Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

Genome Research

The peopling of South America and the trans-Andean gene flow of the first settlers

Alberto Gómez-Carballa[1,2,#], Jacobo Pardo-Seco[1,2], Stefania Brandini[3], Alessandro Achilli[3], Ugo A. Perego[3], Michael D. Coble[4], Toni M. Diegoli[5,6], Vanesa Álvarez-Iglesias[1], Federico Martinón-Torres[2], Anna Olivieri[3], Antonio Torroni[3], Antonio Salas[1,#]

[1]Unidade de Xenética, Departamento de Anatomía Patolóxica e Ciencias Forenses, Instituto de Ciencias Forenses, Facultade de Medicina, Universidade de Santiago de Compostela, and GenPoB Research Group, Instituto de Investigaciones Sanitarias (IDIS), Hospital Clínico Universitario de Santiago, Galicia, Spain

[2]Grupo de Investigación en Genética, Vacunas, Infecciones y Pediatría (GENVIP), Hospital Clínico Universitario and Universidade de Santiago de Compostela, Galicia, Spain

[3]Dipartimento di Biologia e Biotecnologie, Università di Pavia, Pavia, Italy

[4]Applied Genetics Group, National Institute of Standards and Technology, Gaithersburg, MD, USA

[5]Office of the Chief Scientist, Defense Forensic Science Center, Ft. Gillem, GA, USA

[6]Analytical Services, Inc., Arlington, VA, USA

[#] Both authors contributed equally to this work

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

*Correspondence: Antonio Salas; Unidade de Xenética, Departamento de Anatomía Patológica e Ciencias Forenses, Universidade de Santiago de Compostela, Galicia, Spain. Tel: +34–981–582327; Fax: +34–981–580336; E-mail: antonio.salas@usc.es

Running title: Gene flow from and across the Andean area

Keywords: Native Americans; mitogenomes; mitochondrial DNA haplogroups; genome-wide data; settlement of South America; phylogeography.

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

Abstract

Genetic and archaeological data indicate that the initial Paleoindian settlers of South America followed two entry routes separated by the Andes and the Amazon rainforest. The interactions between these paths and their impact on the peopling of South America remain unclear. Analysis of genetic variation in the Peruvian Andes and regions located South of the Amazon River might provide clues on this issue. We analyzed mitochondrial DNA variation at different Andean locations and >360,000 autosomal SNPs from 28 Native American ethnic groups to evaluate different trans-Andean demographic scenarios. Our data reveal that the Peruvian Altiplano was an important enclave for early Paleoindian expansions and point to a genetic continuity in the Andes until recent times, which was only marginally affected by gene flow from the Amazonian lowlands. Genomic variation shows a good fit with the archaeological evidence, indicating that the genetic interactions between the descendants of the settlers that followed the Pacific and Atlantic routes were extremely limited.

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

Introduction

The study of the human settlement and spread into the American double continent has received great attention in the literature (Torroni et al. 1993; Forster et al. 1996; Schurr and Sherry 2004; Tamm et al. 2007; Achilli et al. 2008; Fagundes et al. 2008; Gilbert et al. 2008; Perego et al. 2009; O'Rourke and Raff 2010; Perego et al. 2010; Bodner et al. 2012; Reich et al. 2012; Achilli et al. 2013; Raghavan et al. 2015; Llamas et al. 2016). It is generally agreed that the ancestors of Native Americans were probably of southern Siberian origin. The remains found at the Yana site represent unequivocal evidence of the *Homo sapiens* presence in the Arctic region ~30 thousand years ago (kya), shortly before the Last Glacial Maximum (LGM) (Pitulko et al. 2004). However, recent archaeological evidences revealed human presence in the Arctic much earlier than previously thought (~45 kya; SK mammoth and Bunge-Toll sites)(Pitulko et al. 2016). During the LGM humans survived in glacial refuge areas (Torroni et al. 2001; Achilli et al. 2004; Gamble et al. 2004; Gómez-Carballa et al. 2012; Pala et al. 2012). One of these was Beringia, a land bridge located between the north-eastern continental region of Asia and the north-western extreme of North America, which was largely unglaciated during the Pleistocene. During the LGM, the ancestors of Paleoindians remained almost isolated, possibly up to 6–8 kya, in eastern Beringia (Raghavan et al. 2014; Raghavan et al. 2015; Llamas et al. 2016) or, as paleoecological data much more strongly indicate, in south-central Beringia (Elias and Crocker 2008; Hoffecker et al. 2014; Hoffecker et al. 2016). In that period of global cooling and glacial advance, genetic drift, novel mutations and limited admixture with newly

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

arrived East Asian groups extensively re-shaped their genetic variation (Perego et al. 2009; Perego et al. 2010).

Genetic and archaeological data support the hypothesis that a small population group entered the American continent from Beringia 15-18 kya, in concomitance with the improvement of climatic conditions (Achilli et al. 2008; Goebel et al. 2008; Perego et al. 2009; Bodner et al. 2012; Battaglia et al. 2013; Llamas et al. 2016). More complex scenarios based on genetic and linguistic evidence envision instead two, three or maybe more migration waves from Beringia (Torroni et al. 1992; Perego et al. 2009; Reich et al. 2012; Achilli et al. 2013). The initial incursion of Paleoindians into America was followed by a southward expansion wave along the ice-free Pacific coastal line (Tamm et al. 2007; Fagundes et al. 2008; Bodner et al. 2012), which crossed Mesoamerica and reached the northern latitudes of South America. Here, Paleoindians split, one subset followed the Pacific and colonized the Andean region, and another one expanded along the Atlantic (Reich et al. 2012). The rapid coastal migration movements played a major role in the colonization of the whole South American sub-continent, and genetic age estimates fit well with the archaeological evidence of human presence in the Southern Cone at the Monte Verde site (Chile) at least ~14.5 kya (Morlan 1990; Dillehay et al. 2015; Brandini et al. 2018).

The initial isolation period in Beringia caused both the loss and emergence of maternal lineages, affecting local haplogroup frequencies, as well the molecular differentiation from their East Asian ancestors. At least 16 Asian or Beringian mtDNA founders contributed to the initial settlement of America (Achilli et al. 2008; Perego et al. 2010; Hooshiar Kashani et al. 2012; Achilli et

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

al. 2013). As recorded in both autosomal (Reich et al. 2012; Llamas et al. 2016) and uniparental (Torroni et al. 1993; Bodner et al. 2012; de Saint Pierre et al. 2012b; Roewer et al. 2013) markers, shortly after the initial entry into the double continent, populations began to diverge, favored by small effective population sizes and long isolation periods, generally interrupted by only brief periods of limited gene flow. Phylogeographic analyses of mtDNA variation have shed light on large-scale demographic aspects of the Paleoindian colonization, but also on minor events that took place in the continent during the initial spread (Perego et al. 2009; Perego et al. 2010; Bodner et al. 2012; de Saint Pierre et al. 2012a; de Saint Pierre et al. 2012b; Achilli et al. 2013). The analysis of geographically restricted sub-haplogroups added further support to the scenario of an initial split of the first Paleoindian settlers of South America into two main groups whose expansion routes followed different paths - along the Pacific and the Atlantic coastlines - and highlighted the possibility of trans-Andean events of gene flow at different latitudes (Bodner et al. 2012). Finally, there is recent mtDNA evidence that present-day Peru and Ecuador were first populated by Paleoindian settlers of North American ancestry in the short time frame of 16.0–14.6 kya (Brandini et al. 2018). Despite this recent insight, the colonization process and subsequent population movements that led to the peopling of South America remain poorly known.

The aim of the present study is to shed light on the demographic movements occurred in the region after the initial colonization wave, identify and characterize trans-Andean gene flow, and explore its implications at the continental level.

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

Results

Phylogeography of Andean mtDNA lineages

Haplogroup B2 represents ~23% of the mtDNAs in South America, but its frequency varies across the sub-continent (Salas et al. 2009). In the Andean region, B2 ranges from moderate frequencies in the general population from Ecuador (20%) and in some Ecuadorian Natives, e.g. Cayapa (40%) (Rickards et al. 1999), to very high frequencies (up to 90%) in some Quechua and Aymara populations in the Andean regions of Bolivia and Peru (Bert et al. 2001; Barbieri et al. 2011; Gayà-Vidal et al. 2011). Several selected B2 sub-haplogroups were investigated here in order to further characterize their internal variation and shed light into the Paleoindian settlement of the South American continent.

We named as B2ai a new B2 sub-haplogroup that is defined by the stable transition C16168T (phylogenetic weight = 5.6 in PhyloTree 17 (Weissensteiner et al. 2016)) on top of the B2 mutational motif (**Figure 1A**). B2ai appears almost exclusively in the Peruvian and Bolivian Andes with approximately the same frequency in both Peru and Bolivia (~4%) (**Figure 1B)**, and it probably arose in this area ~8.4 kya (7.8-9.1) (**Figure 1C**) where the molecular diversity indices are also high (**Supplemental Table S1**). Control-region (CR) networks show a star-like pattern in both regions, mirroring recent population growth. Only one Argentinian mitogenome belongs to B2ai, clustering with one ancient Peruvian mitogenome within B2ai2a together with one ancient Peruvian mitogenome. CR sequences provide also further information (**Figure 1D**): haplogroup B2ai has spread into Ecuador and

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

northern Chile. In ancient DNA samples, there are a total of two mitogenomes and 20 CR sequences (with moderate diversity), all of them from Peru.

We identified a second new B2 sub-clade, B2aj (motif A8853G–G15883A–C16188T; **Figure 2**). Its molecular divergence suggests an origin slightly earlier than B2ai, ~10.4 kya (9.1-11.8). According to mitogenomes and CR sequences, B2aj is mainly distributed in Peru and Bolivia, but it is also abundant in the Argentinean Andes. A Peruvian origin for B2aj can be postulated taking into account that its molecular diversity is highest in Peru (*HD* = 0.7530; π = 0.0047) relative to other Andean countries where, while this lineage is found at a higher frequency (e.g. 20.4% in Bolivia *vs*. 10.8% in Peru), it shows much lower diversity (e.g. Bolivia *HD* = 0.5227; π = 0.0026; **Supplemental Table S1**). The network of B2aj CR sequences (**Figure 2B**) is clearly star-like, suggesting an important N_e growth in the Andes, especially in Bolivia (except around the Titicaca basin which is close to the present-day border with Peru; see below). B2aj spread most successfully into North Andean Argentina probably because of its high frequency in Bolivia. The interpolated frequency map shows that B2aj is almost exclusively present in the Andean region plus the above-mentioned Argentinean component (**Figure 2C**). One of its sub-clades (B2aj1a1a; TMRCA: 0.8 kya) might have originated recently in Argentina (**Figure 2A**). Overall, phylogenetic patterns suggest that the gene flow of B2aj was more favored southeastwards (Bolivia to Argentina) than in the opposite direction (Bolivia to Peru) (**Figure 2D**). Argentinean mitogenomes share several B2aj sub-clades with Bolivian samples, and these common sub-clades have TMRCAs ranging from 1.7 (B2aj4a2) to 5.2 (B2aj3c) ky, suggesting a continuous gene flow between these neighboring Andean regions. Ancient

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

B2aj samples were found in the Titicaca region (Tiwanaku; dated 0.8 kya) and in the northwestern Argentinean province of Jujuy (Quebrada and Puno; dated ~0.6-1 kya).

B2y mtDNAs are defined by the transition C16261T (phylogenetic weight = 4.2 in PhyloTree 17 (Weissensteiner et al. 2016)) (**Figure 3**). This is an ancient clade that most likely arose along the Pacific coast of North America ~14.3 kya (13.0-15.7) during the initial settlement of the continent. This scenario is supported by the extent of diversity from the root CR haplotype observed in Native samples from the US (**Figure 3C**) (Malhi et al. 2003; Kemp et al. 2010). The phylogeographic pattern of B2y suggests that its founding haplotype most likely entered South America following an expansion along the Pacific coast that reached Northern Chilean latitudes. B2y is only marginally found in the French Guiana (Mazieres et al. 2011) and in the Gavião (Brazilian Amazon forest (Ward et al. 1996)) (**Figure 3B**). The interpolated frequency map of B2y lineages (**Figure 3D**) shows a frequency peak in North America, high frequencies in the Andean area and a moderate presence in North Argentina. The sub-clade B2y2a1a most likely arose in Peru (**Figures 3A** and **3C**) 6.6 kya (5.3-7.9) and later spread into other Andean locations (Bolivia), and in particular into the Gran Chaco (north-eastern Argentina) (Cabana et al. 2006; Sevini et al. 2013), which was reached passing through the Jujuy province (north-western Argentina) (Cardoso et al. 2013). Ancient B2y DNAs were exclusively found in Peruvian samples.

B2b is defined by the transition G6755A and probably arose ~16.6 kya (15.6-17.5), a little earlier than B2y. Phylogeographic patterns of B2b mtDNAs suggest that this clade most likely arose in the North of Meso-America (Brandini

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

et al. 2018). The greatest proportion of mitogenomes are from South America (93.3%), mostly from the Andean region, especially from Peru (46.6%); **Figures 4A** and **4B**. It seems that B2b successfully spread through both the Pacific and the Atlantic coasts. The Pacific route was probably led by B2b11 ~14.7 kya (13.5-15.9), with B2b14 expanding slightly later (12.8 kya [11.0-11.6]); both most likely initially settled in Peru as indicated by the high CR diversity values relative to surrounding regions (e.g. Bolivia) (**Supplemental Table S1**). CR data show that several B2b14 haplotypes are shared between Peruvians and Bolivians, suggesting intense gene flow at these latitudes of the Andes; a few B2b14 mtDNAs were also detected in northern Argentineans, with a singleton in Buenos Aires that might signal recent migration (**Figure 4B; Figure 4C; Figure 4E**). In addition, four CR B2b14 sequences from ancient specimens were found in Pacapaccari, a location in the fringe area between the Andean highlands and the Pacific coast of Peru.

The geographic features of B2b11 are slightly different from those of B2b14. B2b11 is mainly restricted to the central regions of Peru with indications of limited gene flow southwards, reaching the highlands of Cusco and Andahuaylas in southern Peru (**Figure 4F**). Ancient B2b11 samples are only found along the coast of central Peru (Huaca Pucllana and Pueblo Viejo; **Figure 4A**). It is likely that B2b11 originated in the central coastal regions of Peru in close association with the Wari, Ychsma and Inca cultures that sequentially occupied that area from AD 600 to 1,500 (**Supplemental Table S2**).

The age of the B2b3 branch (15.8 ky) indicates that it probably played a part in the pioneering migration wave that moved southward from Mesoamerica

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

and then along the Atlantic façade of South America. The colonization of the Atlantic coast by B2b3a carriers occurred later than their B2b Andean counterparts, ~8.3 kya (7.0-9.5). The phylogeny based on the CR data shows clear signals of population expansion in the region with at least one main founder haplotype (**Figure 4B**). The origin of B2b3a (and B2b3a1 too) could be placed close to the Amazon River delta, as witnessed to by its high frequency in the region (**Figure 4D**) as well as the detection of a 4 ky old sample from Pirabas (North of Pará state) carrying the HVS-I motif T16189C–T16217C–T16249C–A16312G–C16344T (Ribeiro-dos-Santos et al. 1996). The sub-clade B2b3a2 arose in Brazil ~3.9 kya (3.3-4.8) and likely spread southwards following the Atlantic as testified by the presence of a B2b3a2 member in Uruguay (Guarani; (Sans et al. 2015)) and Argentinians from Río de la Plata (**Figure 4B**). The sub-clade B2b3a1 appeared earlier (6.8 kya [5.5-8.2]), probably in Brazil, although there is only one mitogenome in Uruguay representing its phylogenetic root (**Figure 4A**).

Most interesting is the derived sub-lineage B2b3a1a that clearly moved northwards from Venezuela/Brazil to the Caribbean ~6.0 kya (5.1-6.8) and solidly settled in Puerto Rico (**Figure 4B**), representing 100% of the B2 haplotypes found in this island (The 1000 Genomes Project Consortium 2012; Vilar et al. 2014). Outside Puerto Rico, there are only members of B2b3a1a in the northern coast of Venezuela (probably a crossing point to the Caribbean from Brazil) (Castro de Guerra et al. 2012; Brandini et al. 2018), in Cuba (Mendizabal et al. 2008) and the Dominican Republic (Tajima et al. 2004). The two mitogenomes and one CR sequence sampled in USA (**Figure 4A** and

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

Figure 4B, respectively) belonging to B2b3a1a might represent a recent arrival, probably from Puerto Rico, as recently suggested (Brandini et al. 2018).

Bayesian-based inferences of gene flow in the Andes

Of the nine demographic models tested, the one that received the highest probability (0.825) was the Stepping Stone linear model Peru -> Bolivia -> NW-Argentina, followed by a lineal model (probability = 0.175) that additionally considered bidirectional gene flow between Bolivia and Argentina (Peru -> Bolivia <-> NW-Argentina). However, according to the latter model, the number of migrants involved in a South to North movement from Northwest Argentina to Bolivia would have been very low (**Supplemental Table S3**).

Gene flow models involving populations from Andean Bolivia and northern Argentina (NW-Argentina and NE-Argentina; **Supplemental Table S4)** were also evaluated. Again, the unidirectional gene flow Bolivia -> NW-Argentina -> NE-Argentina was the best fitting model (probability = 1.0). We also observed twice as high gene flow for Bolivia -> NW-Argentina than for NW-Argentina -> NE-Argentina.

Genome-wide autosomal data and trans-Andean gene flow

More than 360K autosomal SNPs were analyzed in Native South Americans in order to shed light into the role played by the Andes in the demography of the sub-continent. We first computed IBS and F_{ST} distances between Peruvians (PEL-1000G) and the rest of southern Native Americans. An initial MDS plot of IBS values showed South Americans clustering together when plotted against populations from Europe and Africa (**Supplemental Figure S1**). When exploring only South American samples in a MDS plot, two populations

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

appeared as genetic outliers, the Surui and the Karitiana (**Supplemental Figure S2**); for this reason, they were removed from subsequent analyses. Dimension 1 of the MDS in **Figure 5A** clusters all Natives from the Andes (Quechua, Peru, Aymara and Diaguita) on one side of the plot, supporting their close relationships but also their differentiation or demographic isolation from the other native groups. Dimension 2 shows at one pole of the plot the Southern Cone populations (Yaghan, Chilote, Chono and Hulliche) and at another pole those living in the northern extreme of the sub-continent (Arhuaco and Kogi), whereas the remaining populations scatter in between.

Figure 5B shows a map of interpolated F_{ST} values between Peruvians and other Native American samples from all across South America, highlighting the close genetic proximity between Peruvians and neighboring Andean populations as well as the relevant genetic distances with the rest of the sub-continent. The map of interpolated IBS values shows the same pattern (**Supplemental Figure S3**).

Admixture analysis consistently suggests a shared ancestry between all Andean populations (**Figure 5C**). IBS distances between Peru and other populations in South America were also geographically interpolated in order to visualize the relationships between Andean populations and the rest. The map (**Figure 5B**) displays a virtual genetic continuity in the Andes indicating a substantial genetic isolation of this mountainous area from the rest of the South American native populations.

Both mtDNA and autosomal data point to a limited trans-Andean genetic diffusion from Peru (and neighboring populations) towards the East of South

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

America. In order to formally test this hypothesis, we first defined population sets as follows:

- $X_{WEST\text{-}ANDEAN}$: represents populations PEL-1000G, Aymara, and Quechua, which live on the western side of the Andes;
- Y: represents the Chibchan-Paezan and Equatorial-Tucanoan linguistic families of the eastern path of South American spread;
- $X_{EAST\text{-}ANDEAN}$: represents the populations that are located in between $X_{WEST\text{-}ANDEAN}$ and Y-populations; that is, populations such as the Chane, Guarani, Kaingang, Toba, Diaguita, and Wichi, which live immediately to the East of the Andes mountain range at sub-Amazon latitudes (it is at this latitude that genetic exchange could have been most favored).

Then, we computed *D*-statistics as follows:

A) *D*(Y, $X_{EAST\text{-}ANDEAN}$, $X_{WEST\text{-}ANDEAN}$, OUT): statistically tests if there was gene flow between $X_{EAST\text{-}ANDEAN}$ populations and populations Y (East South America); **Figure 5D**.

B) *D*($X_{WEST\text{-}ANDEAN}$, $X_{EAST\text{-}ANDEAN}$, Y, OUT): tests if there was gene flow between $X_{EAST\text{-}ANDEAN}$ and $X_{WEST\text{-}ANDEAN}$; **Figure 5D.**

The analyses indicate that the Diaguita (NW-Argentina) is the only population of northern Argentina with a statistically significant input from populations living in the Peruvian/Bolivian Andes (Aymara and Quechua).

Furthermore, all populations living in the eastern side of the Andes (northern Argentina: Wichi, Diaguita, Toba, Chane, etc.) show a statistically significant contribution from populations representing the eastern path of the

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

South American spread when using PEL-1000G as a reference population, a signal that is diluted when using Aymara and Quechua.

Dating the time of admixture between Andean Natives and Diaguita

Although diffusion of genetic variation from the Andes into southern latitudes has been limited, statistical models indicate that there was moderate gene flow with populations living in NW-Argentina, the Diaguita. The track length distribution of genomic segments inherited from different ancestral populations allows to date the admixture time. This analysis suggests that a main admixture event involved the ancestors of modern Diaguita, with an Equatorial-Tucanoan component from the southern latitudes of South America (using Guahibo, Wayuu, Piapoco, Guahibo, Palikur, and Parakana as surrogate population sets) and a Peruvian (PEL-1000G) component, an event that occurred in two main pulses ~21 and 22 generations ago (525-550 ya; using 25 years per generation) (**Figure 5E**).

Discussion

Phylogeographic data inferred from mitogenomes highlight the demographic importance of the Andean region during the South American settlement process and suggest a history of evolutionary success and high-altitude adaptation. Goldberg et al. (2016) have recently reconstructed the spatiotemporal patterns of human population growth in South America using archaeological information and suggested two main demographic periods: (a) an initial one involving a rapid colonization but low N_e and lasting ~8 kya, and (b) a period of population growth along the Pacific coastline (in Peru, Chile and Ecuador) beginning ~5 kya and coinciding with the Neolithic transition (and the onset of sedentism). In

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

Figure 6 we show the location of the mDNA lineages analyzed in the present study together with the interpolated archaeological data from Goldberg et al. (2016) (see also Figure 1 of (Goldberg et al. 2016)). The map shows the great fit between archaeological and mtDNA evidence, then suggesting that B2 is a good marker of the demographic transformation occurred in the Andes. The highest frequencies of B2 in South America are by far in the Andean regions of Peru, Bolivia, and northern Argentina, reaching frequencies up to 70-90% in some Aymara and Quechua Andean populations (Bert et al. 2001; Barbieri et al. 2011; Heinz et al. 2013; Heinz et al. 2015). The pre-Columbian mtDNA data available follow the same pattern: B2 reaches high frequencies in the Peruvian highlands, ranging from 42% to 75% (Shinoda et al. 2006; Baca et al. 2012; Baca et al. 2014). Ancient and contemporary data indicate a genetic continuity in the Central Andes that is compatible with a high N_e and intense gene flow in this area for a long time period (8-9 ky), with limited introgression from surrounding regions (Batai and Williams 2014; Cabana et al. 2014). Moreover, the observed genetic patterns in the Central Andes can be tentatively linked to the population expansions of the most relevant Andean empires: the Tiwanaku, the Wari, and more recently the Incas (Fehren-Schmitz et al. 2014).

The good fit between archaeological and genetic data finds further support in the EBSPs of N_e population growth. B2ai shows a sudden increase of N_e starting ~10 kya, while South American B2y and B2aj show a more progressive growth in a recent time range (from 4 to 6 kya). These lineages are located almost exclusively in the Andes and probably emerged and diversified in the Andean area soon after the initial colonization wave (**Figure 7A**). Whilst population growth inferred from B2ai suggests the existence of an important

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

population expansion starting already during the Andean Preceramic periods II to III (a fact that is not highlighted in the archaeological record; **Figure 7A**), lineages such as South American B2y and B2aj experienced a great population growth starting ~5 kya, thus perfectly matching the beginning of the major demographic expansion in the Andes.

Autosomal data satisfy this demographic scenario, indicating a strong genetic affinity between the populations of the Andean region and their isolation with respect to other American regions (**Figure 5**).

Furthermore, our data also clarify the admixture patterns occurring immediately to the East of the Andes at northern Argentinean latitudes. The genome of present-day Diaguitas seems to be the consequence of a single event or multiple admixture events that occurred ~525-550 ya between a Peruvian-Central Andean component and an ancestral one of Equatorial-Tucanoan origin. The time estimate agrees remarkably well with the known expansion of the Incas into Diaguita territories occurring >500 ya, at the time of the Inca Emperor Túpac Inca Yupanqui (Diaguita is most likely a Quechua voice imposed by the Incas (Dillehay and Netherly 1998)).

By integrating all the available evidence, it is possible to conceive a demographic model for the peopling of South America (**Figure 7B**). A few waves of Paleoindians could have reached the northern edge of South America ~16-15 kya. These settlers advanced across two main colonization routes, following the Atlantic and the Pacific coasts. The populations that settled the Peruvian Andes were demographically very successful, and experienced rapid population growth from at least 8 kya to the present day, which led to the incubation of important genetic variation in the highlands. Substantial gene flow

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

existed in the territory spanning present-day Peru and Andean Bolivia. Population movements were most favored in a North to South direction, not bidirectionally. The Paleoindians living at this latitude of the Andes had very limited genetic exchange with the East of the sub-continent, and even with populations living to the South (with the exception of those that settled the narrow Andean fringe running along the present-day Chilean territories). Probably, the Amazon forest acted as a geographical barrier to gene flow between the Pacific and the Atlantic expansion waves. The Atlantic migration wave played a major role in the colonization of the non-Andean regions. Although demographic growth was likely more moderate here than in the Peruvian Andes in ancient times, it might have been more relevant in more recent times as suggested by the N_e estimated from the Brazilian B2b3 clade, with a quick growth starting 2.5 kya (**Supplemental Figure S4**). At this time, population movements along the Atlantic coast could have been important not only in a North to South direction, but also with an opposite component that might have even involved movements towards the Caribbean. It is haplogroup B2b3a1a that best testifies to this continental-Caribbean connection. Thus, while the ancestral nodes of B2b3 were found in the Atlantic coast of South America (reaching at least Uruguay), we observed the derived B2b3a1a sub-haplogroup in Puerto Rico at relatively high frequencies (here, 100% of B2 haplotypes are B2b3a1a; representing 8.3% of the Native American component), most likely arriving there ~6 kya (almost ~10 kya after the occurrence of B2b3 in continental South America). B2b3a1a could have been part of the Taino's legacy in the modern Caribbean population, representing an ancestral northern South American input in the Greater Antilles. In turn, this

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

lineage possibly arose (from an ancestral Brazilian node) around the Orinoco Basin (Venezuela) not earlier than 6 kya. Alternatively, B2b3a1a could have arisen locally in the Tainos from Puerto Rico, from where it could have spread across the Caribbean and to northern Venezuela. A recent paper on an ancient B2 Lucayan Taino sample from the Bahamas (calibrated to AD 776-992) raises the possibility of a pre-Columbian arrival of haplogroup B2 to the Greater Antilles (Schroeder et al. 2017). This Taino specimen shares ancestry with modern Puerto Ricans and is related to Arawakan speakers from northern South America.

Overall, genetic and archaeological data point to the Andes as the main demographic source for South America in pre-historic and historical times. The homogeneous ecological environment of the highlands most likely favored the incubation of prosperous civilizations and cultures, a complex set of circumstances that was not as favorable in other continental regions. Adaptation to high altitude might have pushed these populations to persist and prosper there. Such a scenario would explain the very limited inter-play between the two main routes of colonization of the sub-continent and the scarce genetic exchange with other South American populations.

Methods

Samples, DNA extraction and amplification

We selected a total of 146 mitogenomes from ancient and modern populations belonging to haplogroup B2 (**Supplemental Table S5**); 72 were newly sequenced in this study, while the rest were retrieved from GenBank (https://www.ncbi.nlm.nih.gov/genbank/; *n* = 59), and from The 1000 Genomes Project (www.1000Genomes.org; *n* = 15; hereafter, 1000G) using the GATK

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

Toolkit from the 1000G repository and processed as in Gómez-Carballa et al. (2015). We also collected samples from natives living in the northernmost Andean provinces of Argentina: 45 Kollas from Salta and Jujuy, and 24 Diaguitas from Catamarca (**Supplemental Table S6**). DNAs were extracted following standard phenol-chloroform protocols. Written informed consent was obtained under protocols approved by the Asociación de Bioquímicos de Córdoba (Argentina), and the Universidad Católica de Córdoba (Argentina).

In addition, we compiled 822 CR sequences from the literature belonging to the targeted B2 sub-clades (**Supplemental Table S2**). Sub-haplogroup affiliations were inferred from the available sequence data (generally HVS-I and/or HVS-II).

PCR amplification, sequencing and SNP analysis was carried out following the protocol described in **Supplemental Text**.

Statistical analyses on mtDNA data

Diversity indices were computed as indicated in **Supplemental Text**.

mtDNAs were classified into haplogroups and checked for data quality using HaploGrep 2 (Weissensteiner et al. 2016). CR data from **Supplemental Table S2** were used to build the networks in **Figures 1, 2, 3 and 4 (**sequence range nps 16090-16362). Subsequently, the maximum likelihood (ML) approach implemented in PAMLX 1.3.1 {Xu, 2013 #5500} was used to estimate the Time to the Most Recent Common Ancestor (TMRCA) for the main phylogenetic nodes as in Brandini et al. (2018). The MP phylogenetic tree was built referring to the haplogroup nomenclature of PhyloTree (van Oven and Kayser 2009). The 72 novel B2 mitogenomes, together with other 570 Native American mitogenomes previously employed in Brandini et al. (2018) (**Supplemental**

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

Table S7), were used to estimate sub-haplogroup ages (**Supplemental Table S8)** (details in **Supplemental Text)**.

Extended Bayesian Skyline Plots and trans-Andean gene flow analysis were carried out as explained in **Supplemental Text**. Maps of interpolated frequencies were built as detailed in **Supplemental Text**.

Demography inferred from genome autosomal data

SNP data were compiled and intersected from two main genome repositories: Native American data available from Reich et al. (2012) and data from 1000G. SNP genotypes were masked for the Native American component using PCAdmix (Brisbin et al. 2012); and BEAGLE 3.3.2 (Browning and Browning 2007) was used to prepare the necessary SNP unphased and imputed data for PCAdmix. The final dataset consisted of 362,765 SNPs representing Native American variation in a total of 28 South American ethnic groups (**Supplemental Table S9**). We initially computed identity-by-state (IBS) values from SNP data using PLINK (Purcell et al. 2007). Multidimensional scaling (MDS) was carried out on a matrix of pair-wise identity-by-state (IBS) values, using the function *cmdscale* from the library *stats* from R (http://www.r-project.org; (R core development team 2011). Admixture components of Native American populations were estimated using ADMIXTURE (Alexander et al. 2009). *D*-statistics were computed to formally test for genetic admixture using ADMIXTOOLS (Patterson et al. 2012). An artificial outgroup was built as in Pardo-Seco et al. (2016). Timing of admixture was obtained from the analysis of tract length of chromosomal segments of different ancestries using the *Tracts* (Gravel 2012). IBS and F_{ST} distances were also displayed in geographic interpolated maps (**Supplemental Text).**

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

Data access

Newly generated mitogenome sequences have been submitted to the NCBI GenBank (https://www.ncbi.nlm.nih.gov/genbank/) under accession numbers MG637053 - MG637124.

Description of supplemental data file

Supplemental data file includes a supplemental text, four figures and eleven tables.

Conflicts of interest

There are not conflicts of interest.

Acknowledgements

We thank Juan Carlos Jaime for help, support, and access to samples. This study received support from the Instituto de Salud Carlos III (Proyecto de Investigación en Salud, Acción Estratégica en Salud): project GePEM ISCIII/PI16/01478/Cofinanciado FEDER) (to AS); project ReSVinext ISCIII/PI16/01569/Cofinanciado FEDER (to FM-T); Consellería de Sanidade, Xunta de Galicia (RHI07/2-intensificación actividad investigadora, PS09749 and 10PXIB918184PR), Instituto de Salud Carlos III (Intensificación de la actividad investigadora 2007–2012, PI16/01569) (to FM-T); Fondo de Investigación Sanitaria (FIS; PI070069/PI1000540) del plan nacional de I + D + I and 'fondos FEDER' (to FM-T); 2016-PG071 Consolidación e Estructuración REDES 2016GI-1344 G3VIP (Grupo Gallego de Genética Vacunas Infecciones y Pediatría, ED341D R2016/021) (to AS and FM-T); the University of Pavia strategic theme "Towards a governance model for

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

international migration: an interdisciplinary and diachronic perspective" (MIGRAT-IN-G) (to AA, AO and AT); the Italian Ministry of Education, University and Research: Progetti Futuro in Ricerca 2012 (RBFR126B8I) (to AA and AO); the Progetti Ricerca Interesse Nazionale 2012 (to AA and AT); and the ERC Consolidator Grant: CoG-2014, No. 648535 (to AA).

References

Achilli A, Perego UA, Bravi CM, Coble MD, Kong QP, Woodward SR, Salas A, Torroni A, Bandelt HJ. 2008. The phylogeny of the four pan-American mtDNA haplogroups: implications for evolutionary and disease studies. *PLoS One* **3**: e1764.

Achilli A, Perego UA, Lancioni H, Olivieri A, Gandini F, Hooshiar Kashani B, Battaglia V, Grugni V, Angerhofer N, Rogers MP et al. 2013. Reconciling migration models to the Americas with the variation of North American native mitogenomes. *Proc Natl Acad Sci U S A* **110**: 14308-14313.

Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, Scozzari R, Cruciani F, Zeviani M, Briem E, Carelli V et al. 2004. The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *Am J Hum Genet* **75**: 910-918.

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* **19**: 1655-1664.

Baca M, Doan K, Sobczyk M, Stankovic A, Weglenski P. 2012. Ancient DNA reveals kinship burial patterns of a pre-Columbian Andean community. *BMC Genet* **13**: 30.

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

Baca M, Molak M, Sobczyk M, Weglenski P, Stankovic A. 2014. Locals, resettlers, and pilgrims: a genetic portrait of three pre-Columbian Andean populations. *Am J Phys Anthropol* **154**: 402-412.

Barbieri C, Heggarty P, Castri L, Luiselli D, Pettener D. 2011. Mitochondrial DNA variability in the Titicaca basin: Matches and mismatches with linguistics and ethnohistory. *Am J Hum Biol* **23**: 89-99.

Batai K, Williams SR. 2014. Mitochondrial variation among the Aymara and the signatures of population expansion in the central Andes. *Am J Hum Biol* **26**: 321-330.

Battaglia V, Grugni V, Perego UA, Angerhofer N, Gomez-Palmieri JE, Woodward SR, Achilli A, Myres N, Torroni A, Semino O. 2013. The first peopling of South America: new evidence from Y-chromosome haplogroup Q. *PLoS One* **8**: e71390.

Bert F, Corella A, Gene M, Pérez-Pérez A, Turbón D. 2001. Major mitochondrial DNA haplotype heterogeneity in highland and lowland Amerindian populations from Bolivia. *Hum Biol* **73**: 1-16.

Bodner M, Perego UA, Huber G, Fendt L, Röck AW, Zimmermann B, Olivieri A, Gómez-Carballa A, Lancioni H, Angerhofer N et al. 2012. Rapid coastal spread of First Americans: Novel insights from South America's Southern Cone mitochondrial genomes. *Genome Res* **22**: 811-820.

Brandini S, Bergamaschi P, Cerna MF, Gandini F, Bastaroli F, Bertolini E, Cereda C, Ferretti L, Gómez-Carballa A, Battaglia V et al. 2018. The Paleo-Indian entry into South America according to mitogenomes. *Mol Biol Evol* **35**: 299-311.

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

Brisbin A, Bryc K, Byrnes J, Zakharia F, Omberg L, Degenhardt J, Reynolds A, Ostrer H, Mezey JG, Bustamante CD. 2012. PCAdmix: principal components-based assignment of ancestry along each chromosome in individuals with admixed ancestry from two or more populations. *Hum Biol* **84**: 343-364.

Browning SR, Browning BL. 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.

Cabana GS, Lewis CM, Jr., Tito RY, Covey RA, Caceres AM, Cruz AF, Durand D, Housman G, Hulsey BI, Iannacone GC et al. 2014. Population genetic structure of traditional populations in the Peruvian Central Andes and implications for South American population history. *Hum Biol* **86**: 147-165.

Cabana GS, Merriwether DA, Hunley K, Demarchi DA. 2006. Is the genetic structure of Gran Chaco populations unique? Interregional perspectives on native South American mitochondrial DNA variation. *Am J Phys Anthropol* **131**: 108-119.

Cardoso S, Palencia-Madrid L, Valverde L, Alfonso-Sanchez MA, Gomez-Perez L, Alfaro E, Bravi CM, Dipierri JE, Pena JA, de Pancorbo MM. 2013. Mitochondrial DNA control region data reveal high prevalence of Native American lineages in Jujuy province, NW Argentina. *Forensic Sci Int Genet* **7**: e52-55.

Castro de Guerra D, Figuera Pérez C, Bravi CM, Saunier J, Scheible M, Irwin J, Coble MD, Rodriguez-Larralde A. 2012. Sequence variation of

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

mitochondrial DNA control region in North Central Venezuela. *Forensic Sci Int Genet* **6**: e131-133.

de Saint Pierre M, Bravi CM, Motti JM, Fuku N, Tanaka M, Llop E, Bonatto SL, Moraga M. 2012a. An alternative model for the early peopling of southern South America revealed by analyses of three mitochondrial DNA haplogroups. *PLoS One* **7**: e43486.

de Saint Pierre M, Gandini F, Perego UA, Bodner M, Gómez-Carballa A, Corach D, Angerhofer N, Woodward SR, Semino O, Salas A et al. 2012b. Arrival of Paleo-indians to the Southern Cone of South america: new clues from mitogenomes. *PLoS One* **7**: e51311.

Dillehay TD, Netherly P. 1998. *La frontera del estado inca*. Abya-Yala, Quito, Ecuador.

Dillehay TD, Ocampo C, Saavedra J, Sawakuchi AO, Vega RM, Pino M, Collins MB, Scott Cummings L, Arregui I, Villagran XS et al. 2015. New archaeological evidence for an early human presence at Monte Verde, Chile. *PLoS One* **10**: e0141923.

Elias SA, Crocker B. 2008. The Bering Land Bridge: a moisture barrier to the dispersal of steppe-tundra biota. *Quat Sci Rev* **27**: 2473-2483.

Fagundes NJ, Kanitz R, Eckert R, Valls AC, Bogo MR, Salzano FM, Smith DG, Silva WA, Jr., Zago MA, Ribeiro-dos-Santos AK et al. 2008. Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. *Am J Hum Genet* **82**: 583-592.

Fehren-Schmitz L, Haak W, Machtle B, Masch F, Llamas B, Cagigao ET, Sossna V, Schittek K, Isla Cuadrado J, Eitel B et al. 2014. Climate

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

change underlies global demographic, genetic, and cultural transitions in pre-Columbian southern Peru. *Proc Natl Acad Sci U S A* **111**: 9443-9448.

Fehren-Schmitz L, Harkins KM, Llamas B. 2017. A paleogenetic perspective on the early population history of the high altitude Andes. *Quaternary Int* **461**: 25-33.

Forster P, Harding R, Torroni A, Bandelt H-J. 1996. Origin and evolution of Native American mtDNA variation: a reappraisal. *Am J Hum Genet* **59**: 935-945.

Gamble C, Davies W, Pettitt P, Richards M. 2004. Climate change and evolving human diversity in Europe during the last glacial. *Philos Trans R Soc Lond B Biol Sci* **359**: 243-253; discussion 253-244.

Gayà-Vidal M, Moral P, Saenz-Ruales N, Gerbault P, Tonasso L, Villena M, Vasquez R, Bravi CM, Dugoujon J-M. 2011. mtDNA and Y-chromosome diversity in Aymaras and Quechuas from Bolivia: Different stories and special genetic traits of the Andean Altiplano populations. *Am J Phys Anthropol* **145**: 215-230.

Gilbert MT, Jenkins DL, Gotherstrom A, Naverán N, Sánchez JJ, Hofreiter M, Thomsen PF, Binladen J, Higham TF, Yohe RM, 2nd et al. 2008. DNA from pre-Clovis human coprolites in Oregon, North America. *Science* **320**: 786-789.

Goebel T, Waters MR, O'Rourke DH. 2008. The late Pleistocene dispersal of modern humans in the Americas. *Science* **319**: 1497-1502.

Goldberg A, Mychajliw AM, Hadly EA. 2016. Post-invasion demography of prehistoric humans in South America. *Nature* **532**: 232-235.

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

Gómez-Carballa A, Olivieri A, Behar DM, Achilli A, Torroni A, Salas A. 2012. Genetic continuity in the Franco-Cantabrian region: New clues from autochthonous mitogenomes. *PLoS One* **7**: e32851.

Gómez-Carballa A, Pardo-Seco J, Amigo J, Martinón-Torres F, Salas A. 2015. Mitogenomes from The 1000 Genome Project reveal new Near Eastern features in present-day Tuscans. *PLoS One* **10**: e0119242.

Gravel S. 2012. Population genetics models of local ancestry. *Genetics* **191**: 607-619.

Heinz T, Álvarez-Iglesias V, Pardo-Seco J, Taboada-Echalar P, Gómez-Carballa A, Torres-Balanza A, Rocabado O, Carracedo A, Vullo C, Salas A. 2013. Ancestry analysis reveals a predominant Native American component with moderate European admixture in Bolivians. *Forensic Sci Int Genet* **7**: 537-542.

Heinz T, Cárdenas JM, Álvarez-Iglesias V, Pardo-Seco J, Gómez-Carballa J, Santos C, Taboada-Echalar P, Salas A. 2015. The genomic legacy of the transatlantic slave trade in the Yungas valley of Bolivia. *PLoS One* **10**: e0134129.

Hoffecker JF, Elias SA, O'Rourke DH. 2014. Anthropology. Out of Beringia? *Science* **343**: 979-980.

Hoffecker JF, Elias SA, O'Rourke DH, Scott GR, Bigelow NH. 2016. Beringia and the global dispersal of modern humans. *Evol Anthropol* **25**: 64-78.

Hooshiar Kashani B, Perego UA, Olivieri A, Angerhofer N, Gandini F, Carossa V, Lancioni H, Semino O, Woodward SR, Achilli A et al. 2012. Mitochondrial haplogroup C4c: a rare lineage entering America through the ice-free corridor? *Am J Phys Anthropol* **147**: 35-39.

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

Kemp BM, Gonzalez-Oliver A, Malhi RS, Monroe C, Schroeder KB, McDonough J, Rhett G, Resendez A, Penaloza-Espinosa RI, Buentello-Malo L et al. 2010. Evaluating the Farming/Language Dispersal Hypothesis with genetic variation exhibited by populations in the Southwest and Mesoamerica. *Proc Natl Acad Sci U S A* **107**: 6759-6764.

Llamas B, Fehren-Schmitz L, Valverde G, Soubrier J, Mallick S, Rohland N, Nordenfelt S, Valdiosera C, Richards SM, Rohrlach A et al. 2016. Ancient mitochondrial DNA provides high-resolution time scale of the peopling of the Americas. *Sci Adv* **2**: e1501385.

Malhi RS, Mortensen HM, Eshleman JA, Kemp BM, Lorenz JG, Kaestle FA, Johnson JR, Gorodezky C, Smith DG. 2003. Native American mtDNA prehistory in the American Southwest. *Am J Phys Anthropol* **120**: 108-124.

Mazieres S, Callegari-Jacques SM, Crossetti SG, Dugoujon JM, Larrouy G, Bois E, Crubezy E, Hutz MH, Salzano FM. 2011. French Guiana Amerindian demographic history as revealed by autosomal and Y-chromosome STRs. *Ann Hum Biol* **38**: 76-83.

Mendizabal I, Sandoval K, Berniell-Lee G, Calafell F, Salas A, Martínez-Fuentes A, Comas D. 2008. Genetic origin, admixture, and asymmetry in maternal and paternal human lineages in Cuba. *BMC Evol Biol* **8**: 213.

Morlan RE. 1990. Monte Verde. A Late Pleistocene Settlement in Chile. Vol. 1, Palaeoenvironment and Site Context. Tom D. Dillehay. Smithsonian Institution Press, Washington, DC, 1989. xxiv, 306 pp., illus. $49.95. Smithsonian Series in Archaeological Inquiry. *Science* **249**: 937-938.

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

O'Rourke DH, Raff JA. 2010. The human genetic history of the Americas: the final frontier. *Curr Biol* **20**: R202-207.

Pala M, Olivieri A, Achilli A, Accetturo M, Metspalu E, Reidla M, Tamm E, Karmin M, Reisberg T, Hooshiar Kashani B et al. 2012. Mitochondrial DNA signals of late glacial recolonization of Europe from near eastern refugia. *Am J Hum Genet* **90**: 915-924.

Pardo-Seco J, Heinz T, Taboada-Echalar P, Martinón-Torres F, Salas A. 2016. Mapping the genomic mosaic of two 'Afro-Bolivians' from the isolated Yungas valleys. *BMC Genomics* **17**: 207.

Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. 2012. Ancient admixture in human history. *Genetics* **192**: 1065-1093.

Perego UA, Achilli A, Angerhofer N, Accetturo M, Pala M, Olivieri A, Hooshiar Kashani B, Ritchie KH, Scozzari R, Kong QP et al. 2009. Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. *Curr Biol* **19**: 1-8.

Perego UA, Angerhofer N, Pala M, Olivieri A, Lancioni H, Kashani BH, Carossa V, Ekins JE, Gomez-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.

Pitulko VV, Nikolsky PA, Girya EY, Basilyan AE, Tumskoy VE, Koulakov SA, Astakhov SN, Pavlova EY, Anisimov MA. 2004. The Yana RHS site: humans in the Arctic before the last glacial maximum. *Science* **303**: 52-56.

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

Pitulko VV, Tikhonov AN, Pavlova EY, Nikolskiy PA, Kuper KE, Polozov RN. 2016. Paleoanthropology. Early human presence in the Arctic: Evidence from 45,000-year-old mammoth remains. *Science* **351**: 260-263.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**: 559-575.

R core development team. 2011. R: A Language and Enviroment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Raghavan M, Skoglund P, Graf KE, Metspalu M, Albrechtsen A, Moltke I, Rasmussen S, Stafford TW, Jr., Orlando L, Metspalu E et al. 2014. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**: 87-91.

Raghavan M Steinrucken M Harris K Schiffels S Rasmussen S DeGiorgio M Albrechtsen A Valdiosera C Avila-Arcos MC Malaspinas AS et al. 2015. Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* **349**: aab3884.

Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, Ray N, Parra MV, Rojas W, Duque C, Mesa N et al. 2012. Reconstructing Native American population history. *Nature* **488**: 370-374.

Ribeiro-dos-Santos ÂKC, Santos SE, Machado AL, Guapindaia V, Zago MA. 1996. Heterogeneity of mitochondrial DNA haplotypes in Pre-Columbian Natives of the Amazon region. *Am J Phys Anthropol* **101**: 29-37.

Rickards O, Martinez-Labarga C, Lum JK, De Stefano GF, Cann RL. 1999. mtDNA history of the Cayapa Amerinds of Ecuador: detection of

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

additional founding lineages for the Native American populations. *Am J Hum Genet* **65**: 519-530.

Roewer L, Nothnagel M, Gusmão L, Gomes V, González M, Corach D, Sala A, Alechine E, Palha T, Santos N et al. 2013. Continent-wide decoupling of Y-chromosomal genetic variation from language and geography in native South Americans. *PLoS Genet* **9**: e1003460.

Salas A, Lovo-Gómez J, Álvarez-Iglesias V, Cerezo M, Lareu MV, Macaulay V, Richards MB, Carracedo Á. 2009. Mitochondrial echoes of first settlement and genetic continuity in El Salvador. *PLoS One* **4**: e6882.

Sans M, Mones P, Figueiro G, Barreto I, Motti JM, Coble MD, Bravi CM, Hidalgo PC. 2015. The mitochondrial DNA history of a former native American village in northern Uruguay. *Am J Hum Biol* **27**: 407-416.

Schroeder H, Sikora M, Gopalakrishnan S, Cassidy LM, Delser PM, Sandoval-Velasco M, Schraiber JG, Rasmussen S, Homburger JR, Ávila-Arcos MC et al. 2017. Origins and genetic legacies of the Caribbean Taino. *Proc Natl Acad Sci U S A*; **submitted**.

Schurr TG, Sherry ST. 2004. Mitochondrial DNA and Y chromosome diversity and the peopling of the Americas: evolutionary and demographic evidence. *Am J Hum Biol* **16**: 420-439.

Sevini F, Yao DY, Lomartire L, Barbieri A, Vianello D, Ferri G, Moretti E, Dasso MC, Garagnani P, Pettener D et al. 2013. Analysis of population substructure in two sympatric populations of Gran Chaco, Argentina. *PLoS One* **8**: e64054.

Shinoda K, Adachi N, Guillen S, Shimada I. 2006. Mitochondrial DNA analysis of ancient Peruvian highlanders. *Am J Phys Anthropol* **131**: 98-107.

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

Tajima A, Hamaguchi K, Terao H, Oribe A, Perrotta VM, Baez CA, Arias JR, Yoshimatsu H, Sakata T, Horai S. 2004. Genetic background of people in the Dominican Republic with or without obese type 2 diabetes revealed by mitochondrial DNA polymorphism. *J Hum Genet* **49**: 495-499.

Tamm E, Kivisild T, Reidla M, Metspalu M, Smith DG, Mulligan CJ, Bravi CM, Rickards O, Martinez-Labarga C, Khusnutdinova EK et al. 2007. Beringian standstill and spread of Native American founders. *PLoS ONE* **2**: e829.

The 1000 Genomes Project Consortium. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**: 56-65.

Torroni A, Bandelt HJ, Macaulay V, Richards M, Cruciani F, Rengo C, Martinez-Cabrera V, Villems R, Kivisild T, Metspalu E et al. 2001. A signal, from human mtDNA, of postglacial recolonization in Europe. *Am J Hum Genet* **69**: 844-852.

Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, Vullo CM, Wallace DC. 1993. Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am J Hum Genet* **53**: 563-590.

Torroni A, Schurr TG, Yang CC, Szathmary EJ, Williams RC, Schanfield MS, Troup GA, Knowler WC, Lawrence DN, Weiss KM et al. 1992. Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. *Genetics* **130**: 153-162.

van Oven M, Kayser M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* **30**: E386-E394.

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

Vilar MG, Melendez C, Sanders AB, Walia A, Gaieski JB, Owings AC, Schurr TG, Genographic C. 2014. Genetic diversity in Puerto Rico and its implications for the peopling of the Island and the West Indies. *Am J Phys Anthropol* **155**: 352-368.

Ward RH, Salzano FM, Bonatto SL, Hutz MH, Coimbra CEA, Santos RV. 1996. Mitochondrial DNA polymorphism in 3 Brazilian Indian tribes. *Am J Hum Biol* **8**: 317-323.

Weissensteiner H, Pacher D, Kloss-Brandstätter A, Forer L, Specht G, Bandelt H-J, Kronenberg F, Salas A, Schonherr S. 2016. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res* **44**: W58-W63.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**: 1586-1591.

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

Legend to the Figures

Figure 1. (A) Maximum parsimony tree of B2ai complete mitogenomes. ML coalescence ages (in ky) are shown below haplogroup names. **(B)** Interpolated frequencies of B2ai mtDNAs in America; red dots indicate sampling points. **(C)** Main migration routes of B2ai as inferred from phylogeographic data. **(D)** Network of B2ai CR haplotypes obtained from population surveys (listed in **Supplemental Table S2**). The position of the revised Cambridge reference sequence (rCRS) is indicated for reading sequence motifs. White circles in the phylogenies indicate Andean geographic origin and grey circles unknown origin. In panels A and D, mtDNA mutations are indicated along the branches of the phylogenies; they are transitions unless a suffix A, C, G, or T is included to indicate a transversion. Other suffixes indicate: insertions (+), synonymous substitutions (s), mutational changes in tRNA (-t), mutational changes in rRNA (-r), non-coding variants located in the mtDNA coding region (-nc). Amino acid replacements are in round brackets, underlined mutations indicate homoplasies in the same tree, while the prefix "@" indicates a back mutation. Haplotypes from ancient samples are in boxes (and red numbers), while haplotypes from present-day samples are in circles. Geographic and/or ethnic origins of the mitogenomes in panel A are provided in **Supplemental Table S5**, while the two-letter codes inside circles in panel D refer to ethnic origins (according to **Supplemental Table S2**) with numbers indicating the number of mtDNAs. Boxes with blue borders in panel A indicate sub-haplogroups that are new relative to the reference phylogeny in PhyloTree Build 17. Mutational hotspot variants at positions 16182, 16183, and 16519, as well as variation around position 310 and length or point heteroplasmies were not considered for the

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

phylogenetic reconstruction. Mitochondrial DNA sequence data for comparisons, phylogenetic trees, and maps of interpolated frequencies were obtained from the literature (**Supplemental Table S10**).

Figure 2. (A) Maximum parsimony tree of B2aj complete mitogenomes. **(B)** Network of B2ab CR haplotypes obtained from population surveys (listed in **Supplemental Table S2**). **(C)** Interpolated frequencies of B2aj mtDNAs in America. **(D)** Main migration routes of B2aj as inferred from phylogeographic data. See the legend of **Figure 1** for more details on the phylogenies of panels A and B.

Figure 3. (A) Maximum parsimony tree of B2y complete mitogenomes. The B2y Andean sub-lineages (B2y2 and B2y3) are highlighted in the dotted rectangle **(B)** Main migration routes of B2y as inferred from phylogeographic data. **(C)** Network of B2y CR haplotypes obtained from population surveys (listed in **Supplemental Table S2**) **(D)** Interpolated frequencies of B2y mtDNAs in America. See the legend of **Figure 1** for more details on the phylogenies of panels A and C.

Figure 4. (A) Maximum parsimony tree of Bb3, B2b11 and B2b14 complete mitogenomes. An asterisk as prefix indicates a position located in an overlapping region shared by two mtDNA genes. For the sake of clarity, the mitogenomes belonging to B2b (× B2b3, B2b11, B2b14) are simply indicated as rectangles in the left side of the tree without accounting for their phylogeny (listed in **Supplemental Table S11**); the colors of rectangles indicate the geographic origin (as shown in the legend) and the inner numbers correspond to the number of detected mitogenomes. **(B)** Phylogenetic network of B2b3, B2b11 and B2b14 CR haplotypes obtained from population surveys (listed in

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

Supplemental Table S2). **(C)** Main migration routes of B2b, B2b3, B2b11 and B2b14 as inferred from phylogeographic data. **(D)** Interpolated frequencies of B2b3 mtDNAs in America. **(E)** Interpolated frequencies of B2b14 mtDNAs in America. **(F)** Interpolated frequencies of B2b11 mtDNAs in America. See the legend of **Figure 1** for more details on the phylogenies of panels A and B.

Figure 5. **(A)** MDS based on IBS values obtained from different Native American populations from South America. **(B)** Interpolated F_{ST} values between Peruvian and other Native American samples. **(C)** Admixture analysis for the best cross validation value K = 2, showing the admixture components shared between different populations from South America; E-T = Equatorial-Tucanoan, C-P = Chibchan-Paezan. **(D)** *D*-statistics computed to formally test for admixture between West-Andean and East-Andean populations. **(E)** Time of the main admixture event occurring in the Diaguita between an ancestral component from Equatorial-Tucanoan (green) populations located to the East of the Andes, and Peruvian variation (red).

Figure 6. **(A)** Age in years (see colored legend inset) and location of archaeological sites in South America according to Goldberg et al. (2016). **(B)** Interpolated Kernel density map (units are expressed as the square decimal degree) from the archaeological data in Goldberg et al. (2016); dots indicate the sampling points of ancient specimens and the availability of mtDNA data. While red dots indicate mtDNA lineages other than those analyzed in the present study, other colors signal the presence and number of the B2 lineages analyzed here (see the legend inset).

Figure 7. **(A)** EBSP analysis of the mtDNA lineages explored in the present study. The figure shows the population growth experienced during the main pre-

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

historical periods of the South American settlement. The EBSP of B2y considers only mtDNAs observed in South America. **(B)** Demographic scenario for the peopling of South America as suggested by the results of the present study.

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

Supplemental Material

Supplemental Text.

Supplemental Figure S1. MDS of Native American samples from South America vs. other reference populations from Africa and Europe taken from 1000G.

Supplemental Figure S2. MDS of Native American samples from South America.

Supplemental Figure S3. Map of interpolated IBS values between Peruvians and other Native American samples from South America. Dots indicate sample geographical sources.

Supplemental Figure S4. EBSP for haplogroup B2b3 obtained from HVS-I data. Generation time (g) is expressed in years because calibration uses a fixed mutation rate (that is expressed in mutations/site/years).

Supplemental Table S1. Diversity indices for B2 sub-haplogroups in different geographic areas.

Supplemental Table S2. CR data (from the literature) of samples belonging to the B2 sub-haplogroups evaluated in the study. In addition, control region data from mitogenomes are also included.

Supplemental Table S3. Andean migration models and their probabilities. Θ = mutation scale population sizes; M = mutation scale immigration rates (sub-indexes are: P = Peru; B = Bolivia; A_NW = North-western Argentina; in brackets is the number of immigrants per generation or N_m).

Supplemental Table S4. Migration models from/to northern Argentina and their probabilities. Θ = mutation scale population sizes; M = mutation scale immigration rates (sub-indexes are B = Bolivia; A_NW = North-western

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

Argentina; A_NE = North-eastern Argentina; in brackets is the number of immigrants per generation or N_m).

Supplemental Table S5. List of complete genomes used to build the phylogenetic trees of **Figures 1** to **4**. Information on the geographic origin and ethnic affiliation of samples is included. References of the revised literature are included.

Supplemental Table S6. HVS-I sequence (from np 16024 to np 16569) and SNP data from Kolla (N = 45) and Diaguita (N = 24) of northern Argentina.

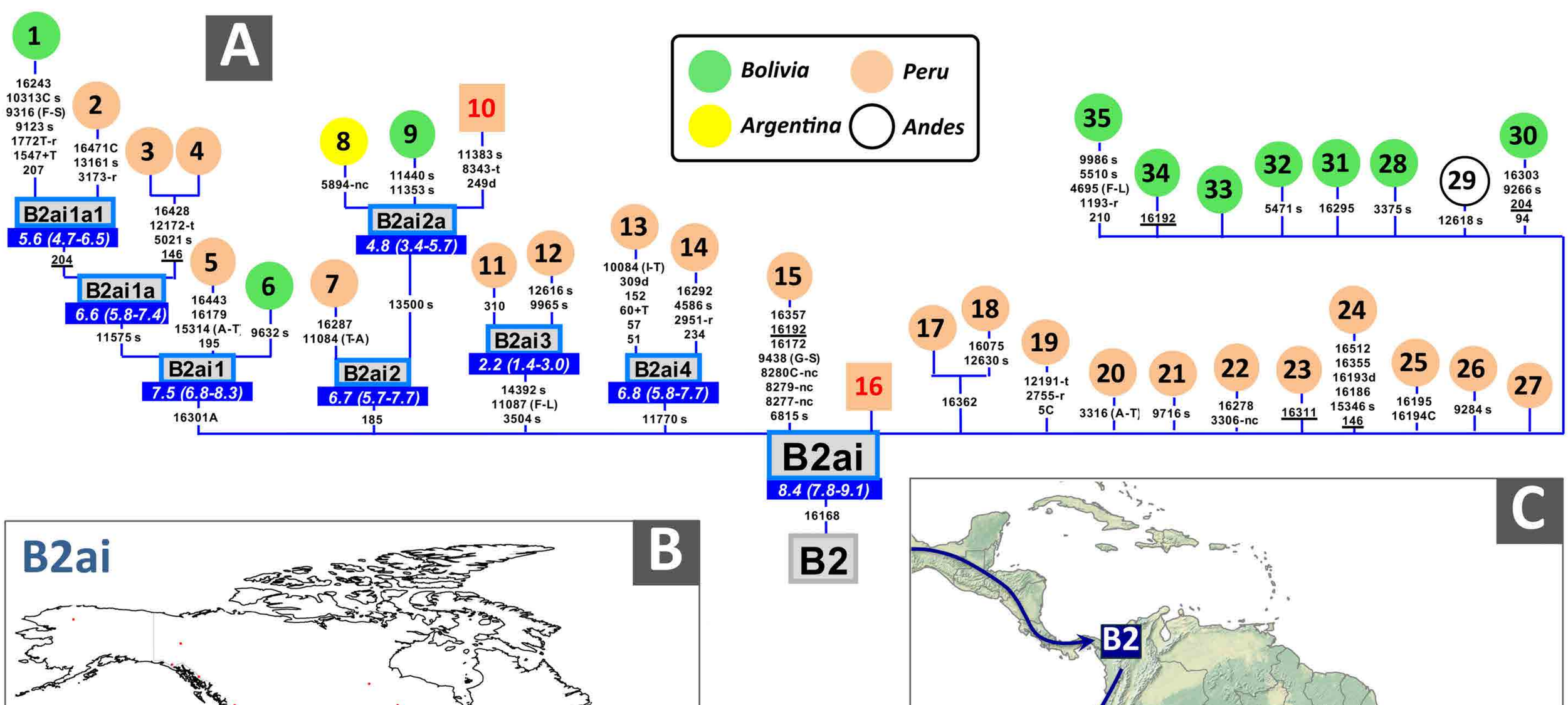
Supplemental Table S7. List of the 570 mitogenomes used, together with the 72 B2 mitogenomes newly obtained in this study, for the ML analysis.

Supplemental Table S8. ML TMRCA estimates for B2 clades and sub-clades.

Supplemental Table S9. Reference populations used for the analyses of autosomal SNPs.

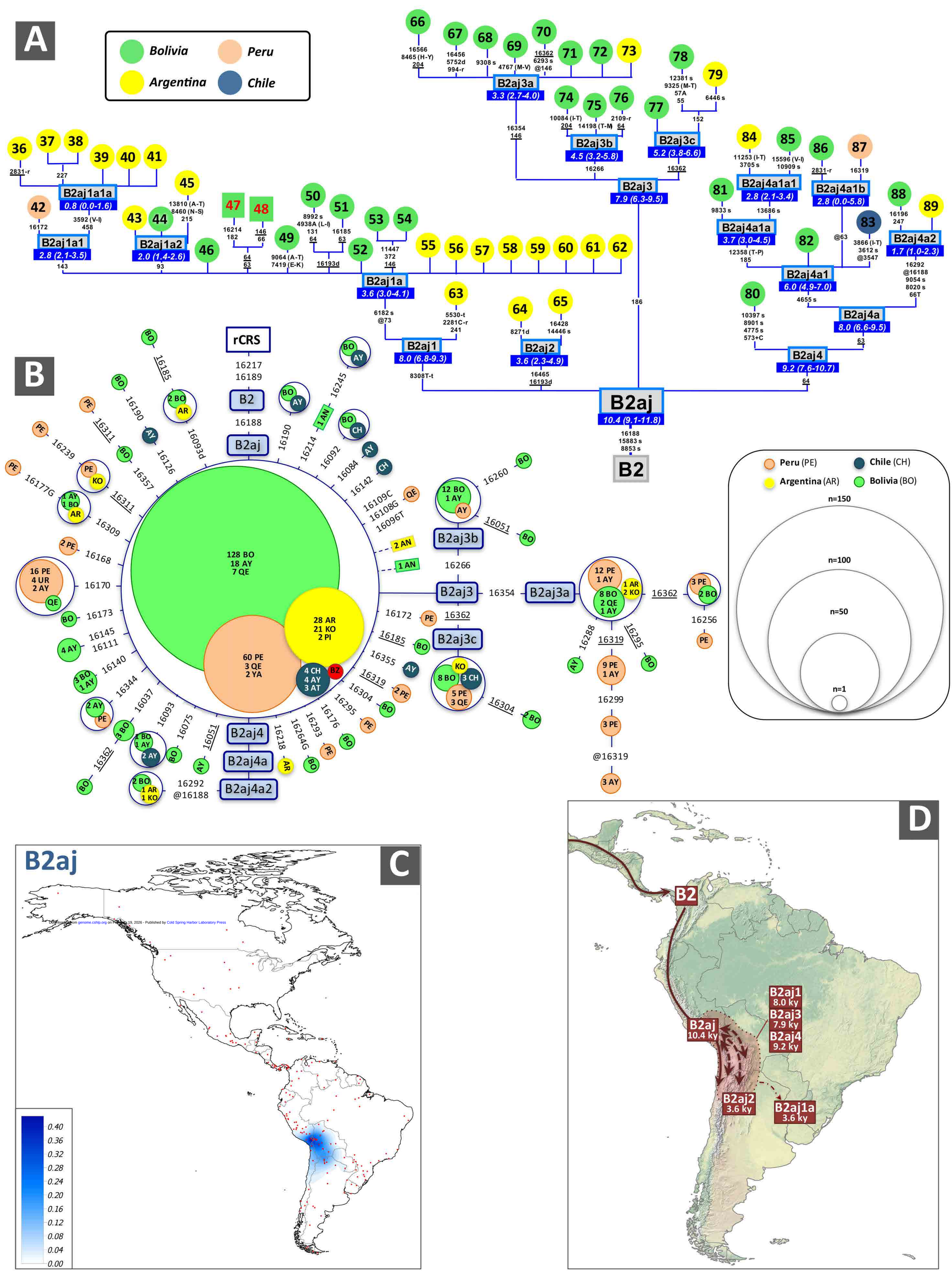
Supplemental Table S10. List of articles containing mtDNA sequence data used in the present paper (e.g. interpolated frequency maps of Figures 1 to 4).

Supplemental Table S11. List of B2b mitogenomes assessed in this study (other than the B2b3, B2b11 and B2b14 mitogenomes shown in panel A of **Figure 4**).



Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press



A

Andes
Brazil
Argentina
USA
Peru
Mexico
Unknown

Andes

99
16168
5442 (F-L)
1627A-t
1555-r
207
B2y2a1a
6.6 (5.3-7.9)

100
16335
14659 s
310
@75

102
8278A-nc
5460 (A-T)
3593 (V-A)
73C

103
16193d
12690 s
10306C (N-T)
7999 s
6270 (G-stp)
5164 (R-H)
326C
318-323d
316C
310
146

104
8989 (A-T)

105
106
15483 (S-L)
107
B2y2a2a
0.7 (0.4-1.1)
8691 s

16212
10595 s
1719-r
65G
101
B2y2a1
10.2 (8.9-11.6)
75

B2y2a2
4.4 (3.1-5.7)
16362
6839 s
3645 s

B2y2a
12.6 (11.3-13.9)
16319

98
16311
15454 s
10192 (S-F)
4952 s
1260-r
195
103T
65+CG
57-58d

96
1949-r
97
14581 s
5899+C
3907 (A-T)
234
B2y3
8.6 (6.5-10.7)
16168

B2y2
13.5 (12.2-14.9)
709-r

90
16278
16124
11218 s
146

91
16284
12771 s
12390 s
11125 s
10101 s
6471 (I-V)
152

92
59

93
16145
15244 s
8962 (T-A)
146

94
@16217
6221 s
146
95
16325
16311
9563 s
195
B2y1
10.5 (8.5-12.4)
3480 s

B2y
14.3 (13.0-15.7)
16261

B2

B

B2y
14.3 ky
B2y
B2y2
13.5 ky
B2y2a
12.6 ky
B2y3
8.6 ky
B2y
B2y2 B2y2a
B2y2a1
10.2 ky
B2y2a1a
6.6 ky
B2y2a
B2y2a2
4.4 ky
B2y2a1a
B2y2a
B2y2a1a
B2y

C

Peru
Argentina
Brazil
Chile
French Guiana
Bolivia
USA
Mexico

n=30
n=15
n=1

rCRS
16217
16189

B2
16261

B2y
2 AN
16187+C
5 TO
1 HU
1 AO
16325
US
2 AR
PE
BO
16311
16247
AR
2 AO
1 TO
16352
5 GA
16284
PE
16145
TO
16187
2 AO
10 JE
1 TO
5 ZU
2 PA
1 QE
2 AY
1 PE
4 BO
1 QE
ZU
B2y1
16124 16278
ME
16342
QE
16197
16223
16227
KI
16168
B2y3
1 QE
3 PE
16186
TO
16142
AR
@16189
B2y2
3 TO
1 AO
1 PP
1 AN
16319
B2y2a
7 AN
2 BO
2 AY
2 PI
2 YA
1 PE
16266
BO
16299
1 AN
@16189
TO
16254
AY
16362
B2y2a2
AR
3 PE
1 QE
BO
B2y2a1a
16212
16362
2 YA
16168
PE
16343T
YA
16138C
2 YA
16270
AY
16327A
YA
16242G
YA
16129
16266
BO
2 YA
16176A
YA
16335
15 CR
3 PI
1 AR
19 YA
1 PE
3 CK
2 BO
AR
Andes
Andes
North America

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

D

B2y

0.110
0.100
0.090
0.080
0.070
0.060
0.050
0.040
0.030
0.020
0.010
0.000

A

Bolivia
Argentina
Brazil
USA
Peru
Chile
Uruguay
Puerto Rico
Ecuador
Venezuela
Colombia
Mexico

N=7
N=1
N=1
N=2
N=3
N=3
N=3
N=4
N=64
N=65

Andes

110 111
16422
5580-nc
109
B2b14a1
1.6 (0.0-5.0)
108
9938 s
293
B2b14a
4.1 (0.9-7.3)
15344 s
14221 s
6791 s
112
252
152
B2b14b
8.2 (6.2-10.1)
16274
56
55
113
15994-t
B2b14b1
3.5 (2.5-4.5)
16362
3666 s
204
114
16278
13014 s
4914 s
B2b14
12.8 (11.0-14.6)
16319
16176

Andes

116
15954
B2b11b1a
1.9 (1.1-2.8)
117
@4013
115
2224-r
16249
1735-r
318
210
B2b11b1
8.3 (6.5-10.2)
239
B2b11b
11.3 (9.4-13.1)
15080 (T-A)
118
11807
10653 (A-T)
3666 s
173
119
16316
8095 s
6620 s
5581-nc
522
59
57
120
62T
55+T
121
16353
14182 s
11252 (I-V)
122
16438
14059 (I-V)
12127 s
11016 (S-N)
8994 s
243
B2b11
14.7 (13.5-15.9)
4013 (T-I)
123
B2b11a2
2.8 (2.0-3.7)
14634 (M-V)
13708 (A-T)
131
124
B2b11a2a
1.4 (0.9-1.9)
14873A (L-I)
125
1709-r
B2b11a
13.3 (12.0-14.5)
15784 s
126
16222
16145
14034 s
6026 s
1007-r
198
B2b11a1
10.8 (9.5-12.2)
16516
60
127
@60
128
129
130
16224
5417 s
B2b11a1a1
3.4 (2.0-4.9)
131
9084 s
8673 s
709-r
B2b11a1a
9.1 (7.7-10.5)
16124
16184
15313 s
3915 s
195

Atlantic side

Caribbean

132
13780 (I-V)
64
63
133
@455
134
16188
326C
318-323d
316C
310
135
9531 (T-A)
452d
136
463+CC
137
138
462-463d
139
16188
140
16193d
472C
471
470C
466
464C
141
16193d
463d
453
142
13634 (G-D)
B2b3a1a1
4.0 (3.3-4.8)
454
B2b3a1a
6.0 (5.1-6.8)
16092
B2b3a1
6.9 (5.5-8.2)
4012 (T-A)
460
455
143
@5899+C
144
10604 s
1719-r
@16344
145
4695 (F-L)
B2b3a2
3.9 (3.3-4.8)
16168
* 8552 (F-S)/(S-P)
5899+C
B2b3a
8.3 (7.0-9.5)
16344
16312
16249
15784 s
4232 (I-T)
3918 s
271
152
146
16497
15106 s
8251 s
195
B2b3
15.8 (14.9-16.9)
13708 (A-T)

B2b
16.6 (15.6-17.5)
6755
B2

B

Atlantic side

Andes

PE
16103
6 PE
B2b11a1a
16124
3 PE
16224
B2b11a1a1
B2b11
16249
B2b11b1a
PE
rCRS
16217
16189
B2
B2b
B2b3
16344
16312
16249
UR
B2b3a
16168
B2b3a2
@16344
AR
16215
5 BZ
3 KA
GU
AR
3 BZ
BZ
16266
BZ
16292
BZ
16261
13 BZ
3 KA
AR
1 AN
16093
1 BZ
2 KA
BZ
16166
16178
BZ
16095
3 KA
16229A
BZ
16179
BZ
@16249
12 RA

Andes

4 AN
5 BO
4 QE
2AY
KO
4 PE
2 QE
1 CP
16319
16176
B2b14
16357
PE
16092
BO
16178
2 PE
16274
B2b14b
QE
16362
B2b14b1
1 BO
1 AY
1 PE
1 QE
16278
BO
16093
1 PE
1 QE
AR
@16362
QE

16092
B2b3a1a
22 PU
DO
CU
3 VZ
3 US
16188
PU

Caribbean

Peru
Argentina
Brazil
Puerto Rico
Uruguay
Venezuela
Cuba
USA
Bolivia
Dominican Republic

n=20
n=10
n=5
n=1

C

B2b
16.6ky
B2b3a1a
B2b3
15.8 ky
B2b11
14.7 ky
B2b3a1a
6.0 ky
B2b3a
8.3 ky
B2b11a
13.3 ky
B2b11b
11.3 ky
B2b14
B2b14b
8.2 ky
B2b14
12.8 ky
B2b3a1
6.9 ky
B2b3a2
3.9 ky
B2b14b1
3.5 ky
B2b11b1
8.3 ky
B2b11a1a
1.9 ky
B2b14a
4.1 ky
B2b11a2
2.8 ky
B2b11a1
10.8 ky
B2b14a1
1.6 ky
B2b11a2a
1.4 ky
B2b11a1a
9.1 ky
B2b11a1a1
3.4 ky
B2b14
B2b14b

D

B2b3

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

0.40
0.36
0.32
0.28
0.24
0.20
0.16
0.12
0.08
0.04
0.00

E

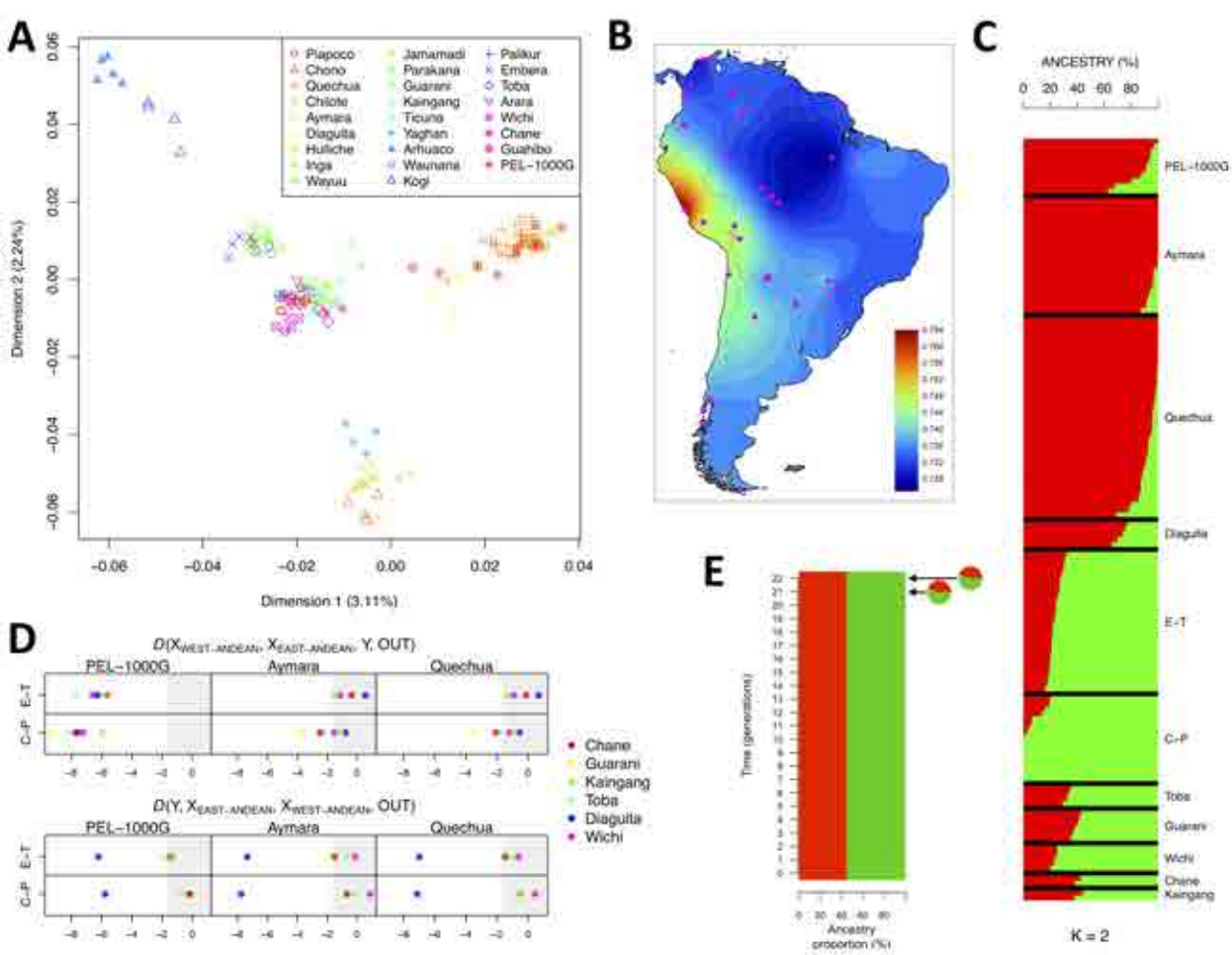
B2b14

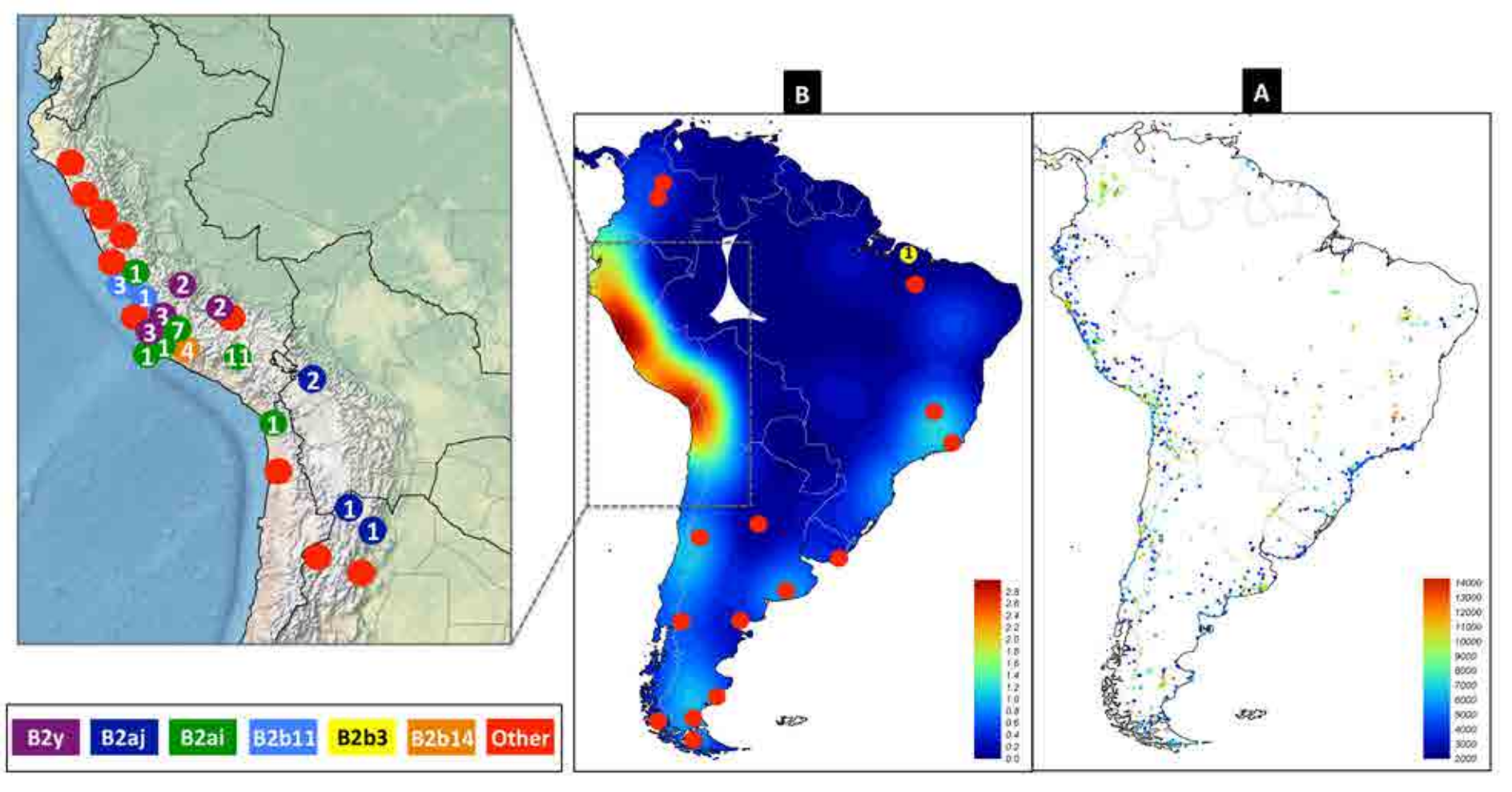
0.018
0.016
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000

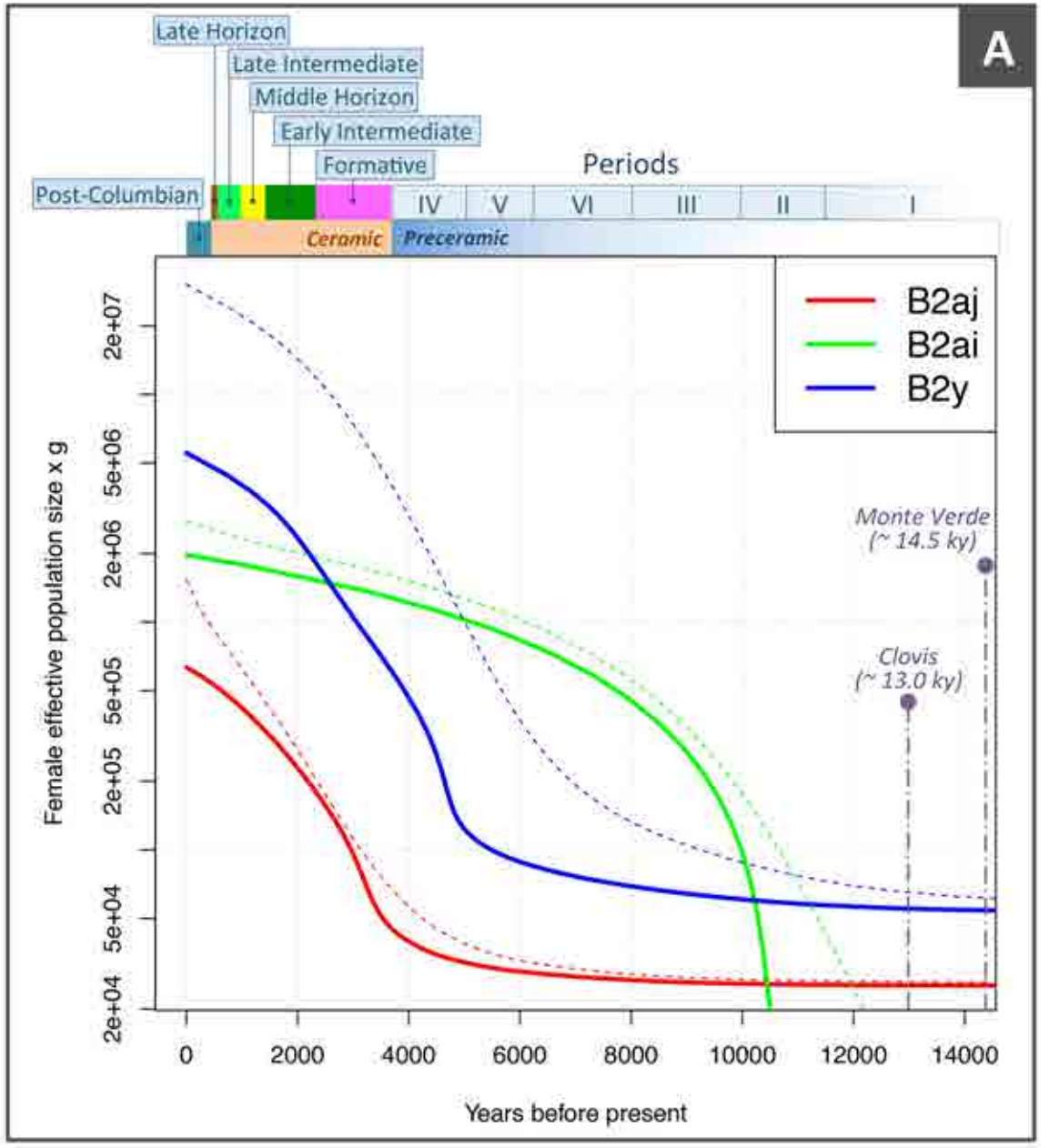
F

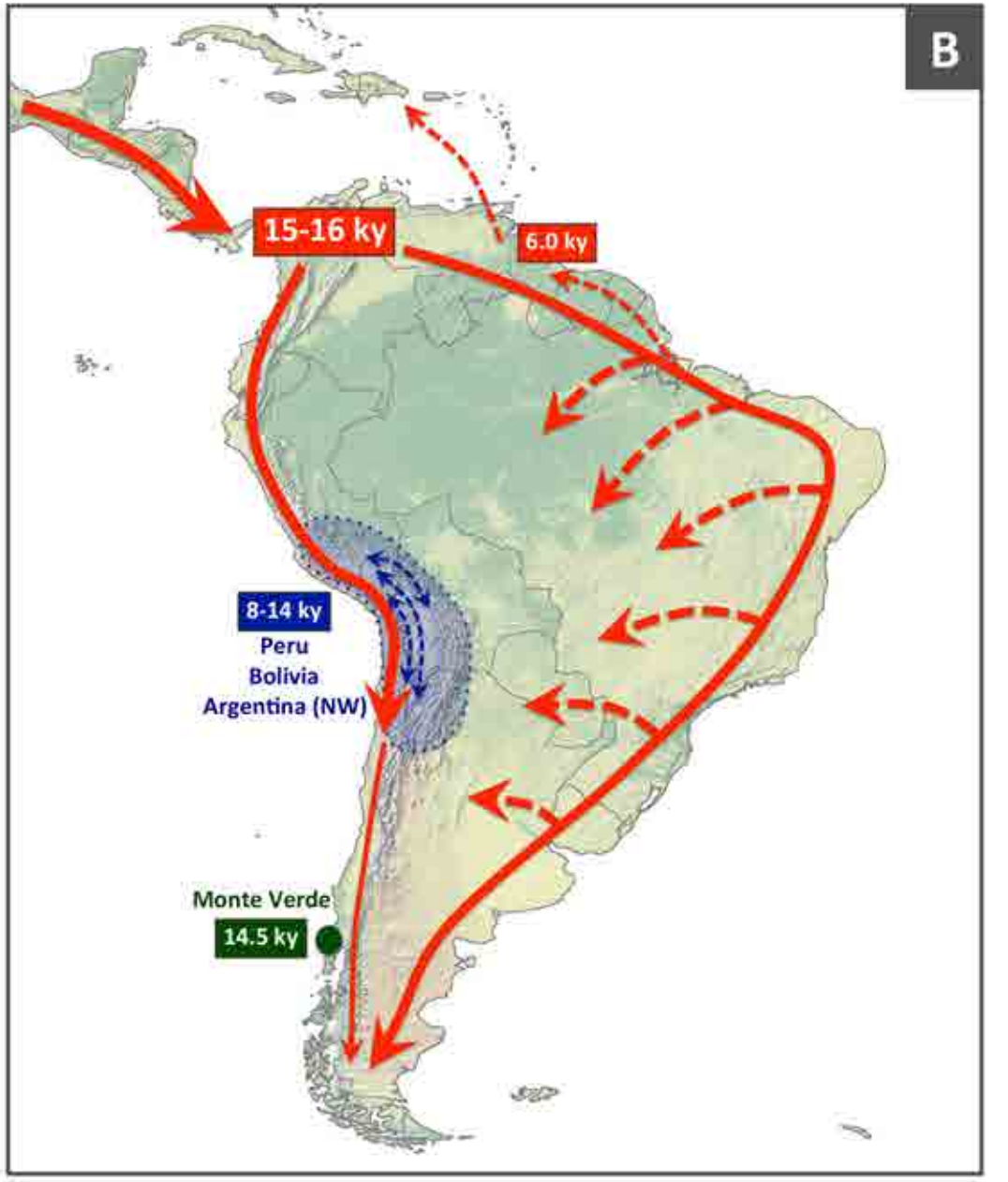
B2b11

0.036
0.032
0.028
0.024
0.020
0.016
0.012
0.008
0.004
0.000


A
Piapoco
Chono
Quechua
Chilote
Aymara
Diaguita
Huilliche
Inga
Wayuu
Jamamadi
Parakana
Guarani
Kaingang
Ticuna
Yaghan
Arhuaco
Wauyana
Kogi
Palikur
Embera
Toba
Arara
Wichi
Chane
Guahibo
PEL−1000G
Dimension 2 (2.24%)
Dimension 1 (3.11%)
B
C
ANCESTRY (%)
0 20 40 60 80
PEL−1000G
Aymara
Quechua
Diaguita
E−T
C−P
Toba
Guarani
Wichi
Chane
Kaingang
K = 2
D
D(X_WEST−ANDEAN, X_EAST−ANDEAN, Y, OUT)
D(Y, X_EAST−ANDEAN, X_WEST−ANDEAN, OUT)
PEL−1000G
Aymara
Quechua
E−T
C−P
Chane
Guarani
Kaingang
Toba
Diaguita
Wichi
E
Time (generations)
Ancestry proportion (%)


B
A
B2y
B2aj
B2ai
B2b11
B2b3
B2b14
Other


A
Late Horizon
Late Intermediate
Middle Horizon
Early Intermediate
Formative
Periods
Post-Columbian
IV
V
VI
III
II
I
Ceramic
Preceramic
B2aj
B2ai
B2y
Monte Verde
(~ 14.5 ky)
Clovis
(~ 13.0 ky)
Female effective population size x g
2e+07
5e+06
2e+06
5e+05
2e+05
5e+04
2e+04
0
2000
4000
6000
8000
10000
12000
14000
Years before present


B
15-16 ky
6.0 ky
8-14 ky
Peru
Bolivia
Argentina (NW)
Monte Verde
14.5 ky

Downloaded from genome.cshlp.org on February 19, 2026 - Published by Cold Spring Harbor Laboratory Press

The peopling of South America and the trans-Andean gene flow of the first settlers

Alberto Gomez-Carballa, Jacobo Pardo-Seco, Stefania Brandini, et al.

Genome Res. published online May 7, 2018
Access the most recent version at doi:10.1101/gr.234674.118

Supplemental Material http://genome.cshlp.org/content/suppl/2018/05/16/gr.234674.118.DC1

P<P Published online May 7, 2018 in advance of the print journal.

Accepted Manuscript Peer-reviewed and accepted for publication but not copyedited or typeset; accepted manuscript is likely to differ from the final, published version.

Creative Commons License This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see http://genome.cshlp.org/site/misc/terms.xhtml). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**

To subscribe to *Genome Research* go to:
https://genome.cshlp.org/subscriptions

Published by Cold Spring Harbor Laboratory Press