

Biogeographic Perspectives on Human Genetic Diversification

Tábita Hünemeier ^{1,2,*}

¹Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, SP, Brazil

²Population Genetics Department, Institute of Evolutionary Biology (IBE - CSIC/Universitat Pompeu Fabra), 08003 Barcelona, Spain

*Corresponding author: Email: hunemeier@usp.br or tabita.hunemeier@csic.es.

Associate editor: Claudia Russo

Abstract

Modern humans originated in Africa 300,000 yr ago, and before leaving their continent of origin, they underwent a process of intense diversification involving complex demographic dynamics. Upon exiting Africa, different populations emerged on the four other inhabited continents, shaped by the interplay of various evolutionary processes, such as migrations, founder effects, and natural selection. Within each region, continental populations, in turn, diversified and evolved almost independently for millennia. As a backdrop to this diversification, introgressions from archaic species contributed to establishing different patterns of genetic diversity in different geographic regions, reshaping our understanding of our species' variability. With the increasing availability of genomic data, it has become possible to delineate the subcontinental human population structure precisely. However, the bias toward the genomic research focused on populations from the global North has limited our understanding of the real diversity of our species and the processes and events that guided different human groups throughout their evolutionary history. This perspective is part of a series of articles celebrating 40 yr since our journal, *Molecular Biology and Evolution*, was founded (Russo et al. 2024). The perspective is accompanied by virtual issues, a selection of papers on human diversification published by *Genome Biology and Evolution* and *Molecular Biology and Evolution*.

Key words: human diversity, genomics, biogeography, underrepresented populations.

Introduction

Human populations differ genetically according to their geographic continent of origin (Mallick et al. 2016; Bergström et al. 2020). This diversity also encompasses cultural and linguistic aspects, resulting from the expansion and diversification of these different populations throughout the great human expansion (Henn et al. 2012), which began in Africa ~300,000 yr ago. The traditional view of the emergence of our species points to a single population of *Homo sapiens* originating within the African continent, likely in the eastern or southern regions of Africa. However, it is challenging to reconcile this view with new evidence found in other regions of Africa, such as the fossils from Jebel Irhoud, Morocco (Hublin et al. 2017). A model based on the complete genomes of present-day populations suggests an initial human population split between 120,000 and 135,000 yr ago. In this model, two or more genetically differentiated populations of ancestral *H. sapiens* would have intermixed for hundreds of thousands of years, giving rise to a genetically diverse and weakly structured ancestral African population due to intense gene flow (Ragsdale et al. 2023). This ancient weakly structured model elucidates polymorphism patterns that were previously thought to arise from the contributions of archaic hominins in Africa.

Additionally, other studies indicate admixing events between pre-out-of-Africa *H. sapiens* populations and “ghost” populations (Lorente-Galdos et al. 2019; Durvasula and Sankararaman 2020), with significant genetic contributions from these unsampled populations to the human species. Our species originated in Africa during this period and expanded and diversified within the African continent for over 100,000 yr before migrating to other continents. Although there were likely numerous attempts to migrate out of Africa during the period of exploration and settlement of the African continent, the first human population only managed to leave Africa toward the Near and Middle East ~100,000 yr ago. They subsequently migrated to Europe and Oceania 45,000 yr ago (Lipson and Reich 2017), Asia 40,000 yr ago (Wang et al. 2021), and the Americas 15,000 yr ago (Reich et al. 2012).

Insights from Paleogenomics

Our knowledge of extinct species and past civilizations has been largely based on fossil and archaeological data. The possibility of recovering genetic information from ancient biological remains, such as bones and teeth, opened new

Received: October 13, 2023. Revised: January 31, 2024. Accepted: February 07, 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Open Access

perspectives for studying human evolutionary history, providing answers that could not be obtained by other areas of knowledge that study our origins.

Nearly 2 million years before the beginning of human expansion, *Homo erectus*, our direct ancestor, had already left Africa and successfully populated the Eurasian territory, reaching also Southeast Asia (Carotenuto et al. 2016). As a result of this expansion, other species derived from this hominid emerged outside of Africa, such as *Homo neanderthalensis* in Europe and the Denisovans in Asia. These species later may have been essential for the success of *H. sapiens* outside of Africa, as evidenced by several studies published in the last decade (Green et al. 2010; Meyer et al. 2012; Huerta-Sánchez et al. 2014; Racimo et al. 2017a, 2017b; Dannemann et al. 2017; Zeberg and Pääbo 2020; McArthur et al. 2021).

The information retrieved from paleogenomic studies has highlighted the impact of the biogeographic distribution of archaic species on the evolution and distribution of our species (Dannemann 2021). For example, with the publication of the first Neanderthal genome, it was demonstrated that ~2% of the genomes of present-day Europeans and Asians originate from an ancestral interbreeding event between *H. sapiens* and *H. neanderthalensis*, which occurred after our species left Africa (Green et al. 2010). The sequencing of another archaic genome would further complicate what we thought about our evolutionary history. Some regions of the Denisovan genome, an unknown species excavated in a cave in Siberia, were found in the genomes of Aboriginal Australians, Melanesians, and other populations from islands in Southeast Asia (Meyer et al. 2012).

The presence of introgressed DNA sequences from extinct archaic species reconstructs the biogeographic landscape of human populations as different introgressions occurred in different regions with different archaic species. Thus, this archaic genetic reservoir present in our species' genome influences both population structure studies and natural selection, in which introgressed regions likely facilitated local adaptation, especially in immunity-related and metabolic functions, helping the response to pathogens out of Africa (Racimo et al. 2017a, 2017b; Gouy et al. 2020). This opens up a fruitful new field for developing methods to detect introgression and measure the impact of these introgressed segments on the human genome (Huang et al. 2022; Zhang et al. 2023).

The Complex Dynamics of Human Diversity

During the remarkable human intercontinental migration, several factors contributed to the diversification of our species, such as founder effects, local adaptation, gene flow, and genetic introgression from archaic hominins (Jobling et al. 2013). In the last decade, large-scale genomic studies have become more accessible, leading to unprecedented advances in identifying human genetic variability and enabling a fine-scale understanding of genetic variation patterns within and among populations. Population genomics studies reveal a biogeographic

pattern in the distribution of human diversity, with a decrease in diversity as populations move away from the African continent. This pattern is explained by the long period of diversification of the African population within their continent, maintaining a larger effective population size and having more time for recombination events across genome regions, reducing the linkage disequilibrium between genetic variants, and allowing for new mutations (Bergström et al. 2020). Additionally, there were successive population founder effects caused by the migrations between continents, where only a portion of African diversity would have been carried to the Near and Middle East. In this region, a new population increase in the number of individuals and the generation of autochthonous diversity would have occurred. This pattern of variability reduction, followed by an increase in population size and the emergence of autochthonous diversification, would have been repeated in each migratory event to different continental regions, culminating in the Americas, the last continent to be populated by our species. In this context, founder effects and gene flow are consequences of the impact of varying degrees of geographical isolation on genetic differentiation among human populations, leading to the accumulation of local allele frequency differences (Jay et al. 2012). For instance, most geographically restricted genetic variants reflect novel mutations that occurred after, or shortly before, the diversification of present-day human groups, with >99% of alleles private to most non-African regions being the derived rather than the ancestral allele (Bergström et al. 2020). As expected, private alleles in Africa have a higher proportion of ancestral alleles, reflecting old variants lost outside Africa. High-frequency private African variants are found in archaic genomes like Neanderthals or Denisovans. Gene flow from archaic groups to present-day non-African ancestries likely occurred before their diversification in each continental region. Oceania is an exception, with ~35% of private variants shared with the Denisovan genome. About 20% of common variants outside Africa and absent inside Africa are shared with Neanderthals and Denisovans, while ~80% are likely derived from novel mutations, indicating that novel mutations have played a more significant role than archaic admixture in introducing new variants into present-day human populations (Bergström et al. 2020).

As a consequence of being the last step in the human journey, the Indigenous populations from America present the lowest genetic diversity among continental populations due to the limited time for diversification within the American continent and during the Beringia Standstill (Reich et al. 2012; Castro e Silva et al. 2022). The Beringia Standstill refers to the genetic isolation period of the proto-American population, comprising around 8,000 yr and restricted to a land bridge that connected Asia to America during the last glaciation (Hoffecker et al. 2014). This period was crucial for the emergence of Native American diversity. Although there are few studies available on the genomics of Native American populations, unique high-frequency alleles that can be traced back to an

origin in Beringia were identified (Niedbalski and Long 2022), with 99.9% of these variants representing derived alleles, encompassing more than 24,000 Native American-specific variants. This highlights the evolutionary significance of major bottleneck events, such as the Beringia standstill, in the evolution of group-specific alleles in continental populations. It also emphasizes the importance of delving deeper into the study of these specific continental variants to assess their impact on differential susceptibility to diseases.

As last happened in Beringia and America, during and after the events of human expansions, diversity also increased within each continent as these regions were populated and shaped by various evolutionary processes, many of which occurred concurrently over time within each continent. With the resolution achieved by genomic studies in the last decade, subcontinental populations in Africa, America, Asia, Europe, and Oceania are easily distinguishable in genomic studies, and, in general, isolation by distance is the pattern of subcontinental population differentiation (Peter et al. 2020). Humans do not mate randomly, and individuals living in the same region and speaking the same language tend to mate with each other more frequently than with individuals from distant regions or cultures. This leads to the differentiation between populations over time due to the effects of genetic drift. Genetically isolated ethnic groups are also more susceptible to assortative mating, a system where physically similar individuals tend to marry each other, and high rates of endogamy due to cultural or social factors (Robinson et al. 2017).

It is worth noting that humans are highly similar at the genetic level (Serre and Pääbo 2004), and the delineation of subcontinental or even continental population structure in genetic studies is commonly enforced by selecting genetic markers that emphasize differences among populations. This is because most of the diversity is shared among humans, and the majority of human diversity is found within these subcontinental groups. Furthermore, within this context, the concept of population is arbitrary, just as the populations chosen for comparison in different genetic analyses are arbitrary. In this text, I am using the term “population” to refer to biogeographic groups with a common evolutionary history, considering the existence of a diversity gradient that connects geographic groups, as well as individuals within those groups, due to their common recent African ancestry.

Overlooked Human Genetic Diversity

Despite the increasing availability of genomes for human population studies, there is still a significant imbalance among biogeographic groups or continents studied. Populations of European origin represent 86.3% of genomic studies, followed by 5.9% from East Asia, 1.1% from Africa, 0.8% from South Asia, and 0.08% from Hispanics and Latinos (admixed populations from Latin America; Fatumo et al. 2022). Therefore, it is not surprising that the biogeographical approach to European populations

has been more accurate. Given a reasonable number of genetic markers and populations studied, it is possible to predict the geographical origin of an individual in Europe with quite precision (Lao et al. 2008; Novembre et al. 2008; Elhaik et al. 2014). Other continental groups, such as Native Americans (or Indigenous Americans), are usually not presented in these statistics due to the limited number of studies available. This lack of information about continental and, consequently, subcontinental populations not only introduces bias into biomedical studies but also results in a critical lack of information essential for understanding human evolution at global and local levels. There are various explanations for the current scenario, such as most authors being based in the global North, which is predominantly of European descent. Additionally, the cost of these studies often serves as a significant barrier to conducting research in the global South, where these understudied populations are native. Another likely obstacle to sampling certain populations is the long history of violence against Native and admixed populations from America and Oceania, for instance. In this regard, sampling efforts should always encompass strong engagement with historically vulnerable populations, accompanied by a sincere respect for cultural perspectives on their origin and diversity. Following this, different countries have various ethical considerations that researchers must adhere to, with the primary focus of scientific work always being transparency about the study objectives and the interests of local communities.

It is noteworthy that even within the global North, there are neglected populations in genetic and evolutionary studies, such as ethnic minorities like the Roma. The Roma population is an instructive example of how the genetic population profile can change during an intense migration process. The early Roma ancestors experienced a population bottleneck as they began their Diaspora from Northwestern India. This bottleneck reduced their effective population size significantly. During their migration and settlement in Europe, they intermixed with Middle Eastern and European populations, which diluted their ancestral gene pool (Bianco et al. 2020). In this sense, while the multiple founder effects suffered by these populations have reduced their genetic diversity, the extensive gene flow that occurred during their migration process has counteracted the increase in mutational load with traceable ancestry-specific patterns in the runs-of-homozygosity segments (Font-Porterías et al. 2021).

Genomic studies involving underrepresented populations have revealed important aspects of their evolutionary histories. Some significant examples of the importance of research involving neglected populations come from studies with Indigenous populations from America and their descendants, such as admixed individuals from America. These studies have demonstrated the intricate patterns and processes of diversification found in this continental region.

America is a unique continent as it spans all latitudes, from north to south, presenting a wide range of ecoregions

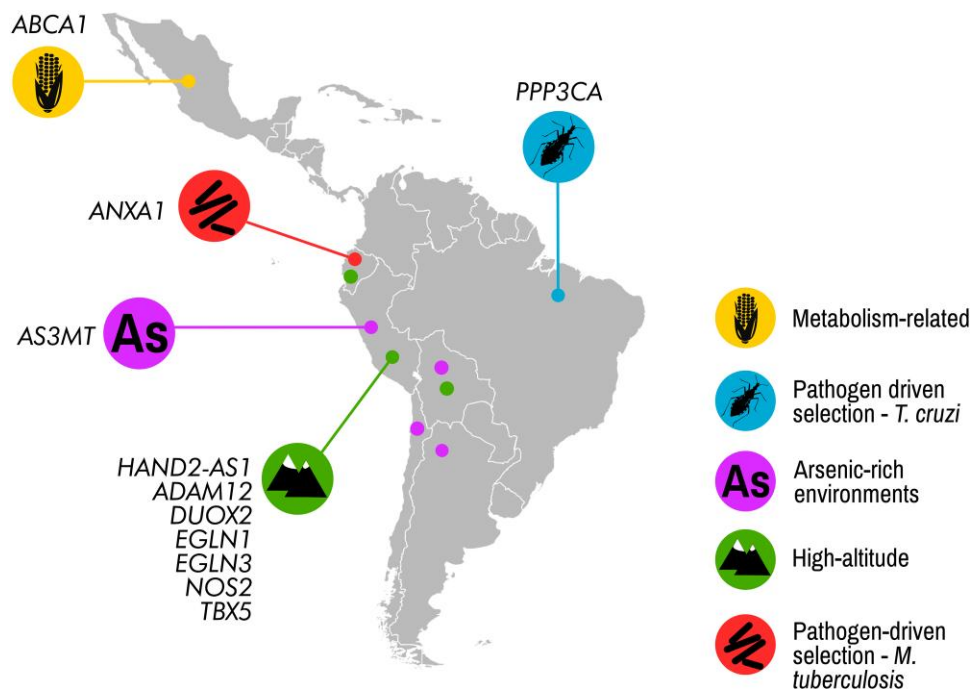


Fig. 1. Natural selection in Native Americans. Geographic distribution of locally adaptive traits in Indigenous populations from Latin America, labeled by the phenotype or selection pressure, and the genes under selection.

that posed significant challenges for the initial settlement of the continent. Indigenous American populations are mainly the result of a main migration from the Beringia region into the continent, which began ~15,000 yr ago. With the end of the last glaciation, the simultaneous flooding of Beringia, and the melting of polar ice caps in the northern hemisphere, the migration of these Beringian populations into the American continent became possible. It is estimated that the continent was rapidly populated within about 2000 yrs. This means the first Americans underwent various selective pressures in different environments, including high-altitude areas, deserts, and dense forests (Fig. 1).

For instance, Mexican indigenous populations show different signals of natural selection related to coping with local environmental and cultural conditions in the past, including those related to metabolic and infectious diseases, with an impact on the health of the current Mexican population (Ojeda-Granados et al. 2022). A parallel exists in Peruvian populations settled in various ecosystems across South America (Andean highlands, Pacific Coast, and Amazonian rainforest). In these populations, individuals from high-altitude regions exhibit signs of local adaptation in genes related to hypoxia, pigmentation, and cardiovascular function, while lowland Peruvian natives, whether coastal or Amazonian, show adaptations related to pathogen responses (Caro-Consuegra et al. 2022). Another example related to high-altitude adaptation and its consequences for the public health of present-day Andean populations (indigenous and admixed) is the high prevalence of preeclampsia in high-altitude Peruvian populations (Nieves-Colón et al. 2022). In these regions, 20% of women experience this complication during pregnancy, three times more than the global average. In this scenario, genes such as ADAM12, which in the past were crucial for survival at high altitudes, likely also lead to a secondary effect related to

higher risks during pregnancy in individuals with high Andean native ancestry. Present-day highlanders from Ecuador also presented a singular precontact adaptation in response to tuberculosis infection (Joseph et al. 2023). It is well known that there were precontact lineages of tuberculosis in America (Mackowiak et al. 2005; Bos et al. 2014); however, the incidence of this disease remains very high in this continent, representing one of the greatest public health issues in Latin American countries (Woodman et al. 2019). Examples of genetic adaptation to a pathogen in the past can help us better manage diseases caused by the same pathogen in current populations, as they help us understand the cellular and molecular mechanisms of combating or resisting infection, which can be population-specific. In the Brazilian Amazon, for instance, present-day local Indigenous populations are less susceptible to Chagas disease infection due to a natural selection event that occurred in the past driven by a trypanosomiasis outbreak (Couto-Silva et al. 2023). Although the Brazilian population has the lowest Native American ancestry in Latin America, around 7% (Ruiz-Linares et al. 2014), regions like the Amazon are an exception within the country, with a significant portion of the population being either Native American or having high Native American ancestry. Therefore, functional mutations in native populations can have a significant impact on the local admixed population. These studies illustrate the historical challenges these populations faced in adapting to different subcontinental ecoregions, and how the ancient adaptation to this new environment has an impact on present-day populations.

Another adaptive process to which indigenous populations in the Americas were subjected is gene-culture co-evolution, where the agriculturalist niche constructed by populations undergoing the Neolithic revolution alters the environment in such a way that it becomes a selective

factor (Laland et al. 2010). The most well-known example of this process is the adaptation to milk consumption in Northern Europe and some regions of Africa. Mexican populations stand out as the most common carriers of an autochthonous and functionally adaptive variant identified in America thus far (Acuña-Alonzo et al. 2010). The 230C variant within the ABCA1 gene underwent natural selection in Mexican populations during the shift from a hunter-gatherer to an agriculturalist lifestyle during the Neolithic revolution, directly linked to corn agriculture in the region (Hünemeier et al. 2012). Individuals harboring the mutation during this transitional period exhibited a thrifty genotype, reflecting an energy metabolism better suited to feast-famine cycles. Presently, native or admixed Mexicans carrying the 230C allele show an increased susceptibility to diseases, such as diabetes, obesity, high blood pressure, and elevated cholesterol. Considering the high percentage of indigenous ancestry in admixed Mexicans (56%; Ruiz-Linares et al. 2014), this allele represents the first case of precision medicine tailored to Natives and Latin American populations. The early diagnosis of the mutation can aid in the nutritional management of individuals, preventing diseases such as type 2 diabetes and obesity (Ávila-Arcos 2022).

Different studies focused on Amazonian populations have elucidated how the settlement of this region occurred and the significant genetic diversity found in each area, establishing connections between geography, language, genetics, and culture among South American Indigenous populations (Arias et al. 2018; Gneccchi-Ruscone et al. 2019; Barbieri et al. 2019; Castro e Silva et al. 2020, 2021, 2022). These studies demonstrated the great mobility of past native populations in South America, known for their present-day strong population structure and prevalence of small groups of individuals. Genetic data showed connections between Amazonian and Andean populations over time, indicating a lack of division and isolation between these two regions (Barbieri et al. 2019). Furthermore, the South American pre-contact population appears to have been formed by large groups, reaching more than 5 million individuals in the Amazon region alone. The impact of European colonization was devastating across the continent but especially severe on the Atlantic coast of South America, where 99% of the population was wiped out. Another interesting point from recent demographic studies is that in the Amazon region, this population decline began five centuries before European colonization and was worsened by the arrival of Europeans (Castro e Silva et al. 2022). The causes of this precontact population collapse are unclear; however, archaeological and paleobotanical data corroborate the genetic data (Bush et al. 2021; Schmidt et al. 2023).

In addition to being the last continent to be peopled, America was also the target of the greatest demographic expansion in human history, which led to the displacement of millions of Europeans and Africans to the American continent. This massive movement of people led to the encounter of continental populations isolated for millennia, such as Native Americans, with other continental populations.

This ongoing process caused a genetic shuffling of the diversity established after the out-of-Africa event, admixing populations of different cultural and geographical origins in America.

Due to the scarcity of genomic data from Indigenous and African populations, many studies utilize admixed populations from America as proxies to understand Native American and African genetic variability and their impact on the genetic profiles of contemporary admixed populations (Mychaleckyj et al. 2017; Mas-Sandoval et al. 2019; Mendoza-Revilla et al. 2022). Various methods have been developed to identify pre- and postcontact selection patterns in the Americas, while others have focused on assessing the influence of different ancestries in these admixed populations. In both cases, the results are crucial for understanding the development of metabolic diseases in Latin populations and the immunogenetic landscape of populations before and after contact with Europeans and Africans. Furthermore, studies investigating the demographic history of African populations that arrived in the Americas as enslaved individuals (Fortes-Lima et al. 2017; Ongaro et al. 2019; Gouveia et al. 2020) have added immense value not only by increasing the diversity of genomic databases but also by facilitating the understanding of two distinct aspects of human history: the earliest African migrations and the impact of the colonization of the American continent.

A promising strategy to enhance studies in human population genomics is commonly referred to as machine-learning (ML) methods. ML methods encompass a series of computer algorithm applications that refine their performance through experience (Mitechell 1997). These approaches have gained popularity over the last few years for their efficacy in identifying population structure, making demographic inferences, and discerning signals of natural selection (Huang et al. 2023). Deep learning, a sophisticated class of machine-learning algorithms, has exhibited state-of-the-art performance across numerous applications involving large-scale population data (Korfmann et al. 2023). However, given their learning-by-experience nature, machine-learning methods require extensive datasets for accurate performance. Consequently, they are generally unsuitable for studies involving sparsely sampled populations or small population groups. A potential alternative in such instances is leveraging machine-learning architectures to construct artificial genomes (Yelmen et al. 2021). These synthetic genomes have the potential to provide crucial population-specific information not affiliated with an actual group but rather with a cohort of artificial individuals subjected to specific demographic events characterizing them. In this approach, access to information detailing the evolutionary history of these populations, along with high-coverage reference genomes, is essential. This information allows specific individuals and populations spanning contemporary and archaic periods to be artificially constructed. Currently, the use of artificial genomes is in its early stage, with most methods predominantly focused on their construction for medical studies where financial or ethical considerations constrain actual data usage. Nevertheless,

considering the swift advancement of these methodologies and the lack of data for several human populations, it emerges as a prospective key approach for studying present-day populations, comprehending their historical trajectories, and modeling their evolutionary futures.

Conclusion

The study of human genetic diversification through biogeographic lenses has deepened our comprehension of our species' history, unveiling the intricate migration patterns, adaptation, and genetic admixture shaping human diversity. However, current studies are still immensely limited by the overrepresentation of populations living in the global North, which masks the real contribution of other populations, at a local and global level, to the formation and maintenance of the diversity of our species. In this sense, addressing the underrepresentation of diverse populations in genomic studies through sampling or artificial genomes is a matter of scientific equity and a critical step toward a more comprehensive understanding of human genetic diversity and evolution.

Acknowledgments

I am grateful to Prof. David Comas for his fruitful suggestions and comments and to Dr Virginia Ramallo for the graphical assistance.

Funding

The study was supported by FAPESP 21/06860-8 (Brazil) and RYC2020/030381-I MCIN/AEI (Spain).

Data availability

The manuscript has no new data or even new analysis, as it is a perspective piece.

References

- Acuña-Alonso V, Flores-Dorantes T, Kruit JK, Villarreal-Molina T, Arellano-Campos O, Hünemeier T, Moreno-Estrada A, Ortiz-López MG, Villamil-Ramírez H, León-Mimila P, *et al.* A functional ABCA1 gene variant is associated with low HDL-cholesterol levels and shows evidence of positive selection in Native Americans. *Hum Mol Genet.* 2010;**19**(14):2877–2885. <https://doi.org/10.1093/hmg/ddq173>.
- Arias L, Schröder R, Hübner A, Barreto G, Stoneking M, Pakendorf B. Cultural innovations influence patterns of genetic diversity in Northwestern Amazonia. *Mol Biol Evol.* 2018;**35**(11):2719–2735. <https://doi.org/10.1093/molbev/msy169>.
- Ávila-Arcos MC. The gene variant that helped put Latinxs in the 1000 genomes project. *Nat Rev Genet.* 2022;**23**(12):712–713. <https://doi.org/10.1038/s41576-022-00538-w>.
- Barbieri C, Barquera R, Arias L, Sandoval JR, Acosta O, Zurita C, Aguilar-Campos A, Tito-Álvarez AM, Serrano-Osuna R, Gray RD, *et al.* The current genomic landscape of Western South America: Andes, Amazonia, and Pacific Coast. *Mol Biol Evol.* 2019;**36**(12):2698–2713. <https://doi.org/10.1093/molbev/msz174>.
- Bergström A, McCarthy SA, Hui R, Almarri MA, Ayub Q, Danecek P, Chen Y, Felkel S, Hallast P, Kamm J, *et al.* Insights into human genetic variation and population history from 929 diverse genomes. *Science.* 2020;**367**(6484):eaay5012. <https://doi.org/10.1126/science.aay5012>.
- Bianco E, Laval G, Font-Porterias N, García-Fernández C, Dobon B, Sabido-Vera R, Stefanovska ES, Kučinskis V, Makukh H, Pamjav H, *et al.* Recent common origin, reduced population size, and marked admixture have shaped European roma genomes. *Mol Biol Evol.* 2020;**37**(11):3175–3187. <https://doi.org/10.1093/molbev/msaa156>.
- Bos KI, Harkins KM, Herbig A, Coscolla M, Weber N, Comas I, Forrest SA, Bryant JM, Harris SR, Schuenemann VJ, *et al.* Pre-Columbian mycobacterial genomes reveal seals as a source of new world human tuberculosis. *Nature.* 2014;**514**(7523):494–497. <https://doi.org/10.1038/nature13591>.
- Bush MB, Nascimento MN, Akesson CM, Cárdenes-Sandí GM, Maezumi SY, Behling H, Correa-Metrio A, Church W, Huisman SN, Kelly T, *et al.* Widespread reforestation before European influence on Amazonia. *Science.* 2021;**372**(6541):484–487. <https://doi.org/10.1126/science.abf3870>.
- Caro-Consuegra R, Nieves-Colón MA, Rawls E, Rubin-de-Celis V, Lizárraga B, Vidaurre T, Sandoval K, Fejerman L, Stone AC, Moreno-Estrada A, *et al.* Uncovering signals of positive selection in Peruvian populations from three ecological regions. *Mol Biol Evol.* 2022;**39**(8):msac158. <https://doi.org/10.1093/molbev/msac158>.
- Carotenuto F, Tsikaridze N, Rook L, Lordkipanidze D, Longo L, Condeci S, Raia P. Venturing out safely: the biogeography of *Homo erectus* dispersal out of Africa. *J Hum Evol.* 2016;**95**:1–12. <https://doi.org/10.1016/j.jhevol.2016.02.005>.
- Castro e Silva MA, Ferraz T, Bortolini MC, Comas D, Hünemeier T. Deep genetic affinity between coastal Pacific and Amazonian natives evidenced by Australasian ancestry. *Proc Natl Acad Sci U S A.* 2021;**118**(14):e2025739118. <https://doi.org/10.1073/pnas.2025739118>.
- Castro e Silva MA, Ferraz T, Hünemeier T. A genomic perspective on South American human history. *Genet Mol Biol.* 2022;**45**(3 suppl 1):e20220078. <https://doi.org/10.1590/1678-4685-gmb-2022-0078>.
- Castro e Silva MA, Nunes K, Lemes RB, Mas-Sandoval A, Amorim CEG, Krieger JE, Mill JG, Salzano FM, Bortolini MC, da Costa Pereira A, *et al.* Genomic insight into the origins and dispersal of the Brazilian coastal natives. *Proc Natl Acad Sci U S A.* 2020;**117**(5):2372–2377. <https://doi.org/10.1073/pnas.1909075117>.
- Couto-Silva CM, Venturini G, Castro e Silva MA, Pereira LV, Comas D, Pereira A, Hünemeier T. Indigenous people from Amazon show genetic signatures of pathogen-driven selection. *Sci Adv.* 2023;**9**(10):eabo0234. <https://doi.org/10.1126/sciadv.abo0234>.
- Dannemann M. The population-specific impact of neandertal introgression on human disease. *Genome Biol Evol.* 2021;**13**(1):evaa250. <https://doi.org/10.1093/gbe/evaa250>.
- Dannemann M, Prüfer K, Kelso J. Functional implications of Neandertal introgression in modern humans. *Genome Biol.* 2017;**18**(1):61. <https://doi.org/10.1186/s13059-017-1181-7>.
- Durvasula A, Sankararaman S. Recovering signals of ghost archaic introgression in African populations. *Sci Adv.* 2020;**6**(7):eaax5097. <https://doi.org/10.1126/sciadv.aax5097>.
- Elhaik E, Tatarinova T, Chebotarev D, Piras IS, Calò CM, De Montis A, Atzori M, Marini M, Tofanelli S, Frnacalacci P, *et al.* Geographic population structure analysis of worldwide human populations infers their biogeographical origins. *Nat Commun.* 2014;**5**(1):3513. <https://doi.org/10.1038/ncomms4513>.
- Fatumo S, Chikowore T, Choudhury A, Ayub M, Martin AR, Kuchenbaecker K. A roadmap to increase diversity in genomic studies. *Nat Med.* 2022;**28**(2):243–250. <https://doi.org/10.1038/s41591-021-01672-4>.
- Font-Porterias N, Caro-Consuegra R, Lucas-Sánchez M, Lopez M, Giménez A, Carballo-Mesa A, Bosch E, Calafell F, Quintana-Murci L, Comas D. The counteracting effects of demography on functional genomic variation: the roma paradigm. *Mol Biol Evol.* 2021;**38**(7):2804–2817. <https://doi.org/10.1093/molbev/msab070>.

- Fortes-Lima C, Gessain A, Ruiz-Linares A, Bortolini MC, Migot-Nabias F, Bellis G, Moreno-Mayar JV, Restrepo BN, Rojas W, Avendaño-Tamayo E, *et al.* Genome-wide ancestry and demographic history of African-descendant maroon communities from French Guiana and Suriname. *Am J Hum Genet.* 2017;**101**(5):725–736. <https://doi.org/10.1016/j.ajhg.2017.09.021>.
- Gneccchi-Ruscione GA, Sarno S, De Fanti S, Gianvincenzo L, Giuliani C, Boattini A, Bortolini E, Di Corcia T, Sanchez Mellado C, Dávila Francia TJ, *et al.* Dissecting the pre-Columbian genomic ancestry of native Americans along the Andes–Amazonia divide. *Mol Biol Evol.* 2019;**36**(6):1254–1269. <https://doi.org/10.1093/molbev/msz006>.
- Gouveia MH, Borda V, Leal TP, Moreira RG, Bergen AW, Kehdy FSG, Alvim I, Aquino MM, Araujo GS, Araujo NM, *et al.* Origins, admixture dynamics, and homogenization of the African gene pool in the Americas. *Mol Biol Evol.* 2020;**37**(6):1647–1656. <https://doi.org/10.1093/molbev/msaa033>.
- Gouy A, Excoffier L. Polygenic patterns of adaptive introgression in modern humans are mainly shaped by response to pathogens. *Mol Biol Evol.* 2020;**37**(5):1420–1433. <https://doi.org/10.1093/molbev/msz306>.
- Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH, *et al.* A draft sequence of the neandertal genome. *Science.* 2010;**328**(5979):710–722. <https://doi.org/10.1126/science.1188021>.
- Henn BD, Cavalli-Sforza LL, Feldman MW. The great human expansion. *Proc Natl Acad Sci U S A.* 2012;**109**(44):17758–17764. <https://doi.org/10.1073/pnas.1212380109>.
- Hoffecker JF, Elias SA, O'Rourke DH. Out of Beringia? A shrub tundra refugium on the Bering land bridge may have played a pivotal role in the peopling of the Americas. *Science.* 2014;**6174**(6174):979–980. <https://doi.org/10.1126/science.1250768>.
- Huang X, Kruis P, Kuhlwlilm M. sstar: a python package for detecting archaic introgression from population genetic data with S*. *Mol Biol Evol.* 2022;**39**(11):msac212. <https://doi.org/10.1093/molbev/msac212>.
- Huang X, Rymbekova A, Dolgova O, Lao O, Kuhlwlilm M. Harnessing deep learning for population genetic inference. *Nat Rev Genet Jan.* 2023;**25**(1):61–78. <https://doi.org/10.1038/s41576-023-00636-3>.
- Hublin JJ, Ben-Ncer A, Bailey S, Freidline SE, Neubauer S, Skinner MM, Bergmann I, Le Cabec A, Benazzi S, Harvati K, *et al.* New fossils from Jebel Irhoud, Morocco and the pan-African origin of *Homo sapiens*. *Nature.* 2017;**546**(7657):289–292. <https://doi.org/10.1038/nature22336>.
- Huerta-Sánchez E, Jin X, Asan, Bianba Z, Peter BM, Vinckenbosch N, Liang Y, Yi X, He M, Somel M, *et al.* Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature.* 2014;**512**(7513):194–197. <https://doi.org/10.1038/nature13408>.
- Hünemeier T, Amorim CE, Azevedo S, Contini V, Acuña-Alonzo V, Rothhammer F, Dugoujon JM, Mazières S, Barrantes R, Villarreal-Molina MT, *et al.* Evolutionary responses to a constructed niche: ancient Mesoamericans as a model of gene-culture coevolution. *PLoS One.* 2012;**7**(6):e38862. <https://doi.org/10.1371/journal.pone.0038862>.
- Jay F, Sjödin P, Jakobsson M, Blum MG. Anisotropic isolation by distance: the main orientations of human genetic differentiation. *Mol Biol Evol.* 2012;**30**(3):513–525. <https://doi.org/10.1093/molbev/mss259>.
- Jobling M, Hollox E, Hurler M, Kivisild T, Tyler-Smith C. *Human evolutionary genetics*. 2nd ed. New York: Garland Science; 2013.
- Joseph SK, Migliore NR, Olivieri A, Torroni A, Owings AC, DeGiorgio M, Ordóñez WG, Aguilú JJO, González-Andrade F, Achilli A, *et al.* Genomic evidence for adaptation to tuberculosis in the Andes before European contact. *iScience.* 2023;**26**(2):106034. <https://doi.org/10.1016/j.isci.2023.106034>.
- Korfmann K, Gaggiotti OE, Fumagalli M. Deep learning in population genetics. *Gen Biol Evol.* 2023;**15**(2):evad008. <https://doi.org/10.1093/gbe/evad008>.
- Laland K, Odling-Smee J, Myles S. How culture shaped the human genome: bringing genetics and the human sciences together. *Nat Rev Genet.* 2010;**11**(2):137–148. <https://doi.org/10.1038/nrg2734>.
- Lao O, Lu TT, Nothnagel M, Junge O, Freitag-Wolf S, Caliebe A, Balasckova M, Bertranpetit J, Bindoff LA, Comas D, *et al.* Correlation between genetic and geographic structure in Europe. *Curr Biol.* 2008;**18**(16):1241–1248. <https://doi.org/10.1016/j.cub.2008.07.049>.
- Lipson M, Reich D. A working model of the deep relationships of diverse modern human genetic lineages outside of Africa. *Mol Biol Evol.* 2017;**34**(4):889–902. <https://doi.org/10.1093/molbev/msw293>.
- Lorente-Galdos B, Lao O, Serra-Vidal G, Santpere G, Kuderna LFK, Aruana LR, Fadhlouli-Zid K, Pimenoff VN, Soodyall VN, Zalloua P, *et al.* Whole-genome sequence analysis of a Pan African set of samples reveals archaic gene flow from an extinct basal population of modern humans into sub-Saharan populations. *Genome Biol.* 2019;**20**(1):77. <https://doi.org/10.1186/s13059-019-1684-5>.
- Mackowiak PA, Blos VT, Aguilar M, Buikstra JE. On the origin of American tuberculosis. *Clin Infect Dis.* 2005;**41**(4):515–518. <https://doi.org/10.1086/432013>.
- Mallick S, Li H, Lipson M, Mathieson I, Gymrek M, Racimo F, Zhao M, Chennagiri N, Nordenfelt S, Tandon A, *et al.* The Simons genome diversity project: 300 genomes from 142 diverse populations. *Nature.* 2016;**538**(7624):201–206. <https://doi.org/10.1038/nature18964>.
- Mas-Sandoval A, Arauna LR, Gouveia MH, Barreto ML, Horta BL, Lima-Costa MF, Pereira AC, Salzano FM, Hünemeier T, Tarazona-Santos E, *et al.* Reconstructed lost native American populations from Eastern Brazil are shaped by differential Jê/Tupi ancestry. *Gen Biol Evol.* 2019;**11**(9):2593–2604. <https://doi.org/10.1093/gbe/evz161>.
- McArthur E, Rinker DC, Capra JA. Quantifying the contribution of Neanderthal introgression to the heritability of complex traits. *Nat Commun.* 2021;**12**(1):448. <https://doi.org/10.1038/s41467-021-24582-y>.
- Mendoza-Revilla J, Chacón-Duque JC, Fuentes-Guajardo M, Ormond L, Wang K, Hurtado M, Villegas V, Acuña-Alonzo V, Jaramillo C, Arias W, *et al.* Disentangling signatures of selection before and after European colonization in Latin Americans. *Mol Biol Evol.* 2022;**39**(4):msac076. <https://doi.org/10.1093/molbev/msac076>.
- Meyer M, Kircher M, Gansauge M-T, Li H, Racimo F, Mallick S, Schraiber JG, Jay F, Prüfer K, De Filippo C, *et al.* A high-coverage genome sequence from an archaic denisovan individual. *Science.* 2012;**338**(6104):222–226. <https://doi.org/10.1126/science.1224344>.
- Mitchell T. *Machine learning*. New York: McGraw Hill; 1997, 414p.
- Mychaleckyj JC, Havt A, Nayak U, Pinkerton R, Farber, Concannon P, Lima AA, Guerrant RL. Genome-wide analysis in Brazilians reveals highly differentiated native American genome regions. *Mol Biol Evol.* 2017;**34**(3):559–574. <https://doi.org/10.1093/molbev/msw249>.
- Niedbalski SD, Long JC. Novel alleles gained during the Beringian isolation period. *Sci Rep.* 2022;**12**(1):4289. <https://doi.org/10.1038/s41598-022-08212-1>.
- Nieves-Colón MA, Badillo Rivera KM, Sandoval K, Dávalos VV, Lencinas LEE, Mendoza-Revilla J, Adhikari K, González-Buenfil R, Chen JW, Zhang ET, *et al.* Clotting factor genes are associated with preeclampsia in high-altitude pregnant women in the Peruvian Andes. *Am J Hum Genet.* 2022;**109**(6):1117–1139. <https://doi.org/10.1016/j.ajhg.2022.04.014>.
- Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A, Indap A, King KS, Bergmann S, Nelson MR, *et al.* Genes mirror geography within Europe. *Nature.* 2008;**456**(7218):98–101. <https://doi.org/10.1038/nature07331>.
- Ojeda-Granados C, Abondio P, Setti A, Sarno S, Gneccchi-Ruscione GA, González-Orozco E, De Fanti S, Jiménez-Kaufmann A, Rangel-Villalobos H, Moreno-Estrada A, *et al.* Dietary, cultural, and pathogens-related selective pressures shaped differential adaptive evolution among native Mexican populations. *Mol Biol Evol.* 2022;**39**(1):msab290. <https://doi.org/10.1093/molbev/msab290>.
- Ongaro L, Scliar MO, Flores R, Raveane A, Marnetto D, Sarno S, Gneccchi-Ruscione GA, Alarcón-Riquelme ME, Patin E,

- Wangkumhang P, *et al.* The genomic impact of European colonization of the Americas. *Curr Biol.* 2019;**29**(23):3974–3986.e4. <https://doi.org/10.1016/j.cub.2019.09.076>.
- Peter BM, Petkova D, Novembre J. Genetic landscapes reveal how human genetic diversity aligns with geography. *Mol Biol Evol.* 2020;**37**(4):943–951. <https://doi.org/10.1093/molbev/msz280>.
- Racimo F, Gokhman D, Fumagalli M, Ko A, Hansen T, Moltke I, Albrechtsen A, Carmel L, Huerta-Sánchez E, Nielsen R. Archaic adaptive introgression in TBX15/WARS2. *Mol Biol Evol.* 2017a;**34**(3):509–524. <https://doi.org/10.1093/molbev/msw283>.
- Racimo F, Marnetto D, Huerta-Sánchez E. Signatures of archaic adaptive introgression in present-day human populations. *Mol Biol Evol.* 2017b;**34**(2):296–317. <https://doi.org/10.1093/molbev/msw216>.
- Ragsdale AP, Weaver TD, Atkinson EC, Hoal EG, Möller M, Henn BM, Gravel S. A weakly structured stem for human origins in Africa. *Nature.* 2023;**617**(7962):755–763. <https://doi.org/10.1038/s41586-023-06055-y>.
- Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, Ray N, Parra MV, Rojas W, Duque C, Mesa N, *et al.* Reconstructing Native American population history. *Nature.* 2012;**488**(7411):370–374. <https://doi.org/10.1038/nature11258>.
- Robinson M, Kleinman A, Graff M, Vinkhuyzen AAE, Couper D, Miller MB, Peyrot WJ, Abdellaoui A, Zietsch BP, Nolte IM, *et al.* Genetic evidence of assortative mating in humans. *Nat Hum Behav.* 2017;**1**(1):0016. <https://doi.org/10.1038/s41562-016-0016>.
- Ruiz-Linares A, Adhikari K, Acuña-Alonzo V, Quinto-Sanchez M, Jaramillo C, Arias W, Fuentes M, Pizarro M, Everardo P, de Avila F, *et al.* Admixture in Latin America: geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals. *PLoS Genet.* 2014;**10**(9):e1004572. <https://doi.org/10.1371/journal.pgen.1004572>.
- Russo CAM, Eyre-Walker A, Katz LA, Gaut BA. Forty years of inferential methods in the Journals of the Society for Molecular Biology and Evolution. *Mol Biol Evol.* 2024;**41**(1):msad264. <https://doi.org/10.1093/molbev/msad264>.
- Schmidt MJ, Goldberg SL, Heckenberger M, Fausto C, Franchetto B, Waltling J, Lima H, Moraes B, Ddorshow WB, Toney J, *et al.* Intentional creation of carbon-rich dark earth soils in the Amazon. *Sci Adv.* 2023;**9**(38):eadh8499. <https://doi.org/10.1126/sciadv.adh8499>.
- Serre D, Pääbo S. Evidence for gradients of human genetic diversity within and among continents. *Genome Res.* 2004;**14**(9):1679–1685. <https://doi.org/10.1101/gr.2529604>.
- Wang CC, Yeh HY, Popov AN, Zhang HQ, Matsumura H, Sirak K, Cheronet O, Kovalev A, Rohland N, Kim AM, *et al.* Genomic insights into the formation of human populations in East Asia. *Nature.* 2021;**591**(7850):413–419. <https://doi.org/10.1038/s41586-021-03336-2>.
- Woodman M, Haeusler IL, Grandjean L. Tuberculosis genetic epidemiology: a Latin American perspective. *Genes (Basel).* 2019;**10**(1):53. <https://doi.org/10.3390/genes10010053>.
- Yelmen B, Decelle A, Ongaro L, Marnetto D, Tallec C, Montinaro F, Furtlehner C, Pagani L, Jay F. Creating artificial human genomes using generative neural networks. *PLoS Genet.* 2021;**17**(2):e1009303. <https://doi.org/10.1371/journal.pgen.1009303>.
- Zeberg H, Pääbo S. The major genetic risk factor for severe COVID-19 is inherited from neanderthals. *Nature.* 2020;**587**(7835):610–612. <https://doi.org/10.1038/s41586-020-2818-3>.
- Zhang X, Kim B, Singh A, Sankararaman S, Durvasula A, Lohmueller KE. MaLAdapt reveals novel targets of adaptive introgression from neanderthals and denisovans in worldwide human populations. *Mol Biol Evol.* 2023;**40**(1):msad001. <https://doi.org/10.1093/molbev/msad001>.