The Creation of Metropolitan Amazonia: Genetic Consequences of Migration and Urbanization

By

Randy E. David
M.A., State University of New York at Binghamton, Binghamton, NY, 2011
Graduate Certificate, State University of New York at Binghamton, Binghamton, NY, 2011
B.S., State University of New York at Binghamton, Binghamton, NY, 2009

2019

Submitted to the Graduate Degree Program in Biological Anthropology and the Faculty of the Graduate School of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

_____
Chairperson: Michael H. Crawford, Ph.D.

_____
Bartholomew Dean, Ph.D.

_____
Dennis H. O'Rourke, Ph.D.

_____
Maria E. Orive, Ph.D.

_____
Jennifer Raff, Ph.D.

Date of Defense: July 22, 2019

The dissertation committee for Randy E. David certifies that this is the approved version of the following dissertation:

The Creation of Metropolitan Amazonia: Genetic Consequences of Migration and Urbanization

_____

Chairperson: Michael H. Crawford, Ph.D.

Date Approved: September 3, 2019

Abstract

Uniparental genetic markers were analyzed in a sample of 182 individuals from the Peruvian Amazonian city of Yurimaguas. The site of extensive pre-Columbian and modern migratory fluctuation, contemporary Yurimaguas is characterized by considerable population diversity. Individually unique migratory and demographic profiles were combined with population genetic methodology to cultivate a co-constitutive view of the establishment of a modern Amazonian urban center. Yurimaguas, like many cities in the Global South, continues to experience the multifarious effects of accelerated migration and urbanization on population architecture. Biodemography and population genetics are used to assess consequences and evaluate potential contributing environments and events. A nuanced, interdisciplinary perspective through both a populational and molecular lens on the creation of today's urban populations is critical to informing the following: (1) our evolutionary development, historical interactions, and prior movements, (2) the relationship between genetic diversity and the evolutionary forces of gene flow, genetic drift, and natural selection, and (3) human migratory and mating behavior.

Key Terms: Ancestry-Informing Markers (AIMs), Biodemography, Human Migration, Non-Recombining DNA, Peruvian Amazon, Population Genetics, Urbanization

This dissertation is dedicated to my family, a reservoir of infinite encouragement, support, and love.

Acknowledgements

I am eternally obliged to Dr. Michael H. Crawford for his unwavering leadership and character. I am most grateful to Dr. Bartholomew Dean for his humanity and devotion. I would like to thank Dr. Dennis H. O'Rourke for his generosity. I am appreciative of the assistance of Drs. Maria E. Orive, Jennifer Raff, and Justin Tackney. Colleagues and friends Akbar Amat, Cyrus Azizi, Dr. Kristine Beaty, Dr. Melody Ratliff-Wood, Dr. Norberto Baldi-Salas, Dr. Fazel Hadadifard, Dr. Abdelmajid Hannoum, Dr. Joshua Homan, and Dr. Gabriel de Oliveira have made my time at the University of Kansas an inspiring period of scholarship and growth. I would like to thank my support system dating back to my first day in higher education, including: Emily Barile, Dr. Nicole Barile, Ezechiel Bly, Paul Choe, William Farber, David Fong, Valbona Gashi, Dr. Ronald Gomez, Dr. Amelia Harris, Kyle Hoellger, John Keith Keaveny, Jonathan Kim, Dr. Zachary Logozio, Max Mileaf, Dr. Tyler J. Miranda, Craig Morgan, Dr. Deanna Scoca, Aryssa Wernle, Kymel Yard, and Dr. Daniel Zheng. Lastly, I would like to thank the people of Yurimaguas, Peru for their acceptance, sincerity, and patience.

This dissertation was completed thanks to the generous support of the following:

- Carol Clark Grant, University of Kansas
- Dean's Office Research Excellence Fund, Graduate Student and Postdoctoral Fund Award, University of Kansas
- GENO 2.0 (Genographic Project) Grant, National Geographic Society
- Graduate Studies Research Support Award, University of Kansas
- Graduate Studies Summer Fellowship, University of Kansas
- Program in Biological Anthropology, University of Kansas
- University of Kansas Institute for Policy and Social Research, Center for Migration Research, Graduate Student Grant Program
- University of Kansas Institute for Policy and Social Research, Center for Migration Research, Seed Grant Program

Table of Contents

List of Figures

List of Tables

Chapter I: Introduction

Much of the research on the history of migration in the Americas has underscored the continent's complex pattern of peopling (Crawford, 1998; Gnecchi-Ruscone *et al*., 2019; Harris *et al*., 2018; Lewis *et al.*, 2007; Lewis, Tito, Lizárraga, & Stone, 2004). Scholarship on the Indigenous centers of power at the time of European contact has included the Peruvian Andes, and the Valley of Mexico (Harris *et al*., 2018; Romero-Hidalgo *et al*., 2017; Santana *et al*., 2014). Studies on modern transnational migrations emphasize the U.S.–Mexico border (Vila, 2000). In contrast, significantly less attention has been accorded to assessing population movements in other key areas of the Americas, most notably Upper Amazonia (Alexiades, 2009; de Jong, Lye, & Abe, 2006). To wit, scant scholarship has been dedicated to exploring the biodemographic and population genetic consequences of migration to Amazonia's inchoate metropolitan centers. This is unexpected given that urbanization—the dynamic, cosmopolitan process of people coming together *en masse* to form villages, towns, and cities—has been a defining aspect of both colonial and postcolonial Amazonia (David & Dean, in press; Dufour & Piperata, 2004). While clearly consequential in terms of natural resource exploitation, public health initiatives, economic opportunities, and regional development trajectories, population mobility in Amazonia is seldom studied.

Given recently renewed interest in neotropical archaeological and historical records, as well as its diverse biomes and sociocultural variation, Peruvian Amazonia offers an exceptional opportunity for the study of the *longue durée* of human migration (Hornborg & Hill, 2011; Pärssinen, Schaan, & Ranzi, 2009). Useful comparative information for understanding what some have termed the "urban revolution"—arguably the most consequential disruption known to humanity, is furthermore attainable (Lefebvre, 2014). Though often emplaced in popular Western imaginaries as bucolic, nearly four-fifths (79%) of the Peruvian populace reside in census-designated urban areas, such as the historically important colonial city of Yurimaguas (World Bank, 2014)

Most scholarship on Peruvian intranational migration has emphasized population movements amongst Indigenous peoples in the Andean highlands (*sierra*), and among European migrants in large conurbations along the Pacific Coast (*costa*) (Cabana *et al.,* 2014; Lindo *et al*. 2018; Little, Thomas, & Garruto, 2013; Messina *et al*., 2008; Rademaker *et al*., 2014). Despite its status as *terra incognita*, the lowland tropical forested region (*selva baja*), specifically Peru's Lower Huallaga River Valley, holds particular import for applied studies on human migration and neotropical urbanization.

The Huallaga River Valley is significant precisely because it flanks an area critical to understanding the peopling of South America: the border between the Andean Cordillera and the Amazon Basin. A geographically consequential crossroads between the tropical Andean intermontane (*selva alta)* and lowland rainforests, Yurimaguas and the Lower Huallaga Valley are characterized by an extensive degree of ethnic diversity, including both Indigenous and immigrant populations (David & Dean, in press). Despite centuries of cataclysmic disruptions associated with European-inspired colonization and extractive industry, language remains an important marker of contemporary collective identity, and perhaps not surprisingly language mixing in Indigenous lowland South America has often been opposed on ideological grounds (Aikhenvald, 2010). Many Indigenous societies reside in the *selva baja* region surrounding the city of Yurimaguas, including: Quechua-speaking populations (Kichwa Lamista, Kichwa del Pastaza); Jivaroan speakers (Achuar [Shiwiar, Jivaro], Aguaruna [Awajún], Kandozi, and Wampis); Tupi-Guaraní speakers (Kukama-Kukamira [Cocama-Cocamilla]); Cahuapanan speakers (Chayahuita [Shawi] and Jebero); Arawakan speakers (Chamicuro); and the Urarina (Kachá), a linguistic isolate (Justice, Dean, & Crawford, 2012). Foreign immigrants to the region brought Spanish (Castilian), Basque (Euskara), Italian, Arabic, Portuguese, Cantonese, Japanese, Hebrew, English, and other languages (Cook, 2016; Dean, 2009; Harrison, 1995; Takenaka, 2004). Though language is a critical aspect of cultural identity, one must keep in mind that essentialist notions of ethnic identities linking language, culture, and biology have obfuscated the actual distribution of population-based genetic markers in Amazonia (Hornborg & Hill, 2011). Indeed, the contemporary nature of Amazonian ethnolinguistic diversity underscores the fluid, dialectic

relationship among population identity, language, genetics, demographics, and the disruptions associated with “post-contact” encounters (Dean & Levi, 2003; Merriwether, Kemp, Crews, & Neel, 2000).

Like neighboring *mestizo* (individuals of admixed European and Indigenous ancestry) communities, Indigenous societies of the Lower Huallaga Valley have been dramatically impacted by transformations in migration patterns, increasingly underwritten by global interconnections (Tapayuri Murayari, 2012; Wang *et al*., 2008). In light of Upper Amazonia’s recent experience of accelerated urbanization, this dissertation will emphasize the population genetic and biodemographic consequences of both historical and contemporary migrations (Dean, Silverstein, Reamer, & Homan, 2011). Rural-to-urban migration in Latin America has been associated with education level, economic status, disease prevalence, fertility, and morbidity (Dufour & Piperata, 2004).The genetic implications of migration and urbanization in Latin America have likewise been found to be diverse and extensive: introducing foreign haplotypes, increasing molecular diversity indices and admixture measures, creating novel genetic amalgamations, causing subpopulational microdifferentiation, and affecting the rates of specific disease variants. Understanding the complex interrelationship between migratory processes, biodemography, evolutionary theory, and population genetics is integral to elucidating the sophisticated underpinnings of these genetic consequences and ensuring long-term population health.

The city of Yurimaguas provides an ideal model population for the study of these relationships. Through the collection and statistical analysis of migratory and fundamental demographic data, coupled with the procurement and laboratory-based analysis of uniparental genetic markers, this research evaluates the biodemographic and population genetic consequences of the movement of peoples.

The research questions and their associated hypotheses for this study include the six primary foci of interest:

•*Research Question 1:*

What is the uniparental marker-based continental-ancestry composition of Yurimaguas? Moreover, is there evidence of sex-skewed gene flow?

Hypothesis 1:

I hypothesize that distributions in both maternal and paternal markers will reflect a majority of Indigenous ancestries, followed by European ancestries, and lastly a small number of East Asian and Sub-Saharan African ancestries. I hypothesize that compared to maternal ancestry, paternal ancestry will reflect a much greater percentage of European lineages (but still a majority of Indigenous paternal lineage). Compared to paternal ancestry, maternal ancestry will reflect a much greater percentage of Indigenous lineages. Therefore, statistically significant populational sex-skewed gene flow will be observed. Both maternal and paternal ancestries will reflect approximately equal proportions of East Asian and Sub-Saharan African lineages.

- *Research Question 2:*

  What is the pattern of genetic diversity in Yurimaguas?

Hypothesis 2:

I hypothesize that Yurimaguas will exhibit very high genetic diversity indices, in particular, nucleotide diversity, in comparison to neighboring populations. I moreover hypothesize that individuals with Indigenous maternal lineages will possess greater diversity than those with European maternal lineages. All demographic and migratory population subdivisions will also exhibit high genetic diversity.

- *Research Question 3:*

  Is there evidence to suggest that the Yurimaguas sample is in population size equilibrium versus demographic expansion, according to neutrality tests and mismatch distribution analyses? Is there evidence of amalgamation?

Hypothesis 3:

I hypothesize that Yurimaguas will show evidence for demographic expansion according to both neutrality test calculations (particularly Fu's F_S), and mismatch distribution analysis. I furthermore predict a statistically nonsignificant result for Chakraborty's test of population amalgamation.

• *Research Question 4:*

Does the distribution of maternal continental-ancestry differ significantly according to demographic variables? Likewise, does the distribution of paternal continental-ancestry differ significantly according to demographic variables?

Hypothesis 4:

I hypothesize that maternal continental-ancestry distribution will not differ significantly according to any of the tested demographic variables. I hypothesize that paternal continental-ancestry, however, will differ significantly according to the demographic variable of living environment (peri-urban or urban), but not according to any of the other tested demographic variables.

•*Research Question 5:*

Does the distribution of maternal continental-ancestry differ significantly according to migratory variables? Likewise, does the distribution of paternal continental-ancestry differ significantly according to migratory variables?

Hypothesis 5:

I hypothesize that maternal continental-ancestry will differ significantly according to two of the migratory variables, namely history of migration (yes or no), and number of migration episodes experienced (1, 2, or 3+). I hypothesize that paternal continental-ancestry will differ significantly according to both number of migration episodes experienced (1, 2, or 3+), and migration duration (0–5 years, 6–10 years, or 11+ years).

• *Research Question 6:*

Is there evidence to suggest that demographic subdivisions of the Yurimaguas sample can be genetically detected at the nucleotide level using mitochondrial DNA (mtDNA) sequence data? Is there evidence to suggest that migration-based subdivisions of the Yurimaguas sample can be genetically detected at the nucleotide level using mitochondrial DNA sequence data?

Hypothesis 6:

I hypothesize that subdivisions of the Yurimaguas sample will be able to be genetically detected at the nucleotide level according to the demographic variable of language type (those who speak an autochthonous language or those who do not speak an autochthonous language). I moreover hypothesize that subdivisions of the Yurimaguas sample will be able to be genetically detected at the nucleotide level according to one of the migration-based variables, history of migration (yes or no).

Chapter II: Historical Background

The social topography of Peruvian Amazonia has a number of contours, peaks, turns, and troughs that provide much-needed context for understanding genetic variation in a regional population. In isolation, information regarding genetic variability within and between populations has limited import. Without a familiarity of both a region's ethnohistory and the contemporary issues facing its peoples, interpretation of biological, including genetic, variability is constrained. Biodemography and population genetics both seek to explicate the processes that shape the molecular structure of populations, using, in this case, uniparental genetic markers. This chapter recounts the history of Amazonia, with a particular focus on Yurimaguas and the Lower Huallaga Valley, critical areas in Peru's Loreto Region.

Geography

The Amazon Basin

The Amazon Basin is the most biodiverse and largest tract of contiguous tropical forest in the world (Field Museum, 2013). The "basin" itself is a wide trough nestled between the Guiana Highlands to the north and the Brazilian Highlands to the south (Lathrap, 1970). The Amazon Basin is further defined by the peaks of the Andean Cordillera to the west, and the Atlantic coast to the east. By the Eocene Epoch, Amazonia was "physiognomically recognizable as rainforest and taxonomically allied to modern neotropical rainforests" (Burnham & Johnson, 2004, p. 1595). Today, the Amazon Basin encompasses circa 6,700,000 Km^2, of which ~570,000,000 hectares (ha) is tropical forest (approximately half of the world's remaining tropical forests), and ~6,600 Km of river systems (World Wildlife Fund, 2017). The Amazon River is among the longest in the world, and in terms of annual discharge is far greater than any other river system, five times more than that of the Congo River, and twelve times greater than the Mississippi River (Lathrap, 1970). The Amazon River is the main artery in a network of over a thousand tributaries, including the

Huallaga, Marañón, and Paranapura, that drain into the Atlantic Ocean from their genesis in the Andes (Carmichael, Hugh-Jones, Moser, & Taylor, 1985). It is along these tributaries that the majority of the original inhabitants of the *selva baja* first settled.

The Lower Huallaga River Valley

The Lower Huallaga River Valley is located in the far western borderlands of Amazonia, adjacent to the Andean *selva alta* (also known as the *chachapoya*, cloud forest, piedmont, or *yungas*). The *selva alta* acts as a transition zone from the high plateaus and mountain peaks of the Andes, to the heavily forested and mostly riverine environments typical of the lowland forest. It's location and natural topography have long conferred it import regarding human migration, communication, and trade patterns (Justice *et al*., 2012). The western boundary of the Huallaga Valley is perhaps most pronounced at the *Pongo de Manseriche* rapids of the Marañón River (Dean, 2009). Once traversing this series of cascades, one has effectively entered the lowland forest. Further west are the forested foothills of the eastern Andean slopes. Due east is a vast extension of the lowland forest, interspersed with numerous waterways. Within the *selva baja* there are interfluvial zones*, i.e*. upland areas characterized by billabongs and smaller waterways that rarely flood. There are also lowland riverine zones that frequently flood and experience the deposition of river silts. Consequently, lowland riverine zones typically have much more fertile soils, leading to a greater diversity of both flora and fauna, and sustaining much larger human populations (Meggers, 1971).

Perhaps most crucial to human history in the Lower Huallaga Valley is the ameliorated angle of ascent of the Nazca tectonic plate that forms its western border, known as the "Peru Flat" (Justice *et al*., 2012). This, along with a lack of volcanoes in the area, and the relatively low altitude of Andean peaks to the immediate west, attracted peoples for millennia as a region of trans-Andean circulation of humans, information, flora, fauna, and manufactured goods (Myers, 1974; Ramos & Folguera, 2009). There is archaeological evidence for such traversals from the

1440's when the Inca, led by Pachacutec, led militia into the region, then known as *Antisuyo*, in pursuit of member-groups of the Confederation of Chanca, a contemporary enemy (Dean, 2009).

Pre-Columbian Period

Arrival in Lowland South America

Despite modern evidence of human occupation in the *selva baja* dating back millennia, historical records concerning pre-Columbian Upper Amazonia are limited due to a dearth of written documents and scant archaeological undertakings in the region. The earliest corroborated human occupation in South America dates to approximately 14,600 cal BP, from the Monte Verde II site in southern Chile (Dillehay, 1989, 1997). Archaeological findings suggest that humans reached South America relatively quickly, perhaps one to two thousand years after departing eastern Beringia, where they may have settled for up to 9,000 years (Homburger *et al*., 2015; Lamas *et al*., 2016; Raghavan *et al*., 2015). The earliest physical evidence of human occupation in the Amazon Basin dates to ~9,950 cal BP and is predicated on the discovery of a freshwater snail shell midden (present with charcoal and animal bones), located within present-day Bolivia (southwestern Amazonia) (Lombardo *et al.,* 2013). Petroglyphs have been located adjacent to the Lower Huallaga Valley at locations including Casa de Cumpanamá (near Puerto Libre), Bello Horizonte, and Utcurarca, the earliest of which has been dated to ~2,950 cal BP (Homan, 2014; Orefici, 2013). Ceramic potsherd has been unearthed beside the Cachiyacu River, dating to ~950 cal BP (Rivas, 2003).

Language

The Indigenous peoples of the Amazon Basin speak a variety of languages, primarily derived from four main families: Arawakan, Tupian, Tukanoan, and Kariban (Kaufman, 1990). While these language families are dispersed on a continental level, Arawakan remains prevalent

primarily in the northwestern Amazon Basin (particularly the eastern Andean foothills); Tupian along the coast of present-day Brazil and in south-central Amazonia; Tukanoan in the Ucayali Basin of Upper Amazonia; and Kariban in the Guiana Highlands and parts of northern Colombia (Carmichael *et al*., 1985; Kaufman, 1990; Rodrigues, 1999). There also exist a number of language isolates, which by definition have no distinguishable link to any established language families in the region. Examples include, but are not limited to Jibaroan, Cahuapanan, and Shirixana (Lathrap, 1970). Across Amazonia there are ~90 distinct autochthonous languages and ~300 dialects spoken (Hudson, 1997). An understanding of the dispersal of these languages is integral for framing later analyses whereby language, specifically "autochthonous language" versus "no autochthonous language" is utilized as a potential distinguishing variable for demographic and genetic association tests. Although indigenous languages in the region are highly diverse, such classifications can potentially act as a proxy uniting Indigenous peoples versus non-Indigenous immigrant communities.

Early "Selva Baja" Culture

The earliest settlers of lowland Peruvian Amazonia were likely small, scattered communities that hunted, scavenged, and collected for sustenance, in addition to employing basic systems of root cultivation and slash-and-burn agriculture. There was likely a heavy reliance on lake, river, and marine resources (Lathrap, 1970). They were the descendants of migrants who crossed Beringia, expanding into the Americas ~17,500 to ~14,600 cal BP (Waters, 2019). Early Peruvian Amazonian settlements were established in the alluvial plains adjacent to tributaries and are exemplified by archaeological sites at Tutishcainyo and Shakimu (Lathrap, 1970). Recent evidence suggests that inhabitants quickly progressed from root cultivation and slash-and-burn agriculture to increasingly complex systems of production. These cultures created geometrically-shaped earthworks composed of circles, rectangles, and composite shapes created by excavating ditches in clay-heavy soil and building earthen walls with the excavated material (Pärssinen *et al*., 2009; Ranzi, 2003; Schaan *et al*., 2007). Their purpose has been suggested to

be for defense and/or possibly ceremony (Pärssinen *et al.,* 2009). Among the approximately 200 discovered earthworks, some measure hundreds of meters in length, with most varying widely in their estimated dates of origin. Nevertheless, the vast majority have been attributed to between ~1,050 cal BP and ~650 cal BP (Pärssinen *et al*., 2009). Their intricate, arithmetical designs, and the massive construction necessary for their creation are indeed evidence for the presence of complex societies in this region prior to the arrival of Europeans.

The Chavín, a theocratic and highly stratified, originally Andean society, were well established by ~1,500 BCE, and have been hypothesized to have displaced the earlier cultures of the *selva baja*, the likely engineers of the aforementioned earthworks (Starn, Degregori, & Kirk, 2005b). The Chavín were themselves displaced by cultures including the Hupa-Iya and Yarinacocha, evidenced by paradigmatic shifts in ceramic production. Serial displacement of peoples and widespread inter-group warring has been proposed to be the result of a struggle for access to fertile, alluvial lands (Lathrap, 1970).

From a Colony to an Independent Republic

European Exploration of the Upper Amazon

The Amazon River was first discovered by non-Indigenous peoples in the year 1500 CE by Spanish sea-captain Vicente Yáñez Pinzón, who entered the flooded river, and named it "*La Mar Dulce*" or "The Sweet Sea" (Carmichael *et al*., 1985). In that year, an estimated 5 million Indigenous peoples occupied the Amazon Basin (Park, 1992). Along the major rivers, estuaries, and coastlines of lowland Amazonia is where the majority of populations were settled at the time of contact, most likely the result of fertile lands and the possibility of water transportation. Despite both popular and scholarly renditions depicting Amazonia as uncharted and unspoiled terrain inhabited by simplistic peoples with a limited degree of sophistication, many of the earliest accounts of the region and its peoples support accounts of complex and burgeoning social formations. The first Europeans to this area describe an environment of great social

sophistication (Lathrap, 1970; Carmichael *et al.,* 1985). Evidence of advanced ceramic-making, intricate iconographies, stratified political systems, highly organized transportation systems, workforce specialization, and sociocultural hierarchies lend credence to the sophistication of local inhabitants at the time of European contact (Dean, 2009; Roosevelt, 1999).

By the 1530's, the Spanish, in pursuit of gold, slave labor, and natural resources, began funding explorations into the reaches of Upper Amazonia, in present-day Ecuador and Peru (Stanfield, 1998). Alonso de Mercadillo and Diego Núñez reached the Huallaga and Marañón River basins by years 1538 and 1539, respectively (Jiménez de la Espada, 1897). By 1542, Francisco de Orellana became the first European to travel the length of the Amazon River (Smith, 1994). In 1560 Pedro de Ursua, Governor of El Dorado and Omagua, departing from the Port of Trujillo was able to navigate the waters of the Huallaga, and eventually reach the Marañón River confluence by September of that year, not far from where the city of Yurimaguas would be founded (Morey Alejo & Sotil García, 2000). It was following these early European explorations that a major cultural shift marked by resource exploitation, proselytization, and colonial hegemony began, altering environmental, social, and economic lifestyles practiced since the early Holocene. Societies based primarily on fishing, hunting, foraging, and limited agroforestry became full sedentary agrarians with a focus on domestication (Hornborg & Hill, 2011; Politis, 2009). Shortly after European contact, the Spanish and Portuguese languages began replacing native tongues, becoming second, only to Quechua (specifically the lowland Kichwa variant) as the *lingua franca* of the Lower Huallaga Valley (Dean, 2009).

The Extraction of Native Labor Power, and Mission Life

The Spanish established an extensive system of *encomiendas* (land grants given to Spanish colonists in the Americas conferring the right to require tribute and labor from Indigenous inhabitants*)* whereby Indigenous peoples and their lands were granted to *encomenderos*, or Spanish labor (typically agrarian) profiteers. Indigenous peoples were forced to pay tax to *encomenderos* through a process of labor power extraction, as well as provide for their own

subsistence (Homan, 2014). Concomitantly, *encomenderos* were tasked with Christianizing the Indigenous slave laborers under their control (Trujillo, 1981). Those individuals that were not able to pay sufficient tribute were severely penalized.

In 1626 the Archbishop of Lima, while visiting the city of Huánuco, baptized the chief of the Panatahua people, who controlled a significant portion of the Huallaga Valley (Santos-Granero, 1985). Five years later, Franciscan clergy began converting the entire Panatahua people, as well as the remainder of surrounding Indigenous peoples. In 1636, Jesuit Father Lucas de la Cueva first arrived in the city of San Francisco de Borja, along the Upper Marañón River, and laid the foundation for what would become known as the Maynas Missions (on the Aipena River near present-day Jeberos) (Reeve, 1993).

Indigenous peoples residing in the regions' missions attended daily mass, obliged to undergo religious instruction, and were forced to provide unpaid services related to maintenance, including construction of buildings, cleaning, and food provisioning (*mitayo*). Such individuals were subjugated by monopolizing, and later directly controlling access to farming tools, in addition to blocking access to natural resources, such as salt deposits. Indigenous peoples were not only economically subjugated, but culturally and psychologically manipulated. Groundwork for proselytization began with a series of concessions made by the Roman Catholic Church to the Spanish Crown. In 1508 the Church conceded the Right of Patronage to the Crown. Allegations of grave abuses against Indigenous peoples taking part in the *encomienda* system began to be brought to the forefront of American religious society by Father Bartolomé de las Casas in 1539 (Santos-Granero & Barclay, 1998). In 1542, King Carlos I passed the "*Leyes Nuevas*," or the "New Laws of the Indies for the Good Treatment and Preservation of the Indians," which above all, outlawed the *encomienda* system in Spanish territorial claims in the Americas. In 1573, King Philip II established the "Statute on New Discoveries and Settlements," calling for a peaceful religious conversion of Indigenous populations based on the Gospels rather than violence, coercion, and servitude (Santos-Granero, 1992). Such legislations, however, did not fully curtail the already entrenched culture of abuse. Slave raiding occurred into the late 17th century, carried out mostly by settlers, profiteers, and pseudo-*encomenderos*, bringing further destabilization to

the Huallaga Valley region. By this time, low profit margins in illicit, yet extant Amazonian *encomiendas,* in tandem with uprisings by Jivaroan and Quijos peoples, led to a gradual phasing out of violent labor extractive processes in favor of a system of *repartimiento* (a system of labor whereby Indigenous peoples were compelled to work a stipulated proportion of the year for Spanish-owned businesses, for low, or no wages) whereby the treatment of laborers was more congruent with that of indentured servants than slaves (Homan, 2014).

The Spread of Disease

Native peoples were highly susceptible to diseases that resulted from the "Columbian exchange," with outbreaks occurring across much of the Americas for centuries (Crosby, 1972). Missionaries often forbade Indigenous converts from leaving mission grounds to avoid such contagions (Bianchine & Russo, 1995). Between years 1718 and 1739, the mean size of a nuclear family among five Yanesha missions declined from 5.48 to 3.75. The mean number of surviving children per family unit decreased from 2.25 to 1.19. Perhaps most strikingly, the percentage of youths aged 0–14 years in the total population fell from 52% to 32%, retarding population growth for decades (Santos-Granero, 1987).

Foreign communicable diseases began to decimate Indigenous populations almost immediately after their arrival in the Americas (Reff, 1991). One of the first documented epidemics, a smallpox outbreak, occurred in year 1519 and spanned Mexico to Peru, killing an estimated 200,000 Indigenous Peruvians alone (Dobyns, 1966; Oveido, 1851). During the next 80 years, 17 epidemics were recorded in either the disputed territories (1520–1542) or conquered lands (1542 onward) of the Viceroyalty of Peru. Most notable, perhaps, was a simultaneous epidemic of smallpox and measles that occurred between years 1585 and 1591. In 1586 alone, approximately 20% of the population of Lima died (Crosby, 1972; Dobyns, 1963). In addition to measles and smallpox, influenza, typhus, chickenpox, and diphtheria decimated Indigenous populations across the Viceroyalty (Cook, 1973; Crosby, 1972; Stearn & Stearn, 1945). A horrific diphtheria epidemic hit the Andean city of Cusco in 1614, and numerous measles outbreaks

occurred throughout much of Peru thereafter, particularly in the years 1618–1619, 1628, and 1634–1635. By year 1621, smallpox epidemics were occurring with regular frequency along Amazon River settlements, and by 1642 smallpox epidemics began reaching the Atlantic coast at Recife, Pernambuco, Brazil. Smallpox continued to spread, affecting the capital of *Brasil Colonial*, São Salvador de Bahia de Todos os Santos (present-day Salvador, Bahia), and eventually reaching the burgeoning city of Rio de Janeiro in year 1669 (Bianchine & Russo, 1995). Also in 1669, a measles epidemic struck the Maynas Mission of Loreto, northeast of the Huallaga Valley, causing some 20,000 in the vicinity to perish. Untold amounts of further deaths occurred when the epidemic reemerged in a nearby Indigenous village located along the Ucayali River one year later. By 1681 an epidemic of smallpox that originated in San Francisco de Quito (present-day Quito, Ecuador) one year earlier, reached the Huallaga Valley settlements, likely killing many thousands more (Dobyns, 1963).

Scholars have hypothesized that during the initial stages of the Spanish conquest of Peru (years 1532 to 1628) the Indigenous American population decreased to only 18% of its original total—from approximately six million to 1.1 million inhabitants—representing a historic population bottleneck (Lipschutz, 1966). Data related to the number and/or percentage of Indigenous deaths during the Spanish conquest of Peru are debated due to wide variances in population estimates at time of contact. In 1767 King Carlos III ordered all Spanish Jesuits to vacate “New World” territories. This effectively ended the vast network of missions in Peru, and the Huallaga Valley in particular (Maynas Mission), resulting in many Indigenous communities retreating back to largely inaccessible areas of the *selva baja*. There they largely continued their traditional, dispersed lifestyles, away, from the ever-encroaching edifices of the European world (Golob, 1982).

The Political and Economic Foundation of Loreto Region

The historical factors discussed thus far are predicated on commonalities between the Peruvian Amazon and adjacent regions related to environment, culture, and history of colonization. Although such parallels continued to exist into the 19th century, political and

economic systems specific to Loreto Region began to shape the history of Yurimaguas and the Lower Huallaga Valley much more than broader, continent-wide impacts.

Following the Jesuits' 1767 expulsion from Maynas Province (comprised of present-day Loreto, Ucayali, and San Martín Regions), the area became increasingly isolated politically, economically, and socially from the Spanish administrative centers of Quito and Lima. Under the request of Governor Francisco de Requena, control over Maynas Province was transferred from the *Audiencia de Quito* to the *Audiencia de Lima* (Dean, 2009). Although the vacuum left by the Jesuits was partially filled by scant Franciscan missionaries, they too, were shortly expelled (Santos-Granero & Barclay, 2000).

Peru, South America's last viceroyalty declared independence from Spain on July 28, 1821, although warring continued through December of 1824. Independence was won through military action, largely the consequence of outsiders taking up the reigns of anti-colonialism across the continent. In the case of Peru, this included Argentine José de San Martín and Venezuelan Simón Bolívar. Despite the newly established Republic of Peru, many divisions based on ethnic origins prevailed (Dean, 2019). A minority of European descendants controlled the vast majority of land, power, and prosperity. Between years 1825 and 1841 the federal Peruvian government experienced 24 regime changes. Colonial Peru gained its earliest experience in the international market with the minimal procurement of guano in the mid-18th century. By the mid-1850's, the Peruvian Republic was making approximately 20 million USD per annum across all international economic sectors (Starn, Degregori, & Kirk, 2005a).

In the mid-19th century the Peruvian government began to focus on Loreto Region as a national "frontier" (*frontera*) territory where natural resources could be exploited for profit (Dean, 2019). The government made a concerted effort to ensure that Loreto became solely Peruvian territory, despite claims by Brazil, Ecuador and Colombia. Peru signed a treaty with Brazil in 1851 that settled frontier border disputes and opened the Region to river navigation by both countries. This had the *de facto* effect of weakening the claims of both Colombia and Ecuador to the territory. In 1864, the Peruvian Republic established a naval port at the commercial outpost of Iquitos, which would ultimately become the Regional capital. Subsequent to the advent of steam

powered navigation in the Region in 1853, but prior to fully international navigation rights in 1868, items including cotton, hammocks, quinine, rubber, salted fish, sarsaparilla, turtle eggs, and vanilla bean were exported to a global market, while footwear, liquor, textiles, and tools were imported on Peruvian and Brazilian vessels (Herrera, 1905). During this time, urban centers of the Upper Amazon including Yurimaguas, Iquitos, and Tarapoto expanded into centers of commercial and sometimes political importance. Spurred by a boom in foreign rubber demand, Loreto was further prepared by the Lima government for insertion into the international marketplace. The State purchased riverboats from Great Britain and the United States, built a military garrison on the Amazon River at Leticia (present-day Colombia), explored navigable tributaries in the region, and encouraged foreign trade *en masse* (Santos-Granero & Barclay, 2000).

From the early development of Loreto Region (~1850), until approximately 1900, the Peruvian government experienced extensive problems, from bankruptcy in 1874, to the War of the Pacific in which a Great Britain-backed Chile defeated Peru and her ally, Bolivia in 1881 (Starn *et al*., 2005a). Despite such impediments, State presence in Peruvian Amazonia and Loreto Region in particular continued to burgeon during the latter half of the 19th century (Dean, 2019).

The Rubber Boom

During the 1850's and 1860's the exports of Loreto remained relatively diverse for a society that had only recently begun to shed its colonial era economic relations. Rubber forever changed the economy of Loreto. During the mid-19th century rubber remained a rather meager commodity, with an average total export of less than five metric tons (Mg) per annum (Pennano, 1988). By 1884, however, only two years after the construction of the first customhouse in Iquitos, rubber exportation had increased to 540 Mg per annum, and would continue to grow until the advent of World War I in 1914 (Santos-Granero & Barclay, 2000; Weinstein, 1983).

Logarithmic growth in rubber exportation was fueled by a demand from the United States and Western European nations. In the former half of the 19th century, rubber products were used for the manufacture of underwater telegraph cables, waterproof footwear, and other, mostly limited, specialty items. The skyrocketing demand for rubber in the latter half of the century was largely the byproduct of the popularization of vulcanized rubber, first patented by Charles Goodyear decades earlier, in 1844. Vulcanization chemically enhances the durability of rubber, in addition to ameliorating its natural susceptibility to temperature swings, making it useful in novel industrial and consumer applications (Pennano, 1988). By the late 19th century vulcanized rubber became integral in roofing, paving, cabling, the creation of machine fittings, and most importantly, the manufacture of bicycle (1884) and later, automobile (1895) tires (Dean, 1987; Pennano, 1988). By 1886, rubber accounted for roughly 90% of Loreto's export value, establishing itself as the tentpole product of Loreto's now undiversified export portfolio (Palacios, 1890)

The rubber boom initially stimulated a migratory labor force from as far away as the Andean highlands and Pacific coast. This was largely the result of a poorly developed local labor market. Rubber barons often used coercive recruitment and retention methods as standard procedure with local Indigenous groups (Dean, 1999). In 1908, "*las razas"* or *"*the races*"* of Loreto were listed as 70% white, 20% *mestizo*, 9% Asian, and 1% African American (Coriat G, 1943). Indigenous populations were either not included in the census (which was common practice in parts of South America at the time) or were grouped with other races, most likely *mestizos* or Asians. Historically, admixed Indigenous peoples in Loreto preferred being classified as *mestizo*, due to the mobility conferred socially to people of mixed rather than completely Indigenous descent (Justice *et al*., 2012).

Recruitment and retention methods used by rubber barons included a system of debt-peonage, whereby an interest-accumulating purchase of merchandise by an individual was forgiven through backbreaking work in the rubber industry (Dean, 2004). Local Indigenous communities were often raided, with women and children becoming domestic servants or concubines. Some native laborers were recruited simply under the false guise of a fair and mutual exchange of goods or services. An epicenter of such coercion tactics was the Putumayo River

Basin, where Indigenous peoples were victims of brutal conditions perpetrated by exceptionally cruel rubber barons (Taussig, 1984). It was in this socioeconomic context, where organized labor was scarcer than rubber, that Indigenous workers became themselves commodified, their debts transferred, bought, and sold (Dean, 1999, 2009). Violent action and forced servitude became the basis by which the majority of local Indigenous peoples entered the rubber economy and had their labor power mobilized. By 1914, World War I had commenced, drastically limiting transatlantic navigation. This, in conjunction with an economic downturn in the rubber industry in 1910 and the establishment of colonial rubber plantations in Southeast Asia, retarded Peruvian rubber production and stunted the progress of the national economy (Santos-Granero & Barclay, 2000).

The Modern Era (Mid-20th Century to Present)

Peruvian Amazonia in the last ~70 years can be principally defined within the context of extractive economies. Oil production in the Peruvian Amazon has its roots in the 1920's, when Standard Oil, and subsequently Sacony Vacuum Oil, Shell Oil, and International Petroleum initiated hydrocarbon explorations that would eventually lead to the oil boom of the 1970's (Rumrrill, Davila Herrera, & Garcia, 1986). In 1952, legislation was passed that had the *de facto* effect of increasing oil drilling opportunities in the region. Four years later, oil was indeed discovered in the Ucayali Basin, along the southern edge of Loreto Region. As a consequence, numerous multinational petroleum companies, the plurality of which were headquartered in the United States, began to establish roots in the region. In 1967, a second major oil discovery was made in the nearby Ecuadorian Amazon. As a consequence of resulting competition, the Peruvian government created further legislation aimed at benefitting foreign corporations. This attracted a second major wave of international investors to Loreto Region during the first half of the 1970's. Actual oil production was nonetheless relatively slow until 1971, when Petro-Perú, a state-run oil company, discovered oil reserves in the Corrientes River Basin in northwestern Loreto Region, along the border with Ecuador. Over the following eight years, private petroleum

companies invested 632 million USD, and Petro-Perú provided another 1.1 billion USD for prospecting, production, and refinement (Santos-Granero & Barclay, 2000). With no new significant oil discoveries after 1971, most private petroleum companies withdrew from Peru by 1976. Occidental Petroleum Corporation and Petro-Perú became the majority producers in the Region, stimulating external trade and fluvial transport, increasing labor markets, and ultimately increasing domestic migration to the region. In 1976 the Peruvian government granted Loreto Region a 10% share of annual oil production within its borders, and with production steadily growing in the region until 1979, Loreto's extractive economy prospered (Santos-Granero & Barclay, 2000).

During the 1990's Alberto Fujimori's presidency began to privatize communal Indigenous lands in the *selva baja*. Neoliberal policies that promoted tax abatements and land concessions and turned a "blind eye" to environmental destruction benefitted multinational corporations involved in primary-product extraction practices (Dean, 2002, 2016). At the start of the 1990's the Fujimori government passed legislative initiatives such as the "Law of Organic Hydrocarbons," which promoted a restoration of international petroleum investment in the region. Such policies eventually led to the privatization of Petro-Perú. Much of northern Peruvian Amazonia was consequently carved into "concessionary blocks," in a further effort to promote foreign hydrocarbon exploration and investment. Companies such as Chevron, Mobil, Shell, Exxon, and Occidental eagerly took part in this new lease on Peruvian Amazonia, to the extent of which by 1998, 21 million hectares of the *selva baja* had been apportioned (La Torre López, 1998). State resources were distributed not to communities native to the lands, but were instead allocated in accordance with cronyism and political patronage that fit with long-established systems of clientelism in the region (Dean, 2002).

While the oil boom brought wealth to a limited number of Peruvian industrialists, it also fostered the metastasis of significant environmental and societal issues. By the early 1970's massive land clearing projects had begun in the acid infertile soils surrounding Yurimaguas. Corporations, capitalizing on the region's political vulnerability, caused extensive damage to the ecosystem's long-term sustainability. The clearing of forests with heavy machinery destroyed

topsoil, increased the density of soils to unnatural levels, and ultimately caused an overall reduction in soil health metrics (Alegre, Cassel, Bandy, & Sanchez, 1986). The planting of nonnative, monotype crops decreased faunal and floral health by disrupting the fragile, natural equilibrium of the diverse lowland forest. The presence of the North Carolina Tropical Soils Research Program lasted for a decade, and in many ways became the predecessor of many extensive, industrialized land-clearing projects funded by multinational corporations (Alegre *et al*., 1986). Such projects not only impoverished the natural environment but destroyed the habitat of many of the region's Indigenous peoples, who historically made their homes amongst forested lands. This issue was compounded further by deforestation for the sole purpose of tropical hardwood exportation. Federal regulations were designed towards the end of the 20th century with extractive industry in mind, to develop the *selva baja* economically. Such regulations often came at the exclusion of local and in particular, Indigenous peoples, exhibiting a blatant disregard for communal welfare and the depletion of forest resources. A highly profitable suite of foreign, primary-product, extractive corporations continues to operate and lobby for policy in Loreto as of the 21st century (Dean, 2004).

The history of the Peruvian Amazon for the last ~70 years plays a major role in the current causes of migration in the region. While broader global forces do cause people to emigrate to and from the region, it is local economic and environmental factors that are the major contributors, informing specifically who and how individuals migrate. An understanding of these factors can help in framing the demographic and genetic consequences of such migratory processes.

Loreto Region and the City of Yurimaguas, Today

Loreto is one of 25 Regions that comprise modern Peru (in addition to the Province of Lima, that exists independent of a Region). Within Loreto are eight provinces—Alto Amazonas, Datem del Marañón, Loreto, Mariscal Ramón Castilla, Maynas, Requena, Ucayali, and Putumayo, which are themselves divided into a total of 53 districts. Loreto is by far the largest

Region of Peru, encompassing 368,851.95 Km^2, or almost 30% of Peru's total land area (Gore Loreto, 2017). It is also the most northerly Region of Peru, sharing a border with Ecuador, Colombia, Brazil, and the Peruvian Regions of San Martín, Amazonas, Huánuco, and Ucayali. The capital, and most populous city of Loreto is Iquitos, followed next by Yurimaguas. Overall, however, Loreto is one of the most sparsely populated Regions of Peru. This is primarily due to its heavy forestation, numerous lagoons and flood plains, lack of employment and education opportunities, and the virtual absence of any system connecting it to metropolises in Andean and coastal Peru.

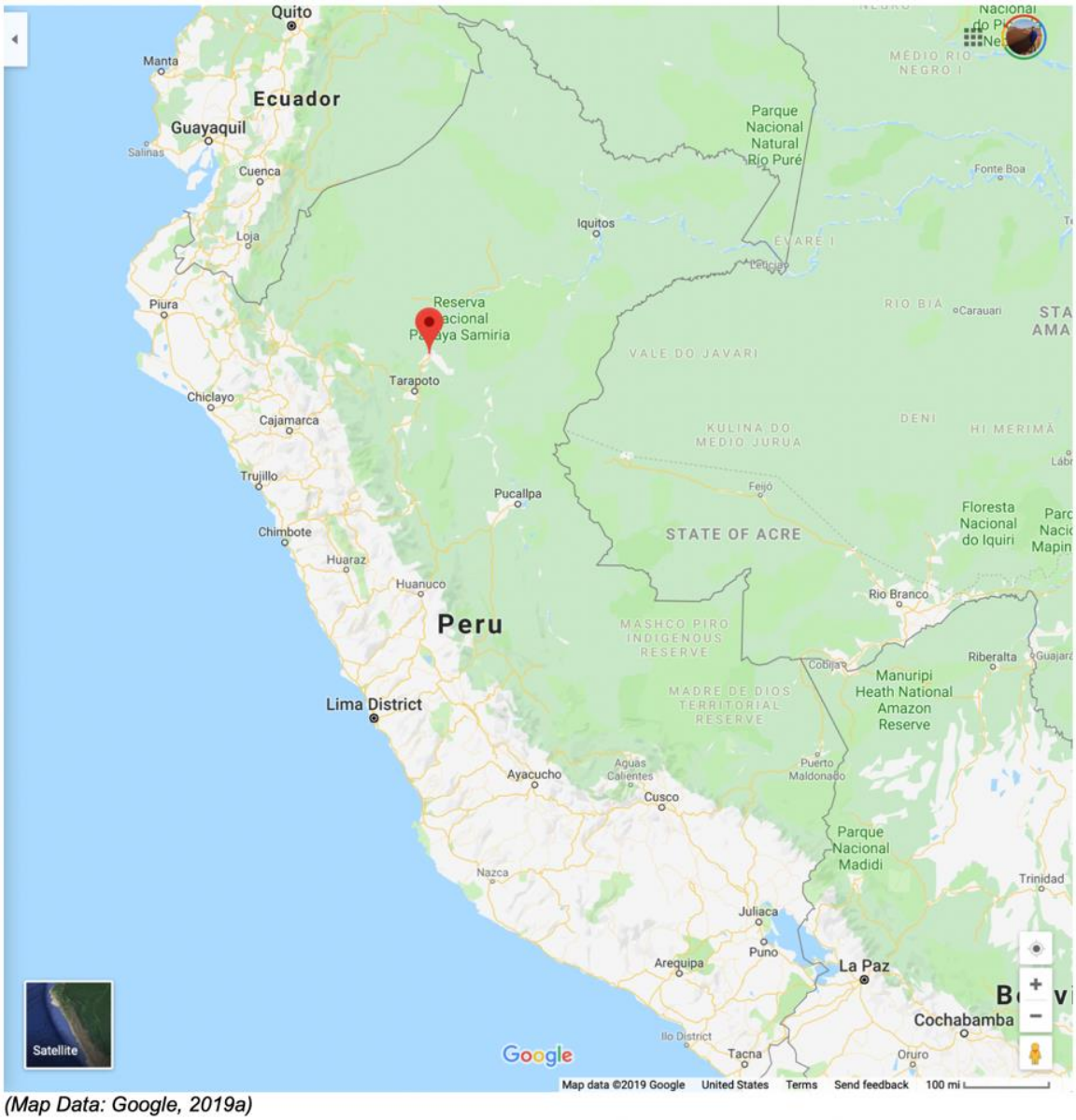
The most recent estimate of Loreto's population is 1,006,953, with a Region-wide population density of 2.7 persons per Km^2 (2012) (Gore Loreto, 2017). Between the 1993 and 2007 federal censuses the population of Loreto grew by 1.8%, making it the seventh fastest growing Region in Peru during that period (Instituto Nacional de Estadística e Informática [INEI], 2007). Loreto's population density increased 6% between the 1940 and 2007 federal censuses (INEI, 2007). Loreto produces 2.51% of Peru's GDP, has a poverty rate of approximately 63% (2011) and has a life expectancy at birth of 72.7 years (2016) (Gore Loreto, 2017). The major trans-Andean highway—*IIRSA Norte*—connects Yurimaguas with its two western terminal cities, Paita and Chiclayo, effectively permitting ground transportation from Yurimaguas to the rest of the country. Yurimaguas is located ~30 Km from the southwestern Loreto border with San Martín and is not connected by roadway to the rest of Loreto. This reality has caused Yurimaguas to share, economically and socially, more in common with San Martín Region than Loreto Region. Because there are no roads connected to the national highway system east of Yurimaguas, most transport in the region occurs via riverboat, or *lancha,* along the Huallaga and its navigable tributaries. Yurimaguas, the second most populous city in Loreto Region had an estimated population of 63,427 for 2015 (Instituto Nacional de Estadística e Informática [INEI], 2012). According to the 2007, *XI Censo de Población y VI de Vivienda*n (Eleventh Population and Sixth Household Census), 92.51% of Loreto inhabitants speak Spanish (Castilian) as their natal tongue, followed by Quechua, at 0.72%, Asháninka (Arawakan language family), at 0.03%, and Aymara (similar to Quechua), at 0.02%. Those who speak any autochthonous language as their natal tongue

comprise 7.34% of the total Regional population (0.04% speak a foreign language [not including Spanish] as their natal tongue and 0.11% are deaf/mute) (INEI, 2007). It is worth noting that spoken Aymara is not aboriginal to Loreto, but rather the south of Peru, as well as highland Bolivia and Chile, indicative of historical migrations from the central Andes to Upper Amazonia.

Yurimaguas is located at the confluence of the Huallaga and Paranapura Rivers, in the Lower Huallaga River Valley, Alto Amazonas, Loreto, Peru (GPS coordinates: 5.9008° S, 76.1129° W). After the widespread destruction by Spanish conquistadors of Indigenous chieftaincies in the Upper Amazon during the mid-16th century, Jesuit missionaries, with support from the Spanish Crown, began establishing religious outposts, or *reducciones*. These backwater colonial communities began to fill the power vacuum left behind by the ancient chieftaincies in the riverine environments of the region. Jéberos and Lagunas were among the first urban centers, founded along the Huallaga River in 1640 and 1670, respectively (Santos-Granero & Barclay, 1998). Named after two local Indigenous peoples, the Yuris and the Omaguas, the mission outpost of “Yurimaguas” was officially founded by Jesuits in 1709 (Rhoades & Bidegaray, 1987). Early migrants to Yurimaguas were often Indigenous refugees seeking protection from the abuses of colonists and *encomenderos* in neighboring areas. Following the Spanish Crown’s 1767 decree calling for all Jesuit missionaries to vacate American territorial claims, many of the outposts they had established during the prior century came under the control of Franciscan or secular clergymen. During this period of late colonial rule, Catholic missions fell into neglect and disrepair. Indigenous peoples that were unable to avoid foreign occupiers, including soldiers, *hacendados* (*hacienda* [plantation or factory] owners), traders, and priests, were frequently enslaved or incorporated into the local peasantry. Native lands were converted into *haciendas*, Indigenous cultures were threatened, and distinctive lifeways were systematically dismantled or destroyed. This process continued until the arrival of various forms of extractive capitalism reached the *selva baja* from the Andes and Pacific coast in the early to mid-19th century (Dean, 2009, 2019). Along with rubber, the territory surrounding Yurimaguas produced a number of agroextractive products, including timber, hides, cascarilla, chicle, barbasco, and cotton (David & Dean, 2017).

By 1864, the population of Yurimaguas was ~2,600 residents (Her Majesty's Consuls, 1864). Disease, extractive economies, debt-peonage, religious proselytization, and steam navigation proved highly disruptive to both the livelihood and settlement patterns of local Indigenous peoples in the Lower Huallaga Valley. Despite this, some Indigenous populations have been able to maintain a distinctive cultural identity into the present-day, due in part to the lack of a strong State presence in this region of Peruvian Amazonia. The proliferation of road networks in the adjacent *selva alta* due west of Yurimaguas began in the 1930's and continued through the 1960's. Eventually, thoroughfares, albeit rudimentary, connected Yurimaguas to neighboring population centers, such as Tarapoto (San Martín Region), and Moyobamba (San Martín Region). Built in the 1960's, the aforementioned *Carretera* (highway) *IIRSA Norte* connected the Lower Huallaga Valley with Andean and coastal Peru (including Lima) reaching the city of Yurimaguas in 1967. Automobiles, trucks and motorcycles became the main modes of transportation, largely supplanting many of the regions' ancient riverways, portage networks, and trails.

Figure 2.1. Location of Yurimaguas within Peru and wider Amazonia.



(Map Data: Google, 2019a)

Figure 2. Satellite image of the greater Yurimaguas area, at the confluence of the Huallaga and Paranapura Rivers.

(Map Data: Google, 2019b)

In 1974, the military-led government of President Juan Velasco passed legislature entitled "*Ley de Comunidades Nativas y de Promoción Agropecuaria de Regiones de Selva y Ceja de Selva"* (Law of Native Communities and Agricultural Promotion of Rainforest and Rainforest-Vicinity Regions). This law permitted the communal titling of ancestral lands to Indigenous groups that agreed to form permanent settlements, while concurrently creating a platform for further extractive processes in the region (García, 1995). Four years later, the law was amended by President General Morales-Bermúdez whose military junta passed legislature known as "*Ley de Comunidades Nativas y de Desarrollo Agrario de la Selva y Ceja de Selva"* (Law of Native Communities and Agrarian Development of the Rainforest and Rainforest-Vicinity). The amended law set restrictions on Indigenous communities while guaranteeing communal rights to Indigenous lands used in an agrarian fashion. The government's enforcement of such laws, with assistance from the evangelical, "Summer Institute of Linguistics" (now known as SIL International), and their partner organization, Wycliffe Bible Translators (now known as Wycliffe

Global Alliance), accelerated the migration of Indigenous peoples from dispersed communities in the *selva baja*, to urban centers, including Yurimaguas (Epps, 2005; Homan, 2014). Intranational migration from the *selva alta*, Andean highlands, and Pacific coast brought émigrés via the *Carretera IIRSA Norte*, which in turn led to further market expansion and economic growth, leading to a cycle of migration to the city that continues to the present-day (Justice *et al*., 2012).

Justice *et al*. described Yurimaguas as a patchwork of ethnic diversity, comprised of regional native communities, Indigenous communities from the *sierra* and *costa*, and foreign migrants (2012). Represented Indigenous communities in the greater area include Quechua-speaking populations (Kichwa Lamista, Kichwa del Pastaza); Jivaroan speakers (Achuar [Shiwiar, Jivaro], Aguaruna [Awajún], Kandozi, and Wampis); Tupi-Guaraní speakers (Kukama-Kukamira [Cocama-Cocamilla]); Cahuapanan speakers (Chayahuita [Shawi] and Jebero); Arawakan speakers (Chamicuro); and the Urarina (Kachá) (speak a linguistic isolate) (Justice *et al*., 2012). Following the rubber boom of the late 19th to early 20th centuries, Ashkenazi Jews, Portuguese, North Americans (Canada and the United States), Spaniards, Italians, British, Chinese, Japanese, and others migrated to the Lower Huallaga Valley in search of economic opportunity (David & Dean, in press). Many current residents of Yurimaguas are the descendants of these immigrants.

Migration to many mid-sized Peruvian cities began to accelerate, and outpace the growth of larger agglomerations, such as Lima, between the 1970's and 1990's (Dufour & Piperata, 2004). In the case of Yurimaguas, migration continues to burgeon, as this once isolated backwater expands its commercial influence in the region, with trade links to the largest Peruvian cities via the *Carretera IIRSA Norte*, and to the rest of the *selva baja*, including Iquitos, not to mention Brazil and Colombia, via Huallaga—Marañón—Amazon River transportation. The recent building of a new, 110 m long, 40-ton capacity suspension bridge now permits the passage of transoceanic ships to the equally new international port, *Puerto Yurimaguas—Nueva Reforma* (Desarrollo Peruano, 2014). Infrastructural improvements such as these have made Yurimaguas a major transshipment point for goods crossing from eastern Amazonia to Oceania, Asia and Australia.

The *Instituto Nacional de Estadística e Informática* (INEI) reported in 2007 that 26.5% of the population of Yurimaguas District migrated to the area in only the two years prior. Furthermore, INEI reported that urban centers in the size class of Yurimaguas went from comprising 19.8% of the Peruvian populace in 1993, to 22.2% in 2007. With the largely unorganized, extraordinary growth of Yurimaguas has come mass transient populations, and the formation of associated *barriadas*, or makeshift settlements that peripherally ring the city (Dean *et al*., 2011) The majority of these communities are comprised of peoples displaced by neoliberal policies and the consequences of extractive industries related to palm oil, logging, coca cultivation, and mass monocrop agriculture (Dean *et al*., 2011; Justice *et al*., 2012). In 1981, the population of Yurimaguas District was 36,417, 63% of which resided in an urban environment (Rhoades & Bidegaray, 1987). By 2007, the population had nearly doubled, to 63,345, 77% of which lived in an urban environment (INEI, 2007). Yurimaguas District includes the peri-urban community of Munichis, one of two sampling locations of this dissertation. Based on economic shifts, on-going social conflicts, and political turmoil, the demographic characteristics of Yurimaguas are bound to continue to evolve over time.

Chapter III: Scientific Literature Review

In recent years concerted efforts have been made to include Peruvian populations in genomic research. One example is the Peruvian Genome Project, in which 280 genomes (of principally Indigenous ancestry) from a wide range of biomes were analyzed (Guio *et al*., 2018; Harris *et al*., 2018). Findings suggest that the first peoples to reach Peru did so approximately 12,000 cal BP (Harris *et al*., 2018). This generally fits the hypothesis that the ancestors of modern Native Americans diverged from Asian populations approximately 23,000 cal BP, and then paused, likely in eastern Beringia for an extended period (~2 to ~9 thousand years) before populating the Americas (Hoffecker *et al.*, 2014; Lamas *et al*., 2016; Raghavan *et al*., 2015; Tamm, *et al.*, 2007). The first peoples would have moved across the American landscape rapidly, reaching the Monte Verde site in southern Chile ~14,500 cal BP, or possibly even as early as ~18,500 cal BP (Dillehay, 1997; Dillehay *et al*., 2008, 2015; Lamas *et al*., 2016). One would, however, expect the earliest migrants to have reached Peru before ~14,000 cal BP if travelling along the Atlantic coast (Bodner *et al*., 2012; Fagundes *et al*., 2008; Wang *et al*., 2007). The difficulty in establishing an accurate timeline is largely based on the lack of evidence to suggest how South America was populated after people traversed the Isthmus of Panama.

A significant consensus supports the theory that a population in Panama supplied the first peoples that would continue across the isthmus to populate South America (Gravel *et al*., 2013; Lewis *et al*., 2007; Rothhammer & Dillehay, 2009). It has been suggested by Skoglund *et al*. (2015) and Luiselli, Simoni, Tarazona-Santos, Pastor, and Pettener (2000) that upon reaching the Andean Cordillera, two separate migration paths were established. This division would have effectively created a boundary between the two subpopulations whose ancestors both settled the Isthmus of Panama. A mating barrier may be the reason for clear differences in population history and genetic characteristics between Andean and Amazonian populations (Harris *et al*., 2018; Reich *et al*., 2012; Rodriquez-Delfin, Rubin-de-Celis, & Zago, 2001). Gnecchi-Ruscone *et al*. (2019) provides genomic evidence in support of this finding. The authors found that all non-Andean South American Indigenous populations derive from a single population. They moreover

describe high levels of “intra-group homogeneity” and “substantial…genetic drift” in Peruvian and other South American Indigenous populations, suggestive of relative isolation (p. 1254, 2019).

There is, despite differences between Andean and Amazonian populations, evidence to suggest that pre-Columbian admixture across the Andean Cordillera has occurred, particularly in the intermontane, and seemingly in an easterly direction (Barbieri *et al*., 2014; Goméz-Carballa *et al*., 2018; Guevara, Palo, Guillén, & Sajantila, 2016; Sandoval *et al*., 2013). Evidence of pre-Columbian gene flow between South American populations in some cases may have been a product of Incan policies, such as *mitmaq*, or “enforced people transfer” (Harris *et al*., 2018; Schjellerup, 2005). Some scholars have supported a tripartite pattern of migration—the Atlantic coast, Andean Cordillera, and lowland Amazonia (Rothhammer & Dillehay, 2009). A west to east gradient in decreasing genetic diversity has been noted in both mitochondrial and Y-chromosomal DNA (Fuselli *et al*., 2003; Lewis *et al*., 2004, Tarazona-Santos *et al*., 2001).

Harris *et al*. (2018) found that the vast majority of European admixture in present-day Peruvians occurred ~300 years after the arrival of the Spanish, during the period of Peruvian Independence (post-1824), with a limited number of Indigenous populations. Prior to this (during the Viceroyalty of Peru), there was evidence of only very limited European admixture. The authors also noted that prior to the Incan Empire (~1116 CE) all Andean populations clustered closely together according to identity-by-descent analysis. By the end of the Incan Empire (~1438 CE), four main Native American population clusters were discernable, with the Chopccas population at the center. By Peruvian independence, the aforementioned Indigenous population clusters were still present, though the Chopccas were no longer at the center. The authors interpret this change as being associated with a shift in power centers away from the Andes, to the coast (2018).

Further research in Peru analyzing genome-wide SNP data has recently been conducted by Homburger *et al*. (2015). Peruvian individuals were found to have a greater percentage of Indigenous ancestry compared to Argentinians, Chileans, Colombians, and Ecuadoreans ($P < 0.001$). Homburger *et al*. (2015) additionally found a significant number of Peruvians that exhibited greater than 25% East Asian ancestry. The authors attribute this result to admixture

with Chinese and Japanese migrants that arrived in Peru during the 19th and 20th centuries. This finding is similar to those made by Sandoval *et al*. (2013), who found significant levels of recent East Asian ancestry in the intermontane city of Lamas. This dissertation reported one individual of mitochondrial haplotype B4a1c3 (most frequently found in Japan), and one individual of mitochondrial haplotype E1a1a (most frequently found in maritime Southeast Asia) (Bilal *et al*., 2008; Soares *et al*., 2008; Tabbada *et al*., 2010; Takenaka, 2004). Homburger *et al*. (2015) reported that native South American haplotypes cluster into two groups, those from the central Andes (Incan/Quechua and Aymara) and all others. Peruvians in general tend to cluster with the central Andean group, although Homburger *et al*.'s (2015) sampling methodology within Peru may have affected this outcome.

A limited number of ancient DNA (aDNA) studies have been conducted in Peru, albeit all outside of Amazonia. Fehren-Schmitz *et al*. (2010) analyzed mitochondrial DNA in 218 pre-Columbian individuals from southwestern Andean and coastal Peru. The authors found the sample to have the following mitochondrial haplogroup distribution: 2.9% A, 11.5% B, 26.9% C, and 58.7% D. Shinoda, Adachi, Guillen, and Shimada (2006) analyzed the mitochondrial DNA of 35 pre-Columbian individuals from the Machu Picchu area (Cusco Region). They found that among the study sample haplotype frequencies are very similar to those found in modern Quechuan and Aymaran populations from the Peruvian Andes. The reported haplogroup distribution was as follows: 8.6% A, 65.7% B, 22.9% C, and 2.9% D (2006). Kemp, Tung, and Summar (2009) analyzed the mitochondrial DNA of 74 samples from the Wari people (600-1,000 CE) of the central Andes and compared the resultant haplotypes to people who lived in the region shortly after the collapse of the Wari Empire. The authors found no evidence of a statistically significant difference in haplotype distribution, concluding that there was no exodus or other major migratory event in the area immediately following the collapse. The Wari-period samples exhibited the following mitochondrial haplogroup distribution: 29% A, 50% B, 14% C, and 7% D. The post-Wari samples exhibited the following mitochondrial haplogroup distribution: 17% A, 22% B, 56% C, and 6% D (2009). Posth *et al*. (2018) conducted a genome-wide analysis of ancient DNA from 49 Indigenous Central and South Americans. The authors provided evidence for two

previously unknown genetic exchanges between ancient North and South America, noting a distinct connection between the genomic characteristics of North American Clovis-culture and ancient South Americans (2018). They found that ancient Peruvians that populated the central Andes after ~4,200 cal BP show evidence of allele sharing with ancient individuals from the California Channel Islands in North America (2018). Ancient DNA studies reveal that prior to the Columbian exchange, all reported haplogroups in Peru were indeed either A, B, C, or D. There does appear to be a fair amount of variability in the overall distribution of these haplogroups across the South American biomes.

Conducting Research in the Lower Huallaga River Valley

Limited genetic or biodemographic research has been conducted among the peoples of the Lower Huallaga Valley (Justice *et al.*, 2012). The lack of scholarship concerning this region and the Peruvian *selva baja* at large seems particularly at odds with the widespread attention given to the peopling of coastal, highland, and even intermontane Peruvian populations (Cabana *et al.,* 2014; Lindo *et al*. 2018; Little *et al*. 2013; Messina *et al*., 2008; Rademaker *et al*., 2014). Potential reasons for the paucity of research in this region include concerns for fieldworker safety and study participant accessibility. The Huallaga Valley itself has been the site of widespread narcoterrorism, civil unrest, and violent government intervention, not to mention the presence of extensive bureaucratic “red tape” (Dean, 2013). Furthermore, those involved in local extractive industries are often quick to defend their trade by violent means (Dean, 2004; Taussig, 1984). The region is a fertile ground for Indigenous rights movements that have resulted in aggressive government action. Such was the case in Bagua (Amazonas Region) when in 2009 native peoples organized in opposition to recent laws passed by the Alan García government that allowed foreign corporations access to Amazonian lands for extractive purposes. The native peoples, largely led by *Asociación Interétnica de Desarrollo de la Selva* (*AIDESEP*), or “Rainforest Development Inter-Ethnic Association,” felt that the legislation was in violation of prior Indigenous rights agreements. A period of civil disobedience ensued that eventually erupted in

violence at “Devil’s Curve,” leaving over thirty, including both Indigenous protesters and police, dead after one day (Romero, 2009a). Violence continued the next day at a Petro-Perú facility when a disputed number of more individuals were killed (Romero, 2009b). Besides concerns of violence, transportation options into the interior of the region are limited, with many locals residing in remote areas that require extensive planning, resources, and local knowledge to access. Despite the aforementioned obstacles, a slow but steadily growing body of genetic research has been developed in the Peruvian *selva baja*. Recent research in the region has been diverse, encompassing various populations, methodologies, and genetic markers.

While they often recognize their Indigenous ancestry, many individuals residing in the Peruvian *selva baja* self-identify as *mestizo*. This is not surprising given the social mobility conferred to *mestizo* rather than Indigenous peoples, and the fact that the tropical highland areas of San Martín Region that form the western border of the Huallaga Valley were one of the first regions of Amazonia to be colonized by the Spanish. Moyobamba (which is located in this Region) was founded a mere decade after Francisco Pizzaro’s arrival to Peru in year 1532. The majority of population genetic studies conducted in the Lower Huallaga Valley vicinity have concerned the peopling of the intermontane transition zone that forms the valley’s western border. These studies provide a generally robust description of the genetic profile of local transition zone populations, but generally lack comparative genetic data on populations residing in the Lower Huallaga Valley proper.

Genetic Consequences of Migration and Urbanization in Peruvian Amazonia

This section provides an overview of the biodemographic and population genetic research that has thus far been conducted in the Lower Huallaga Valley and adjacent regions of the Peruvian Amazon. Particular focus is given to studies with substantial migration and/or urbanization components. The wider comparative importance of these studies will be examined in the Discussion chapter.

Bisso-Machado, Bortolini, and Salzano (2012) assessed mitochondrial haplogroup distribution across Amazonia (55 populations, *n* = 2,410) and reported 97% Indigenous ancestry. The authors noted the extensive variability in haplogroup distribution across Amazonia. One such population that has been the focus of a number of genetic research studies in the region is the intermontane Lamas (or Kichwa Lamista) peoples. The Lamas people live in San Martín Region, outside of the city of Tarapoto, in the piedmont that forms the western border of the Lower Huallaga Valley. In a study by Moscoso *et al*. (2006) the human leukocyte antigen (HLA) class I (A and B) and II (DRB1 and DQB1) genetic profiles of 83 Lamas individuals were sequenced. The authors discovered a high frequency of alleles HLA-DRB1*0901 (commonly found at high frequencies among southern Asian populations) and HLA-B*48 (commonly found among Na-Dene, Siberian, and Inuit populations) (Arnaiz-Villena *et al*., 2000; Gomez-Casado *et al.,* 2003). Although they are generally rare among Indigenous peoples, population genetics has demonstrated that both of these alleles are found in significant frequencies among Aymara (Arnaiz-Villena *et al.,* 2005) and Quechua populations (Martinez-Laso *et al.,* 2005). Such an association indicates a close affinity between the Lamas sample and Peruvian highland populations, particularly those from southern Peru, adjacent to the Bolivian border. Despite such findings, genetic distance-based neighbor-joining dendrograms point to the Lamas population being most closely related to populations with their origins in lowland Amazonia, such as the Wayuu (Colombia), Cayapa (Ecuador), and Guaraní (Brazil). As a result, it can be argued that the Lamas population shares a genetic history with both Andean and lowland Amazonian populations. Based on relatively low HLA class II heterozygosity measures, the Lamas population was determined not to have experienced extensive admixture with outside populations (Moscoso *et al.*, 2006).

Sandoval *et al.* (2013) studied the Lamas population by analyzing biallelic insertion-deletion polymorphisms (INDELS). They describe an “intermediate level” (p.627) of non-Indigenous, primarily European ancestry. Interestingly, they also note significant levels of recent East Asian admixture, proposed to be due to the arrival of Chinese (1849) and Japanese (1899) migrant workers to lowland and intermontane Peruvian Amazonia (Lausent-Herrera, 2011;

Takenaka, 2004). Over 87,000 Chinese, primarily of Cantonese (contemporary Guangdong Province) heritage, entered Peru between the years 1859 and 1874 (Klarén, 1986; Sánchez-Albornoz, 1974).

The dissonance in non-Indigenous admixture proportions reported between the Moscoso *et al*. (2006) and Sandoval *et al*. (2013) studies may be due to discrepant sampling methodology, or perhaps genuine changes in the population that occurred over the 7-year time span between the two studies being published. When compared with the Chachapoyas, a geographically adjacent intermontane transition zone population, both populations exhibited similar and significant levels of European ancestry, as well as a shared Indigenous ancestry based on insertion-deletion events. Although Sandoval *et al.* determined the Lamas population to be related to numerous lowland Amazonian populations, such as the Surui (Brazil) and Piapoco (Colombia), they were found to be more closely related to neighboring intermontane populations (2013).

In 2016, Guevara *et al*. analyzed mitochondrial and Y-chromosomal genetic markers in 382 individuals from the Chachapoyas population, in addition to three neighboring populations—the Huancas, Jivaro, and a Cajamarca-based population. Based on mitochondrial markers, the Chachapoyas population exhibited 89% Indigenous haplotypes. This was, however, a relatively low level of indigeneity when compared with the three neighboring populations analyzed. Subclades B2 and C1 were the most common subclades, respectively, except in the case of the Huancas, in which the most common subclades were A2 and B2, suggestive of pre-Columbian gene flow. Guevara *et al.* found haplotype diversity *(H)* in maternal markers to be consistent, and relatively high among all four populations, with $H = 0.9286$ in the Huancas, $H = 0.9343$ in the Jivaro, $H = 0.9715$ in Cajamarca, and $H = 0.9671$ in the Chachapoyas (see *Table 5.14*) (2016). The Chachapoyas population exhibited small, but significant genetic distance measures (pairwise genetic differences) with the Huancas and Jivaro ($\Phi_{ST} = 0.0450$ and 0.0446, respectively; $P \leq 0.010$), but no significant difference with the Cajamarca population ($\Phi_{ST} = 0.0042$; $P = 0.238$).

Guevara *et al.* also found that Indigenous haplotypes were less common in paternal versus maternal genetic markers across all four populations (2016). The frequency of the Indigenous paternal haplotype Q ranged widely, from a frequency of 48% in Cajamarca, 60% in

the Chachapoyas, 92% in the Huancas, and 96% in the Jivaro. Each population therefore indicates evidence of non-Indigenous paternal gene flow. All four Guevara *et al.* populations also exhibited high paternal haplotype diversity measures, $H = 0.9170$ in the Jivaro, $H = 0.9333$ in the Huancas, $H = 0.9974$ in the Chachapoyas, and $H = 1.0000$ in the Cajamarca population (2016). This trend supports the previously-established pattern of haplotype diversity decreasing with distance from the Atlantic coast in South America (Fuselli *et al*., 2003; Lewis *et al*., 2004, Tarazona-Santos *et al*., 2001). In a similar, but more pronounced trend than in the maternal data, the Chachapoyas population exhibited relatively greater genetic distance measures from both the Cajamarca population ($\Phi_{ST} = 0.0706$; $P = 0.0052$) and the Jivaro ($\Phi_{ST} = 0.0996$; $P < 0.0001$), but no significant genetic distance from the Huancas ($\Phi_{ST} = 0.0181$; $P = 0.1946$). Guevara *et al.* conclude that the Chachapoyas population does not clearly cluster with Andean or lowland Amazonian populations (2016). Meanwhile, Sandoval *et al.* provides evidence for "intermediate levels" (p. 627) of non-Indigenous ancestry among the Chachapoyas, both European and East Asian, similar to what was reported among the Lamas peoples (2013).

In 2014 Barbieri *et al.* studied the Arawak-speaking Yanesha that reside in the intermontane Pasco and Junín Regions, towards the southern end of the Huallaga Valley. Their work reveals a history of Indigenous, populational sex-skewed gene flow defined by a greater proportion of Andean male versus Andean female ancestry. The authors attribute this trend to the historic Incan expansions in the region prior to the Spanish conquest (2014). European-based, populational sex-skewed gene flow was also discovered. A disproportionate frequency of European lineages (particularly Italian [central, northeastern, and Sicilian] and Austrian [Tyrolian]) were discovered in paternal versus maternal genetic markers. The authors propose that Italian gene flow was the result of prodigious migration into the area that occurred following the War of the Pacific, and that Austrian gene flow was the result of the establishment of eastern European immigrant colonies in the nearby Pozuzo Valley during the late 19th century (2014). All eleven Yanesha villages sampled were described as "genetically homogenous" (p. 606) across both paternal and maternal markers, with negligible differences in geographically-differentiated AMOVA measures (Barbieri *et al*., 2014). After dividing the eleven villages into high-altitude

(1,200 m – 1,800 m) and intermediate-altitude (~300 m) groups, the authors were able to detect differentially significant haplogroup profiles (2014). There was a greater proportion of Indigenous haplotypes found amongst both maternal and paternal lineages in the Yanesha than in the Chachapoyas (the intermontane population found to the north in Amazonas Region). The proportion of Indigenous haplotypes was greater amongst maternal lineages. This is indicative of non-Indigenous, populational sex-skewed gene flow (this is in addition to the Quechua-speaking, Andean, populational sex-skewed gene flow already determined to be present in the population). The 30% non-Indigenous paternal haplogroups observed were Eurasian and African in origin, likely reflecting Spanish colonization and the subsequent national incorporation of the region.

Both Scliar *et al.* (2014) and Kersulyte *et al.* (2010) conducted research among the Shimaa peoples, a population due south of the Huallaga Valley that also reside within the Andes-Amazon intermontane (Cusco Region). Whereas Scliar *et al.* based their analysis on ten genomic regions, Kersulyte *et al.* conducted research into the strains of the gastric pathogen, *Helicobacter pylori*. Scliar *et al.* analyzed 11 Quechua individuals as a reference population (2014). Based on Bayesian models, they concluded that the Shimaa peoples were most likely descended from central-Andean, Quechua highlanders, branching off and relocating to the intermontane ~5,300 cal BP. All of the single nucleotide polymorphisms (SNP's) found among the Shimaa were also found in the 11 Quechua reference individuals. The Kersulyte *et al.* study focuses on the molecular effects of transcontinental migration (2010). The study determined that the *Helicobacter pylori* strains found in the Shimaa sample were the result of early Asian migrations into South America. The genetic sequences found among Shimaa *Helicobacter pylori* strains, from a phylogenetic perspective, were related to, but not "intermingled" (Results section, para. 1) with modern East Asian samples. When the Shimaa *Helicobacter pylori* strains were compared with those found in an "Indigenous" Lima-based population, a much greater extent of European *Helicobacter pylori* ancestry was found, suggestive of a displacement of the ancestral Asian-descended *Helicobacter pylori* strains with presumably more fit, European or European-hybrid (due to recombination) type strains. The apparent lack of competitiveness in Indigenous strains has been suggested to be due to a loss in vigor associated with the effects of genetic drift

("deleterious allele fixation") during early Indigenous founder events, or a relative inability of Indigenous *Helicobacter pylori* strains to acquire foreign DNA during infection events (causing them to be less capable of responding to variable host conditions). Despite interesting conclusions, the study of pathogen diversity as related to host demographic history is tenuous. If credible, the findings of Kersulyte *et al.* (2010) speak to the future of Indigenous Amazonian public health concerns. Foreign immigration ultimately leads to host admixture and in some cases the vertical transmission of bacteria, such as *Helicobacter pylori*. Such biological implications of migration are only likely to become more pronounced in the region, due to ever-increasing population movements and urbanization.

In a 2017 study by Di Corcia *et al.*, four Peruvian Amazonian populations were analyzed based on mitochondrial and Y-chromosomal genetic markers and compared with Indigenous populations from across South America. The Asháninka, Cashibo and Shipibo reside in the Ucayali Basin, outside of Pucallpa, along the southern border of the Huallaga Valley. The Huambisa live along the northern border of the Huallaga Valley, in the Marañón River Basin. Based on mitochondrial (control region and one INDEL) evidence, all 162 sampled individuals from these four populations exhibited either Indigenous haplogroup A, B, C, or D. Haplogroup C was the most frequent amongst the Cashibo (70%) and Shipibo (45%); haplogroup D was the most frequent among the Asháninka (35%); among the Huambisa, haplogroup B was the most frequent (63%). The Huambisa population has its closest genetic affinity with central-Andean populations and the intermontane Yanesha (Di Corcia *et al*., 2017). Genetic diversity measures indicate high mitochondrial heterogeneity among the Asháninka, Shipibo, and Huambisa. A significantly lower heterogeneity was reported in the Cashibo, who have been suggested to have lived in isolation for a longer period of time.

Di Corcia *et al*. employed genetic distance measures (pairwise genetic differences), to construct an nMDS plot (multi-dimensional scaling map of "*n*" dimensions) for the HVR1 section of the mitochondrial control region, in a dataset of 50 Indigenous populations (2017). Based on this plot, the Huambisa clustered with Andean populations, the Asháninka and Shipibo clustered with Amazonian populations (from lowland Brazil, Bolivia, and Peru), and the Cashibo remained

very isolated, in the middle of the plot, nearest to the Xavante people of Mato Grosso State, Brazil. The most common Y-chromosome haplogroup among the four studied populations was Q (100%), and in particular the Indigenous Q1a3a subclade. According to Y-chromosomal DNA diversity indices, the Cashibo are the least heterogenous population, as they were also according to mitochondrial DNA diversity measures. An nMDS plot based on R_{ST} (a pairwise genetic distance) measurements between Y-STR's (Y-chromosomal short tandem repeats) suggests a genetic stratification pattern across South America that differs significantly from the mitochondrial DNA-based nMDS plot. In the Y-STR nMDS plot, the Huambisa remain relatively isolated, but are closer to Amazonian rather than Andean populations. The Shipibo and Cashibo fall within a nebulous cluster of lowland Brazilian populations, while the Asháninka are located at approximately the middle of the plot, but slightly closer to Andean populations. As perhaps expected, the intermontane Yanesha, also fall into this region of the plot. A Y-chromosome descent clusters (DC) analysis found that although genetic structure largely follows the Andes-Amazon divide, there were instances of successful Y-lineage transmission across these biomes during the pre-Columbian period (Di Corcia *et al*., 2017).

Maziéres *et al*. (2008) assessed maternal ancestry in the Matsiguenga of southern Peruvian Amazonia (*n* = 38) and found 100% Indigenous ancestry (assuming that all individuals with A4 and B4 subclades had Indigenous second-degree subclades) with subclade distributions as follows: 5.3% A4, 92.1% B4, and 2.6% D1. The authors moreover reported a Y-chromosome haplogroup distribution (*n* = 28) of 90.4% Indigenous haplogroup Q (Q3* 80.7% and Q* 9.7%), and 9.6% non-Indigenous Y haplogroups (2008). Although the Matsiguenga reside in the south of Peru, they are still a lowland Amazonian population. Because comparative data is generally lacking in this geographic region, they were included here.

Justice *et al*. (2012) studied the genetic structure of specifically peri-urban *barriada* residents ringing the city of Yurimaguas. Similar to studies conducted in nearby locales, the population was found to exhibit a high frequency of Indigenous maternal haplogroups (100% Indigenous, 21% A2, 33% B, 35% C1, and 11% D). These results may support a lack of non-Indigenous maternal gene flow, or they may indicate that Yurimaguas *barriada* residents have a

greater proportion of Indigenous ancestry compared to other areas of the city, such as the historically European city center. Residence patterns across modern municipalities have been shown in numerous contexts to be associated with socioeconomic, ethnic, religious, and other statuses (Marrett, 1973). Despite an extensive documented history of intercultural relations, non-Indigenous migration into the Lower Huallaga Valley has not resulted in significant levels of maternal gene flow based on the *barriada* population sampled. The analysis of maternal markers in the Yurimaguas *barriada* population underscores the Indigenous component of the rapid rural-to-urban migration in the Lower Huallaga Valley. According to mitochondrial haplogroup stratification, the migration of peoples from rural lands to the modern, urban center of Yurimaguas has most likely resulted in maternal gene flow between previously isolated Indigenous groups, rather than intercontinental gene flow.

Justice *et al*. (2012) hypothesized that findings reflect the fusion of formerly genetically-isolated Indigenous populations (such as Achuar or Chayahuita) who have recently migrated to Yurimaguas in search of improved employment, education, and/or social mobility opportunities (2012). The admixture of formerly isolated populations results in a sharp increase in variation in a relatively short period of time. The newly introduced genetic variants are furthermore likely to appear in lower frequency within the context of a diverse and large population, such as the authors found in the *barriadas* of Yurimaguas.

Certainly, Yurimaguas and the Huallaga Valley are under dramatic migratory pressures. While the impact of European conquest and subsequent colonization resulted in profound genetic change throughout Peruvian Amazonia, recent analysis of genetic diversity in the Yurimaguas *barriada* population highlights the significance of more recent Indigenous migration and urbanization. While pre- and post-contact activities undoubtedly did set the stage for modern biogenetic and environmental change in the *selva baja*, the region's recent intra and intercontinental patterns of migration and urbanization appear to have been largely neglected by researchers. Most studies of the region either focus on historical changes, or depict Amazonian societies as uniform and in stasis, rather than plastic and in flux. Studies cited in this section will

be discussed further in the Discussion chapter, for a comparison with the findings of this dissertation.

Chapter IV: Methods and Materials

Collection of Data and DNA Samples

Throughout August of 2015, interview-based data and buccal DNA swab samples were collected from individuals in two communities of Yurimaguas District, Loreto, Peru. One collection location was *Hospital Santa Gema de Yurimaguas*, located in the central business district of the city of Yurimaguas (*Progreso Nº 305–307, Yurimaguas—Alto Amazonas—Loreto*). The second was *Hospital Santa Gema de Yurimaguas—Munichis* (*Carretera Yurimaguas, Munichis—Alto Amazonas—Loreto*), located in a small peri-urban community within Yurimaguas District, but outside the borders of the city proper. Government hospital clinics were utilized as a means of data collection due to their familiarity, geographic centrality, and inclusive philosophy towards all strata of the community. Such clinics were furthermore used as a host site that offered greater sterility and privacy than other potential venues, where participants from the surrounding communities could come for the stated purpose of participating in this study. Investigators responsible for the collection of data and DNA samples were Randy E. David and Dr. Bartholomew Dean, both of the University of Kansas and Laboratory of Biological Anthropology (LBA). Prior to data and DNA sample collection, permissions were obtained from the University of Kansas Institutional Review Board, Human Subjects Committee Lawrence (STUDY00141454), and the *Gobierno Regional de Loreto, Dirección General de Salud – Loreto, Hospital Santa Gema de Yurimaguas.* Potential participants were provided a detailed informational pamphlet (provided in both English and Spanish) and informed of the goals and possible risks associated with the study. Ultimately, participants were asked to sign an informed consent form and were distributed the contact information of investigator (and author of this dissertation) Randy E. David, and head of the LBA, Dr. Michael H. Crawford.

Demographic data and migration histories were ascertained via standardized one-on-one, in-person interviews. “Migration” was defined as either permanent relocation, or relocation for a period of ≥ 1 year for any reason other than an immediate existential threat. Refugees, for

example, did not fit this criterion. “Residence” was defined as living for ≥ 1 year in a single municipality. Migratory variable-based associations were not standardized by any given factor (*e.g*. age) since one of the primary goals of this research was to establish the presence or absence of fundamental demographic/migratory associations with genetic traits, *given* the particular substructure of the model Yurimaguas population. Standardization would moreover decrease the statistical power of the study design. Lastly, age in particular was not found to be statistically associated with a history of migration (calculated but not shown), calling into question whether a simple standardization would even have a discernable effect.

Participants were instructed to skip any inquiries that they felt uncomfortable answering. Responses were recorded by hand via pre-designed survey sheets and were designed to include both structured response formats (producing categorical, ordinal, and interval data), and unstructured response formats (producing fully descriptive qualitative or numeric data). Responses were subsequently coded and compiled into a spreadsheet. Each interview took approximately 30 minutes to complete. Interviews were generally conducted in a reserved, secluded corner of the waiting/family room of the hospital with both the interviewer and interviewee seated across from one another. In addition to interview-based data, each participant, after rinsing their mouth with provided bottled water, had buccal cells collected by the gentle scraping of the inner cheek, using Isohelix™ SK–1 swab kits. According to kit directions, swabs were transferred to pre-sterilized, screw-top tubes for safe storage, and had an Isohelix™ “dri-capsule” added to them for long-term preservation. After the collection of 182 samples, swabs/tubes were packed and shipped via Serpost/FedEx (Lima, Peru/Memphis, TN) express to the LBA, University of Kansas, Lawrence, Kansas, where they were transferred to freezers (–20 °C) for permanent storage. Samples did not have to be maintained at a particular temperature during the shipping process. Due to an interest in migratory processes, individuals were not kept from participating based on ancestral claims. Only those over 18 years of age were asked to participate. Of the 182 sampled participants, 79 (43.3%) identified as male, 97 (53.3%) identified as female, and 6 (3.3%) chose not to identify as either male or female. The overall mean age of the sample population was 41.5 years (40.7 years for females, and 42.8 years for males).

Laboratory Methodology

DNA Extraction and Purification

DNA was extracted and purified from the collected Isohelix™ SK–1 buccal swabs using Qiagen (Hilden, Germany) QiAamp® DNA Mini Kits according to the provided spin protocol. Extracted DNA in solution was relocated to a freezer (–20 °C) for storage. All parts of the extraction and purification protocol were carried out in the Anthropology Department, University of Kansas, Lawrence, KS. Gilson (Middleton, WI) PIPETMAN® precision microliter pipettes were calibrated before the commencement of labwork. This method resulted in genomic DNA yields ranging from 4.20 ng/μl to 182.00 ng/μl, with a mean of 27.10 ng/μl (150 μl per sample). Genomic DNA quantitation was completed using a Life Technologies (Carlsbad, CA) Qubit 3.0 Fluorometer.

Mitochondrial DNA

All mitochondrial labwork was carried out in the Anthropology Department, University of Kansas, Lawrence, KS. Forward and reverse primers were designed by author of this dissertation, Randy E. David using Primer3 Input 0.4.0 software (Koressaar & Remm, 2007; Untergasser *et al.,* 2012), and ordered via Integrated DNA Technologies (IDT®) (Coralville, IA). Primers were prepared using a standard desalting protocol and were resuspended in Sigma-Aldrich (St. Louis, MO) Tris-EDTA buffer solution, pH 8.0 (TE_{-3}) to 100 μM for storage stock, and 10 μM for utilization. The resultant amplicon was 1,263 bases and covered the mitochondrial revised Cambridge reference sequence (rCRS) control region, positions 1–576 (HVR2) and 16,024–16,569 (HVR1) (NCBI GenBank® accession #: J01415.2 / NC_012920.1, *Homo sapiens* mitochondrion, complete genome) in full (Andrews *et al*., 1999).

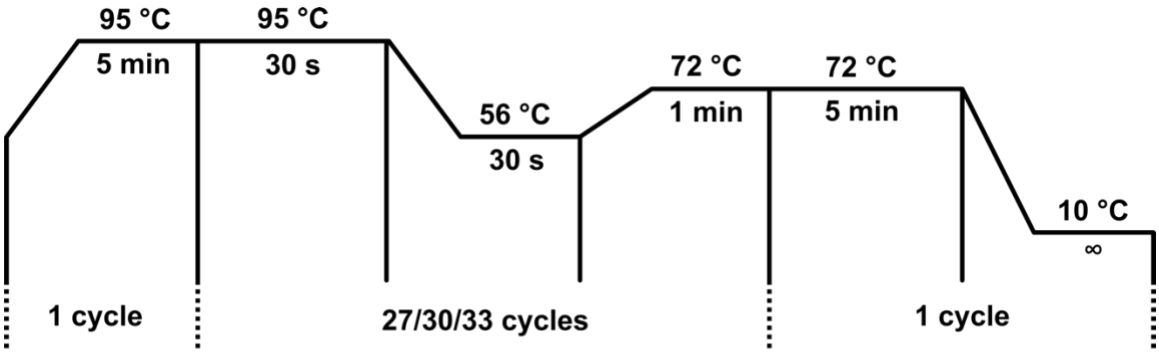
Table 4.1. Mitochondrial control region PCR primers.

Primer Sequence	Direction	Number of Bases	T_m (°C)	GC Content (%)
5'- GTC TTG TAA ACC GGA GAT GAA -3'	Forward	21	56.80	42.86
5'- TGT ATT GCT TTG AGG AGG TAA G -3'	Reverse	22	56.29	40.91

Direction of primer (Direction), number of bases that compose each primer (Number of Bases), melting temperature (T_m), combined guanine and cytosine primer constitution (GC Content). T_m was calculated according to the following equation: $T_m = (2[A + T] + 4[G + C])$ °C.

Polymerase chain reaction (PCR) was conducted to amplify particular DNA fragments for subsequent analysis. PCR was conducted using Thermo Fisher Scientific (Waltham, MA) Dream Taq™ Hot Start DNA Polymerase, and prepared in a Labconco (Kansas City, MO), Delta Series, Purifier Class II biosafety cabinet. Each reaction was 50 μl, and consisted of 5 μl of Dream Taq™ buffer (10x) (containing 20 μM MgCl2), 5 μl of dNTP mix (2 mM each), 1 μl of forward primer (at 10 μM), 1 μl of reverse primer (at 10 μM), 5 μl of template DNA solution from the Qiagen extraction (with a mean concentration of 27.10 ng/μl genomic DNA), 0.25 μl (1.25 U) of Dream Taq™ Hot Start DNA Polymerase, and 32.75 μl of Corning (Corning, NY) molecular biology grade (nuclease-free) water. PCR preparation and execution were conducted via the supplied Thermo Fisher Scientific Dream Taq™ Hot Start DNA Polymerase protocol. Four Corning nuclease-free water controls were included in the process to detect possible cases of contamination. PCR reactions were conducted on an Applied Biosystems by Life Technologies Pro Flex™ PCR System (thermal cycler) (Carlsbad, CA). The annealing temperature was set to 56.00 °C, only slightly below the lower of the two primers' melting temperature of 56.29 °C. This is not as low of an annealing temperature as is often typical according to Innis, Gelfand, Sninsky, and White (1990) and Devor (2004), who suggest a setting some 2.00 °C to 5.00 °C below the lower of the two primers' melting temperatures. The thermal cycler profile for all samples is shown in *Figure 4.1*.

Figure 4.1. MtDNA PCR thermal cycler profile.



There are three PCR steps, each divided by a dotted line. Step 1: template denature, step 2: primer annealing, step 3: DNA synthesis and hold.

Samples that were previously analyzed using a Qubit 3.0 Fluorometer and found to possess a genomic DNA concentration of ≤ 10 ng/μl after extraction underwent 33 annealing cycles. Those with a concentration of > 10, but ≤ 50 ng/μl underwent 30 annealing cycles. Lastly, those samples found to have a concentration of > 50 ng/μl of genomic DNA underwent 27 annealing cycles.

PCR products were tested for successful amplification using electrophoresis. A 1% agarose gel was polymerized by mixing 330 ml of Sigma-Aldrich 1x Tris-Borate-EDTA (TBE) with 3.3 g Lonza (Basel, Switzerland) Sea Kem® LE agarose, and then stained using 19.8 μl (3 μl/50 ml) of Biotium (Fremont, CA) Gel Green® nucleic acid gel stain. The gel was left to set at room temperature (23 °C) for one hour. Four μL of PCR product were added to 3 μL of New England Bio Labs (Ipswich, MA) gel loading dye blue (6x). Six μL of the subsequent mixture was then loaded into each gel well. Electrophoresis was conducted using a Thermo Fisher Scientific, Owl™ Easy Cast™ wide-format horizontal electrophoresis system, model D3, with supplied 100-well comb. One-hundred and twenty volts (~500 mA) were passed across the gel for 50 minutes, using a Life Technologies Power Ease® 90 W power supply. Negative controls using Corning nuclease-free water as a substitute for PCR product were used to detect contamination. Thermo Fisher Scientific Gene Ruler™ 1 kb plus DNA Ladder, in 2 μl and 1 μl aliquots, was used for the sizing and quantitation of PCR products. To confirm successful electrophoresis, lane products

were illuminated using blue light, then photographed, using a Corning Axygen® Gel Documentation System—BL.

Mitochondrial DNA PCR products were purified, and Sanger sequenced (dideoxy method) by Genewiz (South Plainfield, NJ) (Sanger, Nicklen, & Coulson, 1977). One hundred and sixty-seven samples were prepared, packed, and shipped, according to Genewiz protocol. Those samples that did not amplify correctly, of which there were 15, were omitted from sequencing. Specifically, 20 μl of each PCR product was loaded onto a Thermo Fisher Scientific 0.2 ml, 96-well, semi-skirted PCR plate at a concentration of 15–25 ng/μl. Each product was sequenced twice, once using the previously designed forward primer, and once using the previously designed reverse primer. Twenty μl of each primer was supplied to Genewiz, at a concentration of 5 μM. Reactions and primers were then shipped via FedEx express (with dry ice) to: Genewiz—Boston, Attn: Single Pass, 733 Concord Avenue, Cambridge, MA 02138, for sequencing. Associated paperwork was provided to Genewiz via electronic mail.

Non-Recombining, Y-Chromosomal DNA

All non-recombining, Y-chromosomal (NRY) DNA labwork was carried out in the Molecular Biosciences Department, University of Kansas, Lawrence, KS. Twenty-eight ancestry-informing markers were chosen (according to van Oven, Ralf, & Kayser, 2011), and divided into two multiplex genotyping assays for hierarchical detection of worldwide paternal lineages (see *Figure 4.5*) (International Society of Genetic Genealogy [ISOGG], 2019). PCR primers were designed by van Oven *et al*. (2011) and PCR was carried out in 6 μL reactions of Applied Biosystems (Foster City, CA) Gold Buffer (10x), Applied Biosystems $MgCl_2$, 2.5 mM of each Gene Amp™ dNTP (Applied Biosystems), 1.67 units of Applied Biosystems Ampli Taq Gold DNA Polymerase, approximately 2 ng (1 ng/μL concentration) of genomic DNA template, and IDT® forward and reverse PCR primers in concentrations stipulated in *Tables 4.2* and *4.3*.

Table 4.2. NRY multiplex 1 PCR and single-base extension primers.

Locus	Mutation	PCR amplification			Single-base extension				
		Primer sequences (5′-3′)	Conc. (μM)	Amplicon size (bp)	Primer sequence (5′–3′) (5′ aspecific tail in lowercase italics)	Conc. (μM)	Length (nt)	Orientation	Alleles (dye)
M91	ins T	F CAAAAATCCCCCTACATTGC R GCAGTGCCCTTCCAAATAAA	0.600 0.600	144/143	*g* CTACAGTAGTGAACTGATTAAAAAAAA	0.300	28	R	a (yellow), i (green)[a]
M60	ins T	F TCTTTACATTTCAAAATGCATGACT R GAGAAGGTGGGTGGTCAAGA	0.600 0.600	128/129	*ct(gact)$_6$* TAACCACTGTGTGCCTGAT	0.600	45	R	a (yellow), i (green)[a]
M145	G->A	F GCATACTTGCCTCCACGACT R CCTCCCACTCCTTTTTGGAT	0.200 0.200	96	*ct(gact)$_3$gac* TAGGCTAAGGCTGGCTCT	0.450	35	R	G (yellow), A (red)
M174	T->C	F TCTCCGTCACAGCAAAAATG R AGGAGAAGGACAAGACCCATC	0.450 0.450	178	*ct(gact)$_5$g* ATACCTTCTGGAGTGCCC	0.100	41	F	T (red), C (yellow)
M96	G->C	F TGAGCTGTGATGTGTAACTTGG R CACCCACTTTGTTGCTTTGT	0.200 0.200	117	*act(gact)$_{10}$gac* TGGAAAACAGGTCTCTC ATAATA	0.200	69	F	G (blue), C (yellow)
M216	C->T	F CCTCAACCAGTTTTTATGAAGCTA R TTCTAAATCTGAATTCTGACACTGC	0.100 0.100	102	*ct(gact)$_6$g* CTGCTAGTTATGTATACCTGTT GAAT	0.075	53	R	C (blue), T (green)
M89	C->T	F CAGCTTCCTGGATTCAGCTC R CACTTTGGGTCCAGGATCAC	0.200 0.200	105	*ct(gact)$_{13}$ga* AACTCAGGCAAAGTGAGAGAT	0.300	77	R	C (blue), T (green)
M282	A->G	F TGTGCAACCTCAACTTTGCTT R TGTGATCAACTTCTTTCCCTCA	0.750 0.750	106	*t(gact)$_{15}$* GAAAGCAAAATCTCAATATGATAA	1.000	85	F	A (green), G (blue)
P257	G->A	F ACCCCTCAGTCTCTCCGATT R TCATCTCCAACCCCCATCT	0.200 0.200	71	*(gact)$_9$g* ATTATCCCACTGCATTTCTG	0.300	57	F	G (blue), A (green)
M69	T->C	F GGAGGCTGTTTACACTCCTGA R TCTCCCCTTAGCTCTCCTGTT	0.300 0.300	87	*(gact)$_{10}$g* GGCTGTTTACACTCCTGAAA	0.150	61	F	T (red), C (yellow)
M522	G->A	F TCCAATTCCCATGTCCTCTC R CAGTGCAGAAAATCACGGTAGA	0.100 0.100	109	*t(gact)$_{11}$* CTACTACGCCTCTCTTGTCC	0.075	65	F	G (blue), A (green)
M258	T->C	F TTCAGGATTTGTCAAGGATGG R GCTATGACTAAGAGGGATTCCAA	0.200 0.200	108	*t(gact)$_3$gac* GGGATTCCAAGTTCCCA	0.300	33	R	T (green), C (blue)
M304	A->C	F TTGTAACAAACAGTATGTGGGATTT R CGTCTTATACCAAAATATCACCAGTT	0.200 0.200	88	*act(gact)$_{11}$ga* TTATACCAAAATATCACC AGTTGT	0.300	73	R	A (red), C (blue)
M9	C->G	F CTGCAAAGAAACGGCCTAAG R AACTAAGTATGTAAGACATTGAA CGTTTG	0.100 0.100	90	*t(gact)$_7$g* CGGCCTAAGATGGTTGAAT	0.100	49	F	C (yellow), G (blue)

(Reproduced from van Oven et al., 2011)

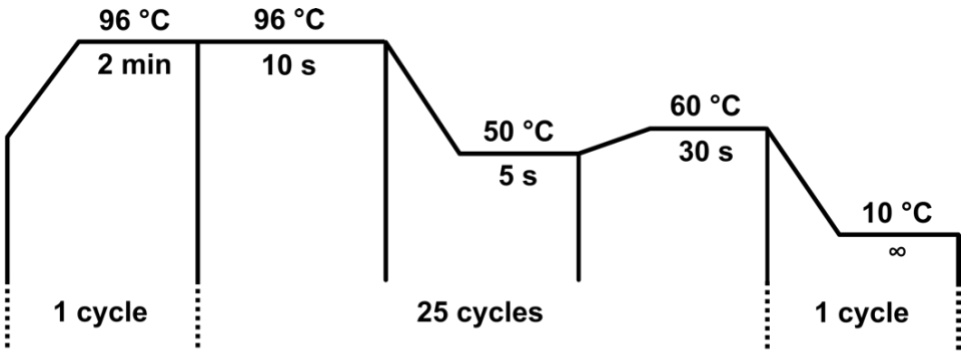
Table 4.3. NRY multiplex 2 PCR and single-base extension primers.

Locus	Mutation	PCR amplification				Single-base extension				
		Primer sequences (5′–3′)		Conc. (μM)	Amplicon size (bp)	Primer sequence (5′–3′) (5′ aspecific tail in lowercase italics)	Conc. (μM)	Length (nt)	Orientation	Alleles (dye)
M9	C->G	F	CTGCAAAGAAACGGCCTAAG	0.300	90	*t(gact)$_7$g* CGGCCTAAGATGGTTGAAT	0.100	49	F	C (yellow), G (blue)
		R	AACTAAGTATGTAAGACATTGAACGTTTG	0.300						
M526	A->C	F	TAGAGGCAGGGTGTTGCTCT	0.300	100	*ct(gact)$_{10}$ga* TGTCATCAGGCTGAATCATAC	0.450	65	F	A (green), C (yellow)
		R	TACTTTGGGAGGCTGCTGTT	0.300						
M147	ins T	F	CCTGAATAAGCTGGTGAAAGAAA	0.500	114/115	*ct(gact)$_{11}$ga* CCTGTCTCTGAAAGAAAAAAA	1.000	69	R	a (yellow), i (green)[a]
		R	GGAGACCCTGTCTCTGAAAGAA	0.500						
P308	C->T	F	GCTACCAATACCCCCAAAGA	0.050	108	*gactgac* GAAATGATTAAGTAAGTGCCTTCT	0.150	31	R	C (blue), T (green)
		R	CCTGGAATATGGCACGAAAT	0.050						
P79	T->C	F	TTGCTTAGTATAATGTCTTTCATGCTC	0.500	101	*ct(gact)$_4$g* TGCTCATTCGCATCTTTG	1.000	37	F	T (red), C (yellow)
		R	AAATGAGGCTAATCAATGGAACA	0.500						
P261	G->A	F	TCCTAGAAGGTAACCCACTACCC	0.500	93	*t(gact)7gac* TTTTTGTTTTTATTAATGAATGCTA	1.000	57	R	G (yellow), A (red)
		R	TGTGCATATGTTATCCACCATGT	0.500						
P256	G->A	F	TCTTGGTTTTCCCATTGACC	0.200	91	*t(gact)$_{13}$ga* TGCCCTACACTAGATAGAAAGG	0.150	77	F	G (blue), A (green)
		R	CATCTCCCAACTTGTCTGTGC	0.200						
M231	G->A	F	AACAACATTTACTGTTTCTACTGCTTTC	0.300	119	*act(gact)$_3$g* CGATCTTTCCCCCAATT	0.450	33	R	G (yellow), A (red)
		R	TTCACATCATCCAGTACAGCAA	0.300						
M175	5 bp del	F	CCCAAATCAACTCAACTCCAG	0.300	101/96	*t(gact)$_{10}$* CACATGCCTTCTCACTTCTC	0.600	61	F	a (red), d (green)[a]
		R	TTCTACTGATACCTTTGTTTCTGTTCA	0.300						
M45	G->A	F	CATCGGGGTGTGGACTTTA	0.400	109	*act(gact)$_6$g* AATTGGCAGTGAAAAATT ATAGATA	0.750	53	F	G (blue), A (green)
		R	CCTCAGAAGGAGCTTTTTGC	0.400						
M242	C->T	F	AAAAAGGTGACCAAGGTGCT	0.400	46	*ct(gact)$_6$g* CGTTAAGACCAATGCCAA	0.100	45	R	C (blue), T (green)
		R	AAAAACACGTTAAGACCAATGC	0.400						
M207	A->G	F	GGGGCAAATGTAAGTCAAGC	0.300	83	*(gact)$_{14}$g* AATGTAAGTCAAGCAAGA AATTTA	0.300	81	F	A (green), G (blue)
		R	TCACTTCAACCTCTTGTTGGAA	0.300						
M412	G->A	F	GGCACTCCTCCGTCATCTT	0.300	114	*ct(gact)$_{16}$* GGGTACAATCTGATGAGGC	0.300	85	F	G (blue), A (green)
		R	GGTGAAGTGGACCCTATCCA	0.300						
P202	T->A	F	AAACTTCCCAGTTTGTGGTTC	0.300	125	*ct(gact)$_{12}$ga* CCAGTTTGTGGTTCTTTGTTA	0.300	73	F	T (red), A (green)
		R	TGATCCCTTAATTAATAGCAAGACC	0.300						
P326	T->C	F	TTCAGATATCAGGCCGCTTT	0.200	61	*t(gact)$_3$* CCTAAGCAGAGGAAAATA GTACAG	0.150	37	R	T (green), C (blue)
		R	GAGCTGTCAGCCTGCCTAAG	0.200						

(Reproduced from van Oven et al., 2011)

Reactions were constructed according to optimized reagent ratios informed by the Applied Biosystems protocol, and prepared in a Labconco Delta Series, Purifier Class II biosafety cabinet. DNA templates were first diluted to a concentration of 1 ng/μL using Sigma-Aldrich Tris-EDTA buffer solution, pH 8.0 (TE_{-4}). Primers were prepared using a standard desalting protocol and were resuspended in Sigma-Aldrich Tris-EDTA buffer solution, pH 8.0 (TE_{-3}) to 100 μM for a storage stock, and 40 μM for utilization. Two Corning nuclease-free water controls were included in the process to detect possible cases of contamination. PCR reactions were conducted on an Applied Biosystems by Life Technologies Pro Flex™ PCR System (thermal cycler). The thermal cycler profile for all samples is shown in *Figure 4.2*.

Figure 4.2. NRY PCR thermal cycler profile.

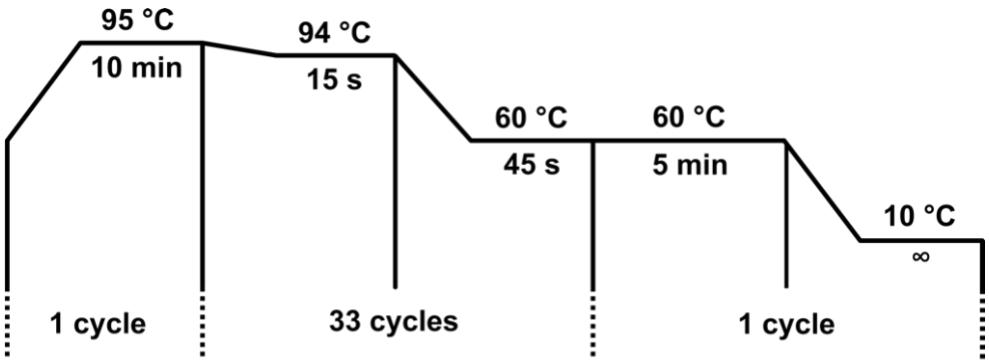


There are three PCR steps, each divided by a dotted line. Step 1: template denature, step 2: primer annealing, step 3: DNA synthesis and hold.

PCR products were purified by adding 2 μL of Thermo Fisher Scientific Exo SAP-IT™ reagent, for a total volume of 8 μL, followed by incubation at 37 °C for 30 minutes and 80 °C for 15 minutes.

Single-base primer extension was carried out in 6 μL reaction volumes, consisting of 1 μL purified PCR product, 1 μL of Applied Biosystems SNaPshot™ Ready Reaction Mix, and 4 μL extension primers in concentrations stipulated in *Tables 4.2* and *4.3*. The thermal cycler profile for all samples is shown in *Figure 4.3.*

Figure 4.3. NRY single-base extension thermal cycler profile.



There are three PCR steps, each divided by a dotted line. Step 1: template denature, step 2: primer annealing, step 3: DNA synthesis and hold.

Single-base extension products were purified by adding 1 unit of USB® Shrimp Alkaline Phosphatase (SAP) to the 6 μL product and incubating at 37 °C for 45 minutes and 75 °C for 15 minutes. Reactions were prepared for capillary electrophoresis (CE) by creating a mixture of 5 μL single-base extension product, 4.85 μL Thermo Fisher Scientific Hi-Di™ Formamide, and 0.15 μL Thermo Fisher Scientific Gene Scan™ 120 LIZ™ Dye Size Standard, for a total volume of 10 μL. Positive (3 μL SNaPshot™ Ready Reaction Mix, 1.2 μL Applied Biosystems control DNA template, 0.6 μL Applied Biosystems control primer mix, 1.2 μL Corning nuclease-free water, for a total volume of 6 μL) and negative (3 μL SNaPshot™ Ready Reaction Mix, 0.6 μL Applied Biosystems control primer mix, 2.4 μL Corning nuclease-free water, for a total volume of 6 μL) single-base extension controls were also prepared. Reactions were then incubated at 95 °C for 5 minutes.

Products were loaded onto a Thermo Fisher Scientific Micro Amp™ Optical 96-Well Reaction Plate and added to an Applied Biosystems 3130xl Genetic Analyzer for capillary electrophoresis. Extended fragments were separated using Thermo Fisher Scientific POP-7™ Polymer and a Thermo Fisher Scientific DS-02 Matrix Standard Kit (including five dye-labeled oligonucleotides). Capillary electrophoresis was conducted using the following conditions: 10 s injection time at 1.2 kV, followed by a 500 s run time at 15.0 kV.

Software-Based Genetic Analyses

Mitochondrial DNA

Upon receiving sequencing results from Genewiz files were converted from ab1 format, and the associated electropherograms were viewed using FinchTV Version 1.5.0 (Geospiza Research Team, 2012). Sequences were edited (aligned and trimmed) and bases were called (against the rCRS) using Gene Codes Corporation's (Ann Arbor, MI) Sequencher® 5.4.6 software. Consensus sequences were created from the forward and reverse sequence calls and were prepared for mitochondrial haplogroup assignation (including the identification of SNP's and conversion to VCF format), using MtDNAprofiler (Yang, Lee, Yang, & Shin, 2013).

Mitochondrial haplogroup assignation (including a check for "phantom mutations" and haplogroup discordance) was performed using HaploGrep version 2.1.1 (Kloss-Brandstätter *et al*., 2011; Weissensteiner *et al*., 2016) in conjunction with Phylotree Build 17 (van Oven, 2015; van Oven & Kayser, 2009). Haplogroups were confirmed using EMMA, a software-based algorithm and assignation tool predicated on the location and frequency of private mutations found in the control region of the mitogenome. Most mitochondrial "haplogrouping" tools rely on only traditional ancestry-defining mutations and ignore potentially informative private mutations (Röck, Dür, van Oven, & Parson, 2013). EMMA utilizes comparative genomes from EDNAP (the European DNA Profiling Group) of the Forensic MtDNA Population Database (EMPOP) and has been confirmed to have a powerful and accurate effect on haplogroup assignation, with both complete and even partial mitogenome control region sequences.

Haplotypes are chosen based on the appearance of various unique genetic markers that are statistically associated. Haplogroup is therefore defined as the clade in which members that share a haplotype belong. Haplogroup cladistics infer a recent common ancestry based on genetic similarity at specific genomic positions. Because of this association, haplogroup distribution in modern, diverse populations informs both ancient and recent migratory events. Haplogroup patterns have furthermore been associated, via deciphering of the mitogenome, with

various health conditions, including osteoarthritis (Rego-Pérez, Fernández-Moreno, Fernández-López, Arenas, & Blanco, 2008), ischaemic attack and stroke (Chinnery, Elliott, Syed, & Rothwell, 2010), type II diabetes mellitus (Fuku *et al*., 2007), and obesity (Guo *et al*., 2005), among others.

The defining mitochondrial DNA subclades of the Indigenous peoples of the Americas include A2, B2, C1b, C1c, C1d, C4c, D1, D2a, D3, D4h3a, X2a, and X2g (van Oven, 2015). Despite its status as a founding mitochondrial subclade, A2 is found in significant proportions and with the greatest diversity, in northeastern Asia, particularly Siberia, indicative of the spread of peoples across Beringia into the Americas (O'Rourke & Raff, 2010; Volodko *et al*., 2008). Haplogroup B is most commonly found in northeastern Asia, eastern Asia, and southeastern Asia. Due to its great diversity measures in China, it has been proposed to have originated in the east or southeast of mainland Asia (Yao, Kong, Bandelt, Kivisild, & Zhang, 2002). Subclade B2, unlike subclade A2, is found solely among Indigenous peoples of the Americas. Haplogroup C is frequently found in Eurasia and northeastern Asia. It has been proposed to have originated in southern Siberia, in the Lake Baikal region (Derenko, *et al*., 2010). Subclades C1b, C1c, C1d, and C4c are particular only to the Americas. Haplogroup D is frequently found in northern, northeastern and central Asia, and Indigenous populations of the Americas. It is less commonly found in eastern Europe, and southwestern Asia. Haplogroup D is believed to have arisen is southern Siberia, in close proximity to where Haplogroup C originally diversified (Derenko *et al*., 2010). Subclades D1 and D4h3a in particular are found throughout a substantial portion of Indigenous populations. Subclades D2a and D3 are primarily restricted to arctic and sub-arctic populations. Haplogroup X is found in many geographic regions of the world, particularly North Africa, Europe, and the Levant. It is found in its greatest frequency among the Druzes, a genetic population isolate currently living in Israel, Lebanon, Jordan, and Syria (but proposed to be the descendants of Ismaeli Muslims in North Africa) (Reidla *et al*., 2003). Subclades X2a and X2g are primarily restricted to the sub-arctic and surrounding areas and are therefore not of great consequence to research in South America (Perego *et al*., 2009; Raff & Bolnick, 2015).

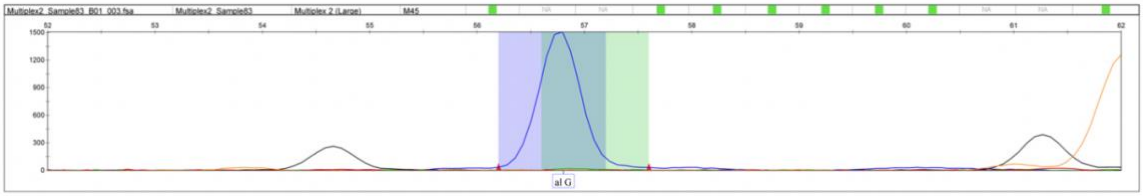
Indigenous haplogroup distribution can be utilized to make inferences about pre-Columbian migrations, including the presence or absence of mating barriers. This is of particular

import in the Lower Huallaga Valley, which is adjacent to the Andes-Amazon transition zone, and a noted region of historical migration and trade. Linguistic and archaeological evidence can either dovetail with genetic findings, lending credence to migratory hypotheses, or it can detract. West Eurasian haplogroups that potentially represent gene flow events into South America include H, T, U, V, K, I, J, and W (van Oven, 2015). East Eurasian haplogroups potentially include E, F, G, M (greatest frequency in East Eurasia, but possesses a wide geographic range), Y, and Z (2015). Sub-Saharan African haplogroups potentially include L0, L1, L2, L3, L4, L5, and L6 (2015). Australo-Melanesian haplogroups potentially include P, Q, and S (2015). The presence or absence of such lineages can either support or refute historical, linguistic, or cultural commentaries on the region's history of migration from the colonial period, onward.

Non-Recombining, Y-Chromosomal DNA

Single-base extension products were analyzed using Genemapper® Software 5, from Applied Biosystems. Samples were added to a newly created project, and optimal analysis parameters and table settings were selected. A sizing analysis was performed using Thermo Fisher Scientific Gene Scan™ 120 LIZ™ Dye Size Standard to ensure correct scaling of electropherogram peaks. A nested system of kits, panels, markers, bin sets, and bins were created. Each marker created corresponds to a fragment size range for each of 28 ancestry-informing markers (AIMs). Each bin corresponds to a fragment size and dye color (according to *Tables 4.2* and *4.3*) that define a given allele within a marker. Refer to *Figure 4.4* for an illustration of both markers and bins.

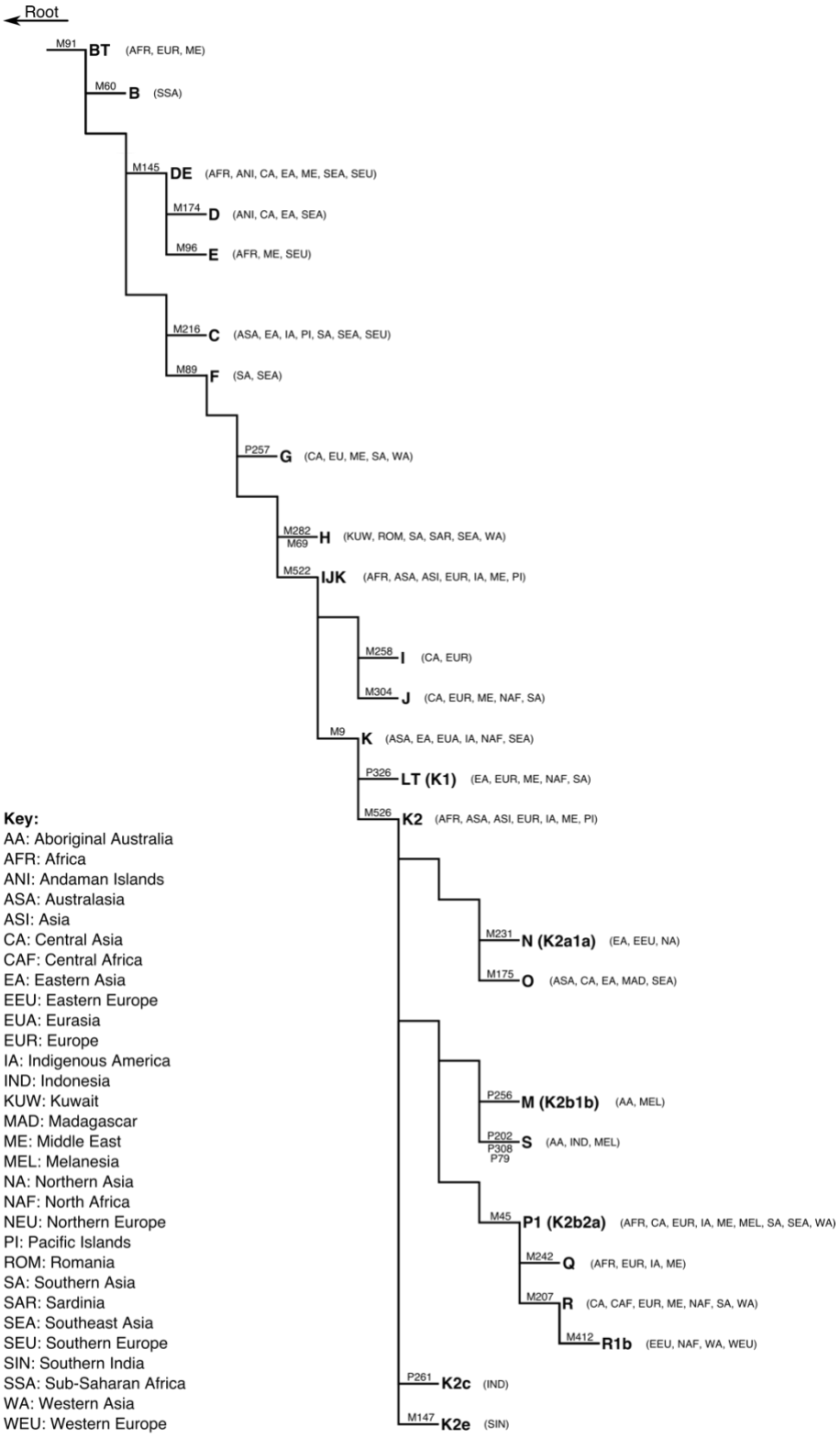
Figure 4.4. Marker and bin creation using Applied Biosystems Genemapper® Software 5.



This electropherogram depicts a G (blue) allele at approximately 56.8 nt (nucleotides) for the M45 locus, multiplex 2. The fragment size range (marker) corresponds to the blue and green rectangular areas. The horizontal spread of each is set according to the location along the x-axis (length, in nucleotides [nt]) of peaks, across all samples. The y-axis measures relative fluorescence units (rfu), exhibiting in this case an intensity of approximately 1,500 rfu.

Samples were hierarchically genotyped for either multiplex 1 or both multiplexes 1 and 2, depending on the M9 locus allele. If a C→G mutation occurred, the sample would be genotyped for multiplex 2 (see *Table 4.3*). If a mutation at this locus had not occurred, the sample would only be genotyped for multiplex 1 (see *Table 4.2*). Once all samples were fully genotyped, Y-chromosome haplogroup assignation was conducted in accordance with *Figure 4.5*.

Figure 4.5. Y-chromosome phylogenetic tree "trunk" using 28 ancestry-informing SNPs.



(Haplogroup YTree v7.03.00, 2019; ISOGG, 2019; Tiirikka & Moilanen, 2015)

The geographic associations of the 28 AIMs utilized are listed in *Figure 4.5*. The distribution of Y-chromosome haplogroups reported is discussed in the Results chapter. Analyzing the distribution of Y-chromosome haplogroups can generate a clearer understanding of the peopling of Yurimaguas and the Lower Huallaga Valley. Such information is integral to elucidating the directionality of historical mating episodes in the region, the effects of sociocultural factors (such as structural racism), and possibly even ancestry-related Y-chromosome-linked health conditions (such as Swyer syndrome and Y-chromosome infertility) (Baxter & Vilain, 2013; Krausz & Casamonti, 2017; Liu, Hu, Guo, & Sun, 2016). Mating directionality is of particular import to the framing of this is study, since many of the region's denizens have historically been migrant workers from around the region, country, and world. The majority of migrant workers have been overwhelmingly men. Sex-skewed populational gene flow can also be the consequence of earlier conquests into the Lower Huallaga Valley by the Spanish Empire in the 16th and 17th centuries, who were also overwhelmingly men. Directional mating may moreover have an effect on the modern creation of urban centers through demographic and migratory processes.

Statistical Analyses

Interview-based data was collected one-on-one and in-person, then later transferred to spreadsheets (R software, "xlsx" package) for further processing. Descriptive and inferential statistics were calculated via Arlequin version 3.5, R version 3 and IBM SPSS Statistics version 25.0 (Excoffier & Lischer, 2010; IBM Corporation, 2017; R Core Team, 2013). Contingency tables, choropleths, and graphs were created to represent collected demographic and migratory data, in addition to various tests of association at the nucleotide, haplogroup, and continental-ancestry levels. Pictorial representations were augmented for visual clarity using the vector graphics editor, Inkscape 0.92.3. Geographic distances were calculated via Google Maps. Distance by automobile was used as a default (Map Data: Google, 2018a). If a given locale was inaccessible via roadway, then direct orthodromic distance was calculated (Map Data: Google,

2018b). Unless explicitly stated otherwise, the population of Yurimaguas is inclusive of the peri-urban area of Munichis. Below is an accounting of statistical methods employed, organized by research question.

Demographic and Migratory Trends in the Yurimaguas Sample

Fisher's exact tests were performed using SPSS 25.0 to assess independence between two or more nominal variables, *i.e*. if the proportions of a given nominal variable vary according to the proportions of another nominal variable or variables (Fisher, 1922). The null hypothesis for Fisher's exact tests was that the proportions of one variable are independent of the proportions of another (1922). Fisher's exact test is more accurate than Pearson's chi-square test of independence when cell numbers are small (McDonald, 2014). Therefore, for the purposes of this study, any contingency comparison that possessed a single cell with $n \leq 5$ was analyzed using a Fisher's exact test. Fisher's exact test was furthermore employed when data was unbalanced, and contingency cells were skewed in one direction (often resulting in low [≤ 5] *n* values per cell). Lastly, Fisher's exact test was used when comparing demographic variables, for example: age interval versus living environment (*Figure 5.5*), and likelihood of migration history versus birth Region (*Table 5.2*). An exact, 2-sided significance was provided for each test.

A Pearson's chi-square test of independence was performed using SPSS 25.0 to assess independence between two or more nominal variables, *i.e*. if the proportions of a given nominal variable vary according to the proportions of another nominal variable or variables (Pearson, 1900). The null hypothesis for Pearson's tests was that the proportions of one variable are independent of the proportions of another (McDonald, 2014). Pearson's test is more appropriate than Fisher's exact test when the sample size, *n*, as well as the individual values in each contingency table cell are relatively large (McDonald, 2014). For the purposes of this study, any contingency comparison that possessed all cells with $n > 5$, unless explicitly stated otherwise, was analyzed using a Pearson's test. If a dataset was relatively small, exact, 2-sided significance

was used rather than asymptotic 2-sided significance. Exact significance, although computationally intensive, is more accurate than asymptotic significance, which assumes a large and normally-distributed data set (Fisher, 1925). Pearson's chi-square test was used when comparing numerous categorical demographic datasets, including language type versus living environment (*Figure 5.6*), sex versus living environment (*Figure 5.7*), likelihood of migration history versus language type (*Figure 5.8*), and likelihood of migration history versus sex (*Figure 5.9*). A chi-square statistic, it's associated *P*-value, and degrees of freedom were provided.

The Kruskal-Wallis *H* (mean rank) test is a non-parametric calculation appropriate when assessing one nominal variable and one ranked (measurement) variable, to determine whether mean ranks are the same across all nominal variable-based groups. It is frequently used instead of a one-way ANOVA when the measurement variable does not meet the assumption of normality (McDonald, 2014). The null hypothesis of a Kruskal-Wallis *H* test is that the mean ranks of the different nominal variable-based groups are equivalent (Kruskal & Wallis, 1952). This test was performed using SPSS 25.0 to assess association between number of migration episodes, modeled as the dependent variable, and four separate demographic variables, *i.e.* sex, language type, living environment, and age interval, modeled as the independent variables (*Table 5.4*).

An Independent samples t-test is used when there is one nominal variable (with two categories) and one measurement variable, and one wants to assess whether the mean of the measurement variable differs significantly between the two nominal variable-based categories (Student [Gosset, W. S.], 1908). The null hypothesis is that the mean of the measurement variable is equal across both categories. The test assumes that the observations within each nominal variable category are normally distributed (although in actuality the test is not sensitive to departures from this assumption if distributions of both groups similar). Three independent samples t-tests were performed using SPSS 25.0 to test for association between migration distance and the demographic variables of sex, language type, and living environment (*Table 5.6*).

A one-way ANOVA was employed using SPSS 25.0 for *Table 5.6* to assess association between migration distance and the demographic variable of age interval. This test is appropriate when one nominal variable divides a measurement variable into two or more groups (McDonald, 2014). A one-way ANOVA analyzes the means between these groups, testing whether they are significantly different from one another (2014). The null hypothesis is that means are equal between groups of the measurement variable (2014). Lastly, A one-way ANOVA assumes that observations within each measurement variable group are distributed normally and works by comparing the variance among the means of the different groups, to the mean variance within each group (Howell, 2010; McDonald, 2014).

Research Question 1: What is the uniparental marker-based continental-ancestry composition of Yurimaguas? Moreover, is there evidence of sex-skewed gene flow?

A Z-test for two proportions was performed using R version 3 software to assess whether there is a significant association between maternal versus paternal Indigenous/non-Indigenous ancestry proportions, and in doing so, address the presence or absence of populational sex-skewed gene flow (*Table 5.10*). This test is appropriate for the comparison of two proportions on either side of a given nominal variable, particularly to address whether the said proportions are significantly different from each other, and the degree of their difference. The null hypothesis is that the two proportions are identical to each other. The Z-test assumes that samples are independent of each other, and that simple random sampling was used for the selection of each sample (Sprinthall, 2011). Compared to t-tests, Z-tests are more appropriate when $n > 30$, as data points begin to approach a normal distribution (2011).

An omnibus (in this case 2x3) Fisher's exact test was performed using SPSS 25.0 to assess independence between maternal-continental ancestry and paternal continental-ancestry (*Table 5.11*). As previously stated, Fisher's exact tests assess whether the proportions of a given nominal variable vary according to the proportions of another nominal variable or variables

(Fisher, 1922). A Fisher's exact test was utilized rather than a Pearson's test because maternal continental-ancestry was unbalanced, with the category of "other" possessing much lower contingency cell n values compared to the alternative categories of "Indigenous" and "European." Because the omnibus Fisher's exact test was found to be statistically very significant (*P* = 0.00487**), three post hoc Fisher's exact tests were performed to identify specific, pairwise associations between the stipulated ancestry categories (*Tables 5.11 and 5.12*). A Bonferroni correction for multiple comparisons was used to adjust the resultant *P*-values. For table consistency, Fisher's exact tests were used for all comparisons even if contingency cell values were > 5.

Research Question 2: What is the pattern of genetic diversity in Yurimaguas?

Haplotype diversity (*H*), nucleotide diversity (π), and $\hat{\theta}_S$ *(Theta[S])* were calculated using Arlequin 3.5 for the following: (1) the total Yurimaguas sample (*Table 5.13*), (2) the demographic variables of sex, language type, living environment, age interval, and maternal ancestry (*Table 5.15*), and (3) the migratory variables of history of migration, number of migration episodes, migration duration, and migration distance (*Table 5.16*). Additionally, the mean number of pairwise differences (Π) was calculated using Arlequin 3.5 for the total Yurimaguas sample (*Table 5.13*).

Haplotype diversity (*H*) represents the likelihood that two haplotypes, randomly sampled from a given sample, are different from one another (Excoffier & Lischer, 2010). Haplotype diversity (*H*) and its sampling variance are calculated according to the following equations:

Diversity (*H*):

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

Sampling variance:

$$\mathrm{V}(\hat{H}) = \frac{2}{n(n-1)}\left\{2(n-2)\left[\sum_{i=1}^{k} p_i^3 - (\sum_{i=1}^{k} p_i^2)^2\right] + \sum_{i=1}^{k} p_i^2 - (\sum_{i=1}^{k} p_i^2)^2\right\}$$

(where *n* represents the total number of gene copies in the sample, *k* represents the total number of haplotypes in the sample, and *pi* represents the "sample frequency to the *i*-th haplotype" [Excoffier & Lischer, 2010, manual p. 117; Nei, 1987])

Nucleotide diversity (π) represents the likelihood that two nucleotides, chosen at random from a DNA sequence, are different from one another (Excoffier & Lischer, 2010). As a metric of populational polymorphism, nucleotide diversity (π) is a proxy for the introduction of diversity through evolutionary forces such as gene flow. Nucleotide diversity (π) and its sampling variance are calculated according to the following equations:

Diversity (π):

$$\hat{\pi}_n = \frac{\sum_{i=1}^{k} \sum_{j<i} p_i p_j \hat{d}_{ij}}{L}$$

Sampling variance:

$$\mathrm{V}(\hat{\pi}_n) = \frac{n+1}{3(n-1)L}\hat{\pi}_n + \frac{2(n^2+n+3)}{9n(n-1)}\hat{\pi}_n^2$$

(where $\hat{d}_{ij}$ represents the number of differences observed between haplotypes *i* and *j*, *k* represents the total number of haplotypes present in the sample, p_i represents the incidence of haplotype *I*, p_j represents the incidence of haplotype *j*, *n* represents the sample size, and *L* represents the number of sequence loci [Nei, 1987; Tajima, 1983])

Watterson's $\theta\hat{}_S$ *(Theta[S])* is a genetic diversity metric based on the proportion of observed versus expected number of segregating (polymorphic) sites (*S*), given a neutral evolution scheme. Its value is determined under the infinite-site model and takes into account the number of segregating sites (*S*), and sample size (*n*), for a given segment of uniparental DNA (Excoffier & Lischer, 2010; Watterson, 1975). $\theta\hat{}_S$ calculations provide insight into departures from neutrality, and are calculated according to the following equation:

$$\theta = \frac{S}{a_1}$$

where:

$$a_1 = \sum_{i=1}^{n-1} \frac{1}{i}$$

The variance of $\theta\hat{}_S$ can be calculated according to the following equation (Tajima, 1989):

$$\mathrm{V}(\hat{\theta}_S) = \frac{a_1^2 S + a_2 S^2}{a_1^2 (a_1^2 + a_2)}$$

where:

$$a_2 = \sum_{i=1}^{n-1} \frac{1}{i^2}$$

(Watterson, 1975).

Mean number of pairwise differences *(Π)* is defined as the mean number of nucleotide differences among all pairwise haplotype comparisons in a given sample (Excoffier & Lischer, 2010). Mean number of pairwise differences (*Π*) and its total variance are calculated according to the following equations:

MNPD (Π):

$$\hat{\pi} = \frac{n}{n-1}\sum_{i=1}^{k}\sum_{j=1}^{k} p_i p_j \hat{d}_{ij}$$

Total variance:

$$\mathrm{V}(\hat{\pi}) = \frac{3n(n+1)\hat{\pi} + 2(n^2+n+3)\hat{\pi}^2}{11(n^2-7n+6)}$$

(where $\hat{d}_{ij}$ represents the number of differences observed between haplotypes *i* and *j*, *k* represents the total number of haplotypes present in the sample, p_i represents the incidence of haplotype *I*, p_j represents the incidence of haplotype *j*, and *n* represents the sample size [Tajima, 1983, 1993b])

Research Question 3: Is there evidence to suggest that the Yurimaguas sample is in population size equilibrium versus demographic expansion, according to neutrality tests and mismatch distribution analyses? Is there evidence of amalgamation?

Tajima's and Fu's tests of selective neutrality as well as Chakraborty's test of amalgamation were performed using Arlequin 3.5 on the total Yurimaguas sample (*Table 5.17*). Additionally, Tajima's and Fu's tests were performed for (1) the demographic variables of sex, language type, living environment, age interval, and maternal ancestry (*Table 5.15*), and (2) the migratory variables of history of migration, number of migration episodes, migration duration, and migration distance (*Table 5.16*) (although both are reported under Research Question 2 in the Results chapter). A Bonferroni correction was applied to all demographic and migratory variable subcategories to adjust for multiple comparisons.

Tajima's test of selective neutrality compares "two estimators of the mutation parameter *θ (theta),"* where $\theta = 2Mu$, and *M* is equal to *N* in "haploid populations of effective size *N*" (Excoffier

& Lischer, 2010, p. 144). It is predicated on the infinite-site model without recombination. The calculation is as follows:

$$D = \frac{\hat{\theta}_\pi - \hat{\theta}_S}{\sqrt{Var(\hat{\theta}_\pi - \hat{\theta}_S)}}$$

$$\hat{\theta}_\pi = \hat{\pi} \text{ and } \hat{\theta}_S = S / \sum_{i=0}^{n-1} (1/i)$$

(where *S* represents the number of segregating sites in the sample, and *n* represents the sample size)

The statistical significance of Tajima's *D* is determined via the production of random samples using a coalescent simulation algorithm, provided the hypothesis of population equilibrium and selective neutrality (Excoffier & Lischer, 2010). The *P*-value is therefore calculated as the proportion of random F_S statistics ≤ the observation. Tajima's test is less sensitive to demographic expansion than Fu's test (Excoffier & Lischer, 2010; Hudson, 1991; Tajima, 1989, 1993a, 1996). Tajima's *D*, unlike Fu's F_S, is a measure of expected versus observed low frequency polymorphisms. Significantly non-zero values of Tajima's *D* indicate deviations from the null model of selective neutrality and constant population size; changes in population size can confound the test for neutrality. For example, a positive value indicates a decreased number of low-frequency (and high-frequency) polymorphisms compared to the expected, which may result from selection, or may result from a population size bottleneck or founding event (Ashfaq *et al*., 2014). A negative value indicates a relative excess of low frequency polymorphisms compared to the expected. Again, this may result from selection, or may signify a possible population expansion (2014).

Fu's test of selective neutrality is based on the infinite-site model without recombination. It assesses the "probability of observing a random, neutral sample with a number of alleles similar or smaller than the observed value given the observed number of pairwise differences, taken as

an estimator of *θ"* (Excoffier & Lischer, 2010, manual p. 145; Fu, 1997). The aforementioned probability is calculated according to the following equation:

$$S' = \Pr(K \geq k_{obs} \mid \theta = \hat{\theta}_{\pi})$$

The test statistic is then the logit of this probability (logit *S'*) (Excoffier & Lischer, 2010):

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

(where *k* represents the total number of haplotypes in the sample, and $\hat{\theta}_{\pi}$ represents the mean number of nucleotide differences between two sequences [Fu, 1997]).

The statistical significance of Fu's F_S is determined via the production of random samples using a coalescent simulation algorithm, given the hypotheses of both population size equilibrium (constant population size) and selective neutrality. The *P*-value is therefore calculated as the proportion of random F_S statistics ≤ the observation. Fu's test is more sensitive to demographic expansion than Tajima's test of selective neutrality (Excoffier & Lischer, 2010; Fu, 1997; Hudson, 1991). Fu's test can be moreover described as a method to assess an excess versus limited number of alleles, with respect to a "mutation-drift equilibrium expectation" (Fuselli *et al*., 2003, p. 1683). A positive outcome indicates a limited number of alleles, while a negative outcome indicates an excessive number of alleles. Either of these results may arise due to selection or due to deviations from the assumption of constant population size. In the absence of selection, a positive F_S value is expected if a population bottleneck or founding event has occurred. A negative F_S value is expected if a population has experienced a period of expansion, or alternatively, may be due to genetic hitchhiking (Ashfaq *et al*., 2014).

Chakraborty's test of population amalgamation (neutrality/homogeneity test) also assumes the infinite-alleles model (Excoffier & Lischer, 2010). It is used to measure the "probability of observing a random neutral sample with a number of alleles similar or larger than the observed value ([Pr($K \geq k_{obs}$)])…given the mutation parameter inferred from the homozygosity

([θ_{Hom}])" (Excoffier & Lischer, 2010, manual p. 144). Simply put, it is the reported likelihood of observing a given number of alleles in a DNA segment, given the calculated homozygosity (Excoffier & Lischer, 2010).

A mismatch distribution analysis (demographic expansion model) was performed using Arlequin 3.5 on the total Yurimaguas sample; 90%, 95%, and 99% confidence intervals were generated for the mismatch distribution under the null hypothesis of demographic expansion (*Figure 5.16*). Distributions were plotted using SPSS 25.0. A sum of square deviations (SSD) and Harpending's raggedness index (*rg*) was calculated, each with their associated *P*-values. A mismatch distribution can be defined as a comparison of the observed distribution of nucleotide differences between haplotypes, versus an expected distribution of nucleotide differences between haplotypes (Excoffier & Lischer, 2010). Smooth, unimodal distributions are generally found among populations that have recently undergone a demographic expansion. Ragged distributions (and those with statistically significant *rg* indices) are generally found among populations in demographic equilibrium, reflecting the stochastic shape of gene trees (Excoffier & Lischer, 2010; Rogers & Harpending, 1992; Slatkin & Hudson, 1991). This association, however, is not a steadfast rule; one must consult the SSD *P*-value for a statistical comparison of "observed" and "expected" curves. Mismatch distributions were calculated according to Schneider & Excoffier (1999). SSD is a measure of goodness-of-fit between expected and observed distributions, "under the hypothesis that the estimated parameters are the true ones, by simulating *B* [defined below] samples around the estimated parameters" (Excoffier & Lischer, 2010, manual p. 123). A statistically significant SSD value signifies that the "observed" curve deviates significantly from the curve that is "expected" under the expansion model. Therefore, a significant SSD value indicates that the "observed" distribution does not fit the expansion model. Demographic expansion parameters, τ, θ_0, and θ_1 were estimated from a generalized, non-linear, least-square approach (Excoffier & Lischer, 2010; Schneider & Excoffier, 1999). Mismatch distribution provides insight into the population history of greater Yurimaguas. It allows for the potentially diagnosis of populational subpatterning if mismatch distributions of constituent

haplogroups are compared to the mismatch distribution of the total sample. It reflects past demographic processes of genetic drift/founding events, gene flow, amalgamation, and natural selection.

Mismatch distribution *SSD P*-value:

$$P = \frac{\text{number of } SSD_{sim} \text{ larger or equal to } SSD_{obs}}{B}$$

(where *SSD* represents sum of square deviations, and *B* represents a large number of random samples based on a coalescent algorithm [set by software], according to estimated demography [Schneider & Excoffier, 1999])

Harpending's raggedness index (*rg*):

$$r = \sum_{i=1}^{d+1} (x_i - x_{i-1})^2$$

(where *d* represents the "maximum number of observed differences between haplotypes," and *x* represents the observed relative frequencies of the mismatch classes" [Excoffier & Lischer, 2010, manual p. 124; Harpending, 1994])

Research Question 4: Does the distribution of maternal continental-ancestries differ significantly according to demographic variables? Likewise, does the distribution of paternal continental-ancestries differ significantly according to demographic variables?

Five separate multinomial logistic regression analyses were performed using R version 3 software and represented in *Table 5.18*. Multinomial logistic regression is a particular form of logistic regression utilized when a multiclass, dependent variable is categorical in nature, and can be predicted according to a group of independent variables (known as "effects," "explanators," or "features") that may also be categorical (Greene, 2012). As is the case with all forms of logistic regression, independent variables need not be statistically independent of each other (Sperandei,

2014). In *Table 5.18*, maternal continental-ancestry was modeled as the dependent variable, and age interval, sex, language type, history of migration, and living environment, were respectively modeled as the "effects" or independent variables of the five regression analyses. For each regression a goodness-of-fit calculation, "*G*" was provided, in addition to the chi-square calculation and degrees of freedom. Multinomial logistic regression was used to assess the contribution of various demographics on uniparental ancestry. It allows "effects" to be assessed simultaneously and is ideal in the case of *Table 5.18* since many of the recorded demographic variables are not mutually exclusive.

Fisher's exact tests were performed using SPSS 25.0 for *Tables 5.19* and *5.20* to assess association between demographic variables and continental-ancestry, both maternal and paternal, respectively. Demographic categories included sex (tested against maternal continental-ancestry only), language type, living environment, and age interval. All subcategories were mutually exclusive and exhaustive. Despite statistical significance being noted between language type and paternal continental-ancestry (*Table 5.20*; *P* = 0.048*), no post hoc analysis was necessary because the comparison was pairwise to begin with (autochthonous vs. no autochthonous X Indigenous vs. non-Indigenous).

Research Question 5: Does the distribution of maternal continental-ancestries differ significantly according to migratory variables? Likewise, does the distribution of paternal continental-ancestries differ significantly according to migratory variables?

Three separate multinomial logistic regression analyses were performed using R version 3 software and represented in *Table 5.21*. Maternal continental-ancestry was modeled as the dependent variable, and number of migration episodes, migration duration, and migration distance were respectively modeled as the "effects" or independent variables of the three regression analyses. For each regression a goodness-of-fit calculation, "*G*" was provided, in addition to the chi-square calculation and degrees of freedom. Multinomial logistic regression was

used to assess the contribution of various migratory categories on uniparental ancestry. As in *Table 5.18*, it allows "effects" to be assessed simultaneously.

Tables 5.22 and *5.23* analyze the association between migratory variables and maternal and paternal continental-ancestry, respectively. Specifically, a Pearson's test was performed using SPSS 25.0 to assess the association between history of migration and continental-ancestry. A Kruskal-Wallis *H* (mean rank) test was performed using SPSS 25.0 to assess the association between number of migration episodes and continental-ancestry. Lastly, two one-way ANOVAs were performed using SPSS 25.0 to assess, independently, the association between both migration distance and migration duration, with continental-ancestry. All migratory and ancestry subcategories were mutually exclusive and exhaustive.

Research Question 6: Is there evidence to suggest that demographic subdivisions of the Yurimaguas sample can be genetically detected at the nucleotide level using mitochondrial DNA sequence data? Is there evidence to suggest that migration-based subdivisions of the Yurimaguas sample can be genetically detected?

Phi$_{ST}$ (Φ_{ST}), an analog of Sewell Wright's fixation index (F_{ST}), was designed to detect statistical microdifferentiation in mitochondrial DNA sequences. Statistically significant differentiation represents non-interbreeding populations, and the possibility of evolutionary change (Excoffier *et al*., 1992; Wright, 1978). Pairwise genetic distance and phi-statistic (Φ_{ST}) calculations were generated (at the nucleotide level) across demographic (*Table 5.24*: Sex, *Table 5.25*: Language type, *Table 5.26*: Living environment, and *Table 5.27*: Age interval) and migratory (*Table 5.28*: History of migration, *Table 5.29*: Number of migration episodes, *Table 5.30*: Migration duration, *Table 5.31*: Migration distance, and *Table 5.32*: Three locales most commonly migrated to) population subdivisions. A Φ_{ST} calculation and its associated *P*-value, as well as mean number of pairwise differences between categories (Π_{XY}) and corrected mean number of pairwise differences between categories (Corrected Π_{XY}) was provided for each subcategory comparison. Additionally, a mean number of pairwise differences within category calculation (Π_X)

was provided for each table with a single comparison (*Tables 5.24, 5.25, 5.26*, and *5.28*) in the table itself, and in the caption for tables with more than a single comparison. A Bonferroni correction was applied to all Φ_{ST} *P*-values in tables that included more than one comparison (*Tables 5.27, 5.29, 5.30, 5.31,* and *5.32*).

The null hypothesis of the Φ_{ST} calculation is that there is no differentiation in DNA sequences between populational subdivisions. Significant *P*-values represent subpopulations that can be detected statistically. *P*-values are grounded in a permutational approach (10,000 permutations), rather than an assumption of normality (Excoffier & Lischer, 2010). Phi-statistic (Φ_{ST}) calculations can elucidate the microdifferentiation process (specifically, when, and according to what variable(s) does differentiation occur), eventually providing insight into the formation of novel lineages through nonrandom mating and the introduction of variation through gene flow. They can moreover inform us of when a population may be in panmixis. Phi-statistics (Φ_{ST}) are similar to other gene frequency-based ANOVA approaches, however phi-statistics (Φ_{ST}) additionally account for number of mutations separating molecular haplotypes, defined as follows for a pair of haplotypes "x" and "y":

$$\hat{d}_{xy} = \sum_{i=1}^{L} \delta_{xy}(i)$$

(where *L* gives the length of the haplotype (number of nucleotide loci) and $\delta_{xy}(i)$ is a Kronecker function that equals 1 if the nucleotides at the *i*-th locus are identical for haplotypes "x" and "y", or otherwise equals 0 [Excoffier & Lischer, 2010; Excoffier, Smouse, & Quattro, 1992]).

Storage of Genetic Data

Genetic information will be stored for future research purposes on an encrypted database administered by the Information Technology (IT) Department of the University of Kansas. Data of this kind is exceedingly rare in the Upper Amazon and can act as an important guidepost for South American uniparental marker research in the future. Secure data storage will furthermore

permit others to confirm results. Interested parties will be granted access to deidentified data after having their credentials confirmed by a member of the LBA, and upon agreeing to the final stipulations of the approved IRB application (STUDY141454). Uploading genetic information to an open-source platform such as Genbank® is not recommended in this particular situation (Benson *et al*., 2013). Open-source data can be used for a plethora of different forms of genetic research. Consent was granted from study participants and approved by the University of Kansas Institutional Review Board, Human Subjects Committee Lawrence, and the *Gobierno Regional de Loreto, Dirección General de Salud* for a narrow range of research related to migration, population diversity, and disease susceptibility. Moreover, given the explicit mentioning of community names and Indigenous populations in this dissertation, unapproved future research could inadvertently compromise the privacy of participants, or bring negative attention to community members.

Chapter V: Results

Demographic and Migratory Trends in the Yurimaguas Sample

Figure 5.1. Age and sex distribution.

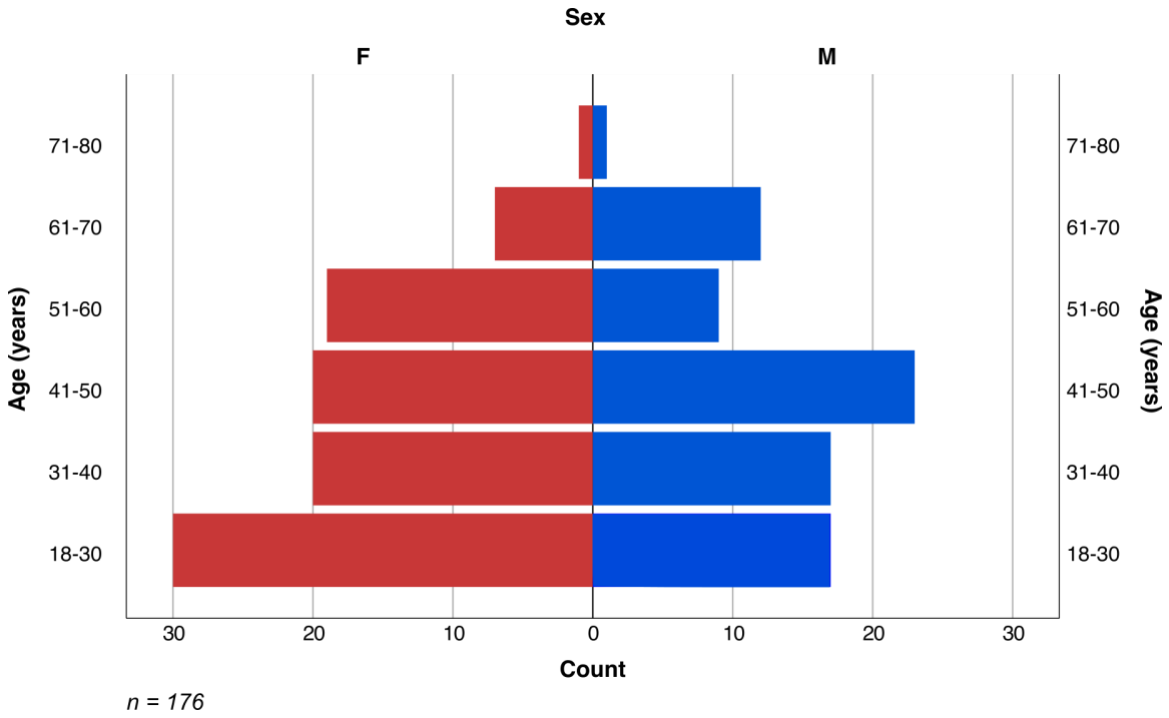


Figure 5.1 depicts the distribution of study participants' age and sex. Among 182 sampled individuals, 79 (43.3%) identified as male, 97 (53.3%) identified as female, and six (3.3%) chose not to identify as either male or female. Those that did not identify with a masculine or feminine identity were excluded from *Figure 5.1*, and all further analyses. The overall mean age of the sample population was 41.5 years (males: 42.8 years; females: 40.7 years). The youngest study participants were 18 years old (*n* = 2), while the oldest was 78. Generally, the study sample depicted a distribution of sex and age supported by recent census data (INEI, 2007).

Table 5.1. Nested sex, language type, living environment, and age distribution.

← Nested Demographics →					Count	Table *n* %
Sex	Language Type	Living Environment	Age (years)			
Female	No Autochthonous	Peri-urban	18-30	8	4.8%
			31-40	3	1.8%
			41-50	3	1.8%
			51-60	3	1.8%
			61-70	2	1.2%
		Urban	18-30	21	12.7%
			31-40	13	7.9%
			41-50	12	7.3%
			51-60	10	6.1%
			61-70	2	1.2%
			71-80	1	0.6%
	Autochthonous	Peri-urban	61-70	1	0.6%
		Urban	31-40	2	1.2%
			41-50	2	1.2%
			61-70	1	0.6%
Male	No Autochthonous	Peri-urban	18-30	2	1.2%
			31-40	1	0.6%
			41-50	2	1.2%
			51-60	1	0.6%
			61-70	3	1.8%
		Urban	18-30	12	7.3%
			31-40	15	9.1%
			41-50	17	10.3%
			51-60	7	4.2%
			61-70	8	4.8%
	Autochthonous	Peri-urban	71-80	1	0.6%
		Urban	41-50	2	1.2%
			61-70	1	0.6%

n = 156

Table 5.1 is an accounting of nested demographics, including sex (female vs. male), language type (those who do not speak an autochthonous language vs. those who speak at least one autochthonous language) living environment (peri-urban vs. urban), and age interval, within the study sample. A frequency “Count” and a “Table *n* %” (percent that a nested category count comprises the entire table count) is provided for each row, emphasizing the demographic frequency of each specific crosstabulated category.

Figure 5.2. Languages spoken.

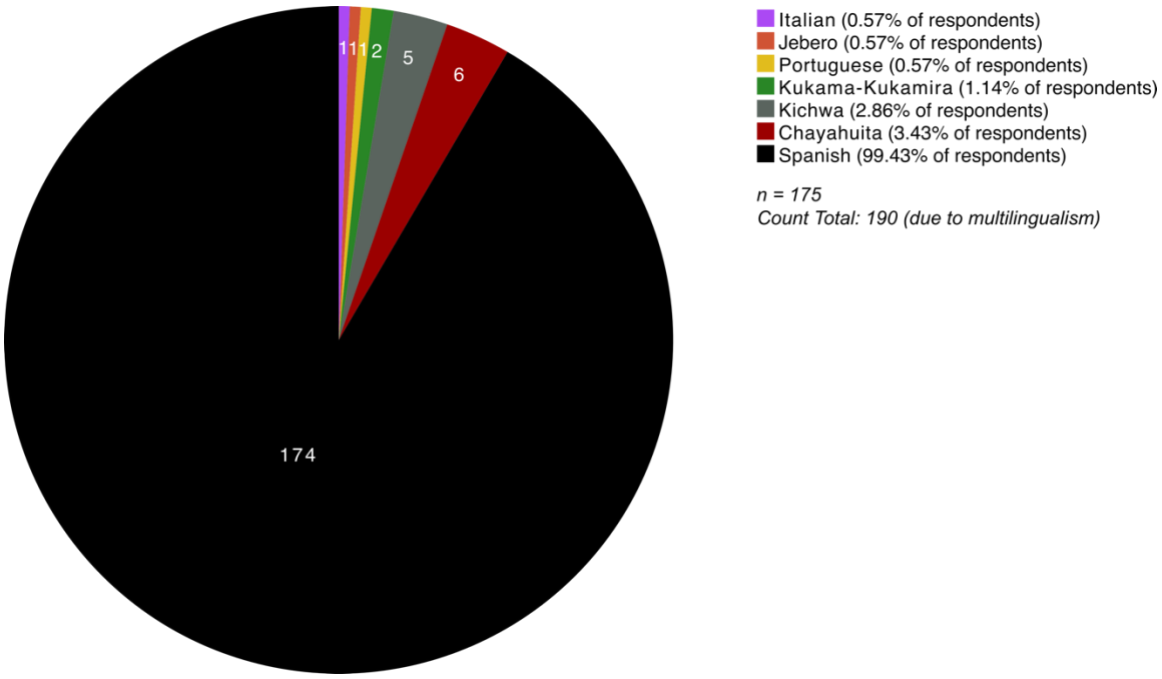
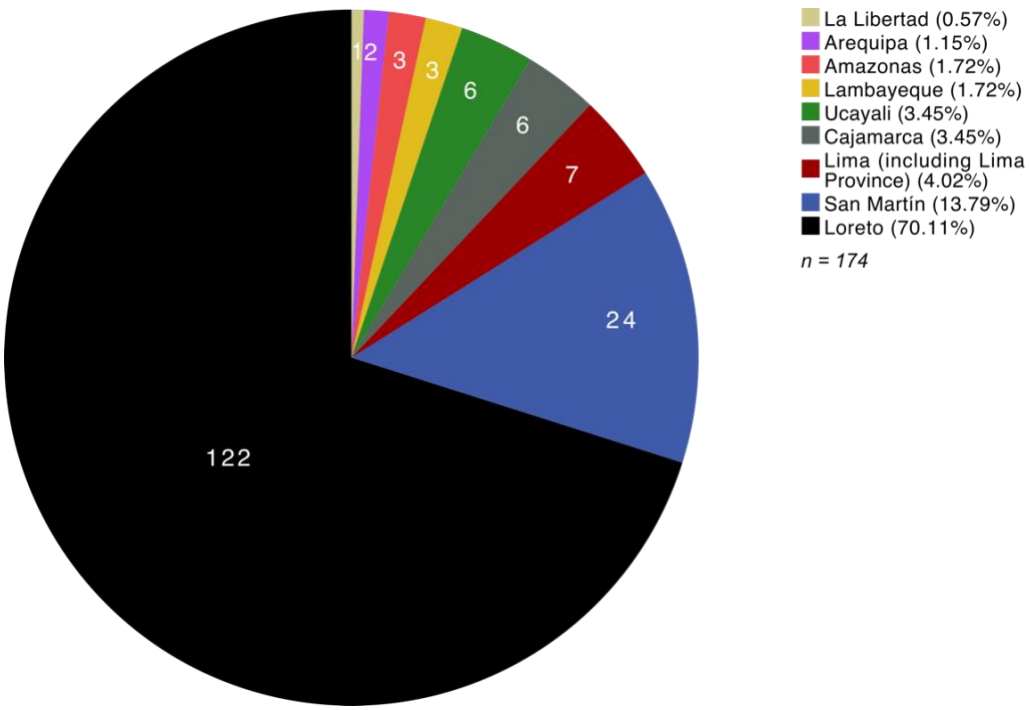


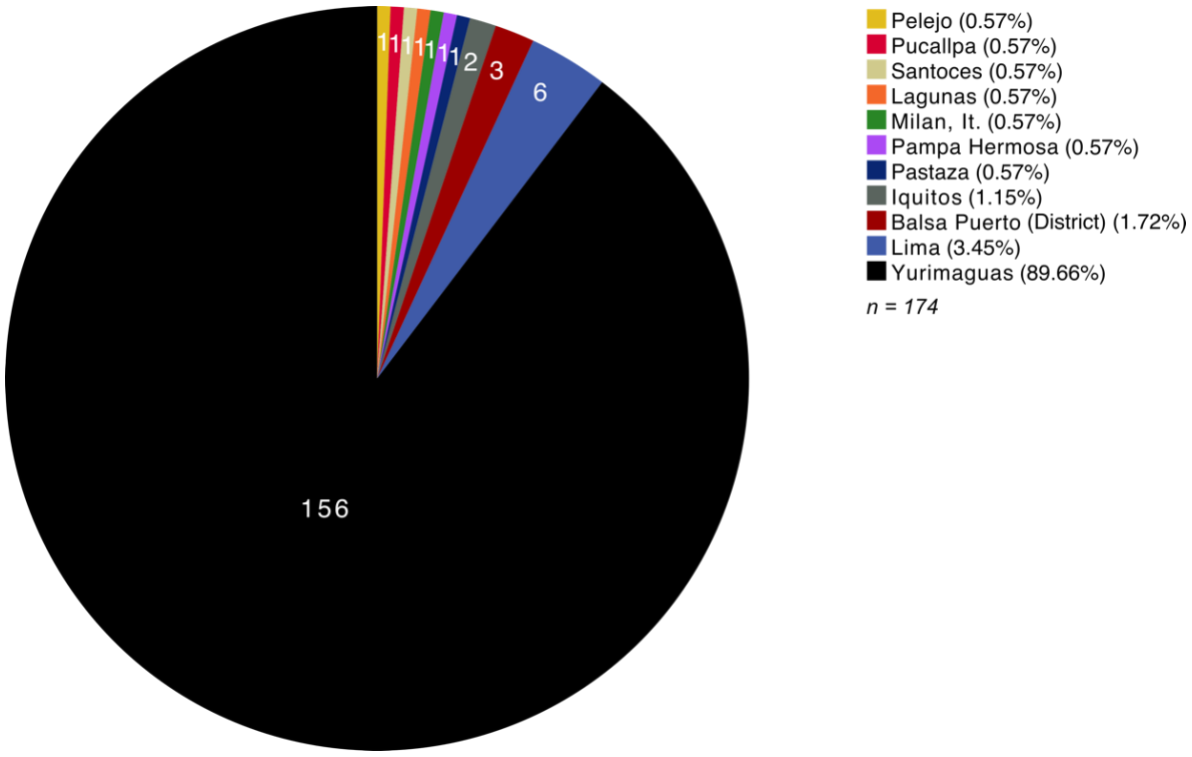
Figure 5.2 presents the languages spoken by study participants. Thirteen of 175 individuals (7.4%) speak an autochthonous American language (one individual spoke both Kichwa and Chayahuita). Amongst autochthonous language speakers there is significant linguistic diversity, with four different tongues represented: Jebero, Kukama-Kukamira, Kichwa, and Chayahuita. Only one individual spoke solely an autochthonous language, which was Chayahuita. All other autochthonous language speakers additionally speak a non-autochthonous language, typically Spanish (the term “Spanish” when referring to language, denotes the Castilian [*Castellano*] language, as opposed to the regional languages of Spain, such as Catalan [*Catalá*] or Basque [*Euskara])*. Only two individuals spoke a non-autochthonous language other than Spanish; these were Italian and Portuguese. Both of these individuals also spoke Spanish. Thirteen individuals were bilingual, and one was trilingual. Because only one study participant did not speak Spanish, 99.4% of the study sample are classified as Hispanophones. Typically, if an individual spoke both Spanish and an autochthonous American language, the autochthonous language was the natal tongue.

Figure 5.3. Regions of birth.



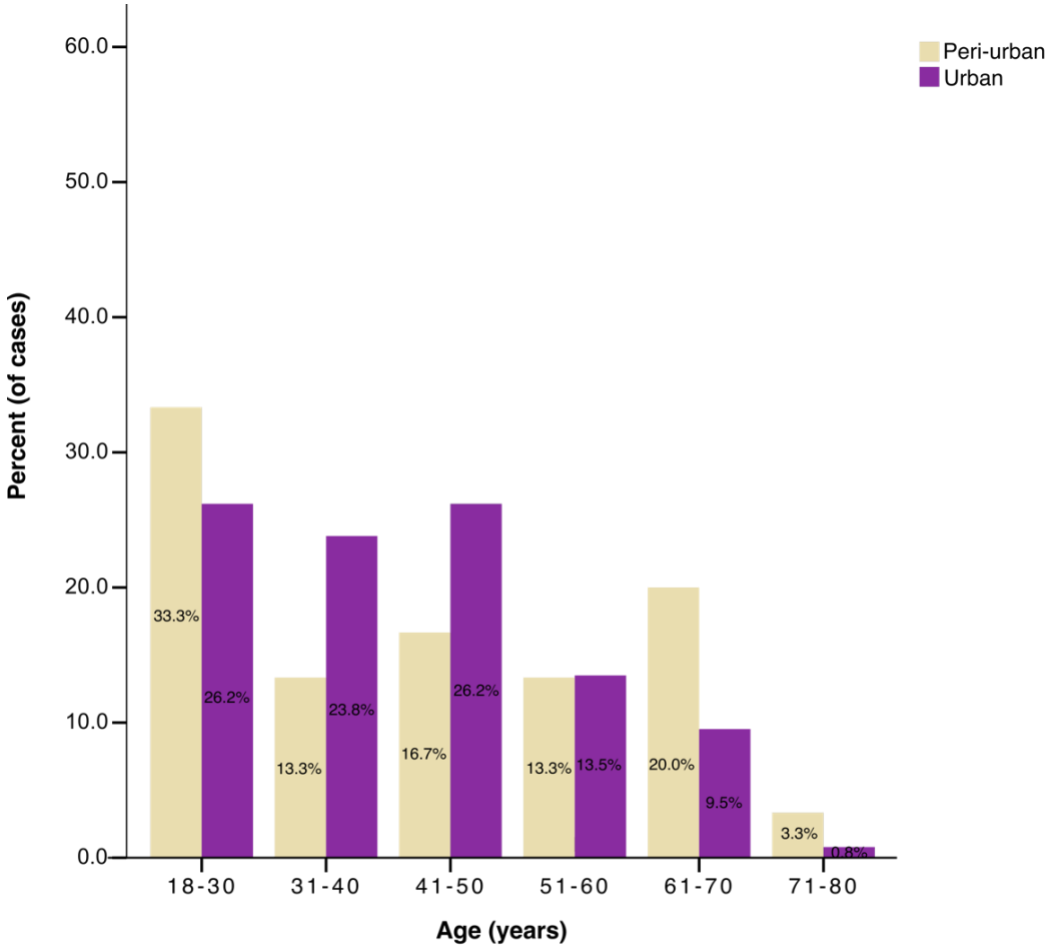
Approximately 70% of study participants were born in Loreto Region. An additional ~14% were born in neighboring San Martín Region. The remaining ~16% of respondents were born in 7 of 23 other Peruvian Regions, namely the coastal Regions due west of Loreto (La Libertad and Lambayeque), the Andean Regions immediately west or south of Loreto (Amazonas, Cajamarca, and Ucayali), or in the Regions that include the major conurbations of Arequipa (pop. 920,047) or Lima (pop. 10,072,000) (INEI, 2012). The Lima metropolitan area, the largest urban agglomeration in Peru, comprises an estimated 40.4% of the total Peruvian populace (United Nations, 2016).

Figure 5.4. Majority-time cities of residence.



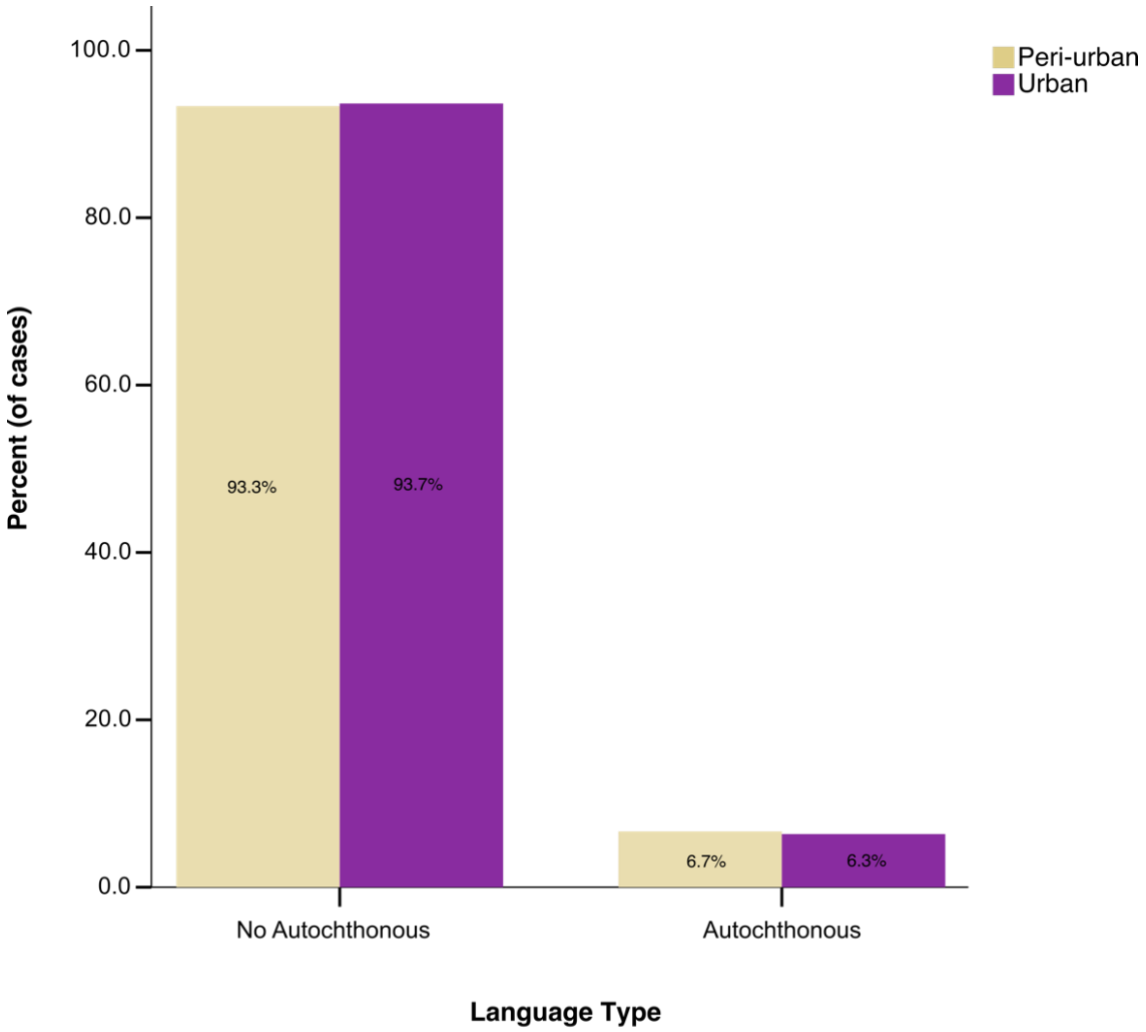
Despite the fact that the present research took place in Yurimaguas District, not all study participants were majority-time residents of the greater Yurimaguas area. Approximately 10% of the study sample reported to reside outside of the Yurimaguas metropolitan area, in locales such as Lima (3.45%), Balsa Puerto District (1.72%), and Iquitos (1.15%) greater than 50% of the time. Many individuals in Upper Amazonia engage in cyclical migration, rotating between residing in two or more locales. Some individuals never spend a majority of their time in any one locale if their migratory cycle encompasses more than two municipalities. All participants of this study reported spending over 50% of their time in a single locality. Cyclical migration patterns have the potential to confound certain population genetic metrics that assume that defined (closed) populations are being sampled.

Figure 5.5. Living environment (peri-urban vs. urban), by age interval.



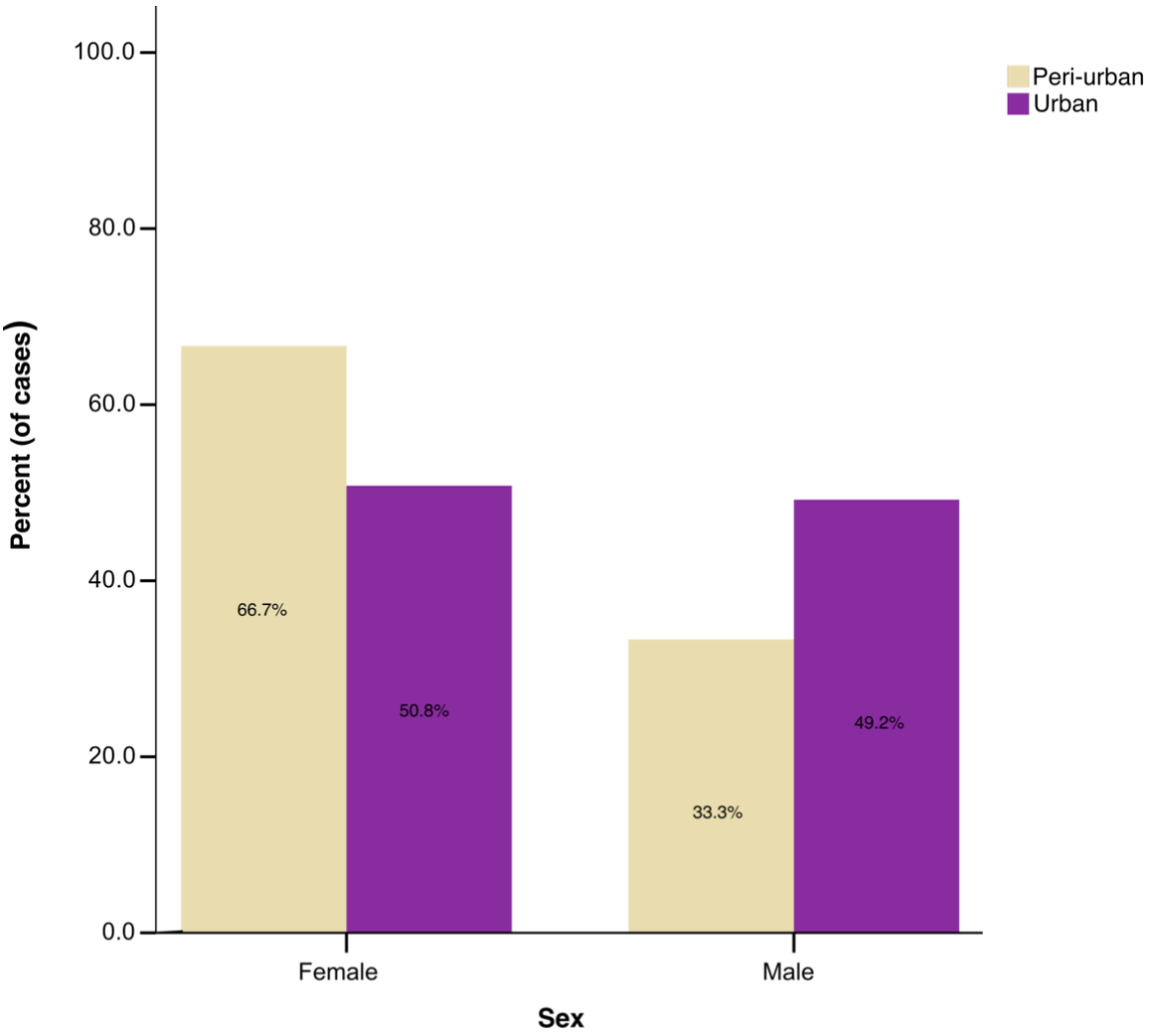
n = 156
Fisher's exact test of independence between age interval and living environment (peri-urban vs. urban), value: 6.399, exact significance (2-sided): P = 0.248. No significant association found.

Figure 5.6. Living environment (peri-urban vs. urban), by language type.



n = 156
Pearson's chi-square test for independence between language type and living environment (peri-urban vs. urban), X^2 = 0.004, df: 1, exact significance (2-sided): P = 1.000. No significant association found.

Figure 5.7. Living environment (peri-urban vs. urban), by sex.



n = 156
Pearson's chi-square test for independence between sex and living environment (peri-urban vs. urban, X^2 = 2.457, df: 1, exact significance (2-sided): P = 0.154. No significant association found.

Figures 5.5, *5.6*, and *5.7* illustrate the association between living environment (those that live in the peri-urban community of Munichis versus the urban city of Yurimaguas), and age interval, language type, and sex, respectively. Among the figures, the subcategories of language type (*e.g*. no autochthonous language and autochthonous language), age interval (*e.g.* 18–30, 31–40, 41–50, 51–60, 61–70, and 71–80), and sex (*e.g.* female and male), independently add to 100%. Consequently, the crucial patterns in each figure are those across a given categorical

descriptor (*i.e.* all of those categorical descriptors within the peri-urban environment *or* all of those categorical descriptors within the urban environment), rather than a comparison within the categorical descriptors themselves.

Figure 5.5 depicts a decreasing proportion of individuals of advanced age compared with younger individuals, in both peri-urban Munichis and urban Yurimaguas. This is at least partially due to an increased likelihood of death with increasing age. Munichis possesses a greater percentage of individuals at age extremes compared with Yurimaguas, which has a greater percentage of intermediate-aged residents. Given the absence of statistical significance, this is likely due to sampling aberrations. *Figure 5.6* depicts both peri-urban Munichis and urban Yurimaguas having nearly the same ratio of individuals that do not speak an autochthonous language and those that do (~94% and ~6%, respectively, in both living environments). Evidence suggests a near complete lack of association between language type and living environment. *Figure 5.7* shows that the study sample of Munichis has a greater female to male ratio than Yurimaguas. The ratio within Munichis is approximately 2:1, female to male, while it is approximately 1:1 in Yurimaguas. This finding makes for an interesting trend, although again, it is not statistically significant ($P = 0.086$).

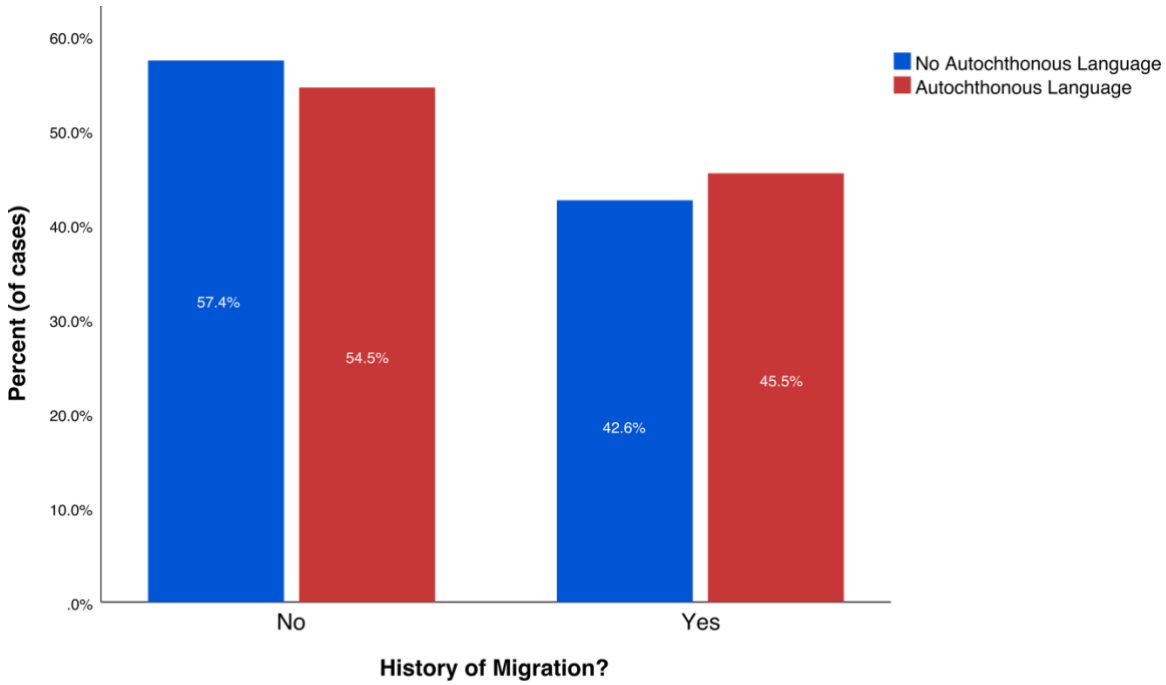
Table 5.2. Likelihood of migration history, by birth Region.

			History of Migration?		
			Yes	No	Total
Birth Region	Amazonas	Count:	2	1	3
		Percentage:	66.7%	33.3%	100.0%
	Arequipa	Count:	2	0	2
		Percentage:	100.0%	0.0%	100.0%
	Lima (including Lima Province)	Count:	2	5	7
		Percentage:	28.6%	71.4%	100.0%
	Cajamarca	Count:	3	3	6
		Percentage:	50.0%	50.0%	100.0%
	La Libertad	Count:	0	1	1
		Percentage:	0.0%	100.0%	100.0%
	Lambayeque	Count:	2	1	3
		Percentage:	66.7%	33.3%	100.0%
	Loreto	Count:	52	75	127
		Percentage:	40.9%	59.1%	100.0%
	San Martín	Count:	11	14	25
		Percentage:	44.0%	66.0%	100.0%
	Ucayali	Count:	4	1	5
		Percentage:	80.0%	20.0%	100.0%
Total:		Count:	78	101	179

Fisher's exact test of independence between history of migration (yes vs. no) and Region of birth, value: 8.373, exact significance (2-sided): P = 0.369. No significant association found.

Table 5.2 illustrates the frequency ("Count" and row "Percentage") of having experienced at least one migration episode, according to one's birth Region. Approximately 41% of those born in Loreto have a history of migration. Approximately 44% of the native-born population of neighboring San Martín Region have a history of migration. All of the other listed Regions have relatively low *n* values, therefore their reported results should be interpreted with reservation. An omnibus Fisher's exact test reported no statistical significance.

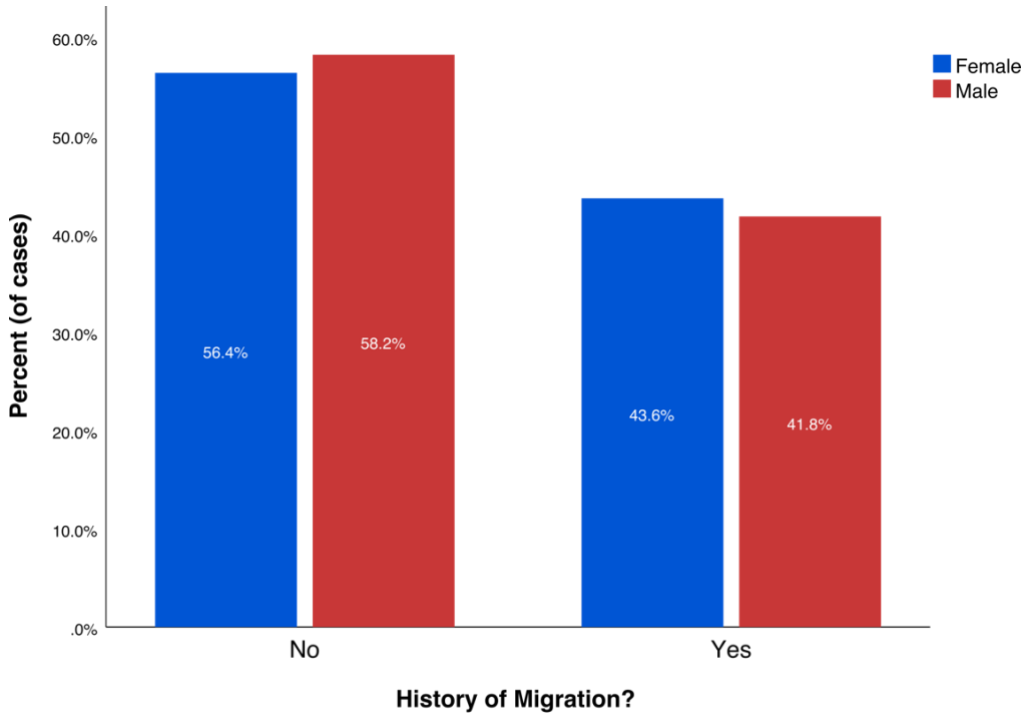
Figure 5.8. Likelihood of migration history, by language type.



n = 173

Pearson's chi-square test for independence between history of migration (yes vs. no) and language type, X^2 = 0.034, df: 1, exact significance (2-sided): *P* = 1.00. No significant association found.

Figure 5.9. Likelihood of migration history, by sex.



n = 173
Pearson's chi-square test for independence between history of migration (yes vs. no) and sex, X^2 = 0.060, df: 1, exact significance (2-sided): *P* = 0.878. No significant association found.

Figure 5.8 portrays the likelihood that study respondents participated in at least one migration episode, by language type. There is no discernable trend between migration history and language type, as defined herein. Instead, there appears to be near equal rates of migration history between autochthonous language speakers and those that do not speak an autochthonous language. *Figure 5.9* illustrates the likelihood that study respondents participated in at least one migration episode, by sex. Like *Figure 5.8*, there is no observable trend between female and male proportions of migration. Approximately equal ratios of females and males have a history of migration.

Table 5.3. Quantitative migration statistics.

Distance Between Birthplace and Current Residence (Km)		Migration Duration (Frequency Count)					
Mean	Standard Deviation	1-5 years	6-10 years	11-15 years	16-20 years	21-25 years	26-30 years
194	404	32	5	9	3	1	1

Number of Migratory Episodes		Distance Between Migration Locale(s) and Birthplace (Km)		Migration Distance (Km)	
Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
2	1	649	394	646	410

All distances were calculated by default according to the Google Maps automobile feature (Map Data: Google, 2018a). If roadway travel was not possible, direct orthodromic distance was was calculated instead (Map Data: Google, 2018b).

Table 5.3 provides various migration statistics *amongst* those that reported having engaged in at least one migration episode. Respondents were found to travel hundreds of kilometers for education and/or employment opportunities. The typical migrant has relocated to two different locales in their lifetime (both of which they spent a minimum of 1 year in), and most frequently spent between one- and five-years total, as a migrant. “Migration Distance (Km),” the last category of *Table 5.3*, was further partitioned by sex, language type, living environment, and age in *Tables 5.5* and *5.6*.

Table 5.4. Number of migration episodes by sex, language type, living environment, and age interval.

		Number of Migration Episodes										Total	
		1		2		3		4		7			
		Count	Row Total *n* %	Count	Row Total *n* %	Count	Row Total *n* %	Count	Row Total *n* %	Count	Row Total *n* %	Count	Row Total *n* %
Sex (*P* = 0.080)	Female	26	70.3	6	16.2	3	8.1	2	5.4	0	0.0	37	100.0
	Male	13	44.8	10	34.5	4	13.8	1	3.4	1	3.4	29	100.0
Language Type (*P* = 0.411)	No Autochthonous	37	60.7	14	23.0	6	9.8	3	4.9	1	1.6	61	100.0
	Autochthonous	2	40.0	2	40.0	1	20.0	0	0.0	0	0.0	5	100.0
Living Environment (*P* = 0.153)	Peri-urban	13	76.5	2	11.8	1	5.9	0	0.0	1	5.9	17	100.0
	Urban	24	52.2	13	28.3	6	13.0	3	6.5	0	0.0	46	100.0
Age Interval (years) (*P* = 0.756)	18-30	13	59.1	5	22.7	3	13.6	1	4.5	0	0.0	22	100.0
	31-40	12	66.7	3	16.7	2	11.1	1	5.6	0	0.0	18	100.0
	41-50	6	50.0	4	33.3	1	8.3	1	8.3	0	0.0	12	100.0
	51-60	3	75.0	1	25.0	0	0.0	0	0.0	0	0.0	4	100.0
	61-70	4	44.4	3	33.3	1	11.1	0	0.0	1	11.1	9	100.0
	71-80	1	100.0	0	0.0	0	0.0	0	0.0	0	0.0	1	100.0

Four Kruskal-Wallis H (mean rank) tests were performed to assess association between number of migration episodes (dependent variable) and (1) sex, (2) language type, (3) living environment,and (4) age interval (years). No significant association was found.

Table 5.4 presents the number of migration episodes that study respondents (who reported having engaged in migration) experienced, across the demographic variables of sex, language type, living environment, and age interval. Each row provides a frequency "Count," as well as a "Row Total *n* %." All included study participants reported one through four migration episodes, except one individual, who reported seven. Across all categories, the trend was a decreasing frequency of individuals who experienced a greater number of migration episodes. Generally, males experienced a disproportionately greater number of migration episodes than females. This trend approaches but fails to reach statistical significance (H: 3.060, df: 1, *P* = 0.080). Language type (H: 0.676, df: 1, *P* = 0.411), living environment (H: 2.040, df: 1, *P* = 0.153) and age interval (H: 2.636, df: 5, *P* = 0.756) do not exhibit statistically significant associations with number of migration episodes either.

Table 5.5. Migration distance by sex, language type, living environment, and age interval (nested).

Nested Demographics					Migration Distance (Km)		
					Mean	Standard Deviation	Row *n*
Sex	Language Type	Living Environment		Age Interval (years)			
Female	No Autochthonous	Peri-urban		18-30	1,076	454	8
				31-40	805	10	3
				51-60	798	-	3
				61-70	50	-	2
		Urban		18-30	555	457	20
				31-40	547	345	12
				41-50	548	775	12
				61-70	398	-	2
				71-80	167	-	1
	Autochthonous	Urban		31-40	335	-	2
Male	No Autochthonous	Peri-urban		18-30	448	634	2
				41-50	88	-	2
				61-70	658	85	3
		Urban		18-30	705	338	12
				31-40	189	267	14
				41-50	593	557	16
				51-60	332	-	7
				61-70	789	492	8
	Autochthonous	Urban		41-50	528	113	2
				61-70	250	-	1

0 - 300 Km *301 - 600 Km* *601 - 900 Km* *901 - 1,200 Km*

Table 5.5 depicts migration distance (the last column of *Table 5.3*), partitioned by sex, language type, living environment, and age interval. Each row provides the mean distance,

standard deviation, and *n* value for each nested demographic category. It is important to note that each individual may have engaged in multiple migration episodes, and that the "Row *n*" value corresponds to number of migration episodes in that category, rather than the number of individuals.

Table 5.6. Migration distance by sex, language type, living environment, and age interval.

		Migration Distance (Km)		
		Mean	Standard Deviation	Row *n*
Sex (*P* = 0.800)	Female	669	431	32
	Male	639	424	24
Language Type (*P* = 0.359)	No Autochthonous	669	432	53
	Autochthonous	435	179	3
Living Environment (*P* = 0.047*)	Peri-urban	856	415	13
	Urban	585	420	41
Age Interval (years) (*P* = 0.638)	18-30	777	451	19
	31-40	578	417	16
	41-50	642	469	10
	51-60	643	269	3
	61-70	548	387	8
	71-80	84	118	2

0 - 300 Km *301 - 600 Km* *601 - 900 Km*

Three independent samples t-tests were performed to assess association between migration distance (Km) and (1) sex, (2) language type, and (3) living environment. A one-way ANOVA was performed to assess association between migration distance (Km) and age interval (years). Statistical significance was found for living environment (peri-urban vs. urban).

According to *Table 5.6* there is no statistically significant difference in migration distance by sex (t = –0.254, df: 54, P = 0.800), language type (t = 0.025, df: 54, P = 0.359), or age interval (F = 0.637, df [between groups]: 4, df [within groups]: 51, P = 0.638). When modeled as an independent test, there is, however, a significant difference between those that reside in peri-urban versus urban living environments (t = –2.030, df: 52, P = 0.047*). If an individual experienced more than one migration episode, the mean distance between them was taken. It is important to note that combined demographic subcategories (sex: male + female, language type: no autochthonous + autochthonous, living environment: peri-urban + urban, and age: 18–30 + 31–40 + 41–50 + 51–60 + 71–80) exhibit a slightly different mean distance from the total sample

population, due to study respondents omitting certain demographic questions, and hence being precluded from this analysis.

Table 5.7. Number of emigrants from each locale, to Yurimaguas.

Locale	Count
Bagua Grande, Amazonas	1
Balsa Puerto (District), Loreto	1
Barranquita, San Martín	2
Cajaruro, Amazonas	1
Celendín, Cajamarca	1
Chancay, Lima	1
Chazuta, San Martín	1
Chiclayo, Lambayeque	2
Huimbayoc (District), San Martín	1
Iquitos, Loreto	1
Jaén, Cajamarca	1
Jeberos, Loreto	3
Lagunas, Loreto	10
Lamas, San Martín	5
Lima, Lima Province	3
Moyobamba, San Martín	2
Naranjos, San Martín	1
Nauta, Loreto	1
Nueva Cajamarca, San Martín	1
Pampa Hermosa, Loreto	1
Papaplaya, San Martín	2
Paucarpata, Arequipa	2
Picota, San Martín	1
Pucallpa, Ucayali	4
Requena, Loreto	1
San Lorenzo, Loreto	2
Santa Cruz de Succhabamba, Cajamarca	5
Saquena, Loreto	1
Schucshuyacu, Loreto	1
Soritor, San Martín	1
Tabalosos, San Martín	1
Tarapoto, San Martín	3
Tocache, San Martín	1
Tomas (District), Lima	1
Trujillo, La Libertad	1
Total:	67

Figure 5.10. Choropleth of locales emigrated from (to Yurimaguas).

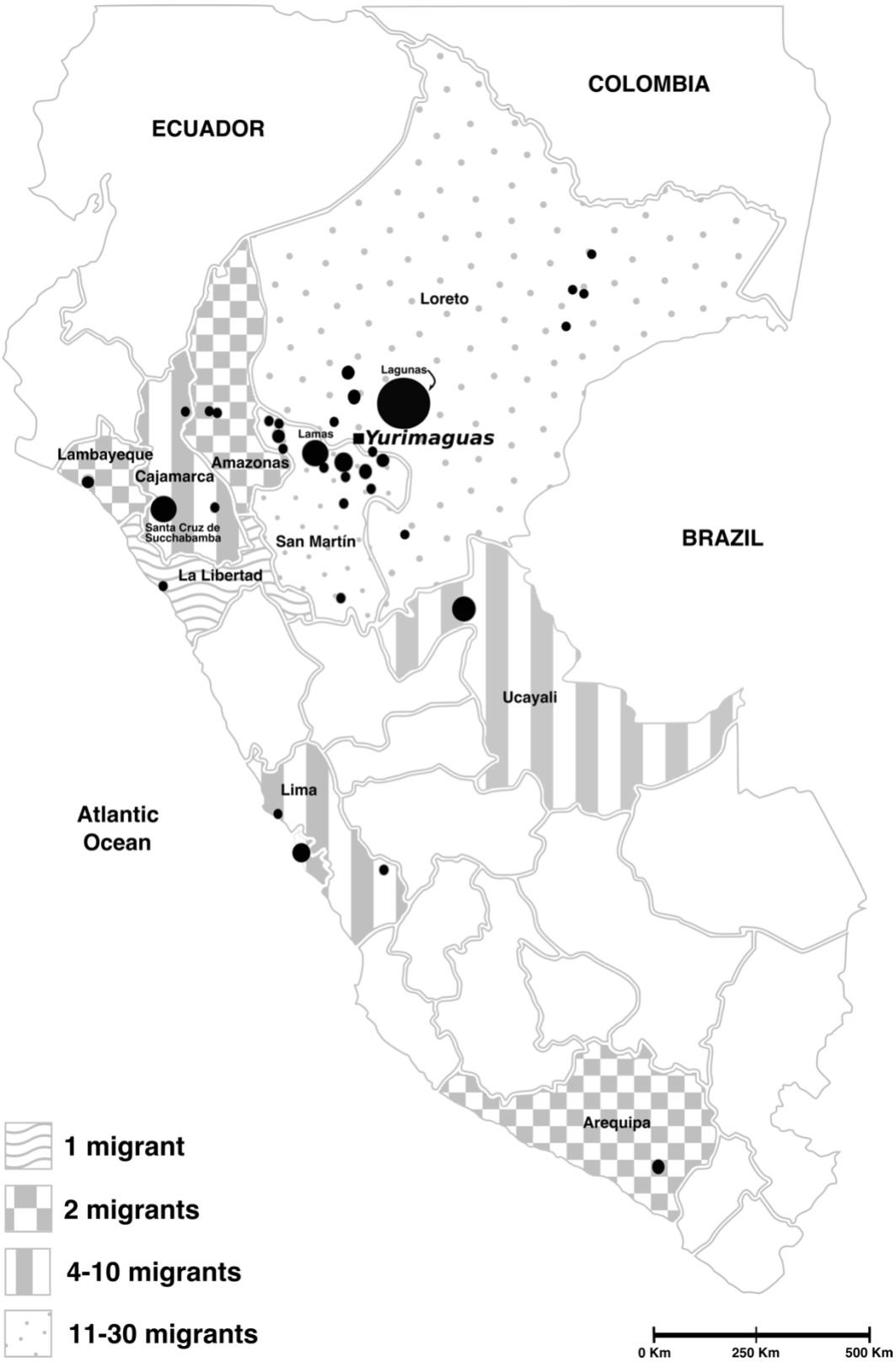


Table 5.7 and *Figure 5.10* display the cities (and in three cases, districts) from which study participants emigrated to Yurimaguas. Sixty-seven emigration episodes were recorded (between 67 individuals) and 35 locales. The three most common locales emigrated from, in descending order, were Lagunas, Loreto Region (*n* = 10; 89 Km orthodromic distance [road travel is not possible]), Lamas, San Martín Region (*n* = 5; 152 Km via roadway), and Santa Cruz de Succhabamba, Cajamarca Region (*n* = 5; 708 Km via roadway). All three are rural villages, associated primarily with Indigenous cultures. The former two are within Peruvian Upper Amazonia. Lagunas is in the *selva baja* to the northeast of Yurimaguas, and Lamas in the *selva alta* to the southwest. Santa Cruz de Succhabamba is in the Andean *sierra* due west of Yurimaguas. The most common Peruvian Regions migrated from, were, in descending order, Loreto Region: 22, San Martín Region: 22, Cajamarca Region: 7, Lima Region (including Lima Province): 5, Ucayali Region: 4, Amazonas Region: 2, Arequipa Region: 2, Lambayeque Region: 2, and La Libertad Region: 1. Each tally corresponds to a migration episode, not an individual. There is no statistically significant association between the Region that one emigrates from to Yurimaguas, and maternal or paternal continental-ancestry (data calculated but not shown).

Table 5.8. Number of emigrants from Yurimaguas, to each locale.

Locale	Count
Andoas, Loreto	1
Arequipa, Arequipa	1
Balsa Puerto (District), Loreto	3
Bellavista, San Martín	1
Cajamarca, Cajamarca	1
Chiclayo, Lambayeque	3
Chulucanas (District), Piura	1
Cusco, Cusco	2
Ica, Ica	2
Iquitos, Loreto	12
Lagunas, Loreto	2
Lambayeque, Lambayeque	1
Lima, Lima Province	24
Moyobamba, San Martín	2
Ecuador (Nation)	1
Italy (Nation)	1
Mórrope, Lambayeque	1
Nauta, Loreto	1
Nueva Cajamarca, San Martín	3
Pisco, Ica	1
Piura, Piura	2
Pucallpa, Ucayali	5
Puno, Puno	1
Requena, Loreto	2
Rioja, San Martín	3
San Fernando, Lima Province	1
San Lorenzo, Loreto	2
Saposoa, San Martín	1
Schucshuyacu, Loreto	1
Shanusi, Loreto	1
Tarapoto, San Martín	15
Tingo María, Huánuco	1
Trujillo, La Libertad	3
Total:	102

Figure 5.11. Choropleth of locales migrated to (from Yurimaguas).

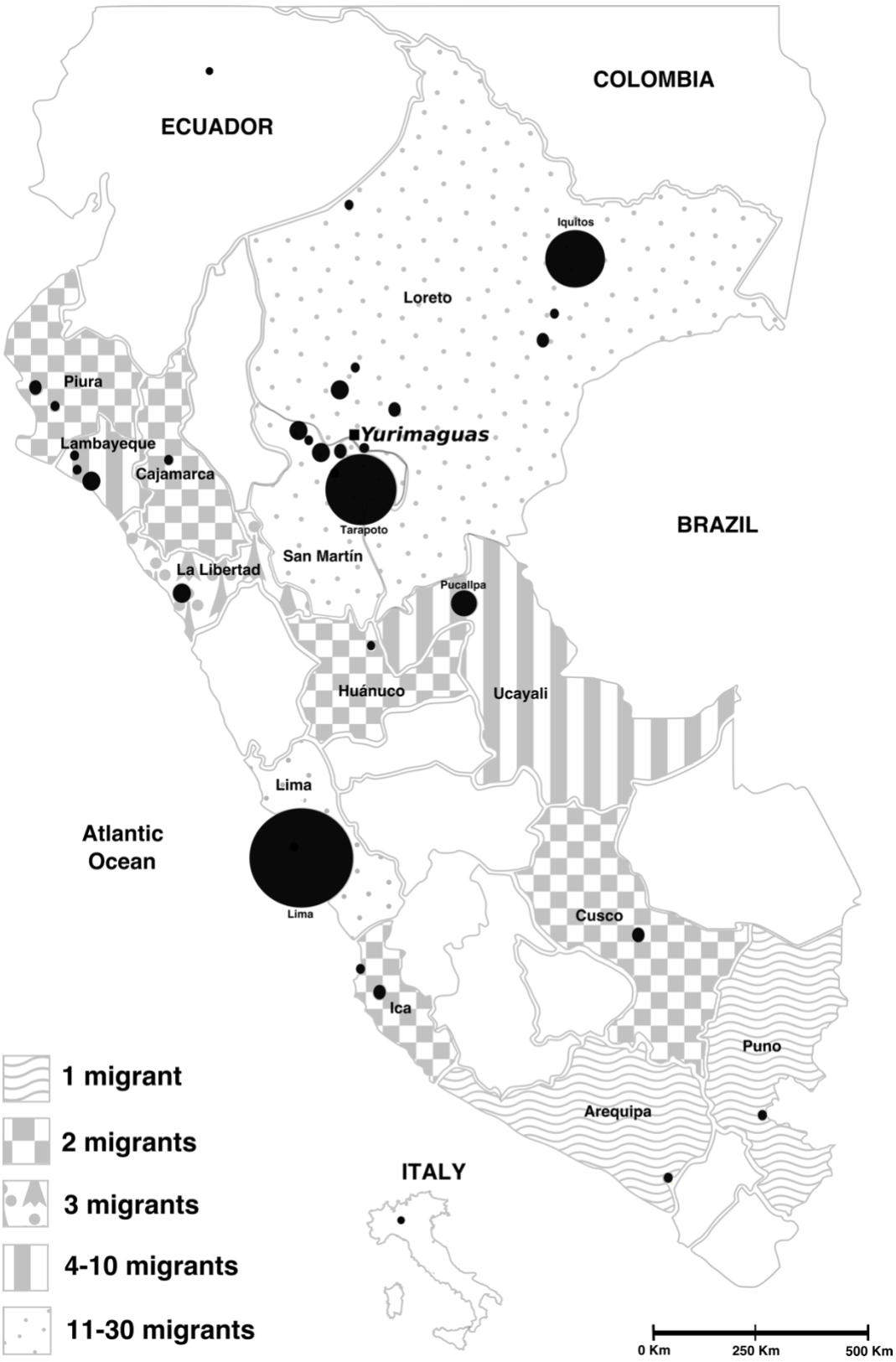
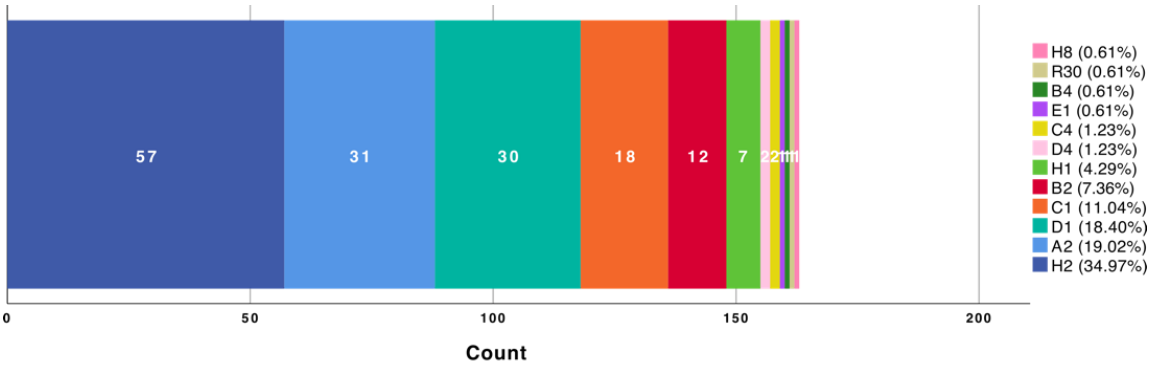


Table 5.8 and *Figure 5.11* display the cities (and in four cases, districts/nations) that study participants migrated to, from Yurimaguas. The dots corresponding to the cities of Bellavista (San Martín Region), San Fernando (Lima Province), and Saposoa (San Martín Region), are hidden behind the dots of prodigious migration locales. One hundred and two migration episodes were recorded between 65 individuals and 33 locales. It is important to note that if one was to calculate the mean number of migration episodes of respondents who reported having engaged in at least one migration episode, based on the aforementioned statistics, it would equal 1.57. This statistic is however artificially low, as a number of study respondents omitted the destination locales of their migration episodes and were hence excluded from the compilation of both *Table 5.8* and *Figure 5.11*. The true mean number of migration episodes among those who reported having engaged in at least one migration episode is 2.0 (see *Table 5.3*). The mean number of migration episodes amongst all study participants is 0.58. Across all study participants, 78 of 179 (43.6%), reported having engaged in at least one migration episode.

The three most common locales emigrated to, were, in descending order, Lima, Lima Province (*n* = 24; 1,096 Km via roadway), Tarapoto, San Martín Region (*n* = 15; 120 Km via roadway), and Iquitos, Loreto Region, (*n* = 12; 397 Km orthodromic distance [road travel is not possible]). All three are major conurbations, the latter two of which are in Peruvian Upper Amazonia. Tarapoto is in the *selva alta,* or intermontane region to the southwest of Yurimaguas, and Iquitos is in the *selva baja* to the northeast. Lima is located on the arid *costa* of west-central Peru. Two individuals sought opportunity outside of Peru, one in Quito, Ecuador and another in Milan, Italy. The most common Peruvian Regions that respondents migrated to were, in descending order, Lima Region (including Lima Province): 26, San Martín Region: 25, Loreto Region: 24, Lambayeque Region: 5, Ucayali Region: 5, Ica Region: 3, La Libertad Region: 3, Piura Region: 3, Cusco Region: 2, Arequipa Region: 1, Cajamarca Region: 1, Huánuco Region: 1, and Puno Region: 1. Each tally corresponds to a migration episode, not an individual. There is no statistically significant association between the Region that one migrates to, from Yurimaguas, and maternal or paternal continental-ancestry distributions (data calculated but not shown).

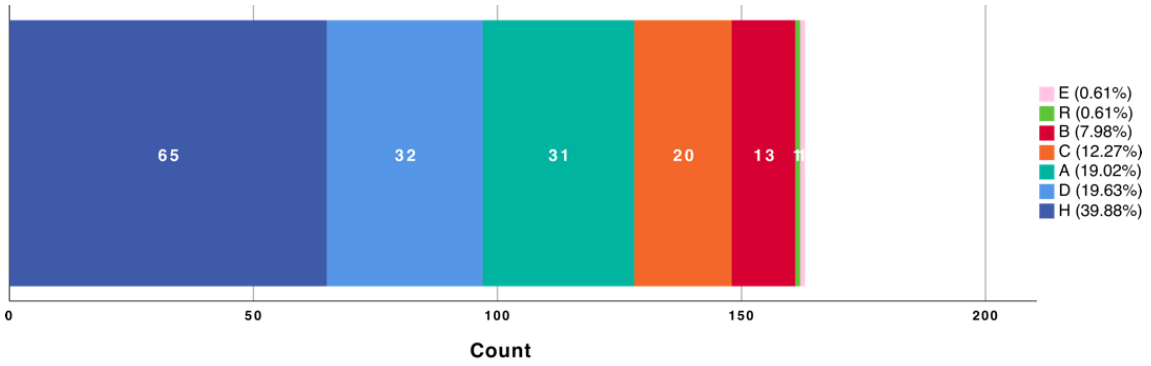
Research Question 1: What is the uniparental marker-based continental-ancestry composition of Yurimaguas? Moreover, is there evidence of sex-skewed gene flow?

Figure 5.12. Mitochondrial primary subclade distribution.



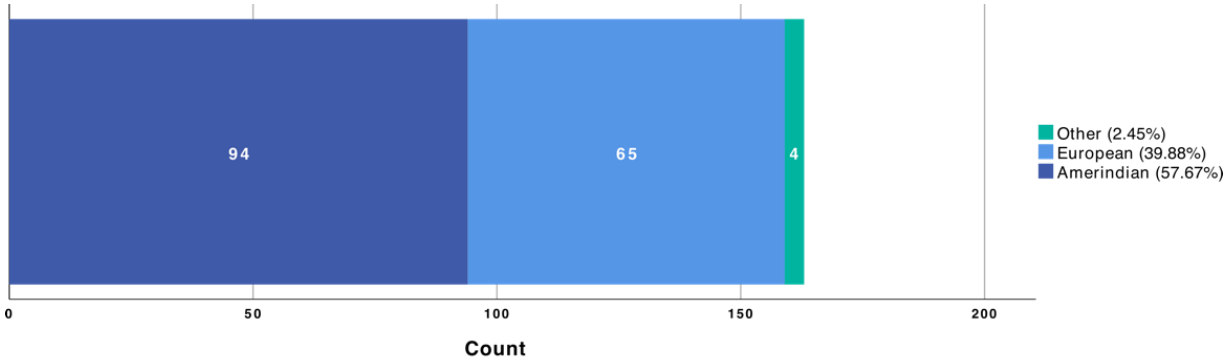
n = 163

Figure 5.13. Mitochondrial haplogroup distribution.



n = 163

Figure 5.14. Maternal continental-ancestry distribution.



n = 163

Seven mitochondrial haplogroups are present among study participants. These are A, B, C, D, E, H, and R. Haplogroup H comprises the plurality, at 39.88% (*n* = 65). Haplogroups E and R are the rarest, each representing only a single individual, or 0.061% of the study sample, each. Haplogroups A, B, C, and D are intermediate in frequency, with proportions ranging from 7.98% to 19.63% of study participants.

There are twelve first-degree subclades (one ancestry-informing marker more specific than haplogroup) found in the study sample. Subclade A2 is found almost exclusively within Indigenous American populations. Small pockets of subclade A2 have been found in the Chukotko Autonomous Okrug and Kamchatka Krai, in far eastern Russia, evidence of early migrations from this region across Beringia (O'Rourke & Raff, 2010; Volodko *et al*., 2008). Indigenous Americans that possess subclade A2 include both Inuit and Na-Dene populations, in addition to Indigenous South Americans (Kumar *et al.*, 2011; Tamm *et al*., 2007). Subclade B2, unlike A2, is found only in Indigenous populations (Just, Diegoli, Saunier, Irwin, & Parsons, 2008; Starikovskaya *et al.*, 2005; Tabbada *et al*., 2010; Tamm *et al*., 2007;). Haplogroup B4 can be divided into subclades B4a and B4b. Within the latter is subclade B2. The former is primarily associated with individuals of eastern Asian origin. The only individual found to have a B4 haplogroup possesses a B4a1c3 haplotype. This particular haplotype is most strongly associated with Japanese populations (Bilal *et al*., 2008; Takenaka, 2004) and is not considered Indigenous American.

Subclade C1, particularly C1b, C1c, and C1d, are Indigenous American subclades (Fagundes *et al*., 2008). C1 subclades found in the study sample are C1b (*n* = 10) and C1c (*n* = 3). Five individuals were assigned a first-degree subclade of C1, and were assumed (for continental ancestry), based on the location of sampling, to be one of the Indigenous American subclades of C, most likely C1b or C1c. There are two individuals that possess subclade C4 in the study sample. One possesses subclade C4b, whereas the other couldn't be assigned to a greater level of specificity. C4c is the only subclade of C4 known to be Indigenous American in origin (Hooshiar Kashani *et al.*, 2012). Since that samples were taken in a region of the Americas with a high proportion of indigeneity, it is plausible, if not likely, that this individual possesses a

C4c subclade. Tamm *et al*. (2007) found evidence of subclade C4c in an Ijka-speaker from Colombia, relatively near to the Upper Amazonian sampling location of this study. Malhi *et al*., (2010) found an Indigenous individual with subclade C4c in British Columbia, Canada. Hooshiar Kashani *et al*. (2012) found subclade C4c in numerous Native American groups across North America, including the Chippewa, Cherokee, and Dakota Sioux. The other C4 sample in this study was assigned mitochondrial subclade C4b. This individual's mitochondrial haplotype is either not Indigenous American or has an unresolved Indigenous mitochondrial haplotype. It may be possible that the aforementioned individual of subclade C4 (greater specificity not assigned), does not possess a C4c subclade, but rather, the non-Indigenous subclade C4b (which may be considered an unresolved Indigenous American mitochondrial haplotype). Subclade C4b has been found in Northern Asian populations, including the Buryats, Koryaks, Evens, and Yukaghirs (Derenko *et al*., 2007; Derenko *et al*., 2010; Federova *et al.,* 2013; Starikovskaya *et al*., 2005).

D1, D2a, D3, and D4h3 are the only haplogroup D subclades that are Indigenous to the Americas (Bodner *et al*., 2012; van Oven, 2015). In this study, all D haplogroups are of the D1 subclade except for two, that are D4. One of these is specifically D4h3a, and is Indigenous American in origin (Perego *et al.*, 2009). D4h3a was first discovered in the Cayapa of Ecuador, and later found among Alaska natives (Kemp *et al*., 2007; Rickards, Martínez-Labarga, Lum, De Stefano, & Cann, 1999). The second individual with subclade D4 in this study was not able to be assigned to a greater specificity. One study respondent exhibits subclade E1a1a. This subclade has been identified primarily in maritime Southeast Asia. It has been found in significant numbers amongst individuals from Mindanao, the Visayas, and Luzon, Indonesia (Tabbada *et al*., 2010), as well as Taiwan and southern mainland China (Soares *et al*., 2008).

Subclade H1 is possessed by seven study participants. Six of these possess subclade H1bs, and one possesses subclade H1cd. Both H1bs and H1cd are primarily associated with the Franco-Cantabrian region of southern France and northern Spain. There is evidence to support that these subclades in particular are statistically more common amongst Basque peoples from this region than non-Basque peoples (Behar *et al*., 2012a). Subclade H2 is the most commonplace of all mitochondrial subclades in this study. Other than one individual that was not

able to be assigned beyond the H2a2a subclade, all other individuals with the H2 subclade are categorized as H2a2a1 (either H2a2a1a, H2a2a1c, H2a2a1d, H2a2a1g, or undefined [H2a2a1]). Subclade H2a2a1a has been identified in both Sweden and Caucasian Americans (Greenspan, 2005a; Greenspan, 2005b). H2a2a1c, H2a2a1d and H2a2a1g have yet to be associated with any specific region of the world. Subclade H2a in general is thought to have arisen in western Europe (Behar *et al*., 2012b; Brotherton *et al*., 2014). It is common today, however, in eastern Europe and the Caucasus, particularly around the Caspian Sea, and to a lesser extent in central Asia. It has been found in Spain (particularly amongst Basque peoples), Austria, Finno-Ugric speakers of the Volga-Ural region, Finns, Estonians, Eastern Slavs, and Turks (Álvarez-Iglesias, *et al.* 2009; Behar *et al*., 2012b; Brotherton *et al*., 2014; Fendt *et al*., 2011; Loogväli *et al*., 2004).

One individual in the study sample possesses subclade H8a. H8a is associated with the Druzes of Israel/Jordan/Lebanon/Syria (considered European continental ancestry in further analyses) (Shlush *et al*., 2008). Subclade R30a1b is possessed by one study participant. This subclade has been shown to be most frequently associated with southern Indian populations (Palanichamy *et al*., 2004; Rani *et al*., 2010). For the frequency of specific haplotypes, refer to *Table 5.9*.

Subclades were divided and represented graphically based on three ancestral associations, European, Indigenous, or other. The division of subclades by ancestry was done at the most specific level attained by haplogroup assignation. This degree of specificity is imperative, since individuals from disparate regions of the world may share a given haplogroup or first-degree subclade. The study sample exhibits 57.67% (*n* = 94) Indigenous maternal ancestry, 39.88% (*n* = 65) European maternal ancestry, and 2.45% (*n* = 4) maternal ancestry from another region of the world (East Asia, northern Asia (possibly), Southeast Asia, and South India).

Table 5.9. Count and percentage of mitochondrial haplotypes.

Haplotype	Count	Percentage
A2+(64)	1	0.6
A2+(64)+@153	15	9.2
A2+(64)+16129	1	0.6
A2aa	6	3.7
A2af	2	1.2
A2ag	1	0.6
A2ao	1	0.6
A2d1a	2	1.2
A2o	1	0.6
A2v1+152	1	0.6
B2	4	2.5
B2a	2	1.2
B2a4	1	0.6
B2b	1	0.6
B2b4	2	1.2
B2l	2	1.2
B4a1c3	1	0.6
C1	5	3.1
C1b	2	1.2
C1b+16311	1	0.6
C1b1	1	0.6
C1b2	5	3.1
C1b5a	1	0.6
C1c2	3	1.8
C4	1	0.6
C4b	1	0.6
D1	13	8.0
D1a	1	0.6
D1a2	2	1.2
D1d2	2	1.2
D1f	3	1.8
D1f2	1	0.6
D1g1b	1	0.6
D1h	1	0.6
D1l	1	0.6
D1j	1	0.6
D1j1a2	4	2.5
D4	1	0.6
D4h3a9	1	0.6
E1a1a1	1	0.6
H1bs	6	3.7
H1cd	1	0.6
H2a2a	1	0.6
H2a2a1	10	6.1
H2a2a1a	1	0.6
H2a2a1c	19	11.7
H2a2a1d	1	0.6
H2a2a1g	25	15.3
H8a	1	0.6
R30a1b	1	0.6
Total:	163	100.0

Table 5.9 includes the complete list of mitochondrial haplotypes present in the study sample, by frequency “Count,” and “Percentage” of the total. The most common mitochondrial haplotype is H2a2a1g (*n* = 25), followed by H2a2a1c (*n* = 19), and A2+(64)+@153 (*n* = 15).

Figure 5.15. Y-chromosome haplogroup distribution.

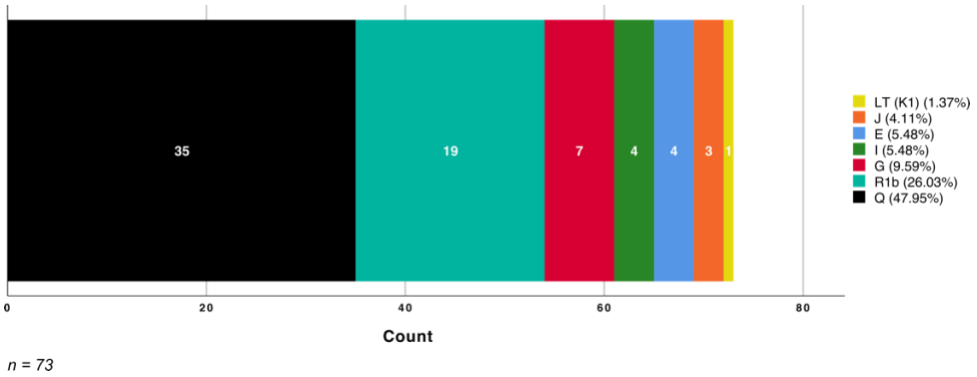


Figure 5.15 illustrates the frequency of Y-chromosome haplogroups in the study sample. Seventy-three of 79 male samples were successfully genotyped, for a success rate of 92.41%. Seven haplogroups are represented, including LT (K1), J, E, I, G, R, and Q. The most common is haplogroup Q (*n* = 35), comprising 47.95% of genotyped male participants. This is followed by haplogroup R (specifically subclade R1b) (*n* = 19), which comprises 26.0% of participants, and haplogroup G (*n* = 7), which comprises 9.59% of study participants. For the ancestry associations of each of the aforementioned Y-chromosome haplogroups see *Figure 4.5.* Because the goal of this study was to determine global Y-chromosome haplogroups and their association with demographic and migratory variables, in some cases clades were not refined to a specificity that precludes all but one continental ancestry. Y-chromosome haplogroups often include more than one continent-level ancestral origin. Follow-up research would be beneficial to clarifying Y-chromosome ancestry proportions.

Table 5.10. Populational sex-skewed gene flow analysis.

Mitochondrial (n = 163)		Y-Chromosome (n = 73)			
Indigenous Haplogroups	Non-Indigenous Haplogroups	Indigenous Haplogroups	Non-Indigenous Haplogroups	*Z*-Score	Significance
57.67%	42.33%	47.95%	52.05%	1.387	$P = 0.165$

A Z-test for two proportions was performed to assess association between mitochondrial haplogroups (indigenous vs. non-indigenous) and Y-chromosome haplogroups (indigenous vs. non-indigenous). No significant association was found.

Table 5.10 illustrates the frequency of Indigenous versus non-Indigenous haplogroups in both mitochondrial and Y-chromosomal DNA. As has been shown previously, mitochondrial (maternally inherited) haplogroups in admixed populations of the Americas are more likely to be Indigenous compared with Y-chromosome haplogroups (Norris *et al*., 2019; Wang *et al*., 2008). This is the product of assortative, sex-skewed gene flow that has occurred historically in the Americas, and is associated with conquest, colonization, warfare, and resource inequality. Despite some weak evidence for this same trend, the association between mitochondrial and Y-chromosome haplogroup proportions is not statistically significant in the study sample (P = 0.165).

Table 5.11. Individual-level sex-skewed gene flow analysis.

Comparison	Test	Significance (2-way)
[a]Status of maternal ancestry (European, Indigenous, and "other") vs. status of paternal ancestry (Indigenous vs. non-Indigenous)	Omnibus (2x3) Fisher's Exact	$P = 0.00487$**
[b]Comparison of specifically European (Indigenous & non-Indigenous paternal ancestry) maternal ancestry proportions, vs. specifically Indigenous (Indigenous & non-Indigenous paternal ancestry) maternal ancestry proportions.	Pairwise (2x2) Fisher's Exact	$P = 0.045$*
[c]Comparison of specifically European (Indigenous & non-Indigenous paternal ancestry) maternal ancestry proportions, vs. specifically "other" (Indigenous & non-Indigenous paternal ancestry) maternal ancestry proportions.	Pairwise (2x2) Fisher's Exact	$P = 0.882$
[d]Comparison of specifically Indigenous (Indigenous & non-Indigenous paternal ancestry) maternal ancestry proportions, vs. specifically "other" (Indigenous & non-Indigenous paternal ancestry) maternal ancestry proportions.	Pairwise (2x2) Fisher's Exact	$P = 0.119$

The converse comparisons of each [b, c, and d] are statistically equivalent. Fisher's exact tests were used for all comparisons even if all contingency cells were > 5, for table consistency. P-values of [b, c, and d] were adjusted using a Bonferroni correction. The comparisons from the first two rows were found to be statistically significant.

The first row of *Table 5.11* represents an omnibus Fisher's exact test of whether the status of maternal ancestry is independent of the status of paternal ancestry (Hg. Q). The latter three rows, *b*, *c*, and *d,* depict the specific, pairwise tests of independence between maternal versus paternal ancestry categories. The omnibus ancestry comparison (row 1) is statistically significant (*P* = 0.00487**), as well as the pairwise comparison of specifically European (or "non-Indigenous" for paternal) ancestry versus Indigenous ancestry (row 2) (*P* = 0.045*). See *Table 5.12* for a detailed description of this statistically significant association. The other two pairwise comparisons (rows 3 and 4) were found to be statistically nonsignificant. Those with Indigenous paternal ancestry have a significantly lower rate of Indigenous maternal ancestry (33.33%), and greater rate of European maternal ancestry (57.58%), compared to those with non-Indigenous paternal ancestry (67.57% and 32.43%, respectively).

Table 5.12. Contingency table evidence of disassortative mating.

Ancestry	Indigenous Maternal	European Maternal
Indigenous Paternal	11	19
Non-Indigenous Paternal	25	12

n = 67. Pairwise Fisher's exact test, P = 0.045 after Bonferroni correction (i.e. non-adjusted P-value multiplied by 3). Statistical significance was found.*

Table 5.12 displays a 2x2 contingency table and its associated Fisher's exact test. Specifically tested was the association between maternal ancestry (Indigenous vs. European) and paternal ancestry (Indigenous vs. non-Indigenous). Results of the Fisher's exact test indicate that a conditional "repulsion" or "disassortative" effect is present, *i.e*. indigenous maternal ancestry is associated with non-Indigenous paternal ancestry, and European maternal ancestry is associated with indigenous paternal ancestry.

Research Question 2: What is the pattern of genetic diversity in Yurimaguas?

Table 5.13. Standard and molecular mitochondrial DNA diversity indices.

n	*K*	% Unique	*H*	Π	π	θ_S
140	110	0.7857	0.9944 (0.0021)	7.2387 (± 3.4114)	0.0201 (± 0.0105)	6.8899 (± 1.9002)

Sample size (n), number of haplotypes (K), K/n (% Unique), haplotype diversity (H), mean number of pairwise differences (Π), nucleotide diversity (π), and Watterson's estimator (θ_S). "±" indicates standard deviation. (Nei, 1987; Nei & Li, 1979; Tajima, 1983,1993; Watterson, 1975).

Table 5.14. Comparative mitochondrial DNA diversity indices.

Population	Living Environment (Pop.)	*n*	*K*	% Unique	*H*	Π	π	Source
Cajamarca, Peru	Urban (< 500K)	34	22	0.647	0.972 (±0.014)	8.832 (±4.176)	0.014 (±0.007)	Guevara *et al.*, 2016
Chachapoya (intermontane Peru)	Indigenous	245	99	0.404	0.967 (±0.005)	9.824 (±4.513)	0.016 (±0.008)	Guevara *et al.*, 2016
Cochabamba, Bolivia	Urban (> 500K)	103	71	0.689	0.981 (±0.007)	6.677 —	0.019 (±0.001)	Taboada-Echalar *et al.*, 2013
Huancas (intermontane Peru)	Indigenous	21	12	0.571	0.929 (±0.033)	10.245 (±4.872)	0.016 (±0.009)	Guevara *et al.*, 2016
Iquique, Chile	Urban (< 500K)	189	90	0.476	0.975 (±0.005)	6.903 (±3.262)	0.017 (±0.009)	Gomez-Carballa *et al.*, 2016
Jivaro (lowland Peru)	Indigenous	46	19	0.413	0.934 (±0.018)	8.014 (±3.792)	0.013 (±0.007)	Guevara *et al.*, 2016
La Paz, Bolivia	Urban (> 500K)	253	121	0.478	0.960 (±0.007)	5.106 —	0.019 (±0.001)	Taboada-Echalar *et al.*, 2013
Lima + 33 urban locales, Peru	Urban (> 500K)	119	70	0.588	0.971 (±0.007)	6.299 (±3.009)	0.016 (±0.008)	Messina *et al.*, 2018
Santiago, Chile	Urban (> 500K)	167	85	0.509	0.969 (±0.007)	6.831 (±3.233)	0.017 (±0.009)	Gomez-Carballa *et al.*, 2016
Tupi-Guarani/Jê Speakers (lowland Brazil)	Indigenous	237	61	0.257	0.959 (±0.005)	5.094 (±2.480)	0.013 (±0.007)	Ramallo *et al.*, 2013; Messina *et al.*, 2018
Uros (Andean Peru)	Indigenous	99	64	0.646	0.986 (±0.004)	6.136 (±2.943)	0.015 (±0.008)	Messina *et al.*, 2018; Sandoval *et al.*, 2013
Yanesha (high-altitude intermontane Peru)	Indigenous	111	42	0.378	0.9509 —	—	0.014 —	Barbieri *et al.*, 2014; Di Corcia *et al.*, 2017
Yungay, Peru	Village	36	20	0.556	0.954 —	—	0.018 —	Lewis *et al.*, 2007
Yurimaguas (*barriada* only), Peru	Urban (< 500K)	52	46	0.885	0.989 —	—	0.022 —	Justice *et al.*, 2012
Yurimaguas, Peru	***Urban (< 500K)***	***140***	***110***	***0.786***	***0.994 (±0.002)***	***7.239 (±3.411)***	***0.020 (±0.011)***	***Present Study***

Sample size (n), number of haplotypes (K), K/n (% Unique), haplotype diversity (H), mean number of pairwise differences (Π), and nucleotide diversity (π). "±" indicates standard deviation. *(Nei, 1987; Nei & Li, 1979; Tajima, 1983,1993).*

Generally, genetic diversity has been found to increase with urbanization and decrease while moving eastward across the South American continent (Fuselli *et al*., 2003). As indicated in *Table 5.14*, when compared with thirteen neighboring populations the Yurimaguas sample has the second greatest percentage of unique haplotypes (78.6%), only after the Yurimaguas *barriada* population (88.5%). Its haplotype diversity (*H*) is the greatest amongst all listed populations (0.994 ± 0.002), with the Yurimaguas *barriada* population a close second (0.989) and the Uros peoples of Lake Titicaca third (0.986 ± 0.004). Yurimaguas' mean number of pairwise differences (Π) is fifth greatest amongst the listed populations (7.239 ± 3.411). Nucleotide diversity (π) in the Yurimaguas sample was the second greatest amongst all listed populations

(0.020 ± 0.011) only after the Yurimaguas *barriada* population. The diversity indices that Justice *et al*. (2012) analyzed within the Yurimaguas *barriada* population appear to be the most similar among those listed in *Table 5.14*, lending credence to the findings of this study. Moreover, Yurimaguas has an extensive pre- and post-Columbian history of gene flow and is currently a significant cosmopolitan center in the region. Nearly half (6 of 14) of the comparative populations are isolated Indigenous groups. Others are relatively exclusive cities (*e.g*. Lima, Peru; La Paz, Bolivia; Santiago, Chile; Iquique, Chile) that are often highly segregated by socioeconomic factors, potentially skewing the sample. Only one of the listed populations, Cajamarca, is both a small urban center (pop. ~226,000, compared to Yurimaguas, pop. ~63,000), and in Peru (INEI, 2012). The comparative populations were chosen by their geographic location (proximity favored), environment (lowland Amazonia favored), and population size (urban favored over Indigenous community).

Table 5.15. Mitochondrial DNA diversity indices and neutrality test values across demographic variables.

Variable		*n*	*K*	% Unique	*H*	π	θ_S	Tajima's *D*	Fu's F_S
Sex	Female	73	60	0.822	0.993 (±0.004)	0.018 (±0.009)	7.406 (±2.229)	-1.366 (*P*: 0.064)	-24.802 (*P*: 0.000***)
	Male	64	55	0.859	0.994 (±0.005)	0.020 (±0.011)	5.076 (±1.658)	-0.761 (*P*: 0.245)	-24.970 (*P*: 0.000***)
Language Type	No Autochthonous	134	105	0.784	0.994 (±0.002)	0.020 (±0.011)	6.945 (±1.925)	-1.348 (*P*: 0.072)	-24.778 (*P*: 0.000***)
	Autochthonous	7	7	1.000	1.000 (±0.076)	0.021 (±0.012)	6.939 (±3.442)	-0.218 (*P*: 0.438)	-1.596 (*P*: 0.099)
Living Environment	Peri-urban	16	15	0.938	0.992 (±0.025)	0.023 (±0.013)	3.918 (±1.720)	-0.108 (*P*: 0.491)	-7.195 (*P*: 0.003)
	Urban	113	85	0.752	0.992 (±0.003)	0.017 (±0.009)	7.358 (±2.072)	-1.395 (*P*: 0.052)	-24.272 (*P*: 0.000***)
Age Interval (years)	18-30	36	33	0.917	0.995 (±0.008)	0.019 (±0.010)	5.064 (±1.822)	-0.865 (*P*: 0.193)	-25.157 (*P*: 0.000***)
	31-40	25	25	1.000	1.000 (±0.011)	0.021 (±0.011)	5.032 (±1.939)	-0.530 (*P*: 0.326)	-20.796 (*P*: 0.000***)
	41-50	36	33	0.917	0.994 (±0.009)	0.021 (±0.011)	4.010 (±1.532)	-0.174 (*P*: 0.482)	-25.011 (*P*: 0.000***)
	51-60	23	22	0.957	0.996 (±0.014)	0.022 (±0.012)	13.005 (±4.576)	-1.954 (*P*: 0.021)	-12.508 (*P*: 0.000***)
	61-70	13	13	1.000	1.000 (±0.030)	0.017 (±0.009)	2.900 (±1.407)	0.623 (*P*: 0.762)	-7.042 (*P*: 0.001)
Maternal Ancestry	Indigenous	72	56	0.778	0.988 (±0.006)	0.012 (±0.006)	19.600 (±5.359)	-1.739 (*P*: 0.015*)	-24.411 (*P*: 0.000***)
	European	63	52	0.825	0.992 (±0.005)	0.018 (±0.010)	5.517 (±1.781)	-1.418 (*P*: 0.053)	-25.317 (*P*: 0.000***)

Sample size (n), number of haplotypes (K), K/n (% Unique), haplotype diversity (H), nucleotide diversity (π), Watterson's estimator (θ_S), Tajima's D (significance at P < 0.05 before correction), and Fu's F_S (significance at P < 0.005 before correction). A significant P-value for Tajima's D, after Bonferroni correction is 0.025 for sex, language type, living environment, and maternal ancestry, and 0.01 for age interval. A significant P-value for Fu's F_S, after Bonferroni correction is 0.0025 for sex, language type, living environment, and maternal ancestry, and 0.001 for age interval. "±" indicates standard deviation. Combined n values for each variable may not add to 149 (the n value of Table 5.12. Standard and molecular mitochondrial DNA diversity indices.) due to participants omitting questions, and the setting of analysis software to exclude sites missing greater than five percent of data. (Fu, 1997; Nei, 1987; Nei & Li, 1979; Tajima, 1983, 1989, 1993; Watterson, 1975).

Table 5.15 divides mitochondrial DNA diversity indices and neutrality tests by sex (female and male), language type (autochthonous and no autochthonous language), living environment (urban and peri-urban), age (18–30 years, 31–40 years, 41–50 years, 51–60years, 61–70 years), and maternal continental-ancestry (Indigenous and European). Mean number of pairwise difference (Π) calculations are not included since they are addressed in detail in *Tables 5.24*, *5.25*, *5.26*, and *5.27*. Tajima's *D* and Fu's F_S statistics had their significance cutoff levels adjusted (listed in caption) using a Bonferroni correction based on the number of subcategories within each variable. By correcting *P*-values by variable, one can more accurately assess if any Tajima's *D* and Fu's F_S statistics are truly significant within each demographic variable.

Females and males generally have very similar indices. The most conspicuous difference is females' greater Watterson's estimator (θ_S) figures, although both are within overlapping standard deviations of each other. Variability in Watterson's estimator (θ_S) signifies differences in diversity that have developed more recently relative to other metrics, such as nucleotide diversity (π) (Fuselli *et al*., 2003; Helgason, Sigureth ardóttir, Gulcher, Ward, & Stefánsson, 2000). Following the overall trend across demographic variables, both females and males possess a negative, statistically nonsignificant Tajima's *D*, and a negative, statistically significant Fu's F_S.

Although differences in diversity indices between those that do not speak an autochthonous language and those that do are apparent, most of these are likely not statistically significant, and have questionable reliability due to the low *n* value of the latter ($n = 7$). Autochthonous language speakers have a noticeably higher percentage of unique haplotypes, and a greater haplotype diversity (*H*). Nucleotide diversity (π) and Watterson's estimator (θ_S) calculations are very similar. Autochthonous language speakers have both a nonsignificant Tajima's *D* and Fu's F_S, while those that do not speak an autochthonous language have a nonsignificant Tajima's *D*, but a significant Fu's F_S.

The greatest differences in diversity indices between those that live in a peri-urban environment versus those that live in an urban environment are based on nucleotide diversity (π), and Watterson's estimator (θ_S). Both the peri-urban and urban community have negative Tajima's *D*'s and Fu's F_S values, which are all nonsignificant except for the urban Fu's F_S.

Among the variable of age, the 51–60-year interval has a very high Watterson's estimator (θ_S) figure (13.005 ± 4.576) compared to other age intervals. Amongst all other diversity indices, all age intervals are similar. Trends in neutrality tests remain consistent except for the 61–70-year age interval, that has a positive Tajima's *D,* and a nonsignificant Fu's F_S.

There is a marked difference in nucleotide diversity (π) and Watterson's estimator (θ_S) between those individuals that were found to have an Indigenous mitochondrial haplogroup versus those that had a European mitochondrial haplogroup. Nucleotide diversity is 50% greater in individuals with European maternal ancestry than those with Indigenous maternal ancestry. Perhaps the greatest difference amongst these subpopulations is in Watterson's estimator (θ_S), which is approximately 75% lower in individuals of European maternal ancestry (5.517 ± 1.781) compared to individuals of Indigenous maternal ancestry (19.600 ± 5.359). Amongst those with Indigenous maternal ancestry there were both negative and statistically significant Tajima's *D* and Fu's F_S values*,* whereas amongst those of European maternal ancestry, neutrality tests reflect the overall trend, a negative but nonsignificant Tajima's *D* and a negative and significant Fu's F_S.

Table 5.16. Mitochondrial DNA diversity indices and neutrality test values across migratory variables.

Variable		*n*	*K*	% Unique	*H*	π	θ_S	Tajima's *D*	Fu's F_S
History of Migration	Yes	54	46	0.852	0.994 (±0.005)	0.021 (±0.011)	4.828 (±1.636)	-0.633 (*P*: 0.274)	-25.003 (*P*: 0.000***)
	No	77	65	0.844	0.994 (±0.004)	0.020 (±0.011)	6.511 (±1.981)	-1.234 (*P*: 0.092)	-25.024 (*P*: 0.000***)
Number of Migration Episodes	1	32	28	0.875	0.992 (±0.0100)	0.017 (±0.009)	5.463 (±1.983)	-1.057 (*P*: 0.182)	-18.764 (*P*: 0.000***)
	2	9	8	0.889	0.972 (±0.064)	0.016 (±0.009)	4.415 (±2.165)	-0.028 (*P*: 0.523)	-1.663 (*P*: 0.136)
	3+	8	8	1.000	1.000 (±0.063)	0.026 (±0.016)	2.700 (±1.479)	-0.415 (*P*: 0.396)	-3.497 (*P*: 0.023)
Migration Duration (years)	0-5	30	28	0.933	0.995 (±0.010)	0.018 (±0.010)	6.311 (±2.270)	-0.882 (*P*: 0.200)	-20.042 (*P*: 0.000***)
	6-10	8	8	1.000	1.000 (±0.063)	0.029 (±0.018)	3.085 (±1.649)	-0.582 (*P*: 0.270)	-3.243 (*P*: 0.029)
	11+	12	11	0.917	0.985 (±0.040)	0.018 (±0.010)	3.974 (±1.851)	0.316 (*P*: 0.651)	-3.410 (*P*: 0.045)
Migration Distance (Km)	0-500	17	16	0.941	0.993 (±0.023)	0.016 (±0.009)	4.437 (±1.884)	-0.584 (*P*: 0.326)	-8.601 (*P*: 0.000***)
	501-1,000	16	15	0.938	0.992 (±0.025)	0.025 (±0.014)	3.315 (±1.501)	-0.302 (*P*: 0.402)	-9.458 (*P*: 0.001*)
	1,001+	9	9	1.000	1.000 (±0.052)	0.015 (±0.009)	4.047 (±2.010)	-1.023 (*P*: 0.181)	-3.715 (*P*: 0.013)

Sample size (n), number of haplotypes (K), K/n (% Unique), haplotype diversity (H), nucleotide diversity (π), Watterson's estimator (θ_S), Tajima's D (significance at $P < 0.05$ before correction), and Fu's F_S (significance at $P < 0.005$ before correction). A significant P-value for Tajima's D, after Bonferroni correction is 0.025 for history of migration, and 0.0167 for number of migratory episodes, migration duration, and migration distance. A significant P-value for Fu's F_S, after Bonferroni correction is 0.0025 for history of migration, and 0.0017 for number of migratory episodes, migration duration, and migration distance. "±" indicates standard deviation. Combined n values for each variable may not add to 149 (the n value of Table 5.13. Standard and molecular diversity indices for the total sample population) due to participants omitting questions, and the setting of analysis software to exclude sites missing greater than five percent of data. (Fu, 1997; Nei, 1987; Nei & Li, 1979; Tajima, 1983, 1989, 1993; Watterson, 1975).

Table 5.16 displays mitochondrial DNA diversity indices and neutrality tests by history of migration (yes or no), number of migration episodes (1, 2, or 3+), migration duration (0–5 years, 6–10 years, or 11+ years) and migration distance (0–500 Km, 501–1,000 Km, or 1,001+ Km). Mean number of pairwise difference (Π) calculations are not included since they are addressed in detail in *Tables 5.28, 5.29, 5.30, 5.31, and 5.32.* Tajima's *D* and Fu's F_S statistics had their significance cutoff levels adjusted (listed in caption) using a Bonferroni correction based on the number of subcategories within each variable.

Across all of the above migratory subcategories the greatest percentage of unique haplotypes and haplotype diversity (*H*) is found amongst those that experienced 3+ migration episodes, those with a migration duration of 6–10 years, and those with a migration distance of 1,001+ Km (all of which had "% unique" and *H* values equal to 1.000). The lowest percentage of unique haplotypes is found in those that do not have a history of migration (0.844). The lowest haplotype diversity (*H*) is found amongst those that experienced 2 migration episodes (0.972 ± 0.064). The greatest nucleotide diversity (π) is found amongst those that had a migration duration of 6–10 years (0.029 ± 0.018). The lowest is found amongst those that migrated 1,001+ Km (0.015 ± 0.009). The greatest Watterson estimator (θ_S) is found amongst those that do not have a history of migration (6.511 ± 1.981). The lowest is found amongst those that experienced 3+ migration episodes (2.700 ± 1.479). All Tajima's *D* values are negative and nonsignificant. All Fu's F_S values are negative, and six of eleven are statistically significant.

Research Question 3: Is there evidence to suggest that the Yurimaguas sample is in population size equilibrium versus demographic expansion, according to neutrality tests and mismatch distribution analyses? Is there evidence of amalgamation?

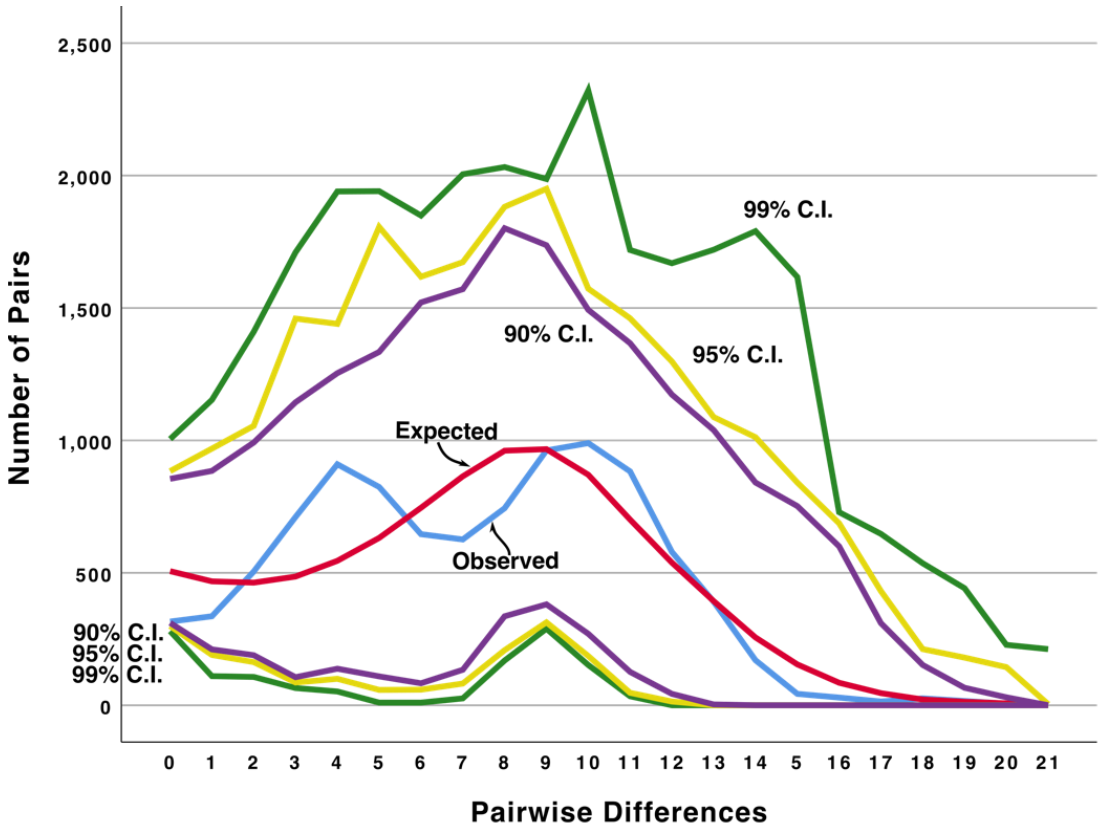
Table 5.17. Tajima's D, Fu's F_S, and Chakraborty's neutrality tests for the study sample.

Tajima's *D*	Fu's F_S	Chakraborty's Test of Amalgamation
-1.311 ($P = 0.080$)	-24.757 ($P = 0.000$***)	$P = 0.093$

Tajima's D significant at $P < 0.05$, Fu's F_S significant at $P < 0.005$, Chakraborty's test significant at P (k or more alleles) < 0.05 *(Chakraborty et al., 1988; Fu, 1997; Tajima, 1989).*

Negative Tajima's *D* and Fu's F_S values indicate a relatively high proportion of low frequency polymorphisms, which can signify either purifying selection or demographic expansion (Fuselli *et al*., 2003, p. 1688; Tajima, 1989; Fu, 1997). Whereas both tests had negative values, only Fu's F_S was found to be significant ($P = 0.000$***). A nonsignificant value for Chakraborty's test indicates a lack of evidence for population amalgamation (Chakraborty, Smouse, & Neel, 1988). This does not mean that the Yurimaguas population is not an amalgamation, but simply that given the sample size, number of potential populations encompassed, and the genetic divergence of constituent populations, it cannot be established with a statistical cutoff of $P = 0.05$. When the sample is divided to reflect only those with Indigenous maternal ancestry and only those with European maternal ancestry, Chakraborty's test statistic is nonsignificant for both ($P = 0.137$ and $P = 0.413$, respectively).

Figure 5.16. Mismatch distribution (demographic expansion model) for the study sample.



A mismatch distribution (demographic expansion model) was conducted for mtDNA in the study sample. 90%, 95%, and 99% confidence intervals were included. The sum of square deviations was calculated as 0.0050 (P = 0.6400), and Harpending's raggedness index (rg) was calculated as 0.0044 (P = 0.9800). No significance was found.

Mismatch distributions, or comparative distributions of pairwise differences, are based upon a generalized least-square approach (Schneider & Excoffier, 1999). A population with a stationary population size will display a ragged mismatch distribution, resulting in a relatively high sum of square deviations (SSD) and Harpending's raggedness index (*rg*), while an expanding population results in a smooth mismatch distribution with a single peak (Harpending, 1994). *Figure 5.16* visually portrays a bimodal distribution of observed pairwise differences, typically indicating a population in numeric equilibrium, particularly when this distribution varies in its modality from the expected distribution (in this case unimodal versus bimodal). The SSD *P*-value, however, is nonsignificant, indicating that the curve of the "observed" distribution is not significantly different from the curve of the "expected" distribution (Excoffier, 2004; Kusza *et al.*,

2018; Ray *et al*., 2003; Slatkin & Hudson, 1991). This denotes that the hypothesis of population expansion cannot be rejected. The *P*-value of *rg* represents confidence in the shape (raggedness) of the "observed" distribution. Statistical nonsignificance indicates a population that is not in numeric equilibrium (Rogers & Harpending, 1992; Slatkin & Hudson, 1991).

Mismatch distributions amongst the five most prevalent haplogroups in the study sample, A, B, C, D, and H were generated but are not shown. These five haplogroups comprise 98.77% of the study sample. All five "observed" distributions visually appear unimodal in nature and congruent with the expected distribution in all cases. Haplogroups B, C, D, and H all possess nonsignificant SSD *P*-values and Harpending's raggedness index *P*-values. A nonsignificant value indicates that the observed curve does not differ significantly from the curve expected under the demographic expansion model. Nonsignificant *rg* P-values support the theory of population expansion. Haplogroup A possesses both a statistically significant SSD *P*-value (0.0000***) and Harpending's raggedness index *P*-value (0.0100*). Consequently, the mismatch distribution for haplogroup A is consistent with demographic equilibrium (Excoffier, 2004; Rogers & Harpending, 1992; Ray *et al*., 2003; Slatkin & Hudson, 1991).

Research Question 4: Does the distribution of maternal continental-ancestries differ significantly according to demographic variables? Likewise, does the distribution of paternal continental-ancestries differ significantly according to demographic variables?

Figure 5.17. Maternal continental-ancestry distribution across sex, language type, living environment, and age interval.

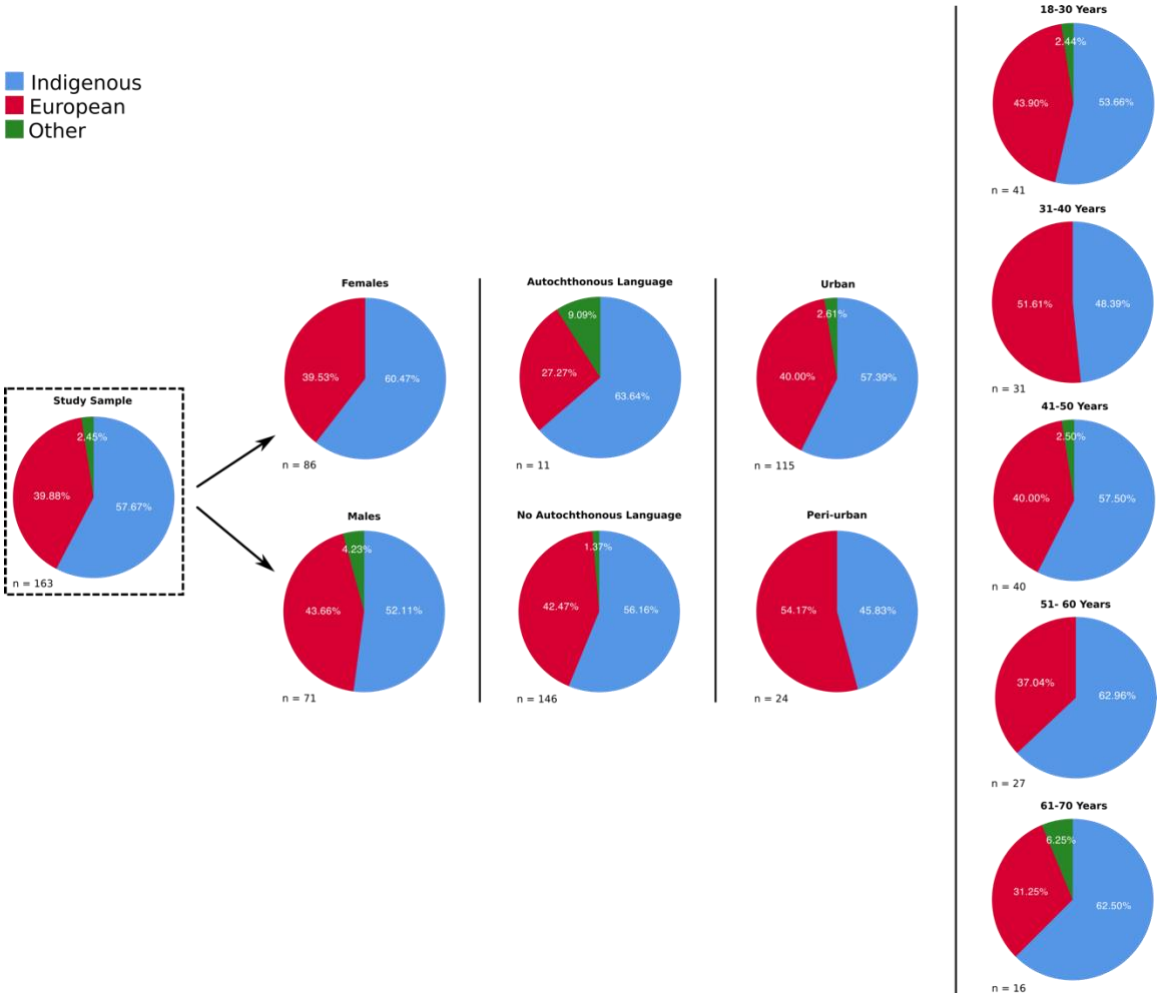


Figure 5.17 illustrates the division of maternal continental-ancestry proportions in the total study sample, as well as by sex (female and male), language type (autochthonous and no autochthonous language), living environment (urban and peri-urban), and age interval (18–30 years, 31–years, 41–50 years, 51–60 years, and 61–70 years). Ancestry was divided into the categories of Indigenous, European, and "other," with an *n* value listed for each demographic subcategory.

Across the study sample, the majority of individuals (57.67%) are of Indigenous ancestry, followed by European (39.88%) and then "other" (2.45%). Females have a greater proportion of Indigenous ancestry (60.47%) compared to their male counterparts (52.11%). Males have a

greater proportion of both European (43.66%, compared to 39.53% in females) and “other” (4.23%, compared to 0.00% in females) ancestry. Autochthonous language speakers have a greater frequency of Indigenous ancestry (63.64%) than those that do not speak an autochthonous language (56.16%). They also have a greater proportion of those with “other” ancestries (9.09%), much greater than both the total study sample (2.45%), and those that do not speak an autochthonous language (1.37%). Those that do not speak an autochthonous language have a greater proportion of individuals with European maternal ancestry (42.47%) than those that do not speak an autochthonous language (27.27%).

Study participants that reside in an urban environment have a greater proportion of Indigenous (57.39%) and “other” (2.61%) ancestries compared to those that live in a peri-urban environment (45.83% and 0.00% respectively), but a lower percentage of individuals with European ancestry (40.00%) compared with both those that reside in a peri-urban community (54.17%) and the total study sample (39.88%). Across all age intervals except 31–40 years, there is a greater proportion of individuals with Indigenous versus European ancestry, similar to the total study sample. The 31–40-year age interval was the category with the highest percentage of individuals of European ancestry (51.61%). The greatest frequency of Indigenous maternal ancestry was in the 51–60-year age interval (62.96%), followed by the 61–70-year interval (62.50%), 41–50-year interval (57.50%), 18–30-year interval (53.66%), and lastly the 31–40-year interval (48.39%). Those with the greatest percentage of “other” ancestries were, in descending order, 61–70-year-olds (6.25%), 41–50-year-olds (2.50%), and 18–30-year-olds (2.44%). The age intervals of 31–40 and 51–60 exhibited 0.00% “other” ancestries. For ancestry associations between demographic subcategories see *Table 5.19*.

Table 5.18. Multinomial logistic regression of maternal continental-ancestry by demographic variables.

	Model Fitting Criteria	Likelihood Ratio		
Effect	Goodness of Fit "*G*" (-2 Log Likelihood)	X^2	df	Significance
Age Interval	85.980	6.204	5	0.287
Sex	82.041	2.265	1	0.132
Language Type	80.069	0.293	1	0.588
History of Migration	79.936	0.160	1	0.689
Living Environment	82.180	2.404	1	0.121

Five separate multinomial logistic regressions were conducted whereby "Effects" were modeled as the independent variables of (1) age interval, (2) sex, (3) language type, (4) history of migration, and (5) living environment. Maternal continental-ancestry was modeled as the dependent variable for all five regressions. No significance was found.

Table 5.18 represents a series of five separate multinomial logistic regressions whereby demographic factors are modeled as independent variables—"Effects," and maternal continental-ancestry is modeled as the dependent variable. Model fitting criteria and likelihood ratio figures, including chi-square values, degrees of freedom, and *P*-values, are listed. The "G" and chi-square value associated with each "effect" is predicated on the ability of that factor to predict maternal ancestry (Indigenous, European, or "other"). It is important to note that the regression for "age" has more degrees of freedom than all of the other independent variables. Consequently it's "G" and chi-square values should not be compared directly to the other "effects." Instead, *P*-values ("Significance") should be directly compared. Despite variability across the various categories in maternal ancestry proportions, none of the effects tested were found to predict maternal ancestry to a statistically significant degree.

Table 5.19. Association between maternal continental-ancestry (European, Indigenous, or "other") distributions within demographic variables.

Demographic Variable	Test Type	Test Value	Significance
Sex	Fisher's Exact	3.859	$P = 0.122$
Language Type	Fisher's Exact	3.724	$P = 0.152$
Living Environment	Fisher's Exact	1.581	$P = 0.401$
Age Interval	Fisher's Exact	8.361	$P = 0.644$

Four separate Fisher's exact tests were conducted to assess independence between maternal continental-ancestry and (1) sex, (2) language type, (3) living environment, and (4) age interval. No significance was found.

Table 5.19 analyzes the association between Indigenous, European, and “other” maternal continental-ancestries within the demographic categories of sex, language type, living environment, and age interval. Distributions were compared *within* a given demographic variable, between each of the demographic subcategories, *i.e.* (female vs. male, autochthonous language vs. no autochthonous language, peri-urban vs. urban, 18–30 years vs. 31–40 years vs. 41–50 years vs. 51–60 years vs. 61–70 years). There is no statistically significant ancestry distribution within any of the listed demographic variables. Consequently, no post hoc analyses were performed.

Figure 5.18. Paternal continental-ancestry distribution across sex, language type, living environment, and age interval.

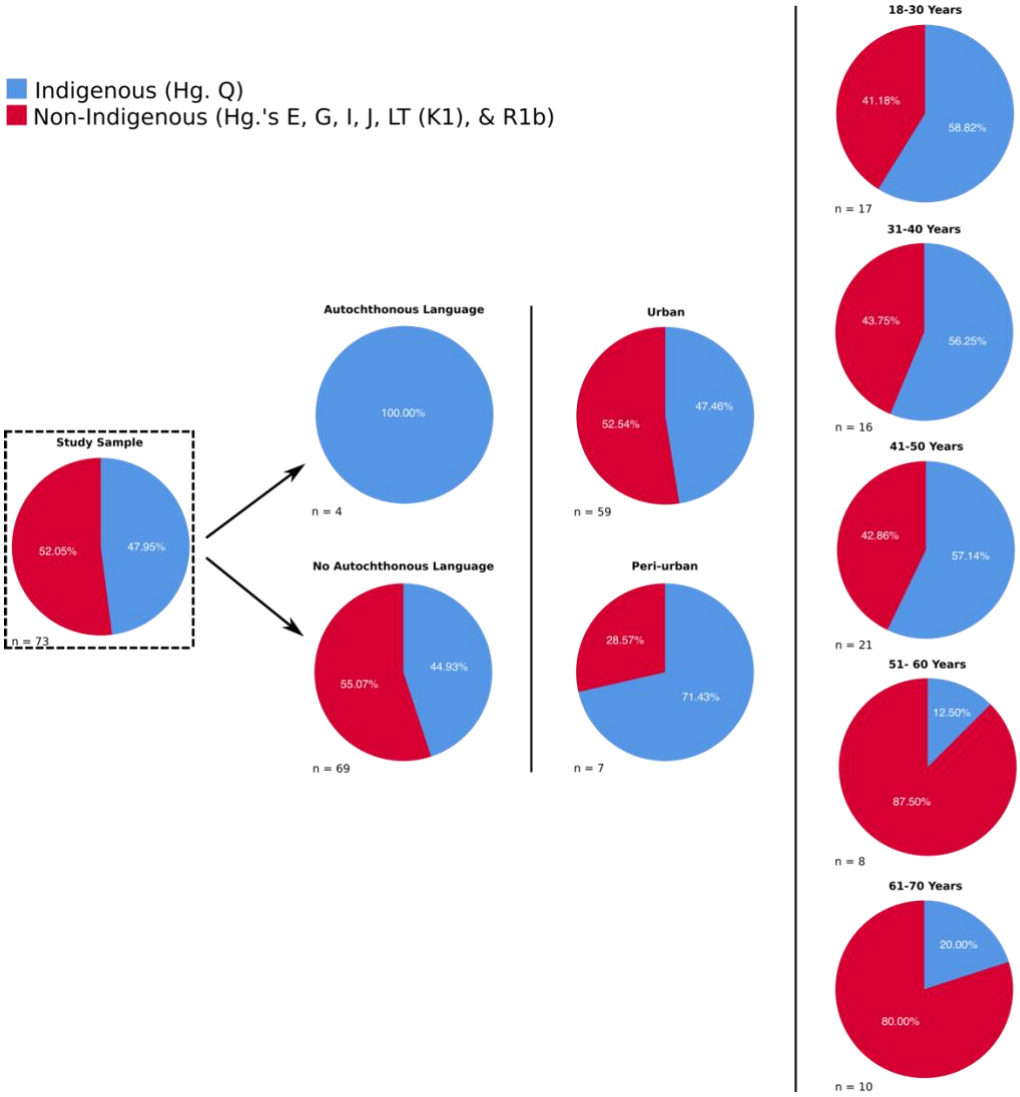


Figure 5.18, like *Figure 5.17,* divides ancestry proportions of the study sample by demographic categories, however by paternal continental-ancestry rather than by maternal continental-ancestry. The total study sample is 47.95% Indigenous paternal ancestry (Haplogroup Q), and 52.05% non-Indigenous paternal ancestry (Haplogroups E, G, I, J, LT (K1), and R1b). Worth noting is the 100% frequency of haplogroup Q among autochthonous language speakers, compared to 44.93% frequency of haplogroup Q in those that do not speak an autochthonous language. Those individuals that live in a peri-urban environment exhibit 28.57% non-Indigenous

ancestry, compared to 52.54% of those that live in an urban environment. Lastly, those in the 51–60- and 61–70-year age interval are 87.50% and 80.00% likely to exhibit non-Indigenous ancestry, respectively, compared to 52.05% in the total study sample.

Table 5.20. Association between paternal continental-ancestry (Indigenous or non-Indigenous) distributions within demographic variables.

Demographic Variable	Test Type	Test Value	Significance
Language Type	Fisher's Exact	4.595	$P = 0.048$*
Living Environment	Fisher's Exact	1.438	$P = 0.427$
Age Interval	Fisher's Exact	9.942	$P = 0.058$

Three separate Fisher's exact tests were conducted to assess independence between paternal continental-ancestry and (1) language type, (2) living environment, and (3) age interval. Significance was determined between paternal continental-ancestry and language type.

Table 5.20 analyzes the association between Indigenous and non-Indigenous paternal continental-ancestries *within* the demographic categories of language type, living environment, and age interval. There is a significant association between paternal ancestry and language type (those that speak an autochthonous language and those that do not) ($P = 0.048$*). No post hoc analyses were performed since the initial Fisher's exact test was pairwise (autochthonous vs. no autochthonous X Indigenous vs. non-Indigenous). See *Figure 5.18* for paternal ancestry proportions.

Research Question 5: Does the distribution of maternal continental-ancestries differ significantly according to migratory variables? Likewise, does the distribution of paternal continental-ancestries differ significantly according to migratory variables?

Figure 5.19. Maternal continental-ancestry distribution across migratory variables.

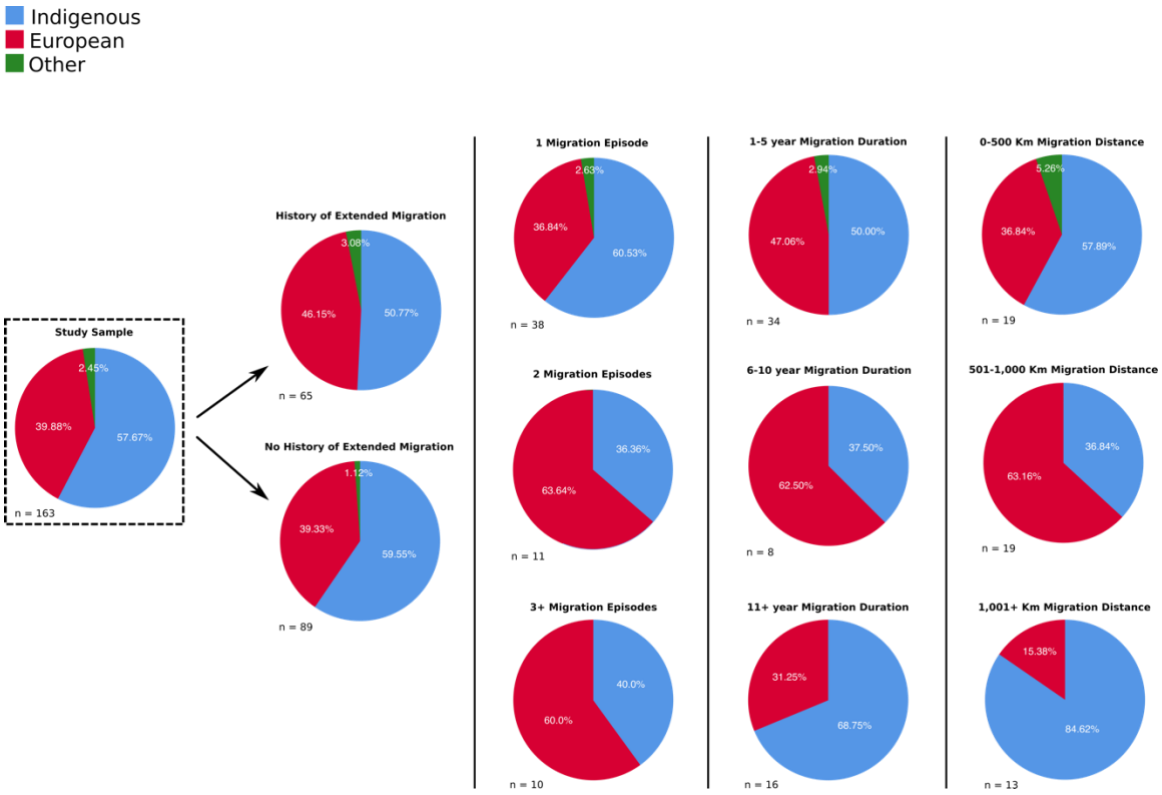


Figure 5.19 illustrates the maternal ancestry distribution of the study sample compared with maternal ancestry distributions within a number of migratory subcategories. The maternal ancestry distribution of the total sample is equal to the distribution listed in *Figure 5.17*. Subcategories “11+ year Migration Duration,” and “1,001+ Km Migration Distance have a conspicuously large percentage of Indigenous ancestry (68.75% and 84.62%, respectively), whereas subcategories “2 Migration Episodes,” “501–1,000 Km Migration Distance,” and “6–10-year Migration Duration” have a conspicuously large percentage of European ancestry (63.64%, 63.16%, and 62.50%, respectively).

Table 5.21. Multinomial logistic regression of maternal continental-ancestry by migratory variables.

	Model Fitting Criteria	Likelihood Ratio		
Effect	Goodness of Fit "*G*" (-2 Log Likelihood)	X^2	df	Significance
Num. of Migration Episodes	33.400	2.130	2	0.345
Migration Duration	32.412	1.142	3	0.767
Migration Distance	34.571	3.301	2	0.192

Three separate multinomial logistic regressions were conducted whereby "Effects" were modeled as the independent variables of (1) number of migration episodes, (2) migration duration, and (3) migration distance. Maternal continental-ancestry was modeled as the dependent variablefor all five regressions. No significance was found.

Table 5.21, like *Table 5.18,* illustrates a series of separate multinomial logistic regressions. In this case, however, migratory instead of demographic factors are modeled as the independent variables—"Effects." Maternal continental-ancestry is still the dependent variable. Model fitting criteria and likelihood ratio figures, including chi-square values, degrees of freedom, and *P*-values are listed. It is important to note that the regression for "Migration Duration (years)" has more degrees of freedom than the other two independent variables or "Effects". Consequently it's "G" and chi-square values should not be compared directly to the other two. Like *Table 5.18*, *P*-values ("Significance") should be directly compared. None of the factors considered gave statistically significant predictions of maternal continental-ancestry.

Table 5.22. Association between maternal continental-ancestry (European or Indigenous) distributions within migratory variables.

Migration Variable	Test Type	Test Value	df	Significance
History of Migration	Pearson χ^2	$\chi^2 = 0.922$	1	$P = 0.337$
Number of Migration Episodes	Kruskal-Wallis *H* (mean rank)	$H = 2.841$	1	$P = 0.092$
Migration Distance[a]	One-way ANOVA	$F = 1.301$	52	$P = 0.259$
Migration Duration[b]	One-way ANOVA	$F = 1.322$	57	$P = 0.255$

One Pearson's chi-square test, one Kruskal-Wallis H (mean rank) test, and two one-way ANOVA's were conducted to assess association between maternal continental-ancestry and, respectively, (1) history of migration, (2) number of migration episodes, (3) migration distance, and (4) migration duration. No significance was found. If a participant experienced multiple migration episodes, mean distance[a] and total duration[b] were used. Only Amerindian and European ancestry was considered due to low n values for those of "other" ancestry.

Table 5.22 demonstrates the relationship between maternal continental-ancestry *within* four different migratory variables, *i.e.* (history of migration vs. no history of migration; 1 migration

episode vs. 2 migration episodes, vs. 3+ migration episodes; by numeric migration distance [Km]; by numeric migration duration [years]). Pearson's chi-square test was used to analyze history of migration, a Kruskal-Wallis *H* (mean rank) was used for number of migration episodes, and a one-way ANOVA was used for both migration distance and migration duration. No migratory subcategory was found to be statistically associated with maternal continental-ancestry. See *Figure 5.19* for maternal ancestry proportions.

Figure 5.20. Paternal continental-ancestry distribution across migratory variables.

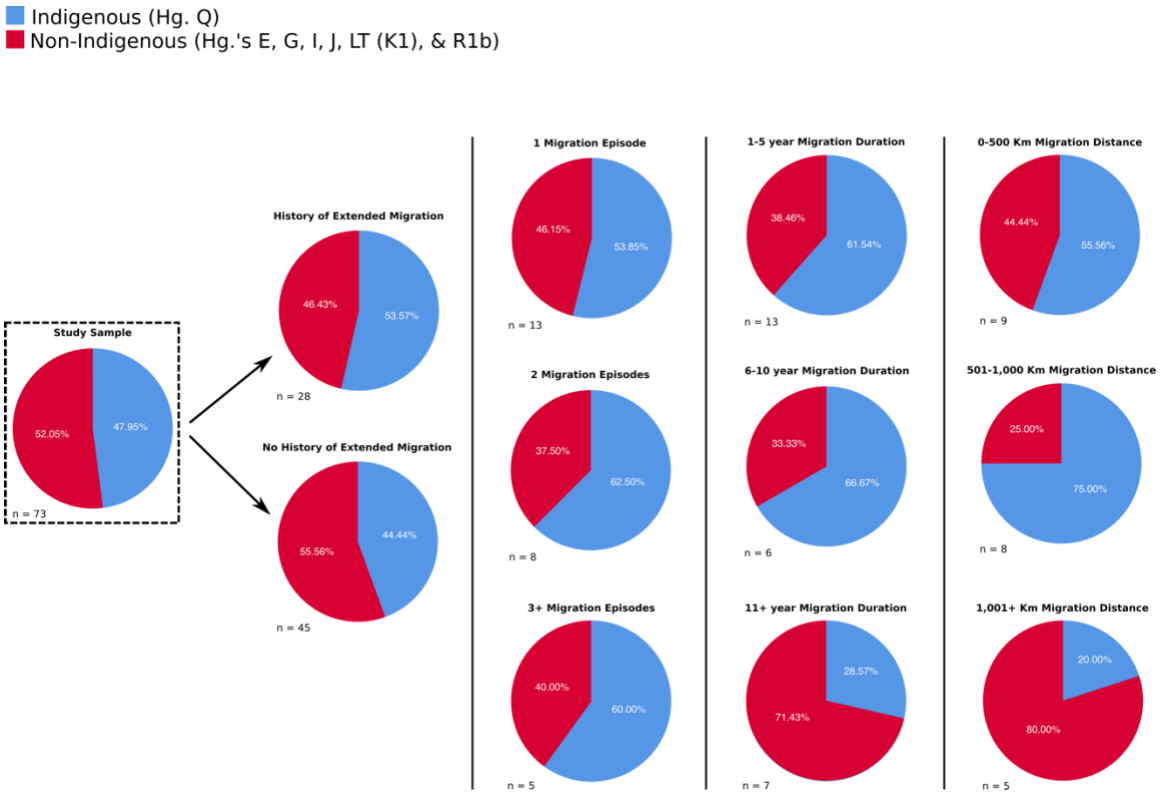


Figure 5.20 represents the paternal continental-ancestry distribution of the total study sample versus paternal continental-ancestry distribution *within* various migratory subcategories. Some migratory subcategories, such as "11+ year Migration Duration" and "1,001+ Km Migration Distance" have a relatively high degree of non-Indigenous (71.43% and 80.00%, respectively) ancestry. Those in the subcategory of "501–1,000 Km Migration Distance" are more likely to have an Indigenous paternal ancestry (75.00%).

Table 5.23. Association between paternal continental-ancestry (Indigenous or non-Indigenous) distributions within migratory variables.

Migration Variable	Test Type	Test Value	df	Significance
History of Migration	Pearson χ^2	$\chi^2 = 0.576$	1	$P = 0.448$
Number of Migration Episodes	Kruskal-Wallis *H* (mean rank)	$H = 0.065$	1	$P = 0.799$
Migration Distance[a]	One-way ANOVA	$F = 1.251$	21	$P = 0.277$
Migration Duration[b]	One-way ANOVA	$F = 2.508$	25	$P = 0.126$

One Pearson's chi-square test, one Kruskal-Wallis H (mean rank) test, and two one-way ANOVA's were conducted to assess association between paternal continental-ancestry (Amerindian vs. non-Amerindian) and, respectively, (1) history of migration, (2) number of migration episodes, (3) migration distance, and (4) migration duration. No significance was found. If a participant experienced multiple migration episodes, mean distance[a] and total duration[b] were used.

Table 5.23, like *Table 5.22*, depicts a lack of statistical association in ancestry between subcategories in each of four migratory variables. *Table 5.23*, however, refers to paternal rather than maternal ancestry. Pearson's chi-square test was used to analyze history of migration, a Kruskal-Wallis *H* (mean rank) test was used for number of migration episodes, and a one-way ANOVA was used for both migration distance and migration duration. No migratory subcategory was found to be statistically associated with paternal continental-ancestry. See *Figure 5.20* for paternal ancestry proportions.

Research Question 6: Is there evidence to suggest that demographic subdivisions of the Yurimaguas sample can be genetically detected at the nucleotide level using mitochondrial DNA sequence data? Is there evidence to suggest that migration-based subdivisions of the Yurimaguas sample can be genetically detected?

Table 5.24. Sex (female and male) Φ_{ST} comparison.

Φ_{ST}	Φ_{ST} *P*-value[a]	Π_{XY}	Π_X (Female)	Π_X (Male)	Corrected Π_{XY}
0.0020	0.2703 (±0.0429)	7.2999 ($P = 0.3273$)	7.2462	7.3239	0.0148 ($P = 0.2818$)

[a]10,000 permutations. Φ_{ST} (significance at $P < 0.05$), mean number of pairwise differences between categories (Π_{XY}) (significance at $P < 0.05$), mean number of pairwise differences within category (Π_X), corrected mean pairwise differences between categories (Corrected Π_{XY}) (significance at $P < 0.05$). "±" indicates standard deviation. (Excoffier, Smouse, & Quattro, 1992).

Table 5.25. Language type (no autochthonous and autochthonous) Φ_{ST} comparison.

Φ_{ST}	Φ_{ST} *P*-value[a]	Π_{XY}	Π_X (No Autochthonous)	Π_X (Autochthonous)	Corrected Π_{XY}
0.0000	0.3514 (±0.0478)	7.2537 (P = 0.7727)	7.4672	7.0000	0.0201 (P = 0.2909)

[a]*10,000 permutations. Φ_{ST} (significance at $P < 0.05$), mean number of pairwise differences between categories (Π_{XY}) (significance at $P < 0.05$), mean number of pairwise differences within category (Π_X), corrected mean pairwise differences between categories (Corrected Π_{XY}) (significance at $P < 0.05$). "±" indicates standard deviation. (Excoffier, Smouse, & Quattro, 1992).*

Table 5.26. Living environment (peri-urban and urban) Φ_{ST} comparison.

Φ_{ST}	Φ_{ST} *P*-value[a]	Π_{XY}	Π_X (Peri-urban)	Π_X (Urban)	Corrected Π_{XY}
0.0478	0.0270* (±0.0139)	8.5332 (P = 0.0091**)	8.4417	7.8602	0.3823 (P = 0.0455*)

[a]*10,000 permutations. Φ_{ST} (significance at $P < 0.05$), mean number of pairwise differences between categories (Π_{XY}) (significance at $P < 0.05$), mean number of pairwise differences within category (Π_X), corrected mean pairwise differences between categories (Corrected Π_{XY}) (significance at $P < 0.05$). "±" indicates standard deviation. (Excoffier, Smouse, & Quattro, 1992).*

Table 5.27. Age interval (18–30 years, 31–40 years, 41–50 years, 51–60 years, and 61–70 years) Φ_{ST} comparisons.

Age Interval Comparison	Φ_{ST}	Φ_{ST} *P*-value[a]	Π_{XY}	Corrected Π_{XY}
18–30 & 31–40	0.0000	0.6126 (±0.0327)	7.4178 (*P* = 0.2091)	-0.0956 (*P* = 0.6546)
18–30 & 41–50	0.0000	0.7117 (±0.0583)	7.0702 (*P* = 0.7182)	-0.0758 (*P* = 0.7182)
18–30 & 51–60	0.0000	0.3604 (±0.0606)	7.3056 (*P* = 0.2364)	-0.0056 (*P* = 0.4273)
18–30 & 61–70	0.0000	0.6667 (±0.0653)	6.8056 (*P* = 0.6636)	-0.1598 (*P* = 0.6727)
31–40 & 41–50	0.0000	0.6306 (±0.0470)	7.5311 (*P* = 0.2273)	-0.0949 (*P* = 0.6727)
31–40 & 51–60	0.0030	0.3153 (±0.0434)	7.8157 (*P* = 0.3546)	0.0245 (*P* = 0.3091)
31–40 & 61–70	0.0018	0.2613 (±0.0344)	7.4800 (*P* = 0.9455)	0.0346 (*P* = 0.2546)
41–50 & 51–60	0.0000	0.7027 (±0.0379)	7.3309 (*P* = 0.5636)	-0.0929 (*P* = 0.7182)
41–50 & 61–70	0.0081	0.2973 (±0.0408)	7.1453 (*P* = 0.5091)	0.0672 (*P* = 0.2818)
51–60 & 61–70	0.0000	0.3423 (±0.0504)	7.2241 (*P* = 0.5727)	-0.0191 (*P* = 0.3455)

[a] *10,000 permutations. Φ_{ST} (significance at $P < 0.05$), mean number of pairwise differences between categories (Π_{XY}) (significance at $P < 0.05$), corrected mean pairwise differences between categories (Corrected Π_{XY}) (significance at $P < 0.05$), mean number of pairwise differences within category (Π_X): 18–30 = 7.0333, 31–40 = 7.9933, 41–50 = 7.2587, 51–60 = 7.5889, 61–70 = 6.8974. Bonferroni correction applied to P-values. "±" indicates standard deviation. (Excoffier, Smouse, & Quattro, 1992).*

Tables 5.24, 5.25, *5.26*, and *5.27* display Φ_{ST} calculations between subcategories of the demographic variables of sex, language type, living environment, and age interval, respectively. *Tables 5.24* through *5.27* furthermore include mean number of pairwise difference calculations within each demographic subcategory (Π_X), between demographic subcategories (Π_{XY}), and corrected (corrected Π_{XY}). The mean number of pairwise differences within each migratory subcategory (Π_X) for *Table 5.27* (Age Interval) is listed beneath the table. *Phi*$_{ST}$ (Φ_{ST}) calculations (including *P*-values) are used to address subpopulational differentiation at the nucleotide level. It is important to note, however, that the Φ_{ST} values reported in this study are for subpopulation *samples*, and therefore are subject to sampling error. Φ_{ST} values reported as 0.0000 (and a *P*-value of 1.0000) in this study therefore *may* represent a single panmictic subpopulation, however the standard deviation must be referred to for a range of statistical possibility. Indices of 0.00 to

0.05 in a subpopulation represents little differentiation, 0.05 to 0.15 represents moderate differentiation, 0.15 to 0.25 represents great differentiation, and indices greater than 0.25 represent very great genetic differentiation (Wright, 1978). For all analyses considered, only living environment provided values that differed significantly from zero. Peri-urban versus urban residents (*Table 5.26*) have a Φ_{ST} value of 0.0478 (P = 0.0270*). Because the P-value is statistically significant, genetic differentiation between these subpopulations is indicated.

Table 5.28. History of migration (yes and no) Φ_{ST} comparison.

Φ_{ST}	Φ_{ST} P-value[a]	Π_{XY}	Π_X (History of Migration)	Π_X (No History of Migration)	Corrected Π_{XY}
0.0000	0.7748 (±0.0389)	7.3050 (P = 0.7182)	7.3445	7.3701	-0.0524 (P = 0.7727)

[a]10,000 permutations. Φ_{ST} (significance at $P < 0.05$), mean number of pairwise differences between categories (Π_{XY}) (significance at $P < 0.05$), mean number of pairwise differences within category (Π_X), corrected mean pairwise differences between categories (Corrected Π_{XY}) (significance at $P < 0.05$). "±" indicates standard deviation. (Excoffier, Smouse, & Quattro, 1992).

Table 5.29. Number of migration episodes (1, 2, and 3+) Φ_{ST} comparisons.

Comparison (Number of Episodes)	Φ_{ST}	Φ_{ST} P-value[a]	Π_{XY}	Corrected Π_{XY}
1 & 2	0.1834	0.0270* (±0.0091)	7.7743 (P = 0.0273*)	1.3699 (P = 0.0273*)
1 & 3+	0.1993	0.0000*** (±0.0000)	7.5938 (P = 0.0000***)	1.5147 (P = 0.0000***)
2 & 3+	0.0000	1.0000 (±0.0508)	6.1250 (P = 1.0000)	-0.2718 (P = 1.0000)

[a]10,000 permutations. Φ_{ST} (significance at $P < 0.05$), mean number of pairwise differences between categories (Π_{XY}) (significance at $P < 0.05$), corrected mean pairwise differences between categories (Corrected Π_{XY}) (significance at $P < 0.05$), mean number of pairwise differences within category (Π_X): 1 Episode = 6.0867, 2 Episodes = 6.7222, 3+ Episodes = 6.0714. Bonferroni correction applied to P-values. "±" indicates standard deviation. (Excoffier, Smouse, & Quattro, 1992).

Table 5.30. Migration duration (0–5 years, 6–10 years, and 11+ years) Φ_{ST} comparisons.

Comparison (Migration Duration)	Φ_{ST}	Φ_{ST} *P*-value[a]	Π_{XY}	Corrected Π_{XY}
0–5 years & 6–10 years	0.0000	1.0000 (±0.0511)	7.1708 (*P* = 1.0000)	-0.1446 (*P* = 1.0000)
0–5 years & 11+ years	0.0000	1.0000 (±0.0317)	6.4472 (*P* = 1.0000)	-0.1312 (*P* = 1.0000)
6–10 years & 11+ years	0.0199	0.6486 (±0.0364)	6.7917 (*P* = 0.3000)	0.1001 (*P* = 0.7638)

[a]*10,000 permutations. Φ_{ST} (significance at P < 0.05), mean number of pairwise differences between categories (Π_{XY}) (significance at P < 0.05), corrected mean pairwise differences between categories (Corrected Π_{XY}) (significance at P < 0.05), mean number of pairwise differences within category (Π_X): 0–5 years = 7.2023, 6–10 years = 7.4286, 11+ years = 5.9546. Bonferroni correction applied to P-values. "±" indicates standard deviation. (Excoffier, Smouse, & Quattro, 1992).*

Table 5.31. Migration distance (0–500 Km, 501–1,000 Km, and 1,001+ Km) Φ_{ST} comparisons.

Comparison (Migration Distance)	Φ_{ST}	Φ_{ST} *P*-value[a]	Π_{XY}	Corrected Π_{XY}
0–500 Km & 501–1,000 Km	0.1420	0.0270* (±0.0091)	8.7206 (*P* = 0.0273*)	1.2343 (*P* = 0.0273*)
0–500 Km & 1,001+ Km	0.0000	1.0000 (±0.0515)	6.8235 (*P* = 1.0000)	-0.0155 (*P* = 1.0000)
501–1,000 Km & 1,001+ Km	0.1891	0.0000*** (±0.0000)	9.1667 (*P* = 0.0273*)	1.7972 (*P* = 0.0000***)

[a]*10,000 permutations. Φ_{ST} (significance at P < 0.05), mean number of pairwise differences between categories (Π_{XY}) (significance at P < 0.05), corrected mean pairwise differences between categories (Corrected Π_{XY}) (significance at P < 0.05), mean number of pairwise differences within category (Π_X): 0–500 Km = 6.9559, 501–1,000 Km = 8.0167, 1,001+ Km = 6.7222. Bonferroni correction applied to P-values. (Excoffier, Smouse, and Quattro, 1992).*

Table 5.32. Three locales most commonly migrated to (Lima, Tarapoto, and Iquitos) Φ_{ST} comparisons.

Comparison (Cities Migrated to)	Φ_{ST}	Φ_{ST} *P*-value[a]	Π_{XY}	Corrected Π_{XY}
Lima & Tarapoto	0.0000	1.0000 (±0.0244)	5.5192 (*P* = 1.0000)	-0.2936 (*P* = 1.0000)
Lima & Iquitos	0.0000	1.0000 (±0.0451)	5.0156 (*P* = 1.0000)	-0.0880 (*P* = 1.0000)
Tarapoto & Iquitos	0.0000	1.0000 (±0.0333)	5.1346 (*P* = 1.0000)	-0.1818 (*P* = 1.0000)

[a]10,000 permutations. Φ_{ST} (significance at P < 0.05), mean number of pairwise differences between categories (Π_{XY}) (significance at P < 0.05), corrected mean pairwise differences between categories (Corrected Π_{XY}) (significance at P < 0.05), mean number of pairwise differences within category (Π_X): Lima = 5.6000, Tarapoto = 5.5192, Iquitos = 5.0156. Bonferroni correction applied to P-values. "±" indicates standard deviation. (Excoffier, Smouse, & Quattro, 1992).

Tables 5.28, *5.29*, *5.30*, *5.31*, and *5.32*, like *Tables 5.24* through *5.27*, depict Φ_{ST} values among subpopulations within the Yurimaguas sample. Unlike *Tables 5.24* through *5.27*, subpopulations are predicated on migratory rather than demographic variables. The mean number of pairwise differences within each migratory subcategory (Π_X) is listed beneath each table. *Table 5.28* displays genetic differentiation statistics according to whether an individual has or does not have a history of migration. These two subpopulations appear to be genetically undifferentiated, with a Φ_{ST} value of 0.0000, although the associated *P*-value of this statistic is nonsignificant (*P* = 0.7748).

Table 5.29 provides genetic differentiation statistics among study participants that engaged in either one, two, or three+ migration episodes. A statistically significant differentiation was found between those individuals that engaged in one versus two, and one versus three+ migration episodes, but not between those individuals that engaged in two versus three+ migration episodes. The Φ_{ST} value between participants that engaged in one versus two migration episodes is 0.1834 (*P* = 0.0270*). The Φ_{ST} value between participants that engaged in one versus three+ migration episodes is 0.1993 (*P* = 0.000***). Lastly, the Φ_{ST} value between individuals that engaged in two versus three+ migration episodes is 0.0000 (*P* = 1.0000),

indicating panmixia. These results corroborate the presence and statistical significance of genetic differentiation between individuals that engaged in a varying number of migration episodes. Unlike *Table 5.28* which includes individuals that both have and have not engaged in at least one migration episode, *Table 5.29* includes only those individuals that have.

Table 5.30 analyzes genetic differentiation according to the amount of time study participants spent as a migrant. Lengths of migration episode(s) were divided into three categories, 0–5 years, 6–10 years, and 11+ years. Subpopulations 0–5 years versus 6–10 years, and 0–5 years versus 11+ years both exhibit no genetic differentiation at the nucleotide level (Φ_{ST}: 0.0000, P = 1.0000; Φ_{ST}: 0.0000, P = 1.0000). A genetic differentiation, albeit nonsignificant (P = 0.6486), was detectable between those who migrated 6–10 years and those who migrated 11+ years (Φ_{ST}: 0.0199).

Table 5.31 divides the total study sample into subpopulations according to migration distance, 0–500 Km, 501–1,000 Km, and 1,001+ Km. A statistically significant genetic differentiation was found between individuals that migrated 0–500 Km versus 501–1,000 Km (Φ_{ST}: 0.1420; P = 0.0270*), and 501–1,000 Km versus 1,001+ Km (Φ_{ST}: 0.1891; P = 0.0000***). The subpopulations of 0–500 Km and 1,001+ Km were found to be in panmixis (Φ_{ST}: 0.0000; P = 1.0000).

Table 5.32 compares Φ_{ST} statistics between those individuals that migrated to Lima, Tarapoto, or Iquitos, the three locales most commonly emigrated to from Yurimaguas. The three comparisons all exhibit Φ_{ST} values of 0.0000 (P = 1.0000), indicating that their sampled allele frequencies did not differ significantly, and they may be in panmixis. Thus, there is no statistically significant difference in the mitochondrial control region between those who migrated from Yurimaguas to Lima, Tarapoto, or Iquitos.

Chapter VI: Discussion

This chapter addresses the overall findings of this study and compares them to related research whenever possible. The overarching theme is a focus on the genetic consequences of migration and urbanization in an emerging Amazonian population center. The goal of this discussion chapter is *not* to compare Peruvian or *selva baja* populations, but to highlight the consequences of migratory and urbanization processes in Yurimaguas as a whole, or in various subpopulations. Finally, this chapter provides a depth of view seldom employed on the processes that lead to population genetic transformations within a developing cosmopolitan center.

Demography

Urbanization rates have continually increased, with 54% of people now living in cities worldwide, and 80% in Latin America specifically (United Nations, 2014). Migration is frequently a consequence (and contributing factor) of climate change, systemic violence, and institutionalized resource inequality (Black, Natali, & Skinner, 2005; Brown, 2008; Clemens, 2017). Scholarship on the consequences of human migration are multifarious. This dissertation addresses biodemographic and population genetic trends resulting from the large-scale, rural-to-urban movement of peoples in the Peruvian Amazon. The concluding chapter addresses the significance and wider applicability of this research's findings.

Migration-based statistics have been vastly understudied in Amazonia. Census-derived migratory statistics primarily concern the net, international movement of peoples, with an emphasis on refugee status and global rather than local urbanization metrics. Conspicuously, limited attention is given to *regional* movements of people and the procurement of *local* indicators of population wellbeing. The vast majority of influences on public welfare nonetheless concern these regional, rather than global migratory trends (Wickramage, Vearey, Zwi, Robinson, & Knipper, 2018). Census data only provide a loose framework of demographic and migratory trends.

The total number of refugees, asylum-seekers, and internally displaced persons in the world is over 50 million (2014) (UN, 2014). Peru, with a total population of 32.2 million in 2017, experienced a net migration of –179,540 (*i.e.* emigration) that same year (most recent data). Specific data on internally displaced persons in Peru is not available. In 2017 the international migrant population of Peru (total number of individuals residing in Peru that were born outside of Peru) was 93,780 (50.07% male, 49.93% female), or 0.29% of the Peruvian populace. Migrants to Peru today come most frequently from the following countries: Argentina (11,639), United States (10,100), Chile (7,362), Colombia (6,086), Bolivia (5,865), Spain (5,202), Brazil (4,981), China (4,188), Ecuador (4,090), Italy (3,429), and Venezuela (3,318) (UN, 2017).

Descriptive population statistics, including sex/age distribution, language(s) spoken, region of birth, majority-time city of residence, and various quantitative migratory measurements were gathered to inform a catholic understanding of the initial demographic structure of Yurimaguas. One consequential, yet understudied demographic facet is cyclical migration. Cyclical migration has become particularly commonplace amongst lowland Peruvians who seek educational and/or employment opportunities in urban centers while attempting to maintain ties, familial or otherwise, to their ancestral homeland in the surrounding *selva baja*. Families frequently become divided, with mothers sometimes choosing to raise their children in urban centers, with the hope of providing them a higher quality education (Dean *et al.*, 2011). Meanwhile, many men periodically migrate to the surrounding backcountry to work in the timber, palm oil, or coca industries (David & Dean, 2017). In some cases, a man will take up minority-time residence between numerous properties that his extended family has established over generations, maintaining an outward appearance of settlement. Community outsiders and practitioners of illicit trades are quick to seize lands if a clear presence is not periodically established. Although not as common, men may seek part-time work in urban centers, particularly if they have attained a relatively advanced degree of schooling and are therefore more likely to be employed in government, transportation, manufacturing, and commerce (David & Dean, 2017).

In addition to descriptive statistics, numerous inferential statistical analyses were conducted. The elucidation of demographic population substructure prior to conducting genetic

analyses is crucial to understanding the significance and wider context of potential genetic differentiation and/or associations. This study demonstrated that within the greater Yurimaguas area, neither age interval (*Figure 5.5*), language type (*Figure 5.6*), nor sex (*Figure 5.7*) has a statistically significant association with living environment (peri-urban vs. urban). Almost identical ratios of autochthonous language speakers and those who do not speak an autochthonous language reside across both peri-urban and urban living environments. *Figure 5.7* displays an approximately 2:1 female to male ratio in peri-urban Munichis, and an approximately 1:1 ratio in urban Yurimaguas proper. Either the female to male ratio in Munichis favors women due to much greater death rates among men in the community, or perhaps more likely, the relative emigration of males from Munichis to elsewhere in the country is much greater than it is in Yurimaguas proper. Furthermore, it is possible that relatively more women than men have migrated to Munichis from Yurimaguas and surrounding areas, although this seems unlikely as this would create a depression in the overall percentage of women in Yurimaguas. Despite these trends, a statistically significant difference in sex ratio was not reported, perhaps due to an insufficient sample size (n = 156). The conspicuous sex imbalance in peri-urban Munichis speaks to wider concerns of migration patterns, and a need for more region-based studies in the Lower Huallaga Valley and low- and middle-income countries experiencing similar migration patterns (World Bank, 2017).

Research assessed whether birth Region (*Table 5.2*), language type (*Figure 5.8*), or sex (*Figure 5.9*) is associated with one's likelihood of having a prior history of migration, but consequently found no evidence in support of the association. Sex has the potential to affect migration likelihood if there is an absence of parity in employment and/or educational opportunities, as well as in parenting expectations. Local cities in the Lower Huallaga Valley and beyond offer greater access to social services and healthcare. Although there are wide variations by neighborhood, overall, urban areas in Latin America offer greater access to social services (Dufour & Piperata, 2004). Depending on historical settlement patterns and access to wealth, respondents more closely associated with Indigenous culture (by language, for example) may migrate to urban centers, at differential rates than those of primarily European ancestry.

European-descended individuals in Upper Amazonia have historically resided in cities built on the foundations of early Spanish outposts and missions. Indigenous peoples have generally distanced themselves from such centers of European dominion, particularly as proselytism and the *encomienda* system gave way to debt-peonage and forced acculturation. Cities are places that receive modern technologies and lend themselves to diversity and more liberal philosophies—*i.e.* metropolitanism (Wirth, 1938). Cities present the possibility of novel, exciting lifeways, where new forms of self-actualization, greater freedom, and access to wider social networks and socioeconomic mobility are possible (David & Dean, in press). Ironically, cities have been a refuge from the extensive pollution and environmental degradation that has plagued the traditional homelands of regional Indigenous societies. Despite the many possible reasons to suspect that migration, within a broader rural-to-urban context, may correlate with certain population demographics, perhaps surprisingly, this study has shown no evidence to corroborate this. There is a similar proportion of both females and males who have an established record of prior migration, 43.6% and 41.8%, respectively (*Figure 5.9*). Likewise, those who speak an autochthonous language migrate at similar rates (45.5%) as those who do not (42.6%) (*Figure 5.8*).

As previously mentioned, this study quantitatively describes migration episodes undertaken by study participants. Of those who have a history of migration, the mean individual experienced two separate episodes. The mean distance traveled was 646 Km. Most migrants (62.75%) spent between 1 and 5 years in residence upon migrating (*Table 5.3*). “Distance to Migration Locale (Km)” was partitioned based on a nested series of demographics and was furthermore tested for associations between various demographic variables. Those who migrated the furthest were females that do not speak an autochthonous language, live in a peri-urban environment, and are between the ages of 18 and 30 years (mean: 1,076 Km; $n = 8$) (*Table 5.5*). A statistically significant association ($P = 0.047$) was discovered between living environment (peri-urban versus urban) and distance to migration locale (Km) (*Table 5.6*). This finding suggests that individuals from different sections of Yurimaguas have variable migratory patterns. The cause(s) of this association need to be developed more fully to assess why specifically those in

peri-urban versus urban sections of the greater Yurimaguas area migrate different distances, and if this trend is found among other population centers. Alternatively, sex, language type, and age were not found to be associated with migration distance.

Tables 5.8 and *5.7* list the locales that individuals emigrated *to* (from Yurimaguas) and *from* (to Yurimaguas), respectively. *Figures 5.11* and *5.10* illustrate the patterns of these migration episodes. Geography and kin migration both play significant roles in the migratory patterns and genetic composition of peoples worldwide (Cavalli-Sforza, Menozzi, & Piazza, 1994; Fix, 1978; Rogers, 1990). Migrants relocating *to* Yurimaguas (*Figure 5.10*) came from more communities, and in less numbers per community, than those relocating *from* Yurimaguas (*Figure 5.11*). This trend supports a rural-to-urban flow of migration, whereby individuals relocate in small numbers from a wide-range of pastoral communities to regional cities, such as Yurimaguas, eventually aggregating in a limited number of megalopolises. As of 2018, approximately 78% of the Peruvian populace lived in urban centers (United Nations, 2018b). This pattern corresponds with a shift from agrarian to industrial lifeways, the consequence of growth in regional economic activities and decreasing access to arable lands in the Andean highlands, resulting in systemic land grabbing and urban "invasions" in the Lower Huallaga Valley (David & Dean, in press). Lima, the most populous city in Peru (pop. 10,072,000) (INEI, 2012), had three migrants relocate *to* Yurimaguas, however, it absorbed 24 migrants *from* Yurimaguas. Lima comprises 31.9% of the total Peruvian populace and 41.0% of urban Peru (United Nations, 2018a). The community that provided the most migrants (10) to Yurimaguas, Lagunas (pop. 12,827) (INEI, 2005), only had two residents migrate *from* Yurimaguas. No study participants came to Yurimaguas from abroad, although two individuals from Yurimaguas relocated abroad, one to Quito, Ecuador, and a second to Milan, Italy.

The majority of migrants arriving in Yurimaguas generally come from intermontane or Amazonian Regions, including San Martín, Loreto, Amazonas, and Ucayali, although noteworthy numbers of migrants also come from the Andean Region of Cajamarca and the coastal Region of Lima (including Lima Province). The migratory flow of peoples from coastal, Andean, and intermontane Peru to the lowland forested Lower Huallaga Valley was a pattern of great

precolonial importance for human movements, trade, and communication networks (Justice *et al*., 2012). The cities that account for the most migrants *to* Yurimaguas are Lagunas, Loreto (10 migrants, 89 Km [orthodromic distance] northeastward), Lamas, San Martín (5 migrants, 152 Km west-southwestward), Santa Cruz de Succhabamba, Cajamarca (5 migrants, 741 Km west-southwestward), and Pucallpa, Ucayali (4 migrants, 820 Km south-southeastward) (*Table 5.7*). The cities that account for the most migrants *from* Yurimaguas are Lima, Lima (24 migrants, 1,098 Km southwestward), Tarapoto, San Martín (15 migrants, 131 Km southwestward), Iquitos, Loreto (12 migrants, 397 Km [orthodromic distance] northeastward), and Pucallpa, Ucayali (5 migrants, 820 Km south-southeastward) (*Table 5.8*). Migration commonly entails family units moving together. This is evident from the chain migration of five kin who moved from Santa Cruz de Succhabamba to Yurimaguas.

The number of migration episodes participants engaged in ranged from 1 to 7 (*Table 5.4*). Although there was no statistically significant difference by sex, language type, living environment, and age, there was one demonstrable trend of interest. In addition to a general decrease in the frequency of respondents who engaged in an increasing number of migration episodes, males experienced a greater number of migration episodes than females. For instance, of those females who have a history of migration (n = 37), 70.3% engaged in only a single episode, compared to 44.8% in males (n = 29). The likelihood of having experienced 2 migration episodes decreases by a factor of 4.33 in females (to 16.2%) but only by a factor of 1.3 in males (to 34.5%). Only 8.1% of females engaged in 3 migration episodes versus 13.8% in males. This trend, though nonsignificant, may be due to males engaging in cyclical migration at greater rates than females.

Population Genetics

Mitochondrial DNA was found to be 57.67% Indigenous, 39.88% European, and 2.45% of "other" ancestry (*Figure 5.14*). All four of the founding South American haplogroups, A, B, C, and D, were present in the study sample. Justice *et al*. (2012) studied a small sample population from

specifically the *barriada* neighborhoods that ring Yurimaguas. They found 100% Indigenous maternal ancestry. This is a markedly greater percentage of Indigenous ancestry than the Yurimaguas sample from this study, although their study size was much smaller (*n* = 52), and they focused on specifically *barriada* residents that recently arrived from the surrounding *selva*. Nevertheless, the following proportion of haplogroups/subclades was reported in the *barriada*: 21% A2, 33% B, 35% C1, and 11% D. Messina *et al*. (2018) found that among Lima and 33 additional Peruvian cities (*n* = 140), 90.2% of individuals exhibited Indigenous haplogroups, at the following proportions: 6.1% A, 51.5% B, 15.2% C and 17.4% D. Among these, Messina *et al.* (2018) found subclade A2 to compose 6.1%, B2 18.2%, B4 33.3%, C1 15.2%, D1 13.6%, and D4 3.8% of the study sample. Comparatively, this study found subclade A2 to compose 19.0%, B2 7.4%, B4 0.6% (but 0.0% Indigenous B4), C1 11.0%, D1 18.4%, and D4 1.2% of the study sample (*Figure 5.12*). There are large discrepancies in subclade proportions between our Yurimaguas sample, and Lima (plus other urban centers) particularly concerning Indigenous subclades A2, B2, and B4 (Messina *et al*., 2018). The Yurimaguas sample has a conspicuously greater percentage of subclade A2 and lower percentage of subclades B2 and B4. Compared to the mitochondrial haplogroups/subclades reported by Justice *et al*. (2012), this study reported a markedly lower percentage of Indigenous haplogroup B and subclade C1. The Yurimaguas *barriada* population exhibited a significant absence of shared mitochondrial haplotypes with both Andean and distant Amazonian populations, prompting Justice *et al.* to suggest that Indigenous ancestry in the Yurimaguas *barriada* is most likely from populations of the nearby *selva* who have not as of yet been comprehensively sampled and incorporated into databases for comparative purposes. This speaks to the ongoing need to work with populations from this region, so that genetic architecture can be further elucidated.

Guevara *et al*. (2016) found the nearby Chachapoya population (intermontane Amazonas Region, Peru; *n* = 277) to possess 88.8% Indigenous maternal ancestry (13.7% A, 35.4% B, 22.0% C, and 17.7% D). They also sampled the Huancas (intermontane Amazonas Region, Peru; *n* = 21), Jivaro (lowland Amazonas Region, Peru; *n* = 47), and Cajamarca (intermontane Amazonas Region, Peru; *n* = 37) peoples. The Huancas displayed 100%

Indigenous maternal ancestry (42.9% A, 42.9% B, 4.7% C, and 9.5% D) while the Jivaro exhibited 97.9% (12.8% A, 59.6% B, 25.5% C, and 0.0% D), and the Cajamarca 91.9% (10.8% A, 32.4% B, 32.4% C, and 16.3% D). Subclades B2 and C1 were the most prevalent amongst the Chachapoya, Jivaro and Cajamarca, but not the Huancas, in which the two most common subclades were A2 and B2. Di Corcia *et al*. (2017) studied the Asháninka, Cashibo, and Shipibo (lowland forested Ucayali Basin, Ucayali Region, Peru), as well as the Huambisa (lowland forested Marañón Basin, Loreto Region, Peru). All 162 individuals were found to have Indigenous maternal ancestry (2017). The Asháninka (*n* = 20) Cashibo (*n* = 60), Shipibo (*n* = 60) and Huambisa (*n* = 22) had the following mitochondrial haplogroup compositions (2017):

• Asháninka: 20.0% A, 20.0% B, 25.0% C, 35.0% D

• Cashibo: 0.0% A, 30.0% B, 70.0% C, 0.0% D

• Shipibo: 31.0% A, 18.0% B, 45.0% C, 5.0% D

• Huambisa: 22.0% A, 63.0% B, 5.0% C, 0.0% D

Maziéres *et al*. assessed mitochondrial and non-recombining Y-chromosomal genetic markers in the Matsiguenga of southern Peruvian Amazonia (*n* = 38) and found a mitochondrial subclade distribution of 5.3% A4, 92.1% B4, and 2.6% D1 (100% Indigenous maternal ancestry) (2008). Bisso-Machado *et al*. (2012) assessed mitochondrial haplogroup distribution across all of Amazonia (55 populations, *n* = 2,410) and found the following haplogroup distribution: 20% A, 21% B, 31% C, 25% D, and 3% “other”.

Worth noting is the variation found in minority ancestries. Messina *et al.* (2018) reported 7.1% Sub-Saharan African and 0.8% East Asian matrilineal ancestry in urban Peru. Yurimaguas exhibited 0.0% Sub-Saharan African ancestry and 1.2% East Asian ancestry. Sandoval *et al*. (2013) also described the presence, but unreported levels of East Asian ancestry in Amazonia, particularly in the cities of Pucallpa (Ucayali Region, Peru), Chachapoyas (Amazonas Region, Peru) and Iquitos (Loreto Region, Peru). Discrepancies in maternal ancestry may denote a dissonance in population history, namely the evolutionary forces of gene flow and genetic drift, and as exemplified by the Yurimaguas *barriada* sample, population substructure. These forces

have wrought unique molecular, and hence haplotypic trademarks that manifest at the population level.

Greater, however nonsignificant proportions of non-Indigenous ancestry were revealed in the paternal, Y-chromosome lineages of participants than the maternal, mitochondrial lineages. This is evidence to refute the widespread occurrence of sex-skewed gene flow at the population level. This finding does not necessarily refute directional mating, however, that potentially dates back to Spanish colonization, national incorporation of the region, and various economic boom cycles. The Yurimaguas sample possessed 47.95% Indigenous haplogroup Q (*Figure 5.15*). Guevara *et al.* (2012) reported Y-chromosome haplogroup Q at a frequency of 60.2% among the Chachapoya (*n* = 113), 91.7% among the Huancas (*n* = 12), 95.8% among the Jivaro (*n* = 24), and 98.0% among the Cajamarca (*n* = 25). Barbieri *et al.* (2014) discovered 70% Indigenous, and 30% European (mainland Italy, Sicily, and Tyrol, Austria) and Sub-Saharan African Y-chromosome haplogroups among the intermontane Yanesha peoples. Di Corcia *et al*. (2017) found haplogroup Q to be present with 100% frequency in the Asháninka, Cashibo, Shipibo, and Huambisa peoples. Lastly, Maziéres *et al*. (2008) reported a Y-chromosome haplogroup distribution (*n* = 28) of 90.4% haplogroup Q (Q3* 80.7% and Q* 9.7%), and 9.6% other non-Indigenous Y haplogroups.

Following Q, haplogroup R (specifically subclade R1b) is most prevalent in the Yurimaguas sample, at 26.03%. This result is consistent with findings from admixed South American populations, where R1b is the most common non-Indigenous subclade (Malhi *et al*., 2008). R1b is found at its greatest frequency in western Europe, but is also present in eastern Europe, North Africa, and western Asia (ISOGG, 2019). While the overwhelming majority of scholarship contends that subclade R1b is found with regularity amongst Indigenous populations due to past admixture with Europeans (Malhi *et al.*, 2008; Salzano & Sans, 2014), some have suggested that the presence of R1b is not due to postconquest gene flow, but instead due to it being a founding South American subclade (Raghavan *et al*., 2014). In agreement with the International Society of Genetic Genealogy (2019), Seielstad (2000), and Bolnick, Bolnick, and

Smith (2006) this study takes the position that evidence of Y-chromosome subclade R1b as an artefact of European admixture has been substantially corroborated.

Bisso-Machado *et al*. (2012) compared paternal continental-ancestry composition (Indigenous Q vs. non-Q) in a compilation of South American populations and should be referred to comparatively. It is important to note, however, that the populations compiled by Bisso-Machado *et al*. (2012) are not from urban centers, but rather traditional pastoral Indigenous communities. Bortolini *et al.* (2003) compared the proportion of Y-chromosome haplogroup Q to Y-chromosome haplogroup R. In reference to the 23 populations that Bortolini *et al*. (2003) compiled, Yurimaguas had the lowest frequency of haplogroup Q (~48%) and highest frequency of haplogroup R (~26%). The next lowest frequency of haplogroup Q was amongst the Witoto peoples of southern Colombia (75%). The next highest frequency of haplogroup R was however amongst the Kaigang people of southern Brazil (14%). Like Bisso-Machado *et al*. (2012), the populations that Bortolini *et al*. (2003) sampled were not urban, but rather traditional, pastoral, Indigenous communities. The relatively high frequency of haplogroup R in Yurimaguas is likely the product of extensive European emigration and gene flow into the Lower Huallaga Valley compared to surrounding populations. Bisso-Machado *et al*. (2012) noted a conspicuous increase in non-Q haplogroups as one travels south through Amazonia. One exception to this rule, however, is the Yanomámi (*n* = 39), that exhibit 62% non-Q Y-chromosome haplogroups. This, again, is likely due to non-Indigenous admixture, from a confirmed history of mating episodes (Bailliet *et al*., 2009; Blanco-Verea, Jaime, Bríon, & Carrecedo, 2010; Marrero *et al*., 2007). Other Amazonian populations with a relatively high incidence of non-Q haplogroups include the Colla (65%, *n* = 63, location: intermontane of western Bolivia, northern Chile, northern Argentina and southern Peru), Mapuche (64%, *n* = 105, location: *sierra* and *pampas* [plains] of southern Chile and southwestern Argentina), and the Diaguita (63%, *n* = 24 location: semi-arid northern Chile, and northwestern Argentina) (Bailliet *et al*., 2009; Bisso-Machado *et al*. 2012; Blanco-Verea *et al*., 2010; Toscanini *et al*., 2011).

The Yurimaguas sample is 57.67% Indigenous maternal ancestry and 47.95% Indigenous paternal ancestry (*Table 5.10*). This pattern of lower percentages of Indigenous

haplotypes in paternal versus maternal genetic markers is common throughout Latin America and other regions of the world with colonial pasts (Carvajal-Carmona *et al*., 2000; Norris *et al*., 2019; Wang *et al*., 2008). Despite this pattern being reported in Yurimaguas statistically, the aforementioned difference in indigeneity proportions is not significant, and therefore does not support populational sex-skewed gene flow (Z-test for two proportions, Z-score: 1.387, *P* = 0.165). For comparative purposes, Guevara *et al*. calculated the Chachapoya population to be 88.8% Indigenous maternal ancestry and 60.2% Indigenous paternal ancestry (Z-test for two proportions, Z-score: 6.476, *P* = 0.000***). Maziéres *et al.* (2008) reported 100% Indigenous maternal ancestry versus 90.4% Indigenous paternal ancestry amongst the Matsiguenga (Z-test for two proportions, Z-score: 2.065, *P* = 0.038*). Although not all Amazonian populations exhibit statistically significant evidence for population-based sex-skewed gene flow, many do. Yurimaguas differs from many of these populations because it is a thriving urban center that has a consistent influx of diverse peoples. The urbanization process attracts peoples to Yurimaguas for diverse reasons. This alters the composition of the population in many ways, one of which is a dampening of many extreme trends through the injection of variability. It is difficult to say if population-based sex-skewed gene flow is less common across Amazonian urban centers than traditional pastoral communities overall, as very little research has addressed this issue in growing cities.

Despite a lack of evidence for population-based sex-skewed gene flow, there is a statistically significant association between maternal ancestry and paternal ancestry (*P* = 0.00487**), at the individual-level (*Table 5.11*). When subdivided into pairwise comparisons (and adjusted using a Bonferroni correction) European (or non-Indigenous for paternal ancestry) versus Indigenous ancestry proportions were significantly associated (*P* = 0.045*) (*Table 5.12*). These findings demonstrate that the Yurimaguas sample exhibits a trend which may be indicative of sex-skewed gene flow beyond the “population level”—which does not typically standardize for the difference between males and females regarding the pre-admixture number of individuals of Indigenous ancestry versus European (or non-Indigenous) ancestry. This trend can be characterized as disassortative mating, or a predilection toward exogamy. Because colonial

powers generally send a greater proportion of men than women to conquered lands, the resultant male population, if random mating is assumed, will have a proportionately greater representation of European ancestry than the female population. Assuming that all individuals contribute to the gene pool, and that they do so through a non-assortative mechanism, one would assume that European lineage would be represented at higher rates through Y-chromosome haplogroups (paternal lineage) than mitochondrial haplogroups (maternal lineage). Because we do not have an accurate understanding of the proportion of males versus females with Indigenous versus European ancestry prior to admixture, we cannot test if the current trend in population-based gene flow is statistically different from what would be expected, given the assumption of random mating. What is possible, however, is assessing disassortative mating, or sex-skewed gene flow, based on conditional probability, at the individual level.

At the individual level, those with Indigenous paternal ancestry are less likely than those with non-Indigenous paternal ancestry to show Indigenous, rather than European maternal ancestry, *i.e.* a "repulsion" or "disassortative" effect (*Tables 5.11* and *5.12*). Likewise, those of Indigenous paternal ancestry are more likely to show European maternal ancestry than those of non-Indigenous paternal ancestry. Given the aforementioned assumptions about colonialism, this result suggests that individuals of non-Indigenous (likely European) paternal ancestry were contributing to the gene pool more so than those of Indigenous ancestry, but that in instances where individuals of Indigenous paternal ancestry *were* contributing, they were likely taking part in disassortative mating along with Europeans. Therefore, disassortative mating, after accounting for parity in pre-admixture ancestry in males versus females, appears to be present, not only between European males and Indigenous females, but also between Indigenous males and European females. This finding highlights how mate choice fundamentally affects population structure and human evolution (Norris *et al*., 2019).

Genetic diversity indices in the Yurimaguas sample reflect a relatively high percentage of unique haplotypes (*K*/*n*), as well as a relatively high haplotype diversity (*H*), rather high mean number of pairwise differences (Π), and relatively high nucleotide diversity (π) (*Table 5.13*). While Tajima's *D* neutrality test and Chakraborty's test of amalgamation were both nonsignificant, Fu's

F_S test was found to have significantly negative values (P = 0.000***) (*Table 5.17*). A concomitant relatively high haplotype diversity and nucleotide diversity suggests that a population has either been undergoing a prolonged period of expansion or is the product of the recent amalgamation of previously isolated neighboring populations.

Nucleotide diversity (π) generally "lags behind" haplotype diversity (H) following sudden population expansion. This may be due to selection, or perhaps population expansion—a great diversity of haplotypes may develop that only differ from each other by a limited number of pairwise differences (Helgason *et al*., 2000). Alternatively, populations that have been isolated for extended periods that then aggregate *en masse* will generally exhibit a pattern of both great haplotype diversity, and great nucleotide diversity. An exception to this can occur when extensive kin migration is present to the extent that genetic diversity can be reduced, although this is uncommon (Fix, 1978). Fu's F_S is particularly sensitive to either demographic expansion or purifying selection, and for these data is statistically significant, suggesting that the Yurimaguas population is not in neutral equilibrium (with respect to selection and/or demography) (Ramos-Onsins & Rozas, 2002). Mismatch distribution for the study sample visually reveals a bimodal distribution. A nonsignificant SSD P-value of 0.6400, however, indicates that the observed curve does not deviate significantly from the curve expected under the demographic expansion model (*Figure 5.16*) (Excoffier, 2004; Kusza *et al*., 2018; Ray *et al*., 2003; Slatkin & Hudson, 1991). This indicates that the hypothesis that the population is expanding cannot be rejected (Rogers & Harpending, 1992; Slatkin & Hudson, 1991).

Despite only sampling the *barriada* population, Justice *et al*. (2012) arrived at the conclusion that Yurimaguas represents a "patchwork" of previously isolated, recently combined demes. These demes are most likely Indigenous groups from the surrounding *selva* that have begun to arrive in Yurimaguas as part of a broader rural-to-urban migration flow, in addition to mostly European émigrés that previously settled in urban Peru. The genetic findings and modern history of Yurimaguas as a port city of immigrants, dovetails with the pre-Columbian history of trans-Andean migration, trade, and communication in the Lower Huallaga Valley (Myers, 1974; Ramos & Folguera, 2009). Additionally, the Inca, whose territory extended to include the Huallaga

Valley, in the late fifteenth and early sixteenth centuries employed a policy of *mitmaq*, in which groups of people from one part of the Incan empire were forcibly transplanted to other parts of the empire. This may have planted discernable populations among others, introducing Indigenous-Indigenous gene flow immediately prior to Spanish exploration and settlement in the Lower Huallaga Valley (Schjellerup, 2005). This "uniquely Peruvian Amazon" amalgamation of both national and international relocatees was formed through a complex history of migration. Now that the Yurimaguas population has been described based on uniparental genetic markers, migration—the major contemporary process contributing to Yurimaguas' distinctiveness, can be explored further.

Table 5.15 provides diversity indices across a number of demographic subdivisions of the Yurimaguas sample. Of particular interest is the comparison between individuals of Indigenous versus European maternal continental-ancestry. The European-descended population appears to be generally more diverse, with a greater percentage of unique haplotypes, and greater haplotype and nucleotide diversity metrics. After Bonferroni correction, both Tajima's *D* and Fu's F_S neutrality tests were negative and statistically significant for the Indigenous population. Both tests were negative but only Fu's F_S test was statistically significant for the European subpopulation (after Bonferroni correction). These findings suggest that the Indigenous contingent was undergoing demographic expansion and/or purifying selection. This result corresponds to the B, C, and D mitochondrial haplogroup mismatch distributions that are also consistent with population expansion. The European subpopulation shows a neutrality test pattern similar to the total Yurimaguas sample. A statistically significant Fu's F_S value support demographic expansion and/or purifying selection. The mismatch distribution for haplogroup H was interestingly, consistent with demographic equilibrium rather than population expansion.

All but one demographic subcategory across sex, language type, living environment, and age interval exhibited both negative Tajima's *D* and Fu's F_S values (*Table 5.15*). The 61–70–year age group had a Tajima's *D* of 0.623, Eight of 11 demographic subcategories exhibited statistically significant Fu's F_S values. No subcategories exhibited a statistically significant Tajima's *D* value. Both Tajima's *D* and Fu's F_S values were negative for all history of migration,

number of migration episodes, migration duration, and migration distance subcategories except for one, those who migrated for a period of 11+ years (*Table 5.16*). Six of 11 subcategories had statistically significant Fu's F_S values. None of the subcategories had a statistically significant Tajima's *D* value.

Figure 5.17 depicts maternal continental-level ancestry proportions in the study sample, plus eleven demographic subcategories. The figure was designed to illustrate ancestry proportions according to the following three comparison types:

- study sample versus one or more demographic subcategories (*i.e*. male, autochthonous language speakers, 51–60-year-olds, *etc.*)
- two or more demographic subcategories within different broader variables (*i.e*. female vs. 18–30-year-olds)
- a demographic subcategory and its complementary subcategory (or subcategories) within the same broader variable (*i.e*. urban vs. peri-urban)

The first two of the above comparison types have overlapping data points and are therefore statistically dependent. The last comparison type comprises mutually exclusive, exhaustive, and consequently, statistically independent demographic subcategories. To assess subpopulational ancestry association, independent variables were utilized. Corresponding to *Figure 5.17* is *Table 5.19*, which reports the results of association tests between independent demographic subcategories. None of the subcategories within the demographic variables of sex, language type, living environment and age interval were found to be statistically significant.

It is likely that at larger sample sizes (*n*) greater statistical power may produce significant results specifically for subcategories with relatively low *n* values. Because this research focuses on a single urban area, with a random sample, certain demographic and migratory categories do not have sufficiently high sample sizes, because the Yurimaguas population itself has a smaller proportion of individuals with these demographic or migratory characteristics. Given the difficulty in collecting large genetic sample sizes in the Peruvian Amazon, and the implicitly low frequency of certain subcategories, a large effort would have to be made to collect enough samples, and therefore have a great enough overall *n* to then partition the sample and still have enough

statistical power to test association between these subcategories. Power analyses have reduced utility, since little is known of the proportion of individuals in, for example peri-urban vs. urban environments, or that migrate 1, 2, or 3+ times (after determining the proportion of those that have a history of migration at all), prior to sample and data collection. Census data are of some, albeit limited practical value, although many census-based statistics are not conducted at the city- but rather the region- or country-level, without accounting for substructure. This often makes the attainment of desired sample sizes based on census-data largely unreliable. Many migratory statistics are not tabulated at all. Lastly, census-data in general can be highly inaccurate and outdated, particularly when migration is occurring so quickly that 26.5% of the population arrived in Yurimaguas only in the prior two years, a statistic that ironically *is* based on census data. The greatest resource for future studies of this type in Yurimaguas or similar cities in the Global South may be the data produced by this research, which can be viewed as a pilot.

Figure 5.18, like *Figure 5.17*, depicts continent-level ancestry distributions in the total Yurimaguas sample as well as the demographic subcategories modelled within language type, living environment and age. Unlike *Figure 5.17*, *Figure 5.18* presents paternal continental-ancestry instead of maternal continental-ancestry. The corresponding table, *5.20*, lists a statistically nonsignificant association for both living environment and age interval, however a statistically significant association for language type of $P = 0.048$. This comparison itself is pairwise (language type: peri-urban vs. urban / paternal continental-ancestry: Indigenous vs. non-Indigenous). Haplogroup Q was considered Indigenous. Non-Indigenous haplogroups included haplogroups E, G, I, J, LT (K1), and R (specifically subclade R1b) (see *Figure 4.5* for Y-chromosome haplogroup assignation). The urban environment comprised a significantly lower proportion of Indigenous paternal ancestry (47.46%) than the peri-urban environment (71.43%), and conversely, a significantly greater proportion of non-Indigenous paternal ancestry (52.54% vs. 28.57% in the peri-urban environment). Interestingly, the association between living environment and continental ancestry was only found in paternal, and not maternal lineages. This result lends itself to the interpretation that fathers' ancestry is paramount in determining peri-urban versus urban settlement configurations. Future sociocultural analyses in conjunction with

further genetic parsing could potentially deconstruct the absence in parity between maternal and paternal ancestry for this statistical association. Bolnick, Bolnick, and Smith (2006) found that among Indigenous North American populations sociocultural factors can be highly associated with Y-chromosome variation, an effect greater than language or geography. They found that postmarital residence patterns, *e.g.* patrilocal or matrilocal settlement, had a strong association with uniparental genetic structure and gene flow directionality (2006). A similar effect could be responsible for explaining why living environment is statistically associated with paternal ancestry but not maternal ancestry.

This result fits with historical patterns of peopling in the region since Spanish colonization. Many of today's cities in the Peruvian *selva*, including Yurimaguas, were established as small outposts by either Jesuit missionaries or explorers. The nearby city of Moyobamba, in the intermontane only 96 Km (orthodromic distance) from Yurimaguas, is the oldest city in the Peruvian Amazon (Moore, 2018). The Spanish utilized early settlements as a base to explore the region further, search for resources, extract labor from local peoples, and convert Indigenous populations to Christianity (Santos-Granero, 2000). During the rubber boom, and subsequent search for hydrocarbons, European immigrants from abroad and *mestizos* from the coast relocated to Amazonian cities, further establishing the viewpoint that cities in the modern state of Peru are for the non-Indigenous (Santos-Granero, 2000). Indigenous peoples that have recently taken part in the wider Peruvian rural-to-urban migration appear to have differentially relocated not to the city proper, but rather the peri-urban area, including *barriadas,* where the cost of living is generally cheaper. Unfortunately, applicable census-based data are not available to support or refute this association.

The fractured nature of the peri-urban environment in South America is such that industrial areas are adjacent to arable plots of land. For individuals who recently relocated for increased educational and economic opportunities, these fertile plots of land, or *chacras*, are money-saving and familiar. Moreover, many peoples relocating to Yurimaguas from the surrounding *selva* have limited experience with the national Peruvian economy and are therefore not as likely to possess sufficient wealth to rent, on an ongoing basis, an apartment in a well-

constructed, modern building in the city proper. The peri-urban environment allows individuals to construct their own domicile, cheaply, quickly, and without the use of highly technical tools, using corrugated pig steel, plywood, cement, flawed bricks, and other easily accessible building materials. Justice *et al*. (2012) found 100% Indigenous maternal ancestry among 52 individuals in the Yurimaguas *barriada*, providing a degree of support for this finding.

Maternal ancestry proportions were found to exhibit no evidence of statistically significant association when divided by migration-based variables, including history of migration, number of migration episodes, migration duration, and migration distance (*Figure 5.19 and Table 5.22).* *Figure 5.20* illustrates paternal ancestry proportions by quantitative migratory variables. Despite the existence of a number of markedly aberrant proportions, there is again no evidence to support statistical association between any of the tested migratory metrics (*Table 5.23*).

Tables 5.24, 5.25, 5.26, and *5.27* assess nucleotide variation in DNA sequences across various subpopulations created heuristically from demographic data. There is a reported differentiation between females and males (0.0020), although it is not statistically significant. The variable of living environment did produce a discernable, (and statistically significant) Φ_{ST} of 0.0478 (*Table 5.26*). Although this level of differentiation is considered small, the significant *P*-value of 0.0270* indicates that the Φ_{ST} differs significantly from zero for the sample collected. As one might expect there was also a statistically significant mean number of pairwise differences between the two subcategories (Π_{XY}: 8.5332, *P* = 0.0091**). The peri-urban versus urban environment was shown to be statistically associated with paternal continental-ancestry distribution discussed earlier in this chapter. Maternal continental-ancestry was not statistically associated with peri-urban versus urban environment (nor was mitochondrial subclade, statistical test generated but not shown). *Table 5.26,* however, shows that despite continent-level maternal ancestry distribution lacking a statistically significant association with living environment, the two subpopulations can be differentiated statistically at the nucleotide level using Φ_{ST} *calculations*. This speaks to the need for a formal system whereby one can compare not only the Π between parallel subpopulations or nested population levels (such as in an AMOVA), but that the percentage of ancestry-informing markers (formally used for haplogroup assignation) in Π

calculations between these parallel subpopulations or nested population levels can simultaneously be recorded. This would allow for the calculation of percentage of Π that are ancestry-informing (specifically used in haplogroup assignation) versus those that are not. Such a system would allow the determination of microdifferentiation, such as Φ_{ST}, are partially founded in ancestry differences. *Table 5.27* reports a Φ_{ST} for ten pairwise age intervals. Three comparisons show only small amounts of differentiation that are statistically nonsignificant after a Bonferroni correction was applied, indicating that intergenerational genetic drift does not appear to be occurring (Fix, 1978).

Tables 5.28, 5.29, 5.30. 5.31, and *5.32* describe subpopulational differentiation based on migratory variables. *Table 5.29* exhibits a statistically significant Φ_{ST} between those who migrated once versus those who migrated twice *and* those who migrated once versus those who migrated 3+ times. The Φ_{ST} values of 0.1834 (*P* = 0.0270* ± 0.0091) and 0.1993 (*P* = 0.0000*** ± 0.0000), respectively, suggest a "great" amount of differentiation. Those who migrated twice versus those who migrated 3+ times may be panmictic with each other (Φ_{ST} equal to zero, *P* = 1.0000 ± 0.0508). It appears that those who migrated once may be somewhat of a genetic outlier. Nevertheless, number of migration episodes experienced is not statistically associated with maternal continental-ancestry (*Figure 5.19* and *Table 5.22*). Neither is number of migration episodes experienced associated with mitochondrial subclade (statistical test generated but not shown). Therefore, genetic subdivisions based on number of migration episodes do not appear to be associated with mitochondrial ancestry-informing markers.

Table 5.31 evaluates genetic differentiation by migration distance (Km). Those who migrated 0–500 Km versus those who migrated 501–1,000 Km did show evidence of being significantly differentiated from each other, with moderately high Φ_{ST} value of 0.1420 and *P*-value of 0.0270* ± 0.0091 after applying a Bonferroni correction. Those who migrated 501–1,000 Km versus those who migrated 1,001+ Km had a "great" discernable genetic differentiation (Φ_{ST}: 0.1891) and possessed a highly significant *P*-value of 0.0000*** ± 0.0000. Those who migrated 0–500 Km versus those who migrated 1,001+ Km possessed a Φ_{ST} value of 0.0000 and a *P*-value of 1.0000 ± 0.0515. Therefore, those who migrated 501-1,000 Km may be a genetic outlier.

This finding suggests that genetic subdivisions may be associated with the specific locales that individuals relocate to, using distance as a proxy. *Table 5.32* analyzes genetic differentiation between the thee locales most commonly migrated to from Yurimaguas, including Lima (1,001+ Km) Tarapoto (0-500 Km), and Iquitos (0-500 Km). As expected, no genetic differentiation was detected. There were unfortunately no individual locales within the 500-1,000 Km range with a large enough sample size to test for genetic differentiation. Many of the locales in this distance interval are smaller communities. Perhaps there is an association between genetics and relocation to a larger urban center versus a smaller more traditional community. This was unfortunately not able to be tested given the constraints of this study. The alternative, that actual migration distance is directly associated with genetic differentiation is possible but would be very difficult to test as one would need multiple relocation locale that are a similar distance from Yurimaguas, however have different characteristics among themselves. Migration distance, like number of migration episodes, was not significantly associated with maternal continental-ancestry (*Figure 5.19* and *Table 5.22*) or subclade (statistical test generated but not shown).

Based on *Tables 5.28* through *5.32*, statistically significant genetic differentiation is possible by the number of migration episodes that one experiences, and their migration distance. No discernable differentiation was reported for the other three migratory variables, *i.e*. history of migration, migration duration, and locales most commonly migrated to.

Chapter VII: Conclusions

The genetic consequences of urbanization and migration in the emerging metropolitan area of Yurimaguas have been assessed. Field collection of data and samples in this unique area of the neotropics has permitted the analysis of biodemographics and population genetics related to both historic and modern migratory trends in this rapidly expanding, dynamic population center in the Peruvian *selva*. Instead of comparing a cross-section of Yurimaguas to pastoral Indigenous communities to the East, or metropolises on the Pacific coast, two commonly employed study designs in this region, this dissertation addressed the internal features and evolutionary processes that have shaped and continue to shape populations *today*.

Yurimaguas represents the future population architecture, both demographically and genetically, of many burgeoning conurbations in South America and the Global South. In conjunction with demographics, migration data, and population genetic indices, descriptive and inferential statistical analyses were utilized to objectively illustrate the processes ultimately responsible for the diverse structure of a highly-applicable model population. Uniparental genetic markers are a unique and useful medium for tracing the evolutionary history of human population cycles through a number of demographic events. The concomitant analysis of individual-specific migratory data alongside such markers is seldom seen. Scientists often attempt to elucidate historic demographic patterns or make inferences about microevolutionary processes based on tenuous or unsubstantiated historical and demographic claims. Although demographic and migratory findings are often displayed in tandem with genetic results, assuming an interdisciplinary approach, rarely are the two methodologies synthesized in a truly co-constitutive manner.

This study set out to address six research questions. In the introductory chapter (Chapter I) a hypothesis was provided for each of these questions. Throughout the study the various methodologies used, and the resultant findings were reported. Here, in the conclusion, a summation statement will be provided for each of the initial research questions. They are as follows:

- *Research Question 1:*

  What is the uniparental marker-based continental-ancestry composition of Yurimaguas? Moreover, is there evidence of sex-skewed gene flow?

The Yurimaguas population has a majority of Indigenous maternal ancestry and a majority of European paternal ancestry. My initial hypothesis of both markers being majority Indigenous was therefore only half correct. Founding indigenous South American mitochondrial subclades A2, B2, C1 (C1b and C1c), D2, and D4h3a were represented, as well as the European haplogroups of H1 and H2, at high frequencies. Indigenous haplogroup Q was the most common paternal haplogroup reported, followed by European haplogroup R (R1b). A trend toward sex-skewed gene flow was present, although not statistically significant, perhaps due to novel patterns of urbanization. Whereas East Asian ancestry was detected in limited cases, as I hypothesized, Sub-Saharan African ancestry was not.

- *Research Question 2:*

  What is the pattern of genetic diversity in Yurimaguas?

Yurimaguas is highly diverse compared to neighboring populations. Subdivisions of the Yurimaguas population based on demographic and migratory factors varied, although they were also relatively high, and generally not very different from each other. In accordance with my original hypothesis, Yurimaguas does specifically possess a very high nucleotide diversity. My hypothesis that those of Indigenous maternal ancestry will have greater diversity indices than European individuals was rejected.

- *Research Question 3:*

  Is there evidence to suggest that the Yurimaguas sample is in population size equilibrium versus demographic expansion, according to neutrality tests and mismatch distribution analyses? Is there evidence of amalgamation?

My hypothesis that the Yurimaguas population is demographically expanding was correct. Significant values of Fu's F_S were consistent with a pattern of population expansion, although the

possibility of signal due to selection cannot be ruled out. Nonsignificant SSD and Harpending's raggedness index (*rg*) *P*-values for the sample mismatch distribution are moreover consistent with demographic population expansion. Chakraborty's test of population amalgamation produced a nonsignificant *P*-value.

- *Research Question 4:*

  Does the distribution of maternal continental-ancestries differ significantly according to demographic variables? Likewise, does the distribution of paternal continental-ancestries differ significantly according to demographic variables?

My hypothesis that maternal continental-ancestry would not differ significantly according to demographic variables was correct. None of the tested demographics provided statistically significant values. My hypothesis that paternal continental-ancestry would significantly differ according to living environment was correct. None of the other demographics were significantly associated, as I hypothesized.

- *Research Question 5:*

  Does the distribution of maternal continental-ancestries differ significantly according to migratory variables? Likewise, does the distribution of paternal continental-ancestries differ significantly according to migratory variables? I hypothesize that paternal continental-ancestries will differ significantly according to both number of migration episodes experienced, and migration duration.

I was incorrect in my hypothesis that maternal continental-ancestries would be significantly associated with history of migration, and the number of migration episodes experienced. These variables were not significantly associated. As I originally hypothesized, the other migratory variables were not associated with maternal continental-ancestries. My hypothesis that paternal continental-ancestry would be statistically associated with number of migration episodes and migration duration was rejected. I was, however, correct in hypothesizing that all of the other

tested migratory variables would show no significant association with paternal continental-ancestry.

- Research Question 6:

  Is there evidence to suggest that demographic subdivisions of the Yurimaguas sample can be genetically detected at the nucleotide level using mitochondrial DNA sequence data? Is there evidence to suggest that migration-based subdivisions of the Yurimaguas sample can be genetically detected?

My original hypothesis that a discernable genetic differentiation at the nucleotide level in mitochondrial sequence data would be able to be observed between those who speak an autochthonous language and those who do not was incorrect. These two subcategories did not show a significant pattern of genetic differentiation. Meanwhile, the demographic subcategories of peri-urban versus urban living environment was shown to be genetically differentiated to a statistically significant degree. My hypothesis that subpopulations based on a history of migration could be genetically detected at the nucleotide level was incorrect. My hypothesis that subdivisions based on “number of migration episodes,” “migration duration (years),” “migration distance (Km)” and the “three most common locales migrated to” could not be detected was partially correct. Subdivisions based on “Number of migration episodes” could be genetically detected (between two of three comparisons), subdivisions based on “migration duration (years)” could not be detected, subdivisions based on “migration distance (Km)” could be detected (between two of three comparisons), and subdivisions based on the “three locales most commonly migrated to” could not be detected.

Biodemography and population genetic research is crucial to the following areas:

(1) elucidating our evolutionary development, historical interactions, and prior movements
(2) understanding how genetic diversity is affected through the evolutionary forces of gene flow, genetic drift, and natural selection.
(3) explaining human behaviors, particularly migratory and mating behavior

As humans continue to aggregate at unprecedented levels, more attention in the field of human genetics must be paid to today's demographic and migratory processes. Because all modern populations derive from past demographic and evolutionary changes, it is vital that we incorporate our understanding of today's human population architecture as we attempt to determine what environmental conditions will affect evolution and create our future populations. It is critical that we integrate migration data with genetic findings, some of which may affect population health and disease prevention. Research has recently shown diversity's effect on disease susceptibility and equity in medical genetics (specifically how ancestry-informed demography affects rare variant analysis and genomic architecture, skewing necessary sample sizes and ratios in study design) (Kessler & O'Connor, 2017). Urbanization has been shown to be associated with both communicable and non-communicable disease, including a 4.3-fold increase in type II diabetes mellitus among urban versus rural Peruvians (Ruiz-Alejos *et al*., 2018). Population genetic structure, the demographic/migratory processes that contribute to this structure, and the health implications that result in part from this structure, are clearly intertwined, and must be treated as such in future research. This dissertation has addressed the intersection of demography (specifically migration), and population genetic structure. Future research should build off of this foundation, collecting more samples, and incorporating health-relevant data.

Works Cited

Aikhenvald, A. (2010). *Language contact in Amazonia*. New York, NY: Oxford University Press.

Alegre, J. C., Cassel, D. K., Bandy, D., & Sanchez, P. A. (1986). Effect of land clearing on soil properties of an ultisol and subsequent crop production in Yurimaguas, Peru. In R. Lal, P. A. Sanchez, & R. W. Cummings Jr. (Eds.), *Land clearing and development in the tropics* (pp.163-180). Rotterdam, Nertherlands: A. A. Balkema.

Alexiades, M. N. (2009). Mobility and migration in Indigenous Amazonia — an introduction. In M. N. Alexiades (Ed.), *Contemporary ethnoecological perspectives* (pp. 1-46). New York, NY: Berghahn Books.

Álvarez-iglesias, V., Mosquera-Miguel, A., Cerezo, M., Quintáns, B., Zarrebeitia, M. T., Cuscó, I., … Salas, A. (2009). New population and phylogenetic features of the internal variation within mitochondrial DNA macro-haplogroup R0. *PLoS ONE*, 4(4), e5112.

Andrews, R. M., Kubacka, I., Chinnery, P. F., Lightowlers, R. N., Turnbull, D. M., & Howell, N. (1999). Reanalysis and revision of the Cambridge Reference Sequence for human mitochondrial DNA. *Nature Genetics*, 23(2), 147.

Arnaiz-Villena, A., Siles, N., Moscoso, J., Zamora, J., Serrano-Vela, J. I., Gomez-Casado, E., … Martinez-Laso, J. (2005). Origin of Aymaras from Bolivia and their relationship with other Amerindians according to HLA genes. *Tissue Antigens*, 65(4), 379-390.

Arnaiz-Villena, A., Vargas-Alarcon, G., Granados, J., Gomez-Casado, E., Longas, J, Gonzales-Hevilla, M., … Martinez-Laso, J. (2000). HLA genes in Mexican Mazatecans, the peopling of the Americas and the uniqueness of Amerindians. *Tissue Antigens*, 56(5), 405-416.

Ashfaq, M., Hebert, P. D. N., Mirza, M. S., Khan, A. M., Mansoor, S., Ghulum, S. S., Zafar, Y. (2014). DNA barcoding of *Bemisia tabaci* complex (hemiptera: aleyrodidae) reveals southerly expansion of the dominant whitefly species on cotton in Pakistan. *PLoS ONE*, 9(8), e104485.

Bailliet, G., Ramallo, V., Muzzio, M., García, A., Santos, M. R., Alfaro, E. L., … Demarchi, D. A. 2009. Brief communication: restricted geographic distribution for Y-Q* paragroup in South America. *American Journal of Physical Anthropology*. 140(3). pp. 578-582.

Barbieri, C., Heggarty, P., Yao, D. Y., Ferri, G., De Fanti, S., Sarno, S, … Pettener, D. (2014). Between Andes and Amazon: The genetic profile of the Arawak-Speaking Yanesha. *American Journal of Physical Anthropology*, 155(4), 600-609.

Baxter, R. M. & Vilain, E. (2013). Translational genetics for diagnosis of human disorders of sex development. *Annual Review of Genomics and Human Genetics*, 14, 371-392.

Behar, D. M., Harmant, C., Manry, J., van Oven, M., Haak, W., Martinez-Cruz, B., … Quintana-Murci, L. (2012a). The Basque paradigm: Genetic evidence of a maternal continuity in the Franco-Cantabrian region since pre-Neolithic times. *American Journal of Human Genetics*, 90(3), 486-493.

Behar, D. M., van Oven, M., Rosset, S., Metspalu, M., Loogväli, E.-L., Silva, N. M., … Villems, R. (2012b). A “Copernican” reassessment of the human mitochondrial DNA tree from its root. *The American Journal of Human Genetics*, 90(4), 675-684.

Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J, & Sayers, E. W. (2013). GenBank. *Nucleic Acids Research*. 41(Database Issue), D36-42.

Bianchine, P. J., Russo, T. A. (1995). The role of epidemic infectious disease in the discovery of America. In G. A. Settipane (Ed.), *Columbus and the New World: Medical implications* (pp. 11-18). Providence, RI: Ocean Side Publications.

Bilal, E., Rabadan, R., Alexe, G., Fuku, N., Ueno, H., Nishigaki, Y, … Tanaka, M. (2008). Mitochondrial DNA haplogroup D4a is a marker for extreme longevity in Japan. *PLoS ONE*, 3(6), e2421.

Bisso-Machado, R., Bortolini, M. C., & Salzano, F.M. (2012). Uniparental genetic markers in South Amerindians. *Genetics and Molecular Biology,* 35(2), 365-387.

Black, R., Natali, C., & Skinner, J. (2005). *Migration and inequality*. World Bank Group. Equity and development, world development report 2006: Background papers. Retrieved from http://siteresources.worldbank.org/INTWDR2006/Resources/477383-1118673432908/Migration_and_Inequality.pdf

Blanco-Verea, A., Jaime, J. C., Bríon, M., & Carracedo, A. (2010). Y-chromosome lineages in native South American populations. *Forensic Science International: Genetics*, 4(3), 187-193.

Boas, F. (1912). Changes in the bodily form of descendents of immigrants. *American Anthropologist*, 14(3), 530-562.

Bodner, M., Perego, U. A., Huber, G., Fendt, L., Röck, A.W., Zimmermann, B., … Parson, W. (2012). Rapid coastal spread of First Americans: Novel insights from South America's Southern Cone mitochondrial genomes. *Genome Research*, 22(5), 811-820.

Bolnick, D. A., Bolnick, D. I. & Smith, D.G. (2006). Asymmetric male and female genetic histories among Native Americans from Eastern North America. *Molecular Biology and Evolution*. 23(11), 2161–2174.

Bonferroni, C. E. (1936). Teoria statistica delle classi e calcolo delle probabilitá. *Pubblicazioni del R. Istituto Superior di Scienze Economiche e Commerciali di Firenze, Florence, R Istituto Superior di Scienze Economiche e Commerciali, 8*, 3-62.

Bortolini, M.-C., Salzano, F. M., Thomas, M. G., Stuart, S., Nasanen, S. P. K., Bau, C. H. D., … Ruiz-Linares, A. (2003). Y-chromosome evidence for differing ancient demographic histories in the Americas. *American Journal of Human Genetics*, 73(3), 524-539.

Brotherton, P., Haak, W., Templeton, J., Brandt, G., Soubrier, J., Adsler, C. J., … The Genographic Consortium. (2014). Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. *Nature Communications*, 4(1764). DOI: 10.1038/ncomms2656

Brown, O. (2008). *Migration and climate change*. IOM migration research series, no. 31. Geneva: International Organization for Migration. Retrieved from https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=8&cad=rja&uact=8&ved=2ahUKEwi7s_Wpjt3iAhVHiqwKHVNDAhoQFjAHegQICRAC&url=https%3A%2F%2Fwww.ipcc.ch%2Fapps%2Fnjlite%2Fsrex%2Fnjlite_download.php%3Fid%3D5866&usg=AOvVaw0fkNew5DKuhr0oV_u8pHMT

Burnham, R. J., & Johnson, K. R. (2004). South American palaeobotany and the origins of neotropical rainforests. *Philosophical Transactions of the Royal Society of London B:*

Biological Sciences, 359(1450), 1595-1610.

Cabana, G. S., Lewis Jr., C. M., Tito, R. Y., Covey, R. A., Cáceres, A. M., De La Cruz, A. F., … Stone, A. C. (2014). Population genetic structure of traditional populations in the Peruvian central Andes and implications for South American population history. *American Journal of Human Biology*, 86(3), 147-165.

Carmichael, E., Hugh-Jones, S., Moser, B., Taylor, D. (1985). *The hidden peoples of the Amazon*. London, England: British Museum Publications Ltd.

Carvajal-Carmona, L. G., Soto, I. D., Pineda, N., Ortiz-Barrientos, D., Duque, C., Ospina-Duque, J, … Ruiz-Linares, A. (2000). Strong Amerind/White sex bias and a possible Sephardic contribution among the founders of a population in northwest Colombia. *American Journal of Human Genetics*, 67(5), 1287-1295.

Cavalli-Sforza, L. L., Menozzi, P., & Piazza, A. (1994). *The history and geography of human genes.* Princeton, NJ: Princeton University Press.

Chakraborty, R., Smouse, P. E., & Neel, J. V. (1988). Population amalgamation and genetic variation: Observations on artificially agglomerated tribal populations of Central and South America. *American Journal of Human Genetics*, 43(5), 709-725.

Chinnery, P. F., Elliott, H. R., Syed, A., & Rothwell, P. M. (2010). Mitochondrial DNA haplogroups and risk of transient ischaemic attack and ischaemic stroke: a genetic association study. *The Lancet: Neurology*, 9(5), 498-503.

Clemens, M. A. (2017). *Violence, development, and migration waves: Evidence from Central American child migrant apprehensions.* CDG working paper 459. Washington, DC: Center for Global Development. Retrieved from https://www.cgdev.org/publication/violence-development-and-migration-waves-evidence-central-american-child-migrant

Cook, K. P. (2016). *Forbidden passages: Muslims and Moriscos in colonial Spanish America.* Philadelphia, PA: University of Pennsylvania Press.

Cook, S. F. (1973). The significance of disease in the extinction of the New England Indians. *Human Biology*, 45, 485-508.

Coriat G, J. E. (1943). *El Hombre del Amazonas y Ensayo Monográfico de Loreto (Segunda Edición*). Lima, Perú: Librería Coriat — Imprenta.

Crawford, M. (1998). *The origins of Native Americans: Evidence from anthropological genetics*. New York, NY: Cambridge University Press.

Crosby, A. W. (1972). *The Columbian exchange: Biological and cultural consequences of 1492.* Westport, CT: Greenwood Publishing Group.

David, R. E., & Dean, B. (2017). *Yurimaguas and the Huallaga River Valley: A biocultural approach to disruptive patterns of migration and urbanization in Peruvian Amazonia (an ideal location for a migration-based, biocultural approach to obesity)*. Second International Human Migration Conference (key speakers): What can Genomic and Cultural Diversity tell us about Migration?. Mexico City, Mexico.

David, R. E., & Dean, B. (in press). A sociogenomic approach to migration and urbanization in Peruvian Amazonia (in press). In M. H. Crawford, M. de Lourdes Muñoz Moreno (Eds.), *Human migration: Biocultural perspectives* New York, NY: Oxford University Press.

de Jong, W., Lye, T.-P., Abe, K.-I. (2006). *The social ecology of tropical forests: Migration, populations and frontiers.* Kyoto, Japan: Kyoto University Press.

Dean, B. (1999). Intercambios ambivalentes en la Amazonía: Formación discursiva y la violencia del patronazgo. *Anthropológia*, 17, 85-115.

Dean, B. (2002). State power and Indigenous peoples in Peruvian Amazonia: A lost decade, 1990-2000. In D. Maybury-Lewis (Ed.), *The politics of ethnicity; Indigenous peoples in Latin American States*. Cambridge, MA: Harvard University Press.

Dean, B. (2004). Ambivalent exchanges: The violence of patronazgo in the Upper Amazon. In M. Anderson (Ed.), *Cultural construction of violence: victimization, escalation, response*. West Lafayette, IN: Purdue University Press.

Dean, B. (2009). *Urarina society, cosmology, and history in Peruvian Amazonia.* Gainesville, FL: University Press of Florida.

Dean, B. (2013). Epilogue. In M. Schlecker & F. Fleischer (Eds.), *Ethnography of support encounters* (pp. 195-211). New York, NY: Palgrave Mcmillan.

Dean, B. (2016). Freedom. In N. Salazar & K. Jayaram (Eds.), *Keywords of mobility: Critical anthropological engagements* (pp. 56-74). Oxford, England: Berghan Books.

Dean, B. (2019). *The State and the Awajún: Frontier expansion in the Upper Amazon, 1541-1990*. San Diego, CA: Cognella Academic Publishing.

Dean, B., & Levi, J. (2003). Introduction. In B. Dean & J. Levi (Eds.), *At the Risk of Being Heard: Identity, Indigenous Rights, and Postcolonial States* (pp. 1-44). Ann Arbor, MI: University of Michigan Press.

Dean, B., Silverstein, S., Reamer, M., & Homan, J. (2011). The new urban jungle: Contesting territories and identities in the Upper Amazon. *Cultural Survival Quarterly*, 35(2), 34-45.

Dean, W. (1987). *Brazil and the struggle for rubber.* Cambridge, United Kingdom: Cambridge University Press.

Derenko, M., Malyarchuk, B., Grzybowski, T., Denisova, G., Dambueva, I., Perkova, M., … Zakharov, I. (2007). Phylogeographic analysis of mitochondrial DNA in northern Asian populations. *American Journal of Human Genetics,* 8(5), 1025-1041.

Derenko, M., Malyarchuk, B., Grzybowski, T., Denisova, G., Rogalla, U., Perkova, M, … Zakharov, I. (2010). Origin and post-glacial dispersal of mitochondrial DNA haplogroups C and D in northern Asia. *PLoS ONE*, 5(12), e15214.

Desarrollo Peruano. (2014). *Infraestructura Peruana: Puente Paranapura*. Desarrollo Peruano. Retrieved June 12, 2018 from http://infraestructuraperuana.blogspot.com/2014/04/puenteparanapura.html

Devor, E. J. (2004). *Polymerase chain reaction (PCR): Theory, method, and practice.* Coralville, IA: Integrated DNA Technologies.

Di Corcia, T., Sanchez Mellado, C., Davila Francia, T. J., Ferri, G., Sarno, S., Luiselli, D., & Rickards, O. (2017). East of the Andes: The genetic profile of the Peruvian Amazon populations. *American Journal of Physical Anthropology*, 163(2), 328-338.

Dillehay, T. D. (1989). *Monte Verde: A late Pleistocene settlement in Chile, vol. 1.* Washington, DC: Smithsonian Institution Press.

Dillehay, T. D. (1997). *Monte Verde: A Late Pleistocene Settlement in Chile, Vol. 2.* Washington, DC: Smithsonian Institution Press.

Dillehay, T. D., Ocampo, C., Saavedra, J., Sawakuchi, A. O., Vega, R. M., Pino, M., … Dix, G. (2015). New Archaeological evidence for an early human presence at Monte Verde, Chile. *PLos ONE*, 10(11), e0141923.

Dillehay, T. D., Ramírez, C., Pino, M., Collins, M. B., Rossen, J., & Pino-Navarro, J. D. (2008). Monte Verde: Seaweed, food, medicine, and the peopling of South America. *Science*, 320(5877) 784-786.

Dobyns, H. F. (1963). An outline of Andean epidemic history to 1720. *Bulletin of the History of Medicine*, 37(6), 493-515.

Dobyns, H. F. (1966). An Appraisal of techniques with a new hemispheric estimate. *Current Anthropology*, 7(4), 395-416.

Dufour, D. L., & Piperata, B. A. (2004). Rural-to-urban migration in Latin America: An update and thoughts on the model. *American Journal of Human Biology*, 16(4), 395-404.

Epps, P. L. (2005). Language endangerment in Amazonia: The role of missionaries. In J. Wolgemuth, T., Dirksmeyer (Eds.), *Bedrohte vielfalt: Aspects of language death.* Berlin, Germany: Weinßensee Verlag.

Excoffier, L. (2004). Patterns of DNA sequence diversity and genetic structure after a range expansion: Lessons from the infinite-island model. *Molecular Ecology*, 13(4), 853-864.

Excoffier, L. & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10(3), 564-567.

Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes-application to human mitochondrial-DNA restriction data. *Genetics*, 131(2), 479-491.

Ewans, W. J. (1972). The sampling theory of selectively neutral alleles. *Theoretical Population Biology*, 3(1), 87-112.

Fagundes, N. J., Kanitz, R., Eckert, R., Valls, A. C., Bogo, M. R., Salzano, F. M., … Bonatto, S. L. (2008). Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. *American Journal of Human Genetics,* 82(3), 583-592.

Federova, S. A., Reidla, M., Metspalu, E., Metspalu, M., Rootsi, S., Tambets, K., … Villems, R. (2013). Autosomal and uniparental portraits of the native populations of Sakha (Yakutia): Implications for the peopling of northeast Eurasia. *BMC Evolutionary Biology,* 13:127.

Fehren-Schmitz, Reindel, M., Cagigao, E. T., Hummel, S, & Herrmann, B. (2010). Pre-Columbian

population dynamics in coastal southern Peru: A diachronic investigation of mtDNA patterns in the Palpa Region by ancient DNA analysis. *American Journal of Physical Anthropology*, 141, 208-221.

Fendt, L., Niederstätter, H., Huber, G., Zelger, B., Dünser, M., Seifarth, C., … Parson, W. (2011). Accumulation of mutations over the entire mitochondrial genome of breast cancer cells obtained by tissue microdissection. *Breast Cancer Research and Treatment*. 128(2). 327-336.

Field Museum. (2013). *Field museum scientists estimate 16,000 tree species in the Amazon.* American Association for the Advancement of Science. Retrieved from https://www.eurekalert.org/pub_releases/2013-10/fm-fms101413.php

Fisher, R. A. (1922). On the interpretation of chi-square from contingency tables, and the calculation of P. *Journal of the Royal Statistical Society*, 85(1), 87-94.

Fisher, R. A. (1925). *Statistical methods for research workers*. Edinburgh, Scotland: Oliver and Boyd.

Fix, A. G. (1978). The role of kin-structured migration in genetic microdifferentiation. *Annals of Human Genetics*, 41(3), 329-339.

Fu, Y.-K. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, 147(2), 915-925.

Fuku, N., Park, K. S., Yamada, Y., Nishigaki, Y., Cho, Y. M., Matsuo, H., … Tanaka, M. (2007). Mitochondrial haplogroup N9a confers resistance against type 2 diabetes in Asians. *The American Journal of Human Genetics*, 80(3), 407-415.

Fuselli, S., Tarazona-Santos, E., Dupanloup, I., Soto, A., Luiselli, D, & Pettener, D. (2003). Mitochondrial DNA diversity in South America and the genetic history of Andean highlanders. *Molecular Biology and Evolution*, 20(10) 1682-1691.

García, P. (1995). *Territorios indígenos y la nueva legislación agraria en el Perú; Copenhagen, Denmark: IWGIA.* Lima, Perú: Grupo de Trabajo Racimos de Ungurahui.

Geospiza Research Team. (2012). *FinchTV version 1.5.0*. 2004-2012 Geospiza Inc., Seattle WA. Retrieved from http://www.geospiza.com

Gnecchi-Ruscone, G. A., Sarno, S. De Fanti, S., Gianvincenzo, L., Giuliani, C., Boattini, A., … Pettener, D. (2019). Dissecting the Pre-Columbian genomic ancestry of Native Americans along the Andes-Amazonia divide. *Molecular Biology and Evolution*, 36(6), 1254-1269.

Golob, A. (1982). *The Upper Amazon in historical perspective* [Ph.D. dissertation]. New York, NY: City University of New York.

Gomez-Carballa, A., Moreno, F., Álvarez-Iglesias, V., Martinon-Torres, F., Garcia-Magarinos, M., Pantoja-Astudillo, J. A., … Aguirre-Morales, E. (2016). Revealing latitudinal patterns of mitochondrial DNA diversity in Chileans. *Forensic Science International: Genetics*, 20(C), 81-88.

Goméz-Carballa, A., Pardo-Seco, J., Brandini, S., Achilli, A., Perego, U. A., Coble, M. D., … Salas, A. (2018). The peopling of South America and the trans-Andean gene flow of the first settlers. *Genome Research*, 28, 767-779.

Gomez-Casado, E., Martinez-Laso, J., Moscoso, J., Zamora, J., Martin-Villa, M., Perez-Blas, M, … Lopez-Santalla, M. (2003). Origin of Mayans according to HLA genes and the uniqueness of Amerindians. *Tissue Antigens,* 61(6), 425-436.

Gore Loreto. (2017). *Gobierno Regional de Loreto*. Retrieved June 28, 2018 from http://www.regionloreto.gob.pe

Gravel, S., Zakharia, F., Moreno-Estrada, A., Byrnes, J. K., Muzzio, M., Rodroguez-Flores, J. L., … Bustamante, C. D. (2013). Reconstructing Native American migrations from whole-genome and whole-exome data. *PLoS Genetics*, 9(12), e1004023.

Greene, W. H. (2012). *Econometric analysis (seventh Edition)*. Boston, MA: Pearson Education.

Greenspan, B. (2005a). GenBank accession EU157923 (September 18, 2007). In D.A., Benson, I., Karsch-Mizrachi, D. J., Lipman, J., Ostell, D. L., Wheeler. *Nucleic Acids Research,* 36(Database Issue), D25-D30.

Greenspan, B. (2005b). Genbank accession EU795361 (June 6, 2009). In D. A., Benson, I., Karsch-Mizrachi, D. J., Lipman, J., Ostell, D. L., Wheeler. *Nucleic Acids Research,* 36(Database Issue), D25-D30.

Guevara, E. K., Palo, J. U., Guillén, S., Sajantila, A. (2016). MtDNA and Y-chromosomal diversity in the Chachapoya, a population from the northeast Peruvian Andes-Amazon divide. *American Journal of Human Biology*, 28(6), 857-867.

Guio, H., Poterico, J. A., Levano, K. S., Cornejo-Olivas, M., Mazzetti, P., Manassero-Morales, G., … Abarca-Barriga, H. Genetics and genomics in Peru: Clinical and research perspective. *Molecular Genetics & Genomic Medicine*, 6, 873-886.

Guo, L.-J., Oshida, Y., Fuku, N., Takeyasu, T., Fujita, Y., Kurata, M, … Tanaka, M. (2005). Mitochondrial genome polymorphisms associated with type-2 diabetes or obesity. *Mitochondrion*, 5(1), 15-33.

Haplogroup YTree v7.03.00. (2019). *YFull: Y-SNPs*. Retrieved March 3, 2019 from https://www.yfull.com/tree/

Harpending, H. C. (1994). Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Human Biology*, 66(4), 591-600.

Harris, D. N., Song, W., Shetty, A. C., Levano, K. S., Cáceres, O., Padilla, C., … Borda, V. (2018). Evolutionary genomic dynamics of Peruvians before, during, and after the Inca Empire. *Proceedings of the National Academy of Sciences USA*. 115(28). E6526-e6535.

Harrison, R. (1995). The language and rhetoric of conversation in the Viceroyalty of Peru. *Poetics Today*, 16(1), 1-27.

Helgason, A. S., Sigureth ardóttir, S., Gulcher, J. R., Ward, R., & Stefánsson, K. (2000). MtDNA and the origin of the Icelanders: Deciphering signals of recent population history. *American Journal of Human Genetics*, 66(3), 999-1016.

Her Majesty's Consuls. (1864). *Commercial Reports Received at the Foreign Office Between July 1st, 1863 and June 30th*. London, England: Harrison and Sons.

Herrera, J. F. (1905). Estado Comercial de la Región Peruana del Amazonas el Año 1872 por Don José Fermín Herrera. In C. Larrabure I Correa (Ed.), *Colección de leyes, decretos,*

resoluciones y otras documentos oficiales referents al Departamento de Loreto, vol. 17 (pp. 104-536)*.* Lima, Perú: Oficina Tipográfica de la Opinión Nacional: Gremios 441.

Hooshiar Kashani, B., Perego, U. A., Olivieri, A., Angerhofer, N., Gandini, F., Carossa, V., … Torroni, A. (2012). Mitochondrial haplogroup C4c: A rare lineage entering America through the ice-free corridor? *American Journal of Physical Anthropology*, 147, 35-39.

Hoffecker, J. F., Elias, S. A., O'Rourke, D. H. (2014). Out of Beringia? *Science,* 343(6174), 979-980.

Homan, J. (2014). The Cordillera Escalara and the Pueblo Shawi: Ethnohistory. In N. Pitman, C. Vriesendorp, D. Alvira, J. A. Markel, M. Johnston, E Ruelas, … & S. Heilpern*.* (Eds.), *Peru: Cordillera-Escalera-Loreto, Rapid Biological and Social Inventories Report 26* (pp. 167-175). Chicago, IL: The Field Museum.

Homburger, J. R., Moreno-Estrada, A., Gignoux, C. R., Nelson, D., Sanchez-E., Ortiz-Tello, P., … Gravel, S. (2015). Genomic insights into the ancestry and demographic history of South America. *PLoS Genetics*, 11(12), e1005602.

Hornborg, A., & Hill, J.D. (2011). *Ethnicity in ancient Amazonia: Reconstructing past identities from archaeology, linguistics, and ethnohistory.* Boulder, CO: University Press of Colorado.

Howell, D. C. (2010). *Statistical methods for psychology (seventh edition)*. Belmont, CA: Wadsworth Cengage Learning.

Hudson, R. 1997. *Brazil: a country study, area handbook series.* Washington, DC: Federal Reserve Division, Library of Congress.

Hudson, R. R. (1991). Gene genealogies and the coalescent process. In D. J. Futuyma, & J. D. Antonovics (Eds.), *Oxford surveys in evolutionary biology (vol. 7)* (pp. 1-44). New York, NY: Oxford University Press.

IBM Corporation. (2017). IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corporation.

Innis, M. A., Gelfand, D. H., Sninsky, J. J., & White, T. J. (1990). Optimization of PCRs. In M. A. Innis, D. H. Gelfand, J. J. Sninsky, & T.J. White (Eds.), *PCR Protocols: A Guide to Methods and Applications* (pp. 3-12). London, England: Academic Press.

Instituto Nacional de Estadística e Informática (INEI). (2007). *Perfiles de sociodemográficos, resultados censos de los nacionales.* Lima, Perú: Instituto Nacional de Estadística e Informática.

Instituto Nacional de Estadística e Informática (INEI). (2012). *Perú: Población estimada al 30 de Junio y tasa de crecimiento de las ciudades capitales, por Departamento, 2011 y 2015*. Lima, Perú: Instituto Nacional de Estadística e Informática.

International Society of Genetic Genealogy (ISOGG). (2019). Y-DNA haplogroup tree 2019, ver. 14.77. Retrieved March 10, 2019 from https://isogg.org/tree/index.html

Jiménez de la Espada, M. (1897). *La jornada del Capitán Alonso Mercadillo a los Indios Chupachos é Iscaicingas*. Madrid, España: Boletín de la Real Sociedad Geográfica 37.

Just, R. S., Diegoli, T. M., Saunier, J. L., Irwin, J. A., Parsons, T. J. (2008). Complete mitochondrial genome sequences for 265 African American and U.S. 'Hispanic' individuals. *Forensic Science International: Genetics*, 2(3), e45-e48.

Justice, A., Dean, B., & Crawford, M. H. (2012). Molecular consequences of migration and urbanization in Peruvian Amazonia. In M. H. Crawford, & B. C. Campbell (Eds.), *Causes and consequences of human migration* (pp. 449-472). Cambridge, United Kingdom: Cambridge University Press.

Kaufman, T. (1990). Language history in South America, what we know. In D. L. Payne (Ed.), *Amazonian linguistics*. Austin, TX: University of Texas Press.

Kemp, B. M., Malhi, R. S., McDonough, J., Bolnick, D. A., Eshleman, J. A., Rickards, O., … Smith, D. G. (2007). Genetic Analysis of Early Holocene Skeletal Remains from Alaska and its Implications for the Settlement of the Americas. *American Journal of Physical Anthropology,* 132(4), 605-621.

Kemp, B. M., Tung, T. A., & Summar, M. L. (2009). Genetic continuity after the collapse of the Wari empire: Mitochondrial DNA profiles from Wari and post-Wari populations in the ancient Andes. *American Journal of Physical Anthropology*, 1-12. DOI: 10.1002/ajpa.21037

Kersulyte, D, Kalia, A, Gilman, RH, Mendez, M, Herrera, P, Cabrera, L, … Berg, D. E. (2010). *Helicobacter pylori* from Peruvian Amerindians: Traces of Human Migrations in Strains from Remote Amazon, and Genome Sequence of an Amerind Strain. *PLoS ONE*, 5(11), e15076.

Kessler, M. D., & O'Connor, T.D. (2017). Accurate and Equitable Medical Genomic Analysis Requires an Understanding of Demography and its Influence on Sample Size and Ratio. *Genome Biology*, 18(42), 1-3.

Klarén, P. F. (1986). *The origins of modern Peru, 1880-1930*. In L. Bethell (Ed.), The Cambridge history of Latin America (pp. 587-640). Cambridge, United Kingdom: Cambridge University Press.

Kloss-Brandstätter, A, Pacher, D, Schönherr, S, Weissensteiner, H, Binna, R, Specht, G, Kronenberg, F. (2011). HaploGrep: A Fast and Reliable Algorithm for Automatic Classification of Mitochondrial DNA Haplogroups. *Human Mutation*, 32(1), 25-32.

Koressaar, T., & Remm, M. (2007). Enhancements and Modifications of Primer Design Program Primer3. *Bioinformatics*, 23(10), 1289-1291.

Krausz, C., & Casamonti, E. (2017). Spermatogenic failure and the Y chromosome. *Human Genetics*, 136(5), 637-655.

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260), 583-621.

Kumar, S., Bellis, C., Zlojutro, M., Melton, P. E., Blangero, J., & Curran, J. E. (2011). Large scale mitochondrial sequencing in Mexican Americans suggests a reappraisal of Native American origins. *BMC Evolutionary Biology,* 11:293.

Kusza, S., Suchentrunk, F., Pucher, H., Mar, K. U., Zachos, F. E. (2018). High levels of mitochondrial genetic diversity in Asian elephants (*Elephas maximus*) from Myanmar. *Hystrix, the Italian Journal of Mammalogy*, 29(1), 152-154.

Lamas, B., Fehren-Schmitz, L., Valverde, G., Soubrier, J., Mallick, S, Rohland, N., … Haak, W. (2016). Ancient mitochondrial DNA provides high-resolution time scale of the peopling of the Americas. *Science Advances*, 2(4), e1501385.

Lasker, G. W., & Mascie-Taylor, C. G. N. (1988). *The framework of migration studies*. In C. G. N. Mascie-Taylor, & G. W. Lasker (Eds.), Biological aspects of human migration (pp. 1-13). Cambridge, United Kingdom: Cambridge University Press.

La Torre López, L. (1998). *Sólo queremos vivir en paz!: Experiencias petroleras en territorios indígenas de la Amazonía Peruana*; *Copenhagen, Denmark: IWGIA*. Lima, Perú: Grupo de Trabajo Racimos de Ungurahui.

Lathrap, D. W. (1970). *The Upper Amazon*. London, England: Thames and Hudson.

Lausent-Herrera, I. (2011). The Chinatown in Peru and the changing Peruvian Chinese communities. *Journal of Chinese Overseas*, 7(1), 69-113.

Lefebvre, H. (2014). The urban revolution. Minneapolis, MN: University of Minnesota Press. Retrieved from https://muse.jhu.edu/book/28861

Lewis, C. M., Lizarraga, B., Tito, R. Y., Lopez, P. W., Iannacone, G. C., Medina, A., … Stone, A. C. (2007). Mitochondrial DNA and the peopling of South America. *Human Biology,* 79(2), 159-178.

Lewis, C. M., Tito, R. Y., Lizarraga, B., Stone, A. C. (2004). Land, language, and loci: MtDNA in Native Americans and the genetic history of Peru. *American Journal of Physical Anthropology*, 127(3), 351-360.

Lindo, J., Haas, R., Hofman, C., Apata, M., Moraga, M., Verdugo, R. A., … Di Rienzo, A. (2018). The genetic prehistory of the Andean highlands 7000 years BP though [sic] European contact. *Science Advances*, 4(11), eaau4921.

Lipschutz, A. (1966). La despoblación de las Indias después de la Conquista. *América Indígena,* 26(3), 229-247.

Little, M. A., Thomas, R. B., & Garruto, R. M. (2013). A half century of high-altitude studies in anthropology: Introduction to the plenary session. *American Journal of Human Biology*, 25(2), 148-150.

Liu, X. G., Hu, H. Y., Guo, Y. H., & Sun, Y.P. (2016). Correlation between Y chromosome microdeletion and male infertility. *Genetics and Molecular Research,* 15(2), 1-8.

Lombardo, U., Szabo, K., Capriles, J. M., May, J., Hutterer, R., Lehndorff, E., … Veit, H. (2013). Early and middle Holocene hunter-gatherer occupations in western Amazonia: The hidden shell middens. *PloS ONE*, 8(8), e72746

Loogväli, E. L., Roostalu, U., Malyarchuk, B. A., Derenko, M. V., Kivisild, T., Metspalu, E., … Villems, R. (2004). Disuniting uniformity: A pied cladistic canvas of mtDNA haplogroup H in Eurasia. *Molecular Biology and Evolution*, 21(11), 2012-2021.

Luiselli, D., Simoni, L., Tarazona-Santos, E., Pastor, S., & Pettener, D. (2000). Genetic structure of Quechua-speakers of the Central Andes and geographic patterns of gene frequencies in South Amerindian populations. *American Journal of Physical Anthropology*, 113(1), 5-17.

Malhi, R. S., Gonzalez-Oliver, A., Schroeder, K. B., Kemp, B. M., Greenberg, J. A., Dobrowski, S. Z., … Brovko, T. (2008). Distribution of Y chromosomes among native North Americans: A study of Athapaskan population history. *American Journal of Physical Anthropology*, 137(4), 412-424.

Malhi, R. S., Cybulski, J. S., Tito, R. Y., Johnson, J., Harry, H., & Dan, C. (2010). Brief communication: Mitochondrial haplotype C4c confirmed as a founding genome in the Americas. *American Journal of Physical Anthropology,* 141(3), 494-497.

Map Data: Google. (2018a). *Distance calculations in Peruvian Amazonia: Automobile feature*. Retrieved October 10, 2017 from https://www.google.com/maps/@38.9696879,-95.2406428,15z

Map Data: Google. (2018b). *Distance calculations in Peruvian Amazonia: Direct orthodromic distance*. Retrieved October 10, 2017 from https://www.google.com/maps/@38.9696879,-95.2406428,15z

Map Data: Google. (2019a). *Location of Yurimaguas within Peru and wider Amazonia*. Retrieved May 25, 2019 from https://www.google.com/maps/place/Yurimaguas,+Peru/@-5.9012202,-80.6028216,6z/data=!4m5!3m4!1s0x91b9ea9bad83c245:0x979dbf93f1a22c35!8m2!3d-5.9007717!4d-76.1129048

Map Data: Google. (2019b). *Satellite image of the greater Yurimaguas area, at the confluence of the Huallaga and Paranapura Rivers*. Retrieved May 25, 2019 from https://earth.google.com/web/@5.89609332,76.11923532,144.76101506a,21047.44225058d,35y,0h,0t,0r

Marrero, A. R., Silva-Junior, W. A., Bravi, C. M., Hutz, M. H., Petzl-Erler, M. L., Ruiz-Linares, A., … Bortolini, M. C. (2007). Demographic and evolutionary trajectories of the Guaraní and Kaigang natives of Brazil. *American Journal of Physical Anthropology*, 132(2), 301-310.

Marrett, C. B. (1973). Social stratification in urban areas. In A. H. Hawley, & V. P. Rock (Eds.), *Segregation in residential areas: Papers on racial and socioeconomic factors in choice of housing* (pp. 172-188). Washington, DC: National Academy of Sciences.

Martinez-Laso, J., Siles, N., Moscoso, J., Zamora, J., Serrano-Vela, J. I., Ira-Cachafeiro, J., … Arnaiz-Villena, A. (2005). Origin of Bolivian Quechua Amerindians: Their relationship with other American Indians and Asians according to HLA genes. *European Journal of Medical Genetics*, 49(2), 169-185.

Maziéres, S., Guitard, E., Crubézy, E., Dugoujon, J.-M., Bortolini, M. C., Bonatto, S. L., … Salzano, F. M. (2008). Uniparental (mtDNA, Y-chromosome) polymorphisms in French Guiana and two related populations – implications for the region's colonization. *Annals of Human Genetics*, 72(1), 145-156.

McDonald, J. H. (2014). *Handbook of biological statistics (third edition)*. Baltimore, MD: Sparky House Publishing.

Meggers, B. (1971). *Amazonia: Man and culture in a counterfeit paradise*. Chicago, IL: Aldine Publishing Company.

Merriwether D. A., Kemp B. M., Crews, D. E., Neel, J. V. (2000). Gene flow and genetic variation in the Yanomama as revealed by mitochondrial DNA. In C. Renfrew (Ed.) *America Past, America Present: Genes and Languages in the Americas and Beyond* (pp. 89-124). Cambridge, United Kingdom: McDonald Institute for Archaeological Research.

Messina, F., Di Corcia, T., Ragazzo, M., Sanchez Mellado, C., Contini, I., Malaspina, P., … Jodice, C. (2018). Signs of continental ancestry in urban populations of Peru through autosomal STR loci and mitochondrial DNA typing. *PLoS ONE*, 13(7), e0200796

Moore, D. (2018). Following ghosts in northern Peru: In the footsteps of 19th century travelers on the old Moyobamba Route. Independently published.

Morey Alejo, H., & Sotil García, G. D. (2000). *Panorama histórico de la Amazonía Peruana: Una visión desde la Amazonía*. Iquitos, Perú: Municipalidad Provincial de Maynas.

Moscoso, J., Seclen, S., Serrano-Vela, J. I., Villena, A., Martinez-Laso, J., Zamora, J., … Arnaiz-Villena, A. (2006). HLA genes in Lamas Peruvian-Amazonian Amerindians. *Molecular Immunology*, 43(11), 1881-1889.

Myers, T. (1974). Spanish contacts and social change on the Ucayali River, Peru. *Ethnohistory*, 21(2), 135-157.

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

Nei, M., & Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences USA*, 76(10), 5269-5273.

Nordborg, M. (2003). Coalescent theory. In D. Balding, M. Bishop, & C. Canninga (Eds.), *Handbook of statistical genetics, second Edition*. New York, NY: John Wiley & Sons Ltd.

Norris, E. T., Rishishwar, L., Wang, L., Conley, A. B., Chande, A. T., Dabrowski, A. M., … King Jordan, I. (2019). Assortative mating on ancestry-variant traits in admixed Latin American populations. *Frontiers in Genetics*, 10:359.

O'Rourke, D. H., & Raff, J. (2010). The human genetic history of the Americas: The final frontier. *Current Biology*, 20(4), R202-R207.

Orefici, G. (2013). Ocupación humana, fases culturales y iconográfica de los petroglifos del Balsapuerto y Yurimaguas. In R. Bustamante, A. Lozano, G. Navarro, G. Orefici, & S. Rivas (Eds.), *Lo que las Piedras Cuentan: Cumpanamá y los Petroglifos de Balsapuerto* (pp. 162-183). Lima, Perú: Terra Nuova.

Oveido, F. D., Valdes, G. (1851). *Historia general y natural de las Indias, islas, y tierra firme del mar océano*. Madrid, España: Imprenta de la Real Academia de la Historia.

Palacios, S. (1890). *Segundo informe del presidente de la comision especial, coronel Samuel Palacios Mendiburu*. In C. Larrabure I Correa (Ed.), *Colección de leyes, decretos, resoluciones y otras documentos oficiales referents al Departamento de Loreto, vol. 4* (pp. 431-469). Lima, Perú: Oficina Tipográfica de la Opinión Nacional: Gremios 129.

Palanichamy, M. G., Sun, C., Agrawal, S., Bandelt, H.-J., Kong, Q.-P., Khan, F., … Zhang, Y.-P. (2004). Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia. *American Journal of Human Genetics,* 75(6), 966-978.

Park, C. C. (1992). *Tropical Rainforests*. Abingdon, England: Routledge.

Pärssinen, M., Schaan, D., & Ranzi, A. (2009). Pre-Columbian geometric earthworks in the Upper Purus: A complex society in western Amazonia, *Antiquity,* 83(322), 1084-1095.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine*, 5(50), 157-175.

Pennano, G. (1988). *La Economía del Caucho*. Iquitos, Perú: Centro de Estudios Teológicos de la Amazonía.

Perego, U. A., Achilli, A., Angerhofer, N., Accetturo, M., Pala, M., Olivieri, A., (2009). Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. *Current Biology*, 19(1), 1-8.

Politis, G. 2009. *Nukak: Ethnoarchaeology of an Amazonian People (first edition)*. Abingdon, England: Routledge.

Posth, C., Nakatsuka, N., Lazaridis, I., Skogland, P., Mallick, S., Lamnidis, T. C., … Reich, D. (2018). Reconstructing the deep population history of Central and South America. *Cell*, 175(5), 1185-1197.

R Core Team. (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org/

Rademaker, G., Hodgins, G., Moore, K., Zarrillo, S., Miller, C., Bromley, G. R., … Sandweiss, D. H. (2014). Paleoindian settlement of the high-altitude Peruvian Andes. *Science*, 346(6208), 466-469.

Raff, J. A., & Bolnick, D. A. (2015). Does mitochondrial haplogroup X indicate ancient trans-Atlantic migration to the Americas? A critical re-evaluation. *PaleoAmerica*, 1(4), 297-304.

Raghavan, M., Skoglund, P., Graf, K. E., Metspalu, M., Albrechtsen, A., Moltke, I., … Willerslev, E. (2014). Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature*, 505 (7481), 87-91.

Raghavan, M., Steinrücken, M., Harris, K., Schiffels, S., Rasmussen, S., DeGiorgio, M., … Willerslev, E. (2015). Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science*, 349(6250), aab3884.

Ramallo, V., Bisso-Machado, R., Bravi, C., Coble, M. D., Salzano, F. M., Hünemeier, T., & Bortolini, M. C. (2013). Demographic expansions in South America: Enlightening a complex scenario with genetic and linguistic data. *American Journal of Physical Anthropology*, 150(3), 453-463.

Ramos, V. A, & Folguera, A. (2009). Andean flat-slab subduction through time. In J. B. Murphy, J. D. Keppie, & A. J. Hynes (Eds.), *Ancient orogens and modern analogues 327* (pp. 31-54). London, England: The Geologic Society of London.

Ramos-Onsins, S. E., & Rozas, J. (2002). Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, 19(2), 2092-2100.

Rani, D. S., Dhandapany, P. S., Nallari, P., Govindaraj, P., Singh, L., Thangaraj, K. (2010). Mitochondrial DNA haplogroup 'R' is associated with Noonan syndrome of South India. *Mitochondrion*, 10(2), 166-173.

Ranzi, A. (2003). Geoglifos: Patrimonio cultural do Acre. In M. Parssinen, & A. Korpisaari (Eds.), *Western Amazonia — Amazonia Occidental* (pp. 135-172). Helsinki, Finland: Renvall Institute Publications.

Ray, N., Currat, M., & Excoffier, L. (2003). Intra-deme molecular diversity in spatially expanding populations. *Molecular Biology and Evolution*, 20(1), 76-86.

Reeve, M.-E. (1993). Regional interaction in the western Amazon: the early colonial encounter and the Jesuit years: 1538-1767. *Ethnohistory,* 41(1), 106-138.

Reff, D. T. (1991). Disease, depopulation, and culture change in northwestern New Spain, 1518-1764. Salt Lake City, UT: University of Utah Press.

Rego-Pérez, I., Fernández-Moreno, M., Fernández-López, C., Arenas, J., & Blanco, F. J. (2008). Mitochondrial DNA haplogroups: Role in the prevalence and severity of knee osteoarthritis. *Arthritis and Rheumatology,* 58(8), 2387-2396.

Reich, D., Patterson, N., Campbell, D., Tandon, A., Mazieres, S., Ray, N., … Ruiz-Linares, A. (2012). Reconstructing Native American population history. Nature, 488(7411), 370-374.

Reidla, M, Kivisild, T, Metspalu, E, Kaldma, K, Tambets, K, Tolk, H-V., … Villems, R. (2003). Origin and diffusion of mtDNA haplogroup X. *The American Journal of Human Genetics*, 73(5), 1178-1190.

Rhoades, R. E., & Bidegaray, P. (1987). *The Farmers of Yurimaguas: Land Use and Cropping Strategies in the Peruvian Jungle.* Lima, Perú: International Potato Center.

Rickards, O., Martínez-Labarga, C., Lum, J. K., De Stefano, G. F., Cann, R. L. (1999). MtDNA history of the Cayapa Amerinds of Ecuador: Detection of additional founding lineages for the Native American populations. *American Journal of Human Genetics*, 65(2), 519-530.

Rivas, S. (2003). *Los asentamientos prehispánicos de la cuenca del Río Cachiyacu, Amazonía Peruana*. Lima, Perú: Instituto Cultural RVNA.

Röck, A. W., Dür, A., van Oven, M., Parson, W. (2013). Concept for estimating mitochondrial DNA haplogroups using a maximum likelihood approach (EMMA). *Forensic Science International: Genetics*, 7(6), 601-609.

Rodrigues, A. D. (1999). *Tupí*. In R. M. W. Dixon, & A. Y. Aikhenvald (Eds.), *The Amazonian languages (Cambridge language surveys)* (pp. 107-124). New York, NY: Cambridge University Press.

Rodriguez-Delfin, L. A., Rubin-de-Celis, V. E., & Zago, M. A. (2001). Genetic diversity in an Andean population from Peru and regional migration patterns of Amerindians in South America: Data from Y chromosome and mitochondrial DNA. *Human Heredity*, 51(1-2), 97-106.

Rogers, A. R. (1990). Group selection by selective emigration: The effects of migration and kin structure. *The American Naturalist*, 135(3), 398-413.

Rogers, A. R., & Harpending, H. (1992). Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, 9(3), 552-569.

Romero, S. (2009a). Americas: Fatal clashes erupt in Peru at roadblock. *The New York Times*. June 5, 2009. Retrieved July 7, 2019 from https://www.nytimes.com/2009/06/06/world/americas/06peru.html?mtrref=www.google.com&gwh=8DE723BE16CA6530841748E701A4FA1D&gwt=pay

Romero, S. (2009b). Americas: 9 hostage officers killed at Peruvian oil facility. *The New York Times*. June 6, 2009. Retrieved July 7, 2019 from

https://www.nytimes.com/2009/06/07/world/americas/07peru.html?mtrref=www.google.com&gwh=DABD5B9D6F1C726A4A2C25D24256D71A&gwt=pay

Romero-Hidalgo, S., Ochoa-Leyva, A., Garcíarrubio, A., Acuña-Alonzo, V., Antúnez-Argüelles, E., Balcazar-Quintero, M., … Soberón, X. (2017). Demographic history and biologically relevant genetic variation of Native Mexicans inferred from whole-genome sequencing. *Nature Communications*, 8(1005), DOI: 10.1038/s41467-017-01194-z

Roosevelt, A. (1999). Maritime, highland, forest dynamic. In F. Salomon, S. Schwartz (Eds.), *The Cambridge history of the Native peoples of the Americas (vol. 3)* (pp. 264-349). Cambridge, United Kingdom: Cambridge University Press.

Rothhammer, F., & Dillehay, T. D. (2009). The late Pleistocene colonization of South America: An interdisciplinary perspective. *Annals of Human Genetics*, 73(5), 540-549.

Ruiz-Alejos, A., Carrillo-Larco, R. M., Miranda, J. J., Anderson, C. A. M., Gilman, R. H. Smeeth, L., and Bernabé-Ortiz, A. (2018). Addressing the impact of urban exposure on the incidence of type 2 diabetes mellitus: The PERU MIGRANT study. Scientific Reports, 8(1), 5512. https://doi.org/10.1038/s41598-23812-6.

Rumrrill, R., Davila Herrera, C., & Garcia, F. B. (1986). *Yurimaguas, capital histórica de la Amazonía Peruana*. Yurimaguas, Perú: Consejo Provincial de Alto Amazonas.

Salzano, F. M., & Sans, M. (2014). Interethnic admixture and the evolution of Latin American populations. *Genetics and Molecular Biology*, 37(1 supplement), 151-170.

Sánchez-Albornoz, N. (1974). The population of Latin America: A history. Berkeley, CA: University of California Press.

Sandoval, J. R., Salazar-Granara, A., Acosta, O., Castillo-Herrera, W., Fujita, R., Pena, S. D. J., & Santos, F. R. (2013). Tracing the genomic ancestry of Peruvians reveals a major legacy of pre-Columbian ancestors. *Journal of Human Genetics*, 58(9), 627-634.

Sanger, F., Micklen, S., & Coulson, A. R. (1977). DNA sequencing and chain-terminating inhibitors. *Proceedings of the National Academy of Sciences USA,* 74(12), 5463-5767.

Santana, C., Noris, G., Meraz-Ríos, M. A., Magaña, J. J., Calderon-Aranda, E. S. Muñoz, Mde. L., & Gómez, R. (2014). Genetic analysis of 17 Y-STRs in a Mestizo population from the Central Valley of Mexico. *Human Biology*, 86(4), 289-312.

Santos-Granero, F. (1985). Crónica breve de un etnocidio o la génesis del mito del “gran vacío Amazónico.” *Amazonía Peruana*, 6(11), 9-38.

Santos-Granero, F. (1987). Epidemias y sublevaciones en el desarrollo demográfico de las misiones Amuesha del Cerro de la Sal, siglo XVIII. *Histórica*, 11(1), 25-53.

Santos-Granero, F. (1992). *Etnohistoria de la Alta Amazonía: Siglos XV-XVIII*. Quito, Ecuador: Abya-Yala.

Santos-Granero, F., & Barclay, F. (1998). *Selva Central, history, economy, and land use in Peruvian Amazon.* Washington, DC: Smithsonian Institution Press.

Santos-Granero, F., & Barclay, F. (2000). *Tamed Frontiers: Economy, society, and civil rights in Upper Amazonia*. Boulder, CO: Westview Press.

Schaan, D., Pärssinen, M., Ranzi, A., & Piccoli, J. (2007). Geoglifos da Amazônia ocidental:

Evidência de complexidade social entre povos da terra firme. *Revista de Arqueología*. 20, 67-82.

Schjellerup, I. (2005). *Incas y Españoles en la Conquista de los Chachapoya*. Lima, Perú: Fondo Editorial PUCP.

Schneider, S., & Excoffier, L. (1999). Estimation of demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: Application to human mitochondrial DNA. *Genetics*, 152(3), 1079-1089.

Scliar, M., Gouveia, M., Benazzo, A., Ghirotto, S., Fagundes, N., & Leal, T. (2014). Bayesian inferences suggest that Amazon Yunga natives diverged from Andeans less than 5000 YBP: Implications for South American prehistory. *BMC Evolutionary Biology*, 14:174.

Seielstad, M. (2000). Asymmetries in the maternal and paternal genetic histories of Colombian populations. *American Journal of Human Genetics*, 67(5), 1062–1066.

Shlush, L. I., Behar, D. M., Yudkovsky, G., Templeton, A., Hadid, Y., Basis, F., … Skorecki, K. (2008). The Druze: A population genetic refugium of the Near East. *PLoS ONE*, 3(5), e2105.

Skoglund, P., Mallick, S., Bortolini, M. C., Chennagiri, N., Hünemeier, T., Petzl-Erler, M. L., … Reich, D. (2015). Genetic evidence for two founding populations of the Americas. *Nature*, 525(7567), 104-108.

Slatkin, M., & Hudson, R. R. (1991). Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics*, 129(2), 555-562.

Smith, A. (1994). *Explorers of the Amazon*. Chicago, IL: University of Chicago Press.

Soares, P., Trejaut, J. A., Loo, J. H., Hill, C., Mormina, M., Lee, C. L., … Richards, M. B. (2008). Climate change and postglacial human dispersals in Southeast Asia. *Molecular Biology and Evolution*, 25(6), 1209-1218.

Sperandei, S. (2014). Understanding logistic regression analysis. *Biochemia Medica*, 24(1), 12-18.

Sprinthall, R. C. (2011). *Basic statistical analysis (ninth edition)*. Upper Saddle River, NJ: Pearson Education.

Stanfield, M. E. (1998). *Red rubber, bleeding trees: Violence, slavery and empire in the Upper Amazon, 1850-1933.* Albuquerque, NM: University of New Mexico Press.

Starikovskaya, E. B., Sukernik, R. I., Derbeneva, O. A., Volodko, N. V., Ruiz-Pesini, E., Torroni, A., … Wallace, D. C. (2005). Mitochondrial DNA diversity in Indigenous populations of the southern extent of Siberia, and the origins of Native American haplogroups. *Annals of Human Genetics*, 69(Pt. 1), 67-89.

Starn, O., Degregori, C. I., & Kirk, R. (2005a). Introduction. In O. Starn, C. I. Degregori, & R. Kirk (Eds.), *The Peru Reader: History, culture, politics (second edition)* (pp. 1-12). Durham, NC & London, United Kingdom: Duke University Press.

Starn, O., Degregori, C. I., & Kirk, R. (2005b). The ancient civilizations. In O. Starn, C. I. Degregori, & R. Kirk (Eds.), *The Peru Reader: History, culture, politics (second edition)* (pp. 13-16). Durham, NC & London, United Kingdom: Duke University Press.

Stearn, E. W., & Stearn, A. E. (1945). *The effect of smallpox on the density of the Indigenous.* Boston, MA: Humphries.

Student (Gossett, W. S.). (1908). The probable error of a mean. *Biometrika*, 6(1), 1-25.

Tabbada, K. A., Trejaut, J., Loo, J. H., Chen, Y. M., Lin, M., Mirazón-Lahr, M., … De Ungria, M. C. (2010). Philippine mitochondrial DNA diversity: A populated viaduct between Taiwan and Indonesia?. *Molecular Biology and Evolution*, 27(1), 21-31.

Taboada-Echalar, P., Álvarez-Iglesias, V., Heinz, T., Vidal-Bralo, L., Gómez-Carballa, A., Catelli, L., … Salas, A. The genetic legacy of the pre-colonial period in contemporary Bolivians. *PLoS ONE*, 8(3), e58980.

Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics*, 105(2), 437-460.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics,* 123(3), 585-595.

Tajima, (1993a). Measurement of DNA polymorphism. In N. Takahata, & A. G. Clark (Eds.), *Mechanisms of molecular evolution: Introduction to molecular paleopopulation biology* (pp. 37-59). Sunderland MA: Sinauer Associates.

Tajima, F. (1993b). Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics*, 135(2), 599-607.

Tajima, F. (1996). The amount of DNA polymorphism maintained in a finite population when the neutral mutation rate varies among sites. *Genetics*, 143(3), 1457-1465.

Takenaka, A. (2004). The Japanese in Peru: History of immigration, settlement, and racialization. *Latin American Perspectives*, 31(3), 77-98.

Tamm, E., Kivisild, T., Reidla, M., Metspalu, M., Smith, D. G., Mulligan, C. J., … Malhi, R. S. (2007). Beringian standstill and spread of Native American founders. *PLoS ONE*, 2(9), e829.

Tapayuri Murayari, B. (2012). Preface. In B. Dean, & J. Homan (Eds.), *Anthropological illuminations of the varieties of human experience (second edition)* (pp. vii). Dubuque, IA: Kendall Hunt Publishing.

Tarazona-Santos, E., Carvalho-Silva, D. R., Pettener, D., Luiselli, D., De Stefano, G. F. Labarga, C. M., … Santos, F. R. (2001). Genetic differentiation in South Amerindians is related to environmental and cultural diversity: Evidence from the Y chromosome. *American Journal of Human Genetics*, 68(6), 1485-1496.

Taussig, M. (1984). Culture of terror, space of death: Roger Casement's Putumayo report and the explanation of torture. *Comparative Studies in Society and History,* 26(3), 467-497.

Tiirikka, T., & Moilanen, J. S. (2015). Human Chromosome Y and Haplogroups; introducing YDHS Database. *Clinical Translational Medicine*, 4:20.

Toscanini, U., Gusmão, L., Berardi, G., Gomes, V., Amorim, A., Salas, A., & Raimondi, E. (2011). Male lineages in South American native groups: Evidence of M19 traveling south. *American Journal of Physical Anthropology*, 146(2), 188-196.

Trujillo, V. (1981). *La legislación eclesiástica en el virreinato del Perú durante el siglo XVI.* Lima, Perú: Lumen Publishing.

Tukey, J. (1949). Comparing individual means in the analysis of variance. *Biometrics*, 5(2), 99-114.

United Nations. (2014). UNHCR: The UN refugee agency. *Global trends: Forced displacement in 2014*. Retrieved from https://www.unhcr.org/enus/statistics/country/556725e69/unhcr-global-trends-2014.html

United Nations. (2016). Department of economic and social affairs, population division. *The world's cities in 2016*. Retrieved from https://www.un.org/en/development/desa/population/publications/pdf/urbanization/the_worlds_cities_in_2016_data_booklet.pdf

United Nations. (2017). Department of economic and social affairs, population division. *Trends in international migrant stock: The 2017 revision*. https://www.un.org/en/development/desa/population/migration/index.asp

United Nations. (2018a). Department of economic and social affairs, population division. *The world's cities in 2018*. Retrieved from https://www.un.org/en/events/citiesday/assets/pdf/the_worlds_cities_in_2018_data_booklet.pdf

United Nations. (2018b). Department of economic and social affairs, population division. *World urbanization prospects*. Retrieved from https://population.un.org/wup/Country-Profiles/

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, Y., Faircloth, B. C., Remm, M., Rozen, S. G. (2012). Primer3—new capabilities and interfaces. *Nucleic Acids Research*, 40(15), e115.

van Oven, M. (2015). PhyloTree Build 17: Growing the human mitochondrial DNA tree. *Forensic Science International, Genetics Supplement Series 5, e392-e394.*

van Oven, M., Ralf, A., & Kayser, M. (2011). An efficient multiplex genotyping approach for detecting the major worldwide human Y-chromosome haplogroups. *International Journal of Legal Medicine*, 125(6), 879-885.

van Oven, M., & Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human Mutation*, 30(2), e386-e394.

Vila, P. (2000). *Crossing borders, reinforcing borders: Social categories, metaphors, and narrative Identities on the U.S.-Mexico frontier*. Austin, TX: University of Texas Press.

Volodko, N. V., Starikovskaya, E. B., Mazunin, I. O., Eltsov, N. P., Naidenko, P. V., Wallace, D. C., & Sukernik, R. I. (2008). Mitochondrial genome diversity in Arctic Siberians, with particular reference to the evolutionary history of Beringia and Pleistocenic peopling of the Americas. *The American Journal of Human Genetics,* 82(5),

Wang, S., Lewis, C. M., Jakobsson, M., Ramachandran, S., Ray, N., Bedoya, G., … Ruiz-Linares, A. (2007). Genetic variation and population structure in Native Americans. *PLoS Genetics*, 3(11), e185.

Wang, S., Ray, N., Rojas, W., Parra, M. V., Bedoya, G., Gallo, C., … Linares, A. (2008). Geographic patterns of genome admixture in Latin American mestizos. *PLoS Genetics*, 4(3), e1000037.

Waters, M. R. (2019). Late Pleistocene exploration and settlement of the Americas by modern humans. *Science*, 365(6449), eaat5447.

Watterson, G. A. (1975). On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology,* 7(2), 256-276.

Weinstein, B. (1983). *The Amazon rubber boom, 1850-1920*. Palo Alto, California: Stanford University Press.

Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.-J., … Schönherr, S. (2016). HaploGrep 2: Mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Research*, 44(W1), w58-w63.

Wickramage, K., Vearay, J., Zwi, A. B., Robinson, C., & Knipper, M. (2018). Migration and health: A global public health research priority. *BMC Public Health*, 18:987.

Wirth, L. (1938). Urbanism as a way of life. *The American Journal of Sociology*. 44(1), 1-24.

World Bank. (2014). *Urban population (% of total)*. United Nations Population Division, world urbanization prospects: 2014 revision. World Bank open data. Retrieved June 17, 2018 from https://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS

World Bank. (2017). *Net migration*. United Nations Population Division, world population prospects: 2017 revision. World Bank open data. Retrieved September 13, 2018 from https://data.worldbank.org/indicator/SM.POP.NETM?year_high_desc=tru

World Wildlife Fund. (2017). *Places: Amazon*. Retrieved February 23, 2017 fromhttp://www.worldwildlife.org/places/amazon

Wright, S. (1978). *Evolution and the genetics of populations: Variability within and among natural populations.* Chicago, IL: University of Chicago Press.

Yang, I. S., Lee, H. Y., Yang, W. I., & Shin, K. J. (2013). MtDNAprofiler: A Web application for the nomenclature and comparison of human mitochondrial DNA sequences. *Journal of Forensic Sciences*. 58(4), 972-980.

Yao, Y.-G., Kong, Q.-P., Bandelt, H.-J., Kivisild, T., & Zhang, Y.-P. (2002). Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *American Journal of Human Genetics*, 70(3), 635-651.