

Article: Discoveries

Dietary, Cultural and Pathogens-related Selective Pressures Shaped Differential Adaptive Evolution Among Native Mexican Populations

Claudia Ojeda-Granados^{1,2}, Paolo Abondio¹, Alice Setti^{1,3}, Stefania Sarno¹, Guido Alberto Gnecci-Ruscone^{1,4}, Eduardo González-Orozco⁵, Sara De Fanti⁶, Andres Jiménez-Kaufmann⁵, Héctor Rangel-Villalobos⁷, Andrés Moreno-Estrada⁵, Marco Sazzini^{1,6*}

¹Laboratory of Molecular Anthropology & Centre for Genome Biology, Department of Biological, Geological and Environmental Sciences (BiGeA), University of Bologna, Bologna, Italy

²Department of Molecular Biology in Medicine, Civil Hospital of Guadalajara “Fray Antonio Alcalde” & Health Sciences Center, University of Guadalajara, Jalisco, Mexico

³current address: Laboratory of Molecular Virology, Department of Cellular, Computational and Integrative Biology (CIBIO), University of Trento, Povo-Trento, Italy

⁴Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany

⁵National Laboratory of Genomics for Biodiversity (LANGEBIO), UGA, CINVESTAV, Iraputato, Guanajuato, Mexico

⁶Interdepartmental Centre Alma Mater Research Institute on Global Challenges and Climate Change, University of Bologna, Italy

⁷Instituto de Investigación en Genética Molecular, Centro Universitario de la Ciénega, Universidad de Guadalajara, Ocotlán, Jalisco, Mexico

*Author for correspondence: Marco Sazzini, Dept. of Biological, Geological and Environmental Sciences (BiGeA), University of Bologna, Italy, +39 051 2094314, e-mail: marco.sazzini2@unibo.it

Abstract

Native American genetic ancestry has been remarkably implicated with increased risk of diverse health issues in several Mexican populations, especially in relation to the dramatic changes in environmental, dietary and cultural settings they have recently undergone. In particular, the effects of these ecological transitions and Westernization of lifestyles have been investigated so far predominantly on Admixed individuals. Nevertheless, indigenous groups, rather than admixed Mexicans, have plausibly retained the highest proportions of genetic components shaped by natural selection in response to the ancient milieu experienced by Mexican ancestors during their pre-Columbian evolutionary history. These formerly adaptive alleles/haplotypes have the potential to represent the genetic determinants of some biological traits peculiar to the Mexican people and a reservoir of loci with potential biomedical relevance. To test such a hypothesis, we used high-resolution genomic data to infer the unique adaptive evolution of 15 Native Mexican groups selected as reasonable descendants of the main pre-Columbian Mexican civilizations. A combination of haplotype-based and gene-network analyses enabled us to detect genomic signatures ascribable to polygenic adaptive traits evolved by the main genetic clusters of indigenous Mexican populations to cope with local environmental and/or cultural conditions. Some of them were also found to play a role in modulating the susceptibility/resistance of these groups to certain pathological conditions, thus providing new evidence for diverse selective pressures having contributed to shape current biological and disease-risk patterns in present-day Native and Mestizo Mexican populations.

Introduction

Mexican populations represent a paradigmatic example of human groups that have recently experienced dramatic shifts in their environmental, dietary and cultural settings, which are expected to have had relevant social and biological consequences, especially for those communities that have still maintained high proportions of Native American (NA) ancestry. In fact, NA genetic background has been implicated with increased risk of a variety of health outcomes in Mestizo (i.e. Admixed) individuals, in whom the effects of recent ecological transitions and Westernization of lifestyles have been predominantly investigated (Rivera et al. 2004; Aguilar-Salinas et al. 2011; Weissglas-Volkov et al. 2013; Ko et al. 2014).

Conversely, few studies have focused so far on Native Mexican populations (Acuña-Alonzo et al. 2010; Lara-Riegos et al. 2015; Romero-Hidalgo et al. 2017), who are instead the most suitable groups to be considered for inferring the pre-Columbian evolutionary history of the ancestors of modern Mexican people (Moreno-Estrada et al. 2014; Ávila-Arcos et al. 2020). Interestingly, these studies revealed a complex genetic structure of the main indigenous groups that have long-inhabited the present-day Mexico's territory, which is tightly related to the well-known historical, cultural and demographic processes they have experienced (Moreno-Estrada et al. 2014; Romero-Hidalgo et al. 2017; Ávila-Arcos et al. 2020). In particular, a northwest-southeast gradient of Mexican variation has been described, along with an overall differentiation in the proportions of ancestry components between northern, central and southern populations, which is clearly distinguishable in both Native and geographically-neighboring Mexican Mestizo groups (Rubi-Castellanos et al. 2009; Moreno-Estrada et al. 2014). Major genetic divergence between Aridoamerican (i.e. northern) and Mesoamerican (i.e. central/southern) indigenous populations has been previously suggested to be ascribable also to the contrasting climate and ecological conditions characterizing these Mexican macro-regions (Gorostiza et al. 2012; González-Martin et al. 2015). Nonetheless, the adaptive evolution of Native Mexican ancestors in response to local environmental and/or cultural selective pressures has been scarcely investigated so far (Ávila-Arcos et al. 2020). Unfortunately, this has limited also the identification of the genetic determinants of important biological features of present-day Native and Mestizo Mexican people. In particular, variation underlying adaptive traits that have been shaped by natural selection in relation to the ancestral milieu of these human groups might represent an important reservoir of alleles/haplotypes with potential biomedical relevance (Sazzini et al. 2016; Ávila-Arcos et al. 2020; Sazzini et al. 2020; Landini et al. 2020). In fact, the rapid and substantial modifications occurred in modern Mexican societies might have turned some of these biological adaptations into unfavorable (i.e. dis-adaptive) traits, thus contributing to differential disease susceptibility among populations with varying proportions of NA ancestry. To test such a hypothesis, we used high-resolution genomic data to infer the adaptive evolution of 15 Native Mexican groups selected as reasonable descendants of the main pre-Columbian Mexican civilizations. For this purpose, we imputed genome-wide data from the Native Mexican Diversity Panel (NMDP) (Moreno-Estrada et al. 2014) and we applied a combination of haplotype-based and gene-network analyses to detect genomic signatures ascribable to the occurrence of selective events under a realistic approximation of a polygenic

adaptation model (Gnecchi-Ruscione et al. 2018; Sazzini et al. 2020). Accordingly, we identified different adaptive traits peculiar to the main genetic clusters of Native Mexican populations, thus providing new evidence for diverse selective pressures having contributed to shape their current biological and disease-risk patterns.

Results

After genotypes imputation of the NMDP and multiple pre- and post-imputation quality control (QC) filtering (see Materials and Methods), we obtained a “high-quality imputed dataset” including 4,875,751 single nucleotide variants (SNVs) characterized for 271 individuals from 15 Native Mexican populations (supplementary table S1, Supplementary Material online). These subjects were selected according to their high proportions of NA ancestry based on results from ADMIXTURE analysis (fig. 1, and supplementary fig. S1, Supplementary Material online).

To test the concordance of the imputed data with patterns of genomic variation previously described for the examined populations, we compared results obtained by means of Principal Component Analyses (PCA) performed before (Moreno-Estrada et al. 2014) and after the imputation procedure (supplementary fig. S2, Supplementary Material online). In both cases, the well-known northwest to southeast cline of Mexican genetic variation was observed (supplementary fig. S2, Supplementary Material online).

Depicting Patterns of Haplotype Sharing Among Native Mexican Populations

We then explored the fine-scale genetic structure of these populations by applying the haplotype-sharing clustering approach implemented in the CHROMOPAINTER/fineSTRUCTURE pipeline. This first enabled us to remove six outlier individuals that did not show close genetic affinity with those belonging to their population of origin and may be thus representative of misleading mismatches between their self-reported and genomic ancestry. By considering only branches of the obtained fineSTRUCTURE dendrogram showing posterior probability above 90%, we then identified three main clusters of Native Mexican populations: the Northern Mexican cluster (NMC), the Central Mexican cluster (CMC), and the Southern Mexican cluster (SMC) (fig. 2). This finding was already attested by previous studies (Gorostiza et al. 2012; Moreno-Estrada et al. 2014), but our fineSTRUCTURE analysis pinpointed a different distribution of some groups within the identified clusters. In detail, the NMC included Seri (SER) and Rarámuri (also known as Tarahumara, TAR) populations from northern Mexico. The CMC was made up of people from Central West and Central East Mexico and belonging to the Huichol (HUI), Nahua (NAH), Nahua Jalisco (NAJ), Nahua Puebla (NAP) and Purepecha (PUR) populations. The SMC instead grouped populations from South (i.e. Triqui, TRQ; Mazatec, MAZ; Zapotec, ZAP) and South East (i.e. Tojolabal, TOJ; Tzotzil, TZT; Lacandon, LAC; Maya, MYA) Mexico, also integrating the Totonaca (TOT) from Central East Mexico (fig. 2). In particular, an appreciable genetic differentiation was observed between NMC populations and CMC or SMC ones ($F_{st} = 0.023$, $p < 10^{-6}$; $F_{st} = 0.029$, $p < 10^{-6}$, respectively), while CMC and SMC turned out to be considerably less differentiated ($F_{st} = 0.007$, $p < 10^{-6}$).

However, when considering NMC populations, a significant differentiation between SER and TAR ($F_{st} = 0.074$, $p < 10^{-6}$) was also evident.

Differential Genomic Signatures of Positive Selection Across Native Mexican Clusters

We then computed the Number of Segregating sites by Length (nSL) statistics and we used the obtained nSL genome-wide distributions to perform gene-network analyses (see Materials and Methods) in order to infer the adaptive evolution of each cluster by assuming that populations belonging to the same genetic group have presumably experienced similar environmental and/or cultural selective pressures. The sole exception was represented by NMC populations, who were analyzed separately due to their significant genetic differentiation and because they have long maintained appreciably different lifestyles despite having encountered similar environmental conditions in Aridoamerica (MacWilliams et al. 2008; Rentería Valencia 2009).

Seri Population

Results obtained by selection scans seems to be consistent with the above-mentioned hypothesis. In detail, three gene networks made up of 18 genes were found to be significantly enriched for selective events in the SER population (fig. 3a, supplementary table S2, Supplementary Material online). Seven of these genes are known to be implicated in the *Starch and sucrose metabolic pathway* and, according to the Kyoto Encyclopedia of Genes and Genomes (KEGG) database, they also play a role in the *Galactose metabolism* and *Fructose and mannose metabolism* by contributing to the degradation/utilization of other simple sugars. Moreover, they also participate in the regulatory steps of the *Glycolysis/Gluconeogenesis* and *Insulin signaling pathway*. In particular, the *SI* and *TREH* genes are highly expressed in the brush border of the small intestine, where the final stage of carbohydrate digestion occurs, including that of sucrose, trehalose and starch intermediates (UniProt Consortium 2019).

Conversely, the *HK2*, *HK1*, *HKDC* and *GCK* genes encode proteins of the hexokinase family that catalyze the phosphorylation of hexoses to produce D-fructose-6-phosphate or D-glucose-6-phosphate, the first and ratelimiting step in most glucose metabolism pathways (UniProt Consortium 2019). Interestingly, as reported in the KEGG database, these latter loci are also involved in the pathogenesis of type II diabetes.

The other two significant gene networks observed in the SER population included 11 genes primarily involved in the *Ras signaling* and *MAPK signaling* pathways, which turned out to be functionally related to the *RASGRF1* gene (fig. 3a, supplementary table S2, Supplementary Material online). Signaling through the Ras/Raf/MAPK cascade responds primarily to growth factors and mitogens to ultimately regulate cell proliferation, differentiation, and apoptosis (Morrison 2012). The loci that showed the most relevant selection signatures in these gene networks were *RASGRP3* and *RASGRF1*. They encode for guanine nucleotide exchange factors that activate Ras proteins by switching its Ras-bound GDP to GTP, *GNG2* encoding the gamma subunit of heterodimeric G proteins and acting as an upstream regulator of *RASGRP3* and *RASGRF1*. Moreover, *CACNA2D3*, *CACNA1E* and *CACNA1D* encode for subunits of calcium voltage-gated channels involved in

calcium-dependent processes, such as muscle contraction, hormone or neurotransmitter release, gene expression, cell motility, cell division and apoptosis (UniProt Consortium 2019) (fig. 3a, supplementary table S2, Supplementary Material online).

Rarámuri Population

Two gene networks belonging to the *Thyroid hormone signaling* and *Mucin type O-Glycan biosynthesis* pathways and including 20 genes were found to have plausibly experienced multiple selective events in the TAR population (fig. 3b, supplementary table S3, Supplementary Material online).

The thyroid hormones (TH) T3 (triiodothyronine) and T4 (thyroxine) play an important role in the organism's growth and development, as well as in the regulation of glucose and lipid metabolism, thermogenesis and energy homeostasis, with T3 being considered the main active form and T4 instead its precursor. Interestingly, the identified putative adaptive genes are primarily involved in the alternative non-canonical and more rapid pathway of TH and encode for Na⁺/K⁺-ATPases (e.g. *ATP1B3*, *ATP1B1*, *ATP1A2* and *ATP1A4*), TRβ isoforms (*THRB*), a subunit of PI3K proteins (*PIK3R5*), and a calcineurin inhibitor (*RCAN2*) (fig. 3b) (Moeller and Broecker-Preuss 2011). Through the non-canonical actions of TH, regulation of gene expression is achieved by an initial non-transcriptionally mediated change in cell physiology. For example, TH can induce modifications in the phosphorylation and activation patterns of the PI3K pathway, as well as in the modulation of availability of second messengers and of plasma membrane ion pump activity (e.g. Na⁺/K⁺-ATPase, Na⁺/H⁺ exchanger and Ca⁺⁺-ATPase) (Cordeiro et al. 2013; Senese et al. 2014). These alternative mechanisms play a relevant role in energy metabolism by regulating glucose and triglyceride concentrations, body temperature, locomotor activity, and heart rate (Hönes et al. 2017). Indeed, TRβ isoforms (*THRB*) are expressed mainly in cardiac and skeletal muscle, brain and adipose tissue.

Candidate genes belonging to the *Mucin type O-Glycan biosynthesis* pathway instead participate to Oglycosilation of mucins, a posttranslational modification mediated by glycosyltransferase enzymes and consisting in the addition of monosaccharides (*O*-glycans) to the hydroxyl group of serine or threonine residues. Mucins represent a family of large and heavily N-acetylgalactosamine (*O*-GalNAc) glycosylated proteins highly expressed by epithelial tissues that produce gel-like secretions to lubricate and protect airways, urogenital and gastrointestinal tracts against physical, chemical or biological insults. Mucin-type *O*-glycans can also serve as surface receptors for adhesion molecules, mediate the interaction with pathogens including bacterial, fungal and viral infections, and locally modulate the inflammatory response (Bergstrom and Xia 2013; Duarte et al. 2016). Among *O*-glycans, *O*-GalNAc is especially relevant for mucin formation and genes showing selection signatures in the TAR population encode for several GalNAc transferases (*GALNT* genes) that fine-tune mucin biosynthesis (fig. 3b).

Populations Belonging to the Central Mexican Cluster

As regards CMC populations, gene networks belonging to the *Glutamatergic*, *Serotonergic* and *Dopaminergic synapse pathways* showed widespread signatures of positive selection distributed across 36 highly interrelated and overlapping genes mediating dependence on addictive substances such as alcohol (fig. 3c, supplementary table S4, Supplementary Material online). Glutamate, serotonin (5-Hydroxytryptamine, 5-HT) and dopamine are indeed important neurotransmitters in the central nervous system, with glutamate representing the major excitatory one and being involved in synaptic plasticity, memory retention and learning, while serotonin regulates behavioral and physiological functions, such as sleep and wake states, emotions, and hemostasis, among others. Dopamine, instead, controls locomotor and visceral functions, motivational and rewarding behaviors, and has been implicated in addiction to activities and substances providing excitement and pleasure (Kanehisa and Goto 2000; Niyonambaza et al. 2019). Interestingly, genes showing some of the most relevant selection signatures in CMC populations encode proteins that modulate inhibitory mechanisms of these systems by suppressing the excess release of neurotransmitters. In detail, Group II (*GRM2*, *GRM3*) and III (*GRM8*, *GRM4* and *GRM7*) metabotropic glutamate receptors (mGluRs), the serotonin 5-HT_{1B} (*HTR1B*) receptor, and the dopamine receptor D₂ (*DRD2*) are G_{i/o} type G-protein coupled autoreceptors. They are known to suppress excess glutamate, 5-HT, and dopamine release via inactivation of the adenylate cyclase/cAMP/protein kinase pathway and the activation of G-protein-coupled inwardly rectifier K⁺ channels –GIRKs (*KCNJ3* and *KCNJ5*), leading to membrane hyperpolarization and suppression of neuronal excitability. In addition, the splice variant of the G- α_{i2} protein (*GNAI2*) regulates the cell surface density of DRD₂ by sequestering them as an intracellular reservoir (López-Aranda et al. 2007). Among the other genes showing smoother selection signatures in these gene sub-networks some encoding the alpha, beta and gamma subunits of G-proteins (*GNAIs*, *GNAs*, *GNBs* and *GNGs*), the members of the adenylate cyclase family (*ADCYs*) and a member of the protein kinase cAMP family (*PRKACB*) were observed. *ITPR2*, *PLCB1* and *PLCB2* genes are instead specifically involved in mobilization of intracellular calcium (Ca²⁺) stores. However, the mechanism of action of the named neurotransmitters, especially their excitatory neurotransmission, also requires intracellular Ca²⁺ mobilization via their coupling to G-proteins that activates phospholipases C (*PLCBs*) to form inositol 1,4,5-trisphosphate (IP₃). IP₃ then binds to the IP₃ calcium release channels receptors (*ITPR2*) in the endoplasmic reticulum to increase eventually cytosolic Ca²⁺ levels. When cross-checking the identified putative adaptive loci with information from the KEGG database, genes belonging to these pathways, which are expressed ubiquitously in diverse tissues, were found to contribute the *Alcoholism*, *Morphine addiction*, *Retrograde endocannabinoid signaling* and *Cocaine addiction* pathways. Moreover, they were found to play a role also in immunological and inflammatory responses to viruses and trypanosomes by participating in the *Human cytomegalovirus infection*, *Chagas disease* and *Chemokin signaling* pathways.

Populations Belonging to the Southern Mexican Cluster

SMC populations showed signatures of positive selection very similar to those observed for CMC groups, particularly at genes implicated in immune and inflammatory responses to viruses and trypanosomes (e.g. *ITPR2*, *ADCY4*, *ADCY8*, *PLCB1*, *PLCB2*, *GNAI2*, *GNAQ* and *GNAS*) (fig. 3d, supplementary table S5, Supplementary Material online). In fact, 18 candidate loci were found to participate in the *Glutamatergic synapse* and *Platelet activation* pathways and were characterized by tight functional connections with the *ITPR2* gene, which showed the highest *per gene* nSL value among those computed. In particular, when we framed these putative adaptive loci in the functional cascades annotated in the KEGG database they turned out to be involved also in the *Chemokine signaling*, *Inflammatory mediator regulation of TRP channels*, *Human cytomegalovirus infection* and *Chagas disease (American trypanosomiasis)* pathways, thus confirming their primary role in immunity and inflammation. In fact, although platelets have been long known to act as regulators of hemostasis and thrombosis, their additional role in innate immune reactions against bacteria, viruses and even tumors has been recently demonstrated. Indeed, they have shown to express nine Toll-like receptors differently by gender during pathogen detection, as well as to have the ability to sample the blood environment to present viruses or bacteria to other immune cells (Koupenova et al. 2015; Holinstat 2017). Interestingly, functioning of the *Platelet activation pathway*, including granule secretion and spreading, is known to depend on the increase of intracellular Ca^{2+} concentration, and indeed *ITPR2*, *PLCB1* and *PLCB2* loci, which are found to be putative adaptive genes also in the CMC, are particularly implicated in the mobilization of intracellular Ca^{2+} stores.

Discussion

To address the lack of studies specifically aimed at inferring the adaptive evolution of Native Mexican human groups, we focused on populations selected as reasonable descendants of the main pre-Columbian civilizations that inhabited the geographical regions nowadays attributed to the Mexican territory. For this purpose, we used the NMDP to collect data for groups showing high proportions of NA ancestry and representative of the main gene pools distinguishable within the overall Mexican population (fig. 1) (Moreno-Estrada et al. 2014; Ávila-Arcos et al. 2020). However, previous studies conducted on these indigenous populations relied on SNP-chip data and/or exome variation, which prevented the exploration of the full spectrum of Mexican alleles in terms of both frequency and nature, thus limiting the potential in investigating fine-scale patterns of population structure and polygenic adaptive processes. In the attempt to overcome this issue, we imputed genome-wide genotypes from the NMDP to considerably increase the resolution of the available genomic dataset.

Fine-scale Genetic Structure of Native Mexican Populations

FineSTRUCTURE haplotype-sharing clustering analysis first enabled us to identify three broad groups of genetically homogenous Native Mexican populations that are expected to have shared a remarkable fraction of their evolutionary histories (fig. 2). The observed patterns of fine-scale population structure were found to be

consistent with the putative ancestral geographic location of the bulk of populations that composed these genetic clusters (fig. 1, supplementary fig. S2, Supplementary Material online). The TOT group from Central East Mexico represented the sole exception, being settled near the Gulf Coast, but clustering within the SMC. This peculiar outcome was likely ascribable to the presence of a geographic barrier (i.e. the Sierra Madre Oriental) that segregated TOT from the other CMC groups. Moreover, TOT are known to have historically maintained hostile relationships with neighboring populations such as the Aztecs, who represented the ancestors of present-day Nahua people, instead establishing intense social and commercial exchanges with southern Mexican populations (Báez-Jorge 2009). Such a genetic affinity between TOT and SMC groups is consistent also with osteological, archeological, and ethnohistorical evidence (Ragsdale and Edgar 2015) and is in agreement with previous findings pointing to gene flow between TOT and MYA people from South East Mexico (Moreno-Estrada et al. 2014).

Overall, a major genetic differentiation was observed between NMC and CMC or SMC ($F_{st} \geq 0.023$, $p < 10^{-6}$) than between CMC and SMC ($F_{st} = 0.007$, $p < 10^{-6}$). This suggests a main divergence between populations that have inhabited Aridoamerica and Mesoamerica, the two major geographical and cultural areas of ancient Mexico, which have been long characterized by contrasting climates and biodiversity landscapes. Such a fragmented demographic history of Mexican groups was proposed also by previous studies, which pointed to a genetic structure of Native Mexicans that is more related to their territories' orography rather than to cultural/linguistic differences (Gorostiza et al. 2012; González-Martín and Gorostiza 2013; Moreno-Estrada et al. 2014; Rangel-Villalobos et al. 2016). In particular, our findings are consistent with results from the analysis of exome sequence data of Mexican indigenous populations, including TAR, HUI, TRQ and MYA, which inferred that northern and southern ancestral Native Mexican groups split approximately 7.2 thousand years ago (KYA) and further diverged locally around 6.5 and 5.7 KYA, respectively (Ávila-Arcos et al. 2020). Moreover, more recent differentiation between HUI/PUR (3.4 KYA), NAH/PUR (2.2 KYA), and HUI/NAH (357 YA) seems to support our assumption that these CMC populations largely shared a common evolutionary history since they diverged from NMC and SMC groups (González-Martín and Gorostiza 2013). This pattern is also consistent with the frequency distribution of the four "Pan-American" mtDNA haplogroups (i.e. A2, B2, C1, D1) in Aridoamerican (e.g. Pima) and Mesoamerican indigenous groups (among which HUI, NAH, Otomí, MYA). In fact, geographic discontinuities in their frequencies revealed two major genetic boundaries deeply delimiting Aridoamericans from all Mesoamerican groups, as well as a southern boundary differentiating MYA from central Mesoamerican populations (Gorostiza et al. 2012).

Interestingly, when considering NMC populations, a noteworthy differentiation was observed also between SER and TAR ($F_{st} = 0.074$, $p < 10^{-6}$) in accordance with the lack of evidence for their historical relationships. SER has indeed shown remarkable linguistic and genetic differentiation with respect to other Native Mexican populations, including TAR, and an independent ancestry component (fig.1, supplementary fig. S1, Supplementary Material online) suggesting a high degree of genetic drift due to prolonged isolation (Moreno-Estrada et al. 2014). In fact,

regions inhabited by SER and TAR are separated by a geographic barrier represented by the Sierra Madre Occidental and in addition, among Aridoamericans, SER maintained a semi-nomadic hunter-gatherer lifestyle until the mid-20th century, while there is evidence for TAR ancestors having become mobile farmers since around 1,600 BC (MacWilliams et al. 2008; Rentería Valencia 2009). Therefore, we can speculate that these populations clustered together into the NMC because they were substantially differentiated from both CMC and SMC ones, rather than because they present a considerable genetic affinity.

Based on this body of evidence, we assumed that SER, TAR, CMC and SMC populations might have adaptively evolved independently in response to differential environmental- and/or cultural-related selective pressures. Accordingly, we searched for genomic signatures left by the action of natural selection on these main Native Mexican groups by testing the occurrence of selective events under a realistic approximation of a polygenic adaptation model.

Seri Metabolic Adaptation to a Diet Rich in Simple Carbohydrates

Metabolism of simple carbohydrates (i.e. mono and disaccharides) and regulation of glucose homeostasis via the *Glycolysis/Gluconeogenesis* and *Insulin signaling* pathways emerged as the biological functions that most plausibly evolved adaptively in the SER population (fig. 3a, supplementary table S2, Supplementary Material online). In fact, selective events at genes playing a role in these metabolic processes seems to be highly consistent with the peculiar diet and lifestyle of this Native Mexican group, who has long-relied mostly on the consumption of fruits of succulent plants that are rich in simple sugars.

According to archeological evidence, this semi-nomadic population originally known as Comcáac has indeed inhabited the dry desert, coastal margins and neighboring islands (e.g. the Tiburón and San Esteban islands in the Gulf of California) of the current state of Sonora in Northwestern Mexico since at least 2 KYA. Moreover, they are known to have maintained their hunting-gathering and fishing based lifestyle until about the 1950s (Rentería Valencia 2009). In particular, they used to organize themselves into mobile groups or "bands", moving around in camps on the shore and islands during the dry-cold season (September-May), while occupying the desert areas during the wet-hot period (June-August). During the cold season, their diet was based on the tender stem of two species of cholla cacti (i.e. *Cylindropuntia bigelovii* and *C. fulgida*), which provided up to 80% of the total carbohydrates consumed, as well as on hunting products and sea resources (Álvarez Palma and Casiano 2003; Hernández Santana 2017). Instead, during the hot season they consumed mainly fruits and seeds of various species of columnar cacti, among which the saguaro (*Carnegiea gigantea*), sweet pitaya (*Stenocereus thurberi*), cardon (*Pachycereus pringlei*), sour pitahaya (*Stenocereus gummosus*) and cina cactus (*Stenocereus alamosensis*), from which they also prepared sweet fermented drinks. The sweet pods from mesquite tree and the prickly pears from the cholla and nopal (*Opuntia ficus-indica*) were also highly appreciated during this season (Álvarez Palma and Casiano 2003; Hernández Santana 2017). Therefore, selective events observed at multiple loci contributing mainly to sucrose, fructose and mannose metabolism might have optimized the utilization of

the high amounts of ingested sugars as energy sources or reserves, also reducing the potential metabolic risk associated with glucose overload (Jang et al. 2018; Rabbani and Thornalley 2019). In line with such a hypothesis, among the top candidate adaptive genes identified (supplementary table S2, Supplementary Material online), *HK2*, *HK1* and *HKDC* encode for hexokinases that have a high affinity for glucose and in particular *HK2*, which is highly expressed in the skeletal muscle and adipose tissue, controls the rate-limiting step of glycolysis (Tan and Miyamoto 2015). In contrast, *GCK* is known to regulate the rate-limiting step of glycolysis and glycogenesis in pancreatic beta cells and hepatocytes and its product shows weak affinity for glucose, but becomes effective when glucose is abundant by modulating insulin secretion and storage of glucose excess (UniProt Consortium 2019). Furthermore, the SI enzyme was proven to be critically involved in the final digestion step of sucrose and starch intermediates such as isomaltose, while TREH hydrolyzes the disaccharide trehalosa, which is naturally found in plants, algae, insects and shrimp (Elbein et al. 2003; UniProt Consortium 2019). Other putative targets of selection in SER people, which support the above-mentioned hypothesis, included genes involved in signal transduction cascades (i.e. *Ras* and *MAPK signaling* pathways). In particular, they play a role in the signaling cascade leading to insulin secretion, as is the case of *CACNA1E* and *CACNA1D* genes that are also expressed by pancreatic beta cells in the islets of Langerhans (Vajna et al. 1998). Interestingly, these latter loci, as well as *HK2*, *HK1*, *HKDC* and *GCK*, have been also proved to be involved in the pathogenesis of type 2 diabetes (Muller et al. 2007; Reinbothe et al. 2013).

Semi-nomadic Lifestyle and Exposure to Maize Mycotoxins as the Main Selective Pressures Having Acted on the Rarámuri Population

Non-canonical actions of TH that are particularly linked to the regulation of energy metabolism (Hönes et al. 2017), as well as variation at *GALNT* genes controlling *O*-glycosylation processes during the biosynthesis of mucins (Bergstrom and Xia 2013), were inferred to have been shaped by positive selection in the TAR population (fig. 3b, supplementary table S3, Supplementary Material online). These adaptations might have evolved in response to selective pressures due to the peculiar mobile agricultural lifestyle adopted by the ancestors of this Native Mexican group since around 1,600 BC, when they first settled the mountains and canyons of the southern Chihuahua State (Northwestern Mexico) (MacWilliams et al. 2008). Throughout this geographic area of contrasting relief, temperatures (which drop below -20 °C in winter, while rising up to 40 °C in summer), rainfall and vegetation, TAR are indeed known for their traditional residential mobility and reliance on food stored year-round. In fact, over the year they move from their primary residences on the valleys to their growing-season residences (May - June), which are associated with scattered fields for maize cultivation and maize depots, and to their winter locations (December - February) in rock-shelters located in the walls of small canyons. These seasonal shifts of residence require them to routinely walk extended distances with heavy loads from the maize depots to maintain or replenish their supplies (Hard and Merrill 1992). Furthermore, according to archaeological evidence, by 500-1,000 AD, TAR had already devised various unique methods of maize storage,

such as underground, cave niches, rectangular and cylindrical structures of stone and cement, or vases of braided grass and clay, thus confirming their strong need to store food for long time as the main subsistence strategy (Hernández Xolocotzi 1985).

It is therefore plausible that the detected selective events might have enhanced two major biological functions that appear to be essential for the TAR population. On the one hand, improvement of the regulation of energy metabolism and thermoregulation might have evolved in response to long-distance walks with heavy loads in such up-and-down relief and extreme climatic conditions. On the other hand, gastrointestinal protection through effective mucins formation could represent a possible adaptation to the ingestion of mycotoxins present in the stored food they relied on. In fact, through the rapid non-canonical signaling mediated by the *THRB*, *ATPases*, and *PIK3R5* putative adaptive genes, TH have been shown to increase energy metabolism, body temperature, oxygen consumption, and locomotor activity, as well as to reduce serum triglycerides concentrations by using them as a substrate for energy production (Hönes et al. 2017; Iwen, Oelkrug, and Brabant 2018). In particular, TR β ^{GS} mice lacking canonical TH receptors signaling, but preserving their non-canonical TH actions, have shown higher VO₂ (i.e. volume of oxygen uptake) and distance traveled compared to control mice (Hönes et al. 2017). Such a hypothesis of enhanced energy metabolism of TAR is in line with that proposed by two previous studies reporting an enrichment of genes involved in musculoskeletal functions harboring novel missense/promoter region variants (Romero-Hidalgo et al. 2017) and a candidate adaptive gene (*BCL2L13*) highly expressed in skeletal muscle (Ávila-Arcos et al. 2020). Both findings suggested adaptive musculoskeletal traits underlying the well-known physical endurance of people of TAR ancestry.

Finally, studies evaluating the effects of exposure to mycotoxins, such as those frequently found in maize (e.g. deoxynivalenol, zearalenone, fumonisin B1, and aflatoxin B1) in relation to its storage conditions have demonstrated a marked disruption of intestinal epithelial cells and mucus barrier integrity and function (Robert et al. 2017). In detail, the assessment of the effects exerted by deoxynivalenol and fumonisin B1 on mucus has revealed that they modulate both mucin production in the duodenum and its monosaccharide composition (namely of GalNAc, galactose and N-acetyl-neuraminic acid) by regulating the expression of genes involved in mucin biosynthesis (Wan et al. 2014; Pinton et al. 2015). Accordingly, since TAR derive approximately 75% of their diet from maize and have relied mostly on stored maize (Hard and Merrill 1992), it is likely that they have been long exposed to mycotoxin ingestion, which might have represented a selective pressure that led to enhanced mucin biosynthesis aimed at ensuring gastrointestinal homeostasis.

Adaptation of Central Native Mexican Populations to Addictive Substances

Populations belonging to the CMC were found to have evolved adaptations mostly mediated by functional pathways involved in the dependence on addictive substances, among which alcohol, morphine, and cocaine (fig. 3c, supplementary table S4, Supplementary Material online). Indeed, the most plausible candidate adaptive genes identified in this population group encode autoreceptors for mGluRs (i.e. *GRM2*, *GRM3*, *GRM8*, *GRM4* and

GRM7), serotonin (*HTR1B*) and dopamine (*DRD2*), which spearhead the inhibitory mechanisms of the related neurotransmission systems and that have been associated with addictive phenotypes (Abraham, Salinas, and Lovinger 2017; Roberto and Varodayan 2017). These findings seem to be highly consistent with the traditional consumption of alcohol and psychoactive plants by the ancestors of these Native Mexican populations (Roman et al. 2013). Accordingly, we propose that adaptive changes at the *Glutamatergic*, *Serotonergic* and *Dopaminergic synapse* pathways might have enhanced biological tolerance towards such stimulant compounds and their detrimental effects, which in turn could have led to the development of a phenotype of dependence. In fact, ethnohistorical evidence suggests that routine consumption of stimulant substances by the ancient Wixárika (HUI) and Aztec (Nahua) populations followed the processes of plants domestications in Mesoamerica approximately 5 KYA. During this period, they also incorporated into their culture the use of fermented alcoholic beverages made from maize (*tejuino*) and the agave plant (“*pulque*”), respectively, which were endemic to this region. In particular, alcoholic beverages were regularly consumed for religious, nutritional, and medicinal purposes, to the extent that excessive drinking leading to drunkenness became strictly prohibited and penalized by death (Roman et al. 2013). Interestingly, most of the identified putative adaptive genes are linked to the ethanol-induced changes in interneuron function and synaptic transmission associated with chronic rather than acute alcohol exposure and play a role in the development of ethanol seeking behavior and alcohol use disorders. For example, neuroadaptations of glutamatergic transmission due to chronic ethanol exposure (CEE) and withdrawal include generally enhanced function of ionotropic (iGluRs) and mGluRs glutamate receptors, resulting in increased release but decreased reuptake of glutamate in several brain regions and an extracellular hyperglutamatergic state (Roberto and Varodayan 2017). In particular, downregulation of mGluR2 encoded by *GRM2* in nucleus accumbens (NAc) neurons contributed to the hyperglutamatergic state and has been implicated in alcohol dependence (Meinhardt et al. 2013). CEE also results in diminished serotonergic and dopaminergic transmission through variable effects on 5-HT receptors, hypoactivity of dopamine neurons, and decreased NAc/striatum dopamine release and *DRD2* availability (Abraham, Salinas, and Lovinger 2017). In addition, *in vivo* studies have revealed that *DRD2*/*GIRK*-mediated autoinhibition is uniquely regulated by a Ca^{2+} -dependent desensitization mechanism involving IP_3 -induced Ca^{2+} release and stimulated by mGluRs activation (Perra et al. 2011). In fact, direct molecular targets of ethanol include specific ion channels, such as the *GIRKs*, encoded by the identified putative adaptive *KCNJ3* and *KCNJ5* genes, as well as non-ion-channel targets, such as intracellular signaling molecules encoded by *ADCYs* and *PRKACB* loci (fig. 3c, supplementary table S4, Supplementary Material online) (Abraham, Salinas, and Lovinger 2017). Moreover, the *HSD17B11* (a shortchain alcohol dehydrogenase that metabolizes secondary alcohols and ketones) and *KCNC2* genes have been recently identified as targets of natural selection in the HUI population by a study implementing different methodological approaches with respect to the present one (Ávila-Arcos et al. 2020). This further supports the hypothesis of adaptive traits evolved in response to traditional consumption of stimulant substances, particularly alcohol. This seems to be in line also with findings showing that Nahua and HUI populations exhibit one of the highest frequencies

documented to date (67% and 65%, respectively) of the *DRD2/ANKK1* TaqIA A1 allele (rs1800497), which is associated with diminished *DRD2* expression and alcohol dependence (Panduro et al. 2017). Interestingly, this variant was found to be not associated with excessive alcohol consumption or metabolic disturbances in these Native Mexican groups. Instead, in the Admixed Mexican population from the same geographic region, the A1 allele was associated with heavy alcohol drinking, overconsumption of palatable foods known to have addictive properties (e.g. meat, fried dishes and sugars), as well as increased risk of impaired glucose, triglycerides, and VLDL levels. Moreover, the C allele of the *DRD2* C957T SNP (rs6277) was associated with increased consumption of sugar and serum triglycerides (Ramos-Lopez et al. 2018; RiveraIñiguez et al. 2019). Although the described genetic signatures overall suggest an ethanol-driven selective pressure, some of the identified putative adaptive genes (i.e. *ITPR2*, *PLCB1* and *PLCB2*), which were found to have evolved adaptively in SMC populations as well, are also known to regulate intracellular Ca²⁺ mobilization and to be implicated in immunological and inflammatory responses to cytomegalovirus and trypanosomes. In fact, parasites such as the American *Trypanosoma cruzi* and *Leishmania Mexicana* are quite common in the central-southern regions of Mexico and therefore we cannot rule out the hypothesis that also pathogen-related selective pressures have contributed to trigger the inferred adaptive events.

Adaptations of Southern Native Mexican Populations to Endemic Pathogens

SMC populations were found to share a considerable fraction of their adaptive evolution with CMC groups, specifically due to the putative adaptive genes described above and controlling intracellular Ca²⁺ mobilization, which exert crucial functions in both the *Glutamatergic synapse* and *Platelet activation* pathways (fig. 3d, supplementary table S5, Supplementary Material online). Nonetheless, all these loci turned out to be implicated also in immune and inflammatory responses, especially to cytomegalovirus and *T. cruzi*. Therefore, we can hypothesize that these (or similar) pathogens have plausibly represented the predominant selective pressures having acted on the ancestors of these Native Mexican populations, who have inhabited for millennia the coastal areas of southern Mexico, or the central-eastern coast in the case of the TOT. Indeed, these regions are characterized by a tropical savanna climate and represent the Mexican endemic areas of parasites such as *T. cruzi* and *L. Mexicana*, which are responsible for Chagas disease and localized cutaneous leishmaniasis (LCL) (RojoMedina et al. 2018; Loría-Cervera et al. 2019).

Interestingly, it has been demonstrated that platelets, besides playing a fundamental role in coagulation and vascular integrity, are also capable of controlling and modulating innate immune responses against pathogens, thus contributing to their clearance. This is mainly due to their abundance of membrane receptors, such as Tolllike receptors, which are able to detect pathogen- and damage-associated molecular patterns (Ribeiro, Migliari Branco, and Franklin 2019). In line with these findings and with our hypothesis of SMC adaptations to pathogens, a recent study identified three innate immune pathways, among which that of *Toll-like receptor signaling*, showing evidence of Native American ancestry enrichment polygenic adaptation in an Admixed

Mexican population of Mayan ancestry (Norris et al. 2020). In fact, platelet granule secretion and spread depend on increased intracellular Ca^{2+} concentration and the *ITPR2* gene that exhibited the most relevant selection signature and functional connection with the other putative adaptive loci in these Native Mexican populations (fig. 3d) is pivotal for mobilization of intracellular Ca^{2+} stores. Furthermore, calcium signaling is vital for host cell invasion and life cycle of both *T. cruzi* and *L. Mexicana* parasites. In particular, cell invasion by *T. cruzi* is Ca^{2+} dependent and its infective trypomastigote stages are capable of triggering intracellular free Ca^{2+} transients in its target cells through a mechanism involving the host IP_3R and the TcIP_3R homolog of *T. cruzi* (Docampo and Huang 2015; Hashimoto et al. 2016). Interestingly, infection by these parasites causes a peculiar clinical picture in southern Mexican populations. For example, in a 7-year longitudinal study conducted on subjects of southern Mexican ancestry chronically affected by Chagas disease, it was unexpectedly found that there was no progression of the main Chagas-related cardiomyopathies and no chagasic megaviscera compared to the chronic clinical picture shown by subjects from different regions of South America (Goldsmith et al. 1986). Similarly, variability of the clinical presentation of LCL has been observed, with populations from endemic areas often manifesting asymptomatic infection, suggesting a host-specific immune response capable of controlling parasite replication (Andrade-Narvaez et al. 2016). Taken together, these evidences and the findings obtained by gene network analyses performed on SMC populations suggest that they might have evolved adaptations in response to specific pathogens that are endemic to the regions long inhabited by their ancestors, possibly resulting in the enhancement of their innate immune-mediated mechanisms of clearance of infection.

Conclusions

By imputing high-resolution genomic data for 271 individuals from 15 Mexican indigenous groups showing high proportions of NA ancestry, the present study succeeded in assembling the largest genomic dataset available so far for Native Mexican populations. This was crucial for investigating the as reliable as possible proxy of the full spectrum of Native Mexican variation, enabling us to explore patterns of fine-scale population structure among Mexican people and to reconstruct the unique adaptive histories of their ancestors. We thus provided new evidence for the considerably different selective pressures that have targeted the main genetic clusters identifiable within the overall Mexican population. Interestingly, some of the identified biological adaptations were found to have the potential to play a role in modulating susceptibility/resistance of these populations to certain pathological conditions. For instance, type 2 diabetes in Seri, impairment of gastrointestinal homeostasis after ingestion of mycotoxins in Rarámuri, dependence on addictive substances/foods in Central Mexican groups, and severity of Chagas-related comorbidities and infection by *L. Mexicana* in people from Southern Mexico. Accordingly, these findings represent valuable insights into the origins and the determinants of some peculiar biological traits of both Native and Mestizo Mexican people. Therefore, the obtained results might lay the foundation for future studies aimed at devising regionalized/personalized preventive strategies properly tailored to these human groups, which is considered the best course of action to cope with the health issues

related to the ecological and cultural transitions that have been recently experienced by Mexican populations (Ojeda-Granados et al. 2017; 2020).

Materials and Methods

Analyzed Dataset

The data analyzed in the present study were retrieved from those constituting the NMDP, which were previously generated as described in Moreno-Estrada et al. (2014). This dataset included 350 samples from 15 Native Mexican populations and characterized for 903,800 single nucleotide polymorphisms (SNPs). In detail, two populations were from Northern Mexico (i.e. SER, TAR), six were from Central Mexico (i.e. HUI, NAH, NAJ, NAP, PUR, TOT), and seven were representative of people from Southern Mexico (i.e. MYA, LAC, TOJ, TZT, ZAP, MAZ, TRQ).

Genotypes Imputation

To generate a high-resolution genomic dataset, the NMDP was used to perform genotype imputation. We first applied preliminary QC procedures using PLINK v2.0 (Chang et al. 2015) to retain only autosomal variants, as well as SNPs and samples showing genotyping success rate $> 95\%$. This led to a filtered dataset made up of 347 samples characterized for 779,283 SNPs, whose genomic coordinates were re-annotated according to the human reference genome build 37 (hg19) using the *liftOver* tool (Hinrichs et al. 2006). We then phased this dataset by means of the Eagle v2.3 software (Loh et al. 2016) implemented in the Michigan Imputation Server (MIS) v1.0.4 (Das et al. 2016) and we used the MIS *Minimac3* algorithm to impute genotypes using the 1000 Genomes Project Phase 3 as reference panel (1000 Genomes Project Consortium et al. 2015). The imputed dataset was subsequently filtered to retain only high-quality genotypes by using the *bcftools* v1.8 package (Li et al. 2009) to remove variants showing a squared Pearson correlation coefficient ($r^2 \leq 0.9$) between the imputed genotype probability and the true genotype call, as suggested in Palmer and Pe'er (2016). This enabled us to obtain a high-resolution dataset including 347 samples characterized for 6,375,183 SNVs.

Data Curation

The imputed dataset was subjected to additional QC filters using PLINK v2.0 to retain SNVs and samples showing genotyping success rate $> 95\%$, as well as variants with no deviations from the Hardy-Weinberg Equilibrium ($p > 1.5 \times 10^{-9}$ after Bonferroni correction for multiple testing). In addition, we filtered out SNVs with ambiguous A/T or G/C alleles and multiallelic variants.

The degree of recent shared ancestry (i.e. identity-by-descent, IBD) for each pair of subjects was also obtained by calculating the genome-wide proportion of shared alleles (i.e. identity by state, IBS) to remove closely related individuals. An IBD kinship coefficient ($PI_HAT \geq 0.270$) was thus considered to exclude 76 samples, which

ultimately led to the generation of a “high-quality imputed dataset” that included information for 4,875,751 SNVs characterized on 271 individuals.

Genotype-based populations structure analyses were then used to check for consistency between the imputed and original NMDP data. PCA was computed using the *smartpca* method implemented in the EIGENSOFT package v6.0.1 (Patterson, Price, and Reich 2006), while the ADMIXTURE algorithm (Alexander, Novembre, and Lange 2009) was applied to test $K = 2$ to $K = 12$ population clusters by running 25 replicates with different random seeds for each K and by retaining only those presenting the highest log-likelihood values. In addition, cross-validation errors were calculated for each K to identify the value that best fit with the data.

The “high-quality imputed dataset” was then phased using SHAPEIT2 v2.r790 (Delaneau et al. 2013) by applying default parameters and by using the HapMap phase 3 recombination maps and the 1000 Genomes Project dataset as a reference panel (1000 Genomes Project Consortium et al. 2015). Moreover, to perform haplotype-based selection scans, the “high-quality imputed dataset” was phased again with SHAPEIT2 v2.r790 (Delaneau et al. 2013) but considering information about ancestral and derived alleles as specified in the reconstructed reference human genome sequence (1000 Genomes Project Consortium et al. 2015).

Haplotype-based Clustering Analyses

To identify groups of genetically homogenous Native Mexican populations and to remove potential outlier individuals not representative of their population of origin, we applied the haplotype sharing clustering methods implemented in the CHROMOPAINTER/fineSTRUCTURE package (Lawson et al. 2012) to the phased “high-quality imputed dataset”. In detail, the ChomoPainter v2 algorithm was used to produce a co-ancestry matrix through the reconstruction of the haplotype sharing pattern of each individual performed by considering all other subjects in the dataset as potential “donors” of DNA sequence chunks, but excluding themselves (Lawson et al. 2012). To this end, ten steps of the Expectation Maximisation algorithm were applied to a subset of chromosomes {4,10,15,22} to estimate recombination/switch rate and mutation/emission probability. The obtained values were averaged across the subset of chromosomes, by weighting them by the number of SNVs, and across individuals. The mean recombination/switch rate and mutation/emission probability values were then used to run again the ChomoPainter v2 algorithm on all autosomes by considering $k = 50$, which is a value preferable when examining closely related populations, to specify the number of expected haplotype chunks used to define a genomic region. We thus obtained a co-ancestry matrix by considering the matrix of counts of haplotype chunks shared among subjects, which were then summed across all autosomes. The obtained matrix was submitted to the Markov chain Monte Carlo (MCMC) fineSTRUCTURE v.fs2.1 algorithm (Lawson et al. 2012) by setting 1,000,000 burn-in iterations, followed by 1,000,000 more iterations and the sampling of inferred clustering patterns every 10,000 runs. A further fineSTRUCTURE v.fs2.1 run consisted in 100,000 additional hill-climbing iteration steps aimed at improving posterior probability and at merging the identified clusters in a step-wise fashion. The inferred haplotype sharing patterns were finally visualized as a circular dendrogram with

the iTOL v5 tool (Letunic and Bork 2019), with clades of genetically homogenous populations being defined by considering only those branches of the dendrogram that were characterized by a posterior probability above 90%.

Pairwise F_{st} values between the identified population clusters were then computed according to the Weir and Cockerham method using the *4P* software (Benazzo et al. 2015). Calculation was replicated after pruning the dataset with PLINK v2.0 to maintain only SNVs in linkage equilibrium with each other and by using the functions implemented in the R *StAMPP* package. This enabled us to perform 1,000 bootstraps to obtain confidence intervals and corresponding p -values associated to the computed F_{st} values.

Number of Segregating Sites by Length Statistic

Based on the assumption that the ancestors of Native Mexican people belonging to the same genetically homogeneous population cluster have plausibly shared a remarkable fraction of their evolutionary histories (Moreno-Estrada et al. 2014; Ávila-Arcos et al. 2020), having also experienced similar environmental- and/or cultural-related selective pressures, we searched for genomic signatures of natural selection independently in each population group.

For this purpose, we computed the nSL statistic, which was developed to detect a broad range of selective events ranging from strong positive selection on newly arisen alleles to more slight selection on multiple alleles, including standing variation (Ferrer-Admetlla et al. 2014). This improved the exploration of the full spectrum of adaptive events evolved by Native Mexican groups, including those mediated by genetic variants with a small individual effect size. In detail, nSL scores for each SNV were obtained through the application of algorithms implemented in the *selscan* v.1.1.0b program (Szpiech and Hernandez 2014) by setting a 20,000 bp threshold for gap scale, 200,000 bp as the maximum length a gap can reach, and 4,500 consecutive SNVs as the maximum extension parameter. The obtained unstandardized nSL scores were finally normalized across 100 frequency bins by subtracting to each value the average nSL score computed for the considered bin and by dividing the resulting score by the related standard deviation (Piras et al. 2012).

Identifying Gene Networks Enriched for Selective Events

To test the occurrence of selective events under a realistic approximation of a polygenic adaptation model, we performed gene-network analyses by applying the *signet* algorithm implemented in the R package (Gouy, Daub, and Excoffier 2017). For this purpose, we used the genome-wide distribution of normalized nSL scores previously obtained for each population cluster to associate nSL values to each gene annotated on the human genome by considering SNVs within a range of 50 kb upstream and downstream of the chromosomal location of the genes. Then, we selected the highest nSL score among those computed for all variants ascribable to the same gene as the representative score of that given genetic locus. This *per gene* nSL value was used as input for the *signet* algorithm, which concomitantly considered also information available for all biological pathways

annotated in the National Cancer Institute/Nature Pathway Interaction Database to assign the reconstructed gene networks to known functional pathways (Schaefer et al. 2009). In fact, the assumption beyond this approach was that multiple genes within a pathway and contributing to a certain biological function might interact to shape a given adaptive phenotype, thus being targeted simultaneously by natural selection. Therefore, by using a simulated annealing approach the *signet* algorithm searches for networks of genes that directly interact with each other and that at the same time present pervasive selection signatures (i.e. those whose distribution of scores within the pathway they belong to was shifted significantly towards extreme values). To this end, we set an iterative process of 20,000 iterations, and we generated null distributions of the highest scoring subnetwork (HSS) for each network of a specific size finally considering a significance threshold of 0.05 when comparing the obtained HSS with them. Accordingly, gene networks showing p -values < 0.05 were considered as significantly enriched for selective events occurred at multiple genes contributing to the same biological function and were plotted with Cytoscape v3.6.0 (Shannon et al. 2003). Moreover, we also performed a cross-check of the identified gene networks using the KEGG database (Kanehisa and Goto 2000) to identify additional pathways in which the putative adaptive genes might play an important role. Overall, this enabled us to shortlist the set of genes and functional pathways that have been more plausibly subjected to pervasive positive selection in the different population clusters.

Data availability

The individual-level genotype data re-analyzed and imputed in the present study are available through a data access agreement by contacting the corresponding authors of the original source of the data as stated in Moreno-Estrada et al. 2014.

Acknowledgments

We would like to thank all donors who contributed biological samples for the generation of the NMDP without whom this work would have not been possible. We are also grateful to Pier Massimo Zambonelli (CESIA, University of Bologna) for his IT assistance. C.O.G. was supported by two postdoctoral grants (grant n. 755288, 711247) from the National Council of Science and Technology (CONACyT) of Mexico in collaboration with the PhD Program in Molecular Biology in Medicine of the University of Guadalajara (CONACyT PNPC 000091) and the PhD Program in Integrative Biology, CINVESTAV, Irapuato (CONACyT PNPC 003649). S.D.S. was supported by the FONDAZIONE CASSA DI RISPARMIO IN BOLOGNA (grant n. 2019.0552 to M.S.).

References

1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, et al. 2015. A Global Reference for Human Genetic Variation. *Nature* 526 (7571): 68–74.

<https://doi.org/10.1038/nature15393>.

- Abraham KP, Salinas AG, Lovinger DM. 2017. Alcohol and the Brain: Neuronal Molecular Targets, Synapses, and Circuits. *Neuron* 96 (6): 1223–38. <https://doi.org/10.1016/j.neuron.2017.10.032>.
- Acuña-Alonzo V, Flores-Dorantes T, Kruit JK, Villarreal-Molina T, Arellano-Campos O, Hünemeier T, Moreno-Estrada A, et al. 2010. A Functional ABCA1 Gene Variant Is Associated with Low HDL-Cholesterol Levels and Shows Evidence of Positive Selection in Native Americans. *Hum Mol Genet* 19 (14): 2877–85. <https://doi.org/10.1093/hmg/ddq173>.
- Aguilar-Salinas CA, Canizales-Quinteros S, Rojas-Martínez R, Mehta R, Rodríguez-Guillén R, OrdoñezSanchez ML, Riba L, Tusié-Luna MT. 2011. The Non-Synonymous Arg230Cys Variant (R230C) of the ATP-Binding Cassette Transporter A1 Is Associated with Low HDL Cholesterol Concentrations in Mexican Adults: A Population Based Nation Wide Study. *Atherosclerosis* 216 (1): 146–50. <https://doi.org/10.1016/j.atherosclerosis.2010.10.049>.
- Alexander DH, Novembre J, and Lange K. 2009. Fast Model-Based Estimation of Ancestry in Unrelated Individuals. *Genome Res* 19 (9): 1655–64. <https://doi.org/10.1101/gr.094052.109>.
- Álvarez Palma AM, and Casiano G. 2003. Uso alimenticio y medicinal de las cactáceas entre los seris de Sonora. *Dimensión Antropológica* 27: 81–104.
- Andrade-Narvaez FJ, Loría-Cervera EN, Sosa-Bibiano EI, Van Wynsberghe NR, Andrade-Narvaez FJ, Loría-Cervera EN, Sosa-Bibiano EI, Van Wynsberghe NR. 2016. Asymptomatic Infection with American Cutaneous Leishmaniasis: Epidemiological and Immunological Studies. *Mem Inst Oswaldo Cruz* 111 (10): 599–604. <https://doi.org/10.1590/0074-02760160138>.
- Ávila-Arcos MC, McManus KF, Sandoval K, Rodríguez-Rodríguez JE, Villa-Islas V, Martin AR, Luisi P, et al. 2020. Population History and Gene Divergence in Native Mexicans Inferred from 76 Human Exomes. *Mol Biol Evol* 37 (4): 994–1006. <https://doi.org/10.1093/molbev/msz282>. Báez-Jorge F. 2009. Cempoala, Veracruz. *Arqueología Mexicana* 99 (17): 32–39.
- Benazzo A, Panziera A, Bertorelle G. 2015. 4P: fast computing of population genetics statistics from large DNA polymorphism panels. *Ecol Evol* 5 (1):172-1755. <https://doi:10.1002/ece3.1261>.
- Bergstrom KSB, Lijun X. 2013. Mucin-Type O-Glycans and Their Roles in Intestinal Homeostasis. *Glycobiology* 23 (9): 1026–37. <https://doi.org/10.1093/glycob/cwt045>.
- Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-Generation PLINK: Rising to the Challenge of Larger and Richer Datasets. *GigaScience* 4: 7. <https://doi.org/10.1186/s13742-0150047-8>.
- Cordeiro A, Souza LL, Einicker-Lamas M, Pazos-Moura CC. 2013. Non-Classic Thyroid Hormone Signalling Involved in Hepatic Lipid Metabolism. *J Endocrinol* 216 (3): R47-57. <https://doi.org/10.1530/JOE-120542>.

- Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, Vrieze SI, et al. 2016. Next-Generation Genotype Imputation Service and Methods. *Nat Genet* 48 (10): 1284–87. <https://doi.org/10.1038/ng.3656>.
- Delaneau O, Zagury JF, Marchini J. 2013. Improved Whole-Chromosome Phasing for Disease and Population Genetic Studies. *Nat Methods* 10 (1): 5–6. <https://doi.org/10.1038/nmeth.2307>.
- Docampo R, Huang G. 2015. Calcium Signaling in Trypanosomatid Parasites. *Cell Calcium* 57 (3): 194–202. <https://doi.org/10.1016/j.ceca.2014.10.015>.
- Duarte HO, Freitas D, Gomes C, Gomes J, Magalhães A, Reis CA. 2016. Mucin-Type O-Glycosylation in Gastric Carcinogenesis. *Biomolecules* 6 (3):33 <https://doi.org/10.3390/biom6030033>.
- Elbein AD, Pan YT, Pastuszak I, Carroll D. 2003. New Insights on Trehalose: A Multifunctional Molecule. *Glycobiology* 13 (4): 17R-27R. <https://doi.org/10.1093/glycob/cwg047>.
- Ferrer-Admetlla A, Liang M, Korneliussen T, Nielsen R. 2014. On Detecting Incomplete Soft or Hard Selective Sweeps Using Haplotype Structure. *Mol Biol Evol* 31 (5): 1275–91. <https://doi.org/10.1093/molbev/msu077>.
- Gnecchi-Ruscione GA, Abondio P, De Fanti S, Sarno S, Sherpa MG, Sherpa PT, Marinelli G, et al. 2018. Evidence of Polygenic Adaptation to High Altitude from Tibetan and Sherpa Genomes. *Genome Biol Evol* 10 (11): 2919–30. <https://doi.org/10.1093/gbe/evy233>.
- Goldsmith RS, Zárate RJ, Zárate LG, Kagan I, Jacobson LB, Morales G. 1986. “Estudios Clínicos y Epidemiológicos de La Enfermedad de Chagas En Oaxaca, México, y Un Estudio Complementario de Siete Años: 1. Cerro Del Aire. *Boletín de La Oficina Sanitaria Panamericana (OSP)* 100 (2): 145–69.
- González-Martín A, Gorostiza A. 2013. Tribalización, Tiempo de Divergencia y Estructura Genética En Mesoamérica: Una Aproximación Molecular. *Cuicuilco* 20 (58): 227–48.
- González-Martín A, Gorostiza A, Regalado-Liu L, Arroyo-Peña S, Tirado S, Nuño-Arana I, Rubi-Castellanos R, Sandoval K, Coble MD, Rangel-Villalobos H. 2015. Demographic History of Indigenous Populations in Mesoamerica Based on mtDNA Sequence Data. *PLoS One* 10 (8): e0131791. doi: 10.1371/journal.pone.0131791.
- Gorostiza A, Acunha-Alonzo V, Regalado-Liu L, Tirado S, Granados J, Sámano D, Rangel-Villalobos H, González-Martín A. 2012. Reconstructing the History of Mesoamerican Populations through the Study of the Mitochondrial DNA Control Region. *PLoS One* 7 (9): e44666. <https://doi.org/10.1371/journal.pone.0044666>.
- Gouy A, Daub JT, Excoffier L. 2017. Detecting Gene Subnetworks under Selection in Biological Pathways. *Nucleic Acids Res* 45 (16): e149. <https://doi.org/10.1093/nar/gkx626>.
- Hard RJ, Merrill WL. 1992. Mobile Agriculturalists and the Emergence of Sedentism: Perspectives from Northern Mexico. *Am Anthropol* 94 (3): 601–20. <https://doi.org/10.1525/aa.1992.94.3.02a00040>.
- Hashimoto M, Doi M, Kurebayashi N, Furukawa K, Hirawake M, Mogi H, Ohmiya Y, Takashi Sakurai, Toshihiro Mita, Katsuhiko Mikoshiba, and Takeshi Nara. 2016. Inositol 1,4,5-trisphosphate Receptor Determines

- Intracellular Ca²⁺ Concentration in Trypanosoma Cruzi throughout Its Life Cycle. *FEBS Open Bio* 6 (12): 1178–85. <https://doi.org/10.1002/2211-5463.12126>.
- Hernández Santana G. 2017. Los seris, el mar, el desierto y su sistema lunar. *Arqueología Mexicana* 25 (147): 68–72.
- Hernández Xolocotzi E. 1985. Graneros de Maíz En México. In Xolocotzia. Obras de Efraím Hernández Xolocotzi Tomo 1. Vol. 1. Revista de Geografía Agrícola. México: Universidad Autónoma de Chapingo.
- Hinrichs AS, Karolchik D, Baertsch R, Barber GB, Bejerano G, Clawson H, Diekhans M, et al. 2006. The UCSC Genome Browser Database: Update 2006. *Nucleic Acids Res* 34 (Database issue): D590-598. <https://doi.org/10.1093/nar/gkj144>.
- Holinstat M. 2017. Normal Platelet Function. *Cancer Metastasis Rev* 36 (2): 195–98. <https://doi.org/10.1007/s10555-017-9677-x>.
- Hönes GS, Rakov H, Logan J, Liao X, Werbenko E, Pollard AS, Præstholm SM, et al. 2017. Noncanonical Thyroid Hormone Signaling Mediates Cardiometabolic Effects in Vivo. *PNAS* 114 (52): E11323–32. <https://doi.org/10.1073/pnas.1706801115>.
- Iwen KA, Oelkrug R, Brabant G. 2018. Effects of Thyroid Hormones on Thermogenesis and Energy Partitioning. *J Mol Endocrinol* 60 (3): R157–70. <https://doi.org/10.1530/JME-17-0319>.
- Jang C, Hui S, Lu W, Cowan AJ, Morscher RJ, Lee G, Liu W, Tesz GJ, Birnbaum MJ, Rabinowitz JD. 2018. The Small Intestine Converts Dietary Fructose into Glucose and Organic Acids. *Cell Metab* 27 (2): 351361.e3. <https://doi.org/10.1016/j.cmet.2017.12.016>.
- Kanehisa M, Goto S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 28 (1): 27–30. <https://doi.org/10.1093/nar/28.1.27>.
- Ko A, Cantor R, Weissglas-Volkov D, Nikkola E, Reddy MVPL, Sinsheimer JS, Pasaniuc B, et al. 2014. Amerindian-Specific Regions under Positive Selection Harbour New Lipid Variants in Latinos. *Nat Commun* 5 (1): 3983. <https://doi.org/10.1038/ncomms4983>.
- Koupenova M, Mick E, Mikhalev E, Benjamin EJ, Tanriverdi K, Freedman JE. 2015. Sex Differences in Platelet Toll-like Receptors and Their Association with Cardiovascular Risk Factors. *Arterioscler Thromb Vasc Biol* 35 (4): 1030–37. <https://doi.org/10.1161/ATVBAHA.114.304954>.
- Landini A, Shaobo Y, Gneccchi-Ruscione GA, Abondio P, Ojeda-Granados C, Sarno S, De Fanti S, et al. 2020. Genomic Adaptations to Cereal-Based Diets Contribute to Mitigate Metabolic Risk in Some Human Populations of East Asian Ancestry. *Evol Appl* 00:1–17. Accessed September 1, 2020. <https://doi.org/10.1111/eva.13090>.
- Lara-Riegos JC, Ortiz-López MG, Peña-Espinoza BI, Montúfar-Robles I, Peña-Rico MA, Sánchez-Pozos K, Granados-Silvestre MA, Menjivar M- 2015. Diabetes Susceptibility in Mayas: Evidence for the

Involvement of Polymorphisms in HHEX, HNF4 α , KCNJ11, PPAR γ , CDKN2A/2B, SLC30A8, CDC123/CAMK1D, TCF7L2, ABCA1 and SLC16A11 Genes. *Gene* 565 (1): 68–75.

<https://doi.org/10.1016/j.gene.2015.03.065>.

Lawson DJ, Hellenthal G, Myers S, Falush D. 2012. Inference of Population Structure Using Dense Haplotype Data. *PLoS Genet* 8 (1): e1002453. <https://doi.org/10.1371/journal.pgen.1002453>.

Letunic I, Bork P. 2019. Interactive Tree Of Life (ITOL) v4: Recent Updates and New Developments. *Nucleic Acids Res* 47 (W1): W256–59. <https://doi.org/10.1093/nar/gkz239>.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map Format and SAMtools. *Bioinformatics (Oxford, England)* 25 (16): 2078–79. <https://doi.org/10.1093/bioinformatics/btp352>.

Loh PR, Danecek P, Palamara PF, Fuchsberger C, A Reshef Y, K Finucane H, Schoenherr S, et al. 2016. Reference-Based Phasing Using the Haplotype Reference Consortium Panel. *Nature Genet* 48 (11): 1443–48. <https://doi.org/10.1038/ng.3679>.

López-Aranda MF, Acevedo MJ, Gutierrez A, Koulen P, Khan ZU. 2007. Role of a Galphai2 Protein Splice Variant in the Formation of an Intracellular Dopamine D2 Receptor Pool. *J Cell Sci* 120 (Pt 13): 2171–78. <https://doi.org/10.1242/jcs.005611>.

Loría-Cervera EN, Sosa-Bibiano EI, Van Wynsberghe NR, Torres-Castro JR, Andrade-Narváez FJ. 2019. Preliminary Epidemiological Findings of Leishmania Infection in the Municipality of Tinum, Yucatan State, Mexico. *Parasite Epidemiol Control* 4: e00088. <https://doi.org/10.1016/j.parepi.2019.e00088>.

MacWilliams AC, Hard RJ, Roney JR, Adams KR, Merrill WL. 2008. The Setting of Early Agriculture in Southern Chihuahua. In *Archeology without Borders. Contact, Commerce, and Change in the U.S. Southwest and Northwestern Mexico*, Laurie D. Webster and Maxine E. McBrinn. Boulder, Colorado: University Press of Colorado.

Meinhardt MW, Hansson AC, Perreau-Lenz S, Bauder-Wenz C, Stählin O, Heilig M, Harper C, Drescher KU, Spanagel R, Sommer WH. 2013. Rescue of Infralimbic MGluR2 Deficit Restores Control Over Drug-Seeking Behavior in Alcohol Dependence. *J Neurosci* 33 (7): 2794–2806. <https://doi.org/10.1523/JNEUROSCI.4062-12.2013>.

Moeller LC, Broecker-Preuss M. 2011. Transcriptional Regulation by Nonclassical Action of Thyroid Hormone. *Thyroid Res* 4 (Suppl 1): S6. <https://doi.org/10.1186/1756-6614-4-S1-S6>.

Moreno-Estrada A, Gignoux GR, Fernández-López JC, Zakharia F, Sikora M, Contreras AV, Acuña-Alonzo V, et al. 2014. The Genetics of Mexico Recapitulates Native American Substructure and Affects Biomedical Traits. *Science* 344 (6189): 1280–85.

Morrison DK. 2012. MAP Kinase Pathways. *Cold Spring Harb Perspect Biol* 4 (11). <https://doi.org/10.1101/cshperspect.a011254>.

- Muller YL, Hanson RL, Zimmerman C, Harper I, Sutherland J, Kobes S, International Type 2 Diabetes 1q Consortium, Knowler WC, Bogardus C, Baier LJ. 2007. Variants in the Ca V 2.3 (Alpha 1E) Subunit of Voltage-Activated Ca²⁺ Channels Are Associated with Insulin Resistance and Type 2 Diabetes in Pima Indians. *Diabetes* 56 (12): 3089–94. <https://doi.org/10.2337/db07-0587>.
- Niyonambaza SD, Kumar P, Xing P, Mathault J, De Koninck P, Boisselier E, Boukadoum M, Miled A. 2019. A Review of Neurotransmitters Sensing Methods for Neuro-Engineering Research. *Appl Sci* 9 (21): 4719. <https://doi.org/10.3390/app9214719>.
- Norris ET, Rishishwar L, Chande AT, Conley AB, Ye K, Valderrama-Aguirre A, Jordan IK. 2020. Admixture Enabled Selection for Rapid Adaptive Evolution in the Americas. *Genome Biol* 21 (1): 29. <https://doi.org/10.1186/s13059-020-1946-2>.
- Ojeda-Granados C, Panduro A, Gonzalez-Aldaco A, Sepulveda-Villegas M, Rivera-Iñiguez I, Roman S. 2017. Tailoring Nutritional Advice for Mexicans Based on Prevalence Profiles of Diet-Related Adaptive Gene Polymorphisms. *J Pers Med* 7 (4):16. <https://doi.org/10.3390/jpm7040016>.
- Ojeda-Granados C, Panduro A, Rivera-Iñiguez A, Sepúlveda-Villegas M, Roman S. 2020. A Regionalized Genome-Based Mexican Diet Improves Anthropometric and Metabolic Parameters in Subjects at Risk for Obesity-Related Chronic Diseases. *Nutrients* 12 (3):645. <https://doi.org/10.3390/nu12030645>.
- Palmer C, Pe'er I. 2016. Bias Characterization in Probabilistic Genotype Data and Improved Signal Detection with Multiple Imputation. *PLOS Genet* 12 (6): e1006091. <https://doi.org/10.1371/journal.pgen.1006091>.
- Panduro A, Ramos-Lopez O, Campollo O, Zepeda-Carrillo EA, Gonzalez-Aldaco K, Torres-Valadez R, Roman R. 2017. High Frequency of the DRD2/ANKK1 A1 Allele in Mexican Native Amerindians and Mestizos and Its Association with Alcohol Consumption. *Drug Alcohol Depend* 172: 66–72. <https://doi.org/10.1016/j.drugalcdep.2016.12.006>.
- Patterson N, Price AL, Reich D. 2006. Population Structure and Eigenanalysis. *PLOS Genet* 2 (12): e190. <https://doi.org/10.1371/journal.pgen.0020190>.
- Perra S, Clements MA, Bernier BE, Morikawa H. 2011. In Vivo Ethanol Experience Increases D 2 Autoinhibition in the Ventral Tegmental Area. *Neuropsychopharmacology* 36 (5): 993–1002. <https://doi.org/10.1038/npp.2010.237>.
- Pinton P, Graziani F, Pujol A, Nicoletti C, Paris O, Ernouf P, Di Pasquale E, Perrier J, Oswald IP, Maresca M. 2015. Deoxynivalenol Inhibits the Expression by Goblet Cells of Intestinal Mucins through a PKR and MAP Kinase Dependent Repression of the Resistin-like Molecule β . *Mol Nut Food Res* 59 (6): 1076–87. <https://doi.org/10.1002/mnfr.201500005>.
- Piras IS, De Montis A, Calò CM, Marini M, Atzori M, Corrias L, Sazzini M, Boattini A, Vona G, Contu L. 2012. Genome-Wide Scan with Nearly 700 000 SNPs in Two Sardinian Sub-Populations Suggests Some Regions as Candidate Targets for Positive Selection. *Eur J Hum Genet* 20 (11): 1155–61. <https://doi.org/10.1038/ejhg.2012.65>.

- Rabbani N, Thornalley PJ. 2019. Hexokinase-2 Glycolytic Overload in Diabetes and Ischemia–Reperfusion Injury. *Trends Endocrinol Metab* 30 (7): 419–31. <https://doi.org/10.1016/j.tem.2019.04.011>.
- Ragsdale CS, Heather JHE. 2015. Cultural Interaction and Biological Distance in Postclassic Period Mexico. *Am J Phys Anthropol* 157 (1): 121–33. <https://doi.org/10.1002/ajpa.22701>.
- Ramos-Lopez O, Panduro A, Rivera-Iñiguez I, Roman S. 2018. Dopamine D2 Receptor Polymorphism (C957T) Is Associated with Sugar Consumption and Triglyceride Levels in West Mexicans. *Physiol Behav* 194: 532–37. <https://doi.org/10.1016/j.physbeh.2018.07.004>.
- Rangel-Villalobos H, Martinez-Sevilla VM, Martínez-Cortes G, Aguilar-Velázquez JA, Sosa-Macias M, RubiCastellanos R, Gonzalez-Martin A. 2016. Importance of the geographic barriers to promote gene drift and avoid (pre and post-Columbian) gene flow in Mexican Native groups: evidence from forensic STR loci. *Am J Phys Anthropol* 160 (2): 298–316. doi: 10.1002/ajpa.22969
- Reinbothe TM, Alkayali S, Ahlqvist E, Tuomi T, Isomaa B, Lyssenko V, Renström E. 2013. The Human LType Calcium Channel Cav1.3 Regulates Insulin Release and Polymorphisms in CACNA1D Associate with Type 2 Diabetes. *Diabetologia* 56 (2): 340–49. <https://doi.org/10.1007/s00125-012-2758-z>.
- Rentería Valencia, Rodrigo F. 2009. Habitar el desierto, navegar en el mar. Procesos de transformación y permanencia entre los Seris. *Arqueología Mexicana* 97 (17): 71–5.
- Ribeiro LS, Migliari Branco L, Franklin BS. 2019. Regulation of Innate Immune Responses by Platelets. *Front Immunol* 10:1320. <https://doi.org/10.3389/fimmu.2019.01320>.
- Rivera JA, Barquera S, González-Cossío T, Olaiz G, Sepúlveda J. 2004. Nutrition Transition in Mexico and in Other Latin American Countries. *Nutr Rev* 62 (7 Pt 2): S149–157. <https://doi.org/10.1111/j.17534887.2004.tb00086.x>.
- Rivera-Iñiguez I, Panduro A, Ramos-Lopez O, Villaseñor-Bayardo SJ, Roman S. 2019. DRD2/ANKK1 TaqI A1 Polymorphism Associates with Overconsumption of Unhealthy Foods and Biochemical Abnormalities in a Mexican Population. *Eat Weight Disord* 24 (5): 835–44. <https://doi.org/10.1007/s40519-018-0596-9>.
- Robert H, Payros D, Pinton P, Théodorou V, Mercier-Bonin M, Oswald IP. 2017. Impact of Mycotoxins on the Intestine: Are Mucus and Microbiota New Targets? *J Toxicol Env Heal B* 20 (5): 249–75. <https://doi.org/10.1080/10937404.2017.1326071>.
- Roberto M, Varodayan FP. 2017. Synaptic Targets: Chronic Alcohol Actions. *Neuropharmacology* 122: 85–99. <https://doi.org/10.1016/j.neuropharm.2017.01.013>.
- Rojo-Medina J, Ruiz-Matus C, Salazar-Schettino PM, González-Roldán JF. 2018. [Enfermedad de Chagas en México]. *Gac Med Mex* 154 (5): 605–12. <https://doi.org/10.24875/GMM.18004515>.
- Roman S, Zepeda-Carrillo EA, Moreno-Luna LE, Panduro A. 2013. Alcoholism and Liver Disease in Mexico: Genetic and Environmental Factors. *World J Gastroenterol* 19 (44): 7972–82. <https://doi.org/10.3748/wjg.v19.i44.7972>.

- Romero-Hidalgo S, Ochoa-Leyva A, Garcíarrubio A, Acuña-Alonzo V, Antúnez-Argüelles E, Balcazar-Quintero M, Barquera-Lozano R, et al. 2017. Demographic History and Biologically Relevant Genetic Variation of Native Mexicans Inferred from Whole-Genome Sequencing. *Nat Commun* 8 (1): 1005. <https://doi.org/10.1038/s41467-017-01194-z>.
- Rubi-Castellanos R, Martínez-Cortés G, Muñoz-Valle JF, González-Martín A, Cerda-Flores RM, Anaya-Palafox M, Rangel-Villalobos H. 2009. Pre-Hispanic Mesoamerican Demography Approximates the Present-Day Ancestry of Mestizos throughout the Territory of Mexico. *Am J Phys Anthropol* 139 (3): 284–94. <https://doi.org/10.1002/ajpa.20980>.
- Sazzini M, Abondio P, Sarno S, Gneccchi-Ruscione GA, Ragno M, Giuliani C, De Fanti S, et al. 2020. Genomic History of the Italian Population Recapitulates Key Evolutionary Dynamics of Both Continental and Southern Europeans. *BMC Biol* 18 (1): 51. <https://doi.org/10.1186/s12915-020-00778-4>.
- Sazzini M, Gneccchi-Ruscione GA, Giuliani C, Sarno S, Quagliariello A, De Fanti S, Boattini A, et al. 2016. Complex Interplay between Neutral and Adaptive Evolution Shaped Differential Genomic Background and Disease Susceptibility along the Italian Peninsula. *Sci Rep* 6 (1): 32513. <https://doi.org/10.1038/srep32513>.
- Schaefer CF, Anthony K, Krupa S, Buchoff J, Day M, Hannay T, Buetow KH. 2009. PID: The Pathway Interaction Database. *Nucleic Acids Res* 37 (Database issue): D674–679. <https://doi.org/10.1093/nar/gkn653>.
- Senese R, Cioffi F, de Lange P, Goglia P, Lanni A. 2014. Thyroid: Biological Actions of ‘nonclassical’ Thyroid Hormones. *J Endocrinol* 221 (2): R1–12. <https://doi.org/10.1530/JOE-13-0573>.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res* 13 (11): 2498–2504. <https://doi.org/10.1101/gr.1239303>.
- Szpiech ZA, Hernandez RD. 2014. Selscan: An Efficient Multithreaded Program to Perform EHH-Based Scans for Positive Selection. *Mol Biol Evol* 31 (10): 2824–27. <https://doi.org/10.1093/molbev/msu211>.
- Tan VP, Miyamoto S. 2015. HK2/Hexokinase-II Integrates Glycolysis and Autophagy to Confer Cellular Protection. *Autophagy* 11 (6): 963–64. <https://doi.org/10.1080/15548627.2015.1042195>.
- UniProt Consortium. 2019. UniProt: A Worldwide Hub of Protein Knowledge. *Nucleic Acids Res* 47 (D1): D506–15. <https://doi.org/10.1093/nar/gky1049>.
- Vajna R, Schramm M, Pereverzev A, Arnhold S, Grabsch H, Klöckner U, Perez-Reyes E, Hescheler J, Schneider T. 1998. New Isoform of the Neuronal Ca²⁺ Channel Alpha1E Subunit in Islets of Langerhans and Kidney--Distribution of Voltage-Gated Ca²⁺ Channel Alpha1 Subunits in Cell Lines and Tissues. *Eur J Biochem* 257 (1): 274–85. <https://doi.org/10.1046/j.1432-1327.1998.2570274.x>.

Wan LY, Allen KJ, Turner PC, El-Nezami H. 2014. Modulation of Mucin mRNA (MUC5AC and MUC5B) Expression and Protein Production and Secretion in Caco-2/HT29-MTX Co-Cultures Following Exposure to Individual and Combined Fusarium Mycotoxins. *Toxicol Sci* 139 (1): 83–98. <https://doi.org/10.1093/toxsci/kfu019>.

Weissglas-Volkov D, Aguilar-Salinas CA, Nikkola E, Deere KA, Cruz-Bautista I, Arellano-Campos O, Muñoz-Hernandez LL, et al. 2013. Genomic Study in Mexicans Identifies a New Locus for Triglycerides and Refines European Lipid Loci. *J Med Genet* 50 (5): 298–308. <https://doi.org/10.1136/jmedgenet-2012101461>.

Figures

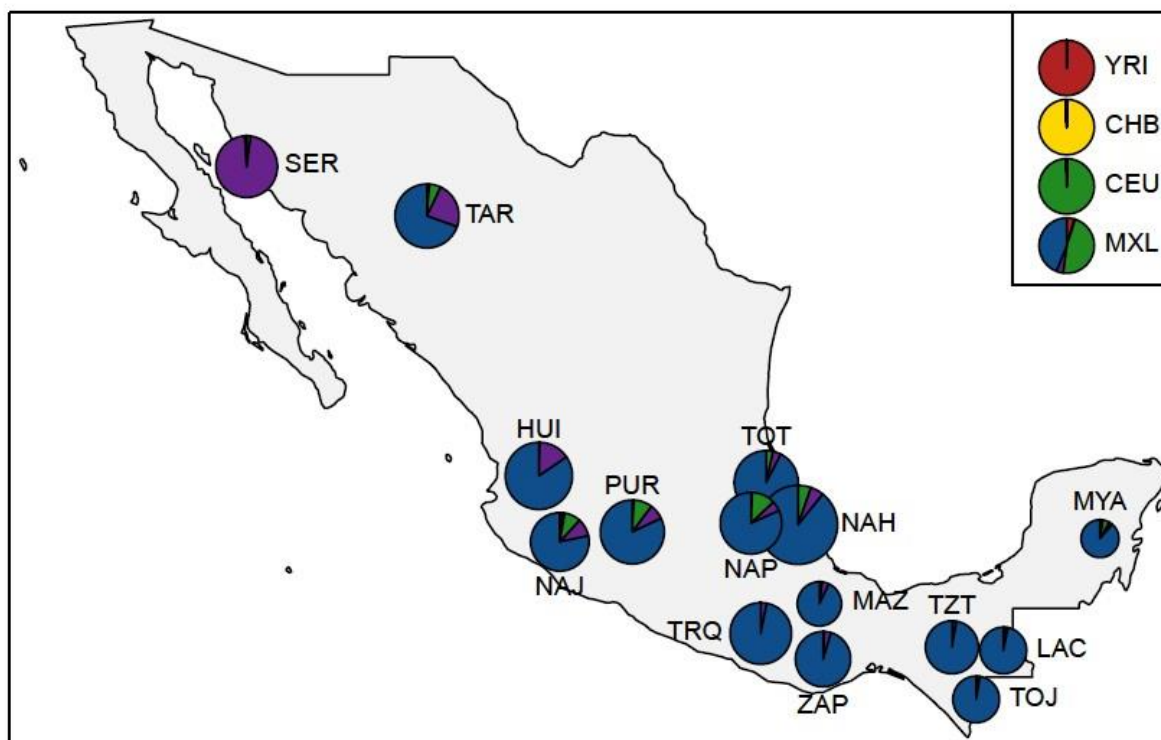


Fig. 1. Average proportions of ancestry components inferred for the 15 Mexican indigenous groups included in the NMDP by means of ADMIXTURE unsupervised clustering analysis at $K = 5$. Evaluation of individual ancestry profiles were used to select subjects representative of the genomic background of their population of origin and showing predominant proportions of Native American ancestry (in dark blue). Individuals of African (YRI), East Asian (CHB), and European (CEU) ancestry, as well as Admixed Mexican-American samples (MXL), were also included to quantify non-Native American genetic fractions observable in the indigenous genomes. SER showed a peculiar ancestry component (in purple) whose high proportion was plausibly ascribable to strong genetic drift experienced due to prolonged isolation, as previously attested (Moreno-Estrada et al. 2014). Pie charts diameters are proportional to the sample sizes of the considered groups (details for each population, full set of K tested and cross-validation errors are reported in supplementary table S1 and supplementary fig. S1, Supplementary Material online). The geographical map has been plotted using the R software (v.4.0.2).

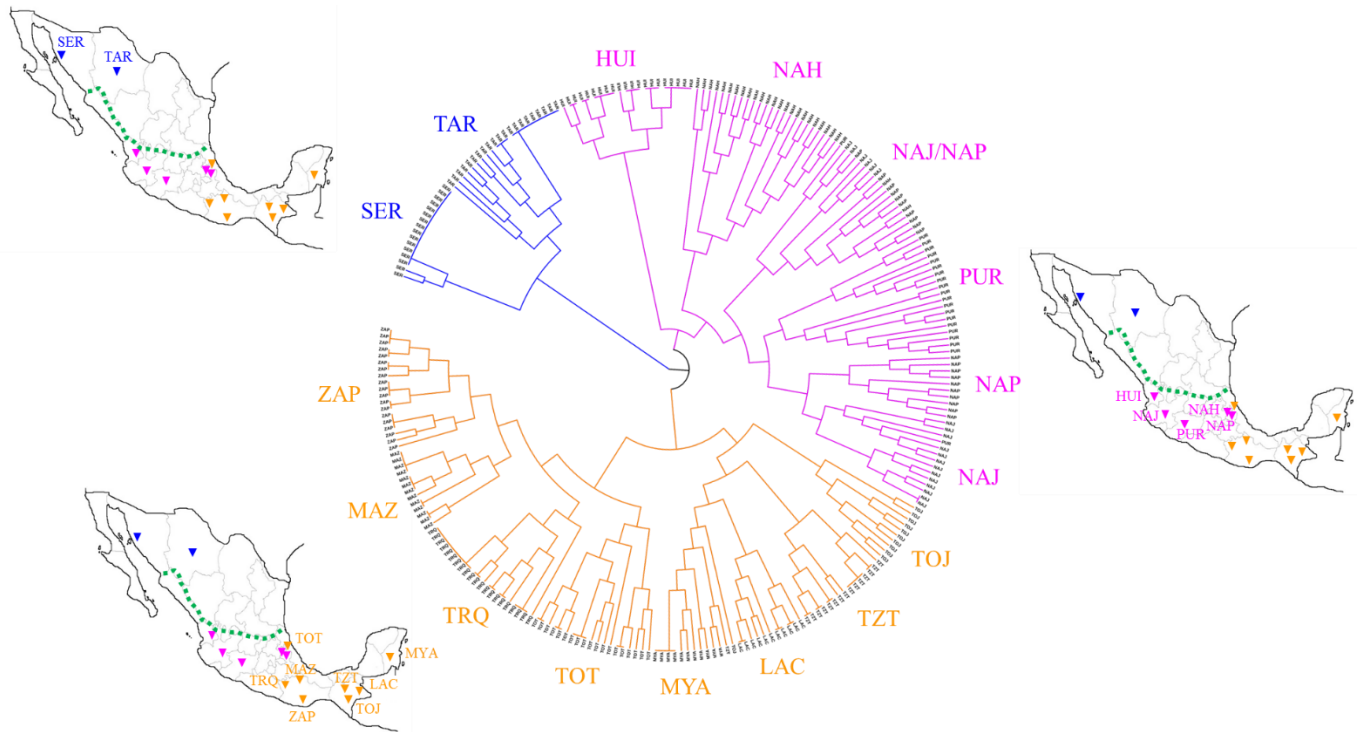


Fig. 2. fineSTRUCTURE hierarchical clustering dendrogram based on haplotype sharing patterns observed among the considered Native Mexican groups and their geographical locations. Population clusters were defined by considering branches of the dendrogram that split with a posterior probability above 90%. Accordingly, the Northern Mexican Cluster (in blue) included SER and TAR populations, the Central Mexican Cluster (in fuchsia) was made up of groups from Central West and Central East Mexico (i.e. HUI, NAH, NAJ, NAP, and PUR), while the Southern Mexican Cluster (in orange) grouped populations from Southern Mexico (i.e. TRQ, MAZ, ZAP), South East Mexico (i.e. TOJ, TZT, LAC, MYA), and Central East Mexico (i.e. TOT). The dotted green line indicates the demarcation between Aridoamerican (northward) and Mesoamerican (southward) regions, which represent the main geographical and cultural areas of ancient Mexico and whose population clusters showed the greatest genetic differentiation according to F_{st} analyses. The geographical maps have been plotted using the R software (v.4.0.2).

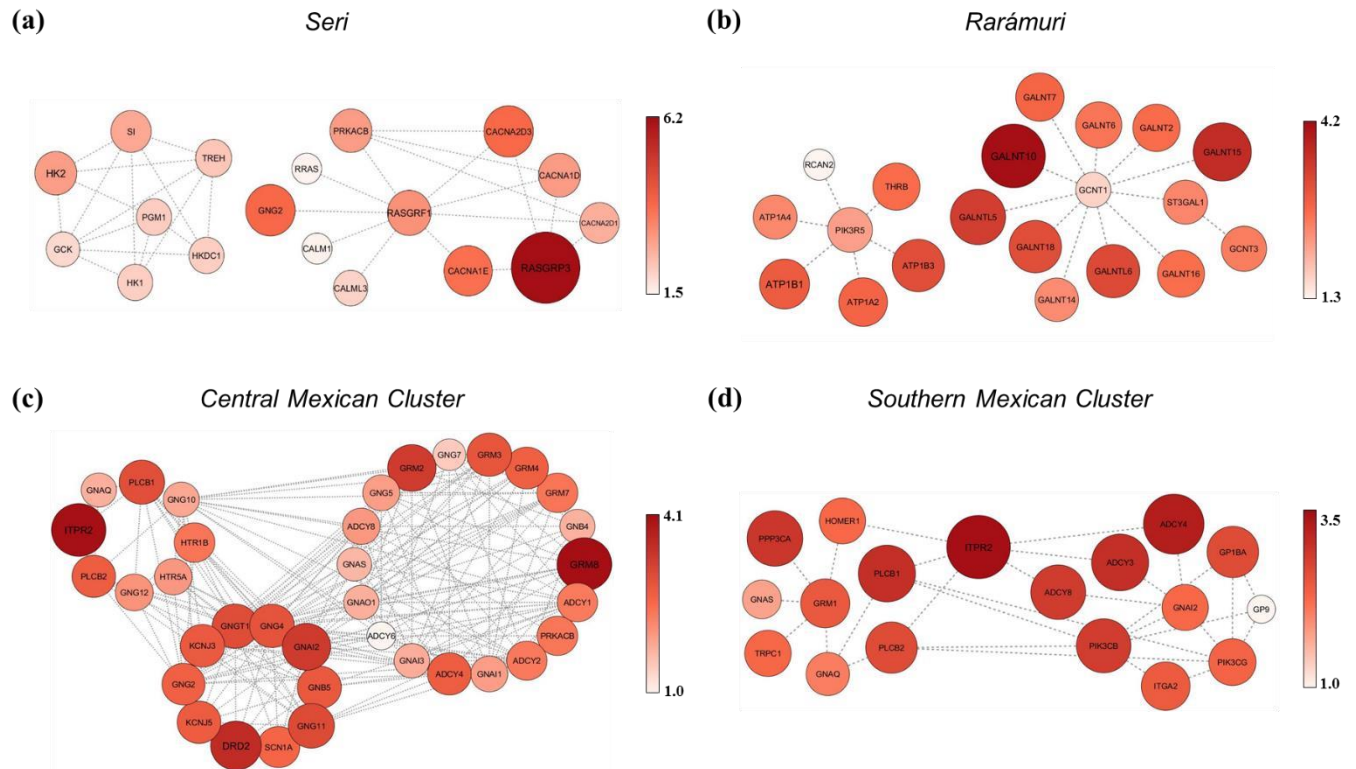


Fig. 3. Representation of gene networks mediating polygenic adaptive events inferred for Seri, Rarámuri and genetic clusters of Native Mexican populations identified according to fineSTRUCTURE analysis. (a) Gene networks belonging to the *Starch and sucrose metabolism*, *Ras signaling* and *MAPK signaling* pathways showing significant footprints of positive selection in the SER population. (b) Gene networks belonging to the *Thyroid hormone signaling* and *Mucin type O-Glycan biosynthesis* pathways having experienced multiple selective events in the TAR population. (c) Highly interrelated gene networks belonging to the *Glutamatergic*, *Serotonergic* and *Dopaminergic synapse* pathways showing widespread selection signatures in Central Mexican Cluster populations. (d) Gene networks belonging to the *Glutamatergic synapse* and *Platelet activation* pathways having mediated adaptive evolution of Southern Mexican Cluster populations. Color intensity of circles is displayed according to the reported color-scale and, along with circles size, is proportional to the *per gene* nSL representative scores used to perform gene-network analyses.