RESEARCH ARTICLE SUMMARY

POPULATION GENETICS

Admixture's impact on Brazilian population evolution and health

Full article and list of author affiliations: https://doi.org/10.1126/science.adl3564

Kelly Nunes†, Marcos Araújo Castro e Silva†, Maíra R. Rodrigues†, Renan Barbosa Lemes†, Patricio Pezo-Valderrama, Lilian Kimura, Lucas Schenatto de Sena, José Eduardo Krieger, Margareth Catoia Varela, Luiz Otávio de Azevedo, Luis Marcelo Aranha Camargo, Ricardo G. M. Ferreira, Henrique Krieger, Maria Cátira Bortolini, José Geraldo Mill, Putira Sacuena, João F. Guerreiro, Celia M. B. de Souza, Francisco V. Veronese, Fernanda S. L. Vianna, David Comas, Alexandre C. Pereira*, Lygia V. Pereira*, Tábita Hünemeier*

INTRODUCTION: Brazil is a vast continental country home to the largest population in Latin America and boasts the world's largest recently admixed population. The colonization process brought ~5 million Europeans to Brazil, alongside the forced migration of at least 5 million Africans and the decimation of Indigenous populations, which once included >10 million people speaking more than 1000 languages. This distinctive historical interplay shaped a complex mosaic of genetic diversity, underscoring the importance of detailed genomic studies. However, similar to other populations in the Global South, the Brazilian population remains notably underrepresented in genomic research, where there is a lack of studies investigating the effects of this population's admixture on its evolution, diversity, and health status.

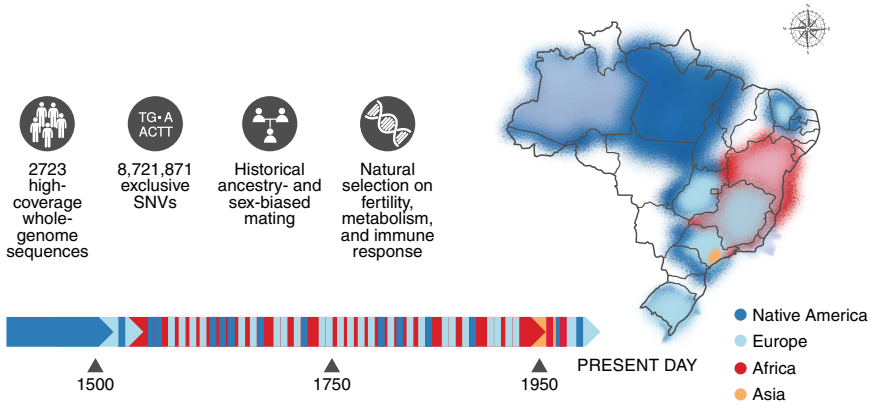
RATIONALE: To address these gaps, we generated 2723 high-coverage whole-genome sequences of the Brazilian population, encompassing urban, rural, and riverine communities from all five geographical regions of Brazil. This dataset reflects a diverse group of ethnic backgrounds, including Afro-Brazilians and descendants of Indigenous people, and provides a comprehensive representation of Brazilian genomic diversity. Advanced methods, such as local ancestry inference and haplotype-based analyses, enabled us to characterize ancestry-specific genomic regions in different time periods and geographic regions and detect signatures of natural selection. Our research highlights admixture's evolutionary and health implications, focusing on the historical and demographic dynamics that shaped Brazilian genomes. Our results contribute to a deeper understanding of how global haplotypes and admixture patterns resulting from an intricate evolutionary history could affect an admixed population's health.

RESULTS: We identified >8 million previously unknown variants, 36,637 of which are putatively deleterious, and elucidated a positive correlation between these deleterious variants and genetic ancestry components. We also showed that the Brazilian population is a tapestry of global haplotypes shaped by nonrandom mating, with the peak of admixture occurring in the 18th and 19th centuries. Multiple or continuous admixture events between Indigenous American, African, and European parental sources have formed Brazilian populations. These processes align with major historical events that have shaped the Brazilian state over the past five centuries. We also identified that, after a prolonged period of sex-biased mating in the initial centuries, a strong pattern of assortative mating has more recently emerged in the Brazilian population, regardless of the region studied. These patterns reveal both the violent dynamics of European colonization and the lasting imprints of this process on contemporary South America. Within the extensive diversity found in Brazil, ancestral-specific genomic regions originating from different populations are unevenly spread across Brazilian regions and historical time frames. This distribution demonstrates the lasting impact of the hundreds of ethnicities that arrived in the country through millions of Europeans and Africans at different times, admixing with and replacing the original Indigenous population. Furthermore, our study identifies several candidate genes that were subject to selection both before and after contact in the Brazilian admixed population. These genes are primarily associated with heightened fertility rates, immune response, and distinctive metabolic traits.

CONCLUSION: Our findings underscore the discernible influence of different ancestral backgrounds on Brazilian admixed individuals' health and genetic makeup. We show that this genetic landscape finds its roots in the evolutionary history of Brazilian Indigenous communities and the intricate demographic interplay stemming from both coerced and voluntary historical immigration to Brazil. □

*Corresponding author. Email: alexandre.pereira@incor.usp.br (A.C.P.); lpereira@usp.br (L.V.P.); hunemeier@usp.br (T.H.) †These authors contributed equally to this work. **Cite this article as K. Nunes *et al.*, *Science* 388, eadl3564 (2025). DOI: 10.1126/science.adl3564**

Brazilian genomic diversity. A comprehensive study of 2723 high-coverage whole-genome sequences from diverse Brazilian regions reveals >8 million newly identified single-nucleotide variants (SNVs), highlighting Brazil's pronounced genomic diversity, shaped by natural selection, nonrandom mating, and continuous admixture pulses since 1500. These findings underscore the influence of ancestral backgrounds on the health and genetic profile of the Brazilian population, rooted in Indigenous history and diverse immigration waves.



POPULATION GENETICS

Admixture's impact on Brazilian population evolution and health

Kelly Nunes[1]†, Marcos Araújo Castro e Silva[1,2]†, Maíra R. Rodrigues[1]†‡, Renan Barbosa Lemes[1]†, Patricio Pezo-Valderrama[3], Lilian Kimura[1], Lucas Schenatto de Sena[1], José Eduardo Krieger[4], Margareth Catoia Varela[5], Luiz Otávio de Azevedo[5], Luis Marcelo Aranha Camargo[6], Ricardo G. M. Ferreira[7], Henrique Krieger[6], Maria Cátira Bortolini[8], José Geraldo Mill[9], Putira Sacuena[10], João F. Guerreiro[10], Celia M. B. de Souza[11], Francisco V. Veronese[11], Fernanda S. L. Vianna[8,11], David Comas[2], Alexandre C. Pereira[4,12]*, Lygia V. Pereira[1]*, Tábita Hünemeier[1,3]*

Brazil, the largest Latin American country, is underrepresented in genomic research despite boasting the world's largest recently admixed population. In this study, we generated 2723 high-coverage whole-genome sequences from the Brazilian population, including urban, rural, and riverine communities representing diverse ethnic backgrounds. We reveal the impressive genomic diversity of Brazilians, identifying >8 million previously unknown variants, including 36,637 predicted deleterious and potentially affecting population health. We found a positive correlation between these deleterious variants and ancestry. Brazilian genomes are a global haplotype mosaic shaped by nonrandom mating, with peak admixture in the 18th and 19th centuries. Within this diversity, ancestry-specific haplotypes exhibit an uneven spatiotemporal distribution. We also identified putatively selected genes in this diverse population, primarily linked to fertility, immune response, and metabolic traits.

The European colonization of the Americas initiated the largest intercontinental population displacement in human history. From the 15th to the 20th centuries, ~5 million Europeans immigrated to Brazil. These European settlers forcibly brought at least 5 million enslaved Africans from different ethnic groups during the first four centuries of colonization (*1*). At the time of European arrival, the original population living in present-day Brazilian territory consisted of >10 million Indigenous individuals speaking more than a thousand Indigenous languages (*2*). Contact with European colonizers led to the decimation of the Indigenous population. It is estimated that the effective population size (N_e) decline in Brazil ranged between 83% inland and 98% on the Atlantic coast (*2*). Currently, the Brazilian Indigenous population is divided into four main genetic groups, mainly related to geography but also to linguistic diversity (*2*). Brazil also has the greatest African diversity found in the Americas, presenting almost all African mitochondrial haplogroups and a great variability of African genetic ancestries (*3*, *4*). There are 56 European languages or dialects officially in use in Brazil today, with German being the second most spoken language (after Portuguese), with 3 million speakers, as of 2022, per IPOL (Instituto de Investigação e Desenvolvimento em Política Linguística), together with 274 Indigenous languages; however, no African language is spoken by the descendants of enslaved individuals, as per IBGE (Instituto Brasileiro de Geografia e Estatística). This demonstrates the cultural richness of the present-day Brazilian population and the impacts of power asymmetry resulting from the colonization period.

However, the genetic heterogeneity resulting from this extensive admixture in Brazil has not been thoroughly characterized. There is a lack of studies investigating the fine-scale population substructure and ancestral sources across the entire country. Existing studies have focused on array data or on specific Brazilian regions (*5*, *6*) and have only broadly characterized the Brazilian population, identifying a limited number of previously unknown variants potentially associated with specific traits in the studied cohorts. Furthermore, no study on natural selection or fine-scale genetic structure based on whole genomes focusing on the Brazilian population has been conducted, leaving a gap in evidence regarding the potential relationships between ancestry, admixture, and relevant medical traits in the Brazilian population.

To address this knowledge gap, we sequenced high-coverage whole genomes from 2723 individuals representing the five Brazilian geographical regions (south, southeast, central, north, and northeast). This diverse sample includes individuals from urban and rural areas, such as riverine communities in the Amazon, and individuals from various ethnic backgrounds, including Afro-Brazilians. We reveal that Brazilian genomes are highly diverse compared with available worldwide genomes, presenting newly identified haplotypes associated with Indigenous American, European, and African ancestries. Brazilians are a mosaic of worldwide haplotypes formed, most extensively during the 18th and 19th centuries, by nonrandom mating. The ancestry-specific tracts derived from the ancestral populations within this highly diverse population are unevenly distributed across Brazilian regions and time periods, affecting its evolution and health and including putatively selected genes associated with higher fertility rates, immune response, and metabolic traits.

Data overview

We examined newly acquired whole-genome sequence data for 2723 individuals (~35× coverage) as part of the DNA do Brasil project (henceforth called DNABR) (*7*). These individuals were recruited from five distinct health study cohorts and represent populations originating from the north, northeast, south, southeast, and center-west regions of Brazil (fig. S1 and tables S1 and S2). We assembled a reference panel of worldwide populations by combining the Human Genome Diversity Project (HGDP) (*8*) and the Simons Genome Diversity Project (SGDP) whole-genome sequencing (WGS) datasets (*9*) to investigate the fine-scale genetic diversity of these populations (table S3). Additionally, we incorporated information from three previously published datasets featuring present-day Indigenous Americans (*2*, *10*, *11*), thereby including an additional 97 individuals (table S3).

Unlocking Brazilians' hidden genetic diversity

We identified 8,721,871 previously unknown (hereafter, "novel") single nucleotide variants (SNVs) not present in the main public datasets, such as the Genome Aggregation Database (gnomAD), the 1000 Genomes Project, the Human Genome Diversity Project, and the Single Nucleotide Polymorphisms Database (dbSNP). The dbSNP includes variants (rare and common) reported in the Trans-Omics for Precision Medicine program (TOPMed). TOPMed, together with gnomAD, represents

[1]Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, SP, Brazil. [2]Institut de Biologia Evolutiva, Departament de Medicina i Ciències de la Vida, Universitat Pompeu Fabra, Barcelona, Spain. [3]Institut de Biologia Evolutiva, CSIC–Universitat Pompeu Fabra, Barcelona, Spain. [4]Laboratório de Genética e Cardiologia Molecular, Instituto do Coração, Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, São Paulo, SP, Brazil. [5]Fundação Oswaldo Cruz, Unidade Rondônia, Rio de Janeiro, RJ, Brazil. [6]Fundação Oswaldo Cruz, Unidade Rondônia, Porto Velho, RO, Brazil. [7]Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, SP, Brazil. [8]Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil. [9]Departamento de Fisiologia, Universidade Federal do Espírito Santo, Vitória, ES, Brazil. [10]Laboratório de Genética Humana e Médica, Instituto de Ciências Biológicas, Universidade Federal do Pará, Belém, PA, Brazil. [11]Hospital de Clínicas de Porto Alegre, Porto Alegre, RS, Brazil. [12]Department of Genetics, Harvard Medical School, Boston, MA, USA. ***Corresponding author. Email: alexandre.pereira@incor.usp.br (A.C.P.); lpereira@usp.br (L.V.P.); hunemeier@usp.br (T.H.)**
†These authors contributed equally to this work. ‡Present address: gen-t Science, São Paulo, SP, Brazil.

highly diverse and variable genome databases, having reported more than 705 million (https://legacy.bravo.sph.umich.edu/freeze8/hg38/) and 270 million variants (*12*), respectively, in >270,000 individuals. Even so, the Brazilian population presents greater diversity, probably owing to the genetic contribution of underrepresented populations, such as Africans and Indigenous Americans. The novel SNVs accounted for 11.16% of all SNVs found in our dataset (78,187,567). Most of these variants were in intronic (50.7%) or intergenic regions (25.3%). However, there were 288,739 novel variants in regulatory regions, 42,101 novel missense mutations, and 1301 novel stop-gain mutations with potential phenotypic effects (fig. S2A and table S1). We further annotated the novel variants for in silico pathogenicity prediction using the PolyPhen (*13*) and SIFT (*14*) methods, specifically focusing on nonsynonymous variants (fig. S2B). We found 36,637 variants scored as "deleterious" or "deleterious low confidence" by SIFT or as "probably damaging" or "possibly damaging" by PolyPhen, 23 (0.06%) of which were identified at a frequency of >1% (polymorphic variants) (table S4).

We also identified 64,769 registers of putative loss of function (pLoF) variants with high-confidence classification, 23,405 different IDs of which 2038 were exclusively found in Brazilians. Approximately 9% of these variants are polymorphic (table S5). For instance, 450 genes containing pLoF variants had associations with metabolic traits, wherein 201 and 222 were related to cholesterol and body mass index (BMI), respectively (fig. S3). Furthermore, we found pLoF variants in 815 genes associated directly (214) or indirectly (601) with infectious diseases, specifically malaria, hepatitis, influenza, tuberculosis, salmonellosis, and leishmaniasis (table S6). We classified genes directly associated as those with a specific function in response, severity of the outcome, or susceptibility to infections; and genes indirectly associated as those identified through association studies (*15*). These diseases impose a considerable burden on the Brazilian population and are associated with high mortality rates in the country (https://data.who.int/countries/076). In this sense, this LoF variant landscape could shed light on the Brazilian population's genetic diversity and has potential implications for disease susceptibility, as metabolic and infectious diseases are the main concerns in Brazilian public health.

To understand the connection of potentially deleterious variants to their ancestral origins, we analyzed how the rare genetic variants, synonymous and deleterious, are related to the ancestral components and the population history. Consistent patterns emerged across ancestral backgrounds for both types of variants (fig. S4). African ancestry exhibited a strong positive correlation with the presence of rare variants, whereas European ancestry showed strong negative correlation. The positive correlation of all variants with African ancestry is expected, owing to the large influx of enslaved Africans brought to Brazil over the course of four centuries, with the vast majority coming from West Africa (*4*). Although a similar tendency has been identified for other Latin American populations (*15*), this is the first time it is being described for Brazilian populations. The occurrence of successive bottlenecks and founder effects, common in Indigenous American communities (*2*, *16*), may have also played a key role in the maintenance of some deleterious variants in the Brazilian population, as well as the arrival of various European ethnicities in different periods, characterized by a small number of immigrants, mainly from Northern Europe and Portuguese islands. There are several examples of genetic diseases of European origin that are highly prevalent in Brazil and rare in Europe, such as Machado-Joseph ataxia (*17*, *18*) and Li-Fraumeni syndrome (*19*–*21*). For instance, we identified two unrelated individuals [kinship coefficient (PI_HAT) < 0.1875] carrying the pathogenic R337H *TP53*/Li-Fraumeni allele, which is extremely rare worldwide (*21*).

The correlation pattern changes considerably when only newly identified variants are considered in the analysis, showing a positive weak correlation between novel deleterious variants and Indigenous American ancestry (Fig. 1 and fig. S5), a negative weak correlation with European ancestry, and no correlation with African or Asian ancestry. This demonstrates the importance of focusing on underrepresented populations in public health research. It is worth mentioning that Brazil's Indigenous American populations are one of the least studied groups globally. Nevertheless, our findings reveal that it is possible to uncover a portion of their genetic diversity by examining admixed individuals' genomes. The observed weak correlation between the proportion of Indigenous American ancestry in individuals and the number of newly identified rare deleterious variants likely reflects the underrepresentation of Indigenous American populations in genetic studies, rather than indicating an inherently higher burden of deleterious alleles associated with any specific ancestry. Nevertheless, previous studies have reported an increase in mutational load with distance from sub-Saharan Africa, inferring a relatively higher accumulation of deleterious variants in Indigenous American populations (*22*).

Nonrandom mating formed the present-day Brazilian population

Principal components analysis (PCA) (*23*) and local ancestry inference (*24*) reveal that Brazilian individuals have different proportions of Indigenous American, African, and European ancestry (Fig. 2, A to C). The exceptions are 10 individuals from São Paulo state with almost 100% East Asian ancestry (Fig. 2C), probably related to the recent Japanese migration to this region. A vast majority of the population sits on a cline of genetic variation between European and African ancestries (Fig. 2, A to C). Global and local ancestry inference results demonstrate that the DNABR dataset is composed of a majority of European ancestries (58.9 to 59.72%; here we report global ancestry results followed by local ancestry inference results), followed by a considerable contribution from African ancestries (27.19 to 27.13%) and a smaller proportion of Indigenous American ancestries (13.36 to 13.14%) (Fig. 2, C and D, and fig. S6). The average ancestral contribution from Indigenous Americans to the DNABR sample is noticeably higher than previously estimated by other studies, where these estimates typically ranged between 7 and 9% (*5*, *25*). This is possibly due to a better representation of individuals from northern Brazilian states in the DNABR sample. The highest proportions of Indigenous American ancestries were observed in the northern region (Fig. 2E and fig. S7). African ancestries are predominant in the northeastern region, and European ancestries are predominant in the southeastern and southern regions.

There is a noticeable lack of individuals with higher Indigenous American and African ancestry (Fig. 2, A to C). This might be explained by the fact that the vast majority of Y-chromosomal lineages in this sample are of European ancestry (71%), whereas the majority of mitochondrial lineages are either African (42%) or Indigenous American (35%) (Fig. 2D and figs. S8 and S9). This pattern is most likely the result of historical asymmetrical mating between European men and Indigenous American and African women, illustrating how the genomic data might help to uncover the history of colonization and its persistent consequences on the genetic diversity of present-day populations. This kind of sex-biased admixture was already observed in other studies of present-day admixed populations in the Americas (*26*); it is predictable given that most European settlers were men, especially in Latin America, and also considering the history of violence during colonization, which likely led to a higher death toll of Indigenous American and African men, along with sexual violence toward women from these groups (*27*, *28*).

Ancestry-assortative mating analysis during the last generations (i.e., the most recent generations) (*29*, *30*) detected a significant positive correlation between all continental ancestry proportions (African, European, and Indigenous American), whether analyzing autosomal or X chromosomal data (fig. S10). This correlation was consistently observed when pooling the entire dataset and when analyzing each federal state independently, with only a few exceptions for the autosomal

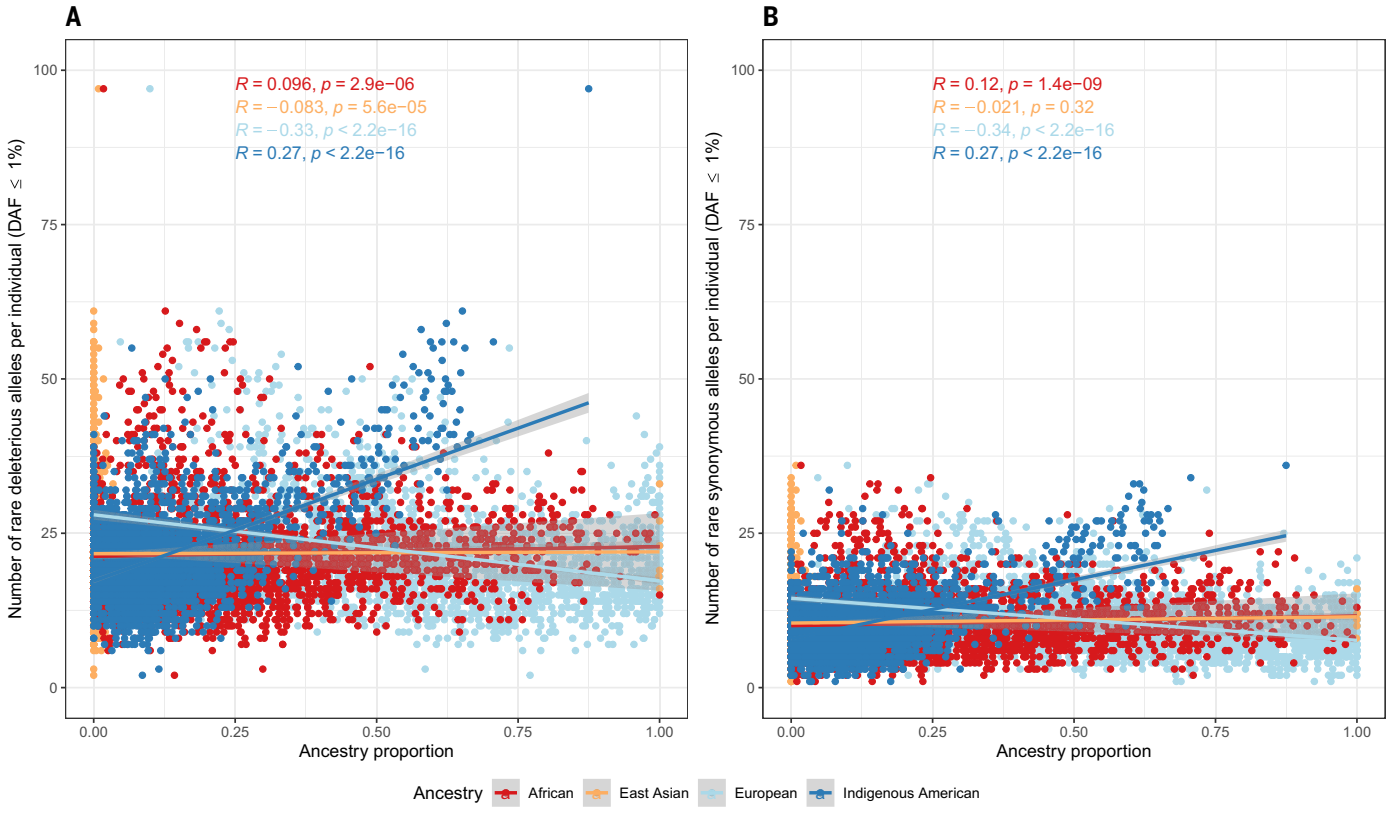


Fig. 1. Ancestry-specific mutation burden of rare new variants. Rare new variants (DAF ≤ 1%) in relation to the global ancestry proportion per individual. The Spearman's *R* coefficient values and corresponding *P* values are color coded according to ancestries. (**A**) New deleterious variants (variants predicted to be damaging by PolyPhen2 and deleterious by SIFT along with splice, stop-loss, and stop-gain variants). (**B**) New synonymous variants.

data (figs. S10 and S11), supporting the occurrence of ancestry-assortative mating in Brazilian populations, even when considering the geographic distribution of the samples. For the autosomes, the strength of the correlation is higher for the African ancestries (Pearson's correlation coefficient $r = 0.68$), followed by the European ($r = 0.63$), and Indigenous American ($r = 0.53$). For the X chromosome, the correlation values are considerably higher, with the highest observed value for Indigenous American ancestries ($r = 0.97$), followed by European ($r = 0.93$) and African ($r = 0.92$). This finding implies that individuals tended to mate with other individuals of similar ancestry over the parental generation. This pattern possibly stretches farther back in time, with potentially substantial implications for the distribution of genetic variation.

The sex-biased admixture, as indicated by the differential ancestry proportions inferred from autosomal and uniparental markers, along with the evidence of ancestry-assortative mating, highlighted by the correlation between inferred ancestry proportions of mating pairs during the parental generation, collectively suggest a temporal shift in the type of nonrandom mating pattern prevalent in Brazil. In the beginning, sex-biased admixture (mating mainly between European men and Indigenous American and African women) had precedence because of the population dynamics of the initial colonization, followed by the dominance of ancestry-assortative mating. The co-occurrence of these two types of nonrandom mating patterns was previously reported in African populations and Hispanic populations from the United States, however, it was not demonstrated in Latin American populations (*26*, *31*). One plausible hypothesis is that this shift occurred as an admixed population gradually formed, prompting changes in mating patterns. Furthermore, these outcomes distinctly reveal that mating patterns are currently influenced by ancestry proportions, a phenomenon likely intertwined with socioeconomic factors, as previously demonstrated in other admixed populations of the Americas (*32*).

Brazilian genomes are a mosaic of worldwide haplotypes

At the finest genetic structure level (*33*), we identified 298 genetic clusters among DNABR individuals and 104 clusters among reference samples. However, we selected a coarser structure level, with 18 recipient genetic clusters (named Brazil A to R) and 41 donor genetic clusters, to make analyses more meaningful and intelligible (Fig. 3). The clustering method (*33*) more effectively captures the patterns of genetic similarity than does a simple geographic clustering of samples (by states of origin), as evidenced by the principal components of variation (both linked and unlinked and data; figs. S12 and S13) and the total variation distance (TVD) (figs. S14 and S15).

As a whole, the major contribution from European ancestries came from southern European populations (Fig. 3, B and C). However, one cluster (Brazil J) that is predominantly composed of individuals from the southern Rio Grande do Sul state shows a European ancestry more related to northern European populations (Fig. 3, A and B, and fig. S15). These results are consistent with historical records suggesting that immigration to Brazil predominantly originated from Portugal and Italy and also with the fact that southern Brazilian states received immigrants from Germany and Northern Italy during the late 19th and early 20th centuries (https://brasil500anos.ibge.gov.br/).

African portions of Brazilian genomes are highly diverse and share more recent ancestors with groups from different subcontinental regions (Fig. 3, B and C, and fig. S15), especially from West and East Africa but also from South and North Africa. This suggests that groups from different subcontinental regions that might have never contacted each other within Africa would have admixed in Brazil,

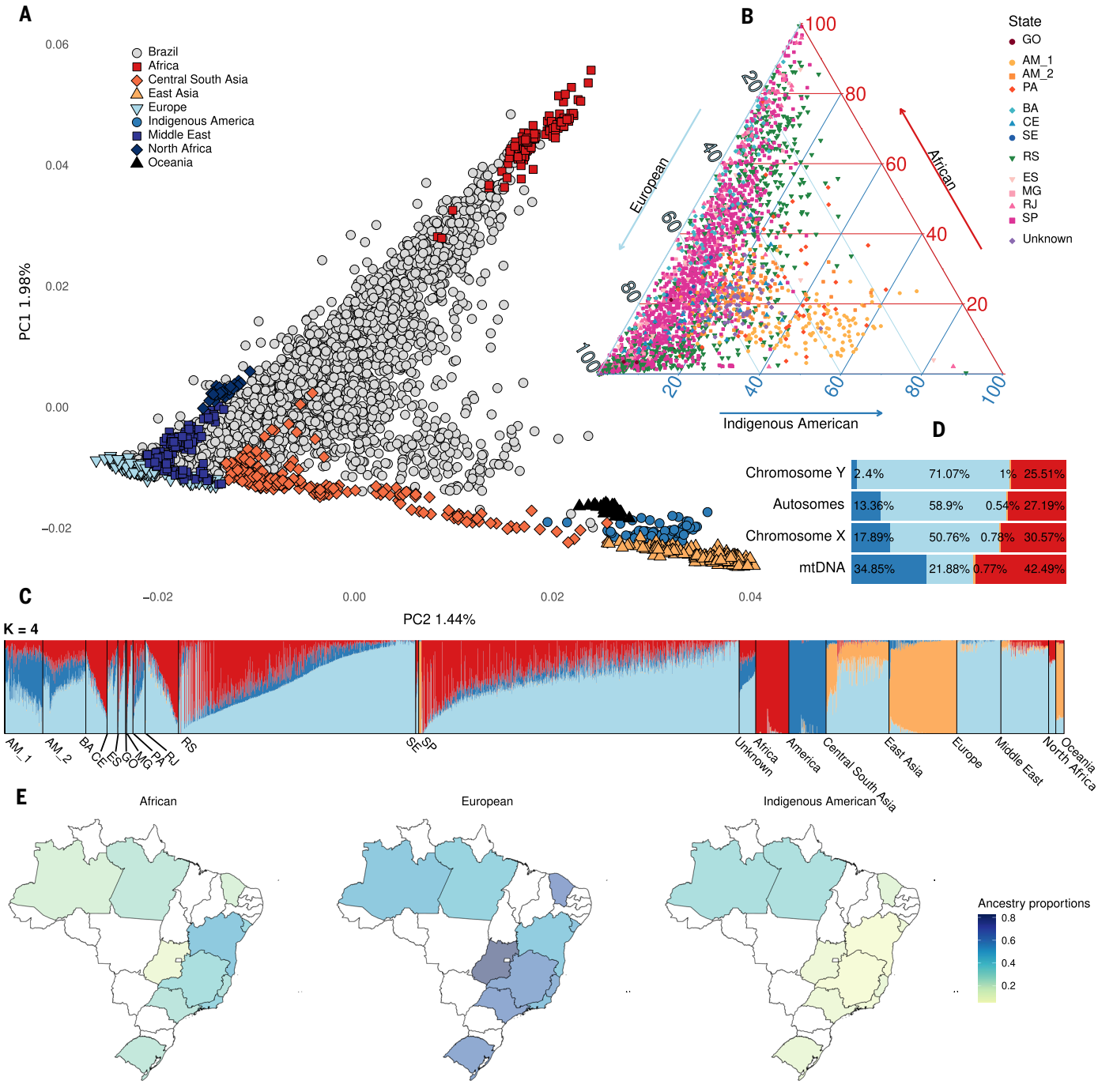


Fig. 2. Patterns of genetic variation and continental-level ancestry proportions. We estimated individual patterns of genetic variation and structure using the LD-pruned autosomal dataset by executing a (**A**) PCA (SNPRelate) (*23*) and (**C**) an unsupervised ADMIXTURE analysis ($K = 4$) (*100*). We inferred local ancestry through GNOMIX (*24*) using the phased autosomal data with SHAPEIT4 (*106*) and plot (**B**) individual averages color coded by state and (**E**) state averages. (**D**) Finally, we estimated average continental-level ancestry proportions for the complete dataset using autosomal and X chromosome markers by performing a supervised ADMIXTURE analysis ($K = 4$) as well as using the Y chromosome and the mitochondrial DNA (mtDNA) by applying SNAPPY (*104*) and HAPLOGREP 3 (*112*), respectively. AM, Amazonas (samples from the Amazonas are categorized into two groups, AM_1 and AM_2, as they were collected from distinct riverine regions within the state); BA, Bahia; CE, Ceará; ES, Espírito Santo; GO, Goiás; MG, Minas Gerais; PA, Pará; RJ, Rio de Janeiro; RS, Rio Grande do Sul; SP, São Paulo; Unknown, no information.

leading to the appearance of Brazilian individuals who are a mosaic of different African populations. Furthermore, clusters with higher African ancestry contributions (Brazil N to R) exhibit larger relative proportions of West and East African components and a larger relative proportion of Bantu components in the case of Brazil P (Fig. 3B and fig. S15).

The patterns of haplotype sharing with present-day Indigenous Americans are at least partially related to geography, as the clusters with a higher proportion of haplotype sharing with Indigenous Americans (Brazil A, B, D, and E) are composed of larger proportions of individuals from northern Brazilian states (Amazonas and Pará; Fig. 3, A and B, and fig. S15). However, the vast majority of these Indigenous American

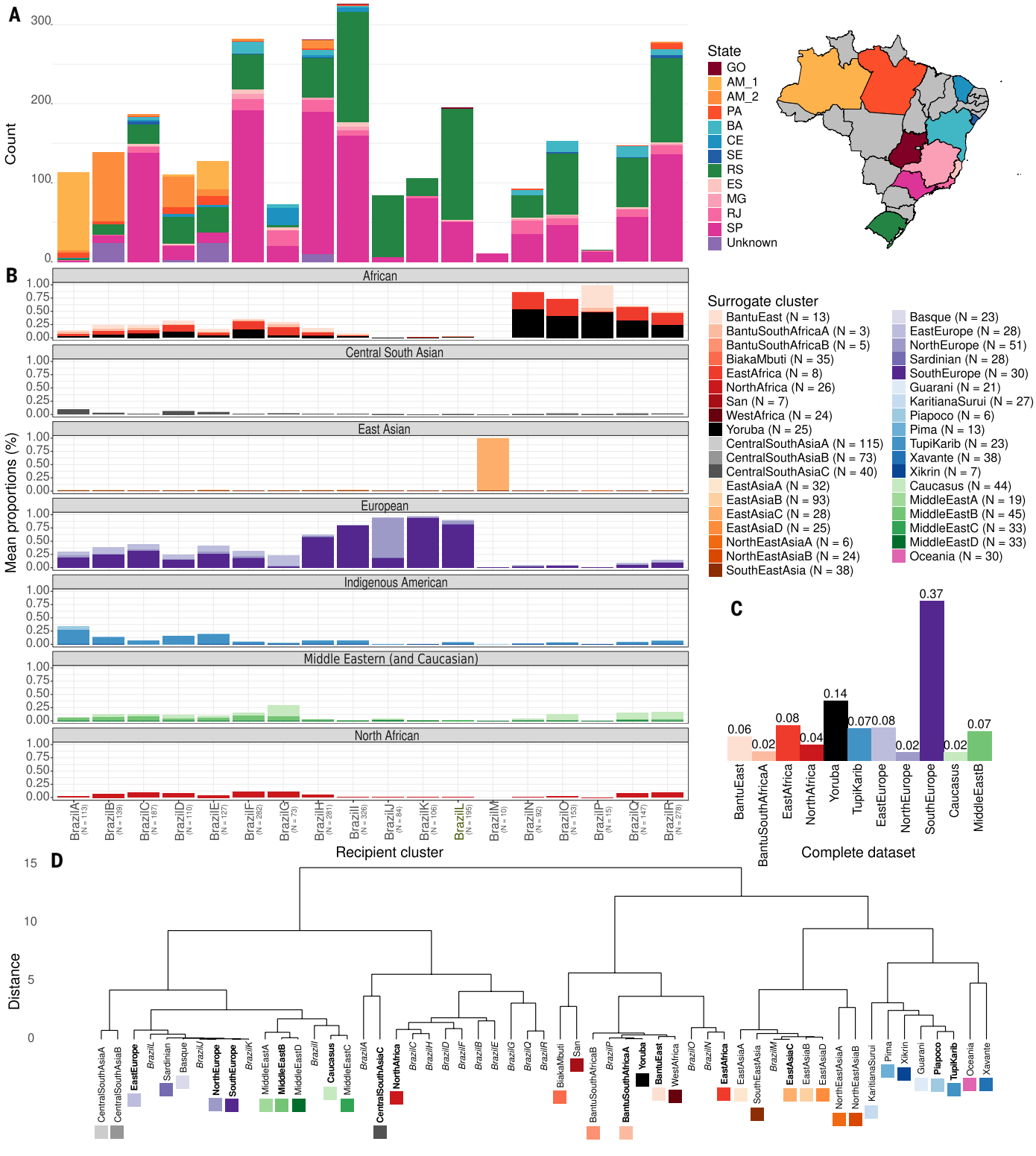


Fig. 3. Fine-scale genetic structure. We used CHROMOPAINTER (*33*) to infer haplotype sharing between individuals and fineSTRUCTURE (*33*) to cluster individuals on the basis of haplotype similarity (18 recipient clusters and 37 surrogate clusters). We then used SOURCEFIND (*107*) to model the clusters' genetic variation as a mixture of the surrogate clusters. Here we show (**A**) how many individuals from each state are included in each cluster and (**B**) the average proportions of contribution from each surrogate cluster obtained in 100 independent runs of SOURCEFIND (median proportions are shown in fig. S15). (**C**) We also show the average proportions of contribution for all DNABR samples (clusters that contribute <1% are not shown). (**D**) Finally, we estimated TVD between all recipient and surrogate clusters [using the method from (*113*)], then we performed hierarchical clustering ("stats" R package) using the estimated TVD values. Surrogate clusters that significantly contribute to the recipient clusters are marked in bold.

haplotypes share the most recent ancestors with individuals from a cluster that contains Tupi and Karib speakers from the Amazonian region (Fig. 3, A to C, and fig. S15). Previous studies did not find genetic differentiation between populations from the Tupi and Karib ethnolinguistic groups (*2*). The Tupi were the predominant group in the Brazilian Atlantic Coast and eastern Brazil before the European invasion. These regions were also where the European settlers and enslaved Africans concentrated, especially during the early period of colonization. Hence, it is plausible that admixture events involved primarily Tupi groups. However, it is important to note that the method's limitations might have prevented the detection of haplotype similarity with other Indigenous groups, such as Jê speakers, who are highly drifted owing to recent inbreeding [as discussed by Moorjani and Hellenthal (*34*)].

Admixture processes peaked during the 18th and 19th centuries

The timeline of admixture events was inferred using three different approaches to examine potential methodological biases and secure more comprehensive results: (i) fastGLOBETROTTER (GT), a haplotype-based method (*35*); (ii) MALDER (MA), an admixture linkage disequilibrium (LD)–based method (*36*); (iii) and TRACTS (TT), a local ancestry–based method (*37*). To ensure comparability and consistency in our inferences and account for the Brazilian population's genetic structure, we applied all three methods to the 18 recipient clusters identified in the fineSTRUCTURE analysis (tables S7 and S8). We also ran TRACTS considering both the whole dataset and regional clusters.

The best-fit TRACTS models suggest an initial admixture event between Indigenous American and European populations about 16 generations ago, succeeded by an African migration pulse eight generations ago, and followed by an additional European migration event five generations ago (table S7). When we analyze TRACTs by geographic region, it is possible to observe some nuances of the admixture process (fig. S17). The admixture process in Brazil began in the northeast and southeast regions, where we find the oldest dates (up to 16 generations ago) and the highest number of migratory waves (fig. S17B). In the south, the process started 10 generations ago and intensified from eight to five generations ago (fig. S17C). Finally, in the north, intercontinental migrations occurred mainly between 10 and 8 generations ago (fig. S17A). Notably, we only infer an African admixture event before the 18th century when individuals from the northeast and southeast regions are pooled together (fig. S17B), which is in line with the earlier presence of plantations in these regions (*38*).

For nearly all of the recipient clusters, at least one admixture event was inferred using the three methods. The inferred dates for the most recent admixture event of all clusters were concentrated between ~5 and ~10 generations before the present (gBP), approximately corresponding to the period between 1750 and 1875 CE, considering a generational time of 25 years (figs. S16 and S17 and table S8). This period corresponds to important demographic events in Brazil, such as the Gold Cycle (1690 CE–1750 CE), which abruptly intensified Portuguese migration to Brazil to the extent that the kingdom of Portugal restricted the free access of Portuguese people to Brazilian lands (*39*). Following the discovery of gold, inland Brazil experienced tremendous urban expansion throughout the 18th century as a result of a sharp rise in migration rates. During the same period, Bandeirantes mercenaries explored the interior of Brazil in search of gold and precious stones, also aiming to capture and enslave Indigenous peoples, resulting in the extermination and displacement of thousands of them (*40*), and expanding Brazilian borders westward beyond the limits established by the Treaty of Tordesillas (1494 CE). Furthermore, the Portuguese royal family fled to Brazil in 1808 CE as a result of the Napoleonic Wars and the invasion of Portugal, establishing Rio de Janeiro as the capital of the Portuguese Empire, which contributed to an expansion of the slave trade. From that time until the abolition of slavery in 1888, >2 million Africans were brought to Brazil, marking the peak of the Atlantic slave trade in the country (*38*), which is in accordance with our genomic results.

During the imperial period (1822–1889 CE), the Brazilian government implemented policies to encourage immigration, particularly from northern Europe to southern and southeastern Brazil. This massive migration was promoted, among other reasons, as a demographic solution to prevent individuals of African descent from becoming the majority of the country's population after the abolition of slavery in 1888 CE, reflecting the prevailing racial ideology of the time (*41*). It is estimated that between 1880 and 1930 CE, ~4 million European immigrants, predominantly from Italy, Spain, Germany, and Portugal, arrived in Brazil (*42*). Our results also align with this intense historical European gene flown, indicating that our sample is representative enough to recover the pivotal Brazilian demographic events.

Thus, Brazilian populations historically were formed by multiple or continuous admixture events between Indigenous American, African, and European parental sources. The peak of admixture also coincides with a period of increased social unrest, including many political and land disputes (e.g., the Ragamuffin War between 1835 and 1845 CE; the Paraguayan War from 1864 to 1870 CE), internal conflicts (e.g., Guaraní War in 1756 CE; Malê revolt in 1835 CE), along with severe climatic disasters such as the Great Drought of 1877–1878 CE (*43*). These events had a great impact on Brazil's demographic structure, resulting in large-scale population displacements across regions and a considerable death toll. Our findings regarding migratory flows and ancestry- and sex-biased mating can be interpreted together to provide a deeper understanding of Brazilian demographic and historical dynamics. However, also as a consequence of this intense demographic history, it is plausible to assume that the inferred admixture events and dates do not fully capture the admixture dynamics involved in forming these populations. In fact, the inferred dates most likely represent a period or a pulse of more intense admixture involving populations that were previously admixed [see (*34*) for a discussion].

Ancestry-specific tracts are unevenly distributed across Brazilian regions and time periods

The patterns of genetic variation and demographic history were analyzed using runs of homozygosity (ROH) and identity-by-descent (IBD) tracts throughout Brazilian populations. Positive correlations were detected in the analysis of the number of ROH (fig. S18, A to C) in European ancestry (Spearman's correlation coefficient $R = 0.68$). In contrast, a robust negative correlation was observed with the African proportion ($R = -0.92$). This indicates that as African ancestry increases, the genomic proportion of ROH decreases, ultimately amplifying individual genetic variation, which is in line with findings from other ROH studies, which have consistently reported that admixed populations, particularly those with African contributions, exhibit fewer ROH and increased genetic diversity (*44*–*49*). These results, together with the distribution of ancestry within Brazil and the average distribution of ROH in Brazilian states (Fig. 2E and fig. S18D), provide evidence that an increase in genetic diversity in the northeast and southeast regions was due to the large proportion of African descendants who historically populated these parts of the country.

Brazilian populations share a greater number of ancestors from earlier time periods, as evidenced by a prevalence of smaller [2- to 4-cM (centimorgan)] and medium-sized (4- to 8-cM) IBD segments [for detailed information on the IBD sharing analysis, refer to (*7*)]. The smaller and medium-sized IBD segments are likely due to the population bottlenecks and founder events involved in their formation. Meanwhile, a distinct pattern emerged (fig. S19) for longer IBD segments (>8 cM), primarily generated after the post-European contact period (*50*, *51*). In this context, as segment size increases, the count of shared segments between individuals decreases, both within individual states and across states (Fig. 4 and fig. S20). This scenario suggests increasing isolation among different

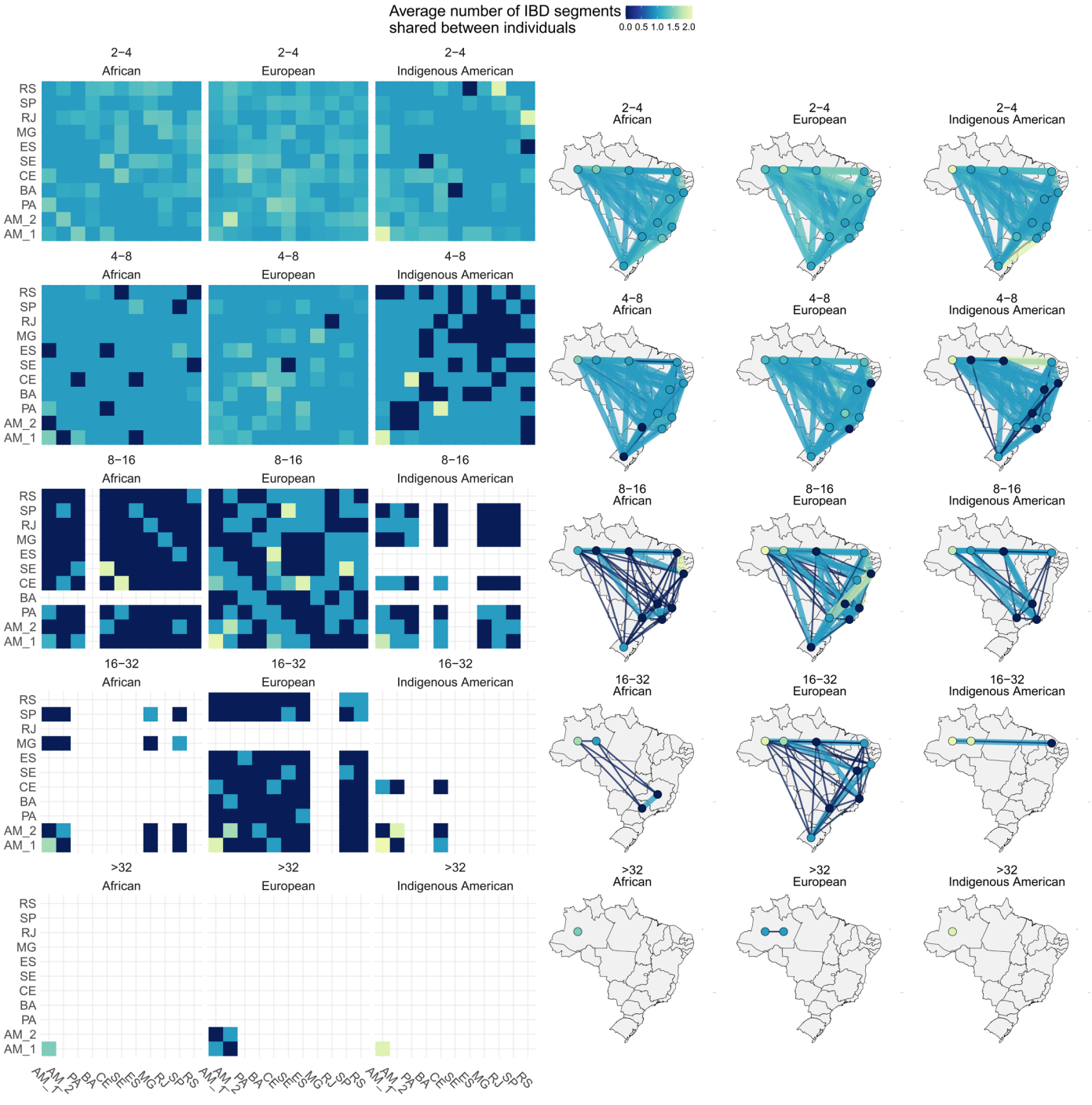


Fig. 4. Ancestry-specific IBD sharing. We used Refined IBD (*114*) and GNOMIX (*24*) to infer the IBD segments and local ancestry, respectively. Based on the local ancestry inference, IBD segments were then classified into African, European, and Indigenous American ancestries and also into different length categories (2–4, 4–8, 8–16, 16–32, and >32). Subsequently, we randomly selected a subset of individuals from each state, matching the smallest sample size of a state (excluding Goiás, which has only four individuals), and the average number of IBD segments shared between individuals from different states in each length category was estimated, and we plotted the natural logarithm of these values to enhance visualization (owing to the presence of outliers). Here, we plot these values in the form of heatmaps (left) and maps (right). For a more detailed description of the analysis and its limitations, refer to (*7*).

regions over time, with fewer shared ancestors. Yet there are relatively more long IBD segments of European origin shared between states compared with other ancestries. The northern state of Amazonas presents a different scenario, characterized by a high number of long IBD segments from the three continental ancestries shared among individuals within the state. This may be attributed to founder events of internal migration, given that Amazonas is the most isolated region of the country (Fig. 4 and fig. S20). Additionally, this pattern is likely influenced by the substantial reduction in the size of the Indigenous American population over the past five centuries, as indicated by the high proportion of medium to long (>8-cM) IBD segments observed in Indigenous American populations (fig. S20).

The Brazilian population as a whole went through a decline in N_e after the start of the colonization period (fig. S21, ~20 gBP red line) for all ancestries [see (*7*) for a definition of N_e, information on how it relates to census size, and analysis limitations]. This is likely related to population bottlenecks caused by the extermination of Indigenous American populations and the founder effects produced by the migration of European colonizers and enslaved Africans forcibly brought to the Americas (*2*, *28*, *52*). In general, the N_e begins to increase at approximately 294 years ago (or 11.75 generations), indicating that the size and genetic diversity of these populations has expanded since then (fig. S21). This date fits with the beginning of the diamond exploration cycle in Brazil, which was followed by the gold cycle. These events began with the massive influx of Portuguese people to Brazil and intensified the trafficking of enslaved Africans more than 10-fold (11).

Conversely, the Amazonas state shows a divergent pattern, where N_e has not stopped decreasing since the European arrival in the Americas, indicating that these populations have been both reduced in the past and inbred in the recent period (fig. S21). The Amazonas state has a low N_e harmonic mean over the past 2500 years, in the same range as the Goiás state, which has a sample size of only four individuals. The highest harmonic mean N_e is observed in São Paulo (813,229.8), which is consistent with the state hosting the largest population size and being the main destination of internal migrants in Brazil in the recent period.

Postcontact selected genes are related to fertility, immune response, and metabolic traits

Signatures of postadmixture natural selection were identified for the three main continental ancestry components (African, European, and Indigenous American) using local ancestry deviation (LAD) (*53*), deep-learning approaches (DL) (*54*), and by comparing allele frequencies in the admixed population with those in their parental populations, as implemented in ADAPTMIX (*55*). Analyses were conducted across the whole dataset (Fig. 5 and Table 1) and within specific geographic regions of Brazil: north (Amazonian; equatorial climate); northeast and southeast (coast; tropical climate), and south (agricultural; temperate climate), enabling the identification of shared selective signals as well as those exclusive to each region (figs. S22 to S24 and tables S9 to S12).

In the northern region, ADAPTMIX (*55*) identified 13 postadmixture loci (fig. S22), including candidate signals associated with *AGAP1* and *HTR1B* genes, which have been associated with height and BMI (*56*, *57*), and *IMPG1*, implicated in respiratory (forced expiratory volume in the first second (FEV1)], metabolic [estimated glomerular filtration rate (eGFR)], and reproductive domains (age at menarche) (*58*, *59*). Other candidates for selection included the zinc finger genes *ZNF83*, *ZNF605*, and *ZNF521*, known for their DNA-binding specificity changes through natural selection (*60*, *61*). The cluster on chromosome 19, which includes *ZNF83*, coincides with a hotspot of copy number variation, suggesting a defense mechanism against endogenous retroviruses (*61*) (table S10).

In the northeastern and southeastern regions, ADAPTMIX also identified candidate genes *IMPG1*, *HTR1b*, and *AGAP1* (fig. S22 and table S11). Using LAD and DL approaches, we found *LINC00871*, previously associated with BMI (*62*), the number of offspring born (*63*), and a suggestive association with longevity in females (*64*) (fig. S23 and table S9). This signal was also found by Cheng *et al.* (*65*) in Mexican populations. However, they associated this signal to the Indigenous American ancestry, whereas in our sample it was linked to the European component. Other candidate genes linked to Indigenous American ancestry include *ACSS1* (associated with cystatin F levels) (*66*), which regulates the cytotoxicity of natural killer (NK) cells (*67*), and *MTMR3*, which plays a role in both the innate immune response (*68*) and metabolic processes such as low-density lipoprotein (LDL) levels (*69*), body weight (*70*), and type 2 diabetes (*71*). We also identified several candidate selection signals along the MHC (major histocompatibility complex) and extended MHC regions in African ancestry. The detection of selection signals in the MHC region is recurrent in the literature and is generally attributed to the large number of genes related to the immune response, particularly the human leukocyte antigen (HLA) gene family. Several studies suggest that natural selection acting in this region could be multilocus, involving both epistasis and genetic hitchhiking (*72*). Finally, the most prominent signal of selection was observed in the *ZNF184/POM121L2* gene region, which has been associated with spermatogenesis (*73*, *74*).

In the southern region, LAD and DL approaches identified the genes *ARHGAP11A* (sleep duration) in African ancestry (*75*) and *ACACA* (associated with menopausal age) in European ancestry (*76*). As in the northeastern and southeastern regions, the candidate gene *ACSS1* (regulating NK cells) was associated with Indigenous American ancestry (fig. S23). Through ADAPTMIX, we detected numerous candidate loci across the genome associated with postadmixture selection (table S12)—notably, genes implicated in immune response, such as *CDH4*, *KBTBD11*, and HLA family genes involved in the immune response. Given the large amount of loci, we evaluated the enrichment of gene sets for biological, cellular, and molecular components, finding substantial evidence for host-pathogen interaction, especially within HLA genes (fig. S24). This is in agreement with other studies of Latin American samples, pointing out that the immune response has been the main selective pressure in admixed populations (*77*–*83*).

Despite the environmental differences among Brazil's geographical regions, we observed that the main postadmixture selective pressures have been directed by traits related to fertility (spermatogenesis, number of offspring born, age at menarche, menopause age), immunity (defense retroviruses, NK regulation, adaptive and innate immune response), and energy metabolism phenotypes (body weight, BMI, LDL levels, sleep duration). These same genomic signatures were also observed when we analyzed the entire dataset (Fig. 5 and Table 1).

Fertility, immunity, and energy metabolism are critical aspects of human biology that directly influence fitness and are potentially strongly subject to natural selection. Fertility is a phenotypic trait with a strong response to selection, and several studies have identified genetic variants associated with hormonal regulation, reproductive health, pregnancy support, and sperm quality, which increase reproductive success (*84*–*86*). Natural selection has strongly targeted the immune response throughout human evolution, favoring genetic variants that improve pathogen recognition, immune response efficiency, and disease resistance (*87*). These adaptations optimize survival and reproductive success in diverse environments. In the Americas, where several pathogens may be distinct from those on other continents, we commonly observe signs of local selective pressures that have acted on the genomes of admixed populations (*55*, *78*, *88*). Genes related to energy metabolism are another recurring target of selection. Studies show that in the past, genes affecting basal metabolic rate, fat storage, and nutrient utilization have been subject to selection pressures (*89*–*92*). Variants that improve energy efficiency and storage are favored, adapting populations to food availability and environmental conditions (*93*). Nowadays, it is recognized that genetic variants improving energy reserves and efficient metabolism can better support reproductive functions and immune response (*94*, *95*).

Conclusions

By leveraging the extensive genetic diversity of the world's largest admixed population through a comprehensive genomic database spanning diverse Brazilian regions, we have demonstrated that fine-scale demographic analyses play a critical role in unraveling diversity's complexity and its implications for population health. Our findings

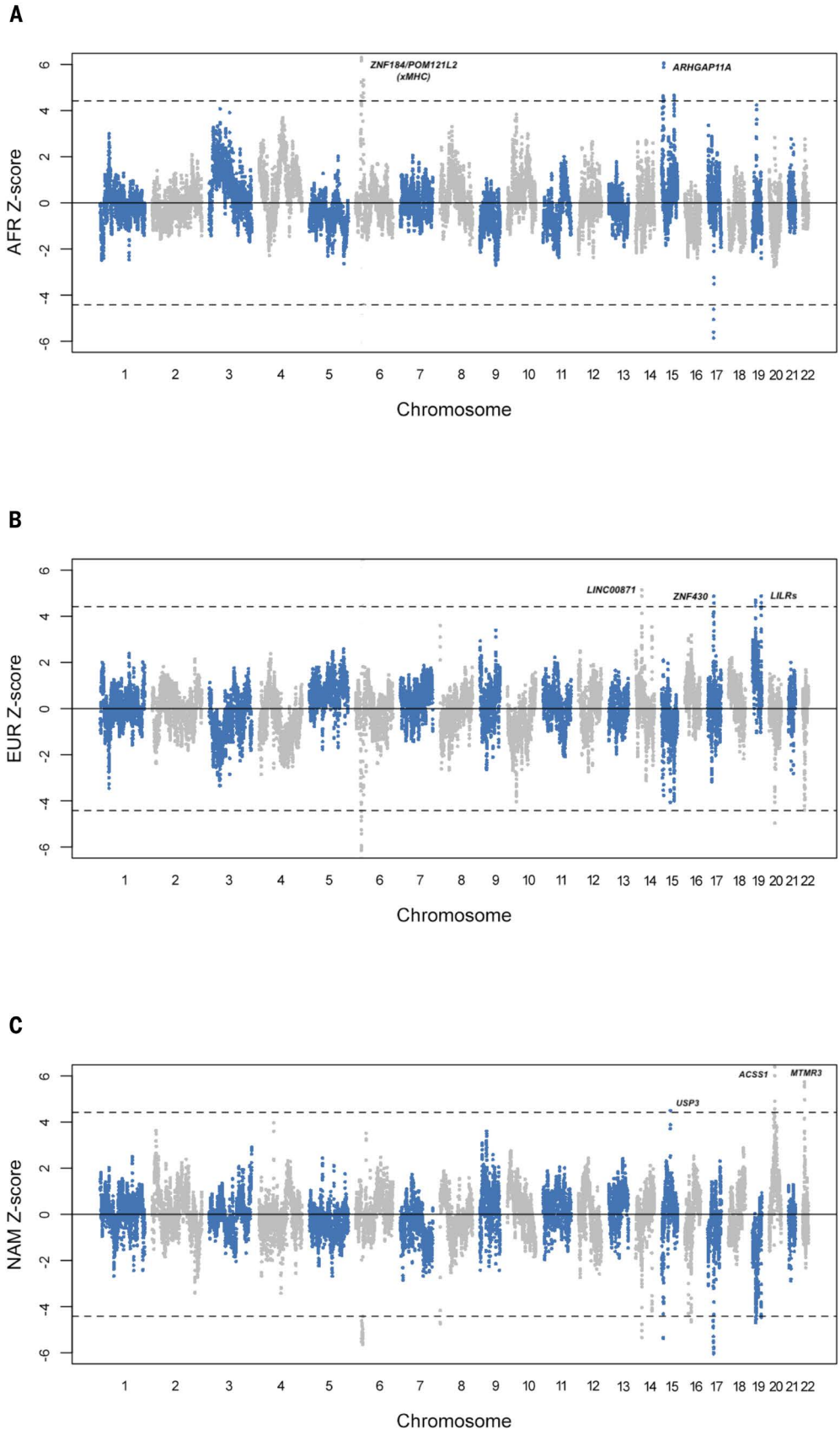


Fig. 5. Genome-wide postadmixture natural selection scan. Genomic variation of African (**A**), European (**B**), and Indigenous American (**C**) ancestries in DNABR samples. The *x* axis denotes the chromosomes, while the *y* axis indicates the number of standard deviations of ancestry in the SNP from the genomic mean. The dashed lines correspond to 4.42 standard deviations (*z*-score) from the mean. The black lines correspond to the genomic mean. The genes identified in the candidate regions are indicated in the figure.

Table 1. Candidate genes with postadmixture selection signatures by *z*-score and deep-learning approaches. Chr, chromosome; bbox, bounding box; LILRs, leukocyte immunoglobulin-like receptors; LRC, leukocyte receptor complex.

Chr	*Z*-score test			Deep-learning test			Target ancestry	Candidate gene/nearest gene
	Start position End position	Block size (max value)	*Z*-score (*P* value)	Start position End position	Block size	Predicted bbox score		
6	27259978 27436536	176 kb	8.16 (3.2×10^{-16})	27256539 27427985	171 kb	0.999	African	*ZNF184/ POM121L2*
15	32520639 32699083	178 kb	8.07 (6.8×10^{-16})	32542539 32721562	179 kb	0.999	African	*ARHGAP11A*
14	46005212 46256913	251 kb	5.14 (2.6×10^{-7})	46011443 46254522	243 kb	0.997	European	*LINC00871*
17	37002394 37147239	144 kb	9.06 (1.3×10^{-19})	37002554 37146392	143 kb	0.999	European	*AATF/ACACA*
19	20963923 21040970	77 kb	4.7 (2.5×10^{-6})	21000776 21053544	52 kb	0.995	European	*ZNF430*
19	54526741 54600262	73 kb	4.88 (1.1×10^{-6})	54673809 54616673	57 kb	0.995	European	LILRs (LRC region)
15	63525359 63798416	273 kb	4.5 (6.7×10^{-6})	63598854 63653908	55 kb	0.994	Native American	*USP3*
20	25044420 25156172	111 kb	6.39 (1.5×10^{-10})	25054327 25098891	44 kb	0.998	Native American	*ACSS1*
22	29829549 29909529	80 kb	5.73 (9.6×10^{-9})	29827910 29907351	79 kb	0.997	Native American	*MTMR3*

underscore the discernible influence of different ancestral backgrounds on Brazilian admixed individuals' health and genetic makeup. We show that this genetic landscape finds its roots in the evolutionary history of Brazilian Indigenous communities and the intricate demographic interplay stemming from both coerced and voluntary historical immigration to Brazil.

Materials and methods summary

High-coverage (average 35×) whole-genome sequences were generated for all samples by DASA Genomica using the Illumina NovaSeq 6000 System and applying the Illumina DNA PCR-Free workflow. Individual whole genomes were produced with GATK's HaplotypeCaller in GVCF mode, and a joint cohort variant call format (VCF) was generated following GATK best practices (JointGenotyping 2.2.0 workflow, GATK docker 4.1.8.0). Autosomal variants were annotated using the Ensembl Variant Effect Predictor (VEP) (*96*) and LOFTEE (*12*) plugin. Novel SNVs were determined by comparison with the following public datasets: dbSNP v.155 (https://ftp.ncbi.nlm.nih.gov/snp/), HGDP (*8*), 1KGP (*97*), and gnomAD v3 genomes (*98*). Ancestral/derived alleles were annotated with VEP's AncestralAllele plugin (*96*), and SNVs with a derived allele frequency (DAF) below a threshold of 1% in the DNABR sample were classified as rare. Rare variants were then annotated with PolyPhen2 (*13*) and SIFT (*99*) using VEP (*96*).

Using BCFTools and keeping only shared variants, the DNABR dataset was merged with HGDP (*8*) and SGDP (*9*) (all mapped to the GRCh38/hg38 reference genome). Next, biallelic SNVs with <5% missing data and no high deviation ($P \leq 10^{-8}$) from the Hardy-Weinberg equilibrium were selected. Variants with a pairwise correlation >20% in a genomic window of 500 kb were also removed before PCA and ADMIXTURE (*100*). The LD-pruning filter and the PCA were executed with the SNPRelate R/Bioconductor package (*23*). For the haplotype-based analysis, we extracted from the WGS dataset the set of variants included in the Axiom Human Origins array (Affymetrix/Thermo Fisher) and combined with three published datasets of present-day Indigenous Americans (*10*, *11*, *101*). PLINK v1.9 (*102*) pairwise IBD inference method was used to estimate the PI_HAT between all pairs of individuals, and PRIMUS (*103*) was applied to identify the maximum set of unrelated individuals.

Inferences of global ancestry from autosomal and X chromosome data were made with ADMIXTURE (*100*). The biological males in the sample were defined through the detection of the Y chromosome, the homozygosity, and the inbreeding coefficient of the X chromosome ($F < 0.2$, female; and $F > 0.8$, male) using PLINK v1.9 (*102*). Y chromosome haplogroup inference was performed with the SNAPPY software (*104*). Mitochondrial DNA haplogroup inferences were performed using the HaploGrep3 v3.3.1.0 software (*105*). Ancestry was assigned to haplogroups according to their most frequent macrogeographic distribution.

Haplotypic phase was inferred with SHAPEIT4 (Segmented HAPlotype Estimation and Imputation Tools version 4; https://odelaneau.github.io/shapeit4/) (*106*) using the 1000 Genomes Project data as a reference panel. We applied CHROMOPAINTER (*33*) to infer haplotype similarity and sharing among samples and fineSTRUCTURE (*33*) to infer genetically homogeneous clusters of individuals. A second CHROMOPAINTER analysis was performed in "donor mode" to generate copying vectors summarizing the haplotype similarity between both donors and surrogates (reference panel samples) and donors and recipients (DNABR samples). Then SOURCEFIND (*107*) was used to model the copying vectors of the recipients as a weighted mixture of the copying vectors from the surrogates. We then applied fastGLOBETROTTER (*35*), MALDER (*108*), and TRACTS (*37*) to infer admixture events and dates.

Local ancestry inference was performed with GNOMIX (*24*), and the ancestry-specific effective population size history was inferred with the IBDNe pipeline (*52*). The inferred IBD segments were also used to estimate the patterns of IBD sharing within and among Brazilian federal states. ANCESTOR was used to infer the ancestry proportions of mating pairs over the parental generation, independently for the autosomes and the X chromosome. ROH were identified with PLINK v1.9 (*109*).

Different approaches were adopted to perform analyses of natural selection. The local ancestry inferences obtained with GNOMIX were used to identify local ancestry deviation, applying the *z*-score test separately for each ancestral component (*53*). We also used a deep-learning object detection strategy (*54*) to identify genomic regions under selection from images generated from population-level local ancestry tract distribution along the chromosomes. ADAPTMIX (*55*) was used to detect

loci under selection and infer whether selection happened after or before admixture. Gene annotation of regions under selection was performed with ANNOVAR (*110*), and biological process enrichment was carried out with all SNPs under positive selection through the Web-based Gene Set Analysis Toolkit (WebGestalt) (*111*).

For a more comprehensive description of the methods used, refer to (*7*).

REFERENCES AND NOTES

1. F. M. Salzano, M. C. Bortolini, *The Evolution and Genetics of Latin American Populations*, vol. 28 of *Cambridge Studies in Biological and Evolutionary Anthropology* (Cambridge Univ. Press, 2005).
2. M. A. Castro e Silva *et al.*, Population histories and genomic diversity of South American natives. *Mol. Biol. Evol.* **39**, msab339 (2022). doi: 10.1093/molbev/msab339; pmid: 34875092
3. T. Hünemeier *et al.*, Niger-Congo speaking populations and the formation of the Brazilian gene pool: mtDNA and Y-chromosome data. *Am. J. Phys. Anthropol.* **133**, 854–867 (2007). doi: 10.1002/ajpa.20604; pmid: 17427922
4. M. H. Gouveia *et al.*, Origins, admixture dynamics, and homogenization of the African gene pool in the Americas. *Mol. Biol. Evol.* **37**, 1647–1656 (2020). doi: 10.1093/molbev/msaa033; pmid: 32128591
5. F. S. G. Kehdy *et al.*, Origin and dynamics of admixture in Brazilians and its effect on the pattern of deleterious mutations. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 8696–8701 (2015). doi: 10.1073/pnas.1504447112; pmid: 26124090
6. M. S. Naslavsky *et al.*, Whole-genome sequencing of 1,171 elderly admixed individuals from Brazil. *Nat. Commun.* **13**, 1004 (2022). doi: 10.1038/s41467-022-28648-3; pmid: 35246524
7. Materials and methods are available as supplementary materials.
8. A. Bergström *et al.*, Insights into human genetic variation and population history from 929 diverse genomes. *Science* **367**, eaay5012 (2020). doi: 10.1126/science.aay5012; pmid: 32193295
9. S. Mallick *et al.*, The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201–206 (2016). doi: 10.1038/nature18964; pmid: 27654912
10. P. Skoglund *et al.*, Genetic evidence for two founding populations of the Americas. *Nature* **525**, 104–108 (2015). doi: 10.1038/nature14895; pmid: 26196601
11. M. A. Castro E Silva *et al.*, Genomic insight into the origins and dispersal of the Brazilian coastal natives. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 2372–2377 (2020). doi: 10.1073/pnas.1909075117; pmid: 31932419
12. K. J. Karczewski *et al.*, The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020). doi: 10.1038/s41586-020-2308-7; pmid: 32461654
13. I. A. Adzhubei *et al.*, A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010). doi: 10.1038/nmeth0410-248; pmid: 20354512
14. P. C. Ng, S. Henikoff, Predicting deleterious amino acid substitutions. *Genome Res.* **11**, 863–874 (2001). doi: 10.1101/gr.176601; pmid: 11337480
15. Z. A. Szpiech *et al.*, Ancestry-dependent enrichment of deleterious homozygotes in runs of homozygosity. *Am. J. Hum. Genet.* **105**, 747–762 (2019). doi: 10.1016/j.ajhg.2019.08.011; pmid: 31543216
16. S. D. Niedbalski, J. C. Long, Novel alleles gained during the Beringian isolation period. *Sci. Rep.* **12**, 4289 (2022). doi: 10.1038/s41598-022-08212-1; pmid: 35277570
17. R. Rodríguez-Labrada *et al.*, Founder effects of spinocerebellar ataxias in the American continents and the Caribbean. *Cerebellum* **19**, 446–458 (2020). doi: 10.1007/s12311-020-01109-7; pmid: 32086717
18. L. S. Sena, J. Dos Santos Pinheiro, M. L. Saraiva-Pereira, L. B. Jardim, Selective forces acting on spinocerebellar ataxia type 3/Machado-Joseph disease recurrency: A systematic review and meta-analysis. *Clin. Genet.* **99**, 347–358 (2021). doi: 10.1111/cge.13888; pmid: 33219521
19. M. I. Achatz, G. P. Zambetti, The inherited p53 mutation in the Brazilian population. *Cold Spring Harb. Perspect. Med.* **6**, a026195 (2016). doi: 10.1101/cshperspect.a026195; pmid: 27663983
20. M. I. W. Achatz *et al.*, The *TP53* mutation, R337H, is associated with Li-Fraumeni and Li-Fraumeni-like syndromes in Brazilian families. *Cancer Lett.* **245**, 96–102 (2007). doi: 10.1016/j.canlet.2005.12.039; pmid: 16494995
21. S. Garritano *et al.*, Detailed haplotype analysis at the *TP53* locus in p.R337H mutation carriers in the population of Southern Brazil: Evidence for a founder effect. *Hum. Mutat.* **31**, 143–150 (2010). doi: 10.1002/humu.21151; pmid: 19877175
22. B. M. Henn *et al.*, Distance from sub-Saharan Africa predicts mutational load in diverse human genomes. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E440–E449 (2016). doi: 10.1073/pnas.1510805112; pmid: 26712023
23. X. Zheng *et al.*, A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**, 3326–3328 (2012). doi: 10.1093/bioinformatics/bts606; pmid: 23060615
24. H. Hilmarsson *et al.*, High resolution ancestry deconvolution for next generation genomic data. bioRxiv 2021.09.19.460980 [Preprint] (2021); https://doi.org/10.1101/2021.09.19.460980.
25. A. Ruiz-Linares *et al.*, Admixture in Latin America: Geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals. *PLOS Genet.* **10**, e1004572 (2014). doi: 10.1371/journal.pgen.1004572; pmid: 25254375
26. K. L. Korunes *et al.*, Sex-biased admixture and assortative mating shape genetic variation and influence demographic inference in admixed Cabo Verdeans. *G3* **12**, jkac183 (2022). doi: 10.1093/g3journal/jkac183; pmid: 35861404
27. S. D. J. Pena, Ed., *Homo brasilis: Aspectos Genéticos, Lingüísticos, Históricos e Socioantropológicos da Formação do Povo Brasileiro* (FUNPEC-RP, 2002).
28. K. Adhikari, J. C. Chacón-Duque, J. Mendoza-Revilla, M. Fuentes-Guajardo, A. Ruiz-Linares, The genetic diversity of the Americas. *Annu. Rev. Genomics Hum. Genet.* **18**, 277–296 (2017). doi: 10.1146/annurev-genom-083115-022331; pmid: 28859572
29. J. Y. Zou, E. Halperin, E. Burchard, S. Sankararaman, Inferring parental genomic ancestries using pooled semi-Markov processes. *Bioinformatics* **31**, i190–i196 (2015). doi: 10.1093/bioinformatics/btv239; pmid: 26072482
30. J. Y. Zou *et al.*, Genetic and socioeconomic study of mate choice in Latinos reveals novel assortment patterns. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 13621–13626 (2015). doi: 10.1073/pnas.1501741112; pmid: 26483472
31. K. Bryc, E. Y. Durand, J. M. Macpherson, D. Reich, J. L. Mountain, The genetic ancestry of African Americans, Latinos, and European Americans across the United States. *Am. J. Hum. Genet.* **96**, 37–53 (2015). doi: 10.1016/j.ajhg.2014.11.010; pmid: 25529636
32. A. Mas-Sandoval, S. Mathieson, M. Fumagalli, The genomic footprint of social stratification in admixing American populations. *eLife* **12**, e84429 (2023). doi: 10.7554/eLife.84429; pmid: 38038347
33. D. J. Lawson, G. Hellenthal, S. Myers, D. Falush, Inference of population structure using dense haplotype data. *PLOS Genet.* **8**, e1002453 (2012). doi: 10.1371/journal.pgen.1002453; pmid: 22291602
34. P. Moorjani, G. Hellenthal, Methods for assessing population relationships and history using genomic data. *Annu. Rev. Genomics Hum. Genet.* **24**, 305–332 (2023). doi: 10.1146/annurev-genom-111422-025117; pmid: 37220313
35. P. Wangkumhang, M. Greenfield, G. Hellenthal, An efficient method to identify, date, and describe admixture events using haplotype information. *Genome Res.* **32**, 1553–1564 (2022). doi: 10.1101/gr.275994.121; pmid: 35794007
36. J. K. Pickrell *et al.*, Ancient west Eurasian ancestry in southern and eastern Africa. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 2632–2637 (2014). doi: 10.1073/pnas.1313787111; pmid: 24550290
37. S. Gravel, Population genetics models of local ancestry. *Genetics* **191**, 607–619 (2012). doi: 10.1534/genetics.112.139808; pmid: 22491189
38. L. M. Schwarcz, H. M. M. Starling, *Brazil: A Biography* (Penguin UK, 2018).
39. T. E. Skidmore, *Brazil: Five Centuries of Change* (Oxford Univ. Press, ed. 2, 2010).
40. H. Langfur, The return of the *Bandeira*: Economic calamity, historical memory, and armed expeditions to the *sertão* in Minas Gerais, 1750-1808. *Americas* **61**, 429–461 (2005). doi: 10.1353/tam.2005.0025
41. L. M. Schwarcz, *The Spectacle of the Races: Scientists, Institutions, and the Race Question in Brazil, 1870–1930* (Farrar, Straus and Giroux, 1999).
42. *Brasil: 500 anos de povoamento* (Instituto Brasileiro de Geografia e Estatística, 2000).
43. G. M. Greenfield, The realities of images: Imperial Brazil and the great drought. *Trans. Am. Philos. Soc.* **91**, 1–148 (2001). doi: 10.2307/3655112
44. T. J. Pemberton *et al.*, Genomic patterns of homozygosity in worldwide human populations. *Am. J. Hum. Genet.* **91**, 275–292 (2012). doi: 10.1016/j.ajhg.2012.06.014; pmid: 22883143
45. F. C. Ceballos, P. K. Joshi, D. W. Clark, M. Ramsay, J. F. Wilson, Runs of homozygosity: Windows into population history and trait architecture. *Nat. Rev. Genet.* **19**, 220–234 (2018). doi: 10.1038/nrg.2017.109; pmid: 29335644
46. R. B. Lemes *et al.*, Inbreeding estimates in human populations: Applying new approaches to an admixed Brazilian isolate. *PLOS ONE* **13**, e0196360 (2018). doi: 10.1371/journal.pone.0196360; pmid: 29689090
47. J. A. Mooney *et al.*, Understanding the hidden complexity of Latin American population isolates. *Am. J. Hum. Genet.* **103**, 707–726 (2018). doi: 10.1016/j.ajhg.2018.09.013; pmid: 30401458
48. M. Sohail *et al.*, Mexican Biobank advances population and medical genomics of diverse ancestries. *Nature* **622**, 775–783 (2023). doi: 10.1038/s41586-023-06560-0; pmid: 37821706
49. N. A. Swinford *et al.*, Increased homozygosity due to endogamy results in fitness consequences in a human population. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2309552120 (2023). doi: 10.1073/pnas.2309552120; pmid: 37847737
50. S. Baharian *et al.*, The Great Migration and African-American genomic diversity. *PLOS Genet.* **12**, e1006059 (2016). doi: 10.1371/journal.pgen.1006059; pmid: 27232753
51. D. N. Harris *et al.*, Evolutionary genomic dynamics of Peruvians before, during, and after the Inca Empire. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E6526–E6535 (2018). doi: 10.1073/pnas.1720798115; pmid: 29946025

52. S. R. Browning *et al.*, Ancestry-specific recent effective population size in the Americas. *PLOS Genet.* **14**, e1007385 (2018). doi: 10.1371/journal.pgen.1007385; pmid: 29795556
53. H. Tang *et al.*, Recent genetic selection in the ancestral admixture of Puerto Ricans. *Am. J. Hum. Genet.* **81**, 626–633 (2007). doi: 10.1086/520769; pmid: 17701908
54. I. Hamid, K. L. Korunes, D. R. Schrider, A. Goldberg, Localizing post-admixture adaptive variants with object detection on ancestry-painted chromosomes. *Mol. Biol. Evol.* **40**, msad074 (2023). doi: 10.1093/molbev/msad074; pmid: 36947126
55. J. Mendoza-Revilla *et al.*, Disentangling signatures of selection before and after European colonization in Latin Americans. *Mol. Biol. Evol.* **39**, msac076 (2022). doi: 10.1093/molbev/msac076; pmid: 35460423
56. K. Watanabe *et al.*, A global overview of pleiotropy and genetic architecture in complex traits. *Nat. Genet.* **51**, 1339–1348 (2019). doi: 10.1038/s41588-019-0481-0; pmid: 31427789
57. L. Yengo *et al.*, Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum. Mol. Genet.* **27**, 3641–3649 (2018). doi: 10.1093/hmg/ddy271; pmid: 30124842
58. G. L. Wojcik *et al.*, Genetic analyses of diverse populations improves discovery for complex traits. *Nature* **570**, 514–518 (2019). doi: 10.1038/s41586-019-1310-4; pmid: 31217584
59. F. R. Day *et al.*, Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nat. Genet.* **49**, 834–841 (2017). doi: 10.1038/ng.3841; pmid: 28436984
60. D. Schmidt, R. Durrett, Adaptive evolution drives the diversification of zinc-finger binding domains. *Mol. Biol. Evol.* **21**, 2326–2339 (2004). doi: 10.1093/molbev/msh246; pmid: 15342798
61. S. Lukic, J.-C. Nicolas, A. J. Levine, The diversity of zinc-finger genes on human chromosome 19 provides an evolutionary mechanism for defense against inherited endogenous retroviruses. *Cell Death Differ.* **21**, 381–387 (2014). doi: 10.1038/cdd.2013.150; pmid: 24162661
62. J. Huang *et al.*, Genomics and phenomics of body mass index reveals a complex disease network. *Nat. Commun.* **13**, 7973 (2022). doi: 10.1038/s41467-022-35553-2; pmid: 36581621
63. N. Barban *et al.*, Genome-wide analysis identifies 12 loci influencing human reproductive behavior. *Nat. Genet.* **48**, 1462–1472 (2016). doi: 10.1038/ng.3698; pmid: 27798627
64. Y. Zeng *et al.*, Sex differences in genetic associations with longevity. *JAMA Netw. Open* **1**, e181670 (2018). doi: 10.1001/jamanetworkopen.2018.1670; pmid: 30294719
65. J. Y. Cheng, A. J. Stern, F. Racimo, R. Nielsen, Detecting selection in multiple populations by modeling ancestral admixture components. *Mol. Biol. Evol.* **39**, msab294 (2022). doi: 10.1093/molbev/msab294; pmid: 34626111
66. G. Thareja *et al.*, Differences and commonalities in the genetic architecture of protein quantitative trait loci in European and Arab populations. *Hum. Mol. Genet.* **32**, 907–916 (2023). doi: 10.1093/hmg/ddac243; pmid: 36168886
67. M. Perišić Nanut, J. Sabotič, U. Švajger, A. Jewett, J. Kos, Cystatin F affects natural killer cell cytotoxicity. *Front. Immunol.* **8**, 1459 (2017). doi: 10.3389/fimmu.2017.01459; pmid: 29180998
68. D. Dewi Pamungkas Putri *et al.*, PtdIns3P phosphatases MTMR3 and MTMR4 negatively regulate innate immune responses to DNA through modulating STING trafficking. *J. Biol. Chem.* **294**, 8412–8423 (2019). doi: 10.1074/jbc.RA118.005731; pmid: 30944173
69. Global Lipids Genetics Consortium, Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1283 (2013). doi: 10.1038/ng.2797; pmid: 24097068
70. S. Sakaue *et al.*, A cross-population atlas of genetic associations for 220 human phenotypes. *Nat. Genet.* **53**, 1415–1424 (2021). doi: 10.1038/s41588-021-00931-x; pmid: 34594039
71. A. Mahajan *et al.*, Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes. *Nat. Genet.* **50**, 559–571 (2018). doi: 10.1038/s41588-018-0084-1; pmid: 29632382
72. D. Meyer, V. R. C. Aguiar, B. D. Bitarello, D. Y. C. Brandt, K. Nunes, A genomic perspective on HLA evolution. *Immunogenetics* **70**, 5–27 (2018). doi: 10.1007/s00251-017-1017-3; pmid: 28687858
73. A. Ziegler, P. S. C. Santos, T. Kellermann, B. Uchanska-Ziegler, Self/nonself perception, reproduction and the extended MHC. *Self Nonself* **1**, 176–191 (2010). doi: 10.4161/self.1.3.12736; pmid: 21487476
74. S. Fan *et al.*, ZNF185-derived peptide induces fertility suppression in mice. *J. Pept. Sci.* **24**, e3121 (2018). doi: 10.1002/psc.3121; pmid: 30270484
75. A. G. Comuzzie *et al.*, Novel genetic loci identified for the pathophysiology of childhood obesity in the Hispanic population. *PLOS ONE* **7**, e51954 (2012). doi: 10.1371/journal.pone.0051954; pmid: 23251661
76. L. Zhang *et al.*, Joint genome-wide association analyses identified 49 novel loci for age at natural menopause. *J. Clin. Endocrinol. Metab.* **106**, 2574–2591 (2021). doi: 10.1210/clinem/dgab377; pmid: 34050765
77. R. Caro-Consuegra *et al.*, Uncovering signals of positive selection in Peruvian populations from three ecological regions. *Mol. Biol. Evol.* **39**, msac158 (2022). doi: 10.1093/molbev/msac158; pmid: 35860855
78. L. Deng, A. Ruiz-Linares, S. Xu, S. Wang, Ancestry variation and footprints of natural selection along the genome in Latin American populations. *Sci. Rep.* **6**, 21766 (2016). doi: 10.1038/srep21766; pmid: 26887503
79. J. Lindo *et al.*, The genetic prehistory of the Andean highlands 7000 years BP though European contact. *Sci. Adv.* **4**, eaau4921 (2018). doi: 10.1126/sciadv.aau4921; pmid: 30417096
80. J. Lindo *et al.*, A time transect of exomes from a Native American population before and after European contact. *Nat. Commun.* **7**, 13175 (2016). doi: 10.1038/ncomms13175; pmid: 27845766
81. A. W. Reynolds *et al.*, Comparing signals of natural selection between three Indigenous North American populations. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 9312–9317 (2019). doi: 10.1073/pnas.1819467116; pmid: 30988184
82. Q. Zhou, L. Zhao, Y. Guan, Strong selection at MHC in Mexicans since admixture. *PLOS Genet.* **12**, e1005847 (2016). doi: 10.1371/journal.pgen.1005847; pmid: 26863142
83. L. Vicuña *et al.*, Postadmixture selection on Chileans targets haplotype involved in pigmentation, thermogenesis and immune defense against pathogens. *Genome Biol. Evol.* **12**, 1459–1470 (2020). doi: 10.1093/gbe/evaa136; pmid: 32614437
84. F. C. Tropf *et al.*, Human fertility, molecular genetics, and natural selection in modern societies. *PLOS ONE* **10**, e0126821 (2015). doi: 10.1371/journal.pone.0126821; pmid: 26039877
85. I. Mathieson *et al.*, Genome-wide analysis identifies genetic effects on reproductive success and ongoing natural selection at the *FADS* locus. *Nat. Hum. Behav.* **7**, 790–801 (2023). doi: 10.1038/s41562-023-01528-6; pmid: 36864135
86. E. A. Brown, M. Ruvolo, P. C. Sabeti, Many ways to die, one way to arrive: How selection acts through pregnancy. *Trends Genet.* **29**, 585–592 (2013). doi: 10.1016/j.tig.2013.03.001; pmid: 23566676
87. L. B. Barreiro, L. Quintana-Murci, Evolutionary and population (epi)genetics of immunity to infection. *Hum. Genet.* **139**, 723–732 (2020). doi: 10.1007/s00439-020-02167-x; pmid: 32285198
88. R. Secolin *et al.*, Distribution of local ancestry and evidence of adaptation in admixed populations. *Sci. Rep.* **9**, 13900 (2019). doi: 10.1038/s41598-019-50362-2; pmid: 31554886
89. M. Fumagalli *et al.*, Greenlandic Inuit show genetic signatures of diet and climate adaptation. *Science* **349**, 1343–1347 (2015). doi: 10.1126/science.aab2319; pmid: 26383953
90. C. E. G. Amorim *et al.*, Genetic signature of natural selection in first Americans. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 2195–2199 (2017). doi: 10.1073/pnas.1620541114; pmid: 28193867
91. S. Fan *et al.*, African evolutionary history inferred from whole genome sequence data of 44 indigenous African populations. *Genome Biol.* **20**, 82 (2019). doi: 10.1186/s13059-019-1679-2; pmid: 31023338
92. A. M. Hancock *et al.*, Human adaptations to diet, subsistence, and ecoregion are due to subtle shifts in allele frequency. *Proc. Natl. Acad. Sci. U.S.A.* **107** (suppl. 2), 8924–8930 (2010). doi: 10.1073/pnas.0914625107; pmid: 20445095
93. F. Luca, G. H. Perry, A. Di Rienzo, Evolutionary adaptations to dietary changes. *Annu. Rev. Nutr.* **30**, 291–314 (2010). doi: 10.1146/annurev-nutr-080508-141048; pmid: 20420525
94. K. Thiele, L. Diao, P. C. Arck, Immunometabolism, pregnancy, and nutrition. *Semin. Immunopathol.* **40**, 157–174 (2018). doi: 10.1007/s00281-017-0660-y; pmid: 29071391
95. E. Tourkochristou, C. Triantos, A. Mouzaki, The influence of nutritional factors on immunological outcomes. *Front. Immunol.* **12**, 665968 (2021). doi: 10.3389/fimmu.2021.665968; pmid: 34135894
96. W. McLaren *et al.*, The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016). doi: 10.1186/s13059-016-0974-4; pmid: 27268795
97. M. Byrska-Bishop *et al.*, High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. *Cell* **185**, 3426–3440.e19 (2022). doi: 10.1016/j.cell.2022.08.004; pmid: 36055201
98. S. Chen *et al.*, A genomic mutational constraint map using variation in 76,156 human genomes. *Nature* **625**, 92–100 (2024). doi: 10.1038/s41586-023-06045-0; pmid: 38057664
99. P. C. Ng, S. Henikoff, SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**, 3812–3814 (2003). doi: 10.1093/nar/gkg509; pmid: 12824425
100. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009). doi: 10.1101/gr.094052.109; pmid: 19648217
101. M. A. Castro e Silva, T. Ferraz, M. C. Bortolini, D. Comas, T. Hünemeier, Deep genetic affinity between coastal Pacific and Amazonian natives evidenced by Australasian ancestry. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2025739118 (2021). doi: 10.1073/pnas.2025739118; pmid: 33782134
102. C. C. Chang *et al.*, Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015). doi: 10.1186/s13742-015-0047-8; pmid: 25722852
103. J. Staples, D. A. Nickerson, J. E. Below, Utilizing graph theory to select the largest set of unrelated individuals for genetic analysis. *Genet. Epidemiol.* **37**, 136–141 (2013). doi: 10.1002/gepi.21684; pmid: 22996348
104. A. L. Severson *et al.*, SNAPPY: Single nucleotide assignment of phylogenetic parameters on the Y chromosome. bioRxiv 454736 [Preprint] (2018); https://doi.org/10.1101/454736.
105. H. Weissensteiner *et al.*, HaploGrep 2: Mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* **44**, W58–W63 (2016). doi: 10.1093/nar/gkw233; pmid: 27084951

106. O. Delaneau, J.-F. Zagury, M. R. Robinson, J. L. Marchini, E. T. Dermitzakis, Accurate, scalable and integrative haplotype estimation. *Nat. Commun.* **10**, 5436 (2019). doi: 10.1038/s41467-019-13225-y; pmid: 31780650
107. J.-C. Chacón-Duque *et al.*, Latin Americans show wide-spread Converso ancestry and imprint of local Native ancestry on physical appearance. *Nat. Commun.* **9**, 5388 (2018). doi: 10.1038/s41467-018-07748-z; pmid: 30568240
108. P.-R. Loh *et al.*, Inferring admixture histories of human populations using linkage disequilibrium. *Genetics* **193**, 1233–1254 (2013). doi: 10.1534/genetics.112.147330; pmid: 23410830
109. S. Purcell *et al.*, PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007). doi: 10.1086/519795; pmid: 17701901
110. K. Wang, M. Li, H. Hakonarson, ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010). doi: 10.1093/nar/gkq603; pmid: 20601685
111. Y. Liao, J. Wang, E. J. Jaehnig, Z. Shi, B. Zhang, WebGestalt 2019: Gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* **47**, W199–W205 (2019). doi: 10.1093/nar/gkz401; pmid: 31114916
112. A. Dür, N. Huber, W. Parson, Fine-tuning phylogenetic alignment and haplogrouping of mtDNA sequences. *Int. J. Mol. Sci.* **22**, 5747 (2021). doi: 10.3390/ijms22115747; pmid: 34072215
113. S. Kerminen *et al.*, Fine-scale genetic structure in Finland. *G3* **7**, 3459–3468 (2017). doi: 10.1534/g3.117.300217; pmid: 28983069
114. B. L. Browning, S. R. Browning, Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics* **194**, 459–471 (2013). doi: 10.1534/genetics.113.150029; pmid: 23535385

ACKNOWLEDGMENTS

We thank all the individuals and communities who participated in the study. We also thank Google Cloud Brasil and Dasa Genômica for support in cloud computing and WGS, respectively; and V. Ramallo (CONICET, Argentina) for graphical assistance. **Funding:** FAPESP 21/06860-8 (T.H.), CSIC 20223AT005 (P.P.-V.); CAPES 88887.505029/2020 (K.N.); Brazilian National Program of Genomics and Precision Health–Genomas Brasil, Departamento de Ciência e Tecnologia da Secretaria de Ciência, Tecnologia e Inovação e do Complexo Econômico-Industrial da Saúde do Ministério da Saúde (Decit/SECTICS/MS) (888379/2019) (T.H., K.N., L.V.P., A.C.P., R.B.L., and L.S.d.S.); European Union's Horizon 2020 Research and Innovation Programme/Marie Skłodowska-Curie COFUND/EUTOPIA 945380 and grant JDC2022-049175-I funded by MICIU/AEI/10.13039/501100011033 and by "European Union NextGenerationEU/PRTR" (M.A.C.e.S.). National Council for Scientific and Technological Development (CNPq/465657/2014-1) supported the project "Instituto Nacional de Epidemiologia Genética," which made possible all the organization, idealization, and collection of samples from Western Amazon area (H.K. and R.G.M.F.). **Author contributions:** Conceptualization: A.C.P., L.V.P., T.H. Methodology: K.N., M.A.C.e.S., M.R.R., R.B.L., P.P.-V., L.K., L.S.d.S. Investigation: K.N., M.A.C.e.S., M.R.R., R.B.L., D.C., L.V.P., A.C.P., T.H. Data acquisition: J.E.K., M.C.V., L.O.d.A., L.M.A.C., R.G.M.F., H.K., M.C.B., J.G.M., P.S., J.F.G., C.M.B.d.S., F.V.V., F.S.L.V. Funding acquisition: L.V.P., A.C.P., T.H. Project administration: L.V.P., T.H. Supervision: T.H. Writing – original draft: K.N., M.A.C.e.S., R.B.L., L.V.P., A.C.P., T.H. Writing – review & editing: T.H. **Competing interests:** L.V.P. is a founder of gen-t Science. **Data and materials availability:** All data needed to evaluate the conclusions in this paper are present in the paper and/or the supplementary materials. The datasets used in this paper have been deposited in the European Genome-phenome Archive and are available for download under accession number EGAD50000001012. They can also be freely accessed at DNA do Brasil Variant Browser.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adl3564
Materials and Methods; Figs. S1 to S26; Tables S1 to S12; References (*115*–*129*); MDAR Reproducibility Checklist

Submitted 17 November 2023; accepted 21 February 2025

10.1126/science.adl3564