Article

Eight millennia of continuity of a previously unknown lineage in Argentina

https://doi.org/10.1038/s41586-025-09731-3

Received: 14 April 2025

Accepted: 9 October 2025

Published online: 05 November 2025

Check for updates

Javier Maravall-López[1,2,3] ✉, Josefina M. B. Motti[4], Nicolás Pastor[5,6,7], María Pía Tavella[8], Mariana Fabra[5,7,8], Pilar Babot[9,10], Mariano Bonomo[11], Silvia E. Cornero[12], Guillermo N. Lamenza[13], Diego Catriel Leon[14,15], Paula C. Miranda de Zela[16], Gustavo G. Politis[11,17], Sofía C. Angeletti[18,19], G. Roxana Cattáneo[5,7,8], Mariana Dantas[5,7,8], Hilton Drube[20,21], Lucia G. Gonzalez Baroni[9,10], Salomón Hocsman[9,10], Andrés D. Izeta[5,7,8], Reinaldo A. Moralejo[11], Verónica Aldazabal[22], Diego M. Basso[23], Cristina Bayón[24], María Guillermina Couso[11,25], Ulises D'Andrea[26], Paula Del Río[12], Germán G. Figueroa[7,8], Romina Frontini[24], Mariela Edith Gonzalez[17], Andrés G. Laguens[5,7], Jorge G. Martínez[10], Pablo G. Messineo[17], Beatriz Nores[26], Daniel E. Olivera[12,27], Gisela M. Sario[5,7,8], Analía Sbattella[28,29], Clara Scabuzzo[30], Aldana M. Tavarone[7,8], Rodrigo Vecchi[24], Kim Callan[31,32], Ella Caughran[31,32], Oscar Estrada[33], Trudi Frost[31,32], Lora Iliev[31,32], Aisling Kearns[31], Jack Kellogg[31,32], Kim-Louise Krettek[34], Ann Marie Lawson[31,32], Matthew Mah[3,31,32], Nihal Manjila[31,32], Adam Micco[31], Iris Patterson[31,32], Lijun Qiu[31,32], Xavier Roca-Rada[35,36,37], Gregory Soos[31,32], Peter A. Webb[35], J. Noah Workman[31,32], Nadin Rohland[31], Nick Patterson[1,3], Iosif Lazaridis[1,31], Lars Fehren-Schmitz[38,39], Cosimo Posth[34,40], Bastien Llamas[35,41,42,43], Swapan Mallick[3,31,32], Darío A. Demarchi[5,7,8], Graciela S. Cabana[44], David Reich[1,3,31,32,45] ✉ & Rodrigo Nores[5,7,8,45] ✉

The central Southern Cone of South America was one of the last regions of the globe to become inhabited by people[1], and remains under-represented in studies of ancient DNA. Here we report genome-wide data from 238 ancient individuals spanning ten millennia. The oldest, from the Pampas region and dating to 10,000 years before present (BP), had distinct genetic affinity to Middle Holocene Southern Cone individuals, showing that differentiation from the central Andes and central east Brazil had begun by this time. Individuals dating to 4,600–150 BP primarily descended from a previously unsampled deep lineage of which the earliest representative is an individual dating to around 8,500 BP. This central Argentina lineage co-existed with two other lineages during the Mid-Holocene and, within central Argentina, this ancestry persisted for thousands of years with little evidence of inter-regional migration. Central Argentina ancestry was involved in three distinct gene flows: it mixed into the Pampas by 3,300 BP and seemingly became the main component there after 800 BP, with central Andes ancestry in northwest Argentina, and with tropical and subtropical forest ancestry in the Gran Chaco. In northwest Argentina, there was an increased rate of close-kin unions by 1,000 BP, paralleling the pattern in the central Andes. In the Paraná River region, a 400 BP individual with a Guaraní archaeological association clusters with Brazilian groups, consistent with Guaraní presence by this time.

The peopling of South America likely followed both the Pacific and Atlantic coasts[1,2]. Genetic differentiation is detectable in ancient genomic data after 9,000 BP in at least three main clusters: central Andes, tropical/subtropical forest or lowlands (including Amazonia), and central Chile, Patagonia and Pampas[3,4]. However, current sampling has major gaps. We focus on the poorly sampled central Southern Cone (CSC)—the territory of central and northern Argentina comprising the Andean mountains in the west to the eastern fluvial plains and southern grassland plains. The CSC has diverse biogeographical regions that we divide for analysis into northwest Argentina (northern and southern Puna, Prepuna and sub-Andean Valleys, including Belén, Aconquija, Hualfín and Ambato); central Argentina (hills, including the southern Pampean Hills of Córdoba and San Luis provinces (collectively called Córdoba Hills here), and plains, including the Laguna Mar Chiquita region, east Córdoba, and the Santiagueña Plains); Gran Chaco (dry and humid); the Paraná River and the adjacent alluvial plains (Middle Paraná–Salado Rivers, upper delta and lower delta); and Pampas (Central Pampean Dunefields, southern Pampas (including Interserrana and Pampas south) and south Salado River). We also studied an ancient individual from Pantanal in present-day Paraguay (Fig. 1a). Our sampling is influenced by the intensity of archaeological research and available samples, providing more resolution in some regions than in others.

A list of affiliations appears at the end of the paper.

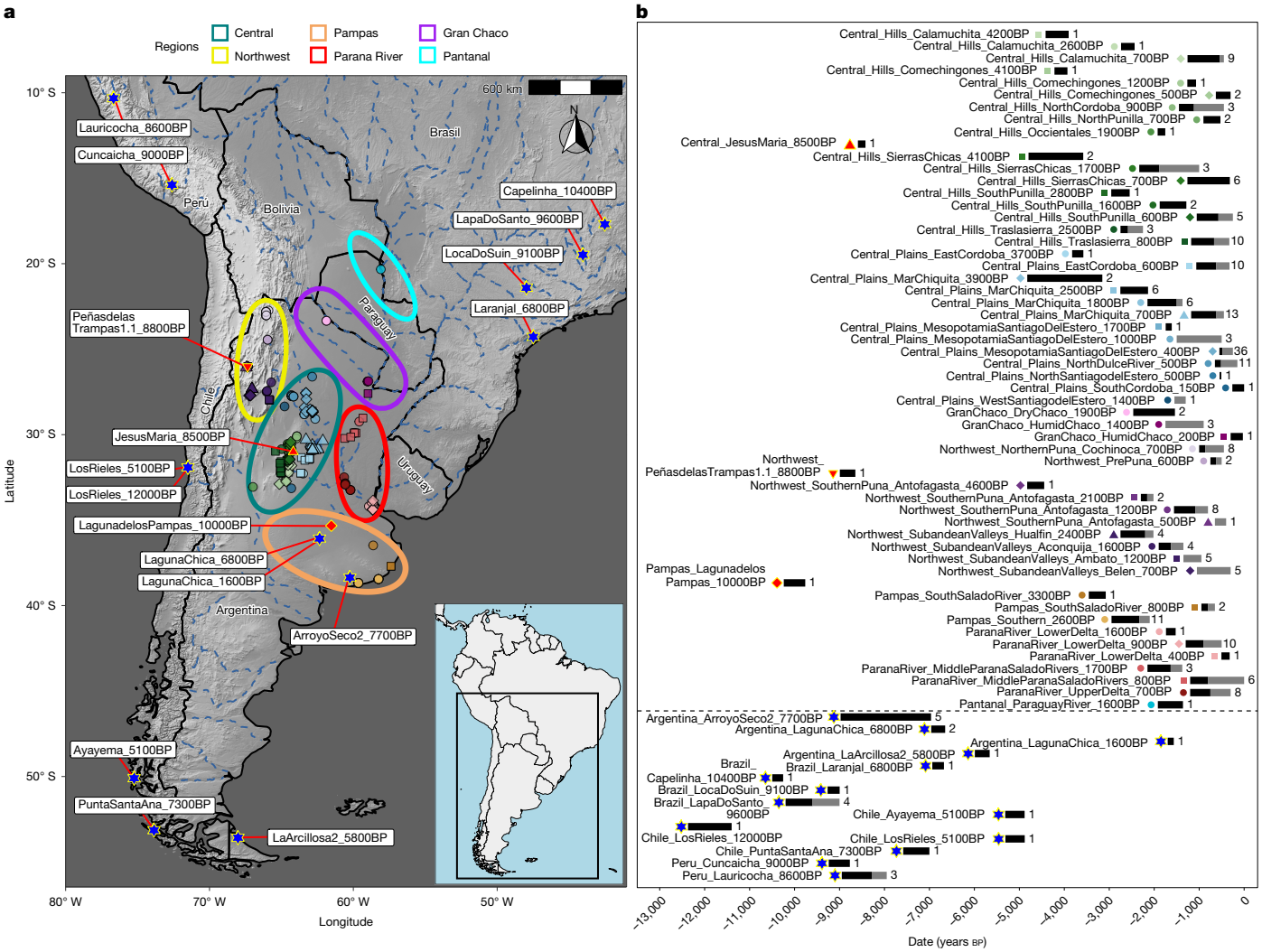


Fig. 1 | Overview of geographical and temporal sampling. a, The geographical distribution of newly reported (black edges) and selected previously published (golden edges) early South American ancient individuals. The map was created in R using open-source data (Methods). **b**, The temporal distribution of newly reported and selected published (below the dashed line) ancient individuals. For each grouping, the number at the right end of the bar indicates the sample size, and the dark fill of the bar indicates the proportion with a direct radiocarbon date.

The CSC has been inhabited since the late Pleistocene, and archaeological research documents multiple influences from the central Andes and the Lowlands[5–8]. The earliest widely accepted site is Arroyo Seco 2 (14,000 BP; all dates throughout are calibrated), in the Pampas. From the late Pleistocene and Early Holocene (13,000–8,200 BP), human presence is well documented in the Pampas, the Puna in northwest Argentina, and the Córdoba Hills in the central region[9]. From 13,300 BP to 11,200 BP, several sites from the Southern Cone are characterized by fishtail projectile points, of which the wide distribution has been interpreted as a signal of a rapid migration across South America, paralleling inferences from ancient genomes[2,10,11].

Humans expanded into a wider range of CSC environments in the Middle Holocene (8,200–4,200 BP). Nevertheless, some areas, such as the Gran Chaco, the central plains and the Paraná River, show less evidence of settlement in this period (Supplementary Information 1–6). These changes occurred at a time of increased temperature known as the Mid-Holocene Hypsithermal[12]; however, the consequences of those environmental fluctuations varied across regions, which may help to explain the uneven distribution of archaeological sites[13,14]. Around 4,500 BP, there was a transition away from hunting and gathering as the sole means of subsistence in the Puna and valleys of northwest Argentina[15].

In the Late Holocene (after 4,200 BP), the CSC harboured communities that ranged from sedentary agropastoralists in the northwest who hunted, foraged and exchanged goods from several ecoregions over long distances through llama caravans[16]; semi-sedentary horticulturists in the Córdoba Hills[17,18] who, in the central plains and Paraná River, adapted to fluvial environments[19–21]; and nomadic hunter–gatherers in the Pampas and Gran Chaco[22–24]. Ethnographic records document wide cultural variation in the CSC at the time of European contact[25,26]: Comechingones (Hênîa and Kâmîare) in the Córdoba Hills; Sanavirones in the Laguna Mar Chiquita area; Diaguitas speaking Cacan in the sub-Andean Valleys; Atacamas speaking Kunza in the Puna; Tonocotés in the Santiagueña Plains; Lules in northwest Santiago del Estero; Chaná-Timbú in the Middle Paraná-Salado shores and Paraná Delta; Guaraní groups speaking Tupí-Guaraní languages who probably arrived by around 700 BP in the Paraná Lower Delta[27]; Wichí speaking a Mataco-Mataguaya language in the southern Gran Chaco; and, in the same area, Mocovíes and Qom (Toba) speaking a Guaycurú language. The introduction of horses and cattle brought about profound changes in the economy and mobility of the Indigenous peoples of the Pampas and Patagonia[24]. Some scholars postulate that the southern Pampas was previously inhabited by groups related to Chon-speaking Patagonian Tehuelches[28]. In the northern Pampas,

Querandí groups were mobile hunter–gatherers whose linguistic affiliation is unclear.

To characterize the genetic structure of the CSC in the Early Holocene, and to test for gene flow and demographic differences across subregions, we screened 344 bone or tooth samples from 310 individuals up to 10,000 BP. The Supplementary Information contains descriptions of Supplementary Data 1–14 (online tables that provide details of these samples and the analyses performed), as well as descriptions of Supplementary Figs. 1–84, and text sections that present archaeological context (Supplementary Information 1–7) and genetic analyses (Supplementary Information 8–13).

We enriched ancient DNA libraries for more than 1.2 million targeted single-nucleotide polymorphisms (SNPs), and added to this off-target sites (not originally targeted by the enrichment protocol but commonly captured because of proximity to targeted SNPs) to arrive at a set of roughly 2 million analysed SNPs (Methods). We obtained new genome-wide data passing quality control from 238 ancient individuals (Fig. 1a,b), with a median of 659,011 SNPs covered at least once (207 individuals with at least 50,000 SNPs covered; Supplementary Data 1). We co-analysed the newly reported individuals with previously reported data for 588 pre-European contact Native/Indigenous Americans (Extended Data Fig. 1 and Supplementary Data 1) using the curation provided by the Allen Ancient DNA Resource (Methods). We defined 'pre-European contact Native/Indigenous American individuals' as those with a date point estimate (from direct radiocarbon dating or archaeological context) before 600 BP. We also included SNP array data from present-day Native Americans[2], restricting to sites intersecting the '1240k' set.

Distinctive genetic drift by 10,000 BP

To understand how the oldest individual, Argentina_Pampas_LagunadelosPampas_10000BP (hereafter, LagunadelosPampas_10000BP) relates to other Early/Middle Holocene South Americans, we computed f_4-statistics of the form (Supplementary Data 2):

$$f_4(\text{Outgroup, Pop1, Pop2, Pop3}), \quad (1)$$

which should not deviate significantly from zero if Pop2 and Pop3 are a true clade (descended without mixture from a common ancestral population) with respect to Pop1. A violation of this test—for which deviation from zero can be expressed as an approximately normally distributed Z-score computed using a genomic block jackknife—indicates a wrong phylogeny or a history that involves gene flow among the tested lineages. These statistics reveal shared drift among LagunadelosPampas_10000BP and Argentina_Central_JesusMaria_8500BP (hereafter, JesusMaria_8500BP), the individuals from southern Patagonia (5,100–7,300 BP) and those from the Argentinian Pampas (7,700–6,800 BP), with respect to both early individuals from the central-east of Brazil (10,400–6,800 BP) and the central Andes (9,000–8,600 BP) (Fig. 2a).

All pairs of JesusMaria_8500BP, southern Patagonia (5,100–7,300 BP) and Argentinian Pampas (7,700–6,800 BP) are symmetrically related to LagunadelosPampas_10000BP, up to the limits of our resolution for statistics unaffected by biases due to using different sequencing technologies (Fig. 2a, Supplementary Information 9 and Supplementary Data 2). The most plausible explanation is that LagunadelosPampas_10000BP belonged to an ancestral Southern Cone population that split from central east Brazil and central Andes groups by 10,000 BP and was geographically in the CSC by that time before differentiating into distinct components. Neither PeñasdelasTrampas1.1_8800BP, from southern Puna in northwest Argentina, nor LosRieles_5100BP from central Chile, showed affinity to LagunadelosPampas_10000BP, so we could not make a definitive statement about their relationship to this individual.

We evaluated the affinities of LagunadelosPampas_10000BP to Anzick, a 12,500 BP individual from present-day Montana, USA, with distinctive genetic affinities to early South Americans relative to later ones[11]. Chile_LosRieles_12000BP showed the strongest affinity ($|Z| < 4.1$), followed by weaker affinity with LagunadelosPampas_10000BP ($|Z| < 2.6$) (Extended Data Fig. 2 and Supplementary Data 2). However, as these three individuals were positioned together as a clade in an outgroup-f_3 neighbour-joining tree (Supplementary Fig. 1), both probably harboured a distinct Anzick-related genetic component. Affinity with Anzick in early South America, and the absence thereof, has been associated with at least two independent migration waves and population replacement[11]. However, the fact that LagunadelosPampas_10000BP also exhibits excess allele sharing with later Southern Cone individuals without a significant genetic affinity towards Anzick, suggests that this individual may have been admixed between a basal Southern Cone lineage and a basal Anzick-associated lineage, and these Anzick-related lineages may therefore not have been completely replaced[29].

We re-examined several other claims of complex relationships between Central and South Americans, studying evidence of asymmetrical relatedness to Mesoamerican-related populations among late Middle Holocene individuals from central Chile and the central Andes[11,30] (Supplementary Data 2). Using qpAdm (Methods), we modelled Chile_LosRieles_5100BP as a mixture of 16.2 ± 3.3% Mesoamerican related and the rest Brazil_LapaDoSanto_9600BP related (Supplementary Information 9 and Supplementary Data 3). However, while asymmetrical relationships to Mesoamerican populations have been interpreted as evidence of a third ancestry movement into the subcontinent, in addition to the differential affinity to Anzick[11], we cannot reject a simple two-source model of diverse early South American populations using qpWave (Methods) ($P > 0.12$) (Supplementary Data 2). This supports the theory that asymmetrical relatedness to Anzick may be better explained by a model of structure on a gradient than two independent pulses[29], with the structured populations differentially related not only to Anzick but also to Mesoamericans.

Affinity between late central Andes individuals and ancient Californians has been interpreted as evidence of a fourth migration pulse into South America[11]. However, late central Andes individuals show stronger genetic affinity to ancient Caribbean individuals than to ancient Californians (Supplementary Data 2) when compared to early central Andes individuals ($Z = 6$). Recent research has documented south-to-north migration in Central America[31], and that California attraction is detectable only when considering Californian populations with Mexican-related gene flow[32]. Thus, the late central Andes signal is plausibly driven by interactions within South America and back-migration spreading up to California.

Three deep lineages in the Mid-Holocene

We combined published data with three individuals dated to before 8,500 BP: LagunadelosPampas_10000BP (Pampas), PeñasdelasTrampas1.1_8800BP (northwest Argentina) and JesusMaria_8500BP (central Argentina) (Fig. 1a,b). Using f_4-statistics, we identified four possible clades of South American Early/Middle Holocene individuals: Brazil, central Andes, Pampas and Southern Patagonia[3,11,29,30,33,34] (Fig. 2b and Supplementary Information 9).

We merged these putative clades into common labels and combined them with remaining individuals that were not identified as part of any clade for automatic population history model exploration. We used the finds_graphs function of ADMIXTOOLS2, which evaluates randomly perturbed admixture graphs until the resulting graph cannot be made to better fit the data. As this search gets trapped in local optima, we performed 100 independent iterations, each starting from a randomly initialized graph, to explore the diversity of equally well-fitting models. We found no evidence that models involving admixture events fit the data significantly better than ones without mixture (Supplementary

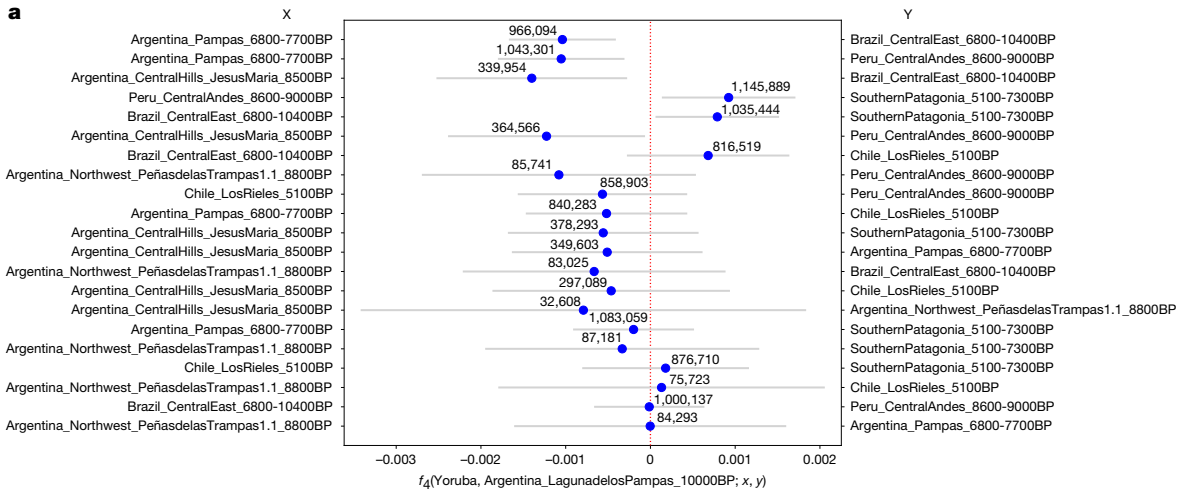


Fig. 2 | Relationships among deep South American lineages. a, The affinities of LagunadelosPampas_10000BP to Early/Middle Holocene South Americans were quantified by f_4 statistics. The bars denote 95% confidence intervals (CIs) (1.96 × s.e.m.) around the mean across genomic-block jackknife pseudoreplicates (f_4 point estimates). The only significantly non-zero statistics (top 6) indicate excess allele sharing with Middle Holocene Southern Cone individuals, with respect to both early individuals from the central east of Brazil (10,400–6,800 BP) and the central Andes (9,000–8,600 BP). At the same time, LagunadelosPampas_10000BP appears symmetrically related to all three of Southern Cone groupings up to the limits of our resolution. The number of SNPs used for each test is shown above each point estimate in the figure. **b**, Distinct lineages in South America by the Middle Holocene. Clades were established using a combination of cladality tests and automatic exploration of population history models. We represent lineages for which we could not robustly favour a particular splitting order as a politomy. Newly reported individuals are shown in bold, and thin evidence for some clades is indicated by star symbols. The square symbol indicates that we detected affinity for Mesoamerican-related populations. We found no evidence of mixture events fitting the data significantly better, although this could be a reflection of low statistical power. LagunadelosPampas_10000BP is absent from the tree owing to its ambiguous positions across well-fitting models.

Information 9 and Supplementary Data 2), and we therefore examined only the nine unique best-fitting models with no admixture (Supplementary Data 2 and Supplementary Figs. 2–10; range of scores, 34.1–43.3; worst residuals, 2.9–4.8). For all of these models, many internal branches had a drift value of either 0, indicating an inability to discern the order of splits, or 1–2, indicating weak support for a branch.

All nine models include a clade with PeñasdelasTrampas1.1_8800BP and central Andes (9,000–8,600 BP), which also agrees with an outgroup-f_3 tree (Supplementary Fig. 1). Eight of the nine support a clade of Chile_LosRieles_5100BP and Middle Holocene Argentinian Pampas (7,700–6,800 BP), with the exception of the worst-fitting one (Supplementary Fig. 8). While none of these clades are rejected by f_4-statistics (Supplementary Data 2), the clades are also not confident, owing to the low inferred drift ancestral to them. The placement of LagunadelosPampas_10000BP was more ambiguous, appearing as an isolated lineage (three models) or grouped with the central Argentina JesusMaria_8500BP (five models) or the Middle Holocene Argentinian Pampas (7,700–6,800 BP) (one model), consistent with its basal position in CSC diversity.

Our results indicate that the CSC harboured at least three deep lineages: a lineage represented by PeñasdelasTrampas1.1_8800BP that appears cladal with the main ancestry component present in the central Andes since 9,000 BP[11,33]; a lineage occupying the Pampas in the Middle Holocene[11], whose earliest representatives are ArroyoSeco2_7700BP; and a central Argentina lineage, whose earliest sampled individual is JesusMaria_8500BP (Fig. 2b and Supplementary Information 9).

Ancestry landscape of the Late Holocene

We computed outgroup-f_3 statistics, measuring shared drift between pairs of populations up to the split from a common ancestor; we use

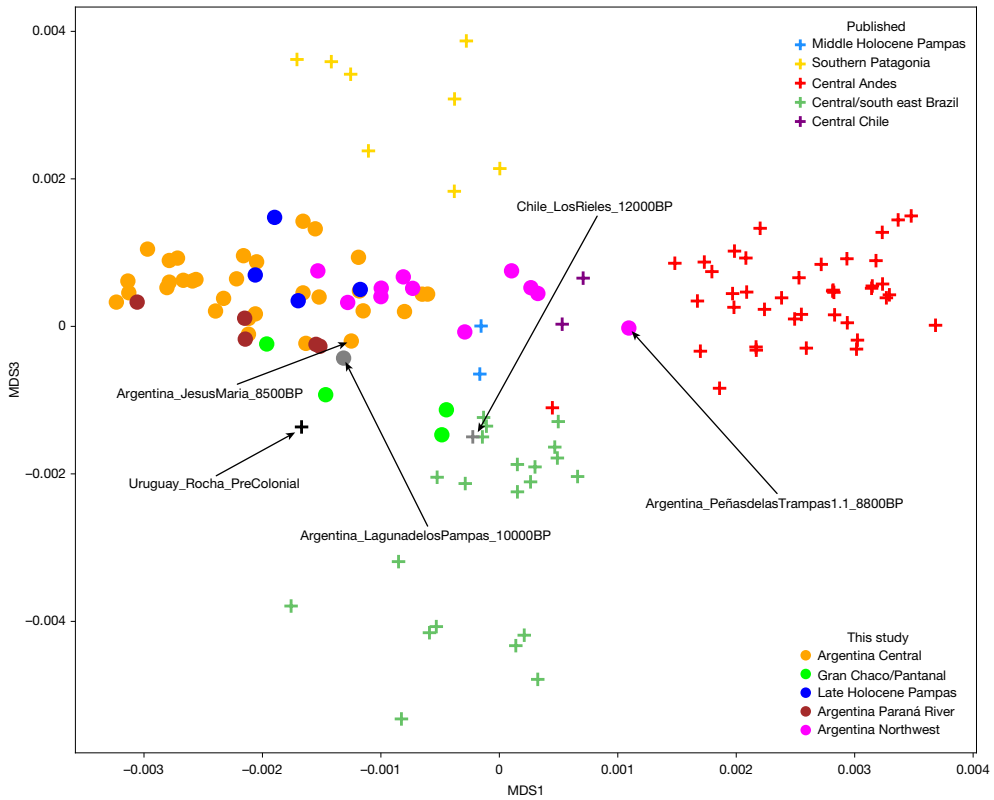


Fig. 3 | Genetic substructure in South America. MDS1 × MDS3 plot of an outgroup-f_3 distance matrix of the form $1/f_3$(Pop1, Pop2; Yoruba), where Pop_i, $i \in \{1, 2\}$ is a newly reported or previously published ancient American context label from present-day Argentina, Chile, Brazil, Uruguay, Peru, Bolivia or Paraguay. We found this more informative than plotting the first against the second component because, in that case, Patagonian populations appeared interspersed with Brazilian populations. Populations sampled in present-day USA, Mexico, Belize, Venezuela and the Caribbean were removed from the plot, as they appeared very distant to the newly reported populations (Supplementary Fig. 11). We caution against over-interpreting the position of the oldest individuals, such as LosRieles_12000BP, who may simply lack much shared drift with the rest.

the inverse as a measure of genetic distance. Dimensionality-reduction techniques such as multidimensional scaling (MDS), developed for distance-based settings, are useful for visualizing affinities. Figure 3 shows the first and the third component of this MDS analysis (Supplementary Fig. 11), in which most new samples form a cluster that includes the oldest central Argentinian, JesusMaria_8500BP. The horizontal axis differentiates central Andes (right) from central Argentina (left); and the vertical axis differentiates southern Patagonia (top) from central east Brazil (bottom). A neighbour-joining tree produces similar patterns (Supplementary Fig. 1).

Late Holocene populations from the northwest are shifted toward central Andes groups, hinting at admixture. In the neighbour-joining tree, the 700–600 BP individuals from northern Puna and Prepuna fall in the central Andes cluster, closest to Bolivian populations. Individuals from the Gran Chaco and Paraguay Pantanal regions shift towards or fall within the cluster of central east Brazilian populations, but not so the 200 BP Gran Chaco individual, who clusters with central Argentina. A 400 BP individual with a Guaraní archaeological association from the Paraná River region also appears in this cluster, probably reflecting the Guaraní expansion[27], but data are too sparse for ancestry component modelling (Supplementary Data 1). All of the remaining samples clustered, with imperfect but consistent separation between Pampas, northwest Argentina, Paraná River and central Argentina individuals, mirroring F_{st} hierarchical clustering (Extended Data Fig. 3 and Supplementary Fig. 12).

To test for genetic affinities, we computed f_4(Outgroup, P1; P2, P3), where P2 represents early Middle Holocene South Americans, P3 represents groups from the study subregion, and P1 represents other ancient groups (Supplementary Data 4). The great majority of CSC individuals show affinity to southern Patagonia, central Andes and Middle Holocene Pampas compared with central-east Brazil, implying that Brazil is probably the deepest split (Supplementary Data 4). Applying a false-discovery rate (FDR) for clade rejection at FDR < 0.05 using the Benjamini–Yekutiel procedure (Z_{BY}; Methods), we highlight six observations (Supplementary Data 4 and Supplementary Information 11). First, northern Puna and Prepuna individuals shared alleles at an excess rate with people of the central Andes ($2.9 < Z_{BY} < 5$), and other northwest Argentina groups have evidence of admixture between central Argentina and central Andes sources (Extended Data Table 1). Second, Late Holocene individuals from central Argentina attract others from the same region ($3 < Z_{BY} < 27.1$) and are a clade with JesusMaria_8500BP, except for excess sharing with Mexicans and ancient Californians ($3 < Z_{BY} < 3.6$) in Argentina_Central_Hills_Calamuchita_4200BP and later, but with no evidence for an increasing trend with time (Supplementary Data 5). This points to a demographic process connecting lower North America all the way to the CSC; although we do not have sufficient sampling from 8,500–4,200 BP to identify the likely sources, it is plausibly the same process that induced Mesoamerican affinity in Chile_LosRieles_5100BP. Third, the Late Holocene individuals from the Paraná River region shared drift with central Argentina ($3 < Z_{BY} < 16.3$). Fourth, individuals from the Gran Chaco, including the 1,400 BP individual from the El Cachapé complex, share alleles with modern groups from the same region, such as Chané, Wichí, Guaraní or Toba ($3 < Z_{BY} < 6.9$); the Paraguay Pantanal individual at 1,600 BP shows a similar signal despite separation by more than 800 km, supporting a 'Chaco-Pantanal' archaeological connection[35]. Fifth, modern Gran Chaco populations are admixed between a central Argentina and a tropical/subtropical forest source (Extended Data Table 1). Sixth,

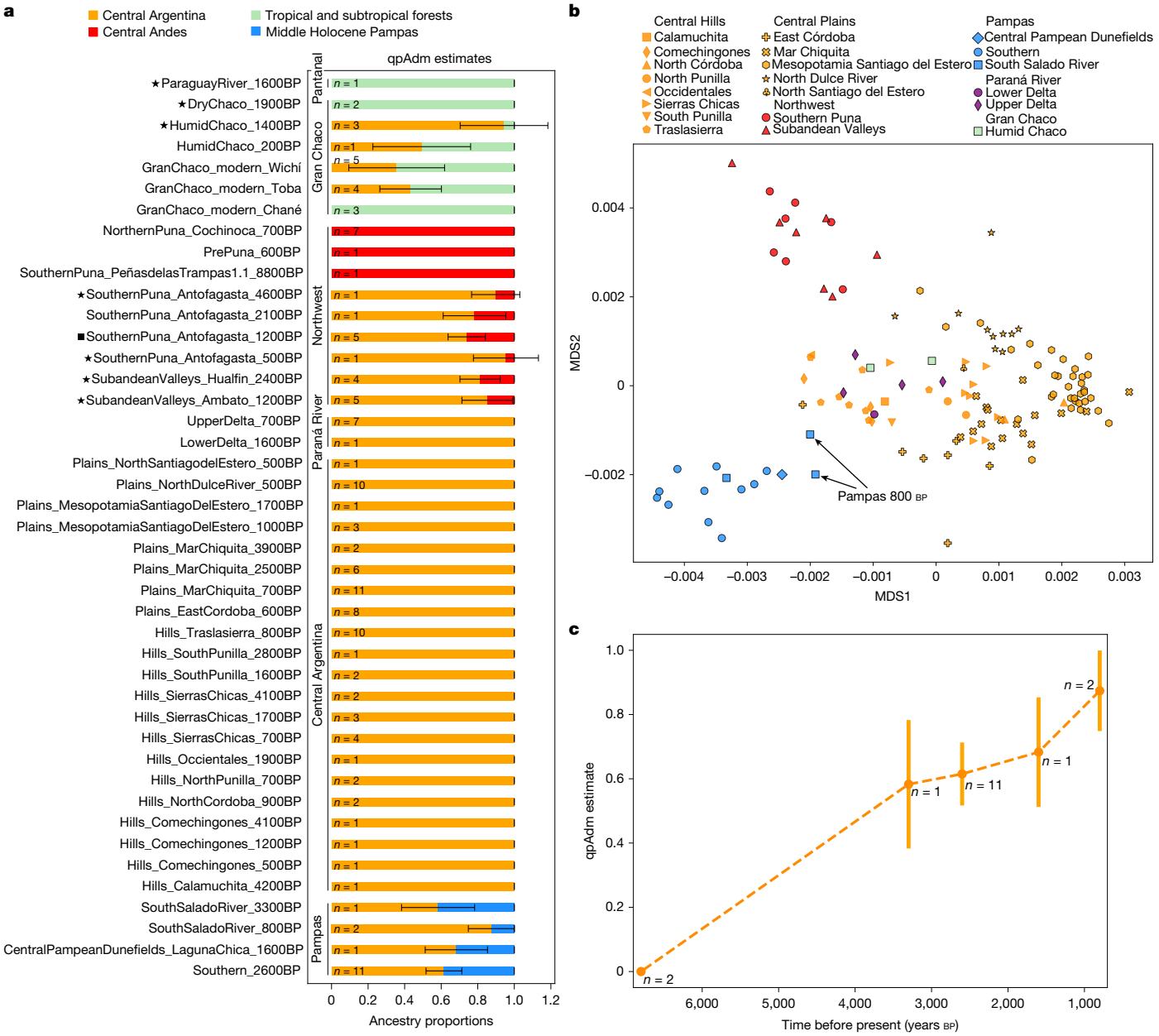


Fig. 4 | Ancestry modelling and fine-scale structure within the CSC reveal three distinct admixture processes. a, qpAdm ancestry component estimates for selected groupings. The bars denote the 95% CIs (1.96 × s.e.m.) around the mean across genomic-block jackknife pseudoreplicates (point estimates). The star symbols indicate instances in which a central-Argentina-only model was also found to fit the flagged grouping label. The square symbol indicates an instance in which a central-Andes-only model was also found to fit the flagged label (details are provided in Supplementary Information 12). Inferences for Gran Chaco and Pantanal were more ambiguous, owing to low sample sizes and coverages. The number of individuals within each grouping is shown within each horizontal bar in the figure. **b**, Fine-scale genetic structure in central Argentina. MDS1 × MDS2 plot of a distance matrix of the form $1/f_3(I_1, I_2;$ Yoruba), where I_i, $i \in \{1, 2\}$ is an individual from a context label estimated to carry primarily central Argentina ancestry. This low-dimensional decomposition revealed two axes of variation, which can be interpreted, in light of the qpAdm results (Fig. 4a), as resulting from admixture between three poles of ancestry: central Argentina, central Andes and Middle Holocene Pampas. Overall, we observe geographically driven clustering maintained over thousands of years. **c**, qpAdm estimates of central Argentina ancestry in the Pampas region over time. The bars denote the 95% CIs (1.96 × s.e.m.) around the mean across genomic-block jackknife pseudoreplicates (point estimates). The 6,800 BP datapoint corresponds to individuals from the LagunaChica site, who appear to be a clade with the 7,700 BP Arroyo Seco individuals (Middle Holocene Pampas). Central Argentina ancestry in the Pampas increased (two-sided $P = 0.0014$ from a Z-test for a significant difference in central-Argentina ancestry proportions in SouthSaladoRiver_800BP with respect to Southern_2600BP). This suggests multiple waves of admixture or continuous gene flow from central Argentina into the Pampas. The number of individuals within each grouping is shown next to each point estimate in the figure.

individuals from the Pampas share drift both with others from the same region ($3 < Z_{BY} < 15.4$) and with central Argentina compared with the Middle Holocene Pampas ($3 < Z_{BY} < 9.8$), with direct evidence of admixture in Late Holocene Pampas (Southern_2600BP and LagunaChica_1600BP) (Extended Data Table 1).

To quantify admixture, we used qpAdm (Methods) (Fig. 4a and Supplementary Data 6–12). We examined what groups were consistent with being simple clades or two-way mixtures of the relevant deep South American lineages (central Argentina, central Andes, Middle Holocene Pampas and tropical/subtropical forest), cyclically assessing models

with respect to the other sources and more distant outgroups, and adding complexity to failing single-source models if needed (Supplementary Information 12). Admixture results were less informative, but shared some broad patterns with the qpAdm conclusions (details are provided in Supplementary Information 13 and Supplementary Figs. 13–17).

Fine structure in central Argentina

We compared the genetic affinity of selected Late Holocene central Argentina populations from 4,200 to 150 BP with the earliest central Argentina individual, JesusMaria_8500BP, with respect to other Early and Middle Holocene South Americans. f_4-statistics are positively skewed, showing excess allele sharing with JesusMaria_8500BP (Extended Data Figs. 4–6 and Supplementary Figs. 18–24) ($Z < 5.54$). Most individuals from central Argentina were consistent with being genetically homogeneous (Supplementary Fig. 81), suggesting continuity in central Argentina going back more than eight millennia, and persisting until at least 150 BP. This extends previous findings based on ancient mitochondrial DNA that detected deep, locally specific mtDNA clades in central Argentina[36]. When we analysed modern admixed central Argentinian individuals[4], we found the same pattern of f_4-statistic skew towards Late Holocene central Argentina individuals (Extended Data Fig. 7; although Early/Middle Holocene comparisons were underpowered, owing to the small overlap between the SNP sets; Supplementary Figs. 25–70), suggesting that the ancestry component represented by JesusMaria_8500BP is the main Native American lineage in the region up to the present day. However, modern individuals that previous work[4] labelled as belonging to the 'central western Argentina' lineage (Calingasta and Río Grande) actually appear genetically closest to ancient individuals from central Chile, Middle Holocene Pampas and Southern Patagonia (Supplementary Figs. 34 and 55), and are therefore not reflecting the deep lineage represented by JesusMaria_8500BP that we characterize here.

To obtain a fine-grained picture of the evolution of the central Argentina lineage, we computed an outgroup-f_3 distance matrix between all pairs of individuals from groupings that were inferred to carry majority central-Argentina-type ancestry (Fig. 4a). We find two axes of variation in Fig. 4b resulting from admixture of the three ancestry poles central Argentina, central Andes and Middle Holocene Pampas. The persistence of these clines for thousands of years with no individuals clustering outside their region suggests isolation by distance, undisrupted by further pulses of cross-regional migration.

We also observed a separation between the Córdoba Hills and the central plains, where we have particularly dense sampling, indicating geographical substructure even at this fine level as also seen in mitochondrial DNA[37]. This is consistent with distinct material culture, diet, physical activity and mortuary practices over the past two millennia between groups that inhabited the Córdoba Hills and the Laguna Mar Chiquita region[21].

Interactions with central Argentina

People of northwest Argentina (northern Puna and Prepuna) in the past millennium were genetically indistinguishable from central Andes individuals. But other northwest groups showed a mostly central Argentina background (Fig. 4a). Northern Puna individuals shared more alleles with Late Holocene groups from Bolivia than with PeñasdelasTrampas1.1_8800BP (Extended Data Fig. 8). Thus, while central Andes ancestry in northwestern Argentina has a deep history, interactions with the southern central Andes continued. The northwest individual dated to 4,600 BP had suggestive, but not unambiguous, evidence of admixture (Fig. 4a and Supplementary Data 12), suggesting these interactions may have been in place by this date. The evidence of central Argentina ancestry in the northwest is paralleled by archaeological evidence linking peoples in Puna, Sub-Andean Valleys and Santiagueña Plains[38] (Supplementary Information 2 and 6). The male individual Northwest_SubandeanValleys_Belen_ElShincaldeQuimivil_500BP, buried within an Inca site, had artifacts indicating a potential central Andean origin, which were interpreted as evidence of relocation in his lifetime under the Mitmaqkuna Inca practice[39]. However, his genetic background is not significantly different from that of other sub-Andean Valley individuals from the same grouping (Supplementary Data 13), so there is in fact no genetic evidence that this individual was a migrant.

Gran Chaco and Pantanal history could be explored only roughly with our data owing to low sample sizes and poor data quality. However, f_3-based analyses cluster them with Brazilian groups, so they are unlikely to have had central-Argentina-type ancestry alone. For Gran Chaco individuals dating to 200 BP (HumidChaco_ElChancho_200BP, clustering with central Argentina in an outgroup f_3-tree) or later (including present-day Toba and Wichí[2]), the only robust model supports a mixture of central Argentina and Amazonian-related sources (Fig. 4a and Extended Data Table 1a). Indeed, most ancient individuals from the Gran Chaco showed significant affinity to modern counterparts, indicating some continuity over two millennia (Supplementary Data 4). The major exception was the Chané, who belong to the Arawak linguistic family and are thought to have migrated more recently to the Gran Chaco and mixed with Chiriguanos (from the Guaraní ethnolinguistic group), and had no evidence of central-Argentina-type ancestry.

Individuals of the Paraná River region showed affinity with central Argentina in f_4-statistics. In fact, most analyses were consistent with these individuals' being simple clades with central Argentina, and failures of this clade test were plausibly due to data artifacts (Supplementary Information 12, Supplementary Data 4 and Supplementary Fig. 82). This finding aligns with archaeological links between the Paraná River region, the Córdoba Hills and the Laguna Mar Chiquita (central plains) populations in the Late Holocene[7,20,21]; other archaeological evidence links the Middle Paraná-Salado Rivers and Santiagueña plains[40,41]. Some Paraná River individuals were associated with the Goya-Malabrigo archaeological complex, characterized by zoomorphic appendages in pottery, earth mound construction and a riverine horticulture subsistence strategy[42]. It has been hypothesized that these traits are a signal of Arawak ethnolinguistic groups spreading along eastern South American rivers[43]. We explored this possibility by comparing the newly reported data with the limited Arawak-related data currently available, that is, both ancient Arawak-associated people from the Ceramic-period Caribbean and modern representatives (Piapoco from northern South America and the geographically closer Chané from Gran Chaco). As we did not find any genetic signal of a specific affinity (Supplementary Data 13), our results do not provide evidence of a large-scale Arawak migration. Arawak influence in the Paraná River region could have been mediated by a small number of individuals or by cultural transmission[42]. Alternatively, a large-scale migration could have occurred, and the absence of the Arawak signal in the Paraná groups could reflect incomplete representation of genomic diversity of Arawak-speaking groups among available samples.

Pampas region individuals from around 6,800 BP do not show affinity with the central Argentina lineage when compared to 7,700 BP Pampas individuals from Arroyo Seco 2 (Supplementary Data 4). Thus, the Arroyo Seco 2 lineage persisted in the region for at least a thousand years without detected interaction with the neighbouring central Argentina lineage. However, Late Holocene Pampas individuals cannot be modelled as a simple clade with the Middle Holocene Pampas or Middle Holocene central Argentina lineages (Extended Data Table 1 and Supplementary Fig. 80). By 3,300 BP, Pampas individuals fit as a mixture of the Middle Holocene central Argentina (58 ± 10%; Fig. 4c and Supplementary Data 12) and Middle Holocene Pampas lineages. Owing to limited sampling, we can place only a lower bound on the beginning of this southward spread of central Argentinian

ancestry at 3,300 BP; we attempted to estimate a date for this mixture (Methods), but it was too noisy. Central Argentina ancestry in the Pampas also continued to increase after 3,300 BP (Fig. 4c; $P = 0.0014$ from a Z-test in SouthSaladoRiver_800BP versus Southern_2600BP), probably reflecting further gene flow from central Argentina into the Pampas. A previous analysis of a 1,600 BP sample from the Laguna Chica site[33] found excess allele sharing between this individual and central Andes populations relative to 6,800 BP Pampas individuals from the same site, which was interpreted as evidence of central-Andes-related gene flow[33] (Supplementary Information 12). However, this was a misinterpretation and, instead, these findings are driven by the then-unsampled central-Argentina lineage. The migrations into the Pampas that we detect are consistent with the observed differentiation between mitochondrial clades from Early/Middle[44] and Late[45] Holocene individuals. Archaeological evidence indicates an increase in population density in the Pampas around 3,500 BP[46], along with the introduction of ceramics and the bow and arrow[24]. Nevertheless, other archaeological connections between these regions are sparse, including evidence of lithic raw material from southern Pampas found in the south of Córdoba province[47], as well as copper necklace beads found in the Pampas[23], potentially sourced from central Argentina.

Kinship and community sizes

We analysed the distribution of runs of homozygosity (ROH) in individuals with sufficient data using hapROH (Methods) and used these distributions to estimate effective community sizes (N_e) (Supplementary Data 14). Communities in central Argentina probably had sizes comparable to those of the central Andes, and both larger than those in the Argentinian Northwest and the Paraná River region. Individuals from the Pampas showed the highest estimated effective population size, plausibly inflated by the inferred history of admixture in that region (Extended Data Table 2).

The cumulative length of ROH segments longer than 20 cM primarily reflects increased parental relatedness, and enabled us to detect significant differences among study regions (Kruskal–Wallis, $P = 0.009$). To identify which region pairs were driving this result, we performed a Conover test—a nonparametric method that compares rank differences between groups—applying an FDR correction at 0.05 to adjust resulting P values (Supplementary Fig. 76). A higher rate of close-kin unions was detected in the Argentinian Northwest than in central Argentina ($P < 0.01$) and Pampas ($P < 0.03$), suggesting differences in mating practices despite close proximity (Extended Data Fig. 9). Given the genetic and cultural connections with the central Andes (Supplementary Information 2), this may reflect a similar phenomenon to what has been reported in that region after the decline of Wari and Tiwanaku societies (1,000 BP)[48]. This was interpreted as the origin or widespread adoption of the ayllu system—a social and political unit bound together by rules of kinship affiliation and reciprocity, with preference of within-group marriage to facilitate cooperation and keep resources within the community. Although the ayllu is not documented in northwest Argentina archaeologically or ethnographically, our findings pointing to a common pattern of close-kin marriage reinforces the evidence of a related process.

In the central region, where we had a large sample size, we tested for an association between time and the cumulative length of ROH between 4 and 12 cM, which reflects background relatedness and thus is informative of community sizes. We found no evidence of population size growth in the past two and a half millennia (Extended Data Fig. 10).

Discussion

Our finding that a 10,000 BP Southern Cone individual shared more alleles with Middle Holocene individuals from the same region than with individuals from central Andes or central eastern Brazil places a lower bound on genetic divergence of Southern Cone people.

We also identify a previously unsampled deep lineage in central Argentina that possessed distinctive genetic drift by 8,500 BP and persisted as the main ancestry component throughout our time transect. This overall genetic homogeneity co-existed with the language diversity observed in the region by the sixteenth century, suggesting that these languages probably developed largely in situ and are not associated with deep genetic structure. This cautions against simplistic extrapolations regarding the mechanisms underlying linguistic and genetic differentiation[49]. We found that the central Argentina lineage is geographically structured along two clines, one reflecting admixture with central-Andes-like ancestry and the other with Middle Holocene Pampas-like ancestry. Individuals clustered with geographically proximate groups, regardless of date, suggesting limited gene flow among communities.

In the Pampas, this deep central Argentina lineage expanded southwards, where it admixed, beginning by at least 3,300 BP, with the distinct Middle Holocene genetic component in that region[11], eventually becoming the dominant ancestry in the Pampas during the last millennium. There is a gap in available data from the Pampas between 6,800 BP and 3,300 BP, and more densely sampled time series would enable a richer characterization of this process.

In northwest Argentina, we document a long-standing presence of central-Andes-type ancestry, at least by around 9,000 BP, and evidence of genetic connectivity between the central Argentina and central Andes lineages potentially as early as 4,600 BP.

We infer an admixture event in the Gran Chaco region involving a tropical/subtropical-forest-like source and the central Argentina lineage. This is consistent with archaeological evidence of increased population movements into the Gran Chaco since about 800 BP[50]. In the Paraná River Lower Delta, a 400 BP individual with a Guaraní-associated archaeological context clustered with populations from Brazil, a region with the largest density of Guaraní sites[27]. We found no evidence of a specific affinity between modern and ancient published Arawak-associated individuals from the Caribbean, north of South America and the Gran Chaco, and the Paraná Delta groups and, therefore, although there is archaeological support for a local adoption of Arawak cultural traits, we were not able to detect a significant migration with our data.

We find a higher rate of close-kin unions in northwest Argentina than in central Argentina, potentially reflecting adoption in northwest Argentina of what in the central Andes was the ayllu social system—a kinship-based organizational structure.

A limitation in our study is sparse sampling of the Mid-to-Early Holocene, and of the Pampas, Gran Chaco and Pantanal regions. However, the genetic structure revealed here provides a basis for correlation to archaeology, and enriches our understanding of an important world region.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-025-09731-3.

1. Rothhammer, F. & Dillehay, T. D. The late Pleistocene colonization of South America: an interdisciplinary perspective. *Ann. Hum. Genet.* **73**, 540–549 (2009).
2. Reich, D. et al. Reconstructing native American population history. *Nature* **488**, 370–374 (2012).
3. Arango-Isaza, E. et al. The genetic history of the Southern Andes from present-day Mapuche ancestry. *Curr. Biol.* **33**, 2602–2615 (2023).
4. Luisi, P. et al. Fine-scale genomic analyses of admixed individuals reveal unrecognized genetic ancestry components in Argentina. *PLoS ONE* **15**, e0233808 (2020).
5. Lorandi, A. M. El desarrollo cultural prehispánico en Santiago del Estero, Argentina. *J. Soc. Am.* **65**, 63–85 (1978).

6. Rex González, A. Dinámica cultural del NO argentino. Evolución e historia en las culturas del NO Argentino. *Antiquitas* **28/29**, 1–15 (1979).
7. Bonomo, M., Cabanillas, E. & Montero, R. Archaeometallurgy in the Paraná Delta (Argentina): composition, manufacture, and indigenous routes. *J. Anthropol. Archaeol.* **47**, 1–11 (2017).
8. Laguens, A. G. El poblamiento inicial del sector austral de las Sierras Pampeanas de Argentina desde la ecología del paisaje. *Anal. Arqueol. Etnol.* **61/62**, 67–106 (2005).
9. Prates, L., Politis, G. G. & Perez, S. I. Rapid radiation of humans in South America after the last glacial maximum: a radiocarbon-based study. *PLoS ONE* **15**, e0236023 (2020).
10. Weitzel, C., Mazzia, N. & Flegenheimer, N. Assessing Fishtail points distribution in the southern Cone. *Quat. Int.* **473**, 161–172 (2018).
11. Posth, C. et al. Reconstructing the deep population history of Central and South America. *Cell* **175**, 1185–1197 (2018).
12. Iriondo, M., Brunetto, E. & Kröhling, D. Historical climatic extremes as indicators for typical scenarios of Holocene climatic periods in the Pampean plain. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **283**, 107–119 (2009).
13. Barrientos, G. & Perez, S. I. Was there a population replacement during the Late mid-Holocene in the southeastern Pampas of Argentina? Archaeological evidence and Paleoecological basis. *Quat. Int.* **132**, 95–105 (2005).
14. Araujo, A. G. D. M., Pilo, L. B., Neves, W. A. & Atui, J. P. V. Human occupation and paleoenvironments in South America: expanding the notions of an "Archaic gap". *Rev. Mus. Arqueol. Etnol.* **15/16**, 3–35 (2006).
15. Hocsman, S. & Babot, M. P. La transición de cazadores-recolectores a sociedades agropastoriles en Antofagasta de la Sierra (Puna de Catamarca, Argentina): perspectivas desde la agencia y las prácticas. *Chungara* **50**, 51–70 (2018).
16. Korstanje, M. A. et al. (eds) *Crónicas Materiales Precolombinas: Arqueología de los Primeros poblados del Noroeste Argentino* (Publicaciones de la Sociedad Argentina de Antropología, 2015).
17. Laguens, A. & Bonnin, M. *Sociedades Indígenas de las Sierras Centrales: Arqueología de Córdoba y San Luis* 2nd edn (Univ. Nacional de Córdoba, 2023).
18. Fabra, M. & Gonzalez, C. Oral health, diet and social change in populations of the region of central Argentina during Late Holocene: bioarchaeological and isotopic evidence. *Latin Am. Antiq.* **30**, 818–835 (2019).
19. Farberman, J. & Taboada, C. ¿"Lules nómades" y "lules sedentarios"?: Sociedades indígenas, Movilidad y Prácticas de Subsistencia en la Llanura Santiagueña Prehispánica y Colonial (Santiago del Estero, Argentina). *Andes* **29**, 1–24 (2018).
20. Bonomo, M. et al. Las poblaciones indígenas prehispánicas del río Paraná Inferior y Medio. *Rev. Mus. La Plata* **4**, 585–620 (2019).
21. Fabra, M. Vivir en los límites del Gran Chaco y la región pampeana: aportes al estudio de las poblaciones de la costa sur de Laguna Mar Chiquita (Córdoba) desde la Arqueología y la Bioantropología. *Rev. Mus. La Plata* **5**, 657–684 (2020).
22. Lamenza, G. N., Calandra, H. A. & Salceda, S. A. Arqueología de los ríos Pilcomayo, Bermejo y Paraguay. *Rev. Mus. La Plata* **4**, 481–510 (2019).
23. González, M. I. & Frère, M. M. Río Salado: espacio de interacción de cazadores-recolectores-pescadores (provincia de Buenos Aires, Argentina). *Rev. Mus. La Plata* **4**, 621–642 (2019).
24. Politis, G. G. & Borrero, L. A. *The Archaeology of the Pampas and Patagonia* (Cambridge Univ. Press) (2024).
25. Censabella, M. *Las Lenguas Indígenas de la Argentina: Una Mirada Actual* (Eudeba, 1999).
26. Sarasola, C. *Nuestros Paisanos Los Indios: Vida, Historia y Destino de las Comunidades Indígenas en la Argentina* (Emecé, 1992).
27. Bonomo, M., Costa Angrizani, R., Apolinaire, E. & Noelli, F. S. A model for the Guaraní expansion in the La Plata Basin and littoral zone of southern Brazil. Quat. Int. **356**, 54–73 (2015).
28. Casamiquela, R. *Rectificaciones y Ratificaciones Hacia una Interpretación Definitiva del Panorama Etnológico de la Patagonia y Área Septentrional Adyacente* (Instituto de Humanidades, Univ. Nacional del Sur, 1965).
29. Ferraz, T. et al. Genomic history of coastal societies from eastern South America. *Nat. Ecol. Evol.* **7**, 1315–1330 (2023).
30. Moreno-Mayar, J. V. et al. Early human dispersals within the Americas. *Science* **362**, eaav262 (2018).
31. Kennett, D. J. et al. South-to-north migration preceded the advent of intensive farming in the Maya region. *Nat. Commun.* **13**, 1530 (2022).
32. Nakatsuka, N. et al. Genetic continuity and change among the Indigenous peoples of California. *Nature* **624**, 122–129 (2023).
33. Nakatsuka, N. et al. A paleogenomic reconstruction of the deep population history of the Andes. *Cell* **181**, 1131–1145 (2020).
34. Nakatsuka, N. et al. Ancient genomes in South Patagonia reveal population movements associated with technological shifts and geography. *Nat. Commun.* **11**, 3868 (2020).
35. Lamenza, G. N. Utilización del análisis multivariante para la sistematización del componente alfarero del Chaco prehispánico. *Arqueol. Iberoam.* **28**, 52–61 (2015).
36. García, A. et al. Ancient and modern mitogenomes from Central Argentina: new insights into population continuity, temporal depth and migration in South America. *Hum. Mol. Genet.* **30**, 1200–1217 (2021).
37. Nores, R., Tavella, P., Fabra, M. & Demarchi, D. Ancient DNA analysis reveals temporal and geographical patterns of mitochondrial diversity in pre-Hispanic populations from Central Argentina. *Am. J. Hum. Biol.* **34**, e23733 (2022).
38. Núñez, L. & Dillehay, T. D. *Movilidad Giratoria, Armonía Social y Desarrollo en los Andes Meridionales: Patrones de Tráfico e Interacción Económica* (Univ. Católica del Norte, 1995).
39. Salceda, S. A. & Raffino, R. A. in *El Shincal de Quimivil* (ed. Rafino, R.) 165–177 (Sarquis, 2004).
40. Crouzeilles, A. L. D. Correlaciones entre la alfarería indígena encontrada en la región de Santa Fe y la de la provincia de Santiago del Estero. *An. Soc. Cient. Argent.* **CXXVIII**, 196–211 (1939).
41. Cornero, S. & Green, A. Los indígenas Corondas del Litoral Bajo Paranaense y su Señor. Un caso de liderazgo en el siglo XVI. *Rev. Soc. Paisajes Aridos Semiaridos* **4**, 10–55 (2022).
42. Politis, G. & Bonomo, M. La entidad arqueológica Goya-Malabrigo (Ríos Paraná y Uruguay) y su filiación Arawak. *Rev. Arqueol.* **25**, 10–46 (2012).
43. Hornborg, A. & Hill, J. D. in *Ethnicity in Ancient Amazonia: Reconstructing Past Identities from Archaeology, Linguistics, and Ethnohistory* (eds Hornborg, A. & Hill, J. D.) 1–27 (Univ. Press of Colorado, 2011).
44. Roca-Rada, X. et al. Ancient mitochondrial genomes from the Argentinian Pampas inform the early peopling of the Southern Cone of South America. *iScience* **24**, 102553 (2021).
45. Motti, J. M. B. et al. Ancient mitogenomes from the Southern Pampas of Argentina reflect local differentiation and limited extra-regional linkages after rapid initial colonization. *Am. J. Biol. Anthropol.* **181**, 216–230 (2023).
46. Martínez, G. et al. Radiocarbon trends in the Pampean region (Argentina). Biases and demographic patterns during the final Late Pleistocene and Holocene. *Quat. Int.* **356**, 89–110 (2015).
47. Heider, G. La gestión de recursos líticos en el Norte de Pampa Seca. *Rel. Soc. Argent. Antropol.* **XLI**, 375–396 (2016).
48. Ringbauer, H., Steinrücken, M., Fehren-Schmitz, L. & Reich, D. Increased rate of close-kin unions in the central Andes in the half millennium before European contact. *Curr. Biol.* **30**, R980–R981 (2020).
49. Hornborg, A. Theorizing ethnolinguistic diversity under globalization: beyond biocultural analogies. *Globalizations* **21**, 990–1008 (2024).
50. Lamenza, G. N. *El Hombre y el Ambiente en el Holoceno Tardío del Chaco Meridional* 183 (Univ. Nacional de Catamarca, Facultad de Ciencias Exactas y Naturales, 2013).

[1]Department of Human Evolutionary Biology, Harvard University, Cambridge, MA, USA. [2]Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA. [3]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [4]Laboratorio de Ecología Evolutiva Humana, Facultad de Ciencias Sociales, Universidad Nacional del Centro de la Provincia de Buenos Aires, Consejo Nacional de Investigaciones Científicas y Técnicas, Quequén, Argentina. [5]Instituto de Antropología de Córdoba (IDACOR), Consejo Nacional de Investigaciones Científicas y Técnicas, Córdoba, Argentina. [6]Departamento de Diversidad Biológica y Ecología, Facultad de Ciencias Exactas, Físicas y Naturales, Universidad Nacional de Córdoba, Córdoba, Argentina. [7]Museo de Antropologías, Facultad de Filosofía y Humanidades, Universidad Nacional de Córdoba, Córdoba, Argentina. [8]Departamento de Antropología, Facultad de Filosofía y Humanidades, Universidad Nacional de Córdoba, Córdoba, Argentina. [9]Centro Científico Tecnológico CONICET NOA Sur, Consejo Nacional de Investigaciones Científicas y Técnicas, San Miguel de Tucumán, Argentina. [10]Grupo de Investigación en Arqueología Andina & Instituto de Arqueología y Museo, Facultad de Ciencias Naturales e Instituto Miguel Lillo, Universidad Nacional de Tucumán, San Miguel de Tucumán, Argentina. [11]División Arqueología, Facultad de Ciencias Naturales y Museo, Universidad Nacional de La Plata, Consejo Nacional de Investigaciones Científicas y Técnicas, La Plata, Argentina. [12]Museo Universitario Florentino y Carlos Ameghino, Facultad de Ciencias Exactas, Ingeniería y Agrimensura, Universidad Nacional de Rosario, Rosario, Argentina. [13]División Antropología, Facultad de Ciencias Naturales y Museo, Universidad Nacional de La Plata, Consejo Nacional de Investigaciones Científicas y Técnicas, La Plata, Argentina. [14]Instituto de Estudios para el Desarrollo Social, Universidad Nacional de Santiago del Estero, Consejo Nacional de Investigaciones Científicas y Técnicas, Santiago del Estero, Argentina. [15]Instituto de Linguística, Folklore y Arqueología, Facultad de Humanidades, Ciencias Sociales y de la Salud, Universidad Nacional de Santiago del Estero, Santiago del Estero, Argentina. [16]Banco Nacional de Datos Genéticos, Buenos Aires, Argentina. [17]Instituto de Investigaciones Arqueológicas y Paleontológicas del Cuaternario Pampeano (INCUAPA), Facultad de Ciencias Sociales, Universidad Nacional del Centro de la Provincia de Buenos Aires, Consejo Nacional de Investigaciones Científicas y Técnicas, Olavarría, Argentina. [18]Centro de Genética Forense, Poder Judicial de la Provincia de Córdoba, Córdoba, Argentina. [19]Centro Científico Tecnológico Córdoba, Consejo Nacional de Investigaciones Científicas y Técnicas, Córdoba, Argentina. [20]Facultad de Ciencias Médicas, Universidad Nacional de Santiago del Estero, Santiago del Estero, Argentina. [21]Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Catamarca, San Fernando del Valle de Catamarca, Argentina. [22]Instituto Multidisciplinario de Historia y Ciencias Humanas, Consejo Nacional de Investigaciones Científicas y Técnicas, Buenos Aires, Argentina. [23]Centro Regional de Estudios Arqueológicos, Facultad de Humanidades y Ciencias Sociales, Universidad Nacional de Jujuy, San Salvador de Jujuy, Argentina. [24]Departamento de Humanidades, Universidad Nacional del Sur, Bahía Blanca, Argentina. [25]Fundación de Historia Natural Félix de Azara, Buenos Aires, Argentina. [26]Junta Municipal de Historia de Río Cuarto, Departamento de Antropología y Prehistoria, Río Cuarto, Argentina. [27]Instituto Nacional de Antropología y Pensamiento Latinoamericano, Consejo Nacional de Investigaciones Científicas y Técnicas, Buenos Aires, Argentina. [28]Facultad de Filosofía y Letras, Universidad de Buenos Aires, Buenos Aires, Argentina. [29]Dirección General de Patrimonio Cultural, Provincia de Santiago del Estero, Santiago del Estero, Argentina. [30]Centro de Investigación Científica y de Transferencia Tecnológica a la Producción, Consejo Nacional de Investigaciones Científicas y Técnicas, Universidad Autónoma de Entre Ríos, Diamante, Argentina. [31]Department of Genetics, Harvard Medical School, Boston, MA, USA. [32]Howard Hughes Medical Institute, Harvard Medical School, Boston, MA, USA. [33]Centre for Anthropobiology and Genomics of Toulouse (CAGT),

CNRS UMR 5288, Université Toulouse III—Paul Sabatier, Toulouse, France. [34]Senckenberg Centre for Human Evolution and Palaeoenvironment, University of Tübingen, Tübingen, Germany. [35]Australian Centre for Ancient DNA, School of Biological Sciences, University of Adelaide, Adelaide, South Australia, Australia. [36]Department of Ecology, Evolution, and Organismal Biology, Brown University, Providence, RI, USA. [37]Center for Computational Molecular Biology, Brown University, Providence, RI, USA. [38]UCSC Paleogenomics, Department of Anthropology, University of California, Santa Cruz, CA, USA. [39]UCSC Genomics Institute, University of California, Santa Cruz, CA, USA. [40]Archaeo- and Palaeogenetics, Institute for Archaeological Sciences, Department of Geosciences, University of Tübingen, Tübingen, Germany. [41]ARC Centre of Excellence for Australian Biodiversity and Heritage, University of Adelaide, Adelaide, South Australia, Australia. [42]Australia Indigenous Genomics, The Kids Research Institute Australia, Adelaide, South Australia, Australia. [43]National Centre for Indigenous Genomics, John Curtin School of Medical Research, Australian National University, Canberra, Australian Capital Territory, Australia. [44]Molecular Anthropology Laboratories, Department of Anthropology, University of Tennessee, Knoxville, TN, USA. [45]These authors jointly supervised this work: David Reich, Rodrigo Nores. ✉e-mail: fmaravalllopez@fas.harvard.edu; reich@genetics.med.harvard.edu; rodrigonores@ffyh.unc.edu.ar

Methods

Genetic data

We produced 504 ancient DNA libraries from 344 distinct skeletal samples (Supplementary Data 1). We used in-solution enrichment for over 1.2 million targeted SNPs, a standard set of genetic markers widely used in ancient DNA studies[51–54], to gather genome-wide data that met standard criteria for ancient DNA authenticity from 238 unique individuals (Supplementary Data 1). To maximize usable information for genetic analysis, we expanded this targeted SNP set with off-target sites (sites not originally targeted by the enrichment protocol but commonly captured because of close physical proximity) to arrive at approximately 2 million SNPs described previously[54]. Individuals were assigned to groups using archaeological, geographical and chronological criteria. The 238 individuals from the CSC were grouped into six biogeographical regions of Argentina and one from the Pantanal region of Paraguay as described in the main text, which we further subdivided for analysis as described in Supplementary Information 1–7. Each individual was assigned to one of the main regions and subregions on the basis of their geographical origin. Individuals from the same subregion were further grouped according to chronological criteria. In a few cases, individuals from the same region and time period were separated into different groups on the basis of distinct cultural or archaeological characteristics (for example, Inca and Guaraní).

The newly reported individuals were co-analysed with genetic data from 588 ancient pre-European contact American individuals[2,11,29–34,55–74] (Extended Data Fig. 1 and Supplementary Data 1), with the data curated as described in the Allen Ancient DNA Resource[75], a publicly available collection of ancient human genome-wide data. For co-analysis purposes, we defined 'pre-European contact American individuals' as those having a mean date (either a direct radiocarbon date or a contextual date from archaeological evidence) before 600 BP. We also included in the analysis previously generated SNP array data from present-day Native American groups[2], restricted to the sites intersecting the 1240k SNP set[52]. No statistical methods were used to pre-determine sample size.

Direct accelerator mass spectrometry ^{14}C bone dates and calibration

We report 35 new direct accelerator mass spectrometry ^{14}C dates obtained from specialized laboratories at Pennsylvania State University (PSUAMS) (n = 13) and the University of Georgia (UGAMS) (n = 22), which we combined with 98 previously reported ^{14}C dates from studies of the newly individuals (Supplementary Data 1). We also integrated archaeological context information to provide information on chronology (Supplementary Information 1–7). Moreover, we made use of 398 previously reported ^{14}C dates for the previously published ancient American individuals whose genetic data we used for co-analysis (Supplementary Data 1). All calibrated ^{14}C ages were calculated using OxCal (v.4.4)[76] with the Southern Hemisphere terrestrial (IntCal20)[77] calibration curves. The marine reservoir effect was not considered, as all individuals analysed in this study had a terrestrial-based subsistence. Calibrated dates are reported in Supplementary Data 1 and Supplementary Information 1–6 as 95.4% CI calibrated radiocarbon ages in BCE–CE format. We also report the date mean in BP, in years before 1950 CE (calculated as the OxCal mu for a direct radiocarbon date, or as the average of the range for a contextual date), as well as the date s.d. in BP (OxCal sigma for a direct radiocarbon date, or the s.d. of the uniform distribution between the two bounds for a contextual date). Individual dates listed under individual IDs correspond to the date mean in BP (years before 1950 CE), rounded to the nearest hundred, except for the individual Argentina Central Plains SouthCordoba BarrioAlberdiRioCuarto150BP. Grouping dates listed under group ID are expressed as the average of the individual date means in BP (years before 1950 CE) of the group members, also rounded to the nearest hundred.

Ancient DNA laboratory work

Tooth or bone powder was prepared in dedicated clean rooms at Harvard Medical School by processing 228 samples corresponding to 201 individuals, and at the University of Tennessee, Knoxville (UTK), using a freezer mill for 108 samples from individual remains. Further wet-laboratory processing for all these samples was conducted at Harvard Medical School. Eight samples from six individuals (including two independent duplicates of individuals powdered at UTK) were analysed at the Australian Centre for Ancient DNA (ACAD). Moreover, for one sample, bone powder was prepared in dedicated clean rooms at University of Tübingen (UT) by abrasing the outer layer of the temporal bone surface before sampling the cochlea from the internal acoustic meatus. Around 50 mg of bone powder was generated using an electric dentist drill. DNA was extracted from powdered samples using a method optimized for retaining small DNA fragments[78–80]. The DNA was converted into sequenceable form using double-stranded or single-stranded library preparation protocols, typically pretreated with uracil-DNA glycosylase (UDG) to minimize cytosine-to-thymine errors common in ancient DNA[81–83], except for DNA processed at UT, which was converted into sequenceable form using single-stranded, double-indexed library preparation protocols with no UDG treatment[82], generating multiple libraries from the same extract. For all double-stranded libraries (except for four prepared at the University of California Santa Cruz), we replaced MinElute columns for reaction clean-ups with magnetic silica beads and Qiagen buffer PB. We then used SPRI beads instead of MinElute columns for PCR cleanup at the end of library preparation[84,85], except for libraries prepared at the University of California Santa Cruz. For libraries prepared at UT, nuclear in-solution capture was performed directly, foregoing shotgun sequencing.

We enriched the libraries for sequences overlapping mtDNA[86] and approximately 1.24 million nuclear targets together (1240k+) through two rounds of enrichment[51–53], except for the four libraries from the University of California Santa Cruz, for which the mtDNA (1 round) and 1240k (2 rounds) enrichments were performed independently. For a number of libraries, including the eight from ACAD, we used the Twist 1.4M capture kit[54,87] instead of the 1240k enrichment, which gives more uniform coverage and targets a larger set of SNPs. For some of the samples, we prepared two libraries simultaneously, and multiplexed them into one capture reaction; double-stranded libraries were captured for a single round, while single-stranded libraries were captured for two consecutive rounds. The unenriched (shotgun) and enriched (mtDNA, 1240k, 1240k+, Twist1.4M) products of double-stranded libraries were indexed and sequenced on the Illumina NextSeq500 instrument for 2 × 76 cycles and 2 × 7 cycles, respectively, or on the Illumina HiSeq X10 or NovaSeq instrument using 2 × 101 cycles and 2 × 7 cycles, respectively, except for the data prepared at UT, which were sequenced on the NovaSeq platform for 2 × 121 cycles and 2 × 8 cycles, respectively. Single-stranded libraries and double-stranded libraries prepared at ACAD were already indexed at the end of library preparation and were sequenced on either the Illumina HiSeq X10 or NovaSeq instrument for 2 × 101 and 2 × 8 cycles. For the single-stranded libraries, we used a custom sequencing read 1 primer CL72. We sequenced the nuclear capture products for about 20 million reads per library (on average 30–40 million reads per captured library in the case of data prepared at UT), and also for typically hundreds of thousands of reads for the unenriched/shotgun library.

Computational processing of sequence data

We merged paired reads overlapping by at least 15 nucleotides (allowing one mismatch) using custom code that concurrently trims adapters (https://github.com/DReichLab/ADNA-Tools), selecting the highest quality base for each nucleotide in the overlap. Non-merging read pairs were discarded. The resulting merged sequences were then mapped to the human genome reference sequence (GRCh37 from the 1000 Genomes Project[88] using the samse command of the Burrows–Wheeler

aligner[89] (v.0.7.15). Duplicate sequences were marked with Picard (command MarkDuplicates) (v.2.17.10; http://broadinstitute.github.io/picard/). For variant calling, we used a pseudohaploid approach, representing each SNP with a single allele representative. We first estimated error rates empirically (assuming that sites monomorphic in 1000 Genomes data[88] are in fact monomorphic). We stratified these error rate estimates by library type, SNP bases (variant and reference), read position, strand, mapping quality and base quality, with the base positions more than 10 bases from the 5′ or 3′ end being considered central and merged. These error rates are determined from the sample BAM, which makes our procedure adaptive. If we simply thresholded on the estimated error, this would introduce bias. For example, at a (C, T) SNP, the estimated error E(C, T) for C→T may be very different from E(T, C) for T→C. Instead, we use a symmetric function S and, for example, at a (C, T) SNP, we calculate $S = \max\{E(C, T), E(T, C)\}$. We threshold S with a parametric value (0.02) and bases with S below threshold go into a pileup of reliable bases. Finally, a random base in the pileup is selected. The actual error achieved is smaller than the threshold which is an upper bound on the error of each potential base that contributes to the pileup. For analysis, we used the SNP set that includes off-target sites apart from the standard 1240k sites and was described previously[54].

Contamination estimation

We evaluated the authenticity of ancient DNA by measuring the damage rate in the first nucleotide, and we flagged individuals as potentially contaminated if the cytosine-to-thymine substitution rate was less than 3% in UDG-treated libraries and less than 10% in non-UDG-treated libraries. Contamination evidence based on mtDNA polymorphism was determined using contamMix[90], while hapConX[91] and ANGSD[92] were used to assess contamination evidence based on X-chromosome polymorphism in males (Supplementary Data 1). These individuals were excluded from analysis, but we report their data. Moreover, we excluded, but still reported, individuals from analysis who were not genetically homogeneous with ancient pre-European contact Native Americans as assessed by either f_4-statistics or qpAdm (Supplementary Information 8 and Supplementary Data 1).

Kinship analyses

We analysed all pairs of individuals to test for evidence of close biological relatedness. In particular, we examined all non-CpG autosomal sites and calculated the mean mismatch rate at all SNPs covered by at least one sequence in both individuals. We compared this to the rate of difference between the two chromosomes within each individual, assuming that they were not closely related[60]. Individuals inferred to have a second-degree or closer relationship with someone else in the dataset (Supplementary Data 1) were excluded from analyses, usually keeping the individual with higher-coverage data (details are provided in Supplementary Information 1.2).

f_4 statistics and outgroup f_3-distance matrices

To compute f_3 and f_4 statistics, we used the qp3pop and qpDstat packages in ADMIXTOOLS[93] (v.7.0.2). When indicated, owing to an extremely large number of tests, we corrected f_4-statistic Z-scores at FDR < 0.05 using the Benjamini–Yekutieli procedure[94] (Z_{BY}) using a custom script available at GitHub (https://github.com/javiermaravall/aDNA_CSC/). Using the outgroupmode: YES parameter, we computed outgroup-f_3 statistics of the form f_3(Pop 1, Pop 2; Yoruba) or f_3(Ind 1, Ind 2; Yoruba). As these quantities measure shared drift with respect to the outgroup up to the split of Pop 1 and Pop 2 (ref. 95), or of Ind 1 and Ind 2, their inverses can be appropriately used to construct a pairwise genetic-distance matrix. We used these matrices to compute neighbour-joining trees using the ape R package (v5.8)[96], rooting them at USA_Ancient_Beringian.SG. To obtain a low-dimensional representation of these objects, we applied MDS to the matrices using the function cmdscale from the R stats package (v.3.6.2)[97,98].

Automatic exploration of population history models

To automatically explore the space of population history models (admixture graphs), we used the R library ADMIXTOOLS2 (refs. 99,100) (v.2.0.0). To extract data, we used function extract f2 setting maxmiss = 0.15. This kept 329,279 SNPs, 293,834 of which were polymorphic among the studied groups. Although the recommended value of this parameter is 0 for automatic population-history model exploration, lower values of allowed missingness resulted in too small numbers of SNPs retained (<30,000). We launched 100 independent iterations of the function find graphs, for each of $n = 0, 1$ admixture events, which starts from a given set of populations and explores admixture graphs until the resulting graph cannot be made to better fit the data. As this search can get trapped in local optima, the execution of a large number of independent iterations, each starting from a randomly initialized admixture graph, enables better characterization of the set of optimally fitting graphs. For each n and each iteration, we recorded the hash (unique topology identifier), score (a measure of fit) and worst residual (Z score for the largest deviation between observed f_4 statistics and the value predicted by the model). For each n, we gathered all final models with a unique hash, and aggregated these across values of n. This resulted in a set of 52 unique models (Supplementary Data 2). To understand whether some elements of this set better fit the data than others, we tested, for each pair of models, whether the scores were significantly different. To this end, we used the functions qpgraph resample multi and compare fits, which perform this test using a combination of holding out data and SNP block bootstrap resampling, to account for both differences in model complexity and potential differences in scores due to chance alone. As these tests indicated no evidence for invoking a higher number of admixture events (Supplementary Data 2), we chose not to explore models with a number of admixture events greater than 1.

Computation of F_{st} values

To compute F_{st} between pairs of groupings, we used smartpca[101] (v.18711), with the flags inbreed: YES, fstonly: YES, fstverbose: YES. We restricted to groupings for which at least 5,000 SNPs were used for all pairwise computations. We computed a complete hierarchical clustering tree with the package linkage from the scipy library[102,103] (v.1.16.0).

Testing cladality and sources of ancestry using qpWave and qpAdm

Determining whether pairs of populations (A,B) and (C,D) form clades can be reframed as evaluating whether a single gene flow event separated the pairs (f_4(A,B,C,D) = 0) or multiple events occurred (f_4(A,B,C,D) ≠ 0). The qpWave method estimates the minimum number of gene flow events between two groups, L and R (sizes n_L and n_R). It uses f_4 statistics $f_4(L_i, L_j; R_m, R_n)$ to quantify shared genetic drift within L and R. If L and R form distinct clades, all f_4 statistics should be zero. It uses f_4 statistics of the form $\frac{n_L(n_L-1)}{2} \cdot \frac{n_R(n_R-1)}{2}$, forming a matrix X. The rank of X indicates the minimum number of gene flow events; a higher rank suggests more events. Practically, X is an $(n_L - 1) \times (n_R - 1)$ matrix of $f_4(L_1, L_i; R_1, R_j)$ statistics. If $n_R > n_L$, the maximum rank is $n_L - 1$, implying at least $n_L - 1$ gene flow events. P values are derived from a χ^2 distribution based on log-likelihood differences between models. Full details are available in the original publication[2]. qpAdm extends this concept to assess the genetic make-up of an additional population T, by comparing gene flow events in L and R with those in L ∪ T and R. If L ∪ T and R show more events than L and R, T has gene flow with R and cannot be modelled solely from L. If both models yield the same rank, T can be modelled from L, allowing estimation of contributions from L to T[53]. For qpWave computations, we used ADMIXTOOLS[93] (v.7.0.2), setting the allsnps: YES. For qpAdm computations, we used ADMIXTOOLS2 (refs. 99,100) (v.2.0.0), setting allsnps=TRUE. To quantify a Mesoamerican contribution into Chile_LosRieles_5100BP, we performed an inverse variance-weighted meta-analysis across

passing models with a Mesoamerican-related source (Supplementary Data 3). Dates of admixture events were estimated using DATES[104] (v.210), but were too noisy.

ADMIXTURE clustering analysis

We used the ADMIXTURE[105,106] (v.1.23) software package to perform an unsupervised assessment of genetic structure among the newly reported individuals, including ancient (Supplementary Data 1) and modern[2] Native Americans for reference. The Karitiana and Surui groups were excluded, to avoid biases that can arise through the presence of highly drifted populations[107]. Input data was prepared using PLINK (v.1.9)[108]. We used the maf 0.01 parameter to remove SNPs with minor allele frequency below 0.01. To prune out genetic markers in strong linkage disequilibrium (LD), we applied the indep-pairwise parameter with the following options: a pairwise r^2 threshold of 0.4, a window size of 200 variants and a step size of 25 variants. For each value $K = 1, \ldots, 12$ of the number of source populations, we ran four random-seed replicates.

Analyses of ROH

To call ROH longer than 4 cM in ancient individuals, we used hapROH[109] (v.0.64). We used the 1000 Genomes Project haplotype panel[88], which includes 5,008 global haplotypes, as our reference panel. We restricted analysis to individuals for whom at least 400,000 SNPs were covered with respect to the 1240k SNP set. As this methodology was calibrated for the 1240k SNP set, not including off-target sites, we downsampled to the 1240k SNP set for this analysis. All analyses were conducted using the default settings of hapROH. To estimate effective population sizes for study subregions (N_e) from ROH distributions, we restricted to individuals with a mean date up to 3,000 BP and with a cumulative sum of ROH segments longer than 20 cM below 50 (to avoid biases due to inbreeding) and used the function MLE ROH Ne() from hapROH (Supplementary Data 14). To test for significant differences among study subregions in the ROH distributions of segments above 20 cM (informative of recent instances of close parental relatedness) we used the Python library SciPy v.1.13.1 (refs. 102,103) to perform a Kruskal–Wallis test (function kruskal()) using the cumulative length of segments in that length range for each individual, which we followed up on with a Conover test for each pair of subregions, performed using the Python library scikit-posthocs[110] v.0.9.0, and correcting P values at FDR < 0.05 (function posthoc_conover() with the parameter p_adjust='fdr_bh'). To test for a significant association between ROH segments in the range 4–12 cM (which are informative of the levels of background relatedness and thus of effective population sizes) and time in the central Argentina region, we regressed the cumulative sum of segments in that length range on mean date, for central Argentina individuals with a mean date below 2,500 BP, using the SciPy library[102] v.1.13.1 (function linregress()).

Map plotting

Figure 1a was generated in R[111] v.4.3.2 using the open-source packages dplyr[112] v.1.1.4, ggforce[113] v.0.4.2, ggnewscale[114] v.0.4.10, ggplot2 (ref. 115) v.3.4.4, ggspatial[116] v.1.1.9, ggstar[117] v.1.0.4, ggrepel[118] v.0.9.5, paletteer[119] v.1.3, raster[120] v.3.6-26, rnaturalearth[121] v.1.0.1, sf[122,123] v.1.0-15, tidyterra v.0.5.2 (ref. 124) and terra[125] v.1.7-71, using Natural Earth (https://www.naturalearthdata.com), GADM (https://gadm.org) and Portal de Información Hídrica de Córdoba-APRHI (https://portal-aprhi.opendata.arcgis.com/) data.

Ethics statement

This study adhered to ethical guidelines for working with human remains drafted both by a diverse and international group of anthropological and paleogenetic scholars[126] and the Argentine Association of Biological Anthropology[127], treating these deceased individuals with respect and using minimally destructive analyses techniques. Our research program involving ancient human remains received approval from the Ethics Committee of the CEMIC (Comité de Etica en Investigación, Centro de Educación Médica e Investigaciones Clínicas 'Norberto Quirno'). Skeletal samples were exported with authorization from the institutions safeguarding them (such as provincial and national museums, universities), obtaining proper permits from each province (for example, Agencia Córdoba Cultura) and the Argentina government (Instituto Nacional de Antropología y Pensamiento Latinoamericano and Customs). In instances in which Indigenous communities were associated with these individuals, analyses were conducted in engagement with these communities (that is, ref. 128, primarily facilitated through interactions between archaeologists and the communities). In the particular case of samples from the Córdoba province, we secured endorsement and support for this research from the Consejo de Comunidades de Pueblos Indígenas de la Provincia de Córdoba, Argentina (Council of Communities of Indigenous Peoples of the Province of Córdoba). As part of our ongoing commitment to responsible and ethical research practices, we summarized the main results of our analyses in a simplified, bulleted text in Spanish describing regional-level population history inferences (Supplementary Information 7), and shared it with Indigenous communities (when present or identified), rural localities, regional Indigenous councils (such as the mentioned Consejo de Comunidades de Pueblos Indígenas de la Provincia de Córdoba) and other stakeholders, including museum directors and curators, landowners and local authorities. We received positive and constructive feedback from them, including comments regarding how the genetic insights could be integrated with their traditional knowledge about their history.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Genotype data for newly reported individuals included in main analyses from this study can be obtained from the Harvard Dataverse repository (https://doi.org/10.7910/DVN/UQVPJQ). The aligned sequences for all individuals are available through the European Nucleotide Archive (PRJEB97713). Previously published data used in our analyses are available as follows: genetic data for modern individuals from Native American groups[2] are available for non-profit research on population history under an interinstitutional data access agreement with the Universidad de Antioquia, Colombia (queries regarding data access should be sent to a.ruizlin@ucl.ac.uk); genetic data for previously published ancient individuals are available at the Allen Ancient DNA Resource (https://doi.org/10.7910/DVN/FFIDCW); 1000 Genomes haplotype reference panel (http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/), human reference genome hg19 (https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000001405.13/); data used for map plotting are available at Natural Earth (https://www.naturalearthdata.com), GADM (https://gadm.org) and Portal de Información Hídrica de Córdoba-APRHI (https://portal-aprhi.opendata.arcgis.com/). Other newly reported data, such as radiocarbon dates and archaeological context information, are included in this Article and its Supplementary Information.

Code availability

Custom scripts and accompanying materials for the appropriate results sections are available at GitHub (https://github.com/javiermaravall/aDNA_CSC/).

51. Fu, Q. et al. An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524**, 216–219 (2015).
52. Mathieson, I. et al. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015).

53. Haak, W. et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).
54. Rohland, N. et al. Three assays for in-solution enrichment of ancient human DNA at more than a million SNPs. *Genome Res.* **32**, 2068–2078 (2022).
55. Barquera, R. et al. Ancient genomes reveal insights into ritual life at Chichén Itzá. *Nature* **630**, 912–919 (2024).
56. Capodiferro, M. R. et al. Archaeogenomic distinctiveness of the Isthmo-Colombian area. *Cell* **184**, 1706–1723 (2021).
57. De la Fuente, C. et al. Genomic insights into the origin and diversification of late maritime hunter-gatherers from the Chilean Patagonia. *Proc. Natl Acad. Sci. USA* **115**, E4006–E4012 (2018).
58. Fernandes, D. M. et al. A genetic history of the pre-contact Caribbean. *Nature* **590**, 103–110 (2021).
59. Flegontov, P. et al. Palaeo-Eskimo genetic ancestry and the peopling of Chukotka and North America. *Nature* **570**, 236–240 (2019).
60. Kennett, D. J. et al. Archaeogenomic evidence reveals prehistoric matrilineal dynasty. *Nat. Commun.* **8**, 14115 (2017).
61. Lindo, J. et al. The genomic prehistory of the Indigenous peoples of Uruguay. *PNAS Nexus* **1**, pgac047 (2022).
62. Lindo, J. et al. The genetic prehistory of the Andean highlands 7000 years BP though European contact. *Sci. Adv.* **4**, eaau4921 (2018).
63. Moreno-Mayar, J. V. et al. Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. *Nature* **553**, 203–207 (2018).
64. Nägele, K. et al. Genomic insights into the early peopling of the Caribbean. *Science* **369**, 456–460 (2020).
65. Nieves-Colón, M. A. et al. Ancient DNA reconstructs the genetic legacies of precontact Puerto Rico communities. *Mol. Biol. Evol.* **37**, 611–626 (2019).
66. Popović, D. et al. Ancient genomes reveal long-range influence of the pre-Columbian culture and site of Tiwanaku. *Sci. Adv.* **7**, eabg7261 (2021).
67. Raghavan, M. et al. The genetic prehistory of the New World Arctic. *Science* **345**, 1255832 (2014).
68. Raghavan, M. et al. Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* **349**, aab3884 (2015).
69. Rasmussen, M. et al. The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* **506**, 225–229 (2014).
70. Rasmussen, M. et al. The ancestry and affiliations of Kennewick Man. *Nature* **523**, 455–458 (2015).
71. Scheib, C. L. et al. Ancient human parallel lineages within North America contributed to a coastal expansion. *Science* **360**, 1024–1027 (2018).
72. Schroeder, H. et al. Origins and genetic legacies of the Caribbean Taino. *Proc. Natl Acad. Sci. USA* **115**, 2341–2346 (2018).
73. Sedig, J. et al. High levels of consanguinity in a child from Paquimé, Chihuahua, Mexico. *Antiquity* **98**, 1023–1039 (2024).
74. Villa-Islas, V. et al. Demographic history and genetic structure in pre-Hispanic Central Mexico. *Science* **380**, eadd6142 (2023).
75. Mallick, S. et al. The Allen Ancient DNA Resource (AADR) a curated compendium of ancient human genomes. *Sci. Data* **11**, 182 (2024).
76. Bronk Ramsey, C. Bayesian analysis of radiocarbon dates. *Radiocarbon* **51**, 337–360 (2009).
77. Hogg, A. G. et al. SHCal20 Southern Hemisphere calibration, 0-55,000 years cal BP. *Radiocarbon* **62**, 759–778 (2020).
78. Dabney, J. et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl Acad. Sci. USA* **110**, 15758–15763 (2013).
79. Korlevic, P. et al. Reducing Microbial and human contamination in dna extractions from ancient bones and teeth. *Biotechniques* **59**, 87–93 (2015).
80. Rohland, N., Glocke, I., Aximu-Petri, A. & Meyer, M. Extraction of highly degraded DNA from ancient bones, teeth and sediments for high-throughput sequencing. *Nat. Protoc.* **13**, 2447–2461 (2018).
81. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. B* **370**, 20130624 (2015).
82. Gansauge, M.-T., Aximu-Petri, A., Nagel, S. & Meyer, M. Manual and automated preparation of single-stranded DNA libraries for the sequencing of DNA from ancient biological remains and other sources of highly degraded DNA. *Nat. Protoc.* **15**, 2279–2300 (2020).
83. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **2010**, db.prot5448 (2010).
84. DeAngelis, M. M., Wang, D. G. & Hawkins, T. L. Solid-phase reversible immobilization for the isolation of PCR products. *Nucleic Acids Res.* **23**, 4742–4743 (1995).
85. Rohland, N. & Reich, D. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Res.* **22**, 939–946 (2012).
86. Maricic, T., Whitten, M. & Pääbo, S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS ONE* **5**, e14004 (2010).
87. Davidson, R. et al. Optimised in-solution enrichment of over a million ancient human SNPs. *Genome Biol.* **26**, 190 (2025).
88. Auton, A. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
89. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
90. Fu, Q. et al. A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* **23**, 553–559 (2013).
91. Huang, Y. & Ringbauer, H. hapCon: estimating contamination of ancient genomes by copying from reference haplotypes. *Bioinformatics* **38**, 3768–3777 (2022).
92. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: analysis of next generation sequencing data. *BMC Bioinform.* **15**, 356 (2014).
93. Patterson, N. et al. Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
94. Benjamini, Y. & Yekutieli, D. The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* **29**, 1165–1188 (2001).
95. Skoglund, P. et al. Genetic evidence for two founding populations of the Americas. *Nature* **525**, 104–108 (2015).
96. Paradis, E., Claude, J. & Strimmer, K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
97. Shepard, R. N. The analysis of proximities: multidimensional scaling with an unknown distance function. I. *Psychometrika* **27**, 125–140 (1962).
98. Kruskal, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* **29**, 1–27 (1964).
99. Maier, R. & Patterson, N. admixtools: inferring demographic history from genetic data. R package version 2.0.0. *GitHub* https://github.com/uqrmaie1/admixtools (2023).
100. Maier, R. et al. On the limits of fitting complex models of population history to *f*-statistics. *eLife* **12**, e85492 (2023).
101. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
102. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
103. Harris, C. R. et al. Array programming with NumPy. *Nature* **585**, 357–362 (2020).
104. Narasimhan, V. M. et al. The formation of human populations in South and Central Asia. *Science* **365**, eaat7487 (2019).
105. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
106. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
107. Lawson, D. J., van Dorp, L. & Falush, D. A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nat. Commun.* **9**, 3258 (2018).
108. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, s13742-015-0047-8 (2015).
109. Ringbauer, H., Novembre, J. & Steinrücken, M. Parental relatedness through time revealed by runs of homozygosity in ancient DNA. *Nat. Commun.* **12**, 5425 (2021).
110. Terpilowski, M. A. scikit-posthocs: pairwise multiple comparison tests in Python. *J. Open Source Softw.* **4**, 1169 (2019).
111. R Core Team. R: a language and environment for statistical computing. *R Foundation for Statistical Computing* https://www.R-project.org/ (2013).
112. Wickham, H., François, R., Henry, L., Müller, K. & Vaughan, D. dplyr: a grammar of data manipulation. R package version 1.1.4. *GitHub* https://github.com/tidyverse/dplyr (2023).
113. Pedersen, T. L. ggforce: accelerating 'ggplot2'. R package version 0.5.0. *GitHub* https://github.com/thomasp85/ggforce (2024).
114. Campitelli, E. ggnewscale: multiple fill and colour scales in 'ggplot2'. R package version 0.5.0. *GitHub* https://eliocamp.github.io/ggnewscale/ (2025).
115. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2016); https://ggplot2.tidyverse.org.
116. Dunnington, D. ggspatial: spatial data framework for ggplot2. *GitHub* https://github.com/paleolimbot/ggspatial (2023).
117. Xu, S. ggstar: Multiple geometric shape point layer for 'ggplot2'. R package version 1.0.4.002. *GitHub* https://github.com/xiangpin/ggstar (2025).
118. Slowikowski, K. ggrepel: automatically position non-overlapping text labels with 'ggplot2'. *GitHub* https://github.com/slowkow/ggrepel (2024).
119. Hvitfeldt, E. paletteer: comprehensive collection of color palettes. R package version 1.3.0. *GitHub* https://github.com/EmilHvitfeldt/paletteer (2021).
120. Hijmans, R. J. raster: geographic data analysis and modeling. R package version 3.6-31. *Spatial Data Science* https://rspatial.org/raster (2025).
121. Massicotte, P. & South, A. rnaturalearth: world map data from Natural Earth. R package version 1.0.1.9000. *GitHub* https://github.com/ropensci/rnaturalearth (2025).
122. Pebesma, E. Simple features for R: standardized support for spatial vector data. *R J.* **10**, 439–446 (2018).
123. Pebesma, E. & Bivand, R. *Spatial Data Science: With Applications in R* (Chapman and Hall/CRC, 2023); https://r-spatial.org/book/.
124. Hernangómez, D. Using the tidyverse with terra objects: the tidyterra package. *J. Open Source Softw.* **8**, 5751 (2023).
125. Hijmans, R. J. terra: spatial data analysis. R package version 1.8-23. *GitHub* https://github.com/rspatial/terra (2025).
126. Alpaslan-Roodenberg, S. et al. Ethics of DNA research on human remains: five globally applicable guidelines. *Nature* **599**, 41–46 (2021).
127. Aranda, C., Barrientos, G. & Del Papa, M. C. Código deontológico para el estudio, conservación y gestión de restos humanos de poblaciones del y pasado. *Rev. Argent. Antropol. Biol.* **16**, 111–113 (2014).
128. Salceda, S. A., Desántolo, B. & Plischuk, M. Espacio de reflexión: el por qué y para quién de la investigación bioantropológica. *Rev. Argent. Antropol. Biol.* **17**, 1–6 (2015).

Acknowledgements We acknowledge the ancient individuals whose data we analysed. We thank the staff at the Consejo de Comunidades de Pueblos Indígenas de la Provincia de Córdoba for endorsing and supporting this research; the members of the local communities for their selfless collaboration during the fieldwork; the museum curators and the many individuals who were directly or indirectly involved in this work for their efforts; N. Adamski, E. Curtis, K. Stewardson and F. Zalzala for ancient DNA laboratory work; T. Wang, B. Sousa da Mota, J. Choin and K. Sirak for providing feedback on an earlier version of this manuscript; and Leonard, S. Ravishankar, G. Purnomo and R. Davidson for discussions and technical guidance. M.D. and G.G.F. were supported by Secretaría de Ciencia y Tecnología of the Universidad Nacional de Córdoba (SECyT-UNC), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET, PIP 2017-2019) and Proyecto de Investigación de UE CONICET (2017-2024). G.M.S. was supported by SECyT-UNC. P.C.M.d.Z. and D.E.O. thank the Antofagasta de la Sierra Archaeological Project (ANS) and all of its members (PIP 11220200103166CO). D.C.L. was supported by Proyecto de investigación trianual (ANPCyT, PICT 2018-2947). R.A.M. was supported by Proyecto I+D UNLP (11/N928), Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT, PICT 2020, 1787). M.F. was supported by PICT 2020, 2701 and PIP 11220200102318CO. P.B., S.H. and L.G.G.B. acknowledge CONICET, PIP 1423, ANPCyT (PICT 3049), Universidad Nacional de Tucumán (PIUNT G/707) and The H. and T. King Grant for Archaeology of the Ancient Americas, administered by the Society for American Archaeology

(grant 202003). M.B., G.G.P. and C.S. were funded by CONICET (PIP 0126), ANPCyT (PICT 0252) and UNLP (N1007). L.G.G.B. acknowledges CONICET Doctoral and Postdoctoral Research Grants. P.G.M., G.G.P. and M.E.G. were supported by National Geographic Society (grant NGS-50543R-18), CONICET (PIP112202101000004CO and PUE no. 0079). G.N.L. was supported by Dinámica cultural prehispánica en el Gran Chaco y ambientes asociados (11/N983) and CONICET; and acknowledges El Quebracho community, E. Boló Bolaño, A. Pusineri, R. Zalazar and the members of the Fundación La Piedad of the Museo Etnográfico Andrés Barbero (Asunción, Paraguay). R.N. was supported by the National Geographic Society, CONICET (PIP 2021-11220200103037CO, PUE 2016 IDACOR and BecExt 2017), ANPCyT (PICT 2020, 3937) and SECyT-UNC. The generation and analysis of ancient DNA data for this study was supported by the National Institutes of Health (R01-HG012287), the John Templeton Foundation (grant 61220), a gift from J.-F. Clin, the Allen Discovery Center programme, a Paul G. Allen Frontiers Group advised programme of the Paul G. Allen Family Foundation and the Howard Hughes Medical Institute (to D.R.). Computations were carried out on the O2 research computing platform at Harvard Medical School.

Author contributions J.M.-L., J.M.B.M., N. Pastor, D.R. and R.N. wrote the manuscript and Supplementary Information with input from all of the co-authors. J.M.-L. performed all genetic analyses under the supervision of D.R. N. Pastor generated several figures and refined all of them. J.M.-L., J.M.B.M., N. Pastor, D.R. and R.N. interpreted the results. N.R., L.F.-S., C.P., B.L., S.M., D.A.D., G.S.C., D.R. and R.N. supervised different aspects of the study. R.N. coordinated sample collection and management. J.M.B.M., M.P.T., M.F., P.B., M.B., S.E.C., G.N.L., D.C.L., P.C.M.d.Z., G.G.P., G.R.C., M.D., H.D., L.G.G.B., S.H., A.D.I., R.A.M., V.A., D.M.B., C.B., M.G.C., U.D., P.D.R., G.G.F., R.F., M.E.G., A.G.L., J.G.M., P.G.M., B.N., D.E.O., G.M.S., A.S., C.S., A.M.T., R.V., O.E. and R.N. contributed anthropological remains and/or contributed to the creation of the archaeological Supplementary Information. K.-L.K., M.M., A.M., X.R.-R., G.S., P.A.W., N. Paterson, I.L., S.M. and D.R. performed bioinformatics data processing. M.P.T., S.C.A., K.C., E.C., T.F., L.I., A.K., J.K., K.-L.K., A.M.L., N.M., I.P., L.Q., X.R.-R., P.A.W., J.N.W. and R.N. carried out wet laboratory work. D.R. and R.N. conceived and co-directed the study.

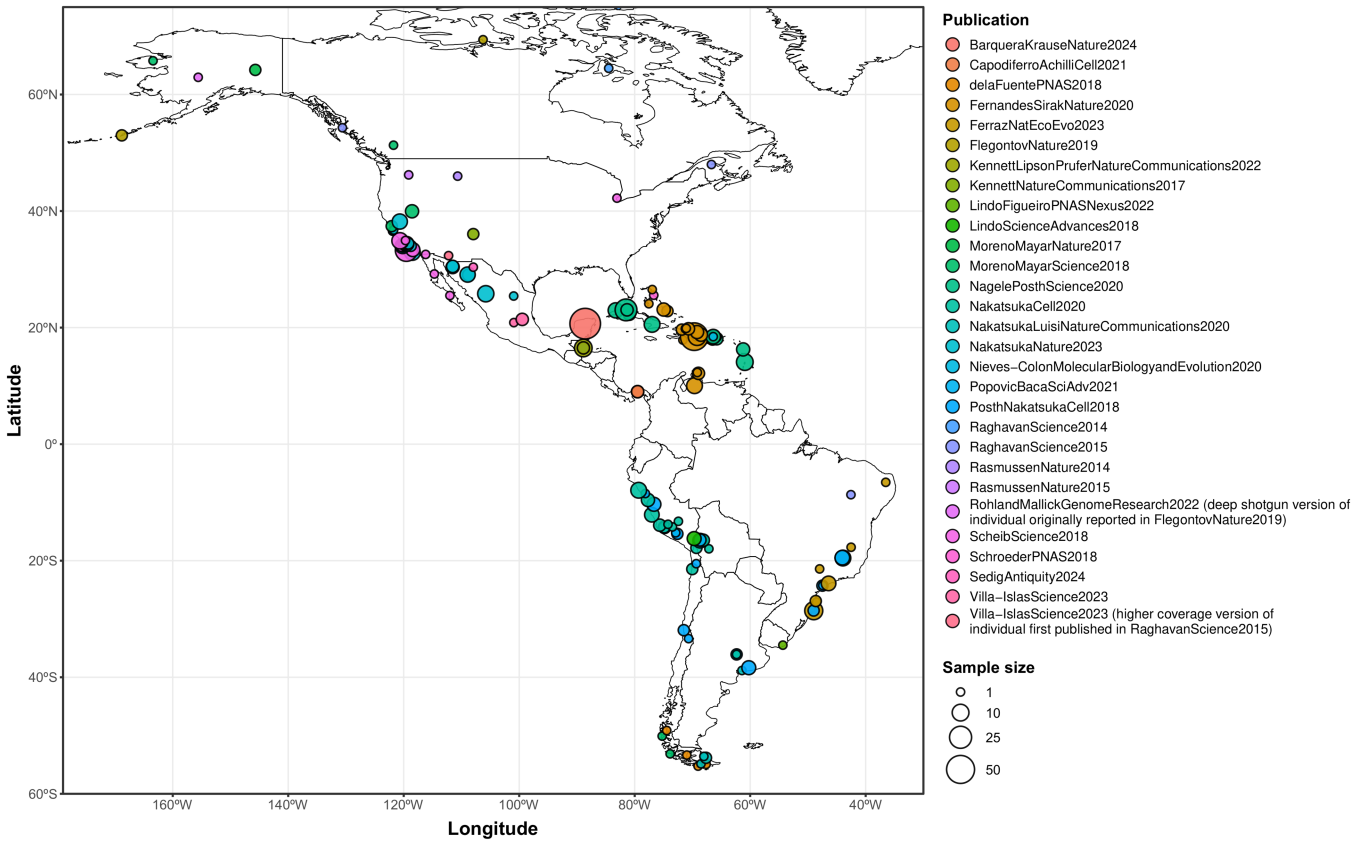
Competing interests The authors declare no competing interests.

Additional information

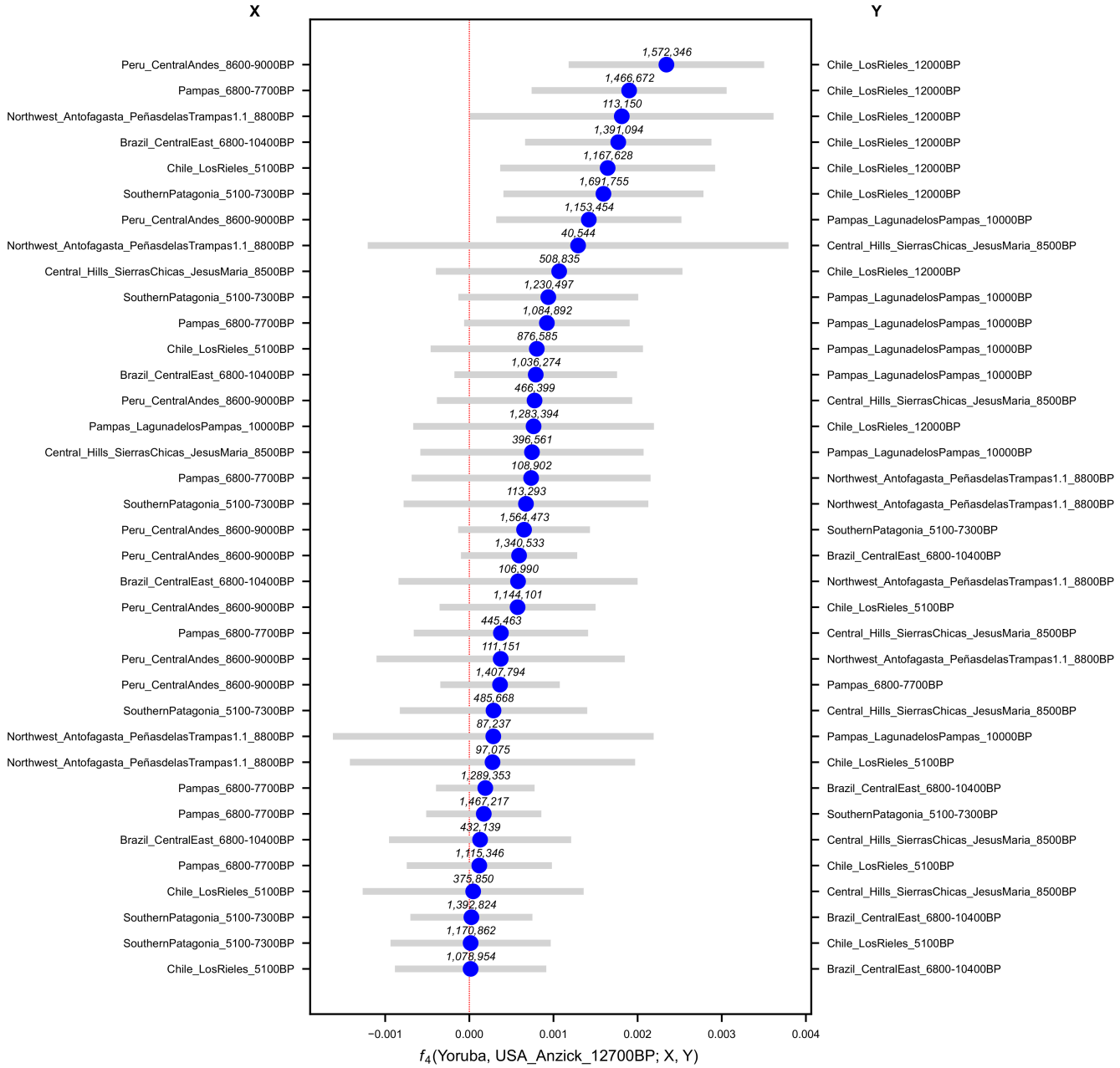
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-025-09731-3.

Correspondence and requests for materials should be addressed to Javier Maravall-López, David Reich or Rodrigo Nores.

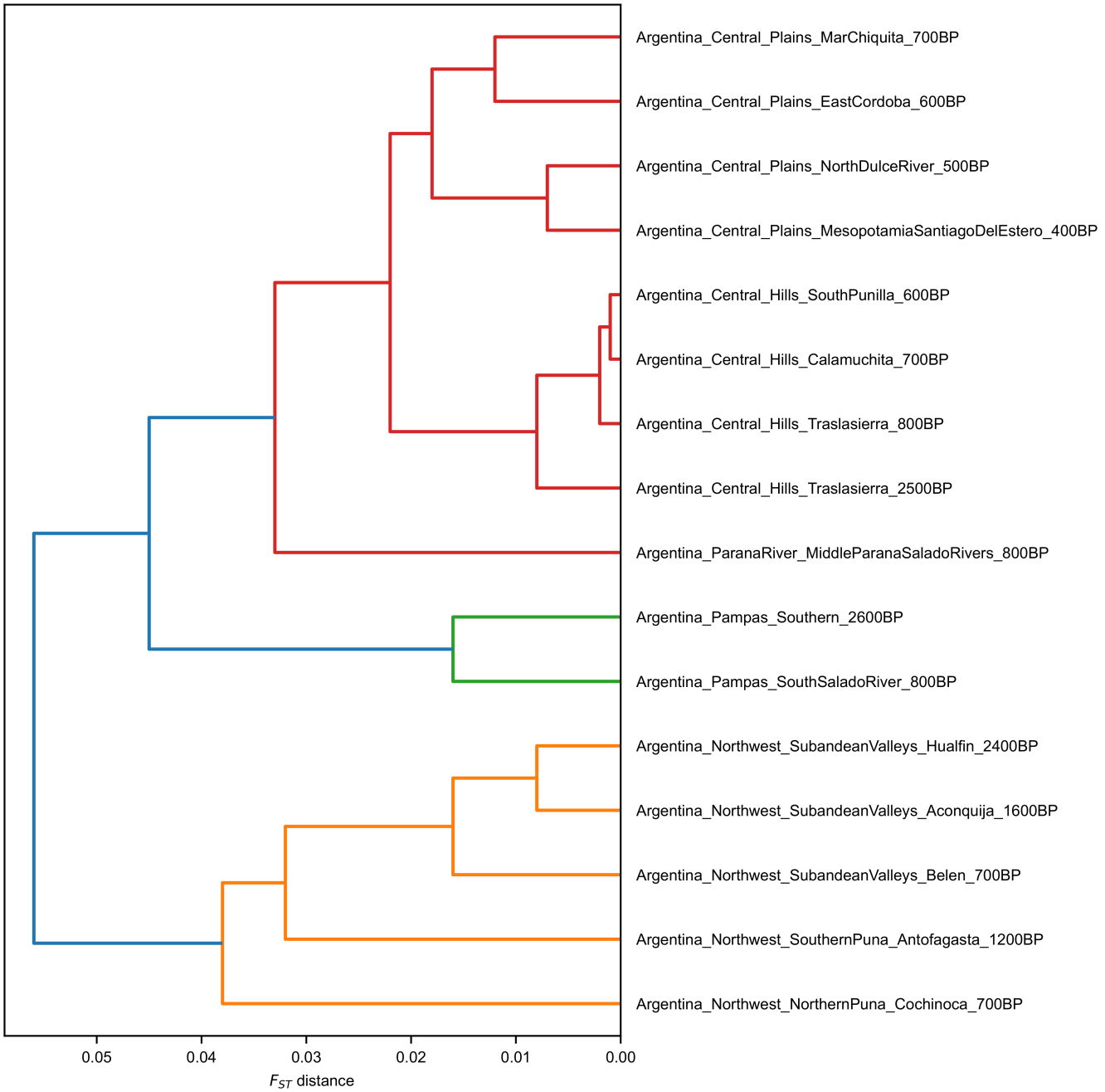
Peer review information *Nature* thanks César Méndez and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.



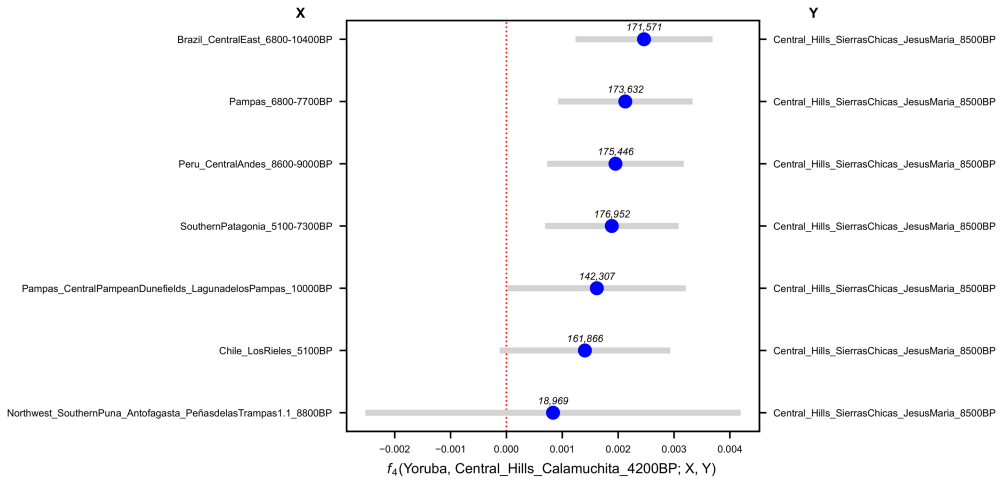
Extended Data Fig. 1 | Geographical origin of previously-published individuals included in co-analysis. Each dot indicates the geographical origin, within North, Central and South America, of a previously-published ancient grouping. Dot colours indicate the original publication, and dot sizes indicate the sample size of the grouping.



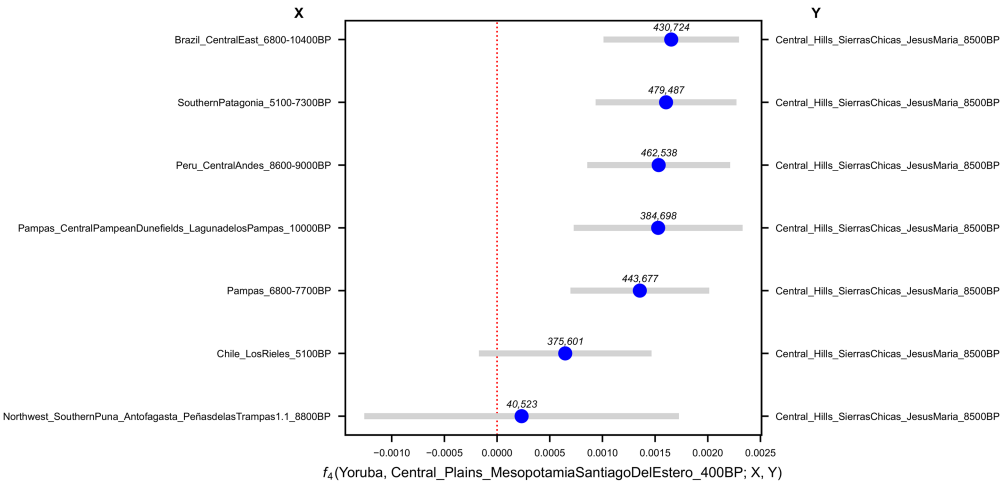
Extended Data Fig. 2 | Affinities of Anzick to Early/Middle Holocene South Americans quantified by f_4 statistics. Bars denote 95% confidence intervals (1.96 × SE) around the mean across genomic-block jackknife pseudoreplicates (point estimate). The number of SNPs used for each test is shown above each point estimate in the figure.



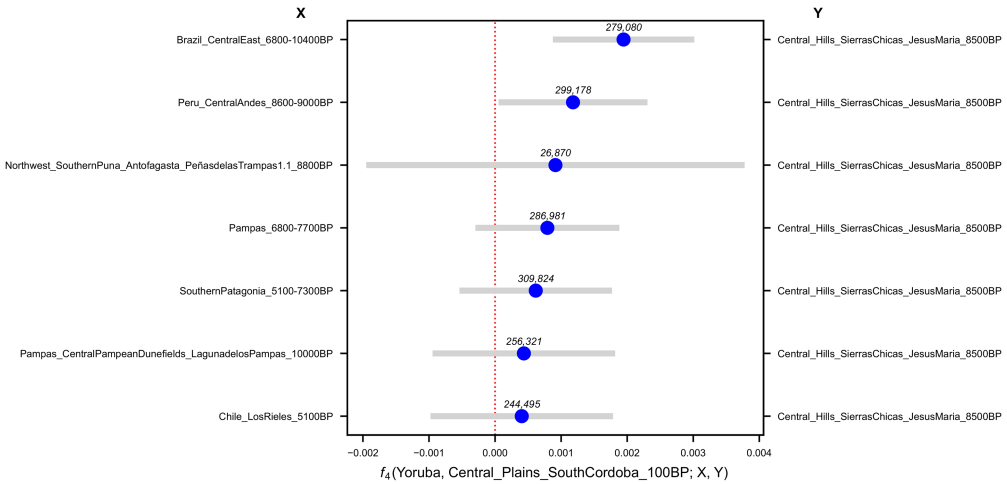
Extended Data Fig. 3 | F_{st} tree for selected groupings. Complete hierarchical-clustering tree from F_{st} distances, restricted to populations for which at least 5000 SNPs were used for all pairwise computations. Colours represent automatically-inferred clusters.



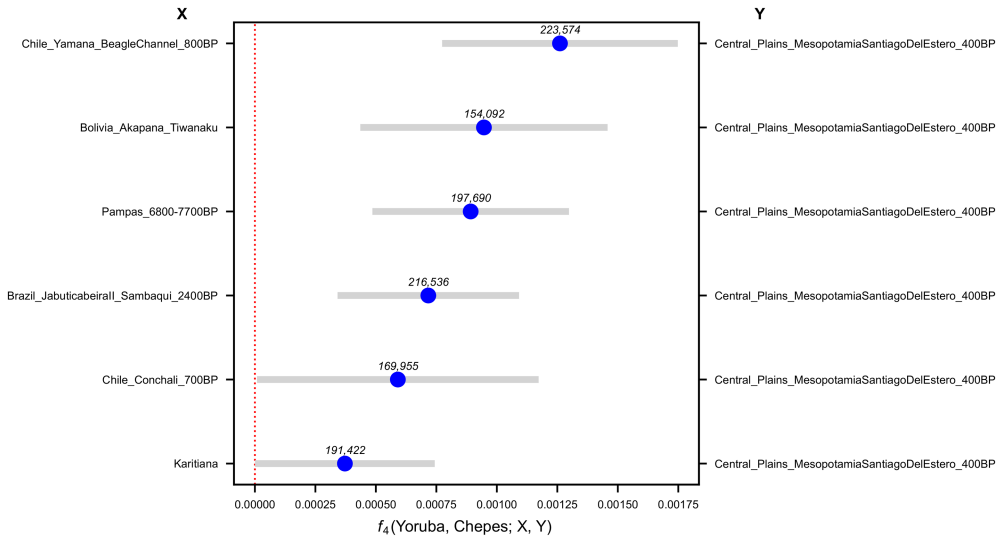
Extended Data Fig. 4 | Affinities of a representative 4200BP Central Argentina population to Early/Middle Holocene South American samples quantified by f_4 statistics. Bars are 95% confidence intervals (1.96 × SE) around the mean across genomic-block jackknife pseudoreplicates. (point estimate). The number of SNPs used for each test is shown above each point estimate in the figure.



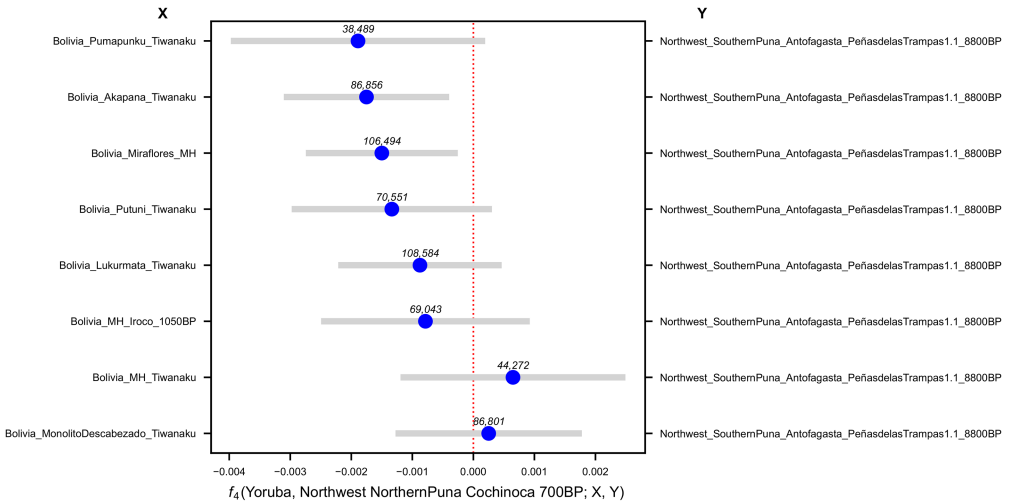
Extended Data Fig. 5 | Affinities of a representative 400BP Central Argentina population to Early/Middle Holocene South Americans quantified by f_4 statistics. Bars are 95% confidence intervals (1.96 × SE) around the mean across genomic-block jackknife pseudoreplicates (point estimate). The number of SNPs used for each test is shown above each point estimate in the figure.



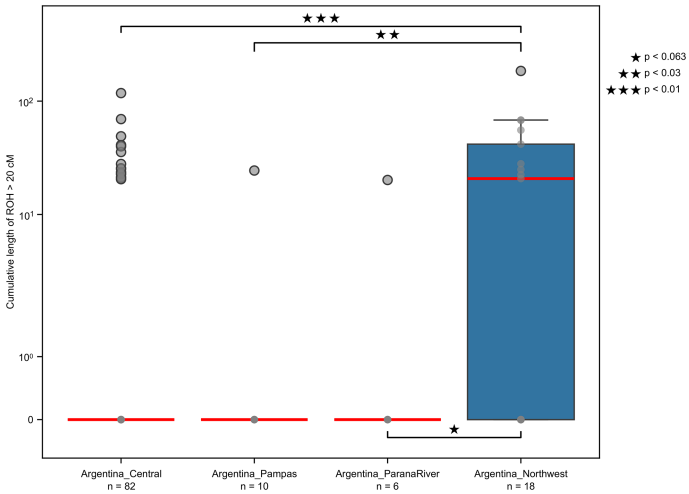
Extended Data Fig. 6 | Affinities of a representative 150BP Central Argentina population to Early/Middle Holocene South Americans quantified by f_4 statistics. Bars are 95% confidence intervals (1.96 × SE) around the mean across genomic-block jackknife pseudoreplicates (point estimate). The number of SNPs used for each test is shown above each point estimate in the figure.



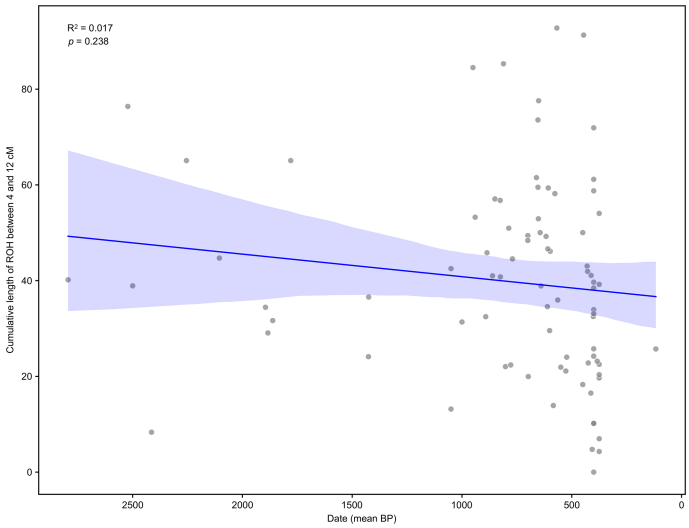
Extended Data Fig. 7 | Affinities of a modern Central Argentina admixed population[4] to Late Holocene South Americans quantified by f_4 statistics. Bars are 95% confidence intervals (1.96 × SE) around the mean across genomic-block jackknife pseudoreplicates (point estimate). The number of SNPs used for each test is shown above each point estimate in the figure.



Extended Data Fig. 8 | Affinities of *Northwest_NorthernPuna_Cochinoca_700BP* to Late Holocene Bolivians quantified by f_4 statistics. Bars are 95% confidence intervals (1.96 × SE) around the mean across genomic-block jackknife pseudoreplicates (point estimate). The number of SNPs used for each test is shown above each point estimate in the figure.



Extended Data Fig. 9 | Differences in the distribution of cumulative length of ROH segments greater than 20 cM for Southern Cone groupings up to 3000BP. Horizontal red lines denote median values (log scale), with boxes showing the interquartile range (IQR) and bars showing 1.5 x IQR Pairwise group comparisons were performed using a Conover's test (two-sided), with correction for multiple comparisons (Benjamini–Hochberg) at FDR = 0.05. Corrected *p*-values for a difference between Northwest Argentina and Central Argentina ($p = 0.00739$), and between Northwest Argentina and Argentina Pampas ($p = 0.0274$), were significant at $\alpha = 0.05$ (see Supplementary Fig. 76 for details). The number of individuals within each grouping is shown below each X axis label in the figure.



Extended Data Fig. 10 | No evidence of population size growth or decline in Central Argentina in the last two and a half millennia. Linear regression of cumulative length of ROH between 4 and 12 cM on date (mean BP), for individuals from Argentina Central at high enough coverage to call ROH (mean BP below 2500). Error bands show 95% confidence intervals around the mean linear regression fit. There is no evidence of a significant association ($p = 0.238$ from a two-sided t-test on the slope coefficient being zero).

Extended Data Table 1 | Selected f_4-statistics revealing three instances of gene flow between Central Argentina and neighbouring regions

- Late Holocene Pampas populations cannot be related with Central Argentina and Middle Holocene Pampas populations via a simple tree.
- Similar evidence of admixture between the Central Argentina and the Central Andes lineages in Northwest Argentina.
- Similar evidence of admixture between the Central Argentina lineage and a Tropical and Subtropical Forests source in the Gran Chaco region.

Outgroup	Pop_1	Pop_2	Pop_3	Z	n_SNPs
Yoruba	Argentina_Pampas_6800-7700BP	Argentina_Central_Plains_MesopotamiaSantiagoDelEstero_400BP	Argentina_Pampas_CentralPampeanDunefields_1600BP	4.65	898597
Yoruba	Argentina_Central_Plains_MesopotamiaSantiagoDelEstero_400BP	Argentina_Pampas_6800-7700BP	Argentina_Pampas_CentralPampeanDunefields_1600BP	7.053	898597
Yoruba	Argentina_Central_JesusMaria_8500BP	Argentina_Pampas_6800-7700BP	Argentina_Pampas_Southern_2600BP	2.613	446224
Yoruba	Argentina_Pampas_6800-7700BP	Argentina_Central_JesusMaria_8500BP	Argentina_Pampas_Southern_2600BP	3.6	446224
Yoruba	Argentina_Central_Plains_MesopotamiaSantiagoDelEstero_400BP	Bolivia_Akapana_Tiwanaku	Argentina_Northwest_SubandeanValleys_Hualfin_2400BP	4.129	922246
Yoruba	Bolivia_Akapana_Tiwanaku	Argentina_Central_Plains_MesopotamiaSantiagoDelEstero_400BP	Argentina_Northwest_SubandeanValleys_Hualfin_2400BP	2.337	922246
Yoruba	Argentina_Central_Plains_MesopotamiaSantiagoDelEstero_400BP	Bolivia_Akapana_Tiwanaku	Argentina_Northwest_SouthernPuna_Antofagasta_1200BP	3.166	1084211
Yoruba	Bolivia_Akapana_Tiwanaku	Argentina_Central_Plains_MesopotamiaSantiagoDelEstero_400BP	Argentina_Northwest_SouthernPuna_Antofagasta_1200BP	2.607	1084211
Yoruba	Argentina_Central_Plains_MesopotamiaSantiagoDelEstero_400BP	Bolivia_Miraflores_MH	Argentina_Northwest_SouthernPuna_Antofagasta_2100BP	3.332	549747
Yoruba	Bolivia_Miraflores_MH	Argentina_Central_Plains_MesopotamiaSantiagoDelEstero_400BP	Argentina_Northwest_SouthernPuna_Antofagasta_2100BP	4.625	549747
Yoruba	Karitiana	Argentina_Central_Plains_MesopotamiaSantiagoDelEstero_400BP	Wichi	2.315	315637
Yoruba	Argentina_Central_Plains_MesopotamiaSantiagoDelEstero_400BP	Karitiana	Wichi	2.728	315637
Yoruba	Piapoco	Argentina_Central_JesusMaria_8500BP	Toba	3.273	131781
Yoruba	Argentina_Central_JesusMaria_8500BP	Piapoco	Toba	3.428	131781

Plains_MiddleSaladoRiver_SantiagodelEstero_400B is a late Central Argentina population that is a clade with *Central_JesusMaria_8500BP* and contains tens of well-covered individuals, increasing power for f_4-statistic computations. Toba and Wichí are modern populations from the Gran Chaco. Karitiana and Piapoco are modern populations from the Northwest Brazilian Amazon and Eastern Colombia, whose ancestry is characteristic of Tropical and Subtropical Forests Native American peoples[2,4]. Blue statistics show that Late Holocene Pampas populations cannot be related with Central Argentina and Middle Holocene Pampas via a simple tree, indicating gene flow between these two lineages. Red statistics show similar patterns for Northwest Argentina context labels in the case of the Central Argentina and Central Andes lineages. Green statistics show similar patterns for modern Gran Chaco populations for Central Argentina and the Forest and Subtropical Forests ancestry components.

Extended Data Table 2 | *hapROH* estimates of effective population size (*Ne*) by region, rounded to the nearest integer

Group	N_e point estimate	Lower bound of 95% CI	Upper bound of 95% CI	n
Argentina_Central	707	650	762	40.0
Argentina_Northwest	438	374	514	6.5
Argentina_Pampas	1100	828	1501	4.5
Argentina_ParanaRiver	518	406	678	3.0
Brazil_Coastal	245	217	278	6.5
CentralAndes	789	683	919	13.0
SouthernPatagonia	174	147	208	2.5

Estimates are obtained by fitting the distribution of runs of homozygosity of individuals from each region with a mean date not older than 3000BP. The estimates indicate that the communities in the Central region of Argentina likely had similar sizes as in the Central Andes, and likely higher than those in the Argentinian Northwest or the Paraná River region. The individuals from the Pampas had the largest effective population size, likely reflecting admixture.

nature portfolio

Corresponding author(s): Javier Maravall-López, David Reich, Rodrigo Nores

Last updated by author(s): Aug 10, 2025

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	OxCal v4.4, https://github.com/DReichLab/ADNA-Tools, BWA v0.7.15, Picard v2.17.10, contamMix v1.0-10, hapConX v1, ANGSD v0.940
Data analysis	ADMIXTOOLS v7.0.2, ape v5.8, stats v3.6.2, ADMIXTOOLS2 v2.0.0, smartpca v18711, SciPy v1.16.0, DATES v210, ADMIXTURE v1.23, PLINK v1.9, hapROH v0.64, SciPy v1.13.1, scikit-posthocs v0.9.0, R v4.3.2, dplyr v1.1.4, ggforce v0.4.2, ggnewscale v0.4.10, ggplot2 v3.4.4, ggspatial v1.1.9, ggstar v1.0.4, ggrepel v0.9.5, paletteer v1.3, raster v3.6-26, rnaturalearth v1.0.1, sf v1.0-15, tidyterra v0.5.2, terra v1.7-71, https://github.com/javiermaravall/aDNA_CSC

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

Data

Policy information about availability of data

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

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

Genotype data for newly-reported individuals included in main analyses from this study can be obtained from the Harvard Dataverse repository at doi.org/10.7910/

DVN/UQVPJQ. The aligned sequences for all individuals are available through the European Nucleotide Archive, accession PRJEB97713. Previously published data used in our analyses are available as follows: genetic data for modern individuals from Native American groups2 are available for non-profit research on population history under an inter-institutional data access agreement with the Universidad de Antioquia, Colombia (queries regarding data access should be sent to a.ruizlin@ucl.ac.uk); genetic data for previously-published ancient individuals is available at the Allen Ancient DNA Resource (doi: 10.7910/DVN/FFIDCW); 1000 Genomes haplotype reference panel (http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/), human reference genome hg19 (https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000001405.13/); data used for map plotting is available at Natural Earth (https://www.naturalearthdata.com), GADM (https://gadm.org) and Portal de Información Hídrica de Córdoba-APRHI (https://portal-aprhi.opendata.arcgis.com/). Other newly reported data, such as radiocarbon dates and archaeological context information, are included in this manuscript, the Supplementary Information, and Supplementary Data files.

Research involving human participants, their data, or biological material

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

Reporting on sex and gender	n/a
Reporting on race, ethnicity, or other socially relevant groupings	n/a
Population characteristics	n/a
Recruitment	n/a
Ethics oversight	n/a

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

Field-specific reporting

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

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

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

Ecological, evolutionary & environmental sciences study design

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

Study description	DNA samples newly-obtained from ancient human remains were co-analyzed with previously-published data from ancient and modern individuals. Using genome-wide SNP genotypes, researchers calculated population genetic statistics that primarily examine allele-sharing patterns to explore historical relationships among populations.
Research sample	We generated new genome-wide data from 238 not previously reported ancient individuals.
Sampling strategy	We produced 504 ancient DNA libraries from 341 distinct skeletal samples. We used in-solution enrichment for over 1.2 million targeted single nucleotide polymorphisms (SNPs), a standard set of genetic markers widely used in ancient DNA studies and commonly referred to as the "1240k SNP set", to gather genome-wide data that met standard criteria for ancient DNA authenticity from 238 unique individuals.
Data collection	DNA from the ancient remains was extracted, sequenced, and processed into SNV genotype calls.
Timing and spatial scale	Ancient individuals lived from 10,000 yBP to 150 yBP in present-day Central and Northern Argentina, and Paraguay.
Data exclusions	103 samples did not produce data of high-enough quality to analyze, either because of low coverage or because of evidence of contamination.
Reproducibility	All attempts to reproduce were successful.
Randomization	No randomization was possible due to the nature of the study, a reconstruction of past events that cannot be repeated.
Blinding	No blinding to dates or geographical origin was possible, due to the criticality of this information for analysis.

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

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

Materials & experimental systems

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

Methods

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

Palaeontology and Archaeology

Specimen provenance	All of the skeletal remains examined in this study were sourced from museum collections or other archaeological archives. We only analyzed the samples after receiving permission from the custodians and the relevant local authorities.
Specimen deposition	The skeletal samples are under the stewardship of the appropriate co-authors or museum collections., and may be accessed by their skeletal code listed in Extended Data Table 0.2.
Dating methods	We report 35 new radiocarbon dates obtained using standard techniques.

☒ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight	This study adhered to ethical guidelines for working with human remains drafted both by a diverse and international group of anthropological and paleogenetic scholars (Alpaslan-Roodenberg, S. et al. 2021, Ethics of DNA research on human remains: five globally applicable guidelines. Nature) and the Argentine Association of Biological Anthropology (Aranda, C., Barrientos, G. & Del Papa, M. C. 2014 Código deontolñogico para el estudio, conservación y gestión de restos humanos de poblaciones del pasado. Revista argentina de antropología biológica) treating these deceased individuals with respect and using minimally-destructive analyses techniques. Our research program involving ancient human remains received approval from the Ethics Committee of the CEMIC (Comité de Ética en Investigación, Centro de Educación Médica e Investigaciones Clínicas 'Norberto Quirno'). Skeletal samples were exported with authorization from the institutions safeguarding them (provincial and national museums, universities, etc.), obtaining proper permits from each province (e.g., Agencia Córdoba Cultura), and the Argentina government (Instituto Nacional de Antropología y Pensamiento Latinoamericano). In instances where indigenous communities were associated with these individuals, analyses were conducted in engagement with these communities (Salceda, S. A., Desántolo, B. & Plischuk, M. Espacio de reflexión: el por qué y para quién de la investigación bioantropológica. Revista argentina de antropología biológica 17, 1–6 (2015).), primarily facilitated through interactions between archaeologists and the communities. In the particular case of samples from the Córdoba province, we secured endorsement and support for this research from the Consejo de Comunidades de Pueblos Indígenas de la Provincia de Córdoba, Argentina (Council of Communities of Indigenous Peoples of the Province of Córdoba).

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

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*