Genome-Wide Selection Scans in Mexican Indigenous Populations Reveal Recent Signatures of Pathogen and Diet Adaptation

Maria Fernanda Miron-Toruno [1,2], Enrique Morett [3], Israel Aguilar-Ordonez [4,*], Austin W. Reynolds [2,*]

[1]Department of Anthropology, Baylor University, Waco, TX 76706, USA

[2]Department of Microbiology, Immunology, and Genetics, University of North Texas Health Science Center, Fort Worth, TX 76107, USA

[3]Departamento de Ingeniería Celular y Biocatálisis, Instituto de Biotecnología, Universidad Nacional Autónoma de México (UNAM), México, Morelos 62210, México

[4]Jefatura de Supercómputo, Subdirección de Bioinformática, Instituto Nacional de Medicina Genomica (INMEGEN), Ciudad de México 14610, México

*Corresponding authors: E-mails: austin.reynolds@unthsc.edu; iaguilar@inmegen.gob.mx.

Accepted: March 05, 2025

Abstract

Whole-genome scans for natural selection signatures across Mexican indigenous populations remain underrepresented in the literature. Here, we conducted the first comparative analysis of genetic adaptation in Mexican indigenous populations using whole-genome sequencing data from 76 individuals representing 27 different ethnic groups in Mexico. We divided the cohort into northern, central, and southern populations and identified signals of natural selection within and across populations. We find evidence of adaptation to pathogenic environments in all our populations, including significant signatures in the Duffy blood group gene in central Mexican indigenous populations. Despite each region exhibiting unique local adaptation profiles, selection signatures on *ARHGAP15*, *VGLL4*, *LINGO2*, *SYNDIG1*, and *TFAP2B* were common to all populations. Our results also suggest that selection signatures falling within enhancers or promoters are usually connected to noncoding features, with notable exceptions like *ARHGAP15* and *GTDC1*. This paper provides new evidence on the selection landscape of Mexican indigenous populations and lays the foundation for additional work on Mexican phenotypic characterization.

Key words: Mexican indigenous populations, evolutionary genomics, adaptive evolution.

Significance

Previous research has identified distinct patterns of genomic adaptation across the different regions of Mexico, highlighting evidence of natural selection within metabolic and immune-related genes. However, the characterization of the Mexican selection landscape from a whole-genome perspective remains unexplored. Here, we conducted the first whole-genome scan for natural selection in 76 Mexican indigenous individuals from 27 different ethnic groups divided into northern, central, and southern populations. Our findings revealed distinct local adaptation profiles for each Mexican region, with different evidence of adaptation to pathogenic environments across these groups. In contrast, all populations had common selection signatures on *ARHGAP15*, *VGLL4*, *LINGO2*, *SYNDIG1*, and *TFAP2B*. This paper provides new evidence on the genetic basis of adaptation of indigenous groups in Mexico. Moreover, it provides a foundation for additional work on Mexican phenotypic characterization.

Introduction

The genomes of Mexican indigenous (MI) populations offer a window into the evolutionary history and adaptation of humans in the Americas. Recent archeological and genetic findings suggest that humans arrived on the American continent during the Late Pleistocene period (Skoglund and Reich 2016; Willerslev and Meltzer 2021). This initial wave of migration, characterized by its wide and rapid expansion, brought human populations to new and often extreme environments that imposed strong selective pressures (Amorim et al. 2017; Willerslev and Meltzer 2021). Due to its geographical position, Mexico has historically served as a natural corridor for human migration, facilitating the movement of populations between north, central, and south America. Understanding the country's selection landscape is crucial to comprehend the genetic makeup of present-day American populations (García-Ortiz et al. 2021; Campelo dos Santos et al. 2022).

Recent advancements in next-generation sequencing technologies, along with the exponential increase of computational power, have provided the opportunity to identify patterns of natural selection in the human genome (Horscroft et al. 2019). In doing so, researchers have begun to understand not only the evolutionary history of populations but also the geographical distribution of genetic loci associated with phenotypic traits (Fan et al. 2016). While many studies have explored natural selection in populations around the world, few have focused on how it has shaped the genomes of present-day Mexican and MI populations (Reynolds et al. 2019; Ávila-Arcos et al. 2020; García-Ortiz et al. 2022; Mendoza-Revilla et al. 2022; Ojeda-Granados et al. 2022; Garcia et al. 2023).

Previous research has revealed distinctive selection patterns across different populations and regions within Mexico. Specifically, metabolic and immune-related genes under putative selection are consistently observed among MI populations. These signatures may have resulted from extended periods of metabolic or pathogenic pressures experienced throughout the population's history, alongside exposure to novel pathogens introduced during La Conquista de México from 1519 to 1521 (Reynolds et al. 2019; Ávila-Arcos et al. 2020; García-Ortiz et al. 2022; Mendoza-Revilla et al. 2022; Ojeda-Granados et al. 2022; Garcia et al. 2023). Selection landscapes that are specific to physical endurance, stature, and metabolic efficiency have been identified in the Rarámuri, Triqui, and Seri indigenous groups, respectively (Ávila-Arcos et al. 2020; Ojeda-Granados et al. 2022). However, these studies have based their inferences on single nucleotide polymorphism (SNP) array and exome-sequencing technologies, thus offering a limited perspective of selection across the whole genome (Goodwin et al. 2016). With a few exceptions (García-Ortiz et al. 2022; Ojeda-Granados et al. 2022; Garcia et al. 2023), studies also limited their findings to ancestry-masked Mexican genomes with variable percentages of MI ancestry.

Here, we present the first comparative analysis of genetic adaptation in MI populations using whole-genome sequencing (WGS) data with high Native American ancestry (>95%). Our study has enabled us to report selection signatures in previously unexplored regions of the genome, showing strong selective signals related to immune processes and providing further insight into possible genetic adaptations to pathogenic pressures. Moreover, we present nf-selection, a Nextflow pipeline for automated detection of recent selection using population branch statistics (PBS) (Yi et al. 2010) and integrated haplotype scores (iHS) (Voight et al. 2006). With nf-selection, we aim to streamline the identification of selection signatures on genome-wide data by simplifying the use of software needed for both tests. The pipeline is publicly available at GitHub: https://github.com/fernanda-miron/nf-selection.

Results

Data Description

We analyzed 76 unrelated MI individuals from the metabolic analysis in an indigenous sample (MAIS) cohort who were whole genome sequenced as part of a previous study (Aguilar-Ordoñez et al. 2021) (Fig. 1a; supplementary table S1, Supplementary Material online). Individuals are representatives of 27 different ethnic groups in Mexico and show an average of 95.7% similarity to populations reported as Native American by ADMIXTURE (Fig. 1c; supplementary fig. S1, Supplementary Material online). The initial publication of these data used a principal component analysis (PCA), and the individuals in this study were divided into three groups corresponding roughly to northern, central, and southern Mexico (recapitulated in Fig. 1b; supplementary fig. S2, Supplementary Material online) (Aguilar-Ordoñez et al. 2021). To further validate the robustness of these groups, we computed within and between group mean weighted F_{ST} (supplementary tables S2 and S3, Supplementary Material online). Our results show significantly smaller F_{ST} values within groups than between groups, thus supporting regional clustering for the selection analyses presented here.

Implementing a Selection Pipeline

To simplify the detection of selection signatures, we developed nf-selection, a Nextflow bioinformatic pipeline that computes two commonly used statistics: PBS and iHS. The PBS is a population differentiation-based test that identifies strong allele frequency differences between three populations by comparing pairwise F_{ST} values (Yi et al. 2010). The iHS is a linkage disequilibrium-based method that measures extended haplotype homozygosity at a given genomic

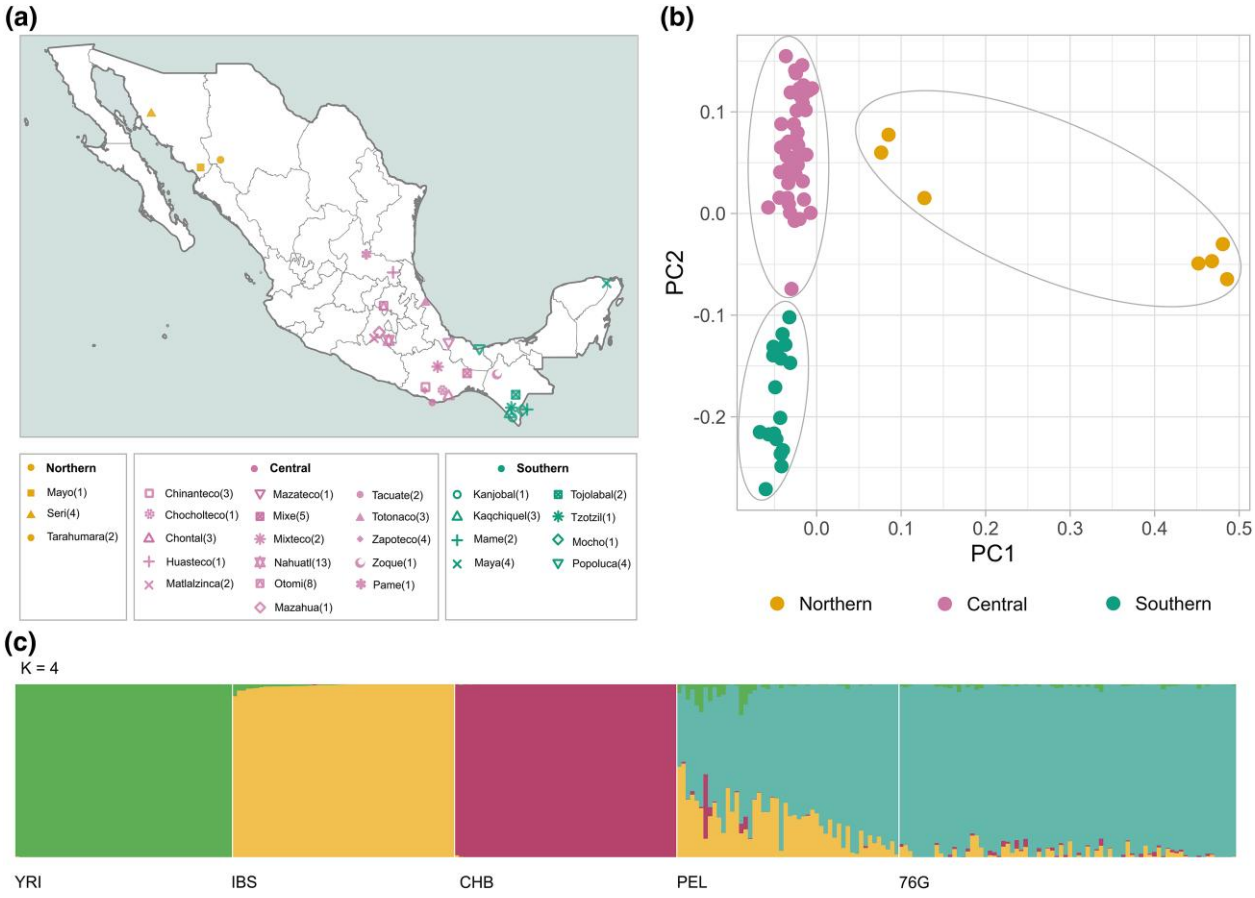

Fig. 1. a) Approximate sampling locations of 27 different ethnic groups in Mexico. Shapes denote the ethnic group, and colors denote their assigned geographic region. *N* indicates the number of individuals from each population. For selection analysis, samples were separated based on northern (yellow), central (pink), and southern regions (Green). b) PCA exhibiting the clustering of northern, central, and southern regions. Gray ovals were added for visual explanation purposes. c) ADMIXTURE analysis for $K = 4$ showing a high proportion of Native American ancestry on the 76 MI samples. References include Yoruba (YRI), Iberians (IBS), Han Chinese (CHB), and Peruvians (PEL) from the 1KGP.

region (Voight et al. 2006). For details on nf-selection implementation, see Materials and Methods. This pipeline is publicly available via GitHub (https://github.com/fernanda-miron/nf-selection) and includes instructions for installation and testing.

For each of our three study populations (northern, central, and southern regions), we used nf-selection to calculate PBS for each autosomal variant using 33 unadmixed Peruvians (PEL) from the 1000 Genomes Project (1KGP) as an ingroup and 50 Han Chinese (CHB) individuals from the 1KGP as an outgroup. iHS was also computed for each population cluster as part of the nf-selection pipeline. Genome-wide *P*-values were calculated separately for PBS using a distribution simulated under a demographic model specific to the Mexican population (supplementary fig. S13, Supplementary Material online). Finally, we cross-referenced the top 1% of *P*-values for both statistics and focused on putative signals of selection detected by both methods to reduce our chances of reporting false positives (Reynolds et al. 2019).

Selection Signatures Analysis

We found 2,758 variants under putative selection for the northern MI populations, 1,097 for the central MI populations, and 1,459 for the southern MI populations (Fig. 2; supplementary figs. S3 to S8 and tables S4 to S6, Supplementary Material online). To prioritize putatively selected genomic regions for further analysis, we calculated the density of significant selection statistics in each region (i.e. the number of variants under selection normalized by gene length) (supplementary tables S7 to S9, Supplementary Material online). Genes with the largest value of selection density for each population are depicted in Table 1 and Fig. 3.

Selection Signals in Regulatory Elements

Next, we looked for selection signals within enhancer and promoter regions using GeneHancer (v. 5.17), a novel database of human enhancers and their inferred target genes, in the framework of GeneCards (Stelzer et al. 2016). This

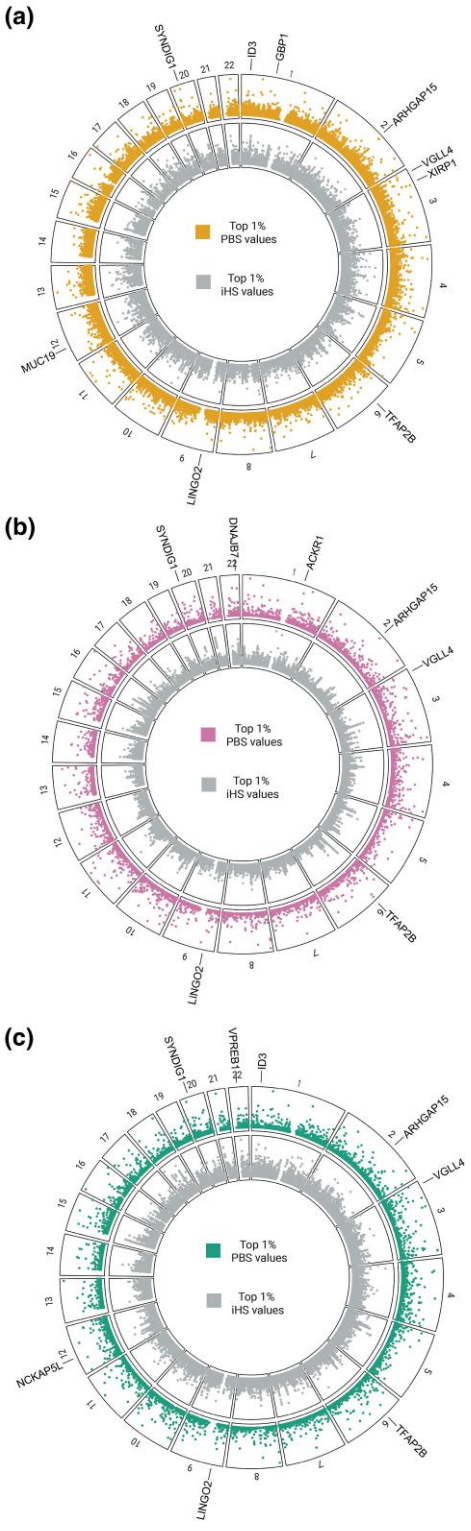

Fig. 2. Significant adaptation signals in MI populations. SNV values within the top 1% distribution of PBS (outer track) and his (inner track) per autosome are plotted. Genes with the greatest selection density, with nonsynonymous variants under putative selection, and shared between the three clusters are labeled. a) Northern MI population. b) Center MI population. c) Southern MI population.

analysis provides two types of results: one for variants per enhancer/promoter and one for variants per genomic feature. An enhancer/promoter element can regulate more than one genomic feature, and each genomic feature can be regulated by more than one enhancer/promoter. This was done independently for each region, and the results were compared across regions.

We find putatively selected variants in 382, 137, and 168 regulatory elements in the northern, center, and southern populations, respectively (supplementary fig. S9, Supplementary Material online). We calculated each regulatory element's selection signal density (supplementary fig. S10, Supplementary Material online) and selected the top ten hits per population (supplementary tables S10 and S11, Supplementary Material online). The top ten enhancers in every region are linked to noncoding elements, such as long noncoding RNAs, microRNAs, and PIWI-interacting RNAs.

Next, we identified putatively selected variants in 21 variable regulatory elements shared between any two populations (supplementary tables S12 and S13, Supplementary Material online). The southern and central populations share three putatively selected SNPs in GH20J024425 (length 1,999 bp) on chromosome 20, which regulate *SYNDIG1* and *GGTLC1*, two interferon-induced transmembrane proteins (Fagerberg et al. 2014). We also identified eight putatively selected SNPs within five enhancers of the gene *MEI1*, known to be involved in meiosis and gamete generation (Fagerberg et al. 2014).

Finally, we detected one enhancer, GH02J143479 (length 4,983 bp) (supplementary fig. S9, Supplementary Material online), containing putatively selected variants across the northern, center, and southern populations, including the SNP chr2-143480377-T-C, which is shared across the three populations. This enhancer regulates the *ARHGAP15* and *GTDC1* genes and the noncoding elements *piR-44085-015*, *lnc-KYNU-7*, and *ENSG0000022865*. Interestingly, GH02J143479 exhibits one SNP under selection (chr2-143480377-T-C), consistently observed across all study populations, with the highest reported allele frequency (0.52) among the American indigenous population in gnomAD v4.0 (Chen et al. 2024).

Gene Annotation and Pathway Analysis

To further explore putative signals of selection identified in the study populations, we annotated significant variants as exonic, intergenic, intronic, etc., using ANNOVAR (Wang et al. 2010) (Table 2). Consistent between the three study populations, most putative selective sites are located within intergenic regions. Additionally, among the variants on gene-coding regions, intronic sites are overrepresented.

Since nonsynonymous mutations change protein sequences and are frequently the object of natural selection (Chu and Wei 2019), we further focused our findings on

Table 1 Top genes under selection in MI populations based on selection density (number of variants under selection normalized by gene length)

Population cluster	Gene	Chromosome	Gene length	Variants	Avg PBS (avg pval)	Average iHS (avg pval)
Northern	*ID3*	1	1.576 kb	chr1-23559007-T-C, chr1-23619424-T-C	0.414 ($2.00e^{-05}$)	3.038 ($2.46e^{-03}$)
Center	*ACKR1*	1	3.194 kb	chr1-159252007-A-G, chr1-159256103-A-T, chr1-159256150-G-A, chr1-159257452-A-G, chr1-159263285-A-G, chr1-159264789-T-C, chr1-159272796-C-T, chr1-159275103-C-T, chr1-159276395-T-C, chr1-159276733-A-G, chr1-159278258-A-G, chr1-159284769-C-T, chr1-159286004-G-A	0.196 ($2.04e^{-02}$)	2.999 ($3.43e^{-03}$)
Southern	*VPREB1*	22	0.736 kb	chr22-22057413-C-T, chr22-22189371-G-A, chr22-22190151-A-C, chr22-22195398-G-A, chr22-22195399-A-G, chr22-22196001-C-G, chr22-22196484-T-C, chr22-22196562-G-C, chr22-22196582-T-A, chr22-22199605-G-A, chr22-22201043-G-A	0.303 ($1.21e^{-03}$)	3.478 ($8.49e^{-4}$)

nonsynonymous variants found in our three study populations. For the northern MI populations, we found four nonsynonymous variants in Guanylate Binding Protein 1 (*GBP1*), Xin Actin Binding Repeat Containing 1 (*XIRP1*), Inhibitor of DNA binding 3 (*ID3*), and *MUC19,* respectively. A nonsynonymous frameshift insertion in DnaJ Heat Shock Protein Family (Hsp40) Member B7 (*DNAJB7*) was found to be under selection in central MI populations, and two nonsynonymous variants were identified in *ID3* and NCK Associated Protein 5 Like (*NCKAP5L*) for southern MI populations.

We conducted a KEGG pathway enrichment analysis using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) (Dennis et al. 2003) to study the pathways in which genes harboring signatures of selection may be overrepresented. However, after correcting for the false discovery rate (FDR < 0.05), we found no significant enrichment pathways for any of the three study populations.

Since adaptations to previous selective pressures may become maladaptations and contribute to disease susceptibility among populations, we also performed a gene–disease association scan using DAVID (supplementary tables S14 to S16, Supplementary Material online). After FDR correction, we found significant associations for renal, psychiatric, metabolic, hematological, chemical dependency, and cardiovascular diseases in all study populations. Immune and neurological diseases were also significant in northern and southern MI populations (supplementary fig. S12, Supplementary Material online).

Shared Signals of Selection Between Populations

Next, we looked for genes with signals of selection among northern, center, and southern Mexican populations by comparing statistically significant results from each analysis. Additionally, to increase the certainty of our analyses, we cross-referenced our results with previously reported signals of selection in MI populations (Reynolds et al. 2019; Ávila-Arcos et al. 2020; García-Ortiz et al. 2022) (Fig. 4).

We found ten genes with putative signatures of selection in all three populations(*CCSER1*, *KIF2B*, *LINC00430*, *LINC01721*, *LINC02546*, *LOC101928516*, *MIR8068*, *PKHD1*, *SLITRK6*, and *WWOX*) that have not been previously reported in the literature and five genes with signals of selection in all three populations that have been previously reported in the literature (*ARHGAP15*, *VGLL4*, *LINGO2*, *SYNDIG1*, and *TFAP2B*). A brief overview and description of the function of each gene is provided in Table 3.

Assessing the Impact of Admixture on Selection Scans

Previously published global admixture analysis of these 76 samples showed minimal European/African admixture

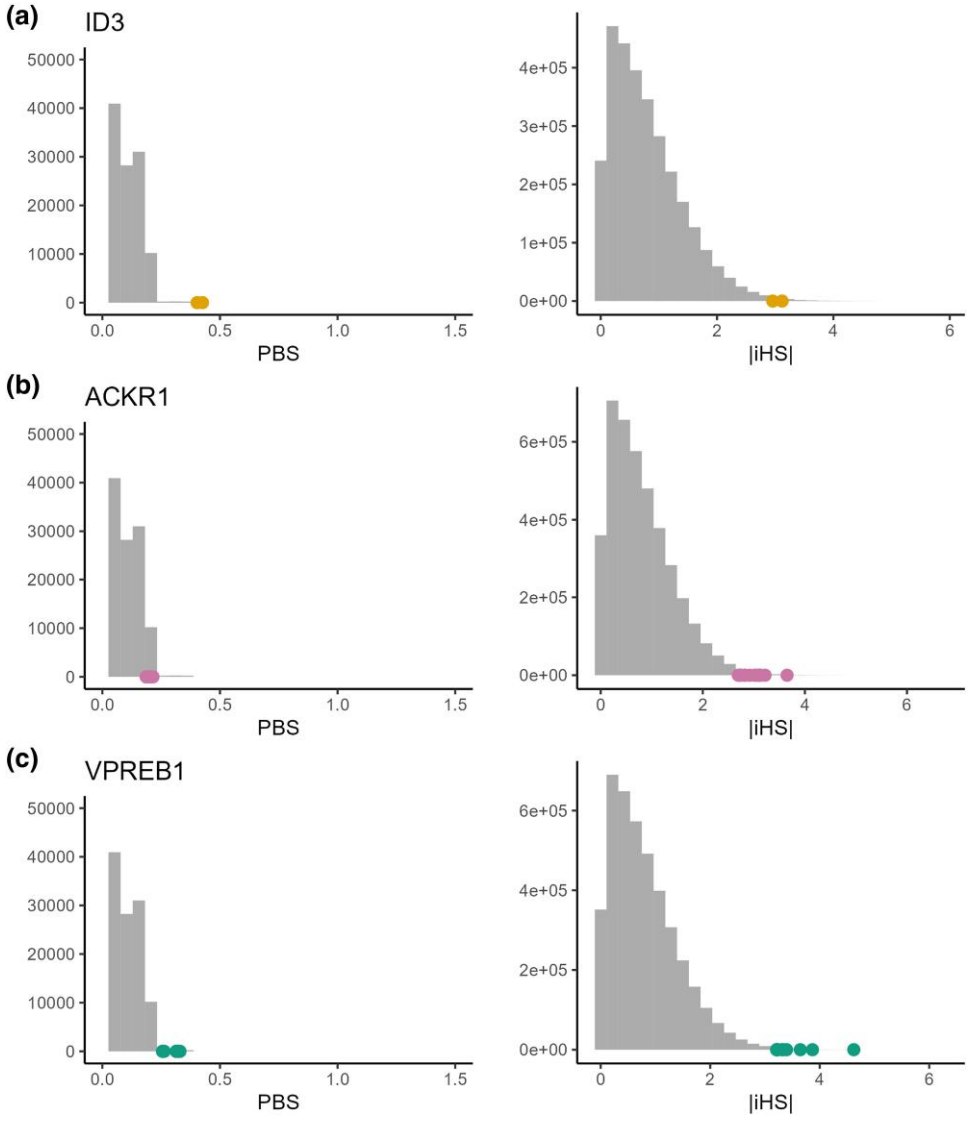

Fig. 3. Genes with the strongest selection signals in MI populations plotted against the simulated distribution of PBS and calculated distribution of iHS. Dots represent the values obtained for each SNP in the focal gene a) ID3 (northern MI population), b) ID3 (center MI population), and c) ID3 (southern MI population).

Table 2 ANNOVAR annotation for variants under hypotheses of selection

Variant type	Northern variants	Central variants	Southern variants
UTR3	25	15	19
UTR5	2	1	2
Downstream	21	8	6
Exonic	3 nonsynonymous SNVs, 3 synonymous SNVs, 1 unknown	2 synonymous SNVs, 1 frameshift insertion	2 nonsynonymous SNVs, 3 synonymous SNVs
Intergenic	1,497	532	783
Intronic	992	444	562
ncRNA exonic	15	6	6
ncRNA intronic	189	72	66
Upstream	10	14	10
Upstream/downstream	0	1	0

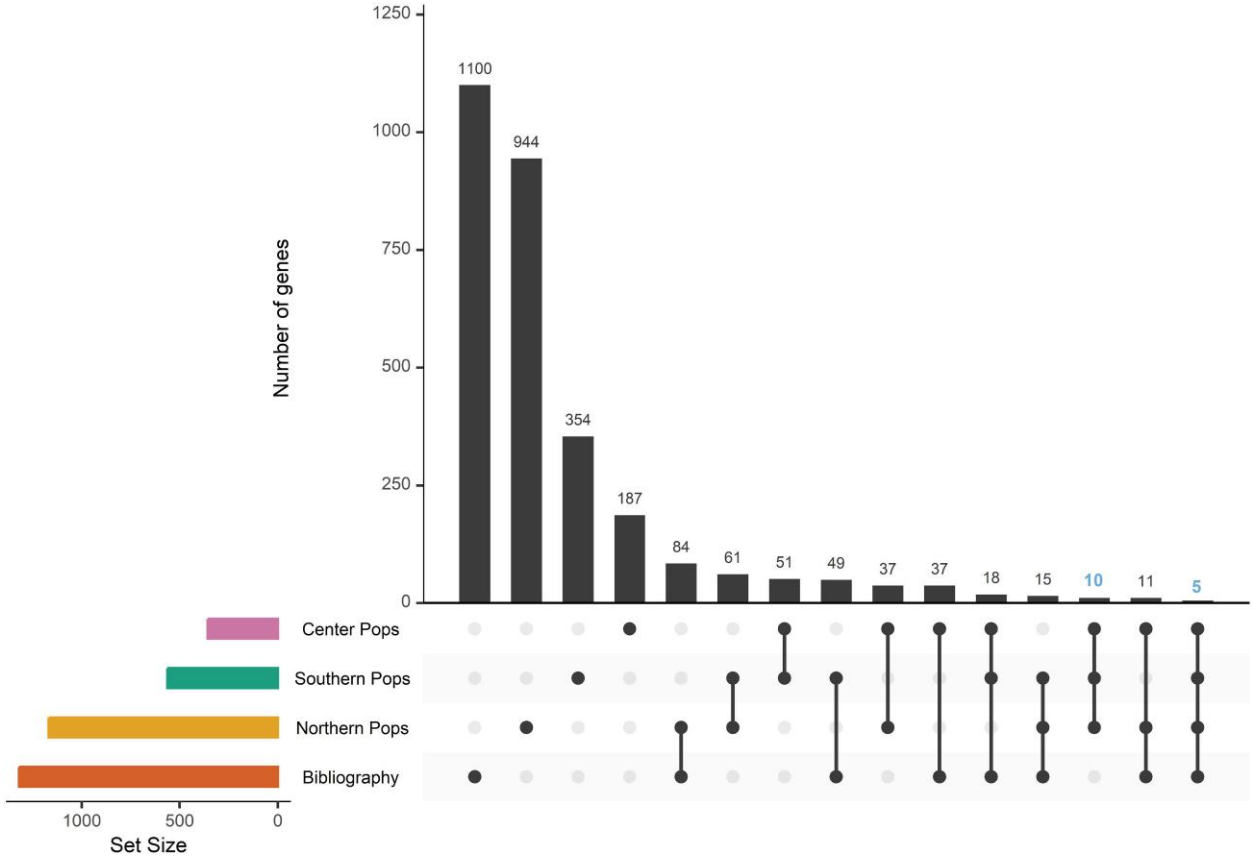

Fig. 4. The intersection of northern, center, and southern MI populations and previously reported genes (bibliography) under putative selection on MI populations. The bar plot (top) shows the number of genes per intersection. The matrix below indicates the intersection represented by each bar. The bar plots on the left show the total number of genes in every population and bibliography. The intersection between three study populations (10) and the intersection between three study populations and the bibliography (5) are depicted in blue.

levels (Aguilar-Ordoñez et al. 2021) (Fig. 1c). To confirm that these were not biasing the variants identified by our selection scans, we conducted local ancestry assignments using GNOMIX. In brief, we trained a model using 1KGP data from YRI, IBS, and unadmixed PEL populations as proxies for African, European, and Native American ancestries, respectively. We filter the results to remove low-confidence ancestry assignments (<90%) and cross-reference the variants under putative selection with the ancestry assignments for those positions. Our results show that putatively selected variants are almost exclusively found in Indigenous American Ancestry segments (northern: 0.99 ± 0.02; central: 0.99 ± 0.01; and southern: 0.99 ± 0.01), confirming that the low levels of admixture reported in the study individuals are not biasing our results (supplementary tables S17 to S19, Supplementary Material online).

Discussion

The genetic structure observed in present-day MI populations reflects a combination of complex demographic and evolutionary events. The results of our study suggest that populations had collectively adapted to metabolic and immune pressures. However, specific selection landscapes were identified in different regions of Mexico.

Northern MI Populations

The northern MI cluster shows nonsynonymous variants putatively under selection on several key genes involved in immune defense (*ID3*, *GBP1*, *XIRP1*, and *MUC19*). Infectious diseases are among the strongest selective pressures shaping the human genome (Karlsson et al. 2014). Demographic events and cultural changes during recent human evolutionary history caused populations to be exposed to new and dangerous pathogens. Specifically, the European Colonization of the Americas represented a major epidemic challenge for MI populations. After the arrival of the Spanish conquistador Hernán Cortés to what is today Mexico, dozens of epidemics swept through the country, leading to a population decline as high as 95% (Vågene et al. 2018). This period also precipitated the introduction of several new pathogens, including the causal agents

Table 3 Genes with significant selection signatures across the three study populations

Gene symbol	Position (GRCh38)	Function	Associated phenotype(s) from GWAS catalog (Sollis et al. 2023)
ARHGAP15	chr2: 143,091,362 to 143,768,352	GTPase activator for the Rho-type GTPases	Lymphocyte count, myeloid white cell count, mean platelet volume
VGLL4	chr3: 11,556,067 to 11,771,350	Negative regulation of cell growth and hippo signaling.	Height, body mass index-adjusted waist–hip ratio, and circumference
LINGO2	chr9: 27,937,617 to 29,213,601	Positive regulation of synapse assembly, a component of membrane	Height, body mass index
SYNDIG1	chr20: 24,469,629 to 24,666,616	Role in postsynaptic development and maturation	Blood protein measurement, Type 2 diabetes mellitus
TFAP2B	chr6: 50,818,355 to 50,847,619	Regulation and transcription of selected genes	Body mass index, weight, height, Char syndrome
CCSER1	chr4: 90,127,394 to 91,605,295	Coiled-Coil Serine Rich Protein	Height, insomnia
KIF2B	chr17: 53,822,927 to 53,825,193	Enable microtubule-binding activity and microtubule motor activity	Body height
LINC00430	chr13: 86,909,585 to 86,937,108	RNA gene	Asthma
LINC01721	chr20: 24,061,057 to 24,226,013	RNA gene	Cystatin, urate, and triglyceride measurements
LINC02546	chr11: 29,586,602 to 29,631,769	RNA gene	Free and bioavailable testosterone levels
LOC101928516	chr6: 74,069,450 to 74,834,752	RNA gene	Height, arthritis
MIR8068	chr11: 28,477,481 to 28,477,548	Posttranscriptional regulation	Bone mineral density
PKHD1	chr6: 51,615,299 to 52,087,615	Promotes ciliogenesis in renal epithelial cells	Body mass index, hair color measurement
SLITRK6	chr13: 85,792,790 to 85,806,683	Regulator of neurite outgrowth required for normal hearing and vision	Body mass index, height, c-reactive protein measurement, Fat body mass
WWOX	chr16: 78,099,400 to 79,212,667	Tumor suppressor gene	Height, multiple sclerosis, SARS, COVID-19

Genes previously documented in the bibliography are shaded in gray for reference.

responsible for well-characterized diseases, such as measles, mumps, smallpox, and influenza (Vågene et al. 2018).

Interestingly, both *GBP1* and *XIRP1* are involved in the immune response to *Salmonella* (Fisch et al. 2019; Urbano et al. 2022). *GBP1* encodes a large GTPase of the dynamin superfamily involved in antimicrobial immunity and cell death. Specifically, *GBP1* promotes the detection of *Toxoplasma* DNA and the immune targeting of *Salmonella* (Fisch et al. 2019). *XIRP1* is a protein-coding gene expressed mainly in fibroblasts and macrophages in response to cytokines and bacterial infections such as *Listeria*, *Shigella*, and *Salmonella* (Urbano et al. 2022). A recent study by Vågene et al. (Acuna-Soto et al. 2002) identified *Salmonella enterica* as a possible cause of the 1545 to 1550 "cocoliztli" epidemic in Southern Mexico. Historically "cocoliztli" epidemics were reported in the northern and central high valleys of Mexico, in addition to Southern Mexico (Acuna-Soto et al. 2002), suggesting that this pathogen may have been a major selective pressure for northern MI populations during the colonial period. While putatively selected variants are linked to the immune response to *Salmonella*, it is crucial to highlight that the exact causal agent of cocoliztli remains unidentified, and caution should be taken in interpreting the results.

Similarly, *MUC19* (Mucin 19) encodes a Mucin family protein that has been involved in the immune response to parasitic and viral infections (Hicks et al. 2000). The *MUC19* gene that we identify as putatively under selection in the northern cluster has been previously identified under selection in both populations from central Mexico (Reynolds et al. 2019) and individuals of Mexican heritage living in Los Angeles (MXL) from the 1KGP (Witt et al. 2023). In a recent article, Villanea et al. (2023) described the presence of a *Denisovan-like MUC19* haplotype at high frequencies in admixed Latin American individuals. Their results suggest that this haplotype served as raw genetic material for positive selection in American populations.

Central MI Populations

The central MI cluster shows variants under selection in the Atypical Chemokine Receptor 1 (*ACKR1*) and *DNAJB7*

genes. *ACKR1*, formerly referred to as DARC (Duffy antigen receptor for chemokines), encodes a glycoprotein expressing the Duffy blood group antigens. The Duffy protein acts as a receptor for the malaria parasites *Plasmodium vivax* and *Plasmodium knowlesi* (Yin et al. 2018). Malaria, a global infectious disease caused by *Plasmodium* spp. and spread to humans by some types of mosquitoes, has been one of the most important selective pressures on the human lineage (Tennessen and Duraisingh 2020). *ACKR1* has previously been reported as a target of directional positive selection in sub-Saharan African populations with resistance to erythrocyte invasion by *P. vivax* (Werren et al. 2021). Interestingly, our center MI cluster shows several variants under putative selection in *ACKR1*.

In Mexico, malaria continues to pose a significant risk for indigenous peoples living in Chiapas and Oaxaca, with *P. vivax* being the dominant *Plasmodium* all over the country (Organización Mundial de la Salud 2018). Based on the World Health Organization Malaria Report, a cumulative total of 8,157 indigenous malaria cases in Mexico were recorded from 2010 to 2022 (World Health Organization 2022). How and when *P. vivax* arrived in the Americas remains highly controversial; however, most theories propose pre- and post-Colonial exposures with Asian and European contributions to the genetic diversity of the parasite (Rodrigues et al. 2018; Wiscovitch-Russo et al. 2019). Our results from the central study cluster suggest that malaria may be driving variation at *ACKR1*. Nevertheless, present-day American indigenous populations do not show evidence of protective phenotypes for malaria—sickle cell trait, glucose-6-phosphate dehydrogenase deficiency, and *ACKR1* negative expression—so further work is required to determine the functional impact, if any, of the putatively selected allele (Rodrigues et al. 2018).

DNAJB7 (DnaJ Heat Shock Protein Family [Hsp40] Member B7) is an intronless gene that belongs to the DNAJ/HSP40 family of proteins, which regulate chaperone activity (Stelzer et al. 2016). Heat shock proteins (Hsps) play a vital role in cell homeostasis under both physiological and stressful conditions (van Eden et al. 2005; Kotlarz et al. 2013). Particularly, the DnaJ Heat Shock Protein Family, or Hsp40, constitutes the largest and most diverse subgroup of Hsp families, playing fundamental roles in neurodegeneration, tumorigenesis, glucose homeostasis, and spermatogenesis (Diane et al. 2021; Bai et al. 2023). *DNAJB7*, a gene with a nonsynonymous mutation under selection in the central MI populations, is a member of the DnaJ family that has been related to the pathogenesis of insulin resistance and Type 2 diabetes (T2D) mellitus (Diane et al. 2021).

Genes related to metabolism and, more specifically, T2D have previously been reported under putative selection in MI populations as a result of adaptation to periodic starvation experienced throughout history (Watve and Yajnik 2007; García-Ortiz et al. 2022). The rapid change in diet and behavior in the contemporary MI populations against a genotype that cannot adapt rapidly creates an evolutionary mismatch expressed as metabolic diseases (Gluckman et al. 2016). Additionally, *DNAJB7* is highly expressed in the testis, suggesting a possible role in male fertility (Uhlén et al. 2015). Considering these observations, selection under *DNAJB7* may result from metabolic or reproductive selective pressures.

Southern MI Populations

The southern MI cluster shows variants under putative selection in the V-Set Pre-B Cell Surrogate Light Chain 1 (*VPREB1*), *NCKAP5L*, and *ID3* genes. *VPREB1* encodes a protein belonging to the immunoglobulin superfamily, expressed at the early stages of immune B-cell development (Stelzer et al. 2016). *NCKAP5L* is a gene that regulates microtubule organization, stabilization, growth, and bunding formation (Stelzer et al. 2016). The gene *ID3* encodes a transcription factor of the helix–loop–helix family reported to be involved in the differentiation and growth of a variety of cell types (18). The coding product of *ID3* controls the formation of effector and memory CD8+ populations, a critical process for adaptive immunity (25). As in northern and center MI populations, pathogenic environments seem to be a major selective pressure for southern MI populations. To the best of our knowledge, *VPREB1* and *ID3* are two genes heavily involved in different immune-related pathways that do not have reports of pathogenic specificity.

On the other hand, *NCKAP5L* is a gene that has been associated with *Spherocytosis* Type 2 (SPH2) (OMIM #616649) (Stelzer et al. 2016), a hereditary disease characterized by the presence of spherical-shaped erythrocytes on the peripheral blood smear and a possible protective phenotype against *Plasmodium falciparum* (Hamosh et al. 2005; Goheen, Campino, and Cerami 2017). The deadliest of the human malaria parasites, *P. falciparum*, presumably entered the Americas after European contact as a result of the trans-Atlantic slave trade (Rodrigues et al. 2018). During the colonial period in Mexico, the slave trade was primarily through the seaports of Veracruz, Campeche, and Guerrero (Laviña 1994; Serna y Herrera 2016) leading to sugar plantations and other extractive industries in the Yucatan and beyond (Zabala 2010). This history may suggest that southern MI populations have had extended exposure and potentially selective pressures from *P. falciparum* infection than other regions of Mexico.

Shared Selection Signatures Between MI Populations

When comparing signals of selection across our study regions, we identified ten (*CCSER1*, *KIF2B*, *LINC00430*, *LINC01721*, *LINC02546*, *LOC101928516*, *MIR8068*, *PKHD1*, *SLITRK6*, and *WWOX*) genes with selection signatures that were present in all our three study populations and five genes that, besides being present, had previous

reports of selection on MI populations (*ARHGAP15*, *VGLL4*, *LINGO2*, *SYNDIG1*, and *TFAP2B*). Our results showed an overrepresentation of genes involved in metabolic and immune pathways.

ARHGAP15 is a gene that encodes an RHO GTPase-activating protein highly expressed in immune reservoir tissues (Fagerberg et al. 2014). Particularly, neutrophils in mice with an *ARHGAP15* knockout showed an increase in bone marrow retention (Campa et al. 2016). Under healthy conditions, most neutrophils are located in the bone marrow (retention). However, during infection, they are rapidly released into the bloodstream to combat pathogens (egress). For both pathways to work, common intracellular signaling elements, such as (GTPases) of the Rho family, are essential. This information confirms the role of *ARHGAP15* in immune cell mobilization. Our results on regulatory elements also show some variants under selection on enhancer elements connected to *ARHGAP15*. Moreover, selection in this gene has previously been reported in MI populations (Reynolds et al. 2019; Ávila-Arcos et al. 2020; García-Ortiz 2022).

Genome-wide association studies (GWASs) showed that different SNPs in *LINGO2* were associated with obesity, T2D, and gestational diabetes mellitus risk (Su et al. 2019). Moreover, *LINGO2* has been related to immunity against helminth infection (Costa et al. 2011; Belle et al. 2019). *VGLL4*, *TFAP2B*, and *WWOX* have been implicated in increased body mass index, insulin resistance, lipid, and triglycerides levels and increased risk for T2D (Dennis et al. 2003; Nordquist et al. 2009; Aldaz, Ferguson, and Abba 2014). Since metabolic phenotypes under selection result from physiological adaptation to the available food supplies, we hypothesize that metabolic adaptation may have occurred either before the divergence of our three clusters. Signals of selection in *LINGO2* and *ARHGAP15* may also be due to selective pressures from the pathogens brought to Mexico by the conquistadors.

Selection in Regulatory Elements

Using the additional data afforded us by WGS data, we investigated signals of selection within enhancer and promoter sequences, to better understand the selective landscape of regulatory elements in MI populations. Our results show that most of the altered enhancer elements of interest are connected to noncoding features, with notable exceptions like *ARHGAP15*, *GTDC1*, *SYNDIG1*, *GGTLC1*, *MEI1*, *ADSL*, *TOB2*, and *NALF1*.

GTDC1 is a glycosyltransferase-like encoding gene highly expressed in the brain, fat, and testis. This gene, predicted to enable glycosyltransferase activity, has been associated with different neurodevelopmental disorders (Aksoy et al. 2017) and *Mycobacterium tuberculosis* infection resistance (Möller et al. 2018). *SYNDIG1* encodes an interferon-induced transmembrane protein, highly expressed in the brain and muscle (Fagerberg et al. 2014). *GGTLC1* encodes a gamma-glutamyl transpeptidase interferon-induced transmembrane, highly expressed in the lung and testis (Stelzer et al. 2016). *MEI1*, targeted by five different enhancers with SNPs under putative selection, is a protein-coding gene predicted to be involved in gamete generation, meiotic spindle organization, and meiotic telomere clustering. Polymorphic alleles of *MEI1* have been linked to human azoospermia, a medical condition characterized by the absence of sperm in the ejaculate due to meiotic arrest (Sato et al. 2006). Furthermore, this gene was previously identified as belonging to the human phenotype ontology (Gargano et al. 2024) category "female infertility." These associations suggest putative selection on regulatory regions associated with reproductive phenotypes.

Conclusions

This analysis represents the first whole-genome selection scan performed on MI populations. Altogether, our study found signals of selection in metabolic and immune-related genes. However, specific selection landscapes were identified in different regions of Mexico. For the northern MI cluster, we found evidence of selection on genes involved in the specific immune response to *Salmonella* and the unspecific immune response to different infectious diseases. Both central and southern clusters exhibit a selection landscape putatively shaped by different malaria strains and other pathogens.

To aid in this and future studies of natural selection, we also presented nf-selection, a pipeline that allows the computing of PBS and iHS to detect selection signals in diploid organisms. The developed pipeline offers a friendly shortcut to the reproducible and organized application of the bioinformatics code required to apply both statistics. Particularly, the use of Nextflow as a workflow orchestration engine in the pipeline provides the possibility of parallelization to speed up analyses. We showed that nf-selection represents a practical tool for the detection of selection signals, and we expect that it will prove useful for the research community.

We recognize several limitations of the current study. Firstly, the number of whole genomes available for analysis is limited for multiple ethnic indigenous groups. This may restrict the representation of the 27 Mexican populations in the findings described. Additionally, the significant genetic drift exhibited by the Seri of the northern MI group poses a challenge in establishing robust conclusions for the northern cluster. Previously, García-Ortiz et al. (2021) conducted effective population size and divergence time estimation for 716 individuals from 60 MI ethnic groups also belonging to the MAIS cohort, including the samples

analyzed in the current study. Those results showed a small long-term N_e that may have contributed to the pattern of greater genetic drift and lower genetic diversity observed in the Seris. Moreover, an inference of autozygosity using runs of homozygosity (ROH) showed that the Seri had the highest proportion of the genome in ROH compared to other MI populations, suggesting that the high genetic differentiation in this population was caused by drift (García-Ortiz et al. 2021). This is supported by our own F_{ST} analyses that show a relatively large genetic distance between the Seri and all other groups included in our analyses. While the demographic model used in nf-selection has allowed for a bottleneck in the simulated populations based on the empirical data, together, this evidence of substantial genetic drift in the Seri necessitates careful consideration when discussing the region-specific results for the northern group.

Secondly, while several signals of selection have been identified on immune genes that may be associated with the recent colonial history of Mexico, future simulation work is needed to estimate the timing and strength of selection on these loci. Finally, as with all studies of genome-wide selection scans, our results only provide the first step to understanding the phenotype resulting in selective pressure. Future work with functional genomics and deep phenotyping will be required to understand the biological consequences of putatively selected variants fully. Despite these limitations, this research provides valuable insight into the adaptation processes that MI populations experienced in the recent past, as well as an easy to use computational pipeline for selection analysis, paving the way for further WGS selection analysis.

Materials and Methods

Samples

Samples in this study were previously reported by Aguilar-Ordoñez et al. (2021). In brief, 76 individuals from the MAIS cohort were whole-genome sequenced. Individuals are representatives of 27 different ethnic groups in Mexico and show a high average of similarity to populations reported as Native American by ADMIXTURE (Aguilar-Ordoñez et al. 2021).

Quality Control and Data Treatment

Whole-genome data in VCF format were filtered with bcftools v1.8 to obtain data from autosomal chromosomes. The file was merged with autosomal whole-genome data from 33 Peruvian individuals (PEL) without any evidence of recent European or African genetic ancestry selected from the 1KGP and with 50 Han Chinese individuals (CHB) from the 1KGP using bcftools v1.8. The merged VCF was filtered to keep positions with at most 0.8% missing data and minor allele frequency (MAF) > 0.05. Using these criteria, the final dataset for selection analysis contained 4,624,975 variants.

Global Ancestry Estimation, PCA, and F_{st} Computation

Whole-genome data in VCF format was filtered with bcftools v1.8 to obtain data from autosomal chromosomes. The file was merged with autosomal whole-genome data from 50 PEL, 50 CHB, 50 Yoruba individuals (YRI), and 50 British individuals (IBS) using bcftools v1.8. The merged VCF was transformed into a PLINK file format and filtered to keep positions with at most 0.8% missing data and MAF ≥ 0.05 using PLINK 1.9 (Purcell et al. 2007). The filtered positions were LD pruned (window size: 50, step size: 10, r^2 threshold: 0.1) and used as input for regional and global PCA analysis with PLINK 1.9. Global ancestry estimation was computed with ADMIXTURE (Alexander et al. 2009) following a ten-run cross-validation approach for $K = 3$ to 7, and results were plotted with PONG (Behr et al. 2016).

To estimate variation within and between cohorts, the final selection dataset was split into individual VCF files for each of the 27 populations. For every pairwise combination, the weighted mean Weir and Cockerham F_{ST} was calculated using vcftools (Danecek et al. 2011). Mean F_{ST} values and Wilcoxon tests were calculated using R to assess significant differences within and between cohorts.

Selection Pipeline

To simplify the detection of selection signatures, we developed nf-selection, a Nextflow pipeline that detects candidate genomic regions and genes under selection by computing two commonly used statistics: PBS and iHS. The analysis starts with haplotype phasing with SHAPEIT 4.2 (Delaneau et al. 2019), followed by ancestral allele annotation using fasta ancestral files and the JavaScript vcfancestralalleles.jar from Jvarkit (https://github.com/lindenb/jvarkit). PBS calculation is implemented in unfiltered ancestral/derived variants with an in-house R script using Wright Fst values obtained from vcftools (Danecek et al. 2011). The calculation of iHS is carried out using the R package rehh (Gautier and Vitalis 2012), using filtered ancestral/derived variants for iHS. Genome-wide *P*-values are calculated separately for PBS using a distribution simulated under a demographic model specific to the population of interest. Genome-wide *P*-values for iHS are retrieved from the rehh output. To reduce false positives, we follow the cross-referenced approach of Reynolds et al. (2019), cross-referencing between the top 1% of PBS and iHS. Cross-referenced SNPs are annotated using ANNOVAR. These results can be used for downstream analyses of gene curation, gene-set overrepresentation, gene networks, etc.

Demographic Model

A demographic model was modified from Reynolds et al. (2019) to calculate expected distributions of PBS under neutral demographic processes for our three study regions using fastsimcoal2 (version fsc27.09). Briefly, authors retrieved previously published estimates for the timing of the Out-of-Africa bottleneck and peopling of the Americas events (timing of the bottleneck coinciding with the peopling of the Americas, time of the population divergence between PEL and the study population, timing of a possible bottleneck for the study population) (Reynolds et al. 2019; Ávila-Arcos et al. 2020) (supplementary fig. S11, Supplementary Material online). Joint site frequency spectrum of the CHB, PEL, and the study populations was calculated using easySFS (https://github.com/isaacovercast/easySFS). We included the divergence time between the PEL and our study population as an open parameter ranging between 480 and 520 generations. Additionally, the timing and severity of the most recent bottleneck were left as an open parameter in the model, as genetic and anthropological evidence for many indigenous populations remains uncharacterized. Unknown parameters for the MI Populations were estimated by running the model 100 times with 1,000,000 iterations per run. The best likelihood run was chosen and used to simulate 1,000,000 sites across 22 chromosomes for 100 CHB, 66 PEL, and 152 MI individuals. The simulation was done 100 times, and variant sites from one randomly selected simulation were chosen to calculate PBS. Values of PBS were used to form a distribution for comparison with the observed empirical values.

Whole-Genome Scan for Selection Signatures

Using our pipeline, we computed PBS and iHS for the three Mexican regions. We calculated the PBS for each autosomal variant in each population using 33 unadmixed Peruvians from the 1KGP as an ingroup and 50 Han Chinese individuals from the 1KGP as an outgroup. The pipeline was run with default settings. Moreover, *P*-values for PBS for each region were calculated using a distribution of PBS values under a neutral demographic model. *P*-values for iHS were extracted from the rehh output. The top 1% of *P*-values for PBS and iHS were identified and cross-referenced to keep only entries that were present in both statistics. This cross-referenced approach has been used previously to avoid false positive results (Reynolds et al. 2019).

Detection of Variants in Enhancers

We detected which SNPs are located in an enhancer or promoter element for each regional selection signal dataset using a pipeline that cross-referenced variant positions with the GeneHancer database (v5.17, downloaded September 1, 2023). The pipeline can be downloaded from https://github.com/laguilaror/nf-vcf2genehancer. In brief, using R scripts, the GeneHancer v5.17 GFF file is converted to a pseudo-BED format to keep the connected genes information, and then the query VCF file is cross-referenced with the pseudo-BED using bedtools intersect (v2.31.0); finally, two summaries of variation are created: one for variants per GeneHancer element and one for variants per genomic feature (a GeneHancer element can regulate more than one genomic feature, and each genomic feature can be regulated by more than one GeneHancer element). The process was run independently for the north, central, and south datasets.

Local Ancestry Assignment

We used GNOMIX (Hilmarsson et al. 2021) for local ancestry inference. Data from YRI, IBS, and unadmixed PEL populations from the 1KGP were used as proxies for African, European, and Native American ancestries for model training. The results were filtered to remove low-confidence ancestry assignments (<90%) and cross-referenced with the final dataset of putatively selected variants for each population. A mean ancestry proportion was calculated for the final results using R.

Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

Acknowledgments

We thank the communities involved in this research and the participants who provided samples for analysis. We thank Olvera-Acosta Ricardo Benjamin and Orozco-Flores Diego from the Expression Analysis Subdirectorate at the National Institute of Genomic Medicine (INMEGEN) for helping organize the data and Denise García Castro, Andrea Arriola-Gamboa, and José de Jesús Mares Guerra for curating the data. This work was partially performed at cluster INMEGEN, which receives technical support from Gómez-Romero Laura.

Data Availability

New genomic data were not produced. Ethics statement and access to data can be found at the original data source (Aguilar-Ordoñez et al. 2021). The Nextflow pipeline to reproduce this work is publicly available at: https://github.com/fernanda-miron/nf-selection.

Literature Cited

Acuna-Soto R, Stahle DW, Cleaveland MK, Therrell MD. Megadrought and megadeath in 16th century Mexico. Emerg Infect Dis. 2002:8(4):360–362. https://doi.org/10.3201/eid0804.010175.

Aguilar-Ordoñez I, Pérez-Villatoro F, García-Ortiz H, Barajas-Olmos F, Ballesteros-Villascán J, González-Buenfil R, Fresno C, Garcíarrubio A, Fernández-López JC, Tovar H, et al. Whole genome variation in 27 Mexican indigenous populations, demographic and biomedical insights. PLoS One. 2021:16(4):e0249773. https://doi.org/10.1371/journal.pone.0249773.

Aksoy I, Utami KH, Winata CL, Hillmer AM, Rouam SL, Briault S, Davila S, Stanton LW, Cacheux V. Personalized genome sequencing coupled with iPSC technology identifies GTDC1 as a gene involved in neurodevelopmental disorders. Hum Mol Genet. 2017:26(2):367–382. https://doi.org/10.1093/hmg/ddw393.

Aldaz CM, Ferguson BW, Abba MC. WWOX at the crossroads of cancer, metabolic syndrome related traits and CNS pathologies. Biochim Biophys Acta. 2014:1846(1):188–200. https://doi.org/10.1016/j.bbcan.2014.06.001.

Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009:19:1655–1664. https://doi.org/10.1101/gr.094052.109.

Amorim CE, Nunes K, Meyer D, Comas D, Bortolini MC, Salzano FM, Hünemeier T. Genetic signature of natural selection in first Americans. Proc Natl Acad Sci U S A. 2017:114(9):2195–2199. https://doi.org/10.1073/pnas.1620541114.

Ávila-Arcos MC, McManus KF, Sandoval K, Rodríguez-Rodríguez JE, Villa-Islas V, Martin AR, Luisi P, Peñaloza-Espinosa RI, Eng C, Huntsman S, et al. Population history and gene divergence in native Mexicans inferred from 76 human exomes. Mol Biol Evol. 2020:37(4):994–1006. https://doi.org/10.1093/molbev/msz282.

Bai S, Hu M, Yu L, Chen L, Zhou J, Wu L, Xu B, Jiang X, Zhang X, Tong X, et al. DNAJB7 is dispensable for male fertility in mice. Reprod Biol Endocrinol. 2023:21(1):32. https://doi.org/10.1186/s12958-023-01086-6.

Behr AA, Liu KZ, Liu-Fang G, Nakka P, Ramachandran S. pong: fast analysis and visualization of latent clusters in population genetic data. Bioinformatics. 2016:32(18):2817–2823. https://doi.org/10.1093/bioinformatics/btw327.

Belle NM, Ji Y, Herbine K, Wei Y, Park J, Zullo K, Hung L-Y, Srivatsa S, Young T, Oniskey T, et al. TFF3 interacts with LINGO2 to regulate EGFR activation for protection against colitis and gastrointestinal helminths. Nat Commun. 2019:10(1):4408. https://doi.org/10.1038/s41467-019-12315-1.

Campa CC, Germena G, Ciraolo E, Copperi F, Sapienza A, Franco I, Ghigo A, Camporeale A, Di Savino A, Martini M, et al. Rac signal adaptation controls neutrophil mobilization from the bone marrow. Sci Signal. 2016:9(459):ra124. https://doi.org/10.1126/scisignal.aah5882.

Campelo dos Santos AL, Owings A, Sullasi HSL, Gokcumen O, DeGiorgio M, Lindo J. 2022. Genomic evidence for ancient human migration routes along South America's Atlantic coast. Proc Biol Sci 289(1986): 20221078. https://doi.org/10.1098/rspb.2022.1078.

Chen S, Francioli LC, Goodrich JK, Collins RL, Kanai M, Wang Q, Alföldi J, Watts NA, Vittal C, Gauthier LD, et al. A genomic mutational constraint map using variation in 76,156 human genomes. Nature. 2024:625:92–100. https://doi.org/10.1038/s41586-023-06045-0.

Chu D, Wei L. Nonsynonymous, synonymous and nonsense mutations in human cancer-related genes undergo stronger purifying selections than expectation. BMC Cancer. 2019:19(1):359. https://doi.org/10.1186/s12885-019-5572-x.

Costa C, Germena G, Martin-Conte EL, Molineris I, Bosco E, Marengo S, Azzolino O, Altruda F, Ranieri VM, Hirsch E. The RacGAP ArhGAP15 is a master negative regulator of neutrophil functions. Blood. 2011:118(4):1099–1108. https://doi.org/10.1182/blood-2010-12-324756.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. The variant call format and VCFtools. Bioinformatics. 2011:27(15):2156–2158. https://doi.org/10.1093/bioinformatics/btr330.

Delaneau O, Zagury J-F, Robinson MR, Marchini JL, Dermitzakis ET. Accurate, scalable and integrative haplotype estimation. Nat Commun. 2019:10(1):5436. https://doi.org/10.1038/s41467-019-13225-y.

Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol. 2003:4(9):R60. https://doi.org/10.1186/gb-2003-4-9-r60.

Diane A, Abunada H, Khattab N, Moin ASM, Butler AE, Dehbi M. Role of the DNAJ/HSP40 family in the pathogenesis of insulin resistance and Type 2 diabetes. Ageing Res Rev. 2021:67:101313. https://doi.org/10.1016/j.arr.2021.101313.

Fagerberg L, Hallström BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, Habuka M, Tahmasebpoor S, Danielsson A, Edlund K, et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics*. Mol Cell Proteomics. 2014:13(2):397–406. https://doi.org/10.1074/mcp.M113.035600.

Fan S, Hansen MEB, Lo Y, Tishkoff SA. Going global by adapting local: a review of recent human adaptation. Science. 2016:354(6308):54–59. https://doi.org/10.1126/science.aaf5098.

Fisch D, Bando H, Clough B, Hornung V, Yamamoto M, Shenoy AR, Frickel E-M. Human GBP1 is a microbe-specific gatekeeper of macrophage apoptosis and pyroptosis. EMBO J. 2019:38(13):e100926. https://doi.org/10.15252/embj.2018100926.

Garcia OA, Arslanian K, Whorf D, Thariath S, Shriver M, Li JZ, Bigham AW. The legacy of infectious disease exposure on the genomic diversity of indigenous southern Mexicans. Genome Biol Evol. 2023:15(3):evad015. https://doi.org/10.1093/gbe/evad015.

García-Ortiz H, Barajas-Olmos F, Contreras-Cubas C, Cid-Soto MÁ, Córdova EJ, Centeno-Cruz F, Mendoza-Caamal E, Cicerón-Arellano I, Flores-Huacuja M, Baca P, et al. The genomic landscape of Mexican indigenous populations brings insights into the peopling of the Americas. Nat Commun. 2021:12(1):5942. https://doi.org/10.1038/s41467-021-26188-w.

García-Ortiz H, Barajas-Olmos F, Contreras-Cubas C, Reynolds AW, Flores-Huacuja M, Snow M, Ramos-Madrigal J, Mendoza-Caamal E, Baca P, López-Escobar TA, et al. Unraveling signatures of local adaptation among indigenous groups from Mexico. Genes (Basel). 2022:13(12):2251. https://doi.org/10.3390/genes13122251.

Gargano MA, Matentzoglu N, Coleman B, Addo-Lartey EB, Anagnostopoulos AV, Anderton J, Avillach P, Bagley AM, Bakštein E, Balhoff JP, et al. The human phenotype ontology in 2024: phenotypes around the world. Nucleic Acids Res. 2024:52(D1):D1333–D1346. https://doi.org/10.1093/nar/gkad1005.

Gautier M, Vitalis R. Rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. Bioinformatics. 2012:28(8):1176–1177. https://doi.org/10.1093/bioinformatics/bts115.

Gluckman P, Beedle A, Buklijas T, Low F, Hanson M. Principles of evolutionary medicine. 2nd ed. Oxford (New York): Oxford University Press; 2016.

Goheen MM, Campino S, Cerami C. The role of the red blood cell in host defence against falciparum malaria: an expanding repertoire of evolutionary alterations. Br J Haematol. 2017:179(4):543–556. https://doi.org/10.1111/bjh.14886.

Goodwin S, McPherson JD, Richard McCombie W. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet. 2016:17(6):333–351. https://doi.org/10.1038/nrg.2016.49.

Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian inheritance in man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res. 2005:33(Database issue):D514–D517. https://doi.org/10.1093/nar/gki033.

Hicks SJ, Theodoropoulos G, Carrington SD, Corfield AP. The role of mucins in host–parasite interactions. Part I—protozoan parasites. Parasitol Today. 2000:16(11):476–481. https://doi.org/10.1016/s0169-4758(00)01773-7.

Hilmarsson H, Kumar SA, Rastogi R, Bustamante CD, Mas-Montserrat D, Ioannidis AG. High resolution ancestry deconvolution for next generation genomic data. bioRxiv. https://doi.org/10.1101/2021.09.19.460980, 2021, preprint: not peer reviewed.

Horscroft C, Ennis S, Pengelly RJ, Sluckin TJ, Collins A. Sequencing era methods for identifying signatures of selection in the genome. Brief Bioinform. 2019:20(6):1997–2008. https://doi.org/10.1093/bib/bby064.

Karlsson EK, Kwiatkowski DP, Sabeti PC. Natural selection and infectious disease in human populations. Nat Rev Genet. 2014:15(6):379–393. https://doi.org/10.1038/nrg3734.

Kotlarz A, Tukaj S, Krzewski K, Brycka E, Lipinska B. Human Hsp40 proteins, DNAJA1 and DNAJA2, as potential targets of the immune response triggered by bacterial DnaJ in rheumatoid arthritis. Cell Stress Chaperones. 2013:18(5):653–659. https://doi.org/10.1007/s12192-013-0407-1.

Laviña J. Somos Indios y somos Negros, somos Mexicanos: la población Afromestiza de la Costa de Guerrero. Hist Fuente Oral. 1994:11:97–106. https://www.jstor.org/stable/27753426.

Mendoza-Revilla J, Chacón-Duque JC, Fuentes-Guajardo M, Ormond L, Wang K, Hurtado M, Villegas V, Granja V, Acuña-Alonzo V, Jaramillo C, et al. Disentangling signatures of selection before and after European colonization in Latin Americans. Mol Biol Evol. 2022:39(4):msac076. https://doi.org/10.1093/molbev/msac076.

Möller M, Kinnear C, Orlova M, Kroon E, van Helden PD, Schurr E, Hoal EG. Genetic resistance to mycobacterium tuberculosis infection and Disease. Front Immunol. 2018:9. https://doi.org/10.3389/fimmu.2018.02219.

Nordquist N, Göktürk C, Comasco E, Eensoo D, Merenäkk L, Veidebaum T, Oreland L, Harro J. The transcription factor TFAP2B is associated with insulin resistance and adiposity in healthy adolescents. Obesity (Silver Spring). 2009:17(9):1762–1767. https://doi.org/10.1038/oby.2009.83.

Ojeda-Granados C, Abondio P, Setti A, Sarno S, Gnecchi-Ruscone GA, González-Orozco E, De Fanti S, Jiménez-Kaufmann A, Rangel-Villalobos H, Moreno-Estrada A, et al. Dietary, cultural, and pathogens-related selective pressures shaped differential adaptive evolution among native Mexican populations. Mol Biol Evol. 2022:39(1):msab290. https://doi.org/10.1093/molbev/msab290.

Organización Mundial de la Salud. 2018. E-2020 Ficha Del País: México. Organización Mundial de la Salud; 2018. [accessed 2024 Jul] https://apps.who.int/iris/handle/10665/272774.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007:81(3):559–575. https://doi.org/10.1086/519795.

Reynolds AW, Mata-Míguez J, Miró-Herrans A, Briggs-Cloud M, Sylestine A, Barajas-Olmos F, Garcia-Ortiz H, Rzhetskaya M, Orozco L, Raff JA, et al. Comparing signals of natural selection between three indigenous North American populations. Proc Natl Acad Sci U S A. 2019:116(19):9312–9317. https://doi.org/10.1073/pnas.1819467116.

Rodrigues PT, Valdivia HO, de Oliveira TC, Alves JM, Duarte AMRC, Cerutti-Junior C, Buery JC, Brito CFA, de Souza Jr. JC, Hirano ZMB, et al. Human migration and the spread of malaria parasites to the New World. Sci Rep. 2018:8. https://doi.org/10.1038/s41598-018-19554-0.

Sato H, Miyamoto T, Yogev L, Namiki M, Koh E, Hayashi H, Sasaki Y, Ishikawa M, Lamb DJ, Matsumoto N, et al. Polymorphic alleles of the human MEI1 gene are associated with human azoospermia by meiotic arrest. J Hum Genet. 2006:51(6):533–540. https://doi.org/10.1007/s10038-006-0394-5.

Serna y Herrera JDdl, Negros, mulatos y pardos en la historia de Veracruz. Negros, mulatos y pardos en la historia de Veracruz. Arqueol Mex. 2016:119:52–57.

Skoglund P, Reich D. A genomic view of the peopling of the Americas. Curr Opin Genet Dev. 2016:41:27–35. https://doi.org/10.1016/j.gde.2016.06.016.

Sollis E, Mosaku A, Abid A, Buniello A, Cerezo M, Gil L, Groza T, Güneş O, Hall P, Hayhurst J, et al. The NHGRI-EBI GWAS catalog: knowledgebase and deposition resource. Nucleic Acids Res. 2023:51(D1):D977–D985. https://doi.org/10.1093/nar/gkac1010.

Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, Stein TI, Nudel R, Lieder I, Mazor Y, et al. The GeneCards suite: from gene data mining to disease genome sequence analyses. Curr Protoc Bioinformatics. 2016:54(1):1.30.1–1.30.33. https://doi.org/10.1002/cpbi.5.

Su T, Ren Q, Lu Y, Tai W, Zhu Y, Li Z, Wen J, Hu L, Zhang L, Ma J. A genetic variant in LINGO2 contributes to the risk of gestational diabetes mellitus in a Chinese population. J Cell Physiol. 2019:234(5):7012–7018. https://doi.org/10.1002/jcp.27454.

Tennessen JA, Duraisingh MT. Three signatures of adaptive polymorphism exemplified by malaria-associated genes. Mol Biol Evol. 2020:38(4):1356–1371. https://doi.org/10.1093/molbev/msaa294.

Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, et al. Tissue-based map of the human proteome. Science. 2015:347(6220):1260419. https://doi.org/10.1126/science.1260419.

Urbanoo R, Park E-S, Tretina K, Tunaru A, Gaudet RG, Hu X, Wang D-Z, MacMicking JD. 2022. Human XIRP1 is a macrophage podosome protein utilized by listeria for actin-based motility. bioRxiv 2022.08.28.505595. https://doi.org/10.1101/2022.08.28.505595, 28 August 2022, preprint: not peer reviewed.

Vågene ÅJ, Herbig A, Campana MG, Robles García NM, Warinner C, Sabin S, Spyrou MA, Andrades Valtueña A, Huson D, Tuross N, et al. *Salmonella enterica* genomes from victims of a major sixteenth-century epidemic in Mexico. Nat Ecol Evol. 2018:2(3):520–528. https://doi.org/10.1038/s41559-017-0446-6.

van Eden W, van der Zee R, Prakken B. Heat-shock proteins induce T-cell regulation of chronic inflammation. Nat Rev Immunol. 2005:5(4):318–330. https://doi.org/10.1038/nri1593.

Villanea FA, Peede D, Kaufman EJ, Añorve-Garibay V, Witt KE, Villa-Islas V, Zeloni R, Marnetto D, Moorjani P, Jay F, et al. 2023. The MUC19 gene in denisovans, Neanderthals, and modern humans: an evolutionary history of recurrent introgression and natural selection. bioRxiv 2023.09.25.559202. https://doi.org/10.1101/2023.09.25.559202, 25 September 2023, preprint: not peer reviewed.

Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. PLoS Biol. 2006:4(3):e72. https://doi.org/10.1371/journal.pbio.0040072.

Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010:38(16):e164. https://doi.org/10.1093/nar/gkq603.

Watve MG, Yajnik CS. Evolutionary origins of insulin resistance: a behavioral switch hypothesis. BMC Evol Biol. 2007:7(1):61. https://doi.org/10.1186/1471-2148-7-61.

Werren EA, Garcia O, Bigham AW. Identifying adaptive alleles in the human genome: from selection mapping to functional validation.

Hum Genet. 2021:140(2):241–276. https://doi.org/10.1007/s00439-020-02206-7.
Willerslev E, Meltzer DJ. Peopling of the Americas as inferred from ancient genomics. Nature. 2021:594(7863):356–364. https://doi.org/10.1038/s41586-021-03499-y.
Wiscovitch-Russo R, Narganes-Stordes Y, Cano RJ, Toranzos GA. Origin of the new world *Plasmodium vivax*: facts and new approaches. Int Microbiol. 2019:22(3):337–342. https://doi.org/10.1007/s10123-018-00053-1.
Witt KE, Funk A, Añorve-Garibay V, Fang LL, Huerta-Sánchez E. The impact of modern admixture on archaic human ancestry in human populations. Genome Biol Evol. 2023:15(5):evad066. https://doi.org/10.1093/gbe/evad066.
World Health Organization. World malaria report 2022. 2022. [accessed 2024 Mar 7]. https://www.who.int/publications-detail-redirect/9789240064898.
Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZXP, Pool JE, Xu X, Jiang H, Vinckenbosch N, Korneliussen TS, et al. Sequencing of 50 human exomes reveals adaptation to high altitude. Science. 2010:329(5987):75–78. https://doi.org/10.1126/science.1190371.
Yin Q, Srivastava K, Gebremedhin A, Taye Makuria A, Flegel WA. Long-range haplotype analysis of the malaria parasite receptor gene ACKR1 in an East-African population. Hum Genome Var. 2018:5(1):1–7. https://doi.org/10.1038/s41439-018-0024-8.
Zabala P. The African presence in Yucatan: sixteenth and seventeenth centuries. In: Tiesler V, editor. Natives, Europeans, and Africans in colonial Campeche: history and archaeology. Florida (USA): University Press of Florida; 2010. p. 152–174. https://doi.org/10.5744/florida/9780813034928.003.0008.

Associate editor: Emilia Huerta-Sanchez