DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY
Office for Scholarly Communication

Genetic Diversity in Urban and Rural Indigenous Mexico

Citation
López, John Emilio William. 2016. Genetic Diversity in Urban and Rural Indigenous Mexico. Master's thesis, Harvard Extension School.

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:33797263

Terms of Use
This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

Genetic Diversity in Urban and Rural Indigenous Mexico

John Emilio William López

A Thesis in the Field of Biotechnology

For the Degree of Master of Liberal Arts in Extension Studies

Harvard University

March 2016

Copyright 2016 John Emilio William López

Abstract

The goal of this study was to examine the genetic diversity of indigenous (and semi-indigenous) populations in Mexico and determine if any genetic variation correlated to culture *and* language.

Mexico's indigenous populations have not been extensively studied using contemporary genetic typing systems; the ones conducted used too few loci, or very small sample sizes. Mexico has 291 living native languages derived from twelve separate and distinct linguistic families divided into 68 major linguistic groups, and an indigenous culture just as diverse. Previous notions that ancient Mexican civilizations were "obliterated" were wrong—they were simply transformed—and many ideas of pre-colonial indigenous-Spanish relations were also recently determined 'incorrect', thus presenting us with redefined variables to explore, genetically.

To pursue the potential link between genetic diversity, native culture *and* language preservation, I selected small indigenous towns in Mexico where native languages were still commonly spoken, and native culture conservation deemed high. I used STR profiles and mtDNA sequence data, in conjunction with corresponding questionnaire data, to compare and contrast genetic profiles of individuals grouped according to Mexican town based on culture and native language preservation. This information was then compared to previously published population studies of various peoples.

The chief results emerging from this study were that *our* sampled populations: (1) had the highest *intra*-population variance in North America, (2) had the lowest *inter*-population variance in North America, (3) maintained low mtDNA haplotype

polymorphism while yielding high mtDNA haplogroup diversity, and (4) exhibited no traces of haplotype X2a, which one typically expects to find in native North American populations.

This study confirmed that Mexican haplotypes are mainly derived from the common Native American haplogroups of A2, B2, C1, and D1, and that indigenous American populations exhibit low haplogroup polymorphism when compared to Caucasian populations. It also forces reexamination of previously held notions of the genetics of Mexico. Historical records described that many European women migrated to Mexico, thus why large numbers of mtDNA with European lineage are expected to be found in Mexico. Our findings suggested the opposite. Native Mexicans preserved their native mtDNA lineage, and Europeans did not contribute much genetic influence; genetic drift was the main driver of mtDNA diversity. Similarly, another study showed that there was no significant link between indigenous Mexicans communities who maintained a high proportion of culture and language preservation—our entire study suggested this to be unfounded.

Table of Contents

List of Tables vi

List of Figures vii

I. Introduction 1

II. Materials & Methods 20

Sample Collection.. 20

DNA Amplification and STR Genotyping 22

Previously Published Data.. 23

Computational Resources 23

Statistical Analysis 24

III. Results 27

STR Intra-population Diversity 27

STR Inter-population Diversity 29

Haplogroup and Haplotype Distributions 32

IV. Discussion 36

V. Appendix 48

References 63

List of Tables

Table 1. Characteristics of Alleles of the 15 Core Loci …………………………….……..…. 3

Table 2. Number of Native Language Speakers in Mexico …………………….………..….. 18

Table 3. Previously Published Population Studies of Various Cohorts …………...….. 23

Table 4. Allelic Frequencies of Cuernavaca …………………….……………...…….. 48

Table 5. Allelic Frequencies of Puebla ………………………..…………….……… 49

Table 6. Allelic Frequencies of Taxco ……………………………..………………….. 50

Table 7. Allelic Frequencies of Xochicalco ……………………………..…………..…. 51

Table 8. Allelic Frequencies of UNAM Students ………………………......………. 52

Table 9. Parameters of Interest for Population Genetics ………………….......………. 53

Table 10. Inter- and Intra-Population Genetic Variance ………………...………..…… 54

Table 11. F_{st} values for 15 STR loci …………………………………......…………. 32

Table 12. Significant pairwise values for 15 STR loci …………………......…………. 32

Table 13. F_{st} values for mtDNA haplotypes ……………………...……...…………. 33

Table 14. Significant pairwise values for mtDNA haplotypes ……….........…………. 33

Table 15. Number of samples per mtDNA haplogroup ……………………………..…. 34

Table 16. Number of Samples per mtDNA-lineage origin ……………………………. 35

Table 17. Summary statistics for mtDNA …………………………….………………….. 35

List of Figures

Figure 1. Languages in Mexico spoken by more than 100k people …………...…...….. 11

Figure 2. Languages in Mexico spoken by 20k to 100k people ………………......…... 12

Figure 3. Languages in Mexico spoken by less than 20k people ……………………..….. 12

Figure 4. Neighbor-Joining tree of sampled Native Mexican groups ……………...….. 30

Figure 5. Dendogram of Native Mexican groups …………………………...…...….. 31

Figure 6. Lineage origins of mtDNA for Native Mexican groups ……….……….…… 55

Chapter I
Introduction

The mechanisms of evolution, heredity, and phenotypic variation have always been controversial ones, and since the time of Darwin people always had varying schools of thought on the matter. Some of his contemporaries believed in the “classical hypothesis,” and others in the “balance hypothesis,” but what *was* scientifically unanimous was that evolution *did* occur (Hamilton, 2009). Mendel’s seminal paper *Versuche über Pflanzenhybriden* published in 1866, provided the first insight on how interbred populations evolved thus laying the foundation for all future formulas of population genetics, and it was on Mendel’s work that Hardy and Weinberg developed their equilibrium formula (Weinberg, 1908; Hardy, 1908). The two proposed that within a randomly mating population, genotype frequencies at any given locus will remain constant, thereby allowing scientists to predict a population’s genotype frequency from its allele frequency (Hamilton, 2009; Goodwin *et al*., 2011); though this only held true if certain conditions were fulfilled: (1) populations must be very large; (2) populations must be isolated from other population; (3) no mutations; (4) random mating; and (5) no natural selection.

In real-world human populations the Hardy-Weinberg equilibrium (HWE) conditions are rarely met, and this caveat required development of mathematical practices to account for these deviations (Hamilton, 2009). Even though advances were aggressively made to account for HWE deviations, it was the advances in DNA technology that finally allowed us to reliably measure genetic variation.

DNA technologies, for the investigation of genetic diversity and human migration patterns, have increased exponentially; those of particular note are: the development of the Polymerase Chain Reaction (PCR), for amplification of specific genome regions; the discovery of Short Tandem Repeat (STR) sequences, to identify and classify allelic variation; and the development of mitochondrial DNA (mtDNA) sequencing, to study matrilineal inheritance (Butler, 2006).

Short Tandem Repeats are genetic markers found in non-coding regions throughout the human genome. They range in size from two to six base-pairs, and those pairs repeat anywhere from five to 5000 times (Goodwin *et al.*, 2011), but most STRs are di-, tri-, tetra- or penta-nucleotide repeats of 20 to 50 times. They were first used for human identification in the early 1990s, and are the perfect candidate for population studies because they are polymorphic, multi-allelic, easy to amplify, and there is a consensus on the 13-15 loci used as markers (Edwards *et al.*, 1991; Edwards *et al.*, 1992; Butler, 2006). This set of STR markers have been referred to as the "core STR loci" and was selected because they are located on different chromosomes (or unlinked if on the same chromosome), and are expected to behave independently—in accordance with Mendel's second law of independent assortment (see Table 1).

Mutations at STR loci are relatively common—leading to high levels of polymorphism—but have little effect within a gene pool since their mutation rates average below 0.2% per generation (Goodwin *et al.*, 2011).

Most human-population genetic studies are based on random samples of 100-200 unrelated individuals. The STR allelic frequencies are calculated from the STR profiles and then compared to known frequencies of specified population groups.

TABLE 1. *Characteristics of alleles observed in 15 core STR loci.*

Locus Name	Chromosome Location	Repeat Motif	Allele Range	Variant Number
CSF1PO	5q33.1 c-fms proto-oncogene, 6th intron	TAGA	5--6	22
FGA	4q31.3 αfibrinogen, 3rd intron	CTTT	12.2--51.2	114
TH01	11p15.5 tyrosine hydroxylase, 1st intron	TCAT	3--14	22
TPOX	2p25.3 thyroid peroxidase, 10th intron	GAAT	4--16	23
VWA	12p13.31 von Willebrand Factor, 40th intron	[TCTG][TCTA]	10--25	20
D3S1358	3p21.31	[TCTG][TCTA]	8--21	30
D5S818	5q23.2	AGAT	7--18	20
D7S820	7q21.11	GATA	5--16	26
D8S1179	8q24.13	[TCTG][TCTG]	7--20	24
D13S317	13q31.1	TATC	5--16	18
D16S593	16q24.1	GATA	5--16	22
D18S51	18q21.33	AGAA	7--39.2	51
D21S11	21q21.1	Complex [TCTA][TCTG]	12--41.2	42
D2S1338	2q35	[TGCC][TTCC]	15--28	28
D19S433	19q12	AAGG	9--17.2	33

Source: Butler, 2006. These 495 alleles were current as of April 2015. For up to date information see http://www.cstl.nist.gov/biotech/strbase/var_tab.htm.

Published STR studies have revealed that high polymorphism occurs *among* human populations as well as *within* closely related ones (Vargas-Alarcon *et al.*, 2007; Ibarra-Rivera *et al.*, 2008; Reich *et al.*, 2009; Sandoval *et al.*, 2009). For example, a recent study of India's population (Reich *et al.*, 2009), showed that different groups of Indians had diverged so significantly over time that they were now more genetically similar to foreign global populations than to their own.

In the above study, researchers genotyped 132 individuals from India, from 25 different population groups, and found that two ancient and genetically-divergent populations were ancestral to most of today's Indians (Reich *et al.*, 2009). The first group called the Ancestral North Indians (ANI), are genetically similar to Middle Easterners, Central Asians, and Europeans, whereas the second group, called Ancestral South Indians (ASI), are genetically distinct from the ANI. The measure of population differentiation and genetic divergence between populations, also known as fixation index (F_{ST}), of the 19 Indian groups was averaged to be 0.0109, and the average F_{ST} (calculated

in a recent study of 23 groups of Europeans (Lao *et al.*, 2008)), was 0.0033, implying that there is more allele frequency variation within India than there is within all of Europe. This high differentiation among the Indian population is hypothesized to result from strong founder effects due to culture practices of endogamy (the human practice of marrying within a specific ethnic, class, or social group), over many generations. These results also suggest that we should see higher rates of recessive inherited diseases in India. Consequently, we should begin to factor in "population stratification" as a confounder in gene mapping studies (Reich *et al*., 2009).

Present day Mexico has one of the richest ethnic *and* linguistic diversities on the North American continent (Gonzalez-Martin *et al.*, 2008), yet only 40 years ago there was absolutely *no* study of pre-colonial Mexico (Kepics *et al*., 2005). Scholars believed that ancient Mexican civilizations were primitive, and that their abundance of culture was attributed to the Spanish-colonial era, but in fact they were very advanced. Ranked on par with the ancients of Egypt, Mesopotamia, and China, the Mesoamerican civilizations boasted kingdoms, empires, cities, monuments, writing, art, belief systems, and metallurgy (Willey, 1965).

"Meso-America" literally means "Middle-America" but colloquially refers to the civilized parts of Mexico and Central America in pre-colonial times. Irrefutable evidence showed that hunters occupied North America by 11 000 BCE, and in 1967 archaeologists discovered a site southeast of Mexico City that contained a well-crafted obsidian blade that was radiocarbon dated to 21 000 BCE. Another excavation in Puebla, Mexico uncovered remains of hunters who captured and fed on mammoths and other extinct animals using an extensive variety of well-made worked-point tools such as knives,

burins, perforators, and scrapers. They were carbon dated to 21 800 BCE (Willey, 1965). Suffice it to say that advanced humans have been living in Mesoamerica for a very long time.

The natural environment was very different for the first Mesoamericans. The ice sheets and caps in the northern regions of North America caused lower temperatures, and local volcanoes actively spread ash over thousands of miles, but around 7000 BCE the temperature increased allowing sustainability of agriculture. Attempted plant cultivation in Mesoamerica was dated as far back as 7000 BCE (Willey, 1965).

Cave evidence from Oaxaca, Mexico placed Mesoamerican plant cultivation as far back as 8900 BCE, and further evidence showed that the *selection* and planting of seeds for chili peppers, corn, and one variety of squash, occurred at 6500 BCE (Coe, 1987; Willey, 1965). Mesoamerican farming of chili peppers, corn, all squash, beans, and cotton was fully established by 1500 BCE (Willey, 1965). Pottery was dated as far back as 2300 BCE and is revealing because it's typically a strong indicator of a group's commitment to permanent settlement due to the fragility, and difficulty to transport ceramics (Coe, 1987; Willey, 1965).

Social change was visualized in Mesoamerica around 1200 BCE to 900 BCE when the building of stone pyramids and sculptures began. Such acts are indicative of people who developed greater political order, and this new group of highly organized people were named the Olmec. The Olmec eventually broke apart into four separate but powerful civilizations that practiced their own regional customs, traditions, and effectively defended and governed themselves; they were known as the Maya, Zapotec, Totonac, and Teotihuacán civilizations (Coe, 1987; Willey, 1965).

Zapotec and Mixtec are considered linguistic groups that contributed the most to Mesoamerican civilization because of their large and powerful empires that reigned from 2000 BCE to the arrival of the Spanish (Coe, 1987). The Zapotec lived in scattered villages throughout Oaxaca and Puebla, with no central location, but in 900 BCE they leveled-off hillside land in Oaxaca to build their capital. In 300 BCE archaeological evidence from Oaxaca showed that cultural influences from southeast Mexico started to permeate Zapotec culture (Willey, 1965).

Kepics (2005) research uncovered that not only were civilizations fully established and thriving in Mesoamerica 1500 years before the arrival of Columbus, but that they were very complex too. Upon their arrival, the Spaniards recorded that the New World they stumbled upon was already very similar to Europe. They were amazed at the bustling ports connected by river systems, and surprised that the indigenous natives had a complex economic trade system in place. Cacao was their trading standard, used as a form of currency; it was traded for goods that could not be found locally, such as salt from distant mines, fruit from better climates, and seafood from the coasts. Since no records exist documenting the time when to Spaniards set their first step on North America to when colonization was underway, Kepics (2005) postulated that most of the notions we held of pre-colonial native-Spanish relations were assumed, and incorrect. One example being the widely believed notion that Spaniards admixed with indigenous natives, or that natives intermarried with natives from other population groups. But if the transportation, economic, and political practices of native civilizations were analogous to those of the Europeans, is it possible that native civilizations practiced endogamy like their European contemporaries?

Anthropologists speculated that Mesoamerican indigenous native people intermixed and intermarried extensively with the European conquistadors over generations so that the amount of distinct native DNA left in pre-Columbian Mexico's gene pool was drastically minimized (Salazar-Flores, *et al.*, 2009), but we now know that this idea is not correct (Vargas-Alarcon *et al.*, 2007; Ibarra-Rivera *et al.*, 2008). While the Spanish conquest undoubtedly led to interbreeding and genetic admixture, the majority of Mexico's indigenous (and semi-indigenous) populations have not been extensively studied using contemporary genetic typing systems (Martinez-Cortes *et al.*, 2010). Those populations that have been studied were the subject of very few genetic studies (Sandoval *et al.*, 2009), and of those studies either only one locus was analyzed, or extremely small sample sizes were available (Moreno-Estrada *et al*., 2014). In order to develop a more accurate, and in-depth understanding of ancient Mesoamerican cultures and their admixture practices, more population genetic studies are needed.

In one recent study, scientists set out to compare DNA profiles of randomly sampled native groups throughout Mexico in attempt to gain a better understanding of the degree to which native Mexicans are genetically related. DNA sequencing of the variable regions in the mitochondrial genome (mtDNA) of 477 individuals from 11 native Mexican groups (Triqui, Tarahumara, Purépecha, Otomi, Mixtec, Nahua Xochimilco, Nahua Zitlala, Nahua Ixhuatlancillo, Nahua Necoxtla, Maya, and Pima), revealed maternal admixture rates of 89.1% Native American, 5.4% European, and 4.5% African. These results suggest that native Mexico's mtDNA-diversity was driven by genetic drift, and that native Mexican mtDNA sequences were preserved, with relatively little admixture from European and African mtDNA (Sandoval *et al.*, 2009).

In another study that targeted the native Mexican group known as Mayans, researchers sought to determine whether the cultural similarities observed among their distant and different communities translated into similarities in their DNA. Mayan descendants of today are spread out over Mexico, Guatemala, Belize, El Salvador, and Honduras; these groups speak one or more of 28 different languages (due to a long history of warring with each other), yet their rituals, artisanship, and architecture distinctly group them together as Mayan (Ibarra-Rivera *et al.*, 2008). DNA was sampled, in the form of blood and buccal swabs, from four major Mayan populations (K'iche, Kakchikel, Campeche, and Yucatan), PCR amplified at 15 autosomal STR loci, and analyzed for degree of genetic relatedness (genealogical and geographical information was also collected for each individual). When compared to the data from previously published studies of other Mayans and Mesoamericans, the Mayans turned out to be more genetically similar to each other than to any Mesoamerican group (Ibarra-Rivera *et al.*, 2008). This suggests that while the Mayan Empire imposed strong political divisions and boundaries to keep many of their communities segregated, genetic material was nevertheless still transferred alongside their cultural exchanges.

In a similar study of a native Mexican group called the Nahuas (who originally migrated from modern day USA), researchers wanted to discover how they ranked in degree of relatedness to the more ancient native Mexicans. The Nahuas are classified as Mexican ethnics that speak Nahuatl. They migrated from a city called Aztlan (in modern day Arizona or New Mexico), around the 12th or 13th century AD, and at that time called themselves Aztecs—later self-changed to Mexitin or Mexicas. Upon their arrival to modern day Mexico they attacked and captured many persons from other ethnic groups

such as the Mayans, Mixtecans, and Zapotecans. The Nahuas then integrated with those they conquered, and later amalgamated all said lands and founded their headquarters at what is now modern-day Mexico City (Vargas-Alarcon *et al.*, 2007). The researchers collected and sequenced mtDNA of 85 Nahuatl speaking individuals (taken from isolated groups and having all four grandparents born in the same area and speaking Nahuatl), then compared it to Amerindians and other various global populations to assess genetic relatedness. Unexpectedly, they found that the Nahuas were genetically similar to the most ancient of Mexicans (Mayans, Mixtecans, and Zapotecans), suggesting that either the Nahua/Aztec Empire imposed their Nahuatl language to many scattered groups throughout Mexico, or that they were in fact living in modern-day Mexico long before the assumed immigration date of 12th to 13th century AD. These findings were at odds with the conventionally accepted theory that the Aztecs migrated to Mexico, from modern day USA (Vargas-Alarcon *et al.*, 2007).

Even though a general study of native Mexicans showed that the majority of the population had underlying genetic similarity throughout the county (Sandoval *et al.*, 2009), the recent studies of parts of India demonstrate that adherence to cultural practices such as endogamy can be enough to divide such a densely populated country into two significantly different branches of ancestral DNA (Reich *et al*., 2009). The Mayan study illustrated how scattered native communities can preserve their DNA distinct from surrounding native communities simply by preserving their culture (Ibarra-Rivera *et al.*, 2008). In the case of the Nahuas we see that one native population can have extremely high genetic relatedness to another native population (so much so that it challenges current migration theory), simply by sharing a common native language (Vargas-Alarcon

et al., 2007). Therefore, to examine the degrees of genetic relatedness among some of Mexico's native groups, a sound starting point was to select previously unstudied native Mexican populations with known preservation of native culture *and* language.

The notion that the ancient civilizations of Mesoamerica were "obliterated" during colonial times was wrong, rather they were simply transformed (Kepics *et al*., 2005). Knowing this we may still locate towns, to our avail, that adhere to native lifestyles. Mesoamerica was described as the perfect example of how historians got "history" incorrect, and needed the help of archaeology to truly explain what transpired (Kepics *et al*., 2005). Population genetic studies can do for archaeology what archaeology did for history—that is, to further clarify the past.

To pursue the potential link between genetic diversity and preservation of native culture *and* language, I felt it logical to select small indigenous towns in Mexico where native languages were well known to still be commonly spoken, as well as being recognized for having high native culture conservation. Since we now know that such practices can translate into native DNA conservation (Vargas-Alarcon *et al.*, 2007; Ibarra-Rivera *et al.*, 2008), I chose Mesoamerican native towns that adamantly practiced such language *and* culture conservation, such as the Nahuatl, Zapotec, and Mixtec (Campbell, 1997).

Mexico boasts a vastly diverse mosaic of native languages comprised of twelve separate and distinct linguistic families, divided into 68 major linguistic groups that include 291 living languages, and seven extinct ones (Gonzalez-Martin *et al.*, 2008). Currently there are 120 communities that still speak native languages: 15 of these languages have over 100,000 native speakers, see Figure 1; 13 with 20,000 to 100,000

speakers, see Figure 2; and 34 with fewer than 20,000 speakers, see Figure 3. Some native populations even integrated with each other so much so that they founded new languages such as Chontal de Tabasco, Chontal Oax, and Maya Yucateco, new cultures (Campbell and Kaufman, 1976; Campbell, 1997), and quite possibly new genetic lineages—inadvertently.

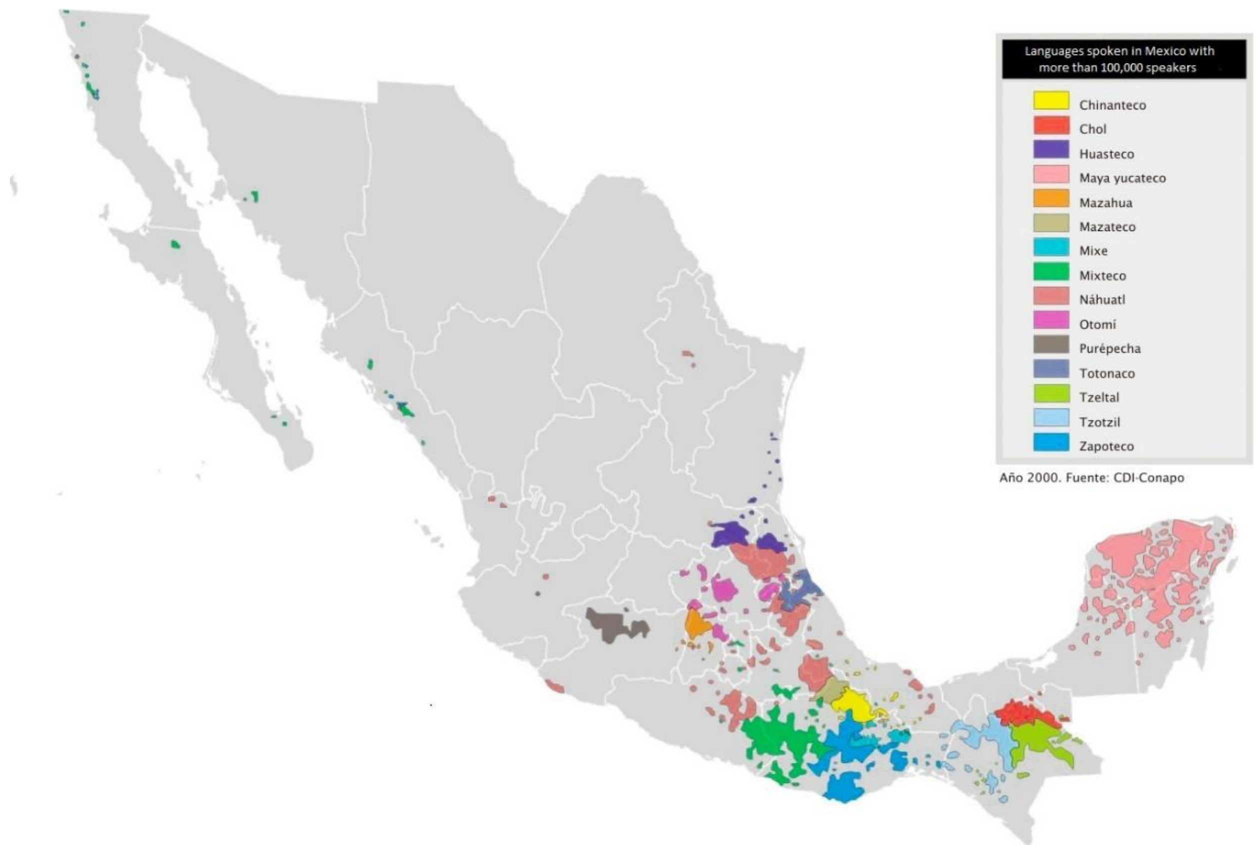

Figure 1. Languages in Mexico. Geograpical locations of native languages spoken by more than 100,000 speakers. (2000) Retrieved June 6, 2011, from http://www.conapo.gob.mx/

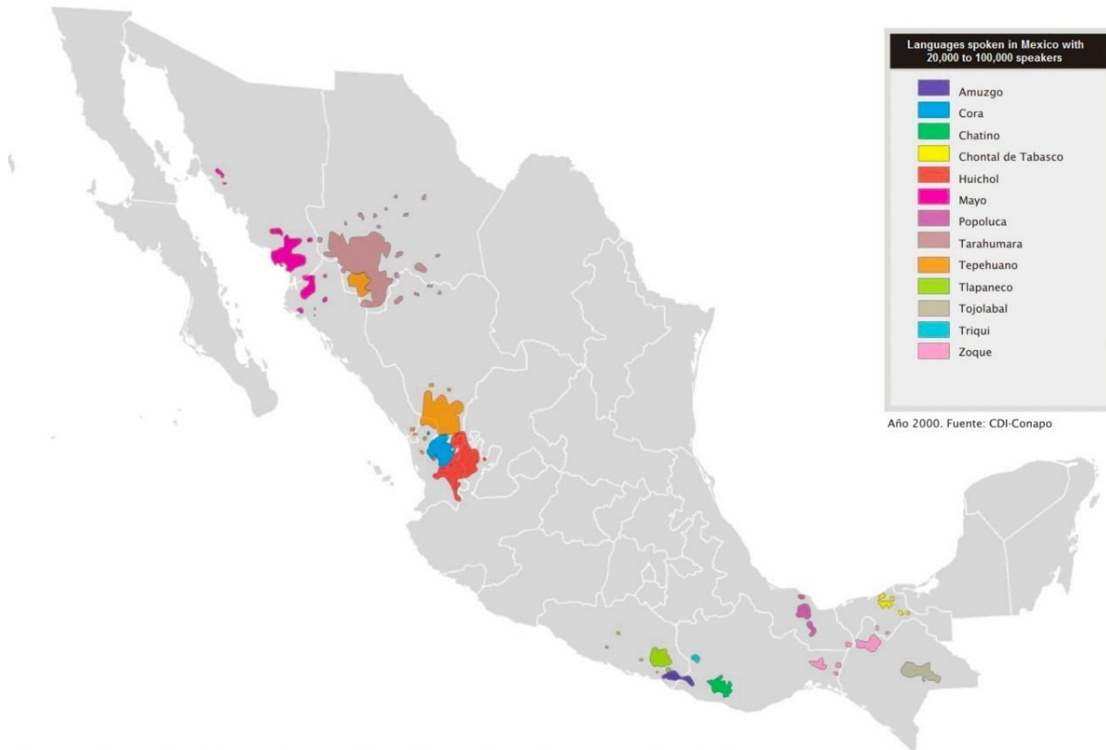

Figure 2. Languages in Mexico. Geograpical locations of native languages spoken by 20,000 to 100,000 speakers. (2000) Retrieved June 6, 2011, from http://www.conapo.gob.mx/

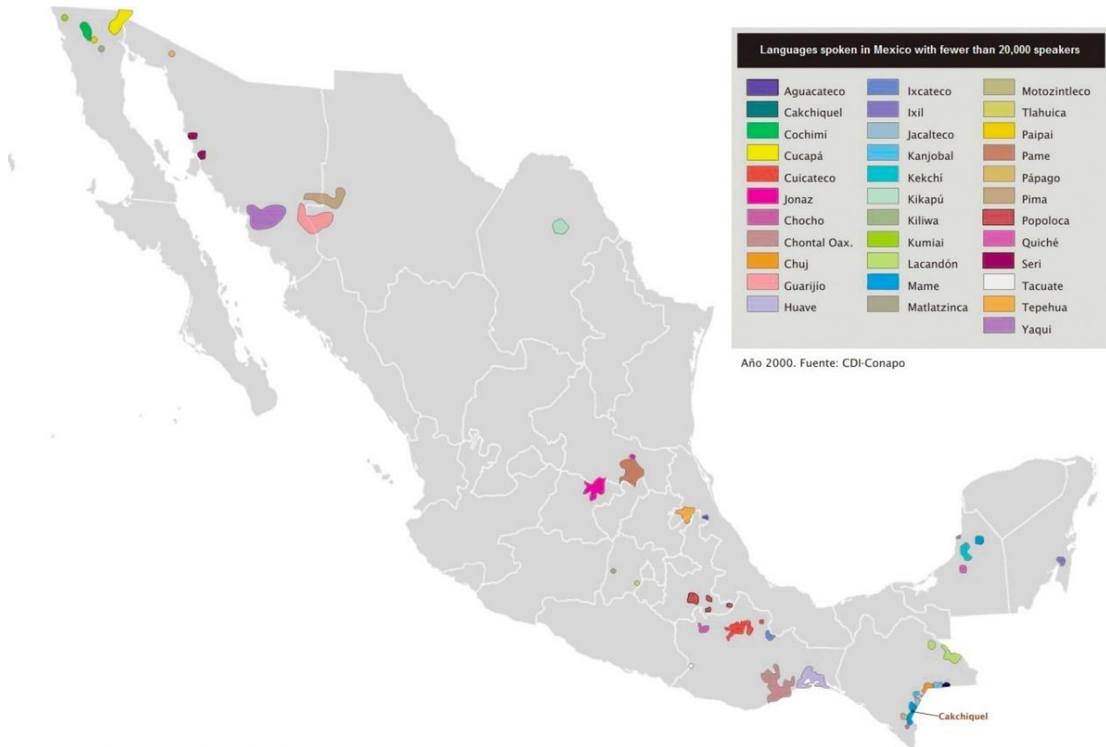

Figure 3. Languages in Mexico. Geograpical locations of native languages spoken by fewer than 20,000 speakers. (2000) Retrieved June 6, 2011, from http://www.conapo.gob.mx/

Nahuatl is the most widely spoken native language in Mexico, with a speaking population of 1.5 million in the states of Distrito Federal, and the Mexican states of Durango, Guerrero, Hidalgo, Mexico, Michoacán, Morelos, Puebla, and Veracruz. It has been a documented language since the 7th century (Suarez, 1983), but lexicostatistical data dates it to around 500 CE (Kaufman, 2001). Nahuatl was the official standardized language of the Aztecs (Vargas-Alarcon *et al.*, 2007), although it was likely that a small percentage spoke other languages as well (Kaufman, 2001). Nahuatl is recognized by the Mexican government as an official language, and among the best studied languages in the Americas (Frawley *et al.*, 2002). Kaufman (2001), stated that Nahuatl originally came from the north, but when, and by whom, was still debatable. He explained that Nahuatl was broken down into three major divisions: Central, Eastern, and Western; among these three there are 15 minor types. The Central, or more prevalent type, also known as General Nahuatl, is spoken in the Valley of Mexico, Morelos, and Puebla. When studying the internal diversification of the entire Nahuatl language, Kaufman (2001), determined that the dispersal of General Nahuatl, from Central Mexico, occurred around 900 CE. He was also convinced that Mesoamerican languages *never* borrow from each other, therefore any borrowing one finds is evidence of serious language—and cultural—contact.

Modern historical and anthropological studies of Mesoamerica have always been macro-regional, never micro-regional (Kepics *et al*., 2005), so a focus on micro-regional sectors of native populations should yield intriguing results. We achieved this by selecting towns that were known sub-populations of a larger native civilization.

The Aztecs became central Mexico's dominant power group in 1428 when they won a decisive battle over the former land dwellers (Willey, 1965), and during this time the Valley of Mexico (also known as Central Mexico), was ruled equally by five independent states; one of the most influential states was named Cuauhnahuac (Smith, 1986). Cuauhnahuac is a Nahuatl phrase meaning "by the trees," this was later renamed to the Spanish phonetic equivalent "Cuernavaca," meaning "Cow's Horn". Smith's (1986) research uncovered that Cuauhnahuac was the oldest province outside the Valley of Mexico (the Aztec empire's central stronghold), and, as such, had a long time to develop. He found that according to Nahuatl native history, Cuauhnahuac was once a 125 mi^2 city-state that expanded in size and power to become the center of a 950 mi^2 conquest-state by 1438. This advancement was achieved by native nobility increasing their power through trade, politico-military cooperation, and marriage alliances. Intermarriage between natives from Cuauhnahuac and those in Mexico Valley became common practice among elites, including almost every member of Aztec royal family (Smith, 1986). This native-elite practice was akin to what the aristocracy and nobility of Europe practiced at the time, and because this led to nearly all of Europe's royal families now possessing some degree of genetic relatedness to each other, it may be possible that native royals possess a similar degree of relatedness among themselves.

At an international conference only 100 years ago Saville (1896) presented some of the very first descriptions ever heard of a native Mexican town located in the state of Morelos, just 12 miles northeast of Cuernavaca (Morelos State capital). This town is named Tepoztlán, and the Temple discovered atop this mountain-city brought it immediate global attention. It's amazing to think that Tepoztlán was unknown even to

the City of Mexico—until the discovery of the Temple. Tepoztlán is situated on elevated ground between rugged cliffs that divide Cuernavaca Valley & Cuautla Valley, it has strategic/defensive views of both Valleys, is on a plain that inclines from west to east, and the north and south regions are also protected by rugged mountains. Saville (1896) documented that the people of Tepoztlán were lineal descendants of the Aztecs, were regarded as speaking the purest Nahuatl in all of Mexico, and closely adhered to ancient customs. He also noted, at the time, that the town had a population comprised of almost pure "aboriginal blood". Tepoztlán is located at the extreme northeast limit of Cuernavaca Valley; on the southwest border of Cuernavaca Valley is another indigenous town named Xochicalco.

Xochicalco was founded in the 7th century on a series of hills, and is a prime example of a fortified city that uses the natural landscape for defense. This town was named a protected UNESCO World Heritage Site because it represents the important period of transition from the old political breakdown of the native empires to the re-grouping of culture from all over Mesoamerica.

Xoxocotla is another town that has large Nahuatl speaking populations, but recent interviews of 1000 locals uncovered that there were increased governmental pressures for them to speak Spanish instead of Nahuatl. This unofficial edict resulted in the current generation receiving very little Nahuatl instruction from their elders (Cantu-Bolan, 2008).

When Spanish colonists set up their administrative headquarters in Mexico City, they casually left native populations out on the periphery of their colony, unreachable and essential untouched. Oaxaca, Chiapas, Yucatan, and Guatemala were labeled as the "south periphery", and the most these peripheral natives were said to experience were

redrawn Spanish jurisdictions that caused minimal migration (Kepics *et al*., 2005). If these natives were left relatively undisturbed over the last few centuries, it is highly possible that little admixture with foreigners occurred, and that they therefore closely resemble the genetic profiles of their indigenous ancestors. It may even be possible to link subpopulations to local geography, as is the case in Europe.

According to the Mexican National Commission (1904), Taxco was amalgamated into the State of Guerrero in 1849. It is a 429 mi^2 town situated among some of the highest peaks and deepest ravines. It is located over 5849 feet above sea level and has a very temperate climate; the natives called this region "the temperate zone" because there are never any severe winters or summers. The weather is so uniform that one can barely notice the change in seasons, which very only by a few degrees year round. Aside from this beautiful climate, the soils are so fertile that all surrounding areas have an abundance of fruit trees, and the water is pure enough to drink directly from the source. Taxco sits on the western slope of the Great Mexican Range, which is considered an extension of the Andes Mountains. Taxco boast a remarkable irregularity of the ground that impresses all; some places in town have 180 foot drops, while directly adjacent to the main plaza are spots of terrain that hover 345 feet above. Old Mexican history claims that a tribute of pure gold bricks that were mined in Taxco was given to Montezuma. Native miners came from all over, and their customs and traditions remain visible in town. Taxco shows signs of ancient native settlements and the surrounding areas contain vestiges of old native towns.

Kepics' (2005) study unveiled that Chiapas was heavily invested in the Mesoamerican economic system, yet the history of Chiapas is the least studied and

understood in Mesoamerica. Very little is known about the pre-colonial period, and few studies have been undertaken. The pre-colonial population estimates put it at around 275,000 natives. Chiapas has very diverse ethno-linguistics because it was a natural bottleneck of migratory populations who travelled and ultimately settled in North, South, and Central America, and many smaller segments of those groups decided to stay in Chiapas. Modern visitors are in awe of the variation of populations in Chiapas; some native groups are highly visible, while others are less so due to their adoption of western dress and culture.

Zapotec and Mixtec are the most widely spoken languages in Oaxaca and Mexico State; the most populous region in Mesoamerica, and flanking the known Nahuatl speaking region on both sides. Zapotec and Mixtec are spoken by a combined 1.5 million individuals, and are grouped together for this study because they were both derived from the ancient precursor language Oto-manguean (Campbell & Kaufman, 1976). Their first linguistic branch-off took place when Mayans left central Mexico to settle the Gulf of Mexico (around 2200 BCE), and resulted in enough divergence that linguists classify Zapotec and Mixtec as officially distinct from one another (Campbell, 1997).

Boone (2010), states that the Mixtec and Aztecs were similar in their organization of politics, rulers, and territory, but were culturally different aside from that. Oaxaca and South Puebla are more culturally Mixtec than Aztec. Small details that highlight cultural differences between the Mixtec and Aztecs are that the Aztecs recorded all their records on hide, or native paper, whereas the Mixtec exclusively used cloth. In storytelling the Mixtec like to convey messages, ideas, and morals via allegory, whereas the Aztecs relayed facts. Manuscripts from different native cultures used different dating systems;

the Aztecs designated the year by a hieroglyph enclosed in a rectangular cartouche, and the Mixtec used an “A-O” sign where rays (A’s) and circles (O’s) interlocked in various manners to represent the year. But during the Spanish colonization both conventions started to appear on native documents, signaling a beginning of mix of those cultures because the Mixtec never used the Aztec convention, and vice versa. The Nahua people passed off royal succession to any male living kin, such as brothers, uncles, nephews, or cousins, whereas those in Oaxaca passed it off only to their eldest male heir, potentially having different effects on how endogamy was practiced and as a result how admixture occurred.

While some indigenous Mexican languages are now extinct (Alaguilac, Coahuilteco, Comecridan, Cotoname, Cuitlatec, and Solano), many others are still widely spoken (Table 2) (Campbell, 1997).

TABLE 2. *Number of native language speakers in Mexico (in thousands).*

Language	Number of Speakers	Language	Number of Speakers	Language	Number of Speakers
Aguacateco	< 20	Ixil	< 20	Popoluca	20--100
Alaguilac	0	Jacalteco	< 20	Purépecha	> 100
Amuzgo	20--100	Jonaz	< 20	Quiché	< 20
Cakchiquel	< 20	Kanjoal	< 20	Seri	< 20
Chatino	20--100	Kekchí	< 20	Solano	0
Chinanteco	> 100	Kikapú	< 20	Tacuate	< 20
Chocho	< 20	Kiliwa	< 20	Tarahumara	20--100
Chol	> 100	Lacandón	< 20	Tepehuano	20--100
Chontal de Tabasco	20--100	Mame	< 20	Tepehuano	< 20
Chontal Oax	< 20	Matlatzinca	< 20	Tlahuica	< 20
Chuj	< 20	Maya Yucateco	> 100	Tlapaneco	20--100
Coahuilteco	0	Mayo	20--100	Tojolabal	20--100
Cochimí	< 20	Mazahua	> 100	Totonaco	> 100
Comecridan	0	Mazateco	> 100	Triqui	20--100
Cora	20--100	Mixe	> 100	Tzeltal	> 100
Cotoname	0	Mixteco	> 100	Tzotzil	> 100
Cuicateco	< 20	Motozintleco	< 20	Yaqui	< 20
Cuitlatec	0	Nahuatl	> 100	Zapoteco	> 100
Cupacá	< 20	Otomi	> 100	Zoque	20--100
Guarijio	< 20	Paipai	< 20		
Huasteco	> 100	Pame	< 20		
Huave	< 20	Pápago	< 20		
Huichol	20--100	Pima	< 20		
Ixcateco	< 20	Popoloca	< 20		

Source: Campbell, 1997. Languages in Mexico, 2015. Retrieved March 15, 2015, from http://en.wikipedia.org/wiki/Languages_of_Mexico.

If culture and language has underlying relations to genetic diversity then we might expect see it in the form of STR allele frequency differences, and possibly in mtDNA sequence differences. We chose to examine the STR profiles (consisting of the 15 core STR loci), and mtDNA of 1000 individuals from the carefully selected towns of Cuernavaca, Oaxaca, Puebla, Taxco, Tepoztlán, and Xochicalco (each town selected for the reasons mentioned earlier).

Chapter II
Materials and Methods

Using STR profiles and mtDNA sequence data, in conjunction with corresponding questionnaire data, I compared and contrasted the genetic profiles of individuals grouped according to Mexican town (based on culture and native language preservation), in attempt to determine if there was any significant genetic variation in these populations. This information was then compared to previously published population studies of various peoples. I expected that some of the isolated towns with high native culture and language conservation would be more genetically distinct from the rest of the general Mexican population.

The purpose of this project was to use STR and mtDNA sequence data to identify how genetically diverse and distinct are the native groups of Mexico. Public and private data were used to calculate STR profile frequencies per town, across all individuals we sampled, and globally. In order to gain better understanding of how Mexican natives fit into the global arena, genetically, I compared the collected data to previously published studies on Eskimo, European, Asian, North-, and South American populations. Statistical analysis was used to determine the significance of genetic variation for each of the variables mentioned above.

Sample Collection

The first stage of this project consisted of the collection of buccal swab DNA samples in-field, along with genealogical and geographical information, from randomly

selected individuals in Cuernavaca, Morelos; Oaxaca, Oaxaca; Puebla, Puebla; Taxco, Guerrero; Tepoztlán, Morelos; and Xochicalco, Morelos. All samples were volunteered, anonymous, and identifiable only by number.

To collect the DNA samples we used the Bode Buccal DNA Collector. Bode certified that all collectors are free of human DNA and DNase. The collectors have three main components: a handle base, with blank label for name and number; a flat swab tip, lined with FTA® collection paper; and a slider cover, to protect the collected DNA.

Once human volunteers were located, we asked brief questionnaires regarding maternal and paternal lineages, had them sign a witnessed consent form, then labeled the Bode Buccal DNA Collector and questionnaire sheet with the same identification number; no names were used.

To collect DNA using the Bode Buccal DNA Collector we first removed the new collector from its sealed pouch, holding the handle at the base. We then moved the slider back if necessary to expose the FTA® collection paper. We then had the volunteer open their mouth, and with our thumb on the back area of the collector marked "Thumb" we placed the white FTA® collection paper side flat against the inside of their cheek. Having the FTA® collection paper pressed against the inner cheek we firmly dragged the collector towards the lips and out of the mouth (similar to the "popping" of the cheek with a finger that children do). The bulging out of the cheek during collection is a good indicator that DNA is being collected. This action was repeated 15 to 20 times depending on the dryness or wetness of the subject's mouth. Bode recommends this action be repeated 7 times, but through experience we found that subjects living at higher altitudes have dryer mouths than normal thus requiring more dragging action to sufficiently collect

enough DNA on the FTA® collection paper. Once the DNA swab has been collected we push the slider cover towards the tip, covering the FTA® collection paper, labeled the handle with an identification number corresponding to the volunteer's questionnaire, placed the collector back into the original plastic pouch, and sealed it completely with tape.

DNA Amplification and STR Genotyping

Once collected, the DNA was extracted at the UNAM campus in Cuernavaca using Promega and Qiagen DNA extraction kits, then amplified via PCR. The amplified DNA was analyzed through an ABI Prism 310 sequencer to profile 15 STR loci: D8S1179, D21S11, D7S820, CSF1P0, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, vWA, TPOX, D18S51, D5S818, FGA, and Amelogenin. Once individual STR profiles were produced they were paired with their corresponding questionnaires, and the STR typing results were stored in digitized format to facilitate categorizing the data by town and by geographic location.

Whole genomic mtDNA was extracted using Qiagen DNA extraction kits, followed by PCR amplification. DNA strands for hypervariable region (HVR)-1 (position 16024-16365), and HVR-2 (position 73-340) was sequenced using the fluorescent cycle-sequencing method (Guardado-Estrada *et al.*, 2009), for 1000 individuals. This mtDNA sequencing was performed by our collaborators at the Armed Forces DNA Identification Laboratory (AFDIL) in Rockville, Maryland.

The goal was to determine if these sampled populations emerged as genetically distinct to other Mexican populations.

Previously Published Data

The collected data were compared to 26 data sets from previously published population studies (see Table 3).

TABLE 3. *Published population studies of Eskimo, European, Asian, South-, Meso-, and North american groups.*

Population	Abbreviation	Location	Number of loci	Reference
Spain (Andalusia)	AND	Europe	13	Perez-Miranda et al. (2005)
Spain (Guipuzcoa)	GUI	Europe	13	Perez-Miranda et al. (2005)
Japan	JAP	Asia	15	Hashiyada et al. (2003)
Korea	KOR	Asia	15	Kim et al. (2003)
China (Chao Shan)	CCS	Asia	15	Hu et al. (2003)
China (Shaanxi Han)	CSH	Asia	15	Wang et al. (2005)
Inupiat	INU	North America	13	Budowle et al. (2002)
Yupík	YUP	North America	13	Budowle et al. (2002)
Athabaskan	ATH	North America	13	Budowle et al. (2002)
Mestizo (Mexico)	MES	Meso America	15	Hernandez-Gutierrez et al. (2005)
Metstitlán	MET	Meso America	15	Gorostiza et al. (2006)
Choles	CHO	Meso America	15	Sanchez et al. (2005)
Yucatán	YUC	Meso America	15	Ibarra-Rivera et al. (2008)
Campeché	CAM	Meso America	15	Ibarra-Rivera et al. (2008)
K'iché	KIC	Meso America	15	Ibarra-Rivera et al. (2008)
Kakchikel	KAK	Meso America	15	Ibarra-Rivera et al. (2008)
Huasteca	HUA	Meso America	15	Barrot et al. (2005)
Otomi (Sierra Madre)	OSM	Meso America	15	Barrot et al. (2005)
Otomi (Ixmiquilpan Valley)	OIX	Meso America	15	Barrot et al. (2005)
Conchagua	CON	Meso America	13	Lovo-Gomez et al. (2006)
Izalco	IZA	Meso America	13	Lovo-Gomez et al. (2006)
Panchimalco	PAN	Meso America	13	Lovo-Gomez et al. (2006)
Kichwa	EKI	South America	13	Gonzales-Andrade et al. (2006)
Colombian Andes	CAN	South America	13	Paredes et al. (2003)
Colombian South Andes	CSA	South America	13	Paredes et al. (2003)

Source: Gonzales-Martin *et al*., 2008

Computational Resources

The data analyses were performed on a Lenovo ThinkPad X220 notebook computer with Intel® Core™ i7-2620M processor (dual-core, 2.70GHz, 4MB Cache), 16 gigabytes of 1333MHz DDR3 RAM, 410 gigabyte dual solid state drive, and the Microsoft® Windows® 7 Professional 64-bit operating system. Genetic profile data was entered into spreadsheets using Microsoft Excel® 2013. The software used for statistical

analysis included: PowerStats v1.2 (Promega, 2011), Arlequin v3.5.1.2 (Excoffier & Lischer, 2010), NTSYSpc v2.2 (Rohlf, 2005), and PHYLIP v3.695 (Felsenstein, 2015).

Statistical Analysis

The STR allelic frequencies were calculated by counting the number of copies of each allele in the population and dividing them by the sum of all alleles in the population. This was done for the whole data set, and for each individual town. To account for sampling errors (e.g., if a sample group was small), we used the Balding correction method (Goodwin *et al.*, 2011).

The expected genotype frequency for each STR locus per individual was calculated by using HWE along with allelic frequencies (calculated earlier). Once all genotype frequencies were determined the product rule was implemented to obtain an individual's multi-locus profile frequency. This calculation is permissible because alleles at distinct loci are inherited independently of each other, in accord with Mendel's Second Law of Independent Assortment.

PowerStats v1.2 was used to calculate Gene Diversity Index (GDI), Matching Probability (MP), Power of Discrimination (PD), Polymorphic Information Content (PIC), Power of Exclusion (PE), and Typical Paternity Index (TPI). I averaged locus by locus values to calculate the Combined Matching Probabilities (CMP), Combined Power of Discrimination (CPD), and Combined Power of Exclusion (CPE). Arlequin v3.5.1.2 was used to estimate the expected and observed heterozygosities (H_e and H_o), per loci. All this information was used to assess the strength of the markers used in discriminating between individuals, and to ultimately analyze the *intra*-population STR diversity.

To analyze the *inter*-population STR diversity I used the results from correspondence analysis (CA), run on data from all available loci (using NTSYSpc v2.2), and neighbor-joining (NJ) trees, created based on the fixation index (F_{ST}) distances of all populations and admixture estimates (using PHYLIP v3.69). These results allowed me to compare frequencies and allelic distributions, thus revealing the amount of potential genetic contribution per population.

Genetic homogeneity between pairs of populations were analyzed using a conservative Bonferroni correction that minimized the chance of "false positives." Pairs that had a P value over α were deemed as genetically homozygous (Bonferroni, 1935).

mtDNA HVR-1 and HVR-2 sequences for all individuals were compared to the revised Cambridge Reference Sequence (rCRS) using Applied Biosystems Seq Scape v2.5 software. All identified polymorphism were saved as a Microsoft Office 2013 Access database in order to construct and identify haplotypes, and calculate allelic and haplotype frequency. Arlequin v3.5.1.2 software was used to calculate pairwise differences, nucleotide diversity, haplotype diversity indexes, mismatch distribution, *Tau* (τ) and Tajima's D neutrality test (Excoffier & Lischer, 2010). Each sample was classified by their characterized haplogroups (Amerindian, European, or African), and TCS v1.21 software was used to construct a minimum spanning network with control region sequences via the statistical parsimony algorithm (Clement *et al.*, 2005). To determine statistical significance of the proportion of mtDNA haplogroup ancestry among the groups, a χ^2 test was performed using Minitab Statistical Software, and the Student's t-test was used to evaluate the statistical significance of comparison of mean

pairwise differences, and nucleotide diversity. All mtDNA sequencing was conducted by AFDIL.

In summary, this thesis used previously published population-study data in conjunction with newly attained first-hand data to determine how genetically diverse are some important native groups in contemporary Mexico.

Chapter III

Results

Once all STR and mtDNA computations were completed, I used statistical analysis to quantify the genetic diversity of Mexican native groups among themselves, and globally. Previously published data was used to aid in these comparisons, and the results are described below.

STR Intra-population Diversity

Allelic frequencies, observed (H_O) and expected (H_E) heterozygosities, Hardy-Weinberg equilibrium (HWE) *P*-values, and parameters of interest for population genetics (GDI, MP, PD, PIC, PE, TPI), for the towns of Cuernavaca, Puebla, Taxco, Xochicalco, and the UNAM student population were computed (see Appendix A). Allelic frequencies at 0.50 and greater are indicated in italics. Six loci deviated significantly from HWE expectations at $\alpha = 0.05$ (CSF1PO, D2S1338, D19S433, D18S51, and two in FGA) in Cuernavaca, one (D18S51) in Puebla, and one (D21S11) in the UNAM student population. Once the Bonferroni correction ($\alpha = 0.0033$) was applied, no loci diverged from HWE expectations.

Analysis of the allelic frequencies demonstrates that there is little difference between Taxco, and Xochicalco, with regards to their total number of alleles across 15 loci. Taxco has the smallest total number of allelic types (101) followed by Xochicalco (102), UNAM student population (105), Puebla (111), and Cuernavaca (133) (Table 9).

The parameters of interest for population genetics, including average heterozygosities, are summarized in Appendix B. The probability that two randomly selected unrelated individuals we sampled would have identical genotypes at all loci is called the Combined Matching Probability (CMP), and the measure of how likely it is that two randomly selected unrelated individuals we sampled would have different genotypes at all loci is called the Combined Power of Discrimination (CPD). According to the CMP and CPD values, the population in Cuernavaca had the strongest discriminative power of all loci analyzed. The power of a test to eliminate a certain percentage of the population from being genetically related to an individual at all loci for our sampled individuals is call the Combined Power of Exclusion (CPE), and these values were strongest for Taxco. Heterozygosity averages were used for analysis against other geographic groups because all the native Mexican populations sampled were of different sizes.

Intra-population variance (H_S), represents the average heterozygosity found within a population, and can illuminate details of that sampled population's structure and/or history. For example, if a population had very low heterozygosity (i.e. very little genetic variability), compared to what was expected for that population under Hardy-Weinberg equilibrium, we would be confident that said population suffered the effects of a small population size; we could then attribute this to various scenarios, such as inbreeding, population bottlenecks, or some other metapopulational dynamic that severely reduced the population's access to genetic variation. Calculated H_S values are summarized in Appendix B. When compared to each other, the Native Mexican population (comprised of Cuernavaca, Puebla, Taxco, Xochicalco, and UNAM students),

have the fourth lowest H_S value (H_S = 0.7382). The lowest overall H_s value belonged to the North Americans (H_S = 0.7236), followed by the Mesoamericans not sampled in this study (H_S = 0.7333), and the Native Americans (H_S = 0.7344). The European population had the highest overall H_S value (H_S = 0.7812), followed by the Asians (H_S = 0.7703), and the South Americans (H_S = 0.7623).

STR Inter-population Diversity

One method of measuring population differentiation is to use Nei's statistic (G_{ST}), which is a measure of genetic differentiation used to describe the total amount of variation observed among populations, over multiple loci (Nei, 1977). To calculate the G_{ST} we first need to determine the average heterozygosity within a subpopulation (the H_S value mentioned above), as well as the average heterozygosity within a total population (H_T), also known as inter-population variance. G_{ST} and H_T values are shown in Table 10. The lowest G_{ST} value among native populations were those of the South Americans (G_{ST} = 0.0078), followed by the Native Mexicans (G_{ST} = 0.0170), North Americans (G_{ST} = 0.0260), and the Mesoamericans not sampled in this study (G_{ST} = 0.0268). When comparing the H_T values of all populations, the North Americans had the lowest (H_T = 0.7429), followed by the Native Mexicans (H_T = 0.7510), then the Mesoamericans not sampled in this study (H_T = 0.7535), then the Native Americans (H_T = 0.7675), and South Americans. The Highest H_T values were for the Europeans (H_T = 0.7877), followed by the Asians (H_T = 0.7739).

In order to visually illustrate DNA data in two dimensional form we use an agglomerative clustering method—called Neighbor-Joining—to create phylogenetic trees. The distances between all nodes of this tree are calculated, and representative of the genetic relatedness between each taxa. In the Neighbor-Joining (NJ) tree of the sampled native Mexican populations (Figure 4.), we see that the Xochicalco, and Puebla populations segregate from the Taxco, Cuernavaca, and UNAM cluster; Taxco and UNAM branch off from a common node. These characteristics become more evident in the dendogram (Figure 5).

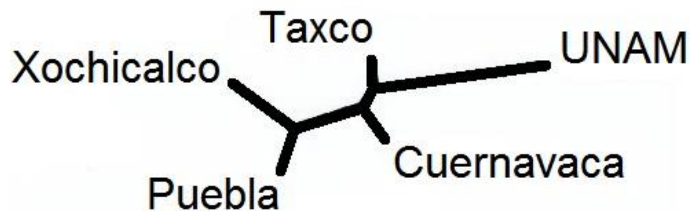

Figure 4. Neighbor-joining tree of sampled Native Mexican populations.

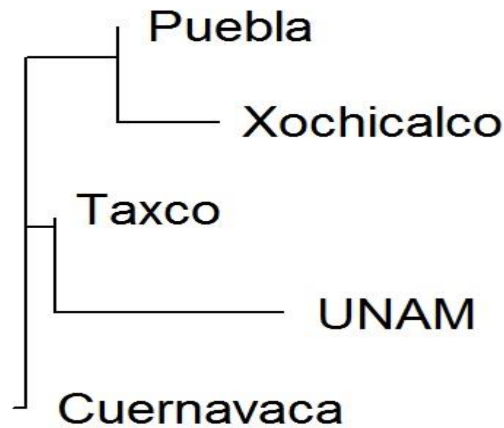

Figure 5. Dendogram of sampled Native Mexican populations.

The F_{ST} values for all sampled native Mexican populations include data from 15 STR loci (Table 11). The highest F_{ST} values were for the pair-wise comparisons of the Xochicalco-UNAM, Cuernavaca-UNAM, and Taxco-UNAM populations (F_{ST} = 0.01384, F_{ST} = 0.00577, and F_{ST} = 0.00517, respectively). The lowest F_{ST} values were for the Cuernavaca-Taxco (F_{ST} = -0.00148), Cuernavaca-Puebla (F_{ST} = 0.00109), Puebla-Xochicalco (F_{ST} = 0.00116), Puebla-Taxco (F_{ST} = 0.00177), Cuernavaca-Xochicalco (F_{ST} = 0.00254), and Taxco-Xochicalco (F_{ST} = 0.00446) pairs. When tested for significance at the 0.05 level only the Xochicalco-UNAM, and Cuernavaca-UNAM pairs were significant (Table 12).

TABLE 11. *Fst values for 15 STR loci*

	Cuernavaca	Puebla	Taxco	Xochicalco	UNAM
Cuernavaca					
Puebla	0.00109				
Taxco	-0.00148	0.00177			
Xochicalco	0.00254	0.00116	0.00446		
UNAM	0.00577	0.0055	0.00517	0.01384	

TABLE 12. *Significant pairwise values for 15 STR loci*

	Cuernavaca	Puebla	Taxco	Xochicalco	UNAM
Cuernavaca					
Puebla	-				
Taxco	-	-			
Xochicalco	-	-	-		
UNAM	+	-	-	+	

+, significance at the 0.05 level

Haplogroup and Haplotype Distributions

The mtDNA haplotype F_{ST} values for all sampled native Mexican populations, including sequence differences were tabulated (Table 13). The highest F_{ST} values were for the pair-wise comparisons of Xochicalco-Xoxocotla (F_{ST} = 0.0577), Cuernavaca-Xochicalco (F_{ST} = 0.04088), Chiapas-Xochicalco (F_{ST} = 0.0359), Tepoztlán-Xochicalco (F_{ST} = 0.03086), Chiapas-Xoxocotla (F_{ST} = 0.02828), Taxco-Xoxocotla (F_{ST} = 0.02755), Puebla-Xoxocotla (F_{ST} = 0.02748), Puebla-Chiapas (F_{ST} = 0.02592), Cuernavaca-Chiapas (F_{ST} = 0.02655), Taxco-Chiapas (F_{ST} = 0.01931), and Cuernavaca- Taxco (F_{ST} = 0.01648). The lowest F_{ST} values were for the pair-wise comparisons of Cuernavaca-Tepoztlán (F_{ST} = -0.00656), Puebla-Xochicalco (F_{ST} = -0.00021), Tepoztlán-Xoxocotla (F_{ST} =0.0001), Taxco-Xochicalco (F_{ST} = 0.00059), Puebla-Taxco (F_{ST} = 0.00366),

Puebla-Tepoztlán (F_{ST} = 0.00499), Cuernavaca-Xoxocotla (F_{ST} = 0.00614), Taxco-Tepoztlán (F_{ST} = 0.00706), Cuernavaca-Puebla (F_{ST} = 0.00983), and Tepoztlán-Chiapas (F_{ST} = 0.01496). When tested for significance at the 0.05 level, Cuernavaca-Taxco, Cuernavaca-Xochicalco, Cuernavaca-Chiapas, Puebla-Chiapas, Puebla-Xoxocotla, Taxco-Chiapas, Taxco-Xoxocotla, Xochicalco-Chiapas, Xochicalco-Xoxocotla, and Chiapas-Xoxocotla all tested as significant (Table 14).

TABLE 13. *Fst values for mtDNA haplotypes (including sequence differences)*

	Cuernavaca	Puebla	Taxco	Tepoztlán	Xochicalco	Chiapas	Xoxocotla
Cuernavaca							
Puebla	0.00983						
Taxco	0.01648	0.00366					
Tepoztlán	-0.00656	0.00499	0.00706				
Xochicalco	0.04088	-0.00021	0.00059	0.03086			
Chiapas	0.02655	0.02592	0.01931	0.01496	0.0359		
Xoxocotla	0.00614	0.02748	0.02755	0.0001	0.0577	0.02828	

TABLE 14. *Significant pairwise values for mtDNA haplotype*

	Cuernavaca	Puebla	Taxco	Tepoztlán	Xochicalco	Chiapas	Xoxocotla
Cuernavaca							
Puebla	-						
Taxco	+	-					
Tepoztlán	-	-	-				
Xochicalco	+	-	-	-			
Chiapas	+	+	+	-	+		
Xoxocotla	-	+	+	-	+	+	

+, significance at the 0.05 level

The mtDNA lineages of all the sampled native Mexican populations were almost exclusively comprised of Native American haplogroups (Table 15), and mainly restricted to haplotypes A2, B2, C1, D1, and D4, as defined by Mannis van Oven (2008). Haplogroup A lineages were most prevalent with haplotype A2 making the largest contribution in all sampled populations. The greatest amount of A2 haplotype was found in Xoxocotla (71.88%), followed by Xochicalco (65.22%), Cuernavaca (59.62%), Tepoztlán (59.52%), Chiapas (57.29%), and Puebla (49.09%). Haplotype A4 (of Asian origin), was only found in the Cuernavaca, and Tepoztlán populations (5.13%, and

4.76%, respectively). Haplogroups B, and C were found at lower but still substantial frequencies. Haplotype B2 was most present in Puebla (29.09%), Taxco (28.74%), Xochicalco (21.74%), Chiapas (21.61%), Cuernavaca (20.51%), Tepoztlán (19.05%), and Xoxocotla (15.63%). Haplotype C1 was most present in Chiapas (20.61%), Taxco (17.24%), Tepoztlán (11.90%), Puebla (9.09%), Cuernavaca (8.97%), Xochicalco (8.70%), and Xoxocotla (3.13%). Haplogroup D was the least prevalent Native American haplogroup detected. Haplotype D1 was found highest in Xoxocotla (9.38%), followed by Taxco (4.60%), Xochicalco (4.35%), and Puebla (3.64%); and found lowest in Chiapas (0.05%), Cuernavaca (1.28%), and Tepoztlán (2.38%). Haplotype D4 was rarely detected and found in only two sampled populations, Puebla (7.27%), and Cuernavaca (1.28%). Haplotype X2a was completely absent in all sampled native Mexican populations.

Table 15: *Number of samples per mtDNA haplogroup*

Distribution/Origin of Haplogroup	mtDNA Haplogroup	Cuernavaca	Chiapas	Puebla	Taxco	Tepoztlán	Xochicalco	Xoxocotla
Native American	A2	93	114	27	39	25	15	23
Asian	A4	8	0	0	0	2	0	0
Native American	B2	32	43	16	25	8	5	5
Native American	C1	14	41	5	15	5	2	1
Native American	D1	2	1	2	4	1	1	3
Native American	D4	2	0	4	0	0	0	0
West Eurasian	HV0	1	0	0	1	0	0	0
West Eurasian	HV1	0	0	0	0	0	0	0
West Eurasian	J1b	0	0	0	0	1	0	0
West Eurasian	J2b	0	0	0	1	0	0	0
West Eurasian	K1a	0	0	0	1	0	0	0
African	L0a	1	0	0	0	0	0	0
African	L3d	0	0	0	1	0	0	0
West Eurasian	R0	2	0	1	0	0	0	0
West Eurasian	U3a	1	0	0	0	0	0	0
West Eurasian	U4	0	0	0	0	0	0	0

All sampled populations had over 90% of their mtDNA originating from Native American lineages, with the exception of the UNAM student population at 83.33%, and the population with the greatest diversity of mtDNA lineages was that of Cuernavaca

(Table 16). Xochicalco, Xoxocotla, and Chiapas all had 100% of their mtDNA originating from Native American lineages, followed my Taxco (95.40%), Tepoztlán (92.86%), Puebla (90.91%), and Cuernavaca (90.38%). Asian lineage was found in Puebla (7.27%), Cuernavaca (6.41%), and Tepoztlán (4.76%). Taxco exhibited the highest proportions of African and West Eurasian mtDNA lineages, but all of the African and West Eurasian lineages found in Taxco and Cuernavaca were haplotypes that were not shared within or between populations (Table 15). African lineage was only found in Taxco (1.15%), and Cuernavaca (2.56%). West Eurasian was found in Taxco (3.45%), Cuernavaca (2.56%), Tepoztlán (2.38%), and Puebla (1.82%) (see Appendix C).

TABLE 16. *Number of samples per mtDNA-lineage origin*

mtDNA Lineage Origin	Cuernavaca	Puebla	UNAM	Taxco	Tepoztlán	Xochicalco	Xoxocotla	Chiapas
Native American	141	50	20	83	39	23	32	199
Asia	10	4	0	0	2	0	0	0
Africa	1	0	1	1	0	0	0	0
Near East/Europe	4	1	3	3	1	0	0	0
Total	156	55	24	87	42	23	32	199

A summary of the statistics of interest for mtDNA, such as Sample Size, Random Match Probability, Number of Haplotypes, Haplotypes Shared, and Average Pairwise Difference is included (Table 17).

TABLE 17. *Summary statistics for mtDNA*

	Cuernavaca	Puebla	Taxco	Tepoztlán	Xochicalco	Xoxocotla	Chiapas
Sample Size	156	55	87	42	23	32	199
Random Match Probability	0.70%	1.35%	0.52%	0.81%	1.19%	4.03%	4.68%
Haplotypes	136	45	76	37	19	21	58
Haplotypes Shared	13	7	6	4	3	7	28
Average Pairwise Difference	11.84	13.52	14.21	11.92	13.26	11.19	13.05

Chapter IV

Discussion

Technological advancements that allow us to study human genetic diversity have increased remarkably in recent years. We are now able to amplify DNA relatively quickly using PCR; identify STRs to study individuals and their polymorphisms, in accordance with Mendel's second law of independent assortment; and we are living at a time when the computational power needed to store and analyze DNA sequences increases while the costs of doing so decrease.

In this study, I compared and contrasted individual genetic profiles of native indigenous groups from select Mesoamerican towns in Mexico to determine their degree of genetic relatedness to each other, and globally. Limited population genetic research of the Mesoamerican people has been performed even though these groups belong to some of the most diverse, and advanced civilizations. Some of these civilizations ranked as more advanced than ancient Egyptians, Mesopotamians, or Chinese. Archaeologists found evidence of Mesoamerican tool-bearing hunters carbon that dated to 24,000 years ago, and since that time Mesoamerican civilizations were formed, destroyed, amalgamated, transformed, conquered, and colonized. Such a rich history should provide a plethora of genetic information, especially when none had been studied yet. To attain insight into the Mesoamerican genetic landscape we travelled to Mexico to obtain DNA samples from over 1000 individuals; we then extracted, amplified, sequenced, and analyzed their STR and mtDNA using the most current methods. Unknown prior to this work, was the degree to which the populations of Cuernavaca, Puebla, Taxco,

Xochicalco, Xoxocotla, Tepoztlán, and Chiapas were genetically related, as many of these populations have never been sampled.

The major points of interest emerging from this study are that *our* sampled populations: (1) had the highest *intra*-population variance in North America, (2) had the lowest *inter*-population variance in North America, (3) maintained low mtDNA haplotype polymorphism while yielding high mtDNA haplogroup diversity, and (4) exhibited no traces of haplotype X2a, which one typically expects to find in native North American populations.

Due to the DNA collection during this project, I was able to quantitatively measure the genetic variation of several select towns, totaling over 1000 individuals. I observed genetic variation across and within these populations, on a regional and global scale. I also observed variations of mtDNA lineage origins within and among the populations. The reasons for such variations are probably the result of linguistic *and* cultural practices of these native Mexican populations. These practices could also be the reason why some genetic conservation, rather than variation, was observed. The collective data suggested that genetic admixture was not a specific sole property of native language *or* culture, but rather a dynamic process of both. While determination of the exact ratio of this combination is beyond the scope of this thesis, it remains a viable topic for further research. Nonetheless, I postulated that native groups who closely adhered to language *and* culture conservation would vary—genetically—from other native groups, and from the general Mexican population.

To test this hypothesis, I examined if genetic variation correlated to variations of a town's culture *and* language. To find towns of different native cultures I researched the

published historical and archaeological literature in attempt to extrapolate from them a list of towns that were located within Mesoamerica, yet had sufficiently different native culture; the criteria of native language history was very helpful in distinguishing these towns. In some cases it was evident that current state borders within Mexico created by modern governments could have been very deceiving in identifying appropriate towns. For example, Oaxaca and Puebla are two large and different states located in central Mexico, but if we selected towns located within Oaxaca and placed them in a different category than those from Puebla we would have undermined the integrity of our variables because Oaxaca and the southern regions of Puebla—not the northern parts—were once unified, for centuries, under various empires and ancient civilization, so much so that indigenous natives from north Puebla and south Puebla have less in common, culturally, than those from Oaxaca and southern Puebla (Boone, 2010). The above example is also true of many native towns in Mesoamerica because native alliances, and jurisdictions were a better indicator of culture-similarity than were the borders drawn up by the Spaniards, whose sole intent was to divvy up real estate for their own elites to govern (Kepics *et al.*, 2015), or the borders currently imposed by the modern Mexican government. Coupling these realities with the fact that Mesoamerican studies were inaccurate until approximately 40 years ago, one can begin to see how much categorical confusion could have compounded with historical errors, and perpetuate into genetic analysis errors. As it turns out, the collective of our sampled populations had the highest intra-population variance ($H_S = 0.7382$) of all North American populations (Table 10), and globally ranked as the median for intra-population variance. This was a very significant fact because it meant that within our sampled groups, which came from a

relatively small geographical radius, individuals were genetically *very* different from each other. This was in direct contrast to the decrease in genetic variation one was expected to find among native populations living in close proximity. Current dogma stated that in the North American continent one was expected to find limited genetic diversity within any single native group; the reason for this was attributed to the founder effect. It was also believed that populations that were farther away from their genetic place of origin (such as native Mesoamericans that had genetic lineages descending from Asia), and maintain reproductive isolation, would experience increased bottlenecks that dwindled their gene pool diversity (Iberra-Rivera et. al., 2008). This was not the case in our study. The results of this thesis supported the recent historical and archaeological findings that Mesoamerica was indeed a vast mosaic of different native groups, each with their own unique cultures that in turn constituted one larger civilization. Ancient native civilizations were not simply primitive people whose main source of culture was that which they derived via Spanish colonialism, rather they already possessed a very complex and intricate societal system, and way of life.

When comparing independent Native American groups there is typically a high inter-population variance expected, but the sampled native populations in our study had the lowest inter-population diversity in North America and the second lowest inter-population diversity (G_{ST} = 0.0170) globally, meaning they were all more closely related to each other than to other native populations in North America, and globally. This could be resultant from native town populations sharing more culture-practices *and* languages with other allied towns, rather than with a town that was simply geographically adjacent. We now know that elites from specific towns practiced various forms of endogamy in

order to remain powerful and increase their societal importance among the larger empire they were a part of (Kepics *et. al.*, 2005; Smith, 1986), so it is not unreasonable to suggest that if these elites practiced endogamy maybe their townsfolk, whom were farther down the societal scale, replicated such practices as well.

When comparing total heterozygosity globally, our native Mexican population had the second lowest value ($H_T = 0.7510$). Therefore it is logical to assume that the native groups we sampled were very closely related to each other so much that they were almost genetically identifiable as a single group, yet they were still significantly different from each other within their own population. This aligned with the recent historical and archaeological findings on Mesoamerican civilizations when you consider that some towns once had very strong relations with their allied towns, facilitating cross-culture, language infiltration, and possibly admixture. I reason that if this practice of endogamy occurred among various "town pairs" within an empire (Kepics *et al.*, 2005), then it is possible that this was a more widespread phenomenon that could have occurred to a sufficient degree that it resulted in genetic similarity within an empires or civilization, but not sufficiently enough to dilute the genetics of smaller founding native groups within these empires.

Delving further into the genetic relatedness between towns, we looked at the NJ-tree and dendogram (Figures 4 and 5, respectively), which demonstrated that the populations of Xochicalco and Puebla formed a cluster that was segregated away from Taxco and Cuernavaca. Although Taxco and Cuernavaca were geographically farther apart from each other than they were from Xochicalco and Puebla individually, they demonstrated more genetic relatedness probably because they maintained a stronger

relationship of trade and commerce with each other than with the other mentioned towns. This trade also existed in ancient times, but for different reasons. Alliances were forged by the ruling class for survival and power that led to admixture among them, and not with towns that simply shared geographic proximity. Today, Taxco is heavily associated with the mining and crafting of silver due to local silver mines, and Cuernavaca is a popular weekend and vacation town for the affluent of Mexico City. Currently frequent travel and migration exists between the two populations, but for tourism and the purchase of luxury goods, thus continuing the same alliances they once had, but for very different reasons.

Analysis of mtDNA haplotypes are very effective at elucidating the evolution and origins of polymorphism, as well as the migratory patterns of human populations. The five major mtDNA haplogroups that are found in the indigenous native populations of North and South America are A, B, C, D, and X.

It is believed that haplogroup A derived from Asia 50,000 years BCE, with the largest population, and greatest variety of haplogroup A found in east Asia, but with the highest frequencies found among the indigenous natives of the Americas (Fagundes *et al.*, 2008). Subgroup A2 is most commonly found among indigenous Americans from as far north as the Arctic Inuit to as south as Central America (Tanaka *et al.*, 2004); it's also the most common haplogroup found within the northernmost natives of Siberia (Volodko *et al.*, 2008).

Haplogroup B derived from Asia 50,000 years BCE. It is commonly said that since the indigenous ancestors of the Americas migrated from Siberia through Beringia into the Western hemisphere, one expected to find haplogroup B scattered throughout the

populations of Siberia and Beringia. This is not the case; haplogroup B (and X), have not yet been found among any Siberian tribes (Fagundes *et al.*, 2008). Subgroup B2 is restricted to the Americas, with a close phylogenetic analogue—B41b—found in populations of southern Asia and the Pacific Islands (Tabbada *et al.*, 2010; Peng *et al.*, 2011).

Haplogroup C is believed to have originated around the Caspian Sea region 60,000 years BCE, and is found in high frequency within populations of Siberia and northeast Asia; subgroup C1 is only found among indigenous native populations of the Americas, and in Asia (Volodko *et al.*, 2008).

Haplogroup D originated in Asia 48,000 years BCE. It is found in northeast Asia and Siberia, in central Asia where it is the second most prevalent mtDNA haplogroup, and found with low frequency in southwest Asia and Europe (Soares *et al.*, 2009; Pimenoff *et al.*, 2008; Comas *et al.*, 2004). Subgroup D1 is a branch stemming from D4, and is found throughout the Americas. Subgroup D4 is one of two principal branches of Haplogroup D (the other being D5'6), and is the most frequent mtDNA haplogroup found in the modern populations of Japan, Korea, Okinawa, and regions of northern China (Maruyama *et al.*, 2003; Lee *et al.*, 2006; Zheng *et al.*, 2011). Other sub-branches of D4 which are phylogenetic analogues of D1 are predominantly found in the Arctic populations, spanning from Siberia to the Inuit of Canada and Greenland (Helgason *et al.*, 2006).

Haplogroup X originated 30,000 years BCE when it branched off into X1 and X2 from haplogroup N (Soares *et al.*, 2009). X1 is restricted to North and East Africa, but X2 has had widespread dispersal since the last ice age 21,000 years BCE, and is found

with more prevalence in Mediterranean Europe. Subgroup X2a is typically found in North America but not South America (Perego *et al.*, 2009). Of the five major indigenous American haplogroups—A, B, C, D, and X—only X is not associated with Asia. In fact, only one small population in Siberia was found to have haplogroup X present, but further study revealed that this population likely settled in their location less than 5000 years BCE (Derenko *et al.*, 2001). The presence of X2 in the Americas, while remaining virtually absent throughout Asia, has been the prime reason why debate continues on varied hypotheses with no unified consensus on how X2 migrated to the New World; some have argued that haplogroup X2 was part of a founding South American population (Fagundes, 2009).

Previous studies have demonstrated that indigenous North American populations exhibit low haplogroup polymorphism when compared to Caucasian populations (Vargas-Alcaron et al., 2007), and that native Mexican haplotypes are mainly derived from the common Native American haplogroups of A2, B2, C1, and D1 (Sandoval et. al., 2009). These results were supported by our study as well; the majority of haplotype lineages (over 90%), were restricted to A2, B2, C1, D1, and D4.

The study of native Mexican Mayans showed they had an A2 average of 84%, C1 average of 8%, and B2 average of 4% (Ibarra-Rivera *et al*., 2008). Our study, comprised of mainly native Mexican Nahuas, resulted in an A2 average of 60%, C1 average of 11%, and B2 average of 23%. So, while the same haplotypes were identified between these two native groups (A2 making the bulk contribution in both), the ratio to which they were expressed differed significantly. This is probably due to differences in their founding populations (i.e. Beringia vs. Pacific (Mizuno *et al.*, 2014; Sandoval *et al.*, 2009);

discussed later). The greatest diversity was found in Cuernavaca, and probably a result of this city's urbanization, the large national and state university student and faculty populations, and its proximity to Mexico City.

Haplogroup X2a is typically found in native North American populations and believed to have derived from West Eurasia (Perego *et al.*, 2009). Its method of introduction into native North American populations is a topic of debate because X2a is not found in Central or South American populations. The prevailing theory is that X2a moved into North America via Beringia (the ice-free corridor connecting Asia to North America), and found a foothold among founding Native Americans. It is also theorized that peoples arriving to the Americas via the Pacific Ocean—and not Beringia—became the founding populations for Mesoamerica (Mizuno *et al*., 2014). It is speculated that the Pacific founders colonized the new land so effectively that they prevented the X2a mtDNA haplogroup from spreading into the south from the northern lands (Sandoval *et al*., 2009). Our findings do not reject this theory because haplotype X2a was completely absent from all our sampled native Mexicans, which were all geographically located in Mesoamerica. This lack of X2a also indirectly contributes to the previous study which showed that a cohort of natives from the American SW have a greater genetic resemblance to South Americans rather than to Mesoamericans, despite geography (Mizuno *et al*., 2014).

African populations were introduced to the Americas in the 18th century by European colonizers. These African groups mixed with local populations and resulted in a genetically non-studied "mulatto" cohort. It is suggested that the lack of genetic studies for this population is probably due to their minimal impact on the Mexican population—

diversity wise (Garcia-Ramos *et al*., 2003). The only mtDNA of African origin found in our study was restricted to Taxco and Cuernavaca, averaging 2% of the population, and consistent with expected results.

Current anthropological theory illustrated that founding Asian populations travelled through Beringia and migrated south along the Pacific coast, throughout the western hemisphere, reaching the southernmost tip of South America, and dispersed their genes via founding populations along their way. One is therefore expected to find mtDNA haplogroups of Asian/Western Eurasian origins; we found this in 6% of the sampled population, restricted to Puebla, Cuernavaca, and Tepoztlán. Of these haplotypes none were shared within or between any of the sampled individuals. This is likely the result of a few sampled individuals having a recent descendent of Asian/Western Eurasian origin, and why D4 was found only in 4% of the population, and then only in either Puebla or Cuernavaca.

Historical records described that most European women who migrated to Mexico settled in Mexico City, thus the reason for the recent expansion of haplogroup A2 and C1 and why a large number of mtDNA with European lineage was expected to be found (Guarado-Estrada *et al*., 2009). Our findings showed the opposite, suggesting that many native Mexicans preserved their native mtDNA lineage, that Europeans did not contribute as much genetic influence as previously thought, and that genetic drift was the main driver of mtDNA diversity. Nevertheless, I suggest that the over-representation of European males during the Spanish conquest may result in different findings, if Y-STR is analyzed—a worthwhile future study.

Clearly, STR markers are an important tool for human identification, and that the use of a core set of 15 STR loci has facilitated national and international data sharing, but it would be prudent to add additional markers to the 15 STR that comprise the "core set" to enhance detail. Our study did not encounter any technical problems, but improvements to commercial DNA extractions kits would also be welcomed because amplification is achieved using small amounts of DNA, and there have been reports of STR's instability and high mutation rate when using such small amounts (Dios *et al.*, 2001). For this single reason Single Nucleotide Polymorphism (SNP) markers have been explored as a likely successor to STR markers. Greater data recovery from poor or degraded DNA samples are possible when using SNP markers because only one nucleotide needs to be measured, as opposed to the sometimes hundreds of nucleotides that STRs require. SNPs also exhibit extremely low mutation rates when compared to that of STRs—approximately 100,000 lower (10^{-8} vs. 10^{-3})—thus being more stable in regards to inheritance (Butler *et al.*, 2007). These two reasons alone demonstrate why SNP markers would prove superior for forensic applications such as predicting identity, ancestry, paternity, kinship, or phenotypic traits, and for mtDNA testing.

There are currently a limited amount of genetic studies published on native Mexican populations, so more comparative DNA data would certainly be helpful to more completely assess correlations between native Mexican genetic conservation and cultural *and* language conservation. Thus I believe it paramount to increase the amount of sampled individuals and populations, so that the pool of genetic data for future studies will facilitate, refine, and increase accuracy of analytics, and in turn insight of founding populations.

Even though native Mexicans have a high proportion of culture and language preservation among their communities, one study showed that there was no significant link between the two (Sandoval *et al*., 2009). Our study would suggest this to be untrue, but an explanation for this discrepancy could be that geographical barriers also play a significant role in a population's genetic isolation and integration in conjunction with culture and language, or that not studying the sociology of ancient civilizations led to improper categorizations of native groups. It is possible that our observations were a combined result of the hidden variable of topography *as well as* culture and language. There are many studies that have theorized about the role geographical barriers play on migration patterns of early civilizations, and this theory could potentially be applied to Mesoamerica due to its vast and imposing network of mountain ranges, but to substantiate this theory one must conduct a systemic study while controlling for all the above variables. Future research could also include analysis demonstrating admixture dates, methods of admixture, and the genetic makeup *pre*-admixture.

At the very least, our study has contributed to the global DNA data pool, and will aid future forensic, evolutionary, migratory, and population studies of Mexico and Mesoamerica. And we have now used contemporary genetic tools to contribute and help progress the study of Mesoamerica in the same fashion that archaeologists came to the aid of historians, only decades ago, adding an new layer of analysis and meaning that clarified what really happened centuries ago to the lost civilizations of Mexico.

Chapter V

Appendix A

Parameters of interest for intra-population diversity of sampled towns in Mexico

TABLE 4. *Cuernavaca allelic frequencies (n = 123)*

Allele	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA
4						0.004									
5.3						0.004									
6						0.325									
7			0.008			0.378								0.077	
8			0.094			0.033	0.041	0.004				*0.508*			
9			0.029	0.002		0.033	0.317	0.093				0.045		0.049	
9.3						0.224									
10	0.057		0.176	0.207			0.118	0.280				0.016	0.004	0.020	
11	0.049		0.361	0.276			0.138	0.297				0.262	0.008	*0.553*	
11.2													0.008		
12	0.142		0.246	0.431			0.187	0.256		0.049		0.168	0.073	0.232	
12.2										0.033			0.004		
13	0.346		0.078	0.065	0.020		0.114	0.069		0.183			0.134	0.071	
13.2										0.183					
14	0.256		0.008		0.024		0.085			0.215	0.024		0.171	0.009	
14.2										0.041					
15	0.122				*0.516*					0.106	0.045		0.167		
15.2										0.138					
16	0.020				0.305				0.020	0.024	0.382		0.114		0.004
16.2										0.012					
17	0.008				0.081				0.142	0.008	0.386		0.126		
17.2										0.008					
18					0.045				0.049		0.130		0.061		0.008
18.2															
19					0.008				0.276		0.028		0.041		0.065
20									0.138		0.004		0.024		0.057
21									0.012				0.033		0.073
22									0.106				0.024		0.089
23									0.150				0.004		0.114
23.2															0.016
24									0.081				0.004		0.154
24.2															
25									0.020						0.167
26									0.004						0.195
27															0.045
28		0.077													0.008
29		0.195													
29.2		0.004													
30		0.280													
30.2		0.004													
31		0.057													
31.2		0.150													
32															
32.2		0.171													
33.2		0.057													
34.2		0.004													
43.2															0.004
GDI	0.774	0.819	0.763	0.690	0.631	0.699	0.810	0.754	0.841	0.850	0.685	0.642	0.883	0.629	0.874
H_O	0.748	0.789	0.672	0.650	0.577	0.715	0.837	0.699	0.870	0.829	0.756	0.684	0.894	0.626	0.837
H_E	0.777	0.822	0.766	0.693	0.633	0.702	0.813	0.757	0.844	0.854	0.687	0.649	0.887	0.631	0.878
P- value	0.750	0.659	0.455	0.025	0.052	0.841	0.341	0.291	0.490	0.120	0.125	0.584	0.865	0.975	0.023
MP	0.083	0.059	0.089	0.157	0.198	0.149	0.071	0.103	0.050	0.048	0.187	0.191	0.031	0.175	0.037
PD	0.917	0.941	0.911	0.843	0.802	0.851	0.929	0.897	0.950	0.952	0.813	0.809	0.969	0.825	0.963
PIC	0.740	0.800	0.730	0.640	0.570	0.640	0.790	0.071	0.820	0.830	0.630	0.590	0.870	0.590	0.860
PE	0.506	0.578	0.386	0.356	0.264	0.453	0.670	0.427	0.734	0.654	0.520	0.352	0.784	0.323	0.670
TPI	1.980	2.370	1.530	1.430	1.180	1.760	3.080	1.660	3.840	2.930	2.050	1.420	4.730	1.340	3.080

H_O, observed heterozygosity; H_E, expected heterozygosity; GDI, Gene Diversity Index; MP, Matching Probability; PD, Power of Discrimination; PIC, Polymorphic Information Content; PE, Power of Exclusion; TPI, Typical Paternity Index.

TABLE 5. *Puebla allelic frequencies (n = 56)*

Allele	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA
4															
5.3															
6						0.393									
7						0.393								0.018	
8			0.071	0.009		0.036	0.054					0.473			
9			0.054	0.063		0.036	0.277	0.071				0.018		0.080	
9.3						0.143									
10	0.036		0.232	0.205			0.205	0.259				0.009		0.116	
11	0.045		0.375	0.295			0.143	0.268				0.232		*0.509*	
11.2															
12	0.143		0.241	0.384			0.170	0.375		0.036		0.250	0.116	0.196	
12.2															
13	0.402		0.027	0.045			0.098	0.027		0.188		0.018	0.098	0.071	
13.2										0.125					
14	0.241				0.027		0.054			0.241	0.054		0.116	0.009	
14.2										0.054					
15	0.116				0.491					0.116	0.036		0.179		
15.2										0.170					
16	0.018				0.304					0.054	0.330		0.134		
16.2										0.018					
17					0.134				0.116		0.304		0.196		0.009
17.2															
18					0.036				0.045		0.214		0.008		0.018
18.2															
19					0.009				0.277		0.045		0.045		0.045
20									0.116		0.018		0.009		0.036
21									0.018				0.027		0.143
22									0.089						0.179
23									0.188						0.008
23.2															
24									0.098						0.214
24.2															
25									0.045						0.143
26									0.009						0.116
27															0.018
28		0.071													
29		0.170													
29.2															
30		0.304													
30.2		0.009													
31		0.098													
31.2		0.063													
32															
32.2		0.143													
33.2		0.134													
34.2		0.009													
43.2															
GDI	0.743	0.822	0.739	0.718	0.647	0.668	0.817	0.715	0.839	0.841	0.746	0.659	0.866	0.677	0.858
H_O	0.750	0.839	0.839	0.696	0.571	0.714	0.911	0.714	0.821	0.839	0.750	0.643	0.786	0.661	0.857
H_E	0.750	0.829	0.745	0.724	0.653	0.674	0.824	0.721	0.847	0.849	0.753	0.665	0.874	0.683	0.865
P-value	0.809	0.503	0.865	0.720	0.108	0.840	0.432	0.429	0.549	0.637	0.644	0.705	0.049	0.655	0.209
MP	0.108	0.064	0.131	0.121	0.195	0.189	0.078	0.146	0.054	0.057	0.108	0.180	0.046	0.140	0.052
PD	0.892	0.936	0.869	0.879	0.805	0.811	0.922	0.854	0.946	0.943	0.892	0.820	0.954	0.860	0.948
PIC	0.710	0.800	0.700	0.670	0.590	0.610	0.790	0.660	0.820	0.820	0.700	0.600	0.850	0.640	0.840
PE	0.510	0.674	0.674	0.423	0.258	0.451	0.817	0.451	0.639	0.674	0.510	0.345	0.573	0.370	0.709
TPI	2.000	3.110	3.110	1.650	1.170	1.750	5.600	1.750	2.800	3.110	2.000	1.400	2.330	1.470	3.500

H_O, observed heterozygosity; H_E, expected heterozygosity; GDI, Gene Diversity Index; MP, Matching Probability; PD, Power of Discrimination; PIC, Polymorphic Information Content; PE, Power of Exclusion; TPI, Typical Paternity Index.

TABLE 6. *Taxco allelic frequencies (n = 22)*

Allele	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA
4															
5.3															
6						0.386									
7						0.364								0.068	
8			0.136			0.023						0.341			
9				0.045		0.023	0.295	0.023						0.091	
9.3						0.205									
10	0.091		0.227	0.205			0.068	0.341				0.045		0.045	
11	0.023		0.409	0.182			0.205	0.273				0.432		*0.523*	
11.2										0.023					
12	0.114		0.159	*0.500*			0.227	0.273		0.023		0.182	0.091	0.182	
12.2										0.023					
13	0.318		0.068	0.045			0.159	0.068		0.114			0.091	0.091	
13.2										0.114					
14	0.273			0.023	0.091		0.045	0.023		0.227	0.023		0.227		
14.2										0.023					
15	0.114				*0.545*					0.182	0.023		0.136		
15.2										0.205					
16	0.023				0.227						*0.523*		0.114		
16.2										0.068					
17	0.045				0.068				0.114		0.227		0.182		
17.2															
18					0.068				0.045		0.091		0.091		
18.2															
19									0.273		0.114		0.023		0.045
20									0.227				0.023		0.045
21													0.023		0.114
22									0.045						0.068
23									0.227						0.159
23.2															
24									0.068						0.159
24.2															
25															0.136
26															0.182
27															0.068
28		0.068													0.023
29		0.136													
29.2															
30		0.318													
30.2															
31		0.045													
31.2		0.227													
32															
32.2		0.182													
33.2		0.023													
34.2															
43.2															
GDI	0.787	0.788	0.732	0.671	0.633	0.676	0.787	0.729	0.801	0.841	0.653	0.662	0.857	0.671	0.871
H_O	0.818	0.818	0.682	0.545	0.636	0.636	0.955	0.864	0.818	0.909	0.727	0.545	0.727	0.636	0.955
H_E	0.806	0.807	0.749	0.686	0.648	0.691	0.806	0.746	0.819	0.860	0.668	0.678	0.877	0.686	0.891
P-value	0.715	0.662	0.625	0.271	0.200	0.938	0.780	0.212	0.910	0.948	0.667	0.316	0.227	0.152	0.456
MP	0.095	0.099	0.128	0.165	0.211	0.161	0.116	0.078	0.087	0.070	0.190	0.170	0.062	0.186	0.074
PD	0.905	0.901	0.872	0.835	0.789	0.839	0.884	0.822	0.913	0.930	0.810	0.826	0.938	0.814	0.926
PIC	0.760	0.760	0.690	0.630	0.590	0.610	0.750	0.680	0.770	0.820	0.610	0.600	0.840	0.640	0.860
PE	0.633	0.633	0.401	0.230	0.337	0.337	0.908	0.722	0.633	0.814	0.472	0.230	0.472	0.370	0.908
TPI	2.750	2.750	1.570	1.100	1.380	1.380	11.000	3.670	2.750	5.500	1.830	1.100	1.830	1.470	11.000

H_O, observed heterozygosity; H_E, expected heterozygosity; GDI, Gene Diversity Index; MP, Matching Probability; PD, Power of Discrimination; PIC, Polymorphic Information Content; PE, Power of Exclusion; TPI, Typical Paternity Index.

TABLE 7. *Xochicalco allelic frequencies (n = 22)*

Allele	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA
4															
5.3															
6						0.341									
7						0.386								0.045	
8	0.023		0.091				*0.680*					*0.614*			
9			0.023	0.023		0.045	0.364					0.023		0.136	
9.3						0.227									
10	0.091		0.205	0.250			0.068	0.318					0.023	0.023	
11	0.091		0.364	0.364			0.205	0.295			0.023	0.205		*0.705*	
11.2															
12	0.136		0.159	0.273			0.205	0.341		0.045		0.159	0.068	0.068	
12.2															
13	0.364		0.114	0.091			0.068	0.045		0.114			0.114	0.023	
13.2										0.159					
14	0.205		0.045		0.023		0.023			0.205	0.023		0.182		
14.2										0.068					
15	0.045				*0.659*					0.114	0.023		0.159		
15.2										0.205					
16	0.045				0.182					0.023	0.318		0.114		
16.2										0.068					
17					0.114				0.023		0.273		0.136		
17.2															
18					0.023				0.114		0.250		0.114		0.045
18.2															
19									0.318		0.068		0.045		0.227
20									0.068		0.023				
21									0.091						0.068
22									0.136				0.023		0.068
23									0.182						0.023
23.2															
24									0.068				0.023		0.273
24.2															
25															0.136
26															0.159
27															
28		0.091													
29		0.205													
29.2															
30		0.273													
30.2															
31		0.045													
31.2		0.091													
32		0.023													
32.2		0.182													
33.2		0.091													
34.2															
43.2															
GDI	0.786	0.823	0.777	0.722	0.519	0.681	0.770	0.693	0.816	0.853	0.755	0.556	0.876	0.477	0.818
H_O	0.773	0.773	0.864	0.727	0.409	0.682	0.682	0.773	0.864	0.909	0.909	0.545	0.864	0.409	0.818
H_E	0.804	0.842	0.795	0.739	0.531	0.697	0.788	0.709	0.835	0.873	0.773	0.569	0.896	0.488	0.837
P- value	0.442	0.630	0.415	0.390	0.129	0.529	0.424	0.929	0.902	0.877	0.669	0.328	0.354	0.275	0.737
MP	0.095	0.074	0.116	0.157	0.302	0.186	0.099	0.182	0.079	0.066	0.145	0.273	0.066	0.331	0.083
PD	0.905	0.926	0.884	0.843	0.698	0.814	0.901	0.818	0.921	0.934	0.855	0.727	0.934	0.669	0.917
PIC	0.760	0.800	0.750	0.670	0.480	0.620	0.740	0.063	0.790	0.840	0.710	0.500	0.860	0.450	0.790
PE	0.549	0.546	0.722	0.472	0.120	0.401	0.401	0.549	0.722	0.814	0.814	0.230	0.722	0.120	0.633
TPI	2.200	2.200	3.670	1.830	0.850	1.570	1.570	2.200	3.670	5.500	5.500	1.100	3.670	0.850	2.750

H_O, observed heterozygosity; H_E, expected heterozygosity; GDI, Gene Diversity Index; MP, Matching Probability; PD, Power of Discrimination; PIC, Polymorphic Information Content; PE, Power of Exclusion; TPI, Typical Paternity Index.

TABLE 8. *Student allelic frequencies (n = 21)*

Allele	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA
4															
5.3															
6						0.333									
7				0.024		0.357								0.048	
8	0.024		0.024			0.024	0.119					0.429			
9			0.024			0.048	0.214	0.119				0.119		0.024	
9.3						0.238									
10	0.048		0.357	0.286			0.071	0.167				0.048	0.048	0.048	
11	0.048		0.333	0.190			0.143	0.283				0.286		0.381	
11.2															
12	0.095		0.214	0.333			0.310	0.262		0.119		0.119	0.143	0.405	
12.2										0.048					
13	0.310		0.024	0.143			0.095	0.214		0.214	0.024		0.238	0.095	
13.2										0.071					
14	0.333		0.024	0.024	0.024		0.048			0.190	0.095		0.167		
14.2										0.048					
15	0.119				0.476					0.190	0.143		0.095		
15.2										0.119					
16	0.024				0.310						0.262		0.167		
16.2															
17					0.143				0.238		0.214		0.048		
17.2															
18					0.048				0.048		0.190		0.071		
18.2															
19									0.214		0.071				0.119
20									0.119						0.095
21									0.024				0.024		0.143
22									0.071						0.071
23									0.167						0.143
23.2															
24									0.071						0.167
24.2		0.024													
25									0.024						0.238
26									0.024						0.024
27		0.024													
28		0.095													
29		0.167													
29.2															
30		0.238													
30.2															
31		0.071													
31.2		0.071													
32															
32.2		0.262													
33.2		0.048													
34.2															
43.2															
GDI	0.764	0.824	0.713	0.749	0.654	0.702	0.807	0.787	0.841	0.844	0.814	0.704	0.848	0.677	0.846
H_O	0.714	0.857	0.667	0.619	0.762	0.810	0.905	0.714	0.857	0.810	0.762	0.762	0.905	0.571	0.905
H_E	0.783	0.844	0.731	0.768	0.670	0.719	0.827	0.806	0.861	0.864	0.834	0.721	0.869	0.693	0.866
P-value	0.687	0.912	0.409	0.511	0.131	0.153	0.437	0.716	0.752	0.573	0.058	0.079	0.062	0.148	0.626
MP	0.098	0.008	0.143	0.111	0.256	0.206	0.107	0.093	0.070	0.070	0.107	0.197	0.102	0.175	0.079
PD	0.902	0.921	0.857	0.889	0.744	0.794	0.893	0.907	0.930	0.930	0.893	0.803	0.898	0.825	0.921
PIC	0.730	0.800	0.660	0.710	0.600	0.640	0.780	0.750	0.820	0.820	0.790	0.660	0.830	0.620	0.830
PE	0.451	0.709	0.379	0.314	0.530	0.617	0.805	0.451	0.709	0.617	0.530	0.530	0.805	0.258	0.805
TPI	1.750	3.500	1.500	1.310	2.100	2.630	5.250	1.750	3.500	2.630	2.100	2.100	5.250	1.170	5.250

H_O, observed heterozygosity; H_E, expected heterozygosity; GDI, Gene Diversity Index; MP, Matching Probability; PD, Power of Discrimination; PIC, Polymorphic Info-rmation Content; PE, Power of Exclusion; TPI, Typical Paternity Index.

Appendix B

Parameters of genetic interest and heterozygosity of global populations

TABLE 9. *Parameters of population genetics interest*

	N	No. of alleles	CMP	CPD	CPE	Avg. H_E (15 loci)	Avg. GDI (15 loci)
Cuernavaca	123	133	4.01413E+15	0.99999999999999974599	0.9999913312	0.759	0.756
Puebla	56	111	1.06803E+15	0.99999999999999907858	0.9999813290	0.764	0.757
Taxco	22	101	4.14119E+13	0.99999999999997606158	0.9999994068	0.761	0.744
Xochicalco	22	102	2.05721E+13	0.99999999999921238488	0.9999971544	0.745	0.728
Students	21	105	1.01524E+14	0.99999999999983382552	0.9999990889	0.790	0.772

N, number of individuals; CMP, Combined Matching Probablilities; CPD, Combined Power of Discrimination; CPE, Combined Power of Exclusion; GDI, Gene Diversity Index.

TABLE 10. Inter- and Intra-population genetic variance

	Native Mexicans			Meso Americans†			South Americans			North Americans‡			Asians			Europeans			Native Americans			Average of All Populations		
	G_{ST}	H_S	H_T	G_{ST}	H_S	H_T	G_{ST}	H_S	H_T	G_{ST}	H_S	H_T	G_{ST}	H_S	H_T	G_{ST}	H_S	H_T	G_{ST}	H_S	H_T	G_{ST}	H_S	H_T
D8S1179	0.0072	0.7707	0.7763	0.0208	0.7427	0.7585	0.0040	0.7875	0.7906	0.0105	0.7157	0.7232	0.0041	0.8426	0.8461	0.0092	0.7979	0.8053	0.0197	0.7449	0.7599	0.0106	0.7717	0.7800
D21S11	0.0109	0.8153	0.8243	0.0547	0.8006	0.8469	0.0031	0.8369	0.8395	0.0205	0.7896	0.8061	0.0061	0.8052	0.8102	0.0064	0.8289	0.8342	0.0600	0.8047	0.8561	0.0234	0.8116	0.8310
D7S820	0.0119	0.7448	0.7537	0.0169	0.7343	0.7469	0.0111	0.7559	0.7644	0.0228	0.7724	0.7905	0.0041	0.7560	0.7591	0.0085	0.8085	0.8154	0.0280	0.7458	0.7673	0.0148	0.7597	0.7710
CSF1PO	0.0192	0.7099	0.7238	0.0533	0.6924	0.7314	0.0050	0.7215	0.7251	0.0130	0.7181	0.7275	0.0043	0.7255	0.7287	0.0256	0.7235	0.7426	0.0561	0.7019	0.7436	0.0254	0.7133	0.7318
D3S1358	0.0141	0.6168	0.6256	0.0867	0.6601	0.7228	0.0050	0.7215	0.7251	0.0233	0.6528	0.6684	0.0021	0.7102	0.7117	0.0070	0.7730	0.7784	0.1417	0.6453	0.7518	0.0410	0.6828	0.7120
THO1	0.0034	0.6852	0.6876	0.0215	0.6415	0.6556	0.0079	0.7102	0.7159	0.0734	0.4764	0.5141	0.0114	0.6739	0.6816	0.0127	0.7961	0.8063	0.0877	0.6201	0.6798	0.0290	0.6576	0.6773
D13S317	0.0141	0.7980	0.8094	0.0152	0.8033	0.8157	0.0185	0.8105	0.8257	0.0273	0.7690	0.7906	0.0059	0.8027	0.8074	0.0023	0.7764	0.7782	0.0345	0.7977	0.8262	0.0169	0.7939	0.8076
D16S539	0.0171	0.7357	0.7485	0.0132	0.7492	0.7592	0.0045	0.7852	0.7887	0.0409	0.7039	0.7339	0.0048	0.7736	0.7773	0.0048	0.7581	0.7617	0.0245	0.7456	0.7643	0.0154	0.7502	0.7619
vWA	0.0252	0.7306	0.7495	0.0148	0.7140	0.7247	0.0076	0.7188	0.7243	0.0245	0.7669	0.7862	0.0049	0.7925	0.7964	0.0058	0.7827	0.7872	0.0198	0.7253	0.7399	0.0146	0.7473	0.7583
TPOX	0.0277	0.6445	0.6629	0.0218	0.6333	0.6474	0.0047	0.6244	0.6274	0.0269	0.6578	0.6760	0.0048	0.6303	0.6443	0.0013	0.6769	0.6777	0.0456	0.6364	0.6668	0.0215	0.6434	0.6575
D18S51	0.0119	0.8660	0.8764	0.0069	0.8533	0.8592	0.0086	0.8490	0.8564	0.0371	0.8152	0.8466	0.0022	0.8562	0.8581	0.0049	0.8618	0.8660	0.0171	0.8453	0.8600	0.0126	0.8495	0.8604
D5S818	0.0404	0.6258	0.6521	0.0097	0.6514	0.6578	0.0143	0.7203	0.7307	0.0148	0.7143	0.7250	0.0040	0.7889	0.7921	0.0062	0.7099	0.7143	0.0179	0.6753	0.6876	0.0149	0.6980	0.7085
FGA	0.0223	0.8533	0.8728	0.0141	0.8572	0.8695	0.0068	0.8686	0.8745	0.0171	0.8542	0.8691	0.0023	0.8567	0.8587	0.0126	0.8619	0.8729	0.0187	0.8584	0.8748	0.0134	0.8586	0.8703
All Loci	0.0170	0.7382	0.7510	0.0268	0.7333	0.7535	0.0078	0.7623	0.7683	0.0260	0.7236	0.7429	0.0046	0.7703	0.7739	0.0083	0.7812	0.7877	0.0432	0.7344	0.7675	0.0192	0.7490	0.7637

† This group does not include populations from this study.; ‡ This group does not include populations from Mexico.

Appendix C

Graphic representation of mtDNA lineage origins for Mexico

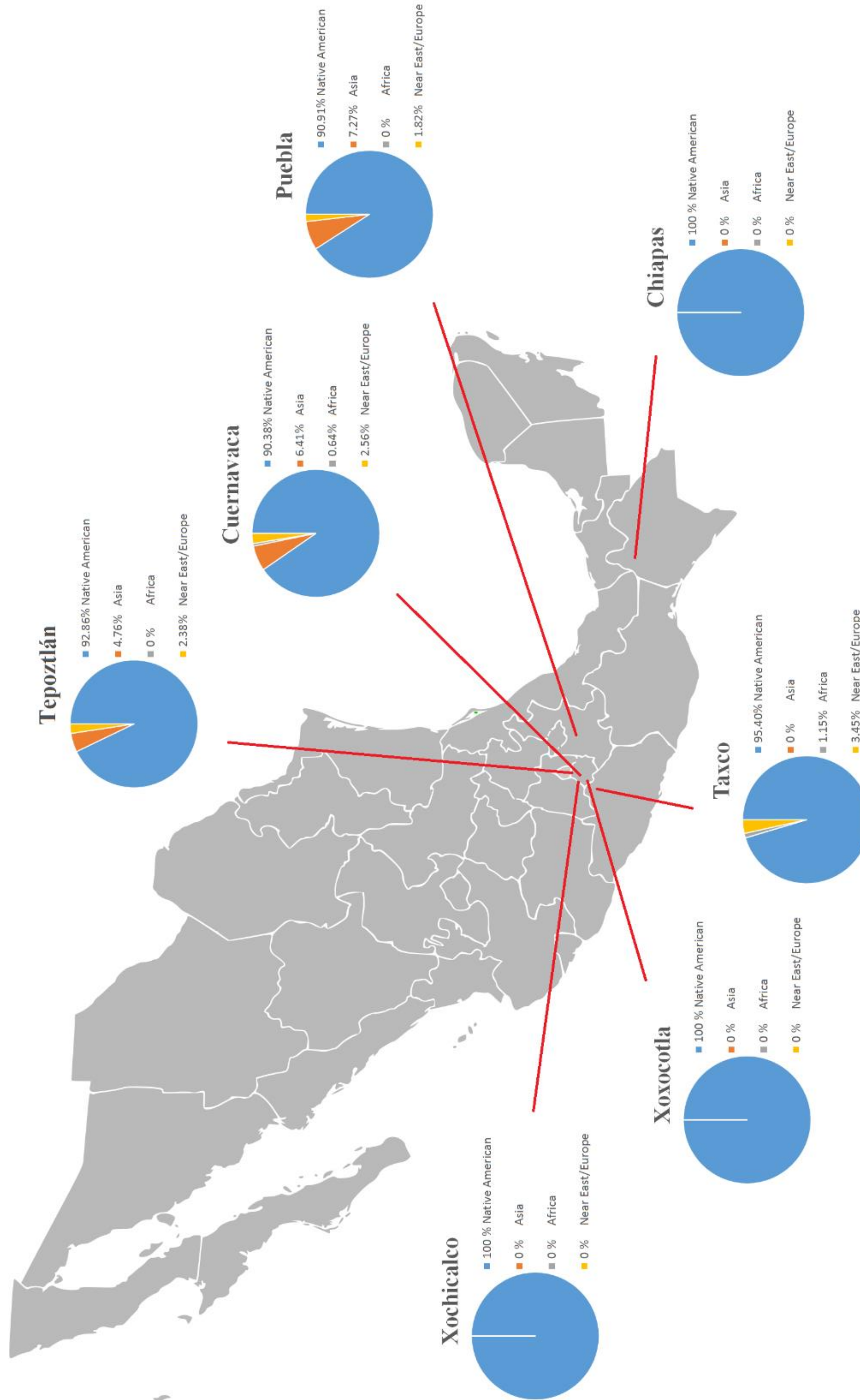

Figure 6. Lineage origins of mtDNA for sampled Native Mexican populations.

Appendix D

Formulae and types of analysis

Hardy-Weinberg Equilibrium (HWE) is a principle that states genotype and allele frequencies in a population will remain constant from generation to generation when random mating occurs, and there are no evolutionary influences (such as mate choice, mutation, selection, genetic drift, gene flow, and meiotic drive). Since these influences typically occur in real populations, the HWE principle applies in an ideal scenario. It is described as

$$p^2 + 2pq + q^2 = 1\,,$$

where p and q represent each of two alleles at a single locus (Weinberg, 1908; Hardy, 1908).

Fixation index (F_{ST}), is the measure of genetic difference between two populations, using genetic polymorphism data. In the Reich *et al*., 2009 study, F_{ST} was defined as

$$F_{ST} = \frac{N}{D},$$

where

$$N = p_1(q_2 - q_1) + p_2(q_1 - q_2)$$

$$D = p_1q_2 + q_1p_2 = N + p_1q_1 + p_2q_2$$

In *this* thesis, F_{ST} was computed using the Arlequin v3.5.1.2 software, which used the formula

$$F_{ST} = \frac{f_0 - \bar{f}}{1 - \bar{f}},$$

where f_0 is the probability of identity by descent of two different genes drawn from the same population, and $\bar{f}$ is the probability of identity by descent of two genes drawn from two different populations (Excoffier & Lischer, 2010).

Gene Diversity Index (GDI), is a measure of the degree of genetic polymorphism in a population, and is defined as

$$h = 1 - \sum p_i^2 ,$$

where p_i is the population frequency of each allele at locus *i* (Nei, 1987).

Matching Probability (MP), is the probability that two randomly selected unrelated individuals would have identical genotypes, and is represented as

$$pM = \sum_{i=a}^{n} \sum_{j \geq 1}^{n} P_{ij}^2 ,$$

where *i* and *j* represent the frequencies of all possible alleles *a* through *n*, and P_{ij} represents the frequencies of all possible genotypes (Huston, 1998).

Power of Discrimination (PD), is a measure of how likely it is that two randomly selected unrelated individuals will have different genotypes, and is defined as

$$P_d = 1 - pM ,$$

where pM is the matching probability (Huston, 1998).

Polymorphic Information Content (PIC), is a measure of the informativeness of a genetic marker for linkage studies, and is defined as

$$PIC = 1 - \sum_{i=1} Pi^2 - \sum_{i=1} \sum_{j=i+1} Pi^2 \, Pj^2 \, ,$$

where $\sum Pi^2$ is the sum of each squared allele of the ith frequency in the population (Botstein *et al.*, 1980).

Power of Exclusion (PE), is the power of a test to eliminate a certain percentage of the population from being genetically related to an individual, and is defined as

$$PE = \, h^2(1 - 2hH^2) \, ,$$

where $h + H = 1, h = \, n_h/n$ and n_h is the number of individual observations with two alleles and n is the total number of individuals (Huston, 1998).

Typical Paternity Index (TPI), is a likelihood ratio between the chances of observing data in a parentage trio if an alleged father passed down an allele, compared to the likelihood of observing the data if a random individual passing down the same allele in question, and is defined as

$$PI_{typical} = \frac{1}{2H}$$

(Huston, 1998).

Combined Matching Probability (CMP), is the probability that two randomly selected unrelated individuals would have identical genotypes at all loci, and is defined as

$$pM_{combined} = 1 - \prod_{i=1}^{n} (1 - \, pM_i)$$

(Huston, 1998).

Combined Power of Discrimination (CPD), is a measure of how likely it is that two randomly selected unrelated individuals will have different genotypes at all loci, and is defined as

$$P_{d\ combined} = 1 - \prod_{i=1}^{n} (1 - P_{di})$$

(Huston, 1998).

Combined Power of Exclusion (CPE), is the power of a test to eliminate a certain percentage of the population from being genetically related to an individual at all loci, and is defined as

$$PE_{combined} = 1 - \prod_{i=1}^{n} (1 - PE_i)$$

(Huston, 1998).

Expected Heterozygosity (H_e), is represented by

$$H_e = 1 - \sum_{i=1}^{n} q_i^2 ,$$

where q_i is the frequency of the ith of n alleles at a locus (Lodish *et al.*, 2000).

Observed Heterozygosity (H_o), is represented by

$$H_o = \frac{\sum_{i=1}^{n} (1\ if\ a_{i1} \neq a_{i2})}{n},$$

where n is the number of individuals in the population, and a_{i1}, a_{i2} are the alleles of individual i at the target locus (Lodish *et al.*, 2000).

Correspondence analysis (CA), is a statistical visualization method for picturing the observed association of two traits; inferring whether certain levels of one characteristic are associated with some levels of another (Greenacre, 2007).

Neighbor Joining is an agglomerative clustering method to create phylogenetic trees based on DNA data. The algorithm requires knowledge of the distance between each pair of taxa, in matrix form. It is calculated as

$$Q\,(i,j) = (n-2)d(i,j) - \sum_{k=1}^{n} d(i,k) - \sum_{k=1}^{n} d(j,k)\,,$$

where $d(i,j)$ is the distance between taxa i and j, and n is the taxa. Distance from the above pair to a new node is calculated by

$$\delta(f,u) = \frac{1}{2}\,d(f,g) + \frac{1}{2(n-2)}\left[\sum_{k=1}^{n} d(f,k) - \sum_{k=1}^{n} d(g,k)\right],$$

and

$$\delta(g,u) = d(f,g) - \delta(f,u)\,,$$

where taxa f and g are the paired taxa and u is the newly created node. The branches joining f and u and g and u, and their lengths, $\delta(f,u)$ and $\delta(g,u)$ are part of the tree which is gradually being created. For the remaining taxa we calculate the distance of the new node by

$$d(u,k) = \frac{1}{2}[d(f,k) + d(g,k) - d(f,g)]\,,$$

where *u* is the new node, *k* is the node which we want to calculate the distance to and *f* and *g* are the members of the pair just joined (Nei, 1987).

Bonferroni correction is an adjustment made to critical *P* values, in order to avoid false positives, when multiple dependent or independent statistical tests are being calculated simultaneously on a single data set. The statistical power of the study is then assessed based on the *modified P* value. It is calculated as

$$P_{mod} = \frac{\alpha}{m},$$

where α is the critical *P* value, and *m* is the number of comparisons being made (Bonferroni, 1935).

The **Tau** (τ) test measures the degree of similarity between two sets of rankings (given to the same sets of items), and assesses the significance of this similarity. It is defined as

$$\tau = \frac{n_c - n_d}{n_0},$$

where n_c is the number of concordant pairs, n_d is the number of discordant pairs, and $n_0 = n\,(n-1)/2$ (Kendall, 1938).

Tajima's *D* neutrality test is a statistical test for natural selection; illustrating the allele frequency distribution of nucleotide sequence data based on the differences between the average number of pairwise differences between sequences, and the number of segregating sites in the sample. It compares two estimators of the mutation parameter

theta ($\theta = 2Mu$, with $M = 2N$ in diploid populations or $M = N$ in haploid populations of effective size N), and is defined as

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

where $\hat{\theta}_\pi = \hat{\pi}$ and $\hat{\theta}_S = S/\sum_{i=0}^{n-1} 1/i$, and S is the number of segregation sites in the sample (Tajima, 1989).

Chi-squared test (χ^2) is a statistical test applied to sets of categorical data that calculates how likely any observed differences between data sets arose by chance. It is typically used to test a null hypothesis and is defined as

$$\chi^2 = \sum_{i-1}^{n} \frac{(O_i - E_i)^2}{E_i},$$

where χ^2 is the cumulative test statistic; asymptotically approaching a χ^2 distribution, O_i is the number of observations of type *i*, E_i is the expected frequency of type *i*, an *n* is the number of cells in the table (Greenwood & Nikulin, 1996).

G_{ST} is a measure of genetic differentiation used to describe the total amount of variation observed among populations, over multiple loci. It is defined as

$$G_{ST} = 1 - \frac{H_S}{H_T},$$

where H_S is the average heterozygosity within subpopulations, and H_T is the average heterozygosity within the total population (Ryman & Leimar, 2009).

References

Bonferroni, C. E. (1935). Il calcolo delle assicurazioni su gruppi di teste. In Studi in onore del Professore Salvatore Ortu Carboni (13-60). Rome, IT: Bardi.

Boone, E. H. (2010). *Stories in red and black: Pictorial histories of the Aztecs and Mixtecs.* Austin, TX: University of Texas Press.

Botstein, D., White, R. L., Skolnick, M., & Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*, *32*, 314.

Butler, J. M. (2006). Genetics and genomics of core short tandem repeat loci used in human identity testing. *Journal of Forensic Sciences, 51*, 253-265.

Butler, J. M., Coble, M. D., & Vallone, P. M. (2007). STRs vs SNPs: Thoughts on the future of forensic DNA testing. *Forensic Science, Medicine, and Pathology, 3,* 200-205.

Campbell, L. (1997). *American Indian languages: The historical linguistics of native America*. New York, NY: Oxford University Press.

Campbell, L., & Kaufman, T. (1976). A linguistic look at the Olmec. *American Antiquity, 41*, 80-89.

Cantu-Bolan, B. (2008). Linguistic vitality and language shift of Nahutlahtolli in Xoxocotla, Morelos (Mexico). *International Sociological Association*. UNAM, Mexico. 2008. Conference Paper.

Clement, M., Derington, J., Woolley, S., & Posada, D. (2005). Phylogenetic network estimation using statistical parsimony (Version 1.21) [Software]. Available from http://darwin.uvigo.es/software/tcs.html

Coe, M. D. (1987). *The Maya* (4th ed.). New York, NY: Thames and Hudson.

Comas, D., Plaza, S., Wells, R. S., Yuldaseva, N., Lao, O., Calafell, F., & Bertranpetit, J. (2004). Admixture, migrations, and dispersals in Central Asia: evidence from maternal DNA lineages. *European Journal of Human Genetics*, *12*, 495-504.

Derenko, M. V., Grzybowski, T., Malyarchuk, B. A., Czarny, J., Miścicka-Śliwka, D., & Zakharov, I. A. (2001). The presence of mitochondrial haplogroup X in Altaians from South Siberia. *American Journal of Human Genetics*, *69*, 237.

Dios, S., Luis, J. R., Carril, J. C., & Caeiro, B. (2001). Sub-Saharan genetic contribution in Morocco: Microsatellite DNA analysis. *Human Biology, 73*, 675-688.

Edwards, A., Civitello, A., Hammond H. A., & Caskey, C. T. (1991). DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *American Journal of Human Genetics, 49*, 746-756.

Edwards, A., Hammond, H. A., Jin, L., Caskey, C. T., & Chakraborty, R. (1992). Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics, 12*, 241-253.

Excoffier, L., & Lischer, H. (2010). Arlequin: An integrated software package for population genetics data analysis (Version 3.5.1.2) [Software]. Available from http://cmpg.unibe.ch/software/arlequin35

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

Felsenstein, J. (2015). PHYLIP (Version 3.695) [Software]. Available from http://evolution.genetics.washington.edu/phylip.html

Frawley, W., Hill, K. C., & Munro, P. (2002). *Making dictionaries.* Los Angeles, CA: University of California Press.

García-Ramos, G., Téllez-Zenteno, J. F., Zapata-Zúñiga, M., Yamamoto-Furusho, J. K., Ruiz-Morales, J. A., Villareal-Garza, C., . . . Granados, J. (2003). HLA class II genotypes in Mexican Mestizo patients with myasthenia gravis. *European Journal of Neurology, 10*, 707-710.

González-Martín, A., Gorostiza, A., Rangel-Villalobos, H., Acunha, V., Barrot, C., Sánchez, C., . . . Calderón, R. (2008). Analyzing the genetic structure of the Tepehua in relation to other neighbouring Mesoamerican populations: A study based on allele frequencies of STR markers. *American Journal of Human Biology, 20*, 605-613.

Goodwin, W., Lincare, A., & Hadi, S. (2011). *An introduction to forensic genetics*. Oxford, UK: Wiley-Blackwell.

Greenacre, M. (2007). *Correspondence analysis in practice* (2nd ed.). London, UK: Chapman & Hall.

Greenwood, P. E., & Nikulin, M. S. (1996). *A guide to chi-squared testing* (Vol. 280). New York, NY: John Wiley & Sons.

Guardado-Estrada, M., Juarez-Torres, E., Medina-Martinez, I., Wegier, A., Macias, A., Gomez, G., . . . Berumen, J. (2009). A great diversity of Amerindian mitochondrial DNA ancestry is present in the Mexican mestizo population. *Journal of Human Genetics, 54*, 695-705.

Hamilton, M. B. (2009). *Population genetics.* Oxford, UK: Wiley-Blackwell.

Hardy, G. H. (1908). Mendelian proportions in a mixed population. *Science, 28*(706), 49-50.

Helgason, A., Pálsson, G., Pedersen, H. S., Angulalik, E., Gunnarsdóttir, E. D., Yngvadóttir, B., & Stefánsson, K. (2006). mtDNA variation in Inuit populations of Greenland and Canada: migration history and population structure. *American Journal of Physical Anthropology*, *130*, 123-134.

Huston, K. A. (1998). Statistical analysis of STR data. *Profiles DNA*, *1*(3), 14-15.

Ibarra-Rivera, L., Mirabal, S., Regueiro, M. M., & Herrera, R. J. (2008). Delineating genetic relationships among the Maya. *American Journal of Physical Anthropology, 135*, 329-347.

Kauffman, T. (2001). The History of the Nawa language group from the earliest times to the sixteenth century: Some initial results. Theory and Methods in Linguistic Reconstruction. University of Pittsburgh, Pittsburgh. March 1998. Workshop.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika, 30,* 81-93.

Kepecs, S., & Alexander, R. T. (2005). *The postclassic to Spanish-era transition in Mesoamerica.* Albuquerque, NM: University of New Mexico Press.

Lao, O., Lu, T. T., Nothnagel, M., Junge, O., Freitag-Wolf, S., Caliebe, A., . . . Kayser, M. (2008). Correlation between genetic and geographic structure in Europe. *Current Biology, 18*, 1241-1248.

Lee, H. Y., Yoo, J. E., Park, M. J., Chung, U., Kim, C. Y., & Shin, K. J. (2006). East Asian mtDNA haplogroup determination in Koreans: Haplogroup-level coding region SNP analysis and subhaplogroup-level control region sequence analysis. *Electrophoresis*, *27*, 4408-4418.

Lodish, H., Berk, A., Zipursky, S., Matsudaira, P., Baltimore, D., & Darnell, J. (2000). *Molecular cell biology* (4th ed.). New York, NY: W. H. Freeman.

Martinez-Cortes, G., Nuno-Arana, I., Rubi-Castellanos, R., Vilchis-Dorantes, G., Luna-Vazquez, A., Coral-Vazquez, R. M., . . . Rangel-Villalobos, H. (2010). Origin and genetic differentiation of three Native Mexican groups (Purepechas, Triquis and Mayas): Contribution of CODIS-STRs to the history of human populations of Mesoamerica. *Annals of Human Biology, 37*, 801-819.

Maruyama, S., Minaguchi, K., & Saitou, N. (2003). Sequence polymorphisms of the mitochondrial DNA control region and phylogenetic analysis of mtDNA lineages in the Japanese population. *International Journal of Legal Medicine*, *117*, 218-225.

Mendel, G. (1866). Versuche über pflanzenhybriden. *Verhandlungen des Naturforschenden Vereines in Brunn, 4*(3), 44.

Mexican National Commission (1904). Historical data concerning the mining district of Taxco de Alarcon, State of Guerrero. Universal Exposition. St. Louis, Missouri. 1904. Exposition.

Mizuno, F., Gojobori, J., Wang, L., Onishi, K., Sugiyama, S., Granados, J., . . . Ueda, S. (2014). Complete mitogenome analysis of indigenous populations in Mexico: Its relevance for the origin of Mesoamericans. *Journal of Human Genetics, 59*, 359-367.

Moreno-Estrada, A., Gignoux, C. R., Fernandez-Lopez, J. C., Zakharia, F., Sikora, M., Contreras, A. V., . . . Bustamante, C. D. (2014). The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. *Science, 344*, 1280-1285.

Nei, M. (1977). F-statistics and analysis of gene diversity in subdivided populations. *Annals of human genetics*, *41*, 225-233.

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

Peng, M. S., He, J. D., Liu, H. X., & Zhang, Y. P. (2011). Tracing the legacy of the early Hainan Islanders-a perspective from mitochondrial DNA. *BMC Evolutionary Biology*, *11*, 46.

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

Pimenoff, V. N., Comas, D., Palo, J. U., Vershubsky, G., Kozlov, A., & Sajantila, A. (2008). Northwest Siberian Khanty and Mansi in the junction of West and East Eurasian gene pools as revealed by uniparental markers. *European Journal of Human Genetics*, *16*, 1254-1264.

Promega (2011). PowerStats (Version 1.2) [Software]. Available from http://www.promega.com/geneticidtools/powerstats

Reich, D., Thangaraj, K., Petterson, N., Price, A. L., & Singh, I. (2009). Reconstructing Indian population history. *Nature, 461*, 489-494.

Rohlf, F. J. (2005). NTSYSpc [computer software]. Setauket: Exeter.

Ryman, R., & Leimar, O. (2009). G_{ST} is still a useful measure of genetic differentiation – A comment on Jost's D. *Molecular Ecology, 18,* 2084-2087.

Salazar-Flores, J., Dondiego-Aldape, R., Rubi-Castellanos, R., Anaya-Palafox, M., Nuño-Arana, I., Canseco-Avila, L. M., . . . Rangel-Villalobos, H. (2010). Population structure and paternal admixture landscape on present-day Mexican-Mestizos revealed by Y-STR haplotypes. *American Journal of Human Biology, 22*, 401-409.

Sandoval, K., Buentello-Malo, L., Penaloza-Espinosa, R., Avelino, H., Salas, A., Calafell, F., . . . Comas, D. (2009). Linguistic and maternal genetic diversity are not correlated in Native Mexicans. *Human Genetics 126*, 521-531.

Saville, M. H. (1896). The temple of Tepotzlan, Mexico. *Bulletin of the American Museum of Natural History, 8*, 221-226.

Secretariat of the Interior, Mexico. (2000). [Interactive map of indigenous languages of Mexico June 6, 2011]. *National Commission for the Development of Indigenous Peoples; National Institute of Indigenous Languages*. Retrieved from http://www.conapo.gob.mx/

Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Röhl, A., . . . Richards, M. B. (2009). Correcting for purifying selection: an improved human mitochondrial molecular clock. *The American Journal of Human Genetics*, *84*, 740-759.

Smith, M. E. (1986). The role of social stratification in the Aztec empire: A view from the provinces. *American Anthropologist, 88*, 70-91.

Suarez, J. A. (1983). *The Mesoamerican indian languages.* Cambridge, UK: Cambridge University Press.

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

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

Tanaka, M., Cabrera, V. M., González, A. M., Larruga, J. M., Takeyasu, T., Fuku, N., . . . Shinoda, K. I. (2004). Mitochondrial genome variation in eastern Asia and the peopling of Japan. *Genome Research*, *14*, 1832-1850.

UNESCO. (2015). "Archaeological Monuments Zone of Xoxhicalco". *United Nations Educational, Scientific and Cultural Organization World Heritage Centre*. UNESCO. Accessed October 12, 2015 < http://whc.unesco.org/en/list/939>

Van Oven, M., & Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human Mutation*, *30*, E386-E394.

Vargas-Alarcon, G., Moscoso, J., Martinez-Laso, J., Rodriguez-Perez, J. M., Flores-Dominguez, C., Serrano-Vela, J. I., . . . Arnaiz-Villena, A. (2007). Origin of Mexican Nahuas (Aztecs) according to HLA genes and their relationships with worldwide populations. *Molecular Immunology 44*, 747-755.

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*, *8*, 1084-1100.

Weinberg, W. (1908). Über vererbungsgesetze beim menschen. *Molecular and General Genetics MGG*, *1*, 440-460.

Willey, G. R. (Ed.). (1965). *Handbook of middle American Indians: Archaeology of southern Mesoamerica* (Vols. 2-3). Austin, TX: University of Texas.

Zheng, H. X., Yan, S., Qin, Z. D., Wang, Y., Tan, J. Z., Li, H., & Jin, L. (2011). Major population expansion of East Asians began before neolithic time: Evidence of mtDNA genomes. *PLoS One*, *6*(10).

Works Consulted

Applied Biosystems. (2001). AmpFlSTR identifiler PCR amplification kit user's manual. Foster City, CA: Applied Biosystems.

Balding, D. J., Bishop, M., & Cannings, C. (Eds.). (2007). *Handbook of statistical genetics* (3rd ed.). Hoboken, NJ: John Wiley & Sons.

Barrot, C., Sanchez, C., Ortega, M., Gonzalez-Martin, A., Brand-Casadevall, C., Gorostiza, A., . . . Gene, M. (2005). Characterization of three Amerindian populations from Hidalgo State (Mexico) by 15 STR-PCR polymorphisms. *International Journal of Legal Medicine, 119,* 111–115.

Bhaskar, A., & Song, Y. S. (2009). Multi-locus match probability in a finite population: A fundamental difference between the Moran and Wright–Fisher models. *Bioinformatics, 25*, i187-i195.

Bosch, E., Calafell, F., Perez-Lezaun, A., Clarimon, J., Comas, D., Mateu, E., . . . Bertranpetit, J. (2000). Genetic structure of north-west Africa revealed by STR analysis. *European Journal of Human Genetics, 8*, 360-366.

Budowle, B., Chidambaram, A., Strickland, L., Beheim, C. W., Taft, G. M., & Chakraborty, R. (2002). Population studies on three native Alaska population groups using STR loci. *Forensic Science International, 129*, 51–57.

Butler, J. M. (2005). *Forensic DNA typing: Biology, technology, and genetics of STR markers.* Burlington, MA: Elseiver Academic Press.

Calderon-Garciduenas, A. L., Rivera-Prieto, R. A., Ortiz-Lopez, R., Rivas, F., Barrera-Saldana, H. A., Penaloza-Espinosa, R. I., . . . Cerda-Flores, R. M. (2008). Genetic structure of Mexican Mestizo women with breast cancer based on three STR loci. *American Journal of Human Biology, 20*, 191-193.

De Knijff, P. (2007). Hidden secrets of your ancestors. *European Journal of Human Genetics, 15*, 509-510.

Gonzalez-Andrade, F., Sanchez, Q. D., & Martinez-Jarreta, B. (2006). Genetic analysis of the Amerindian Kichwas and Afromerican descendents populations from Ecuador characterized by 15 STR-PCR polymorphisms. *Forensic Science International, 160,* 231–235.

Gorostiza, A., Gonzalez-Martın, A., Lopez Ramırez, C., Sanchez, C., Barrot, C., Ortega, M., . . . Gene, M. (2007). Allele frequencies of the 15 AmpFlSTR Identifiler loci in the population of Metztitlan (Estado de Hidalgo), Mexico. *Forensic Science International, 166,* 230–232.

Hashiyada, M., Itakura, Y., Nagashima, T., Nata, M., & Funayama, M. (2003). Polymorphism of 17 STRs by multiplex analysis in Japanese population. *Forensic Science International, 133*, 250–253.

Hernandez-Gutierrez, S., Hernandez-Franco, P., Martinez-Tripp, S., Ramos-Kuri, M., & Rangel-Villalobos, H. (2005). STR data for 15 loci in a population sample from the central region of Mexico. *Forensic Science International, 151*, 97–100.

Hu, S. P., Yu, X. J., Liu, J. W., & Cai, K. L. (2005). Analysis of STR polymorphisms in the Chao Shan population in South China. *Forensic Science International, 147*, 93–95.

Jobling, M. A., & Tyler-Smith, C. (2003). The human Y chromosome: An evolutionary marker comes of age. *Nature Reviews Genetics, 4*, 598-612.

Kim, Y. L., Hwang, J. Y., Kim, Y. J., Lee, S., Chung, N. G., Goh, H. G., . . . Kim, D. W. (2003). Allele frequencies of 15 STR loci using AmpFlSTR Identifiler kit in a Korean population. *Forensic Science International, 136,* 92–95.

Lovo-Gomez, J., Salas, A., & Carracedo, A. (2007). Microsatellite autosomal genotype data in four indigenous populations from El Salvador. *Forensic Science International, 170*, 86–91.

Mesa, M. S., Arroyo-Pardo, E., Martinez-Labarga, C., Ocana, M. A., Arroyo, M., & Bandres, F. (2001). Variability of six STR loci in populations from central Spain. *Human Biology, 73*, 745-753.

Paredes, M., Galindo, A., Bernal, M., Avila, S., Andrade, D., Vergara, C., . . . Carracedo, A. (2003). Analysis of the CODIS autosomal STR loci in four main Colombian regions. *Forensic Science International, 137,* 67–73.

Perez-Miranda, A. M., Alfonso-Sanchez, M. A., Kalantar, A., Garcıa-Obregon, S., de Pancorbo, M. M., Pena, J. A., . . . Herrera, R. J. (2005). Microsatellite data support subpopulation structuring among Basques. *Journal of Human Genetics, 50*, 403–414.

Wang, Z. Y., Yu, R. J., Wang, F., Li, X. S., & Jin, T. B. (2005). Genetic polymorphisms of 15 STR loci in Han population from Shaanxi (NW China). *Forensic Science International, 147*, 89–91.