-	-		-	-	-	**5**	-	-	-	**This Study**
Wayuú	**16**	**7 (44%)**	-	-	-	-	-	-	**3**	-	-	**4**	-	-	-	-	**This Study**
Ijka	**19**	**14 (74%)**	-	-	**14**	-	-	-		-	-	-	-	-	-	-	**This Study**
Emberrá	44	23 (52%)	-	2	-	-	2	2	9	-	2	-	-	2	2	4	8
Wounan	31	14 (45%)	-	-	-	-	1	2		-	-	-	-	1	7	5	8
SM Chibcha Com.	**45**	**34 (76%)**	-	-	**20**	**7**	-	-		-	-	-	7	-	-	-	**This Study**

Table 9: mtDNA HVSI shared haplotypes between Central and South American populations
Ns*=Number of individuals who share mtDNA haplotypes with other populations
References (1) = Santo *et al* 1994, (2) = Batista *et al.* 1995, (3)= Kolman *et al.* 1996, (4) = Ward et al. 1996 (5) = Easton *et al.* 1996, (6) = Torroni *et al.* 1993, (7)= Ginther *et al.* 1993, (8) = Kolman & Bermingham 1997.

Population	Number of [a] Individuals	# haplotypes	s^b	h^c	π^d	Tajima's D	Fu's Fs
Arsario	18	6	12	0.7974	0.014	1.25 (NS)[e]	2.09 (NS)
Ijka	19	5	5	0.4620	0.002	-1.43 (NS)	-1.86 (NS)
Santa Marta Chibcha comb.[f]	45	13	20	0.7646	0.011	-0.551(NS)	-1.57(NS)
Emberá	44	20	23	0.9419	0.018	0.459(NS)	-4.38(NS)
Wounan	31	14	29	0.9118	0.02	-0.273 (NS)	-1.01 (NS)
Ngöbe	46	7	12	0.7633	0.012	1.68 (NS)	3.39 (NS)
Kuna	63	7	10	0.5919	0.009	1.52(NS)	2.77(NS)
Huetar	29	7	12	0.7094	0.010	0.413(NS)	1.18 (NS)
Wayuú	16	7	15	0.8417	0.017	0.913(NS)	1.29(NS)
Mapuche	38	13	20	0.9123	0.017	0.78(NS)	-0.12 (NS)
Xavante	25	4	10	0.6767	0.009	0.438(NS)	3.72 (NS)
Zoro	29	8	16	0.7586	0.011	-0.202(NS)	0.850(NS)
Gavião	28	7	16	0.8624	0.012	0.083(NS)	1.97(NS)
Yanamamõ	53	20	28	0.8266	0.011	-1.27 (NS)	-7.13**

Table 10: Haplotype and Nuclear Diversity Measure in Central and South American populations

[a] Data from Kolman & Bermingham 1997, Kolman *et al.* 1995, Santos *et al.* 1994, Batista *et al.* 1995, Ginther *et al.* 1993, Easton *et al.* 1996, Torroni *et al.* 1993, Ward *et al.* 1996, This study

[b] Number of nucleotide variant sites

[c] Haplotype diversity

[d] Nucleotide diversity

[e] Not Significant

[f] Santa Marta Chibchan speakers combined contains 19 Ijka, 18 Arsario, and 8 Kogí

*$p < 0.05$

** $p < .01$

Measures of Selective Neutrality

Haplotype and nucleotide diversity values were calculated for all of the study populations as well as for five Central and five South American populations for which comparative data were available in the literature and is summarized in Table 10. The eight Kogí sequenced individuals were included with the other two Santa Marta groups in the SM Chibcha combined populations in order to increase the sample size. The Ijka have the lowest haplotype diversity (0.4620) of any the groups studied. The Arsario have a haplotype diversity of 0.7974, which is the highest of any of the Chibchan-speaking groups. The combined Santa Marta Chibchan speaking groups have a haplotype diversity of 0.7646. The Wayuú have the highest haplotype diversity (0.8417) of any of the study populations and is consistent with other populations from South America. The haplotype diversity for South and lower Central American populations ranges from 0.4620 (Ijka) to 0.9419 in the Emberá. All of the groups have low nucleotide diversity values with the Ijka having the lowest at 0.002 and the Chocoan-speaking Wounan exhibiting the highest at 0.02. Two measures of selective neutrality (Tajima's D and Fu's Fs) were also undertaken to determine if natural selection was acting on these populations or if other evolutionary forces were occurring. Tajima's D values are not statistically significant for any of the populations shown in Table 10, indicating that population expansion does not appear to be acting on these groups and that other evolutionary forces may be having an impact. The only population that possesses a significant Fu's Fs value is the Yanamamõ.

R-Matrix

Two separate R-matrices were used in the analyses of these data. The first of these compares the four study populations to one another using mtDNA haplogroup diversity along with data obtained from the literature for four blood groups (MNS, RH, Diego, and Duffy) (Yunis *et al.* 1994). The second R-matrix compared the four study population with mtDNA haplogroup variation in 36 existing Central and South American populations. A list of these populations, their geographic locations, mtDNA haplogroup frequencies and original references are included Table 7. The criterion used for compiling this list of populations was a minimum sample size of twenty individuals and representative of a wide geographic area in South America. If a population was referenced in more than one publication, then sample sizes were pooled and haplogroup frequencies were calculated. The geographic distribution of these populations is shown in Figure 5.

The first R-matrix compares the four study groups using eleven allelic markers (mtA, mtB, mtC, MS, Ms, R1, R2, Jka, Jkb, Fya, Fyb) and is shown in Figure 6. This matrix accounts for 92% of the variation present in the four study populations. The first axis accounts for 55% of the variation present in the samples and the 2nd axis represents 37% of the variation. In the plot of variation the first axis separates the three Chibchan populations from the Wayuú. The second axis separates the Wayuú and the Ijka from the Arsario and Kogí. This indicates that the Arsario are intermediate between the Kogí and the Ijka which makes sense based on ethnographic information regarding this population.

In order to ascertain the relationship of the four study populations to thirty-six other populations in lower Central and South America a second R-matrix analysis was conducted with mtDNA RFLP frequency data for the four major American specific haplogroups (mtA, mtB, mtC, mtD) as well as other undetermined haplogroups that were present in these groups (mtE). The results of this analysis are shown in Figure 7. This matrix explained 75% of the total variation present in the populations. The first axis accounted for 47% of the total variation and the 2nd axis accounted for 28% of the variation. There are three distinct clusters, with one being composed exclusively of Chibchan populations. This cluster was caused by the high frequency of haplogroup A present in these populations. Haplogroup B is clustering the populations from central and western South America. Haplogroups C and D are influencing the plots of populations in the southern cone of the continent.

Median Joining Networks

Two different median joining network analyses were run for these data. The first analysis combines the four study populations (Kogí, Ijka, Arsario, and Wayuú) into a single network connected through mutation defining haplotypes. In order to compare these populations to other Chibchan populations, sequence haplotype data were collected from the literature for comparison. A total of three Chibchan and two Chocoan speaking populations were found to have a sufficient sample size for comparison. These include the Kuna (Batista *et al.* 1996), the Ngöbé (Kolman *et al.* 1995), the Huetar (Santos *et al.* 1994), and the Emberá and the Wounan (Kolman and Bermingham 1997). Therefore, the final network constructed includes the four study

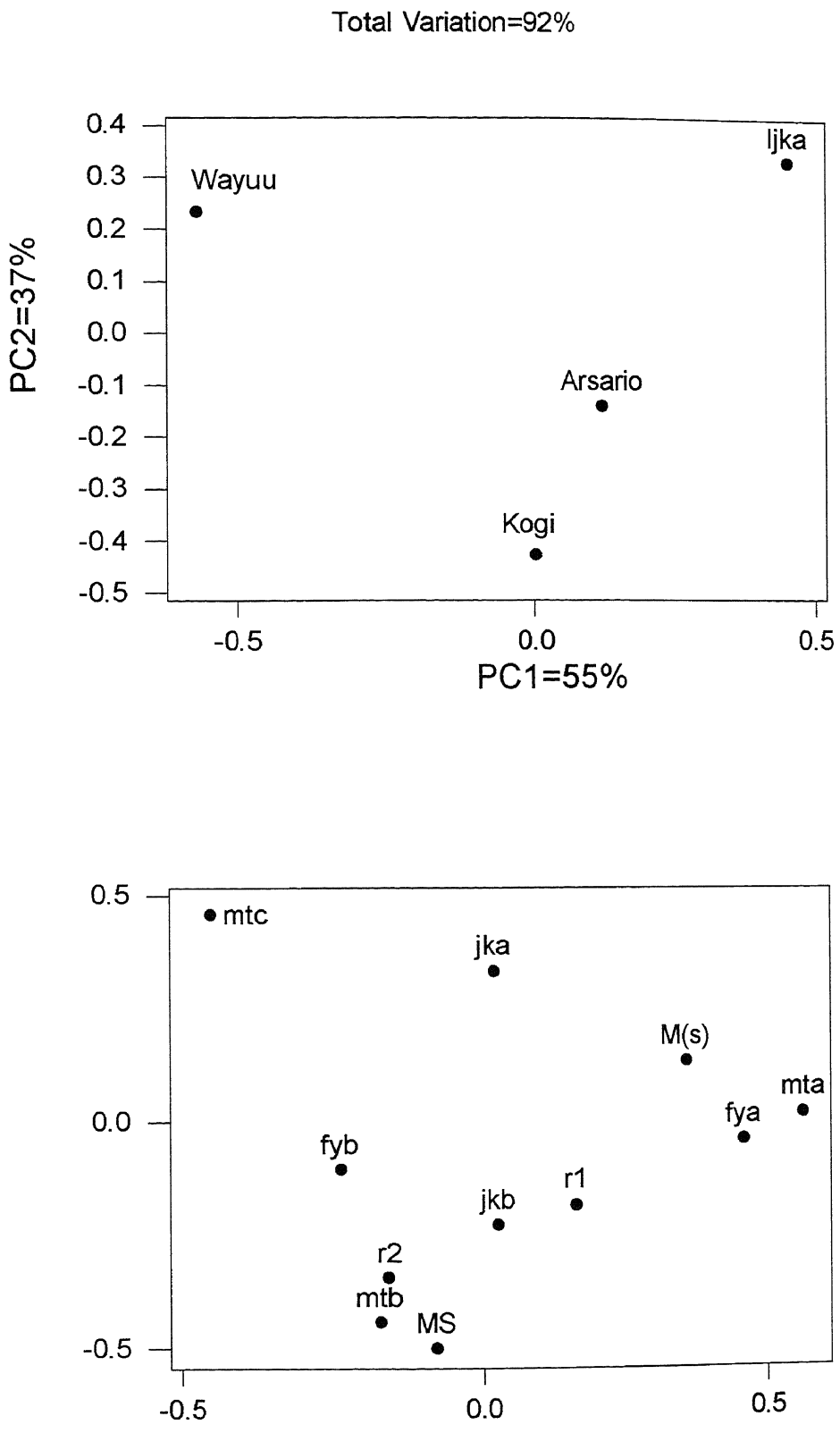
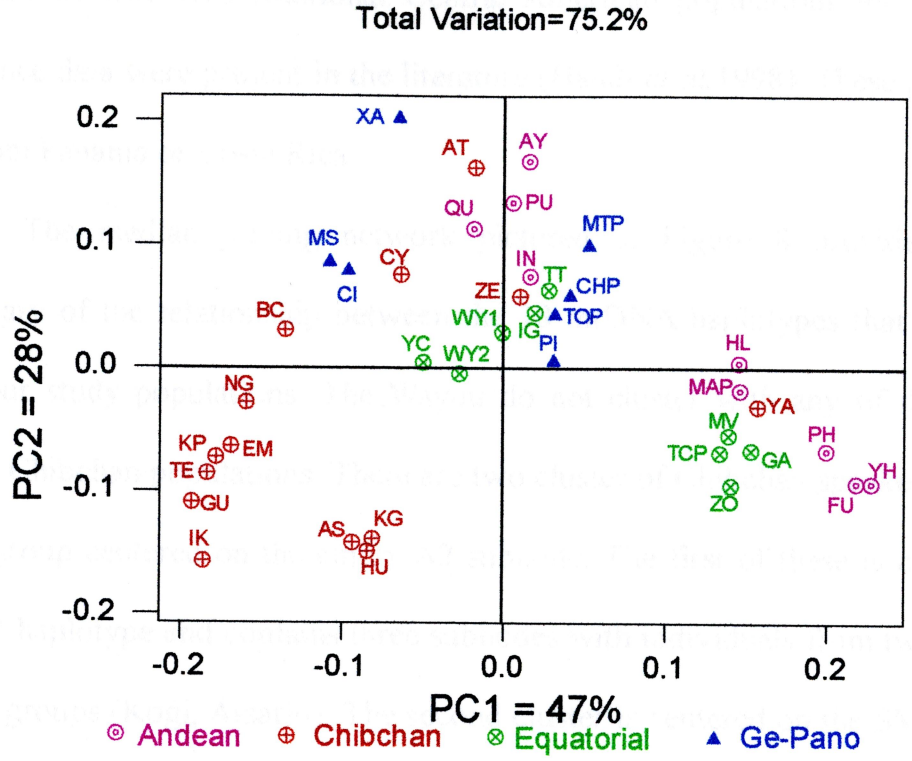


Figure 6: Top graph shows R-Matrix (populations) and bottom graph shows S-Matrix (alleles) for the four study populations



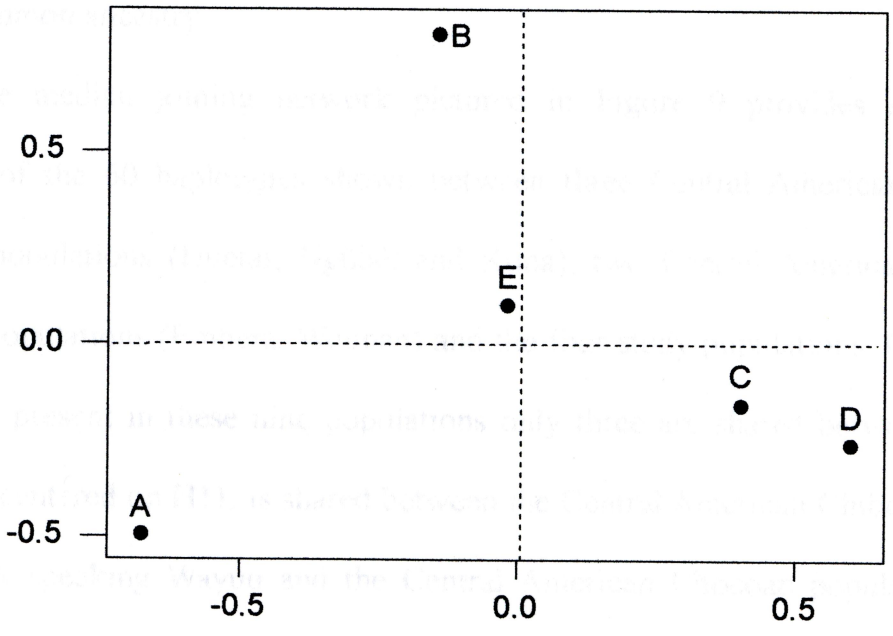


Figure 7: Top graph shows R-matrix (populations) and bottom graph shows S-matrix (alleles) from 39 South American indigenous populations

populations with five additional Central American populations for which mtDNA sequence data were present in the literature (Handt *et al.*1998). These populations are all from Panama or Costa Rica

The median joining network pictured in Figure 8 provides a graphical summary of the relationship between the 18 mtDNA haplotypes that were found in the four study populations. The Wayuú do not cluster with any of the three Santa Marta Chibchan populations. There are two cluster of Chibchan speakers within the A haplogroup centered on the empty A2 subnode. The first of these is centered on the SMA3 haplotype and contains three subnodes with individuals from two of the Santa Marta groups (Kogí, Arsario). The second cluster is centered on the SMA7 haplotype and contains five subnodes with at least one individual from each of all three Santa Marta populations represented. This result indicates that these populations probably shared common ancestry.

The median joining network pictured in Figure 9 provides a graphical summary of the 60 haplotypes shown between three Central American Chibchan speaking populations (Huetar, Ngöbé, and Kuna), two Central American Chocoan speaking populations (Emberá, Wounan) and the four study populations. Of the sixty haplotypes present in these nine populations only three are shared between groups. The node, centered on H11, is shared between the Central American Chibchan group, the Arawak speaking Wayuú and the Central American Chocoan populations. The haplotype NG4 is shared between the Central American Chibchan and the Central American Chocoan speakers. Haplotype EW22 is shared between the Wayuú and the

Central American Chocoan speakers. All three of these nodes are shared between several Native American groups and represent several founding Native American haplotypes (B1, A2, and C2, respectively). The Santa Marta Chibchan populations do not share nodes with any of the other three groups. The most starlike structure belongs to haplogroup B centered on node H11 and is representative of a recently expanding population. Haplogroup A2 centered on NG4 demonstrates the most haplotype diversity with three satellite starlike clusters (SMA7, SMA3, NG2) and several large single nodes representative of founder effect based on the short branch length (1 mutational unit) and the large nodes with no satellite branches with little or no population overlap between groups..

A subsequent network analysis of 104 haplotypes conducted with all populations found in Table 10 did not differ significantly from the second network shown in Figure 9 (data not shown). Only 14 of the 104 haplotypes were shared between populations and most of these were found in neighboring populations already pictured in Figure 9. The addition of the five South American populations only added two shared nodes on the haplogroup B node H11 and the haplogroup C2 node and several single nodes within haplogroups C and D.

AMOVA

Two separate analyses of molecular variance (AMOVAs) were conducted for this study. The first AMOVA includes the four study populations, separated into two groups, the Santa Marta Chibchan speakers (Kogí, Arsario, Ijka) and the Arawak speaking Wayuú. The second AMOVA includes data from all of the ten South

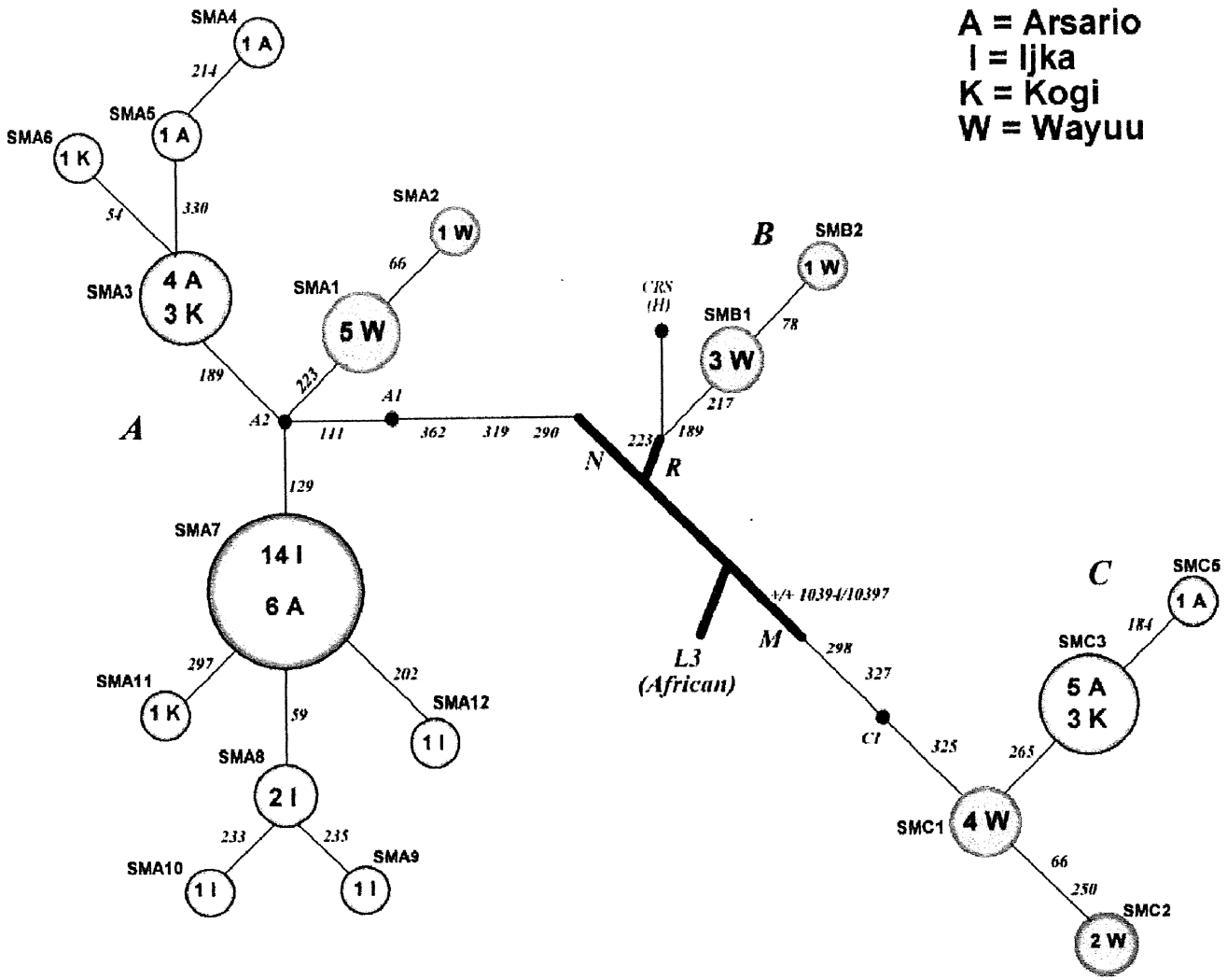


Figure 8: MJ network of four study populations: Circle size represents number of samples in each haplotype

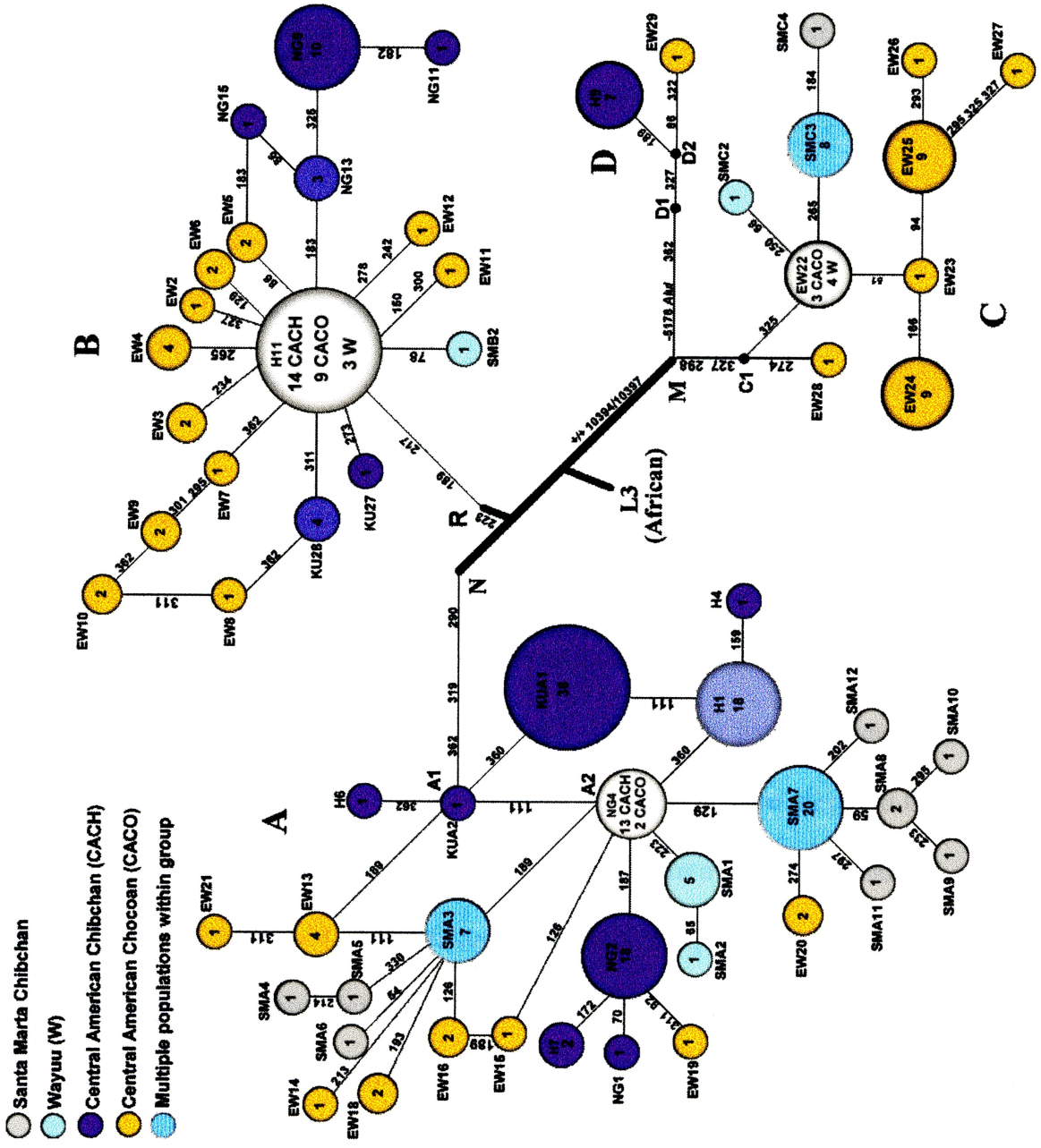


Figure 9: MJ network for four study populations, three Central American Chibchan and two Central American Chocoan populations: Size of node indicates number of individuals and number on branches represent mtDNA mutation location (nt.16050-16383). Major mtDNA haplogroups are A1=KUA2, A2=NG4, B1=H11, C2=EW22, d1=empty node d1

American populations shown in Table 7. These separated into seven different groups based on their linguistic affiliation as designated by Ruhlen (1991): Central American Chocoan (Emberá, Wounan), Central American Chibchan (Kuna, Ngöbé, Huetar), Central American Chibchan (Ijka, Arsario), South American Chibchan (Yanamamõ), Equatorial-Tucanoan (Wayuú, Zoro, Gavião), Ge-Pano-Carib (Xavante), and Andean (Mapuche).

The results of the AMOVA between the four study populations separated into two groups (Chibchan and Non-Chibchan) are shown in Table 11. The amount of variation among groups, 15.05%, was not considered statistically significant at the α=0.05 level. There was statistically significant variation among populations within the two groups and this accounted for 18.20% of the variation. The greatest amount of variation was within populations and accounted for 66.75% of all of the variation within the groups. This was also considered statistically significant. The lack of variation between groups and the high level of diversity among populations within groups may have been observed because only two groups were used and there may be significant mtDNA diversity between the three Santa Marta Chibchan speaking populations. When an AMOVA was run separating the Ijka out from the Kogí and the Arsario statistically significant variation was found among groups but not within groups indicating the possibility that the Arsario and the Kogí are more closely related than either is to the Ijka.

Source of variations	D.F.	Sum of Square	Variance Components	Percentage of Variation	Φ Statistic	p Value
Among Groups	1	20.152	0.42346	15.05%	Φct 0.15051	0.25 (NS)
Among populations within groups	2	18.272	0.51192	18.20%	Φsc 0.21419	>0.0001
Within populations	57	107.052	1.87810	66.75%	Φst 0.33246	>0.0001
Total	60	145.475	2.81348			

Table 11: AMOVA between two groups (Chibchan and Non-Chibchan): Four study populations.

Source of variations	D.F.	Sum of Square	Variance Components	% of Variation	Φ Statistic	p Value
Among Groups	6	298.47	0.57720	18.53%	Φct .18532	>0.0001
Among populations within groups	6	84.204	0.37903	12.17%	Φsc 0.14938	>0.0001
Within populations	425	917.297	2.15834	69.30%	Φst 0.30702	>0.0001
Total	437	1299.973	3.11458			

Table 12: AMOVA for 7 groups based on linguistic affiliation as defined by Ruhlen (1992). Central American Chocoan (Emberá, Wounan), Central American Chibchan (Ngöbé, Kuna, Huetar), Santa Marta Chibchan (Ijka, Arsario), South American Chibchan (Yanamamö), Equatorial-Tucanoan (Wayuú, Zoro, Gavião), Ge-Pano-Carib(Xavante), and Andean (Mapuche)

The results of the AMOVA for all of the South and Central American populations are shown in Table 12. Groups are based on linguistic affiliation as defined by Ruhlen (1991) (Table 7). The amount of variation among groups accounted for 18.53 % of the variation, the amount of variation among populations within groups accounted for 12.17% of the total variation, and the amount of variation within populations accounted for 69.30% of the variation. All three were considered statistically significant at α>0.001 level. The lowest amount of diversity

was shown among groups indicating that there is probably differentiation between groups based on linguistic affiliation.

Time estimates

Mismatch analyses were conducted for all of the comparison populations that are listed in Table 10 and compared to the four study populations. The first analysis was applied to the HVS-I sequences data for the four study populations (Ijka, Arsario, Wayuú, and Santa Marta Chibchan combined). Due to the low number of sequenced Kogí individuals, all of the Santa Marta individuals were combined into a single entity in order to increase the statistical power for that group. The τ values for all populations are shown in Table 13 as well as age ranges based on the 95% upper and lower τ confidence interval values. The τ values range from 2.18 for the Ijka to 9.98 for the Santa Marta Chibcha combined population. Time estimates based on these values range from 20,274 - 92,814 YBP. Four mismatch distributions for the study populations are shown in Figure 10. The only study population to demonstrate a unimodal mismatch distribution is the Ijka (Figure 10.b). They show a spike beginning at 2 mutational units then a dip at 1 mutational unit, then a sharp peak again between 0 and 1 unit of mutational time, possibly indicating a recent population expansion during the last 10,000 years. The other three study populations all demonstrate multimodal distributions indicating several population expansions and contractions over the last 100,000 years. The Arsario (Figure 10.a) show a slightly trimodal distribution with a sharp peak occurring at mutational unit eight and then two smaller peaks, with one occurring at two units of mutational time and another at

zero units of mutational time. The Santa Marta Chibcha combined group (Figure 10.c) exhibits a mismatch distribution that is similar to the Arsario with potentially three peaks occurring. The Wayuú (Figure 10.d) also demonstrate a trimodal peak, however, they have two peaks between five and nine mutational units and one small peak at zero units of mutational time. All of the other South American populations also showed multimodal distributions However, with the exception of the Xavante and the Zoro (data not shown), the only other populations to have frequencies above 0.2 between 0 and 1 units of mutation are the three Central American Chibchan speaking populations and the Santa Marta populations.

In order to determine whether or not stochastic effects and the distribution of haplogroups in these populations may be having an underlying effect on the mismatch distributions, a subsequent mismatch analysis was run on individuals of haplogroup A for the Santa Marta Chibchan populations and two clusters of haplogroup A. The first cluster was defined as individuals who had the 129 G-A transition and the second cluster was based on the 189 T-C transition. The Wayuú were excluded because they were clearly a separate population and all of their haplotypes occurred in clusters of two for each haplogroup. Individuals of haplogroup C from the Santa Marta populations were also excluded because more than two haplotypes are necessary to get accurate results. The τ values and date ranges are shown in Table 10 and the mismatch distribution for all three groups is shown in Figure 11. The τ value for all the Santa Marta Chibchan speaking groups is 2.6 which dates to 24,180 years bp, this date is consistent with other estimates for the presence of haplogroup A in the New

World, especially haplogroup A2 (Schurr & Sherry 2004, Saillard *et al.* 2000). The mismatch distribution for this haplogroup is also bimodal and may be representative of the two clusters of haplotypes found in the median joining network shown in Figure 7. Both of these clusters are unimodal with lower τ values. The distribution for SM haplotype A7 has the higher value and is similar to the Ijka mismatch distribution shown in Figure 10, except that it plateaus between 2 and 1 unit of mutational time. The distribution for SM haplotype A3 has the lowest τ value of 0.97 and dates to 9,021 years bp, which is the lowest estimate for any of the dates given using mismatch analysis.

A second coalescent dating technique using the ρ-statistic of Saillard *et al.* (2000) was used because the dates given by the mismatch analysis were not all within the range for the peopling of the Americas. This indicated a drawback to the mismatch technique because it does not allow for the date to be rooted in a phylogenetic framework. A significant advantage of this technique is that it allows the researcher to root the tree with an ancestral node and determine the ancestral nodes to be dated. Two subnodes from Figure 8 were chosen to be dated using this technique because they contained more than one descendent subnode and more than one of the study populations. The first subnode was centered on SMA7 and gave a date of 6,985 (+/- 3,557) years BP. The second cluster was centered on the SMA3 haplotype and contains individuals from the Kogí and Arsario. The ρ statistic gave a date of 8,072 (+/-4,943) years BP for this node. The dates given for both of these nodes are consistent with other dates previously reported for Chibchan ethnogenesis, which

range between 6,000 and 10,000 years bp (Kolman *et al.* 1995, Kolman and Bermingham 1997, Torroni *et al.* 1994). An additional analysis was done on the A2 node shown in Figure 9. Based on these data the ρ statistic (1.11) gives a date of 22,422 (+/- 8,206) years BP for this haplogroup which is consistent with other estimates for the presence of haplogroup A in the Americas (Schurr and Sherry 2004). Haplogroups C and D centered on empty nodes C1 and D1 demonstrate the least amount of diversity. The branches for these haplogroup A nodes are deeper and there are several large nodes (KUA1, H1, SMA7, NG2, SMA3) with a only a small number of individual branching nodes indicating populations that have gone through a recent bottleneck in regards to mtDNA.

Time estimates for Mismatch and ρ values for South American Populations				
Population	**τ**	**95% CI**	**Time Est.**	**Date Range**
Arsario	**9.57**	3.8-19.88	89,001	35,340-184,884
Ijka	**2.18**	0.247-4.43	20,274	2,297-41,119
Santa Marta Chibcha comb	**9.98**	4.184-15.98	92,814	38,911-148,614
Emberá	**9.11**	4.21-13.85	84,723	39,153-128,805
Wounan	**9.95**	4.9-14.59	92,535	45,570-235,687
Ngöbé	**9.92**	4.26-16.72	92,256	39,618-155,496
Kuna	**8.14**	2.9-13.7	75,702	26,970-127,410
Huetar	**6.83**	2.25-14.82	63,519	20,925-137,826
Wayuú	**8.28**	3.75-13.75	77,004	34,875-127,875
Mapuche	**7.93**	4.23-11.22	73,749	39,339-104,346
Xavante	**9.52**	4.19-14.52	88,536	38,967-135,036
Zoro	**7.02**	2.23-12.52	65,286	20,739-116,436
Gavião	**8.24**	2.94-15.92	76,632	27,342-148,056
Yanamamõ	**6.22**	2.08-11.27	57,846	19,344-104,811
Santa Marta Haplogroup A	**2.6**	0.48-6.34	24,180	4,464-58,962
Santa Marta Haplotype A3	**0.97**	0.00-3.96	9,021	0-36,828
Santa Marta Haplotype A7	**2.25**	0.188-4.21	20,925	1,748-39,153
Median Joining Cluster	**ρ**			
Santa Marta Haplotype A3	**0.4**	0.15-0.65	8,072	3,129-13,015
Santa Marta Haplotype A7	**0.35**	0.17-0.52	6,985	3,428-10,542

Table 13: Mismatch τ and ρ values for South American populations. Mismatch time estimates are based on 1 mutational event every 9,300 years (Ward *et al.* 1991), ρ values are based on 1 mutational event every 20,180 years (Saillard *et al.* 2000)

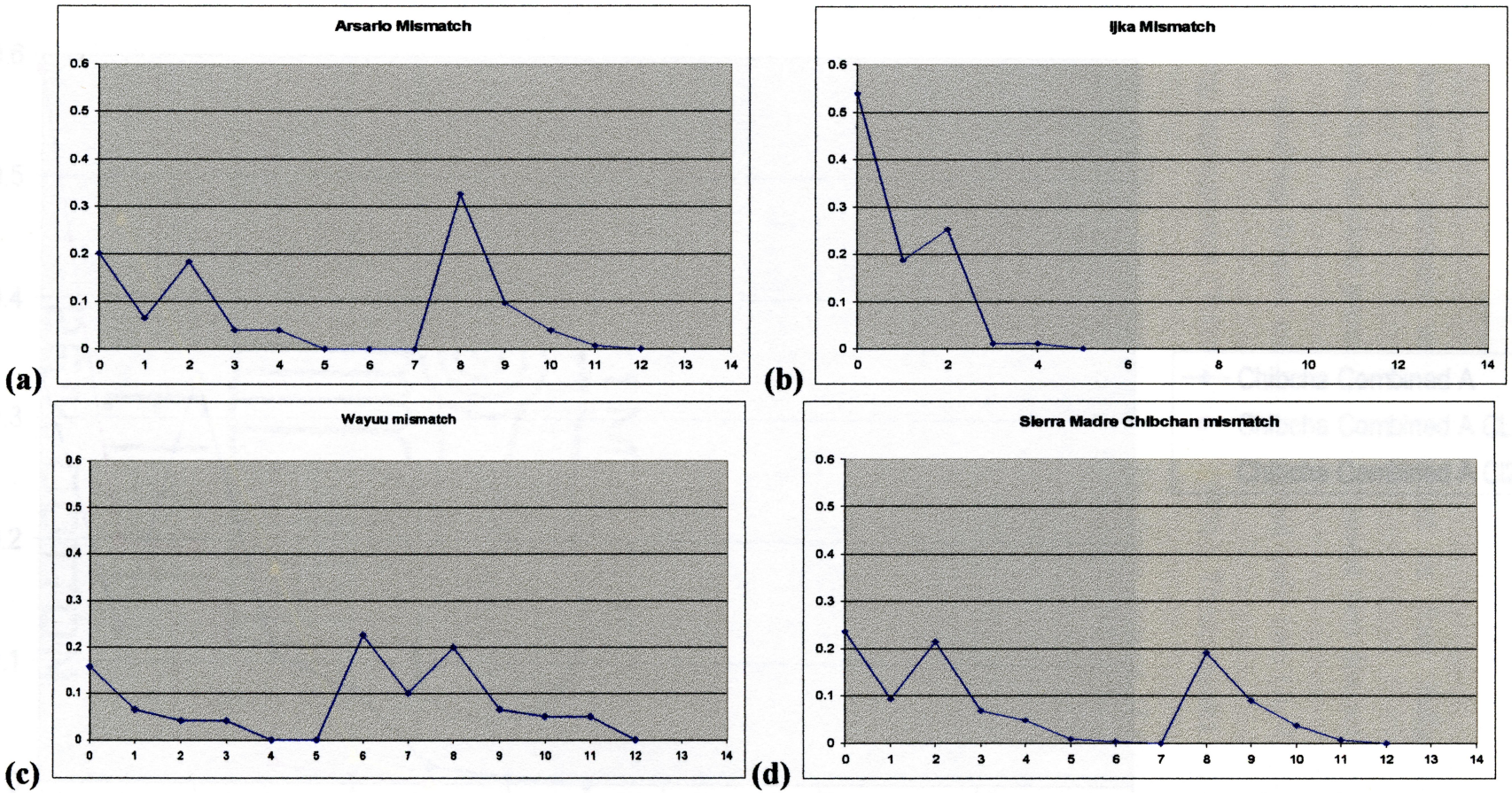


Figure 10: Mismatch distributions for three study population and pooled Chibchan populations. τ values are shown in Table 7 and dates and error ranges are shown in Table 10. Mutational units are shown on the X-axis and equal ~ 9,300 yrs.

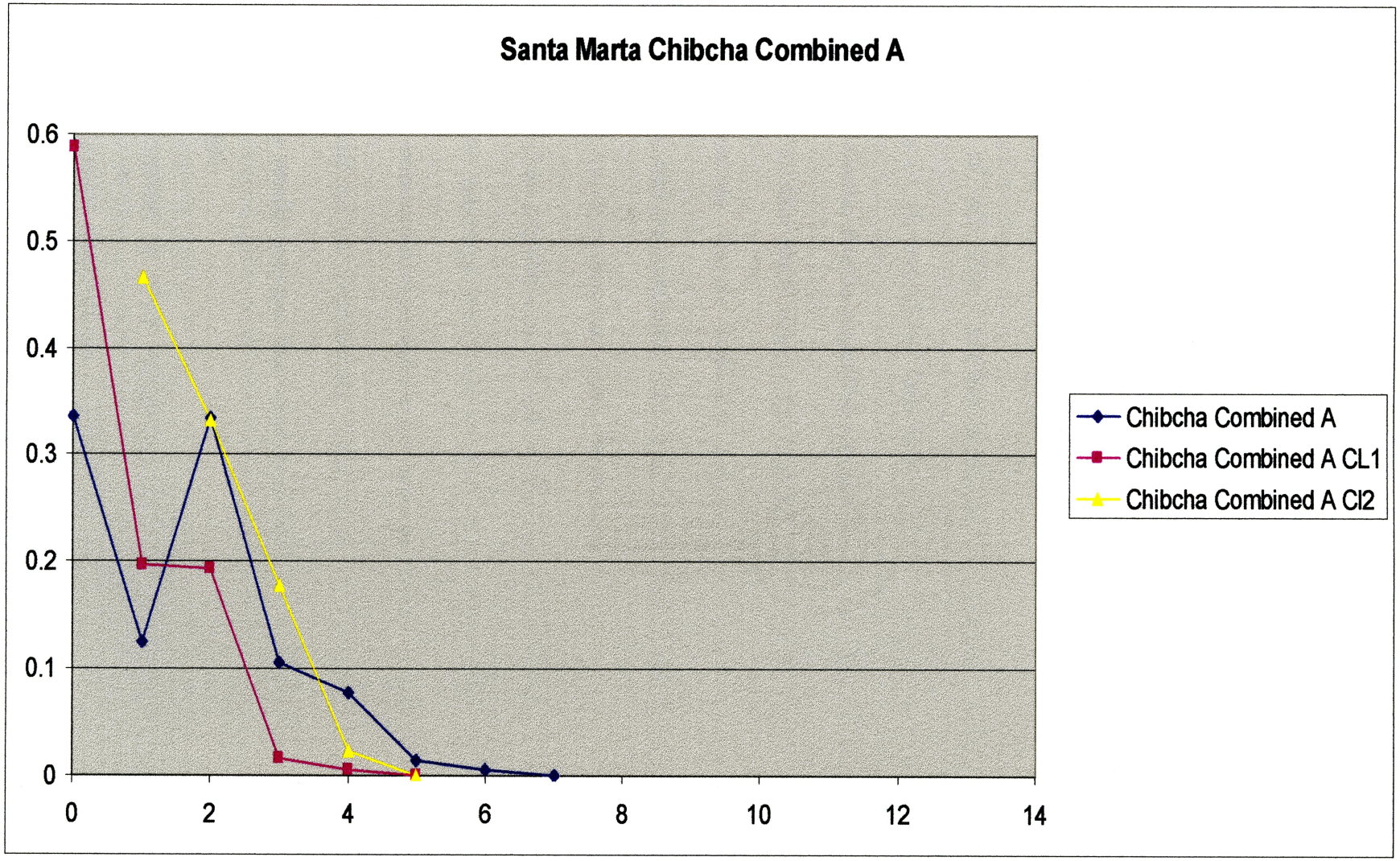


Figure 11: Mismatch distribution for Santa Marta Chibchan speaking individuals of Haplogroup A.

Chapter V: Discussion

Mitochondrial markers were examined in four indigenous Colombian populations who live in close proximity but represent two distinct linguistic families. The Chibchan speaking Arsario and Kogí carried only two (A, C) of the four New World founding haplogroups (Torroni *et al.* 1993b). These results are consistent with other published mtDNA haplogroup data for these groups (Keyeux *et al.* 2002, Briceño *et al.* nd). The Chibchan-speaking Ijka contained three of the founding haplogroups (A, B, and C) but they had over 90% A. One previous study of the Ijka also found similar results regarding the presence of the three haplotypes (Briceño *et al.* nd) and another study found only haplogroups A and C (Keyeux *et al.* 2002).The Arawak-speaking Wayuú exhibited three (A, B, and C) of the four founder haplogroups and are consistent with other mtDNA haplogroup research conducted on this population (Keyeux *et al.* 2002, Mesa *et al.* 2000, Briceño *et al.* nd). The Santa Marta Chibchan populations did not exhibit any non-Native American haplogroups, indicating no maternal admixture in these populations. The Kogí and Arsario had previously demonstrated no admixture, but the Ijka had shown admixture from both European and African populations (Yunis *et al.* 1994, Salzano & Callegari-Jacques 1988). This would indicate that admixture into this population was generated from male lineages.

In contrast to other indigenous New World populations the Santa Marta Chibchan groups demonstrate a distinctive pattern of mtDNA haplogroups. The majority of individuals in these populations belong to haplogroup A and this is shared

with other Chibchan populations from lower Central America. Where the Santa Marta Chibchan-speakers differ is in the presence of haplogroup C, which is absent in Central American groups. Kolman and Bermingham (1997) suggested that haplogroup C was not present in Chibchan groups throughout their genetic history, but these data appear to refute that. The Santa Marta populations also differ in the absence of haplogroup B, which is found in moderate frequencies in the Central American populations. Only two Central American Chibchan (Huetar, Boruca) speaking groups are known to demonstrate small frequencies of haplogroup D, and this was found in one or a few individuals from one community (Santos *et al.* 1994, Torroni *et al.* 1994). A potential explanation for this is that the four major Native American haplogroups were present at the beginning of Chibchan genetic divergence and then subsequently lost through genetic drift. All of these populations have gone through a significant depopulation within the last four hundred years due to European contact and the different distributions shown here could be representative of that. However, this explanation is unlikely as all Native American populations went through these severe contact depopulation events. The only difference for these depopulation events is the duration of time since they occurred. Indigenous populations in the Caribbean, along the Caribbean coast of lower Central and northern South America were the first to be colonized by Europeans and therefore suffered the most profound losses. Nevertheless, several of these groups contain at least four of the five founding haplogroups and higher mtDNA diversity then seen in Chibchan populations (Schurr & Sherry 2004, Kolman *et al.* 1995).

The indigenous populations of Colombia offer an alternate explanation to the theory of genetic drift. In Colombia there are differing haplotype frequencies (A/D cline) between east and west (Keyeux *et al.* 2002) created by the Andean cordillera that diagonally splits the country. This breaks the Colombian groups into two broad groups. The first group contains populations characterized by high frequencies of haplogroup A who are culturally related to the Central American populations though linguistic affiliation, biology, and geography (i.e., Chibchan or Arawak) located on the northwestern side of the Andes. The Santa Marta Chibchan populations and the Wayuú clearly fit into this group. The second group contains populations with high frequencies of haplogroup D who are more closely related to populations from the rest of South America along the southeastern side of the Andes. Haplogroup C is seen as being stable throughout the continent with similar fluctuations within each subcontinental region. Haplogroup B is absent from both the north and south of the continent but is stable throughout the interior. While genetic drift may explain these distributions in isolated populations, it assumes an absence of selective pressure over all the founding Native American mtDNA haplogroups. In order for this to be a plausible explanation there would need to be either a gradual cline or at least random distribution as opposed to abrupt genotypic changes caused by a significant geographic barrier (Keyeux *et al.* 2002).

An alternate explanation for this discontinuous gradient of haplogroup frequencies is that South America was settled by two waves of human migrants. The first small groups of Paleoindians would have migrated into South America, moved

south and over time gradually evolved into several heterogeneous groups. The second group would have maintained closer contact with their Central American genetic cousins (Keyeux *et al.* 2002, Fox 1996). If this second migration represented a Chibchan diaspora that occurred early during the peopling of the Americas it is likely that the Andean cordillera may have the slowed the movement of people further south. As the population size increased it may have backed up into the isthmus, causing it to become more densely populated, effectively blocking more migrants entering from the north (O'Rourke *et al.* 1992). Ethnohistoric data support this argument. The two largest Chibchan speaking populations present at contact were from Colombia and included the ancient Tairona whose maximum population size has been estimated at 468,000 individuals (Langeback 2003) and the extinct Muisca from near Bogotá whose numbers were estimated to be 500,000 (Hoopes & Fonseca 2003) at Spanish contact. Linguistic evidence also supports this argument. Colombia contains the second largest number of indigenous populations in South America representing 80 different language families (Yunis *et al.* 1994). The geographic distribution of these languages is more numerous and fragmented on the southeastern side of the Andes, whereas on the northwestern side there are fewer languages and most populations having cultural and biological relationships with either Central American (Chibchan & Chocoan) or Caribbean (Arawak) populations.

Today, the largest Chibchan-speaking populations inhabit lower Central America and number approximately 250,000 (Hoopes & Fonseca 2003). One of the largest of these contemporary populations (~125,000 pop.) is the Kuna. This

population is thought to have immigrated into Panama from Colombia during the 1600s and is known to have migrated from the Pacific coast to the Caribbean coast in the mid-1800s due to pressure from Chocoan speaking populations (Kolman *et al.* 1995). The Central American Chibchan population size at contact is unknown but may have been roughly 500,000 (Hoopes personal comm.). One can assume that the population in the region began to expand when plant domesticates were introduced into the region around 5,000 YBP (Hoopes & Fonseca 2003, Constenla 1991) and that this only increased over time until the time of contact. This would indicate the presence of a sizable population at contact which would have blocked population movement along the Andes from the South into Mesoamerica. What is less known is the size of populations on the northern Chibchan linguistic frontier and what were the Chibchan relationships with Caribbean populations?

Population Relationships

Study Populations

Both the R-matrix (Figure 6) and the median joining (MJ) network (Figure 8) indicate that the Santa Marta Chibchan populations share a genetic relationship with each other and not with the Wayuú. In the R-matrix the Kogí and the Arsario cluster close together. These two groups also share two mtDNA haplotypes (SMA3, SMC3) indicating shared matrilineages. The Ijka are separated from the other two groups in the R-matrix but also share a matrilineage with the Arsario (SMA7). These results indicate that the Arsario may be intermediate between the two groups as they are the only group who share matrilineages with the other two as well as sharing cultural

traits with both the Kogí and the Ijka. Y-chromosome data shows a different relationship for these populations where all three groups share at least two patrilineages (Guarino *et al.* 1999). These data are consistent with ethnographic information regarding at least, the Kogí, who practice uxorilocal marriage customs (husband moves to the wife's home) (Reichel-Dolmatoff 1950). Ethnographic information regarding the other two Santa Marta Chibchan populations is sparse but it can be assumed that these groups also share this cultural custom as it is practiced in two other Chibchan-speaking populations from lower Central America (Kuna, Huetar). However, the Chibchan-speaking Ngöbé practice virolocal customs (wife moves to the husband's home) and also demonstrate low mtDNA diversity levels (Kolman *et al.* 1995). An alternate explanation for this may be that geographic proximity may be more important than marriage customs. In this case it would appear that the former explanation is the most parsimonious as all three groups descend from the same ancestral population as well as inhabit the same geographic area.

The Wayuú are genetically distinct from the Santa Marta Chibchan populations and this is the result of a number of factors. They inhabit the vast arid Guajiro peninsula where they are exposed to more extreme climatic conditions so they are more likely to inhabit the Caribbean or Atlantics coast then the arid interior where the Santa Martas border the peninsula. They belong to the largest existing linguistic family in Colombia (Arawak) and their population size is also large (40,000-80,000 in Colombia) compared to the other groups (Yunis *et al.* 1994) They do not share any mtDNA haplotypes with the Chibchan populations (Figure 8) and

they are not clustered with any of the groups in the R-matrix (Figure 6). They also do not share any Y-chromosome haplotypes with the Santa Marta populations (Guarino *et al.* 1999) and are known to be admixed based on blood group data with both Africans and Europeans (Yunis *et al.* 1994, Salzano & Callegari-Jacques 1988). Culturally, they are also distinct from Santa Marta Chibchan populations by being seasonal pastoralists as opposed to seasonal horticulturists. So, while the four study populations are geographic neighbors, there are several key environmental, cultural, and biological differences creating genetic schisms between the two clusters based on linguistic affiliation.

South American and lower Central American Chibchan populations

Based on the R-matrix (Figure 7) there appears to be a relationship between northern South American and lower Central American Chibchan-speaking populations. These two groups cluster close together and do not appear to be associated with other linguistic stocks from South America. These groups are separated out by a high frequency of haplogroup A and a virtual absence of haplogroup D, similar to the Colombian mtDNA pattern found by Keyeux *et al.* (2002). The Kogí and the Arsario cluster with the remnant Huetar population from Central Costa Rica. The Ijka are genetically distinct from other populations and this may be due to their high frequency of haplogroup A which separates them from other Chibchan populations. Three Chibchan populations are found outside the main cluster and these include the Yanamamõ, the Zenu, and the Atacemeño. The linguistic affiliation all these populations, as defined by Ruhlen (1991), are debatable. A few

linguists place the Yanamamõ in the Chibchan linguistic phylum (Ruhlen 1991) while others disagree and believe that the Yanamamõ language belongs with other Amazonian languages (Constenla 1991). Also, the Yanamamõ practice a marriage custom that may be called linguistic exogamy, meaning that a small number of females are acquired as wives through raiding from neighboring linguistically distinct tribes (Neel & Chagnon 1968). Due to small population size and population dynamics (including differential fertility and mortality) of many Amazonian cultures, this may have a profound impact on the number and type of maternal lineages present in the group, therefore making them genetically distinct in regards to mtDNA. The Zenu are an acculturated group whose language is extinct at this time so their exact linguistic affiliation is unknown but some argue that due to geographic location that it should be Chibchan while others argue that it is a Carib language (Mesa *et al.* 2000) or Chocoan (Constenla 1991). The Atacemeño are from Chile and cluster with other Andean groups (Aymara, Quechua, Pueños) indicating that in their case geographic proximity is a more plausible explanation than linguistic affiliation.

Further examination of this relationship shows that this does not extend to the mtDNA haplotype level. The MJ network shown in Figure 9 demonstrates that while the Chibchan population may cluster together based on mtDNA haplogroup data they are genetically distinct at the haplotype level. The Santa Marta Chibchan populations do not share any mtDNA haplotypes with other Chibchan or Chocoan population from Central America. This indicates that while they may share similar haplogroup frequencies they do not share recent genetic admixture at the mtDNA level. So, if

they all went through an ancestral Chibchan bottleneck, then significant time depth would be required for this population fragmentation to have occurred and that founder effect would have to have had an acute effect on their genetic history at the maternal level.

Support for Chibchan population founder effect

Evidence for this Chibchan population founder effect is seen in this MJ network (Figure 9). All Chibchan populations are characterized by a single large node that is population specific whereas the Chocoan nodes (EW3, EW7, etc.) are characterized by a large number of singletons. The Chibchan nodes include the Kuna (KUA1), Huetar (H1), Ngöbé (NG2), and the Santa Marta Chibchan populations (SMA3, SMA7). These nodes are all one evolutionary mutational unit away from one of the two ancestral Native American haplogroup A nodes (haplogroup A1=KUA2, haplogroup A2=NG4). The haplogroup A2 is characterized by the 16111T mutation that is specific to New World populations. Four of these nodes are within one mutational unit of this, indicating that all of these populations arose in the New World within the last 10,000 years. This would appear to indicate that each of these Chibchan populations was influenced by founder effect within their own specific matrilineages. If this was a random evolutionary event, one would expect to see several empty nodes or small nodes between the ancestral populations. If this were a genetic bottleneck, then one would expect all of these populations to descend through the same mutational pathway. An argument against this second interpretation would be that due to the severe depopulation that occurred after European contact that

variation within these populations was lost and therefore does not show up in the present mtDNA haplotype data set. However, as previously mentioned, all New World populations differentially experienced this bottleneck and not all of these populations demonstrate this reduced haplotype diversity. The evidence for a bottleneck is more difficult to refute as the majority of Chibchan nodes radiate out from a single interior indicating that all of these populations recently went through the same genetic bottleneck. The impact of maternal founder effect is demonstrated because none of these populations, with the exception of the Santa Marta groups, actually share these large haplotype nodes with other Chibchan populations. Other evidence for this is seen in the lowered mtDNA haplotype diversity, measures of selective neutrality, and mismatch distribution.

The three Santa Marta Chibchan populations also demonstrate reduced levels of mtDNA haplotype and nucleotide diversity relative to the Wayuú. The Ijka have the lowest mtDNA haplotype diversity (0.462) of any of the populations from lower Central American or South American populations shown in Table 10. The mtDNA diversity levels of the Arsario are also lower than other populations but are consistent with levels from other Chibchan populations from lower Central America. Some researchers have suggested that the reduced amount of mtDNA diversity seen in Chibchan populations is evidence of a population bottleneck at the onset of Chibchan genetic history (Kolman & Bermingham 1997, Batista *et al.* 1995, Kolman *et al.* 1995). The average mtDNA haplotype diversity for all six Chibchan populations (Arsario, Ijka, SM Comb., Ngöbé, Kuna, Huetar) shown in Table 10 is 0.6814

whereas the average for the six South American (Wayuú, Mapuche, Xavante, Zoro, Gavião, Yanamamõ) populations is 0.813 and the average for the two Chocoan Central American populations is 0.9268. The only population to show similarly low mtDNA haplotype diversity levels is the Xavante, and the data used for this population are known to have come from one village in Brazil (Ward *et al.* 1996). An argument for a selective mechanism working on the Chibchan populations can be ruled out because selection would probably act on a particular mitochondrial protein within the coding region of mtDNA and not in the hypervariable control region. Also, the fact that the Chibchan populations inhabit similar geographic areas as the Chocoan speakers indicating that if selective pressure were acting on the Chibchan populations it should also be reflected in Chocoan mtDNA diversity and this is clearly not the case. This indicates that these low mtDNA haplotype diversity levels in the Santa Marta populations support the idea that all Chibchan populations went through the same genetic bottleneck and only a fraction of the original Amerind mtDNA variation present in the initial migration was maintained in these groups.

Further evidence for a population bottleneck occurring in Chibchan populations comes from the two measures of selective neutrality (Tajima's D and Fu's Fs) shown in Table 10. Populations that have expanded will have large negative values. In the case of the Chibchan speaking and South American populations all of the values are either positive or low negative numbers and not statistically significant, indicating populations that are either undergoing drift or recently passed through a bottleneck. Only one of these values, belonging to the Yanamamõ, is negative and

statistically significant and this again may be associated with the aforementioned practice of linguistic exogamy. This differs dramatically from Tajima's D and Fu's Fs values for Siberian and Native North American populations where only one Amerind population (Bella Coola) had a positive Fu's Fs value and only one Na-Dene group (Haida) had a statistically non-significant value (Zlojutro *et al.* in press). This supports both the idea of genetic drift as the primary evolutionary force acting on South American populations (O'Rourke *et al.* 1992) and the idea that Chibchan populations have undergone a genetic bottleneck within the last 10,000 years (Kolman & Bermingham 1997, Barrantes *et al.* 1990). Some researches have argued that statistical measures such as Tajima's D may not be sufficiently robust enough to detect evolutionary events that have 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-Brosou & Excoffier 1996). However, the less conservative Fu's Fs statistic presented here does take into account mutation rate heterogeneity and gives similar results, suggesting that more than likely a population bottleneck did occur in Chibchan populations. Granted, even Fu's Fs may not be sensitive enough to detect recent population expansions or genetic bottlenecks, such as those that have certainly occurred in New World populations over the last 500 years.

Estimated timeframe for Chibchan genetic history

The use of pairwise genetic differences offers a means of approximating the date of Chibchan expansion after the proposed bottleneck. Mismatch distributions

(Figure 10) of the Arsario, Wayuú and Santa Marta Chibchan combined populations all shared a major peak between six and nine units indicating an ancient expansion between 55,000-90,000 YBP, corresponding to the first major expansion of human populations out of Africa (Harpending *et al.* 1993). This date is consistent with several other South American populations (Table 13) indicating a shared common maternal ancestor in the distant past. The Ijka were the only population that did not share this peak and show a jagged peak between zero and four units, suggesting that this population may have developed its distinct genetic characteristic in the New World. A major drawback to time estimates using mismatch distributions is that they do not take into account the order in which mutations may have occurred. If a wide range of populations share a mutation then it can be assumed that this mutation occurred prior to a mutation that is shared by only a single or small number of populations. This problem is solved using the method proposed by Sailard *et al.* (2000) that allows for ancestral and descendent nodes to be chosen from the network in order to estimate time depth using mtDNA coalescent data.

The three Santa Marta Chibchan populations also share a peak (above 0.2 on the Y-axis) between zero and one mutational unit (Figure 10). This peak is shared with several other Chibchan populations from lower Central America including the Kuna, Ngöbé, and Huetar (data not shown) (Batista *et al.* 1995). Previously, some researchers have argued that this major peak was indicative of a shared population bottleneck and used it to estimate that Chibchan genetic history began within the last 10,000 years (Kolman & Bermingham 1997, Batista *et al.* 1995, Kolman *et al.* 1995).

Others have suggested that bimodal or multimodal peaks in pairwise differences for mtDNA may be caused by differential haplogroup frequencies (Zlojutro *et al.* in press). As all of these Chibchan populations are primarily characterized by the presence of two of the five founding haplogroups, caution should be practiced in interpreting the results. Another concern is that there is no way of actually knowing what, if any, evolutionary force is causing this peak and so assuming that it is caused by a bottleneck and subsequent population expansion may be overreaching. Kolman & Bermingham (1997) did separate out the Ngöbé and Kuna by haplogroup A and reported an estimated date for these two populations of 10,900 YBP. This date is consistent with the time estimates given for the haplotype nodes SMA7 (6,985 YBP) and SMA3 (8,072 YBP) given in Table 13 using the ρ-statistic of Sailard *et al.* (2000). These dates are all consistent with other temporal estimates of Chibchan genetic history from classic genetic markers (Barrantes *et al.* 1990), and mtDNA (Kolman & Bermingham 1997) evidence that place the genetic origins of these populations between 7,000-10,000 YBP.

The inclusion of the Santa Marta Chibchan populations with the lower Central American groups indicates that the genetic origins for groups belonging to this language family probably occurred early during the peopling of the New World. This is further confirmed through genetic evidence that suggests populations in these two geographic areas are related but that this relationship was fomented in the distant past. There are also several lines of evidence pointing to a genetic bottleneck occurring in these populations. The fact that all Chibchan-speaking populations demonstrate low

mtDNA haplotype frequency, are characterized by rare unshared mtDNA haplotypes and all have low negative or positive Fu's Fs and Tajima's D values, and all share a mutational peak between zero and one mutational unit suggests that all of these populations experienced the same genetic event. This coupled with the knowledge that neighboring Chocoan populations who would have been subjected to the same ecological pressures as the Chibchan speakers do not share any one of these characteristics gives further support to this argument. This evidence also has notable connotations for both the peopling of the Santa Martas and for the origin of New World populations.

Peopling of the Santa Martas

Of the four proposed models for the initial peopling of the Santa Martas and the rise of stratified cultural complexity in the region, the mtDNA evidence appears to support an *in situ* development for the Santa Marta population as proposed by Bray (1984) and Langeback (2003). The mtDNA evidence presented here suggests that people were in the area long before the development of social complexity in the Santa Marta area. The presence of long term biological continuity in the matrilineages dating back to the Paleoindian time period also suggests that these populations have inhabited the Santa Marta region for a significant amount of time and that they developed their biological distinctiveness *in situ*. Based on the mtDNA coalescent dates ranging from 7,000 – 8,000 YBP and that the Santa Marta populations do not share any mtDNA haplotypes with any of the other South American or Central American groups suggests a long independent maternal

biological history for these groups. The absence of shared haplotypes suggests that the female Chibchan populations were not a genetic pathway for populations moving north out of South America, however a biological influence from Mesoamerica or the Caribbean can not be ruled out as there is a dearth of mtDNA data from these regions, especially those areas that may have been in contact with Chibchan speaking populations from the Santa Martas. The proposed timing of this event also seems to refute the idea that Paleoindian populations already in place in the Santa Marta area were subsumed by Chibchan populations moving out of the Panamanian isthmus due to ecological pressures or carrying newly acquired technologies as based on the dates given here due to the lack of a clear mtDNA relationship between the two cultures.

The mtDNA coalescent dates presented here suggest that these populations have been in the area since the late Pleistocene-Early Holocene. There are no known stratified Paleoindian archaeological sites in the Santa Marta but there are a dozen sites in lower Central America suggesting continuous occupation by humans over the last 10,000 years. Pearson (2002) argues that northern South America is both genetically and technologically heterogeneous compared to areas immediately north and south. During the Paleoindian period this region is characterized by the widespread use of fluting techniques on a variety of different point types and suggests that South America was probably already inhabited prior to the southward expansion of Clovis-like cultures moving into the region. Due to the natural constrictive barrier that lower Central America presents it is possible that first human who colonized Colombia may have buffered subsequent gene flow into the continent and suggests

that the isthmus may have not been an active corridor for human interaction. The extension of Chibchan genetic research to the Santa Marta of the South American continent would suggest that this extended along the northern Caribbean coast line.

Glottochronological evidence for the origin of Chibchan language suggests that the language arose due to increased sedentism as a result of the introduction of agriculture into the region between 6000 -4,000 YBP (Constenla 1991). mtDNA coalescent dates (with standard error) presented here overlap within this range but point to an earlier settlement of the region and imply that the earliest inhabits of the region were not migrating Chibchan-speaking populations and these groups may have adopted the language at a later date. However, the presence of long term biological continuity in the region suggests that these populations were not mobile and this may have facilitated the adoption of the Chibchan language. This indicates a need to understand why these populations are becoming more sedentary during the late Pleistocene – Early Holocene.

In order for this long term maternal biological continuity to be present in these populations for an extended period of time, two traits are necessary. The first trait is a matrilocal residence pattern, thus indicating a reduced migration rate for females between communities. The second trait is that these populations had to be relatively sedentary. This can be explained by the exploitation of marine or riverine resources, such as shellfish. There is archaeological evidence that populations were exploiting the resource within the Santa Marta region (Oyuela-Caycedo 1996). This increased sedentism would had led to an increase in population size and subsequently

necessitated the need for other resources that would had allowed for the adoption of agriculture. Granted, this hypothesis needs further development and evidence from both archaeology and ecology to either support or reject it.

Origin of New World populations

Understanding the microevolutionary role of Chibchan populations in the Americas has important implications for understanding the peopling of the Americas as well as the peopling of South America. The fact that you have endogenous linguistic development over a wide geographic area beginning around 10,000 YBP would indicate that humans would have to have existed in the area for a long time. This gives credence to an "early" migration (~20,000-15,000 YBP) of humans into the New World which is also supported by archaeological evidence from South American Paleoindian sites, such as Monte Verde in Chile, which has confirmed radiocarbon dates of approximately 14,000 YBP (Dillehay 2000). This evidence, coupled with the geographic barrier presented by the Andean cordillera in Colombia and the natural constriction effect of the narrow Panamanian isthmus, may have made resulted in these Chibchan-speaking populations blocking or slowing gene flow into South America as well as forcing gene flow north along the Andes. There is now significant evidence from a variety of scientific disciplines that support this theory including archaeology, (Pearson 2002) biology (Kolman & Bermingham 1997, Kolman *et al.* 1995, Batista 1995, Barrantes *et al.* 1990) and linguistics (Constenla 1991, 1981). This may have "corked the bottle" of South America making it plausible that genetic drift became the primary evolutionary force operating on the small

populations who had already entered the continent and may have contributed to the larger amount of genetic and spatial differentiation that is seen today in South American populations as opposed to North America groups (O'Rourke *et al.* 1992, Schanfield 1992). However, this "cork in a bottle" model that is being presented needs to be extended to other populations that may have had an impact on these Chibchan cultures. These include populations from other areas of Central America especially the countries of Nicaragua and Honduras that form the northern boundaries of the Chibchan linguistic frontier, populations from the Caribbean, populations from Mesoamerica, as well as native groups from North America. If it can be demonstrated clearly that there is no relationship between the north and the eastern frontiers then there would be little doubt that these populations formed a significant homogenous cultural area that impeded gene flow between the two continents.

Chapter VI: Conclusion

Over the last decade increased scientific inquiry has focused on Chibchan-speaking populations from lower Central and northern South America because of their geographic location bridging the two American continents. This research has rejected the traditional notion of the region as a heavily trodden pathway for migrating populations and has suggested long term occupancy and biological continuity dating back 10,000 years. While archaeological and linguistic research has found a relationship between the Central and northern South American Chibchan populations the biological relationship of these groups remains largely unresolved. This thesis characterized mtDNA haplogroup and haplotype diversity in 190 individuals from three Chibchan speaking (Kogí, Ijka, and Arsario) populations from the Santa Marta mountain region and a neighboring Arawak speaking (Wayuú) population from the Guajiro peninsula of northeastern Colombia. This thesis: (1) determines the relationship of the four study populations to each other; (2) determines the relationship of Chibchan-speaking populations to linguistically related populations from lower Central America; (3) compares this Chibchan-speaker relationship with archaeological, linguistic, and other biological evidence; (4) evaluates proposed models for the initial peopling of the Santa Martas; and (5) considers the role of Chibchan-speaking populations the initial peopling of the Americas.

This study demonstrated that the three Chibchan populations from the Santa Marta mountain range in northeast Colombia are not related to the neighboring

Wayuú but are related to each. This result may indicate that all of these populations are descendents of the ancient Tairona who constituted a large population that was destroyed by the Spanish in the early 17th century. The RFLP results of this study show that the Santa Marta Chibchan populations demonstrate a distant relationship with Chibchan-speaking populations from lower Central America. These populations also share low mtDNA haplotype diversity, low negative or positive Tajima's D or Fu's Fs values, a peak between zero and one mutational unit in pairwise difference analysis, and coalescent dates between 7,000-11,000 YBP. These dates are supported by archaeological and linguistic evidence that further corroborates this relationship. Due to the time depth, it is suggested that the Santa Marta Chibchan cultures probably developed in the region and then became stratified and culturally complex as time progressed.

This thesis offers further support for a Chibchan "population plug" that may have impeded or blocked gene flow moving through the Panamanian isthmus. The original reason for this "population plug" may have been due to the geographic barrier caused by the Andes transecting the northern portion of the continent and as the population numbers expanded backed up into the naturally constrictive isthmus, allowing for the development of an endogenous language in the Americas that has important implications for the peopling process. This "cork in a bottle" blocked the movement of genes between the two continents early on and possibly allowed for genetic drift to become the primary evolutionary force working on the South American continent. However, further evidence is needed from both the northern

boundaries of the Chibchan linguistic frontier, the Caribbean, Mesoamerica, as well ancient DNA samples to either support or refute this hypothesis.

Literature Cited:

Anderson S, Bankier AT, Barrell BG, De Bruijn MHL, Coulson AR, Drouin J, Eperon IC, et al. (1981) "Sequence and organization of the human mitochondrial genome" Nature 290:457-465

Angulo, Carlos (1962). "Evidencias de la serie Barrancoide en el norte de Colombia." in: Revista Colombiana de Antropología, 11: 73-88

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

Bailliet G, Rothammer F, Carnese FR, Bravi CM, Bianchi NO (1994) "Founder mitochondrial haplotypes in Amerindian populations" Am J Hum Genet. 54:27-33

Bandelt HJ, Forster P. Rohl A. (1999) "Median-joining networks for inferring intraspecific phylogenies." Mol Biol Evol.(1):37-48.

Bandelt HJ, Forster P, Sykes BC, Richards MB (1995) "Mitochondrial portraits of human populations using median networks" Genetics. 141(2):743-53.

Barbujani G & Sokal R (1990) "Zones of sharp genetic change in Europe are also linguistic boundaries." Proc Natl Acad Sci U S A. 87(5):1816-9.

Barrantes R, Smouse PE, Mohrenwieser HW, Gershowitz H, Azofeifa J. (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

Batista O, Kolman CJ, Bermingham E.(1995). "Mitochondrial DNA diversity in the Kuna Amerinds of Panama" Hum Mol Genet. 4(5):921-9.

Bert F., A Corella, M Gene, A Pérez-Pérez, D. Turbón (2001). "Major Mitochondrial DNA Haplotype Heterogenity in Highland and Lowland Amerindian Populations from Bolivia" Hum Biol 73:1-16.

Bianchi NO, Bailliet G, Bravi C (1995) "Peopling of the Americas as inferred through the analysis of mitochondrial DNA" Brazilian Journal of Genetics 18:661-668

Bieber, H., Bieber, S. W., Rodewald, A., and Barrantes, R. (1996). Microevolution and genetic affinities among six Amerindian tribes of Lower Central America: Comparative genetic study of serum proteins. Hum Biol 68: 929-953

Black, FL (1991)"Reasons for the failure of genetic classifications of South Amerindian populations" Hum. Biol. 63:763-764

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

Bray, W (2003) "Gold, Stone, and Ideology: Symbols of Power in the Tairona Tradition of Northern Colombia" in Quilter, J and JW Hoopes (eds) Gold and Power in Ancient Costa Rica, Panama, and Colombia. Dumbarton Oaks Research Library and Collections, Washington D.C

Bray, W (1984) "Across the Darien Gap: A Colombian View of Isthmian Archaeology".in: Archaeology of Lower Central America. Lange F. & D. Stone eds:. Alburquerque: University of New Mexico Press. 305-40

Briceño I, Gómez A, Lozano P, Umaña A, Mitchell RJ, Papiha, SS (n.d.)"Mitochodrial variation in Colombia: Study of matrilineal lineages among amerindian tribes."abstract for paper in progress.

Briceño I, Bernal JE, Duran S. Ortega J. Umaña A. Gómez A. Paphia SS (1996a) "HLA antigens in Amerindian groups of two different linguistic families from Colombia" European Journal of Immunogenetics 23:21-28

Briceño I. Gómez A., JE Bernal et al (1996b) "HLA-DPB1 polymorphism in seven south American Indian tribes in Colombia" European Journal of Immunogenetics 23:235-240

Cann RL, Stoneking M, Wilson AC (1987) "Mitochondrial DNA and human evolution" Nature. 325(6099):31-6.

Cardoza, SV & HD Correa (2000) "Los Wayuu, Entre Juya ("El que llueve"), mma ("la tierra") y el Desarollo Urbano Regional" Geografía Humana de Colombia Nordeste Indígena (Tomo II) Bogotá, Bank of the Republic

Constenla, UA (1991) Las Lenguas del Área Intermedia: Introducción a su studio areal Editorial de la Universidad de Costa Rica, San José.

Constenla, UA (1981) Comparative Chibchan Phonology Ph.D. dissertation, Department of Linguistics, University of Pennsylvania, Philadelphia.

Crawford, MH (1998) The Origins of Native Americans: Evidence from AnthropologicalGenetic New York, Cambridge University Press

Crawford MH, JT Williams, R. Duggirala (1997) "Genetic Structure of the Indigenous Populations of Siberia" Am J Phys Anthropol 104:177-192

Crawford, MH, R. Duggirala (1992) "Digital Dermatoglyphic Patterns of Eskimo and Amerindian Populations: Relationships between Geographic, Dermatoglyphic, Genetic, and Linguistic Distances" Human Biology 64:683-704

Demarchi DA, GM Panzetta-Dutari, CC Motran, M. López de Basualdo, AJ Marcellino (2001) "Mitochondrial DNA Haplogroups in Amerindian Populations from the Gran Chaco" Am J Phys Anthropol 115:199-203

Devor, EJ (2004) Polymerase Chain Reaction (PCR): Theory, Method, and Practice Integrated DNA Technologies, Corralville, Iowa

Dipperi JE, E. Alfaro, VL Martínez-Marignac, G. Bailliet, CM Bravi, S Cejas, NO Bianchi (1998). "Paternal directional mating in two Amerindian subpopulations located at different altitudes in northwestern Argentina" Hum Biol 70:1001-1010

Easton RD, Merriwether DA, Crews DE, Ferrell RE (1996)."mtDNA variation in the Yanomami: evidence for additional New World founding lineages." Am J Hum Genet. 59(1):213-25.

Excoffier L, Smouse PE, Quattro JM (1992) "Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data" Genetics. Jun;131(2):479-91.

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

Fagundes NJR, SL Bonatto, SM Callegari-Jacques & FM Salzano (2002) "Genetic, Geographic, and Linguistic Variation Among South American Indians: Possible Sex Influence" Am. J. Phys. Anthropol.: 117:68-78

Fox, CL (1996) "Mitochondrial DNA Haplogroups in Four Tribes from Tierra del Fuego-Patagonia: Inferences about the Peopling of the Americas" Hum Biol 68(6):855-871

Francalacci P, Montiel R., Malgosa A (1999). "A Mitochondrial Database: Applications to Problems of Nomenclature and Populations Genetics" in Papiha SS, R. Deka, and R. Chakraborty (eds) Genomic Diversity: Applicatons in Human Population Genetics Kluwer Academic/Plenum Publishers New York pp103-120

Fu, YX (1997) "Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection" Genetics. 147(2):915-25.

Ginther C., D. Corach, GA Penacino, JA Rey, FR Carnese, MH Hutz, A. Anderson, J. Just, FM Salzano MC. King (1993) "Genetic Variation among the Mapuche Indians from the Patagonian region of Argentina: Mitochondrial DNA sequence variation and allele frequencies of several nuclear genes" In: Pena SD, R.Charkaborty, JT Epplen , AJ Jeffreys DNA Fingerprinting: State of the Science Birkhäuser-Verlag, Basel, pp 211–219

Goicoechea AS, Carnese FR, Dejean C, Avena AS, Wiemer TA, Franco MHLP, Callegari-Jacques SM, Estalote AC, Simoes MLMS, Palatnik M, Salzano FM (2001)"Genetic Relationships between Amerindian population of Argentina" Am. J. Phys. Anthropol. 115:133-143

Graven L, Passarino G, Semino O, Boursot P, Santachiara-Benerecetti S, Langaney A, Excoffier L. (1995) "Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka sample." Mol Biol Evol. 12(2):334-45.

Greenberg JH, Turner CG, Zegura SL (1986) "The settlement of the Americas: a comparison of the linguistic, dental and genetic data" Curr. Anthropol 27:477-495

Guarino FD, L. Federele, RAH van Oorschot,I. Briceño, JE Bernal, SS. Papipha, MS Schanfield, RJ Mitchell. (1999) "Genetic Diversity among five native American Tribes of Colombia: evidence from nine autosomal microsatellites" in SS Papiha, R. Deka, R. Chakraborty eds. Genomic Diversity: Application in Human Population Genetics Kluwer Academic, New York pp. 33-51.

Haprending, HC, ST Sherry, AR Rogers, M. Stoneking (1993) "The genetic structure of ancient human populations" Curr. Anthropol. 34: 483-496.

Harpending, H & Rogers A. (1984) ANTANA: A package for multivariate data analysis. Distributed by the authors.

Harpending H. and Jenkins T. (1973) "Genetic Distance among Southern African Populations" In: Crawford, MH and PL Workman (eds): Methods and Theories of Anthropological Genetics. University of New Mexico Press, Albuquerque. pp. 177-199

Hoopes, JW (2005a) "The Emergence of Social Complexity in the Chibchan World of Southern Central America and Northern Colombia, AD 300-600." Journal of World Prehistory, *in press.*

Hoopes, JW (2005b) "Chibchan Cultures of the Caribbean Littoral: Exploring Precolumbian Connections Between Colombia and Costa Rica" *in press*.

Hoopes JW & OM Fonseca "Goldwork and Chibchan Identity: Endogenous Change and Diffuse Unity in the Isthmo-Colombia Area"(2003) in Quilter, J and JW Hoopes (eds) Gold and Power in Ancient Costa Rica, Panama, and Colombia. Dumbarton Oaks Research Library and Collections, Washington D.C., pp 49-89

Horai, S, Kondo R. Hattori YK, Hayashi S, Sonoda S, Tajima K (1993)"Peopling of the Americas, Founded by Four Major Lineages of Mitochondrial DNA" Mol Biol Evol 10(1):23-47

Hudson RR (1990) "Gene genealogies and the coalescent process" in Futuyama & JD Antonivies (eds). Oxford Surveys in Evolutionary Biology. Oxford University Press, New York. 1-44.

Keyeux G, Rodas C, Gelvez N, Carter D. (2003) "Possible migration routes into South America deduced from mitochondrial DNA studies in Colombian Amerindian populations." Hum Biol. 74(2):211-33.

Kolman, CJ., and Bermingham, E. (1997). Mitochondrial and nuclear DNA diversity in the Choco and Chibcha Amerinds of Panama. Genetics 147: 1289-1302.

Kolman CJ, Bermingham E, Cooke R, Ward RH, Arias TD, Guionneau-Sinclair F (1995). "Reduced mtDNA diversity in the Ngöbé Amerinds of Panama" Genetics. 140(1):275-83.

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

Lange, FW & DZ Stone (eds) (1984) The Archaeology of Lower Central America. School of American Research and University of New Mexico Press, Albuquerque.

Langeback, CH (2003) Poblameinto Prehispánico de las Bahías de Santa Marta: Contribucíon al studio del Desarrollo de los Caciazgos Tairona del Norte de Colombia
Universidad de los Andes, Bogatá, Colombia

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, Z., Layrisse, M, Rodríguez-Larriade, A. (1995) "Estudios genéticos en pobliaciones de lengua Chibcha" Boletín del Museo del Oro 38-39: 71-86.

Layrisse, M, Layrisse Z, Wilbert J (1963) "Blood Group Antigen Studies of Four Chibchan Tribes" American Anthropologist 65(1):36-55.

Layrisse, M, Layrisse Z, Wilbert J (1961) "Blood Group Antigens in Goajiro Indians" Am J Phys Anthropol 19:255-62

Layrisse, M & J. Wilbert (1961) "Absence of the Diego Antigen: A Genetic Characteristic of Early Immigrants of South America" Science. 134:1077-1098.

Lorenz JG, & Smith DG (1996)"Distribution of four founding mtDNA haplogroups among Native North Americans." Am J Phys Anthropol. 101(3):307-23.
Luiselli D, Simoni L, Tarzano-Santos E, Pastor S, Pettener D (2000)"Genetic Structure of Quechua-speakers of the central Andes and geographical patterns of gene frequencies in South Amerindian populations" Am. J. of Phys. Anthropol. 113:5-17

Menozzi P, Piazza A, Cavalli-Sforza L. (1978)" Synthetic maps of human gene frequencies in Europeans". Science. 1978 Sep 1;201(4358):786-92.

Merriwether DA, Rothammer F, Ferrel 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 98:411-443

Mesa NR, Mondragon MC, Sote ID, Para Maria, Duque C., Ortiz-Barrientos D, Garcia LF, Velex ID, Bravo ML, et al. (2000) "Autosomal, mtDNA, and Y-Chromosome Diversity in Amerinds: Pre - and Post-Columbian Patterns of Gene Flow in South America" Am J Hum Genet.67:1277-1286

Mitchell RJ, Federle L, Sofro ASM, Papiha SS, Briceno L., Bernal JE (2000) "Further Data on the Microsatellite Locus D12S67 in Worldwide Populations: An Unusual Distribution of D12S67 Alleles in Native America" Hum Biol 72(4):697-705

Moraga ML, Rocco P, Miquel JF, Nervi F, Llop E, Chakraborty R, Rothammer F, 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 113:19-29

Muse, T. (2004) "Lost City" Archaeology Sept/Oct. 57(4):18-23

Neel JV, Chagnon NA (1968) "The demography of two tribes of primitive, relatively unacculturated American Indians" Proc Natl Acad Sci U S A. 59(3):680-9.

O'Rourke DH, Hayes MG, 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, Suarez BK (1992) "Patterns of Genetic Variation in Native America" Hum Biol 64:417-434

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

Oyuela-Caycedo, A.. (1996). "The study of Collector variability in the Transition to sedentary Food Producers in Northern Colombia." Journal of World Prehistory 10(1)

Oyuela-Caycedo, A. (1987). ""Gaira": Una introducción a la ecología y arqueología del litoral de la Sierra Nevada de Santa Marta". Boletín del Museo del Oro 19: 34–55.

Oyuela-Caycedo, A. (1986) "Contribución a la periodización cultural en el litoral del Parque Tairona". Boletín de Arqueología 1 (2): 24-8.

Park, WZ (1965) "Tribes of the Sierra Nevada de Santa Marta, Colombia" in J. Steward ed.Handbook of South American Indians. Bulletin Vol. 143, Bureau of American Ethnology, Smithsonian Institution, Washington D.C.

Pearson, GA (2002) Pan-continental Paleoindian expansions and interactions as viewed from the earliest lithic industries of lower Central America. Ph.D. Dissertation, University of Kansas, Lawrence

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

Quilter, J and JW Hoopes (eds) (2003). Gold and Power in Ancient Costa Rica, Panama, and Colombia. Dumbarton Oaks Research Library and Collections, Washington D.C

Ramos-Onsins SE, Rozas J. (2002) "Statistical properties of new neutrality tests against population growth." Mol Biol Evol. Dec. (12):2092-100

Reichel-Dolmatoff, G. (1965). Colombia. Thames and Hudson. London.

Reichel-Dolmatoff, G. (1950) The Kogi: a tribe of the Sierra Nevada de Santa Marta, Colombia. Vol. 1 Bogotá: El Instituto, 1949-1950. HRAF MS: 342 p. [original: 319 p.], plates: ill., map, Published in English exclusively online at eHRAF.

Rogers AR & Harpending H (1992) "Population growth makes waves in the distribution of pairwise genetic differences." Mol Biol Evol.May;9(3):552-69.

Rickards, O, Matinez-Labarga C. Lum JK De Stefano GF, 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. 65:519-530

Rubicz R, Schurr TG, Babb PL, Crawford MH. (2003) "Mitochondrial DNA variation and the origins of the Aleuts." Hum Biol. 75(6):809-35.

Ruhlen M. (1991). A Guide to the World's Languages: Volume 1: Classification 2nd edition. Stanford: Stanford University Press

Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N. (1985) "Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia." Science. 230(4732):1350-4.

Saillard J, Forster P, Lynnerup N, Bandelt HJ, Norby S (2000) "mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion." Am J Hum Genet. 67(3):718-26.

Saler B (1992) "The Guajiro". In "Borders and Peripheries in Lowland South America," Working Papers on South American Indians 4:37-40, Bennington College, Bennington, Vermont.

Salzano FM (2002) "Molecular variability in Amerindians: widespread but uneven information" An Acad Bras Cienc 74(2):223-63

Salzano FM & SM Callegari-Jacques (1988) South American Indians: A case study in evolution Oxford Clarendon Press

Santos M, Ward RH, Barrantes R. (1994) "mtDNA Variation in the Chibcha Amerindian Huetar from Costa Rica" Hum Biol 66(6):963-977

Schanfield MS. (1992) "Immunoglobulin allotypes (GM and KM) indicate multiple founding populations of Native Americans: evidence of at least four migrations to the New World." Hum Biol. 64(3):381-402.

Schneider, S, Roessli, D, Excoffier, L (2000) Arlequin ver. 2.00: A software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Schneider S, Excoffier L.(1999). "Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA." Genetics. 152(3):1079-89

Schurr TG & ST Sherry (2004) "Mitochondrial DNA and Y chromosome diversity and the peopling of the Americas: evolutionary and demographic evidence" Am J Hum Biol. 16(4):420-39.

Schurr TG, Ballinger SW, Gan YY, Hodge JA, Merriwether DA, Lawrence DN , Knowler WC, Weiss KM, Wallace DC (1990) Amerindian mitochondrial DNAs have a rare Asian mutation at high frequencies suggesting that they derived from four primary maternal lineages" Am J Hum Genet.46:613-623

Silva WA Jr, Bonatto SL, Holanda AJ, Ribeiro-Dos-Santos AK, Paixao BM, Goldman GH, Abe-Sandes K, Rodriguez-Delfin L, Barbosa M, Paco-Larson ML, Petzl-Erler ML, Valente V, Santos SE, Zago MA.(2002)."Mitochondrial genome diversity of Native Americans supports a single early entry of founder populations into America." Am J Hum Genet. 71(1):187-92.

Simonsen KL, Churchill GA, Aquadro CF "Properties of statistical tests of neutrality for DNA polymorphism data." Genetics. Sep;141(1):413-29.

Smith D, Malhi R, Eshleman J, Lorenz J, and Kaestle F (1999)"mtDNA haplogroup X among Native North Americans" Am J Phys Anthropol 110:271-284

Stone AC & M Stoneking (1998) "mtDNA analysis of a prehistoric Oneonta population: implications for the peopling of the New World" Am J Hum Genet 62:1153-1170

Suarez BK, O'Rourke DH, Crouse JD. (1985) "Genetic variation in North Amerindian populations: association with sociocultural complexity." Am J Phys Anthropol. 1985 Jul;67(3):233-9

Tajima, F (1993) "Measurement of DNA polymorphism" in Takahata, N & AG Clark (eds) Mechanisms of Molecular Evolution: Introduction to Molecular Paleopopulation Biology. Japan Scientific Societies Press. 37-59.

Torroni A, Huoponen K, Francalacci P, Petrozzi M, Morelli L, Scozzari R, Obinu D, Savontaus ML, Wallace DC. (1996). "Classification of European mtDNAs from an analysis of three European populations." Genetics. 144(4):1835-50

Torroni A, Neel JV, Barrantes R, Schurr TG, Wallace DC (1994) "Mitochondrial DNA "clock" for the Amerinds and its implications for timing their entry into North America" Proc Natl Acad Sci U S A. 91(3):1158-62.

Torroni A, Sukernik RI, Schurr TG, Starikorskaya YB, Cabell MF, Crawford MH, Comuzzie AG, Wallace DC. (1993a) "mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans" Am J Hum Genet. 53(3):591-608.

Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, Vullo CM, Wallace DC (1993b) "Asian Affinities and continental radiation of the four founding Native American mtDNA's" Am J Hum Genet 53:563-590

Torroni A, Schurr TG, Yang CC, Szathmary EJ, Williams RC, Schanfield MS, Troup GA, Knowler WC, Lawrence DN, Weiss KM, et al.(1992). "Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations" Genetics. 130(1):153-62

Torroni A, Stepien G, Hodge JA, Wallace DC (1990) "Neoplastic transformation is associated with coordinate induction of nuclear and cytoplasmic oxidative phosphorylation genes" J Biol Chem. 265(33):20589-93.

Uribe, CA (2000) "La Ethnografia de la Sierra Nevada de Santa Marta y las Tierras Bajas Adyacentes" in Geografía Humana de Colombia Nordeste Indígena (Tomo II) Bogotá, Bank of the Republic.

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

Vigilant L, Pennington R., Harpending H., Kocher TD, Wilson AC (1989)"Mitochondrial DNA sequences in single hairs from Southern African populations" Proc Natl Acad Sci USA 86:9350-9354.

Wallace, D. C. and Lott, M. T. (2004) "MITOMAP: A Human Mitochondrial Genome Database" http://www.mitomap.org.

Ward, R.H., Salzano, F.M., Bonatto, S.L., Hutz, M.H., Coimbra, C.E.A., and Santos, R.V. (1996) "Mitochondrial DNA polymorphism in three Brazilian Indian tribes". Am J Hum Biol. 8 (3): 317-323

Ward RH, Frazier BL, Dew-Jager K, Paabo S.(1991) "Extensive mitochondrial diversity within a single Amerindian tribe." Proc Natl Acad Sci U S A. 88(19):8720-4

Wilson, D. (1999) Indigenous South Americans of the Past and Present-An Ecological Perspective. Boulder: Westview Press.

Wynn, J (1975) Buritaca Ceramic Chronology: a seriation from the Tairona Area, Colombia. PhD. Dissertation. University of .Missouri.

Yunis, JJ, Ossa H. Salazar M, Delagado MB, Deulofeut R, de la Hoz Antonio, Bing DH, Ramos O. Yunis EJ, Yunis EJ (1994) "Major Histocompatibility Complex II Alleles and Haplotypes and Blood Groups of Four Amerindian Tribes of Northern Colombia" Human Immunology 41:248-258

Zlojutro, M., R. Rubicz, E.J. Devor, et al (in press) "The genetic structure of the Aleuts and Circumarctic populations based on mitochondrial DNA sequences." Am J Phys Anthropol