Identifying selection signatures across Mexican Indigenous populations using a new bioinformatics tool for whole-genome data

Fernanda Miron [1], Austin W. Reynolds [1], Enrique Morett [3], Israel Aguilar [2].

[1] Department of Anthropology, Baylor University; [2] Department of Supercomputing, Instituto Nacional de Medicina Genómica, Mexico;
[3] Institute of Biotechnology, Universidad Nacional Autónoma de México (UNAM).

INTRODUCTION

The exponential increase in sequencing and computational technologies have paved the way to identify genomic regions that appear to be shaped by **natural selection**. Selection signature studies represent a genomic approach to understand **human evolution** and differences in disease risk across a growing number of populations worldwide. However, the limited amount of genomic data impairs our understanding of the genetic underpinnings of health and human diversity in Mexican populations.

Here we present a simple **Nextflow** pipeline for detecting natural selection in whole-genome data using **Population Branch Statistics (PBS)** and **Integrated Haplotype Score (iHS)**. We then apply the pipeline to whole genome data of 76 individuals from 27 Mexican Indigenous Ethnic groups.

WHOLE GENOME SAMPLES

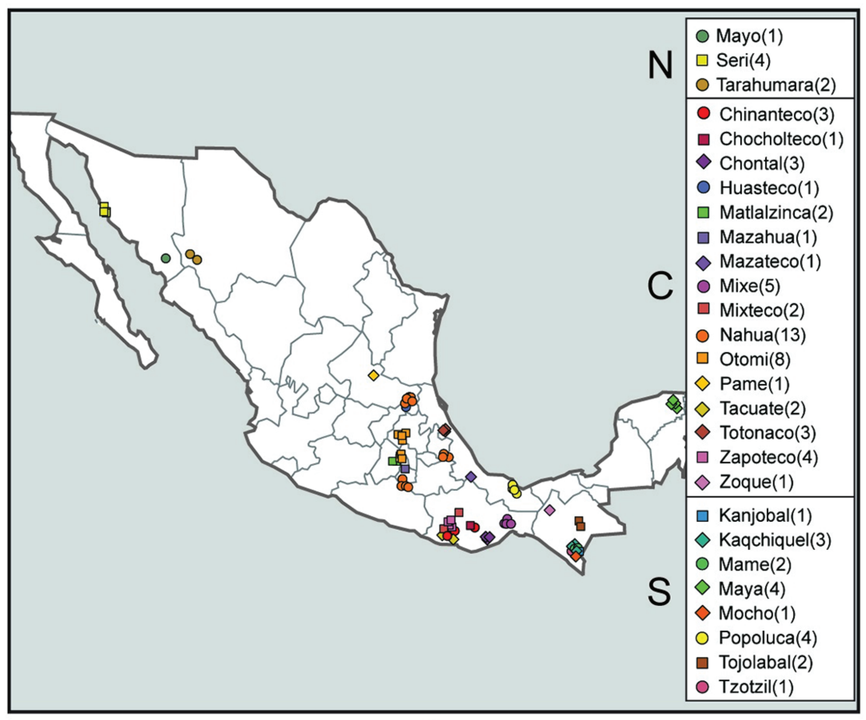
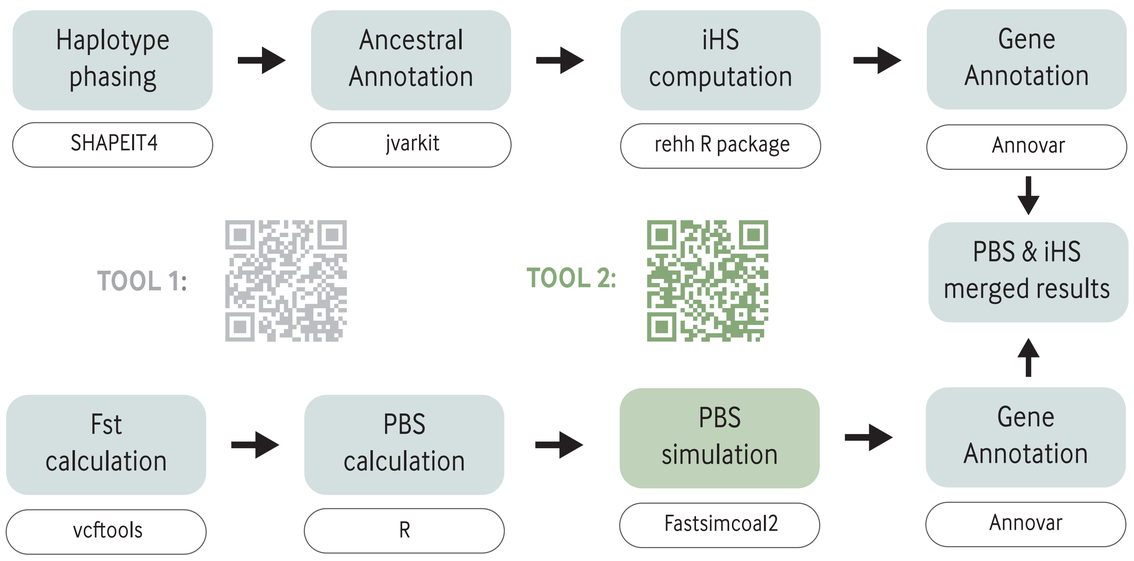


FIGURE 1. Approximate sampling location of 76 unrelated Mexican Indigenous individuals from the Metabolic Analysis in an Indigenous sample (MAIS) cohort. Individuals are representatives of 27 different ethnic groups in Mexico. For selection analysis samples were separated based on Northern (N), Central (C), and Southern Regions (S).

METHODS



1. Reynolds, A. W. et al. Comparing signals of natural selection between three Indigenous North American populations. Proc. Natl Acad. Sci. USA 116, 9312–9317 (2019).
2. Aguilar-Ordoñez, Israel, et al. "Whole genome variation in 27 Mexican indigenous populations, demographic and biomedical insights." PLoS One 16.4 (2021): e0249773.

RESULTS

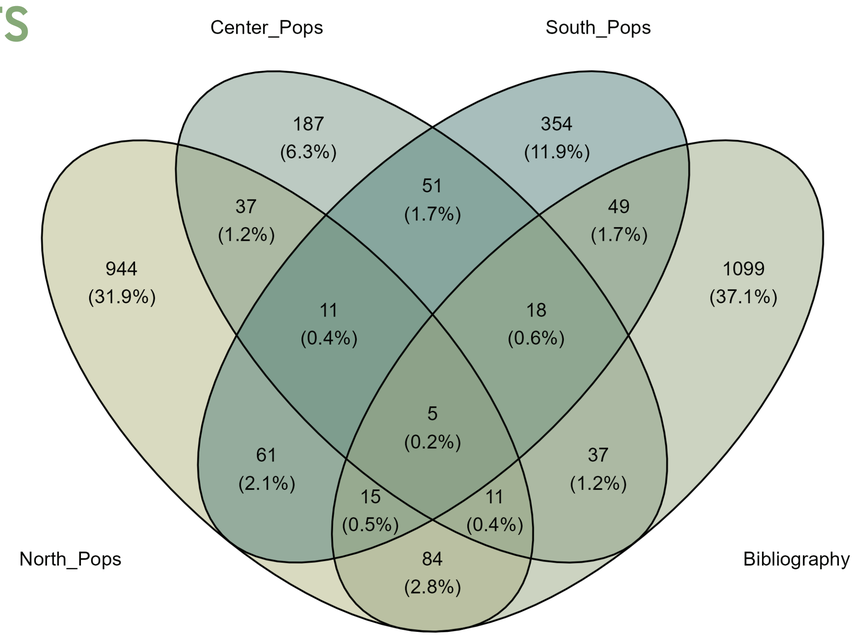


FIGURE 2. Counts of putative selection signals in study populations (**Intersection of the top 1% of each statistic**). Northern Mexican Indigenous populations show 944 exclusive variants under possible selection, while Central and South Mexican Indigenous Populations show 187 and 354 exclusive variants under putative selection respectively.

TABLE 1. Five variants with evidence for natural selection in all study populations and in previous selection analyses. Information from GeneCards database.

Gene	Location (GRCh38)	Function	Phenotypes from GWAS catalog
ARHGAP15	chr2:143,091,362-143,768,352	GTPase activator for the Rho-type GTPases	Lymphocyte count, myeloid white cell count, mean platelet volume
VGLL4	chr3:11,556,067-11,771,350	Negative regulation of cell growth and hippo signaling.	Body height, bmi-adjusted waist-hip ratio and circumference
LINGO2	chr9:27,937,617-29,213,601	Positive regulation of synapse assembly, component of membrane.	Body height, body mass index, smoking initiation
SYNDIG1	chr20:24,469,629-24,666,616	Role in postsynaptic development and maturation	Blood protein measurement, type 2 diabetes mellitus
TFAP2B	chr6:50,818,355-50,847,619	Regulation transcription of selected genes	Body mass index, body weight and height and Char syndrome

TABLE 2. Top selected genes in each study region. Information from GeneCards database

Pop.	Gene	Location (GRCh38)	Selection Density	Function
North	*OR52B6*	chr11:5,580,877-5,581,884	1.092354e-02	Odorant receptor
North	*ID3*	chr1:23,884,417-23,885,992	1.269841e-03	Involved in several cellular processes. Regulates circadian clock CD8+ T cell subsets
Center	*KRT35*	chr17:41,476,710-41,481,151	1.801396e-03	Filament and keratin gene, involved on hair type
Center	*FCER1A*	chr1:159,283,575-159,308,224	7.023231e-04	Responsible for initiating the allergic response, binding of allergen to receptor bound IgE
South	*MCHR1*	chr22:40,679,273-40,682,812	1.695394e-03	Receptor for melanin-concentrating hormone. Related with obesity and energy homeostasis
South	*CYP24A1*	chr20:52,769,985-52,790,525	9.737098e-05	Key role in vitamin D catabolism and calcium homeostasis

CONCLUSION

In line with previous selection studies, we find evidence of selection on genes related to **immunity** and **metabolism** across Mexican Indigenous populations.

3. Yi, Xin, et al. "Sequencing of 50 human exomes reveals adaptation to high altitude." science 329.5987 (2010): 75-78.
4. Voight, Benjamin F., et al. "A map of recent positive selection in the human genome." PLoS biology 4.3 (2006): e72.
5. Ávila-Arcos, María C., et al. "Population history and gene divergence in Native Mexicans inferred from 76 human exomes." Molecular biology and evolution 37.4(2020): 994-1006.