CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS DEL INSTITUTO POLITÉCNICO

UNIDAD IRAPUATO

PROGRAMA EN BIOLOGÍA INTEGRATIVA

Reconstrucción del mestizaje y dinámicas de migración del México Poscolombino

Tesis que presenta

Juan Esteban Rodríguez Rodríguez

Para obtener el grado de

Maestro en Ciencias

en

Biología Integrativa

Director de tesis:

Dr. Andrés Moreno Estrada

Irapuato, Guanajuato, México Agosto 2019

Center for Research and Advanced
Studies of the National Polytechnic Institute

Irapuato Unit

RECONSTRUCTING ADMIXTURE AND MIGRATION DYNAMICS OF POST-COLUMBIAN MEXICO

A thesis by

Juan Esteban Rodriguez-Rodriguez

As requirement for the degree of

Master of Science

in

Integrative Biology

Thesis advisor:

Dr. Andres Moreno-Estrada

Irapuato, Guanajuato, Mexico August 2019

ABSTRACT

Mexico has a considerable population substructure due to historical events and distinct amounts of admixture between ethnic groups, primarily Native Americans, Europeans, Sub-Saharan Africans and, to a lesser extent, East Asians. Genetic substructure in Mexico has been attested previously at a continental degree. However, deeper analyses to explore sub-continental structure remain limited and post-Columbian demographic dynamics within Mexico have not been inferred with genomic data. The availability of genome-wide SNP array data from worldwide and admixed Mexican populations, offers the possibility to characterize the differences in the admixture process across different Mexican states. In this thesis, we explored admixture and demographic differences within Mexico in greater detail. We analyzed the ancestry tract length distribution to infer the timing of admixture in each region, as well as the number of migratory pulses. We observed older admixture timings in the earliest colonial cities and more recent estimates in southern Mexico in agreement with historical records. We characterized the specific origin of the Native American ancestry in Mexico: a widespread Western Native Mexican in Aridoamerican states and a Central Nahua extension to Southern Mexico in Guerrero and Eastern Mexico in Veracruz. Yucatan shows lowland Mayan ancestry, while Sonora exhibited a unique and unattested Northwestern Mexican ancestry. Demographic shifts across time also left a genetic footprint in modern samples. Sonora portrays a limited gene flow with the rest of Mexico. Consequently, a bottleneck is recapitulated in the local European component. While regarding the Native American component, a bottleneck possibly related to the demographic collapse is observed in Aridoamerican cosmopolitan Mexicans. Finally, a significant proportion of East Asian ancestry was observed in samples from Acapulco, Guerrero, which remains an understudied heritage. We pinpointed its origin to Southeast Asia, displaying Indonesian and non-Negrito Filipino affinities. This reveals a surprising genetic remnant from the Manila Galleon slave trade with the Philippines. This unprecedented repot of genetic origins uncovers ethnic identities lost in historical records.

RESUMEN

México muestra una subestructura poblacional humana debido a eventos históricos y diferentes grados de mestizaje entre grupos étnicos, principalmente indígenas americanos, europeos, africanos subsaharianos y, en menor medida, asiáticos del este. La subestructura genética mexicana ha sido estudiada a niveles continentales en el pasado, sin embargo, los estudios genéticos que exploran esta estructura a escalas subcontinentales son escasos, al igual que aquellos que estudian las dinámicas demográficas poscolombinas entre regiones de México. Gracias a la disponibilidad de datos genómicos de microarreglos de poblaciones alrededor del mundo y de mestizos mexicanos, es posible caracterizar las diferencias del mestizaje entre estados de México. Para ello, analizamos los histogramas de los fragmentos de ancestría en genomas de mestizos para inferir tiempos de mestizaje y el número de pulsos migratorios. De modo que encontramos tiempos de mestizaje más antiguos en las primeras ciudades fundadas en la colonia, comparado con estimados más recientes en el sur de México, indicando una concordancia con registros históricos. También determinamos un origen más específico de la ancestría indígena en mestizos: una ancestría indígena del oeste presente en estados aridoamericanos y un componente nahua central extendido por estados como Veracruz y Guerrero. En Yucatán observamos una ancestría maya peninsular, mientras que Sonora muestra un componente del noroeste con un origen étnico desconocido. Los cambios demográficos a través del tiempo también han dejado una marca genética en la actualidad. Por ejemplo, Sonora posee un aislamiento genético con el resto de México y a su vez recapitula un cuello de botella en su componente europeo. Un cuello de botella es observado en el componente indígena de mestizos de Aridoamérica posiblemente relacionado al colapso demográfico. Por último, las muestras de Acapulco, Guerrero presentan una proporción considerable de ancestría asiática, específicamente de origen indonesio y filipino austranesio. Estos hallazgos en poblaciones modernas evidencian la existencia de una huella derivada del tráfico de esclavos de las Filipinas mediado por el Galeón de Manila.

ACKNOWLEDGEMENTS

I am grateful for having such a supportive and loving family. They have always supported my life decisions, even when that meant living far from home in the arid Northern lands of Mexico. They are part of this achievement as it would not have been possible without them.

I would like to thank the Moreno Lab with all the members that have been part of it: Javier, Erika, Pavel, Consuelo, Majo, Lalo, Andrés K., Carmina, María, Mitzi, Sergio, Adriana, Mashaal and Karla. Thanks to Maria Avila's lab in Juriquilla as well. Their suggestions polished my methodology and interpretations, also they helped me improve the way I communicate science.

I thank my master's committee for their observations and constructive questionings: Selene Valverde, Rafael Montiel and Maria Avila. I thank Andrés Moreno for being an incredible mentor, he offered me many opportunities such as congress participations and collaborations with many other projects which helped my academic career. He always guided me through the several stages of the degree.

I deeply thank Alex Ioannidis for his unconditional support. He helped me understand human population genetics, the methods of this field and the critical thought needed to discuss results and propose subsequent analyses. He also developed very useful methods for the lab. Basically, I would have not been able to finish my project without his advice. I appreciate all his personal and career advices, for being an informal mentor but also a friend with whom you can have endless conversations and keep learning new things every day.

I also thank Conacyt for the financial support during my master studies, as well as the funding from the International Centre for Genetic Engineering and Biotechnology (ICGEB) with the training grant "Human Functional Genomics of Native American Variation in Mexico and Latin America".

AGRADECIMIENTOS

Le agradezco mucho a mi familia por apoyarme en todas mis decisiones, incluso cuando tuve que irme de Mexicali para hacer lo que quería. Ellos son parte de mis logros.

Gracias al laboratorio de María Ávila en Juriquilla y al Moreno Lab con todos los que hemos sido parte de este grupo de investigación: Javier, Érika, Pavel, Consuelo, Majo, Lalo, Andrés K., Carmina, María, Mitzi, Sergio, Adriana, Mashaal y Karla. Sus sugerencias mejoraron los métodos e interpretaciones de mi proyecto, así como me enseñaron a comunicar adecuadamente mi proyecto.

Agradezco a mi comité de la maestría por sus observaciones y preguntas: Selene Valverde, Rafael Montiel y María Ávila. Gracias a Andrés Moreno por ser un mentor increíble, me ofreció muchas oportunidades de congresos y colaboraciones en los muchos proyectos que está involucrado, ayudándome en mi formación académica. Siempre me guio en las diversas etapas de este posgrado.

Le quiero agradecer enormemente a Alex Ioannidis por su apoyo incondicional. Gracias a él entiendo mucho de la genética de poblaciones humanas, los métodos del área y, además poseo un sentido crítico para discutir resultados y proponer los análisis posteriores adecuados. También nos ha ayudado con programas muy útiles para el laboratorio. Básicamente, no hubiera completado este proyecto sin su apoyo. Le agradezco por todos los consejos tanto personales como académicos, ya que ha sido más que un mentor no reconocido, sino un amigo de quien se aprende algo nuevo cada día.

Gracias a Conacyt por el apoyo económico durante la maestría, así como le agradezco al International Centre for Genetic Engineering and Biotechnology (ICGEB) por la beca de formación del proyecto “Human Functional Genomics of Native American Variation in Mexico and Latin America”.

CONTENTS

List of Tables

List of Figures

1 INTRODUCTION

1.1 The relevance of genetic studies in human origins

Humans have always attempted to explain their origins. Many scientific disciplines have sought the answer. Even though archeology has provided the most comprehensive information, other areas have also contributed to answer this question. In the case of genetics, the categorization of human populations started by identifying similarities and differences with classical markers. The study of heritable characteristics, such as blood type frequencies, allowed the proposal of some classifications corresponding to now obsolete racial terms (Boyd, 1963).

Later, the genetic role of DNA was discovered (Watson & Crick, 1953), and together with the invention of the polymerase chain reaction (PCR) allowed the targeted amplification of specific sequences (Mullis, 1990). In such way, a gene of interest representing a minuscule portion of the genome, could be transformed into millions of copies. The study of uniparental markers and repeated sequences such as microsatellites or STRs provided some initial insight about affinities between ethnic groups (Mesa et al., 2000; Prugnolle, Manica, & Balloux, 2005). Nevertheless, these represented a limited view of human genetic variation as they relied on very few markers, requiring subsequent sequencing and genotyping efforts to fully grasp the potential of genomics.

The first sequencing techniques exploited fluorescence to identify genotypes, leading to the sequencing of the first human genome (Consortium International Human Genome Sequencing, 2001; Venter et al., 2001). The process demanded the participation of several laboratories and huge expenditures. Eventually, the invention of next-generation sequencing platforms made sequencing more efficient resulting in the considerable decrease of the cost (Levy & Myers, 2016). High-throughput data from these technologies allowed a more precise characterization of variants across the genome.

A cheaper genotyping technology based on microarrays emerged shortly after. As variants were well-studied and human reference genomes became available, hybridization techniques allowed the evaluation of a big volume of genetic markers at lower costs. Nowadays, microarrays with hundreds of thousands of markers are still widely used because of their low cost in studies with large sample sizes. However, they pose some limitations as they only evaluate previously characterized single nucleotide polymorphisms (SNP) common in certain populations, usually those used as discovery panels, such as populations of European and East Asian descent (Consortium, 2010). Because of this, most microarrays have an ascertainment bias towards common variants, and an underrepresentation of rare variants from understudied ethnic groups. This bias has a greater impact in analyses based on the frequency spectrum of variants (SFS), which can be used to infer selective and demographic events, ideally from sequencing data.

Despite this limitation, microarrays still pose a useful and affordable alternative to study the genetic structure of human populations with great resolution, as large collections of samples can be genotyped with millions of SNPs with the newest array models. Millions of markers allow deeper analyses and the separation of closely related human groups (Novembre & Peter, 2016). Even though the small differentiation in humans makes most alleles not to be population-exclusive, bioinformatics tools are able to accurately dissect subtle differences between human groups as they rely on many markers and haplotypes simultaneously (Edwards, 2003). High-throughput data coupled with better algorithms are enabling researchers to answer precise evolutionary questions about adaptation, divergences and mixing. Ultimately, characterizing our past with novel bioinformatics tools, not only contributes to answer long-lasting questions about our origins, but it also has health implications in improving our quality of life through biomedical applications.

1.2 Genetic reconstruction of human migrations

Most peopling events are thousands of years older than the historical period of human evolution. They consisted of population splits, cultural expansions and admixture events that left no written record, other than the genetic profile of ancient and modern populations. Therefore, the early demographic events in human evolution have no historical support, and are rather inferred from morphological and, more recently, genetic analyses since the advent of molecular approaches as described above.

Different disciplines have supported an African origin for modern humans, especially archeology, as the oldest modern human remains have been found in Africa (McDougall, Brown, & Fleagle, 2005; White et al., 2003). The advances of genetic studies have supported this claim, as the deepest uniparental lineages root in Africa (Vigilant, Stoneking, Harpending, Hawkes, & Wilson, 1991), and as human populations outside of Africa portray a reduced genetic diversity and higher linkage disequilibrium across the genome (DeGiorgio, Jakobsson, & Rosenberg, 2009). These phenomena are consistent with a founder effect from an African population that gave origin to all groups out of Africa, i.e. Oceania, Asia, Europe and the Americas. As the world was settled, populations split for periods of thousands of years (Nielsen et al., 2017), leading to long-standing geographic isolation and the accumulation of genetic differences. This resulted in the diversity observed today, with ethnic groups exhibiting unique patterns of variation, according to their histories and inhabited environments.

Giving a detailed description of the peopling of every region of the world is obviously outside the scope of this thesis. Therefore, the following sections will be limited to those regions and groups that became primary contributors to the mixture in present-day Mexican populations.

1.2.1 Origins of Europeans

The migration that gave origin to all Eurasian populations left Africa 60,000 to 50,000 years ago (Mellars, 2006; Rito et al., 2019). This wave was the ancestor of

East Asians and Upper Paleolithic (UP) populations distributed in West Eurasia. The earliest UP industries date 48,000 years ago in the Levant, while East Asians are estimated to have split from the rest of Eurasian populations at least 36,200 years ago (Seguin-Orlando et al., 2014). UP populations exhibited an early genetic substructure that gave origin to Basal Eurasians in the Middle East, West Eurasians in Europe and Ancient North Eurasians (ANE). Modern Europeans have three main ancestral components that derive from this Eurasian substructure (see Figure 1.1).

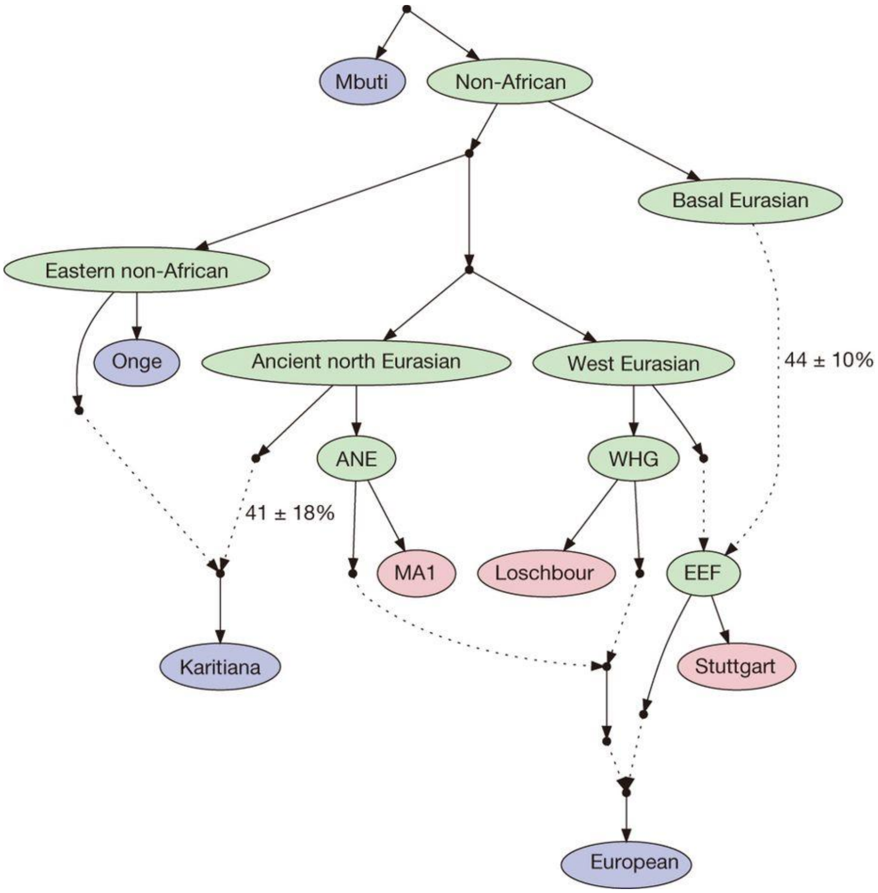


Figure 1.1: The origin of modern Europeans is traced to UP contributions from West Eurasians, Basal Eurasians (which contributed to Early European Farmers) and Ancient North Eurasians (which contributed to the Yamnaya ancestry). They descend from the migration out of Africa (referred as Non-African in the figure) previously mentioned. These sub-structured UP groups originated the three main ancestries in Europe: Western Hunter-Gatherer, Early European Farmer and the Yamnaya. Figure from (Lazaridis et al., 2014).

- The three ancestries in Europe

The study of ancient samples has allowed the characterization of more detailed origins for all modern populations, such as the timing of migrations, population replacements and recently known contributions from extinct populations. On Europe, three ancestral components associated to cultural shifts have been identified: Western Hunter-Gatherer (WHG), West Eurasians present in Europe before the spread of agriculture; Early European Farmers (EEF) related to the Neolithic expansion from the Near East, and the posterior Yamnaya culture associated to steppe pastoralists (Lazaridis et al., 2014). The genetic uniqueness of these three ancestral groups is partially explained by the Basal Eurasian proportion in the EEF and an ANE fraction in the Yamnaya, as represented in Figure 1.1 with EEF and ANE contributions.

In the Early Neolithic around 8,000-7,000 years ago, Near Eastern farmers replaced the local hunter-gatherers, followed by a slight reemerge of WHG ancestry (Brandt, Szécsényi-Nagy, Roth, Alt, & Haak, 2015). The arrival of the Yamnaya pastoralists occurred in the Late Neolithic ~4,500 years ago contributing genetic proportions as high as ~75% in cultures such as the Corded Ware in Germany (see lower section from Figure 1.2) (Haak et al., 2015).

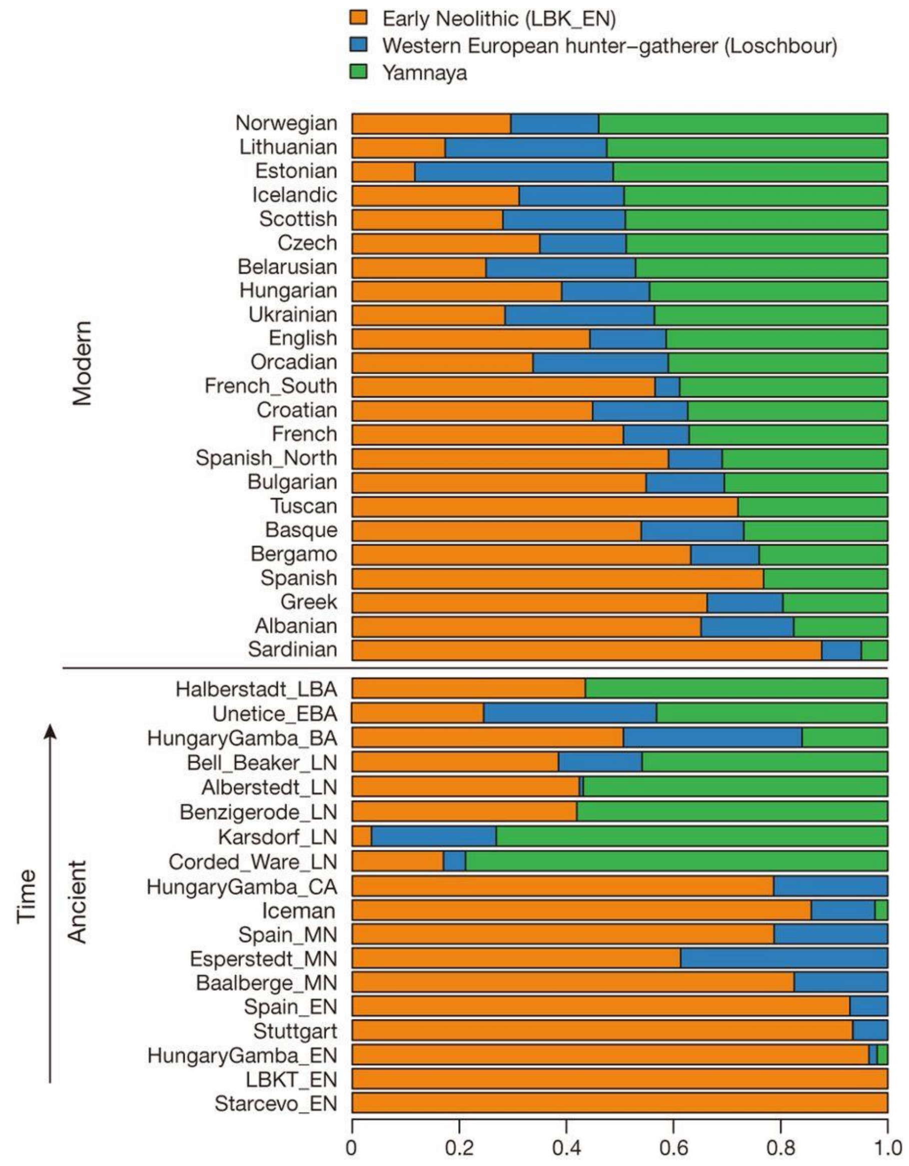


Figure 1.2: Proportions from the three main ancestral populations in Europe. Ancient and modern European populations are shown. The three ancestral populations are: Western European hunter-gatherer (WHG), Early Neolithic (EEF) and the Yamnaya. In the upper section, modern European substructure is illustrated, while in the lower section, genetic shifts across time are shown. Figure from (Haak et al., 2015).

- Present-day European substructure

Subsequent events such as invasions, civilizations and cultural expansions could have generated gene flow within Europe, leading to a generalized persistence of all three components across modern groups on the continent (see upper section from Figure 1.2), in addition to a very small differentiation between Europeans

(Rosenberg et al., 2002). However, the incomplete random mating across Europe produced geographic gradients of the three ancestral components resulting in a heterogeneous genetic profile in modern European populations. Even though this substructure is small, populations between and within countries can still be told apart if enough genetic markers are included. For example, the Sardinian's profile stands out from the rest of Europe as they represent an ancient remnant from the EEF migration that remained mainly isolated ever since (Chiang et al., 2018), showing the highest proportions of EEF in present-day populations. In the same manner, Northern Europeans show a higher Yamnaya ancestry, which coincides with the Corded Ware's historical occupation.

Some intercontinental admixture has also contributed to Europe's genetic background. Mediterranean Europeans have additional intercontinental contributions of more recent origin. A small sub-Saharan African (Moorjani et al., 2011) and North African contributions are present in the Iberian Peninsula, and a Middle Eastern fraction in Italy (Botigué et al., 2013). In the case of Iberia, it was probably related to gene flow during the 8th century CE Moorish Berber conquest of the peninsula (Botigué et al., 2013). The complex origin of Europeans results in specific population footprints that can be differentiated with enough genetic information.

1.2.2 Origins of Native Americans

Initial studies with classical markers and mitochondrial DNA reported a resemblance between Native Americans and East Asians, specifically Siberians (Matson et al., 1968; Wallace, Garrison, & Knowler, 1985). However, ancient samples have elucidated a more complex history for the peopling of the Americas. Native Americans originated from an East Asian population split 36,000 years ago, showing continuous gene flow with them until 25,000 years ago (Moreno-Mayar, Potter, et al., 2018). A successive admixture event resulted in the distinct variation of Native Americans compared to other East Asians (Raghavan, Skoglund, et al., 2014). The sequencing of a 24,000-year-old individual from Mal'ta revealed the divergent ANE component that contributed 25% (Moreno-Mayar, Vinner, et al.,

2018) to all modern Native Americans (Raghavan, Skoglund, et al., 2014). A similar ANE contribution took place in all Native Americans 25,000-20,000 years ago (Raghavan, Skoglund, et al., 2014), suggesting a Beringian standstill model that led to the differentiation of the ancestral population of all Native Americans, referred as ancestral Native Americans (see Figure 1.3). A reduced effective size of 250-2000 individuals has been proposed for this founder population (Fagundes et al., 2018; Gravel et al., 2013).

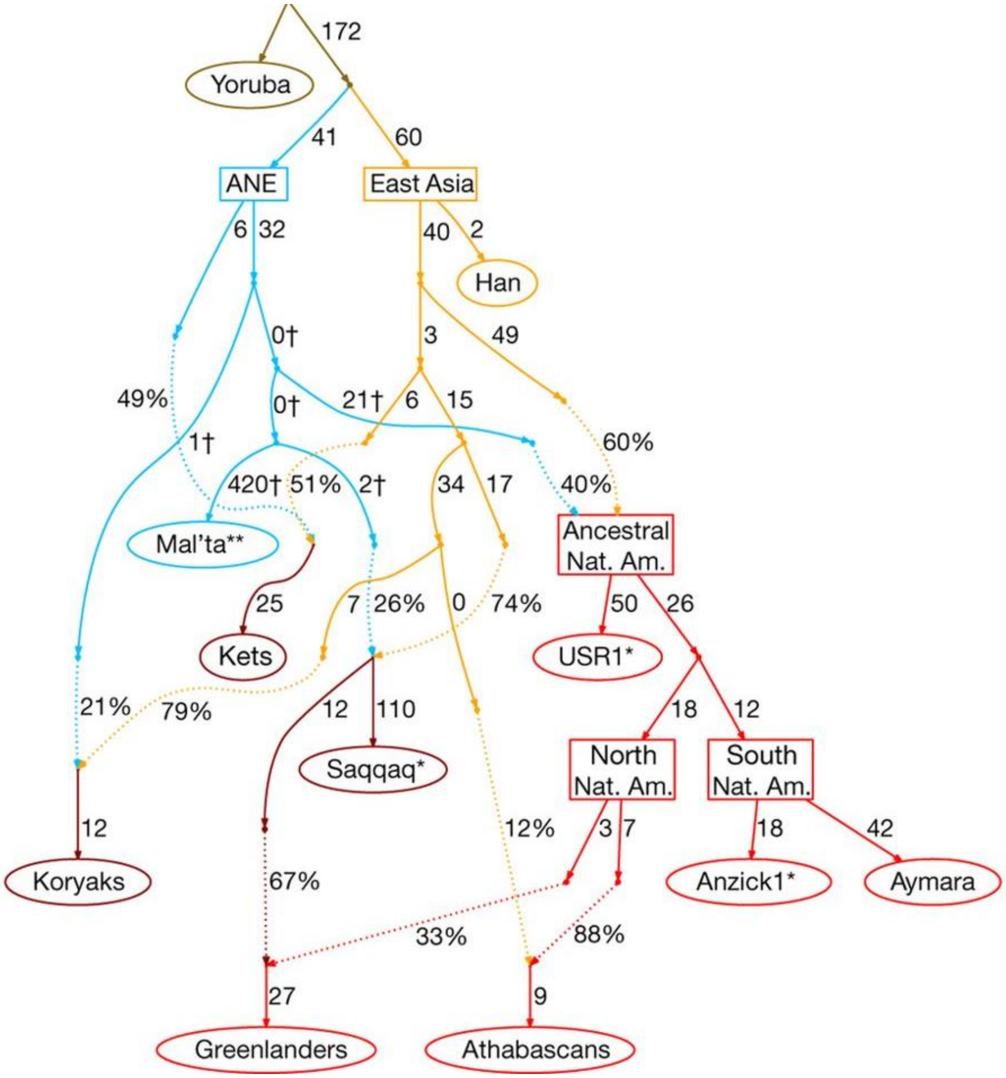


Figure 1.3: Tree diagram shows the dual origin of all Native Americans and the subsequent divergence into two main branches. Native Americans (in red) are depicted as a mixture of ANE (in blue) and East Asia (in yellow). Modern and ancient Native American populations originate from the basal divergence into North Native Americans and South Native Americans. Figure from (Moreno-Mayar, Potter, et al., 2018).

- Initial settlement of the Americas

Ancestral Native Americans diverged 17,500–14,600 years ago into two main basal clades (Moreno-Mayar, Potter, et al., 2018) from which all extant native groups descend. The first clade, the North Native Americans (NNA) contributed to Canadian and USA Natives e.g. Algonquin speakers (Reich et al., 2012), Athabascans and Inuit (Moreno-Mayar, Potter, et al., 2018). The South Native American (SNA) clade had a rapid ~2,000-year expansion over thousands of kilometers (Moreno-Mayar, Potter, et al., 2018), giving origin to numerous natives on Mesoamerica, Central America and South America. Dispersals had a characteristic serial founder expansion, associated to a diversity reduction similar to the out-of-Africa migration (Wang et al., 2007).

- Post-peopling dispersals

After this initial settlement, later migrations shaped the diversity in pre-European contact populations. In North America, some NNA groups admixed with posterior incoming migrations from Siberia, as it is the case of Athabascans and Arctic Natives (Raghavan, DeGiorgio, et al., 2014), while others are a mixture of NNA and SNA groups (Moreno-Mayar, Vinner, et al., 2018). The domestication of crops in Mesoamerica represented an important cultural shift in the Americas at 7,000 BC, according to archeology (Coe & Koontz, 2013), linguistics (Brown, Clement, Epps, Luedeling, & Wichmann, 2014; Diamond & Bellwood, 2003) and genetics (Moreno-Mayar, Vinner, et al., 2018). Recent studies on ancient DNA have proposed a widespread Mesoamerican admixture distributed across the USA Great Basin, Mexico, Central America and South America, by using the modern Mixe people from Southern Mexico as a genetic source. Possibly related to the spread of agriculture, this migration occurred posterior to 8.7 ka, reaching Patagonia until 5.1 ka and the Great Basin at 2 ka. Genetic substructure in Northwestern Mexico, i.e. Pima and Yaqui, have been explained as a possible NNA population with a contribution of this Mesoamerican component (Moreno-Mayar, Vinner, et al., 2018).

Later events have been mainly addressed without genetic information due to the limited ancient DNA studies in the Americas. Instead, cultural expansions have been characterized with archeological complexes, not necessarily involving population replacement, admixture or an actual movement of people. The following relevant cultural innovation after agriculture represented the origin of all Mesoamerican civilizations with their mother civilization: the Olmecs, in 1500-400 BCE (Diehl, 2004). Posterior civilizations, e.g. the extensive Teotihuacan and Classic Maya, may have generated gene flow within and outside Mesoamerica, creating the present-day genetic substructure in Native Mexicans.

Migrations, divergences and bottlenecks in Native American groups have left heterogeneous genetic footprints. For instance, some Native American groups in Mexico are highly isolated, generating a differentiation as large as continental differences measured by Fst. The Seri from the Sonoran Desert and the isolated Mayan Lacandon portray an Fst of 0.136 when compared pairwise, in contrast to a 0.11 Fst from the differentiation between European and Chinese populations (Moreno-Estrada et al., 2014). Most Native American populations in Mexico can be told apart by their genetic profile if enough markers are provided, the same way as other continental populations. Nevertheless, the absence of comprehensive reference panels for underrepresented populations have limited the characterization of the huge diversity in the Americas.

1.3 The Americas in 1492 CE

1.3.1 Native Mexican cultures at European contact

In the Mexican territory, the majority of pre-Hispanic people were congregated in Mesoamerica. When the Spanish disembarked to the Americas, most of Central Mexico was under the rule of the Aztecs or Mexicas administered by a confederation of city-states (Wauchope, Ekholm, & Bernal, 1971), as shown in Figure 1.4. The predominant Native language in Mexican territory since the Spanish arrival to present-day has been the Nahuatl language, whose speakers are denominated Nahua people. Many other ethnic groups have existed in the Mexican territory since then, some under the Aztec rule (whose lingua franca was Nahuatl) and others resisting subjugation, both Nahua and non-Nahua people

(Wauchope et al., 1971). Central Mexican Natives such as the non-Nahua Tarascan Empire to the West and the Nahua Tlaxcaltec people remained independent, the latter are known for siding with the Spaniards against the Aztecs (Wauchope et al., 1971). The Southern Mexican civilization of Tututepec lead by the Zapotec, Mixtec, among others, remained independent (Spores, 1993), as well as the several Mayan provinces in the Southeast, comprising present-day Chiapas, the Yucatan Peninsula and surrounding areas, such as Central America (Steckel & Rose, 2002). Furthermore, not-as-numerous Native peoples also existed north from Mesoamerica, in the region of present-day Northern Mexico. The ethnically heterogeneous groups of nomad and semi-nomad lifestyles were considered by the Aztecs as unconquerable and barbaric, reflected by the general term "Chichimec" (Berdan, 2014).

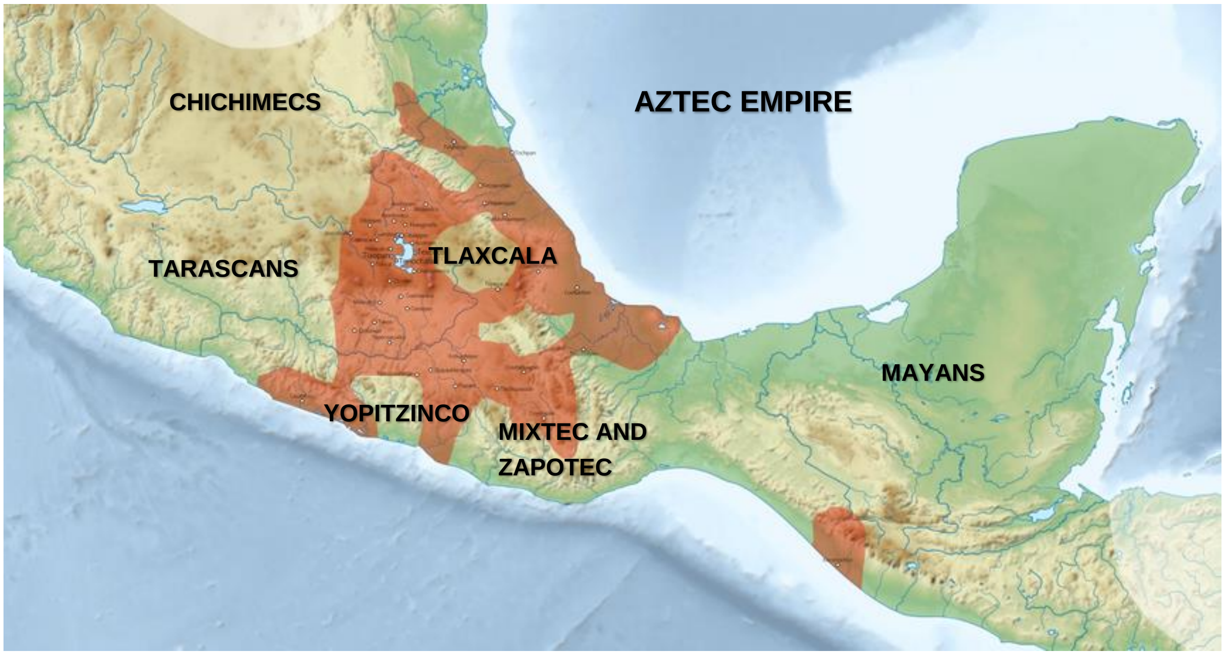


Figure 1.4: Map showing Mesoamerican natives' organization at European contact. The orange area in Central Mexico represents the extent of the Aztec Empire. Other governments covered in this thesis are shown, such as the western Tarascan Empire (Purepecha Empire), Mixtec and Zapotec independent territories in the South, southeastern Mayan states and the Chichimec peoples to the North.
Image modified from:
https://en.wikipedia.org/wiki/Aztec_Empire#/media/File:Aztec_Empire.png

1.3.2 European and African immigration to the Americas

European immigrants mainly came from the Iberian Peninsula, specifically from provinces in Southern Spain such as Andalusia, Extremadura and Castile (Lagunas Rodríguez, 2004). African slaves forcibly brought mainly belonged to West and West-Central African ethnic groups. Several other minorities made their way to the New World at much smaller scales, from diverse origins such as present-day Portugal, Italy, Belgium, France, among other European countries (Lagunas Rodríguez, 2004), as well as other ethnicities like Converso Jews (Hordes, 2005).

1.4 The Spanish Empire's conquest

The Spanish first conquered the capital of the Aztec empire in 1521 CE: Tenochtitlan, now Mexico City (López De Gómara, 2007). Other Central Mexico empires fell shortly after with help of their Native allies. The Tarascan Empire, led by the Purepecha people, lost its independence in 1530 CE. The Purepecha had an important presence in neighboring states during the colony, in states such as Guanajuato. In the Bajio region, comprising Guanajuato, Eastern Jalisco, Aguascalientes and Southern Zacatecas, the Spanish purchased for peace after having lost the Chichimec War in 1590 CE (Cázares, 2000). In other words, semi-nomads were assimilated once they adopted a new sedentary lifestyle with the presents the Spanish provided (herein Native people exclusively from this region will be referred to as Chichimecs). In the case of Southern Mexico, natives were conquered but conserved their culture with minimal admixture. In Guerrero, one of the regions independent of the Aztec Empire was the Kingdom of Yopitzinco, which was defeated and replaced by the predominant Nahua peoples in the colony. In Oaxaca, ethnic groups such as the Zapotec and Mixe persisted culturally. In the Southeast, territory of the Mayan peoples, the Yucatan Peninsula was conquered in 1543 CE and the current capital of the Yucatan State was founded in an inhabited Mayan city. Mayan presence and culture resistance have been considerable ever since, as reflected by the Caste War in 1847 CE (Adams & Macleod, 2000). On the other hand, vast areas in Northwest Mexico had a very late contact, as well as many failed conquest attempts. In Sonora, the first cities were built in 1700 CE, achieving a stable settlement until 1787 CE due to warfare.

Spanish people also reached overseas, conquering a fraction of the Philippines in 1565 CE, which became part of the Spanish Empire. Trade with Asia was accomplished by trade routes between Manila and Mexican ports. The Spanish Empire extended over vast domains, across continents and across North and South America. It comprised present-day USA, Mexico, Honduras, Colombia, Peru, Chile, Argentina, among others, as well as the Philippines and parts of North and sub-Saharan Africa (excluding the vast territories owned by the allied Portuguese).

1.5 Admixture and the caste system

The arrival of the Spanish to the Americas represented a unique moment in history when people from very different backgrounds interacted with each other for the first time. This interaction resulted in admixed offspring from distant ethnic groups and ultimately originated most present-day Mexicans. Complex admixture is evident since the onset of the colonial period with the establishment of the many castes the Spanish created (Lagunas Rodríguez, 2004). The caste system was the result of dealing with the privileges Europeans imposed and the rights an admixed person should have based on that principle. The complexity of the system is shown by all the sorts of combinations and succeeding admixture between the resulting admixed individuals.

1.5.1 Demographic collapse of the Natives

According to some estimations, it is thought that the Native population at pre-contact times was as high as 22 million. In the 16th century most Native populations collapsed as a result of many pandemics unknown for the local people, such as smallpox, typhus, measles, influenza, bubonic plague, cholera, malaria, cocoliztli (an unidentified pandemic), among other diseases (Lagunas Rodríguez, 2004) (see Figure 1.5). The lack of immunity to the new pathogens caused a decimation across the Americas. Moreover, the high death rates were promoted by forced labor, droughts and the collapse of the Native lifestyle due to the introduction of livestock. The native economic system was affected because of the disruption of the pre-contact agricultural systems (Rionda Ramirez, 2002).

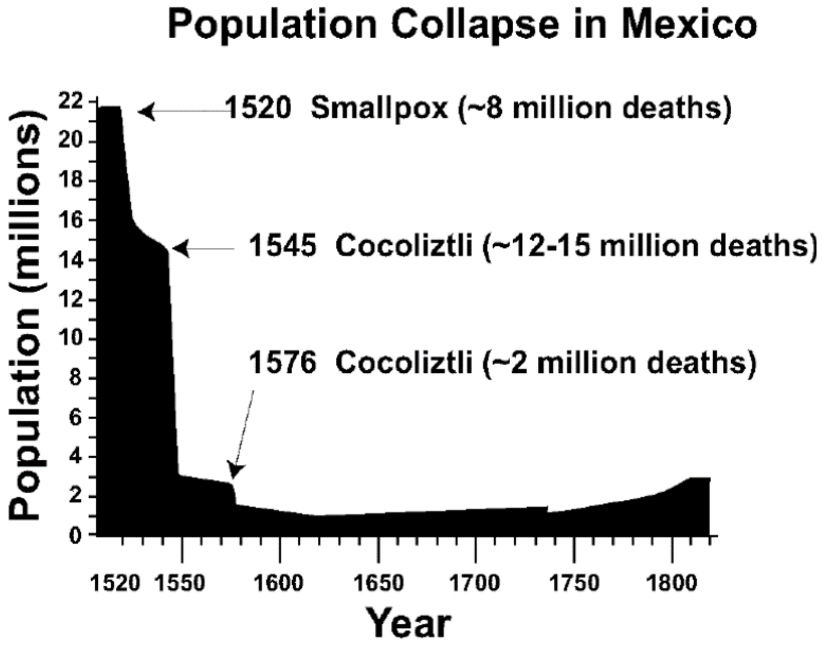


Figure 1.5: Population size fluctuations in Mexican territory since European contact. Major pandemics and the effect on the population size are indicated. Pre-contact population size changes between authors varying from 4.5 million in Mexico to 25 million people only in Central Mexico. Even though authors propose differing numbers, they all agree on the existence of the demographic collapse during this period. Figure from (Acuna-Soto, Stahle, Cleaveland, & Therrell, 2002).

1.5.2 Population shifts during the colony

The demographic disaster had its most critical point at 1646 CE affecting the largest sector at the time: Native Americans. The population at New Spain reached its lowest with ~1,700,000 people. The importation of African slaves was promoted around this period, as a way to compensate the loss of native labor. Then, admixed populations started to increase even though many droughts and a few epidemics continued occurring in the next century (Rionda Ramirez, 2002). Some of the most cosmopolitan cities gave origin to an unprecedented admixture as mining wealth attracted Spanish people and required Native Americans and Africans for labor. Before the collapse, the number of admixed castes in 1570 constituted only the 0.5% of New Spain's population. In the 17th century, the castes growth became considerable, until reaching an important 39.5% in 1810 CE (Lagunas Rodríguez, 2004). Nowadays, most Mexicans self-identify as admixed with the term "mestizo", making Native Americans a minority in Mexico since the end of the 19th century for the first time. Genetic studies show most people in Mexico exhibit certain degree of admixture today (Moreno-Estrada et al., 2014; Ruiz-Linares et al., 2014).

1.6 Understudied ancestries in Mexico: East Asian immigration

1.6.1 The colonial link between Mexico and Asia

Local Native Americans, Spanish Europeans and sub-Saharan Africans had the largest presence in Mexico during the colony. However, many other ethnic groups were residing in colonial Mexico, as is the case of Asian people. They arrived via the Manila Galleon, ships that conducted the trans-Pacific trade with the Philippines by accomplishing round trips every year between 1565 CE and 1815 CE (Seijas, 2014). The largest migration occurred in the 17th century, to compensate the diminished labor force from the demographic collapse (Seijas, 2014). Some Asians travelled deliberately to Mexico, but many others were slaves from Manila, where a third of the population were slaves of many ethnic origins (Seijas, 2014). The main disembarking point was located in Southern Mexico, in the Pacific Coastal port of Acapulco, Guerrero. This genetic contribution may have remained overlooked as Asians were treated as indigenous vassals by law in the 17th century. At the end, they were referred as "Indians" in the same way as Native Americans and they were assimilated in the population (Seijas, 2014).

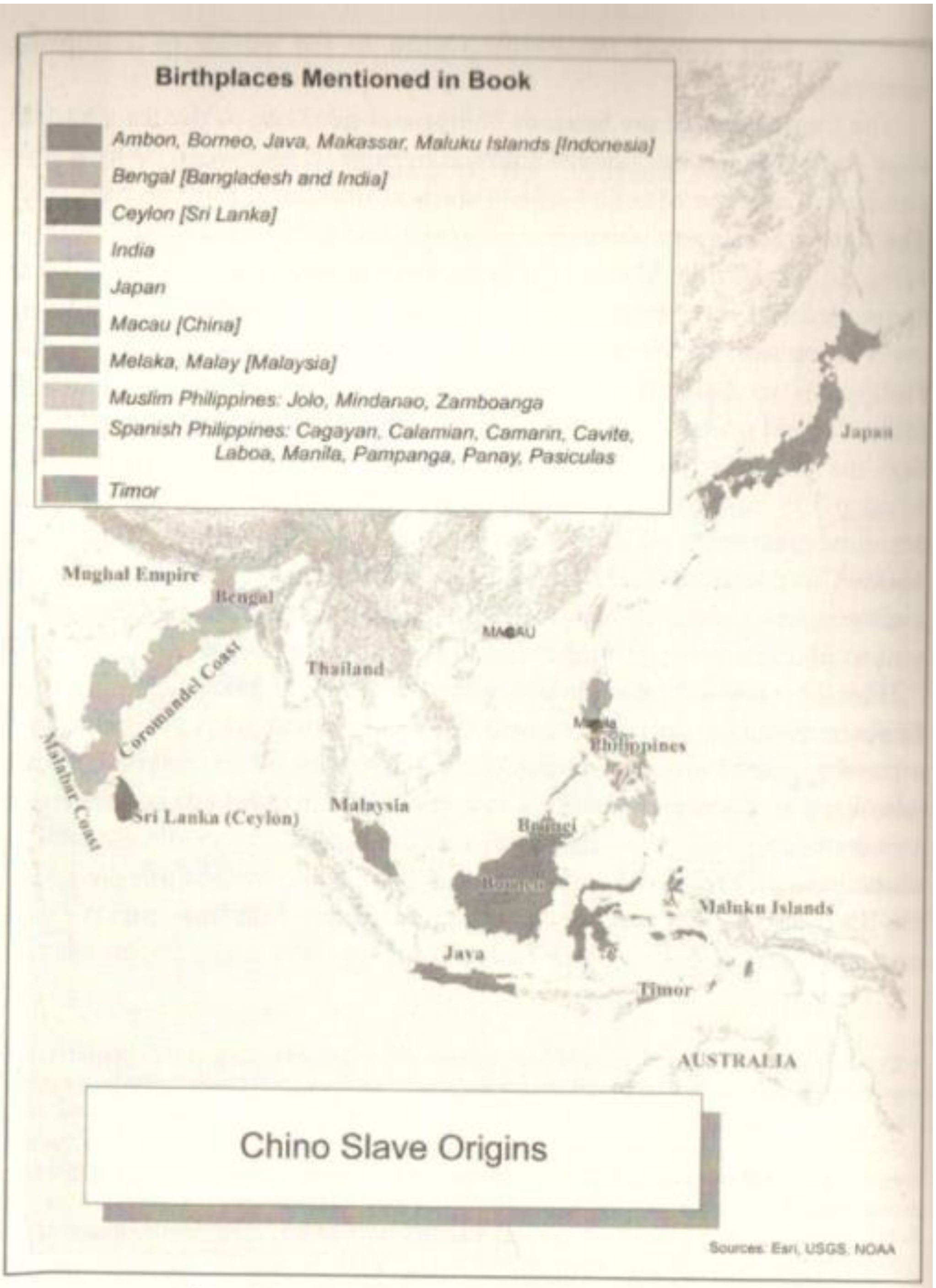
1.6.2 Asian presence in Mexico

Historical records estimate a total of 40,000-120,000 immigrants from Manila in colonial Mexico (Carrillo, 2014). The Spanish wrote they were very numerous in Acapulco, where every home had at least three or up to eighteen Asian slaves (Seijas, 2014). The cultural impact of this migration is evident in Mexico with the usage of terms of Filipino etymology such as "parián" (Guevarra, 2011). Also, the

Filipino beverage “tuba” or coconut wine had an important industry in the colony, which is still traditionally produced in the nearby coastal region of Colima. People from the coast of Guerrero even acknowledge an Asian heritage in the region.

1.6.3 The origin of colonial “chinos”

The city of Manila in colonial times represented an important slave trade center, as Natives had their own slavery system since pre-contact periods. Spanish people did not interfere, on the contrary, they traded with indigenous Filipino elites and the Portuguese. Manila was a multiethnic city with slaves from as far as Africa, India, Indonesia, Japan and many places more (Seijas, 2014). However, when all these overseas peoples arrived in Mexico, they were ambiguously labelled with the term “chino” (Spanish term for Chinese) regardless of their diverse origins. Scant written record was left about their ethnicity, as many Asian slaves were brought illegally. Only 225 of the many thousands of Asians have identifiable historical origins (as shown in Figure 1.6) (Seijas, 2014). Genetic studies have not confirmed the existence of a remnant from this Asian immigration in modern Mexicans, and the exact origin of the many illegal chinos remains unsolved.



Birthplaces Mentioned in Book
Ambon, Borneo, Java, Makassar, Maluku Islands [Indonesia]
Bengal [Bangladesh and India]
Ceylon [Sri Lanka]
India
Japan
Macau [China]
Melaka, Malay [Malaysia]
Muslim Philippines: Jolo, Mindanao, Zamboanga
Spanish Philippines: Cagayan, Calamian, Camarin, Cavite, Laboa, Manila, Pampanga, Panay, Pasiculas
Timor
Mughal Empire
Bengal
Coromandel Coast
Malabar Coast
Sri Lanka (Ceylon)
Thailand
MACAU
Japan
Manila
Philippines
Malaysia
Brunei
Borneo
Maluku Islands
Java
Timor
AUSTRALIA
Chino Slave Origins
Sources: Esri, USGS, NOAA

FIGURE 2.1. Chino Slave Origins. Map prepared by Eric Johnson, Numeric and Spatial Data Services Librarian, Miami University.

Figure 1.6: Birthplace of "chinos" involved in the slave trade during the colony according to historical records. Page from the book (Seijas, 2014).

1.6.4 Post-colonial Asian immigration

Post-colonial Asian immigrations are well-separated in time from the colonial immigration from Manila. Immigration from the Manila Galleon was greatly reduced after 1672 CE, when Asian slavery was actively abolished because of their free legal condition as natives. The next relevant Asian immigration occurred centuries later. It was until 1880 CE and 1910 CE when the President Porfirio Diaz favored Chinese immigration to work at railroad constructions and agriculture in Northern Mexico (Ma & Cartier, 2003). The Chinese population reached its maximum in 1930 CE with 25,000 people, later disrupted by xenophobic movements (Ma & Cartier, 2003). Even though Pre- and Post-colonial Asian immigration involve related peoples at a continental level, they pose discernible migrations due to non-overlapping timings and to distinct Asian ethnicities.

1.7 Genetic studies in present-day Mexicans

Before the arrival of the Spaniards to the Americas, populations remained largely isolated from intercontinental admixture. The divergence of thousands of years produced uniparental heritages, allele frequencies and linkage disequilibrium patterns specific from each continental population, facilitating the identification of each genetic contribution in admixed individuals. Genetic studies have become more detailed as technologies improved progressively. The results are compared with historical records in order to support demographic events with independent datasets.

1.7.1 Studies of admixture complexity

- Latin Americans as a tripartite admixture

First genetic analyses supported the historical perspectives: Latin Americans show African, European and Native American heritages. The profile has been observed as intermediate between the three unadmixed source populations. In the case of Mexico, this has been supported since the employment of classical markers. After the discovery of DNA as a genetic molecule, microsatellites (Felix-López et al., 2006; Hernández-Gutiérrez, Hernández-Franco, Martínez-Tripp, Ramos-Kuri, &

Rangel-Villalobos, 2005) and small subsets of markers, including ancestry informative markers (AIMs) (Joshua Mark Galanter et al., 2012; Kosoy et al., 2009), have confirmed this three-way mixture gradient (Cerda-Flores & Garza-Chapa, 1989; Gorodezky et al., 2001). Uniparental markers, which are non-recombinant, corresponded to the three continental origins of Mexicans (Luna-Vázquez et al., 2008; Rangel-Villalobos et al., 2008; Salazar-Flores et al., 2010). Mexicans have always portrayed a predominant Native American and European ancestry, while the African component is proportionally lower but consistent.

- Differences in continental proportions

Although the two main continental sources are consistent across the country, some substructure has been acknowledged and proved since the studies using a few markers (Licea-Cadena, Rizzo-Juárez, Muñiz-Lozano, Páez-Riberos, & Rangel-Villalobos, 2006; Luna-Vázquez et al., 2008; Salazar-Flores et al., 2010). Northern Mexico displays a more European genetic profile and Southern Mexico a greater Native American ancestry, while some coastal states show a higher sub-Saharan African component, i.e. Veracruz and Guerrero (Moreno-Estrada et al., 2014). This substructure evidences the particular history from each region and suggests contrasting admixture dynamics. As genotyping technologies became much cheaper, a huge number of markers became accessible with SNP microarrays, giving place to more elaborate analyses, such as the inference of demographic dynamics. This is of interest, as admixture between continental groups did not occur in a single event across the Americas. Some regions were conquered first by the Spaniards, while Native assimilation has occurred repeatedly after contact (Lagunas Rodríguez, 2004). Each region had a differential amount of European and sub-Saharan African immigrants, as well as several migrations through the colonial period and afterwards. Demographic dynamics are usually addressed with sequencing data to properly characterize the allele frequency spectrum and to account for microarray's ascertainment bias. Nevertheless, elevated costs from sequencing technologies have encouraged the development of algorithms that employ ancestry tracts distributions derived from microarray data (Pool & Nielsen, 2009). These algorithms rely on the accuracy of local ancestry calls, which can be reliably inferred if enough markers (usually hundreds of thousands) are available, and not necessarily from sequencing data. Ancestry tracts inferences are appropriate to address admixture dynamics in Mexico as they represent recently admixed individuals from divergent ancestry sources, which can be easily differentiated (i.e., properly called) by mathematical algorithms. Moreover, they allow complex admixture models and the inclusion of larger sample sizes due to low-priced genotyping, compared to whole-genome sequencing.

- Differences in sub-continental origins

High-density genotyping data have allowed greater genetic resolution in population studies. As more projects genotype underrepresented populations such as Native Americans and admixed Latin American individuals, we can know more about their genomics patterns with higher precision. For instance, the Native American component of admixed Latin Americans differs between countries. Mexicans cluster with local Mesoamerican Natives while Peruvians group with geographically close Andean Natives (Wang et al., 2008). This variation has been shown to change even within Mexico (Moreno-Estrada et al., 2014). Studies with larger samples have identified a widespread Nahua ancestry, Mayan heritage in the Yucatan Peninsula, and a Southern Native component mainly in Oaxaca (Ruiz-Linares et al., 2014). More comprehensive reference panels from Native populations will show the specific source population from this structured Native component. On the other hand, the large number of markers allows the disentangling of minor admixture events, such as East Asian contributions.

1.7.2 Unresolved questions

- Inferences of demographic dynamics with genetics

Demographic inferences from local ancestry have been inferred for some Latin American countries, including Mexico, showing varying admixture timings (Moreno-Estrada et al., 2013). However, inferences for Mexico were obtained from a single population residing in the United States of America, omitting the regional admixture dynamics within Mexico. Analyses with separate populations per Mexican region are needed to highlight demographic differences and relate them to historical events. These will help elucidate the most relevant historical events that gave origin to the majority of Mexicans nowadays. In addition, population size and migration dynamics can be addressed with this approach.

- Detailed Native American heritage in Mexico

Nahua people are the most numerous Native Mexican group due to historical reasons. This ethnic group, as well as other Central Native groups, contributed to most admixed Mexicans. A deep sampling of Central Native groups, including various Nahua populations, will uncover the specific origin of the Native component across Mexican states. A broader Nahua panel should be considered as they exhibit genetic substructure even though they share the same language and ethnicity (according to unpublished data available in the lab). Advanced bioinformatics tools will be able to help pinpointing a more precise origin of the Native component.

- East Asian heritage in the Pacific Coast of Guerrero

The traditional view of three-way global ancestry analysis excludes details and additional immigrations. Some initial studies have proposed additional admixture sources, however, the lack of power from a limited number of markers had left these discussions as hypotheses without solid evidence (Gorodezky et al., 2001; Silva-Zolezzi et al., 2009). High-throughput genotyping studies have been able to identify more ancestries across Latin America, e.g. Sephardic and non-European Mediterranean (Chacón-Duque et al., 2018). Attempts to characterize East Asian ancestry in Latin American have been described previously (Chacón-Duque et al., 2018; Silva-Zolezzi et al., 2009), but only low proportions were observed. Low proportions prevent the ability to discard false positive contributions due to similarities between East Asians and Native Americans. Meticulous computational tools and new datasets will help characterize this component, originated in historical events that possibly led to East Asian contributions. Particularly, the study of Mexican sampled regions with considerable historical Asian presence results in higher contributions to allow a more thorough analysis, as it is the case of the state of Guerrero. Further and more detailed genetic analyses will reveal the existence of this unexplored component and its sub-continental origin, in addition to its anthropological implications.

2 OBJECTIVES AND JUSTIFICATION

2.1 General objective

Genetically characterize admixture dynamics and demographic processes in different Cosmopolitan Mexican populations since European contact.

2.2 Specific objectives

1. **Estimate the number, size and timing of continental migration** waves in different Mexican states.
2. **Explore the substructure within the Nahua component** by adding newly genotyped samples from Central Mexico, in order to have a more comprehensive Native American reference panel.
3. **Characterize the sub-continental ancestry** of European and Native American components, as well as the understudied East Asian ancestry in Mexico.
4. **Infer population dynamics** such as recent gene flow between Mexican states and effective population sizes across generations.

2.3 Justification

Mexicans exhibit a genetic background as complex as their history. Despite numerous efforts to characterize ancestry in Mexicans, most of the genetic studies conducted so far have not specifically addressed demographic dynamics and less prevalent components in Mexico, which remain to be elucidated with genomic tools. Thus, additional studies are justified and required to resolve more complex and finer regional demographic patterns within Mexico. These patterns can be linked to historical events such as major admixture events and population size fluctuations.

Another limitation from previous studies is the lack of deep sampling in the Central region of Mexico. Nahua groups are the most extended pre-Columbian population, which are also known to have substantially contributed to the genetic pool of present-day Mexicans. Additional Nahua samples must be included in genotyping efforts to increase the diversity of reference panels, providing a more precise resolution in the Native American substructure of admixed Mexicans.

On the other hand, understudied ancestries that have contributed to a lesser extent remain to be explored with novel methods, e.g. East and Southeast Asian components. This justifies the inclusion of extended reference panels to cover the Southeast Asian region in genomic studies like the one presented here. Resolving the population substructure of Mexico could have important implications in biomedical studies that assume more homogeneous genetic profiles in Mexican and other Latin American populations, classifying them as a single group for medical and epidemiological purposes.

3 MATERIALS AND METHODS

3.1 Data description

- Cosmopolitan Mexican genotyping data

In order to study the substructure of admixed Mexicans, a total of 312 cosmopolitan Mexicans was analyzed and genotyped as part of a previous publication (Moreno-Estrada et al., 2014). We utilized the high-density Mexican Genome Diversity Project (MGDP) dataset, which consists of seven cosmopolitan and three Native American populations genotyped with two microarrays in b36/hg18 build. Each individual was genotyped with both Affymetrix 500K and Illumina 550K arrays resulting in more than 900,000 SNPs. A high-density genotyping is fundamental to obtain reliable local ancestry calls, as each genomic window will have more input markers. This accuracy is important when assigning an ancestry at short ancestry tracts, which are prone to errors. The cosmopolitan Mexican dataset includes 49 individuals from Hermosillo, Sonora; 17 from Ciudad Victoria, Tamaulipas; 50 from Zacatecas, Zacatecas; 48 from Guanajuato, Guanajuato; 50 from Xalapa, Veracruz; 50 from Acapulco, Guerrero, and 49 from Mérida, Yucatán (locations shown in Figure 3.1). The cities are among the largest and most important from each Mexican state sampled. Cosmopolitan samples are intended to provide complex admixture signals we aim to characterize. All individuals were asked if their four grandparents were born in the state they were sampled, thus describing complex but regional admixture events. For a subset of the analysis, additional Native American reference populations were required to increase the substructure resolution. Therefore, 49 Nahua samples were newly genotyped for this study.

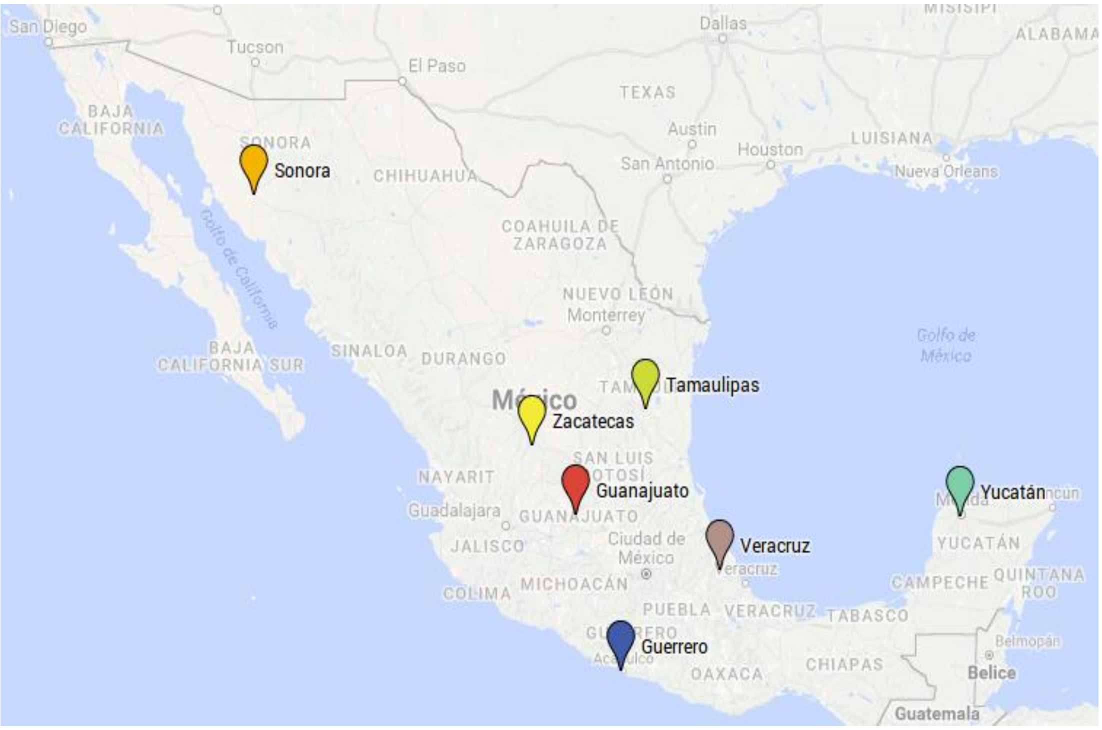


Figure 3.1: Cosmopolitan Mexican sampling locations. Locations point to the cities where the sampling was performed, while the labels indicate the Mexican state they represent.

- New sampled Nahua

The sampling was conducted in collaboration with Dr. Rosenda Peñaloza and Leonor Buentello from the Universidad Autónoma Metropolitana and the Instituto de Investigaciones Antropológicas at UNAM, respectively. The Nahua populations included consist of San Pedro Atocpan and Xochimilco in Mexico City, Necoxtla in Veracruz and Zitlala in Guerrero. All individuals were Nahuatl speakers with local ancestors. Samples were collected with the appropriate informed consent for population genetic studies. DNA was extracted from blood samples and genotyped using the Axiom LAT 1 array (World Array IV chip), which includes 783,856 SNPs common in Latin American populations. Genotyping was coordinated by Dr. Karla Sandoval and Dr. Andres Moreno-Estrada and performed at UCSF in collaboration with Dr. Esteban Burchard. Forty-nine self-identified Nahua were genotyped from a previous study (Joshua M Galanter et al., 2014).

- Lifting over and filtering of the cosmopolitan Mexican dataset

The MGDP dataset was updated from GRCh36/hg18 to GRCh37/hg19 with the LiftOver's executable file (Karolchik et al., 2003), in order to match the build of the reference panels and allow subsequent analyses. Ambiguous SNPs (A/T and C/G) were detected with the snpflip script (source: https://github.com/biocore-ntnu/snpflip) and removed. These SNPs have two alleles which can be a flipped version of each other, thus it is hard to determine which instances are a flipped SNP or a real mutation in the individual without the respective raw microarray data. The inclusion of ambiguous SNPs generates batch effects when strand errors are recurrent in a dataset. The final dataset had 829,278 SNPs, after lifting over, correcting flipped variants and removing all ambiguous variants. We pruned each continental and admixed panel separately with PLINK v1.90b4.4 using the next parameters: a mind of 0.1, geno of 0.05, hwe of 10-3 and me of 0.05 and 0.1. These parameters filtered markers with a high missingness, which would result in errors in the pipeline. A mind flag of 0.1 removes all individuals with >10% of missing data, while a geno of 0.05 excludes all SNPs with a missingness of >5%. A hwe flag of 1e-3 discards biased heterozygote genotypes based on the Hardy-Weinberg equilibrium. The heterozygote frequency should be "2pq", where "p" and "q" represent the frequency of each of the two alleles. Even though in practice most SNPs are in disequilibrium, genotyping errors are intended to be removed with extreme hwe thresholds. The filter is applied to populations separately. An me flag of 0.1 and 0.05 discards genotypes and individuals that do not display a Mendelian inheritance, respectively. Mendelian concordance is compared to the reported pedigrees to discard genotyping errors and false trios.

- Dataset merging

Merges were performed depending on the nature of the algorithms involved. Each analysis required different continental references or an additional sub-continental reference. Table 3.1 shows the specific reference panel and the number of intersecting SNPs for every analysis. This step was done with PLINK's bmerge function by extracting the SNP's by pairs of datasets. Merged datasets contain the intersection of all sets of SNPs, not the union of the sets. Merging is performed in this manner as missing data due to different lists of SNPs will be an obstacle to further analyses. Filtered and merged datasets were the input for all analyses covered in this thesis, which are shown in Figure 3.2.

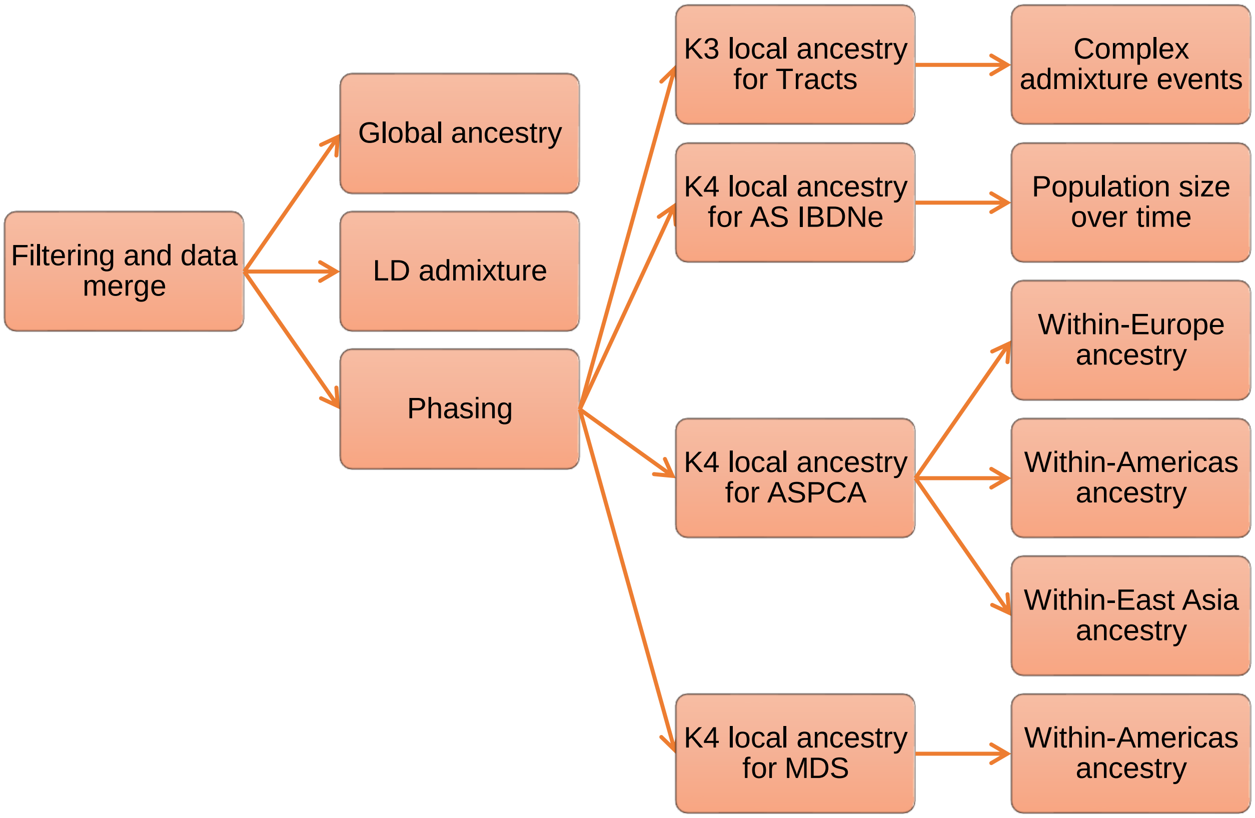

Filtering and data merge
Global ancestry
LD admixture
Phasing
K3 local ancestry for Tracts
K4 local ancestry for AS IBDNe
K4 local ancestry for ASPCA
K4 local ancestry for MDS
Complex admixture events
Population size over time
Within-Europe ancestry
Within-Americas ancestry
Within-East Asia ancestry
Within-Americas ancestry

Figure 3.2: General pipeline for analyses covered in the thesis. It includes global ancestry, LD-based admixture estimation and local ancestry related analyses, such as the determination of admixture timings, population sizes and within-continent ancestry. The intersection of datasets resulted in differing SNP numbers as each analysis required different population subsets. For further details see Table 3.1, a list of populations and the number of SNPs are provided for each analysis.

Table 3.1: Summary of each analysis included in this thesis. It includes the type of analysis, test populations, reference populations and number of markers considered. The algorithm employed in each analysis is shown in parenthesis in the first column. ASPCA analyses show the sub-continental populations included. Source publication from each population are specified with superscript numbers corresponding to the list below.

Analysis (program used)	Target dataset	Reference dataset	Intersection SNP amount
Global ancestry (ADMIXTURE)	Cosmopolitan Mexicans[1]	Continental African: YRI, MSL[2] Europe: IBS, CEU[2] Native: PEL[2] East Asian: KHV, CHB[2]	80,587
LD admixture (MALDER)	Cosmopolitan Mexicans[1]	Continental African: YRI, MSL[2] Europe: IBS, CEU[2] Native: TAR, TOT, ZAPN, TZT[3] East Asian: KHV, CHB[2]	426,915
Complex admixture events (Tracts)	Cosmopolitan Mexicans[1]	Continental African: YRI[2] Europe: CEU[2] Native: TEP, ZAPN, MYA[1]	762,455
Population size over time (AS IBDNe)	Cosmopolitan Mexicans[1]	Continental African: YRI[2] Europe: CEU[2] Native: TEP, ZAPN, MYA[1] East Asian: CHS[2]	430,089
European ASPCA	Cosmopolitan Mexicans[1]	Continental African: YRI[2] Europe: CEU[2] Native: TEP, ZAPN, MYA[1] East Asian: CHS[2] Sub-continental European: SW Europe, S Europe, SE Europe, E Europe, W Europe, C Europe, NW Europe, NNE Europe[4] and Basque[5]	254,463
Native American ASPCA	Cosmopolitan Mexicans[1]	Continental African: YRI[2] Europe: CEU[2] Native: TEP, ZAPN, MYA[1] East Asian: CHS[2] Sub-continental	369,242

		Native American: Tarahumara, Tepehuano, Huichol, Nahua (Jalisco), Purepecha, Totonac, Nahua (Puebla), Trique, Zapotec, Mazatec, Tzotzil and Maya[3]	
East Asian ASPCA	Cosmopolitan Mexicans[1]	Continental African: YRI[2] Europe: CEU[2] Native: TEP, ZAPN, MYA[1] East Asian: CHS[2] Sub-continental East Asian: Japan, Northern China, Southern China, Negrito Philippines, Non-Negrito Philippines, Sumatra, Borneo, Lesser Sunda Islands and Maluku Islands[6]	368,928
Array A Native American MAAS-MDS	Cosmopolitan Mexicans[1]	Continental African: YRI[2] Europe: CEU[2] Native: TEP, ZAPN, MYA[1] East Asian: CHS[2]	819,971
Array B Native American MAAS-MDS	Native Americans from NMDP[3]	Continental African: YRI[2] Europe: CEU[2] Native: TAR, TOT, ZAPS, TZT[3] East Asian: CHS[2]	693,556
Array C Native American MAAS-MDS	Nahua natives[7]	Continental African: YRI[2] Europe: CEU[2] Native: PEL[2] East Asian: CHS[2]	737,815

1.- Mexican Genome Diversity Project (MGDP) from (Moreno-Estrada et al., 2014).
2.- 1000 genomes consortium from (Gibbs et al., 2015).
3.- Native Mexican Diversity Project (NMDP) from (Moreno-Estrada et al., 2014).
4.- Population Reference Sample (POPRES) from (Nelson et al., 2008).
5.- Basque population from (Henn et al., 2012).
6.- Southeast Asian reference panel from (Reich et al., 2011).
7.- New Nahua populations from this thesis (see section 3.1).

3.2 Global ancestry

Global ancestry proportions were estimated for the reference populations and admixed cosmopolitan Mexicans. The analysis was performed with Admixture version 1.3.0 in unsupervised mode without replicates. The proportion of four well-differentiated continental source populations was determined for all samples: sub-Saharan African, European, East Asian and Native American. Each continental signal was estimated with an equal number of samples from each reference panel to avoid cryptic clusters. In order to include the largest number of markers, the cosmopolitan Mexican dataset was merged with the 1000 genomes consortium dataset (Gibbs et al., 2015) exclusively. Individuals from 1000 genomes provided all four continental reference panels. 20 YRI and 20 MSL represented the Sub-Saharan African panel. 20 IBS and 20 CEU made the European panel. 40 PEL with >90% of Native American ancestry represented the Native American panel. Finally, 20 KHV and 20 CHS composed the East Asian panel. A linkage disequilibrium (explained in the next section) pruning step was performed in the merged dataset with the plink indep-pairwise flag, using the parameters 50, 10 and 0.1. In such way, it was performed on markers within 50 kb windows exceeding an r^2 of 0.1 and considering 10 kb steps to allow some overlap between windows. A total of 80,587 SNPs without LD was considered for the Admixture runs. Generally, the exclusion of variants under LD do not change Admixture results. However, LD is not strictly considered in the algorithm and the remaining tens of thousands of SNPs provide enough markers for this continental resolution.

3.3 Admixture timing based on linkage disequilibrium

Continental admixture timings were estimated with Multiple Admixture-induced Linkage Disequilibrium for Evolutionary Relationships (MALDER), which uses recombination patterns. Recombination does not occur across all loci every generation. A pair of markers close to each other in the genome will be co-inherited as recombination is unlikely to happen randomly at the very restricted region between them. This co-inheritance is called linkage disequilibrium (LD) and all human populations show degrees of it depending on their evolutionary relationship. LD can be explained by a shared bottleneck between populations or due to admixture from a source population (Loh et al., 2013). We will focus on the latter case. Depending on the timing and magnitude of the genetic contribution from a source population to a receiver population, a specific LD exponential pattern is portrayed. More recent admixture events show a longer LD as less recombination events have broken the chromosome-size source haplotypes. As more generations pass by, more random recombination events will shorten the LD segments. Algorithms such as ALDER, approximate the exponential decay of LD between admixed and source populations to characterize admixture details. Here, we utilize an extended version of ALDER called MALDER that analyzes multiple admixture events and source populations. This is an appropriate approach for Mexican populations which had a complex admixture with more than two continental sources arriving in multiple waves (Lagunas Rodríguez, 2004).

The default steps of ALDER consist of searching the minimum length of LD blocks to consider for the fit of the exponential distribution at first. However, in our dataset, Native Americans presented a high LD correlation with cosmopolitan Mexicans. To avoid the algorithm to stop at this point because of high correlation, the flag mindis was set to 0.1 cM. Events estimated with this threshold correspond to an average of 15 generations in the past, corresponding to post-contact admixture events. The threshold was calculated with the formula: average generation = 3/(2*length in Morgans). A similar sample size for each reference population was provided to avoid biases on estimations. The prediction with the highest Z-score from each pulse was considered, as each calculated pulse provides amplitudes and error ranges for every pair of reference populations. Once admixture predictions were chosen in this manner, Z-score values lower than 5 were discarded as they represented unreliable admixture events. Furthermore, amplitudes smaller than 1 and wide errors of more than 5 generations were also excluded as they were not informative of a real admixture event. Because of this, some cosmopolitan populations had a single admixture prediction, while in other cases MALDER was able to characterize three admixture timings due to a clear LD correlation for each pairwise reference.

All methods regarding admixture timings provide generation-based estimates. We converted the generation numbers into dates, considering the sampling was done in 2005 and assuming generations of 30 years. Generations of 30 years have been proposed as an accurate estimation for average generation time in humans across societies. It represents the mean of the average generation for both women and men, which have been shown to be lower in women and higher in men (Tremblay & Vézina, 2000).

3.4 Local ancestry

A phasing (or estimation of haplotypes) and local ancestry pipeline was applied to the genotype data, to make demographic inferences and sub-continental PCA's (see pipeline from Figure 3.3). Demographic inferences used the local ancestry block distribution across populations with the Tracts algorithm, while ASPCA required the masking of the inferred blocks of certain ancestries.

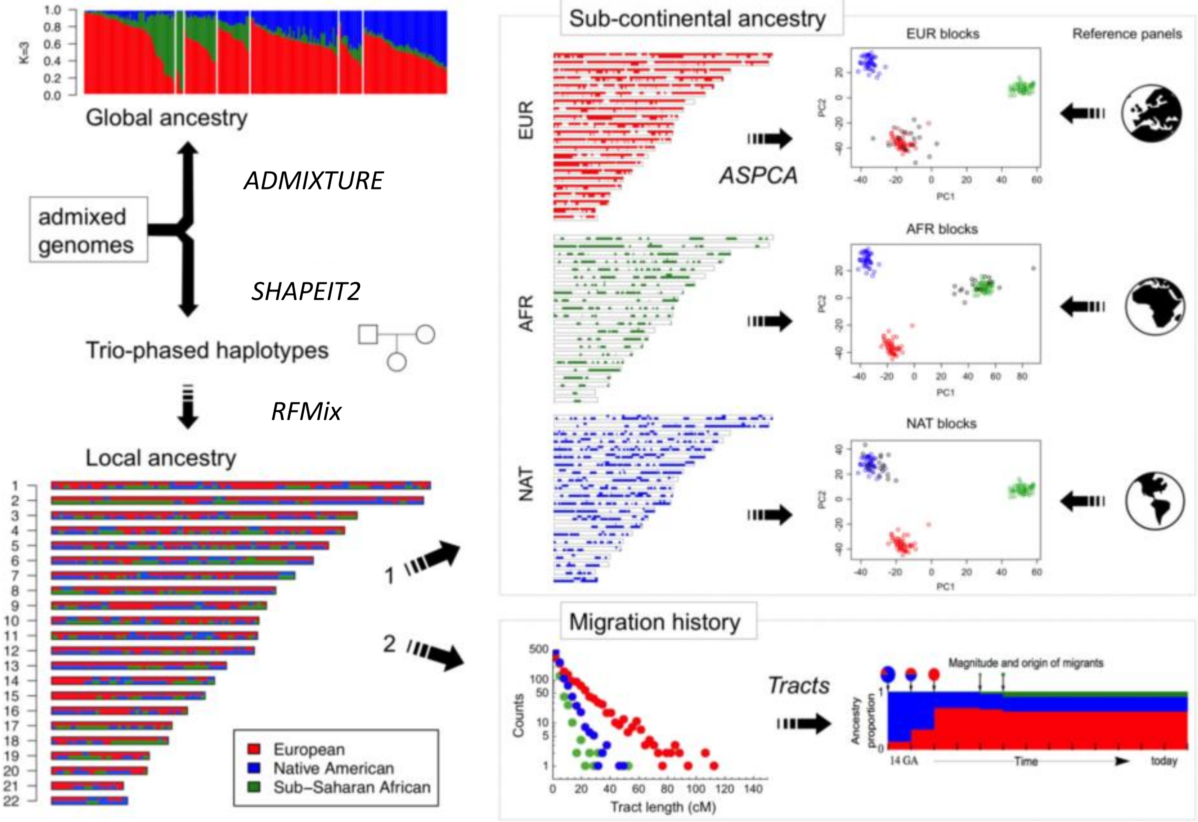


Figure 3.3: Local ancestry-based workflow for ASPCA and Tracts. Phasing was done with SHAPEIT2 and local ancestry calling with RFMix. Masked ancestry tracts provide the

ASPCA's input, while ancestry tracts histograms represent the input for Tracts. Modified from (Moreno-Estrada et al., 2013).

- Phasing with SHAPEIT2

All genotype data was phased with Segmented HAPlotype Estimation & Imputation Tool (SHAPEIT2) (Delaneau, Zagury, & Marchini, 2013) as further analyses rely on haplotypes rather than genotypes. All parameters were set as default, except for the duoHMM flag. The flag is intended to phase complete trios from Native American populations, i.e. 10 Tepehuano and 15 Mayan trios. 25 children and 50 parents comprised a total of 75 individuals to be trio-phased. Each reference panel and the admixed panel were all phased separately.

- Local ancestry with RFMix

Local ancestry was performed with RFMix v1.5.4 on the phased haplotypes, to classify genomic regions at a continental level. RFMix utilizes random forests, a machine learning approach of classification trees. The algorithm uses the reference panels as training data and employs the multiple decision trees that result to classify haplotypes, as shown in Figure 3.4.

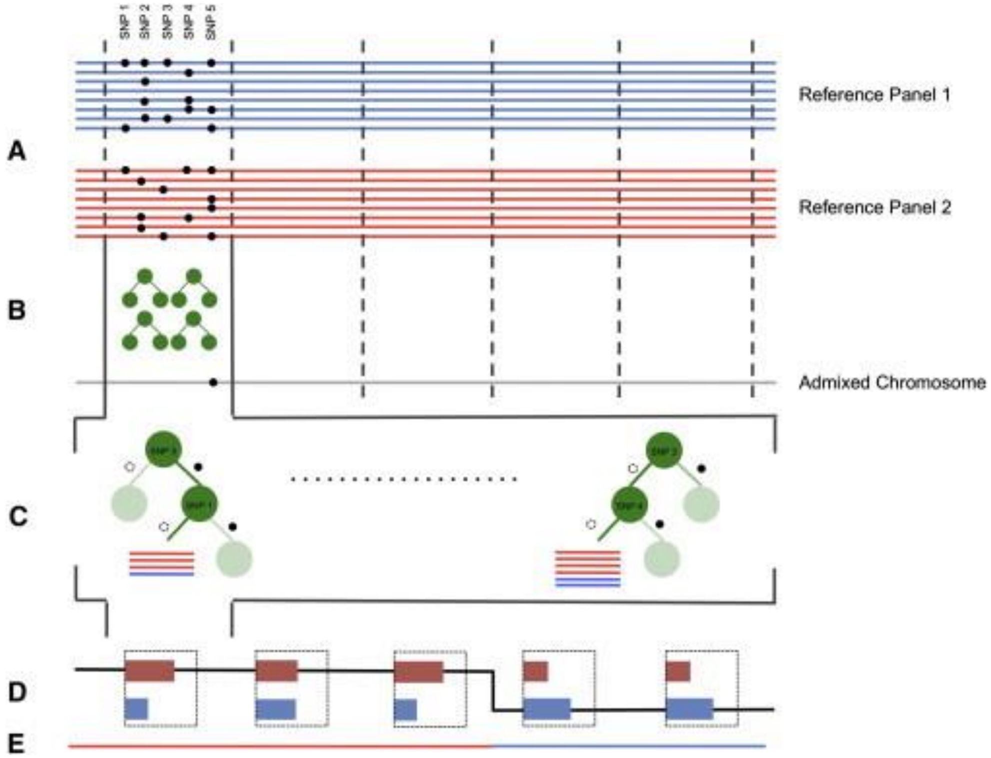


Figure 3.4: RFMix algorithm's workflow. A) Haplogroups specific of each reference panel are identified with classification trees. In order to include enough SNPs, a fixed window size was set (0.2 or 1 cM). B) The constructed trees from training data are later applied to test haplotypes from admixed individuals. C) Each tree has a set of excluded SNPs to achieve a greater diversity in decision trees and as a result, more robust assignments. D) A consensus from all decision trees allows the classification of the test genomic region to a reference ancestry. E) A confidence threshold is applied at the end (>90%). A discrete ancestry is assigned to each SNP, unless no consensus was achieved. Ancestry is reported as unknown if no ancestry passed the threshold. Figure from (Maples et al., 2013).

A rephasing step is required to have more reliable ancestry tracts. In this case, it was executed according to population patterns by setting the RFMix PopPhased flag. It was not performed in a pedigree-based manner, due to the absence of trios and duos in the admixed individuals. Default parameters were set: 0.2 cM window sizes, 8 generations, 100 trees to generate per random forest, zero EM iterations and one for the minimum number of reference haplotypes per tree node. We considered three or four continental reference panels in the local ancestry pipeline. For the demographic inferences, three references were used: Sub-Saharan

African, European and Native American, as Tracts models with four ancestries are non-existent. The rest of the analyses employed an additional East Asian reference panel (as shown in Table 3.1).

3.5 Estimation and timing of admixture events

Recombination occurring every generation generates ancestry tracts histograms with a distinctive exponential distribution. The meticulous study of the effect of admixture parameters on simulated distributions, allows a great resolution on the timing and magnitude of each admixture event. Once an admixture event occurs, ancestry variance in the population decreases from its maximum in an exponential fashion due to the segregation of incoming chromosomes and recombination (see Figure 3.5). Then, the variance diminishes temporarily in a linear way due to linkage disequilibrium. Finally, the exponential decay is resumed due to genetic drift ~25 generations later.

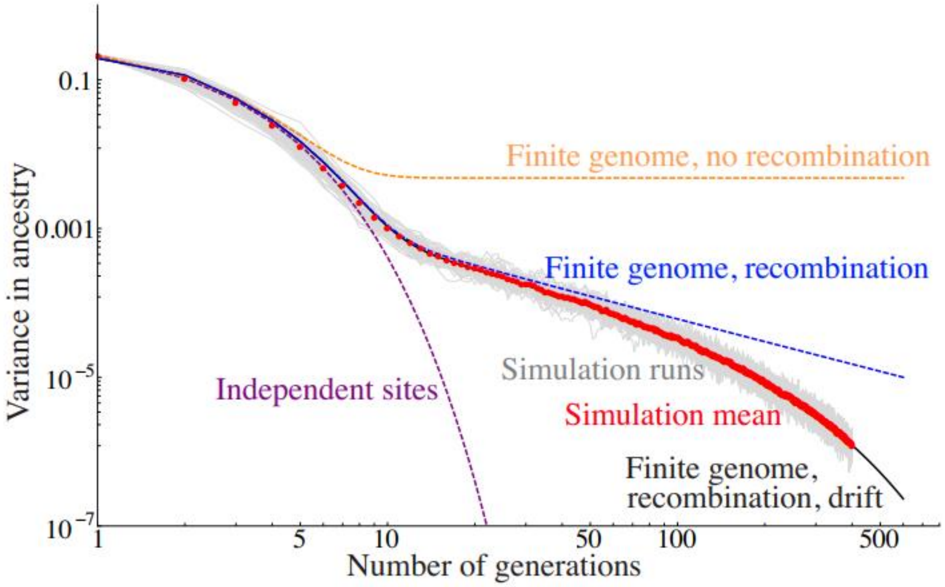


Figure 3.5: Variance changes per generation after a simulated single admixture event. Recombination, genome size and genetic drift shape the variance decrease over time.

The first methods to infer admixture from local ancestry tracts distributions had some limitations (Pool & Nielsen, 2009). A new method called Tracts was developed (Gravel, Stephens, & Pritchard, 2012) following the same approach with some improvements. In contrast to the previous algorithm, Tracts allows medium and large immigration pulses, considers recombination events between same ancestry tracts and ponders end chromosome dynamics (Gravel et al., 2012). Tracts accuracy relies on the quality of the local ancestry calls, especially for short tracts which can be misassigned if the SNP density is low. Therefore, high-density genotyping is highly recommended in order to have enough SNPs even at short ancestry switches. Tracts poses a suitable approach to this dataset as it is not sensitive to ascertainment bias and as it allows complex demographic models involving multiple waves from the same ancestry, compared to LD methods as MALDER.

In order to run Tracts, a priori admixture pulse orders are provided. Each Tracts model syntax consists of migration pulses into the admixed population shown as groups of three letters (as three is the total number of continental ancestries considered in this case). The letter order corresponds to the continental ancestries: European, Native American and African. For example, in the admixture event “ppx”, the “p” letters represent a pulse where two ancestries are contributing, while the “x” represents the absence of that ancestry in that specific pulse. Therefore, “ppx” indicates an admixture event between Europeans and Native Americans, with no contribution of African ancestry.

According to historical data, African slave trade in Latin America took place decades after the main contact between Europeans and Native Americans (Eltis, 2018). Thus, all models tested consist of an initial admixture event between Native Americans and Europeans (ppx), followed by an African pulse (xxp). We tested four Tracts models separately for every cosmopolitan Mexican population. The first model comprises only the previous two pulses (ppx and then, xxp). The second one is the same as the first one, plus another African pulse (ppx, xxp and xxp, again). The third one shows more than one European pulse (ppx, xxp and pxx). Finally, the fourth model has a final dual pulse of European and Native American (ppx, xxp and ppx).

The algorithm estimates the timings and magnitude of the admixture events with the admixture order restriction. The fit of the predicted and real tract distributions is evaluated with a likelihood. Each population has four estimated likelihoods corresponding to the four Tracts models evaluated. A single run of each model was considered for each Mexican state, the one with the best likelihood. Finally, likelihoods were adjusted with Bayesian Information Criterion (BIC), as more complex models usually have better scores for their parameter flexibility (see annexes Table 8.2).

These analyses required K3 local ancestry as the models available are restricted to three-way admixture. East Asian tracts were called as Native American due to genetic proximity. However, they are uncommon enough in the population to affect the Native American timing estimates considerably.

3.6 Gene flow and effective population size

Identity by descent (IBD) analyses can provide an insight into migration, population size, growth and bottlenecks in the past. An IBD fragment is an identical haplotype shared between a pair of individuals. It is called “by descent” because the similarity and size of the fragment can only be explained by the existence of a recent common ancestor inheriting the autosomal haplotype to both individuals, as seen in Figure 3.6. Shared IBD between populations provides information about recent gene flow, while IBD length histograms within a population informs about the effective population across generations.

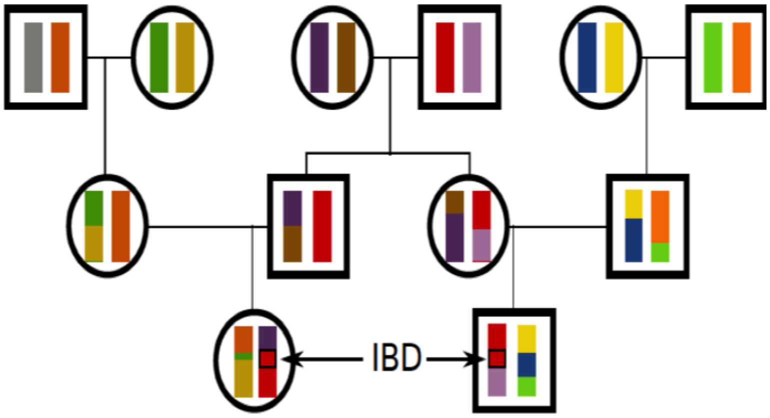


Figure 3.6: Explanation of IBD segments. The diagram illustrates how certain haplotypes (called IBD segments) shared between individuals are explained by a common ancestor few generations ago.

- Recent gene flow between Mexican states

IBD segments larger than 5 cM were obtained with the GERMLINE algorithm (Gusev et al., 2009), as shorter haplotypes have a high false positive rate with this algorithm. Because of the IBD theory, these fragments are reported for each pair of individuals. For example, individuals A and B can share two IBD segments: one IBD fragment of 5 cM in chromosome 21 and another of 15 cM in chromosome 22. Quantification for general patterns of gene flow was done as the sum of the length of all IBD fragments, thus adding a total of 20 cM of shared IBD between individuals A and B. A normalization was applied to obtain the average total shared IBD at a population level in a pairwise manner, e.g. a total average for Sonora-Tamaulipas, for Sonora-Zacatecas, for Sonora-Sonora, etc. Normalization was achieved by adding every pair's total IBD lengths and dividing the result by the number of possible pairwise comparisons between states. The dividing factor in Sonora-Tamaulipas would be *n*m* or 49*17, according to the sample size of each state. While for self-comparisons, such as Sonora-Sonora, the normalization would consist of *n(n-1)/2* or 49*48/2. The result consisted of a matrix of relative gene flow between Mexican states.

Finally, matrices were performed once more in an ancestry specific way. Calculations were made identically, but only IBD fragments with a clear local ancestry assignment were considered for each ancestry specific IBD matrix. IBD segments were filtered using RFMix calls at K4.

- Effective population size across generations

The length of IBD fragments informs about time-related events, as shorter fragments are a result of more recombination events and thus, older events. For example, an overrepresentation of IBD fragments of certain cM could represent a bottleneck. According to the length of those fragments, individuals at that time became more related by sharing more common ancestors. Even if the population size recovered afterwards, the IBD distribution will reveal that past event.

Beagle v4.1 was employed to estimate IBD fragments >= 3.0 cM for this analysis (B. L. Browning & Browning, 2013). Beagle utilizes GERMLINE to estimate shared haplotypes that are long enough to assume they were inherited from a common ancestor and not a random result of recombination. Then, the results are refined with probabilistic approaches that model LD, allowing reliable identification of IBD segments < 5 cM.

In order to estimate effective population across generations, IBDNe (S. R. Browning & Browning, 2015) employed the estimated IBD fragments as an input. This algorithm focuses on recent estimates, compared to other methods like site frequency spectrum (SFS) analyses (which require large sample sizes to make recent estimates). The other methods can model older demographic events such as the out-of-Africa bottleneck, while IBDNe is an ideal approach for post-contact times as it provides estimates between 4 and 50 generations in the past with SNP array data. IBDNe utilizes a coalescent approach that considers the probability q_g of a pair of random haplotypes sharing a common ancestor in a specific number of generations *g* in the past. The number of generations is chosen if an ancestor is not shared in a previous generation *g - 1*. In other words, the oldest generation that can explain the length of the shared IBD segment is taken into account. Finally, the effective population size at certain generation (N_g) is estimated by utilizing the probability q_g in the next solved equation: $N_g = 1/(2q_g)$.

The particular case of Mexicans consists of three very differentiated source populations that recently admixed in the last ~17 generations. As each ancestry represents independent demographic histories and are easily discernible with local ancestry, effective population sizes can be estimated for each ancestry, separately. Ancestry Specific IBDNe (AS IBDNe) considers IBD segments that coincide with local ancestry segments. AS IBDNe (S. R. Browning et al., 2018) was applied to the Native American, European, Sub-Saharan African and East Asian components in cosmopolitan Mexicans (see Table 8.1 for the local ancestry calling). Mexican states were grouped in metapopulations, according to their geographic region and Native American affinities, as AS IBDNe recommends more than 100 samples per population. Three groupings of cosmopolitan states were considered: the first grouping included 215 individuals from all states except for the most distant states of Sonora and Yucatan; the second one considered near northern and northeast states with 115 individuals from Tamaulipas, Zacatecas and Guanajuato, and finally, a Central-Southern estimate with 100 samples from Veracruz and Guerrero.

3.7 Sub-continental ancestry with PCA

Ancestry Specific PCA (ASPCA) allows a within-continent resolution in ancestry. Its resolution makes it possible to differentiate between European ancestries such as Spanish and French, or Native American ancestries such as Nahua and Totonac. This is achieved by masking all other ancestry tracts. For example, only European ancestry in cosmopolitan Mexicans is compared to a European reference panel. Finally, a PCA is performed and the overlap reports the exact within-Europe heritage in admixed individuals. ASPCA will provide a sub-continental resolution for each continental ancestry in cosmopolitan Mexicans. The number of markers included in each ASPCA varies as it depends on the datasets merged. Details from each ASPCA analysis are provided in Table 3.1. Continental references in the table refer to the panel for local ancestry assignment, while sub-continental references were included in the ASPCA along with the properly masked admixed cosmopolitan Mexicans.

- European ASPCA

ASPCA achieved a sub-continental resolution in the European ancestry of cosmopolitan Mexicans. We performed a PCA with 254,463 SNPs. It included 616 cosmopolitan Mexican haplotypes of European descent compared to a sub-continental reference panel of Europe. Almost all admixed individuals had more than 25% of European ancestry to be included in the ASPCA. The sub-continental panel includes 41 populations across Europe and nearby regions. They were classified into eight geographic regions, as well as an additional label for Basque individuals (see supplementary Table 8.1). Basque people are considered separately as they have a different genetic footprint to the rest of Europeans and Iberians. Also, this heritage is important to characterize as historical records report Basque immigration to the Americas during the colonial period (Lagunas Rodríguez, 2004).

- Native American ASPCA

Within-Native American resolution in cosmopolitan Mexicans was analyzed with ASPCA. The PCA employed 369,242 SNPs and compared 601 cosmopolitan Mexican haplotypes of Native American descent to a sub-continental reference panel of Native Americans from Mexico. The reference panel forms part of the Native Mexican Diversity Project (NMDP) (Moreno-Estrada et al., 2014). Cosmopolitan individuals with more than 25% of Native American ancestry were considered in this ASPCA. The sub-continental panel included 15 Native American populations grouped into the following 12 categories. It included 48 Tarahumara haplotypes, 60 Tepehuano haplotypes, 48 Huichol haplotypes, 40 Nahua (from Jalisco) haplotypes, 46 Purepecha haplotypes, 128 Nahua (two populations from Puebla) haplotypes, 48 Totonac haplotypes, 34 Mazatec haplotypes, 90 Zapotec (both Northern and Southern Zapotecs) haplotypes, 48 Trique haplotypes, 42 Tzotzil haplotypes and 126 Mayan (from Campeche and Quintana Roo) haplotypes. Populations often defining the first PC's due to large bottlenecks were removed from the original dataset, i.e. Seri, Lacandon and Tojolabal (Moreno-Estrada et al., 2014). PCA with many related samples or bottlenecked populations affect the first PC's considerably. Individual-specific genetic diversity is overrepresented in these cases, accounting for up to ~94% of human variance (Rosenberg et al., 2002). This individual variance will always outnumber the population-specific variance as population-exclusive alleles are scarce. In order to explore population-specific substructure related samples and heavily bottlenecked populations must be excluded from the PCA. Native American populations with a close geographic location and genetic profile were collapsed into a single population. According to high K admixture analyses (included in lower section of Annexes Figure 8.2), we merged closely-related populations with similar ancestry proportions: two Nahua populations from Puebla, two Zapotec populations and two Mayan populations from the Yucatan Peninsula. In contrast, Nahua people from Jalisco were considered separately as they show a distinctive affinity to nearby Western populations, instead of showing similarities to other Nahua groups in Central Mexico.

Intercontinental admixture in the Native American reference panel has been reported previously and poses a difficulty in the analysis (Rangel-Villalobos et al., 2008). In previous studies, cosmopolitan Mexicans did not completely overlap with Native American populations in an ASPCA (Moreno-Estrada et al., 2014). Even though a genetic differentiation was reported in the Native American component of admixed Mexicans, Central Native Mexican populations showed a considerable unmasked intercontinental admixture that prevented an overlap between admixed and reference individuals. In this study, Native American substructure in cosmopolitan Mexicans accurately matches modern Native American diversity as intercontinental tracts from admixed Native Mexicans were also masked.

Individuals with > 99.9% of Native American global ancestry using Admixture (shown in the upper section of Annexes Figure 8.2), were considered in the reference panel. The rest of Native American individuals were considered as admixed to mask intercontinental admixture. The unadmixed Native American reference panel consisted of 80 individuals or 160 haplotypes: two Tepehuano haplotypes, 24 Huichol haplotypes, 20 Totonac haplotypes, six Mazatec haplotypes, 44 Zapotec haplotypes, 38 Trique haplotypes, 14 Tzotzil haplotypes and two Mayan haplotypes.

- East Asian ASPCA

The ASPCA characterized the specific origin of East Asian ancestry in Mexico. The PCA was performed with a sub-continental reference panel of East and Southeast Asian samples. It included 14 sampled populations classified into nine regions: Japan, Northern China, Southern China, Negrito Philippines, Non-Negrito Philippines, Sumatra, Borneo, Lesser Sunda Islands and Maluku Islands (see Table 8.1 for specific sampling locations).

All samples were collected in four countries: Japan, China, Philippines and Indonesia. Japanese and Chinese samples were obtained from 1000 genomes (Gibbs et al., 2015), while the rest of the samples were genotyped from a previous publication (Reich et al., 2011). Cosmopolitan Mexican individuals with more than 5% of East Asian ancestry were considered, resulting in two Sonoran haplotypes, two Yucatec haplotypes and twelve haplotypes from Guerrero.

3.8 Sub-continental ancestry with a novel method

- New method applied to Native American ancestry

Although the masking of test and reference individuals greatly increased the resolution of sub-continental structure, the employment of new methods allowed a better accuracy and ancestry resolving. The method Multiple Array Ancestry Specific Multidimensional Scaling (MAAS-MDS) appropriately handles missing data, poor SNP intersection between datasets and batch effects (Ioannidis, Bustamante, Feldman, & Moreno-Estrada, 2018), compared to ASPCA. It permitted the inclusion of more comprehensive reference panels and increased genetic markers. A Native American MAAS-MDS was applied to the cosmopolitan Mexican dataset.

- MAAS-MDS allows the inclusion of more reference populations

Central Native Mexican populations have contributed the most to the Native American component of many cosmopolitan Mexicans. The historically large presence of Central indigenous groups and previous ASPCA's suggest this affinity. In this study, we included four additional Nahua populations to better characterize this Native affinity. However, when these new Nahua populations were included, the SNP set intersection for all microarrays resulted in a poor overlap with less than 60,000 markers. Applying the previous ASPCA pipeline would have resulted in unreliable local ancestry assignments and scarce genotyping data for a sub-continental characterization. MAAS-MDS considers high-density local ancestry calls for each microarray separately. High-density local ancestry is achieved by merging each microarray with a continental reference panel from 1000 genomes database.

- Advantages of the MAAS-MDS algorithm

MDS characterizes genetic structure by summarizing genetic distance matrices for each pair of individuals. MAAS-MDS specifically considers the Manhattan distance between individual haplotypes by comparing microarrays in a pairwise manner. In such way, the intersection of a pair of microarrays will have more shared SNPs compared to the intersection of all simultaneous microarrays, allowing better distance estimates. Distances from a pair of microarrays with a limited overlap can be corrected as distances are biased in a linear manner as shown in (Ioannidis et al., 2018). Missing data due to genotyping errors or masked ancestry tracts do not affect the average distance as it only sums non-missing genotypes. If a distance between a pair of individuals cannot be estimated because of a complete lack of overlap between same-ancestry tracts, the distance is inferred by extrapolating the distance with a third individual in a triangular fashion. In summary, MAAS-MDS's approach solves several issues from ASPCA: missing data due to genotyping errors, poor SNP intersections between microarrays, batch effects from masking admixed reference individuals and inaccurate genetic profiles due to very low ancestry proportions.

- Local ancestry provides MAAS-MDS input

Local ancestry analyses were performed separately for three datasets as portrayed in Figure 3.7. All datasets had four continental references for local ancestry calls: sub-Saharan African, European, Native American and East Asian. Reference panel populations varied between local ancestry runs (see Table 3.1 for a summary of each local ancestry run).

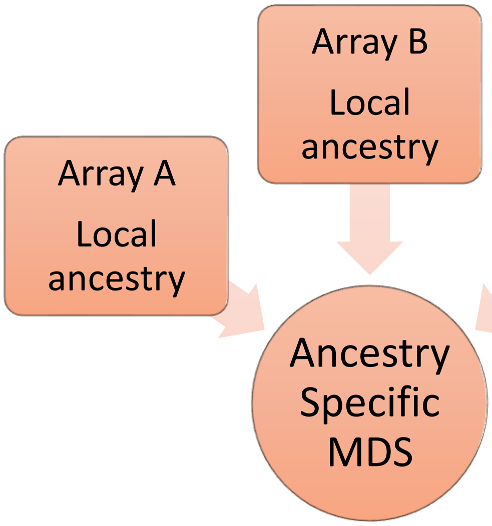
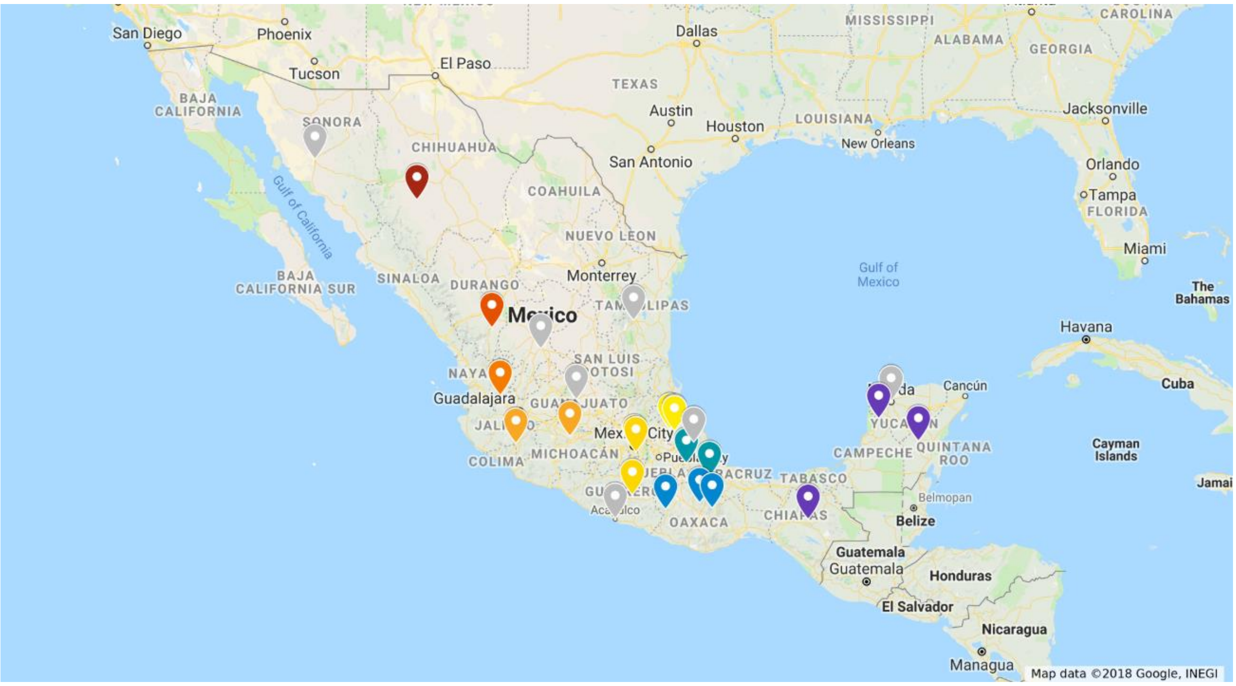


Figure 3.7: Illustration of the ancestry specific MDS's input. Ancestry specific MDS utilize independent local ancestry runs. The local ancestry analyses include the same steps as the workflow in Figure 3.1. Each array was filtered, phased and assigned by continental ancestry separately. Afterwards, local ancestry results were merged and an Ancestry Specific MDS was performed with MAAS-MDS.

Filtering, phasing and local ancestry assignment were performed with the same parameters as previous runs, except for RFMix. EM iterations were set to two, instead of zero, and the -n parameter or minimum number of reference haplotypes per tree node was changed from one to five. The changes allowed EM iterations, which consist of more reliable local ancestry assignments as they are based on previous local ancestry runs. Final local ancestry results were merged with the set union of the SNPs instead of the set intersection as in the previous ASPCA. This generated several missing data as SNPs differed between datasets, which is properly accounted for by MAAS-MDS.

- Description of local ancestry runs per array

The first dataset, array A, consisted of the same seven cosmopolitan Mexican populations. In this case, the Native American reference panel consisted of those three populations from the same microarray: Tepehuano, Northern Zapotec and Maya from Campeche. The second one, array B, represented the twelve Native American populations from NMDP, in such way, intercontinental admixture was masked from the reference panel. To assign Native American ancestry in these Native individuals, four populations were included to represent the genetic diversity from the main geographic regions in Mexico. The Native Mexican component was represented as: Northern Native Mexican with Tarahumara, Central Native Mexican with Totonac, Southern Native Mexican with Zapotec and Southeastern Native Mexican (Mayan) with Tzotzil. These four groups were chosen to have the least intercontinental admixture without considerable bottlenecks. The third panel, array C, included the four newly genotyped Nahua populations from Veracruz, Guerrero and two Mexico City locations. Native American ancestry was assigned with the closest proxy in 1000 genomes to avoid the loss of markers in the merge: Peruvians, the reference panel with the highest Native ancestry proportions (Table 3.1 summarizes the local ancestry runs for Array A, B and C). All populations included for this Native American MAAS-MDS, either as a test or reference population, are shown in the Figure 3.8.



Map data ©2018 Google, INEGI

Figure 3.8: Sampling locations of the populations included in the ancestry specific Native American MAAS-MDS. Cosmopolitan Mexican are shown in gray and Native Mexican with colored pins according to their genetic affinities in Figure 4.6.

4 RESULTS

4.1 Global ancestry with ADMIXTURE

Global ancestry analyses estimate proportions of source populations for each individual. In this thesis, a continental resolution was employed, estimations for European, Native American, African and East Asian affinities were obtained. The classification of these results by Mexican state allows the characterization of ancestry differences in a population-level.

As reported in previous publications, European ancestry is more prevalent in cosmopolitan samples from Northern Mexico, especially in Sonora (61.9% in average). Native American ancestry shows higher proportions in Central and Southern Mexico, with the highest contribution in Guerrero (70.7% in average) according to our populations. Sub-Saharan African ancestry reaches up to 32.3% in individuals from coastal states known for their Afro-Mexican presence (Nieto & Velasquez, 2016), i.e. Veracruz and Guerrero. Novel results in this project were related to East Asian ancestry (see Figure 4.1). East Asian ancestry is estimated as being less than 2% in all cosmopolitan populations. This proportion could represent Native American ancestry misassigned as East Asian as both populations are closely related or a real small contribution. However, some individuals exhibit more than 5% of East Asian global ancestry especially in Guerrero where they reached up to 12.4%. These proportions can only be explained by a relatively recent East Asian immigration. Moreover, a Sonoran and Yucatec individual showed East Asian ancestry greater than 5%. Finally, correlations between ancestries are observed in Guerrero. A positive linear correlation is observed between East Asian and African ancestry, as shown in Figure 4.2.

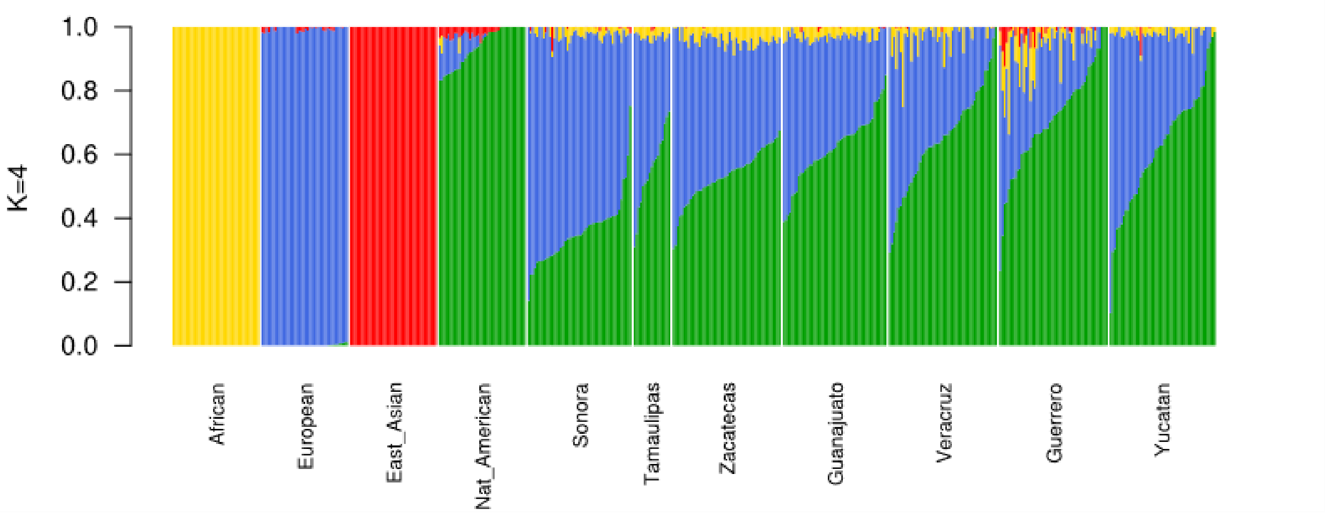


Figure 4.1: Admixture plot with cosmopolitan Mexicans at K4. The ancestral populations coincide with the continental differentiation of the four reference populations included. A proportion of the four components was estimated for each individual.

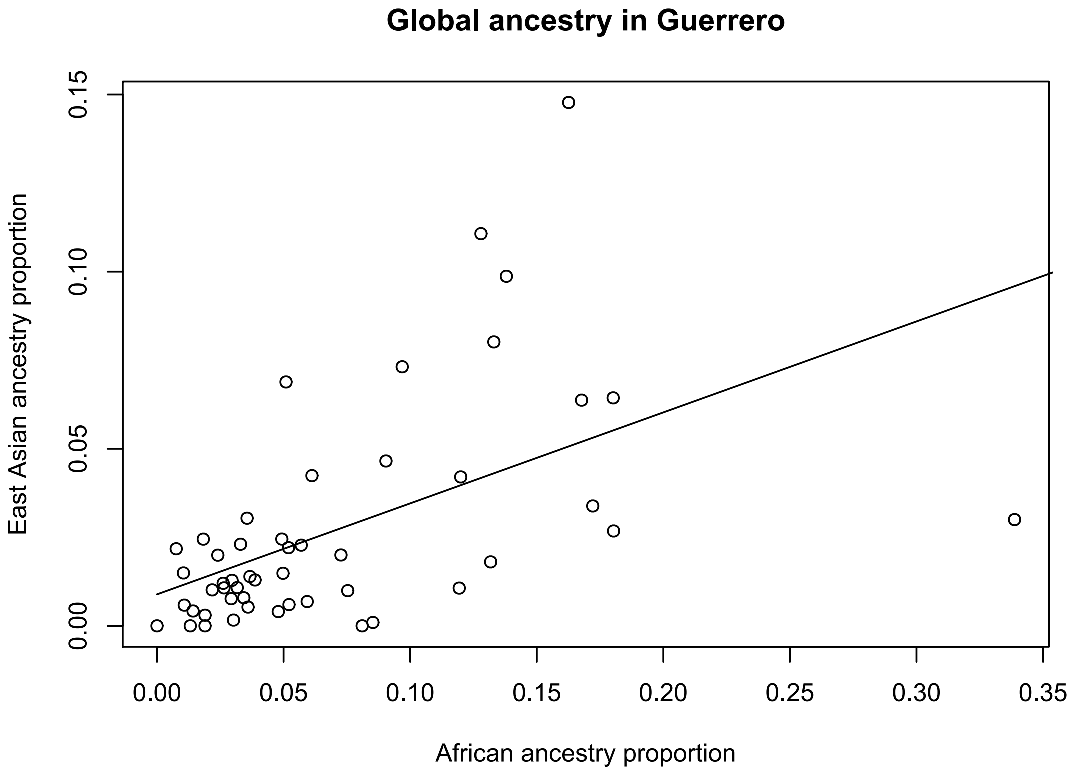


Figure 4.2: Admixture proportions for East Asian and African ancestries in Guerrero. Each point represents an individual plotted by its global ancestry proportions. Sub-saharan African proportions are plotted against East Asian proportions to evidence the linear positive correlation between the two ancestries.

4.2 Simple admixture timings with MALDER

Admixture timings were calculated based on LD patterns. These analyses provide admixture inferences based on genetic data, in order to compare them with historical records.

The earliest admixture timings were reported in the Central states of Guanajuato and Zacatecas, as shown in Table 4.1. The latest events occurred in Sonora, Veracruz and Yucatan. The most common predictions for each pair of admixture sources consisted of Native American and European intermixing. These two components are the most prevalent in cosmopolitan Mexicans. Furthermore, these timings coincide with the admixture timing estimates from Tracts (see the Discussion section).

The oldest admixture timings coincide with cities with the most important mines. These places were known for having an important presence of many ethnic groups as well as admixture between them. The success of mine exploitation attracted several Europeans to manage the sources, meanwhile Africans and Native Americans were brought to work at the mines.

Table 4.1: Admixture timings predicted by MALDER. Estimates are reported in generations for each pair of continental sources with their respective error intervals. A timing prediction was not resolved in all instances.

Population	Admixture event timing in generations		
Mexican state	Native - European	African - Native	African - European
Sonora	9.98 ± 1.35		
Tamaulipas	14.11 ± 2.94		
Zacatecas	12.21 ± 1.73	12.79 ± 1.25	14.93 ± 1.65
Guanajuato	15.27 ± 2.36		
Veracruz	11.96 ± 1.88	12.23 ± 1.60	
Guerrero		12.00 ± 2.59	
Yucatan	8.06 ± 0.61		

4.3 Complex admixture timings with Tracts

Admixture timings were also estimated with Tracts. In contrast to MALDER, this algorithm provides more complex models which are more suitable to the intricate admixture history in Mexico.

The model with the best BIC likelihood in each state was considered. A summary of the results is shown in Figure 4.3. Except for Tamaulipas, the oldest admixture events match the predictions from MALDER. These predictions have older timings as these complex models allow more than one pulse from the same ancestry, while MALDER assumes a single pairwise admixture event to explain the LD patterns. Tracts identified the many admixture events between Native Americans and Europeans that occurred repeatedly across generations. The earliest first timings correspond to Central and Northern Mexican populations, while the most recent first timings are observed in Veracruz and the Southeast with Yucatan.

For instance, all states had two predicted European pulses, while most states also presented a second Native American pulse. Simpler Tracts models with a single European pulse had a worse likelihood and more recent timings, as the model tried to approximate an intermediate timing between the initial and second European pulses. Yucatan was the only state with a second European pulse instead of a second dual pulse of Native American-European ancestry observed in all other states. Second dual pulses had a similar timing between 9 and 10 generations ago in Central and Northern cosmopolitan Mexican populations.

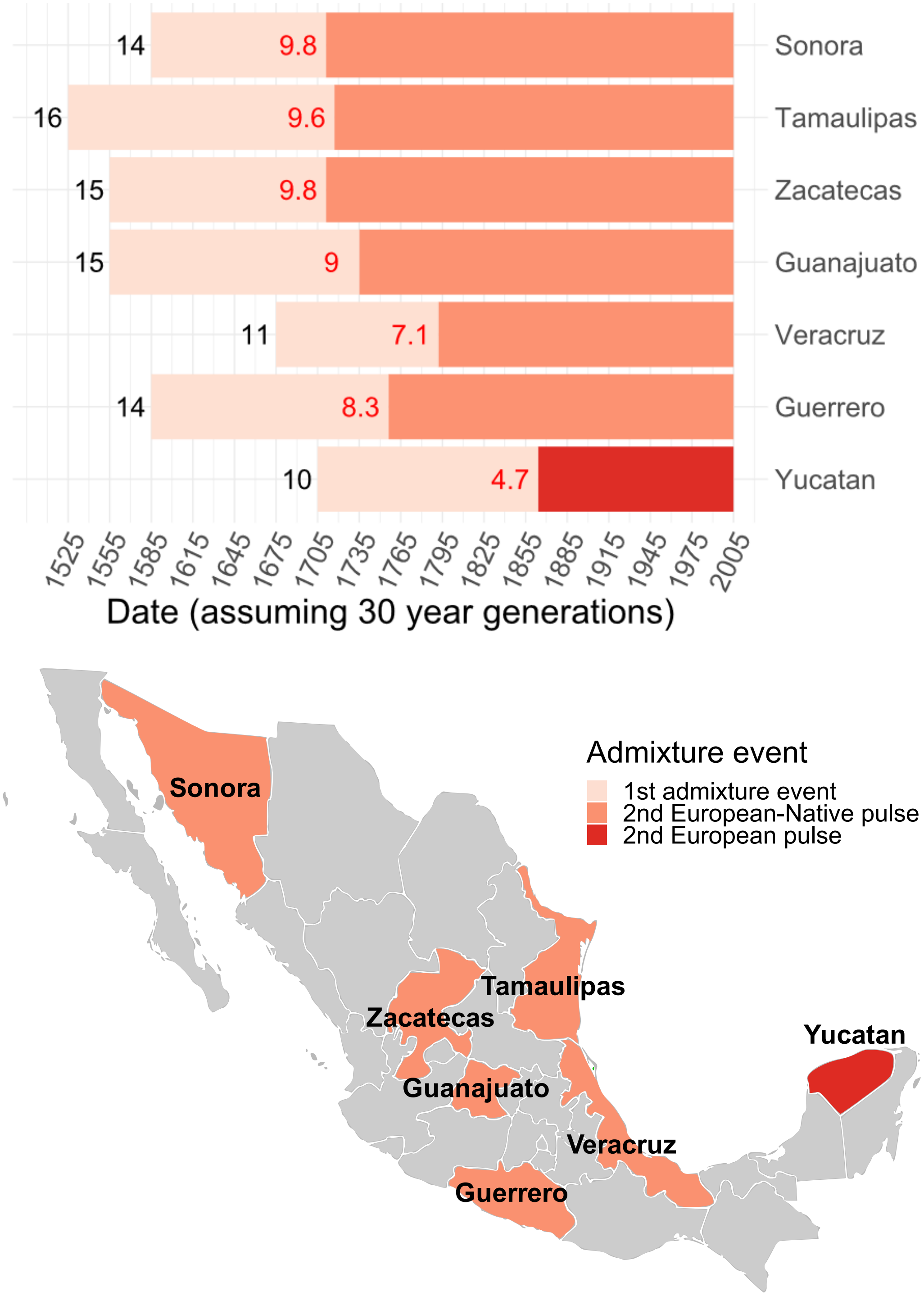


Figure 4.3: Admixture dynamics across Mexico predicted by Tracts. Admixture timings are shown in the upper section. All cosmopolitan Mexican populations exhibited an initial tripartite admixture event succeeded by a second pulse of unadmixed individuals some generations later. The lower section shows the type of second pulse predicted in each state. Most populations had a second dual pulse of Native American/European origin, with only Yucatan showing a better fit with a second European pulse.

4.4 European ASPCA

The European ASPCA shed light on the specific origin of the European ancestry in Mexico. The meticulous comparison of masked haplotypes with a European reference panel, permits a sub-continental identification of this heritage.

The majority of Mexican haplotypes clustered with Iberians (labelled as Southwestern Europeans). The average per state, shown in white labels, does not change in the ASPCA, suggesting an absence of structure in the European component across the country (see Figure 4.4). The homogeneity in the European ancestry contrasts with the substructure found in the Native American component.

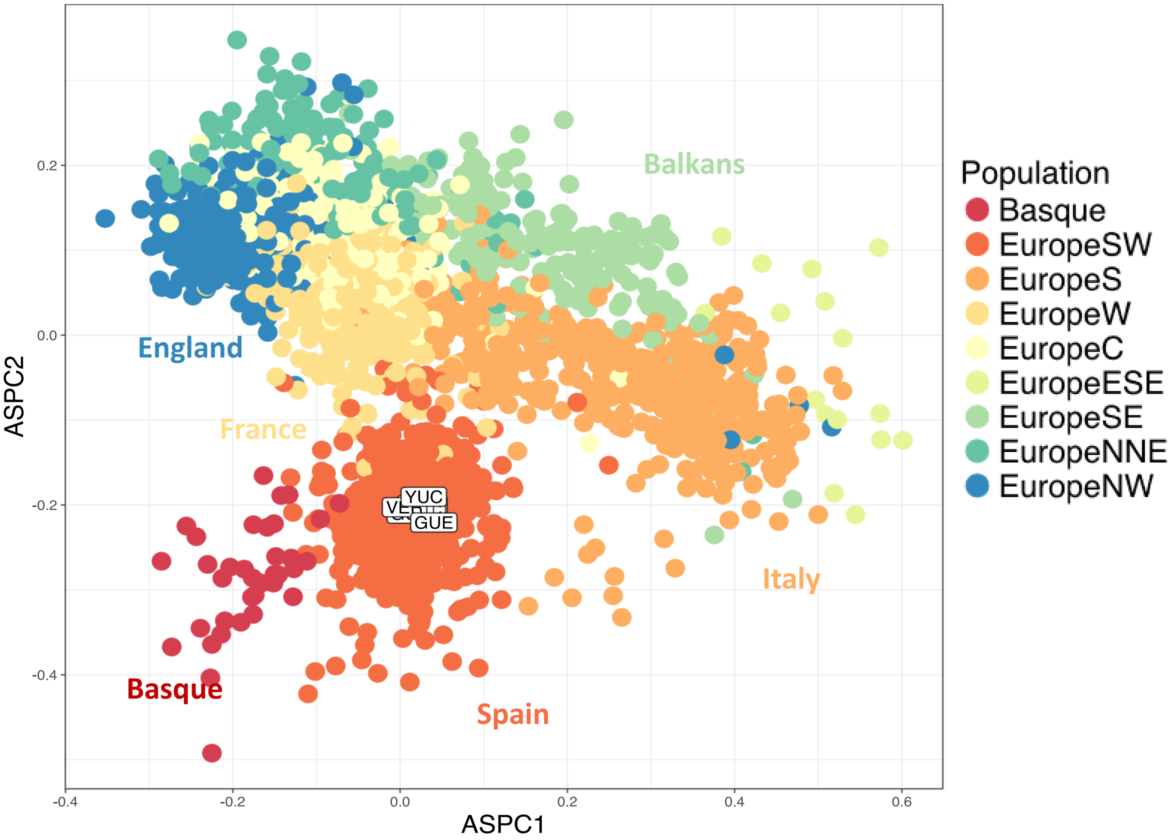


Figure 4.4: European ASPCA showing the average position of each cosmopolitan Mexican population. They cluster with non-Basque Iberian individuals labelled as Europe SW.

4.5 Native American ASPCA and MDS

The study of Native American haplotypes in admixed Mexicans elucidates the precise affinity of this ancestry across Mexican regions. Varied reference panels and methods pinpointed specific native sources in cosmopolitan Mexicans. An ASPCA and MAAS-MDS for Native American haplotypes were applied for this purpose.

The Native ASPCA showed a considerable substructure in the Native American ancestry of cosmopolitan Mexicans. Native American substructure exhibited differences corresponding to geography as previously reported. However, cosmopolitan samples also showed a considerable similarity with modern Native American populations when European and sub-Saharan African ancestries were consistently masked. Most states clustered with Western Native Mexicans or Nahua in Central Mexico, as seen in Figure 4.5.

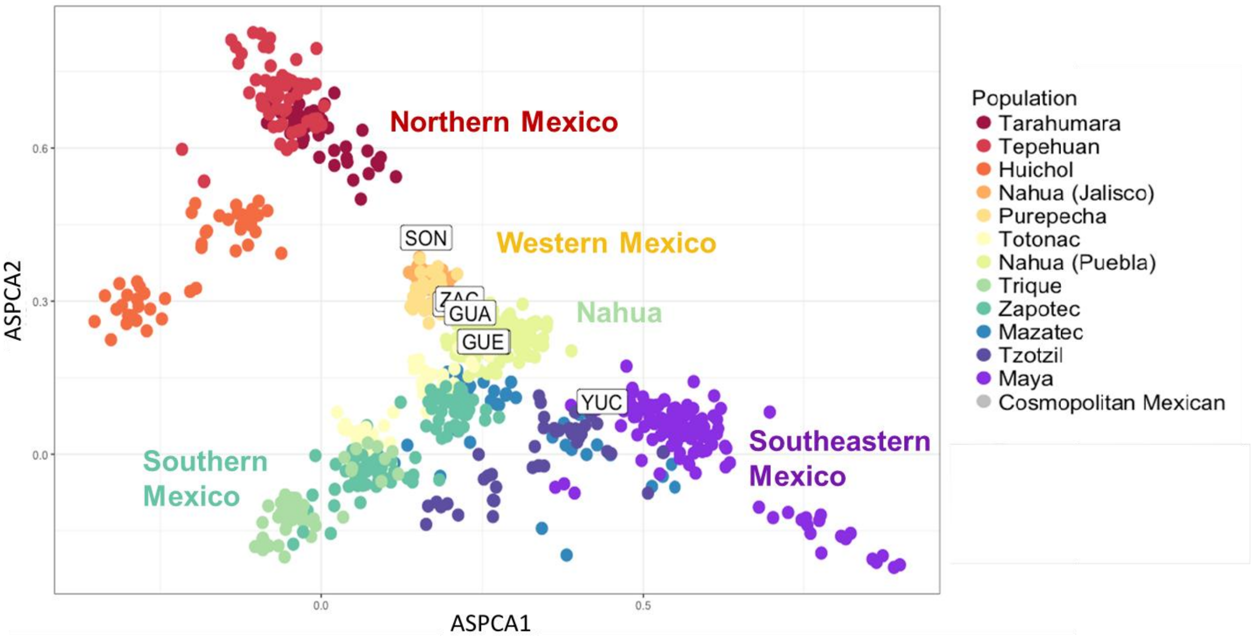


Figure 4.5: Native American ASPCA showing the average position of each cosmopolitan Mexican population. Sonora and Yucatan had the most differentiated heritage portraying a Northern Native Mexican or Mayan affinity, respectively. Most states showed a Western Native Mexican or Central Nahua component. Some substructure is observed as the Central and Southern states of Veracruz and Guerrero have a clear Central Nahua overlap, while states in Aridoamerica show more Western Native Mexican affinity.

The most outer cosmopolitan populations showed distinctive Native heritages. Compared to the rest of Mexicans, Sonora had more affinity to Northern Uto-Aztecan groups, such as the Tepehuano and Tarahumara. Yucatan clustered with modern Mayan populations with whom they share cultural traits in the Southeastern peninsula. However, the Native ASPCA had some difficulties as some reference individuals were admixed. The method generated a batch effect when treating reference panels as admixed in order to mask intercontinental admixture. For example, Tepehuano and Zapotec references group on two different clusters where one consists of the individuals treated as admixed or reference. Ideally, they should cluster in a single group as they belong to the same sampled population.

The Native American MAAS-MDS provided a more comprehensive result as it removed the ASPCA's batch effect and allowed the inclusion of more Nahua populations. The affinities coincide with the Native ASPCA, but they provide a more specific Nahua characterization for Veracruz and Guerrero. They overlap with two clusters: Nahuas from Mexico City and Guerrero and another cluster of Puebla Nahua and Totonacs (see Figure 4.6). In the case of Veracruz, Nahuas from Puebla represent the geographically nearest Nahua population to the sampled city of Xalapa, compared to the Nahua from Veracruz located South, which exhibited a greater affinity with Mazatecs in Oaxaca (locations shown previously in Figure 3.8).

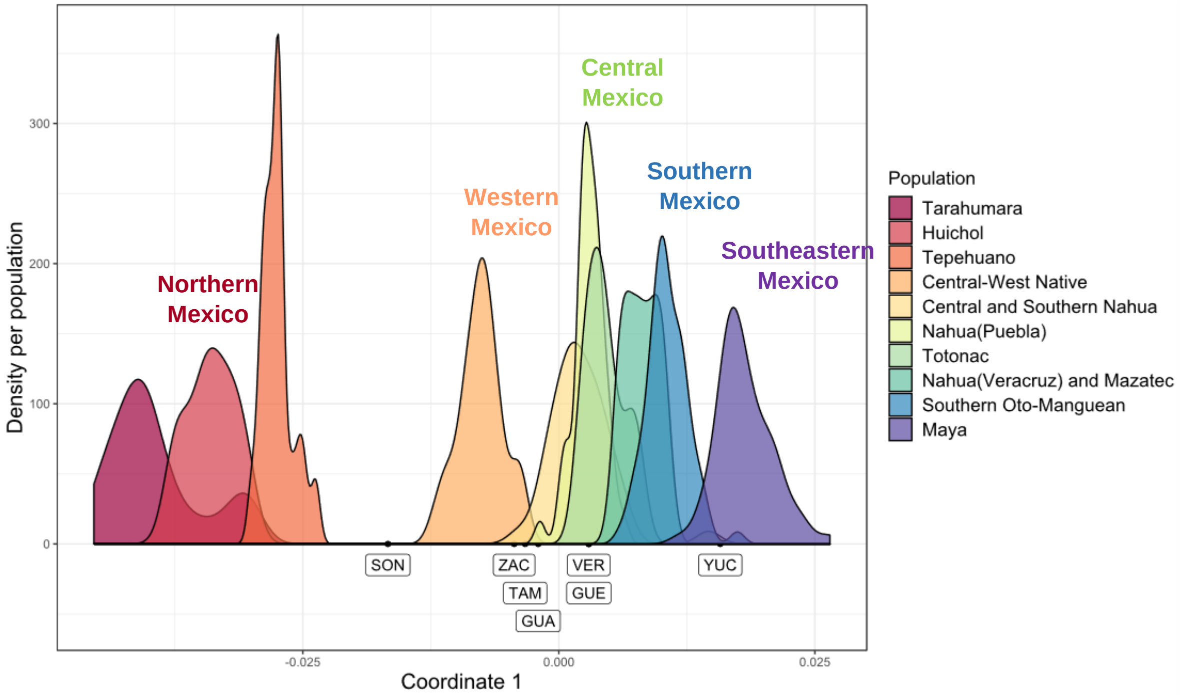


Figure 4.6: Density plot of first coordinate from Native American MAAS-MDS. The average location per cosmopolitan population is shown with points over the X axis and labels below them. The density of the Native American references is shown with colors that moderately coincide with Figure 3.8. Densities per cosmopolitan Mexican population are shown in annexes with Figure 8.3.

4.6 East Asian ASPCA

East Asian ancestry was observed in Guerrero State. In order to characterize the specific source location of the heritage, an East Asian ASPCA was performed.

Cosmopolitan Mexicans with more than 5% of East Asian ancestry were considered in the East Asian ASPCA, resulting in two haplotypes from Sonora, two from Yucatan and twelve from Guerrero. Haplotypes from Sonora and Yucatan overlapped with Southern Chinese populations, while Guerrero had a Filipino and mostly Indonesian affinities. The Indonesian heritage in Guerrero overlaps specifically with Sumatra and Borneo, in contrast to other islands with Melanesian contributions, such as the Lesser Sunda Islands and the Maluku Islands (see Figure 4.7).

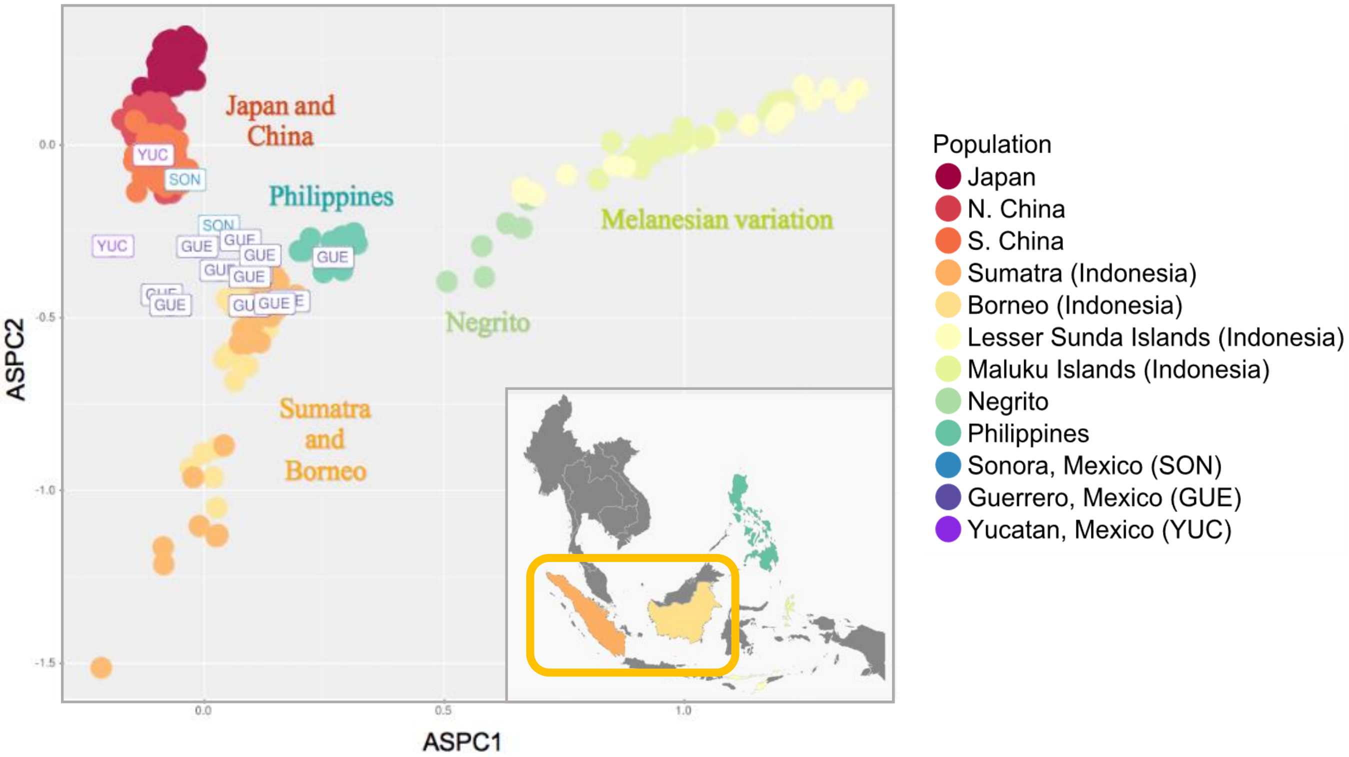


Figure 4.7: East Asian ASPCA showing cosmopolitan Mexican haplotypes. Haplotypes from Sonora and Yucatan cluster with Southern China, while most haplotypes from Guerrero cluster with Southeast Asians. The most recurrent geographic origin of this ancestry in Guerrero is shown with an orange rectangle in the map.

4.7 Gene flow between states with IBD

The average of total shared IBD is shown as a matrix between states. The main diagonal displays the within-population normalized IBD (see Figure 4.8). Estimates represent a general gene flow measure without addressing any timing. In general, geographically close cosmopolitan populations have more shared IBD, while the highest value is always the self-comparison. Sonora is the most isolated state, as it does not share IBD with any other state in the dataset and as it has the highest within-population IBD. Veracruz shows some above-average IBD with Guerrero and Guanajuato, as it is the case of Guanajuato with Tamaulipas.

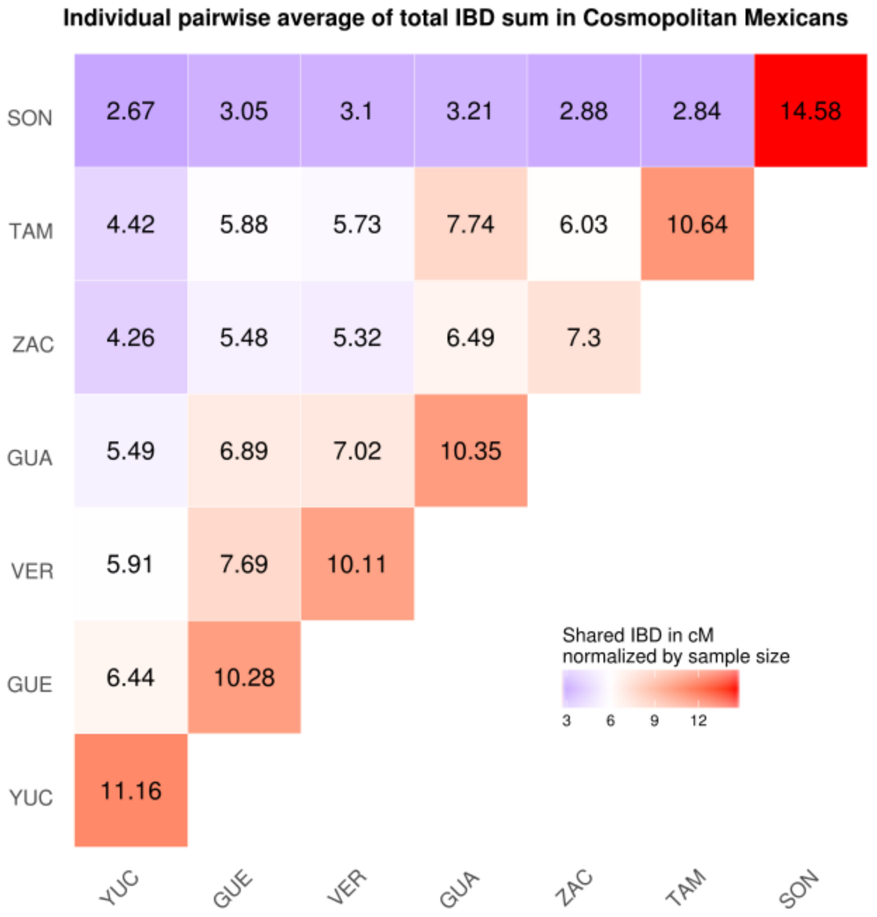


Figure 4.8: Average IBD shared between cosmopolitan Mexican populations. The average of shared IBD sums between a pair of individuals are shown as a matrix, between and within states.

Matrices for ancestry specific IBD were calculated as well, excluding Tamaulipas due to low sample size. Only the Native American and European ASIBD matrices had enough information to illustrate relevant patterns, as seen in Figure 4.9. Regarding Native American ASIBD, Guerrero showed above average affinity with Guanajuato and Veracruz, while Zacatecas showed affinity with Guanajuato. Sonora is clearly isolated in its Native American component to the rest of the country, as is the case of Yucatan in a lesser extent. This matches the Native ASPCA and MAAS-MDS results. The European ASIBD matrix did not show relevant patterns between states. However, it displays Guerrero with a low within-state IBD, in contrast to Sonora with the highest shared IBD.

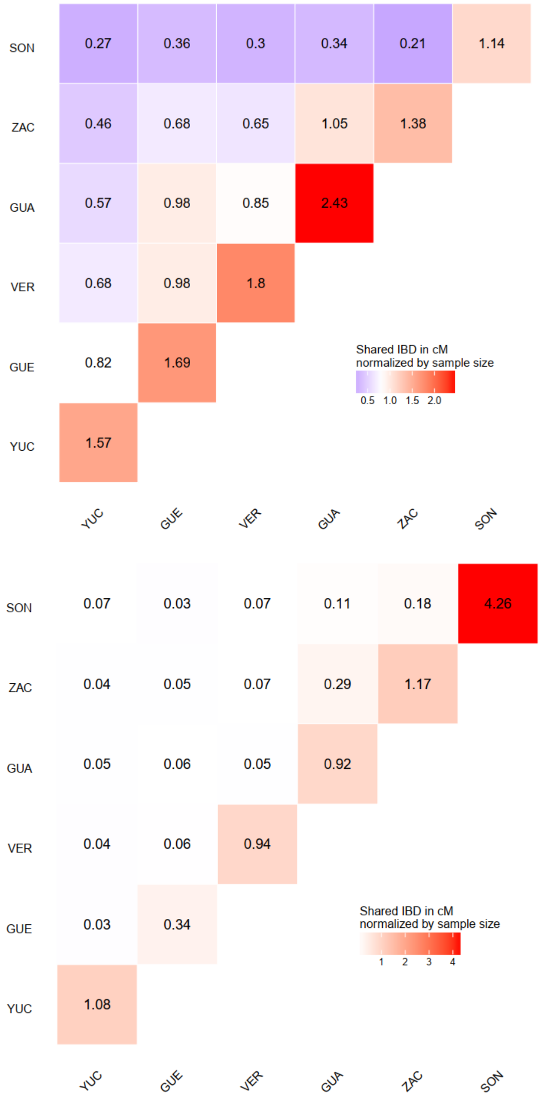


Figure 4.9: Average ancestry specific IBD shared between cosmopolitan Mexican populations. Native American IBD is shown in the left matrix, while European IBD in the right matrix. IBD segments within local ancestry fragments were only considered. IBD sums were normalized by sample size in the same way.

4.8 Effective population size across generations with AS IBDNe

Effective population sizes were estimated in cosmopolitan Mexican metapopulations for each ancestry. Metapopulations were chosen in order to have more than 100 samples per AS IBDNe estimation, a geographical proximity and a

Native American affinity, in order to characterize a shared history. A metapopulation with all seven samples was not included as estimates between ancestries did not differ considerably. Due to a reduced number of African and East Asian IBD fragments, only European and Native American results are shown in Figure 4.10.

Effective population sizes from European IBD differed slightly but always exhibited the same behavior, a constant size followed by an accelerated exponential growth. Predictions from the Native American population size recapitulate bottlenecks or constant population sizes. Subsets including Aridoamerican Mexican states seem to recapitulate a Native American bottleneck 10-15 generations in the past with its lowest point ~12 generations ago (1657 CE). The subset with exclusively Aridoamerican populations (second row from Figure 4.10) showed a reduction in the estimated effective size from ~100,000 to ~31,000 between 12 and 15 generations in the past. Samples from Mesoamerica seem to show a constant size and an exponential growth, as in the European AS IBDNe results. Finally, initial European sizes in Central and Southern Mexico are smaller than Northern Mexico, while Native American sizes in Aridoamerica are slightly lower.

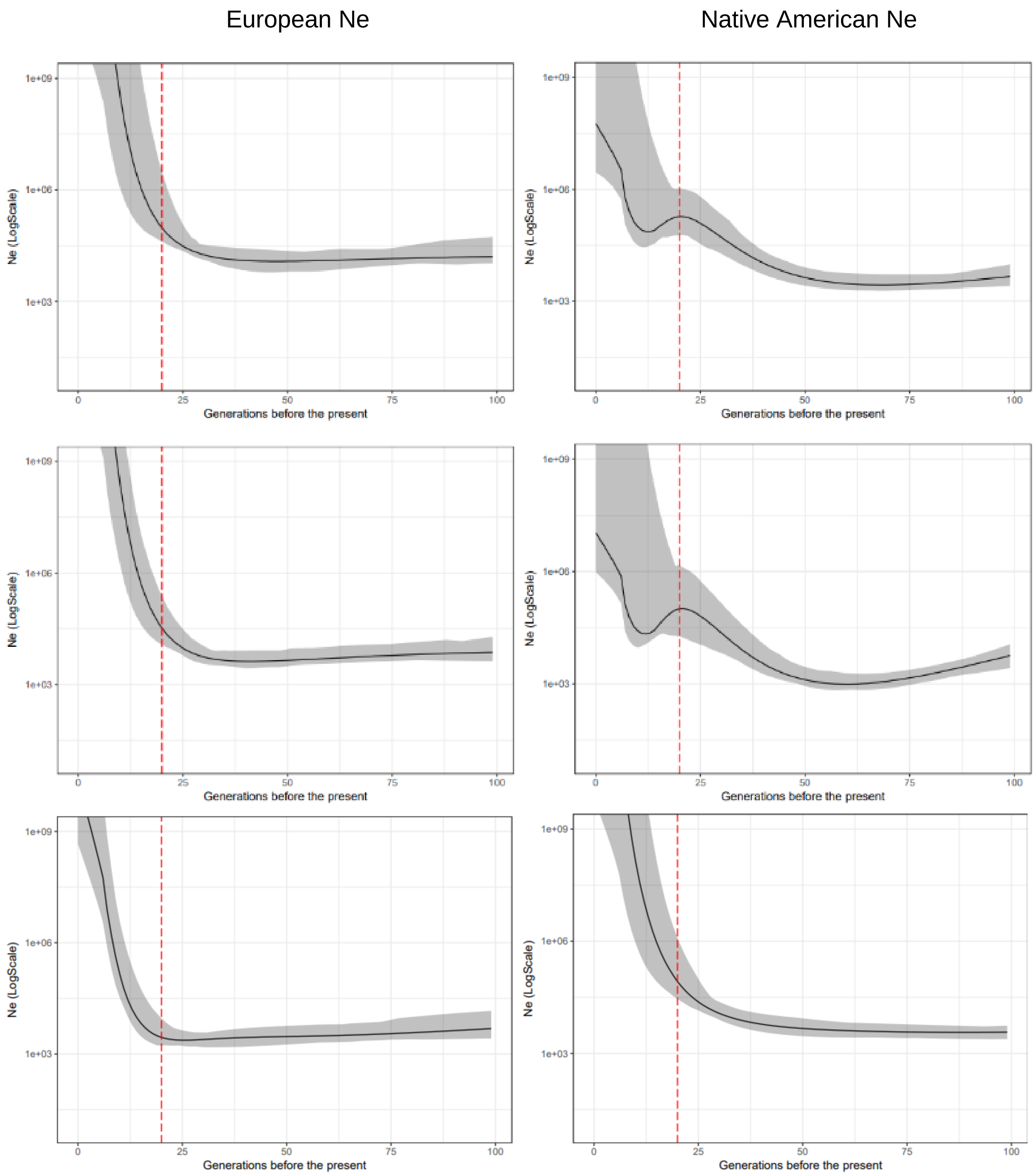


Figure 4.10: Effective population size estimated for each ancestry in cosmopolitan Mexicans. Estimations to 100 generations in the past are provided and 20 generations in the past are indicated with a red dotted line. European (left column) and Native American (right column) ancestry-specific estimates are displayed. The first row corresponds to all states excluding Sonora and Yucatan. Second row corresponds to Tamaulipas, Zacatecas and Guanajuato. The last row includes Veracruz and Guerrero.

5 DISCUSSION

The results presented in this thesis emerged from different patterns of genetic data to provide demographic estimates that can be associated to historical events. Therefore, possible historical interpretations will be proposed by considering all previous analyses together. The following sections will discuss, for instance, the early admixture in Guanajuato City due to the exploitation of the mines, the sub-continental structure from possible post-contact Mesoamerican Native movements into Aridoamerica, the genetic agreement with linguistic clades in Nahua populations, the different Asian immigrations in colonial and post-colonial times, among others. In some cases, interpretations are rather speculative as further analyses must be performed to test the hypotheses directly. Each objective will be addressed corresponding to each section.

5.1 Admixture timings

Admixture occurred at different periods across Mexico. A general pattern is noticed, where Central Mexico shows older admixture timings and Southern-Southeastern Mexico exhibits more recent migration dynamics. Timings agree with important historical events, especially economic activities that attracted people from all ethnic backgrounds.

Table 5.1: Comparison between Tracts and MALDER admixture timings with historical events. All generation times from MALDER and Tracts have been converted assuming 30-year generations and subtracting from 2005 CE, the sampling date. Error intervals were excluded for simplicity. Dates in the last column show the foundation date of the city, where samples were collected. Other important demographic events are shown in parenthesis. The date of Guanajuato shows the onset of the mine exploitation in Guanajuato City. Veracruz shows the increase of trade in Xalapa City that resulted in the immigration of Spanish families and population growth. In Guerrero corresponds to the promotion of Acapulco into a city.

Population	MALDER (Average date of admixture timing)			Tracts		History
Mexican state	Native – European	African - Native	African - European	Native-European first models	Second pulse	Foundation date of city sampled
Sonora	1706 ± 41			1585	1711	1700
Tamaulipas	1582 ± 88			1525	1717	1750
Zacatecas	1639 ± 52	1621 ± 38	1557 ± 50	1555	1711	1546
Guanajuato	1547 ± 71			1555	1735	1548 (1564)
Veracruz	1646 ± 56	1638 ± 48		1675	1792	1521 (1720)
Guerrero		1645 ± 78		1585	1756	1528 (1599)
Yucatan	1763 ± 18			1705	1864	1542

5.1.1 Mine exploitation as the main admixture source in Guanajuato City

Guanajuato City had its first important occupation when the formal exploitation of mines started in 1564 CE. This event led to the substantial encounter and mixing of people from diverse continental origins. They represented European descent people who managed the extracted minerals, and sub-Saharan Africans and Native Americans that forcibly worked in the mines. Admixture in Guanajuato city was unprecedented to that time and our results seem to provide a signal consistent with the mixing of that period, according to LD and complex Tracts models. Furthermore, Native American ancestry in Guanajuato City seems to recapitulate a bimodal affinity (MAAS-MDS from Guanajuato annexes Figure 8.3), possibly related to the main two Native American immigration sources into the state. A Western Native American component brought from Michoacan State with the Purepecha and a more central component from the Valley of Mexico with the Otomi and Nahua peoples during this mining period (Rionda Ramirez, 2002). Admixture during the exploitation of the mines represent the origin of most admixed Mexicans in Guanajuato, and it possibly represented the origin of their Native American substructure. More sophisticated computational tools are needed to address the latter question.

5.1.2 Outer admixture in Tamaulipas

The estimated admixture timing in Tamaulipas is almost two centuries earlier than the city foundation of Ciudad Victoria. This suggests admixture of the sampled individuals occurred previously outside of the city. The immigration of already admixed people would reflect an older estimation. The Western Native Mexican affinity in this Northeastern state could support this hypothesis and provides a hint of a more Central location where admixture could have occurred. However, more analyses remain to be done to verify these claims, as the affinity could be explained by a pre-contact shared substructure from Guanajuato to Tamaulipas or by many possible migration events during the colony.

5.1.3 Recent admixture in Merida

Even though the foundation of Merida took place in the first century of the colonial period, the settlement of the city occurred in an abandoned Mayan city. Admixture could not happen immediately upon Merida’s founding due to the absence of Native peoples. Nevertheless, preliminary analyses from other states suggest recent admixture timings are not exclusive of the Yucatan State. On the contrary, it is a widespread pattern in the south and southeast (see annexes Figure 8.4). A late admixture in the Southeast can be explained by a less severe ruling by the Spanish, as evinced by the cultural persistence of Mayan peoples (Farriss, Setó, & Forstall Comber, 1992).

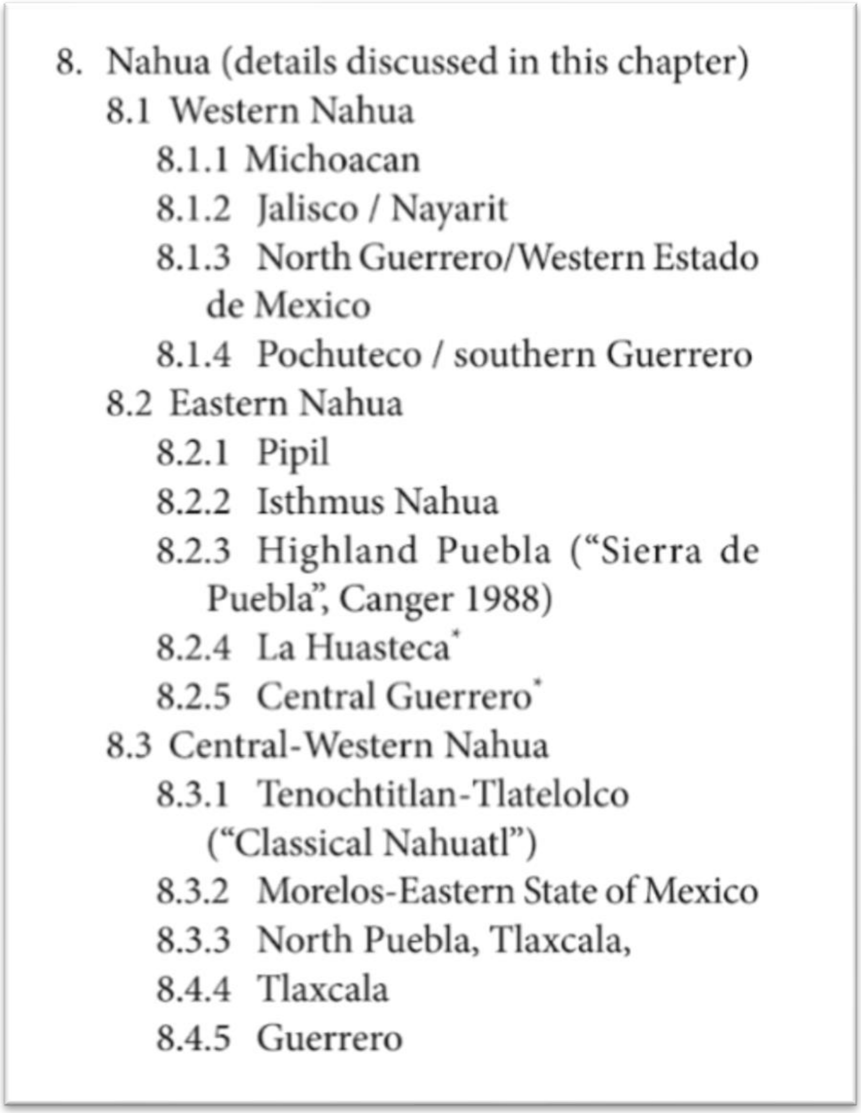
5.2 Substructure in Nahua populations

5.2.1 Genetic and linguistic heterogeneity

In order to compare Nahua populations as a reference panel, we first characterized their genetic substructure. Nahua peoples are ethnically united by their language. However, as linguists have studied dialectal differences and shared language mutations, they have identified three major clades within the Nahuatl language (Dakin & Operstein, 2017). Genetic analyses suggest some coincidences with these Nahua dialects. The native MAAS-MDS reports a close resemblance between Nahua populations from Mexico City and Zitlala (located in Central Guerrero). Both populations resemble the same Nahuatl subdivision: Central-Western Nahuatl (Dakin & Operstein, 2017). Linguistic variants from Jalisco belong to the Western Nahua linguistic clade, while Eastern Nahua are distributed across Puebla and Veracruz (as seen in Figure 5.1).

5.2.2 Genetic affinities from interethnic relationships

Some Nahua populations show genetic affinities with other Native groups, suggesting considerable gene flow with neighboring peoples or a language shift to the colonial lingua franca. Nahua from Puebla show a very close affinity to neighboring Totonacs compared to Nahua from Mexico City, while Nahua from Southern Jalisco are highly divergent from other Nahua, exhibiting similarities with the nearby Purepecha people. The study of the correlation between language and genetics are relevant, as language has been proposed to create genetic barriers. Nevertheless, it does not seem to be the case in Central Mexico, the reduced Fst differentiation between ethnic groups suggests linguistic barriers are not that strong in this Mesoamerican region (according to unpublished data available in the lab). More diverse reference panels from Central Mexico will provide information to identify shared histories and precise profiles between populations. These details would provide complementary evidence for hypotheses regarding language evolution and contact.



8. Nahua (details discussed in this chapter)
    8.1 Western Nahua
        8.1.1 Michoacan
        8.1.2 Jalisco / Nayarit
        8.1.3 North Guerrero/Western Estado de Mexico
        8.1.4 Pochuteco / southern Guerrero
    8.2 Eastern Nahua
        8.2.1 Pipil
        8.2.2 Isthmus Nahua
        8.2.3 Highland Puebla ("Sierra de Puebla", Canger 1988)
        8.2.4 La Huasteca*
        8.2.5 Central Guerrero*
    8.3 Central-Western Nahua
        8.3.1 Tenochtitlan-Tlatelolco ("Classical Nahuatl")
        8.3.2 Morelos-Eastern State of Mexico
        8.3.3 North Puebla, Tlaxcala,
        8.4.4 Tlaxcala
        8.4.5 Guerrero

Figure 5.1: The three main Nahuatl language subclades. All Nahuatl variants are distributed in these three categories. The variants from Jalisco belong to the Western Nahua clade. The variants represented by our sampling locations in Puebla and Veracruz are grouped within the Eastern Nahua branch. Mexico City and Zitlala (Central Guerrero) variants belong or are very influenced by the Central-Western Nahua clade. Figure from (Dakin & Operstein, 2017).

5.3 European, Native American and East Asian substructure in cosmopolitan Mexicans

5.3.1 Iberia as the main European source

Most cosmopolitan individuals cluster with modern Spanish populations. This is expected because most European immigrants were from the Iberian Peninsula as the Mexican territory belonged to the Spanish Empire. No substructure was observed between Mexican states displaying a homogeneous profile, in contrast to the Native American ancestry. Migrations from other European and Mediterranean populations were diluted by the most common non-Basque Iberian heritage due to the genomic approach we employed. Further analyses, perhaps with complete genomes and/or a deeper sampling are required to increase resolution in order to distinguish the components that became undetectable in a predominant Spanish ancestry generalized across the country.

5.3.2 Genetic continuity in Xalapa's Native American ancestry

As evidenced in Figure 4.6, the population from Xalapa possess a mainly Totonac-Pueblan Nahua affinity, which agrees with the closest populations included in our Native American reference panel. Moreover, when Hernan Cortes first arrived to Xalapa in 1521 it was a Totonac village, suggesting a possible Native genetic continuity to present. The specificity of this component is evinced as Nahua populations from the state of Veracruz or Mexico City do not overlap. A pre-contact Totonac genetic continuity in Xalapa remains as a hypothesis, as the underlying substructure between nearby Nahua and non-Nahua populations has not yet been characterized minutely. The similarity could be explained by pre-contact relationships or post-contact dynamics involving recent gene flow or language shift.

5.3.3 The unknown Sonoran Native ancestry

The Native American fraction of Sonora State individuals clusters between Northern and Western Native populations with no clear source population identified (i.e. between the signals of Tepehuano and Nahuas from Jalisco, which have their own genetic signature in contrast to other Nahuas). Possibly the clustering of all individuals from Sonora between these clusters suggest a Native American ancestry from an unsampled population with an intermediate affinity from these two surrounding ethnic groups. To understand the origin of this Native component, we must acknowledge that the current territories of Sinaloa, Sonora and the Baja California Peninsula were a unified region due to their economic organization by the Jesuits during the colony. Native ancestry of cosmopolitan individuals in the region has three candidate sources: the Mesoamerican component from Southern Sinaloa groups in early colonial times which became extinct in the demographic collapse, the many Sonoran Native groups that persist to present-day or an outsider component from Native and admixed individuals, an immigration promoted after the expulsion of the Jesuits in the 18th century (Ortega Noriega, 1985). The latter source is the most feasible according to Tracts results and historical events. Early admixture timings from Tracts between 1585-1615 CE seem to coincide with a signal brought from pre-admixed individuals. Additionally, a local admixture signal could only have been generated until the 18th century, when European presence and native assimilation in Northwestern Mexico became considerable due to the Bourbon Reforms from 1767 CE. Before the reforms, the Jesuit friars were the only Spanish who had contact with the natives, which remained largely isolated from Europeans (Navarrete, 2008). Most native groups in the region were extremely affected by pandemics leading mostly to extinction in Baja California and Sinaloa, while Sonoran Natives persisted longer. Assuming a native heritage from Sonoran Natives, due to the early extinction of other natives, would suppose a much recent admixture timing. Even if this component contributed to the later second dual pulse predicted by Tracts around 1711 CE, it is still not plausible because of the early timing it represents decades before the Bourbon Reforms. Local Sonoran Native contributions by a more recent admixture event are not ruled out, however the algorithm could have ignored this pulse by prioritizing the earliest ones. The origin of the main Native American component in Sonora is still unknown. Here, we propose that admixture first occurred in another place followed by the immigration of admixed peoples carrying this native heritage into Sonora.

5.3.4 Pacification of Northern Mexico: the role of Mesoamerican Natives

At first, Northern Mexico was difficult to occupy by the Spaniards, because of the arid climate, long distances and bellicosity of the hunter-gatherer natives. The conquest of the North was achieved until the Center and South were completely conquered and after the establishment of a stable economy based on the exploitation of silver and gold. The settlement and pacification of the Northern lands were carried out with the help of Mesoamerican Natives, e.g. Tlaxcaltecs, Otomis, Purepechas and Mexicas (Nahua people from the former Aztec Empire) (Lagunas Rodríguez, 2004). The shared affinity between Guanajuato and Northern regions could be explained by this Mesoamerican movement in the pacification process. This is shown with the very close position of cosmopolitan samples from Tamaulipas, Zacatecas and Guanajuato in the Native MAAS-MDS. However, a shared affinity between Mesoamericans and Northern hunter-gatherers is not ruled out as a possible explanation, as the pre-contact genetic profile from Aridoamerica remains uncharacterized. The genotyping or sequencing of modern and ancient Native American groups will shed light in the matter, as many ethnic groups from Northern Mexico have not been analyzed and many more became extinct after European contact.

5.3.5 The Manila Galleon legacy in Guerrero

Southeast Asian ancestry was observed in Mexicans from Guerrero, particularly from the Pacific port of Acapulco, suggesting a genetic remnant from the Manila Galleon, which used Acapulco as the port of disembarkation in Mexico. Limited historical records indicate that the main source of these thousands of chinos was from the Philippines. Genetic results revealed some Filipino affinity, however, most individuals exhibited Western Indonesian ancestry from Borneo and Sumatra. The results pose an unprecedented origin for the numerous Asians that landed in Mexican territory. Moreover, a correlation between East Asian and African ancestries can observed in the global ancestry estimates of the cosmopolitan individuals sampled in Acapulco. This suggests a social cohesion between the two continental groups in Mexico. According to historical records, it could be explained by a shared slave condition, favoring the mating between slaves. Many chino slaves married other slaves, mainly chino, black and mulatto women (Seijas, 2014). These correlations and the discordance with historical records suggest the revealing of an untold history of the Asian slave trade in Mexico, hidden in the genetic footprint of people.

5.3.6 Chinese diaspora in Northern Mexico and Korean immigration in the Yucatan Peninsula

A couple of individuals from other two states were included in the East Asian ASPCA. In contrast to Guerrero, the component had a different affinity. The haplotypes from Sonora and Yucatan clustered with Chinese individuals, instead of Southeast Asians. The origin of this East Asian contribution could be explained with the Chinese and Korean diasporas in post-colonial Mexico. These migrations were promoted during the Porfiriato period, especially between 1880 and 1910 CE. Chinese people worked in agriculture and railroad constructions in the North, while Korean labor was centered in the henequen fields alongside Mayan Natives in the Yucatan Peninsula (Riestra, 1996). Future analyses with larger sample sizes will provide more robust conclusions about the impact of these post-colonial historical migrations in modern Mexican genomes.

5.4 Inferences of population dynamics

5.4.1 The demographic collapse in Aridoamerica

The demographic collapse has been confirmed in all Mexican regions and across the Americas by historical records (Lagunas Rodríguez, 2004), thus it is expected to be reflected in all Native American heritages following an approach like AS IBDNe analysis. Nevertheless, according to our estimations, the bottleneck is only appreciable in the cosmopolitan Mexicans from Aridoamerica (as previously shown in Figure 4.10). In our cosmopolitan populations from Mesoamerica, the Nahua heritage from Veracruz and Guerrero does not exhibit a bottleneck, as reported previously with Mexicans from 1000 genomes database (S. R. Browning et al., 2018). Assuming the accuracy of the estimations, many hypotheses could explain these numbers: a multiethnic origin could have compensated the loss of native genetic diversity; continuous admixture overestimated the effective population size; or the demographic collapse in Mesoamerica was not as severe as historical records suggest. The Aridoamerican bottleneck coincides with the lowest point of the demographic collapse at 1646 CE (Nieto & Velasquez, 2016). However, as genetic bottlenecks can be a result of both a population decline or a founder effect, this Aridoamerican bottleneck could also be explained by a Mesoamerican founder effect related to the pacification process mentioned previously. More comprehensive Native American references from Mexico will clarify these hypotheses. AS IBDNe estimates should be interpreted with caution especially with a reduced sample size. However, higher effective population sizes in the European component of Northern Mexicans and larger sizes in the Native American ancestry of Mesoamerican Mexicans, suggest some agreement with historical records.

5.4.2 Bottleneck in the Sonoran European ancestry

IBD matrices suggest an overrepresentation of shared IBD within Sonora. The high IBD average could be due to a bottleneck or founder effect with a subsequent isolation period. The reduced variance in global ancestry proportions in Sonora supports a separation from incoming migration waves. The ancestry specific analysis suggests a possible founder effect from the European heritage, as the high IBD average is only recapitulated in the European ASIBD matrix in contrast to Native American ASIBD in the Mexican state (Figure 4.9). AS IBDNe failed to identify the timing and magnitude of this possible bottleneck, as only 49 samples were available. Half the minimum sample size recommended for these studies. The finding could have important medical implications as population bottlenecks have been shown to increase genetic drift and higher disease incidences (Risch, Tang, Katzenstein, & Ekstein, 2003).

6 CONCLUSIONS

This thesis demonstrates how genomic analyses can be a powerful tool to reconstruct human evolutionary history. We have applied computational methods to specifically recapitulate the recent population dynamics of the Mexican population dating since European contact. Every region has a particular history that shaped its modern populations. A broad range of results were obtained by re-analyzing previously generated high-density genotyping data from admixed Mexicans and extended reference panels. These shed light on differential admixture timings and dynamics, as well as more precise Native American heritages, past bottlenecks and unknown ancestry contributions.

Sub-continental structure in admixed Mexicans greatly varies in their Native American component, compared to the homogeneous European heritage across the country. Genetic profiles were specifically associated to modern Native Mexicans, providing increased resolution in Central Mexico by incorporating new data from Nahua populations. Moreover, Native American bottlenecks are recapitulated in the ancestry specific component of admixed individuals in some regions. Possibly because of a lower population density, as compared between Aridoamerica and Mesoamerica.

The usual three-way admixture model for Latin American populations seems to be suboptimal in more complex scenarios such as in Guerrero. More than three continental population sources collided and originated the local mixed population, particularly with greater contribution from East Asia. The characterization of this understudied genetic component has very relevant anthropological implications. It represents a historical remnant from the Trans-Pacific trade via the Manila Galleon. The untold birthplace of many Asian slaves was elucidated, whose origins and identities were forgotten due to contraband and assimilation. We confirm the existence of a contemporary heritage and we unraveled its unknown Indonesian origin, largely unregistered in historical records due to illegal slave trade. Probably the remote source is explained by the multiethnicity of the colonial Manila City.

More generally, the study of admixture and genetic ancestry differences has important biomedical implications. Characterizing population-specific patterns will allow the improvement in accuracy of association studies, which can result in personalized medicine applications for disease traits.

7 REFERENCES

Acuna-Soto, R., Stahle, D. W., Cleaveland, M. K., & Therrell, M. D. (2002). Megadrought and megadeath in 16th century Mexico. *Emerging Infectious Diseases*, *8*(4), 360–362. https://doi.org/10.3201/eid0804.010175

Adams, R. E. W., & Macleod, M. J. (2000). THE CAMBRIDGE HISTORY OF THE NATIVE PEOPLES OF THE AMERICAS VOLUME II: MESOAMERICA. *Cambridge University Press.* Retrieved from http://www.cup.org

Berdan, F. F. (2014). Aztec archaeology and ethnohistory. *New York: Cambridge University Press.*

Botigué, L. R., Henn, B. M., Gravel, S., Maples, B. K., Gignoux, C. R., Corona, E., … Bustamante, C. D. (2013). Gene flow from North Africa contributes to differential human genetic diversity in southern Europe. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(29), 11791–11796. https://doi.org/10.1073/pnas.1306223110

Boyd, W. C. (1963). Genetics and the human race. *Science*, *140*(3571), 1057–1064. Retrieved from https://www.jstor.org/stable/1711494

Brandt, G., Szécsényi-Nagy, A., Roth, C., Alt, K. W., & Haak, W. (2015). Human paleogenetics of Europe – The known knowns and the known unknowns. *Journal of Human Evolution*, *79*, 73–92. https://doi.org/10.1016/J.JHEVOL.2014.06.017

Brown, C. H., Clement, C. R., Epps, P., Luedeling, E., & Wichmann, S. (2014). The Paleobiolinguistics of Maize (Zea mays L.). *Ethnobiology Letters*, *5*(0), 52. https://doi.org/10.14237/ebl.5.2014.130

Browning, B. L., & Browning, S. R. (2013). Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics*, *194*(2), 459–471. https://doi.org/10.1534/genetics.113.150029

Browning, S. R., & Browning, B. L. (2015). Accurate Non-parametric Estimation of Recent Effective Population Size from Segments of Identity by Descent. *The American Journal of Human Genetics*, *97*(3), 404–418. https://doi.org/10.1016/J.AJHG.2015.07.012

Browning, S. R., Browning, B. L., Daviglus, M. L., Durazo-Arvizu, R. A., Schneiderman, N., Kaplan, R. C., & Laurie, C. C. (2018). Ancestry-specific recent effective population size in the Americas. *PLOS Genetics*, *14*(5), e1007385. https://doi.org/10.1371/journal.pgen.1007385

Carrillo, R. (2014). Asia llega a América. Migración e influencia cultural asiática en Nueva España (1565-1815). *Asiadémica: Revista Universitaria de Estudios*. Retrieved from https://www.raco.cat/index.php/asiademica/article/download/286846/375066

Cázares, A. C. (2000). El debate sobre la guerra chichimeca, 1531-1585. Volumen II*. El Colegio de Michoacán A.C.* Retrieved from https://books.google.com.mx/books/about/El_debate_sobre_la_guerra_chichimeca_153.html?id=tYVYDULra74C&redir_esc=y

Cerda-Flores, R. M., & Garza-Chapa, R. (1989). Variation in the gene frequencies of three generations of humans from Monterrey, Nuevo León, Mexico. *Human Biology*, *61*(2), 249–261. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2767673

Chacón-Duque, J.-C., Adhikari, K., Fuentes-Guajardo, M., Mendoza-Revilla, J., Acuña-Alonzo, V., Barquera, R., … Ruiz-Linares, A. (2018). Latin Americans show wide-spread Converso ancestry and imprint of local Native ancestry on physical appearance. *Nature Communications*, *9*(1), 5388. https://doi.org/10.1038/s41467-018-07748-z

Chiang, C. W. K., Marcus, J. H., Sidore, C., Biddanda, A., Al-Asadi, H., Zoledziewska, M., … Novembre, J. (2018). Genomic history of the Sardinian population. *Nature Genetics*, *50*(10), 1426–1434. https://doi.org/10.1038/s41588-018-0215-8

Coe, M., & Koontz, R. (2013). Mexico: from the Olmecs to the Aztecs. *University of Michigan.*

Consortium International Human Genome Sequencing. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*(6822), 860–921. https://doi.org/10.1038/35057062

Consortium, T. I. H. 3. (2010). Integrating common and rare genetic variation in diverse human populations. *Nature*, *467*(7311), 52–58. https://doi.org/10.1038/nature09298

Dakin, K., & Operstein, N. (2017). Chapter 1. Language contact in Mesoamerica and beyond. *John Benjamins Publishing Company.* https://doi.org/10.1075/slcs.185.01dak

DeGiorgio, M., Jakobsson, M., & Rosenberg, N. A. (2009). Out of Africa: modern human origins special feature: explaining worldwide patterns of human genetic variation using a coalescent-based serial founder model of migration outward from Africa. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(38), 16057–16062. https://doi.org/10.1073/pnas.0903341106

Delaneau, O., Zagury, J.-F., & Marchini, J. (2013). Improved whole-chromosome phasing for disease and population genetic studies. *Nature Methods*, *10*(1), 5–6. https://doi.org/10.1038/nmeth.2307

Diamond, J., & Bellwood, P. (2003). Farmers and their languages: the first expansions. *Science (New York, N.Y.)*, *300*(5619), 597–603. https://doi.org/10.1126/science.1078208

Diehl, R. A. (2004). The Olmecs : America's first civilization. *Thames & Hudson*. Retrieved from https://books.google.com.mx/books/about/The_Olmecs.html?id=bXd-QgAACAAJ&redir_esc=y

Edwards, A. W. F. (2003). Human genetic diversity: Lewontin's fallacy. *BioEssays*, *25*(8), 798–801. https://doi.org/10.1002/bies.10315

Eltis, D. (2018). A Brief Overview of the Trans-Atlantic Slave Trade, Voyages: The Trans-Atlantic Slave Trade Database. Retrieved June 1, 2019, from https://www.slavevoyages.org/voyage/about

Fagundes, N. N. J. R., Tagliani-Ribeiro, A., Rubicz, R., Tarskaia, L., Crawford, M. M. H., Salzano, F. M., … Bonatto, S. L. (2018). How strong was the bottleneck associated to the peopling of the Americas? New insights from multilocus sequence data. *Genetics and Molecular Biology*, *41*(1 suppl 1), 206–214. https://doi.org/10.1590/1678-4685-

gmb-2017-0087

Farriss, N. M., Setó, J. tr., & Forstall Comber, B. (1992). La sociedad maya bajo el dominio colonialla empresa colectiva de la supervivencia. *Alianza América*. Retrieved from http://www.sidalc.net/cgi-bin/wxis.exe/?IsisScript=sibe01.xis&method=post&formato=2&cantidad=1&expresion=mfn=019597

Felix-López, X. A., Argüello-Garcí-a, R., Cerda-Flores, R. M., Peñaloza-Espinoza, R. I., Buentello-Malo, L., Estrada-Mena, F. J., … Arenas-Aranda, D. J. (2006). FMR1 CGG Repeat Distribution and Linked Microsatellite-SNP Haplotypes in Normal Mexican Mestizo and Indigenous Populations. *Human Biology*, *78*(5), 579–598. https://doi.org/10.1353/hub.2007.0004

Galanter, Joshua M, Gignoux, C. R., Torgerson, D. G., Roth, L. A., Eng, C., Oh, S. S., … Burchard, E. G. (2014). Genome-wide association study and admixture mapping identify different asthma-associated loci in Latinos: the Genes-environments &amp; Admixture in Latino Americans study. *The Journal of Allergy and Clinical Immunology*, *134*(2), 295–305. https://doi.org/10.1016/j.jaci.2013.08.055

Galanter, Joshua Mark, Fernandez-Lopez, J. C., Gignoux, C. R., Barnholtz-Sloan, J., Fernandez-Rozadilla, C., Via, M., … Carracedo, A. (2012). Development of a Panel of Genome-Wide Ancestry Informative Markers to Study Admixture Throughout the Americas. *PLoS Genetics*, *8*(3), e1002554. https://doi.org/10.1371/journal.pgen.1002554

Gibbs, R. A., Boerwinkle, E., Doddapaneni, H., Han, Y., Korchina, V., Kovar, C., … Rasheed, A. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68–74. https://doi.org/10.1038/nature15393

Gorodezky, C., Alaez, C., Vázquez-García, M. N., de la Rosa, G., Infante, E., Balladares, S., … Muñoz, L. (2001). The Genetic structure of Mexican Mestizos of different locations: tracking back their origins through MHC genes, blood group systems, and microsatellites. *Human Immunology*, *62*(9), 979–991. https://doi.org/10.1016/S0198-8859(01)00296-8

Gravel, S., Stephens, M., & Pritchard, J. K. (2012). Population genetics models of local ancestry. *Genetics*, *191*(2), 607–619. https://doi.org/10.1534/genetics.112.139808

Gravel, S., Zakharia, F., Moreno-Estrada, A., Byrnes, J. K., Muzzio, M., Rodriguez-Flores, J. L., … Bustamante, C. D. (2013). Reconstructing Native American Migrations from Whole-Genome and Whole-Exome Data. *PLoS Genetics*, *9*(12), e1004023. https://doi.org/10.1371/journal.pgen.1004023

Guevarra, R. P. (2011). Filipinos in Nueva España: Filipino-Mexican Relations, Mestizaje, and Identity in Colonial and Contemporary Mexico. *Journal of Asian American Studies*, *14*, 389–416. Retrieved from https://muse.jhu.edu/article/456194/summary

Gusev, A., Lowe, J. K., Stoffel, M., Daly, M. J., Altshuler, D., Breslow, J. L., … Pe'er, I. (2009). Whole population, genome-wide mapping of hidden relatedness. *Genome Research*, *19*(2), 318–326. https://doi.org/10.1101/gr.081398.108

Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., … Reich, D. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*, *522*(7555), 207–211.

https://doi.org/10.1038/nature14317

Henn, B. M., Botigué, L. R., Gravel, S., Wang, W., Brisbin, A., Byrnes, J. K., … Comas, D. (2012). Genomic Ancestry of North Africans Supports Back-to-Africa Migrations. *PLoS Genetics*, *8*(1), e1002397. https://doi.org/10.1371/journal.pgen.1002397

Hernández-Gutiérrez, S., Hernández-Franco, P., Martínez-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*(1), 97–100. https://doi.org/10.1016/J.FORSCIINT.2004.09.080

Hordes, S. M. (2005). To the end of the earth : a history of the crypto-Jews of New Mexico. *Columbia University Press.*

Ioannidis, A., Bustamante, C., Feldman, M. W., & Moreno-Estrada, A. (2018). Inverse Problems in the Pacific: Ancestry Deconvolution and Dimensionality Reduction for Reconstructing Human Settlement in Oceania (Stanford University).

Karolchik, D., Baertsch, R., Diekhans, M., Furey, T. S., Hinrichs, A., Lu, Y. T., … University of California Santa Cruz. (2003). The UCSC Genome Browser Database. *Nucleic Acids Research*, *31*(1), 51–54. https://doi.org/10.1093/nar/gkg129

Kosoy, R., Nassir, R., Tian, C., White, P. A., Butler, L. M., Silva, G., … Seldin, M. F. (2009). Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. *Human Mutation*, *30*(1), 69–78. https://doi.org/10.1002/humu.20822

Lagunas Rodríguez, Z. (2004). Población, migración y mestizaje en México: época prehispánica, época actual. *Escuela Nacional de Antropologia e Historia.* Retrieved from https://books.google.com.mx/books/about/Poblacion_migracion_y_mestizaje_en_Mexic.html?id=7wV0ygAACAAJ&redir_esc=y

Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., … Krause, J. (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature*, *513*(7518), 409–413. https://doi.org/10.1038/nature13673

Levy, S. E., & Myers, R. M. (2016). Advancements in Next-Generation Sequencing. *Annual Review of Genomics and Human Genetics*, *17*(1), 95–115. https://doi.org/10.1146/annurev-genom-083115-022413

Licea-Cadena, R. A., Rizzo-Juárez, R. A., Muñiz-Lozano, E., Páez-Riberos, L. A., & Rangel-Villalobos, H. (2006). Population data of nine STRs of Mexican-mestizos from Veracruz (Central South-Eastern, Mexico). *Legal Medicine*, *8*(4), 251–252. https://doi.org/10.1016/J.LEGALMED.2006.04.003

Loh, P.-R., Lipson, M., Patterson, N., Moorjani, P., Pickrell, J. K., Reich, D., & Berger, B. (2013). Inferring admixture histories of human populations using linkage disequilibrium. *Genetics*, *193*(4), 1233–1254. https://doi.org/10.1534/genetics.112.147330

López De Gómara, F. (2007). Historia de la Conquista de México. *Fundación Biblioteca Ayacucho.* Retrieved from https://biblioteca.org.ar/libros/211672.pdf

Luna-Vázquez, A., Vilchis-Dorantes, G., Aguilar-Ruiz, M. O., Bautista-Rivas, A., Pérez-García, A., Orea-Ochoa, R., … Rangel-Villalobos, H. (2008). Haplotype frequencies of the PowerPlex® Y system in a Mexican-Mestizo population sample from Mexico City. *Forensic Science International: Genetics*, *2*(1), e11–e13. https://doi.org/10.1016/J.FSIGEN.2007.08.010

Ma, L. J. C., & Cartier, C. L. (2003). The Chinese diaspora : space, place, mobility, and identity. *Rowman & Littlefield Publishers.* Retrieved from https://books.google.com.mx/books/about/The_Chinese_Diaspora.html?id=Uw_Id2wXjo4C&redir_esc=y

Maples, B. K., Gravel, S., Kenny, E. E., & Bustamante, C. D. (2013). RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. *American Journal of Human Genetics*, *93*(2), 278–288. https://doi.org/10.1016/j.ajhg.2013.06.020

Matson, G. A., Burch, T. A., Polesky, H. F., Swanson, J., Sutton, H. E., & Robinson, A. (1968). Distribution of hereditary factors in the blood of Indians of the Gila river, Arizona. *American Journal of Physical Anthropology*, *29*(3), 311–337. https://doi.org/10.1002/ajpa.1330290308

McDougall, I., Brown, F. H., & Fleagle, J. G. (2005). Stratigraphic placement and age of modern humans from Kibish, Ethiopia. *Nature*, *433*(7027), 733–736. https://doi.org/10.1038/nature03258

Mellars, P. (2006). A new radiocarbon revolution and the dispersal of modern humans in Eurasia. *Nature*, *439*(7079), 931–935. https://doi.org/10.1038/nature04521

Mesa, N. R., Mondragón, M. C., Soto, I. D., Parra, M. V., Duque, C., Ortíz-Barrientos, D., … Ruiz-Linares, A. (2000). Autosomal, mtDNA, and Y-Chromosome Diversity in Amerinds: Pre- and Post-Columbian Patterns of Gene Flow in South America. *The American Journal of Human Genetics*, *67*(5), 1277–1286. https://doi.org/10.1016/S0002-9297(07)62955-3

Moorjani, P., Patterson, N., Hirschhorn, J. N., Keinan, A., Hao, L., Atzmon, G., … Reich, D. (2011). The History of African Gene Flow into Southern Europeans, Levantines, and Jews. *PLoS Genetics*, *7*(4), e1001373. https://doi.org/10.1371/journal.pgen.1001373

Moreno-Estrada, A., Gignoux, C. R., Fernández-López, J. C., Zakharia, F., Sikora, M., Contreras, A. V, … Bustamante, C. D. (2014). Human genetics. The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. *Science (New York, N.Y.)*, *344*(6189), 1280–1285. https://doi.org/10.1126/science.1251688

Moreno-Estrada, A., Gravel, S., Zakharia, F., McCauley, J. L., Byrnes, J. K., Gignoux, C. R., … Bustamante, C. D. (2013). Reconstructing the Population Genetic History of the Caribbean. *PLoS Genetics*, *9*(11), e1003925. https://doi.org/10.1371/journal.pgen.1003925

Moreno-Mayar, J. V., Potter, B. A., Vinner, L., Steinrücken, M., Rasmussen, S., Terhorst, J., … Willerslev, E. (2018). Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. *Nature*, *553*(7687), 203–207. https://doi.org/10.1038/nature25173

Moreno-Mayar, J. V., Vinner, L., de Barros Damgaard, P., de la Fuente, C., Chan, J., Spence, J. P., … Willerslev, E. (2018). Early human dispersals within the Americas. *Science*, *362*(6419), eaav2621. https://doi.org/10.1126/science.aav2621

Mullis, K. B. (1990). The unusual origin of the polymerase chain reaction. *Scientific American*, *262*(4), 56–61, 64–65. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2315679

Navarrete, F. (2008). Pueblos Indígenas del México Contemporáneo. In *Comisión Nacional para el Desarrollo de los Pueblos Indigenas*.

Nelson, M. R., Bryc, K., King, K. S., Indap, A., Boyko, A. R., Novembre, J., … Lai, E. H. (2008). The Population Reference Sample, POPRES: a resource for population, disease, and pharmacological genetics research. *American Journal of Human Genetics*, *83*(3), 347–358. https://doi.org/10.1016/j.ajhg.2008.08.005

Nielsen, R., Akey, J. M., Jakobsson, M., Pritchard, J. K., Tishkoff, S., & Willerslev, E. (2017). Tracing the peopling of the world through genomics. *Nature*, *541*(7637), 302–310. https://doi.org/10.1038/nature21347

Nieto, G. I., & Velasquez, M. (2012). Afrodescendientes en México: una historia de silencio y discriminación. *Consejo Nacional para Prevenir la Discriminación.*

Novembre, J., & Peter, B. M. (2016). Recent advances in the study of fine-scale population structure in humans. *Current Opinion in Genetics & Development*, *41*, 98–105. https://doi.org/10.1016/J.GDE.2016.08.007

Ortega Noriega, S. (1985). Ensayo de periodización sobre la historia socioeconómica del noreste mexicano, siglo XVI a XIX. *Secuencia*, *0*(03), 005. https://doi.org/10.18234/secuencia.v0i03.105

Pool, J. E., & Nielsen, R. (2009). Inference of historical changes in migration rate from the lengths of migrant tracts. *Genetics*, *181*(2), 711–719. https://doi.org/10.1534/genetics.108.098095

Prugnolle, F., Manica, A., & Balloux, F. (2005). Geography predicts neutral genetic diversity of human populations. *Current Biology*, *15*(5), R159–R160. https://doi.org/10.1016/J.CUB.2005.02.038

Raghavan, M., DeGiorgio, M., Albrechtsen, A., Moltke, I., Skoglund, P., Korneliussen, T. S., … Willerslev, E. (2014). The genetic prehistory of the New World Arctic. *Science*, *345*(6200), 1255832. https://doi.org/10.1126/science.1255832

Raghavan, M., Skoglund, P., Graf, K. E., Metspalu, M., Albrechtsen, A., Moltke, I., … Willerslev, E. (2014). Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature*, *505*(7481), 87–91. https://doi.org/10.1038/nature12736

Rangel-Villalobos, H., Muñoz-Valle, J. F., González-Martín, A., Gorostiza, A., Magaña, M. T., & Páez-Riberos, L. A. (2008). Genetic admixture, relatedness, and structure patterns among Mexican populations revealed by the Y-chromosome. *American Journal of Physical Anthropology*, *135*(4), 448–461. https://doi.org/10.1002/ajpa.20765

Reich, D., Patterson, N., Campbell, D., Tandon, A., Mazieres, S., Ray, N., … Ruiz-Linares, A. (2012). Reconstructing Native American population history. *Nature*, *488*(7411), 370–374. https://doi.org/10.1038/nature11258

Reich, D., Patterson, N., Kircher, M., Delfin, F., Nandineni, M. R., Pugach, I., … Stoneking, M. (2011). Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. *American Journal of Human Genetics*, *89*(4), 516–528.

https://doi.org/10.1016/j.ajhg.2011.09.005

Riestra, R. A. (1996). Arquitectura de las Haciendas Henequeneras. *Universidad Autónoma de Yucatán.* Retrieved from https://www.worldcat.org/title/arquitectura-de-las-haciendas-henequeneras/oclc/651464506

Rionda Ramirez, J. I. (2002). Historia demográfica de Guanajuato Periodo precolombino y siglos XVI al XX. *Centro de Investigaciones Humanísticas de la Universidad de Guanajuato*. Retrieved from https://books.google.com.mx/books?id=Jb7kIXxFp-UC&pg=RA1-PA20&source=gbs_selected_pages&cad=3#v=onepage&q&f=false

Risch, N., Tang, H., Katzenstein, H., & Ekstein, J. (2003). Geographic distribution of disease mutations in the Ashkenazi Jewish population supports genetic drift over selection. *American Journal of Human Genetics*, *72*(4), 812–822. https://doi.org/10.1086/373882

Rito, T., Vieira, D., Silva, M., Conde-Sousa, E., Pereira, L., Mellars, P., … Soares, P. (2019). A dispersal of Homo sapiens from southern to eastern Africa immediately preceded the out-of-Africa migration. *Scientific Reports*, *9*(1), 4728. https://doi.org/10.1038/s41598-019-41176-3

Rosenberg, N. A., Pritchard, J. K., Weber, J. L., Cann, H. M., Kidd, K. K., Zhivotovsky, L. A., … Cavalli-Sforza, L. L. (2002). Genetic structure of human populations. *Science (New York, N.Y.)*, *298*(5602), 2381–2385. https://doi.org/10.1126/science.1078311

Ruiz-Linares, A., Adhikari, K., Acuña-Alonzo, V., Quinto-Sanchez, M., Jaramillo, C., Arias, W., … Gonzalez-José, R. (2014). Admixture in Latin America: Geographic Structure, Phenotypic Diversity and Self-Perception of Ancestry Based on 7,342 Individuals. *PLoS Genetics*, *10*(9), e1004572. https://doi.org/10.1371/journal.pgen.1004572

Salazar-Flores, J., Dondiego-Aldape, R., Rubi-Castellanos, R., Anaya-Palafox, M., Nuño-Arana, I., Canseco-Ávila, 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*(3), 401–409. https://doi.org/10.1002/ajhb.21013

Seguin-Orlando, A., Korneliussen, T. S., Sikora, M., Malaspinas, A.-S., Manica, A., Moltke, I., … Willerslev, E. (2014). Paleogenomics. Genomic structure in Europeans dating back at least 36,200 years. *Science*, *346*(6213), 1113–1118. https://doi.org/10.1126/science.aaa0114

Seijas, T. (2014). Asian Slaves in Colonial Mexico: From Chinos to Indians. *Cambridge University Press.*

Silva-Zolezzi, I., Hidalgo-Miranda, A., Estrada-Gil, J., Fernandez-Lopez, J. C., Uribe-Figueroa, L., Contreras, A., … Jimenez-Sanchez, G. (2009). Analysis of genomic diversity in Mexican Mestizo populations to develop genomic medicine in Mexico. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(21), 8611–8616. https://doi.org/10.1073/pnas.0903045106

Spores, R. (1993). Tututepec: A Postclassic-period Mixtec conquest state. *Cambridge University Press*, *4*, 167–174. Retrieved from https://www.cambridge.org/core/journals/ancient-mesoamerica/article/tututepec/422B4FBEAFB7201DE3B655701FD45E27

Steckel, R. H., & Rose, J. C. (2002). *The backbone of history : health and nutrition in the Western Hemisphere*. Cambridge University Press.

Tremblay, M., & Vézina, H. (2000). New estimates of intergenerational time intervals for the calculation of age and origins of mutations. *American Journal of Human Genetics*, *66*(2), 651–658. https://doi.org/10.1086/302770

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., … Zhu, X. (2001). The Sequence of the Human Genome. *Science* (Vol. 291). Retrieved from http://science.sciencemag.org/

Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K., & Wilson, A. (1991). African populations and the evolution of human mitochondrial DNA. *Science*, *253*(5027), 1503–1507. https://doi.org/10.1126/SCIENCE.1840702

Wallace, D. C., Garrison, K., & Knowler, W. C. (1985). Dramatic founder effects in Amerindian mitochondrial DNAs. *American Journal of Physical Anthropology*, *68*(2), 149–155. https://doi.org/10.1002/ajpa.1330680202

Wang, S., Lewis, C. M., Jakobsson, M., Ramachandran, S., Ray, N., Bedoya, G., … Ruiz-Linares, A. (2007). Genetic Variation and Population Structure in Native Americans. *PLoS Genetics*, *3*(11), e185. https://doi.org/10.1371/journal.pgen.0030185

Wang, S., Ray, N., Rojas, W., Parra, M. V., Bedoya, G., Gallo, C., … Ruiz-Linares, A. (2008). Geographic Patterns of Genome Admixture in Latin American Mestizos. *PLoS Genetics*, *4*(3), e1000037. https://doi.org/10.1371/journal.pgen.1000037

Watson, J. D., & Crick, F. H. C. (1953). Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature*, *171*(4356), 737–738. https://doi.org/10.1038/171737a0

Wauchope, R., Ekholm, G. F., & Bernal, I. (1971). *Handbook of Middle American Indians. Volume ten, volume eleven, Archaeology of Northern Mesoamerica*. Austin: University of Texas Press.

White, T. D., Asfaw, B., DeGusta, D., Gilbert, H., Richards, G. D., Suwa, G., & Clark Howell, F. (2003). Pleistocene Homo sapiens from Middle Awash, Ethiopia. *Nature*, *423*(6941), 742–747. https://doi.org/10.1038/nature01669

8 ANNEXES

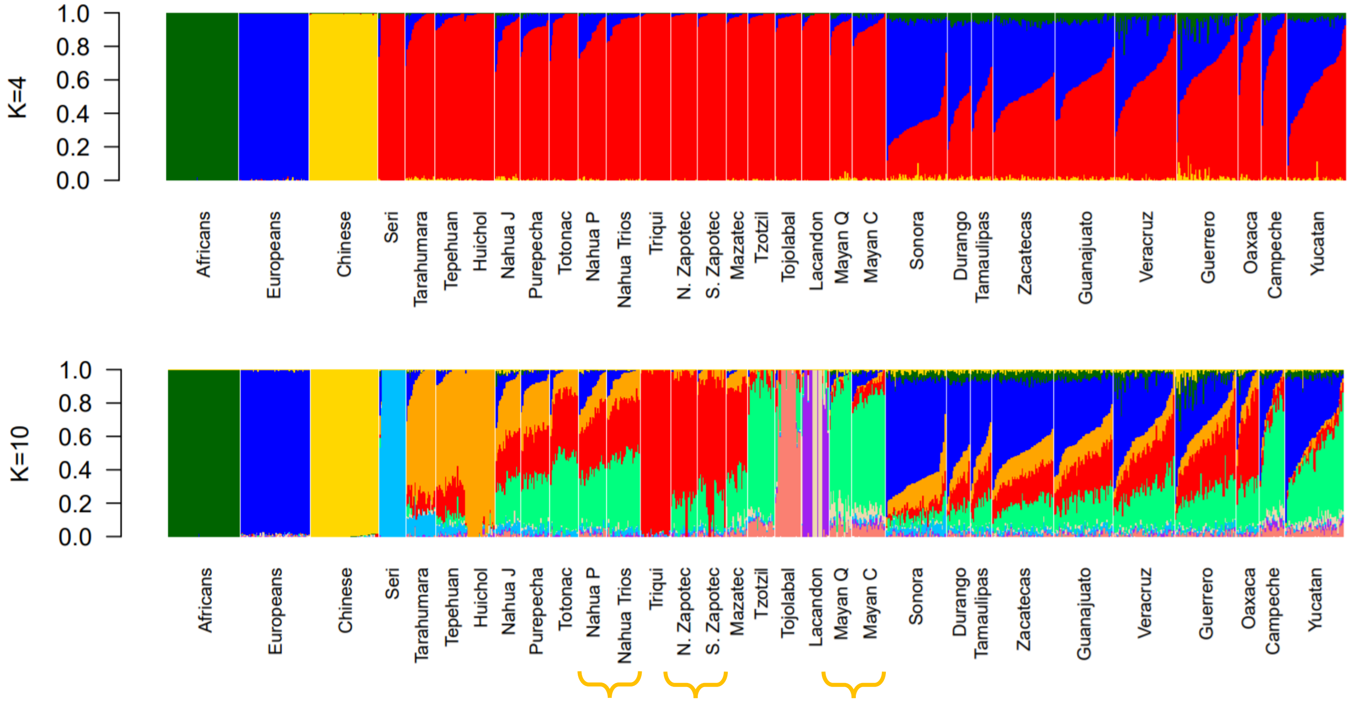
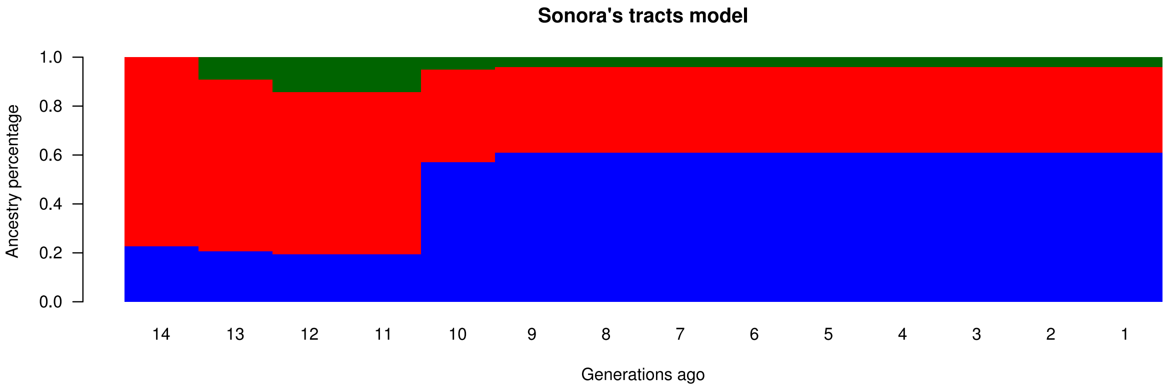
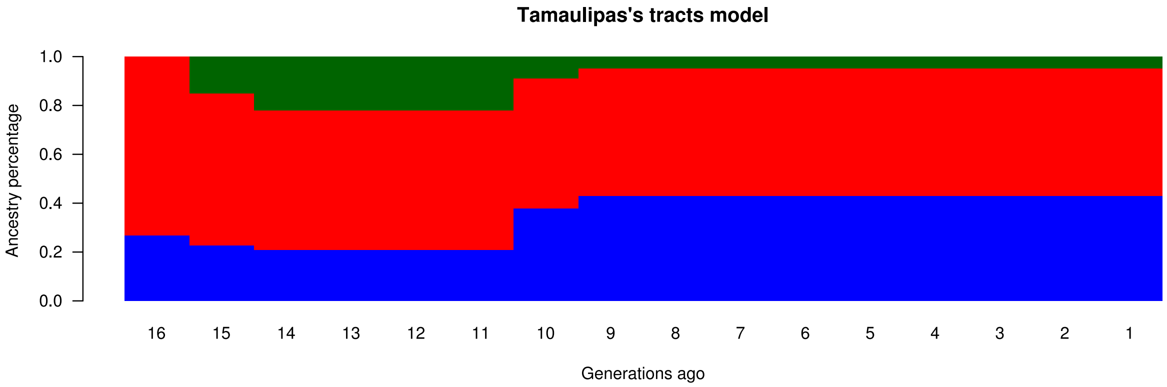


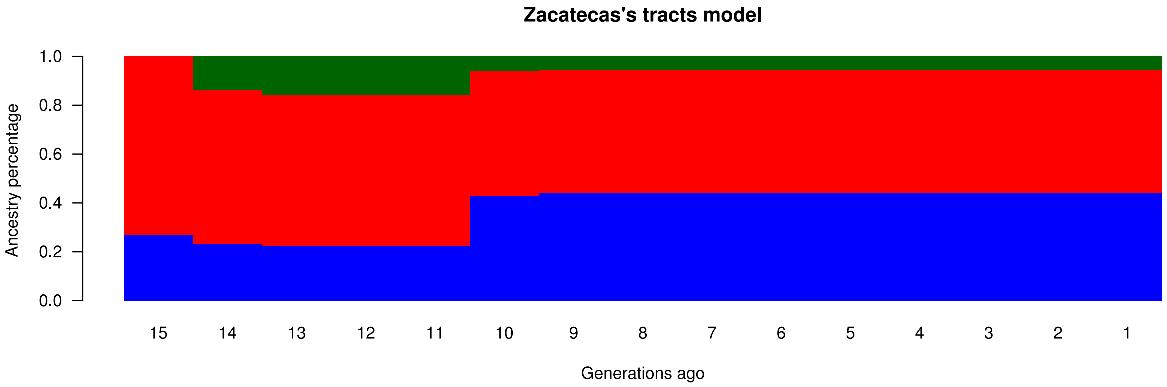
Figure 8.1: Admixture with ten cosmopolitan Mexican populations, four continental reference panels and all Native Mexicans from NMDP. K4 showed a continental resolution in the upper section. K10 showed the lowest cv-error and identifies Native American substructure in the reference panel. Components specific from bottlenecked populations are present in Seri, Huichol, Trique, Tojolabal and Lacandon. The rest of Native groups exhibit more gene flow and can be grouped by their similar profiles. Four main groups are identified: Northwest Natives, Central Natives, Southern Natives and Southeastern Natives. The Native substructure is recapitulated in the cosmopolitan samples, with a Northwest affinity in Sonora and a Southeastern profile in Campeche and Yucatan. Orange braces in the bottom correspond to the merged population categories as they portrayed genetic affinities at K=10 and belonged to the same ethnicity.



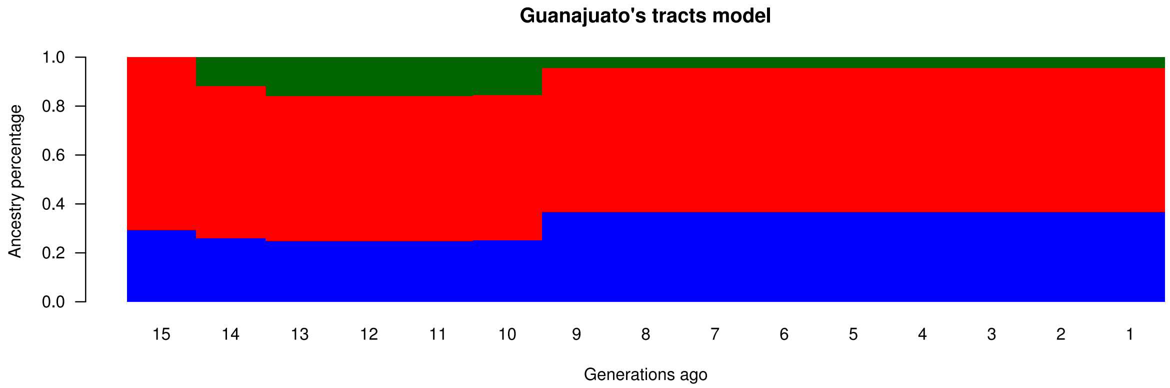
Sonora's tracts model

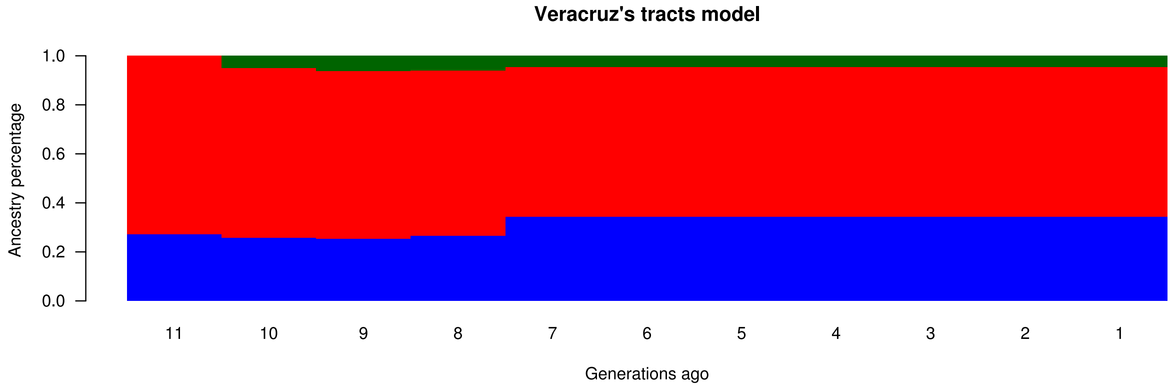


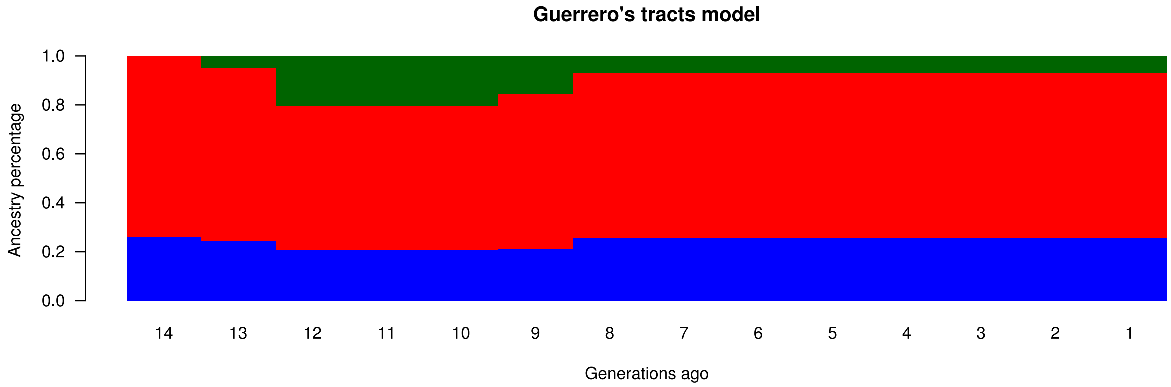
Tamaulipas's tracts model



Zacatecas's tracts model


Guanajuato's tracts model
Ancestry percentage
1.0
0.8
0.6
0.4
0.2
0.0
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
Generations ago


Veracruz's tracts model
Ancestry percentage
1.0
0.8
0.6
0.4
0.2
0.0
11
10
9
8
7
6
5
4
3
2
1
Generations ago


Guerrero's tracts model
Ancestry percentage
1.0
0.8
0.6
0.4
0.2
0.0
14
13
12
11
10
9
8
7
6
5
4
3
2
1
Generations ago

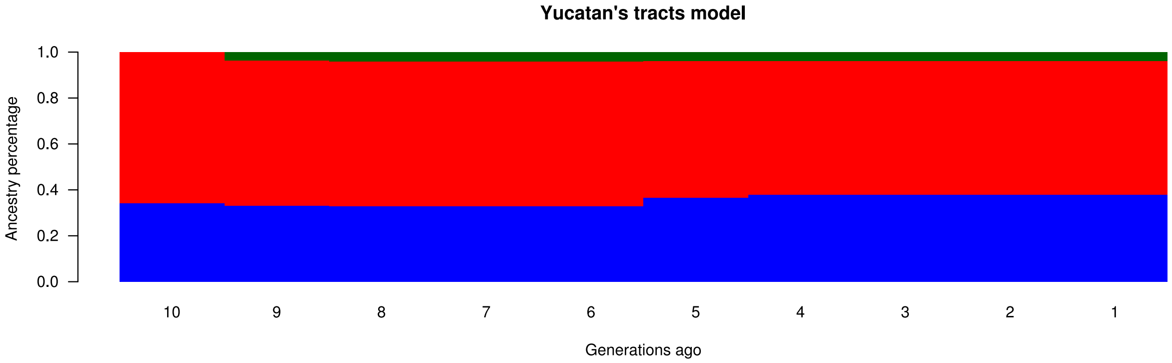


Figure 8.2: Admixture and ancestry proportions across generations. Estimations predicted by the best-likelihood Tracts models for each Mexican state. Dark green represents African ancestry proportions, red, Native, and blue, European.

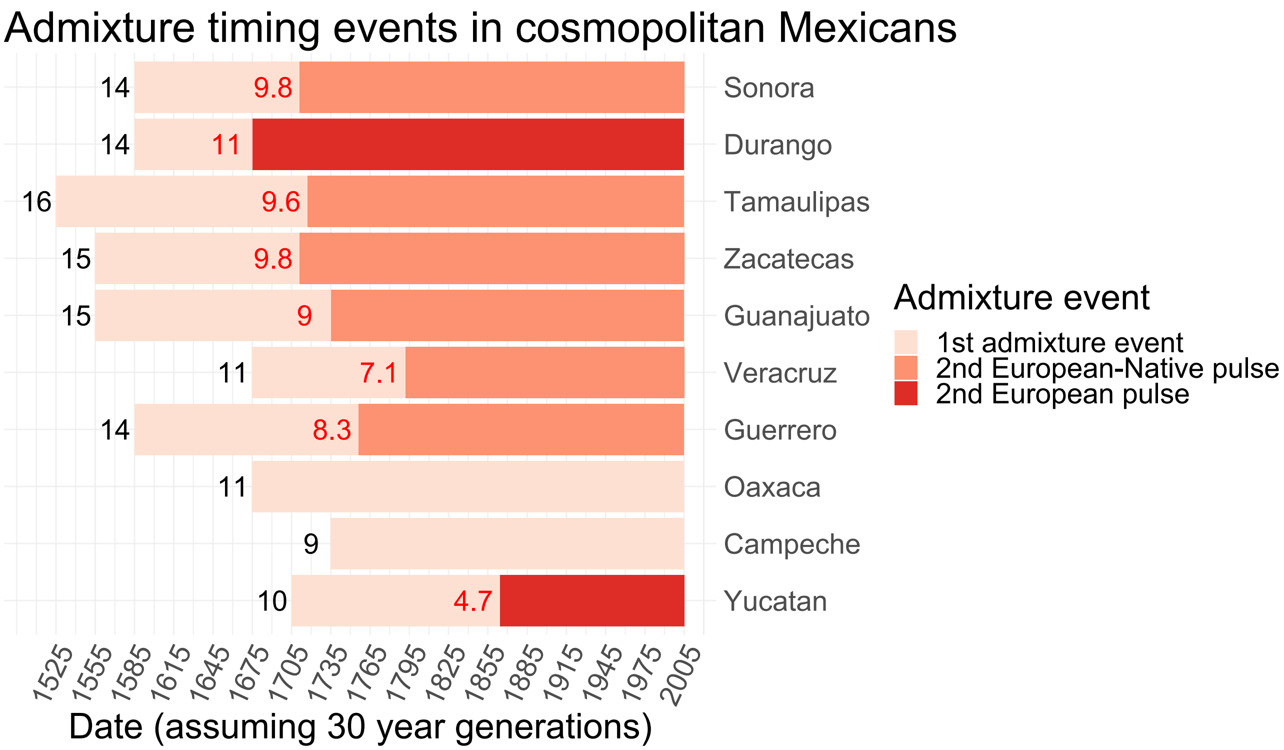
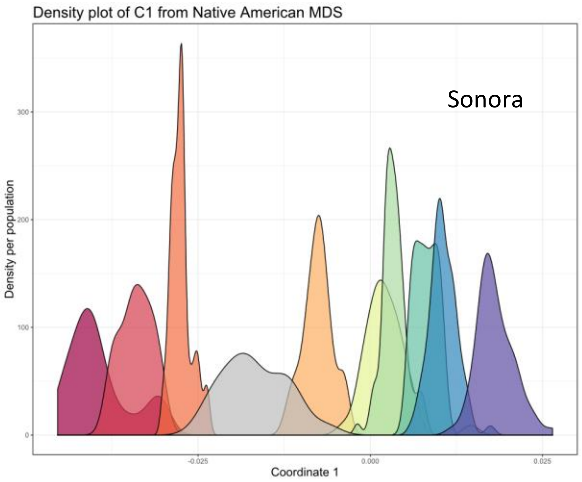
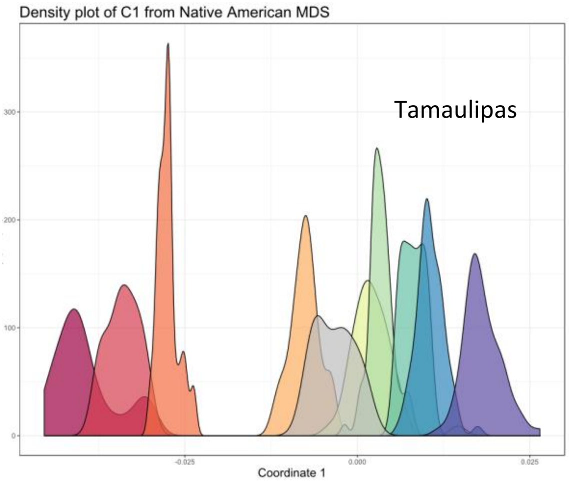
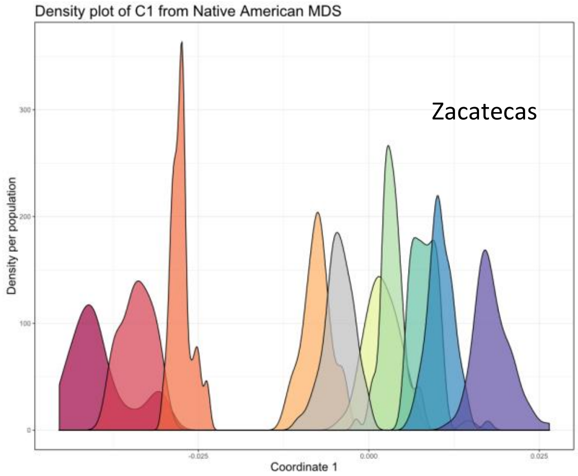
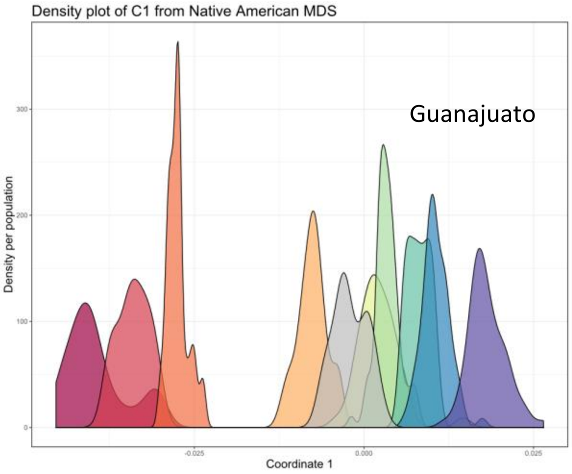


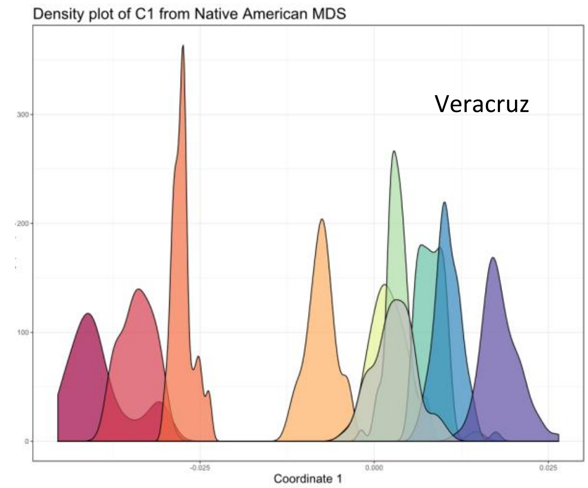
Figure 8.3: Admixture timings are shown in ten cosmopolitan Mexican populations. All cosmopolitan Mexican populations exhibited an initial tripartite admixture event followed by an optional second pulse of unadmixed individuals generations later. More recent admixture timings are observed in the Southeast and neighboring populations. Black numbers show the timing of the initial tripartite admixture event in generations in the past, while red numbers represent second pulse timings.

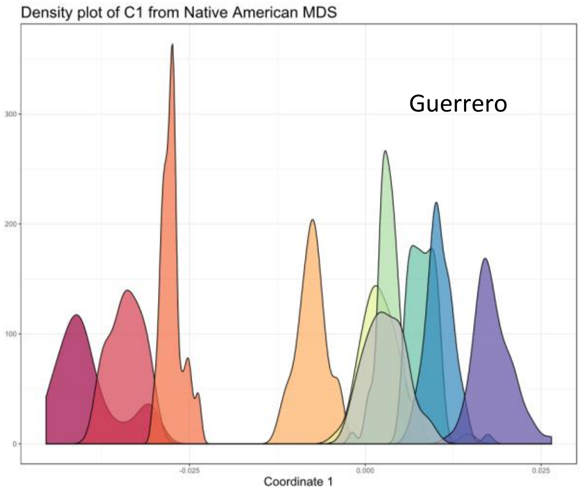

Density plot of C1 from Native American MDS
Sonora
Density per population
300
200
100
0
-0.025
0.000
0.025
Coordinate 1


Density plot of C1 from Native American MDS
Tamaulipas
300
200
100
0
-0.025
0.000
0.025
Coordinate 1


Density plot of C1 from Native American MDS
Zacatecas
Density per population
300
200
100
0
-0.025
0.000
0.025
Coordinate 1


Density plot of C1 from Native American MDS
Guanajuato
Density per population
300
200
100
0
-0.025
0.000
0.025
Coordinate 1


Density plot of C1 from Native American MDS
Veracruz
300
200
100
0
-0.025
0.000
0.025
Coordinate 1


Density plot of C1 from Native American MDS
Guerrero
300
200
100
0
-0.025
0.000
0.025
Coordinate 1

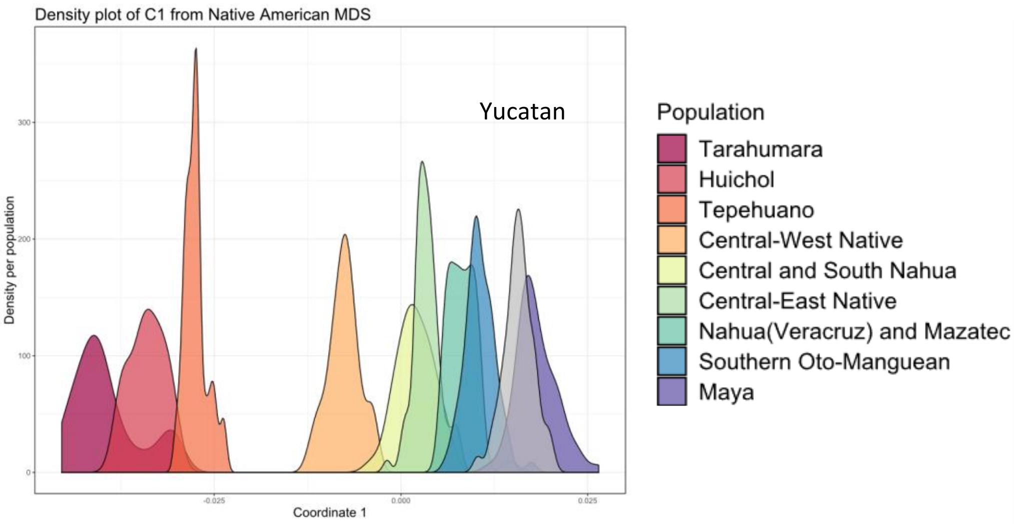


Figure 8.4: Native American MAAS-MDS density per state. Coordinate 1 is shown as a density for reference panels with colors and cosmopolitan Mexicans in gray. The corresponding cosmopolitan Mexican population is specified with a label in each plot.

Table 8.1: Populations considered in the analyses of this thesis, including admixed populations and their reference panels. A specific description of the population, as well as by a simplified label with abbreviations in parenthesis. Sample size, genotyping method or microarray and sampling location are also provided.

Population	Population code (abbreviation)	Sample size	Genotype method	Latitude	Longitude
Cosmopolitan Mexicans from MGDP [1]					
Mexican from Hermosillo, Sonora	**Sonora (SON)**	49	Affymetrix 500K and Illumina 550K	29.07	-110.94
Mexican from Ciudad Victoria, Tamaulipas	**Tamaulipas (TAM)**	17	Affymetrix 500K and Illumina 550K	23.74	-99.14
Mexican from Zacatecas, Zacatecas	**Zacatecas (ZAC)**	50	Affymetrix 500K and Illumina 550K	22.79	-102.59
Mexican from Guanajuato, Guanajuato	**Zacatecas (GUA)**	48	Affymetrix 500K and Illumina 550K	21.01	-101.26
Mexican from Xalapa, Veracruz	**Veracruz (VER)**	50	Affymetrix 500K and Illumina 550K	19.57	-96.90
Mexican from Acapulco, Guerrero	**Guerrero (GUE)**	50	Affymetrix 500K and Illumina 550K	16.88	-99.87
Mexican from Merida, Yucatan	**Yucatan (YUC)**	49	Affymetrix 500K and Illumina 550K	20.98	-89.63
Native American from MGDP [1]					
Tepehuano	**Tepehuano (TEP)**	30	Affymetrix 500K and Illumina 550K	23.48	-104.39
Zapotec (North)	**Zapotec (ZAPN)**	21	Affymetrix 500K and Illumina 550K	17.41	-96.69
Maya in Campeche	**Maya (MYAC)**	45	Affymetrix 500K and Illumina 550K	20.37	-90.05
1000 genomes reference populations [2]					
Yoruba in Ibadan, Nigeria	**Yoruba (YRI)**	108	Full sequencing	7.38	3.95
Mende from Sierra Leone	**Mende (MSL)**	85	Full sequencing	8.46	-11.78
Iberian populations in Spain	**Iberian (IBS)**	107	Full sequencing	40.46	-3.75
Utah Residents (CEPH) with Northern and Western European Ancestry	**Northern Europeans from Utah (CEU)**	99	Full sequencing	52.36*	-1.17*
Japanese in Tokyo, Japan	**Japanese (JPT)**	104	Full sequencing	36.20	138.25
Han Chinese in Beijing, China	**Chinese (CHB)**	103	Full sequencing	39.90	116.41
Han Chinese in South, China	**Southern Chinese (CHS)**	105	Full sequencing	27.63	111.86
Kinh in Ho Chi Minh City, Vietnam	**Vietnamese (KHV)**	99	Full sequencing	10.82	106.63

Punjabi in Lahore, Pakistan	**Punjabi (PJL)**	96	Full sequencing	31.52	74.36
Bengali in Bangladesh	**Bengali (BEB)**	86	Full sequencing	23.68	90.36
Gujarati Indians in Houston, Texas, USA	**Gujarati (GIH)**	103	Full sequencing	22.26*	71.19*
Indian Telugu in the U.K.	**Telugu (ITU)**	102	Full sequencing	18.11*	79.02*
Sri Lankan Tamil in the UK	**Tamil (STU)**	102	Full sequencing	7.87*	80.77*
Mexican Ancestry in Los Angeles CA USA	**Mexican-American (MXL)**	64	Full sequencing	23.63*	-102.55*
Peruvian in Lima, Peru	**Peruvian (PEL)**	85	Full sequencing	-12.05	-77.04
Native Mexican Reference Panel (NMDP) [3]					
Tarahumara	**TAR**	25	Affymetrix 6.0	27.75	-107.17
Huichol	**HUI**	24	Affymetrix 6.0	21.17	-104.08
Nahua in Jalisco	**NAJ**	23	Affymetrix 6.0	19.50	-103.50
Purepecha	**PUR**	23	Affymetrix 6.0	19.75	-101.50
Totonac	**TOT**	25	Affymetrix 6.0	20.00	-97.80
Nahua in Puebla	**Nahua in Puebla (NXP)**	25	Affymetrix 6.0	19.97	-97.62
Nahua trios from Puebla	**Nahua in Puebla (NFM)**	41?	Affymetrix 6.0	19.93	-97.62
Triqui	**TRQ**	25	Affymetrix 6.0	17.18	-97.95
Zapotec (South)	**ZAPS**	24	Affymetrix 6.0	17.23	-96.23
Mazatec	**MAZ**	17	Affymetrix 6.0	18.33	-96.33
Tzotzil	**TZT**	22	Affymetrix 6.0	16.83	-92.67
Maya in Quintana Roo	**MYAQ**	19	Affymetrix 6.0	19.58	-88.58
New Nahua populations (present study) [4]					
Nahua in San Pedro Atocpan, Mexico City	**Nahua from Mexico City (NSP)**	17	Axiom LAT 1	19.20	-99.05
Nahua in Xochimilco, Mexico City	**Nahua from Mexico City (NXO)**	7	Axiom LAT 1	19.26	-99.10
Nahua in Necoxtla, Veracruz	**Nahua from Veracruz (NNX)**	10	Axiom LAT 1	18.80	-97.18
Nahua in Zitlala, Guerrero	**Nahua from Guerrero (ZIT)**	15	Axiom LAT 1	17.69	-99.19
North African reference panel [5]					
Basque from Spain	**Basque**	20	Affymetrix 6.0	42.99	-2.62
POPRES [6]					
Portuguese	**Southwest Europe**	128	Affymetrix 500K	39.40	-8.22
Spain (non-Basque)	**Southwest Europe**	136	Affymetrix 500K	40.46	-3.75
Italy	**South Europe**	214	Affymetrix 500K	41.87	12.57
Sardinian from Italy	**South Europe**	5	Affymetrix 500K	40.12	9.01

Swiss-Italian	**South Europe**	13	Affymetrix 500K	46.82	8.23
Albania	**Southeast Europe**	3	Affymetrix 500K	41.15	20.17
Bosnia-Herzegovina	**Southeast Europe**	9	Affymetrix 500K	43.92	17.68
Bulgaria	**Southeast Europe**	2	Affymetrix 500K	42.73	25.49
Croatia	**Southeast Europe**	8	Affymetrix 500K	45.10	15.20
Greece	**Southeast Europe**	8	Affymetrix 500K	39.07	21.82
Kosovo	**Southeast Europe**	2	Affymetrix 500K	42.60	20.90
Macedonia	**Southeast Europe**	4	Affymetrix 500K	41.61	21.75
Romania	**Southeast Europe**	14	Affymetrix 500K	45.94	24.97
Serbia	**Southeast Europe**	3	Affymetrix 500K	44.02	21.01
Slovenia	**Southeast Europe**	2	Affymetrix 500K	46.15	15.00
Yugoslavia	**Southeast Europe**	41	Affymetrix 500K	43.92	17.68
Cyprus	**East-southeast Europe**	4	Affymetrix 500K	35.13	33.43
Turkey	**East-southeast Europe**	4	Affymetrix 500K	38.96	35.24
Belgium	**Western Europe**	43	Affymetrix 500K	50.50	4.47
France	**Western Europe**	91	Affymetrix 500K	46.23	2.21
Swiss-French	**Western Europe**	125	Affymetrix 500K	46.82	8.23
Austria	**Central Europe**	14	Affymetrix 500K	47.52	14.55
Germany	**Central Europe**	71	Affymetrix 500K	51.17	10.45
Netherlands	**Central Europe**	17	Affymetrix 500K	52.13	5.29
Swiss-German	**Central Europe**	84	Affymetrix 500K	46.82	8.23
Ireland	**Northwest Europe**	61	Affymetrix 500K	53.14	-7.69
Scotland	**Northwest Europe**	5	Affymetrix 500K	56.49	-4.20
United Kingdom	**Northwest Europe**	200	Affymetrix 500K	55.38	-3.44
Czech Republic	**North-northeast Europe**	11	Affymetrix 500K	49.82	15.47
Denmark	**North-northeast Europe**	1	Affymetrix 500K	56.26	9.50
Finland	**North-northeast Europe**	1	Affymetrix 500K	61.92	25.75
Hungary	**North-northeast Europe**	19	Affymetrix 500K	47.16	19.50
Latvia	**North-northeast Europe**	1	Affymetrix 500K	56.88	24.60
Norway	**North-northeast Europe**	3	Affymetrix 500K	60.47	8.47

Poland	**North-northeast Europe**	22	Affymetrix 500K	51.92	19.15
Russia	**North-northeast Europe**	6	Affymetrix 500K	61.52	105.32
Slovakia	**North-northeast Europe**	1	Affymetrix 500K	48.67	19.70
Sweden	**North-northeast Europe**	10	Affymetrix 500K	60.13	18.64
Ukraine	**North-northeast Europe**	1	Affymetrix 500K	48.38	31.17
Reich's reference panel [7]					
Dravidian speakers, South India	**South Asia**	13	Affymetrix 6.0	12.26	77.15
Timor, Indonesia	**Lesser Sunda Islands**	3	Affymetrix 6.0	-9.86	124.33
Roti, Indonesia	**Lesser Sunda Islands**	4	Affymetrix 6.0	-10.74	123.12
Alor, Indonesia	**Lesser Sunda Islands**	2	Affymetrix 6.0	-8.28	124.73
Flores, Indonesia	**Lesser Sunda Islands**	1	Affymetrix 6.0	-8.66	121.08
Besemah, Indonesia	**Sumatra**	10	Affymetrix 6.0	-0.59	101.34
Semende, Indonesia	**Sumatra**	10	Affymetrix 6.0	-3.32	103.91
Ternate, Indonesia	**Maluku Islands**	3	Affymetrix 6.0	0.75	127.36
Hiri, Indonesia	**Maluku Islands**	7	Affymetrix 6.0	0.88	127.32
Kalimantan, Land Dayak, Borneo (Indonesia)	**Borneo**	20	Affymetrix 6.0	0.96	114.55
Barito River region, Borneo (Indonesia)	**Borneo**	23	Affymetrix 6.0	-1.84	114.51
Manobo from Mindanao, Philippines (Austranasian)	**Philippines**	16	Affymetrix 6.0	8.50	123.30
Mamanwa from Mindanao, Philippines (Negrito)	**Negrito**	15	Affymetrix 6.0	8.50	123.30
Southern highlands of Papua New Guinea	**Melanesia**	25	Affymetrix 6.0	-6.31	143.96
Fiji	**Melanesia**	25	Affymetrix 6.0	-17.71	178.07
Polynesia	**Polynesia**	25	Affymetrix 6.0	-16.84	-148.37
Amis from Taitung county, Taiwan	**Taiwan aborigines**	2	Affymetrix 6.0	22.80	121.07
Atayal from Taoyuan and Hsinchu counties, Taiwan	**Taiwan aborigines**	2	Affymetrix 6.0	24.91	121.16
Bunun from Kaohsiung county, Taiwan	**Taiwan aborigines**	2	Affymetrix 6.0	22.63	120.30

Paiwan from Kaohsiung county, Taiwan	**Taiwan aborigines**	2	Affymetrix 6.0	22.63	120.30
Pingpu from Pingtung county, Taiwan	**Taiwan aborigines**	2	Affymetrix 6.0	22.55	120.55
Puyuma from Taitung county, Taiwan	**Taiwan aborigines**	2	Affymetrix 6.0	22.80	121.07
Rukai from Kaohsiung county, Taiwan	**Taiwan aborigines**	2	Affymetrix 6.0	22.63	120.30
Saisiat from Miaoli county, Taiwan	**Taiwan aborigines**	2	Affymetrix 6.0	24.56	120.82
Tsou from Taitung county, Taiwan	**Taiwan aborigines**	2	Affymetrix 6.0	22.80	121.07

* Locations are shown as a population proxy according to ethnic background, not sampling location.

1.- Mexican Genome Diversity Project (MGDP) from (Moreno-Estrada et al., 2014).
2.- 1000 genomes consortium from (Gibbs et al., 2015).
3.- Native Mexican Diversity Project (NMDP) from (Moreno-Estrada et al., 2014).
4.- New Nahua populations from this thesis (see section 3.1).
5.- Basque population from (Henn et al., 2012).
6.- Population Reference Sample (POPRES) from (Nelson et al., 2008).
7.- Southeast Asian reference panel from (Reich et al., 2011).

Table 8.2: Tracts likelihoods for each model tested. A corrected likelihood with BIC is provided for all populations and models according to the number of parameters of each model. The best predicted model for each state is marked with a square.

Likelihood values					
Mexican state	Migration model (European, Native American, African)				
	ppx-xxp	**ppx-xxp-xpx**	**ppx-xxp-xxp**	**ppx-xxp-pxx**	**ppx-xxp-ppx**
Sonora	-265.83	-11490	-276	-257	-251
Tamaulipas	-185.65	-270	-189	-174	-170
Zacatecas	-233.57	-7306	-238	-214	-206
Guanajuato	-276.38	-510	-291	-268	-245
Veracruz	-312.31	-582	-322	-300	-298
Guerrero	-344.22	-373	-352	-328	-310
Yucatan	-357.30	-412	-365	-345	-355