The Genomic Impact of European Colonization of the Americas

Article

Current Biology

The Genomic Impact of European Colonization of the Americas

Highlights

- European and African genomic signature in the Americas shows high complexity
- Sex-biased gene flow occurred between European and American mixing groups
- Admixture is geographically and chronologically correlated with historical records
- Source-specific demographic histories reveal the huge impact of recent admixture

Authors

Linda Ongaro, Marilia O. Scliar, Rodrigo Flores, ..., Mait Metspalu, Luca Pagani, Francesco Montinaro

Correspondence

linda.ongaro@ut.ee (L.O.), francesco.montinaro@gmail.com (F.M.)

In Brief

The complexity of the admixture dynamics that shaped American populations is unveiled by Ongaro et al., where genetic data for more than 12,000 individuals from the continents are investigated. This study evaluates the dramatic impact of events after the colonial era, revealing a spatial and temporal heterogeneity and mirroring historical records.

Ongaro et al., 2019, Current Biology *29*, 3974–3986
December 2, 2019
https://doi.org/10.1016/j.cub.2019.09.076

CellPress

CellPress

Current Biology
Article

The Genomic Impact of European Colonization of the Americas

Linda Ongaro,[1,2,*] Marilia O. Scliar,[3,4] Rodrigo Flores,[1] Alessandro Raveane,[5] Davide Marnetto,[1] Stefania Sarno,[6] Guido A. Gnecchi-Ruscone,[6,7] Marta E. Alarcón-Riquelme,[8] Etienne Patin,[9] Pongsakorn Wangkumhang,[10] Garrett Hellenthal,[10] Miguel Gonzalez-Santos,[11] Roy J. King,[12] Anastasia Kouvatsi,[13] Oleg Balanovsky,[14,15,16] Elena Balanovska,[14,15,16] Lubov Atramentova,[17] Shahlo Turdikulova,[18] Sarabjit Mastana,[19] Damir Marjanovic,[20,21] Lejla Mulahasanovic,[22] Andreja Leskovac,[23] Maria F. Lima-Costa,[24] Alexandre C. Pereira,[25]

(Author list continued on next page)

[1]Estonian Biocentre, Institute of Genomics, Riia 23, Tartu 51010, Estonia
[2]Department of Evolutionary Biology, Institute of Molecular and Cell Biology, Riia 23, Tartu 51010, Estonia
[3]Human Genome and Stem Cell Research Center, Biosciences Institute, University of São Paulo, São Paulo, SP 05508-090, Brazil
[4]Departamento de Genética, Ecologia e Evolução, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, MG 31270-901, Brazil
[5]Department of Biology and Biotechnology "L. Spallanzani", University of Pavia, Pavia 27100, Italy
[6]Department of Biological, Geological and Environmental Sciences, University of Bologna, Bologna 40100, Italy
[7]Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena 07745, Germany
[8]GENYO, Centre for Genomics and Oncological Research, Pfizer/University of Granada/Andalusian Regional Government, Granada 18016, Spain
[9]Human Evolutionary Genetics Unit, Pasteur Institute, UMR2000, CNRS, Paris 75015, France
[10]Department of Genetics, Evolution and Environment and UCL Genetics Institute, University College London, London WC1E 6BT, UK
[11]Department of Zoology, University of Oxford, Oxford OX1 3SZ, UK
[12]Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, CA 94305-5101, USA
[13]Department of Genetics, Development and Molecular Biology, School of Biology, Aristotle University of Thessaloniki, Thessaloniki 54124, Greece
[14]Vavilov Institute of General Genetics, Ulitsa Gubkina, 3, Moscow 117971, Russia
[15]Research Centre for Medical Genetics, Moskvorech'ye Ulitsa, 1, Moscow 115478, Russia
[16]Biobank of North Eurasia, Kotlyakovskaya Ulitsa, 3 строение 12, Moscow 115201, Russia
[17]Department of Genetics and Cytology, V.N. Karazin Kharkiv National University, Kharkiv 61022, Ukraine
[18]Laboratory of Genomics, Institute of Bioorganic Chemistry, Academy of Sciences Republic of Uzbekistan, Tashkent 100047, Uzbekistan
[19]School of Sport, Exercise and Health Sciences, Loughborough University, Loughborough LE11 3TU, UK
[20]Department of Genetics and Bioengineering, Faculty of Engineering and Information Technologies, International Burch University, Sarajevo 71000, Bosnia and Herzegovina
[21]Institute for Anthropological Researches, Zagreb, Croatia

(Affiliations continued on next page)

SUMMARY

The human genetic diversity of the Americas has been affected by several events of gene flow that have continued since the colonial era and the Atlantic slave trade. Moreover, multiple waves of migration followed by local admixture occurred in the last two centuries, the impact of which has been largely unexplored. Here, we compiled a genome-wide dataset of ~12,000 individuals from twelve American countries and ~6,000 individuals from worldwide populations and applied haplotype-based methods to investigate how historical movements from outside the New World affected (1) the genetic structure, (2) the admixture profile, (3) the demographic history, and (4) sex-biased gene-flow dynamics of the Americas. We revealed a high degree of complexity underlying the genetic contribution of European and African populations in North and South America, from both geographic and temporal perspectives, identifying previously unreported sources related to Italy, the Middle East, and to specific regions of Africa.

INTRODUCTION

North and South America were the last continental areas to be colonized by humans. Their peopling was a complex process, which began at least 15 thousand years ago (kya) [1–6]. Nowadays, a substantial proportion of individuals living in the western hemisphere is the result of more recent episodes of admixture, following extensive migrations during and after the European colonial era [7].

Indeed, soon after the European discovery of the continents in 1492, western European powers began to explore and settle the double continent. This process heavily impacted native populations, which were decimated by wars and new pathogens. The Atlantic slave trade, which occurred between the 16th and 19th century and was started by European merchants, added further complexity to the continental genetic landscape.

Historical records attest a general imbalance in the number of incoming males and females especially during the early phase of European colonization. For instance, the first Iberian immigrants

Mauricio L. Barreto,[26,27] Bernardo L. Horta,[28] Nédio Mabunda,[29] Celia A. May,[30] Andrés Moreno-Estrada,[31] Alessandro Achilli,[5] Anna Olivieri,[5] Ornella Semino,[5] Kristiina Tambets,[1] Toomas Kivisild,[32] Donata Luiselli,[33] Antonio Torroni,[5] Cristian Capelli,[11] Eduardo Tarazona-Santos,[4] Mait Metspalu,[1] Luca Pagani,[1,34] and Francesco Montinaro[1,11,35,*]

[22]CeGaT GmbH, Tübingen, Praxis für Humangenetik, Tübingen 72076, Germany
[23]Vinca Institute of Nuclear Sciences, University of Belgrade, M. Petrovica Alasa 12-14, Belgrade 11001, Serbia
[24]Instituto de Pesquisa Rene Rachou, Fundação Oswaldo Cruz, Belo Horizonte, MG 30190-002, Brazil
[25]Instituto do Coração, Universidade de São Paulo, São Paulo, SP 05403-900, Brazil
[26]Instituto de Saúde Coletiva, Universidade Federal da Bahia, Salvador, BA 0110-040, Brazil
[27]Center of Data and Knowledge Integration for Health (CIDACS), Fundação Oswaldo Cruz (FIOCRUZ), Salvador, BA 41745-715, Brazil
[28]Programa de Pós-Graduação em Epidemiologia, Universidade Federal de Pelotas, 464, Pelotas, RS 96001-970, Brazil
[29]Instituto Nacional de Saúde, Distrito de Marracuene, Estrada Nacional N 1, Província de Maputo, Maputo 1120, Mozambique
[30]Department of Genetics & Genome Biology, University of Leicester, Leicester LE1 7RH, UK
[31]National Laboratory of Genomics for Biodiversity (LANGEBIO), CINVESTAV, Irapuato, Guanajuato 36821, Mexico
[32]Department of Human Genetics, KU Leuven, Herestraat 49 - box 602, Leuven 3000, Belgium
[33]Department of Cultural Heritage, University of Bologna, Ravenna Campus, Ravenna 48100, Italy
[34]Department of Biology, University of Padua, Via Ugo Bassi 58B, Padua 35100, Italy
[35]Lead Contact
*Correspondence: linda.ongaro@ut.ee (L.O.), francesco.montinaro@gmail.com (F.M.)
https://doi.org/10.1016/j.cub.2019.09.076

were mostly (>80%) males [8], and the proportion of females, initially 5%–6%, began to increase only in the following decades [7].

Additional migratory waves, mainly from the Southern and Eastern regions of Europe, have added further demographic variability since the end of the 19th century. In fact, it has been estimated that more than 32 million individuals reached the United States at the end of the 1800s and the beginning of the 1900s, and similar estimates are available for other American countries. For example, more than 6 million people moved to Argentina, and more than 5 million people moved to Brazil in the same period [9].

Given their historical and epidemiological implications, migrations to the Americas have been the subject of several genetic studies [10–15]. Most of these have exploited local ancestry (LA) inference algorithms, in which individual genomes are deconvoluted into fragments ultimately tracing their ancestry in populations from different macro-geographic areas. LA approaches have so far provided multiple insights into the composition of several recently admixed populations [16, 17]. However, when closely related populations are involved in the admixture of a specific target group, this approach shows some difficulties in discriminating among sources, leading to spurious or incomplete results.

Several surveys [10, 13, 14, 18] present a continental-wide analysis of the origin and dynamics of the African and European diaspora into the Americas. For instance, Gouveia et al. [18] have recently performed a detailed analysis of the African regional ancestry and its dynamics in several populations from North and South America and the Caribbean. However, a more comprehensive and systematic investigation considering multiple ancestries across the two continents is still missing [10–13].

The recently increased availability of genome-wide data offers now the chance to capture the complexity of historical and demographic events that affected the recent history of the Americas by studying the recent admixture profile of many modern populations.

With this in mind, we have first assembled a genome-wide dataset of 17,722 individuals, including ~12,000 from North, Central, and South America and ~6,000 from Africa, Europe, Asia, and Oceania (Figure S1A; Data S1A and S1B) and then we harnessed haplotype-based and allele frequency methods to (1) reconstruct the fine-scale ancestry composition, (2) evaluate the time of admixture, (3) assess the extent and magnitude of sex-biased gene-flow dynamics, and (4) explore the demographic evolution of different continental ancestries after the admixture.

RESULTS

Clustering of the Donor Individuals

To minimize the impact of within-source (“donors”) genetic heterogeneity in the ancestry characterization process, we grouped the assembled 6,115 individuals (Figure 1; Data S1A and S1B) from 239 population-label donors (from which American individuals are subsequently allowed to copy fragments of genome; see STAR Methods) into 89 genetically homogeneous clusters (Figure S2A; Data S2A) on the basis of haplotype similarities using CHROMOPAINTER and fineSTRUCTURE [19].

African individuals were classified into 40 clusters, with a clear split between sub-Saharans and North and Eastern Africans (Figure S2A). We identified 7 North African clusters, including individuals from Morocco, Egypt, and the Levant. Individuals from Western and South Western Africa (sub-Saharan), i.e., from the major slave-trading regions, are grouped in 14 clusters. East African individuals are distributed across 10 clusters, while the cluster “SouthEastAfrica” includes individuals from Mozambique and Zimbabwe (together with 10 Bantu South Africans). European individuals are differentiated into 36 clusters, mirroring the geographic location of the analyzed samples. We identified two Iberian clusters: “Spain”, which includes mostly Spanish samples (60 Spanish [92.3%], 4 French, and 1 Corsican; Data S1), and “SpainPortugal”, which contains 25 Spanish and all the Portuguese samples (10 individuals). Italians are grouped into four groups, reflecting the peculiarity of Sardinian individuals and the genetic differences among the peninsula [20]. Basques form two region-specific genetic groups: one in France and one in Spain. The British samples fall into two different clusters;

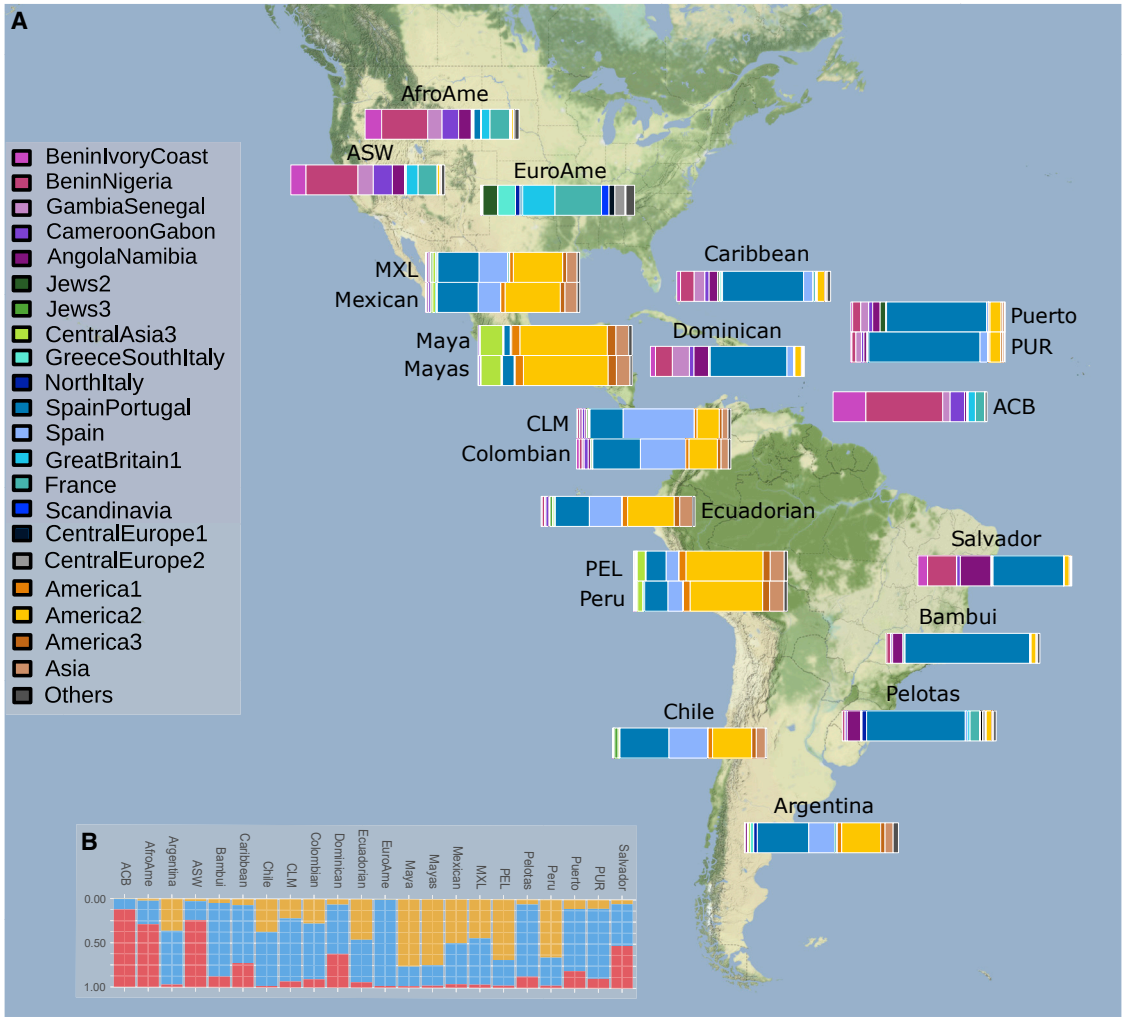

Figure 1. The Ancestral Mosaic of American Populations Reveals a Highly Complex Ancestral Composition

(A) Barplots representing ancestral genetic proportions based on SOURCEFIND results for North and South American populations. We applied CHROMOPAINTER/fineSTRUCTURE and SOURCEFIND to find the ancestral compositions of 22 American populations. Only the contribution for the 21 most representative fineSTRUCTURE clusters (contributing $\geq$2% in at least one recipient population) is reported (Data S2A).

(B) Proportion of continental ancestries for all target populations. Ancestries are represented in red for Africa, blue for Europe, and yellow for America/Asia.

See also Figures S1, S4, and S5 and Data S1, S2A, and S2B. Map tiles by Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under ODbL.

the vast majority of the samples (74 individuals) cluster together with 17 Welsh and 4 additional individuals from Germany and Sweden, while a smaller subset (24 individuals) forms a homogeneous cluster with Orcadians, possibly reflecting the northernmost nature of the group. Central North Eastern Europe is represented by seven clusters containing individuals from multiple countries, such as Lithuania, Poland, Belarus, Hungary, Russia, Germany, Austria, Finland, and Norway. Four distinct groups, including Jewish individuals, were identified (Figure S2A; Data S2A). Native Americans (American populations characterized by more than 95% of autochthonous ancestry) are grouped into three main clusters: one composed only by Brazilian samples (Karitiana and Surui), one including only Wichi, and one composed by several populations (Piapoco from Colombia, Colla from Argentina, and Tepehuano, Zapotec, and Pima from Mexico). East Asian and Oceania individuals are grouped into five clusters, each exclusively containing individuals from the same population (Data S2A).

Our fineSTRUCTURE results (Figure S2A; Data S2A) confirm the worldwide genetic variation pattern already observed by previous studies at the continental scale [20–24].

The Ancestral Mosaic of American Populations

We fit each of the 22 American populations as a mixture of the identified donor groups using SOURCEFIND [25]. Differently from non-negative least squares (NNLSs) approach, SOURCEFIND harnesses a Bayesian algorithm to provide increased resolution in distinguishing true contribution from background noise (see STAR Methods).

The contribution of the 21 most representative clusters (with proportion of no less than 2% in at least one recipient population) to the American admixed populations are reported in Figure 1A

and Data S2B. The same procedure using NNLS provided consistently similar results (Figure S3).

African Ancestries' Distribution Reflects the Complexity of the Slave Trade Dynamics

Sub-Saharan African ancestry was observed at high proportion in African Americans (AfroAme: 69.0% and ASW [Americans of African ancestry in southwest USA]: 74.1%) and Barbadians (ACB [African Caribbeans in Barbados]: 87.1%), with relatively high contribution registered also for the other Caribbean and Brazilian populations (>10%; Figure 1B).

In detail, "BeninNigeria" cluster showed the highest contribution (≥30% of the total) in AfroAme and ACB, while in other Caribbean populations, the contribution of "BeninNigeria" and "GambiaSenegal" clusters is comparable, with average proportion of 6.9% (min = 2.6%, max = 11%) and 6.7% (min = 3.6%, max = 11.1%), respectively.

Moreover, we found contribution from Gambia and Senegal ("GambiaSenegal"; mean = 4.2%, min = 1.3%, max = 11.1%) in Mexico, Caribbean islands, and Colombia, but not in Brazil, Argentina, and Chile, which have a proportion of less than 0.2%, consistent with previous results [18].

In South America, all the analyzed populations show high heterogeneity in African proportions, the highest values in individuals from Salvador (47.8%) [26], reflecting the high number of deported African slaves for sugar production in the Northeastern area of Brazil in the 17th century [27].

In detail, the African cluster contributing the most is related to groups from Angola and Namibia ("AngolaNamibia" cluster) with Salvador (Brazil) having the highest percentage (>20%), which is similar to the contribution from BeninNigeria (~19%) and mirrors the history of African slaves arrivals in Brazil [27] (Figure S4B; Data S2B). Although a non-negligible contribution from East and South East Africa at the end of the slave trade period has been documented [28], none of the analyzed populations showed an East/South East African ancestry fraction larger than 2%. AfroAme and ASW show the highest proportion of this ancestry (1.2% and 0.8%, respectively). Nevertheless, when the ancestry is explored at individual level, samples with more than 5% of East and/or South East African ancestries were identified in more than 1% of individuals from AfroAme (30/2,004), ASW (2/55), Bambui (10/909), and Pelotas (51/3,629) (Figure S5), supporting recent findings [18].

When dissecting the African ancestry into regional sources (Figure S4D), the UPGMA (unweighted pair group method with arithmetic mean) clustering does not strictly mirror geographical/historical patterns. Yet all the Caribbean and circum-Caribbean populations, with the exception of a Colombian sample, cluster together. Similarly, all the Southern American samples, but not Chile, form a private group. Interestingly, ACB is different from any other populations, composed mainly by "BeninIvoryCoast" and "BeninNigeria" clusters.

Complex Variation of European Ancestries' Distribution

European ancestry was observed at high proportion in European Americans (EuroAme), Caribbean Islands (PUR [Puerto Ricans from Puerto Rico] having the highest proportion, 79%), and Mexico (~42% and ~48% for Mexican and MXL [people with Mexican ancestry from Los Angeles, USA], respectively) but also in Southern America, with proportions ranging from 22% in Peru (PEL [Peruvians from Lima]) to ~82% in Bambui.

When the variation of European ancestry in the Americas is evaluated, groups from the United States (EuroAme, AfroAme, and ASW) and Barbados (ACB) are characterized by a substantial proportion of British and French ancestries. On the contrary, in the remaining populations, the most prominent European ancestry was represented by Iberian-related clusters, reflecting the geo-political extent of European occupation during the Colonial Era (Figure 1A). In details, populations from Mexico, Caribbean islands, and South America derive most of their European ancestry from the Iberian Peninsula, represented by two clusters. EuroAme exhibit high levels of heterogeneity, showing not only a high proportion of France and Great Britain but also Greece, South Italy, Central Europe, and Scandinavia, revealing the high variability of European ancestries in the United States, possibly due to secondary movements in the 19th and 20th centuries [29], which involved populations that did not take part in the Colonial Era movements [9]. Moreover, Pelotas (Brazil) is characterized by a high contribution from North Italy (~3%) while Argentina is characterized by contributions from both North and South Italy (2.3% and 2.2%, respectively).

The investigation of the individual ancestry profiles confirmed and further refined the identification of multiple European secondary sources.

In one AfroAme population sample, we identified a high variability of European ancestry, with several individuals characterized by more than 5% ancestry from Northern, Central, and Southern European regions (Figure S5).

Italian ancestry was found at considerable proportion (>5%) in individuals from Colombia (4/98), Caribbean (51/1,112), Dominican Republic (2/27), Ecuador (1/19), Mexico (15/427), Peru (6/153), Puerto Rico (4/99), Argentina (27/133), and Brazil (622/5,779). In fact, Italy has been reported as one of the main sources of migrants to South America during the 19th century, second only to the Spanish and Portuguese influences [30] (Figures S4A and S5).

We estimated the relationship among American populations considering the relative European ancestries proportion by applying a UPGMA clustering approach (Figure S4C). Differences in regional affinities to British and/or French versus Spanish and/or Portuguese ancestries among American populations were observed. Furthermore, within the last group, Spanish and Portuguese ancestries show distinct geographical distributions, consistent with the Treaty of Tordesillas, signed in 1494 to regulate the regional influence of Spain and Portugal in the Americas (Caribbean islands represent an exception; Figure S4C).

Native American Ancestry Distribution

With the exception of Mayan individuals (>65%), Native American ancestry is high in populations from the Southern part of the continent and in Mexico (41%), with the highest values in Peru (59.2% PEL), Ecuador (37%), and Argentina (31%; Figure 1; Data S2B). Interestingly, in both the analyzed AfroAme population samples, we identified a non-negligible proportion of individuals harboring Native-American-related ancestry.

The Contribution of Jewish-Related Ancestry in the Americas

A recent genetic investigation found a non-negligible proportion of ancestry related to Jews and Middle East groups in five populations from Northern and Southern America (Mexico, Colombia, Peru, Chile, and Brazil) [25]. In our analysis,

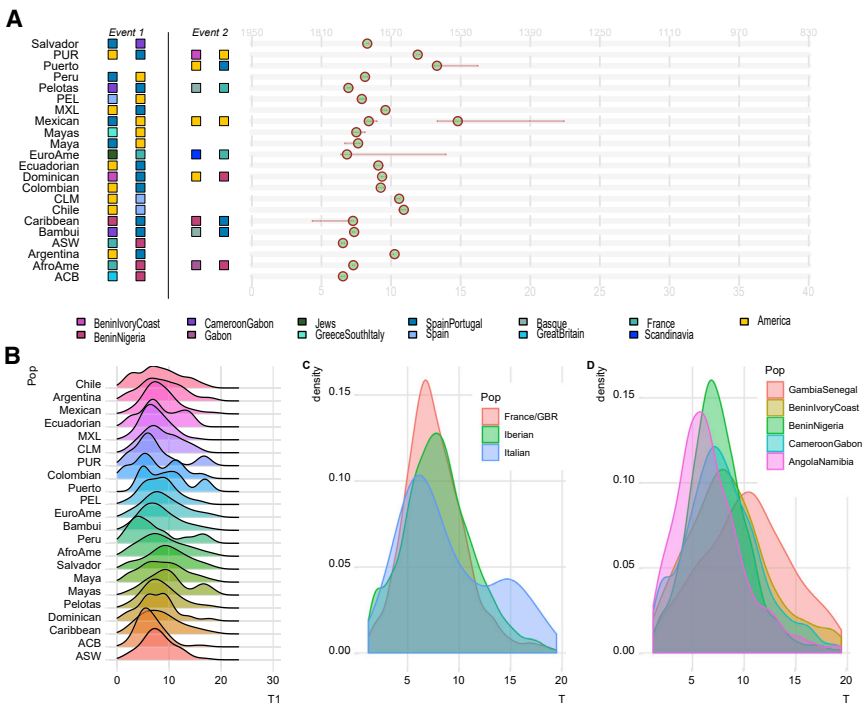

Figure 2. The Admixture History of the Americas, as Inferred by GLOBETROTTER (GT)

(A) Estimates of time and sources of admixture events considering the whole population as target. One or two events of admixture are reported for each population. The closest inferred sources of admixture are represented as colored squares; circles show the corresponding time of admixture estimated by GT. Time is expressed in generations from present (bottom x axis) and years of CE (top x axis). Red bars indicate the 95% CIs of bootstrapped analyses. Only admixture times estimated to have occurred in the last 30 generations are shown. Multiple admixture events may have occurred at the same time.

(B) Distribution of admixture times considering single individuals as targets. We retained only the 2.5%–97.5% distribution of time estimation for each population.

(C) Density of admixture times inferred in events considering France/GBR, Iberian, and Italian clusters as sources, for all the 11,607 admixed American individuals under study.

(D) Density of admixture times inferred in events considering "GambiaSenegal", "BeninIvoryCoast", "BeninNigeria", "CameroonGabon", "Gabon", and "AngolaNamibia" clusters as sources for all the 11,607 admixed American individuals under study. See also Data S2C.

we confirmed the presence of genetic ancestries related to "NorthAfrica", "Levant", "LevantCaucasus", and "Jews" clusters in the same countries, although at a lower proportion than previously estimated (~2.8%). This discrepancy might be due, at least in part, to the fact that our dataset is mostly composed by Brazilian individuals, which have been documented to have a smaller Jewish ancestry [25]. Only 2.5% of analyzed individuals contain more than 5% of Jewish or Middle Eastern ancestry (Salvador: 0.8%, Bambui: 3.2%, Pelotas: 2.9%). In contrast, this proportion is higher in the non-Brazilian populations (CLM [Colombians from Medellin]: 8%, Colombian: 3.8%, Peru: 2.3%, Mexican: 5.4%, MXL: 11%, Chile: 16%, Argentina: 12%). Similar proportions were found for Caribbean populations (ACB: 1.4%, Caribbean: 6.8%, Dominican: 3.7%, Puerto: 3.9%, PUR: 1.4%). Interestingly, we found a relatively high proportion of individuals showing more than 5% contributions close to "Jewish" sources also in AfroAme (3.8%) and in EuroAme (26.7%) (Figure S5).

Inferring the Time of Admixture in American Populations

To provide a temporal dimension to the gene flow among the analyzed populations, we inferred the time of admixture by applying GLOBETROTTER (GT) in two different setups for "population" and "individual" level analyses, as detailed in the STAR Methods section. For both analyses, we focused on admixture events inferred to have occurred in the last 30 generations.

In population-level inferences, all the analyzed groups showed evidence of at least one admixture event as reported in Figure 2A and Data S2C. Specifically, we identified one admixture event in 14 populations (ASW, ACB, Mayas, Maya, PEL, Peru, Salvador, Ecuadorian, Colombian, MXL, Argentina, CLM, Chile, and Puerto) with inferred times spanning between ~6 and ~11 generations ago. The identified sources are related to British or French and Benin-Nigeria in ACB and ASW and Iberian or Southern European and America in Maya, Mayas, PEL, Puerto, Peru, Ecuador, Colombian, CLM, MXL, Argentina and Chile, in line with SOURCEFIND estimates. In contrast, Salvador sources are representative of Iberia and Cameroon-Gabon. Two populations from Caribbean Islands, PUR and Dominican, showed a curve profile that fits better with a single admixture involving more than two sources from Europe, Africa, and America, dated ~9–11 generations ago. The remaining six populations (Mexican, EuroAme, Pelotas, Caribbean, AfroAme, and Bambui) showed signature of at least two admixture events mainly involving American, European, and African sources, with the most recent occurring ~6–8 generations ago.

To assess regional spatiotemporal differences in admixture dynamics, we performed a GT individual analysis (Figures 2B and 2D). For all the analyzed populations, the inferred 2.5%–97.5% time interval had similar boundaries spanning between ~1 and ~20 generations ago (min = 1.18, max = 19.5).

The source-specific admixture time estimates were explored, evaluating the distributions of time inferred considering different European and African signals (Figures 2B and 2C). When the European sources were considered, times involving Iberian clusters were significantly older than those involving British/French ones, which in turn were characterized by dates significantly older than those involving Italian sources (Wilcoxon test; Bonferroni adjusted p value < 0.05).

For the five African sources considered, times inferred for the "SenegalGambia" cluster are significantly older than all the other tested sources (Wilcoxon test; Bonferroni adjusted p value < 0.05). In contrast, times involving "AngolaNamibia" are more recent than all the others (Wilcoxon test; Bonferroni

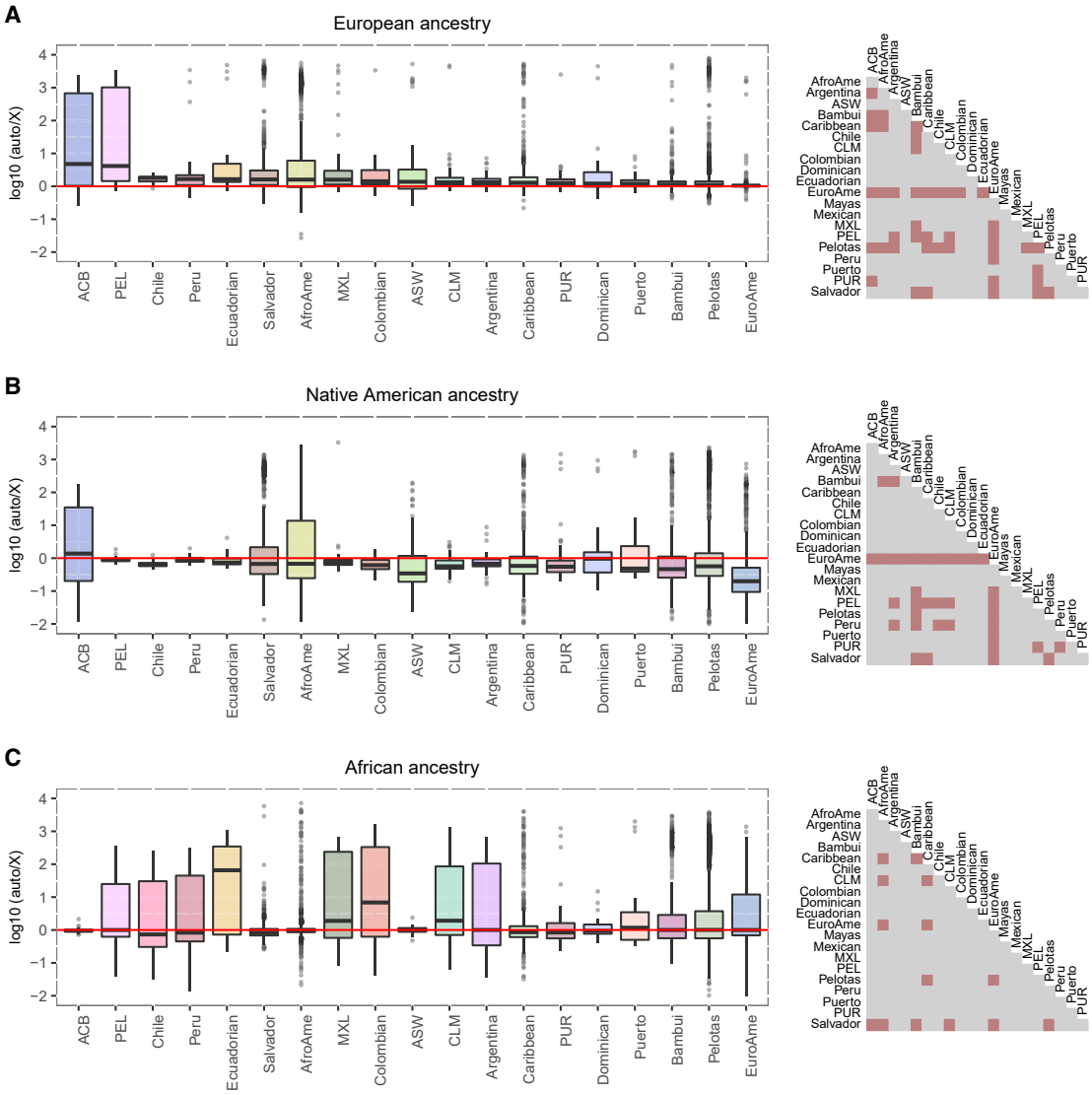

Figure 3. Autosomal versus X Chromosome Ancestry Proportions

Each boxplot shows the log10-scaled ratio of autosomal to X chromosome ancestry proportion for (A) European, (B) Native American, and (C) African continental components as inferred by ADMIXTURE analysis (K = 3) in 19 American populations. Boxplots refer to the interquartile range, and whiskers refer to data points not exceeding the interquartile range more than 1.5 times. All the other data points are considered outliers and shown as dots. See also Data S2D. The heatmap on the right side of every panel shows the significance of Wilcox test between pairs of autosomal versus X distributions, with red color highlighting significant tests ($p < 0.05$ after Bonferroni correction).

adjusted p value < 0.05). Moreover, times involving "BeninIvoryCoast" are significantly older than the one involving "BeninNigeria" and "CameroonGabon". Lastly, times involving CameroonGabon are older than the one involving "BeninNigeria" (Wilcoxon test; Bonferroni adjusted p value < 0.05).

Assessing the Impact of Sex-Biased Admixture of the Americas

To evaluate the impact of sex-biased admixture dynamics in the American populations, we compared the continental ancestry proportions inferred by ADMIXTURE [31] from autosomal data against those estimated for the X chromosome (see STAR Methods). With respect to European ancestry, a paired Wilcoxon test comparing the distribution of autosomal versus X chromosome revealed that the former is significantly higher in all comparisons, suggesting a higher contribution of European males than females in the gene pool of American populations (Figure 3; Data S2D), in agreement with previous continental-scale reports based on more limited data [26, 32, 33]. This observation is further supported by the fact that Native American ancestry estimated from autosomal data is always lower (with the exception of Dominican) than that estimated from the X chromosome. In contrast, when considering the African ancestry, a considerable number of populations do not show any signature of sex imbalance. Indeed, in only eight out of 19 comparisons (ACB, AfroAme, Bambui, Caribbean, EuroAme, Pelotas, PUR, and

Salvador), the autosomal proportion was significantly lower than that inferred from the X chromosome (adjusted p value < 0.05). With the exception of ACB, all these significant differences were associated with sample sizes greater than 100. These results are in contrast with historical records documenting a higher number of disembarked male slaves [28] and might reflect complex admixture dynamics. In fact, gender imbalance in treatment of slaves could lead to different chances to have descendants and therefore explain, at least in part, these results. Alternatively, they could reflect limitations in the approach exploited here, as previously suggested [34]. We repeated the test for two single chromosomes resembling chromosome X in terms of length and number of markers analyzed. Despite the fact that some of the comparisons were no longer significant for chromosome 19, and at a lesser extent for chromosome 7, the overall observed pattern persisted.

We evaluated differences in the distribution of autosomal versus X chromosome continental proportion performing pairwise Wilcoxon test among populations (Figure 3). For all the continental ancestries evaluated, we observed a substantial homogeneity in autosomal/X ratio, suggesting that similar admixture dynamics took place in the whole continent, despite historical, cultural, and geographical differences among populations. In fact, 93%, 79%, and 82% of pairwise comparisons between distributions were not significant (after Bonferroni correction) for sub-Saharan Africa, Europe, and America, respectively. For European ancestry, European American and Pelotas population show significant differences when compared to most of the other groups. These results might be due to more recent and heterogeneous contribution from Europe.

Reconstructing the Ancestry-Specific Demographic Histories of Admixed Populations

To characterize the demographic history of specific continental ancestries, we intersected the results of identity-by-descent (IBD) and LA inferences as in Browning et al. [35]. We excluded from the analysis all the population ancestries in which $\alpha(continent) * n < 50$, where α is the proportion of a specific ancestry as estimated by SOURCEFIND and n is the total number of chromosomes in the analyzed population.

The majority of the studied populations showed, for all the continental ancestries considered, a demographic curve characterized by a decline until approximately 10 generations ago, followed by a general recovery. In a random-mating scenario, all the admixing "ancestries" are expected to present an identical demographic path, scaled by their proportion in the admixed deme. However, when different dynamics (such as assortative mating) occurred, differences in Ne through time for single ancestries could emerge. The correlation of single ancestries on the same population might be explained by the existence of a scenario close to random mating.

This pattern is not universally observed in all the American populations: the Brazilian samples from Bambui showed a general decline in population size for the African and European ancestry, according to previous surveys reporting their low heterogeneity [26]. Conversely, the European ancestry for EuroAme does not show signs of demographic decline, possibly reflecting multiple European waves contributing to this population.

When evaluating the Native American ancestry, the Mexican sample differs from all the others not showing any decrease in the effective population size. A similar behavior was shown when the two samples from Peru were pooled together (Figure S6H) and could reflect admixture among different Native American groups occurred after the European colonization or different demographic histories across various American regions.

For the European ancestry, PUR and CLM showed the most severe decline in effective population size (Figures 4 and S6G).

Interestingly, for the four populations showing a decline-recovery pattern and for which the effective population size for African and European components were available, the African ancestry started to recover later than the European one, with the exception of the Caribbean population. Furthermore, when all the available data points are considered, the time of the last minimum before the recovery is significantly younger for the African ancestry (Wilcoxon test; $p < 0.05$).

We explored the correlation in demographic trajectories among pairs of populations considering two different time intervals: after 30 generations ago and between 30 and 60 generations ago (Figures S6A and S6F). For all the observed ancestries, especially for the European and American one, the overall degree of pairwise correlation is lower in recent times than in the past. These results may suggest that admixture acted as a diversifying factor in terms of past demographic evolutionary trajectories, as opposed to having a homogenizing effect in terms of genetic variability [18]. Data analysis for large sample size groups of diverse American groups may elucidate the overall admixture impact pattern in the continents.

DISCUSSION

Despite being virtually isolated from the rest of the world until 500 years ago, most of the individuals living in the Americas harbor, together with Native American ancestry, a substantial genomic proportion inherited from Europe and Africa. These ancestral mosaics could be explained as the consequence of two major historical processes occurring over the Americas: first, the genocide of indigenous people of America [36] that caused a sharp decrease of the Native American populations and, second, the admixture occurring after the European exploration and colonization, which was followed by African deportation and labor migration that has impacted the American continents in the 19th and 20th centuries.

The investigation of the times of admixture among the two continents revealed that all the analyzed present day American populations are the result of at least one admixture event involving Native American, African, and European sources within the last 6–12 generations, corresponding to 1644 CE and 1812 CE (considering a generation time of 28 years; Figures 1 and 2). However, considering a population as a whole does not fully capture the complexity of its admixture dynamics, characterized by several waves of migration in the last five centuries [7, 28, 37]. One way to partially overcome this limitation is analyzing single individuals rather than populations, evaluating a higher degree of variation in fragment length distributions. Our per-individual time estimations provided several insights into the complexity of admixture in the Americas. It has been previously reported

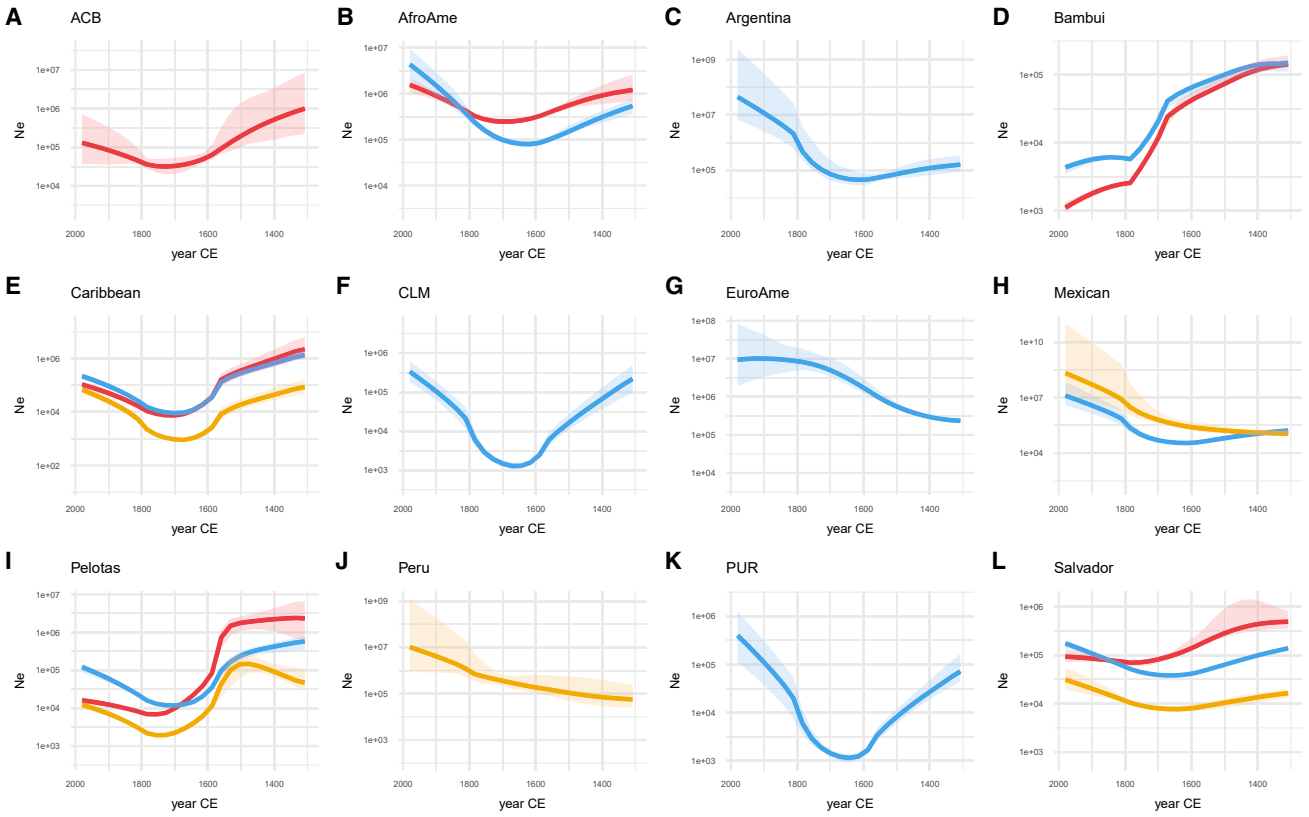

Figure 4. Ancestry-Specific Effective Population Size of American Populations

We combined identity by descent and local ancestry inferences to estimate ancestry-specific population size through time. The x axes show time expressed in years of CE. The y axes show ancestry-specific effective population size (Ne), plotted on a log scale. Solid lines show estimated ancestry-specific effective population sizes (red, African ancestry; blue, European ancestry; yellow, Native American ancestry), with ribbons indicating the 95% CIs. Only the population ancestries, in which α(continent) × n > 50 (where α is the proportion of a specific ancestry and n is the total number of chromosomes in the analyzed population), are represented. See also Figure S6.

that the origin of Africans disembarked in the continents followed a general North-South temporal pattern [38], with slaves from Senegal and Gambia being deported earlier than the ones from more southern areas (https://www.slavevoyages.org). In accordance with historical data, the inferred admixture dates involving populations from Senegal and Gambia are older than the ones involving all the others; this area remained the main slave trade site for the Spanish possessions until 1640 [39]. Similarly, all the dates involving clusters related to Angolan and Namibian individuals are characterized by younger recent admixture times (Figure 2D).

For European sources, the estimated admixture dates characterized by gene flow from Iberia are older than the dates of admixture with France/Great Britain sources, which in turn are older than admixture events involving Italian sources, which, according to historical records, became substantial only in the second half of the 19th century.

Furthermore, we assessed the severe impact of the Atlantic slave trade in several populations under study, with a pattern that reflects historical records [28, 37, 39].

In detail, our analysis revealed that West Central African ancestry is the most prevalent in the American continents, as previously reported [14, 18]. Moreover, we additionally identified a high contribution from Senegal and Gambia in the Caribbean, Mexico, and Colombia in accordance with African slave arrivals predominantly to Spanish-speaking America until 1620s [28].

Subsequently, about 50% of all West African slaves were deported to Dutch, French, and British sugar plantations in the Caribbean. Accordingly, we estimated a high contribution from Benin and Nigeria in all the Caribbean populations and in populations from the United States, in line with the reported slave arrivals.

Among all the analyzed populations, ACB is characterized by the highest sub-Saharan ancestry proportion (~88%), possibly due to the presence of sugar cane industry combined with the relatively low European immigration [37] in the 18th century.

At a microgeographic scale, Barbadians derive their African ancestry from "BeninNigeria" (~50%) and from "BeninIvoryCoast" (~21%; Figure 1A; Data S2B), two of the main source areas reported for the British-mediated slave trade.

In contrast, Brazil showed a peculiar African ancestral composition, characterized by a high proportion of ancestry related to modern-day Angola and Namibia, consistent with the Portuguese settlement in Angola from the beginning of the 17th century. A similar African component is also observed in Argentina,

probably due to the fact that slaves arrived primarily from Brazil via the Portuguese slave trade from Angola [38, 40].

The Atlantic coast of Africa was not the only region involved in slave deportation; in fact, in the last decades of the slave trade period, Mozambique was the third largest supplier of slaves [28]. We found ancestries from Southern East African groups in a non-negligible proportion of individuals from Bambui and Pelotas.

Although similar works [14, 18] analyzed Bantu populations from Southern, Southern Eastern, and Eastern Africa, here, we included Bantu populations from Angola, which has been documented as one of the main regions for slave deportation. Considered together, this study and Gouveia et al. [14, 18] suggest an important role of Southwestern, South, and Southeastern Africa in shaping the African gene pool of populations from the Atlantic Coast of the Southern Cone of South-America.

When we investigated the European contribution to the continents, we confirmed the large impact of Great Britain, France, and Iberian Peninsula for all the tested populations, with a distribution reflecting the geographic occupation of the Americas in the Colonial Era.

Furthermore, our approach revealed the existence of several European secondary sources contributing to a substantial proportion of American populations. Among the others, we have identified ancestry closely related to Italian populations in European Americans from the US, Argentinians, and Brazilian populations [41].

The Italian migration in the Americas has been recently described as one of the largest migrations of the 19th century and has been usually referred to as the ''Italian diaspora'' [42–44]. Although it started soon after 1492, it reached high proportions only in the second half of the 19th century, with more than 11 million individuals migrating toward the continents, largely to the United States, Brazil, and Argentina.

Between 1866 and 1916, approximately 4 million Italians were admitted in the United States. In the 2017 US Census Bureau, nearly 17 million people (5% of global population) were reported as Italian, with proportions spanning from 1.3% to 17.0% in different states.

In Brazil, also thanks to subsidies offered by the society for the promotion of immigration, after 1820, nearly half of all immigrants were Italians, and in 1876, their annual arrival rate became higher than the one from Portugal. These migrations continued steadily until 1902, when a decree of the Italian government put an end to all subsidized emigration to Brazil [45]. We found genetic signals of these migrations, mostly related to North Italy, in all the three Brazilian samples analyzed, with the highest proportion in Pelotas, followed by Bambui and Salvador.

In Argentina, the identified Italian contribution is related both to the northern and southern part of the peninsula, which is in accordance with movements of millions of individuals from Northern (earlier) and Southern (later) Italy registered from the second half of 1800 throughout the 1950s [9, 30]. It has been reported that Italian immigration was the highest (39.4%) compared to the ones from other countries at the beginning of the 20th century [46, 47]. Therefore, at a pan-American level, the distribution of the Italian components is heterogeneous and closely reflects the one reported by historical records.

Differently from the other Brazilian groups analyzed, Pelotas is also characterized by contributions from additional sources, such as Central and North Europe (''GreatBritain1'', ''France'', ''CentralEurope1-2'', and ''Scandinavia'') in accordance with historical records.

Recently, a survey employing similar methods on five Southern American populations identified South and East Mediterranean ancestries across Americas, which has been interpreted as a contribution from Converso Jews [25]. Our analysis of the individual ancestry distribution confirmed the presence of Jews and Levantine ancestries in virtually all the analyzed populations, including those from the Caribbean (Figure S5).

By evaluating the continental ancestry estimates using an allele frequency method, we were able to confirm the sex-biased admixture dynamics, suggesting that a higher number of American females than males have contributed to the modern populations. Conversely, European males had a larger contribution than females from the same continent.

In contrast, for the African ancestry, we observed inconsistent results, with some of, but not all, the populations showing evidence for a higher female contribution, partially discordant with historical reports. A possible explanation might be that the ratio between African males and females is lower than the one observed for the European component, preventing its identification with small sample sizes and suggesting that such patterns (or their absence) should be interpreted with caution, as previously suggested [34]. In addition, it is possible that the different treatment of female and male slaves, and the resulting unbalanced chances of having descendants, may have contributed to add complexity on the admixture dynamics.

All these results confirm that the European and African components are playing an important role in shaping the genetic differentiation of different American groups, although their demographic evolution after the arrival in the ''new world'' is still unknown.

The analysis of ancestry-specific effective population sizes demonstrated that, regardless of their composition, most of the continental ancestries experienced a general decrease until approximately 10 generations ago, after which a general population size recovery was inferred (Figures 4 and S6G).

Interestingly, the recovery of the African population component postdates those of the European one, possibly reflecting the different conditions experienced by African slaves and European settlers.

On the other hand, the effective population size of the Native American component in Mexicans and Peruvians does not show evidence of decrease, in contrast with historical records reporting a general dramatic decline of the Native American population after European colonization.

This observation is in line with Browning et al. [35], in which a smaller reduction in the effective population size of Mexicans for Native American ancestry compared to other populations was observed. This result is also in line with our GLOBETROTTER results, where we found evidence for admixture between two Native-American-related sources approximately 15 generations ago.

It may be possible that the reported decline did not heavily affect the genetic variability of survivor populations or that individuals from different isolated native groups have been put in

contact as a consequence of the European colonization and deportation, as recently suggested for Peruvian populations [48]. This would result in an inflated effective population size estimate, as we observe in our ancestry-specific analysis.

In conclusion, we demonstrated that the European and African genomic ancestries in American populations are composed of several different sources that arrived in the Americas in the last six centuries, dramatically affecting their demography and mirroring historical events. The analysis of high-quality genomes from the American continents, combined with the analysis of ancient DNA and denser sampling, will be crucial to better clarify the genetic impact of these dramatic events. In addition, the fine-scale composition here reported is important for the future development of epidemiological, translational, and medical studies.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- METHOD DETAILS
  - Analyzed data
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - PC Analysis
  - Phasing
  - Clustering of donor populations
  - Painting of the recipient populations
  - Bayesian haplotype-based ancestry estimation (SOURCEFIND)
  - Non-Negative Least Square haplotype-based ancestry estimation
  - Estimation of admixture dates
  - Ancestry-specific effective population size estimation
  - Sex-biased admixture evaluation
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cub.2019.09.076.

ACKNOWLEDGMENTS

We thank Estonian Genome Center and Andres Metspalu for access to the data of Georgian, German, Hungarian, Latvian, Lithuanian, and Ukrainian samples. We thank Doron Behar for early access to the Ashkenazi Jew, Karaite Jew, and Lemba Jew data. We acknowledge Elza Khusnutdinova for early access to Avar, Besserman, Cirkassian, Kabardin, Karelian, Komi, Mordovian, Udmurt, and Vepsa data. We thank Rusudan Khukhunaishvili and Sophiko Tskvitinidze for early access to Adjarian data. We thank Oleg Balanovsky for access to Cossack, Coast Crimean Tatar, Mountain Crimean Tatar, and Steppe Crimean Tatar data. All these samples will be published in dedicated papers. We thank Lena Khushniarevich for Belarusian Tatar samples and Vladimir Ferak for some of the Roma samples. This study received support from the Italian Ministry of Education, University and Research (MIUR): Dipartimenti di Eccellenza Program (2018–2022), Department of Biology and Biotechnology "L. Spallanzani", University of Pavia (to A.A., A.O., O.S., and A.T.) and the Fondazione Cariplo (project no. 2018-2045 to A.A., A.O., and A.T.). This research was supported by the European Union through the European Regional Development Fund (project no. 2014-2020.4.01.16-0030 to L.O., M.M., and F.M.; project nos. 2014-2020.4.01.16-0271 and 2014-2020.4.01.16-0125 to R.F.; and project no. 2014-2020.4.01.16-0024 to D.Marnetto and L.P.). This work was supported by the Estonian Research Council grant PUT (PRG243) to R.F., M.M., and L.P. This work was supported by institutional research funding IUT (IUT24-1) of the Estonian Ministry of Education and Research to T.K. This research was supported by the European Union through Horizon 2020 grant no. 810645 (to M.M.). Computational analyses were performed at the High Performance Computing Center of the University of Tartu; we thank Tuuli Reisberg for assistance in data management. This work was supported by the state assignments of Russian Ministry of Science and Higher Education for the Research Centre for Medical Genetics (to E.B.) and for the Vavilov Institute of General Genetics (0112-2019-0001 to O.B.).

AUTHOR CONTRIBUTIONS

Conceptualization, C.C., L.P., M.M., and F.M.; Methodology and Supervision, F.M., E.T.-S., L.P., and M.M.; Investigation, L.O., F.M., L.P., and R.F.; Resources, R.J.K., A.K., O.B., E.B., L.A., S.T., S.M., A.R., A.A., A.O., O.S., A.T., S.S., G.A.G.-R., D.L., M.E.A.-R., E.P., P.W., G.H., M.G.-S., D.Marjanovic, L.M., A.L., N.M., M.F.L.-C., A.C.P., M.L.B., B.L.H., C.A.M., A.M.-E., and C.C.; Writing – Original Draft, L.O., F.M., L.P., R.F., D.Marnetto, K.T., T.K., M.O.S., and E.T.-S. with support from all the coauthors.

DECLARATION OF INTERESTS

G.H. is a founding member of GenSci.

Received: July 5, 2019
Revised: September 6, 2019
Accepted: September 30, 2019
Published: November 14, 2019

REFERENCES

1. Tamm, E., Kivisild, T., Reidla, M., Metspalu, M., Smith, D.G., Mulligan, C.J., Bravi, C.M., Rickards, O., Martinez-Labarga, C., Khusnutdinova, E.K., et al. (2007). Beringian standstill and spread of Native American founders. PLoS ONE *2*, e829.
2. Hoffecker, J.F., Elias, S.A., O'Rourke, D.H., Scott, G.R., and Bigelow, N.H. (2016). Beringia and the global dispersal of modern humans. Evol. Anthropol. *25*, 64–78.
3. Grugni, V., Raveane, A., Ongaro, L., Battaglia, V., Trombetta, B., Colombo, G., Capodiferro, M.R., Olivieri, A., Achilli, A., Perego, U.A., et al. (2019). Analysis of the human Y-chromosome haplogroup Q characterizes ancient population movements in Eurasia and the Americas. BMC Biol. *17*, 3.
4. Scheib, C.L., Li, H., Desai, T., Link, V., Kendall, C., Dewar, G., Griffith, P.W., Mörseburg, A., Johnson, J.R., Potter, A., et al. (2018). Ancient human parallel lineages within North America contributed to a coastal expansion. Science *360*, 1024–1027.
5. Achilli, A., Perego, U.A., Lancioni, H., Olivieri, A., Gandini, F., Hooshiar Kashani, B., Battaglia, V., Grugni, V., Angerhofer, N., Rogers, M.P., et al. (2013). Reconciling migration models to the Americas with the variation of North American native mitogenomes. Proc. Natl. Acad. Sci. USA *110*, 14308–14313.
6. Moreno-Mayar, J.V., Vinner, L., de Barros Damgaard, P., de la Fuente, C., Chan, J., Spence, J.P., Allentoft, M.E., Vimala, T., Racimo, F., Pinotti, T., et al. (2018). Early human dispersals within the Americas. Science *362*, eaav2621.
7. Bethell, L. (1984). The Cambridge History of Latin America (Cambridge University Press).
8. Boyd-Bowman, P. (1976). Patterns of Spanish emigration to the Indies until 1600. Hisp. Am. Hist. Rev. *56*, 580–604.
9. Baily, S.L., and Míguez, E.J. (2003). Mass Migration to Modern Latin America (Rowman & Littlefield).

10. Moreno-Estrada, A., Gravel, S., Zakharia, F., McCauley, J.L., Byrnes, J.K., Gignoux, C.R., Ortiz-Tello, P.A., Martínez, R.J., Hedges, D.J., Morris, R.W., et al. (2013). Reconstructing the population genetic history of the Caribbean. PLoS Genet. *9*, e1003925.
11. Moreno-Estrada, A., Gignoux, C.R., Fernández-López, J.C., Zakharia, F., Sikora, M., Contreras, A.V., Acuña-Alonzo, V., Sandoval, K., Eng, C., Romero-Hidalgo, S., et al. (2014). Human genetics. The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. Science *344*, 1280–1285.
12. Bryc, K., Durand, E.Y., Macpherson, J.M., Reich, D., and Mountain, J.L. (2015). The genetic ancestry of African Americans, Latinos, and European Americans across the United States. Am. J. Hum. Genet. *96*, 37–53.
13. Homburger, J.R., Moreno-Estrada, A., Gignoux, C.R., Nelson, D., Sanchez, E., Ortiz-Tello, P., Pons-Estel, B.A., Acevedo-Vasquez, E., Miranda, P., Langefeld, C.D., et al. (2015). Genomic insights into the ancestry and demographic history of South America. PLoS Genet. *11*, e1005602.
14. Montinaro, F., Busby, G.B.J., Pascali, V.L., Myers, S., Hellenthal, G., and Capelli, C. (2015). Unravelling the hidden ancestry of American admixed populations. Nat. Commun. *6*, 6596.
15. Han, E., Carbonetto, P., Curtis, R.E., Wang, Y., Granka, J.M., Byrnes, J., Noto, K., Kermany, A.R., Myres, N.M., Barber, M.J., et al. (2017). Clustering of 770,000 genomes reveals post-colonial population structure of North America. Nat. Commun. *8*, 14238.
16. Montinaro, F., Busby, G.B.J., Gonzalez-Santos, M., Oosthuitzen, O., Oosthuitzen, E., Anagnostou, P., Destro-Bisol, G., Pascali, V.L., and Capelli, C. (2017). Complex ancient genetic structure and cultural transitions in Southern African populations. Genetics *205*, 303–316.
17. Pagani, L., Schiffels, S., Gurdasani, D., Danecek, P., Scally, A., Chen, Y., Xue, Y., Haber, M., Ekong, R., Oljira, T., et al. (2015). Tracing the route of modern humans out of Africa by using 225 human genome sequences from Ethiopians and Egyptians. Am. J. Hum. Genet. *96*, 986–991.
18. Gouveia, M.H., Borda, V., Leal, T.P., Moreira, R.G., Bergen, A.W., Aquino, M.M., Araujo, G.S., Araujo, N.M., Kehdy, F.S.G., Liboredo, R., et al. (2019). Origins, admixture dynamics and homogenization of the African gene pool in the Americas. bioRxiv. https://doi.org/10.1101/652701.
19. Lawson, D.J., Hellenthal, G., Myers, S., and Falush, D. (2012). Inference of population structure using dense haplotype data. PLoS Genet. *8*, e1002453.
20. Raveane, A., Aneli, S., Montinaro, F., Athanasiadis, G., Barlera, S., Birolo, G., Boncoraglio, G., Di Blasio, A.M., Di Gaetano, C., Pagani, L., et al. (2019). Population structure of modern-day Italians reveals patterns of ancient and archaic ancestries in Southern Europe. Sci. Adv. *5*, eaaw3492.
21. Tishkoff, S.A., Reed, F.A., Friedlaender, F.R., Ehret, C., Ranciaro, A., Froment, A., Hirbo, J.B., Awomoyi, A.A., Bodo, J.M., Doumbo, O., et al. (2009). The genetic structure and history of Africans and African Americans. Science *324*, 1035–1044.
22. Patin, E., Lopez, M., Grollemund, R., Verdu, P., Harmant, C., Quach, H., Laval, G., Perry, G.H., Barreiro, L.B., Froment, A., et al. (2017). Dispersals and genetic adaptation of Bantu-speaking populations in Africa and North America. Science *356*, 543–546.
23. Busby, G.B.J., Hellenthal, G., Montinaro, F., Tofanelli, S., Bulayeva, K., Rudan, I., Zemunik, T., Hayward, C., Toncheva, D., Karachanak-Yankova, S., et al. (2015). The role of recent admixture in forming the contemporary West Eurasian genomic landscape. Curr. Biol. *25*, 2878.
24. Yunusbayev, B., Metspalu, M., Metspalu, E., Valeev, A., Litvinov, S., Valiev, R., Akhmetova, V., Balanovska, E., Balanovsky, O., Turdikulova, S., et al. (2015). The genetic legacy of the expansion of Turkic-speaking nomads across Eurasia. PLoS Genet. *11*, e1005068.
25. Chacón-Duque, J.C., Adhikari, K., Fuentes-Guajardo, M., Mendoza-Revilla, J., Acuña-Alonzo, V., Barquera, R., Quinto-Sánchez, M., Gómez-Valdés, J., Everardo Martínez, P., Villamil-Ramírez, H., et al. (2018). Latin Americans show wide-spread Converso ancestry and imprint of local Native ancestry on physical appearance. Nat. Commun. *9*, 5388.
26. Kehdy, F.S.G., Gouveia, M.H., Machado, M., Magalhães, W.C.S., Horimoto, A.R., Horta, B.L., Moreira, R.G., Leal, T.P., Scliar, M.O., Soares-Souza, G.B., et al.; Brazilian EPIGEN Project Consortium (2015). Origin and dynamics of admixture in Brazilians and its effect on the pattern of deleterious mutations. Proc. Natl. Acad. Sci. USA *112*, 8696–8701.
27. Klein, H.S., and Luna, F.V. (2009). Slavery in Brazil (Cambridge University Press).
28. Klein, H.S. (1999). The Atlantic Slave Trade (Cambridge University Press).
29. Hatton, T.J., and Williamson, J.G. (1992). What Drove the Mass Migrations from Europe in the Late Nineteenth Century? (National Bureau of Economic Research).
30. Meade, T.A. (2011). A History of Modern Latin America: 1800 to the Present (John Wiley & Sons).
31. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. *19*, 1655–1664.
32. Fortes-Lima, C., Gessain, A., Ruiz-Linares, A., Bortolini, M.C., Migot-Nabias, F., Bellis, G., Moreno-Mayar, J.V., Restrepo, B.N., Rojas, W., Avendaño-Tamayo, E., et al. (2017). Genome-wide ancestry and demographic history of African-descendant Maroon communities from French Guiana and Suriname. Am. J. Hum. Genet. *101*, 725–736.
33. Bryc, K., Auton, A., Nelson, M.R., Oksenberg, J.R., Hauser, S.L., Williams, S., Froment, A., Bodo, J.M., Wambebe, C., Tishkoff, S.A., and Bustamante, C.D. (2010). Genome-wide patterns of population structure and admixture in West Africans and African Americans. Proc. Natl. Acad. Sci. USA *107*, 786–791.
34. Goldberg, A., and Rosenberg, N.A. (2015). Beyond 2/3 and 1/3: the complex signatures of sex-biased admixture on the X chromosome. Genetics *201*, 263–279.
35. Browning, S.R., Browning, B.L., Daviglus, M.L., Durazo-Arvizu, R.A., Schneiderman, N., Kaplan, R.C., and Laurie, C.C. (2018). Ancestry-specific recent effective population size in the Americas. PLoS Genet. *14*, e1007385.
36. Jones, A. (2016). Genocide: A Comprehensive Introduction (Routledge).
37. Curtin, P.D. (1972). The Atlantic Slave Trade: A Census (University of Wisconsin Press).
38. Eltis, D., and Richardson, D. (2010). Atlas of the Transatlantic Slave Trade (Yale University).
39. Barry, B. (1998). Senegambia and the Atlantic Slave Trade (Cambridge University Press).
40. Edwards, E. (2014). Slavery in Argentina (Oxford Bibliographies Online Datasets).
41. Foerster, R.F. (1919). The Italian Emigration of Our Times (Harvard University Press).
42. Bonaffini, L., and Perricone, J. (2014). Poets of the Italian Diaspora: A Bilingual Anthology (Fordham University Press).
43. Wong, A.S. (2006). Race and the Nation in Liberal Italy, 1861-1911: Meridionalism, Empire, and Diaspora (Palgrave Macmillan).
44. Giunta, E., and Sciorra, J. (2014). Embroidered Stories: Interpreting Women's Domestic Needlework from the Italian Diaspora (University Press of Mississippi).
45. Tosi, L. (2002). La tutela internazionale dell'emigrazione. In L'emigrazione italiana, II, Arrivi, P. Bevilacqua, A. De Clementi, and E. Franzina, eds. (Donzelli), pp. 439–456.
46. Brown, J.C. (2010). A Brief History of Argentina (Facts On File).
47. Solberg, C.E. (1970). Immigration and Nationalism, Argentina and Chile, 1890–1914 (University of Texas Press).
48. Harris, D.N., Song, W., Shetty, A.C., Levano, K.S., Cáceres, O., Padilla, C., Borda, V., Tarazona, D., Trujillo, O., Sanchez, C., et al. (2018). Evolutionary genomic dynamics of Peruvians before, during, and after the Inca Empire. Proc. Natl. Acad. Sci. USA *115*, E6526–E6535.

49. 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. Nature *526*, 68–74.

50. Li, J.Z., Absher, D.M., Tang, H., Southwick, A.M., Casto, A.M., Ramachandran, S., Cann, H.M., Barsh, G.S., Feldman, M., Cavalli-Sforza, L.L., and Myers, R.M. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. Science *319*, 1100–1104.

51. Behar, D.M., Yunusbayev, B., Metspalu, M., Metspalu, E., Rosset, S., Parik, J., Rootsi, S., Chaubey, G., Kutuev, I., Yudkovsky, G., et al. (2010). The genome-wide structure of the Jewish people. Nature *466*, 238–242.

52. Behar, D.M., Metspalu, M., Baran, Y., Kopelman, N.M., Yunusbayev, B., Gladstein, A., Tzur, S., Sahakyan, H., Bahmanimehr, A., Yepiskoposyan, L., et al. (2013). No evidence from genome-wide data of a Khazar origin for the Ashkenazi Jews. Hum. Biol. *85*, 859–900.

53. Di Cristofaro, J., Pennarun, E., Mazières, S., Myres, N.M., Lin, A.A., Temori, S.A., Metspalu, M., Metspalu, E., Witzel, M., King, R.J., et al. (2013). Afghan Hindu Kush: where Eurasian sub-continent gene flows converge. PLoS ONE *8*, e76748.

54. Eichstaedt, C.A., Antão, T., Pagani, L., Cardona, A., Kivisild, T., and Mormina, M. (2014). The Andean adaptive toolkit to counteract high altitude maladaptation: genome-wide and phenotypic analysis of the Collas. PLoS ONE *9*, e93314.

55. Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., Karthikeyan, S., Iles, L., Pollard, M.O., Choudhury, A., et al. (2015). The African Genome Variation Project shapes medical genetics in Africa. Nature *517*, 327–332.

56. Kovacevic, L., Tambets, K., Ilumäe, A.M., Kushniarevich, A., Yunusbayev, B., Solnik, A., Bego, T., Primorac, D., Skaro, V., Leskovac, A., et al. (2014). Standing at the gateway to Europe–the genetic structure of Western balkan populations based on autosomal and haploid markers. PLoS ONE *9*, e105090.

57. Kushniarevich, A., Utevska, O., Chuhryaeva, M., Agdzhoyan, A., Dibirova, K., Uktveryte, I., Möls, M., Mulahasanovic, L., Pshenichnov, A., Frolova, S., et al.; Genographic Consortium (2015). Genetic heritage of the Balto-Slavic speaking populations: a synthesis of autosomal, mitochondrial and Y-chromosomal data. PLoS ONE *10*, e0135820.

58. May, A., Hazelhurst, S., Li, Y., Norris, S.A., Govind, N., Tikly, M., Hon, C., Johnson, K.J., Hartmann, N., Staedtler, F., and Ramsay, M. (2013). Genetic diversity in black South Africans from Soweto. BMC Genomics *14*, 644.

59. Migliano, A.B., Romero, I.G., Metspalu, M., Leavesley, M., Pagani, L., Antao, T., Huang, D.W., Sherman, B.T., Siddle, K., Scholes, C., et al. (2013). Evolution of the pygmy phenotype: evidence of positive selection fro genome-wide scans in African, Asian, and Melanesian pygmies. Hum. Biol. *85*, 251–284.

60. Pagani, L., Kivisild, T., Tarekegn, A., Ekong, R., Plaster, C., Gallego Romero, I., Ayub, Q., Mehdi, S.Q., Thomas, M.G., Luiselli, D., et al. (2012). Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool. Am. J. Hum. Genet. *91*, 83–96.

61. Pankratov, V., Litvinov, S., Kassian, A., Shulhin, D., Tchebotarev, L., Yunusbayev, B., Möls, M., Sahakyan, H., Yepiskoposyan, L., Rootsi, S., et al. (2016). East Eurasian ancestry in the middle of Europe: genetic footprints of Steppe nomads in the genomes of Belarusian Lipka Tatars. Sci. Rep. *6*, 30197.

62. Patin, E., Siddle, K.J., Laval, G., Quach, H., Harmant, C., Becker, N., Froment, A., Régnault, B., Lemée, L., Gravel, S., et al. (2014). The impact of agricultural emergence on the genetic history of African rainforest hunter-gatherers and agriculturalists. Nat. Commun. *5*, 3163.

63. Petersen, D.C., Libiger, O., Tindall, E.A., Hardie, R.A., Hannick, L.I., Glashoff, R.H., Mukerji, M., Fernandez, P., Haacke, W., Schork, N.J., and Hayes, V.M.; Indian Genome Variation Consortium (2013). Complex patterns of genomic admixture within southern Africa. PLoS Genet. *9*, e1003309.

64. Raghavan, M., Skoglund, P., Graf, K.E., Metspalu, M., Albrechtsen, A., Moltke, I., Rasmussen, S., Stafford, T.W., Jr., Orlando, L., Metspalu, E., et al. (2014). Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature *505*, 87–91.

65. Schlebusch, C.M., Skoglund, P., Sjödin, P., Gattepaille, L.M., Hernandez, D., Jay, F., Li, S., De Jongh, M., Singleton, A., Blum, M.G.B., et al. (2012). Genomic variation in seven Khoe-San groups reveals adaptation and complex African history. Science *338*, 374–379.

66. Tambets, K., Yunusbayev, B., Hudjashov, G., Ilumäe, A.M., Rootsi, S., Honkola, T., Vesakoski, O., Atkinson, Q., Skoglund, P., Kushniarevich, A., et al. (2018). Genes reveal traces of common recent demographic history for most of the Uralic-speaking populations. Genome Biol. *19*, 139.

67. Tamm, E., Cristofaro, J.D., Mazières, S., Pennarun, E., Kushniarevich, A., Raveane, A., Semino, O., Chiaroni, J., Pereira, L., Metspalu, M., and Montinaro, F. (2019). Genome-wide analysis of Corsican population reveals a close affinity with Northern and Central Italy. Sci. Rep. *9*, 13581.

68. Triska, P., Soares, P., Patin, E., Fernandes, V., Cerny, V., and Pereira, L. (2015). Extensive admixture and selective pressure across the Sahel Belt. Genome Biol. Evol. *7*, 3484–3495.

69. Uren, C., Kim, M., Martin, A.R., Bobo, D., Gignoux, C.R., van Helden, P.D., Möller, M., Hoal, E.G., and Henn, B.M. (2016). Fine-scale human population structure in Southern Africa reflects ecogeographic boundaries. Genetics *204*, 303–314.

70. Yunusbayev, B., Metspalu, M., Järve, M., Kutuev, I., Rootsi, S., Metspalu, E., Behar, D.M., Varendi, K., Sahakyan, H., Khusainova, R., et al. (2012). The Caucasus as an asymmetric semipermeable barrier to ancient human migrations. Mol. Biol. Evol. *29*, 359–365.

71. Ghani, M., Sato, C., Lee, J.H., Reitz, C., Moreno, D., Mayeux, R., St George-Hyslop, P., and Rogaeva, E. (2013). Evidence of recessive Alzheimer disease loci in a Caribbean Hispanic data set: genome-wide survey of runs of homozygosity. JAMA Neurol. *70*, 1261–1267.

72. López Herráez, D., Martínez-Bueno, M., Riba, L., García de la Torre, I., Sacnún, M., Goñi, M., Berbotto, G.A., Paira, S., Musuruana, J.L., Graf, C.E., et al. (2013). Rheumatoid arthritis in Latin Americans enriched for Amerindian ancestry is associated with loci in chromosomes 1, 12, and 13, and the HLA class II region. Arthritis Rheum. *65*, 1457–1467.

73. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience *4*, 7.

74. Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M., and Chen, W.M. (2010). Robust relationship inference in genome-wide association studies. Bioinformatics *26*, 2867–2873.

75. Delaneau, O., Marchini, J., and Zagury, J.F. (2011). A linear complexity phasing method for thousands of genomes. Nat. Methods *9*, 179–181.

76. Lawson, C., and Hanson, R. (1995). Solving Least Squares Problems (Society for Industrial and Applied Mathematics).

77. Hellenthal, G., Busby, G.B.J., Band, G., Wilson, J.F., Capelli, C., Falush, D., and Myers, S. (2014). A genetic atlas of human admixture history. Science *343*, 747–751.

78. Browning, B.L., and Browning, S.R. (2013). Improving the accuracy and efficiency of identity-by-descent detection in population data. Genetics *194*, 459–471.

79. Maples, B.K., Gravel, S., Kenny, E.E., and Bustamante, C.D. (2013). RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. Am. J. Hum. Genet. *93*, 278–288.

80. Browning, S.R., and Browning, B.L. (2015). Accurate non-parametric estimation of recent effective population size from segments of identity by descent. Am. J. Hum. Genet. *97*, 404–418.

81. Hancock, D.B., Levy, J.L., Gaddis, N.C., Bierut, L.J., Saccone, N.L., Page, G.P., and Johnson, E.O. (2012). Assessment of genotype imputation performance using 1000 Genomes in African American studies. PLoS ONE *7*, e50610.

82. Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N., Nordenfelt, S., Tandon, A., et al. (2016). The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature *538*, 201–206.

83. Skoglund, P., Northoff, B.H., Shunkov, M.V., Derevianko, A.P., Pääbo, S., Krause, J., and Jakobsson, M. (2014). Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. Proc. Natl. Acad. Sci. USA *111*, 2229–2234.

84. Leslie, S., Winney, B., Hellenthal, G., Davison, D., Boumertit, A., Day, T., Hutnik, K., Royrvik, E.C., Cunliffe, B., Lawson, D.J., et al.; Wellcome Trust Case Control Consortium 2; International Multiple Sclerosis Genetics Consortium (2015). The fine-scale genetic structure of the British population. Nature *519*, 309–314.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Donor and Target individuals	[49]	https://www.internationalgenome.org/data#download
Donor and Target individuals	[50]	http://hagsc.org/hgdp/files.html
Donor and Target individuals	[11]	https://www.inmegen.gob.mx/
Donor individuals	[51]	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc= GSE21478; GEO: GSE21478
Donor individuals	[52]	http://evolbio.ut.ee/khazar/
Donor individuals	[53]	http://evolbio.ut.ee/afghan/
Donor individuals	[54]	cae31@cam.ac.uk
Donor individuals	[55]	Request to ms23@sanger.ac.uk, eleftheria@sanger.ac.uk, rotimic@mail.nih.gov, and cts@sanger.ac.uk
Donor individuals	[56]	http://evolbio.ut.ee/balkan/
Donor individuals	[57]	http://evolbio.ut.ee/slavic/
Donor individuals	[58]	http://sbimb.core.wits.ac.za/data/SNPgenotyping_01.html
Donor individuals	[59]	
Donor individuals	[16]	https://capelligroup.wordpress.com/data/
Donor individuals	[60]	http://mega.bioanth.cam.ac.uk/data/Ethiopia
Donor individuals	[17]	https://www.ebi.ac.uk/ega/home, lp.lucapagani@gmail.com
Donor individuals	[61]	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc= GSE82309; GEO: GSE82309
Donor individuals	[62]	European Genome-phenome Archive; EGA: EGAS00001000605; https://www.ebi.ac.uk/ega/home
Donor individuals	[22]	European Genome-Phenome Archive; EGA: EGAS00001002078
Donor individuals	[63]	http://www.jcvi.org/cms/research/projects/southern-african-genome-diversity-study/
Donor individuals	[64]	http://evolbio.ut.ee/malta/
Donor individuals	[20]	https://capelligroup.wordpress.com/data/
Donor individuals	[65]	https://www.ebi.ac.uk/arrayexpress/; ARRAYEXPRESS: E-MTAB-1259; https://www.ebc.uu.se/Research/IEG/evbiol/research/Jakobsson/data/
Donor individuals	[66]	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc= GSE108646; GEO: GSE108646
Donor individuals	[67]	https://evolbio.ut.ee/Tamm_2019/
Donor individuals	This paper	http://evolbio.ut.ee/
Donor individuals	This paper	N/A
Donor individuals	[68]	European Genome-Phenome Archive; EGA: EGAS00001001610
Donor individuals	[69]	https://github.com/bmhenn/khoesan_arraydata
Donor individuals	[70]	http://evolbio.ut.ee/caucasus/
Donor individuals	[24]	http://evolbio.ut.ee/turkic/
Target individuals	[33]	GEO: GSE21248
Target individuals	[26]	European Nucleotide Archive (PRJEB9080 (ERP010139) Genomic Epidemiology of Complex Diseases in Population-Based Brazilian Cohorts), under EPIGEN Committee Controlled Access mode; EGA: EGAS00001001245

(*Continued on next page*)

Continued		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
Target individuals	[71]	GEO: GSE33528
Target individuals	IlluminaControlDB (discontinued retrieved april 2013)	N/A
Target individuals	[72]	marta.alarcon@genyo.es
Software and Algorithms		
PLINK	[73]	http://www.cog-genomics.org/plink/1.9/
KING	[74]	http://people.virginia.edu/~wc9c/KING/
ShapeIT2	[75]	https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html
CHROMOPAINTER	[19]	http://www.paintmychromosomes.com/
fineSTRUCTURE	[19]	http://www.paintmychromosomes.com/
SOURCEFIND	[25]	http://www.paintmychromosomes.com/
NNLS pipeline	[76]	https://cran.r-project.org/web/packages/nnls/nnls.pdf
GLOBETROTTER	[77]	http://www.paintmychromosomes.com/
fastGLOBETROTTER		Request to Wangkumhang P. (pongsakornw@gmail.com)
Beagle 4.1	[78]	https://faculty.washington.edu/browning/beagle/b4_1.html
RFMIX	[79]	https://sites.google.com/site/rfmixlocalancestryinference/
IBDNe	[80]	http://faculty.washington.edu/browning/ibdne.html
ADMIXTURE	[31]	https://www.genetics.ucla.edu/software/admixture/

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Francesco Montinaro (francesco.montinaro@gmail.com). This study did not generate new unique reagents.

METHOD DETAILS

Analyzed data

We assembled [11, 16, 17, 20, 22, 24, 26, 33, 49–72, 81–83] a genome-wide dataset of 25,732 worldwide individuals genotyped with different Illumina platforms. Of these, 25,455 were retrieved from publicly available and controlled access resources. In order to increase our resolution in identifying the source of analyzed individuals, we added 277 samples from 35 Eurasian populations. Genotype data for 89 samples are available at http://evolbio.ut.ee/. The remaining samples will be available in dedicated future publications. More detailed information are reported in Data S1A and S1B and Figure S1. The obtained dataset was filtered using PLINK ver. 1.9 [73] to include only SNPs and individuals with a successful genotyping rate >97%, retaining a total of 251,548 autosomal markers.

We used KING to remove one random individual from pairs with kinship parameter higher than 0.0884 [74]. The final dataset was therefore composed of 17,722 individuals from 261 populations [11, 16, 17, 20, 22, 24, 26, 33, 49–72, 81–83] (Data S1A and S1B; Figure S1). Of these, 11,607 individuals belonging to 22 admixed American populations were treated as 'recipients', while the remaining 6,115 samples from 239 source populations were considered 'donors'.

QUANTIFICATION AND STATISTICAL ANALYSIS

PC Analysis

Principal Components Analysis (PCA) was performed on the final dataset using the command --pca from PLINK 1.9. The resulting plot is shown in Figure S1B.

Phasing

Germline phase was inferred using the Segmented Haplotype Estimation and Imputation tool (ShapeIT2) software [75], using the HapMap37 human genome build 37 recombination map.

Clustering of donor populations

As a first step, we clustered the individuals belonging to 'donor' populations into homogeneous groups. First, we used the inferential algorithm implemented in CHROMOPAINTER (v2) [19] to reconstruct each individual's chromosomes as a series of genomic

fragments inherited (copied) from a set of donor individuals, using the information on the allelic state of recipient and donors at each available position. Briefly, we 'painted' the genomic profile of each donor as the combination of fragments received from other donor individuals. We used a value of 288.998 for the nuisance parameters 'recombination scaling constant' (which controls the average switch rate of the HMM) Ne, and 0.00076 for the 'per site mutation rate' M, nuisance parameters, as estimated by 10 iterations of the expectation-maximization algorithm in CHROMOPAINTER. This algorithm finds the local optimum values of these parameters iterating over the data. Given the computational complexity of this process, the estimation of these two parameters was obtained by averaging the values calculated from an analysis performed on a subset of six hundred individuals from all the analyzed populations, with sample sizes mirroring the global composition of the dataset for five randomly selected chromosomes (3, 7, 10, 18 and 22).

Second, we analyzed the painted dataset using fineSTRUCTURE [19], in order to identify homogeneous clusters. We ran the software in three subsequent steps: the first, also called "greedy," infers in a fast way a rough clustering summarizing the relationships among individuals, and it is usually used when the number of samples is large (> 5000 individuals); the second, starting from the greedy clustering, performs 1 million MCMC iterations thinned every 10,000 and preceded by 100,000 burn in iterations. This generated a MCMC file (.xml) that was used, by the third run, to build the tree structure using the option --T 1 [84].

FineSTRUCTURE classified the analyzed individuals into 370 clusters (Figure S2B). In order to increase the interpretability of subsequent analysis we reduced the number of identified groups. In doing so, we iteratively climbed the tree, and lumped pairs of clusters until the minimum pairwise Total Variation Distance (TVD) estimated on the chunkcounts was lower than a given threshold. Taking into consideration the within continents variability and their relevance as sources to American populations, we applied a threshold of 0.04 for sub-saharan African, Asian and Oceanian clusters, 0.03 for North-African, Native American and North-East European clusters and 0.015 for Central, West and South European clusters. After refining, 89 clusters remained (Data S2A; Figure S2A). One cluster composed of less than five individuals was excluded from the following further analysis.

Painting of the recipient populations

We used CHROMOPAINTER, to paint each recipient individual as a combination of genomic fragments inherited by 'donor individuals' pooled using the clustering affiliation obtained as previously described, and with the same nuisance parameters inferred for the donor individuals.

Bayesian haplotype-based ancestry estimation (SOURCEFIND)

We applied a recently developed Bayesian method, SOURCEFIND [25], to estimate the ancestral composition of recipient individuals. Thus, we modeled the copying vector (obtained with CHROMOPAINTER analysis) of each admixed individual as a weighted mixture of copying vectors from the donors. We used as parameters: self.copy.ind = 0, number of total (num.surrogates) and expected (exp.num.surrogates) surrogates equal to 8 and 4 respectively; performing (total number of MCMC iterations) 200,000 iterations thinned every 1,000, and preceded by a burn in step of 50,000. Furthermore, we assigned equally-sized proportions to the surrogates (num.slots = 100). For each recipient individual, we combined 10 independent runs extracting and averaging the estimates with the highest posterior probability, weighted by their posterior probability. The efficacy and reliability of the method has been assessed for a similar scenario through an extensive simulation approach in Chacón-Duque et al. [25].

Non-Negative Least Square haplotype-based ancestry estimation

CHROMOPAINTER provides a summary of the amount of DNA copied from each donor population. We identified the most closely ancestrally related donor population for each admixed population by comparing their copying vectors to copying vectors inferred in the same way for each of the donor clusters, using a slight modification of non-negative least square (NNLS) function in R 3.5.1 [76], and following the approach reported in Montinaro et al. and Leslie et al. [14, 84]. Briefly, this approach identifies copying vectors of donor populations that better match the copying vector of recipient populations as estimated by CHROMOPAINTER. For each recipient population, we decomposed the ancestry of that group as a mixture (with proportions summing to 1) of each sampled potential donor cluster, by comparing the 'copying vector' of donor and recipient populations.

Estimation of admixture dates

In order to provide a temporal characterization of the admixture events in the Americas, we estimated times and most closely related putative sources using population-based and individual-based painting profiles.

In the "population" approach, given the high demand of computational resources requested for the analysis, we have used fastGLOBETROTTER, which, based on GLOBETROTTER [77], implements several optimizations in performance, making it suitable for large datasets. In detail, we first harnessed the painting profiles obtained by CHROMOPAINTER by testing for any evidence of admixture using the options null.ind = 1, prop.ind = 1, and performing 100 bootstrap iterations. For each of the admixture events inferred, we considered only those characterized by bootstrap values for time of admixture between 1 and 400. Subsequently, we estimated time of admixture repeating the same procedure with options null.ind = 0 and prop.ind = 1 (Data S2C).

For the individual analysis we estimated admixture times with GLOBETROTTER, applying the prop.ind = 1, null.ind = 0 approach to the 11,607 target individuals. In order to remove individuals with "unusual" painting profiles, only those falling in the 2.5%–97.5% admixture time confidence interval were retained.

We tested significant differences in times of admixture involving specific African or European clusters by applying a Wilcoxon test using R and setting alternative to "greater."

Ancestry-specific effective population size estimation

In order to estimate ancestry-specific effective population size for the 22 recipient American populations we followed the pipeline presented by Browning et al. [35] (http://faculty.washington.edu/sguy/asibdne/posted_commands.txt). The overall reliability of the method has been previously proved through extensive simulations mimicking the admixture of the Americas, also in the presence of genotyping errors and population structure [35].

We used IBD and LA inferred from genome-wide data as a first step. We inferred IBD segments using the refined IBD algorithm implemented in Beagle 4.1 [78], with the following parameters: ibdcm = 2, window = 400, overlap = 24 and ibdtrim = 12, as suggested in Browning et al. [35]. Subsequently, we ran the merge-ibd-segments.12Jul18.a0b.jar script to remove breaks and short gaps in the inferred IBD segments (gaps shorter than 0.6cM).

We estimated the local ancestry for genomic fragments in the American individuals using RFMIX [79]. As reference populations we used Yoruba (YRI), Gambia (GWDwg) and Mozambique for Africa, Chinese Han (CHB) and Japanese (JPT) for Asia, Spanish (IBS), British (GBR), and Tuscans (TSI) for Europe and Tepehuano, Wichi and Karitiana for Native American ancestry. We used "PopPhased", "-n 5" and "--forward-backward" options as recommended in RFMix manual. Then, we corrected the initial phasing following the modifications of RFMIX and using the rephasevit.py script provided by Browning et al. [35].

We combined the results from IBD analysis and LA assigning to each IBD segment the most probable ancestry.

Subsequently, we calculated the adjusted number of pairs of haplotypes for each ancestry. This is required because two haplotypes can only be in IBD with respect to a given ancestry at genomic positions if both haplotypes have that ancestry. Therefore, in a sample composed by n individuals the ancestry-adjusted number of pairs of haplotypes is equal to:

$$\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} 4p_i p_j$$

*(*where i and j are independent individuals and pi and pj are their proportions of the given ancestry).

Finally, we used the obtained "npairs" to run IBDNe software (version ibdne.07May18.6a4) [35, 80] in default mode, except for filtersample = false.

Sex-biased admixture evaluation

We intersected SNPs from the X chromosome that were present in both our main datasets and in the 1000 Genomes Project samples. Three admixed American groups (Mexican, Maya, and Mayas) were removed because they did not include any genotypes for chromosome X. We revised and imputed sex assignments based on X chromosome data using the --impute-sex command in PLINK. A male or female call is made when the rate of homozygosity is $> 80\%$ and $< 20\%$, respectively. Individuals for which sex imputation was ambiguous were removed and heterozygous SNPs in male X chromosomes were set as missing. After this step, only samples and positions with a genotyping rate $\geq 97\%$ were retained: 5,227 SNPs in a total of 15,353 individuals. The same set of individuals was extracted from the filtered autosomal dataset with 258,720 SNPs. Subsequently, we performed LD pruning (--indep-pairwise 200 50 0.2) in both X chromosome and autosomal datasets, resulting in a total of 2,519 and 116,912 SNPs, respectively. We ran separate unsupervised ADMIXTURE (version 1.3.0 [31]) analysis for the two datasets using K values = 3 and 10 independent runs. We used the option '--haploid = 'male:23' in order to properly treat male individuals and chose the best run according to the highest value of log likelihood. Finally, we performed paired Wilcoxon tests in order to test for significant differences between the ancestry proportions observed in the autosomes versus the X chromosome and used Bonferroni correction for multiple-testing (adjusted p value < 0.05). We evaluated similarities in the autosomal/X chromosome ratio distribution by applying a Wilcoxon distribution, and reported the p value in Figure 3 and Data S2D.

DATA AND CODE AVAILABILITY

The genotype data for 89 samples published in this study are available at http://evolbio.ut.ee/. The remaining 188 samples will be available in dedicated future publications.