nature portfolio

Corresponding author(s): Tábita Hünemeier

Last updated by author(s): Jun 4, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used.
Data analysis	Sarek v.3.5.0, GATK Best Practices, Ensembl Variant Effect Predictor (VEP) v.113, Scikit-Allel v1.3.13, 'SNPRelate' v1.28.0 R package, ADMIXTURE v1.3.0, CLUMPP v.1.1.2, PLINK 1.9, PRIMUS v1.9.0, ShapeIT4 v.4.2.2, RFMix v.1.5.4, 'admixtools' v.2.0.10 R package, IBDNe v.07May18.6a4, Relate v1.2.2, selscan v.2.0, SPrime method, WEB-based GEne SeT AnaLysis Toolkit (WebGestalt 2024). Newly generated code essential for reproducing the results has been deposited on GitHub (https://github.com/macscastro/lamask).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All data needed to evaluate the conclusions of this study are presented in the paper and/or Supplementary Materials. The datasets used in this study have been

deposited in the European Genome-Phenome Archive and are available for download under accession number EGAXXXXXXX. They can also be freely accessed using the IAGDP Variant Browser.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	No analysis in this study was conducted using biological or social information related to sex or gender.
Reporting on race, ethnicity, or other socially relevant groupings	The present study included individuals who self-identify and are concomitantly officially recognized by their countries of origin as Indigenous Americans, and whose mother tongue is a native language of the American continent.
Population characteristics	No demographic or population-related characteristics were used in the study.
Recruitment	Participants were recruited using a stratified linguistic approach to maximize Native American genetic diversity by sampling a small number of unrelated speakers per Indigenous language. This strategy prioritizes breadth of representation rather than proportional sampling. We acknowledge potential self-selection bias due to voluntary participation and limited sample sizes per group. However, as our primary analyses focus on inter-population structure and broad patterns of genetic diversity, these factors are unlikely to substantially affect the main conclusions.
Ethics oversight	Ethical approval for sample collection was provided by the respective local ethics committees in each country: Argentina (Puerto Madryn Zonal Hospital, Resolution No 009/2015; San Carlos de Bariloche Zonal Hospital, Resolution No 1510/2015), Brazil (CONEP, Resolution No. 123, 763, and 4599), Bolivia (Universidad Mayor de San Andrés), Ecuador (Universidad de Las Américas; Consejo Nacional de Ciencia y Tecnología—CONACyT, grant # 69856; Instituto Nacional de Ciencias Médicas y de la Nutrición Salvador Zubirán Ref.: 1518), Mexico (Consejo Nacional de Ciencia y Tecnología—CONACyT, grant # 69856; Instituto Nacional de Ciencias Médicas y de la Nutrición Salvador Zubirán Ref.: 1518, CNIC Salud 2013-01-201471; Committee of Ethics and Research, UADY, official notice no. F-FENC-SAC-14/REV: 04; registry no. 09/17) and Peru (Universidad San Martín de Porres). Individual and/or tribal informed oral consent was obtained from participants who were unable to read or write. The ethics committees also approved the oral consent procedure. Logistical support for sample collection in Brazil was provided by the Fundação Nacional do Índio. All sampling was conducted in accordance with the Declaration of Helsinki and the applicable laws and regulations of each country at the time of sampling.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No formal a priori statistical power calculation was performed, as the primary objective was not to estimate within-population allele frequencies with high precision, but rather to maximize representation of Native American genetic diversity across ethnolinguistic groups. Sample sizes were therefore determined based on the inclusion of a small number of unrelated speakers per sampled language, prioritizing breadth of linguistic and geographic coverage. Given the current scarcity of available Indigenous genomes and the limited representation of many populations in public datasets, this strategy substantially expands existing diversity and is sufficient for analyses focused on inter-population structure and large-scale evolutionary patterns.
Data exclusions	No data exclusions.
Replication	Methods are published and freely available, and all data is available for replication purposes.
Randomization	Individuals were grouped at different levels according to the analyses, such as by language and geographic region.
Blinding	Blinding was not performed in this study. Sample collection was conducted with prior knowledge of the populations, as cultural and linguistic information was fundamental to the sampling strategy and study design. Blinding was not relevant to the primary analyses, which rely on genome-wide data processed through standardized bioinformatic pipelines, thereby minimizing the potential for investigator-driven bias.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.