Epifanía Arango-Isaza, Marco Rosario Capodiferro, María José Aninao, Hiba Babiker, Simon Aeschbacher, Alessandro Achilli, Cosimo Posth, Roberto Campbell, Felipe I. Martínez, Paul Heggarty, Scott Sadowsky, Kentaro K. Shimizu, Chiara Barbieri

The genetic history of the southern Andes from present-day Mapuche ancestry

Current Biology 33, 1-14, June 5, 2023 © 2023 Elsevier Inc

Arango-Isaza et al., 2023, Current Biology 33, 1-14
June 5, 2023 © 2023 Elsevier Inc.
https://doi.org/10.1016/j.cub.2023.05.013

The genetic history of the southern Andes from present-day Mapuche ancestry

Epifanía Arango-Isaza[1,2,*], Marco Rosario Capodiferro[3,4], María José Aninao[5], Hiba Babiker[6], Simon Aeschbacher[1], Alessandro Achilli[4], Cosimo Posth[7,8], Roberto Campbell[9], Felipe I. Martínez[9,10], Paul Heggarty[11], Scott Sadowsky[12], Kentaro K. Shimizu[1,2], Chiara Barbieri[1,2,6,*]

[1] Department of Evolutionary Biology and Environmental Studies; University of Zurich; Zurich, 8057; Switzerland
[2] Center for the Interdisciplinary Study of Language Evolution; University of Zurich; Zurich, 8050; Switzerland
[3] Trinity College Dublin; Dublin 2, Dublin; Ireland
[4] Department of Biology and Biotechnology "L. Spallanzani"; University of Pavia; Pavia, 27100; Italy
[5] Pontifical Catholic University of Peru; Lima; Peru
[6] Department of Linguistic and Cultural Evolution; Max Planck Institute for Evolutionary Anthropology; Leipzig, 04103; Germany
[7] Institute for Archaeological Sciences, Archaeo- and Palaeogenetics; University of Tübingen; Tübingen, 72074; Germany
[8] Senckenberg Centre for Human Evolution and Palaeoenvironment; University of Tübingen; Tübingen, 72074; Germany
[9] Escuela de Antropología; Pontificia Universidad Católica de Chile; Santiago, 6904411; Chile
[10] Center for Intercultural and Indigenous Research, Santiago, 7820436; Chile.
[11] 'Waves' ERC group, Department of Human Behavior, Evolution and Culture; Max Planck Institute for Evolutionary Anthropology; Leipzig, 04103; Germany
[12] Department of Linguistics and Literature; Universidad de Cartagena; Cartagena, 130001; Colombia

* Corresponding authors: Chiara Barbieri barbieri.chiara@gmail.com, Epifanía Arango-Isaza epifaniarango@gmail.com

Summary

The southernmost regions of South America harbor some of the earliest evidence of human presence in the Americas. However, connections with the rest of the continent, and the contextualization of present-day indigenous ancestries remain poorly resolved. In this study, we analyze the genetic ancestry of one of the largest indigenous groups in South America: the Mapuche. We generate genome-wide data from 64 participants from three Mapuche populations in Southern Chile: Pehuenche, Lafkenche, and Huilliche. Broadly, we describe three main ancestry blocks with a common origin, which characterize the Southern Cone, the Central Andes, and Amazonia. Within the Southern Cone, ancestors of the Mapuche lineages differentiated from those of the Far South during the Middle Holocene, and did not experience further migration waves from the north. We find that the deep genetic split between the Central and Southern Andes is followed by instances of gene flow, which may have accompanied the southward spread of cultural traits from the Central Andes, including crops and loanwords from Quechua into Mapudungun (the language of the Mapuche). Finally, we report close genetic relatedness between the three populations analyzed, with the Huilliche characterized additionally by intense recent exchanges with the Far South. Our findings add new perspectives on the genetic (pre)history of South America, from first settlement through to the present-day indigenous presence. Follow-up fieldwork took these results back to the indigenous communities to contextualize the genetic narrative alongside indigenous knowledge and perspectives.

INTRODUCTION

The peopling of the Americas represents the last of the major human migrations, beginning – according to the latest interpretations of the genetic data – no earlier than ~23 thousand years ago (kya) [1]. Historical admixture with European and African individuals and the devastating decline of indigenous populations caused by contact with Europeans hamper our ability to reconstruct ancient demographic history in the Americas. The patterns and timing of the initial migrations remain debated in population genetics and other disciplines [2].

From a genetic perspective, all non-Arctic Native American groups descend from an ancestral population that split into a northern and a southern branch while still within North America [3, 4, 5, 6]. Within the southern branch, an early wave related to a 12,600 BP individual associated with the Clovis culture (Anzick-1) [7] and labeled SNA1 spread rapidly southwards during the late Pleistocene (before 13,000 BP) [8, 9, 10]. A second wave, associated with a North American sample from 10,100 BP (Spirit Cave) and labeled SNA2, also entered South America possibly as early as the Late Pleistocene [9, 10].

The genetic profile of present-day indigenous populations of South America stems predominantly from this second migration wave, which differentiated into three main ancestries characteristic of three broadly defined ecogeographic regions: one primarily represented in the Andes, one in the Amazonian lowlands, and one in the Southern Cone [4, 11]. During the Holocene, further migration waves from North and Central America reached the Andes and Amazonia [12, 8]. Genetic studies in South America have so far focused on Amazonia or the Central Andes. In contrast, only a few recent genome-wide studies have addressed populations of the remaining macro-region, the Southern Cone [11, 13, 14]. In comparison, a rich literature based on uniparental markers revealed the presence of characteristic early-diverging lineages in the Southern Cone [15, 16, 17, 18, 19, 20, 21]. The lack of genome-wide data leaves open questions on how the earliest human migrations reached the south, which routes they took, and how much they interacted with subsequent migration waves. With our study, we focus in particular on southernmost South America, i.e. the Southern Cone more narrowly defined as modern Chile (excluding the northernmost regions), Argentina, and Uruguay.

Archaeological evidence points to the southern regions of South America as having been inhabited from the earliest stages of human settlement of South America. The site of Monte Verde in southern Chile, dated to at least ~14,500 cal BP, bears the earliest widely accepted traces of human presence in the continent [22]. The migration routes taken by the first settlers are debated: some argue that routes along the Pacific or Atlantic coasts are equally likely [15], while others strongly support the Pacific coast as the predominant route [16, 17]. The eastern Southern Cone could have been settled from a trans-Andean [16], Atlantic, or inland route [17]. Ancient DNA (aDNA) data from southern Patagonia shows genetic continuity from 6600 BP onwards, as well as differentiation between sea nomads along the Pacific coast, who relied on marine resources, and foot nomads in eastern Patagonia, who relied on hunting wild guanaco [13, 23]. In the Late Holocene (~ 3000 BP), more complex resource management and the first settlements appeared [24, 25]. More work is needed to link the archaeological and genetic evidence through the Holocene and to recent indigenous history in the southern regions of South America.

Today, the Mapuche represent one of the main indigenous groups of the Southern Cone, with a major presence in Southern Chile and small parts of Argentina. The archaeological record suggests population continuity from the first centuries CE to the groups encountered by Europeans in the 16th century [26]. Central Chile was conquered by the Inca Empire in the 15th century, and then came under Spanish control from 1541. Further south, a conflict known as the Arauco War affected the indigenous populations. By 1641 a

frontier was established, south of which the region known as “Araucanía” remained largely independent until Chilean and Argentinean conquest from the 1860s to 1880s.

Geographic and ecological factors draw the boundaries between several self-identified Mapuche territorial identities such as Lafkenche (people of the sea) along the Pacific coast, Pehuenche (people of the araucaria pine) in the Andean mountains, and Huilliche (people of the south) south of the Toltén river and into the Chiloé archipelago. Some groups in southern Chile, in particular, self-identify as just Huilliche, rather than as a Huilliche subgroup of the Mapuche people, to mark a separation from the broad Mapuche group. Mapudungun is considered a single broad language with various regional varieties (‘geolects’) [27]. Previous population genetics studies [4, 7] have followed an ethnolinguistic categorization of the Americas that placed Mapudungun in a supposed “Andean” macro-group [28]. This framework has been rejected within mainstream linguistics, however, for its lack of sound methodological foundation [29, 30]. In standard classifications, Mapudungun is a language isolate, with no demonstrable shared origin to any other language [31, 32]. Nonetheless, a small number of words (e.g. for *hundred*, *thousand*, *fish*, and perhaps also *sun*) that appear to be borrowings do imply that Mapudungun was once in sporadic contact with Quechua and Aymara from the Central Andes. The borrowings remain phonologically very similar to the source words, suggesting that these contacts do not predate the Inca period by long, if at all [33, 34, 35]. Further exchanges can also be traced with crops such as potato, quinoa, beans, squash, and maize [24, 25], domesticated in the Central Andes and then introduced to these southern regions. Gene-flow events may have accompanied the entry of loanwords and crops from the Central Andes: such demographic contacts could be dated with genetic analysis and provide a time frame for the exchanges.

Despite the distinct role played by the Mapuche in the indigenous history of South America, their internal population structure and relationships with other indigenous groups remain poorly understood. The origins of the Mapuche have been explained variously as 1) a local trajectory in continuity since the earliest occupation [26], 2) a migration from the Central Andes [36], or 3) a migration from the Amazonian rainforest or the Gran Chaco region [37]. Published genome-wide data from one Mapuche and one Huilliche population suggest that they are closely related to other ancient and modern Patagonian populations [11]. Overall, it remains unclear which major migration wave the Mapuche stem from, or whether they were affected by the population movements of the Holocene, with potentially major implications for our understanding of the peopling of the Americas.

To reconstruct the origins of Mapuche genetic ancestry and trace its trajectory within South America, we generated genome-wide SNP-chip data for 64 participants of Mapuche descent from two regions of Araucanía, and from the island of Chiloé. We analyze their genetic make-up, compare it with published data from Native American groups and aDNA, and set our findings alongside linguistic and archeological evidence. Our results reveal in more detail the nature of the genetic connections between ancient and living American populations, describe the recent demographic effect of European contact, and clarify the genetic relationships between different Mapuche groups.

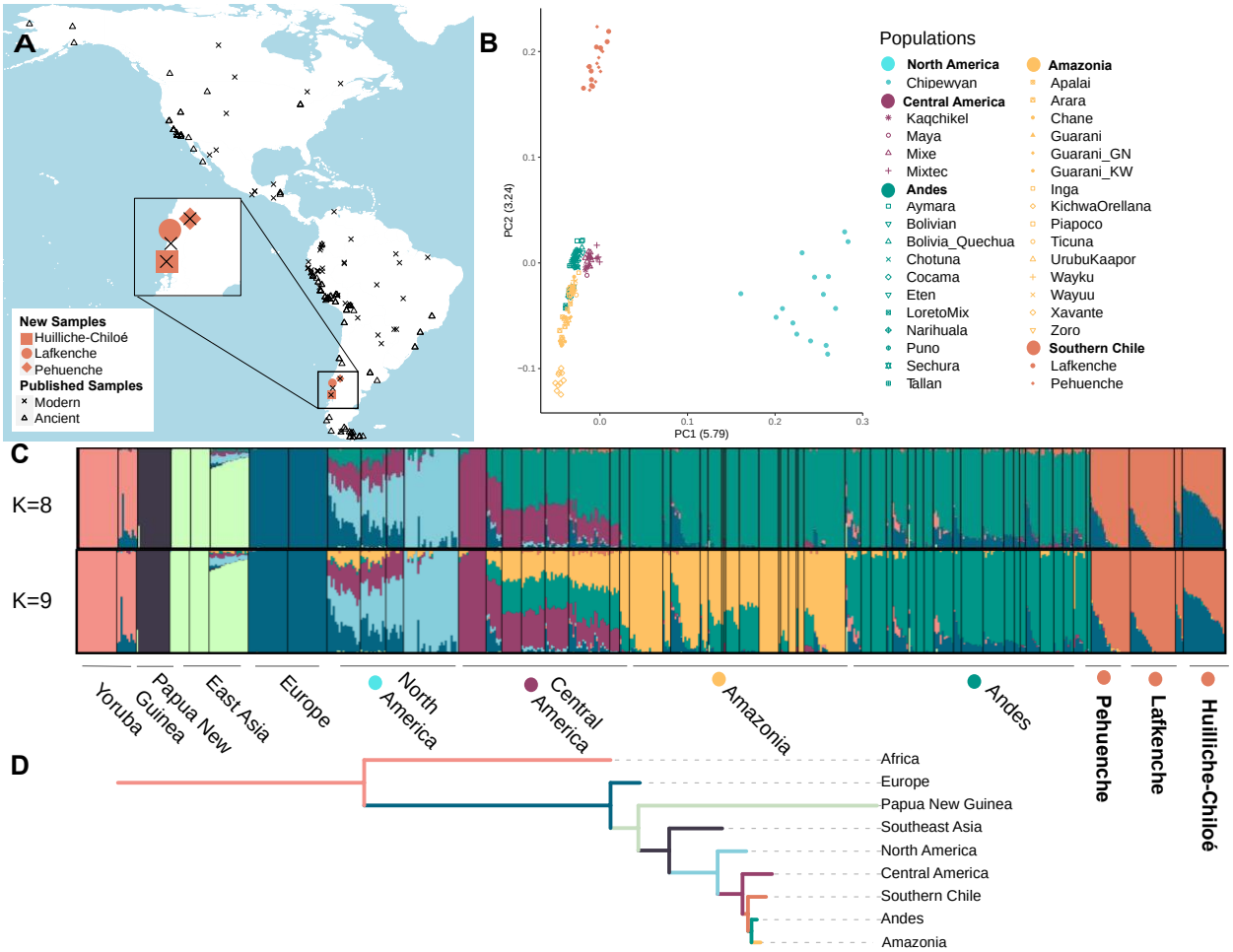


Figure 1. Sampling locations of newly genotyped individuals, reference individuals, and overall genetic patterns in the Americas. (A) Map of the Americas showing the approximate location of the newly generated samples and the selection of individuals from published literature, both modern (crosses) and ancient (triangles). **(B)** PCA analysis including only "unadmixed" individuals from the Americas, defined as having at least 99.9% of Native American component, as inferred by ADMIXTURE at K = 8 (Dataset 1.3; Figure S1.B). Only 20 individuals from the Lafkenche and Pehuenche populations passed this filter and no individuals from the Huilliche-Chiloé population. Color coding corresponds to broad macro-regions of the Americas (see Figure S1.A). **C)** ADMIXTURE run for the global Dataset 1 (Figure S1.B) for K = 8 and K = 9. Runs from K=2 to K=15 are available in Figure S2. For the ADMIXTURE run with Dataset 1.2. see Figure S3. **D)** NJ tree of covariance-derived distances among ancestry components for K = 9 (Dataset 1), computed with OHANA based on the global ADMIXTURE run with the highest likelihood score.

Results

Overall genetic patterns in the Americas

We generated genome-wide data with the Axiom Human Origins SNP array [38] from the following three populations: Pehuenche from the mountains of Araucanía, Lafkenche from the coast of Araucanía (both groups who recognize their ancestry as Mapuche), and Huilliche-Chiloé, a population from the island of Chiloé which in part recognizes its ancestry as Huilliche. The label Huilliche-Chiloé is to distinguish them from Huilliche groups on the mainland (Figure 1.A). We merged the genotypes of these individuals with modern publicly available data from relevant populations (Dataset 1, 2 and 3) [38, 39, 40, 41] and with ancient DNA data (Dataset 3.3; see Figure S1.B for a schematic description of the different datasets used) [6, 7, 8, 11, 12, 13 , 38, 39, 40, 41, 42, 43].To understand the global pattern of genetic relatedness among modern samples (Dataset 1), we used ADMIXTURE [44]. The ADMIXTURE analysis showed the lowest cross-validation (CV) errors for *K* = 8 and *K* = 9 (Figure S2.B). Major ancestry components specific to the Americas (Native American ancestry) start to differentiate from *K* = 6 (Figure S2.A). At *K* = 7 a component emerges that is prevalent in our Mapuche sample, and which we refer to here as the Southern

Chilean (*SC*) component (orange in Figure 1.C). This *SC* component is also sporadically present in the Andes. The Mapuche individuals sampled have a variable percentage of European admixture (dark blue in Figure 1C), ranging from averages of 9.2% in the Lafkenche and 13.3% in the Pehuenche, to 43.4% in the Huilliche-Chiloé (at K = 8). Dataset 1 includes four individuals from a previous publication labeled as Chilote (i.e., from the island of Chiloé) [39], which show a similar admixture profile to our Huilliche-Chiloé sample. We ran an analogous ADMIXTURE analysis with a dataset that included the modern samples from De la Fuente et al., 2018 [11] but with fewer overlapping SNPs (Dataset 1.2; Figure S1.B). The results are consistent with Figure 1; our Mapuche samples are genetically similar to the Pehuenche sample from De la Fuente et al. (Figure S3.A). Populations in this dataset from the Far South (*FS*, defined here as south of ~50°S), namely Yámana and Kawéskar, display a characteristic component at K = 12. The relationship between the ADMIXTURE components is visualized with a Neighbor-Joining (NJ) tree, which supports all South American ancestries branching from each other closely in quick succession (Fig 1.D).

To exclude historical gene flow from Europe, we retained only “unadmixed” individuals (here defined as having 99.9% Native American ancestry as computed by ADMIXTURE at K = 8, Dataset 1.3) and performed a PCA on this subset (Figure 1.B). Here, the first component separates North and South American groups, and the second component separates *SC* and Amazonian groups, while Central American and Andean populations remain close to the Amazonian ones.

Both ADMIXTURE and PCA suggested that the *SC* populations are genetically distinct from the rest of South America. The marked differentiation of the *SC* component could imply an early divergence in the population structure of South America, but could alternatively be due to an overrepresentation of Mapuche individuals in the dataset, or to a recent bottleneck associated to strong genetic drift in small, isolated *SC* groups [45]. An analysis of the degree of consanguinity and the distribution of runs of homozygosity (ROH) (Fig 2) show that our *SC* individuals are not especially high in homozygosity compared to other populations of South America, ruling out recent strong drift as the predominant explanation for this genetic divergence. A mixed scenario with early divergence and moderate drift is also plausible.

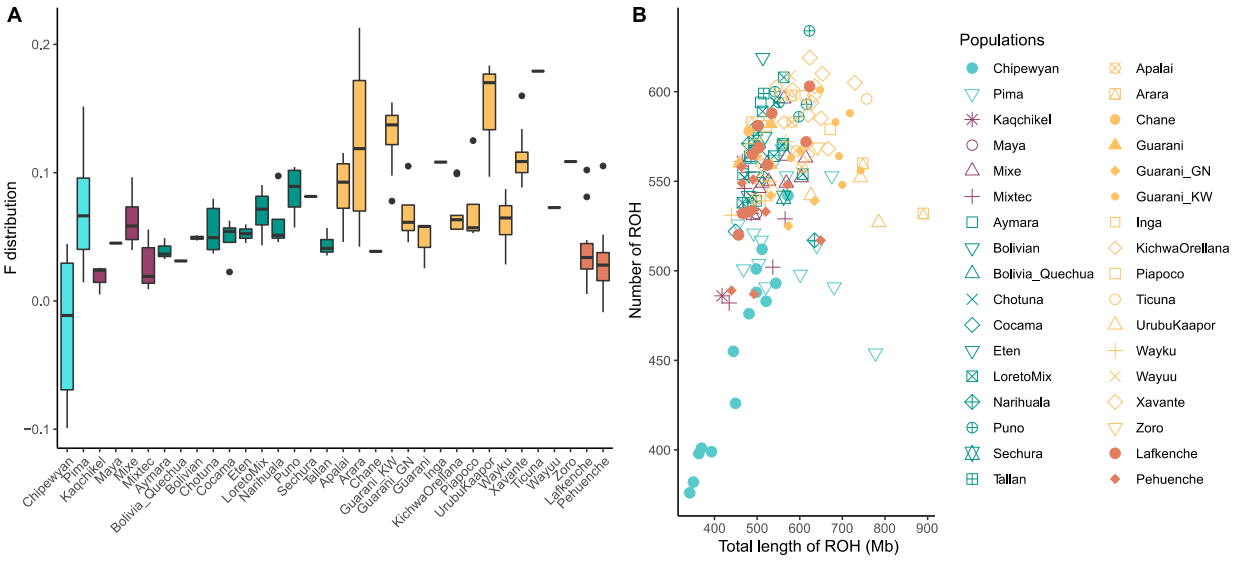


Figure 2: Within-population diversity. (A) Individual values of consanguinity (F) averaged for each population from South America. **(B)** Distribution of ROH fragments. Both analyses are computed on the "unadmixed" individuals (Dataset 1.3)

Recent demography and connectivity

To further investigate the structure of the Mapuche populations in relation to the rest of the continent, we looked at recent demography (~3 ka) and gene-flow patterns using shared Identity by Descent (IBD) blocks, inherited from the same common ancestor. Due to recombination, the length of IBD blocks shared by two populations decays with time since these populations split [46, 47], but admixture may (re)introduce shared IBD blocks [48]. The individuals from the Southern Cone are connected to each other by a dense network of IBD sharing (Figure 3.A). The Mapuche populations, and in particular the two Pehuenche populations from this and a previous study [11], show high shared ancestry with each other, while the Huilliche-Chiloé population shares more blocks with Chilote (from [39]) and with the distant Kawéskar and Yámana (from [11]) than with the neighboring Lafkenche and Pehuenche. Across South America as a whole, three broad networks of shared ancestors roughly correspond to the regions where the three main ancestries from our ADMIXTURE analysis are represented: Andes, Amazonia, and Southern Cone. The three regions share a significant number of blocks, especially between the Andes and Amazonia. In contrast, the southern regions are less integrated into this network of shared IBD blocks, which suggests a higher degree of isolation. We found only one persistent link between our *SC* samples and the northern Andes, and a similar link to the Gran Chaco region (Figure 3.A-B, Figure S4.A-B). This analysis is also performed with fragments of Native American descent, identified with the masking process that filtered out variants of possible African and European origin (Figure S4.D-E). The overall pattern is consistent with Figure 3B, confirming a connection between Mapuche populations and the Andes for fragments smaller than 10 cM.

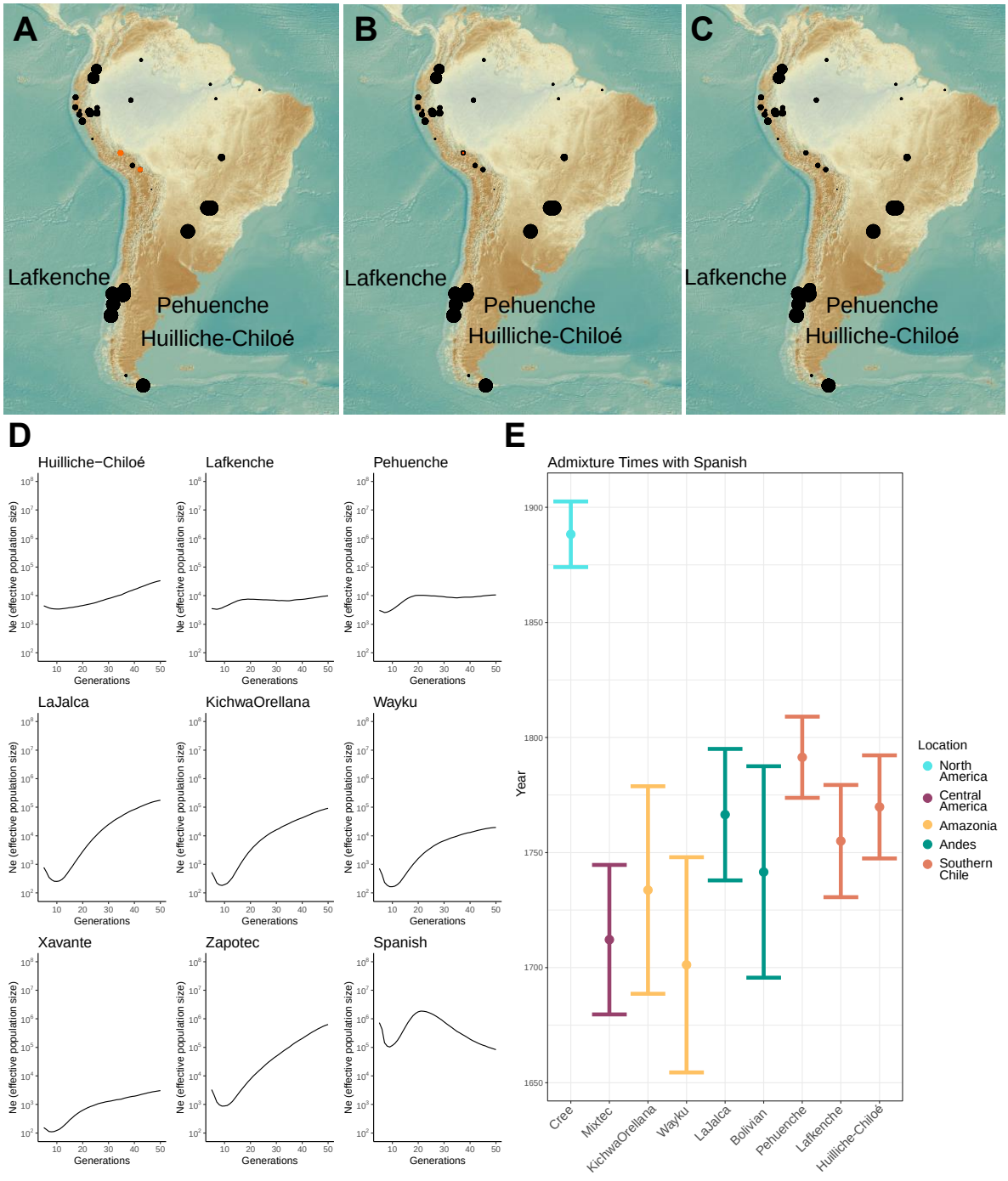


Figure 3: Recent demography and connectivity. IBD sharing probability network among South Americans (Dataset 2.2). The network represents the probability of a pair of individuals from populations A and B sharing an IBD fragment, adjusted by population size. Thicker width and lighter orange color of the lines correspond to higher exchange between populations. The size of the black circles is proportional to sample size. **A)** Shared fragments from 4 to 7 cM. For visualization purposes, only population pairs with a probability of sharing higher than 10% are considered. **(B)** Shared fragments from 7 to 10 cM. For visualization purposes, only population pairs with a probability of sharing higher than 2% are considered. **(C)** Shared fragments longer than 10 cM. **(D)** Variation in effective population size for selected Native American and Spanish populations over the last 50 generations, calculated with IBDNe. **(E)** Estimated admixture times of selected Native American populations with a Spanish source, calculated using ALDER. The error bars represent a 95% confidence interval (generation time: 28 years, only the most recent admixture pulse is reconstructed with this method). Matrix visualization of IBD sharing is available in Figure S4.A. IBD sharing Native American ancestry specific markers is available in Figure S4.B.

The length of the shared IBD blocks has been correlated with the number of generations back to shared ancestors in previous studies of European [46], Asian [49], and Native American populations [50]. In these studies, fragments between 5 and 10 cM have been associated with sharing events occurring 500-1,500 years ago. In a recent study, Ioannidis et al. [51] dated a gene-flow event from indigenous Americans into Polynesians at around

1200 CE and independently matched it with IBD blocks longer than 7 cM. However, it is important to note that patterns of IBD sharing can also be affected by population-specific histories. With this caveat, and considering the data from other studies as an indicative reference, the IBD sharing between *SC* and the Central Andes may date back more than ~500 years ago, as we do not find shared fragments longer than 10 cM.

The analysis of IBD fragments among individuals within a population can also give insights into variations in effective population size (N_e) [52]. We inferred the demography of populations of the Americas represented by large sample sizes (min 7 individuals) and of the Spanish population as a reference from outside the Americas (Figure 3.D). Before ~30 generations ago (~ 840 years), the three Mapuche populations are characterized by a relatively small and constant N_e. Starting around 15 to 20 generations ago, all Native American populations underwent a severe bottleneck which corresponds to European colonial impacts (including pathogens) and the ensuing historically documented population decline, to its lowest point at 10 generations ago. This bottleneck is not strongly visible in the three populations of Mapuche ancestry, contrary to reports by other studies based on simulation methods [53].

Finally, we used ALDER to estimate the date of admixture with Europeans (Figure 3.E). This software is based on linkage-disequilibrium and is most sensitive to recent large admixture events, ignoring minor admixture episodes and multiple admixture pulses that could have occurred previously [54]. The inferred admixture times vary across populations. The Wayku in lowland Peru show the earliest estimate (late 16th century), whereas the Cree in North America show the most recent estimate (late 19th century). The Mapuche populations show an intermediate admixture time to the mid-18th century (Figure 3.E).

The ancient population structure of South America

To focus on indigenous ancestry, we performed masking to filter out variants associated with European or African descent. To check the performance of the masking process, we ran f_4-tests and PCA (Figure S5.A-C and https://github.com/epifaniarango/popgen_with_epi/tree/Local-ancestry-and-masking). We then merged masked individuals with a selection of ancient samples from the Americas into a new dataset (Dataset 3.3, see Methods and Supplementary Data S1.A). On this dataset, we ran ADMIXTURE [44] from K = 2 to K = 10. K = 5 to 7 were associated with the lowest CV errors (Figure S5.D-E), and K = 5 had the least variance between runs. At K = 3, a distinct component emerges that is present primarily in the ancient and modern *SC* samples (Figure S5.E).

We take the results of the ADMIXTURE analysis at K = 5 and explore the genetic relationships across geographic locations and time scales. Most samples older than 6500 BP (i.e., from the Late Pleistocene to the start of the Middle Holocene) harbor all five ancestry components, but at varying proportions (Figure 4.A). Only the two Andean samples have a single predominant component, which persists at high frequency in the Andes through all later periods. Between 6500 and 1500 BP, we observe an increasing differentiation between the central Andes and the Southern Cone (Figure 4.B). The samples from Central Chile and the Far South are structured and non-homogenous, with the *SC* local component at varying proportions. In the most recent period represented by aDNA (1500 BP to ~100BP), the three major ecogeographical regions appear more differentiated, although Amazonia is poorly represented. (Figure 4.C).

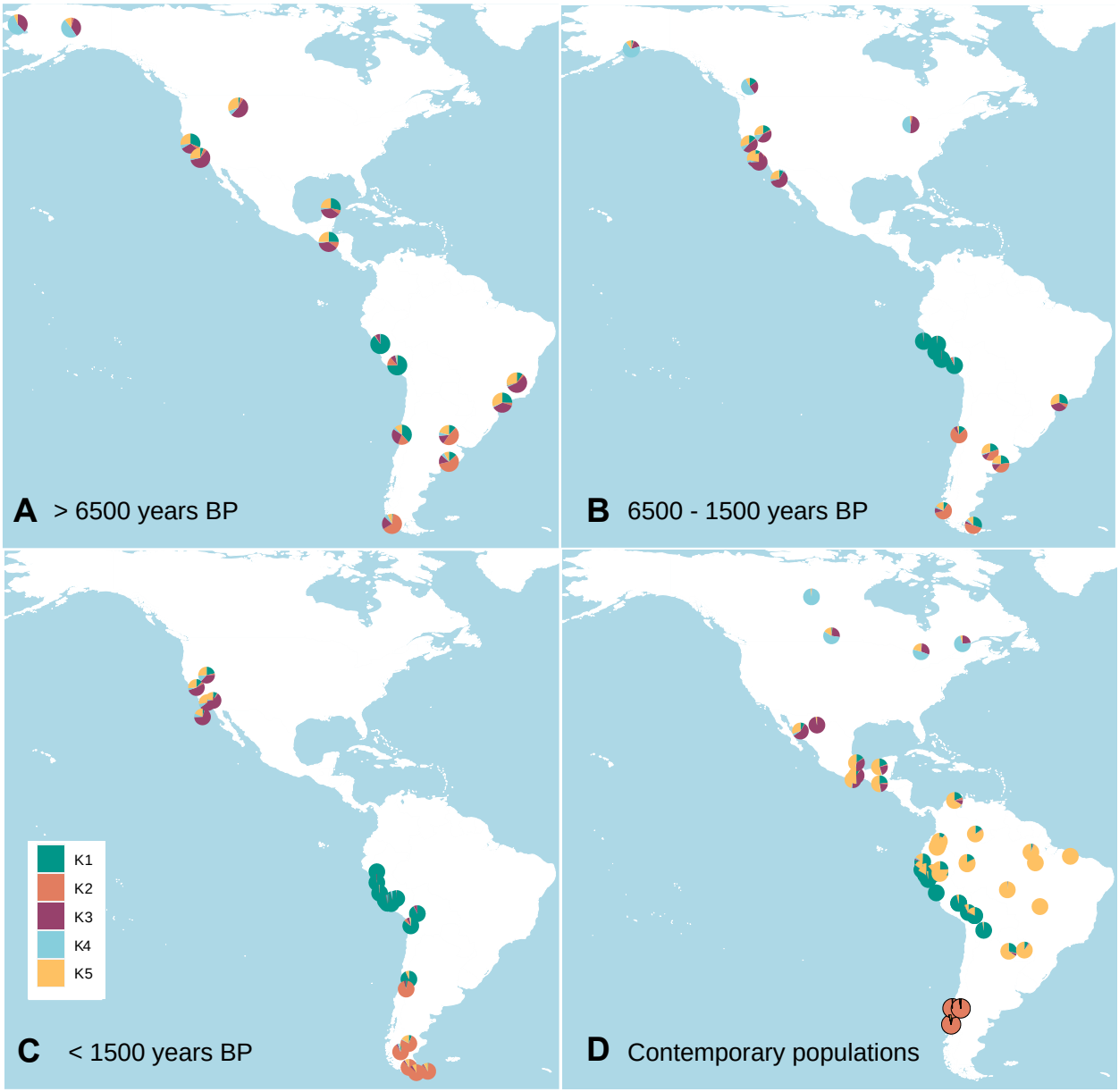


Figure 4: The ancient population structure of South America. ADMIXTURE run with ancient and masked individuals (Dataset 3.3) for K=5. Modern samples are grouped by population (in panel D), and ancient samples by proximity in geographical space and time. **(A)** Up to 6500 BP; **(B)** 6500 to 1500 BP; **(C)** After 1500 BP but not contemporary; **(D)** Contemporary. For details about the masking procedure see Figure S5.

Populations in North and Central America, and the most ancient individuals of the dataset show the highest number of different ancestry components. We do not interpret this strictly as an admixture event between distinct ancestries but as a characteristic of the initially undifferentiated gene pool, which drifted as the populations were migrating southwards. Finally, among present-day populations, the Andes, Amazonia and Southern Cone are clearly distinguished from each other, with various degrees of admixture in the contact zone between Amazonia and the Andes (Figure 4.D).

An alternative perspective on the genetic prehistory of the Americas can be gained from outgroup f_3-statistics [5], which estimate shared genetic drift among two populations relative to an outgroup. We selected commonly used transformations of f_3-statistics (NJ and MDS) to visualize the shared genetic history (Figure 5).

The NJ tree suggests an early split of the North and Central American samples (Figure 5.A), which is in line with the North-to-South population expansion through the Americas [1, 4, 7, 12, 8, 6]. The next groups to branch off independently are the Amazon and the Andean

clades. A further clade includes the most ancient individuals of the dataset, which have been associated with SNA1 [8, 12] (here defined as “Anzick-related”). The Southern Cone samples then divide into three sub-branches: ancient Argentinean samples from the Pampas; ancient Central Chile (*CC*) and modern Southern Chile (*SC*); and ancient Far South (*FS*). The three *SC* populations are closely related and share drift mostly with ancient samples from the Pampas and with *CC* (*Conchalí_700BP* and *LosRieles_5100BP*), and overlap with them in the MDS plot (Figure 5.B).

Population history of the Southern Cone

We used f_4-statistics to study changes in genetic structure over time. F_4-statistics measure the shared genetic drift between a set of four populations/individuals to provide insights into population contact. We explored allele sharing within Mapuche groups using the configuration f_4(Mbuti*, X; Y, Z)*, where *X* is any ancient or modern population, and *Y* and *Z* are two populations taken from Lafkenche, Pehuenche, and Huilliche-Chiloé. As expected, our *SC* samples show high and symmetrical allele sharing among each others (Z-score ~ 0, Supplementary Data S1.C). Of all the populations/individuals tested in position *X*, only the *Conchali_700BP* samples and, to a lesser extent, the *ArroyoSeco2_7,700BP* sample showed shared drift with Lafkenche (Figure 6.A), when Lafkenche is paired together with Huilliche-Chiloé in position *Z* (for details, see https://github.com/epifaniarango/popgen_with_epi/tree/Dstats-plots).

We further explored the relationship between *SC* and Conchalí by computing f_4-statistics of the form f_4(Mbuti, [targeted ancient or modern sample]; *Conchali_700BP*, Modern *SC*). All the three *SC* Mapuche populations consistently show greater affinity to *Conchali_700BP* than to any other ancient or contemporary group (Z-score < |3|, Supplementary Data S1.D), with the exception of *LapaDoSanto_9600BP* which shows affinity to Conchalí, but only when all SNPs are used (Z-score < -3.3).

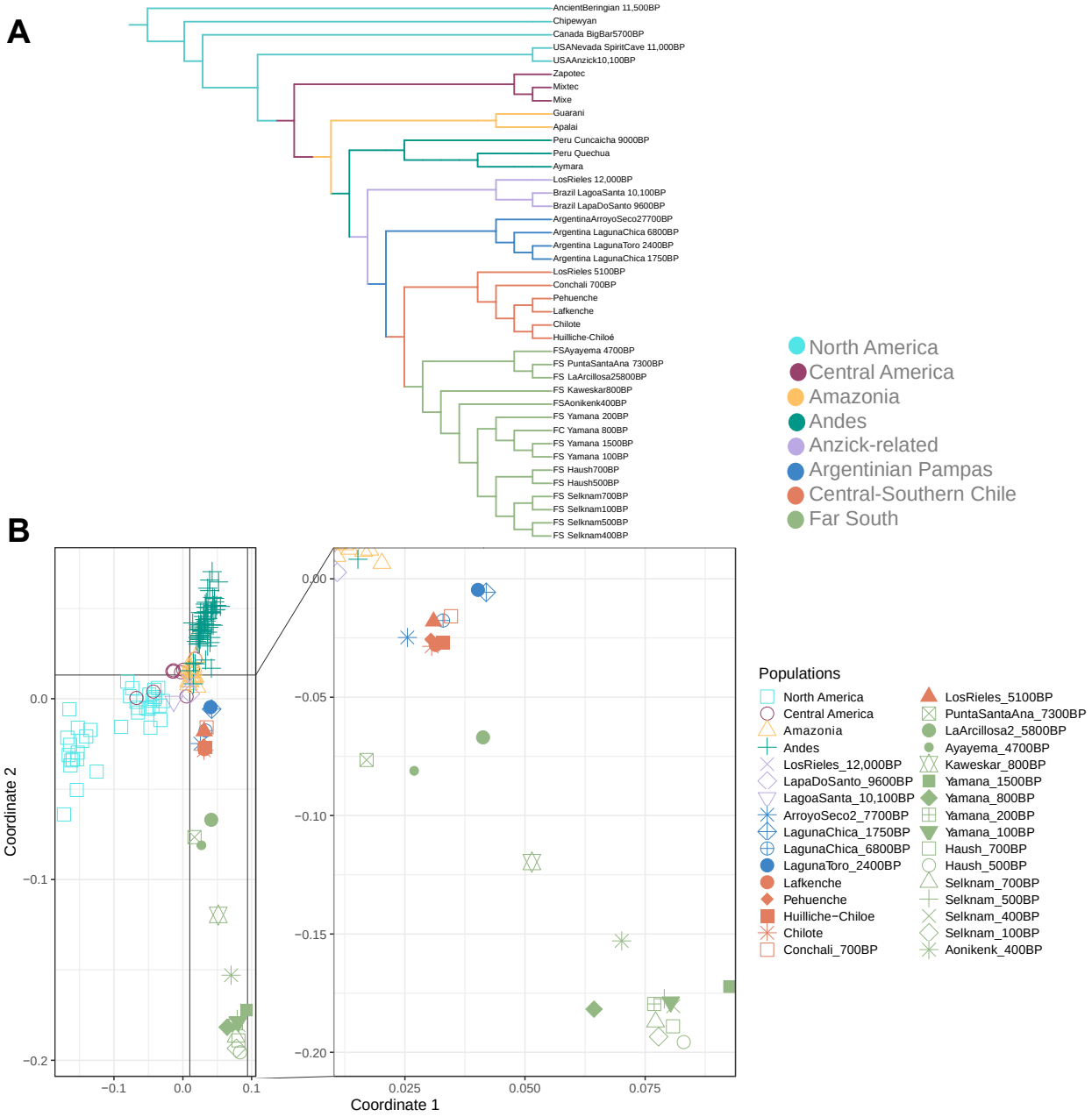


Figure 5: F_3 statistics analyses. (**A)** NJ tree of the matrix of inverted outgroup f_3-statistics (1/ f_3(Mbuti; X, Y)) using Ancient Beringian as an outgroup. Ancient samples are filtered for a minimum of 100k SNPs. The tree is a graphic simplification which does not include all populations/samples and which displays the cladogram without information on branch lengths. The original tree with all samples and branch length to scale can be found in Figure S6. Color code corresponds to broad regions and time transects. (**B)** MDS plot of the matrix of 1- f_3(Mbuti, X, Y). The blow-up to the right zooms in on the Far South, Argentinean, and Central-Southern Chilean samples.

A previous study related *Conchali_700BP* to Late Holocene *FS* samples using the test f_4(Mbuti, *Conchali_700BP*; Middle Holocene *FS*, Late Holocene *FS*) [13]. We repeated this test, also with our Mapuche *SC* populations in the position of Conchalí. We obtained higher f_4 values with our *SC* Mapuche populations than with Conchalí, in most combinations tested (Supplementary Data S1.E, Figure 6.B). We consistently found significant Z-scores (>3.3) in those combinations which involved more recent Late Holocene *FS* individuals, i.e. those dated between 400 and 200 BP. This result suggests that the genetic ancestors of the Mapuche were involved in contacts with the Late Holocene *FS*, either through the same contact event described in literature for Conchalí, or possibly with a further, more recent contact event between recent ancestors of Mapuche and *FS* populations. We also searched for evidence of Late Holocene gene-flow between

SC and other areas of South America (Argentinean Pampas and Central Andes) using f_4(Mbuti, SC/*Conchali_700BP*; [Middle Holocene Argentinean Pampas/Central Andes], [Late Holocene Argentinean Pampas/Central Andes]), but found none (Supplementary Data S1.F-G).

With the configuration f_4*(Mbuti, X; SouthernCone 1, SouthernCone 2)*, where *X* is any ancient or modern sample outside the Southern Cone, we did not find evidence of significant shared drift from other regions (Z-scores < |3|). suggesting a single origin for the Southern Cone populations. Only a few configurations involving the Middle Holocene Pampas as one of the Southern Cone populations are significant (Supplementary Data S1.H, see https://github.com/epifaniarango/popgen_with_epi/tree/Dstats-plots). The Middle Holocene Pampas individuals display a more distinctive genetic profile within the Southern Cone, which could be tentatively related to other substrates.

We also explored the formation of ancestries and admixture events affecting the Southern Cone using admixture graph modeling with qpGraph. We created an initial scaffold using individuals representing the three Southern Cone clades, plus *LapaDoSanto_9600BP* (associated with SNA1), and two outgroups (African Mbuti and Ancient Beringian USR1). The best-fitting topology showed a split among the main three clades Central-Southern Chile (*LosRieles_5100BP*), Pampas (*ArroyoSeco2_7,700BP*) and Far South, with the Far South comprising the two lineages associated to the sea nomads (*Ayayema_4700BP)* and foot nomads (*LaArcillosa2_5800BP)*(Figure S7.A). We then added Late Holocene samples from *CC* and *FS* (Figure S7.C-D). The *FS* sampled are modeled by a strong contribution from the lineages of Central-Southern Chile, confirming the f_4 results (Supplementary Data S1.E). In the last step, we added our modern *SC* Mapuche samples. Their genetic profile stems from a common ancestor with *Conchali_700BP* (Figure S7.E).

The qpGraph scaffolding in some configurations requires admixture edges that are not clearly supported by direct f_4 tests. For example, to model the genetic profile of *Conchali_700BP and LosRieles_5100BP*, an admixture edge is required from an ancestral population close to *LapaDoSanto_9600BP*. To model the Mapuche, an admixture edge is required from a population related to *ArroyoSeco2_7,700BP* , which is supported in the f_4 described above with a Z-score of 3.174 (Figure 6.A, Figure S7.E, Figure S7.G). (See https://github.com/epifaniarango/popgen_with_epi/tree/Dstats-plots for further discussion on the f_4 tests). When adding the Huilliche-Chiloé population, the best configuration requires a substantial (10%) admixture edge from an ancestral population at the root of the South American lineages (Figure S7.H). This effect could be due to the Huilliche-Chiloé having a higher proportion of European ancestry than the other Mapuche populations, which results in higher missingness after the masking. Alternatively, some European ancestries could have eluded the masking, resulting in a signal of gene flow from Eurasia. Nevertheless, configurations where the admixture edge comes from nodes upstream to the entry into the Americas return a worse fit, favoring a Native American source (Figure S7.H). Possible sources of bias in these discrepancies between qpGraph and f_4 statistics include having modern and ancient samples modeled together, the small number of SNPs available when merging ancient samples with masked modern samples, and the presence of ancient samples sequenced with a different technology (Shotgun sequencing vs. SNP capture).

In the scheme of Figure 6.C-D, we reconstruct a possible scenario for the formation of ancestries in the Southern Cone, considering possible ancestry divergence and admixture pulses as reconstructed from f_3, f_4 and qpGraph analysis.

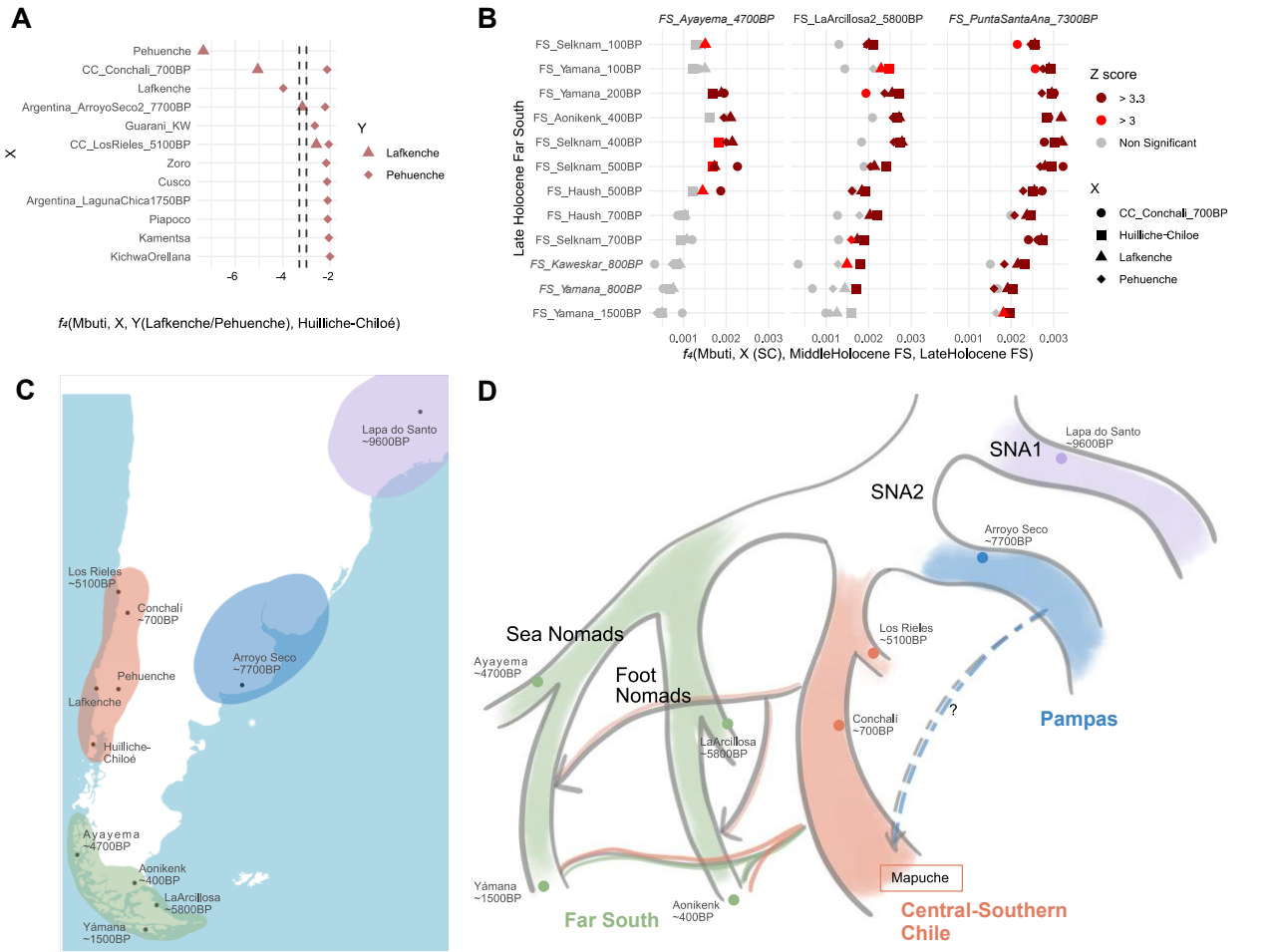


Figure 6: Population history of the Southern Cone. (A) f_4 tests exploring allele sharing between South American samples (ancient and modern) and present time SC Mapuche. In the y axis, only the individuals/populations associated to a Z-score < -2 are displayed. Two vertical dashed line mark the significance thresholds of -3 and -3.3. **(B)** f_4 tests exploring connections between Central-Southern Chile ancient and modern individuals and Late Holocene *FS*. F_4 values are plotted in the x axis. Plot marks are colored by the significance of the test based on Z-scores. Individual names in italics distinguish those samples sequenced with a shotgun technology from the rest of the samples genotyped with a capture approach. **(C)** Geographical location of key samples and possible geographic distribution of main genetic lineages of the Southern Cone. **(D)** Schematic representation of the most important regional lineages shaping the genetic landscape of the Southern Cone, summarized from the f_3, f_4, and qpGraph results in Figure S7. Admixture edges between Late Holocene FS and Central-Southern Chile refer to the f_4 in panel B. The connection between the Pampas and Mapuche, marked with a question mark, refers to the f_3 results in Figure 5B, and is partially supported in the f_4 configuration in panel A.

Discussion

Southern ancestry formed through isolation from the rest of South America

Our findings allow us to trace the formation of the Mapuche ancestry in the wider context of genetic diversity across South America. As previously noted [11], Mapuche ancestry belongs to a genetic cluster characteristic of southern South America. We observe that this ancestry cluster is equally distinct from the two other main clusters, which characterize the Central Andes and Amazonia, respectively (Figure 1.B-D). The split between these ancestry clusters could trace back to the early Holocene: simulation studies dated the split between Mapuche and Andean Aymara at 8750BP [53]. Our IBD sharing profiles between modern samples suggest that the Southern Cone has been developing in relative isolation from other regions (Figure 3.A-C). Small, isolated groups could have persisted until ~1000 years ago, hence their relatively small N_e (Figure 3D). Homozygosity and ROH data suggest that this isolation pattern is not confounded by high levels of recent consanguinity within populations (Figure 2).

A recent publication has proposed that genomic variation within South America during the Pleistocene derives principally from two ancestries, labeled SNA1 and SNA2, and that the genetic ancestry of South America emerges mostly from SNA2 [10]. Between the Middle/Late Holocene, the Central Andes and Amazonia admixed with further migration waves from the California Channel and/or Central America [5, 12, 8]. Based on a f_4 analysis of the single ancient sample from *Ayayema_4700BP*, Moreno-Mayar *et al.* proposed that these waves did not reach the Southern Cone [12]. We can now extend this finding to the whole Southern Cone, consistently over most of the Holocene and into present-day populations (Supplementary Data S1.H-I).

Our results suggest that the ancestry of the Mapuche, and of the rest of Southern Chile and the Far South, originates in local continuity from an early migration wave (SNA2), followed by relative isolation. This consistent genome-wide pattern matches other evidence of relative isolation, such as the presence of early diverging lineages coming from uniparental markers [15, 16, 17]. This finding has important repercussions for our understanding of the cultural and demographic background of these populations. First, it excludes any extensive pan-Andean regional development, as assumed in the putative "Andean" population group claimed in previous analyses [4, 7]. Second, it excludes a major external population source for changes associated with the adoptions of pottery and crops in Central-Southern Chile during the Late Holocene [36, 37]. Third, it corresponds with the status of Mapudungun and the extinct Chono language as language isolates [32]. Mapudungun has long attracted speculation that it could be related to other languages in the Americas, not only those further north in the (Central) Andes, but alternatively to Arawak, Tupí and even Mayan; however, there is no accepted linguistic support for any of those claims [55]. Our finding of the genetic isolation of the Mapuche is compatible with the standard linguistic view that Mapudungun is indeed a language isolate. Finally, the lack of any evident serial founder effect cannot straightforwardly be associated with either a Pacific/Andean or an Amazonian route for first settlement. This does not directly match the scenario of a single major route along the Pacific coast, as hypothesized from archaeology and mtDNA [16, 22].

Interaction between lineages of the Southern Cone

In the finer-scale structure of the Southern Cone through time (Figure 5), our f_3 results distinguish three main lineages: Argentinean Pampas (*ArroyoSeco2_7700BP);* Central-Southern Chile (*LosRieles_5100BP),* including the modern *SC* samples; and *FS* (*LaArcillosa2_5800BP* and *Ayayema_4700BP)* (Figure 5.A, Figure 6C-D and Figure S6). The appearance of population structure in the region could correspond with the warming climatic conditions of the Holocene, which peaked around 8.5 kya [56, 57]. The ancient *CC* individuals (Conchalí in particular) confirm the presence of Mapuche ancestry in regions further north than those where most Mapuche live today.

In the Far South, we confirmed the presence of two sub-lineages associated with the sea nomads and foot nomads, respectively [13]. The first lineage, in the western archipelagos, was genetically related to *Ayayema_4700BP*, and is associated with groups such as the Chono, Kawéskar, and Yámana. The second lineage, in eastern Patagonia, was genetically related to *LaArcilosa2_5800BP*, and is associated to other groups such as the Selk'nam, Haush (also known as the Manek'enk) and Tehuelche, who spoke languages of the Chonan family (note, not related to Chono in the west). Both the sea nomads and foot nomads experienced admixture from the Central Chile ancestry described here, to which present-day Mapuche are closely related (Figure 6.B).

Differences between Mapuche groups: connections with the Central Andes and with the Far South

Multiple lines of evidence suggest past connections between the Central Andes and Central-Southern Chile: 1) cultivated plants that appeared in Central-Southern Chile during the Late Holocene [24, 25], 2) evidence from Quechua loanwords in Mapudungun [33, 34, 35], and 3) historical contact with the southernmost expansion of the Inca Empire**.** We investigated demographic connections with f_4 statistics and IBD block sharing. The f_4 statistics do not support preferential allele sharing with the Central Andes, in contrast to the results of another study on a present-day Huilliche sample [58]. Nevertheless, our IBD analysis (Figure 3 and S4) does show a subtle but robust signal of shared ancestry, but which does not persist into the most recent time-frame considered (IBD fragments over 10 cM). This is compatible with contacts that may have predated the Inca period and may have brought not only crops and loanwords southwards, but a faint genetic legacy too [16].

The three *SC* populations analyzed are genetically closely related to each other. This is in line with Mapuche territorial identities being shaped essentially by geographic residence rather than by different demographic histories, and with their limited linguistic divergence. The fact that together, regional varieties of Mapudungun still form a coherent single language [27] is consistent with a relatively recent common origin, followed by geographical expansion and divergence over a time-scale of the order of many centuries, but not millennia. However, earlier population structure could have been altered by Mapuche relocating southwards first because of the Arauco War and then the so-called "Pacification" of Araucanía [59], as well as by long-term internal migration driven by forced population transfers, economic necessity, imposed restructuring of land ownership, and other factors [60]. In the face of such pressures, local communities also formed alliances that led to the absorption of previously unrelated groups [61, 62]. This fusion of genetically stratified groups would have increased the population diversity of the current Mapuche groups (Figure 2), mirroring the effect of a relatively constant effective population size for the *SC* populations (Figure 3.D). A similar effect of recent ethnogenesis through the fusion of structured populations has been suggested to account for the IBDNe profiles of Mexican populations [63]. Our result contrasts with the decimation of indigenous groups reported in historical sources [61, 64], and with the results obtained by Lindo *et al.* 2018 [53]. The different demographic trends obtained by Lindo *et al.* 2018 [53] could result from the different methods employed, and could be further explored with simulations on a high-resolution dataset.

The IBD analysis shows a high level of shared ancestry between our Pehuenche and Lafkenche samples, suggesting that the two groups differentiated only recently or have continued in close contact for generations. A slightly different genetic profile emerges for Huilliche-Chiloé, distinguished by its high level of IBD sharing with *FS* populations, indicatively dated at 500-1000 BP (Figure 3.A-C). Historical sources report strong migration waves from Chiloé into southern Patagonia for economic reasons, during the 19th century, but these recent migrations cannot explain the sharing of shorter fragments (<7 cM) which date back to an earlier time-frame [65]. A study of mtDNA haplogroups in Chiloé found a composition more similar to ancestries in southern Patagonia than among the Pehuenche and Lafkenche, suggesting a connection to the Chono people, who occupied the Chonos Archipelago, southernmost Chiloé and the coast around the Gulf of Corcovado [66, 67], and are associated with the sea nomads. The putative Chono toponymy further north through Chiloé suggests the Chono and Huilliche could have come into contact there.

Integrating genetic results with indigenous Mapuche perspectives

Our genetic results reconstruct the genetic trajectory of Mapuche ancestors back to the first peopling of the Americas. Set into their archaeological and linguistic contexts, our findings enrich and complement the historical records and local narratives of the indigenous populations of southern Chile. As recent scholarship points out (see [68, 69, 70, 71]), best practice in anthropological and genetic research to fill in the gaps in a region's history entails direct indigenous participation. We involved the local community through the process of data collection in 2019, and with a return visit to discuss our results with different members from the various locations in 2022 (see Methods). These conversations contributed to the drafting our manuscript. Returning results represents a natural extension of the scientific work. It creates trust and accountability between communities, participants, stakeholders, and the scientific community. From our conversations with local partners and participants, we note that the research focus on the pre-Hispanic period was frequently commented on and positively received. We also note that personal involvement of local partners can be culturally charged, as it requires participants to overcome the stigma long associated with indigenous descent, given the historical contexts of long-standing abuse and exploitation by representatives of non-Mapuche cultures. With this work we aim to contribute to an emerging scientific framework that takes more into account of local cultural codes and complex, dynamic social contexts, and which challenges old models of scientific knowledge production.

Acknowledgements

We thank all the voluntary participants in this study from Araucanía and the Island of Chiloé and all the people who discussed the genetic results and gave feedback during the expedition in March-April 2022 (Jaqueline Caniguan, Jaime Haro, Carlos Catrileo, Juan Manuel Huentelican, Carmen Cayun, Julio Chewin, María Isabel Díaz, Carolina Aillapán, Awunwuenu Aillapán, Sixto Cuyul, Juan Carlos Domihual, Romero family, Sonia Catepillan, Guido Brevis, Colegio Adenauer in Melipeuco, Liceo Público Reino de Suecia in Puerto Saavedra, Liceo Galvarino Riveros Cárdenas in Castro Chiloé, Nicolás Montalva, Marcelo González, Fernando Pairican, Piergiorgio Di Giminiani, Alejandra Vidal, Verónica Silva, Eduardo Barrientos).
CB and KKS were supported by the URPP 'Evolution in Action' of the University of Zurich and the NCCR Evolving Language, Swiss National Science Foundation Agreement #51NF40_180888. CB, EAI and KKS were supported by the SNSF Sinergia project 'Out of Asia' (Grant Number 183578). For the return expedition, the Graduate Campus of the University of Zurich (GRC) supported EAI, and the Zurich Latin American Center (LZZ) supported both EAI and CB. AA received support from the Italian Ministry of Education, University and Research (MIUR) for Progetti PRIN2017 20174BTC4R and the Dipartimenti di Eccellenza Program (2018–2022). PH was supported within the ERC Starting Grant 'Waves' (ERC758967). FIM is funded by CIIR-FONDAP 15110006.

Author contributions

CB conceived the study with the support of SS, PH and FIM. EAI performed the genetic analysis. CB and HB performed laboratory analyses. CB and KKS supervised the study. CB and MJA organized the fieldwork expedition and collected samples. CB, MJA and EAI organized the trip to return the genetic results to the communities. MRC provided guidance and expertise with qpGraph and masking analysis. SA provided expertise with admixture analysis. CP provided expertise with f_4 analysis. CB, MRC, CP and AA provided expertise

with the genetic contextualization within South America. SS, PH, RC and FIM provided linguistic, historical, and archaeological contextualization. MJA curated community engagement and data interpretation. EAI wrote the first draft of the manuscript. EAI and CB wrote the final manuscript with major contributions from PH, SS, MRC, SA, MJA, RC, and KKS.

Inclusion and diversity

We support inclusive, diverse, and equitable conduct of research.

Declaration of interests

The authors declare no competing interests.

STAR METHODS

LEAD CONTACT

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Chiara Barbieri (barbieri.chiara@gmail.com).

MATERIALS AVAILABILITY

This study did not generate new unique reagents.

DATA AND CODE AVAILABILITY

- Modern genotype data have been deposited in the European Genome-phenome Archive (EGA; https://ega-archive.org/) with accession number EGAS00001007200 (see key resources table). Given the sensitive nature of the human genetic data generated in this study, these will not be made publicly available, but access to the data will be granted by a Data Access Committee upon agreeing the conditions on the Data Access Agreement Form available upon request.
- Main scripts to reproduce the analyses are available on GitHub (https://github.com/epifaniarango/popgen_with_epi).
- Any additional information required to reanalyze the data reported in this paper, together with other scripts used for analysis and plots, is available from the corresponding authors upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Sample collection

The study involved individuals either of self-declared Mapuche ancestry and/or who lived in regions where Mapuche presence was historically attested. Sampling was conducted in early 2019 in Chile's Araucanía region and on the island of Chiloé. Local authorities such as municipalities, cultural centers and figures such as *lonkos* (traditional leaders of Mapuche communities) were consulted prior to and/or during the sampling. In Araucanía, residents of rural inland Andean regions are grouped as individuals of putative Pehuenche ancestry, those living near the coast are grouped as individuals of putative Lafkenche ancestry. Individuals of putative Huilliche ancestry were recruited in towns across Chiloé. This sampling underrepresents the documented differences between rural communities in

various parts of the island. Exact sampling locations are not disclosed to protect participants' privacy. Participants of both sexes and all ages were recruited. The composition of age and sexes does not influence our analysis of genetic history of the region. Participants agreed to participate after the project's aims had been explained to them extensively, and all signed consent forms. Cultural indicators like grandparents' places of birth, local surnames (often of Mapuche origin), and the native language of parents and grandparents were also noted. The biological sample consisted of ~2 ml of saliva, collected in Oragene tubes (DNAgenotek), and stored with an anonymous code. The research project and sample collection were approved by the Unidad de Ética y Seguridad de Investigación of the Pontificia Universidad Católica de Chile's IRB (project #171009001, decree #1520863561038). All project's steps were performed in compliance with the Declaration of Helsinki. The samples analyzed in this study represent only a small fraction of the population living in the target regions of Araucanía and Chiloé and are only partially representative of these regions' populations and their complex demographic histories.

Ethics and Community Engagement

After finalizing the data analysis, we organized a return expedition trip in early 2022 with the goal of making our results accessible to the participants and the local population. To achieve this, we translated the scientific results into a language accessible to the general public (in Spanish). The presented material was printed in a large format to be displayed conveniently without the aid of screens or video projectors. Additionally, we engaged with local schools in each study area to share our findings with lectures for students and teachers. Our contacts with local stakeholders were crucial for framing of research questions, displaying our results appropriately, and, more generally, being aware of the cultural and social context. The return expedition was conducted before the writing of the present manuscript to incorporate participants' suggestions.

METHOD DETAILS

DNA extraction and genotyping

Before lab processing, a second anonymization step assigned a new random code to each sample to ensure that sample numbers did not follow the sampling chronological order. DNA was extracted from the Oragene kit, following the manufacturer's protocols, in the molecular biology laboratories of the Max Planck Institute for the Science of Human History in Jena, Germany. DNA samples were screened and quantified with a Nanodrop spectrophotometer and Qubit fluorometer and visually assessed by gel electrophoresis. Samples were genotyped at the ATLAS Biolab in Berlin, on the Axiom Human Origins array [38]. Genotyping data were processed using Affymetrix Genotyping Console v4.2.0.26. In total, 64 samples were genotyped for 629,443 SNPs. PLINK v1.90b5.2 [72] was used to calculate the missing genotype rate with the command '*--missing*'. The proportion of missing calls per sample is <0.005. A small fraction of sites on the array are potentially triallelic in specific populations and are included by reporting the SNP several times (either 2 or 3 times) with a different name. Most population genetic analyses are designed based on biallelic sites, so the triallelic SNPs were removed from the final dataset. PLINK was then used to calculate the consanguinity coefficient *F* (i.e., [<observed hom. count> – <expected count>]/[<total observations> – <expected count>]) and Pi_Hat values (degree of relatedness as Proportion of IBD, i.e., P[IBD = 2] + 0.5*P[IBD = 1]) between pairs of individuals, filtering for minimum allele frequencies of 0.05. All pairs of individuals have Pi_Hat values below 0.2, which excludes the presence of first- or second-degree relatives in the dataset.

The final data set comprises 597,167 SNPs.

QUANTIFICATION AND STATISTICAL ANALYSIS

Comparative datasets

For different sets of analysis, we assembled different datasets. Dataset 1 and variations of it were used to study global relationships; samples from the Americas [39, 40, 41, 38] were merged with a selection of reference populations for each continent (Africa represented by Yoruba [38], Asia represented by Han Chinese [38], Southeast Asia and Oceania represented by Ami [39], Atayal [39], and Papuan [38], and Europe by French [38] and Spanish [39]) genotyped with the Axiom Human Origins array [38]. This dataset contains 584 individuals and an average missing call rate of 0.99642. Dataset 1.2 includes the 61 samples from De la Fuente *et al.* 2018 [11], which were genotyped with a different SNP array (Axiom LAT1 platform, Affymetrix). Dataset 1.2 thus consists of 645 individuals and 96,492 filtered SNPs that overlap in the two genotyping platforms. Dataset 1.3 retains only those individuals from Dataset 1 whose Native American ancestry component is higher than 99.9%. Datasets 2 and 2.2 include a selection of individuals from Datasets 1 and 1.2, respectively, for phasing and identity-by-descent analysis. Finally, Dataset 3 was used to study relationships with aDNA. It combines the modern data with a selection of relevant ancient samples [39, 40, 41, 38, 6, 42, 8, 43, 13, 12, 11, 7] downloaded from the Allen Ancient DNA Resource [73] that are compatible with the Human Origins SNP array format. All the modern individuals are masked to focus on indigenous American history and exclude European and African ancestries. Details on each dataset and the analysis for which it is used can be found in the Supplementary Data S1.A and Figure S1.B.

Population structure analysis

ADMIXTURE [44] was used on the modern global dataset (Dataset 1) to estimate the proportions of ancestry components for each individual. Before the analysis, variants were pruned to limit pairwise linkage disequilibrium (LD) r^2 to at most 0.4 among neighboring SNPs, in sliding windows of 200 SNPs (step size: 25) using PLINK (*--indep-pairwise 200 25 0.4*), which left 218,339 SNPs for the analysis. We performed 10 replicates from $K = 2$ to $K = 15$. Results were visualized using Pong version 1.4.7 [74]. The analysis was also performed on Dataset 1.2, which includes more populations from the Southern Cone (Figure S3). With this dataset, 73,212 SNPs were left after LD pruning.

We used OHANA [75] to create a covariance matrix of the ancestry components for each value of K from 2 to 15, from the average of the allele-frequency matrices generated by ADMIXTURE (selecting the run with the highest likelihood) from Dataset 1 (P.matrix). We converted the covariance matrix into a distance matrix and made an NJ tree following the OHANA protocol. We used Itol [76] to rearrange the branches with the Yoruba population as an outgroup and visualize the tree.

To inspect genetic relationships among Native Americans, we selected Native American individuals with no European or African ancestry, according to the previous ADMIXTURE analysis at $K = 8$ (Dataset 1.3). PCA was performed on this dataset using the PLINK option (*--pca*) with LD-pruning.

We also used PLINK on Dataset 1.3 to check for inbreeding coefficients (*--het*) and infer ROHs and the distribution of their lengths (with the setting *--homozyg --homozyg-density 50 --homozyg-gap 100 --homozyg-kb 500 --homozyg-snp 50 --homozyg-window-het 1 --homozyg-window-snp 50 --homozyg-window-threshold 0.05*).

Analysis of Identity-by-Descent fragments

To infer blocks of identity by descent (IBD) shared among populations, we first phased all individuals using Beagle 5.1 [77] without a reference panel and with the following options:

window=20 trim=0.3. Refined IBD software [78] was used to identify the IBD fragments with the same options as above: window=20 trim=0.3. We used three replicates of phasing and IBD analysis to remove breaks and gaps in IBD segments. Afterward, all replicates were merged using the merge-ibd-segments tool using a gap of 0.5 cM (all software versions are available at https://faculty.washington.edu/browning/refined-ibd.html). In the Americas, most populations share fragments over 4 or 5 cM, as also verified in other studies [41, 79]. For this reason, we apply a high cutoff and consider only fragments above 4 cM. We excluded pairwise comparisons between two Guarani populations as this very high proportion of shared fragments would have obscured the remaining continental patterns. We then binned the fragments in three categories: 4-7 cM, 7-10 cM, >10 cM. For each bin, we then calculated the probability of an individual from population A sharing an IBD fragment with an individual from population B. These probabilities were calculated by dividing the number of pairs of individuals from populations A and B who do share fragments by the total number of possible combinations of pairs of individuals from A and B (which is obtained by multiplying the number of individuals in population A by the number of individuals in population B) (following [51]). The same analysis is applied to the same set of IBD fragments, filtering for fragments with more than 80% presence of SNPs of Native American ancestry, as defined during the masking step (see below) – therefore excluding fragments of African, European, or "unassigned" descent. The probability was projected as a network onto a map in which populations are nodes and edges between them are scaled in width by the probability of IBD sharing. Only probabilities above 0.005 are displayed. The raster file for creating the map was downloaded from [80].

Dating and demographic analysis

We used ALDER 1.03 [54] to reconstruct admixture times from linkage disequilibrium patterns. To infer the admixture time with Europeans (primarily Spanish) in each target population, we chose two source populations: one proxy for the Spanish population and one for Native American ancestry. For the Spanish parental population, all Spanish individuals were merged. The best representative for Native American population depends on which of the various ancestries of the Americas is most represented in the target population. We therefore ran all possible combinations of source populations, and then selected the runs with the best Z-score and p-value. We only considered populations with a high population size (>7 individuals).

To estimate variation in effective population size over time, we used IBDNe [52] with the default settings. We used only fragments over 2cm and reconstructed population sizes only for the past 50 generations, as this method is not able to reconstruct older ancestor relationships reliably from SNP array data [52]. A generation time of 28 years was used to convert generations to years.

Local ancestry analysis and masking

Modern samples can be used for generating high-quality genome-wide data with less effort than ancient samples. On the other hand, modern samples from the Americas harbor a range of ancestries, in particular the highly divergent ancestries brought by European impacts since 1492, which act as confounds for our research focus on the prehistory of the native populations of the Americas. Masking is the process of filtering out variants associated with other components, in our case those from outside the Americas (https://github.com/epifaniarango/popgen_with_epi/).

This regional ancestry analysis is a semi-supervised approach, requiring a reference panel for each ancestry of interest, performed with the software RFMIX v1.5.4 [81]. This analysis uses the same phased haplotypes as the IBD analysis. The reference panels for African and European admixture were constituted by Yoruba and Spanish individuals respectively. The Native American Reference panel was built as follows. We first selected individuals

previously identified as “unadmixed” in ADMIXTURE. This set was further filtered with an f_4-statistic of the form f_4 (Unadmixed Native American Population, Target individual; Han, San) [4] designed to detect more subtle European and African admixture in each individual. The selected Unadmixed Native American populations for the f_4 were Karitiana, Mixe, and Xavante. The individuals who passed this filter with a non-significant f_4-statistic were included in the Native American Reference panel (116).
We ran RFMIX with 2 expectation-maximization iterations (-e 2) that also screens the reference panel in the ongoing analysis, as recommended by the authors [81]. Parameters settings were: window size = 0.2 cM, spacing = CFR, node size = 5, and number of generations since admixture = 11, according to the analysis from Homburger et al., 2015 [83] and our ALDER results (*-G 11 -n 5 –forward-backward --use-reference-panels-in-EM -e 2 -w 0.2*). The threshold for local ancestry assignment was a probability level of 0.9. To check the consistency of the method, we compared the global Native American ancestry proportions estimated with RFMIX, using a weighted mean by chromosomes, with the proportions calculated by ADMIXTURE. The correlation is almost linear (>0.9 after performing a Spearman’s correlation test), except for the North American samples, for which the RFMIX estimates of Native American ancestry is sometimes smaller (see https://github.com/epifaniarango/popgen_with_epi/blob/Local-ancestry-and-masking/README.md).

African and European ancestries were then “masked” for each sample. Following the previous analysis of local ancestry, we kept only those SNPs assigned to Native American ancestry above a threshold of >0.9 probability. The remaining SNPs were coded as missing data, and individuals were separated into the two-phased haplotypes in a pseudo-haploidization process. We evaluated various masking strategies using quality checks and confirmed that the commonly used pseudo-haploid masking [10, 51] performs well and retains more SNPs for the analysis.
We then removed individuals with <30% SNPs typed and SNPs with >50% missing genotypes (*--mind 0.7 --geno 0.5* with PLINK 1.9). To check the performance of the masking protocol, we again computed an individual-based f_4-statistic of the form f_4(Unadmixed Native American, Test (Admixed Native American); Han, San) following [4], and compared the results before and after masking (for details, see https://github.com/epifaniarango/popgen_with_epi). The selected unadmixed populations were Kaqchikel, Karitina, Mixe, and Xavante. The results were consistent with a positive f_4 after masking. We also performed PCA visualization to confirm the absence of outliers and no attraction towards European and African individuals (Figure S5.A-C).
After these quality checks, the masked dataset was merged with the Native American Reference panel and the ancient samples (which do not bear traces of European and African admixture), retaining only individuals with more than 100,000 SNPs (Dataset 3.3).
Dataset 3 was used to perform another ADMIXTURE analysis, following the same protocol described above (Figure S5.D-E). We summarized the results at K = 5 as pie-charts on a map, distinguished by time period (Figure 4). Ancestry proportions were averaged among populations. Ancient samples were grouped as populations if they belong to the same archaeological site and time period.

Marker-frequency-based statistics and ancestry modeling
We compute D-statistics with ADMIXTOOLS [38] to analyze fine-scale population dynamics between ancient and modern samples. Genetic affinity in terms of shared genetic drift was quantified with the outgroup f_3-statistic, with the Mbuti as the outgroup, i.e. f_3(Mbuti; Pop1, Pop2), using qp3Pop [38]. For Pop1 and Pop2 we used all possible combinations of individuals or populations from Dataset 3 and then created an f_3-distance matrix. Higher f_3-

values imply higher genetic affinity (more shared genetic drift) between Pop1 and Pop2. The converted dissimilarity matrix 1-f_3 was used to generate an MDS plot using R, and the matrix 1/f_3 to generate an NJ tree with the R package "ape" [82], using Ancient Beringian as the outgroup. The tree was displayed using Itol [76].

f_4-statistics were designed to search for an excess of allele sharing between populations, and were computed with qpDstats using the default parameters: "f4mode: YES", and block jackknife over 5-Mb. Most statistics were computed in the form f_4(Mbuti, Target; X, Y). X and Y are paired only if their data were generated through the same sequencing technology (SNPChip vs. ShotGun sequencing), to minimize bias and attraction effects. In cases where this was not possible, we also compared various configurations of the f_4, in order to exclude possible attraction effects. For robustness, the tests were computed with all available SNPs and verified with transversions to confirm that the signal was not biased by aDNA degradation (Dataset 3.3 and 3.4). Transitions in aDNA data often result from miscoding lesions; selecting for "transversions only" allows us to avoid those errors.

To model the relationships between the various Southern Cone populations, we used qpGraph [38], considering only transversions and using the default settings. The qpGraph combines f_2, f_3, *and* f_4-statistics to check the robustness of the tree topologies that we provide. To reduce bias, we used only those ancient samples genotyped with a capture method close to the SNP chip data used for the modern samples, except for *USR1* and *Ayayema_4700BP*, which were genotyped with shotgun sequencing. We focused on samples from the Southern Cone, and contextualized them with Mbuti as the basal outgroup, and *Ancient_Beringian* and *Brazil_LapaDoSanto_9600BP* as non-SNA2 references. Our basal tree was built with *Mbuti, Ancient_Beringian, Brazil_LapaDoSanto_9600BP, ArroyoSeco2_7700BP, LosRieles_5100BP, LaArcillosa2_5800BP*, and *Ayayema_4700BP* (Figure S7.A). We used the simplest tree topology (without the basal admixture in the ancestors of *ArroyoSeco2_7700BP* and *LosRieles_5100BP)* for building the following topologies. We tested different configurations without *Ayayema_4700BP* to test for biased in the sequencing method (Figure S7.B). We successively added populations in various configurations, keeping only graphs with |Z|<3.5 (following [8]) (Figure S7.B-D). We explore topologies without admixture edges from *LosRieles_5100BP* and *ArroyoSeco2_7700BP* which are not fully supported in f_4 configurations (Figure S7.F-G). When incorporating the modern samples, we merged Pehuenche and Lafkenche individuals into a single larger population labeled "Mapuche" (Figure S7.E).

Supplementary Data S1.A-I: Supporting Information for Individuals in Datasets and Raw D-Statistics. Related to STAR Methods.

https://www.sciencedirect.com/science/article/pii/S0960982223006073#app2

This table provides supporting information about the individuals included in the datasets, including their IDs, study of origin, geographic location, and type of data. Additionally, raw D-statistics are included in the table, which are a measure of the genetic affinity between different populations. This information supports the analyses presented in the main text and provides additional insights into the genetic relationships among the populations. Each tab in the Excel table is named with a letter. **(A)** List of individuals included in the analysis and assignation to different comparative Datasets. **(B)** f_3 statistics combinations. **(C)** f_4 (Mbuti, X; SC, SC) examining the genetic continuity between modern Southern Chile (SC) populations. **(D)** f_4 (Mbuti, X; Conchalí, Z) examining the genetic connection between

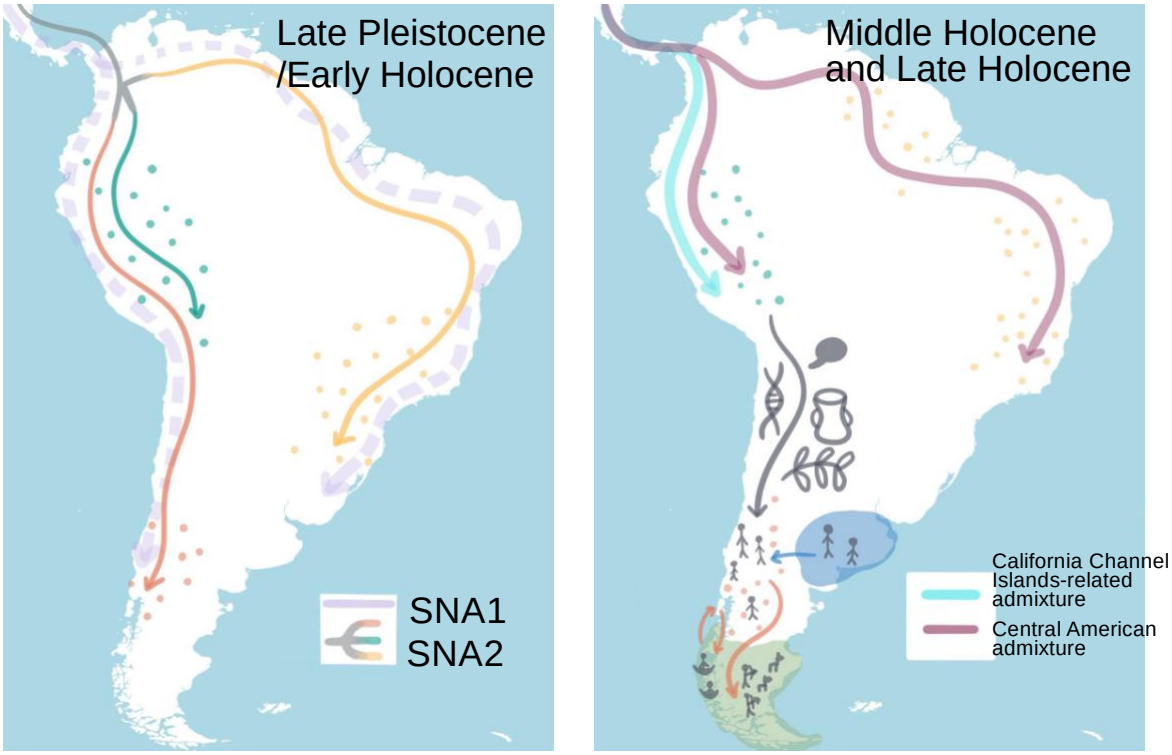
Conchalí and modern SC. We calculate the statistics in two configurations f_4 (Mbuti, SC; Conchali, Z) and f_4 (Mbuti, X; Conchali, SC), to check for any asymmetrical tendency. **(E)** f_4 (Mbuti, SC/Conchali_700BP; Middle Holocene Far South, Late Holocene Far South) inspired by Nakatsuka, N. et al. [S7], to explore if the migration from Central Chile into the Far South also involved the ancestors of contemporary Mapuche populations. **(F)** f_4(Mbuti, SC/Conchali_700BP; Middle Holocene Central Andes, Late Holocene CentralAndes) exploring connections between South-Central Chile and Ancient Central Andes. **(G)** f_4(Mbuti, SC/Conchali_700BP; Middle Holocene Central Argentina, Late Holocene Argentina) exploring connections between South-Central Chile and Ancient Argentina. **(H)** f_4 (Mbuti, X; SCone, SCone) exploring allele sharing outside of the Southern Cone populations. **(I)** f_4 (Mbuti, X; Central-Southern Chile/Far South, Argentinean Pampas) exploring allele sharing outside of Chilean and Argentinean populations.

References

1. Raghavan, M., Steinrücken, M., Harris, K., Schiffels, S., Rasmussen, S., DeGiorgio, M., Albrechtsen, A., Valdiosera, C., Ávila-Arcos, M.C., Malaspinas, A.-S., et al. (2015). Genomic evidence for the Pleistocene and recent population history of Native Americans. Science *349*. 10.1126/science.aab3884.
2. Bennett, M.R., Bustos, D., Pigati, J.S., Springer, K.B., Urban, T.M., Holliday, V.T., Reynolds, S.C., Budka, M., Honke, J.S., Hudson, A.M., et al. (2021). Evidence of humans in North America during the Last Glacial Maximum. Science. 10.1126/science.abg7586.
3. Rasmussen, M., Li, Y., Lindgreen, S., Pedersen, J.S., Albrechtsen, A., Moltke, I., Metspalu, M., Metspalu, E., Kivisild, T., Gupta, R., et al. (2010). Ancient human genome sequence of an extinct Palaeo-Eskimo. Nature *463*, 757–762. 10.1038/nature08835.
4. Reich, D., Patterson, N., Campbell, D., Tandon, A., Mazieres, S., Ray, N., Parra, M.V., Rojas, W., Duque, C., Mesa, N., et al. (2012). Reconstructing Native American population history. Nature *488*, 370–374. 10.1038/nature11258.
5. Raghavan, M., DeGiorgio, M., Albrechtsen, A., Moltke, I., Skoglund, P., Korneliussen, T.S., Grønnow, B., Appelt, M., Gulløv, H.C., Friesen, T.M., et al. (2014). The genetic prehistory of the New World Arctic. Science *345*. 10.1126/science.1255832.
6. Scheib, C.L., Li, H., Desai, T., Link, V., Kendall, C., Dewar, G., Griffith, P.W., Mörseburg, A., Johnson, J.R., Potter, A., et al. (2018). Ancient human parallel lineages within North America contributed to a coastal expansion. Science *360*, 1024–1027. 10.1126/science.aar6851.
7. Rasmussen, M., Anzick, S.L., Waters, M.R., Skoglund, P., DeGiorgio, M., Stafford, T.W., Rasmussen, S., Moltke, I., Albrechtsen, A., Doyle, S.M., et al. (2014). The genome of a Late Pleistocene human from a Clovis burial site in western Montana. Nature *506*, 225–229. 10.1038/nature13025.
8. Posth, C., Nakatsuka, N., Lazaridis, I., Skoglund, P., Mallick, S., Lamnidis, T.C., Rohland, N., Nägele, K., Adamski, N., Bertolini, E., et al. (2018). Reconstructing the Deep Population History of Central and South America. Cell *175*, 1185-1197.e22. 10.1016/j.cell.2018.10.027.
9. Willerslev, E., and Meltzer, D.J. (2021). Peopling of the Americas as inferred from ancient genomics. Nature *594*, 356–364. 10.1038/s41586-021-03499-y.
10. Capodiferro, M.R., Aram, B., Raveane, A., Rambaldi Migliore, N., Colombo, G., Ongaro, L., Rivera, J., Mendizábal, T., Hernández-Mora, I., Tribaldos, M., et al. (2021). Archaeogenomic distinctiveness of the Isthmo-Colombian area. Cell. 10.1016/j.cell.2021.02.040.
11. De la Fuente, C., Ávila-Arcos, M.C., Galimany, J., Carpenter, M.L., Homburger, J.R., Blanco, A., Contreras, P., Dávalos, D.C., Reyes, O., Roman, M.S., et al. (2018). Genomic insights into the origin and diversification of late maritime hunter-gatherers from the Chilean Patagonia. Proc. Natl. Acad. Sci. *115*, E4006–E4012. 10.1073/pnas.1715688115.
12. Moreno-Mayar, J.V., Vinner, L., Damgaard, P. de B., Fuente, C. de la, Chan, J., Spence, J.P., Allentoft, M.E., Vimala, T., Racimo, F., Pinotti, T., et al. (2018). Early human dispersals within the Americas. Science *362*. 10.1126/science.aav2621.
13. Nakatsuka, N., Luisi, P., Motti, J.M.B., Salemme, M., Santiago, F., D'Angelo del Campo, M.D., Vecchi, R.J., Espinosa-Parrilla, Y., Prieto, A., Adamski, N., et al. (2020). Ancient genomes in South Patagonia reveal population movements associated with technological shifts and geography. Nat. Commun. *11*, 3868. 10.1038/s41467-020-17656-w.
14. Luisi, P., García, A., Berros, J.M., Motti, J.M.B., Demarchi, D.A., Alfaro, E., Aquilano, E., Argüelles, C., Avena, S., Bailliet, G., et al. (2020). Fine-scale genomic analyses of admixed individuals reveal unrecognized genetic ancestry components in Argentina. PLOS ONE *15*, e0233808. 10.1371/journal.pone.0233808.
15. García, A., Nores, R., Motti, J.M.B., Pauro, M., Luisi, P., Bravi, C.M., Fabra, M., Gosling, A.L., Kardailsky, O., Boocock, J., et al. (2021). 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. 10.1093/hmg/ddab105.
16. de Saint Pierre, M., Motti, J.M.B., Fuku, N., Tanaka, M., Llop, E., Bonatto, S.L., and Moraga, M. (2012). An Alternative Model for the Early Peopling of Southern South America Revealed by Analyses of Three Mitochondrial DNA Haplogroups. PLOS ONE *7*, e43486. 10.1371/journal.pone.0043486.
17. Roca-Rada, X., Politis, G., Messineo, P.G., Scheifler, N., Scabuzzo, C., González, M., Harkins, K.M., Reich, D., Souilmi, Y., Teixeira, J.C., et al. (2021). Ancient mitochondrial genomes from the Argentinian Pampas inform the early peopling of the Southern Cone of South America. iScience *24*, 102553. 10.1016/j.isci.2021.102553.
18. Sepúlveda, P.B.P., Mayordomo, A.C., Sala, C., Sosa, E.J., Zaiat, J.J., Cuello, M., Schwab, M., Golpe, D.R., Aquilano, E., Santos, M.R., et al. (2022). Human Y chromosome sequences from Q Haplogroup reveal a South American settlement pre-18,000 years ago and a profound genomic impact during the Younger Dryas. PLOS ONE *17*, e0271971. 10.1371/journal.pone.0271971.
19. Perego, U.A., Angerhofer, N., Pala, M., Olivieri, A., Lancioni, H., Kashani, B.H., Carossa, V., Ekins, J.E., Gómez-Carballa, A., Huber, G., et al. (2010). The initial peopling of the Americas: A growing number of founding mitochondrial genomes from Beringia. Genome Res. *20*, 1174–1179. 10.1101/gr.109231.110.

20. De la Fuente, C., Galimany, J., Kemp, B.M., Judd, K., Reyes, O., and Moraga, M. (2015). Ancient marine hunter-gatherers from Patagonia and Tierra Del Fuego: Diversity and differentiation using uniparentally inherited genetic markers. Am. J. Phys. Anthropol. *158*, 719–729. 10.1002/ajpa.22815.
21. Crespo, C.M., Lanata, J.L., Cardozo, D.G., Avena, S.A., and Dejean, C.B. (2018). Ancient maternal lineages in hunter-gatherer groups of Argentinean Patagonia. Settlement, population continuity and divergence. J. Archaeol. Sci. Rep. *18*, 689–695. 10.1016/j.jasrep.2017.11.003.
22. Dillehay, T.D., Ocampo, C., Saavedra, J., Sawakuchi, A.O., Vega, R.M., Pino, M., Collins, M.B., Cummings, L.S., Arregui, I., Villagran, X.S., et al. (2015). New Archaeological Evidence for an Early Human Presence at Monte Verde, Chile. PLOS ONE *10*, e0141923. 10.1371/journal.pone.0141923.
23. Clairis, C. (2021). 20. Indigenous Languages of Tierra del Fuego. In 20. Indigenous Languages of Tierra del Fuego (University of Texas Press), pp. 753–783. 10.7560/775923-021.
24. Planella, M., Falabella, F., Belmar, C., and L., L. (2015). Huertos, chacras y sementeras. Plantas cultivadas y su participación en los desarrollos culturales de Chile central. Rev. Esp. Antropol. Am. *44*. 10.5209/rev_REAA.2014.v44.n2.50727.
25. Roa Solís, C., Funes, H.R., and Campbell, R. (2018). Entre la Pampa y el Pacífico Sur. Evaluando la dispersión más austral de cultígenos en el Cono Sur americano desde la evidencia arqueobotánica y radiométrica de Isla Mocha y Cueva de los Catalanes (Sur de Chile). An. Arqueol. Etnología *73(2):189-220*.
26. Adán, L., Mera, R., Navarro, X., Campbell, R., and Sánchez, M. (2016). Historia prehispánica en la región Centro-Sur de Chile: Cazadores-recolectores holocénicos y comunidades alfareras. In Prehistoria en Chile (Editorial Universitaria).
27. Sadowsky, S., Aninao, M.J., Cayunao, M.I., and Heggarty, P. (2015). Huilliche: ¿geolecto del mapudungun o lengua propia? Una mirada desde la fonética y la fonología de las consonantes. In Lingüística indígena sudamericana: aspectos descriptivos, comparativos y areales, A. Fernández Garay and M. A. Regúnaga, eds. (Editorial de la Facultad de Filosofía y Letras, Universidad de Buenos Aires).
28. Greenberg, J.H. (1987). Language in the Americas (Stanford University Press).
29. Campbell, L. (1988). Review of Language in the Americas, by Joseph Greenberg. Linguist. Soc. Am. *64*, 591–615.
30. Heggarty, P. (2020). Deep time and first settlement: what, if anything, can linguistics tell us? In Rethinking the Andes-Amazonia Divide.
31. Sadowsky, S., Painequeo, H., Salamanca, G., and Avelino, H. (2013). Mapudungun. J. Int. Phon. Assoc. *43*, 87–96. 10.1017/S0025100312000369.
32. Hammarström, H., Forkel, R., Haspelmath, M., and Bank, S. (2022). Glottolog 4.6. https://doi.org/10.5281/zenodo.6578297.
33. Golluscio, L.A. (2009). 41. Loanwords in Mapudungun, a language of Chile and Argentina. In 41. Loanwords in Mapudungun, a language of Chile and Argentina (De Gruyter Mouton), pp. 1035–1071. 10.1515/9783110218442.1035.
34. Pache, M. (2014). Lexical Evidence for Pre-Inca Language Contact of Mapudungun (Mapuche) with Quechuan and Aymaran. J. Lang. Contact *7*, 345–379. 10.1163/19552629-00702005.
35. Moulian, R., Catrileo, M., and Landeo, P. (2015). Afines Quechua en el vocabulario Mapuche de Valdivia. RLA Rev. Lingüíst. Teórica Apl. *53*, 73–96. 10.4067/S0718-48832015000200004.
36. Guevara, T. (1929). Historia de Chile : Chile prehispano (Balcells&Co).
37. Latcham, R.E. (1924). La organización social y las creencias religiosas de los antiguos araucanos (Imprenta Cervantes).
38. Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient Admixture in Human History. Genetics *192*, 1065–1093. 10.1534/genetics.112.145037.
39. Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., Sudmant, P.H., Schraiber, J.G., Castellano, S., Lipson, M., et al. (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature *513*, 409–413. 10.1038/nature13673.
40. Skoglund, P., Mallick, S., Bortolini, M.C., Chennagiri, N., Hünemeier, T., Petzl-Erler, M.L., Salzano, F.M., Patterson, N., and Reich, D. (2015). Genetic evidence for two founding populations of the Americas. Nature *525*, 104–108. 10.1038/nature14895.
41. Barbieri, C., Barquera, R., Arias, L., Sandoval, J.R., Acosta, O., Zurita, C., Aguilar-Campos, A., Tito-Álvarez, A.M., Serrano-Osuna, R., Gray, R.D., et al. (2019). The Current Genomic Landscape of Western South America: Andes, Amazonia, and Pacific Coast. Mol. Biol. Evol. *36*, 2698–2713. 10.1093/molbev/msz174.
42. Moreno-Mayar, J.V., Potter, B.A., Vinner, L., Steinrücken, M., Rasmussen, S., Terhorst, J., Kamm, J.A., Albrechtsen, A., Malaspinas, A.-S., Sikora, M., et al. (2018). Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. Nature *553*, 203–207. 10.1038/nature25173.
43. Nakatsuka, N., Lazaridis, I., Barbieri, C., Skoglund, P., Rohland, N., Mallick, S., Posth, C., Harkins-Kinkaid, K., Ferry, M., Harney, É., et al. (2020). A Paleogenomic Reconstruction of the Deep Population History of the Andes. Cell *181*, 1131-1145.e21. 10.1016/j.cell.2020.04.015.
44. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. *19*, 1655–1664. 10.1101/gr.094052.109.
45. Lawson, D.J., van Dorp, L., and Falush, D. (2018). A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. Nat. Commun. *9*, 3258. 10.1038/s41467-018-05257-7.
46. Ralph, P., and Coop, G. (2013). The Geography of Recent Genetic Ancestry across Europe. PLoS Biol. *11*, e1001555. 10.1371/journal.pbio.1001555.
47. Ringbauer, H., Coop, G., and Barton, N.H. (2017). Inferring Recent Demography from Isolation by Distance of Long Shared Sequence Blocks. Genetics *205*, 1335–1351. 10.1534/genetics.116.196220.
48. Stevens, E.L., Heckenberg, G., Roberson, E.D.O., Baugher, J.D., Downey, T.J., and Pevsner, J. (2011). Inference of Relationships in Population Data Using Identity-by-Descent and Identity-by-State. PLOS Genet. *7*, e1002287. 10.1371/journal.pgen.1002287.
49. Liu, D., Duong, N.T., Ton, N.D., Van Phong, N., Pakendorf, B., Van Hai, N., and Stoneking, M. (2020). Extensive Ethnolinguistic Diversity in Vietnam Reflects Multiple Sources of Genetic Diversity. Mol. Biol. Evol. *37*, 2503–2519. 10.1093/molbev/msaa099.
50. Arias, L., Emlen, N.Q., Norder, S., Julmi, N., Lemus Serrano, M., Chacon, T., Wiegertjes, J., Howard, A., Azevedo, M.C.B.C., Caine, A., et al. (2022). Interpreting mismatches between linguistic and genetic patterns among speakers of Tanimuka (Eastern Tukanoan) and Yukuna (Arawakan). Interface Focus *13*, 20220056. 10.1098/rsfs.2022.0056.
51. Ioannidis, A.G., Blanco-Portillo, J., Sandoval, K., Hagelberg, E., Miquel-Poblete, J.F., Moreno-Mayar, J.V., Rodríguez-Rodríguez, J.E., Quinto-Cortés, C.D., Auckland, K., Parks, T., et al. (2020). Native American gene flow into Polynesia predating Easter Island settlement. Nature *583*, 572–577. 10.1038/s41586-020-2487-2.
52. Browning, S.R., and Browning, B.L. (2015). Accurate Non-parametric Estimation of Recent Effective Population Size from Segments of Identity by Descent. Am. J. Hum. Genet. *97*, 404–418. 10.1016/j.ajhg.2015.07.012.

53. Lindo, J., Haas, R., Hofman, C., Apata, M., Moraga, M., Verdugo, R.A., Watson, J.T., Viviano Llave, C., Witonsky, D., Beall, C., et al. (2018). The genetic prehistory of the Andean highlands 7000 years BP though European contact. Sci. Adv. *4*, eaau4921. 10.1126/sciadv.aau4921.
54. Loh, P.-R., Lipson, M., Patterson, N., Moorjani, P., Pickrell, J.K., Reich, D., and Berger, B. (2013). Inferring Admixture Histories of Human Populations Using Linkage Disequilibrium. Genetics *193*, 1233–1254. 10.1534/genetics.112.147330.
55. Campbell, L. (1997). American Indian Languages: The Historical Linguistics of Native America (Oxford University Press).
56. Lowell, T.V., Heusser, C.J., Andersen, B.G., Moreno, P.I., Hauser, A., Heusser, L.E., Schlüchter, C., Marchant, D.R., and Denton, G.H. (1995). Interhemispheric correlation of late pleistocene glacial events. Science *269*, 1541–1549. 10.1126/science.269.5230.1541.
57. Clapperton, C.M., Sugden, D.E., Kaufman, D.S., and McCulloch, R.D. (1995). The Last Glaciation in Central Magellan Strait, Southernmost Chile. Quat. Res. *44*, 133–148. 10.1006/qres.1995.1058.
58. Vicuña, L., Mikhailova, A., Norambuena, T., Ilina, A., Klimenkova, O., Shchur, V., and Eyheramendy, S. (2021). Genomic insights into the recent population history of Mapuche Native Americans. 2021.11.25.470066. 10.1101/2021.11.25.470066.
59. Boccara, G. (1999). Etnogénesis mapuche: resistencia y restructuración entre los indígenas del centro-sur de Chile (siglos XVI-XVIII). Hisp. Am. Hist. Rev. *79*, 425–461. 10.1215/00182168-79.3.425.
60. Sadowsky, S., and Aninao, M.J. (2019). Internal migration and ethnicity in Santiago. In The Routledge Handbook of Spanish in the Global City (Routledge).
61. Bengoa, J. (2007). Historia de los antiguos mapuches del sur: desde antes de la llegada de los españoles hasta las paces de Quilín : siglos XVI y XVII Catalonia, Ltda.
62. Dillehay, T.D. (2007). Monuments, Empires, and Resistance: The Araucanian Polity and Ritual Narratives (Cambridge University Press).
63. Ongaro, L., Scliar, M.O., Flores, R., Raveane, A., Marnetto, D., Sarno, S., Gnecchi-Ruscone, G.A., Alarcón-Riquelme, M.E., Patin, E., Wangkumhang, P., et al. (2019). The Genomic Impact of European Colonization of the Americas. Curr. Biol. CB *29*, 3974-3986.e4. 10.1016/j.cub.2019.09.076.
64. Bauer, A.J. (1994). La sociedad rural chilena: desde la conquista española a nuestros dias (Andres Bello).
65. Martinic, M. (1999). La inmigración chilota en Magallanes. Apreciación histórica sobre sus causas, características y consecuencias. An. Inst. Patagon. Ser. Cienc. Humanas *27*, 27–47.
66. García, F., Moraga, M., Vera, S., Henríquez, H., Llop, E., Ocampo, C., Aspillaga, E., and Rothhammer, F. (2004). Origen y microdiferenciación de la población del Archipiélago de Chiloé. Rev. Chil. Hist. Nat. *77*, 539–546. 10.4067/S0716-078X2004000300012.
67. Reyes, O. (2020). The Settlement of the Chonos Archipelago, Western Patagonia, Chile (Springer International Publishing) 10.1007/978-3-030-54326-6.
68. Tsosie, K.S., Begay, R.L., Fox, K., and Garrison, N.A. (2020). Generations of genomes: advances in paleogenomics technology and engagement for Indigenous people of the Americas. Curr. Opin. Genet. Dev. *62*, 91–96. 10.1016/j.gde.2020.06.010.
69. Malhi, R.S. (2019). Community-Oriented Research and the Future of Anthropological Genetics. In A Companion to Anthropological Genetics (John Wiley & Sons, Ltd), pp. 37–44. 10.1002/9781118768853.ch3.
70. Hudson, M., Garrison, N.A., Sterling, R., Caron, N.R., Fox, K., Yracheta, J., Anderson, J., Wilcox, P., Arbour, L., Brown, A., et al. (2020). Rights, interests and expectations: Indigenous perspectives on unrestricted access to genomic data. Nat. Rev. Genet. *21*, 377–384. 10.1038/s41576-020-0228-x.
71. Muller, C., and Dortch, J. (2019). Traditional owner participation in genetic research: A researcher perspective. In Interrogating Human Origins (Routledge).
72. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience *4*. 10.1186/s13742-015-0047-8.
73. Datasets | David Reich Lab https://reich.hms.harvard.edu/datasets.
74. Behr, A.A., Liu, K.Z., Liu-Fang, G., Nakka, P., and Ramachandran, S. (2016). pong: fast analysis and visualization of latent clusters in population genetic data. Bioinformatics *32*, 2817–2823. 10.1093/bioinformatics/btw327.
75. Cheng, J.Y., Mailund, T., and Nielsen, R. (2016). Ohana, a tool set for population genetic analyses of admixture components. bioRxiv, 071233. 10.1101/071233.
76. Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. *47*, W256–W259. 10.1093/nar/gkz239.
77. Browning, S.R., and Browning, B.L. (2007). Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering. Am. J. Hum. Genet. *81*, 1084–1097. 10.1086/521987.
78. Browning, B.L., and Browning, S.R. (2013). Improving the Accuracy and Efficiency of Identity-by-Descent Detection in Population Data. Genetics *194*, 459–471. 10.1534/genetics.113.150029.
79. Harris, D.N., Song, W., Shetty, A.C., Levano, K.S., Cáceres, O., Padilla, C., Borda, V., Tarazona, D., Trujillo, O., Sanchez, C., et al. (2018). Evolutionary genomic dynamics of Peruvians before, during, and after the Inca Empire. Proc. Natl. Acad. Sci. *115*, E6526–E6535. 10.1073/pnas.1720798115.
80. Natural Earth - Free vector and raster map data at 1:10m, 1:50m, and 1:110m scales https://www.naturalearthdata.com/.
81. Maples, B.K., Gravel, S., Kenny, E.E., and Bustamante, C.D. (2013). RFMix: A Discriminative Modeling Approach for Rapid and Robust Local-Ancestry Inference. Am. J. Hum. Genet. *93*, 278–288. 10.1016/j.ajhg.2013.06.020.
82. Paradis, E., and Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics *35*, 526–528.
83. J.R. Homburger, A. Moreno-Estrada, C.R. Gignoux, D. Nelson, E. Sanchez, P. Ortiz-Tello, B.A. Pons-Estel, E. Acevedo-Vasquez, P. Miranda, C.D. Langefeld, et al. (2015) Genomic insights into the ancestry and demographic history of South America. PLoS Genet., *11*, p. e1005602, 10.1371/journal.pgen.1005602


Late Pleistocene
/Early Holocene
SNA1
SNA2
Middle Holocene
and Late Holocene
California Channel
Islands-related
admixture
Central American
admixture