www.sciencemag.org/cgi/content/full/science.aav2621/DC1

Supplementary Materials for

Early human dispersals within the Americas

J. Víctor Moreno-Mayar*, Lasse Vinner*, Peter de Barros Damgaard*, Constanza de la Fuente*, Jeffrey Chan*, Jeffrey P. Spence*, Morten E. Allentoft, Tharsika Vimala, Fernando Racimo, Thomaz Pinotti, Simon Rasmussen, Ashot Margaryan, Miren Iraeta Orbegozo, Dorothea Mylopotamitaki, Matthew Wooller, Clement Bataille, Lorena Becerra-Valdivia, David Chivall, Daniel Comeskey, Thibaut Devièse, Donald K. Grayson, Len George, Harold Harry, Verner Alexandersen, Charlotte Primeau, Jon Erlandson, Claudia Rodrigues-Carvalho, Silvia Reis, Murilo Q. R. Bastos, Jerome Cybulski, Carlos Vullo, Flavia Morello, Miguel Vilar, Spencer Wells, Kristian Gregersen, Kasper Lykke Hansen, Niels Lynnerup, Marta Mirazón Lahr, Kurt Kjær, André Strauss, Marta Alfonso-Durruty, Antonio Salas, Hannes Schroeder, Thomas Higham, Ripan S. Malhi, Jeffrey T. Rasic, Luiz Souza, Fabricio R. Santos, Anna-Sapfo Malaspinas, Martin Sikora, Rasmus Nielsen, Yun S. Song†, David J. Meltzer†, Eske Willerslev†

*These authors contributed equally to this work.
†Corresponding author. Email: ewillerslev@snm.ku.dk (E.W.); dmeltzer@smu.edu (D.J.M.); yss@berkeley.edu (Y.S.S.)

Published 8 November 2018 on *Science* First Release
DOI: 10.1126/science.aav2621

This PDF file includes:

Other Supplementary Material for this manuscript includes the following:
(available at www.sciencemag.org/cgi/content/full/science.aav2621/DC1)

Materials and Methods

Samples

The samples analyzed in this study come from individuals recovered from the following sites: Trail Creek Cave 2, Alaska; Big Bar Lake, British Columbia; Lovelock and Spirit caves, Nevada; Sumidouro Cave, Brazil; Ayayema and Punta Santa Ana, Patagonia; and Aconcagua, Argentina. The ages of the specimens range from ~0.5-10.9 ka, and are located throughout the Americas. As might be expected of sites over this expanse of time and space, none display cultural similarities. Hence, there is no a priori evidence to link the specimens.

The **Trail Creek** Caves site (BEN-001) is located on the Seward Peninsula of Alaska, ~25 km south-southwest of Deering, Alaska (approximate coordinates N65.79 W163.41), and is comprised of ~12 caves in a limestone ridge that rises in places is ~25-100 meters above Trail Creek (*76*). The caves were discovered in 1928 by Deering residents Taylor Moto and Alfred Karmun, and relocated and tested in 1948 by geologist David Hopkins. Caves 2 and 9 were then more thoroughly excavated by archaeologist Helge Larsen in 1949-1950 (*76*). The human tooth analyzed in this report was recovered from Cave 2 in 1949.

The human tooth came from Layer III at the 4 m section of the excavation, and was not identified as human until long after Larsen's work (*77*). Pasda identified the tooth as a left I1 (upper left permanent central incisor), slightly worn although the root was not fully formed. The incisal edge of the crown is slightly worn (as seen from the absent mammelons) and the root formation not completed. Root formation for the maxillary central incisor, (di1), is completed at about 2.5 years of age (Table 4 in (*78*)). Due to the root formation not being completed, rather than the root being re-absorbed prior to being shedded, it can be presumed the child died while the tooth was in situ. Therefore, this child can be presumed to have died at ~1.5 years of age (Figure 6 in (*78*), Table 2 in (*79*)). The sex of the child cannot be positively assessed from the size or shape of the tooth. The human tooth was directly dated to 8085 ± 40 radiocarbon years BP (OxA-37281), yielding a calibrated age of 9020 BP (Table S1).

Big Bar Lake is located on the Fraser Plateau of central British Columbia, ~30 km northwest of the town of Clinton, British Columbia, and within the drainage catchment of the Fraser River, which is located west of Big Bar Lake at a distance of ~25 km. The Big Bar Lake archaeological site (**EhRk-4**), from which the human remains for this analysis derive, is located on the southeast corner of the lake (approximate coordinates N51.3 W121.77). The site was first recorded during a 1970 archaeological survey, though the location where the human remains were found – some 10 m southeast of the 1970 survey area – was not recorded at the time. The human remains were first reported in 2002 by First Nations members of the Canoe Creek Band, who found them eroding out of an interment atop a prominent knoll. Excavation of the human remains took place in 2003, with First Nations community members present throughout the work. The bones were found in what appeared to be a shallow pit, were oriented roughly north-south, and the individual was interred resting supinely, facing west toward the lake. Although some skeletal elements were missing, the skeleton was more or less complete and in anatomical position. No cultural objects or grave goods were found with the skeleton (*80*).

The interred individual is a female who died at roughly 60 years of age (see also (*81*)). The skeleton exhibited no signs of unusual disease or trauma, but displayed some degenerative changes associated with her age, including signs of osteoporosis and pre-mortem tooth loss (*80*). The Big Bar Lake individual harbored mtDNA haplogroup **A**, as might be expected of an ancestral Native American. In this respect, this individual differed from the two individuals recovered from the roughly contemporaneous, nearby China Lake site (the two sites are separated by just ~25 km (*80*)), which could only be identified to mtDNA superhaplogroup **M** but excluding haplogroups C and D, a lineage common in East Asia, but otherwise unknown in the Americas (*81*).

Radiocarbon analysis of rib fragments and the left fibula from Big Bar Lake initially produced ages several centuries apart. A re-analysis of the original samples yielded a composite age estimate of 4940 ± 40 radiocarbon years BP, or 5.6 ka for the individual (Table S2). Isotopic analyses of both carbon (Table S2) and $\delta^{15}N$ (+16.2‰ and +16.‰), indicate that the individual had a diet comprised of both terrestrial and marine resources, with the former comprising the bulk of the diet (*80*). In terms of the regional culture history, Cybulski et al. place the individual in "… the transitional period between the Lehman Phase and the Lochnore Phase" (*80*). The earlier Lehman phase groups are an interior group reliant on terrestrial hunting, the later Lochnore phase is marked by heavier reliance on riverine and coastal resources, the latter possibly "a movement of riverine-adapted Salishan peoples up the major drainages to exploit salmon resources" (*80*); see also (*82*).

Spirit Cave is a small rockshelter located in Churchill County, in west-central Nevada (approximate coordinates N37.41 W122.08), above the dry lakebed of Pleistocene Lake Lahontan. Spirit Cave was first investigated by Sydney and Georgia Wheeler in 1940 (*83–85*). The Wheelers recovered in the shelter's shallow, sandy deposits the remains of two bundle burials wrapped in tule reed woven mats (*83*, *86*). The burials lay one on top of another. The upper burial (Burial 1 of the Wheelers; NV State Museum number Ahur 770, and Spirit Cave 2 in this analysis) contained a relatively few disarticulated bones of an adult female. The skeleton of the lower individual (Burial 2 of the Wheelers; NV State Museum number Ahur 2064, and Spirit Cave in this analysis) was virtually complete and partially mummified, and included hair, leather moccasins, a rabbit skin blanket, and, in the intestines, the remains of the last meal eaten. This individual became known as the Spirit Cave mummy (*83*). The mummy was an adult female, the mummy an adult male, perhaps 45 ± 5 years of age at death (*87–89*).

Radiocarbon dating in the 1990s of hair from the Spirit Cave mummy along with fragments of the mat in which the individual was buried yielded ages that ranged from 9350-9460 ^{14}C years BP; dating of the burial atop it returned an age of 9300 ^{14}C years BP. Given the technical advances in radiocarbon dating since those ages were obtained, and to insure precise age control on the samples submitted for genetic analysis, additional dating and isotopic analysis was on three specimens, the results of which are provided in Table S3. The newly-obtained radiocarbon ages are slightly older (by ~200-300 radiocarbon years) than those obtained in the 1990s, but otherwise are broadly consistent, and affirm the Early Holocene age of the Spirit Cave mummy and Spirit Cave 2.

Comparison of Spirit Cave's crania with those in a worldwide data base of craniometric measurements placed Spirit Cave outside the range of any modern population, but those to which he was closest included Europeans (Norse and Zalavar) and in particular the Ainu (*89*, *90*). Importantly, it was not very similar in its cranial form to any modern Native Americans (*89*, *91*). Spirit Cave proved to be one of a number of ancient individuals – including Kennewick and Lagoa Santa individuals – whose craniometric attributes appeared to be sufficiently distinct to warrant the label of 'Paleoamerican', and possibly represent a population who arrived in a migration to the Americas separate from that of later ancestral Native Americans (*90*, *92*, *93*).

Lovelock Cave (NV-Ch-18) is located in a limestone formation in Churchill County, in west-central Nevada (approximate coordinates N39.96 W118.56). It is situated on the edge of Humboldt Lake, an intermittent lake that occupies a portion of the floor of Pleistocene Lake Lahontan, which helped form the cave itself (*94*).

Archaeological excavations and salvage efforts were conducted over several months by L. L. Loud in 1912; over a like span by M. R. Harrington and Loud in 1929; and intermittently in the 1960s by teams led by Robert Heizer (*94*). The site yielded an abundance of archaeological remains (estimated at >20,000 specimens (*94*)). Overall, it is estimated that 45 humans were interred in the cave.

Of the human remains from the site, five individuals we provided for genomic analysis by the Fallon band of the Paiute-Shoshone. All five individuals were adults and, in advance of the genomic analysis were directly radiocarbon dated. All are Late Holocene in age (Table S4). Given their ages, genomes of these individuals can potentially inform on the debated 'Numic Expansion' hypothesis.

The **Lagoa Santa** specimens examined in our analyses are from Sumidouro Cave (approximate coordinates S19.54 W43.94, elevation 660 m asl), located in the Lagoa Santa region of east-central Brazil (Minas Gerais state). The 30 human skeletal remains from this site are the largest known collection of ancient material from a single site in this region, a region that has produced a large number of ancient human remains (*38*, *95*). These were collected by the Danish naturalist Peter W. Lund, who between 1835 and 1845 explored ~800 caves in this karstic limestone region (*96–98*), of which six yielded human remains (*98*), including Sumidouro and the cave of Lapa Vermelha. The great majority of Lund's collection is housed at the Zoology Museum of the University of Copenhagen, the source of the samples analyzed for this study, and for which direct radiocarbon dates are shown in Table S5.

The Lagoa Santa skeletal remains, especially the remains from Sumidouro Cave, were subject of considerable discussion and debate over the century that followed their discovery (as detailed in (*99*). The crania from Lagoa Santa and Australo-Melanesia were characterized as presenting narrow and long neurocrania, prognathic low faces with relatively low and broad orbits and noses. The crania of modern-day Native American and east Asian population presented the opposite pattern. The two morphological patterns were later referred to as 'Paleoamerican' and 'Amerind'. However, these terms can be misleading as they imply associations between morphology, chronology and geography that should not be taken for granted (chronological and geographically neutral alternatives are the terms 'generalized' (i.e. 'Paleoamerican') and 'derived' (i.e. 'Amerindian') morphology.

The two **ancient Patagonian** samples included in this study were obtained from human remains currently housed at Instituto de la Patagonia, Universidad de Magallanes, Chile. The remains were recovered in two archaeological sites dated to the Middle Holocene (c. 8.000-4.000 BP): **Punta Santa Ana**-1 (PSA-1) and **Ayayema** Cave (*100–102*). Available radiocarbon AMS dates were calibrated with CALIB 7.0.4 (*103*, *104*).

PSA-1 is a multicomponent shell-midden located by the shore of the Strait of Magellan (Fig. 1). The human remains have been dated to 6,540 - 6,290 BP (cal. 7,001 - 7,584BP) (*103*, *104*), which are in agreement with other dates obtained for the site (*102*). However, isotopic analyses (*102*) indicate that this individual's diet was rich in marine resources so the dates might have been influenced by a reservoir effect. Bioanthropological analyses of cranial and pelvic traits in this individual (Id# 5832) indicate that the remains correspond to an adult female (35-50 years of age at the time of death).

Cueva Ayayema is located in Isla Madre de Dios, in the central region of the archipelago area. The osteological remains of this individual (Id# 85926) were found on the surface of the cave (*101*), and although they are very incomplete their condition is good. The remains have been dated to 4,520±60BP (cal. 5308 - 4,933BP). As in the case of PSA-1, it is possible that the radiometric dates have been influenced by the reservoir effect. Macroscopic analyses of bioanthropological markers show the remains correspond to an adult male (35-50 years of age at the time of death). Thus, both individuals examined in this study correspond to early marine hunter-gatherer settlers in Fuego-Patagonia.

A frozen mummy was found in Cerro **Aconcagua** (Mendoza province, Argentina) in 1985 by a group of mountaineers. The mummy was only partially unearthed at an altitude of 5,300 meters, at the base of the Pirámide mountain of the Aconcagua. The mummy is a well-preserved body of a seven-year-old boy wrapped in 18 textile pieces and surrounded by anthropomorphic and other small statuettes (*105*, *106*). The scenario resembled the typical Inca ritual burial known as "capacocha" (or "capac hucha"). The age of the mummy was determined to be ~500 years ago, coinciding with the time of maximum expansion of the Inca Civilization towards the southern cone. There exist archaeological indicating that the Aconcagua boy could have been brought from the Central Coast of Peru (*106*).

Two genetic studies on uniparental DNA markers were carried out on the Aconcagua mummy. The mummy's mtDNA belongs to the rare sub-haplogroup C1b*i,* which has been proposed to have a Peruvian Inca origin (*14*). In agreement with anthropological documentation and the mtDNA study, the mummy's Y-chromosome profile was found to be also of most likely Central Peruvian origin (*107*). It shows close genetic affinity to the Choppca, a Quechuan speaking Peruvian population.

The ancient **Andamanese** genome reported in this study was obtained from DNA extracted from a tooth from a skull collected by Alfred Radcliffe-Brown (*108*) from among the Great Andamanese in the early 20th Century and today held as part of the Human Osteological Collections of the Duckworth Laboratory, University of Cambridge.

Community engagement and permissions for genomic analysis

As noted in the main text, we have wherever possible engaged and sought feedback from Indigenous groups in North America linked to the ancestral individuals analyzed in this study using the recommendations for genomics research with Indigenous communities (*11*, *12*). In South America, permissions from the responsible authorities were obtained and, when possible, we engaged with Indigenous communities. Permissions and engagements where possible are as follows:

Trail Creek: A letter expressing the support from the Native Village of Deering for the genetic analysis of these remains is included in Fig. S1.

Big Bar: The human remains from this locality were first reported in 2002 by First Nations members of the Stswe´cemc/Xgat'temc band, who found them eroding out of an interment atop a prominent knoll. Excavation of the human remains took place in 2003, with First Nations community members participating and present throughout the work. Authors RSM and JSC visited the community in 2005 to discuss DNA analysis of the Big Bar ancestor and present-day community members. The community agreed to a partnership on the project. The results were discussed with the community and disseminated to the public in four publications (*18*, *80*, *81*, *109*). In July of 2016, RSM met with Chief Patrick Harry and some Council members of Stswe´cemc/Xgat'temc First Nation to discuss a follow up project on the genomic analysis on remnant tooth from the analysis that was published in (*81*). On August 2nd 2016 RSM discussed the details of the project with Chief Harry and Council Members. Initial genomic analysis was completed in 2017, but RSM had to reschedule plans to visit the community to discuss the results due to forest fires in the area. RSM was able to visit the Stswe´cemc/Xgat'temc First Nation on July 19, 2018 to discuss the genomic results of the Big Bar analysis. Council member Harold Harry (representative of the Stswecem'c/Xgat'tem band) is a co-author on the paper and worked with RSM reviewing the manuscript for publication.

Spirit Cave: In 1997, the Fallon Paiute-Shoshone Tribe headquartered in Fallon, Nevada near Spirit Cave, in 1997 claimed affiliation with the Spirit Cave remains and requested repatriation (*84*, *87*). As the remains were recovered from federal land under the auspices of the Bureau of Land Management (BLM), the BLM in the late 1990s reviewed the NAGPRA-relevant evidence, and concluded that while the preponderance of that evidence indicated the Spirit Cave remains were Native American, they could not be affiliated with any modern individual, tribe, or other group and were therefore deemed culturally unidentified (*87*). The Fallon Paiute-Shoshone Tribe disputed that determination, and after an unsuccessful effort through the NAGPRA Review Committee to have the decision reversed, in 2004 filed a lawsuit against the BLM in U.S. District Court (*84*). The Court ruled in 2006 that the BLM had not fully considered the evidence of cultural affiliation, and directed the BLM to make a further attempt to resolve Spirit Cave's status under NAGPRA.

Little progress was made over the next half dozen years or so, but in early 2015 an agreement was reached between the BLM, the Fallon Paiute-Shoshone Tribe and the Centre for GeoGenetics to undertake a DNA analysis of the Spirit Cave remains (*84*). A letter signed by representatives of the Paiute Shoshone Tribe authorizing the analysis of these ancient human remains is included in Fig. S2. DNA analysis was conducted in

2015, and in January 2016, the results were provided to the BLM and the Fallon Paiute-Shoshone tribe. After a summary and synthesis of all the available evidence now bearing on the affiliation of the Spirit Cave remains, in October 2016 the BLM again concluded that the Spirit Cave individuals are Native American, and therefore subject to NAGPRA, but that the remains still could not be affiliated with any modern tribe (*110*). However, under the statute determining the 'Disposition of culturally unidentifiable human remains' (43 CFR §10.11), the Spirit Cave assemblage could be transferred to the Indian tribe or tribes recognized as aboriginal to the area in which they were found, and on those grounds the BLM determined that Spirit Cave was within the ancestral territory of the Northern Paiute, and the ancestors of the Fallon Paiute-Shoshone Tribe (*110*). On October 18, 2016, the BLM published their intent to transfer control of the remains to the Fallon Paiute-Shoshone Tribe (*111*). Only two inquiries – but no appeals of that determination – were filed during the requisite 30 day appeal period. Accordingly, a transfer of control document was signed in November of 2016, and soon thereafter all human remains and funerary objects from Spirit Cave were repatriated to the Fallon Paiute Shoshone Tribe (*84*, *112*). We note as well that Len George, chairman of the Fallon Paiute-Shoshone Tribe, is a co-author of this paper.

Lovelock Cave: During his visit to meet with the members of the Fallon Paiute-Shoshone Tribe to discuss the analysis of the Spirit Cave remains, co-author Eske Willerslev was also offered the opportunity to sample the remains of the Lovelock Cave individuals. These remains had for many years been curated at a local museum, but had been previously returned to the tribe. The tribe was interested in having those remains examined for DNA as well, and with their permission these were sampled on the occasion of that visit. The remaining parts after DNA testing were subsequently retuned to the tribe. Although the Lovelock Cave specimens had not been recovered from public lands and they were already owned and possessed by the tribe, the Fallon Paiute Shoshone Tribe agreed with Eske Willerslev to authorize the genetic analysis of the Lovelock remains. We note here too that Len George, chairman of the Fallon Paiute-Shoshone Tribe, is a co-author of this paper.

Lagoa Santa: The Lagoa Santa individuals are part of the Lund collection (made in the 1830s-1840s) that are housed at the Natural History Museum of Denmark, under the curatorial direction of by Kurt Kjær. There are no claims for these remains by indigenous groups in Brazil. These sites are not in the proximity of any indigenous lands or populations, and lack association with surviving cultures of today, or from recent pre-colonial times. However, we have already established contacts with other indigenous students and authorities of the region, setting the grounds for a future meeting.

Punta Santa Ana and Ayayema: As was the case with the Lagoa Santa remains, there is no group or groups claiming (or directly associated with) these remains. Permission for the analysis of the remains was provided to co-author Marta Alfonso-Durruty by Susana Simonetti de Groote, Executive Secretary of the National Monuments Council of Chile, as provided in Fig. S3.

Aconcagua: Permission for the sampling of this individual (Fig. S4) was signed by a representative of the 'Equipo Argentino de Antropología Forense,' who holds curatorial responsibility for the Aconcagua sample. The signee, Carlos Vullo, is also a co-author in the present study.

Andaman: The Andamanese genome reported in this study originates from a human tooth sampled with permission from Marta Mirazón Lahr, Director of the Duckworth Laboratory, LCHES, University of Cambridge. The remains belong to an individual from the Great Andamanese population, who traditionally retrieved some of the bones of deceased relatives after a period of exposure, but these were put aside. They were collected by Alfred Radcliffe-Brown, who lived with the Great Andamanese during his fieldwork in the islands in the early 1900, and who subsequently donated them to the University of Cambridge in 1909.

Radiocarbon dating

Three bone samples from Spirit Cave (P43584-P43586) and one tooth sample from Seward Peninsula (P44821) were dated at the Oxford Radiocarbon Accelerator Unit (ORAU; see Table S6 for results), using the methods noted below and further described in references (*113*, *114*). The quality of the extracted collagen and hydroxyproline was monitored by measuring carbon and nitrogen contents (%C and %N), and the atomic ratio of carbon to nitrogen (C:N). C:N values are deemed acceptable at 2.9-3.5 for collagen, and at ~5.0 for hydroxyproline (*113*, *114*).

Collagen dating (coded '**AF***'): Bone samples (P43584-P43586) were first washed with organic solvents (acetone, methanol and chloroform). Collagen was then extracted following decalcification in acid, a base wash, re-acidification, gelatinisation, and ultrafiltration. After freeze-drying, the collagen was combusted, graphitised, and dated by AMS.

Hydroxyproline dating (coded '**HYP**'): Collagen samples (P43584-P43586) were hydrolyzed and the underivatised amino acids separated using preparative High Performance Liquid Chromatography (HPLC). Hydroxyproline was isolated and later combusted, graphitised, and dated by AMS.

Carbonate dating (coded '**OO**'): The tooth sample (P44821) was shot blasted and crushed. It was then etched with 0.01 M HCl (20 mL, 1 hr, 4°C) before rinsing with ultrapure water (20 mL, 3 times). Once etched, the sample was digested with ortho-phosphoric acid and the evolved carbon dioxide collected, graphitised and dated by AMS. Because of likely incorporation of diagenetic carbonate into the tooth enamel and the difficulty in removing all of this diagenetic carbon, this determination should be considered as a minimum age.

We found that in the three cases where we could compare different pretreatment methods on bone the results were statistically indistinguishable. This supports the notion that contamination in bone collagen derived AMS dates is low.

Stable isotope values obtained show that the several Great Basin individuals had a significant reliance upon aquatic resources. The ultrafiltered collagen $\delta^{15}N$ values for the two older Spirit Cave samples are extremely enriched, consistent with the direct dietary evidence of fish remains. $\delta^{13}C$ values are also consistent with aquatic source uptake.

Laboratory procedures

All aDNA procedures for the **Trail Creek, Big Bar, Spirit Cave and Lovelock Cave** remains were performed in dedicated clean lab facilities at the Centre for GeoGenetics, Natural History Museum, University of Copenhagen, according to generally accepted recommendation (*115*, *116*). Dense portions of petrous bone surrounding the cochlea were isolated mechanically and crushed. Tooth crowns were separated from roots. The roots were superficially mechanically cleaned on outer and inner surfaces. Ancient DNA was extracted from crushed cementum-enriched tooth material or petrous bone (Table S7).

DNA extraction and preparation of libraries for Illumina sequencing followed protocols published previously (*1*, *117*, *118*), except extractions of the Trail Creek specimen, which were performed according to (*119*). Dually indexed sequencing libraries were prepared from double-stranded DNA fragments following the strategy originally described elsewhere (*120*), with the modifications presented in (*1*). For some libraries the DNA extracts were pre-treated with USER enzyme as described earlier (*1*) (Table S8). Purified DNA libraries were sequenced (80 bp single-read) on HiSeq4000 instruments at the National High-throughput Sequencing Centre, Natural history museum, Copenhagen, Denmark.

All sample preparation and treatment of the **Lagoa Santa remains** were undertaken in the dedicated clean-laboratory facilities of the Centre for GeoGenetics, University of Copenhagen. Approximately 50-460 mg of either endosteal petrous bone, which surrounds petrous bone canals, or of the most compact mesosteal tissue of the otic capsule was identified macroscopically by drilling thin slices of petrous bone along parasagittal planes. Each sample was carefully broken into smaller pieces by hand until the most compact material was identified. Less optimal material was also extracted in separate tubes. All samples were treated in sterile 5 mL Lo-Bind tubes.

Samples were exposed to a full 30 minutes of decontamination through pre-digestion in 3 mL EDTA (*117*), while being monitored due to the risk of total dissolution of the highly degraded mineralized tissue. Samples were subsequently completely dissolved for 24 hours at 37C in 4.9 mL EDTA and 100 uL proteinase K. DNA was extracted through in-solution silica adsorption, using optimized binding buffers for ultra-degraded DNA (*118*). Next-generation sequencing libraries following standardized ancient DNA protocols (*118*) were built directly on 20 uL DNA extract. Prior to adaptor ligation, the end-repaired DNA extract was purified in MinElute columns using 10X PB Buffer rather than custom binding buffer for ultra-short DNA, in order to slightly increase the average DNA molecule length.

DNA libraries were indexed through a two-step PCR approach described elsewhere (*118*) in which indexes were incorporated in a first amplification of 12 PCR cycles. 5 uL aliquots from the first amplifications were used for a second amplification using Illumina P5 and P7 primers for an additional 4-10 cycles. Libraries were QC on Agilent Bioanalyzer 2100 and sequenced on the Illumina HiSeq 2500.

DNA from two teeth from **Punta Santa Ana** and **Ayayema** was extracted following a modification of (*121*). Both samples were mechanically cleaned to minimize modern contamination with a diamond-coated disk using a drill. After removing the crown, the root was cut into smaller pieces and used for the DNA

extraction. A pre-digestion step was performed in a digestion buffer containing 1.5 mL 0.5M EDTA and 25 µL of Proteinase K (20 mg/mL) for 45 minutes at 48ºC with rotation. The samples were centrifuged and the supernatant removed. New EDTA was added (4 mL) and the samples were incubated at room temperature with rotation for 24 hours. Then Proteinase K was added to a concentration of 0.2 mg/ml and the samples incubated at 48ºC overnight on rotator. After this step, the samples were centrifuged and 3.5 mL of the supernatant was transferred to a 30 kDA centrifugal unit (Milipore), concentrating to 200-250 µL. A special buffer development by Dabney et al (2013) was used for the purification step following the publication guidelines for its preparation. A total of 13 volumes of the buffer were added to each sample and spins were done at 8,000 rpm using a silica column (MinElute PCR purification Kit, Qiagen). The washing step was done using the Qiagen Washing buffer and spin down at 13,000 rpm for one minute. In the elution step, the columns were incubated twice with 35 µL of Qiagen EB at 37º for 10 minutes.

Blunt-end libraries using Illumina-specific adapter and NEBNext DNA library Prep Kits (End repair module: E6050; Quick Ligation module: E6056; and Bst DNA Polymerase, Large Fragment: M0275) were built following the manufacture's instruction with the following modifications. The end-repair step was carried out in 25 µL reactions using 21.25 µL of DNA extract. This was incubated for 20 min at 12C and 15 min at 37C, and purified using PB buffer with QIAGEN MinElute spin columns, and eluted in 18 µl. Next, Illumina-specific adapters according to Meyer and Kircher 2010 were ligated to the end-repaired DNA in 25 µL reactions. The reaction was incubated for 15 min at 20C and purified with PB buffer in QIAGEN MinElute columns, before eluted in 25 µL EB Buffer. The adaptor fill-in reaction was performed in a final volume of 30 mL and incubated for 20 min at 65C followed by 20 min at 80C to inactivate the Bst enzyme. During all the elution steps, the columns were incubated with EB buffer at 37º for 10 minutes before centrifugation.

To assess the amount of DNA libraries in each sample and therefore the optimal number of PCR cycle for library amplification, qPCR was performed using SYBR green MIX (Roche) according to manufacturer's instructions and the same forward and reverse primers used for the index PCR step. The DNA library (10 µl) was then amplified and indexed in a 50 µL PCR reaction containing 0.2 µL of High Fidelity Platinum Taq (Invitrogen), 5 µL of 10x High Fidelity PCR buffer, 2 µL of 50 mM MgSO4, 2 µL of each primer (10 µM, inPE forward primer and indexed reverse primer) and 27.8 µL of water. Thermocycling conditions were 2 minutes at 94º followed by cycles (number according to qPCR values) of 30 seconds at 94º, 30 seconds at 60º and 30 seconds at 72º, and a final extension step of 5 minutes at 72º. The amplified library was purified using the QIAGEN MinElute columns following the manufacturer's instructions and eluted in 20 µL of EB buffer. Libraries were quantified on an Agilent Bioanalyzer 2100 using the High Sensitivity kit and submitted to the Danish National High-throughput DNA Sequencing Center for 80 bp, single read sequencing in an Illumina HiSeq 2500 system.

All the molecular work (pre-amplification) for the **Andaman Islander** genome was conducted in dedicated aDNA clean lab facilities at Centre for GeoGenetics, University of Copenhagen, using strict aDNA guidelines, *e.g.*, refs. (*115*, *116*).

To minimize potential contamination from modern DNA due to previous handling, the outermost surface of the tooth was first removed using the edge of a diamond-dust-coated cutting disk in a mechanical drill. The crown of the tooth was then separated from the root using another cutting disk. Three DNA extractions were prepared from the tooth. The first extraction (MA755) was conducted on dentine powder (250 mg) which was drilled from the core of the root and crown using a pointy drill-bit. The second extraction (MA756) was prepared from the crushed remaining bits of the root (550 mg), consisting mainly of the outer cementum layer. The third extraction (MA770) was conducted on non-digested cementum leftovers after the initial 24-hour digestion step in extraction MA756.

The substrates were incubated for 24 hours at 45Cº in 5 ml digestion buffer containing 4.7 ml 0.5 M EDTA, 50 µL Proteinase K (0.14-0.22 mg/ml, Roche) and 250 µL 10% N-Laurylsarcosyl. Following incubation the samples were spun down and pellets were stored for later re-extraction. A silica-powder-based extraction method (*122*) was used to isolate the DNA from the supernatant. Briefly, the silica binding buffer was prepared by mixing 118.2 g Guanidinium Thiocyanate with 10 mL Tris 1M, 1 mL NaCl 5M, 8mL EDTA 0.5M, 1 g N-Lauryl-Sarcosyl and molecular grade H_2O to a total volume of 200 ml. Then 20 mL of the binding buffer and 100 µl silica suspension was transferred to each sample and adjusted to pH 4-5 with 37% HCL. This was followed by a 3 hour incubation to bind the DNA to the silica. The supernatant was removed and the down-spun silica was resuspended in 1 ml binding buffer and washed twice with 80% cold ethanol. Finally, the DNA was eluted in 60 µl EB buffer (Qiagen). Extraction blanks were included with each round of extractions.

Following extraction, 20 µl of DNA extract was built into a blunt-end library using the NEBNext DNA Sample Prep Master Mix Set 2 (E6070) and Illumina specific adapters (*120*). A total of 10 libraries were built from the three extracts (one from MA755; five from MA756; four from MA770). The libraries were prepared according to manufacturer's instructions, with a few modifications outlined below (see also (*122*)). Because ancient DNA is already highly fragmented, the initial nebulization step was skipped. The end-repair step was performed in 25 µl reactions using 20 µl of DNA extract. This was incubated for 20 mins at 12°C and 15 mins at 37°C, and purified using PN buffer with Qiagen MinElute spin columns, and eluted in 15 µl. Next, Illumina-specific adapters (prepared as in ref. (*120*)) were ligated to the end-repaired DNA in 25 µl reactions. The reaction was incubated for 15 mins at 20°C and purified with PB buffer on Qiagen MinElute columns, before eluted in 20 µl EB Buffer. The adapter fill-in reaction was performed in a final volume of 25 µl and incubated for for 20 mins at 37°C followed by 20 mins at 80°C to inactivate the Bst enzyme. The entire DNA library (25 µl) was then amplified and indexed in a 50 µl PCR reaction, mixing with 5 µl 10X PCR buffer, 4 µl $MgCl_2$ (50 mM), 1 µl BSA (20 mg/ml), 0.5 µl dNTPs (25 mM), 1 µl of each primer (10 µM, inPE forward primer + indexed reverse primer), and 1 µl AmpliTaq Gold DNA Polymerase (Applied Biosystems). Thermocycling conditions were 5 min at 94°C, followed by 12 cycles of 30s at 94°C, 30s at 60°C and 40s at 72°C, and a final 7 min elongation step at 72°C. This was followed by a second PCR reaction (25 µl and 8 cycles) using 5 ul of the 'pre-amplified' library and P5/P7 primers (*120*). The amplified library was purified with PB buffer on Qiagen MinElute columns, before being eluted in 30 µl EB. Library blanks (built on H2O were included as well as libraries built on the extractions blanks.

The 10 DNA libraries were profiled on an Agilent Bioanalyzer 2100, followed by pooling and shot-gun sequencing (100 bp, single read). This was done across multiple runs on Illumina HiSeq 2000 platforms at the National High Throughput DNA Sequencing Centre, University of Copenhagen.

For the **Aconcagua mummy**, DNA was extracted from a piece of lung tissue following standard protocols, optimized for ancient DNA (*117*). In brief, 100 mg of lung tissue was weighed into a 5 ml LoBind Eppendorf tube and incubated in 1 ml of an EDTA-based digestion buffer containing 0.25 mg/mL Proteinase K for 10 min at 45°C. Subsequently, the buffer was replaced with 4 ml fresh digestion buffer, and the samples digested overnight at 45°C. The DNA was then purified using 50 µl of silica pellets and 40 ml of binding buffer containing 5 M Gu-HCl, 100 mM NaOAc, 20 mM NaCl and 30% isopropanol, as described in (*118*). The silica-bound DNA was washed twice in 80% ethanol eluted in 60 µl EB.

Sixteen microliters of the DNA extract were built into Illumina sequencing libraries using a recently developed protocol (*123*), with minor modifications as described in (*124*). Ten microliters of adapter-ligated library DNA were amplified and indexed in fifty microlitre reactions, using a dual indexing approach, as described by (*125*). The optimal number of PCR cycles was determined by qPCR using an MxPro 3000 (Agilent). The amplified libraries were purified using SPRI-beads and quantified on a 2200 TapeStation (Agilent) using High Sensitivity tapes. The amplified and indexed library was then pooled with other libraries based on their molarity and sequenced on a lane of an Illumina HiSeq 2500 run in SR mode. Following the initial screening, the same library was deep-sequenced on 2 lanes of an Illumina HiSeq 4000 run in SR mode.

Data pre-processing, mapping, genotyping and phasing

Base calls were produced using *CASAVA 1.8.2*. and reads were demultiplexed by requiring a perfect match to the 6- and 10-bp index sequences used for each library. *AdapterRemoval v1.5.3* (*60*) was used to trim Illumina adapter sequences, leading Ns (--trimns) and trailing quality 2 runs (--trimqualities --minquality 2) from both single- and paired-end reads. For the latter, overlapping pairs were collapsed (--collapse) into single reads requiring a minimum overlap of 11 bases and only such reads were kept for subsequent analyses, as done in refs. (*1*, *126*). Trimmed reads shorter than 30 base pairs were excluded from further analyses.

Filtered reads were mapped to build 37 of the human reference genome, following ref. (*1*). We used *bwa aln v0.6.2-r126* (*61*) with disabled seeding (-l parameter) (*62*), after which we excluded reads with mapping quality lower than 30. PCR duplicates were removed using *picard-tools MarkDuplicates* (*63*), and local realignment was carried out using *GATK* (*64*). The *MD* tag, as well as the extended *BAQ* were computed for each read using the *calmd* command included in *samtools 0.1.18* (*65*).

We only considered samples with an average depth of coverage greater than 10X for diploid genotype calling, while we sampled one random allele at each site for samples with a lower average depth. For samples for which USER-treated data was available, we only used such data for diploid genotype calling (Table S8). We called diploid genotypes following ref. (*2*), using *samtools mpileup* (*65*) with the -C parameter

set to 50, on sites with a depth of coverage lower than 2 times the average depth and greater than 1/3 the average depth. Variant calls located within 5 bp of each other, with a phred posterior probability lower than 30, or with a significant strand and/or end distance bias (p<1e-4) were filtered out using *bcftools*. Additionally, heterozygous calls with an allelic balance lower than 0.2 were discarded from the following analyses. Note that we used this strategy to call genotypes for all genomes described in (Reference datasets), in order to reduce any potential batch effect.

Diploid genotypes were phased using *impute2* (*66*, *67*), over 5Mb-long windows with the -fill_holes and -no_remove options, following ref. (*127*). We used the 1,000 genomes phased variant panel (phase 3) as a reference and the HapMap recombination rates. Imputed sites not present in the original variant set were set to missing, and the phased dataset was masked using a 35-mer 'snpability' mask with a stringency of 0.5 (*68*), as well as the strict accessible regions from the 1000 genomes project (*69*).

Supplementary Text

Ancient DNA sequence data quality control

We performed a standard quality control for ancient DNA (aDNA) data produced in this study following ref. (*1*).

We assessed the authenticity of ancient DNA sequence data on the basis of the **fragment length distributions** and **nucleotide misincorporation patterns** (*128*). We ran *bamdamage* (*20*) on the library level and observed an enrichment in shorter fragments as well as increased frequencies of C to T substitutions in data derived from non-USER-treated extracts. As expected, the aDNA-characteristic substitution pattern is not evident in data derived from USER-treated extracts (*17*). We note that for each USER-treated extract, we screened at least one library prior to the treatment; all of which yielded a misincorporation pattern characteristic of authentic aDNA. In Table S8, we show the frequency of C to T substitutions in the 5' terminal base, and the mean fragment length for each library. We show examples of these patterns in Fig. S5.

For each library, we estimated the overall and type-specific **error rates** described in ref. (*122*). We used *ANGSD* (*71*) to estimate the number of derived alleles in high quality data produced by the 1000 genomes project (sample NA12778), with respect to the chimp-human pa*ntro2-hg19* multiZ alginments from the UCSC Genome Browser (*129*). We compared this expectation to the number of derived alleles in each library using the same reference, and attribute excess derived alleles to error. As expected, ancient DNA data showed increased error rates compared to the high-quality sample, most likely due to *post-mortem* degradation (Table S8 and Figure S6). Moreover, C to T and G to A substitutions accounted for the majority of the error in data derived from non-USER-treated extracts.

We estimated the amount of **mtDNA contamination** in each library using the approach described in ref. (*70*), and following the procedure in ref (*1*). In brief, we called a majority rule consensus for the mtDNA using *ANGSD* (*71*), based on reads with mapping quality greater than 30, bases with quality greater than 20 and sites with depth greater than 5X. We then remapped all reads to the new consensus in order to call a refined consensus sequence. We aligned the new consensus to a set of 311 worldwide mtDNA sequences using *mafft* (*130*, *131*) and used *contamMix* to estimate the maximum *a posteriori* probablity of the consensus sequence being authentic. In Table

S8, we show the contamination estimates for each library and note that the Gelman-Rubin diagnostic (*132*) suggested convergence for all runs.

For each library derived from male individuals, we used the approach introduced in ref. (*133*) to estimate the **amount of nuclear contamination**, based on the reads mapped to the **X chromosome**. This method relies on the fact that males are hemizygous for X-linked loci (excluding the pseudoautosomal regions). Thus, multiple alleles in such loci could be attributed to error or contamination. In brief, non-consensus base counts are modeled as a function of the non-consensus allele frequency in the contaminant population, an overall error rate estimated from fixed sites and the contamination fraction. The latter is estimated through a maximum likelihood optimization. We used the implementation included in *ANGSD* (*71*) to estimate contamination based on reads with a mapping quality greater than 30, bases with quality greater than 20, sites with depth between 3 and 20 and the HapMap CEU allele frequencies (*134*). In Table S8, we show contamination estimates for each library from male individuals.

We obtained a third **contamination estimate using autosomal data** for samples with an average depth of coverage greater than 10 using *DICE* (*72*), following the strategy used in ref. (*1*). For each high-depth sample, we used the 1000 genomes project allele frequencies for the two-population model, where we set 'YRI' as the anchor population and 'CEU' as the putative contaminant. We ran 100,000 steps of the MCMC algorithm and discarded the first 10% as burn-in. The remaining iterations were summarized using the *coda* R package. In Table S8, we report *DICE* contamination and error rate estimates for high-depth samples.

For each individual, we determined the **chromosomal sex** using the approach detailed in ref. (*135*). We show these results in Table S8.

Y chromosome

Y chromosome analysis in aDNA samples faces several challenges, due not only to issues related to aDNA work *per se* – as ubiquitous *post-mortem* DNA damage profile coupled with low-coverage samples (*136*) – but also due to the inherent low mappability of the Y chromosome, and the difficulty of placing low coverage sequence data on a phylogeny (*137*, *138*). First, we estimated the Y chromosome sequence coverage in all ancient samples in the single copy Y-specific 10 Mb region defined in (*139*) known to be suited for appropriate calling in short-read sequencing. It was achieved using SAMtools 1.3.1 depth function (*65*) (coupled with arguments '-b' referring to the bed file containing the 10 Mb region; '-r Y' to specify the Y chromosome and '-aa' to force SAMtools to output all sites) and found individual samples to vary between coverage 0.01 to 9.55x (Tables S8,9). As three ancient samples presented coverage high enough to be compared with a modern dataset, after unifying sequence dictionaries we called the variants using FreeBayes v1.2.0 (*140*) on a dataset containing the three aDNA sequences and an additional 26 relevant high coverage Y chromosomes from the literature (4 ancient), representing the known diversity of Y haplogroup Q (*141*) (Table S10). To address the particularities of calling variants on ancient Y data, we used arguments '-ploidy 1', '-report-monomorphic', '-i' to ignore indels, '-F 0.9' to expect FreeBayes to only output sites with over 90% variant occurrence on the reads of the same sample, '-C 2' so a variant is only considered if it is seen on two non-clonal reads and '-t' referring to the bed file with the 10 Mb accessible

region of the Y chromosome. We additionally used vcfallelicprimitives function of vcflib to break down complex variants (MNPs) and bcftools v1.3.1 filter function to remove heterozygous sites, as they were likely results of contamination. Due to post-mortem DNA damage, we intended to build a phylogeny using only high-quality transversions, but that would make the definition downstream of Q-M3 impossible, as key variants are all transitions. We then sought to increase the filtering threshold on transition variants, which we did by running an in-house script to set all transitions supported by less than 4 reads to missing, while accepting transversions observed on 2 reads. This resulted on a dataset containing 1321 transversions and 2206 transitions, which we converted to fasta format to build a maximum-likelihood phylogeny using RAxML v8.2.10 (*142*). We counted and included on the input all positions confidently called as reference using the '-stamatakis' ascertainment correction and used the 'ASC_GTRGAMMA' substitution model, with 100 bootstrap replicates. The tree was visualised using the software FigTree and rooted between the Q and R haplogroups (Fig. S7).

The placing of low-coverage ancient samples in a high-coverage phylogeny presents several drawbacks (*137*), and sometimes it is unclear whether samples appear to be basal due to missing genotypes or due to the age of the sample. In this respect, the adequate data curation of defining SNPs and of variants with equivalent phylogenetic signal is paramount, as well as the concomitant validation by ancestral calls in derived positions downstream. We used FreeBayes with similar parameters as used in the dataset of high-coverage samples, but outputting all positions even if supported by only one read ('-C 1'), and then compared to manually-curated high-confidence SNPs from Native Americans (from refs. (*7*) and ISOGG) (Table S9). In this case, we included in the dataset the historical Andaman islander, and other basal K samples from the literature (Table S10), which also have a difficult phylogenetic placement. We then visually inspected the positions of key importance for the placement of the sample to confirm the genotype.

Albeit our strict filtering, the maximum likelihood approach was able to recover the expected tree topology (*7*, *141*, *143*), and the three high-coverage American aDNA samples fell within Q-M3(M848) star-cluster, with an estimated split time of 15 ka (*141*) (Fig. S7). When comparing all samples with the annotated SNP list, 7 out of 9 new aDNA samples presented derived positions downstream of Q-L54, Q-M930 and Q-M3, and for five of those also for Q-M848. One sample (Sumidouro4, chrY coverage: 0.43x) misses all positions for Q-M848, but carries the ancestral allele for three variants at the Q-Y3408 branch, hence it was defined as Q-M3(xQ-Y4308) (Table S9). Another sample (AHUR770c, chrY coverage: 0.05) can be confidently placed at Q-M3, but misses all downstream positions. Of particular interest was the Y chromosome of Lovelock4 individual (chrY coverage: 0.27x), which was negative for Q-L54, becoming the second fully-sequenced non-Q-L54 Y chromosome from the Americas not belonging to haplogroup C (*7*), as the 4.5 ky old Palaeo-eskimo from Greenland (*144*) (but see (*145*)). The Lovelock4 fell basal to Q-YP4011 haplogroup (as anticipated in (*7*)), and is ancestral to most SNPs found in Eastern Europeans Q-YP4004 (Table S9). This suggests a very ancient split, although the low-coverage of this sole sample makes detailed inferences tentative. Fig. S8 shows the proposed topology of the tree after the inclusion of the ancient samples as hashed lines.

Haplogroup Q-M848, downstream of Q-M3 (and Q-L54), is the most frequent paternal lineage of the Americas, vastly outnumbering all other lineages combined. In fact, with the exception of CT-M168, the Out-of-Africa paternal lineage, likely no other paternal lineage was found occupying such vast continental area at high frequency in pre-modern times (*141*). This highly disproportionate distribution in the American continent very likely represents a strong founder effect and a rapid settlement of the diverse environments of the Americas (*146*), occupying previously empty areas, reaching as far south as Chile by 14.5 ka (*147*). Uniparental genetic evidence suggests a coastal expansion taking place around 16 ka (*8*). Such a fast population expansion has been described to be archaeologically instantaneous, but is not unprecedented – in the settlement of Sahul, where humans occupied all resource-rich areas in equally swift fashion (*148*, *149*).

Between 13-15 ka, the Q-M848 phylogeny displays a star-like pattern including individuals from all geographical locations in the Americas, a signal usually interpreted to represent a strong population expansion. These results mirror, and in fact anticipate (*141*), the autosomal data results, where a very close connection is observed between the 10 ky old samples from Lagoa Santa and Spirit Cave, more than 10,000 km apart, suggesting that large distances were covered rapidly. Strikingly, although there is close mtDNA connection (see below) between the 12.6 ka Anzick1 individual from Western Montana (*3*) and Sumidouro5, they present different Y chromosome haplogroups. Anzick1 falls on the Q-CTS1780 branch (Fig. S7 and S8), a different Beringian lineage (*3*), which is rare, but relatively widespread in the continent. It is also worth noting that although genomically we see a cluster formed by Spirit Cave, Lagoa Santa and Anzick1, those populations probably suffered genetic drift on their quick spread on the continent, which may be reflected in different haplogroup frequencies.

Both the Sumidouro and Spirit Cave remains were previously classified as having a plesiomorphic skull morphology (*38*, *39*), a finding with alternative interpretations (*38*–*40*, *42*, *150*, *151*). Regardless, that all ancient individuals with this distinct morphology tested so far belonged to the most common paternal lineage in the continent (Q-M3) favours a scenario where microevolutionary patterns and Holocene gene flow had a role in the skull shape differences between present and ancient (>8 ka) Native Americans (*42*).

The historical Andaman Islander Y chromosome is, to our knowledge, the first occurrence of basal haplogroup P*. Fig. S9 conveys the basic structure of haplogroup K-M9, one of the largest patrilineages of mankind, of whom the majority of Y chromosomes from Europe, East Asia, Oceania and Americas descend from, and with four basal lineages highlighted, three one of them being identified elsewhere (*141*, *152*, *153*) and the historical Andaman Islander (P*). Extensive Y chromosome calls for all relevant positions listed on ref. (*154*) and ISOGG] can be found on Table S9, allowing us to definitely place or re-place 91 SNPs of doubtful phylogenetic position downstream of K2-M256, and to create a comprehensive list of all SNPs defining the deepest branches of K-M9 haplogroup (16 new) (Table S11). Most of those variants now properly define haplogroup P-P295 and P1-M45, which is of global interest for Y chromosome variation as those can be used to understand the ancestral state in the reference Y sequence itself, which falls downstream of P (haplogroup R1).

We estimate the divergence time of the lineages downstream of K-M9 (Fig. S9 and Table S11) using previously published methods (*155*) and the mutation rate defined in (*152*), and found the split between P* and P1 to have taken place around 53.6 ka (CI 95%: 46.8-61.3). This date is chronologically close and coherent with the proposed view in (*156*), where the deep divergences found in Andamanese Y chromosomes are synchronous with the first splits of Out-of-Africa patrilineages, and can be accommodated in a model of a single expansion event.

mtDNA

We first estimated the coverage of the mtDNA of all new aDNA samples using SAMtools v1.3.1 depth function, arguments '-r MT' to target only reads mapping the mtDNA and '-aa' to force SAMtools to output all positions, and found mtDNA coverage to range from 0.27x to 559.72x. We then used ancient DNA aware software Schmutzi (*157*) for assembling and creating a consensus sequence, only allowing positions to be called at quality 20. Schmutzi uses the *post-mortem* deamination patterns to diagnose endogenous (ancient) reads thus avoiding present-day contamination.

For the alignment and haplogroup determination we replicated the standard practice established in the literature (*8*, *158–160*): the 17 mtDNA consensus sequences were visually inspected for alignment errors (due to low-coverage), and problematic positions 309.1C(C), 315.1C, AC indels at 515-522, A16182c, A16183c, 16193.1C(C) and C16519T/T16519C were removed using MEGA7 (*161*) following reference (*158*). We then used Haplogrep (*162*) and Haplofind (*163*) to determine the sample haplogroups for samples with sufficient coverage, and manually checked the genotype for key variants (phylotree.org), confirming the haplogroup call. Samples with too low mtDNA coverage (Sumidouro2: 0.27x; Sumidouro8: 3.66x) for a high definition placement to be made by automatic calling were further analysed for downstream SNPs manually.

To establish the phylogenetic relationships between the ancient samples and the mtDNA diversity in the American continent, we compared them to all relevant sequences from the literature, in a dataset that contained 38 modern and 7 ancient mitogenomes for D4h3a and 197 modern and 20 ancient D1 sequences (Table S10). After using MEGA7 (*161*) on this dataset to correct possible misalignments and removing the sites described in ref. (*158*), we partitioned our dataset in coding, non-coding and control regions, as they are known to evolve at different rates (*164*). We recovered the expected tree topology in a maximum-likelihood analysis using RAxML v8.2.10 (*142*) with 100 bootstrap replicates (Figs. S10 and S11), where all ancient American samples fell together with present or ancient Native American diversity. We then ran a Bayesian phylogenetic analyses using BEAST v2.4 (*165*) for each haplogroup separately, and used the estimated age of the ancient sequences as calibration points. Each partition was assigned an independent relaxed lognormal clock and also independent substitution models, using the mutation rates described in ref. (*164*). We also included priors regarding the known dates of settlement of the American continent. We reasoned that 12.6 ka, the age of Anzick1 (*3*), should be used as a lower bound, as it is the oldest Native American sequenced so far. We performed 200,000,000 steps with sampling every 10,000 generations, and did this twice for each phylogeny, to ensure reproducibility of the Markov Chains, and the Effective Sample Sizes (ESS) values were higher than 200 in both replicates for both phylogenies. The first 10,000,000 steps were not considered when sampling the trees, and split times were all

calculated by BEAST itself. The tree was visualized and edited for clarity on FigTree (Figs. S12 and S13). The approximate geographical locations of all samples were obtained either on Genbank or in their original paper, and was plotted on Figs. S14 and S15 (listed in Table S10).

Of the six Early Holocene human remains from **Lagoa Santa**, all belonged to mitochondrial haplogroup D4, which contains both D1 and D4h3a subhaplogroups. Three samples (Sumidouro4, mtDNA coverage: 26.54x; Sumidouro6, mtDNA coverage: 48.66x; Sumidouro7, mtDNA coverage: 10.24x) could be confidently placed at a basal position at haplogroup D1, carrying all expected variants for D1 and being negative for further downstream placement (Fig. S12). Out of the other three sequences, only Sumidouro5 (mtDNA coverage: 227.31x) could be placed at D4h3a by automated callers, and the placement of Sumidouro8 (mtDNA coverage: 3.6x) was done by visually inspecting for diagnostic variants (Table S9). The high-coverage D4h3a sequence was found to be negative for downstream variants, and we found a G to A transition at position 4769 shared between Anzick1 (*3*), Sumidouro5 and Sumidouro8, but negative for all other sampled D4h3a sequences, as supported by both ML and Bayesian analyses (Figs. S11 and S13). The split time between the Anzick1 and Sumidouro5 mitogenomes were estimated to have taken place around 13.5 ka (upper bound: 15.1 ka, lower bound: 12.6 ka), similarly to as proposed by our demographic modelling. The TMRCA for D4h3a was estimated at 15.3 ka (upper bound: 19.4 ka, lower bound: 13 ka), in agreement with previous estimations (*9*).

Although those two mtDNA lineages belong to what are labelled the Pan-American founding lineages (*166–168*), and even both have been described as likely having a Beringian origin (*169*), their geographic distribution is uneven in the Americas. Today, D1 has a broad distribution, but is more common in South America and peaks in frequency on the Southern Cone (*170*) (Fig. S14); the rare D4h3a seems restricted almost exclusively to the Pacific coast of South America, with few occurrences in Mexico (*9*, *171*) (Fig. S15). Ancient DNA shows a slightly different scenario, as the oldest mitochondria analyzed in the Americas are from North America and belong to D4h3a and D1 haplogroups (*3*, *9*, *172*, *173*) (but see (*174*, *175*)). This discrepancy is, however, consistent with the autosomal profile of North American 'Paleoindians' falling on the SNA branch (*3*). Regarding South America, only one D4h3a mitochondria has been found farther east of the Andes, on Maranhão state of Brazil (*171*), and it is basal on the phylogeny (Figs. S11 and S13), sharing no downstream variants with no other samples, including Lagoa Santa mitogenomes. This result largely mirrors our autosomal results, on which Lagoa Santa has not contributed significantly to present day populations (of our sample), with the exception of some Amazonian populations (Karitiana, Suruí), both of which also have high frequency of D1 mitochondria (D1a, D1e). Sumidouro samples carry no derived position related to those haplogroups, but this may be due to the sequences antiquity, as BEAST estimates the subhaplogroups TMRCA (D1a: 4.5 ka and D1e: 7.9 ka) to be younger than Lagoa Santa itself (Figs. S12).

The co-occurrence at high frequency of haplogroup D1 and D4h3a is not rare only east of the Andes, but also in a continental scale – the only present or past population known to carry high frequencies of both haplogroups are the Fuego-Patagonians (*2*, *43*, *176*), belonging to the highly spatially restricted D4h3a5 and D1g haplogroups. Lagoa Santa mtDNAs carry the ancestral state of the defining variants of

both haplogroups, but this could again be due to the old age of the Brazilian samples for D4h3a5 (D4h3a5 TMRCA: 5.1 ka, upper bound: 8.1 ka, lower bound: 2.9 ka). In the case of D1g, its TMRCA is consistent with it having an origin before entering South America (D1g TMRCA: 19.7 ka – similar estimates were reached using a ML approach (*170*, *177*)), but basal D1 is also found among Fuego-Patagonians (*43*, *177*). These findings give strong support to previously proposed hypothesis of a Late Pleistocene continuity in Patagonia (*170*, *176*, *177*), and to a common genetic stock between Lagoa Santa and Patagonian populations, and affirmed by our whole genome analysis. These mtDNA results propose a large and undocumented population shift east of the Andes, possibly related to the concomitant, and possibly related, spread of a derived skull morphology (*42*).

Previous whole-genome analysis of two Terminal Pleistocene Alaskan burials (*1*, *178*) have shown the existence of Ancient Beringians (AB), which now includes the **TrailCreek** individual. Consistent with this autosomal clustering, TrailCreek mtDNA sequence falls on the root of the haplogroup B2, lacking an expected variant (A3547G), as does one of the Ancient Beringians (URS2) (*179*). In fact, although we were able to reconstruct only 85% of TrailCreek mtDNA sequence due to our sequencing strategy and coverage, it has the same haplotype as the USR2 mtDNA, independently supporting the AB profile of TrailCreek autosomes. The absence of the expected variant (A3547G) that defines haplogroup B2 in USR2 has been argued to be a back-mutation (*179*) as, at the time, the exact relationship of the individual with the rest of Native American diversity was not clear. We now reason here that both the split time between Ancient Beringians and other Native Americans (*1*) and the TMRCA of haplogroup B2 (*179*) are consistent with the mutation taking place after the split, thus making it a defining variant for all Native American B2 mtDNA south of Alaska.

The ancient mitochondrial diversity in the **North American Great Basin** had been previously studied by PCR-based approaches, which found a high proportion of haplogroup D (*180*). According to the original authors, this skewed distribution of haplogroups was incongruent with a scenario of population continuity in the region, in the light of mtDNA haplogroup frequencies of surrounding populations (but see (*181*)). Although we have large gaps in our temporal transect of the region, we confirm the high prevalence of haplogroup D, of which we increased definition to D1, in both 10 ka samples (AHUR_2064, mtDNA coverage: 559.72x and AHUR770c, mtDNA coverage: 135.7x) and in three more recent samples, ca. 2 ka (Lovelock1, mtDNA coverage: 65.83x; Lovelock2, mtDNA coverage: 270.59x and Lovelock4, mtDNA coverage: 34.5x). Those samples, with the exception of Lovelock3, which may harbor extra ancestry, show also an autosomal signal suggesting a deep continuity in the region. Intriguingly, the most recent sample, a 700-year old individual also from the Lovelock Cave, belongs to the B2a1 haplogroup (Lovelock3, mtDNA coverage: 364.3x), which may reflect the arrival of Mesoamerican-related ancestry in the region.

All **Andaman Islanders** tested so far belong to mtDNA haplogroup M31 or M32 (*182*, *183*), a deep branch on the macrohaplogroup M phylogeny. Those two haplogroups, however, are not restricted to the islands, being found also in mainland India (*184–186*), Myanmar (*186*) and even Saudi Arabia (*187*). The historical Andaman Islander sequenced in this study is not only negative for the defining markers of both M31 and M32, but in fact is negative for all markers downstream of haplogroup M, stemming directly from the root (M*). Further comparison between the Andaman

Islander sequence and the literature M* confirmed the basal position of the Andaman Islander mtDNA, and give support to findings using the Y chromosome (this study and (*156*)), in which the deep splits between Andamanese lineages and other global populations are synchronous with the first Out-of-Africa population divides.

Reference datasets

We compiled a whole genome sequencing and a SNP array dataset for comparison with the ancient genome data generated in this study.

Whole genome sequencing data was compiled from multiple studies (*1–3*, *6*, *9*, *10*, *31*, *126*, *127*, *138*, *152*, *188–204*). For high-depth genomes we called diploid genotypes (Data pre-processing, mapping, genotyping and phasing), while we sampled one random allele at each site for low-depth samples (discarding reads with mapping quality below 30 and bases with quality below 20). We excluded sites that deviated significantly from the Hardy-Weinberg equilibrium (p<1e-4, based only on high-depth samples), as well as regions not included in the strict accessible regions from the 1000 genomes project (*69*). In Table S12, we present summary statistics for this dataset.

The **SNP array** reference dataset is identical to the dataset described in SI 6.1 from ref. (*1*). It includes 2,537 contemporary individuals from 167 ethnic groups (enriched in Native Americans), genotyped over 199,285 SNP sites (*4*, *31*). Note that Native American individuals were masked for non-Native American ancestry tracts. We extended this panel with all individuals from the whole genome dataset.

Multidimensional scaling

We explored the relationships between the ancient American genomes and ancient and present-day worldwide populations in the SNP array dataset using multidimensional scaling (MDS). Since not all individuals are represented by diploid genotypes (Reference datasets), we followed a strategy similar to that implemented in ref. (*20*). For each individual, for each site, we sampled one random allele and computed the identity-by-state distances for each pair of individuals using plink (*205*). We then computed an MDS transformation using the *cmdscale* R function. We built three different MDS transformations based on three subsets of the panel: 1. all individuals excluding Africans, 2. Native Americans belonging to the South Native American (SNA) group (*1*), and 3. Native Americans belonging to SNA, except the Pima, Surui, Karitiana and Cabecar, which were outliers in subset 2. For each panel we plot the first two dimensions as well as dimensions 4 and 5 for subsets 1 and 3, respectively (Fig. S16). In what follows we describe the placement of each of the novel genome sequences.

Trail Creek was placed adjacent to USR1, a representative of Ancient Beringians (AB); an early branch off Native Americans who inhabited eastern Beringia as late as ~11.5 ka (*1*). Since Trail Creek and USR1 are approximately close in time and geography, these results support that AB occupied eastern Beringia during the end of the Pleistocene and as late as the early Holocene (Trail Creek age, ~9 ka).

Along the first two dimensions of the MDS transformation, **Big Bar** was placed separate from SNA and together with NNA, including Algonquian, Athabascan, Tsimshian, and Salish speakers among others (Fig. S16A). This placement is similar to that of other ancient genomes from North America such as **939** (British Columbia) (*2*),

ASO (Southwestern Ontario) (*7*), and **Kennewick** (Washington) (*10*). However, Big Bar did not cluster with any of these groups or ancient genomes in higher dimensions. Dimension 5 placed Big Bar in an intermediate position between NNA groups with and without recent Siberian admixture (the former represented by 939) (*1*, *2*, *18*). Meanwhile, ASO and Kennewick were placed in an intermediate position between SNA and NNA (Fig. S16B). These results suggest that, despite their broad genetic similarities and apparent geographic proximity, there are finer genetic structure patterns that differentiate these individuals and populations.

The **remainder of the ancient genomes** presented in this study clustered together with SNA populations. MDS transformations focused on SNA showed that ancient individuals, in particular those with calibrated radiocarbon ages greater than ~5 ka, did not cluster together with any particular SNA population across the first five dimensions. Although the Spirit Cave and Lagoa Santa genomes (~10 ka) were placed adjacent to Ge, Tucanoan and Andean speakers close to zero in the first two dimensions (Fig. S16C,D), this pattern was not repeated in higher dimensions. In Fig. S16E, ancient genomes remain close to zero, together with Anzick1 which has been previously shown to be directly ancestral to some SNA populations (*3*). Both ancient Patagonians (Punta Santa Ana and Ayayema) show similar results. Nonetheless, these individuals were slightly shifted towards present-day Fuego-Patagonians, which potentially reflects greater shared genetic drift with such populations as a result of their more recent sampling times (~7 and ~5 ka). In contrast, previously published historical genomes from Patagonia and Mexico (*2*) were placed adjacent to present-day individuals from those regions. Similarly, the ~0.5 ka Aconcagua individual clustered together with present-day Andean speakers. This observation for ancient genomes dating to the last millennium, together with the fact that the Spirit Cave genome is based solely on USER-treated aDNA extracts (Data pre-processing, mapping, genotyping and phasing) (*17*), suggest that these results are not mainly driven by correlated error patterns caused by *post-mortem* damage. Altogether, these results suggest that even though these ancient individuals (Anzick1, Spirit Cave, Lagoa Santa, Punta Santa Ana and Ayayema) cluster together with SNA, they represent early Native American populations that share little genetic drift with present-day populations. These results are further supported by the clustering analysis of population f_3 statistics.

Admixture

We carried out a model-based clustering analysis to explore the genetic ancestry components of present-day and ancient Native Americans. To do so, we considered a subset of the SNP array dataset (Reference datasets) including all ancient Native American samples shown in Table S12, 40 YRI individuals, 40 French, 40 Han Chinese, 40 Papuans and all present-day Siberians and Native Americans. Additionally, we included the following ancient individuals from Eurasia: UstIshim (*152*), two Hoabinhians (*192*), Tian Yuan (*191*), Yana (*189*), Mal'ta (*31*), Jomon (*192*), four individuals from the Devil's Gate Cave (*189*), three ancient Ekvens (*189*), three individuals from the UstBelaya site (*189*), and the historical Andaman sequenced in this study. We excluded individuals with missingness greater than 95% and for each site, we sampled one random allele in order to account for low depth ancient individuals for which genotype calling is not feasible. Following ref. (*1*), we ran *ADMIXTURE* (*19*) on this subset assuming $K=\{2, ..., 16\}$ ancestral components and for each value of K we selected the run with the best likelihood out of 100 independent replicates. For clarity,

barplots in Figs. 1F and S17 only show three individuals from populations with $n>3$, and single genomes are represented as wider bars.

This analysis recapitulates previously observed broad continental and Native American genetic structure patterns *e.g.* (*1*, *2*). We observed eight Native American-specific ancestry components that are maximized in: Andean speakers from South America ($K=6$), Chibchan and Paezan speakers from Central and South America ($K=7$), the Surui from the Amazon ($K=8$), NNA ($K=10$), the Pima from Mexico ($K=11$), the Karitiana from the Amazon ($K=12$), Eskimo-Aleutian speakers from the Arctic ($K=15$) and the Mixe from Mexico ($K=16$) (Figs. 1F and S17). Six of these groups are classified as SNA (*1*, *3*). In contrast, the populations carrying a majority of the components that arose at $K=10,15$ (NNA and Eskimo-Aleutian speakers, respectively) are classified as NNA (*1*, *3*). These eight components are distributed differentially among ancient Native American genomes.

The ancestry component distribution in **Trail Creek** is qualitatively comparable to that in USR1. Both individuals carry similar proportions of 'Siberian' and Native American components. The latter are divided into approximately equal proportions of NNA and SNA (Figs. 1F and S17). These results are consistent with MDS, which placed these ancient individuals in an intermediate position between Siberians and Native Americans. Thus, we infer that both individuals were members of the 'Ancient Beringian' group, a population that diverged from Native Americans after their divergence from Siberians but before the NNA/SNA split (*1*).

Similar to other contemporaneous and ancient individuals from the Pacific Northwest, the 'NNA' component accounts for the majority of the ancestry in **Big Bar** (*2*, *9*). However, we observed that Big Bar bears a slightly increased proportion of SNA components, such as those present in Andean speakers and the Mixe. We observed that other ancient individuals from North America such as Kennewick (*10*) and ASO (*7*) carry a different distribution, characterized by a greater proportion of SNA components (Figs. 1F and S17). These patterns are consistent with those obtained through MDS, and suggest that complex population history underlies the observed broad patterns not only in the Pacific Northwest, but also in other regions that were once covered by glacial ice (*9*, *18*). In (Admixture graph fitting using qpGraph), we derive more nuanced models for the ancestry of these individuals.

For the **ancient SNA** individuals, we observed different ancestry component distributions consistent with their sampling location and ages. We found that the distribution in the ~12.6 ka Anzick1 individual (*3*) is mirrored in both Spirit Cave individuals, in that they mainly carry a mixture of SNA components. This pattern suggests that Spirit Cave and Anzick1 were likely part of a relatively homogeneous population that occupied mid-latitude North America during the Late Pleistocene and the Early Holocene. In contrast to MDS results, the ancestry component distribution in the Lagoa Santa individuals (also dating back to the Early Holocene), resembles that of Equatorial, Tucanoan, Ge, Pano and Carib speakers. These results suggest that South American groups and the population represented by the Lagoa Santa individuals share ancestry independently from northern groups including Mesoamericans, Spirit Cave and Anzick1. *ADMIXTURE* results were similar for more recent South American genomes. The ancestry component distributions in ancient Patagonians and in the pre-Columbian Aconcagua mummy resemble that in present-day Andean speakers.

f_3 statistics

We investigated the genetic affinity between the populations included in the SNP array dataset (Reference datasets) by computing outgroup f_3-statistics of the form $f_3(San; A, B)$. Assuming there was no independent admixture between the outgroup and either of the ingroups, f_3 is proportional to the shared drift between A and B after the divergence of their common ancestor from the African San (*21*). We computed f_3 as detailed in ref. (*21*) and computed standard errors for each statistic using a weighted block-jackknife procedure over 5Mb blocks.

We computed f_3 for all possible pairs of East Eurasian and Native American populations (Fig. S18) and summarized these results using two approaches. 1. a heatmap for the f_3 statistics, ordered according to the hierarchical clustering procedure implemented in the *heatmap.2 R* function (Fig. S18). 2. we scaled the f_3-statistics using the maximum observed value of f_3 and computed an MDS transformation based on a matrix of 'f_3 distances' between populations (Fig. 1D-E) (*22*). For this approach we considered two population subsets: one including Siberians and Native Americans and one including SNA populations only.

We observed broad concordance between both f_3-based clustering approaches and the IBS distance-based MDS. All three approaches allowed us to identify three main Native American clusters corresponding to the previously described SNA, NNA and AB groups (*1*, *2*). Furthermore, the internal structure of these groups was comparable across three approaches, as well as the placement of the rest of the ancient Native American genomes. As such, we interpret these results following (Multidimensional scaling).

D-statistics

We computed *D*-statistics to formally test specific hypotheses of 'treeness' and gene flow between populations. In brief, if a set of four populations are related by an unrooted tree of the form $(H_1, H_2; H_3, O)$, we expect H_1 and H_2 to be symmetrical to H_3 and O ($D(H_1, H_2; H_3, O)=0$). We expect D to deviate from 0 if: 1. the proposed tree is incorrect; 2. there was independent gene flow between H_3 and H_1 or between H_3 and H_2; 3. data from H_1 and H_2 have differential error rates, thus giving rise to artificial allele frequency correlation between the group with the largest error (*H1* or *H2*) and the most divergent group (*H3* or *O*); or 4. the ancestral population of H_1, H_2 and H_3 was structured up to the time in which H_1 and H_2 diverged, thus resulting in differential allele sharing between H_3 and H_1 or between H_2 and H_1 (*21*, *206–208*). We obtained the standard error of D through a weighted block-jackknife procedure over 5Mb blocks and assessed the statistical significance of $D\neq0$ with a Z-test (*21*). In what follows, we consider that D-statistics with $|Z|>3.3$ (corresponding to a p-value of ~0.001) deviate significantly from $D=0$.

In order to control for and explore the potential bias deriving from differential error rates in the dataset (Ancient DNA sequence data quality control), we computed *D*-statistics following three different approaches. 1. 'standard' *D*-statistics based on called genotypes (Data pre-processing, mapping, genotyping and phasing), as detailed in ref. (*21*); 2. 'standard' *D*-statistics based on randomly sampled alleles (*71*, *206*); and 3. error corrected-*D*-statistics (*23*). For all tests involving the whole genome sequence dataset (Reference datasets), we only considered transversion polymorphisms located within

the strict accessible regions from the 1000 genomes project (*69*), unless otherwise stated. Since transition polymorphisms dominate the SNP array dataset due to the merging of different datasets (Reference datasets) (*2*, *4*), we considered both transition and transversion polymorphisms for these data in the interest of keeping the largest possible number of sites.

In order to **exclude the signal of recent non-American admixture** (*4*) from subsequent analyses, we defined a subset of the whole genome sequence dataset (Reference datasets) containing **'non-admixed' SNA genomes only**. We computed *D*-statistics of the form *D(SNA, Aymara/Mixe; Yoruba/French/Han, Chimp)* to identify SNA individuals with African, European or East Asian admixture. We used the Mixe from Mexico (two individuals from ref. (*6*)) and Aymara from Peru (one individual from ref. (*2*)) as proxies for 'non-admixed' Native Americans and considered that a given SNA individual is 'admixed' if we observed $|Z|>3$ for any test (Fig. S19). Based on this conservative criterion, we excluded a Quechua (LP6005677-DNA_F01), a Zapotec (LP6005677-DNA_D01) and both Mixtec genomes from the following analyses.

We explored the relationship between the ancient genomes presented in this study and the **basal Native American groups** that diverged ~15 ka, likely south of eastern Beringia: North (**NNA**) and South Native Americans (**SNA**) (*1–4*). The former include Inuit, Algonquian, Athabascan, Tsimshian, and Salish speakers (among others) from North America, with a number of these populations carrying mid/late-Holocene admixture from at least two Siberian sources (*1*, *4*, *18*, *33*). The latter include most Native American groups south of Montana, where one of the earliest representatives of this group has been documented (*2*, *3*).

We computed *D*-statistics of the form *D(Aymara, Native American; Ancient, Yoruba)*, where the Aymara from Peru is a proxy for SNA (Figs. S20,21). In order to account for recent Siberian-related admixture into some NNA groups, we computed *D*-statistics of the form *D(Aymara, Native American; Han, Yoruba)*, and took these results into consideration when interpreting the results. For these tests, we used the SNP array dataset described in (Reference datasets). In agreement with results presented in (Multidimensional scaling, Admixture, and f_3-statistics), we observed most of the ancient genomes presented in this study are more closely related to SNA than to NNA, with the exception of Trail Creek and Big Bar (Figs. S20,21). In the following sections, we use these results to aid in the interpretation of particular analyses.

MDS, ADMIXTURE, and f_3-statistics suggest that **Trail Creek** and USR1 are similarly related to other Native Americans. We formally tested whether USR1 shares a significantly larger proportion of alleles with Trail Creek than with other Native Americans by computing *D*-statistics of the form *D(USR1, Native American; Trail Creek, Yoruba)* using both datasets (Reference datasets). While we were not able to reject the null hypothesis ($D=0$) according to our significance threshold ($|Z|>3.3$), we observed that *D* was positive across most tests, thus suggesting that USR1 and Trail Creek share a larger proportion of alleles compared to other Native Americans (Fig. S22). We attribute the observed large standard errors to the low depth of the Trail Creek genome (~0.4X), which was ultimately limited by the little available material and low endogenous DNA fraction in the Trail Creek tooth (Data pre-processing, mapping, genotyping and phasing). In (Admixture graph fitting using qpGraph) we formally test a

model including SNA, NNA and USR1 and show that USR1 and Trail Creek form a clade to the exclusion of other Native Americans.

Clustering analyses (Multidimensional scaling, Admixture, f_3-statistics) showed that the **Big Bar** genome does not represent an 'SNA population'. Yet, although it clusters adjacent to NNA, it carries a slightly increased proportion of an SNA-characteristic ancestry component compared to other Pacific Northwest ancient genomes (Admixture, Figure 1F), and is separated from NNA in higher MDS dimensions (Figs. 1D-E, S16). Furthermore, *D*-statistics of the form *D(Aymara, Native American; Ancient, Yoruba)* suggest that Big Bar is equally related to the SNA Aymara and NNA populations (*e.g.*, Algonquin and Athabascans) (Fig. S20).

To further explore the relationship between Big Bar and other Native Americans using a larger number of informative sites, we computed *D*-statistics of the form *D(Athabascan, Nat. Am.; Big Bar, Yoruba)* using the whole genome dataset (Reference datasets). In addition, we computed *D(USR1, Nat. Am.; Big Bar Yoruba)* using the SNP array dataset (Reference datasets). Based on these tests, we were not able to reject the null hypothesis that Big Bar is equally related to Athabascans and other Native Americans (Fig. S23). This pattern excludes 'Ancient Beringians' (as represented by USR1), as we found Big Bar to form a clade with other Native Americans to the exclusion of USR1 (Fig. S23) (*1*). While other ancient and present-day Pacific Northwest populations have been found to bear recent Siberian-related admixture (*1*, *2*, *18*, *33*), we observed that Big Bar most likely does not carry this genetic signature (*D(Athabascan, BigBar; Koryaks, Yoruba)*=0.021, *Z*=3.59; *D(939, BigBar; Koryaks, Yoruba)*= 0.012, *Z*=1.644). While we did not find the test second test (including 939) to be statistically significant ((|*Z*|<3.3), we consider that the sensitivity of this test is reduced due to 1. the low depth of both ancient samples (Tables S8,12); and 2. the higher error rate in 939 due to the fact that Big Bar is partially USER-treated (Ancient DNA sequence data quality control) while 939 is not (*2*). These results suggest that Big Bar represents a population that diverged from other Native Americans prior to the NNA/SNA divergence, but after 'Ancient Beringians' (*1*). We formally test this model in (Admixture graph fitting using qpGraph), using an approach based on admixture graph fitting.

MDS and *ADMIXTURE* analyses suggest that the relationship between the Kennewick and ancient Southwestern Ontario (**ASO**) genomes and the NNA and SNA groups is complex. Both individuals were placed in an intermediate position between both groups and they carry a larger proportion of SNA-characteristic admixture components compared to NNA populations (Admixture, Figs. 1F,S17). In (Admixture graph fitting using qpGraph), we formally test different models involving these genomes and conclude that both individuals most likely derive from admixture between the two basal Native American branches (NNA and SNA). Although the placement of Kennewick is contentious due to low depth and potentially increased contamination (*10*, *16*), we used error-corrected *D*-statistics (*23*) to confirm the hypothesis that ASO bears ancestry from both NNA and SNA. We interpret the following results in the context of recent findings suggesting that present-day populations south of mid-latitude North America derive from the admixture between two ancestral sources: one represented by the Anzick1 genome and one represented by the ASO population (*7*).

We computed *D*-statistics of the form *D(ancient SNA, present-day SNA; ASO, Yoruba)*, where ancient South Native American groups include the Anzick1, Spirit Cave and Lagoa Santa individuals and the present-day SNA include the Mixe (n=3), Aymara (n=1), Karitiana (n=5) and Surui (n=2). We note that no NNA whole genomes without recent European- and/or Siberian-related admixture are available to date (*2*, *18*, *190*). Therefore, we test the null hypothesis that ASO is equally related to all tested SNA groups (*D*=0) and expect *D* to be inconsistent with 0 if ASO shares a significantly larger proportion of alleles with any of the tested SNA groups. We caution that *D* is expected to be consistent with 0 under a model where ASO is admixed between NNA and SNA, if the SNA group that contributed to the ancestry of ASO is an outgroup to all tested SNA groups. Although we were not able to reject the null hypothesis that ASO is equally related to Anzick1 and present-day populations (|*Z*| between 0.09 and 1.61), we observed that the Lagoa Santa population is most closely related to ASO (|*Z*| between 0.87 and 3.7), followed by the Spirit Cave genome (|*Z*| between 0.5 and 2.49) (Fig. 2F). Together with the best-fitting model presented in (Fig. 2F), these tests suggest that ASO carries admixture from a SNA population most closely related to Lagoa Santa and Spirit Cave that split after Anzick1. Thus, we consider that the apparent ASO-related ancestry in South Americans (*7*) could be explained by shared ancestry between Spirit Cave and/or Lagoa Santa and ASO. In addition, we interpret that the excess allele sharing between ASO and Lagoa Santa compared to present-day South Native Americans is likely the consequence of Mesoamerican-related admixture in such groups. In (Admixture graph fitting using qpGraph) and below, we confirm that SNA groups in South America carry ancestry from two Native American sources; namely, a Mesoamerican-related group and a population most closely related to Lagoa Santa.

MDS analyses showed that the Early Holocene **Spirit Cave and Lagoa Santa** individuals are most closely related to geographically nearby ancient and present-day populations. Yet, as expected from their age, higher MDS dimensions suggest that these individuals (as well as Anzick1) represent early populations that share little genetic drift with present-day populations (Multidimensional scaling, Admixture, f_3-statistics). In (Admixture graph fitting using qpGraph) we infer that the Lagoa Santa and Spirit Cave genomes can be modeled as a clade to the exclusion the Mesoamerican Mixe. We confirmed these results by computing 'standard' and error-corrected *D*-statistics including the present-day Mixe, and the Anzick1, Spirit Cave and Lagoa Santa ancient genomes. For 'standard' *D*-statistics, we only considered the high-depth Lagoa Santa individual for which genotype calling was feasible (Data pre-processing, mapping, genotyping and phasing) and for error-corrected *D*-statistics we considered both the high-depth Lagoa Santa genome and a pool of the five Lagoa Santa individuals. For all tests we used the whole genome dataset (Reference datasets) and restricted the analysis to transversion polymorphisms. Results for all tests are presented in Figs. S24 and S25. In brief, *D*-statistics of the form *D(Anzick1/Spirit Cave, Mixe; Lagoa Santa, Yoruba)* and *D(Mixe, Lagoa Santa; Spirit Cave, Yoruba)* suggest that Lagoa Santa shares a larger proportion of alleles with Anzick1 and Spirit Cave, than with Mixe. Although the only test that produced a statistically significant result was *D(Spirit Cave, Mixe; Lagoa Santa, Yoruba)*, we note that the three tests yield a sign of *D* supporting this inference. Furthermore, based on *D*-statistics of the form *D(Anzick1/Spirit Cave, Lagoa Santa; Mixe, Yoruba)*, we were not able to reject the null hypothesis that Mixe is equidistant to the ancient genomes.

Importantly, these results are also consistent with a scenario in which Mixe bears admixture from an outgroup (Admixture graph fitting using qpGraph). While a model without non-SNA admixture into Mixe is a reasonable fit to the data, we show that a model where Mixe carries 'unsampled admixture' from an outgroup is a significantly better fit (Admixture graph fitting using qpGraph). Furthermore, joint site frequency spectrum-based inference supports a model where the Mesoamerican Mixe carry admixture from a group closely related to Native Americans (Demographic inference).

Various hypotheses regarding genetic continuity and population replacement in the **North American Great Basin** have been proposed. In order to assess the relationship between groups that inhabited this region at different times, we obtained whole genome data from six individuals that lived in the region during the Early Holocene (two Spirit Cave individuals) and the Late Holocene (four Lovelock individuals) (Fig. 1A,B). Both MDS and *ADMIXTURE* analyses suggest that these individuals are closely related. In order to confirm that the more recent Lovelock individuals form a clade with the Early Holocene Spirit Cave individuals to the exclusion of other Native Americans, we computed *D*-statistics of the form *D(Lovelock, Spirit Cave; Native American, Yoruba)* using the whole genome dataset (Reference datasets). For this test we considered the Spirit Cave, Lovelock2 and Lovelock3 high-depth genomes separately and pooled the data for the two low-depth Lovelock ('1' and '4') genomes. For all Lovelock individuals we were not able to reject the null hypothesis that they form a clade with the Spirit Cave individual to the exclusion of other Native Americans. In addition, we observed that the Lovelock individuals are most closely related to each other, relative to other populations (Fig. S26). In (Admixture graph fitting using qpGraph) we infer that while both high-depth Lovelock individuals are most closely related to Spirit Cave, the younger Lovelock3 individual (~0.7 ka) also bears ancestry from a Mesoamerican-related population.

Admixture graphs optimized using *TreeMix* (*24*) and *qpGraph* (*4*, *21*) suggest that **present-day South American groups derive from an admixture event between a population most closely related to Lagoa Santa and a Mesoamerican-related source**. We confirmed these results by computing the three possible *D*-statistics involving the Mesoamerican Mixe, the ancient Lagoa Santa and various SNA groups, using the Yoruba as an outgroup. For 'standard' *D*-statistics, we only considered the high-depth Lagoa Santa individual for which genotype calling was feasible (Data pre-processing, mapping, genotyping and phasing) and for error-corrected *D*-statistics we considered both the high-depth Lagoa Santa genome and a pool of the five Lagoa Santa individuals.

In agreement with MDS and *ADMIXTURE* analyses suggesting that South American groups are closely related to the Lagoa Santa genomes, we observed that *D(SNA, Mixe; Lagoa Santa, Yoruba)* yielded positive results for South American populations. However, *D(SNA, Lagoa Santa; Mixe, Yoruba)* was also positive for such populations, suggesting that one of these groups carries admixture from the other two (Figs. S27,28). Ancient Patagonians are an exception, as we were not able to reject the null hypothesis that they form a clade with the Lagoa Santa population to the exclusion of the Mesoamerican Mixe. We consider that these results are unlikely to be substantially biased due to 'artificial' correlation between error patterns in ancient genomes (*206*) as we observed low- and high-depth ancient genomes with and without a

signal for Mesoamerican-related admixture (*e.g.,* Ayayema vs Taino and Punta Santa Ana vs Aconcagua). Moreover, we observed qualitatively similar results for error-corrected *D*-statistics as well as random allele- and genotype-based 'standard' *D*-statistics. In (Admixture graph fitting using qpGraph) we infer the best-fitting admixture graphs for each of these SNA groups.

Although 'standard' and error-corrected-*D*-statistics follow a similar trend, we only observed significant deviations from *D*=0 when we computed error-corrected-*D*-statistics using the five Lagoa Santa individuals. While this is expected due to the amount of data taken into account in each case, we carried out a simulation study in order to explore the likelihood of observing statistically significant results with the data at hand. We simulated data under the model shown in Fig. S29 using *msprime* (*209*). We assumed fixed mutation and recombination rates $\mu = 1.25 \times 10^{-8}$ and $\rho = 10^{-9}$. Constant population sizes and split times were set to N_{afr} = 20,000 for Africa, N_B = 1,600 for the out of Africa population, N_{Mixe} = 1,500 for Mixe, N_{LSta} = 1,500 for Lagoa Santa, and N_{SA} = 1,500 for South Americans, T_{afr} = 72 ka for Africa, T_{Mixe} = 13 ka for Mixe, following refs. (*1*, *2*), and sampled the Lagoa Santa individual S_{LSta} = 10 ka. We tested all possible combinations of a series of divergence times between Lagoa Santa and South Americans, T_{div} = {8,500, 10,500}, different times of admixture in South America, T_{adm} = {2,500, 5,000, 7,500}, along with admixture proportions of P_{adm} = {0.2, 0.4, 0.6, 0.8 }. We simulated ten replicates of each combination of the last three parameters. In order to mimic the real data (Reference datasets), simulations included eight diploid samples: two Africans, two South Americans, three Mixe and one ancient sample 10 ka (Lagoa Santa). For each individual we simulated 22 autosomes and downsampled the simulated data to include the same number of SNPs found in Yoruba, Lagoa Santa, Mixe and Surui. We computed the three possible *D*-statistics involving Lagoa Santa, South Americans and Mixe, based on the simulated data (Fig. S30,31). While most statistics followed the expected trend for each test, we did not observe statistically significant results ($|Z|$>3.3) in all cases. For example, *D(SouthAmerican, LagoaSanta; Mixe, Yoruba)* yielded results with $|Z|$<3.3 when admixture was recent (2.5 and 5 ka) and when the admixture proportion was low (20 and 40%). These results suggest that the data at hand is not necessarily sufficient in order to produce statistically significant deviations from *D*=0. We note, however, that model testing in (Admixture graph fitting using qpGraph, and Demographic inference) supports a model where present-day South Americans derive from a mixture of a Lagoa Santa- and a Mixe-related population.

Recent studies have shown that some **Amazonian populations share a significantly larger proportion of alleles with Australasians** (Australians, Papuans and island Southeast Asians, Fig. 4C), compared to other Native Americans (*2*, *6*, *16*). While this signal has been suggested to be potentially linked to the distinctive **'Paleoamerican'** skull morphology observed in early human remains, genetic analyses have failed to support such link (*2*, *10*). We leveraged the high-depth whole genome data from two Early Holocene human remains that have been identified as 'Paleoamericans' (Spirit Cave and Lagoa Santa) (Samples, Figure 1), to explore the geographic and temporal extent of this genetic signature. We computed all possible *D*-statistics of the form *D(Nat.Am., Nat.Am.; Eurasian, Yoruba)*, where the Native American groups include all ancient and contemporary Native Americans with no recent Eurasian admixture, for which high-depth whole genome data is available (Fig.

S19); and the Eurasian groups include all ancient and present-day Eurasians in the whole genome dataset (Reference datasets).

For most tests, we were not able to reject the null hypothesis ($|Z|>3.3$) that pairs of Native American groups are equally related to Eurasian populations, except for the following comparisons:
· Most Native Americans are more closely related to Rapanui and Siberian Eskimo-Aleut speakers compared to USR1. These results are expected under the assumption that USR1 represents a population that diverged early from other Native Americans (*1*). Hence, the Native American populations that contributed to the formation of these Siberian and Polynesian groups are more closely related to Native Americans south of Alaska than they are to the early-splitting USR1 (*1*, *210*).
·Athabascans are more closely related to Siberian groups compared to other Native Americans. Similarly, this result is expected as many Na-Dene speakers from the Pacific Northwest have been found to bear recent Siberian-related admixture (*1*, *18*, *33*).
· Amazonian groups such as the Surui and Karitiana are more closely related to Australasians (Australians, Papuans and island Southeast Asians, Figs. 4A,B, S32,33) than are Mesoamerican groups such as the Mixe and the Huichol. This result recapitulates previous findings of excess allele sharing between Amazonians and Australasians (*2*, *6*).

While we were not able to reject the null hypothesis that Spirit Cave and other Native Americans are equally related to Eurasians, we observed that Lagoa Santa yielded similar results to those obtained for the Amazonian Surui. When compared to Mesoamerican groups such as the Mixe and Huichol, we found that Lagoa Santa shares a significantly larger proportion of alleles with Australasian groups ($|Z|$ between 2.17 and 3.86), but not with other Eurasians (Figs. 4A,C, S32,33). Furthermore, we compared the distribution of the observed Z-scores to the expected normal distribution under the null hypothesis and observed that the majority of the outliers corresponded to tests of the form *D(Surui/Lagoa Santa, Mixe; Australasian, Yoruba)* (Fig. S34). In order to minimize the potential bias caused by post-mortem damage in the Lagoa Santa genome, we performed these tests using transversion polymorphisms only and restricted the analysis to the high-depth Lagoa Santa genome for which genotype calling was feasible (Data pre-processing, mapping, genotyping and phasing).

We note that excess error in Lagoa Santa would result in a decreased signal as a consequence of artificial correlation between Lagoa Santa and the outgroup (*206*). We validated these results by computing error-corrected D-statistics of the form *D(Nat.Am., Nat.Am.; Eurasian, Yoruba)* using a set of representative Eurasian and Native American populations. Based on error-corrected D-statistics, we observed similar patterns of excess allele sharing between Lagoa Santa and Australasians, compared to the Mesoamerican Mixe (Fig. S35A). In addition, we confirmed these results by comparing D-statistics of the form *D(Lagoa Santa, Mixe; Eurasian, Yoruba)* obtained separately for the high-depth Lagoa Santa genome (called genotypes) and a pool of the four low-depth Lagoa Santa individuals (random allele). Although D was systematically shifted towards negative values when we considered the low-depth Lagoa Santa individuals, likely as a consequence of increased error, we observed that tests involving Australasians yielded the highest values of D for both Lagoa Santa versions (Fig. S35B).

Additionally, we computed 'contamination-corrected' f_4 statistics to explore the effect of European contamination (<3% in the high-depth Lagoa Santa genome) on the observed signal. In brief, we assumed that the observed f_4 statistics are a linear combination of f_4 obtained for the ancient individual assuming no contamination and f_4 obtained for the 'contaminant', weighted by the contamination fraction c (*4*). Thus, for $f_4(H_1, Ancient; H_3, H_4)$, we write the contamination-corrected statistic as:

$$f_{4corr}(H_1,Ancient;H_3,H_4)=f_{4obs}(H_1,Ancient;H_3,H_4)-cf_4(H_1,Contaminant;H_3,H_4)/(1-c)$$

We computed 'contamination-corrected' f_4 statistics of the form $f_{4corr}(Mixe, Lagoa Santa; Australasian, Yoruba)$ where Australasians are represented by Papuans, Onge, Australians from the West Central Desert and the historical Andaman islander (Figure 4B). We used the French as a proxy for the contaminant and assumed different values of the contamination fraction $c=\{0,0.005,...,0.1\}$. We used the whole genome dataset (Reference datasets) and restricted the analysis to transversion polymorphisms. For all tests, we observed that increasing the assumed contamination fraction produced more negative values of f_4. Thus, we conclude that the observed excess allele sharing between Lagoa Santa and Australasians is most likely not a consequence of European contamination, which would, in any case, dim the observed signal.

Altogether, these results suggest that the previously documented Australasian genetic signature was present in South America during the Early Holocene. However, this signal is not necessarily correlated with the so-called 'Paleoamerican' skull morphology, as we were not able to detect such signature in the Spirit Cave genome.

Admixture graph fitting using Treemix

We built admixture graphs using *TreeMix* (*24*) with two objectives. First, we used *TreeMix* as a complementary approach to that detailed in (*D*-statistics), to identify South Native American individuals with non-American admixture. Second, we used *TreeMix* to explore the relationship between ancient and contemporary SNA populations. For each test described below, we considered a different subset of the whole genome dataset (Reference datasets) and only included transversion polymorphisms where all analyzed populations had at least one individual with non-missing data. For each dataset, we grouped SNPs into 5Mb blocks to account for linkage disequilibrium and fitted admixture graphs with zero to five admixture edges after disabling the built-in sample size correction procedure (-noss parameter) and allowed for a final rearrangement of the tree (-global parameter). For each number of migrations we ran 1000 replicates with different starting seed values and picked the run with the highest likelihood.

In (*D*-statistics) we identified a set of **SNA whole genomes with no detectable non-American admixture** using *D*-statistics. We used *TreeMix* as a complementary approach to further refine this set. To do so, we compiled a dataset including the individuals identified in (*D*-statistics) as well as three outgroups (the Han Chinese, the French and the Yoruba). Initially we considered all Native American genomes individually and ran *TreeMix* assuming no admixture edges in order to determine whether Native American individuals could be pooled according to their population label (Fig. S36a). Overall, we observed that individuals from the same population formed a clade to the exclusion of other populations with the exception of the Quechua. Thus, in the following analyses, we pooled individuals from the same population with

the exception of the Quechua, which we considered separately. We then ran *TreeMix* on the dataset with 'pooled' populations assuming zero to five admixture edges, and found that the best fitting models included admixture edges relating the French with one Zapotec (LP6005443-DNA_A12) and one Quechua individual (LP6005519-DNA_G02) (Fig. S36B-D). Therefore, we also excluded these individuals from further analyses.

We merged the set of 'non-admixed' SNA populations defined above with all available high-depth ancient Native American genomes (Anzick1, USR1, Spirit Cave, Lagoa Santa, Ayayema, Taino, Lovelock2 and Lovelock3) and a set of outgroups (the Han Chinese, the historical Andaman islander and the Yoruba). We ran *TreeMix* assuming zero to five admixture edges to explore the **relationships between these populations**. The tree with zero admixture edges broadly recapitulated the geographic structure of the tested populations (Fig. S37A). However, the best-fitting graphs with admixture involved different placements of Anzick1, Spirit Cave, Lagoa Santa and Ayayema. When these individuals were placed in different clades of the tree, the model included a heavy admixture edge relating these individuals (Fig. S37B,C). However, graphs with more than one admixture edge, placed these individuals in the same clade, together with present-day populations from South America (Fig. S37B,C). In addition, all admixture edges were placed between Native American groups except for the third edge, which goes from the historical Andaman islander into the base of the (Lagoa Santa, Ayayema) clade (Fig. S37C). Together with (Multidimensional scaling and f_3-statistics), these results suggest that the ancient Anzick1, Spirit Cave and Lagoa Santa genomes are closely related, but that this 'link' is not evident in Mesoamerican populations such as the Mixe. Moreover, these results support that the Lagoa Santa population carries a small proportion of Australasian-derived admixture (in this case represented by the historical Andamanese). Importantly, we observed that the ancient Taino genome did not follow the same trend as other ancient genomes and thus consider that these results are unlikely to be substantially biased by the fact that ancient and present-day genomes are analyzed jointly. In (Admixture graph fitting using qpGraph), we formally test different models involving these populations.

Admixture graph fitting using qpGraph

We implemented an *f*-statistics-based admixture graph search strategy in order to model the ancestry of ancient and present-day Native Americans. In particular, we used this approach to explore 1. the phylogenetic placement of the ancient genomes from North America with respect to Ancient Beringians (AB), North (NNA) and South Native Americans (SNA); 2. the relationship between the Late Pleistocene and Early Holocene Native American genomes with respect to the Mesoamerican Mixe; and 3. the formation of present-day South Native American populations. In each case we start by building a 'seed graph' that is a reasonable fit to a small set of key populations from different broad continental ancestries. Then, for a given test population, we use the *admixturegraph* R package (*32*) to enumerate all possible extensions of the seed graph, where the test population is added as either a **'non-admixed'** or an '**admixed**' leaf. For the former, we consider all graphs where the test population branches off a given internal or external edge of the seed graph. For the latter, each extended graph corresponds to one of the possible pairs of edges of the seed graph that will contribute to the ancestry of the test population. We then used *qpGraph* to optimize the parameters (branch lengths and admixture proportions) for each extended graph and evaluate each topology based on 1. its fit score (log-likelihood), 2. the *Z*-score of the worst residual between the observed and predicted *D*-statistics, and 3. the presence of zero-length

internal edges (*30*). Following ref. (*30*), for a given test population, we assess whether the 'admixed' topology is a significantly better fit to the data than the 'non-admixed' topology through a regular likelihood-ratio test, where fit score differences of ~3 and ~4.6 correspond to *p*-values of 0.05 and 0.01 respectively. In addition, when comparing topologies with similar fit, we favor the model without zero-length internal edges.

We explored the ancestry of the ancient **North American genomes** using a 'seed graph' including the following leaves: Yoruba (representing Africans), Mal'ta (Ancient North Eurasians), historical Andaman (Australasians), Han (East Asians), USR1 (Ancient Beringians), Athabascan (North Native Americans) and Spirit Cave (South Native Americans) (Fig. S38). In this graph, we included Native Americans as an admixture between East Asians and Ancient North Eurasians (*31*), and modeled Mal'ta (*31*) as an admixed individual following ref. (*30*). Since we did not include west Eurasians in this graph, we fixed the admixture proportions to those estimated in ref. (*30*). In addition, we included Athabascans as a mixture of a Native American branch (representing North Native Americans) and a population most closely related to East Asians (*1*, *4*, *18*, *33*). We used *qpGraph* (*21*) to fit the seed graph (Fig. S38) to the whole genome dataset (excluding transition polymorphisms) and consider it to be a reasonable fit as the largest *D*-statistic residual corresponds to a *Z*-score of ~1.93.

In agreement with previous sections, we observed that **Trail Creek** forms a clade with USR1 in the best-fitting 'non-admixed' model (Fig. S39). We deem this model (without zero-length internal edges) a reasonable fit to the data as the largest *D*-statistic residual corresponds to a *Z*-score of ~1.6. Moreover, the models where Trail Creek is admixed do not represent a significantly better fit (*p*~0.57). Together with (Multidimensional scaling, Admixture, f_3-statistics, and *D*-statistics), this result suggests that Trail Creek forms a clade with USR1, thus supporting conclusions from (*1*).

D-statistics suggest that the two contemporaneous and geographically nearby **939** and **Big Bar** genomes are differentially related to the North and South Native American branches. While the former forms a clade with present-day Athabascans and carries Siberian-related admixture, the latter does not carry such signal and most likely represents an outgroup to NNA and SNA. Including these genomes in the graph described in Fig. S38, further supported these inferences. In agreement with (*D*-statistics), we were not able to reject a model in which 939 forms a clade with Athabascans (*p*~0.33) (Fig. S41). This result is reproduced in the 2.5 ka 302 British Columbia genome (*p*~0.09) (Fig. S42), but not in the 1.75 ka 443 genome for which an admixed model is better supported (*p*~0.0078) (Fig. S43). In contrast, the best-fitting model (without zero-length internal edges) places Big Bar as an outgroup to NNA and SNA. Importantly, alternative placements of Big Bar in other Native American branches yielded zero-length internal edges or *D*-statistic residuals with $|Z|>3.3$. Moreover, the best-fitting 'admixed' model for Big Bar did not represent a statistically significant likelihood improvement compared to the 'non-admixed' model (*p*~0.099) (Fig. S40). These results suggest that the population history of the Pacific Northwest is complex and involves different episodes of isolation and admixture in geographically nearby locations.

The ancient **ASO** and **Kennewick** genomes have been associated with Native American branches that differ from the more widely spread South Native Americans. ASO have been suggested to be representatives of a population whose split from other

Native Americans antedates that of Anzick1 (*3*, *7*), and the placement of Kennewick remains contentious (*10*, *16*). Interestingly, allele-sharing and *f3*-distance-based MDS place both the ASO and Kennewick genomes in an intermediate position between North and South Native Americans. Moreover, *D*-statistics suggest that ASO bear admixture from a South Native American population. We used the admixture graph search procedure to formally test the relationship between the ASO and Kennewick genomes and other Native Americans. For Kennewick, we did not find 'non-admixed' models without zero-length edges and *D*-statistic residuals with $|Z|<3.3$. In contrast, 'admixed' models without zero-length edges all resulted in a statistically significant likelihood improvement ($p<0.0056$) (Fig. S44). We note that the best-fitting admixed models yield a similar likelihood; therefore we cannot exclude any of the models. However, based on (Multidimensional scaling, Admixture, f_3-statistics) we favor a model in which Kennewick derives from a mixture of SNA and a previously diverging population (likely NNA). Similarly, we found that the best-fitting model without zero-length internal edges for ASO is significantly better than the 'non-admixed' model (p~2.02e-6) (Fig. S45). This topology suggests that ASO derives from a mixture between two deep-splitting Native American populations whose divergence postdates that of USR1. These results are further supported by *D*-statistics and suggest that although South Native Americans likely expanded south of the glacial ice swiftly after diverging from NNA, northern North America remained a region where these two branches exchanged migrants at different times.

Together with **Anzick1** (*3*)**, Spirit Cave** and **Lagoa Santa** represent the earliest SNA individuals for which high coverage whole genome data have been produced to date. We fitted admixture graphs using *qpGraph* to formally test the relationships between these individuals and the Mesoamerican **Mixe**. The latter have been previously used as a reference population for present-day SNA (*6*) and have been shown to derive from one of the earliest splits off that basal branch (*4*). In (Multidimensional scaling, Admixture, f_3-statistics) we show that these individuals can be confidently assigned to SNA, thus we considered a seed graph similar to that described in Fig. S38, excluding USR1 and Athabascans, and where South Native Americans are represented by a (Mixe, Anzick1) clade (Fig. S46A). We found that this model is a reasonable fit to the data as the largest *D*-statistic residual has $|Z|<3.3$ and no zero-length internal edges. We then extended this graph by adding Spirit Cave as a 'non-admixed' leaf, and found that the only model for which we did not observe *D*-statistic residuals with $|Z|>3.3$ and/or zero-length internal edges includes a (Spirit Cave, Anzick1) clade (Fig. S46B). We repeated this procedure for Lagoa Santa, but in this case we forced Lagoa Santa to derive a fraction of its ancestry from a population most closely related to the historical Andaman islander, following results presented in (*D*-statistics). In this case there were two topologies with similar likelihood scores and residuals, without zero-length internal edges; both of which place the Mesoamerican Mixe as an outgroup to other South Native Americans (Fig. S46C).

A more sensitive approach based on the joint site frequency spectrum (Demographic inference) showed that the most likely model is one in which the Mixe form a clade with Lagoa Santa, but carry admixture from an 'unsampled population' whose divergence predates that of South Native Americans. To differentiate between these competing models, we fitted the model shown in Fig. S47A to the data using *qpGraph* and fixed the 'unsampled admixture' proportion into Mixe to a set of values ranging between 0 and 1. Models where the Mixe are admixed involve a significant

improvement in likelihood (p<0.05), compared to models where the Mixe are 'non admixed'. We note that we have little resolution to determine the admixture proportions as different values between 5 and 95% yield similar likelihoods (Figs. 3B, S47B).

To assess the **position of the 'unsampled population'** that contributed a fraction of the Mesoamerican Mixe ancestry, we considered a graph similar to that shown in Fig. S47, but that includes Athabascans as a representative of North Native Americans. We modeled Mixe as a mixture of the Native American component of Lagoa Santa, and explored the likelihood of the 'unsampled admixture' to derive from all possible Native American branches above the divergence of Anzick1. We found that the only model yielding no zero-length internal edges was that in which the 'unsampled population' is placed above the NNA-SNA split (Fig. S48). Importantly, we consider that the 'unsampled admixture' in Mixe is unlikely to come from a non-Native American branch as *D*-statistics of the form *D(Mixe, Spirit Cave; Eurasian, Yoruba)* (*D*-statistics) do not support Eurasian admixture in Mixe.

We formally tested **'non-admixed' and 'admixed' models for ancient and present-day SNA**, where we extended the graph detailed in Fig. S47. For most South American groups, the 'admixed' model involving source populations represented by the Mixe and Lagoa Santa is a significantly better fit than the 'non-admixed' models where they form a clade with either group (Figs. S53-62). Ancient Patagonians such as Ayayema and Punta Santa Ana do not follow this trend. For these genomes, the 'admixed' model is not a significantly better fit, relative to a model where they form a clade with Lagoa Santa (Figs. S63-68). These results suggest that after initial settlement by a population most closely related to Lagoa Santa, groups in South America received gene flow from a Mesoamerican-related population. Yet, results from the Ayayema individual suggest that the arrival of the Mesoamerican-related ancestry postdates 5.1 ka, and/or that it did not reach the remote and isolated region inhabited by the ancestors of the Ayayema individual.

While the Lagoa Santa-Mixe admixture pattern is widespread in South American groups, the best-fitting models for both high-depth **Lovelock** individuals and the present-day **Pima** and **Huichol** differ. For both Huichol and Lovelock2 we were not able to reject the 'non-admixed' models. For Huichol, we were not able to reject the model with a (Mixe, Huichol) clade (p~0.2) (Fig. S52). In the case of Lovelock2, we were not able to reject a simple model where Lovelock2 and Spirit Cave form a clade (p~0.3) (Fig. S49). In contrast, the best-fitting model for Lovelock3 involves a mixture between a branch most closely related to Spirit Cave and a branch most closely related to the Mixe (p~0.004) (Fig. S50). Similarly, the ancestry of the Pima is best represented by a mixture between a SNA component and a Native American population that diverged prior to the split of Anzick1 (p~0.014) (Fig. S51). These results suggest that there was a high degree of genetic continuity in the great basin between ~10 and ~2 ka (the ages of Spirit Cave and Lovelock2). However, Mesoamerican-related admixture entered the region prior to the time in which Lovelock3 lived, but importantly this movement did not result in complete population replacement. Similarly, we observe that different populations interacted in the North American Southwest, where Athabascan groups are estimated to have arrived in the region <1 ka.

We further explored the **fit of the 'admixed' model for South Americans** where they derive from a mixture of a Lagoa Santa- and Mixe-related population. We

fitted the graph in Fig. 3A to South Americans with varying values for the Mesoamerican admixture proportion into South Americans {0,0.05,...,1} and the Australasian contribution into Lagoa Santa {0,0.01,...0.1}. In agreement with Figs. S53-68 and (*D*-statistics), we found that the 'admixed model' with non-zero Australasian-related admixture into Lagoa Santa is a significantly better fit (p<0.05) for most South American groups, except for ancient Patagonians such as Ayayema (Fig. S69). For the latter, we did not observe a significant increase in likelihood when considering models with non-zero Mesoamerican admixture. Additionally, we observe consistent results across South American groups when fixing the Australasian contribution into Lagoa Santa to ~3% (maximum likelihood estimate from qpGraph) (Fig. S70). These results show that while we can reject the 'non-admixed' model for most South Americans, we cannot estimate the Lagoa Santa and Mixe-related admixture proportions with precision. We highlight that these results are unlikely to be substantially influenced by ancient DNA *post-mortem* damage. In addition to excluding transitions from this analysis, we note that different ancient genomes with similar quality show different patterns, *e.g.,* Ayayema forms a clade with Lagoa Santa while the ancient Taino genome is best-modeled as a 'Mixe-Lagoa Santa' mixture.

As a control, we sought to **replicate these results using the Human Origins SNP array dataset** (*6*, *21*). In this case, we included SNP array data for the Yoruba, Han, Mixe and Surui and substituted the historical Andaman for present-day Onge. We merged these data with the called genotypes for Anzick1, Spirit Cave, Lagoa Santa, and Ayayema and sampled a random allele at each position for the low-depth Mal'ta genome. In agreement to the results obtained using the whole genome sequencing dataset, we found that the Mixe are best modeled as deriving a fraction of their ancestry from an 'unsampled population' (Fig. S47), we were unable to reject the (Lagoa Santa-Ayayema) clade, and we found evidence supporting an 'admixed' origin of the Suruí (Fig. S71).

Demographic inference

We **inferred past population sizes** for a subset of populations using *smc++* (version 1.11.1) (*27*), using the data containing all mutations as well as using only transversions. We used the default parameters for running *smc++,* using the '--r' flag to fix the recombination rate to 1.2×10^{-8} per-base per-generation and fixing the per-base per-generation mutation rates to be 1.25×10^{-8} and 4.16×10^{-8} for all mutations and transversions only respectively (*211*, *212*). For modern samples, we filtered out sites where any modern individual in our populations had fewer than ten reads in addition to all sites that failed either the 1000 Genomes strict filter (*69*) or Heng Li's 'SNPable' filter following ref. (*1*). For all other samples, we used the SNPable and strict filters, but restricted to sites where there were at least ten reads in that particular sample. The results, presented in Fig. S72, are mostly unchanged whether using all mutations or only transversions, especially for modern samples or USER-treated ancient samples (e.g. Spirit Cave) but differ substantially for non-USER-treated ancient samples.

We now describe the site frequency spectrum **(SFS)-based demographic inference** procedure. The SFS was restricted to biallelic sites with transversion mutations to ensure data quality associated with ancient DNA decay. To improve statistical power, the data was polarized by chimp genome computed from the human-chimp alignment on the UCSC Genome Browser (*129*). Sites of the genome were further restricted to only those where the human genome and the chimp genome aligned. In addition, sites

were masked as before, using Heng Li's SNPable mask and the 1000 Genomes Strict mask files as well as sites which did not have a minimum of read depth 10 in the modern genomes. All procedures regarding missing data were handled by *momi2* (*25*, *26*) directly.

The strategy that we used for inferring the backbone topology was to first use f-statistics to determine the topology of the samples which were unlikely to contain admixture (e.g. mostly ancient samples). Then, proceed by incorporating additional populations sequentially.

First, we determine the topology of three ancient samples which were likely to experience minimal admixture due to sampling time (Spirit Cave, Lagoa Santa, and Anzick) along with Mixe. We first verify via f4 statistics (*21*) that the Spirit Cave, Lagoa Santa, Anzick, and Mixe samples form a tree topology as shown in Table S13. While one may expect at least two of the observed f4 statistics to be large enough to reject the corresponding null unrooted tree topology, notice that we do not have enough statistical power to reject any of the topologies. This may be due to the internal branch of the true underlying unrooted tree topology(denoted as e in Fig. S73) being rather small in terms of genetic drift (e.g., a star graph). In order to test which unrooted tree topology is correct among the three choices above, we use f3 statistics. F3 statistics provide us with estimates of $a, b, c, d,$ and e for each topology. Each external branch length can be estimated from two f3 statistics estimates. For example, a can be estimated via $\hat{a} = F3(B, C; A)$ or $\hat{a} = F3(B, D; A)$. We can similarly do this for b, c, and d. In order to derive estimates for e, we can estimate $\hat{a} + \hat{b} = F3(C, D; A)$, then solve for $\hat{e}$. By symmetry, we can derive 8 different estimates for $\hat{e}$ for any given topology as plotted in Fig. S74. We expect that all estimates of $\hat{e} \geq 0$ for the true topology, so we can conclude from the f3 statistics that ((S,A),(LS,M)) is the correct unrooted topology. Further evidence that this is the correct tree topology is that the variance of the estimates, $\hat{e}$, is much smaller than those of the other topologies. From here on out, we define the A, B, C, and D populations and their corresponding branches as Spirit Cave, Anzick, Lagoa Santa, and Mixe, respectively.

With an unrooted tree topology in hand, we attempt to derive the rooted tree topology. First, we must confirm that there is no admixture between the ancestral branch, and any of the four populations at hand. The F4(Lagoa Santa, Mixe; Spirit Cave, Ancestral) = 2.266 indicates that the proper rooting of the ancestral branch cannot occur on the *a,b,c,* and *e* branches conditioned on no admixture occurring since the test indicates Mixe and ancestral alleles are more similar than expected. F4(Lagoa Santa, Anzick; Mixe, Ancestral) = 1.698 is suggestive that there is further gene flow and that simply joining the ancestral branch to the *d* branch without modeling any additional gene flow is unlikely to be correct. In order to test all possible admixture events between the ancestral branch and the four populations of interest(Anzick, Lagoa Santa, Spirit Cave, and Mixe), we use qpgraph (*21*) to test all possible pairs of branches(one onto the tree root and another as an admixture event) in which the ancestral branch joined onto as shown in Table S14. The only two branches that have sufficiently small error in terms of least squares yield the same unrooted topology(Spirit Cave and Lagoa Santa; Anzick and Mixe) since *qpgraph* does not directly utilize the ancestral allele. In all other admixture graphs inferred, the F3(Lagoa Santa, Spirit Cave; Mixe) blew up. Explicitly adding the ancestral allele as the chimp population as an input into *qpgraph* yields a lower error for Anzick and Mixe(142976.800) over Spirit Cave and Lagoa

Santa(462987.703). Thus, we conclude that the correct rooted tree topology is shown in Fig. S75.

After fixing the topology of the ancient samples with respect to Mixe, we focus on incorporating additional present-day South Native American populations such as Karitiana, Surui, and Aymara. In order to reduce the complexity of the admixture graph topology, we focus on placing these three populations onto the existing graph topology one at a time instead of jointly. For clarity, we focus on the process of placing the Karitiana population onto the admixture graph; however, the below procedure was also run with Surui and Aymara and resulted in the same graph topology. We start by comparing the f4 statistics of Karitiana among the other populations as shown in Table S15. From the z-scores, we can rule out Karitiana joining onto Anzick and Spirit Cave. However, we are unable to reject Karitiana joining onto Lagoa Santa, Mixe, or the branch ancestral to both of them. Using *qpgraph*, we check all possible placements of Karitiana with and without any admixture events. We find that the most likely in terms of least squares score is that Karitiana is the result of admixture between the Lagoa Santa branch and the Mixe branch that comes after the ancestral admixture as shown in Fig. S76. We verify that this result is not due to overfitting in the *momi2* analysis later.

The ***momi2* fitting procedure** performs maximum likelihood inference to fit the optimization problem to the site frequency spectrum. Using previous results from (*1*), we incorporated the USR1 individual as an outgroup for the Native American populations, but fixed the USR1-related parameters (population size $N_U = 1375$ as inferred by *smc*++ and two admixture events with fixed pulse times set to 15 ka and 18 ka) instead of inferring them afresh. Note, however, that USR1 pulse probabilities were inferred in each model. From the f-statistics results in the previous subsection, we fixed the topology involving the Spirit Cave, Lagoa Santa, Karitiana, and Mixe populations. We choose Karitiana for this analysis due to the large sample size relative to Surui and Aymara for more accurate inference. A generation time of $g = 29$ years and a transversion-only mutation rate of $\mu = 3.85\times10^{-9}$ per bp/gen were assumed throughout. Initialization of the optimization procedure was obtained via inferred estimates from a combination of simpler inferred models in *momi2* and *smc*++ estimates. Ancestral population sizes were estimated using *smc*++ starting from 23 ka and older. We also fixed the population sizes of Lagoa Santa ($N_L = 670$) and Spirit Cave ($N_S = 1100$) to *smc*++ estimates, since high coverage data were available for only a single individual from each of these populations. The remaining population sizes, divergence times, admixture times, and pulse probabilities were all estimated unless otherwise stated.

We began our analysis with a basic 5-population model (called Model 1) without any of the admixture events inferred in the previous section; this model consisted of the USR1, Spirit Cave, Lagoa Santa, Karitiana, and Mixe populations. The only gene flow events included in the model were from USR1 into the ancestors of Native Americans, as inferred by (*1*). Fig. S77A illustrated the inferred model, which had a log-likelihood of –5395883.

Next, we tested for gene flow from the ancestral branch into Mixe, as was found in the previous subsection. In order to model gene flow from the ancestral branch into Mixe in a geographically realistic manner, we introduced an unsampled population, called 'UPopA', which branches off of the ancestral branch and admixes into the Mixe

population. We refer to this 6-population model as Model 2. The inferred model, shown in Fig. S77B, had a log-likelihood of –5395587, which is a substantial improvement over the the log-likelihood of Model 1. Also, we note that the pulse probability is sufficiently higher than 0, which recapitulates the results found by f-statistics.

Finally, we tested for gene flow from Mixe into Karitiana, as was found using *qpgraph*. In order to model gene flow from the geographically distant populations, Mixe and Karitiana, we introduced an additional unsampled population, 'UPopM', which branches off of Mixe and admixes into the Karitiana population. We refer to this 7-population model as Model 3. Our inferred model, illustrated in Fig. S77C, had a log-likelihood of –5395201. This is a substantial improvement over the log-likelihood of Model 2, in agreement with our results using the f-statistics. The inferred parameters can be found in Table S16 along with the corresponding parameters in Fig. S78. In testing Model 3, we found that additional pulses (Fig. S77D) improved the model fit, suggesting that there was likely continuous migration from 'UPopM' to Karitiana.

To quantify the uncertainty of our parameter estimates, we performed **nonparametric bootstrap** on the inferred model. This was done by dividing the data into 100 jackknife blocks from which the SFS was resampled and *momi2* inference was run afresh. Note that we fixed the topology and initialization while performing the bootstrap. This procedure was run for 400 bootstrap iterations with each bootstrap iteration comprising of 10 runs of momi2 for a given resampled dataset with the parameters corresponding to the best likelihood chosen as was done in the original inference procedure above. We removed any bootstrap iterations that converged to degenerate local minima resulting in 228 nondegenerate bootstrap iterations. The results from the bootstrap are shown in Fig. S79 and Table S17. For the most part, the mean of the bootstrap runs agree with our inferred values. The pulse probabilities, divergence times, and ancient sizes are most confidently estimated. The small bias and standard deviation of the pulse probabilities from the unsampled populations strengthen the evidence that admixture occurred from UPopA to Mixe and from UPopM to Karitiana. The parameters with the most variation are those related to the unsampled populations (UPopA and UPopM) which we expect to be difficult to estimate due to the lack of data from the source populations. In addition, since we expect unmodeled continuous migration from UPopM to Karitiana, this likely contributes to additional uncertainty and multiple local maxima in the likelihood surface.

Despite the difficulty in inferring the unsampled population parameters, we note that the temporal ordering of events is largely consistent. In particular, the pulse into Mixe from UPopA is more ancient than the divergence time of Mixe and UPopM in 167 out of 211 runs in which neither of these parameters degenerated to a time close to 0.

We also used a **coalescent-HMM method, diCal** (*28*), to get an independent estimate of several key features of the SFS-based model. We considered several pairs of populations: Aymara and Mixe; Karitiana and Mixe; and Surui and Mixe. To account for ancestry from the branch leading to Lagoa Santa as well as the ancestral admixture into Mixe, we fit the model in Fig. S80. We used the phased data, restricted to sites that pass Heng Li's SNPable filter and also the 1000 Genomes Strict filter (*69*). We used all mutations and assumed a mutation rate of 1.25×10^{-8} per-base per-generation and a recombination rate of 1.2×10^{-8} per-base per-generation. When fitting a model to a

pair of populations we used all individuals from those populations, using all chromosomes. To fit the model, we used a composite likelihood scheme where we took the product over all conditional likelihoods obtained by leaving out one haplotype at a time and computing the likelihood of that haplotype conditioned on all of the others. For computational reasons we enforced that each non-overlapping 1kb region have the same hidden state and we discretized time by having 8 break points placed log-uniformly between 580 years ago and 2.9 Mya. To aid in optimization, we assume a small amount of migration between all demes. To avoid local optima, we randomly initialized each run 15 different starting parameter values and performed 6 EM iterations for each initialization. We then selected the 3 best points, and used them to randomly generate a total of 10 new points, each of which were run for another 6 EM iterations. We report the final parameter values with the highest composite likelihood. This procedure is implemented in diCal, using the flags '--lociPerHmmStep 10000 --minimalMigration 0.001 --compositeLikelihood lol --intervalType logUniform --intervalParams 8,0.001,5 --numberIterationsEM 6 --metaNumIterations 2 --metaKeepBest 3 --metaNumPoints 10'. More details about the method can be found in (*28*).

The results of these inferences are presented in Table S18. They largely agree with the results from the SFS-based analysis. In particularly, both find that Karitiana, Aymara, and Surui are a mixture of an earlier migration and a population more closely related to Mixe. Both methods infer a divergence between Mixe and this later wave of around 5ka, and a divergence between Mixe and the earlier wave of around 15ka. Note that the results using Aymara are somewhat divergent from the results using Karitiana or Surui. This is because Aymara only receives about 5% of their genome from the earlier wave (compared to about 15% in Surui and about 50% in Karitiana). This, coupled with the small sample size for Aymara -a single diploid individual- results in less power to make inferences about the unsampled populations, particularly $T_{adm}, T_{LSAB}, T_{unadm}, T_{un}$.

Figure S1. Support letter for the genetic analysis of the Trail Creek remains, signed by a representative of the Native Village of Deering.

TERMS OF AGREEMENT BETWEEN FPST AND BLM

RECITALS

A. Thousands of years ago, the remains of several people were buried in their final resting place within Spirit Cave, Nevada. Spirit Cave is located on the aboriginal lands of the Fallon Paiute Shoshone Tribe (Tribe) and the Tribe's oral histories and spiritual beliefs provide that Spirit Cave is a sacred place containing the remains of the Tribe's ancient ancestors. The Tribe's religious and spiritual beliefs also provide that human remains, once buried, are not to be disturbed in any way, and that such disturbance can have terrible consequences for the living members of the Tribe.

B. In 1940, archaeologists removed human remains and associated funerary objects from Spirit Cave, including a mummified man, cremated remains found in woven bags, and the bones of at least two other persons (collectively, the Spirit Cave remains). Those remains and objects are currently stored in the Nevada State Museum.

C. The Tribe has been seeking to repatriate the Spirit Cave remains, as well as associated funerary objects, for almost two decades. All of the Spirit Cave remains and associated funerary objects are controlled by the Bureau of Land Management (BLM). The Tribe first submitted a formal request to repatriate the Spirit Cave remains and associated funerary objects to BLM in 1997.

D. In 2000, BLM denied the Tribe's repatriation request for the remains. Although BLM found that the remains were Native American, it concluded that they were not culturally affiliated with the Tribe under the Native American Graves Protection and Repatriation Act (NAGPRA).

E. The Tribe disagreed with BLM's decision and sought the assistance of the NAGPRA Review Committee to resolve the dispute. In 2001, the Review Committee found that the Spirit Cave remains were culturally affiliated with the Tribe and recommended that BLM repatriate them to the Tribe. BLM made a final decision not to repatriate the remains to the Tribe in 2004.

F. The Tribe challenged BLM's decision in the United States District Court for the District of Nevada. In 2006, the Court held that BLM's decision was arbitrary and capricious because it had failed to consider all of the evidence submitted by the Tribe and the findings of the NAGPRA Review Committee. The Court vacated BLM's decision and remanded the matter to BLM to make a new decision based on review and consideration of all of the evidence submitted by the Tribe, the findings of the Review Committee, and any additional evidence that existed at that time.

G. In 2014 (eight years after issuance of the remand order), BLM re-initiated consultation with 26 tribal officials under NAGPRA regarding repatriation of the Spirit Cave remains and associated funerary objects. BLM stated that its review of the evidence indicated that the remains were Native American and were culturally affiliated with the Tribe.

H. Nevertheless, BLM subsequently informed the Tribe that it wished to authorize DNA testing of the mummified Spirit Cave remains in order to confirm its conclusion that they were Native American within the meaning of NAGPRA. It later informed the Tribe that it would also like to authorize DNA testing of the other human remains from Spirit Cave. No other additional testing or examination of the remains or objects has been proposed by BLM.

I. The Tribe has strongly opposed any destructive testing of the Spirit Cave remains, including DNA testing, as a violation of its core religious beliefs. Moreover, the Tribe believes that DNA testing is not authorized by (and would violate) NAGPRA and other applicable law under the circumstances involving the Spirit Cave remains.

J. However, in order to avoid litigation and expedite BLM's long-delayed decision on remand from the United States District Court, the Tribe will not oppose DNA testing of the Spirit Cave remains under the terms specified in this agreement.

K. The Tribe's agreement not to oppose DNA testing in this case should not be interpreted to establish a precedent of any kind and is not intended to authorize DNA or other destructive testing in any situation other than that with respect to the present circumstances of its pending request for repatriation of the Spirit Cave remains.

TERMS

1. BLM may authorize DNA testing of the Spirit Cave remains by Dr. Eske Willerslev in order to help BLM determine whether or not the remains are Native American within the meaning of NAGPRA. BLM will limit testing to the mummified man (catalogued by the Museum as AHUR 2064) and the bones catalogued as AHUR 748 and AHUR 770. BLM will not authorize testing of the cremated remains due to the low probability of extracting DNA from them. No other testing, examination or review of the Spirit Cave remains and objects will be allowed, and BLM will not solicit further testing or examination of the remains or any associated funerary objects without the Tribe's consent.

2. BLM will permit only those portions of the human remains absolutely necessary to conduct DNA testing to be extracted by Dr. Willerslev.

3. Extraction will occur on a date agreeable to the Tribe within 30 days of the signing of this agreement.

4. BLM will provide the Tribe with at least 10 days' written notice of the date for extraction and allow the Tribe to have two representatives present for the extraction.

5. BLM will require Dr. Willerslev to report his conclusion whether the human remains are Native American within 90 days of extraction. The report need not be in a formal, peer reviewed or published format; any reliable written form will suffice.

6. If Dr. Willerslev confirms that the remains are Native American, BLM will issue a final decision to return the Spirit Cave remains and associated funerary objects to the Tribe no later than 30 days after receiving Dr. Willerslev's report.

7. Upon a decision of BLM to return the Spirit Cave remains and associated funerary objects to the Tribe, the time and manner of such return will be determined by the Tribe and honored by BLM.

8. Upon a decision of BLM to return the Spirit Cave remains and associated funerary objects to the Tribe, BLM will take all necessary steps to ensure that the Nevada State Museum has the remains and funerary objects available for transfer at the agreed-upon date and time.

9. Upon a decision of BLM to return the Spirit Cave remains and associated funerary objects to the Tribe, neither BLM nor the Nevada State Museum may retain any portion of the remains or funerary objects.

10. BLM may authorize Dr. Willerslev's team to publish the results of the DNA testing in a scientific journal and in a DNA sequencing database. BLM will not authorize any other publication of the results of the DNA testing.

Fallon Paiute Shoshone Tribe

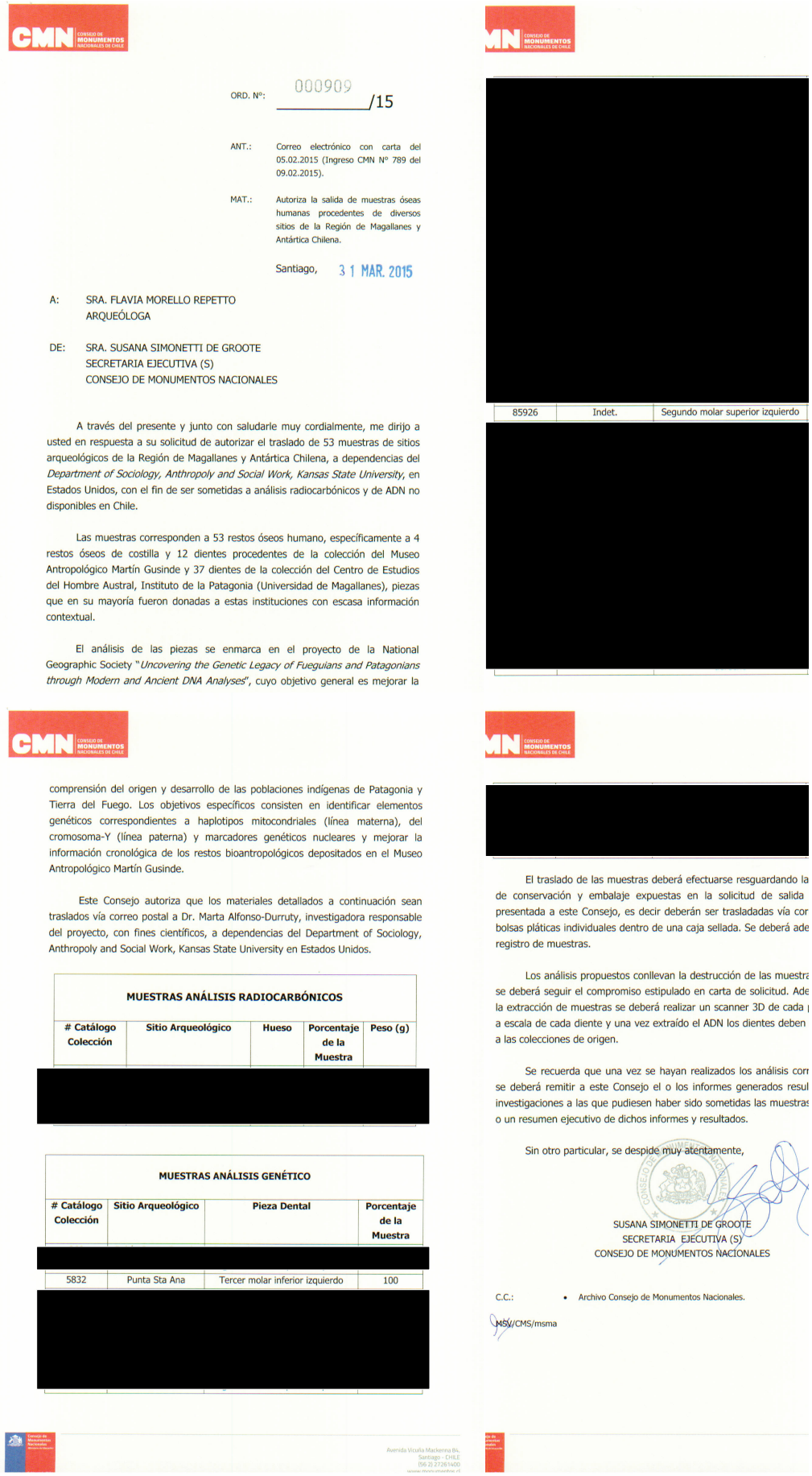
BY: _____ Date: 09/01/2015
Len George, Chairman

Bureau of Land Management

BY: _____ Date: 9/1/2015
Amy Lueders
Acting Assistant Director, Resources and Planning

BY: _____ Date: 1 September 2015
John Ruhs
Acting Nevada State Director

Figure S2. Document authorizing the analysis of human remains from Spirit Cave, signed by representatives of the Fallon Paiute Shoshone Tribe.



ORD. Nº: 000909 /15

ANT.: Correo electrónico con carta del 05.02.2015 (Ingreso CMN Nº 789 del 09.02.2015).

MAT.: Autoriza la salida de muestras óseas humanas procedentes de diversos sitios de la Región de Magallanes y Antártica Chilena.

Santiago, 31 MAR. 2015

A: SRA. FLAVIA MORELLO REPETTO
ARQUEÓLOGA

DE: SRA. SUSANA SIMONETTI DE GROOTE
SECRETARIA EJECUTIVA (S)
CONSEJO DE MONUMENTOS NACIONALES

A través del presente y junto con saludarle muy cordialmente, me dirijo a usted en respuesta a su solicitud de autorizar el traslado de 53 muestras de sitios arqueológicos de la Región de Magallanes y Antártica Chilena, a dependencias del *Department of Sociology, Anthropoly and Social Work, Kansas State University*, en Estados Unidos, con el fin de ser sometidas a análisis radiocarbónicos y de ADN no disponibles en Chile.

Las muestras corresponden a 53 restos óseos humano, específicamente a 4 restos óseos de costilla y 12 dientes procedentes de la colección del Museo Antropológico Martín Gusinde y 37 dientes de la colección del Centro de Estudios del Hombre Austral, Instituto de la Patagonia (Universidad de Magallanes), piezas que en su mayoría fueron donadas a estas instituciones con escasa información contextual.

El análisis de las piezas se enmarca en el proyecto de la National Geographic Society "*Uncovering the Genetic Legacy of Fueguians and Patagonians through Modern and Ancient DNA Analyses*", cuyo objetivo general es mejorar la

[redacted]

85926	Indet.	Segundo molar superior izquierdo

[redacted]

comprensión del origen y desarrollo de las poblaciones indígenas de Patagonia y Tierra del Fuego. Los objetivos específicos consisten en identificar elementos genéticos correspondientes a haplotipos mitocondriales (línea materna), del cromosoma-Y (línea paterna) y marcadores genéticos nucleares y mejorar la información cronológica de los restos bioantropológicos depositados en el Museo Antropológico Martín Gusinde.

Este Consejo autoriza que los materiales detallados a continuación sean traslados vía correo postal a Dr. Marta Alfonso-Durruty, investigadora responsable del proyecto, con fines científicos, a dependencias del Department of Sociology, Anthropoly and Social Work, Kansas State University en Estados Unidos.

MUESTRAS ANÁLISIS RADIOCARBÓNICOS

# Catálogo Colección	Sitio Arqueológico	Hueso	Porcentaje de la Muestra	Peso (g)

[redacted]

MUESTRAS ANÁLISIS GENÉTICO

# Catálogo Colección	Sitio Arqueológico	Pieza Dental	Porcentaje de la Muestra
[redacted]			
5832	Punta Sta Ana	Tercer molar inferior izquierdo	100
[redacted]			

[redacted]

El traslado de las muestras deberá efectuarse resguardando la de conservación y embalaje expuestas en la solicitud de salida presentada a este Consejo, es decir deberán ser trasladadas vía cor bolsas plásticas individuales dentro de una caja sellada. Se deberá ade registro de muestras.

Los análisis propuestos conllevan la destrucción de las muestra se deberá seguir el compromiso estipulado en carta de solicitud. Ade la extracción de muestras se deberá realizar un scanner 3D de cada a escala de cada diente y una vez extraído el ADN los dientes deben a las colecciones de origen.

Se recuerda que una vez se hayan realizados los análisis corr se deberá remitir a este Consejo el o los informes generados resul investigaciones a las que pudiesen haber sido sometidas las muestras o un resumen ejecutivo de dichos informes y resultados.

Sin otro particular, se despide muy atentamente,

SUSANA SIMONETTI DE GROOTE
SECRETARIA EJECUTIVA (S)
CONSEJO DE MONUMENTOS NACIONALES

C.C.: • Archivo Consejo de Monumentos Nacionales.

MSV/CMS/msma

Figure S3. Document authorizing the analysis of human remains from Punta Santa Ana and Ayayema individuals. This document (issued in Spanish) authorizes for the transportation of the remains out of Chile as well as for destructive analyses of the samples. This permit has been granted to Flavia Morello Repetto and Marta Alfonso-Durruty by the National Monuments Council in Chile. Samples not included in this study are not shown.

EAAF EQUIPO ARGENTINO DE ANTROPOLOGÍA FORENSE

Av. del Libertador 8151
C1429BNC-CABA-Argentina
Tel./Fax +5411 5275-0552
E-mail: eaaf@eaaf.org
Web: www.eaaf.org

A quien corresponda

Certificación de cumplimiento de los requisitos éticos del proyecto de investigación titulado:

TÍTULO DEL PROYECTO: **Estudio biológico de la momia Incaica del cerro Aconcagua**

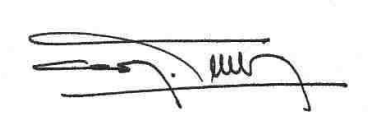
Por la presente hago constar que la muestra (biopsia de pulmón) de la momia del cerro Aconcagua analizada en proyectos genéticos pasados y en el presente proyecto fue extraída por el Profesor Juan Schobinger (Mendoza) y cedida para su estudio a mi persona, quien suscribe, Dr. Carlos Vullo en el año 1985. Dicha muestra fue custodiada desde sus inicios por mí y dispongo de ella para hacer los análisis biológicos/genéticos oportunos.

El objetivo del presente proyecto es el de realizar análisis biológicos desde distintas perspectivas, que incluye por ejemplo el estudio de su genoma parcial o total como el de la microbiota presente en las muestras. Estos estudios están en la misma línea que los que habitualmente se llevan a cabo en el ámbito de la antropología molecular.

La muestra de biopsia de la momia se analiza en estrecha colaboración académica con el investigador Prof. Antonio Salas Ellacuriaga, del laboratorio del Instituto de Ciencias Forenses de la Universidad de Santiago de Compostela (USC; España), con el que se comparte la muestra para su análisis. Así mismo, desde el laboratorio de Santiago de Compostela se pueden tomar también decisiones estratégicas en cuanto a incluir a otros laboratorios colaboradores nacionales o extranjeros que pudieran aportar capacidad de análisis e interpretación de las muestras o con los que se pudieran establecer colaboraciones académicas en el terreno de la antropología molecular, incluyendo el compartir una pieza de la biopsia, ya sea original o procesada (ej. ADN extraído de la biopsia).

SE INFORMA QUE:
Que el proyecto titulado "**Estudio biológico de la momia Incaica del cerro Aconcagua**" ha sido examinado por el Equipo Argentino de Antropología Forense (EAAF) y cumple en su protocolo experimental con todos los aspectos éticos necesarios. Así mismo, el estudio conforma toda la normativa legal y cumple con los preceptos derivados de la declaración de Helsinki en relación los principios éticos para las investigaciones biomédicas.

Es cuanto informo en honor a la verdad y para los fines consiguientes,
Atentamente,



CARLOS M VULLO
Director Científico,
Lab de Genética Forense del EAAF
Responsable en custodia

Figure S4. Document authorizing the analysis of the Aconcagua human remains. This document (issued in Spanish) was signed by a representative of the 'Equipo Argentino de Antropología Forense', who holds responsibility over the Aconcagua sample. The signee, Carlos Vullo, is also a co-author in the present study.

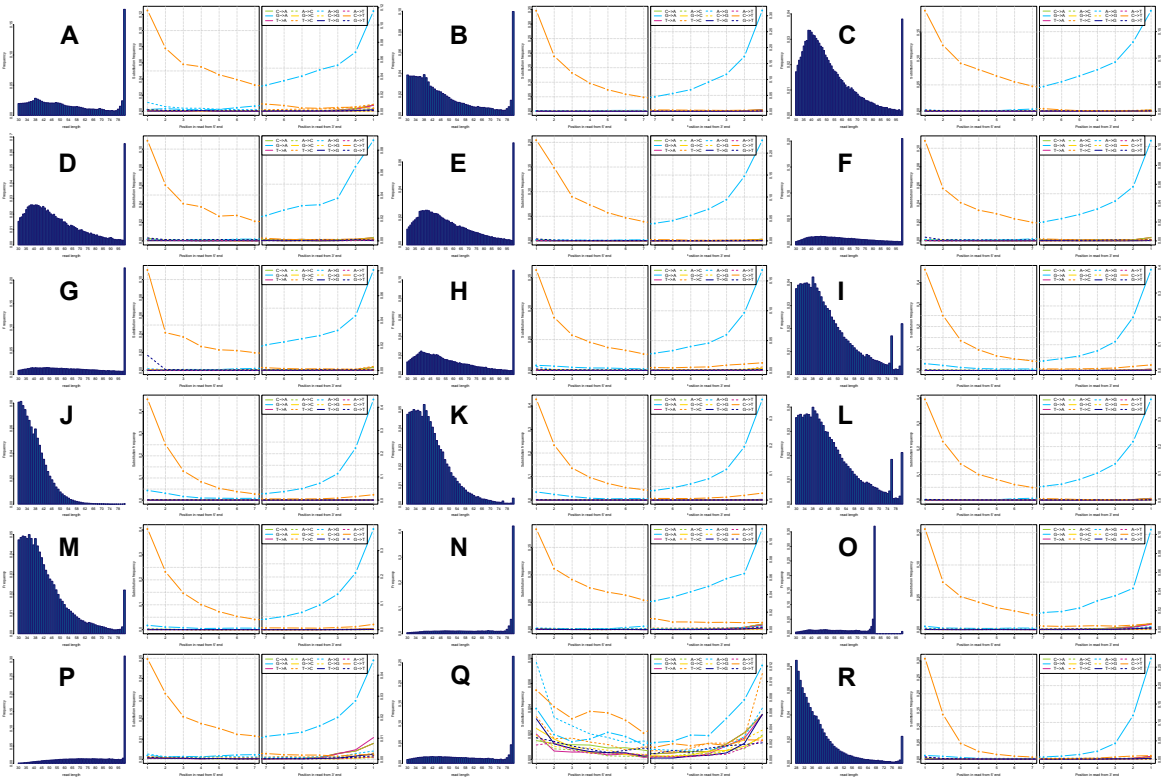


Figure S5. aDNA fragment length distributions and misincorporation patterns. In each panel, we show the length distribution of the mapped trimmed reads after removing PCR duplicates (left) and misincorporation frequencies with respect to the position from the 5' (middle) and 3' (right) ends of the reads. We only show results for one representative non-USER treated library per sample, and we summarize this information for all libraries, in Table S8.

Libraries represented in this figure are: **a.** TrailCreek_198830_E1_L12, **b.** 19651_19651_E1_L1, **c.** AHUR_2064_017374E5_L1, **d.** AHUR770c_017378E1_L1, **e.** Lovelock1_017380PE3_L1, **f.** Lovelock2_017381AE1_L1, **g.** Lovelock3_017382BE2_L1, **h.** Lovelock4_017383E2_L1, **i.** Sumidouro4_PW8E2_L1, **j.** Sumidouro5_PW12E1_L2, **k.** Sumidouro6_PW13E2_L1, **l.** Sumidouro7_PW15E2_L1, **m.** Sumidouro8_DA370E1_L1, **n.** 5832_n5832b1, **o.** A460_E1_L2, **p.** Andaman_MA756_L4, **q.** Aconcagua_L1, and **r.**DS4_DS4_E2_L4.

Figure S6. Relative error rates. Excess derived alleles compared to a high quality sequencing assay from the 1000 genomes project. For each library produced in this study, we show type-specific (bars) and overall error rates (legend).

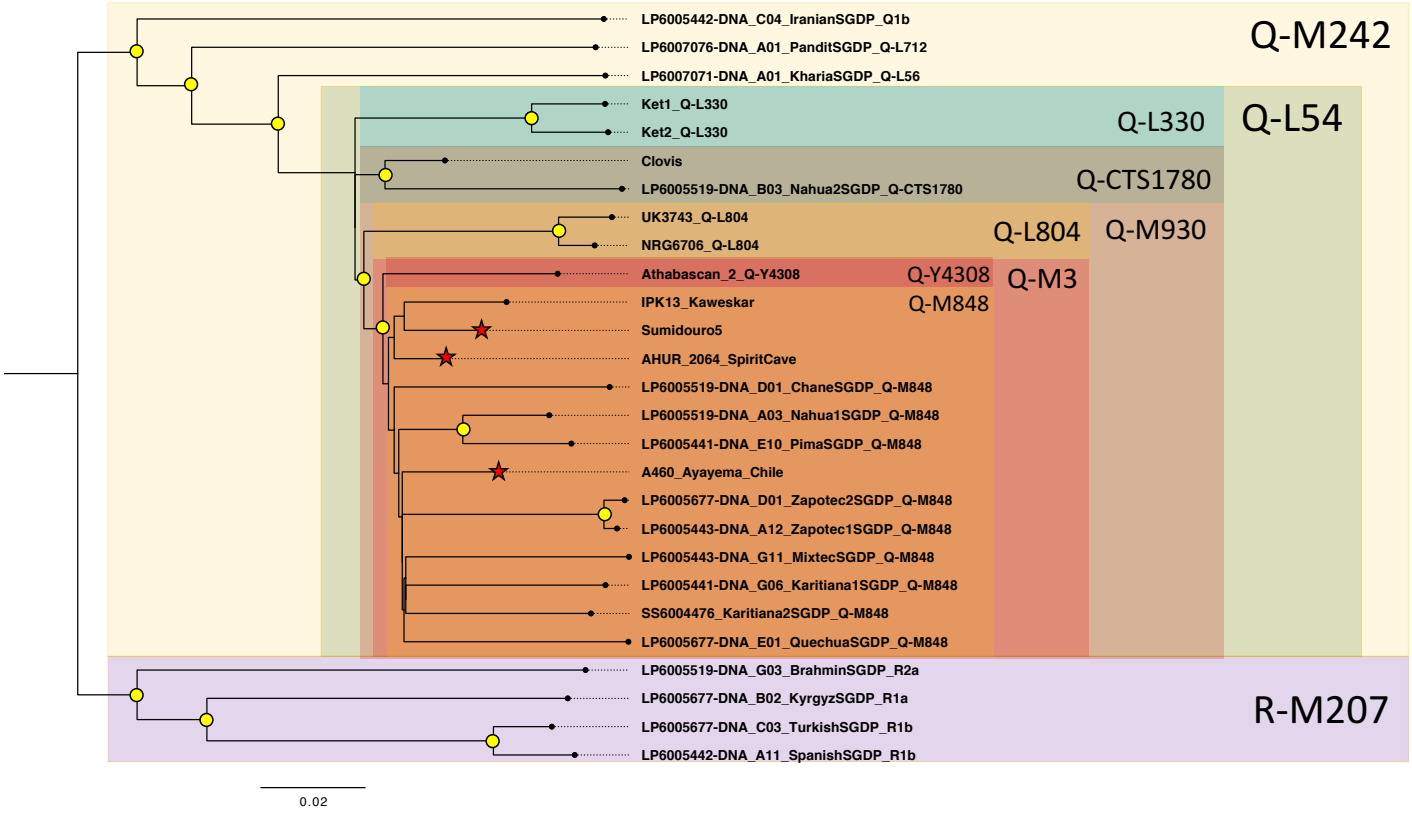


Figure S7. Maximum-likelihood Y chromosome phylogeny. Tree of new high-coverage ancient Y chromosome sequences and a published dataset of relevant modern and ancient Native American and Eurasian Y chromosomes. Hierarchical haplogroup assignment was added on the phylogeny, yellow circles denote bootstrap support >90%, and red stars mark the ancient samples sequenced in this study. See Tables S9 and S10 for a detailed list of samples and references.

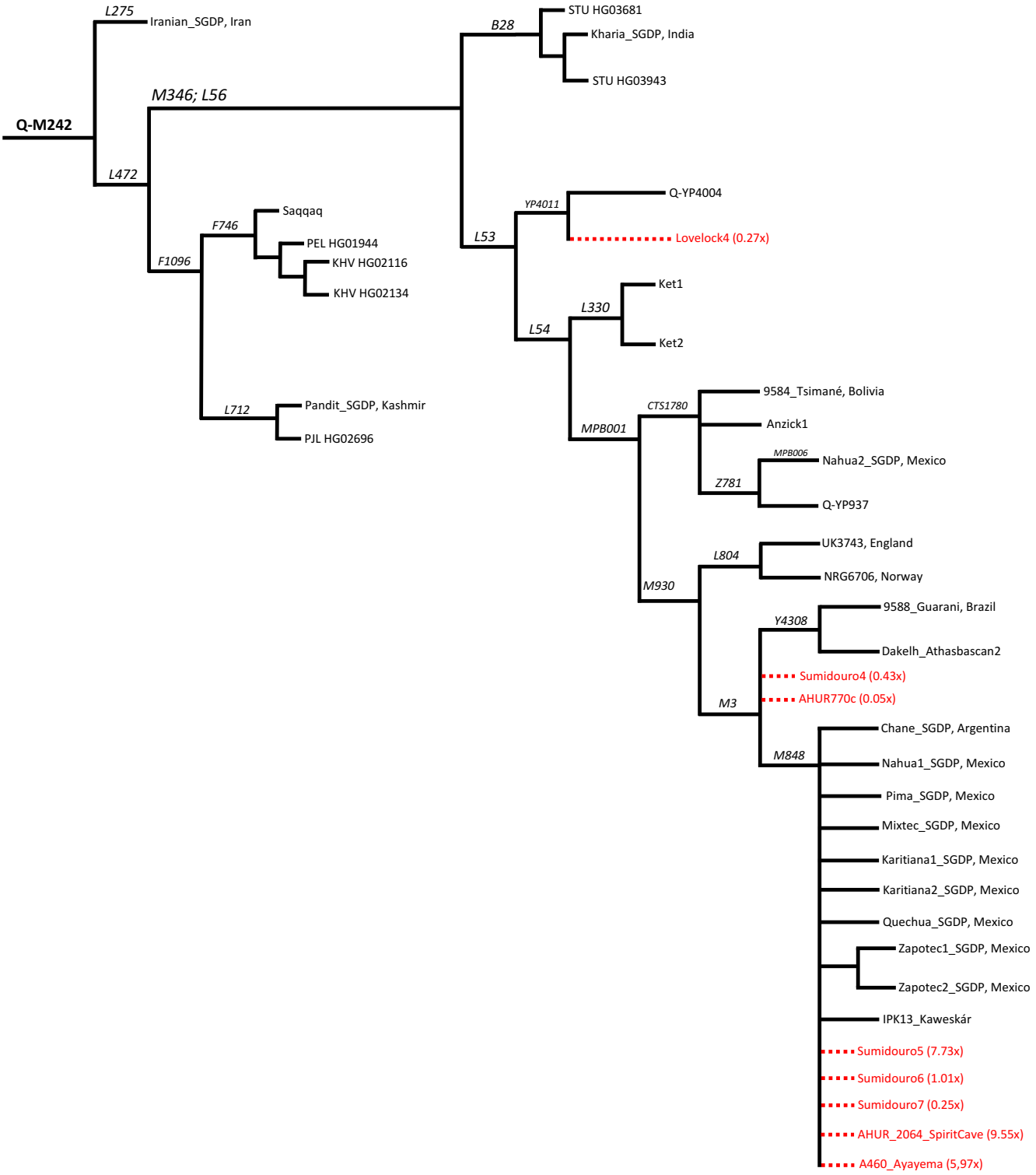


Figure S8. Y chromosome phylogeny including low-coverage samples. Simplified haplogroup Q tree showing the main branches, with the inclusion of new ancient sequences in red and hashed lines. Branch lengths are arbitrary and only added for clarity. Defining SNPs are found on top of branches, and location of high-coverage samples included in this study are shown. Y chromosome coverage of ancient samples can be found between parentheses. See Tables S9 and S10 and (7) for detailed information on samples and Y calls.

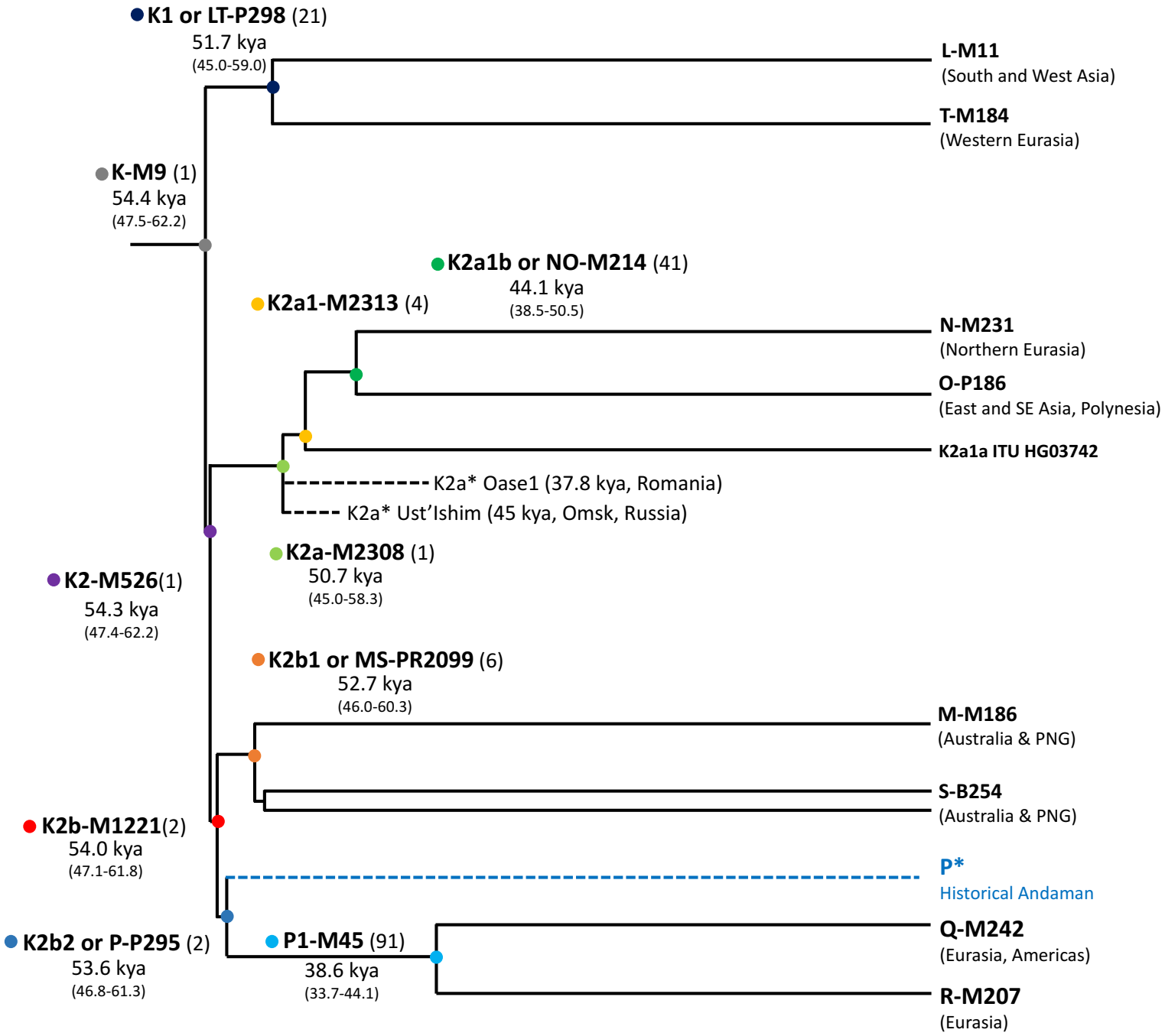


Figure S9. Time-calibrated phylogeny of macro-haplogroup K-M9. Y chromosome tree depicting main lineages descending from K-M9. Ancient basal samples are included as hashed lines. A nomenclature of those deep branches is presented based on a parsimony approach, and the number of SNPs supporting each branch is found between parentheses (all SNPs are detailed in Table S11). Under each time estimate, we show 95% Confidence Intervals taking into the account the uncertainty of the mutation rate and the stochastic nature of mutations in different samples. See Tables S9 and S10 for a detailed list of high-coverage samples used to represent each non-basal haplogroups.

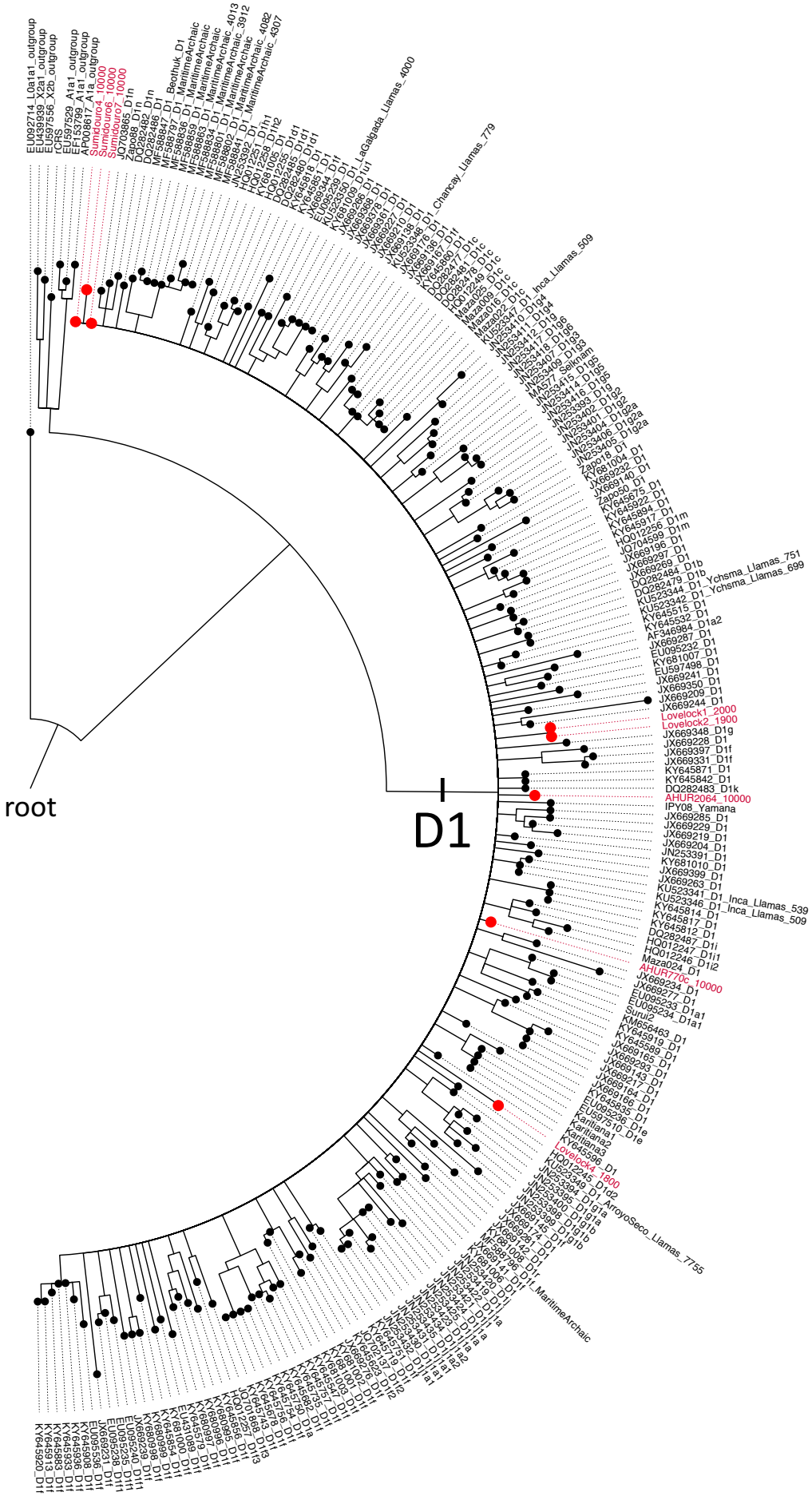


Figure S10. Maximum-likelihood mtDNA haplogroup D1 phylogeny. Tree with 197 modern and 20 ancient mtDNA D1 sequences from the literature, together with 8 ancient mitogenomes sequenced in this study, highlighted in red. See Tables S9 and S10 for detailed information and references of all samples.

Figure S11. Maximum-likelihood mtDNA haplogroup D4h3 phylogeny. Tree with 38 modern and 7 ancient relevant mtDNA D4 sequences from literature, together with Sumidouro5 mitogenome sequenced in this study. Both Sumidouro5 and Clovis were highlighted in red for clarity. See Tables S9 and S10 for detailed information and references of all samples.

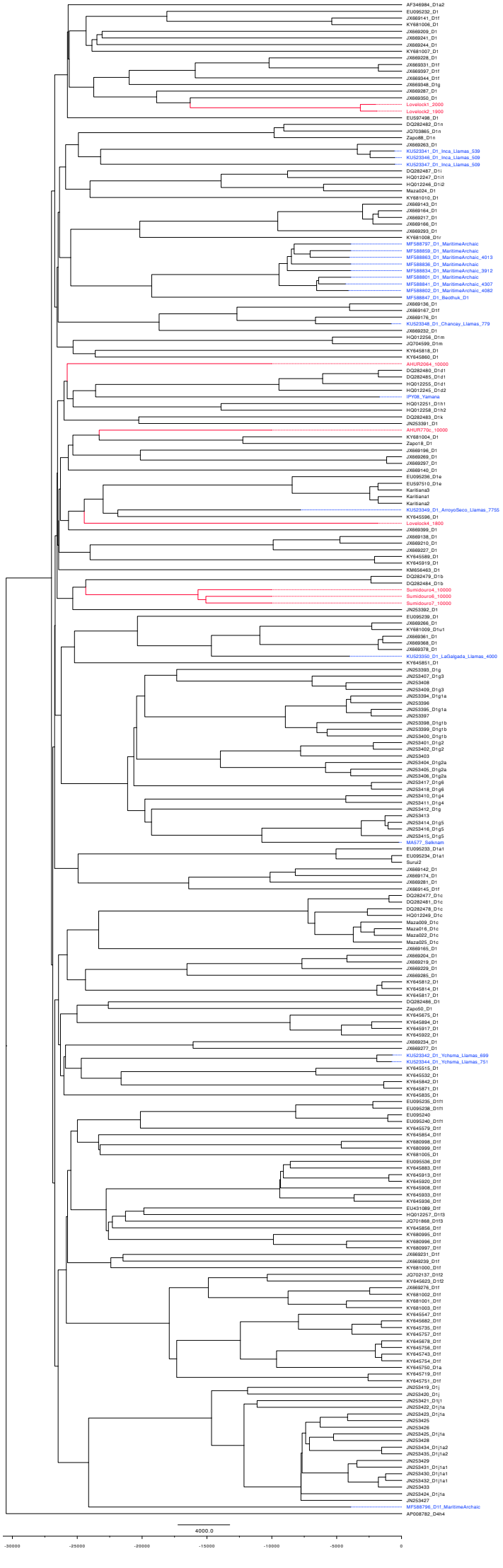

AF346984_D1a2
EU095232_D1
JX669141_D1f
KY681006_D1
JX669209_D1
JX669241_D1
JX669244_D1
KY681007_D1
JX669228_D1
JX669331_D1f
JX669397_D1f
JX669344_D1f
JX669348_D1g
JX669287_D1
JX669350_D1
Lovelock1_2000
Lovelock2_1800
EU597498_D1
DQ282482_D1n
JQ703865_D1n
Zapo88_D1n
JX669263_D1
KU523341_D1_Inca_Llamas_539
KU523346_D1_Inca_Llamas_509
KU523347_D1_Inca_Llamas_509
DQ282487_D1i
HQ012247_D1i1
HQ012246_D1i2
Maza024_D1
KY681010_D1
JX669143_D1
JX669164_D1
JX669217_D1
JX669166_D1
JX669293_D1
KY681008_D1r
MF588797_D1_MaritimeArchaic
MF588859_D1_MaritimeArchaic
MF588863_D1_MaritimeArchaic_4013
MF588836_D1_MaritimeArchaic
MF588834_D1_MaritimeArchaic_3912
MF588801_D1_MaritimeArchaic
MF588841_D1_MaritimeArchaic_4307
MF588802_D1_MaritimeArchaic_4082
MF588847_D1_Bacthuk_D1
JX669136_D1
JX669167_D1f
JX669176_D1
KU523348_D1_Chancay_Llamas_779
JX669232_D1
HQ012256_D1m
JQ704599_D1m
KY645818_D1
KY645860_D1
AHUR2084_10000
DQ282480_D1d1
DQ282485_D1d1
HQ012255_D1d1
HQ012245_D1d2
IPY08_Yamana
HQ012251_D1h1
HQ012258_D1h2
DQ282483_D1k
JN253391_D1
AHUR770c_10000
KY681004_D1
Zapo18_D1
JX669196_D1
JX669269_D1
JX669297_D1
JX669140_D1
EU095256_D1e
EU597510_D1e
Karitiana3
Karitiana1
Karitiana2
KU523349_D1_ArroyoSeco_Llamas_7755
KY645596_D1
Lovelock4_1800
JX669399_D1
JX669138_D1
JX669210_D1
JX669227_D1
KY645589_D1
KY645919_D1
KM656463_D1
DQ282479_D1b
DQ282484_D1b
Sumidouro4_10000
Sumidouro6_10000
Sumidouro7_10000
JN253392_D1
EU095239_D1
JX669266_D1
KY681009_D1u1
JX669361_D1
JX669368_D1
JX669378_D1
KU523350_D1_LaGalgada_Llamas_4000
KY645851_D1
JN253393_D1g
JN253407_D1g3
JN253408
JN253409_D1g3
JN253394_D1g1a
JN253396
JN253395_D1g1a
JN253397
JN253398_D1g1b
JN253399_D1g1b
JN253400_D1g1b
JN253401_D1g2
JN253402_D1g2
JN253403
JN253404_D1g2a
JN253405_D1g2a
JN253406_D1g2a
JN253417_D1g6
JN253418_D1g6
JN253410_D1g4
JN253411_D1g4
JN253412_D1g
JN253413
JN253414_D1g5
JN253416_D1g5
JN253415_D1g5
MA577_Suknuan
EU095233_D1a1
EU095234_D1a1
Surui2
JX669142_D1
JX669174_D1
JX669281_D1
JX669145_D1f
DQ282477_D1c
DQ282481_D1c
DQ282478_D1c
HQ012249_D1c
Maza009_D1c
Maza016_D1c
Maza022_D1c
Maza025_D1c
JX669165_D1
JX669204_D1
JX669219_D1
JX669229_D1
JX669285_D1
KY645812_D1
KY645814_D1
KY645817_D1
DQ282486_D1
Zapo50_D1
KY645675_D1
KY645894_D1
KY645917_D1
KY645922_D1
JX669234_D1
JX669277_D1
KU523342_D1_Ychsma_Llamas_699
KU523344_D1_Ychsma_Llamas_751
KY645515_D1
KY645532_D1
KY645842_D1
KY645871_D1
KY645835_D1
EU095235_D1f1
EU095238_D1f1
EU095240
EU095240_D1f1
KY645579_D1f
KY645854_D1f
KY680998_D1f
KY680999_D1f
KY681005_D1
EU095536_D1f
KY645883_D1f
KY645913_D1f
KY645920_D1f
KY645908_D1f
KY645933_D1f
KY645936_D1f
EU431089_D1f
HQ012257_D1f3
JQ701868_D1f3
KY645856_D1f
KY680995_D1f
KY680996_D1f
KY680997_D1f
JX669231_D1f
JX669239_D1f
KY681000_D1f
JQ702137_D1f2
KY645623_D1f2
JX669276_D1f
KY681002_D1f
KY681001_D1f
KY681003_D1f
KY645547_D1f
KY645682_D1f
KY645735_D1f
KY645757_D1f
KY645678_D1f
KY645756_D1f
KY645743_D1f
KY645754_D1f
KY645750_D1a
KY645719_D1f
KY645751_D1f
JN253419_D1j
JN253420_D1j
JN253421_D1j1
JN253422_D1j1a
JN253423_D1j1a
JN253425
JN253426
JN253425_D1j1a
JN253428
JN253434_D1j1a2
JN253435_D1j1a2
JN253429
JN253431_D1j1a1
JN253430_D1j1a1
JN253432_D1j1a1
JN253433
JN253424_D1j1a
JN253427
MF588796_D1f_MaritimeArchaic
AP008782_D4h4
4000.0
-30000
-25000
-20000
-15000
-10000
-5000
0

Figure S12. Bayesian mtDNA haplogroup D1 phylogeny. Tree with 197 modern and 20 ancient mtDNA D1 sequences from the literature, together with 8 ancient mitogenomes sequenced in this study. New ancient sequences are highlighted in red, while ancient sequences from the literature are in blue. The dates of all ancient samples were added as calibration points. See Tables S9 and S10 for the dates and detailed information and references for all samples.

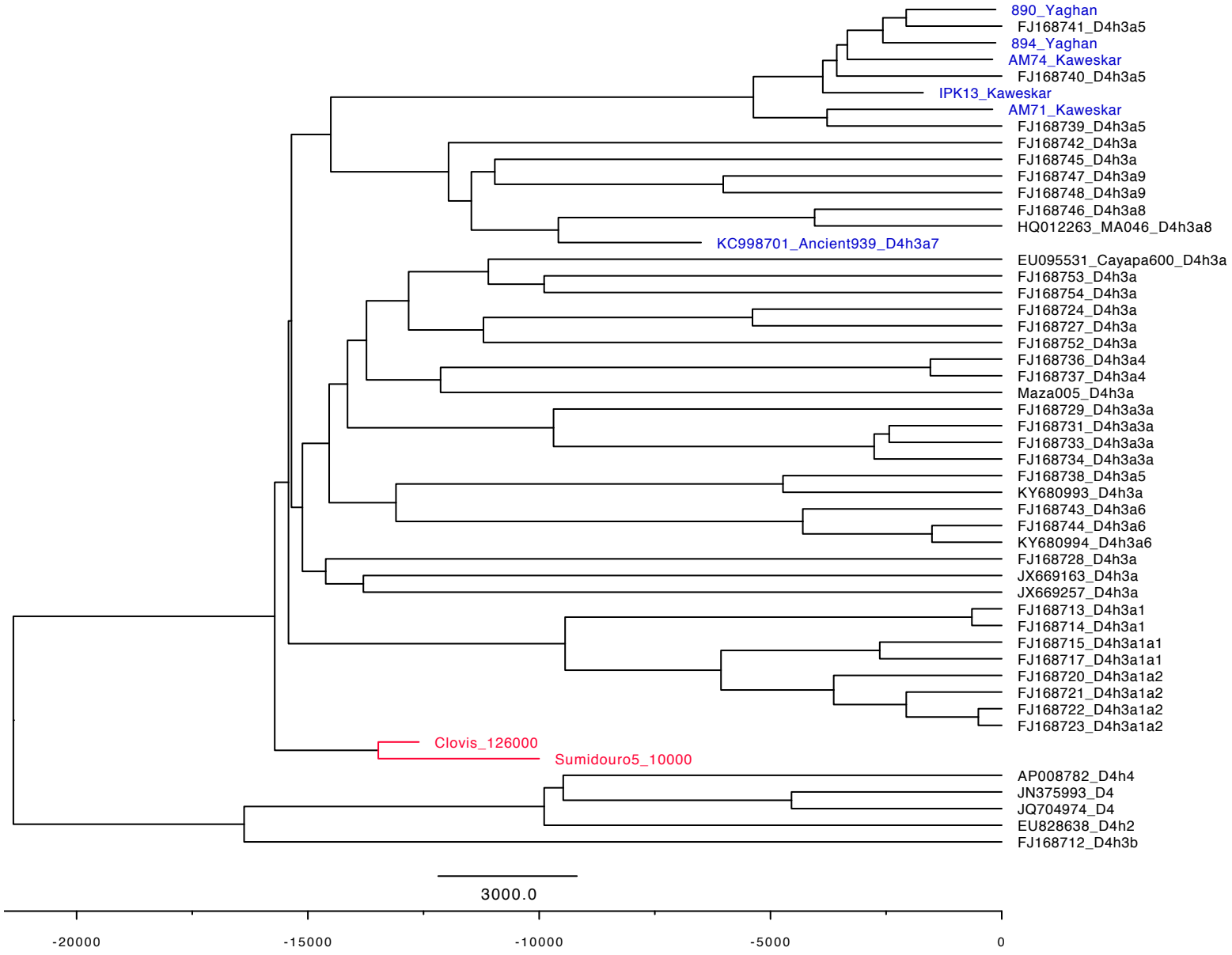


Figure S13. Bayesian mtDNA haplogroup D4h3a phylogeny. Tree with 38 modern and 7 ancient relevant mtDNA D4 sequences from the literature, together with the Sumidouro5 mitogenome sequenced in this study. Both Sumidouro5 and Anzick1 are highlighted in red, while other ancient sequences from the literature are highlighted in blue. The dates of all ancient samples are added as calibration points. See Tables S9 and S10 for the dates and detailed information and references for all samples.

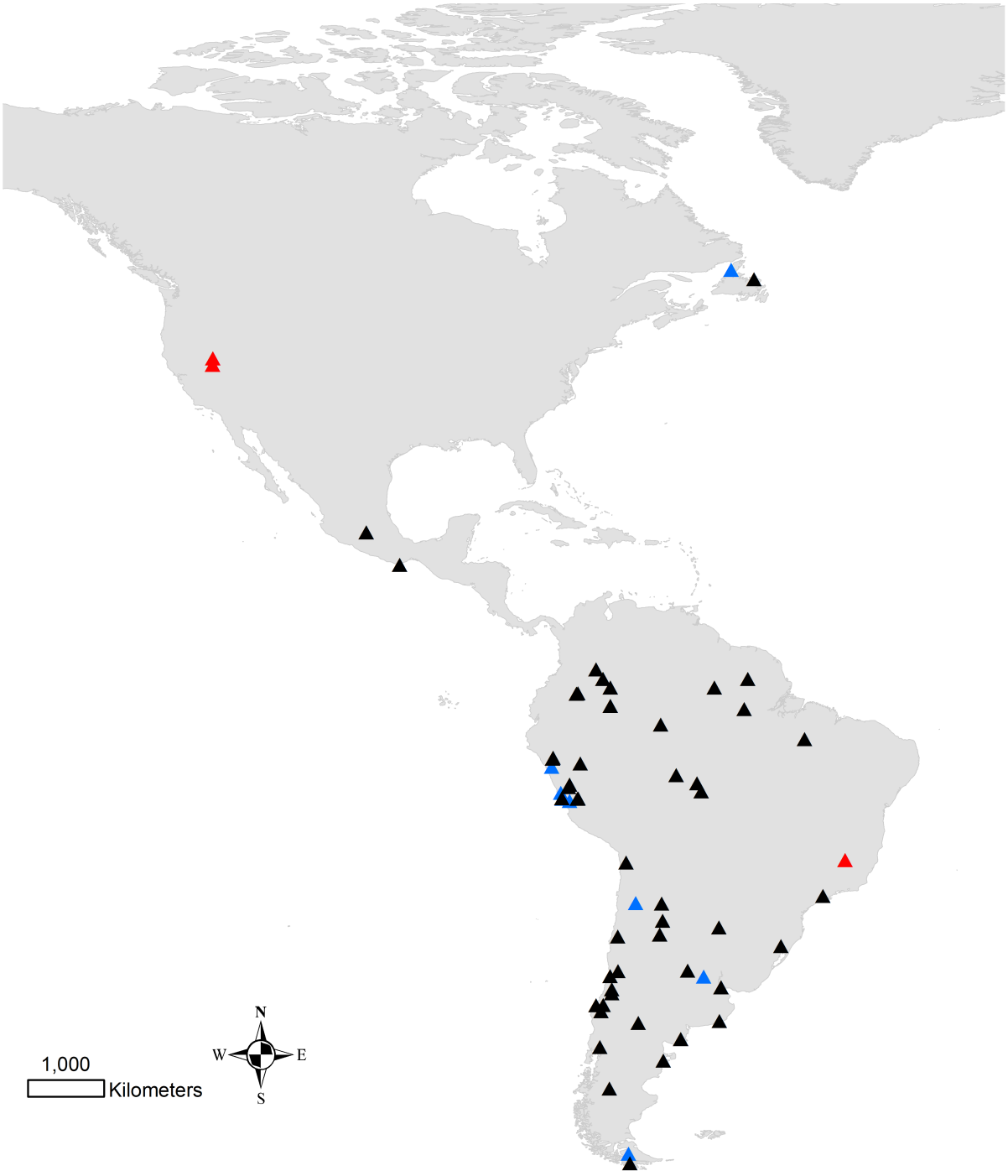


Figure S14. Approximate locations of mtDNA haplogroup D1 samples. New ancient sequences are in red and dated ancient sequences from the literature in blue. The triangles mark the sample location, not sample size. Literature samples with insufficient geographical information available were not plotted. All locations are listed in Table S9.

Figure S15. Approximate locations of mtDNA haplogroup D4h3a samples. New ancient sequences are in red and dated ancient sequences from the literature in blue. The triangles mark the sample location, not sample size. Literature samples with insufficient geographical information available were not plotted. All locations are listed in Table S9.

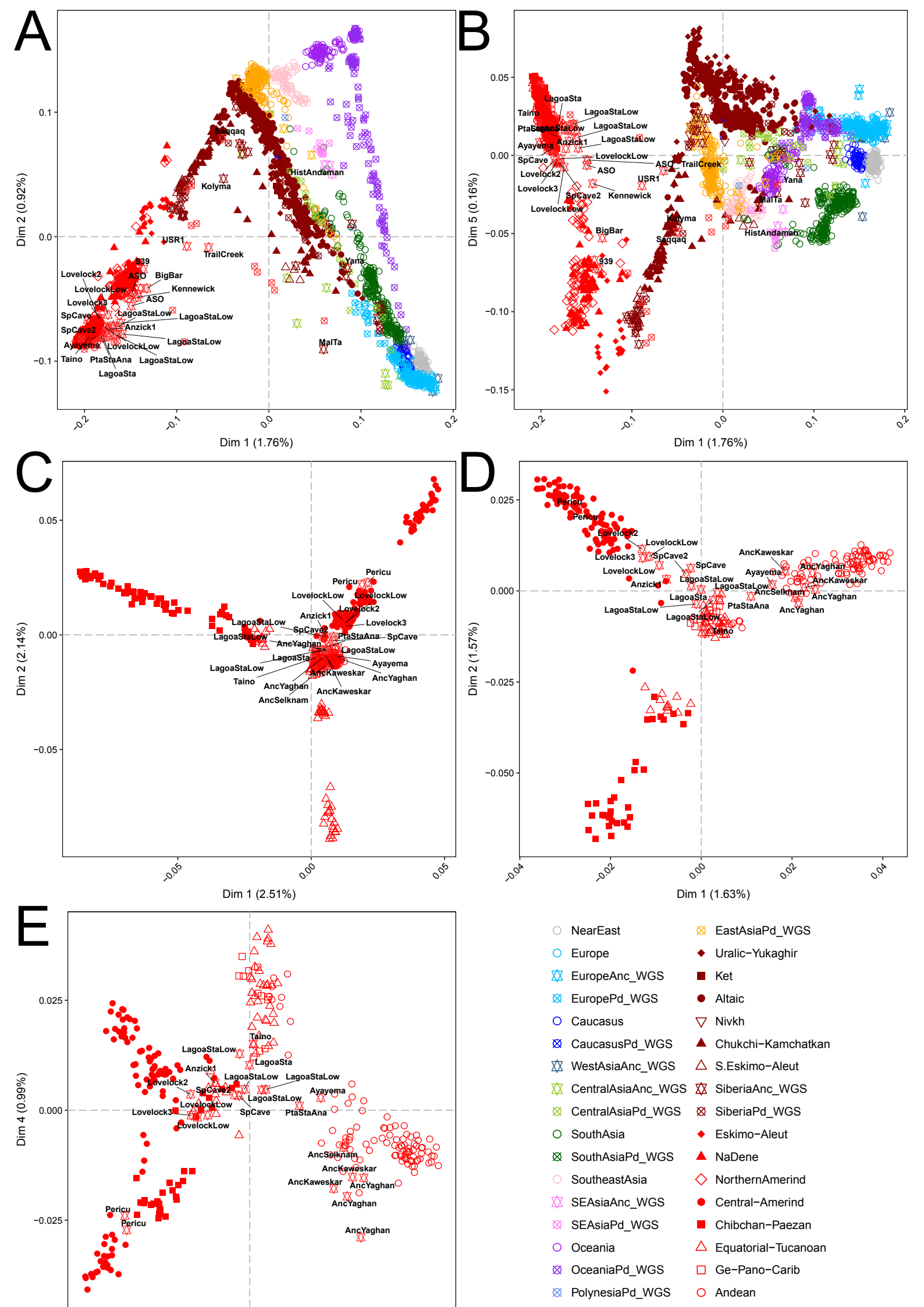

A
Dim 1 (1.76%)
Dim 2 (0.92%)
Saqqaq
Kolyma
HistAndaman
USR1
TrailCreek
Yana
939
Lovelock2
ASO
BigBar
LovelockLow
Kennewick
Lovelock3
SpCave
LagoaStaLow
Anzick1
SpCave2
Ayayema
LovelockLow
LagoaStaLow
Taino
PtaStaAna
LagoaStaLow
LagoaSta
MalTa
B
Dim 1 (1.76%)
Dim 5 (0.16%)
LagoaSta
Taino
LagoaStaLow
PtaStaAna
LagoaStaLow
Ayayema
Anzick1
LagoaStaLow
SpCave
LovelockLow
Lovelock2
ASO
Lovelock3
SpCave2
Kennewick
LovelockLow
ASO
TrailCreek
USR1
Yana
MalTa
Kolyma
BigBar
Saqqaq
939
HistAndaman
C
Dim 1 (2.51%)
Dim 2 (2.14%)
Pericu
Pericu
LovelockLow
LovelockLow
Anzick1
Lovelock2
Lovelock3
LagoaStaLow
SpCave2
PtaStaAna
SpCave
LagoaStaLow
AncYaghan
LagoaStaLow
LagoaSta
Ayayema
LagoaStaLow
AncKaweskar
AncYaghan
Taino
AncYaghan
AncKaweskar
AncSelknam
D
Dim 1 (1.63%)
Dim 2 (1.57%)
Pericu
Pericu
Lovelock2
LovelockLow
Lovelock3
SpCave2
SpCave
LovelockLow
LagoaStaLow
Anzick1
LagoaSta
LagoaStaLow
LagoaStaLow
Taino
AncKaweskar
Ayayema
AncYaghan
AncKaweskar
LagoaStaLow
AncSelknam
AncYaghan
PtaStaAna
AncYaghan
E
Dim 4 (0.99%)
Taino
LagoaStaLow
LagoaSta
Anzick1
LagoaStaLow
LagoaStaLow
Lovelock2
SpCave2
Ayayema
LagoaStaLow
LovelockLow
Lovelock3
SpCave
PtaStaAna
LovelockLow
AncSelknam
AncKaweskar
AncKaweskar
AncYaghan
AncYaghan
Pericu
Pericu
AncYaghan
NearEast
Europe
EuropeAnc_WGS
EuropePd_WGS
Caucasus
CaucasusPd_WGS
WestAsiaAnc_WGS
CentralAsiaAnc_WGS
CentralAsiaPd_WGS
SouthAsia
SouthAsiaPd_WGS
SoutheastAsia
SEAsiaAnc_WGS
SEAsiaPd_WGS
Oceania
OceaniaPd_WGS
PolynesiaPd_WGS
EastAsiaPd_WGS
Uralic−Yukaghir
Ket
Altaic
Nivkh
Chukchi−Kamchatkan
S.Eskimo−Aleut
SiberiaAnc_WGS
SiberiaPd_WGS
Eskimo−Aleut
NaDene
NorthernAmerind
Central−Amerind
Chibchan−Paezan
Equatorial−Tucanoan
Ge−Pano−Carib
Andean

Figure S16. Multidimensional scaling (MDS) of the identity-by-state distance matrix. We computed MDS transformations for different subsets of the SNP array dataset described in (Reference datasets) and plot different dimensions. **A.** Dimensions 1 and 2 for a subset excluding African individuals. **B.** Same subset as a., but we plot dimension 5. **C.** Dimensions 1 and 2 for a subset including SNA individuals only. **D.** Dimensions 1 and 2 for a subset including SNA individuals only, but removing outlier populations from c. (Cabecar, Pima, Surui and Karitiana). **E.** Same subset as d., but we plot dimension 4. Individuals are colored based on their broad continental ancestry, and Siberians and Native Americans are labeled according to their language family. Individuals sequenced in this study, as well as other relevant individuals are labeled on each plot. In the legend, whole genome sequencing samples are labeled 'WGS', ancient samples 'Anc' and present-day samples (Pd).

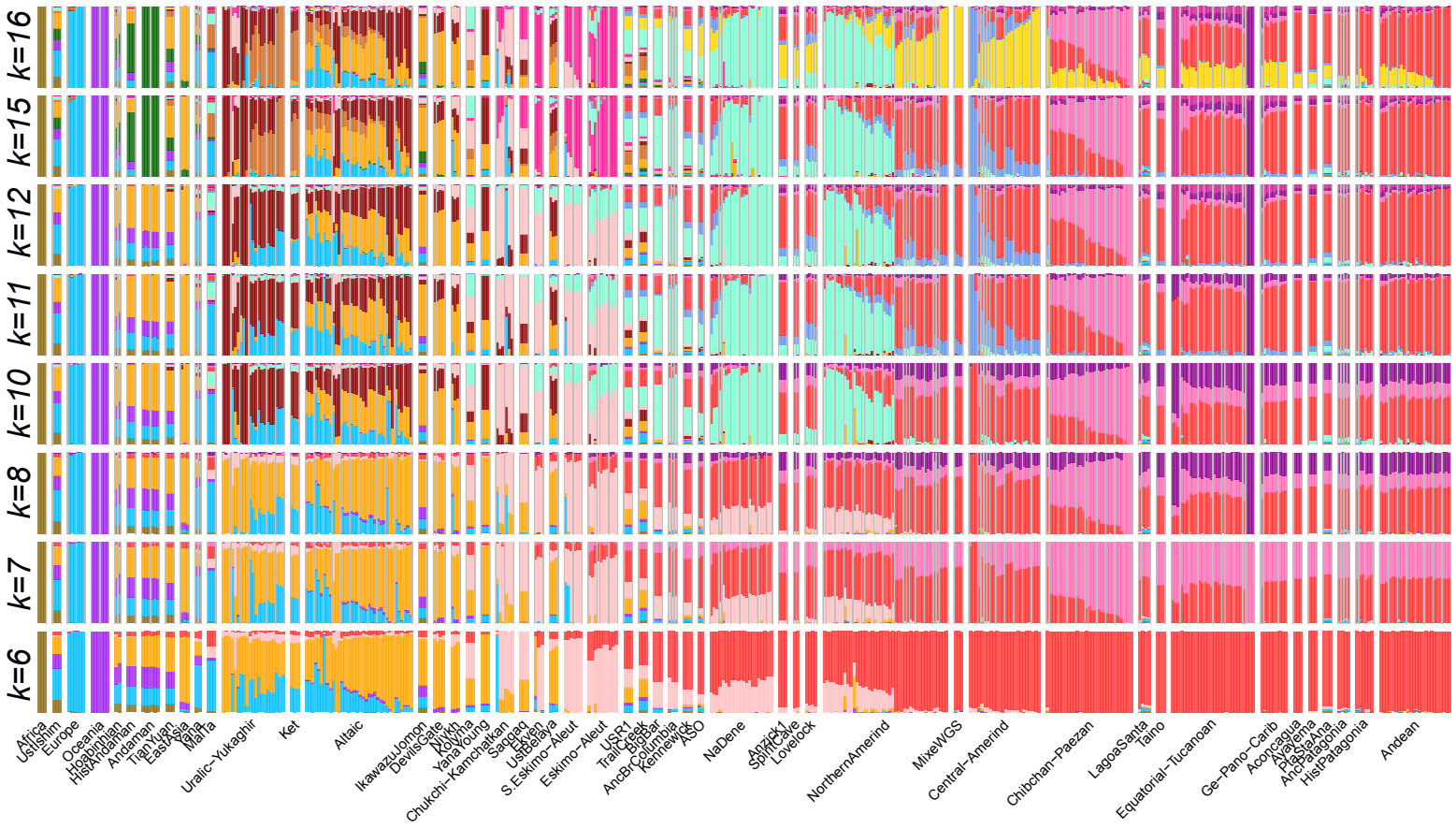


Figure S17. Model-based clustering analysis. We ran ADMIXTURE assuming *K={6,7,8,10,11,12,15,16}* ancestral components, on a subset of the SNP array dataset described in (Reference datasets), comprising 1,500 individuals. Each bar represents an individual, and each color represents the admixture proportion from each of the *K* assumed ancestral components. For clarity, we only show three individuals for populations with three or more individuals and single genomes are represented as wider bars. Siberian and Native American individuals are organized according to their language family.

Figure S18. Genetic drift sharing between ancient and present-day populations in the SNP array dataset. We show a heatmap for f_3-statistics of the form $f_3(San; A, B)$, where A and B are ancient and present-day East Eurasian and Native American populations in the SNP array dataset. Results are ordered according to a hierarchical clustering on the statistic values, and the corresponding dendrogram is shown. Each population is colored based on its broad continental ancestry, and Siberians and Native Americans are labeled according to their language family. Rows and columns corresponding to individuals sequenced in this study, as well as other relevant individuals are labeled.

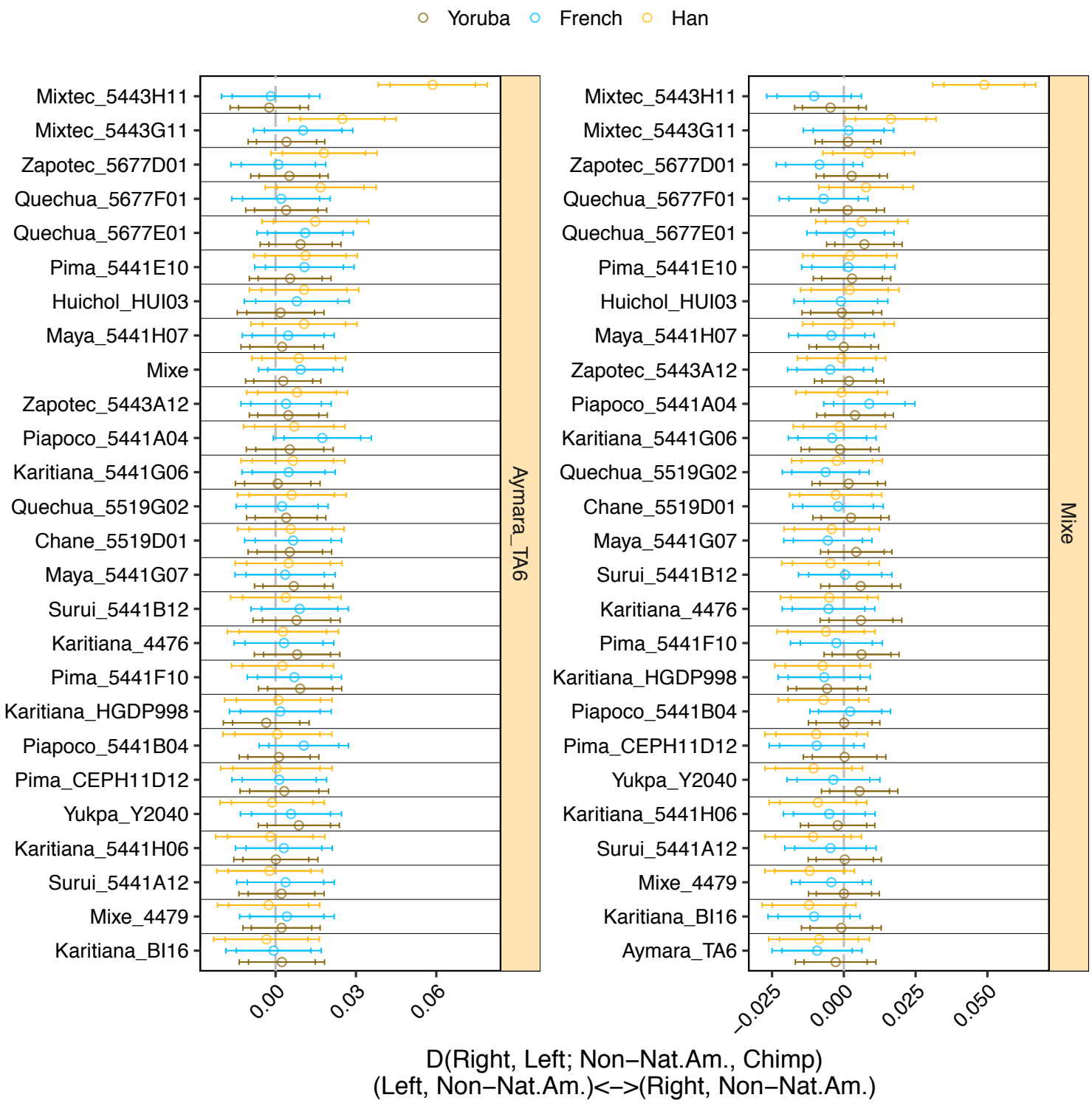
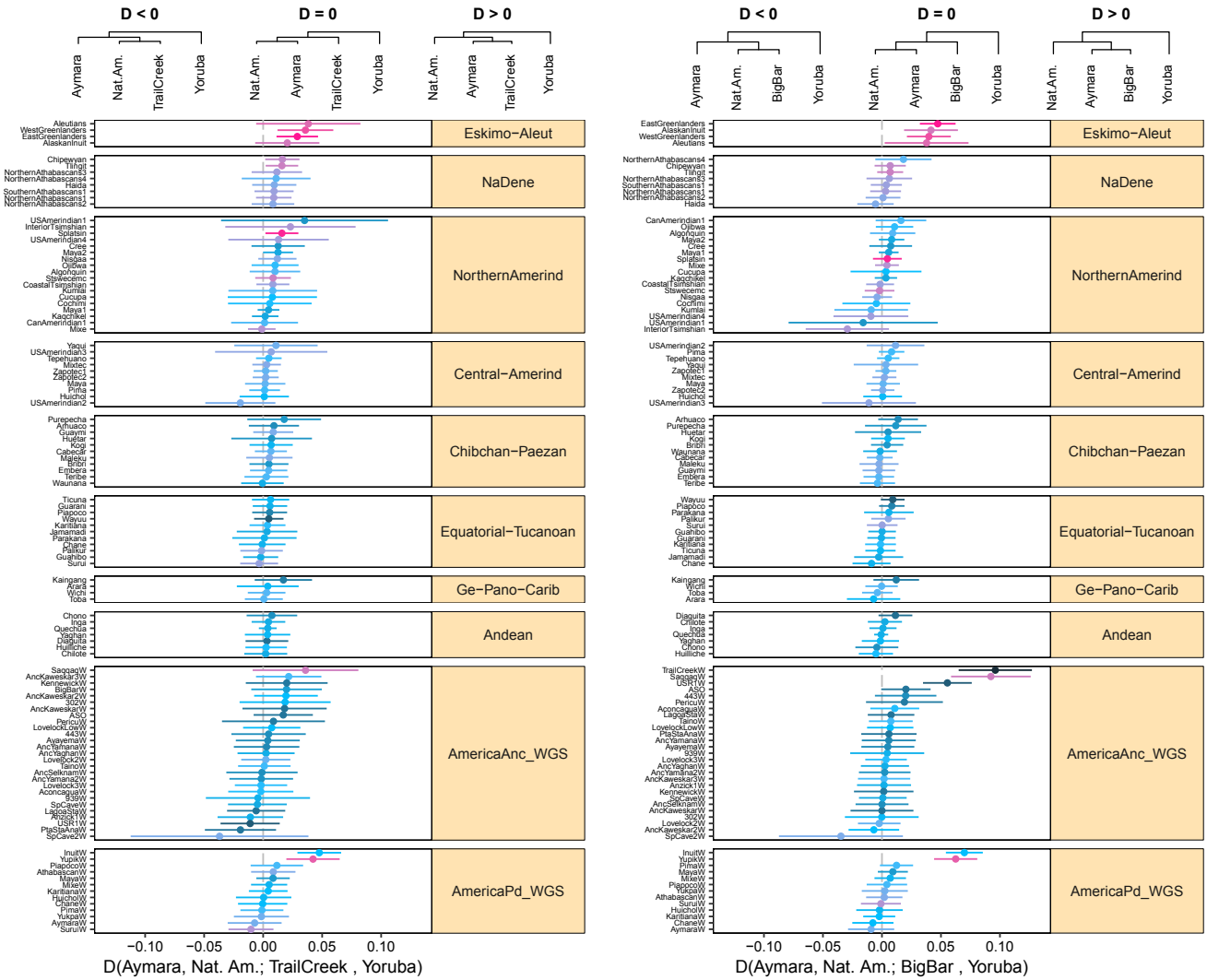


Figure S19. Called genotype-based *D*-statistics to identify 'non-admixed' present-day SNA individuals. We computed *D*-statistics of the form *D(SNA, Aymara/Mixe; Yoruba/French/Han, Chimp)* to identify SNA individuals in the whole genome dataset (Reference datasets) with 'non-Native American' admixture. Points represent *D* statistics and error bars represent one and ~3.3 standard errors (which corresponds to a p-value of ~0.001). The x-axis title shows the direction of potential excess allele sharing, depending on the two alternative hypotheses (D>0 or D<0).



Z−score for D(Aymara, Nat. Am.; Han, YRI)

−5.0 −2.5 0.0 2.5 5.0

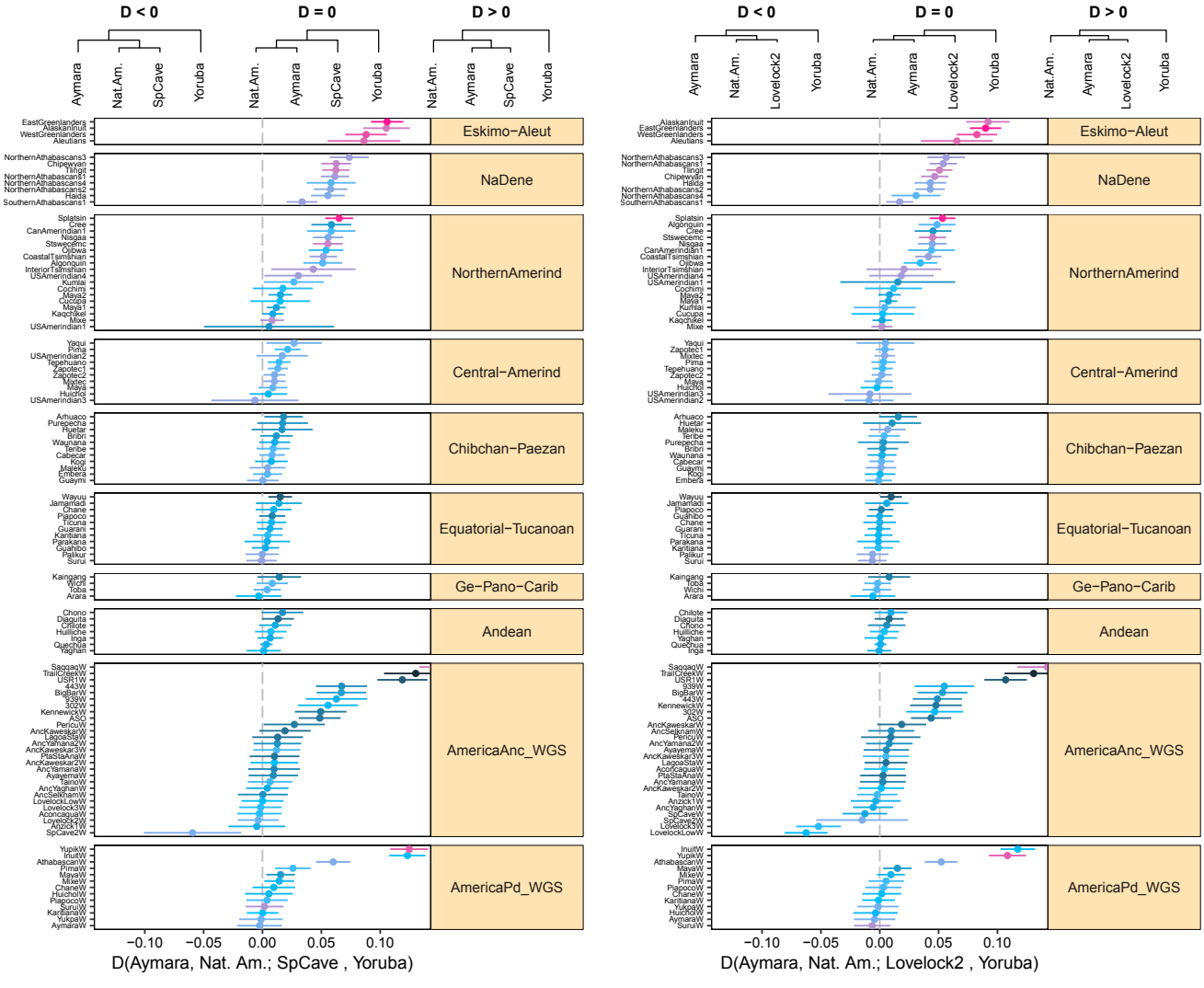
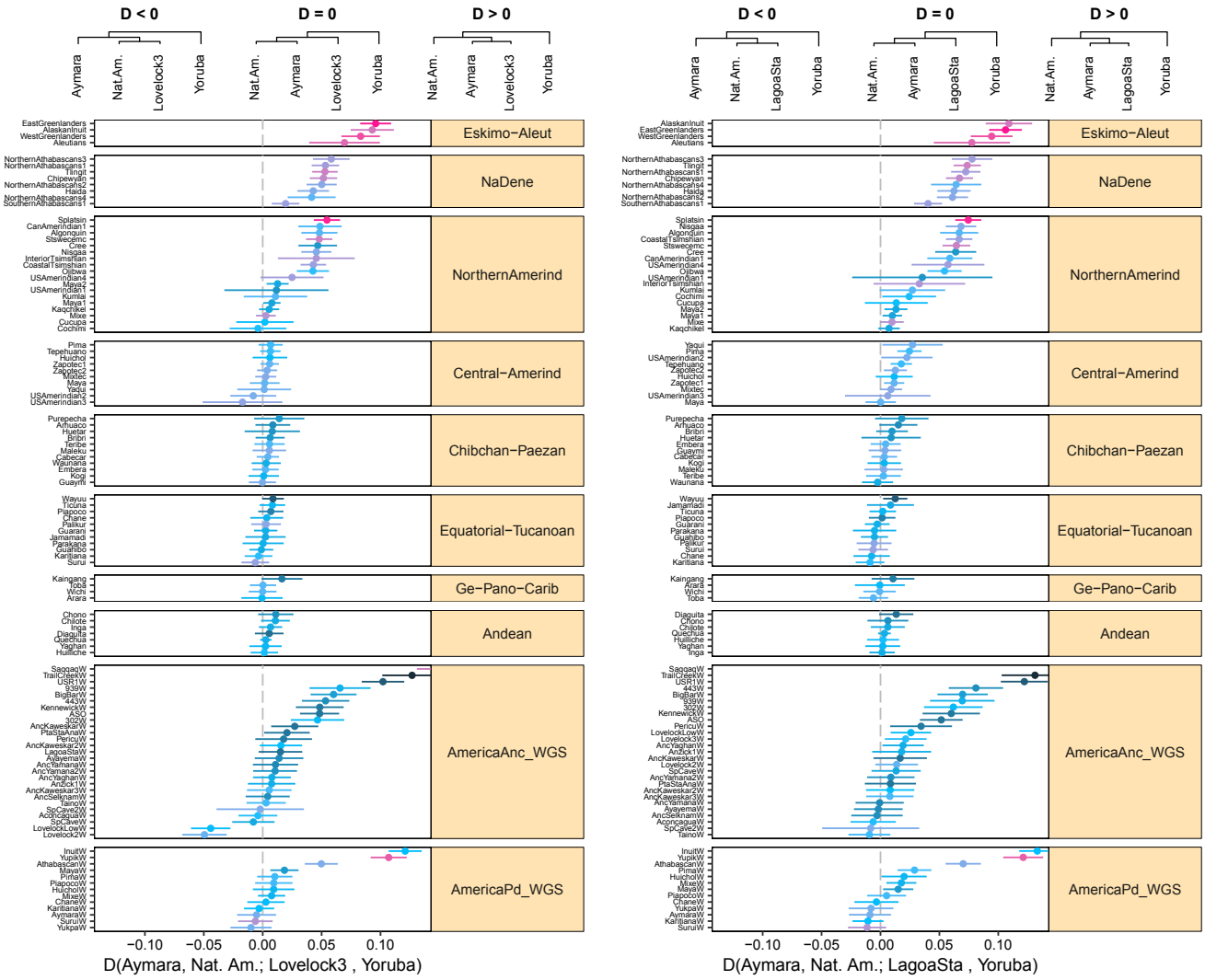


Figure S20. Called genotype-based *D*-statistics exploring the affinity between ancient individuals/populations and the NNA or SNA branches (part 1). We computed *D*-statistics of the form *D(Aymara, Native American; Ancient, Yoruba)* to test the relationship between the ancient genomes presented in this study and the basal NNA and SNA branches. The Aymara represent the latter. We test the null hypothesis (D=0) of whether the ancient individuals represent an outgroup to pairs of Native American populations (Aymara, Native American. In all plots, a schematic representation of the null hypothesis for the Z-test is shown in the middle (D=0) and the two possible outcomes of the alternative hypothesis are shown to the left (D<0) and the right (D>0). Colours correspond to the *Z*-score obtained for *D(Aymara, Native American; Han, Yoruba)*. Points represent D statistics and error bars represent ~3.3 standard errors (which corresponds to a p-value of ~0.001). For these tests, we used the SNP array dataset described in (Reference datasets). Native American populations are arranged by their language family and whole genomes are separated into ancient (Anc_WGS) and present-day (Pd_WGS).



Z−score for D(Aymara, Nat. Am.; Han, YRI)

−5.0 −2.5 0.0 2.5 5.0

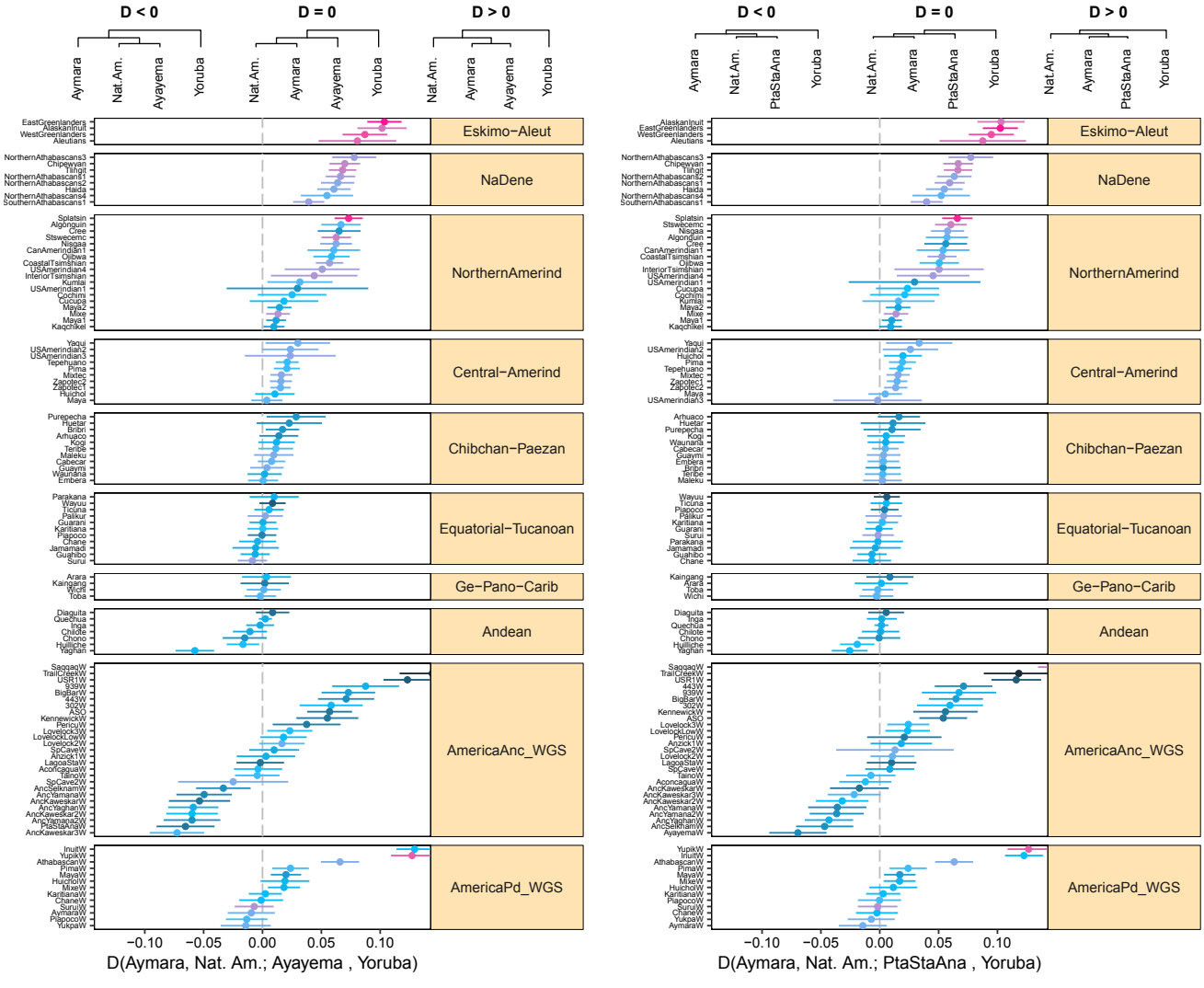


Figure S21. Called genotype-based *D*-statistics exploring the affinity between ancient individuals/populations and the NNA or SNA branches (part 2). Continued from Fig. S20.

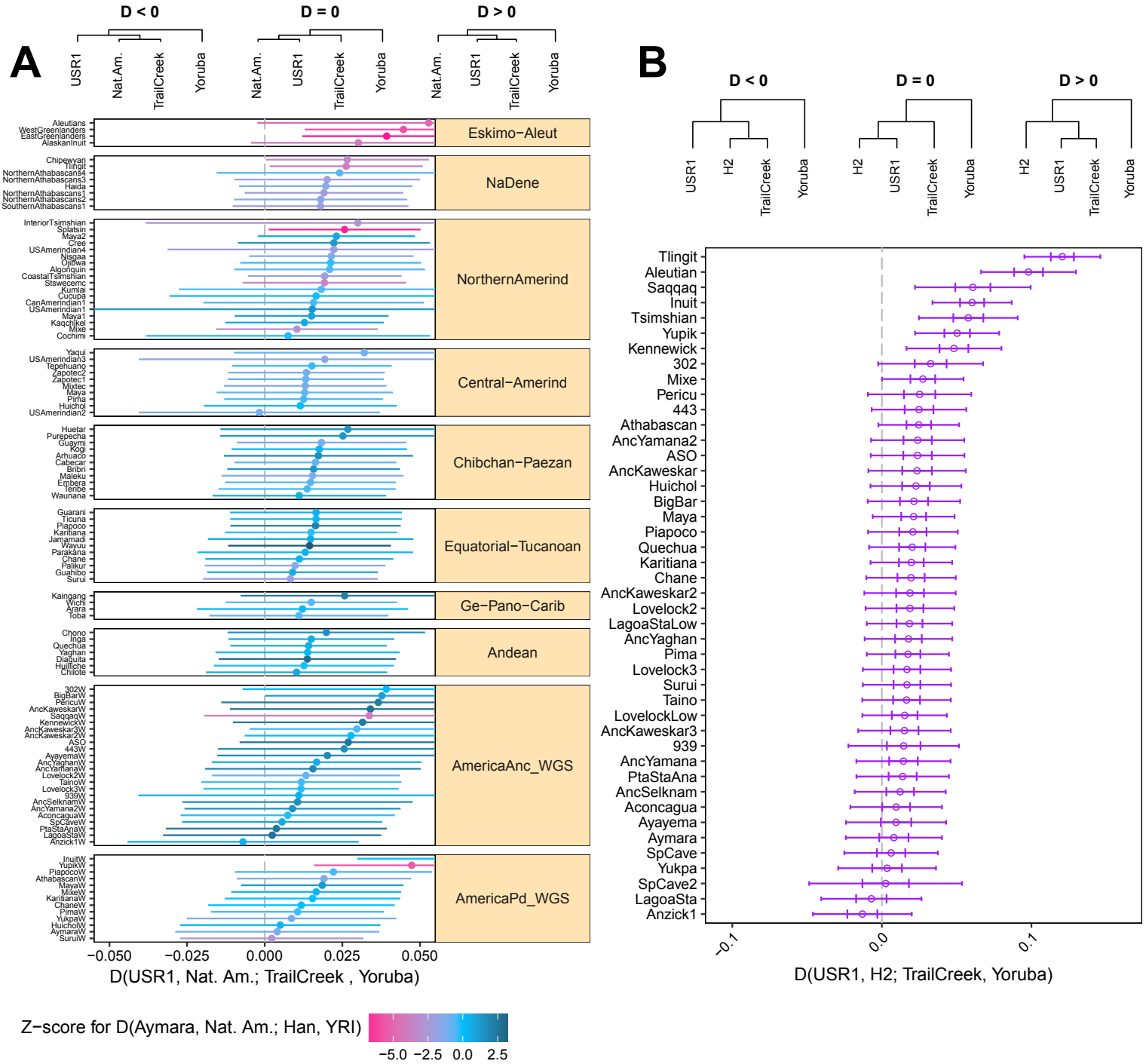


Figure S22. *D*-statistics testing the relationship between the Ancient Beringian Trail Creek and USR1 genomes. We computed *D*-statistics of the form *D(USR1, Native American; Trail Creek, Yoruba)* to test whether Trail Creek is more closely related to USR1 than to other Native Americans. **A.** Results based on the SNP array dataset described in (Reference datasets). Colors correspond to the *Z*-score obtained for *D(Aymara, Native American; Han, Yoruba)*. Native American populations are arranged by their language family and whole genomes are separated into ancient (Anc_WGS) and present-day (Pd_WGS). **B.** Results based on the whole-genome dataset described in (Reference datasets) and excluding transition polymorphisms. Points represent D statistics and error bars represent ~3.3 standard errors (which corresponds to a p-value of ~0.001). For both panels, a schematic representation of the null hypothesis for the Z-test is shown in the middle (D=0) and the two possible outcomes of the alternative hypothesis are shown to the left (D<0) and the right (D>0).

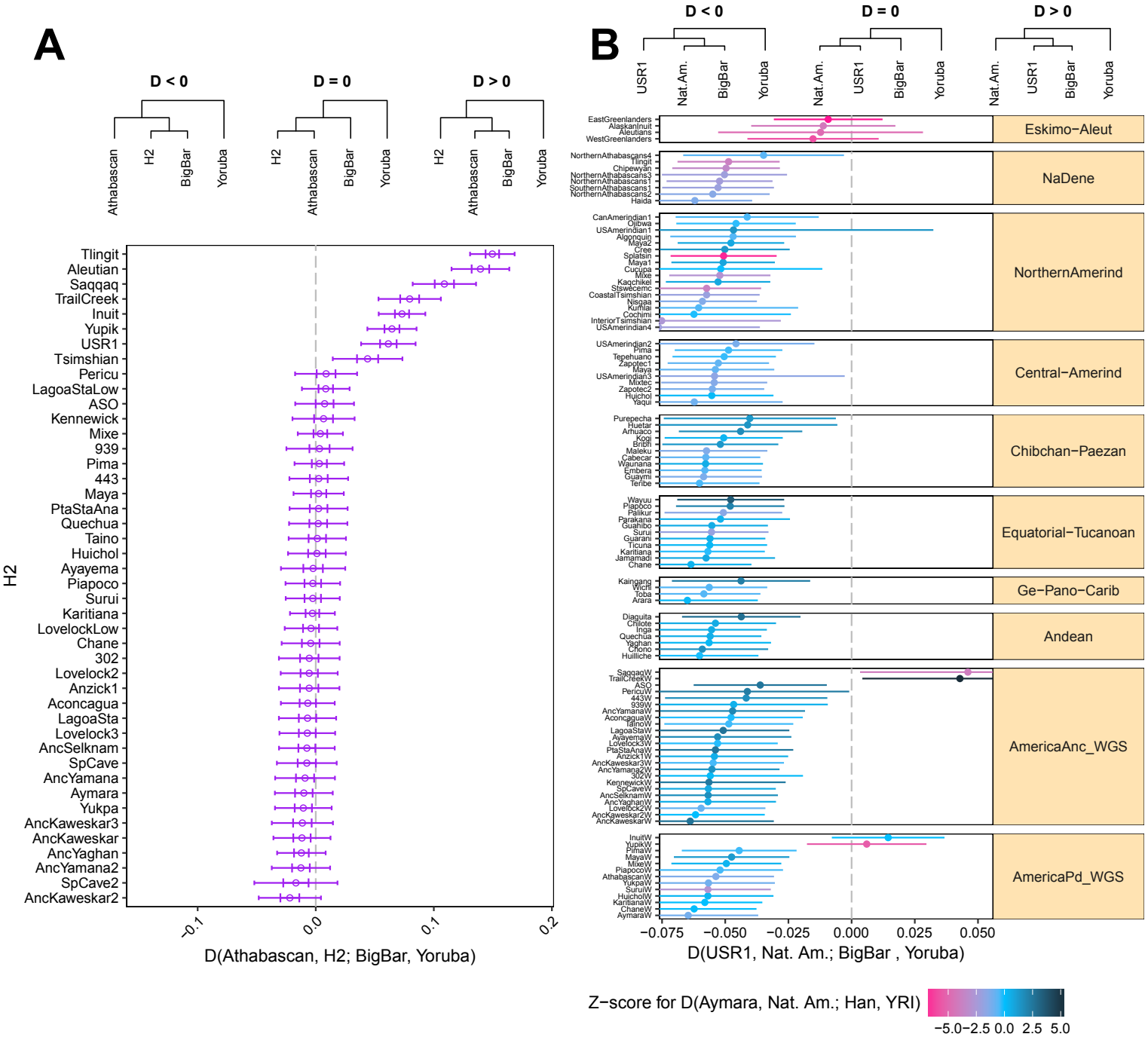


Figure S23. D-statistics testing the relationship between Big Bar and other Native Americans. A. We confirmed the results presented in Fig. S20 regarding the relationship between Big Bar and the NNA/SNA branches. In this case we tested whether Big Bar is differentially related to Athabascans (NNA) and other Native Americans by computing D-statistics of the form *D(Athabascans, Native American; Big Bar, Yoruba)*. **B.** We tested whether Big Bar is more closely related to USR1 than to other Native Americans by computing *D*-statistics of the form *D(USR1, Native American; Big Bar, Yoruba),* using the SNP array dataset described in (Reference datasets). Colors correspond to the *Z*-score obtained for *D(Aymara, Native American; Han, Yoruba)*. Native American populations are arranged by their language family and whole genomes are separated into ancient (Anc_WGS) and present-day (Pd_WGS). For both panels points represent D statistics and error bars represent ~3.3 standard errors (which corresponds to a p-value of ~0.001). A schematic representation of the null hypothesis for the Z-test is shown in the middle (D=0) and the two possible outcomes of the alternative hypothesis are shown to the left (D<0) and the right (D>0).

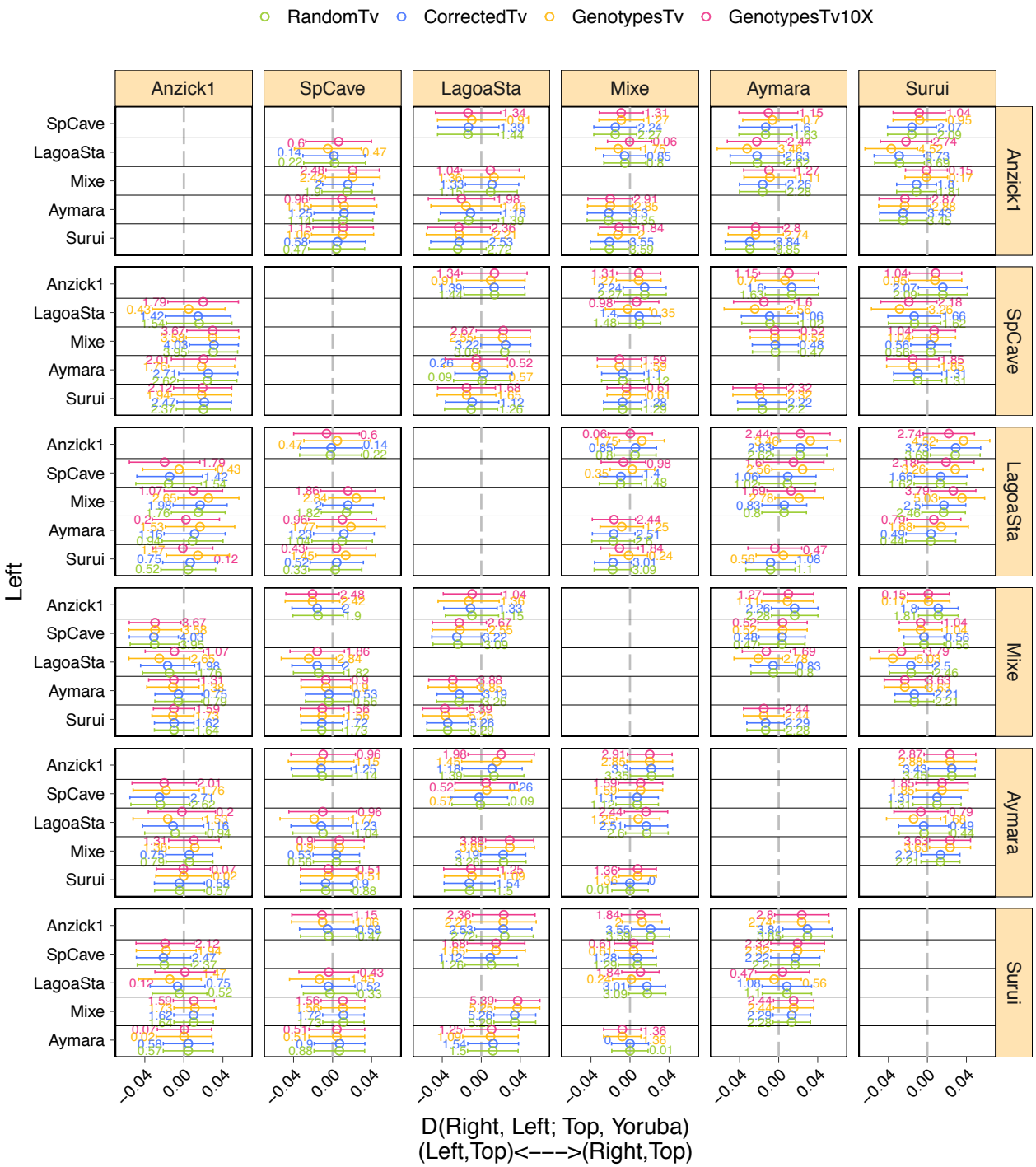


Figure S24. D-statistics testing the relationship between representative South Native American populations (Part 1). We computed *D*-statistics of the form *D(Native American$_i$, Native American$_j$; Native American$_k$, Yoruba)* to assess the relationship between the Native American populations included in the admixture graph shown in Fig. 3A. In the figure, ingroups (*Native American$_i$, Native American$_j$*) are shown in the left and right, and '*Native American$_k$*' is shown in the top panel. In order to control for and explore the potential bias deriving from differential error rates in the Native American genomic data, we computed *D*-statistics based on random alleles excluding transition polymorphisms (green), called genotypes excluding transition polymorphisms (orange) and called genotypes excluding transition polymorphisms but only including sites with a minimum depth of coverage of 10X (red). Additionally, we computed error corrected-*D*-statistics based on transversion polymorphisms (blue). Points represent *D*-statistics and error bars represent ~3.3 standard errors (which corresponds to a p-value of ~0.001). For each test, we show the absolute *Z*-score beside its corresponding *D* value. For all tests, we used genome-wide data. In this case, the Lagoa Santa population is represented by the high-depth 'Sumidouro5' individual.

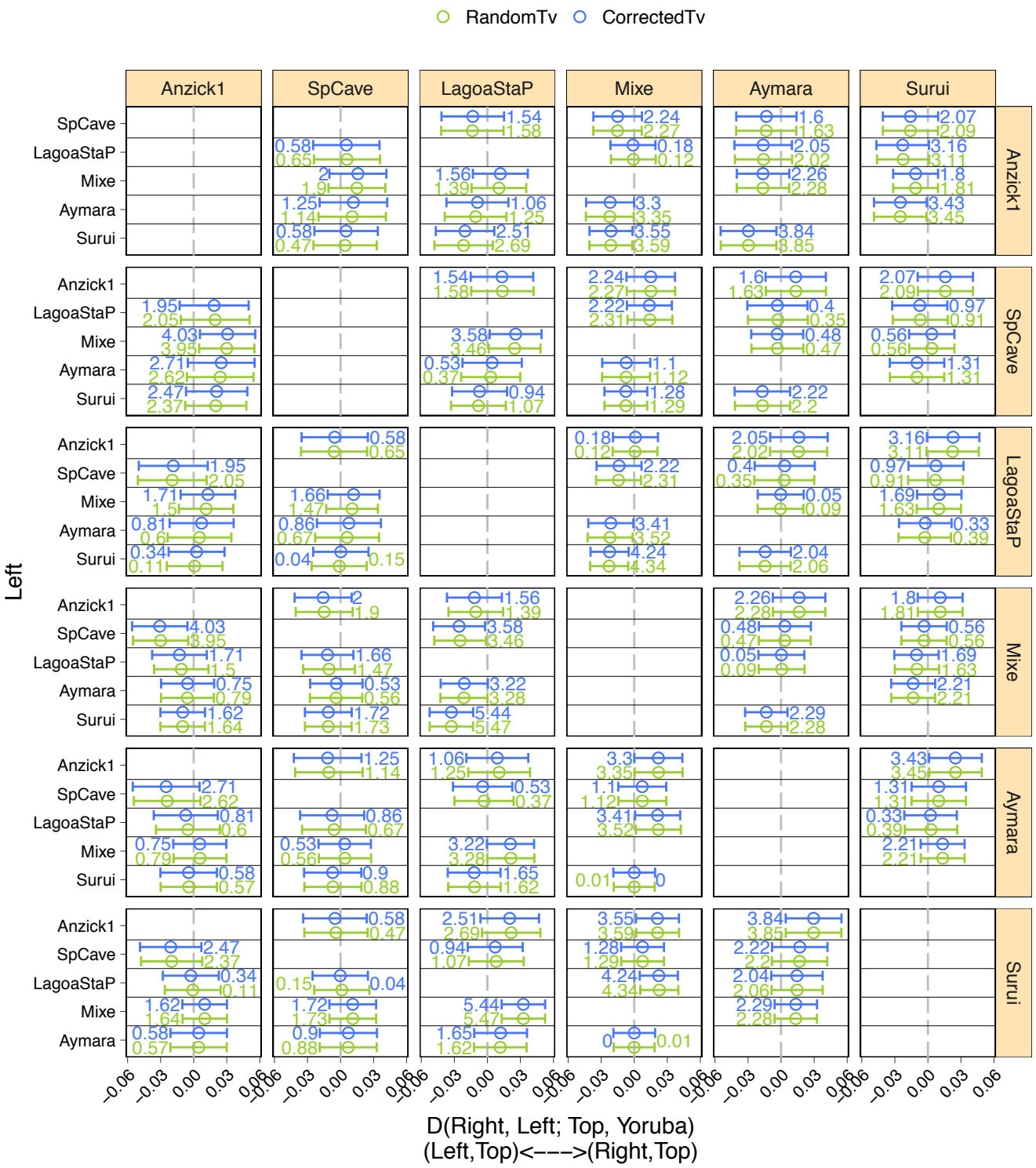


Figure S25. D-statistics testing the relationship between representative South Native American populations (Part 2). This figure is equivalent to Fig. S24. In this case, the Lagoa Santa population is represented by a pool of the five sequenced individuals. Due to low depth of coverage, calling genotypes with confidence for all Lagoa Santa individuals is not feasible. Therefore, we only show results for corrected-*D*-statistics based on transversion polymorphisms (blue) and standard D-statistics based on random alleles excluding transition polymorphisms (green).

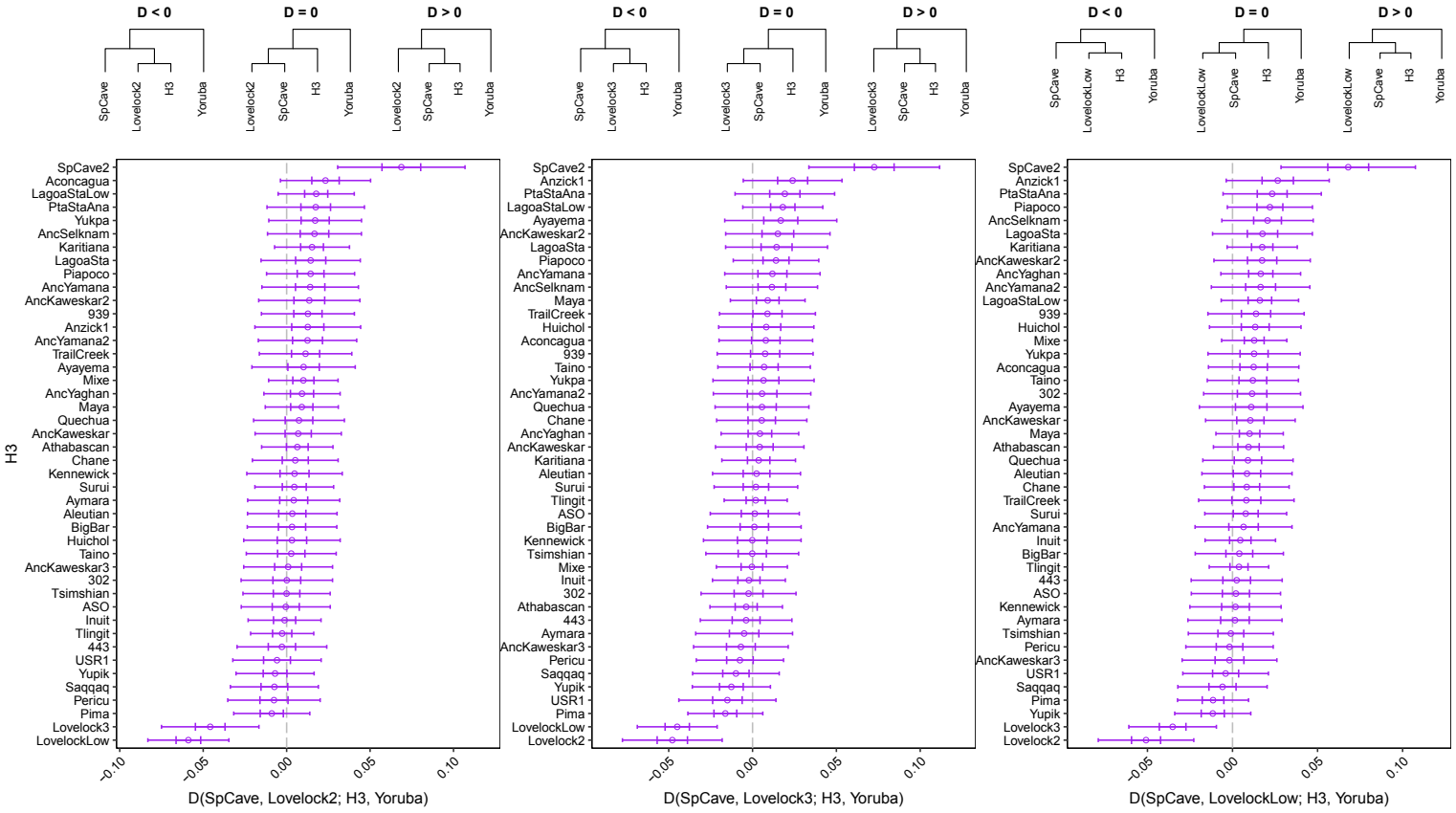


Figure S26. *D*-statistics testing the relationship between Early and Late Holocene North American Great Basin indiviuals. We computed D-statistics of the form *D(Lovelock, Spirit Cave; Native American, Yoruba)* to test whether the Spirit Cave (Early Holocene) and Lovelock (Late Holocene) individuals form a clade with respect to other Native Americans. Due to their variable age and sequencing depth, we tested each Lovelock individual independently except for the low-depth individuals 1 and 4, which we pooled into the 'LovelockLow' population. All tests are based on the whole-genome dataset described in (Reference datasets) and excluding transition polymorphisms. Points represent D statistics and error bars represent ~3.3 standard errors (which corresponds to a p-value of ~0.001). For each panel, a schematic representation of the null hypothesis for the Z-test is shown in the middle (D=0) and the two possible outcomes of the alternative hypothesis are shown to the left (D<0) and the right (D>0).

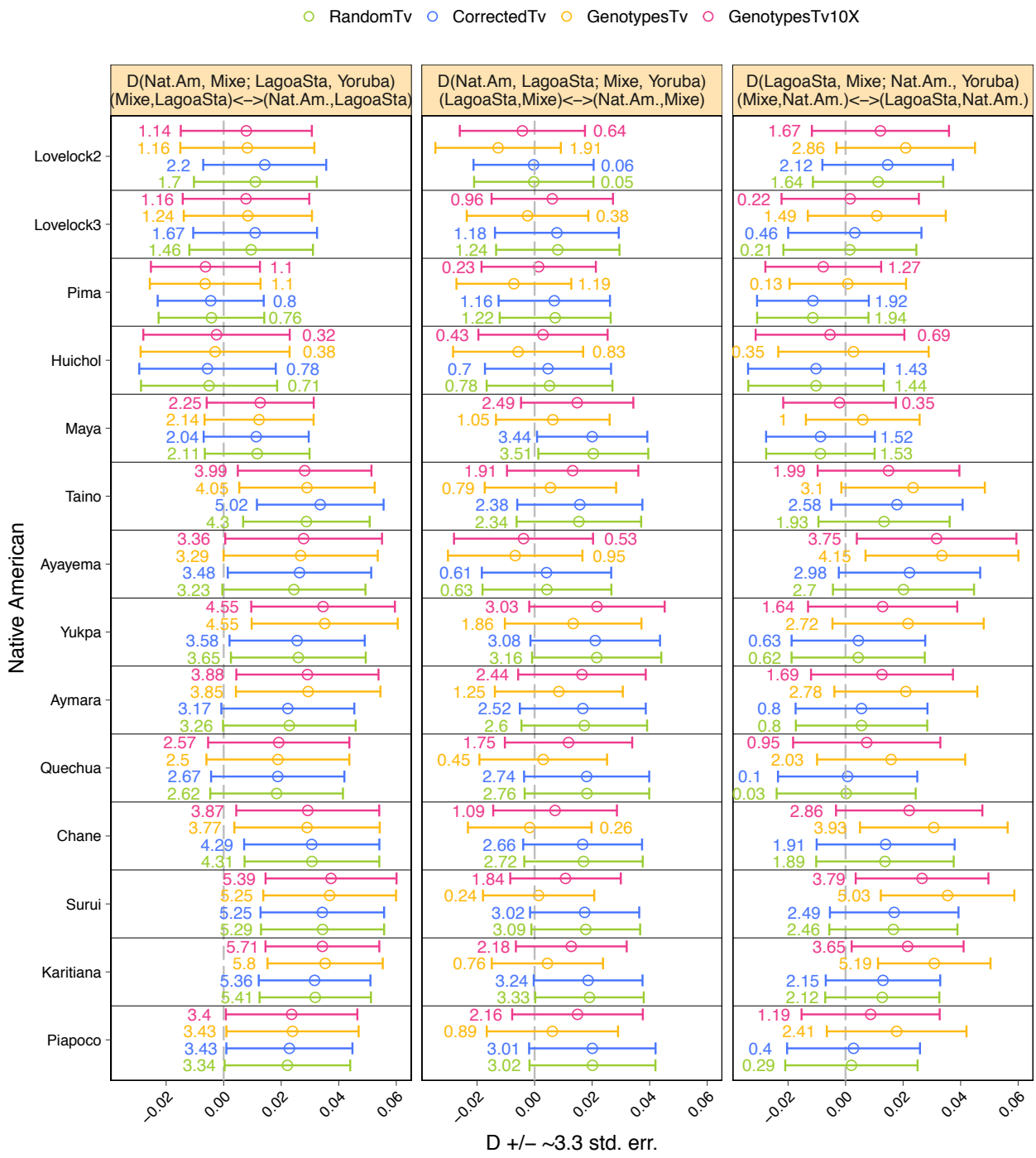


Figure S27. *D*-statistics testing the relationship between Lagoa Santa, Mixe and other SNA genomes (Part 1). We computed D-statistics of the form *D(SNA, Mixe; Lagoa Santa, Yoruba)*, *D(SNA, Lagoa Santa; Mixe, Yoruba)* and *D(Lagoa Santa, Mixe; SNA, Yoruba)* to test whether these populations could be modeled with a simple tree. Each arrangement is shown in the top panel, together with an indication of the pair of populations with excess allele sharing, depending on the sign of *D*. In order to control for and explore the potential bias deriving from differential error rates in the Native American genomic data, we computed *D*-statistics based on random alleles excluding transition polymorphisms (green), called genotypes excluding transition polymorphisms (orange) and called genotypes excluding transition polymorphisms but only including sites with a minimum depth of coverage of 10X (red). Additionally, we computed error corrected-*D*-statistics based on transversion polymorphisms (blue). Points represent *D*-statistics and error bars represent ~3.3 standard errors (which corresponds to a p-value of ~0.001). For each test, we show the absolute *Z*-score beside its corresponding *D* value. For all tests, we used genome-wide data. In this case, the Lagoa Santa population is represented by the high-depth 'Sumidouro5' individual.

Figure S28. *D*-statistics testing the relationship between Lagoa Santa, Mixe and other SNA genomes (Part 2). This figure is equivalent to Fig. S27. In this case, the Lagoa Santa population is represented by a pool of the five sequenced individuals. Due to low depth of coverage, calling genotypes with confidence for all Lagoa Santa individuals is not feasible. Therefore, we only show results for corrected-*D*-statistics based on transversion polymorphisms (blue) and standard D-statistics based on random alleles excluding transition polymorphisms (green).

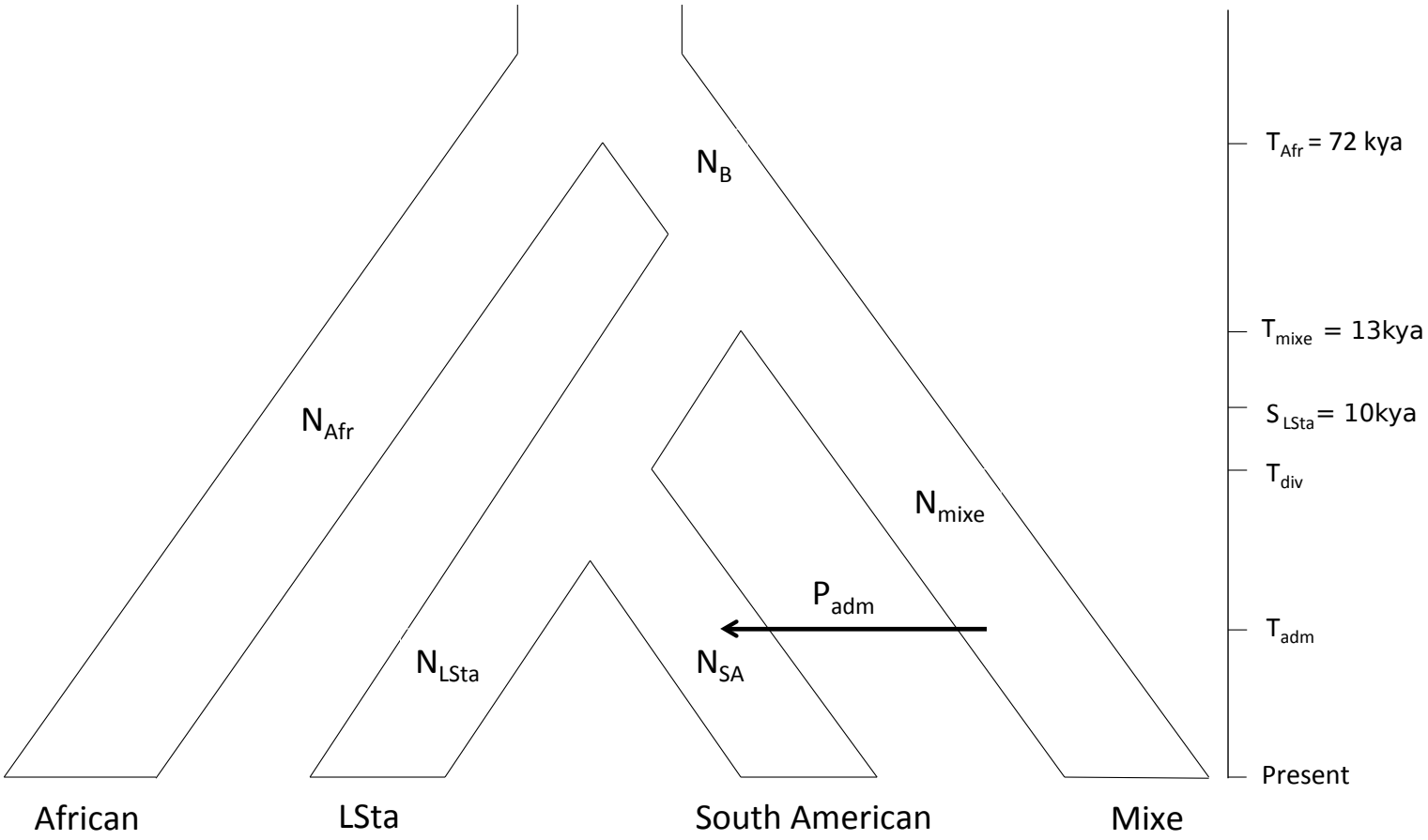


Figure S29. Schematic representation of the demographic model that was used for coalescent simulations testing *D*-statistics sensitivity. Demographic model used for simulations assessing the sensitivity of D-statistics aimed at detecting Lagoa Santa- and Mixe-related admixture in South American populations. Constant population sizes were set to N_{afr}=20,000 for Africans, N_B=1,600 for the out of Africa population, and N_{Mixe}=N_{LSta}=$N_{SouthAme}$=1,500 for Mixe, Lagoa Santa and South Americans respectively Divergence time between Mixe and Lagoa Santa was set to T_{mixe}=13kya. We simulated data across a range of values for the following parameters: the the divergence time (in kya) between Lagoa Santa and South Americans (T_{div}={8,500, 10,500}), the time of admixture from Mixe into South Americans (T_{adm}={2,500, 5,000, 7,500}), and the corresponding admixture proportion (P_{adm}={0.2, 0.4, 0.6, 0.8}). For each model, for each parameter combination, we simulated ten replicates. In order for the simulations to resemble real data, each replicate involves the following number of diploid individuals: 2 Africans, 1 Lagoa Santa, 3 Mixe and 2 South Americans, which were downsampled in order to keep the same number of observed segregating sites.

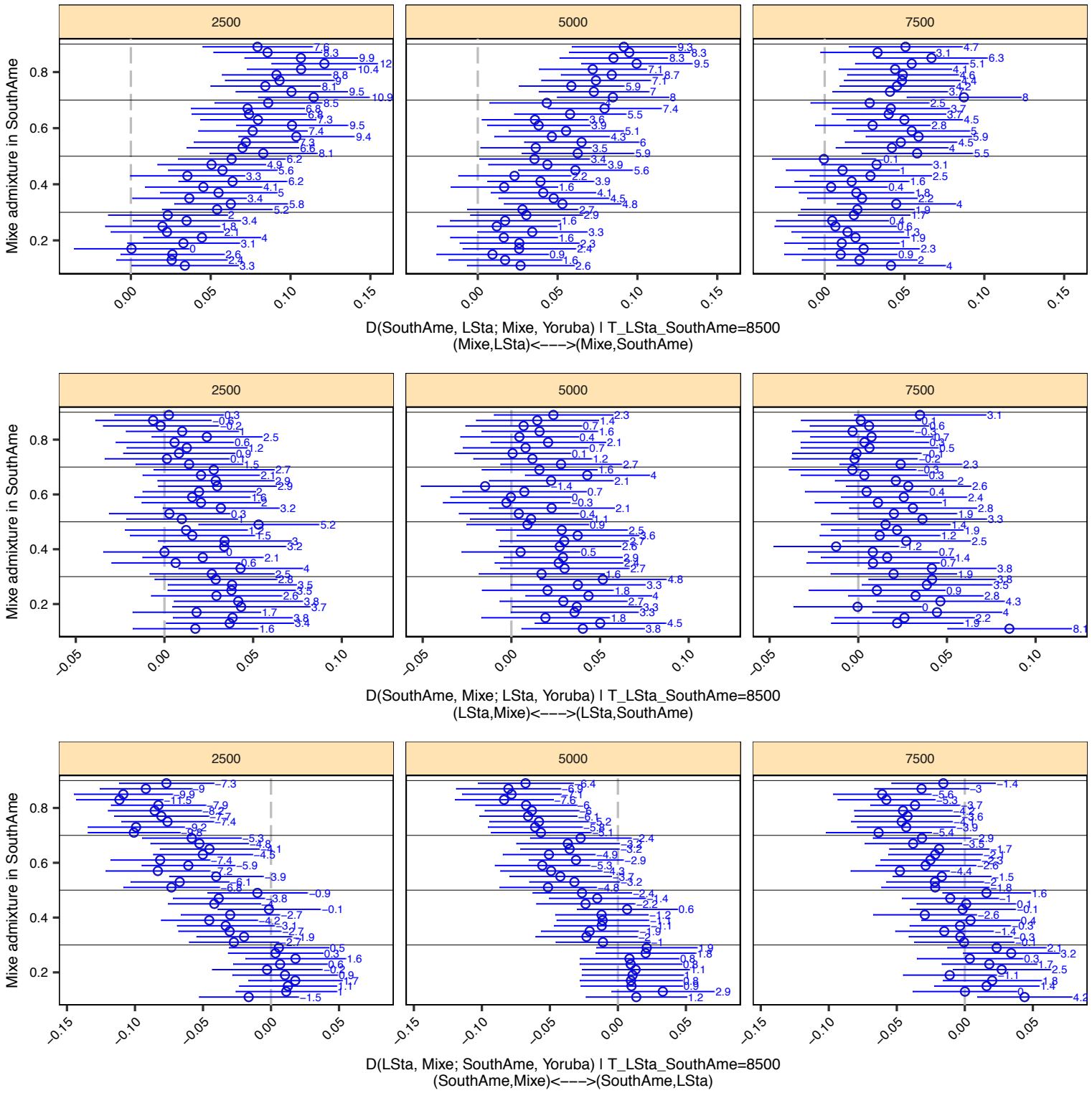


Figure S30. Simulation study results testing the sensitivity of *D*-statistics aimed at detecting Lagoa Santa- and Mixe-related admixture in South American populations (Part 1). We simulated data under the scenario shown in Fig. S29. For each combination of T_{div} (indicated in the *x*-axis), T_{adm} (top labels) and P_{adm} (*y*-axis), for each replicate, we computed *D(South American, Lagoa Santa; Mixe, Yoruba), D(South American, Mixe; Lagoa Santa, Yoruba)* and *D(Lagoa Santa, Mixe; South American, Yoruba)*. Times are given in kya. Points represent *D*-statistics and error bars represent ~3.3 standard errors (which corresponds to a *p*-value of ~0.001). For each test, we show the absolute *Z*-score beside its corresponding *D* value. In this case we show results for simulations where T_{div}=8,500.

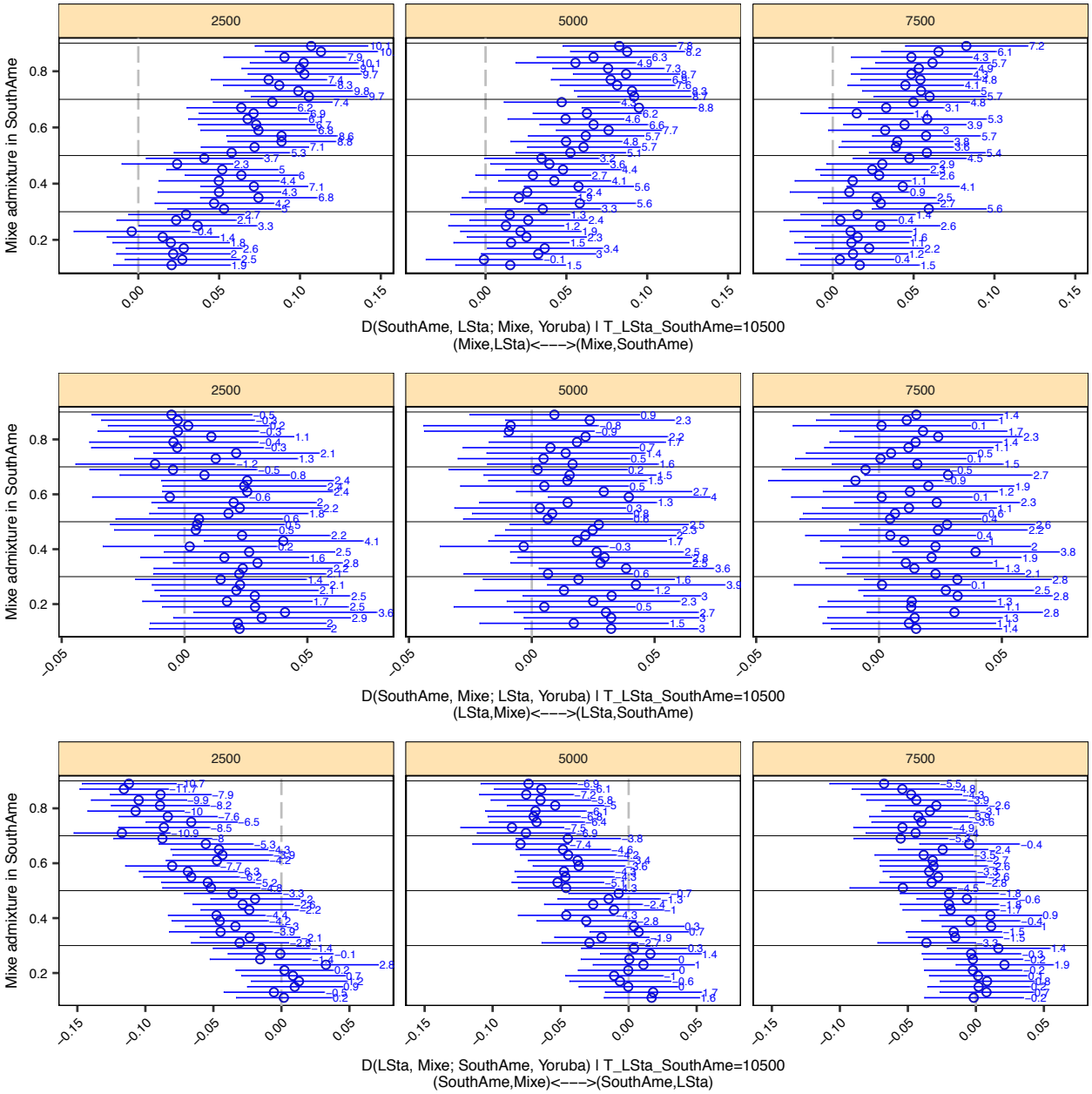


Figure S31. Simulation study results testing the sensitivity of *D*-statistics aimed at detecting Lagoa Santa- and Mixe-related admixture in South American populations (Part 2). This figure is equivalent to Fig. S30. In this case, we show results for T_{div}=10,500.

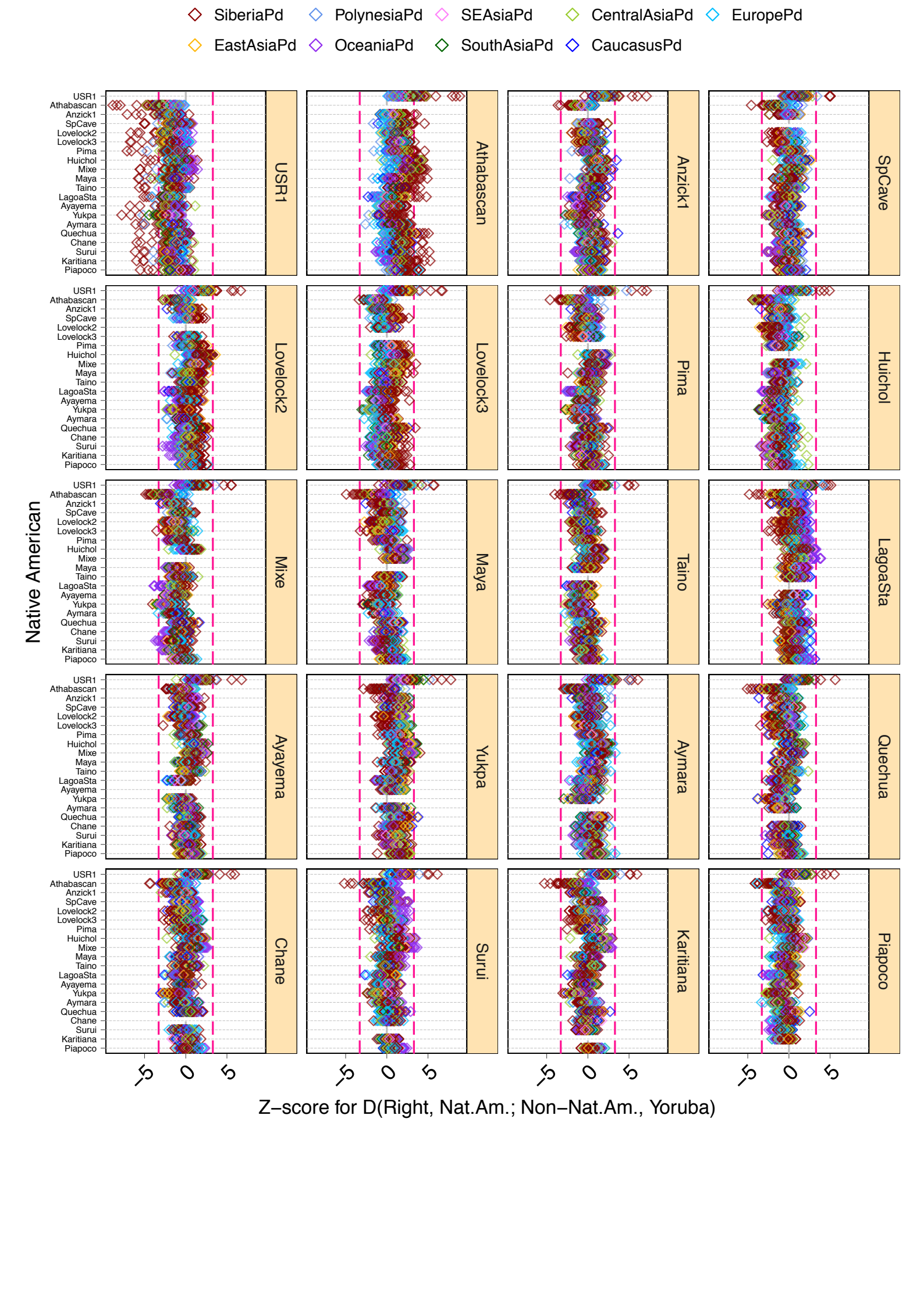

SiberiaPd
PolynesiaPd
SEAsiaPd
CentralAsiaPd
EuropePd
EastAsiaPd
OceaniaPd
SouthAsiaPd
CaucasusPd
Native American
USR1
Athabascan
Anzick1
SpCave
Lovelock2
Lovelock3
Pima
Huichol
Mixe
Maya
Taino
LagoaSta
Ayayema
Yukpa
Aymara
Quechua
Chane
Surui
Karitiana
Piapoco
−5
0
5
Z−score for D(Right, Nat.Am.; Non−Nat.Am., Yoruba)

Figure S32. *Z*-scores for *D*-statistics testing the symmetry between pairs of Native American populations, relative to Eurasian populations (Part 1). We computed *D*-statistics of the form *D(Native American, Native American; Eurasian, Yoruba)* to test whether a given Native American group carries excess 'non-Native American' ancestry compared to other Native Americans. For each statistic, we obtained a *Z*-score (diamonds) based on a weighted block-jackknife procedure over 5Mb blocks. Vertical lines represent ~3.3 and ~-3.3 (which correspond to a p-value of ~0.001). Points are colored based on the broad continental ancestry of the Eurasian group. All tests are based on the whole-genome dataset described in (Reference datasets) and excluding transition polymorphisms. In this case we only show results for present-day Eurasian populations.

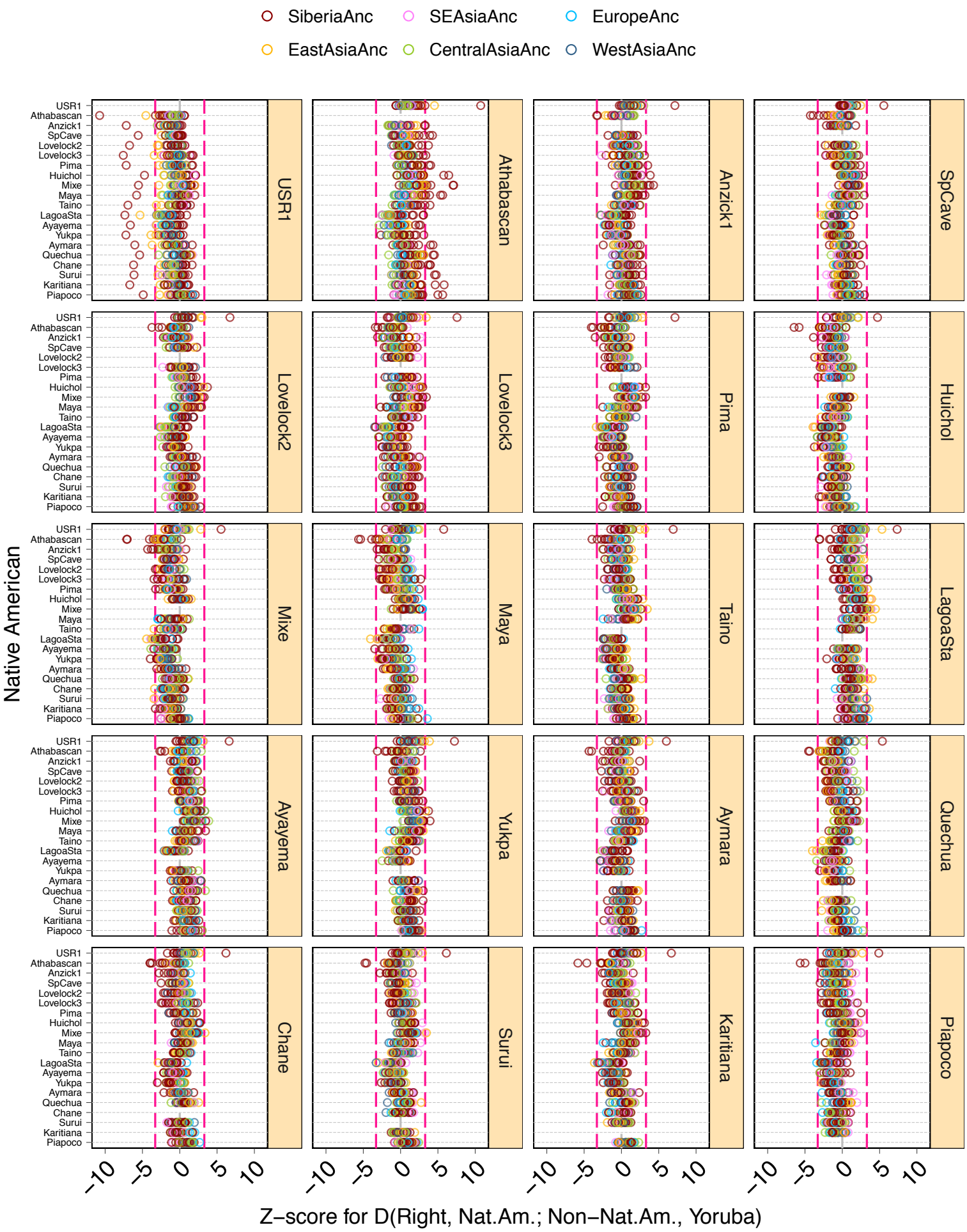


Figure S33. *Z*-scores for *D*-statistics testing the symmetry between pairs of Native American populations, relative to Eurasian populations (Part 2). This figure is equivalent to Fig. S32. In this case, we only show results for Eurasian populations represented by ancient DNA data.

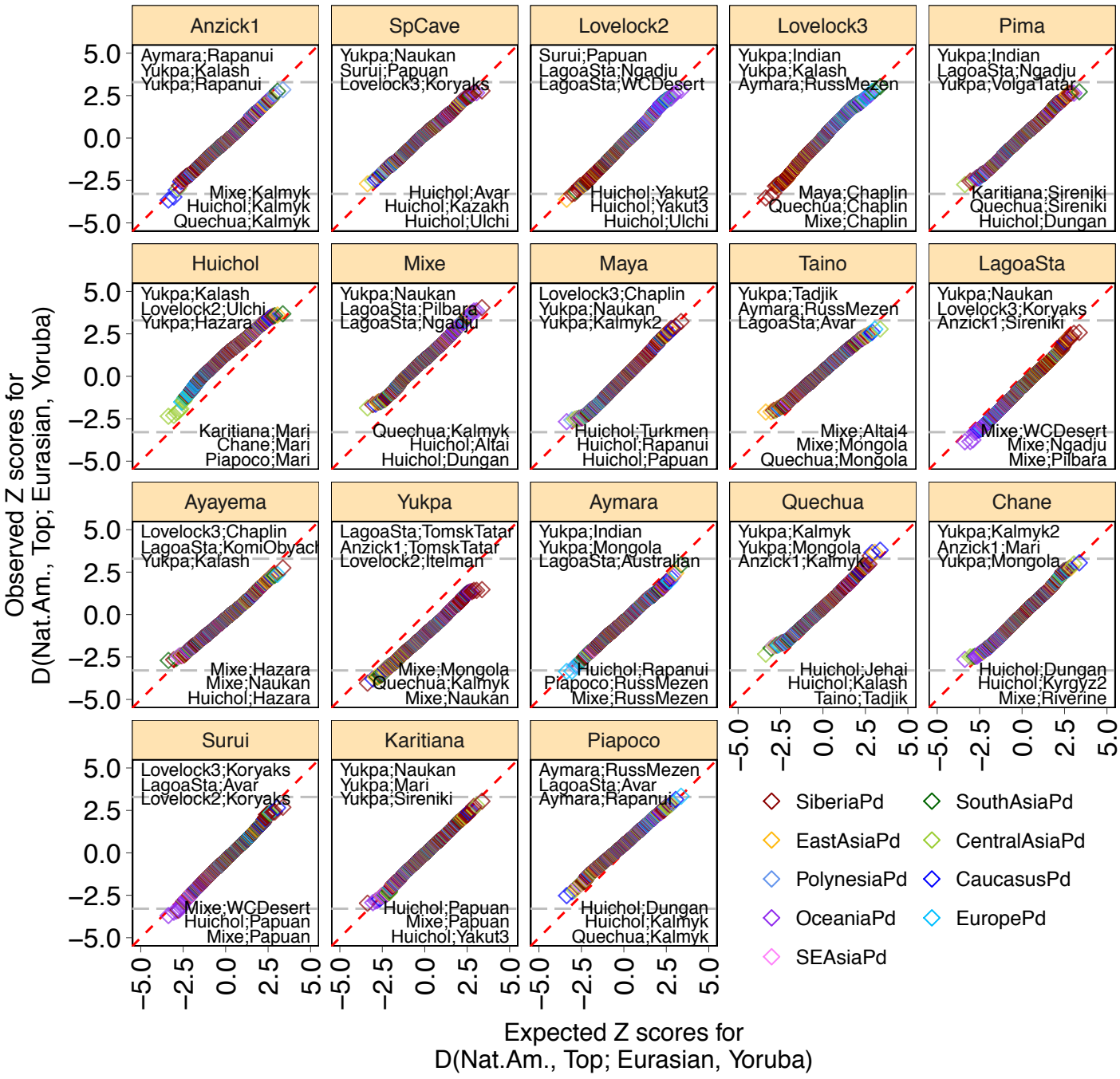


Figure S34. Quantile-quantile plots testing the symmetry between pairs of Native American populations, relative to Eurasian populations. We computed D-statistics of the form *D(SNA, SNA; Eurasian, Yoruba)* and built quantile–quantile plots comparing observed *Z*-scores to the expected normal distribution under the null hypothesis (all SNA populations are equally related to Eurasian populations). For each SNA population indicated in the top labels, we computed all possible *D(SNA, 'Top'; Eurasian, Yoruba)*. In each panel, we show the (SNA;Eurasian) combinations that produced the three highest and lowest *Z*-scores, *e.g.*, *D(Mixe, Lagoa Santa; Pilbara, Yoruba)* produced the minimum *Z*-score among all possible *D(SNA, Lagoa Santa; Eurasian, Yoruba)*. The expected normal distribution under the null hypothesis was computed for each SNA ('Top') population separately. For each statistic, *Z*-scores (diamonds) were obtained based on a weighted block-jackknife procedure over 5Mb blocks. Horizontal lines represent ~3.3 and ~-3.3 (which correspond to a *p*-value of ~0.001). Points are colored based on the broad continental ancestry of the Eurasian group. All tests are based on the whole-genome dataset described in (Reference datasets) and excluding transition polymorphisms. In this case we only show results for present-day Eurasian populations.

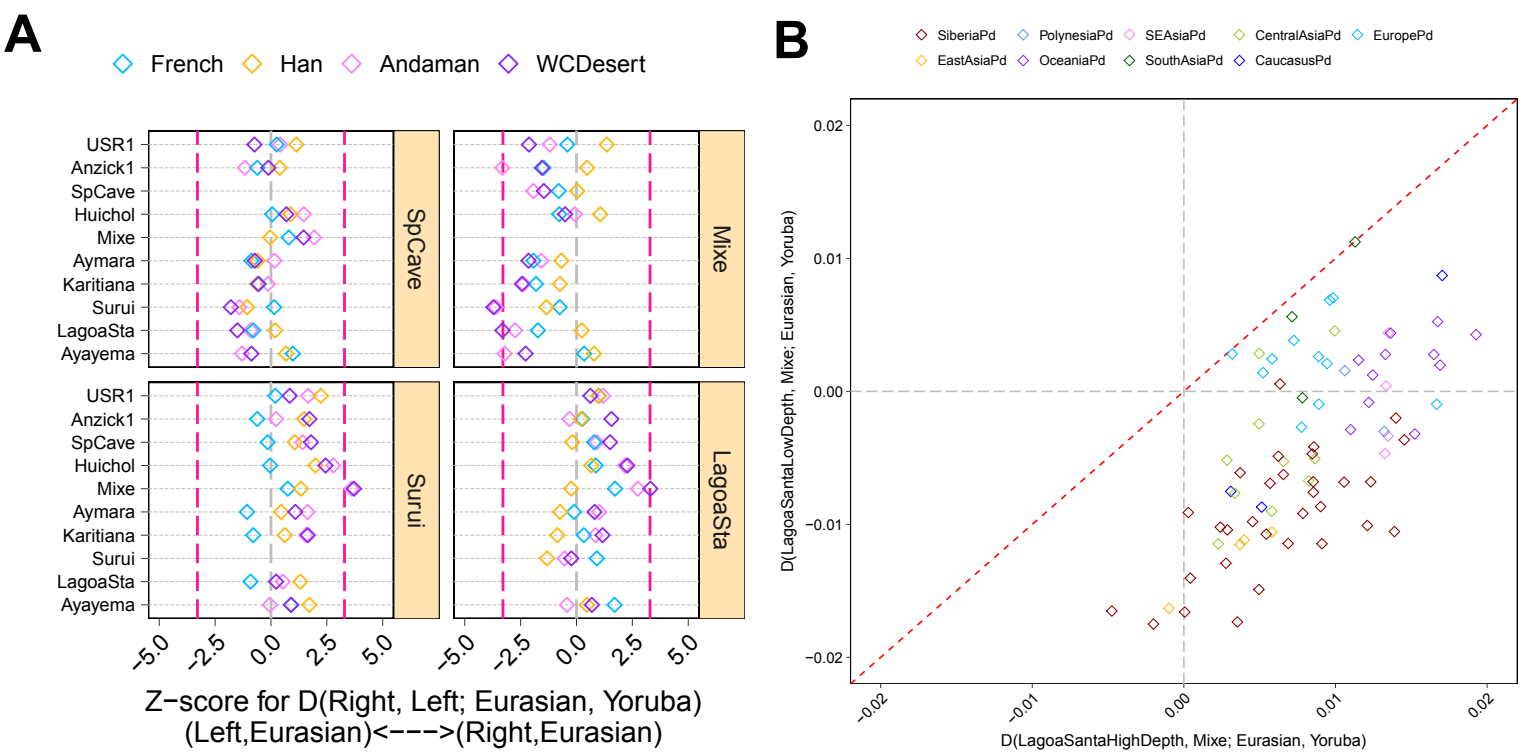


Figure S35. Validation of D-statistics showing Australasian ancestry in Lagoa Santa. A. Error-corrected D-statistics similar to those shown in Figure 5A. In this case we selected representative Eurasian and Native American populations. **B.** D-statistics showing that low- and high-depth Lagoa Santa individuals are similarly related to Eurasians. We computed D-statistics of the form *D(Lagoa Santa, Mixe; Eurasian, Yoruba)* for the high-depth Lagoa Santa genome (called genotypes) and for a pool of the four low-depth Lagoa Santa genomes (random allele). Points are colored based on the broad continental ancestry of the Eurasian group. All tests are based on the whole-genome dataset described in (Reference datasets) and excluding transition polymorphisms. In this case we only show results for present-day Eurasian populations.

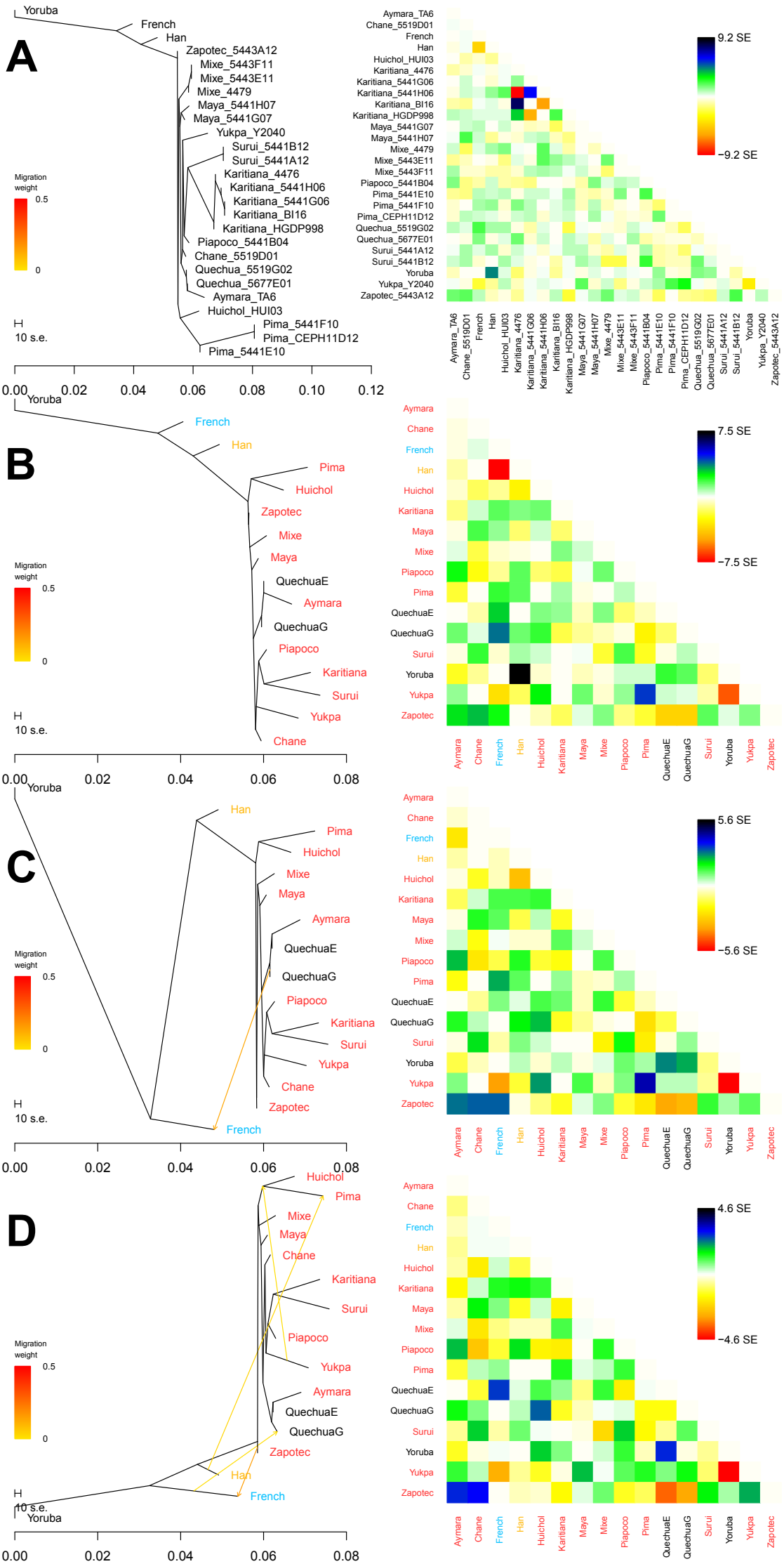

A
Yoruba
French
Han
Zapotec_5443A12
Mixe_5443F11
Mixe_5443E11
Mixe_4479
Maya_5441H07
Maya_5441G07
Yukpa_Y2040
Surui_5441B12
Surui_5441A12
Karitiana_4476
Karitiana_5441H06
Karitiana_5441G06
Karitiana_BI16
Karitiana_HGDP998
Piapoco_5441B04
Chane_5519D01
Quechua_5519G02
Quechua_5677E01
Aymara_TA6
Huichol_HUI03
Pima_5441F10
Pima_CEPH11D12
Pima_5441E10
Migration weight
0.5
0
H
10 s.e.
0.00
0.02
0.04
0.06
0.08
0.10
0.12
9.2 SE
−9.2 SE
B
Yoruba
French
Han
Pima
Huichol
Zapotec
Mixe
Maya
QuechuaE
Aymara
QuechuaG
Piapoco
Karitiana
Surui
Yukpa
Chane
Migration weight
0.5
0
H
10 s.e.
0.00
0.02
0.04
0.06
0.08
7.5 SE
−7.5 SE
C
Yoruba
Han
Pima
Huichol
Mixe
Maya
Aymara
QuechuaE
QuechuaG
Piapoco
Karitiana
Surui
Yukpa
Chane
Zapotec
French
Migration weight
0.5
0
H
10 s.e.
0.00
0.02
0.04
0.06
0.08
5.6 SE
−5.6 SE
D
Huichol
Pima
Mixe
Maya
Chane
Karitiana
Surui
Piapoco
Yukpa
Aymara
QuechuaE
QuechuaG
Zapotec
Han
French
Yoruba
Migration weight
0.5
0
H
10 s.e.
0.00
0.02
0.04
0.06
0.08
4.6 SE
−4.6 SE

Figure S36. TreeMix admixture graphs for identifying 'non-admixed' present-day SNA individuals. A. A maximum-likelihood tree (zero admixture edges) relating present-day SNA individuals from the whole-genome dataset (Reference datasets), for which *D(SNA, Aymara/Mixe; Yoruba/French/Han, Chimp)* did not yield significant deviations from *D*=0 (Fig. S19). In the following analyses, we pooled individuals according to their population label only if they form a monophyletic clade in this tree, *e.g.,* we did not pool Aymara and Quechua. **B-D**. We fitted admixture graphs with zero to five admixture edges to the dataset used in a., after pooling individuals into 'populations'. We show results for zero, one and four admixture edges. Based on these results, we excluded the Zapotec and a Quechua individual ('QuechuaG') from other analyses as they potentially carry admixture from Eurasian groups. On the left, we show the admixture graphs produced by TreeMix. On the right, we show the pairwise residuals after fitting the maximum likelihood graph.

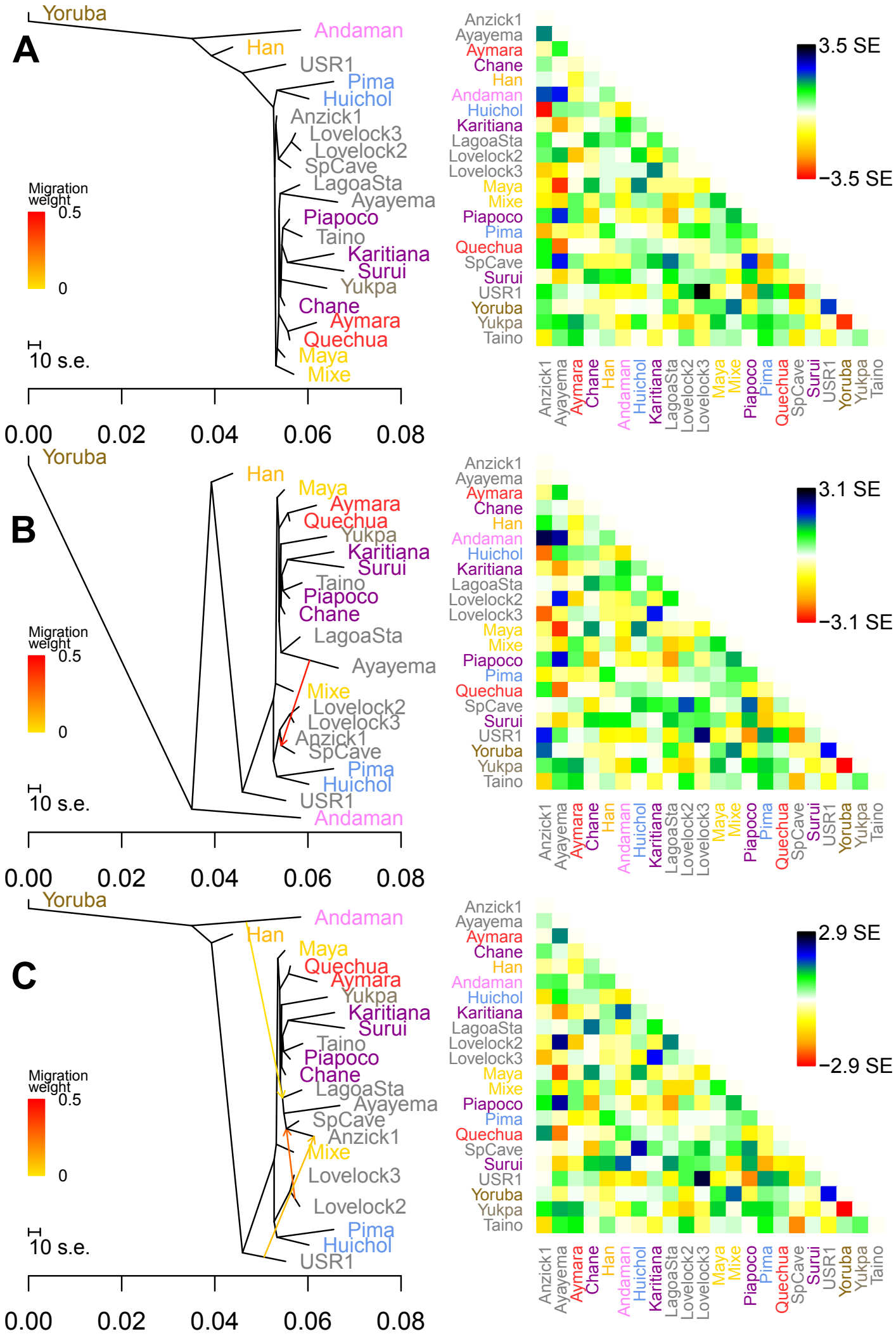

A
Yoruba
Andaman
Han
USR1
Pima
Huichol
Anzick1
Lovelock3
Lovelock2
SpCave
LagoaSta
Ayayema
Piapoco
Taino
Karitiana
Surui
Yukpa
Chane
Aymara
Quechua
Maya
Mixe
Migration weight
0.5
0
10 s.e.
0.00
0.02
0.04
0.06
0.08
3.5 SE
−3.5 SE
B
3.1 SE
−3.1 SE
C
2.9 SE
−2.9 SE

Figure S37. TreeMix admixture graphs relating contemporary and ancient SNA populations. We fitted admixture graphs with zero to five admixture edges to the SNA populations and individuals that we identified as 'non-admixed' based on *D*-statistics of the form *D(SNA, Aymara/Mixe; Yoruba/French/Han, Chimp)* (Fig. S19) and previous TreeMix results (Fig. S36). We show results for 0, 1 and 3 admixture edges. On the left, we show the admixture graphs produced by TreeMix for increasing number of admixture edges. On the right, we show the pairwise residuals after fitting the maximum likelihood graph.

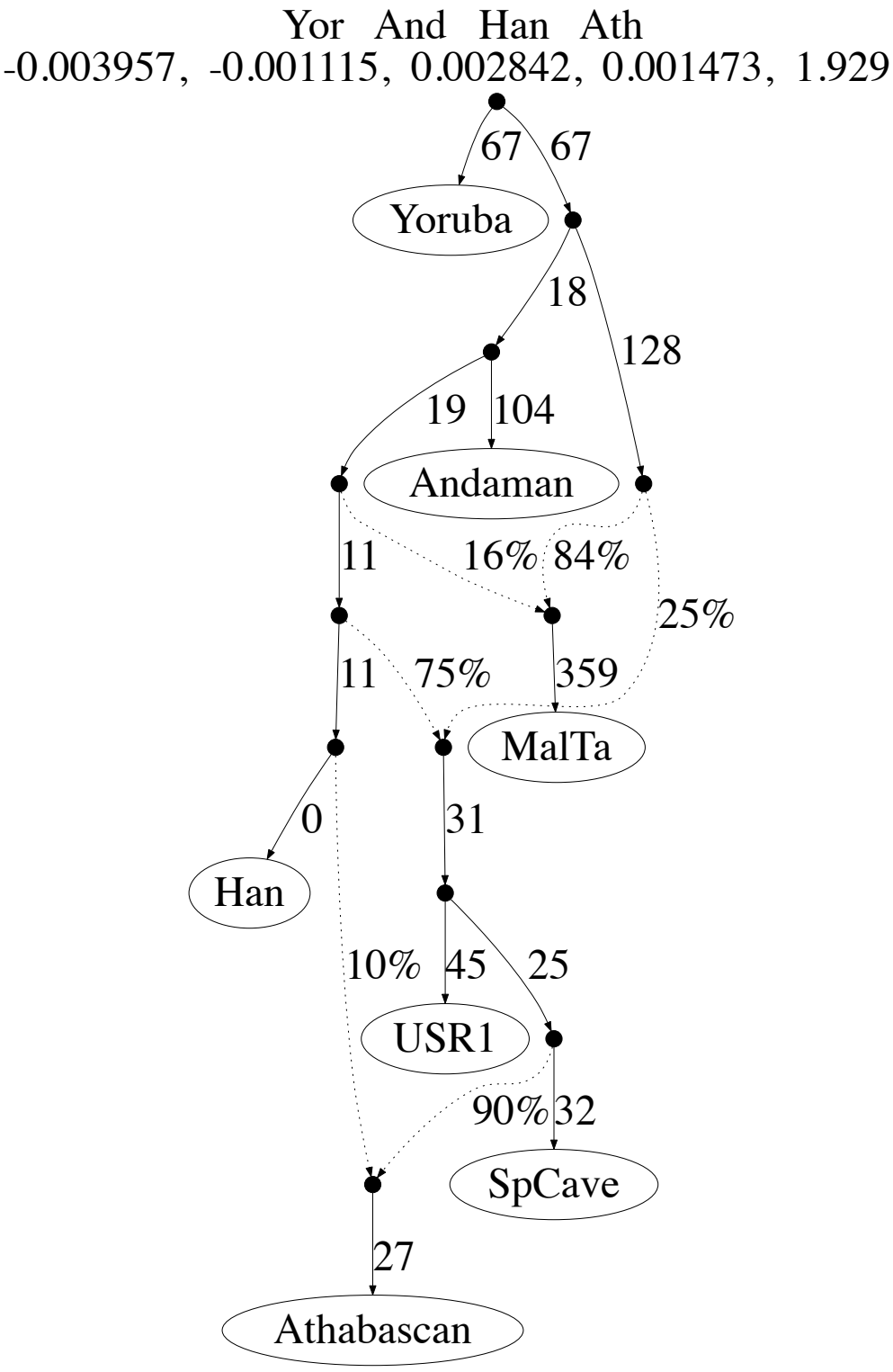


Figure S38. 'Seed graph' used for modeling the ancestry of North American ancient genomes. We fitted an *f*-statistic-based admixture graph modelling the formation of Native Americans, as well as the two basal Native American splits. At the top, we show the four populations leading to the worst *D*-statistic residual after optimizing the model parameters (Yoruba, Andaman; Han, Athabascan), followed by the observed value for such statistic, the expected value under the fitted model, the residual, the standard error of the residuals and a *Z*-score for such residual. Numbers to the right of solid edges represent optimized drift parameters and percentages to the right of dashed edges represent admixture proportions. In order to model the Trail Creek, Big Bar, 939, Kennewick and ASO genomes, we enumerated all possible extensions of this graph where we included each of these genomes both as a 'non-admixed' and an 'admixed' group.

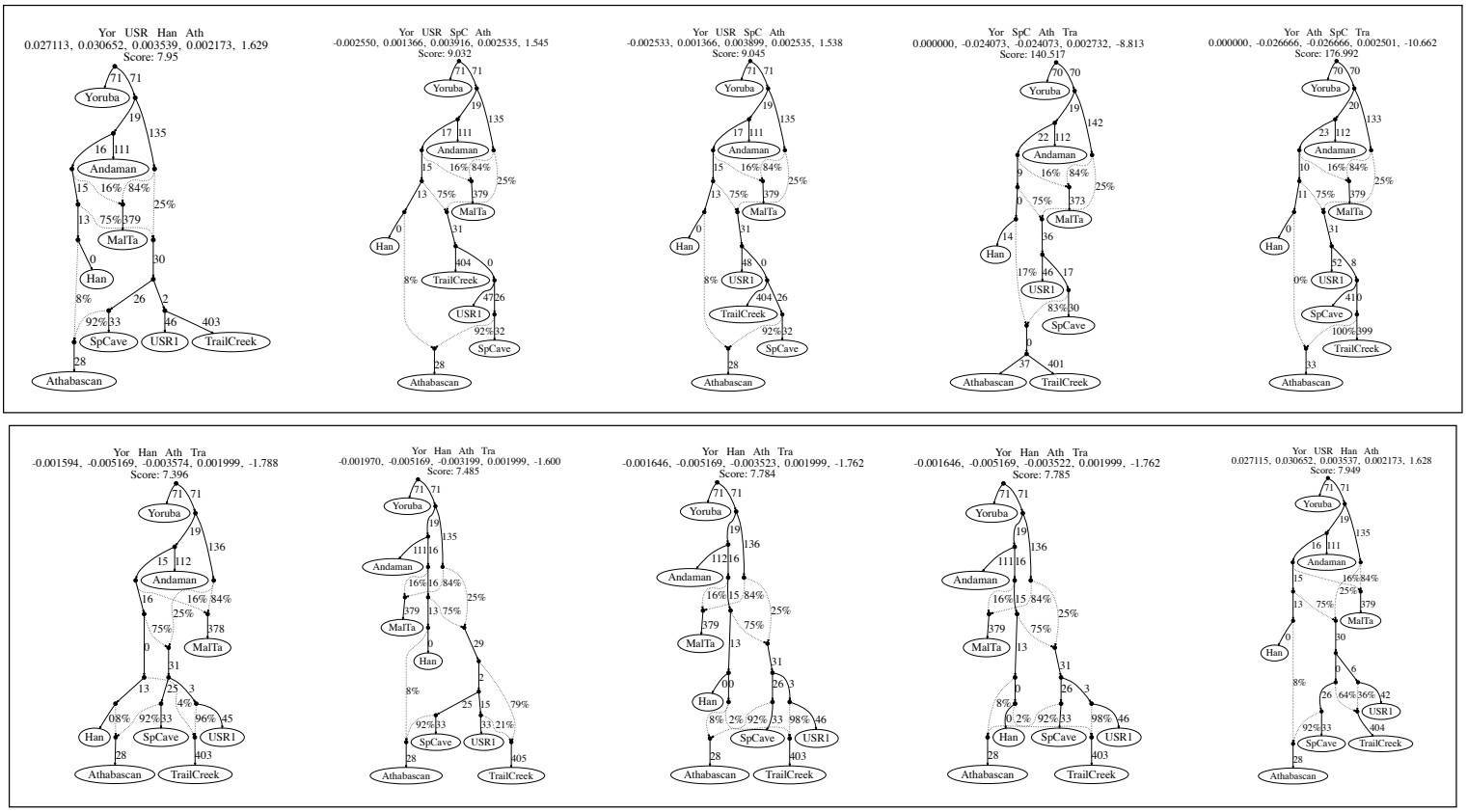


Figure S39. *f*-statistic-based admixture graphs exploring the placement of Trail Creek with respect to other Native Americans. We enumerated all possible extensions of the seed graph shown in Fig. S38 where we added Trail Creek as either a 'non-admixed' or an admixed population, and optimized the parameters for each topology using qpGraph. We show the five best-fitting 'non-admixed' (top) and 'admixed' (bottom) topologies sorted by their fit score. Above each graph, we show three text rows: 1. the four populations leading to the worst D-statistic residual after optimizing the model parameters; 2. the observed value for such statistic, the expected value under the fitted model, the residual, the standard error of the residuals and a Z-score for such residual; and 3. the model fit score. Numbers to the right of solid edges represent optimized drift parameters and percentages to the right of dashed edges represent admixture proportions.

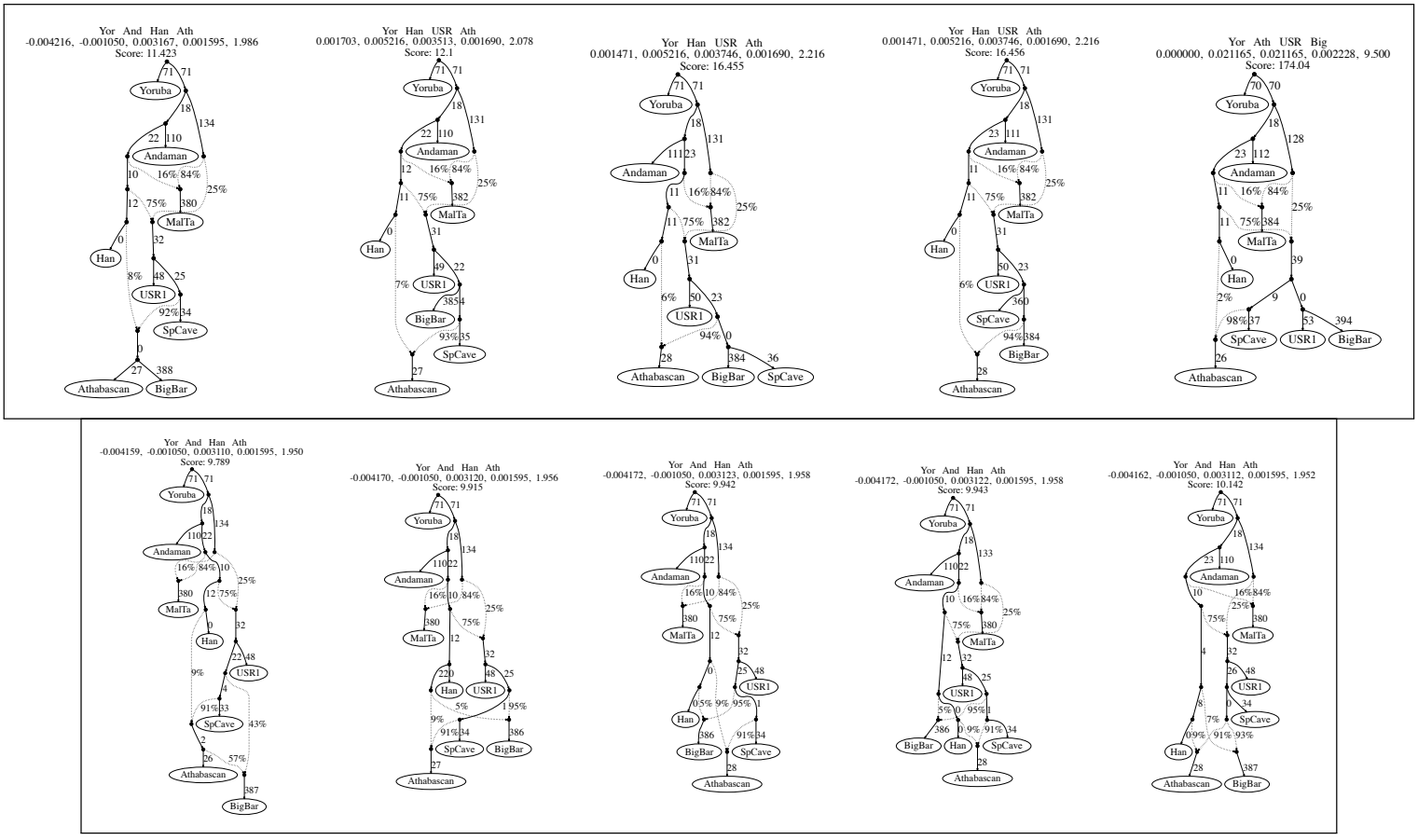


Figure S40. *f*-statistic-based admixture graphs exploring the placement of Big Bar with respect to other Native Americans. We enumerated all possible extensions of the seed graph shown in Fig. S38 where we added Big Bar as either a 'non-admixed' or an admixed population, and optimized the parameters for each topology using qpGraph. We show the five best-fitting 'non-admixed' (top) and 'admixed' (bottom) topologies sorted by their fit score. Above each graph, we show three text rows: 1. the four populations leading to the worst D-statistic residual after optimizing the model parameters; 2. the observed value for such statistic, the expected value under the fitted model, the residual, the standard error of the residuals and a Z-score for such residual; and 3. the model fit score. Numbers to the right of solid edges represent optimized drift parameters and percentages to the right of dashed edges represent admixture proportions.

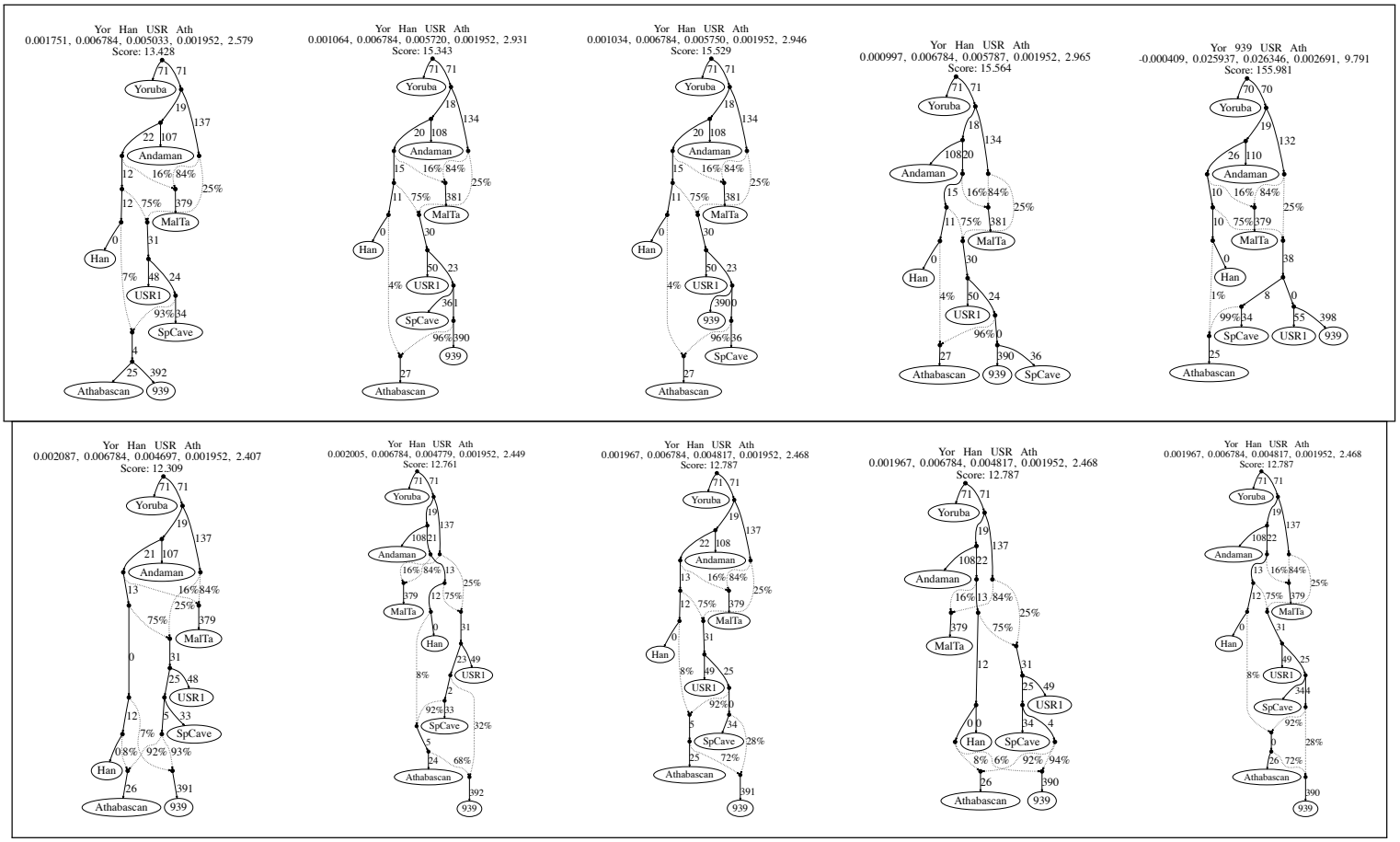


Figure S41. *f*-statistic-based admixture graphs exploring the placement of the ancient British Columbian '939' genome with respect to other Native Americans. We enumerated all possible extensions of the seed graph shown in Fig. S38 where we added 939 as either a 'non-admixed' or an admixed population, and optimized the parameters for each topology using qpGraph. We show the five best-fitting 'non-admixed' (top) and 'admixed' (bottom) topologies sorted by their fit score. Above each graph, we show three text rows: 1. the four populations leading to the worst D-statistic residual after optimizing the model parameters; 2. the observed value for such statistic, the expected value under the fitted model, the residual, the standard error of the residuals and a Z-score for such residual; and 3. the model fit score. Numbers to the right of solid edges represent optimized drift parameters and percentages to the right of dashed edges represent admixture proportions.

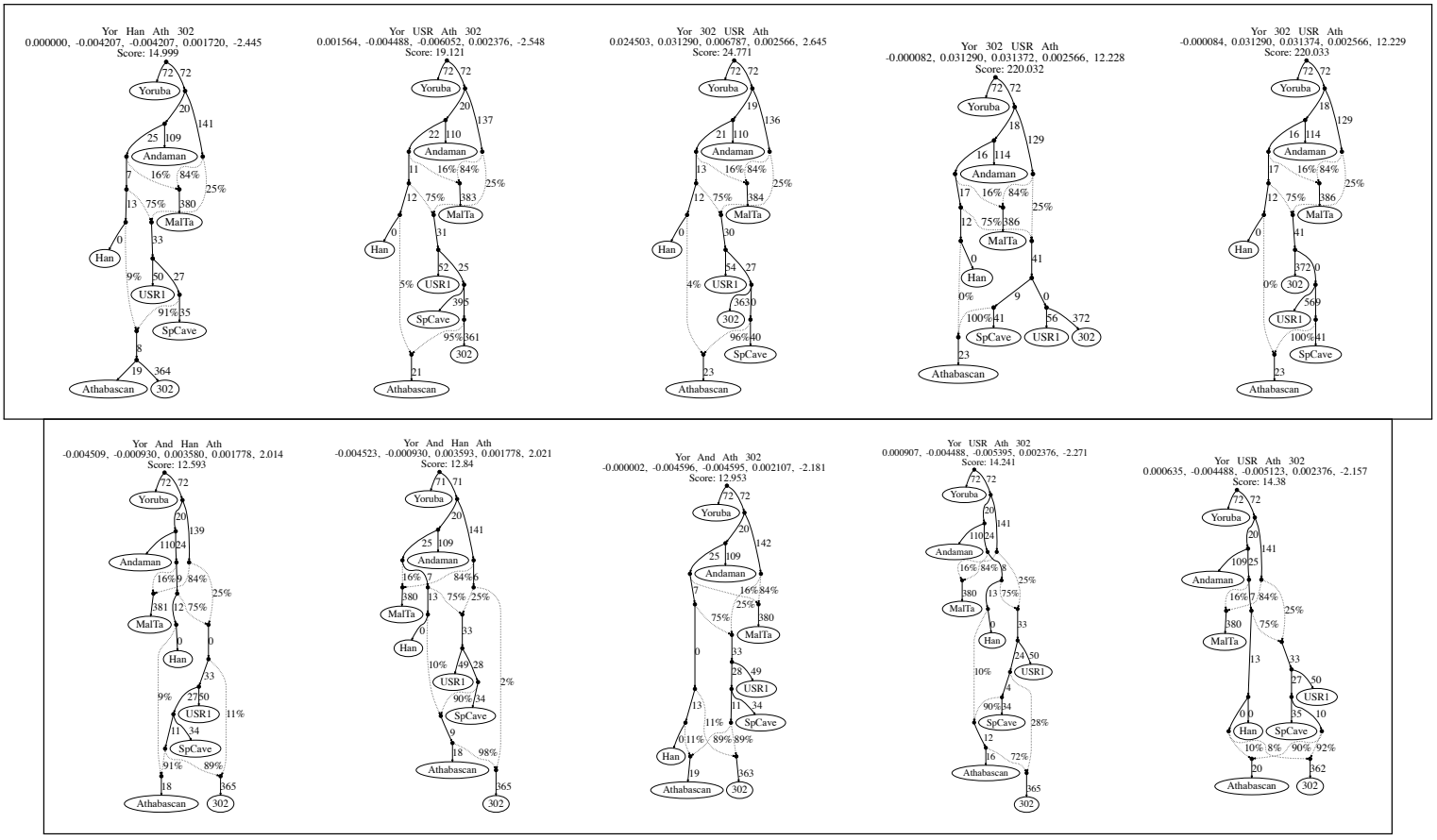


Figure S42. *f*-statistic-based admixture graphs exploring the placement of the ancient British Columbian '302' genome with respect to other Native Americans. We enumerated all possible extensions of the seed graph shown in Fig. S38 where we added 302 as either a 'non-admixed' or an admixed population, and optimized the parameters for each topology using qpGraph. We show the five best-fitting 'non-admixed' (top) and 'admixed' (bottom) topologies sorted by their fit score. Above each graph, we show three text rows: 1. the four populations leading to the worst D-statistic residual after optimizing the model parameters; 2. the observed value for such statistic, the expected value under the fitted model, the residual, the standard error of the residuals and a Z-score for such residual; and 3. the model fit score. Numbers to the right of solid edges represent optimized drift parameters and percentages to the right of dashed edges represent admixture proportions.

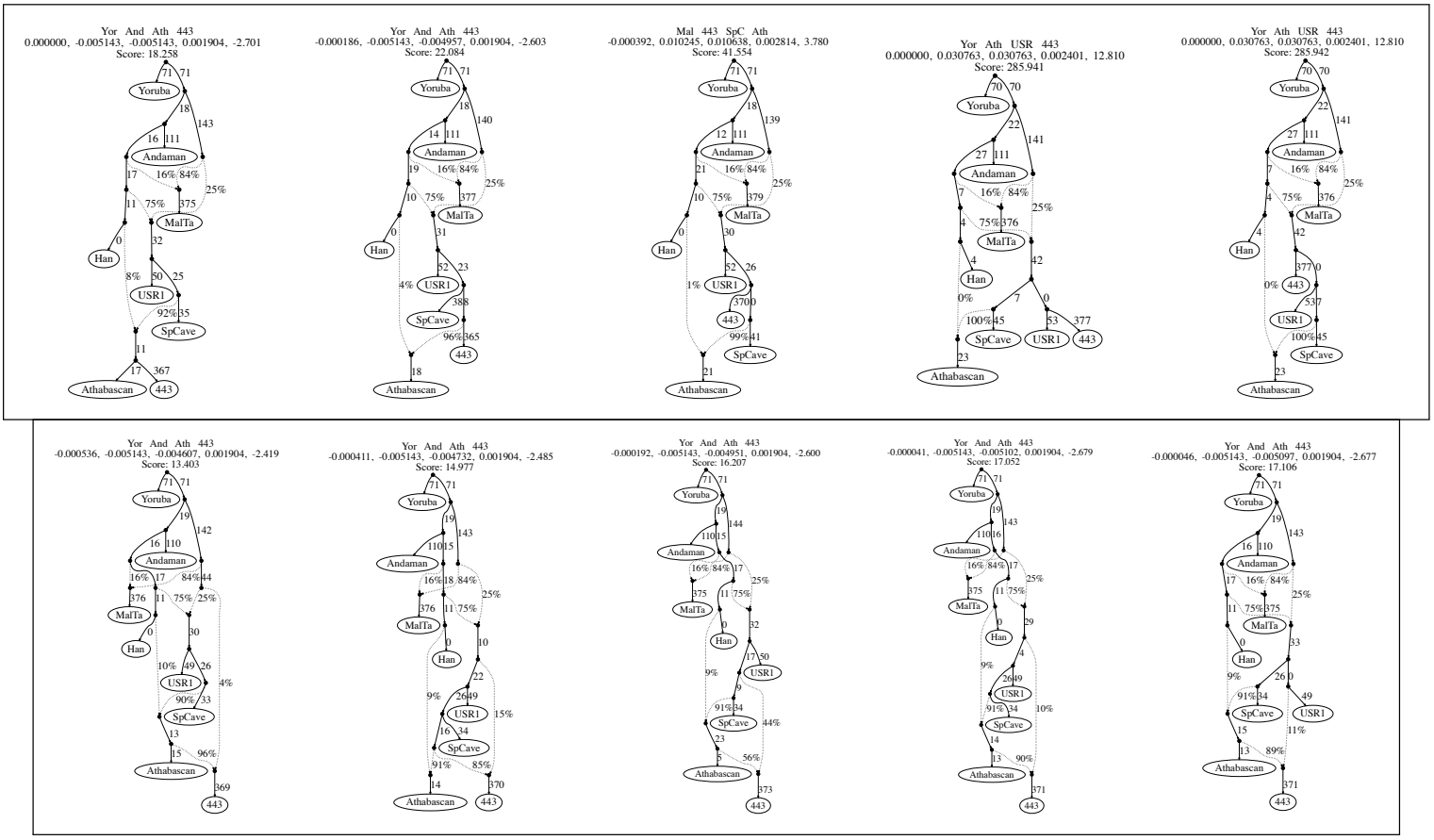


Figure S43. *f*-statistic-based admixture graphs exploring the placement of the ancient British Columbian '443' genome with respect to other Native Americans. We enumerated all possible extensions of the seed graph shown in Fig. S38 where we added 443 as either a 'non-admixed' or an admixed population, and optimized the parameters for each topology using qpGraph. We show the five best-fitting 'non-admixed' (top) and 'admixed' (bottom) topologies sorted by their fit score. Above each graph, we show three text rows: 1. the four populations leading to the worst D-statistic residual after optimizing the model parameters; 2. the observed value for such statistic, the expected value under the fitted model, the residual, the standard error of the residuals and a Z-score for such residual; and 3. the model fit score. Numbers to the right of solid edges represent optimized drift parameters and percentages to the right of dashed edges represent admixture proportions.

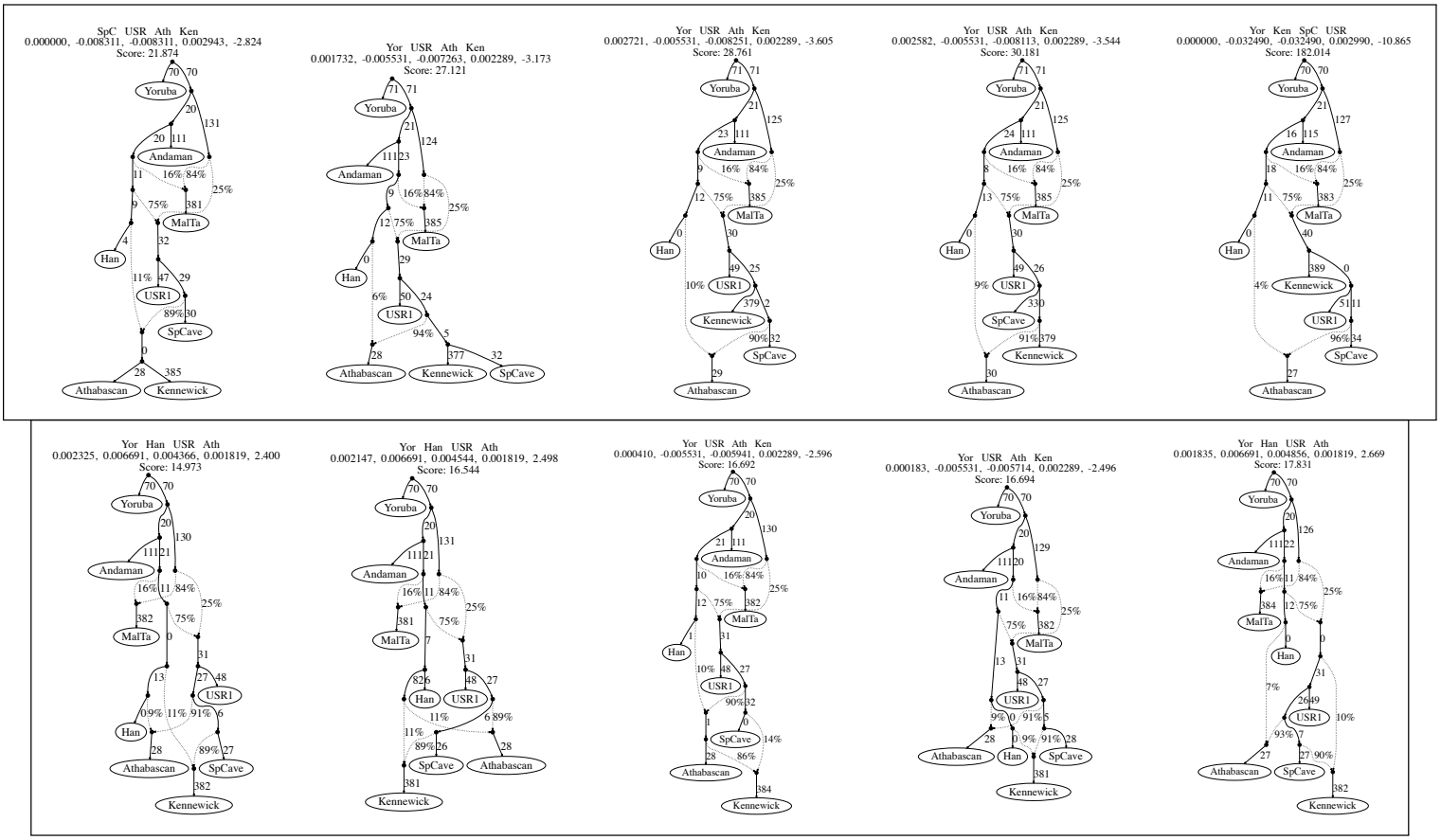


Figure S44. *f*-statistic-based admixture graphs exploring the placement of the Kennewick Man/Ancient One with respect to other Native Americans. We enumerated all possible extensions of the seed graph shown in Fig. S38 where we added Kennewick as either a 'non-admixed' or an admixed population, and optimized the parameters for each topology using qpGraph. We show the five best-fitting 'non-admixed' (top) and 'admixed' (bottom) topologies sorted by their fit score. Above each graph, we show three text rows: 1. the four populations leading to the worst D-statistic residual after optimizing the model parameters; 2. the observed value for such statistic, the expected value under the fitted model, the residual, the standard error of the residuals and a Z-score for such residual; and 3. the model fit score. Numbers to the right of solid edges represent optimized drift parameters and percentages to the right of dashed edges represent admixture proportions.

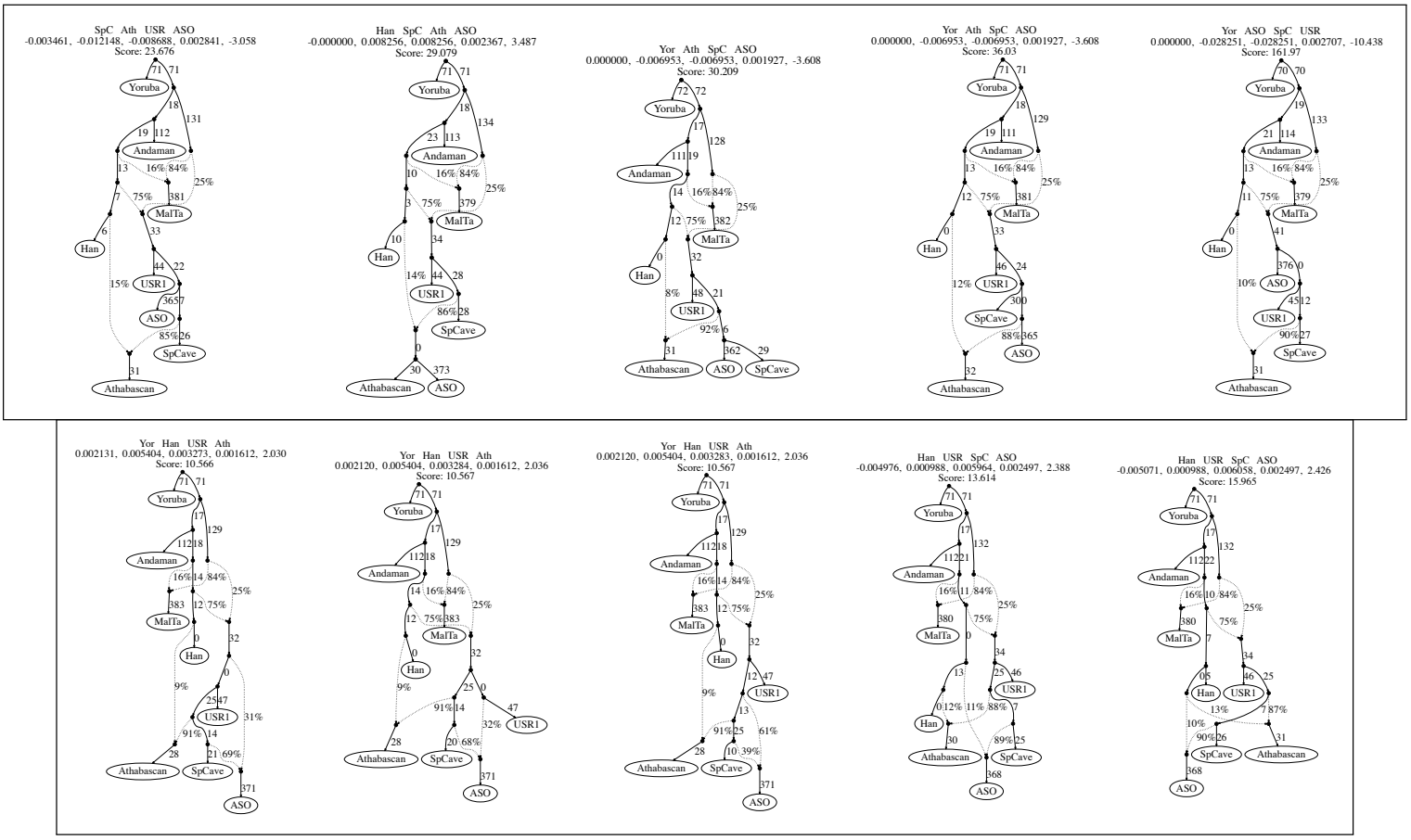


Figure S45. *f*-statistic-based admixture graphs exploring the placement of the Ancient Southwestern Ontario (ASO) genome with respect to other Native Americans. We enumerated all possible extensions of the seed graph shown in Fig. S38 where we added ASO as either a 'non-admixed' or an admixed population, and optimized the parameters for each topology using qpGraph. We show the five best-fitting 'non-admixed' (top) and 'admixed' (bottom) topologies sorted by their fit score. Above each graph, we show three text rows: 1. the four populations leading to the worst D-statistic residual after optimizing the model parameters; 2. the observed value for such statistic, the expected value under the fitted model, the residual, the standard error of the residuals and a Z-score for such residual; and 3. the model fit score. Numbers to the right of solid edges represent optimized drift parameters and percentages to the right of dashed edges represent admixture proportions.

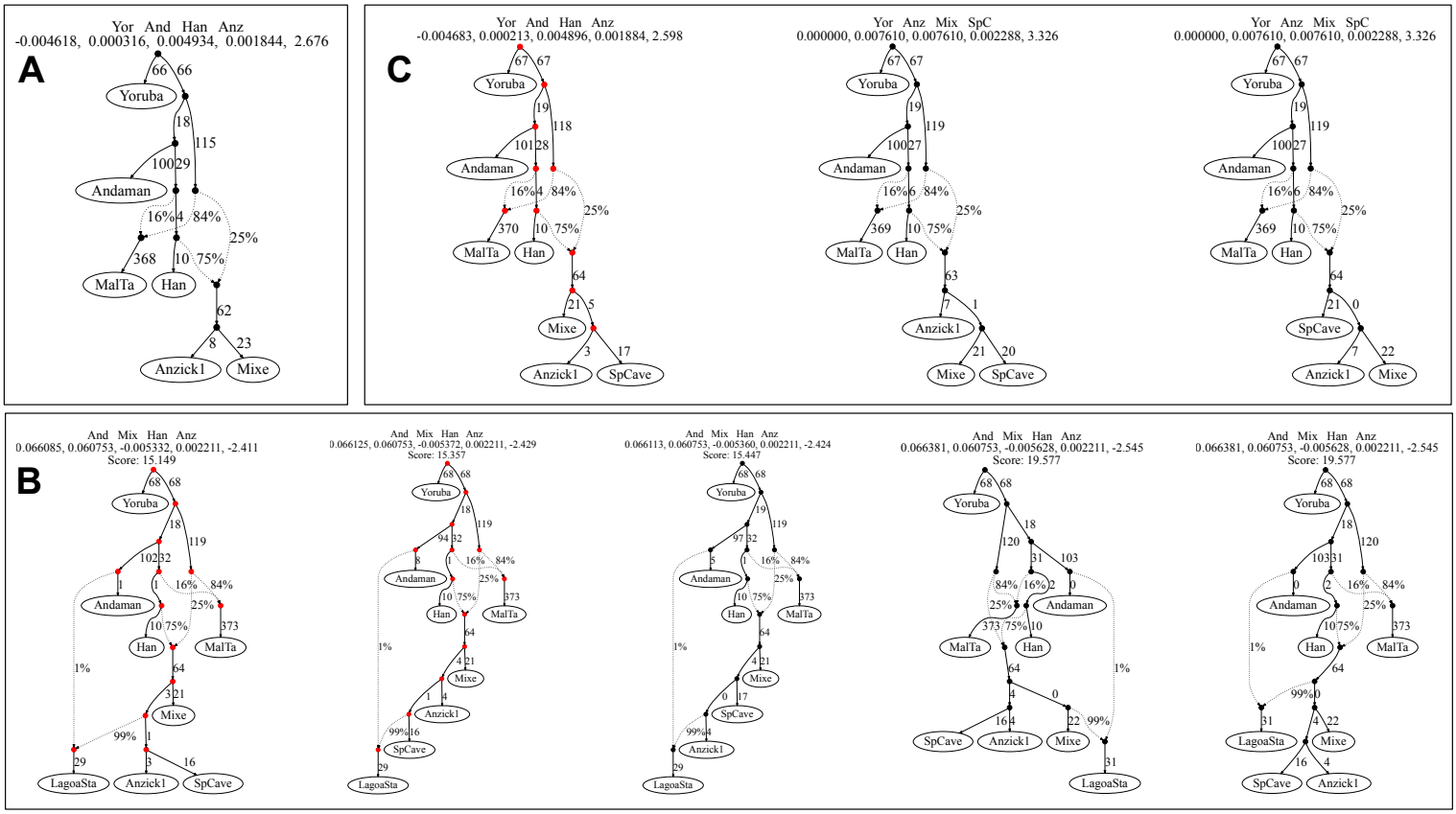


Figure S46. Admixture graphs relating Late Pleistocene and Early Holocene SNA genomes and the Mesamerican Mixe. A. We fitted an *f*-statistic-based admixture graph modelling the formation of Native Americans, represented by the Anzick1 ancient genome and the present-day Mesoamerican Mixe. The former represents an individual who is 'directly ancestral' to many SNA groups, while the latter represent one of the first internal branches within SNA. **B.** We explored the three possible placements of Spirit Cave inside the Native American subgraph, using the graph in a. as a starting point. We show the fitting results for each topology and sort them by their fit score. **C.** Based on the best-fitting model from the previous step (highlighted in red in B.), we enumerated all its possible extensions where we added Lagoa Santa as an admixed leaf. Note that we only considered models where Lagoa Santa derives part of its ancestry from a source most closely related to the ancient Andaman. In other words, we search for the best placement of Lagoa Santa in the Native American subgraph, taking into account its excess allele sharing with Australasians. We show the five best-fitting topologies sorted by their fit score. Both topologies highlighted in red yield no zero-length internal braches and similar likelihoods and D-statistic residuals. Thus, we consider that we cannot discern between these two models with the data at hand. Above each graph, we show three text rows: 1. the four populations leading to the worst D-statistic residual after optimizing the model parameters; 2. the observed value for such statistic, the expected value under the fitted model, the residual, the standard error of the residuals and a Z-score for such residual; and 3. the model fit score. Numbers to the right of solid edges represent optimized drift parameters and percentages to the right of dashed edges represent admixture proportions.

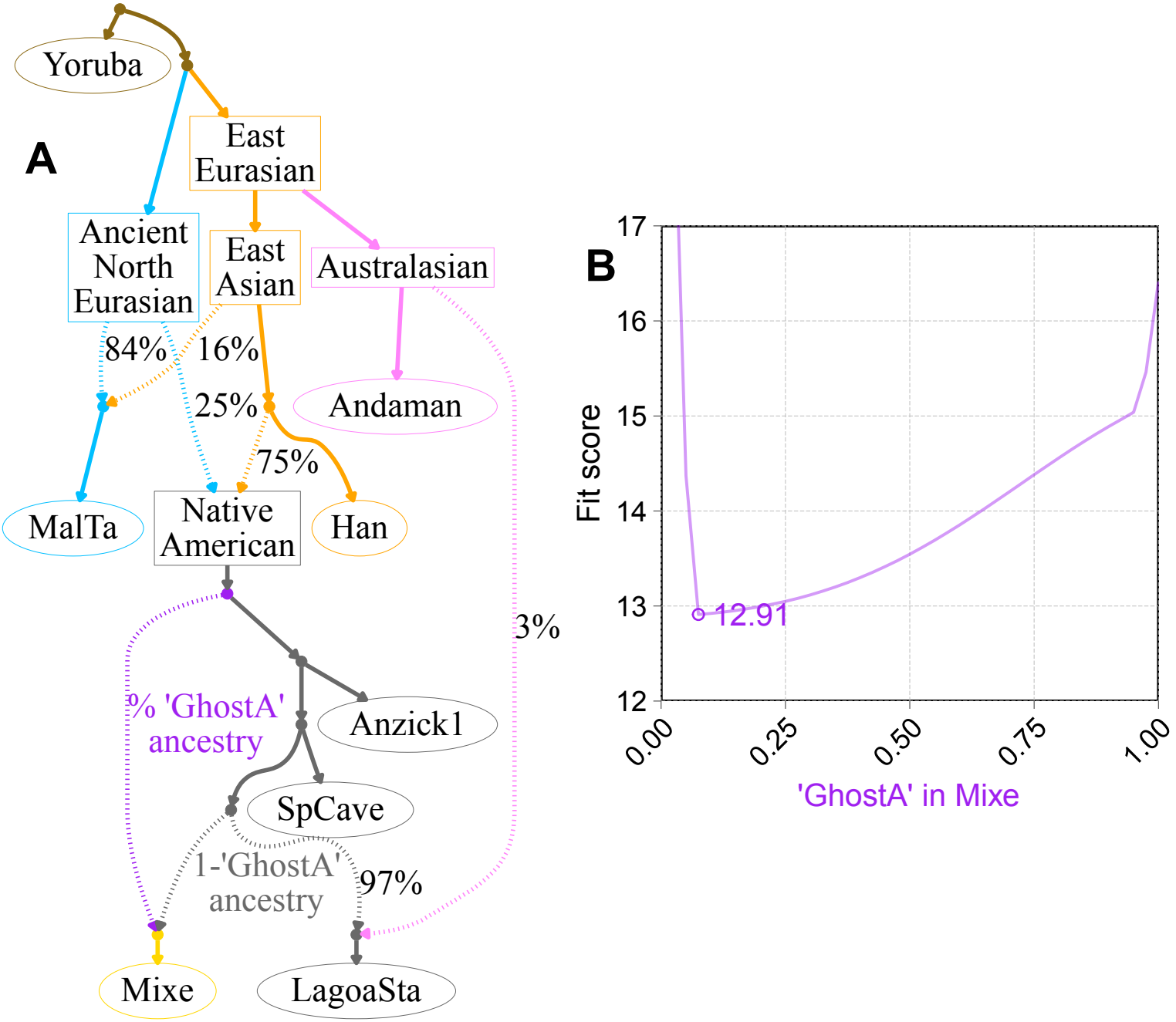


Figure S47. Likelihood scores for a model where the Mesoamerican Mixe bear 'unsampled admixture'. We explored whether a model where the Mesoamerican Mixe form a clade with Lagoa Santa but also bears ancestry from a 'unsampled' Native American population, is a better fit to the data compared to the models explored in Fig. S46 where Mixe is an outgroup to other SNA groups. We fitted the model in **A.**, by fixing the 'unsampled' ancestry proportion in Mixe to different values ranging from 0 to 1. **B.** A plot of the fit score (log-likelihood) as a function of the 'unsampled' admixture in Mixe. While we can reject the cases where the admixture proportion is 0 and 1, we do not observe a statistically significant increase in likelihood when considering intermediate admixture proportions.

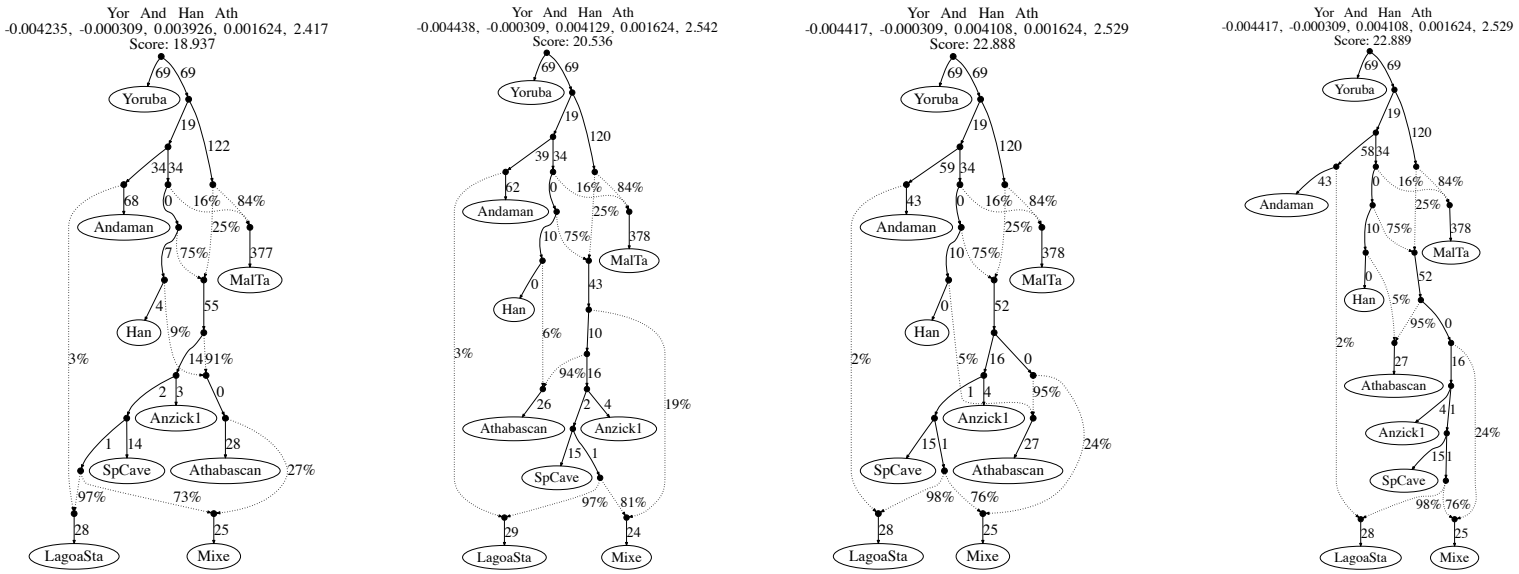


Figure S48. Admixture graphs exploring the position of the 'unsampled population' that contributed a fraction of the ancestry in the Mesoamerican Mixe. We extended the graph in Fig. S47 by including Athabascans as a representative of NNA and fitted graphs with all possible placements of the 'unsampled population' that contributed to Mixe, i.e., as a clade with Athabscans who bear Siberian-related admixture (left), as an outgroup to NNA and SNA (middle left), as a clade with NNA (middle right) and as a branch that diverges below NNA but above Anzick1 (right). While all models yield D-statistic residuals for which Z<3.3 and similar likelihood scores, the only model that does not include zero-length internal branches is that in which the 'unsampled population' is an outgroup to NNA and SNA (middle-left). Above each graph, we show three text rows: 1. the four populations leading to the worst D-statistic residual after optimizing the model parameters; 2. the observed value for such statistic, the expected value under the fitted model, the residual, the standard error of the residuals and a Z-score for such residual; and 3. the model fit score. Numbers to the right of solid edges represent optimized drift parameters and percentages to the right of dashed edges represent admixture proportions.

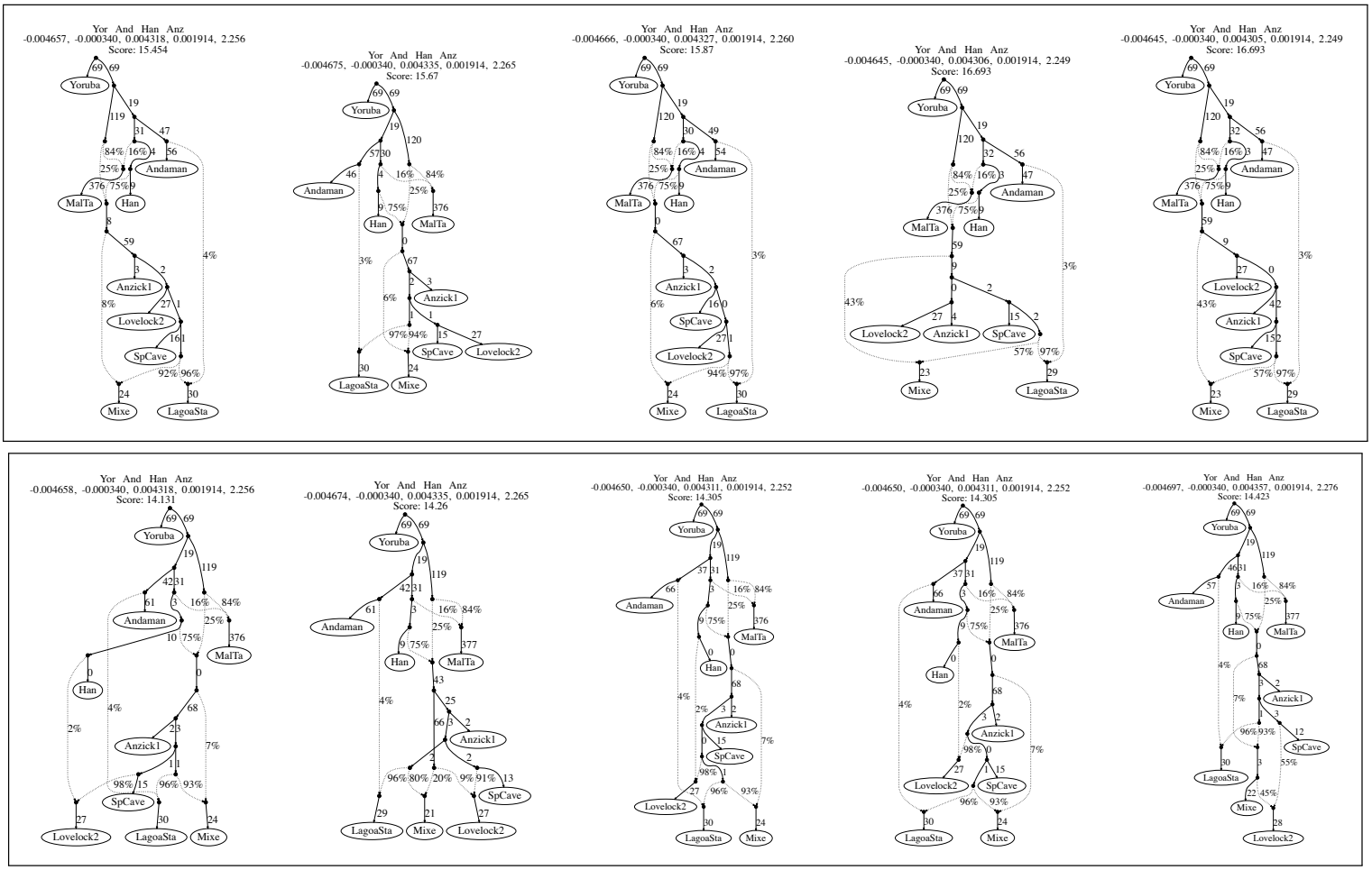


Figure S49. *f*-statistic-based admixture graphs exploring the placement of the Lovelock2 ancient genome from the Great Basin with respect to other Native Americans. We enumerated all possible extensions of the seed graph shown in Fig. 3A where we added Lovelock2 as either a 'non-admixed' or an admixed population, and optimized the parameters for each topology using qpGraph. The top panel shows the five best-fitting 'non-admixed' topologies and the bottom panel shows the 'admixed' topologies, sorted by their fit score. Above each graph, we show three text rows: 1. the four populations leading to the worst D-statistic residual after optimizing the model parameters; 2. the observed value for such statistic, the expected value under the fitted model, the residual, the standard error of the residuals and a Z-score for such residual; and 3. the model fit score. Numbers to the right of solid edges represent optimized drift parameters and percentages to the right of dashed edges represent admixture proportions.

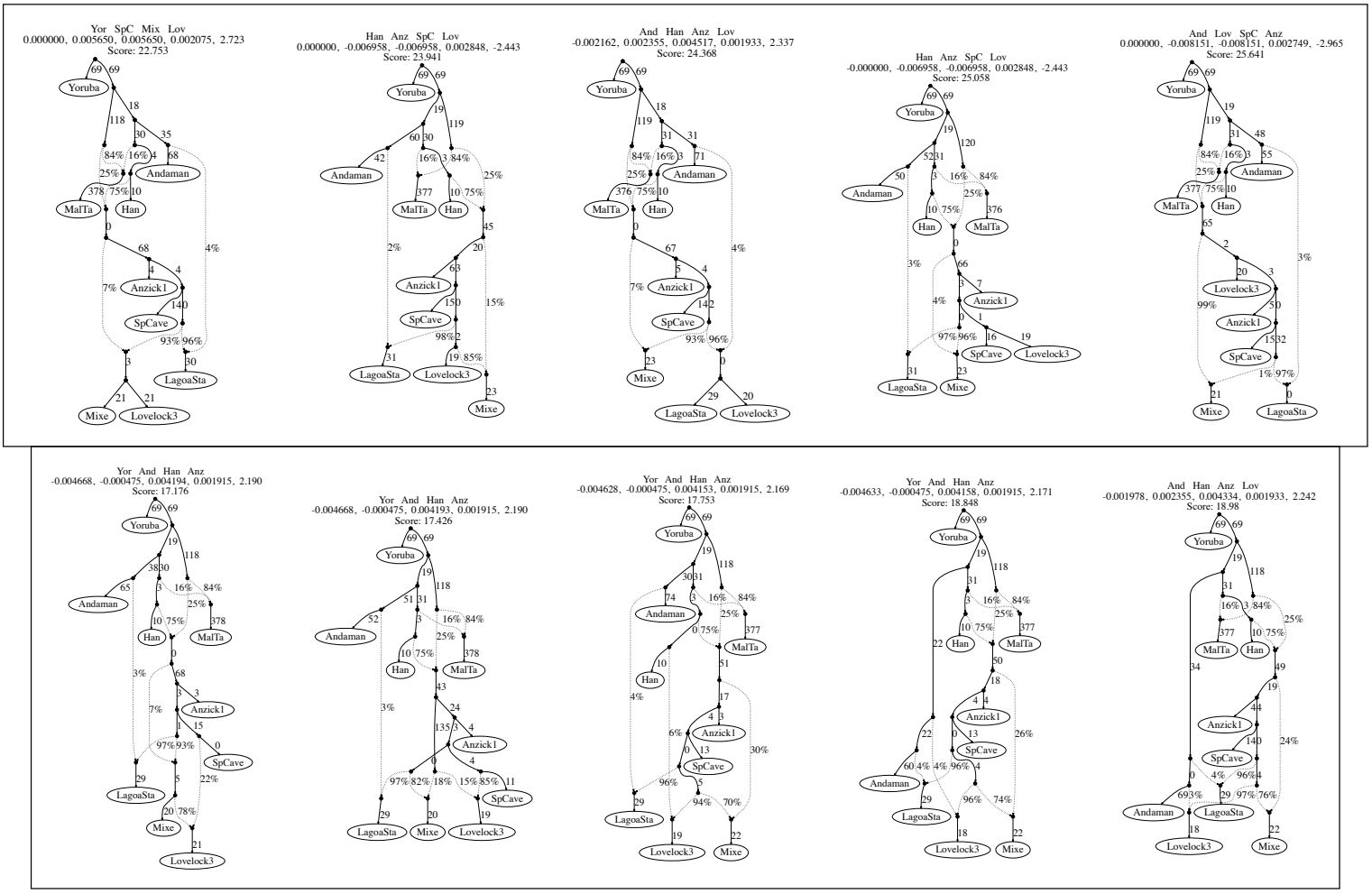


Figure S50. *f*-statistic-based admixture graphs exploring the placement of the Lovelock3 ancient genome from the Great Basin with respect to other Native Americans. We enumerated all possible extensions of the seed graph shown in Fig. 3A where we added Lovelock3 as either a 'non-admixed' or an admixed population, and optimized the parameters for each topology using qpGraph. The top panel shows the five best-fitting 'non-admixed' topologies and the bottom panel shows the 'admixed' topologies, sorted by their fit score. Above each graph, we show three text rows: 1. the four populations leading to the worst D-statistic residual after optimizing the model parameters; 2. the observed value for such statistic, the expected value under the fitted model, the residual, the standard error of the residuals and a Z-score for such residual; and 3. the model fit score. Numbers to the right of solid edges represent optimized drift parameters and percentages to the right of dashed edges represent admixture proportions.

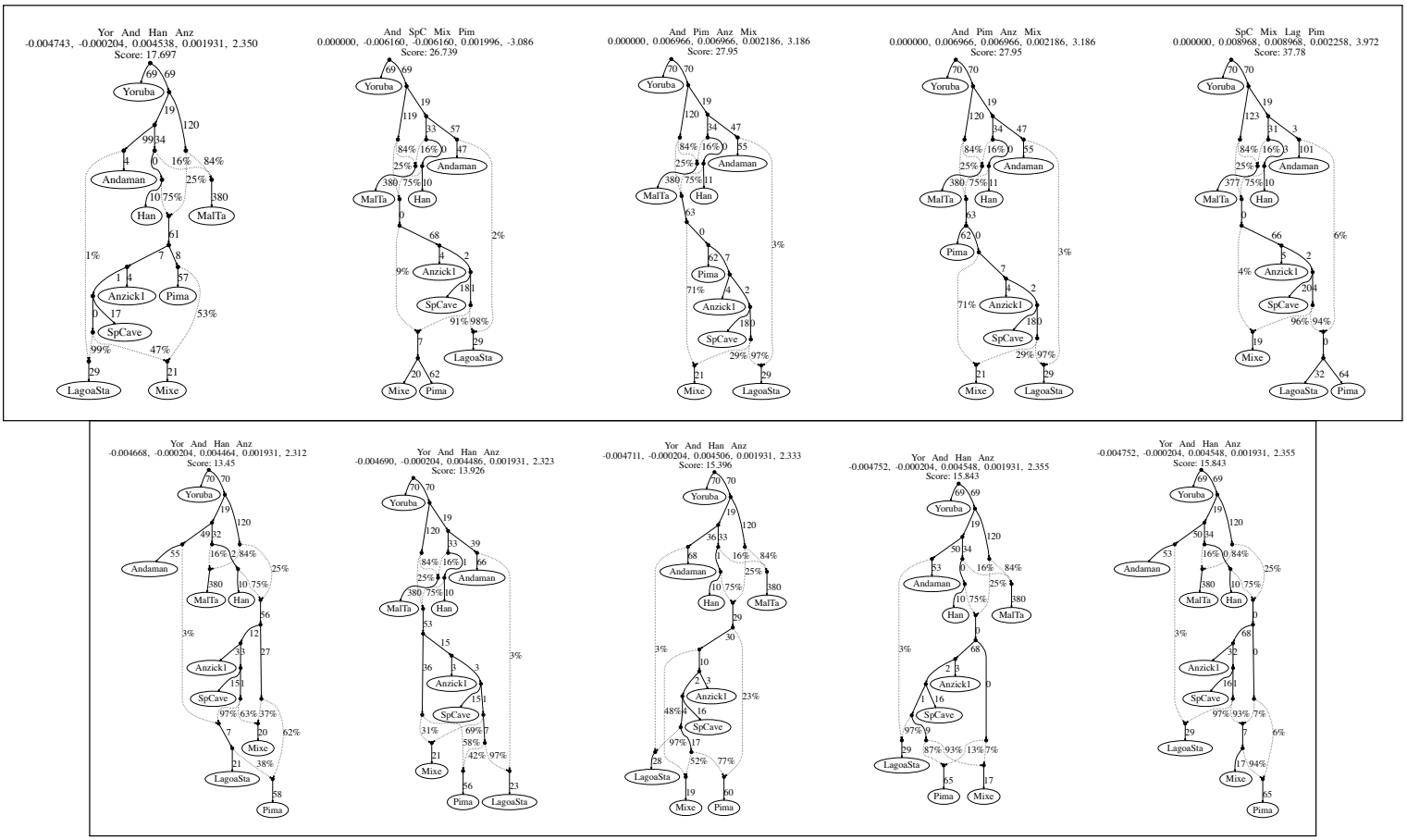


Figure S51. *f*-statistic-based admixture graphs exploring the placement of the present-day Pima from Mexico with respect to other Native Americans. We enumerated all possible extensions of the seed graph shown in Fig. 3A where we added Pima as either a 'non-admixed' or an admixed population, and optimized the parameters for each topology using qpGraph. The top panel shows the five best-fitting 'non-admixed' topologies and the bottom panel shows the 'admixed' topologies, sorted by their fit score. Above each graph, we show three text rows: 1. the four populations leading to the worst D-statistic residual after optimizing the model parameters; 2. the observed value for such statistic, the expected value under the fitted model, the residual, the standard error of the residuals and a Z-score for such residual; and 3. the model fit score. Numbers to the right of solid edges represent optimized drift parameters and percentages to the right of dashed edges represent admixture proportions.

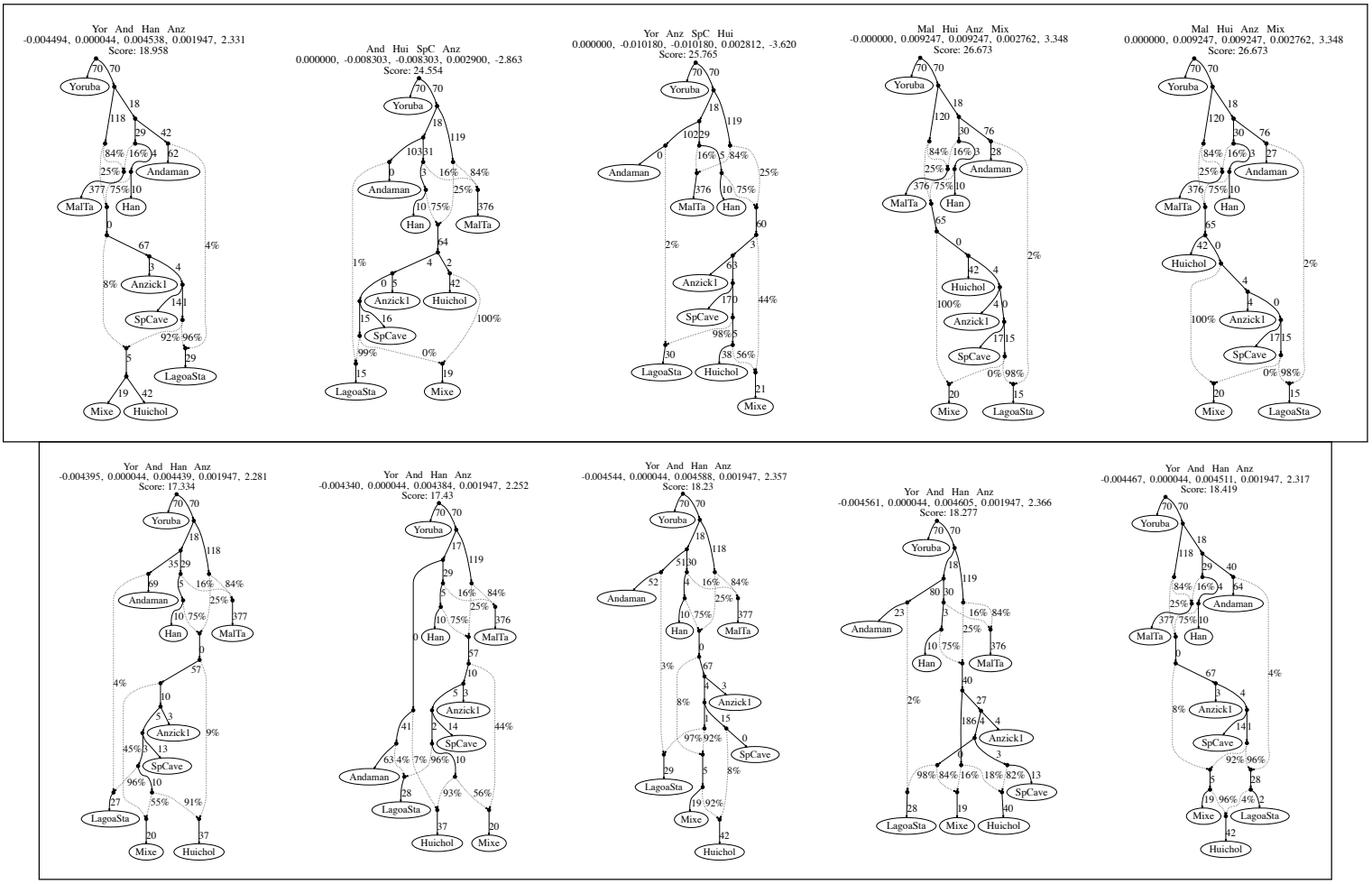


Figure S52. *f*-statistic-based admixture graphs exploring the placement of the present-day Huichol from Mexico with respect to other Native Americans. We enumerated all possible extensions of the seed graph shown in Fig. 3A where we added Huichol as either a 'non-admixed' or an admixed population, and optimized the parameters for each topology using qpGraph. The top panel shows the five best-fitting 'non-admixed' topologies and the bottom panel shows the 'admixed' topologies, sorted by their fit score. Above each graph, we show three text rows: 1. the four populations leading to the worst D-statistic residual after optimizing the model parameters; 2. the observed value for such statistic, the expected value under the fitted model, the residual, the standard error of the residuals and a Z-score for such residual; and 3. the model fit score. Numbers to the right of solid edges represent optimized drift parameters and percentages to the right of dashed edges represent admixture proportions.

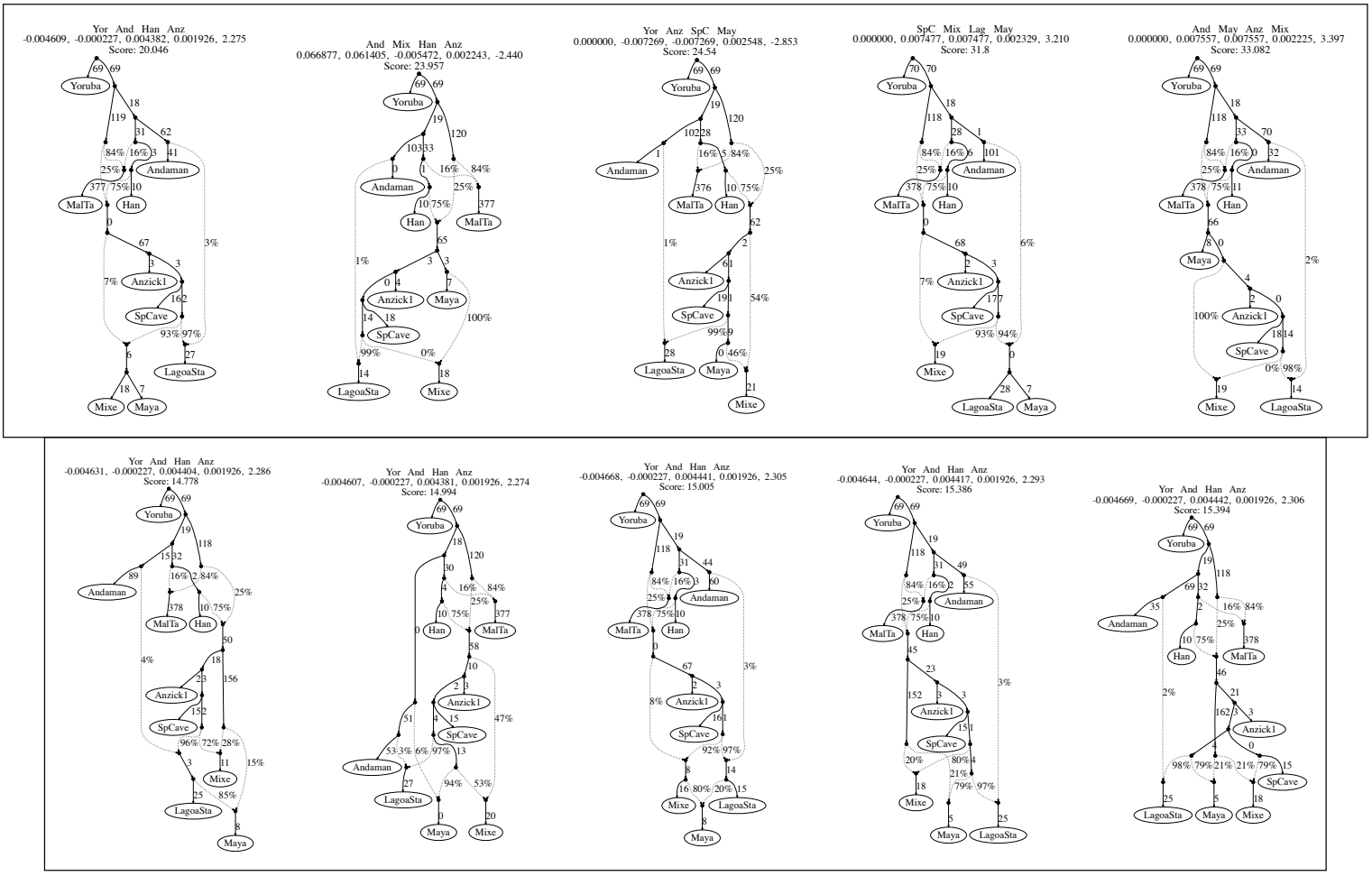


Figure S53. *f*-statistic-based admixture graphs exploring the placement of the present-day Maya from Mexico with respect to other Native Americans. We enumerated all possible extensions of the seed graph shown in Fig. 3A where we added Maya as either a 'non-admixed' or an admixed population, and optimized the parameters for each topology using qpGraph. The top panel shows the five best-fitting 'non-admixed' topologies and the bottom panel shows the 'admixed' topologies, sorted by their fit score. Above each graph, we show three text rows: 1. the four populations leading to the worst D-statistic residual after optimizing the model parameters; 2. the observed value for such statistic, the expected value under the fitted model, the residual, the standard error of the residuals and a Z-score for such residual; and 3. the model fit score. Numbers to the right of solid edges represent optimized drift parameters and percentages to the right of dashed edges represent admixture proportions.

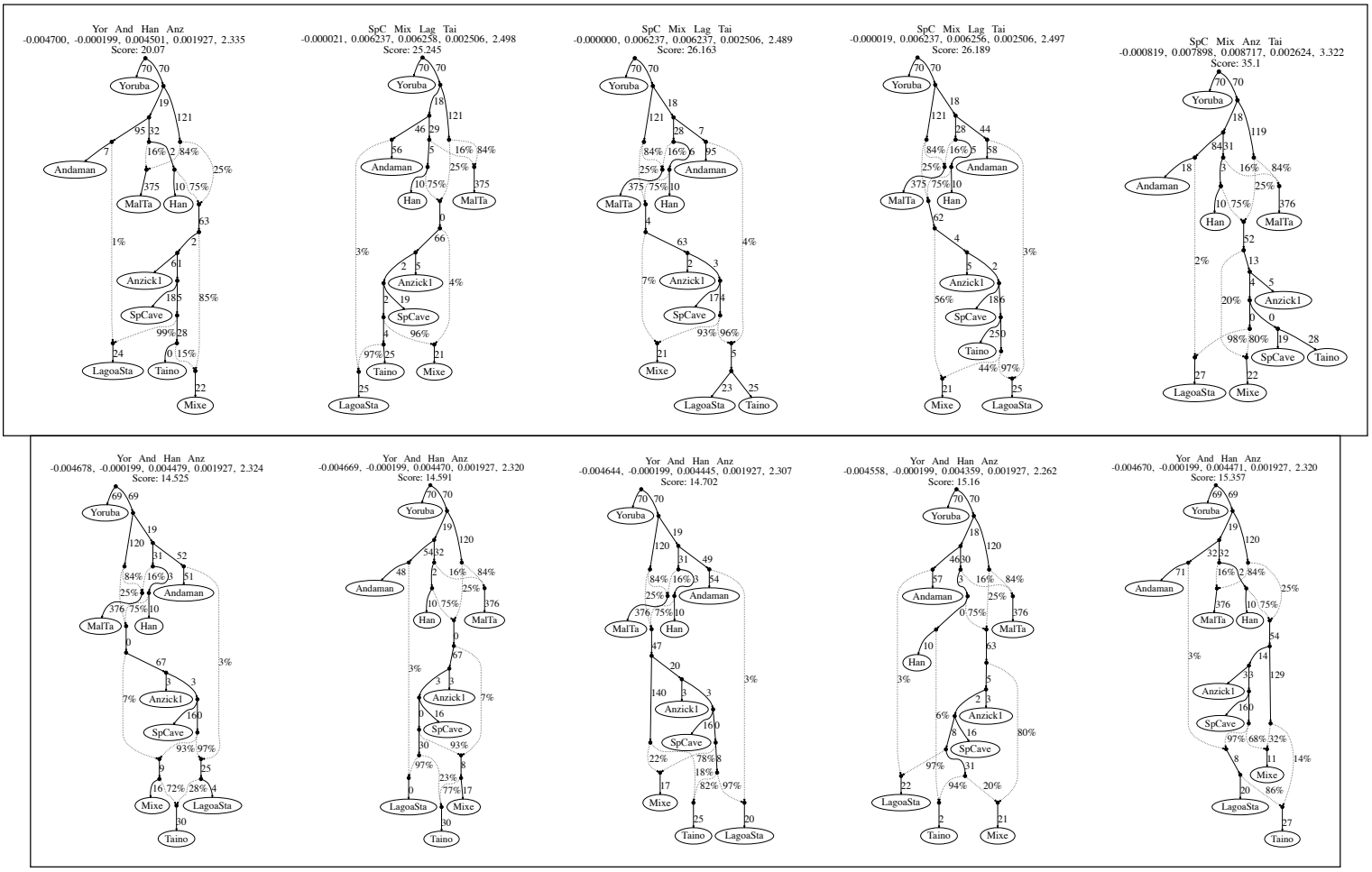


Figure S54. *f*-statistic-based admixture graphs exploring the placement of the Taino ancient genome from the Caribbean with respect to other Native Americans. We enumerated all possible extensions of the seed graph shown in Fig. 3A where we added Taino as either a 'non-admixed' or an admixed population, and optimized the parameters for each topology using qpGraph. The top panel shows the five best-fitting 'non-admixed' topologies and the bottom panel shows the 'admixed' topologies, sorted by their fit score. Above each graph, we show three text rows: 1. the four populations leading to the worst D-statistic residual after optimizing the model parameters; 2. the observed value for such statistic, the expected value under the fitted model, the residual, the standard error of the residuals and a Z-score for such residual; and 3. the model fit score. Numbers to the right of solid edges represent optimized drift parameters and percentages to the right of dashed edges represent admixture proportions.

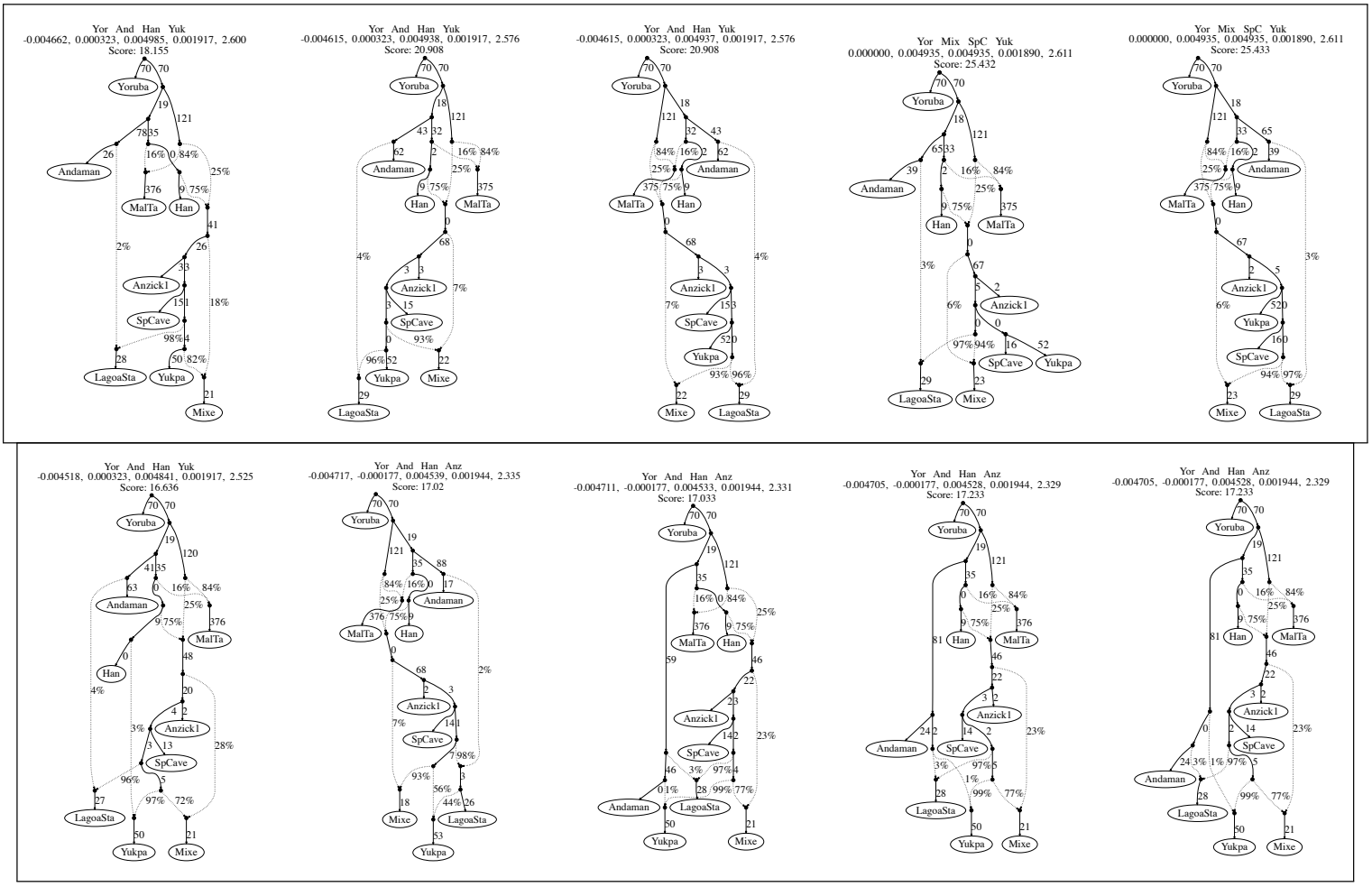


Figure S55. *f*-statistic-based admixture graphs exploring the placement of the present-day Yukpa from Venezuela with respect to other Native Americans. We enumerated all possible extensions of the seed graph shown in Fig. 3A where we added Yukpa as either a 'non-admixed' or an admixed population, and optimized the parameters for each topology using qpGraph. The top panel shows the five best-fitting 'non-admixed' topologies and the bottom panel shows the 'admixed' topologies, sorted by their fit score. Above each graph, we show three text rows: 1. the four populations leading to the worst D-statistic residual after optimizing the model parameters; 2. the observed value for such statistic, the expected value under the fitted model, the residual, the standard error of the residuals and a Z-score for such residual; and 3. the model fit score. Numbers to the right of solid edges represent optimized drift parameters and percentages to the right of dashed edges represent admixture proportions.

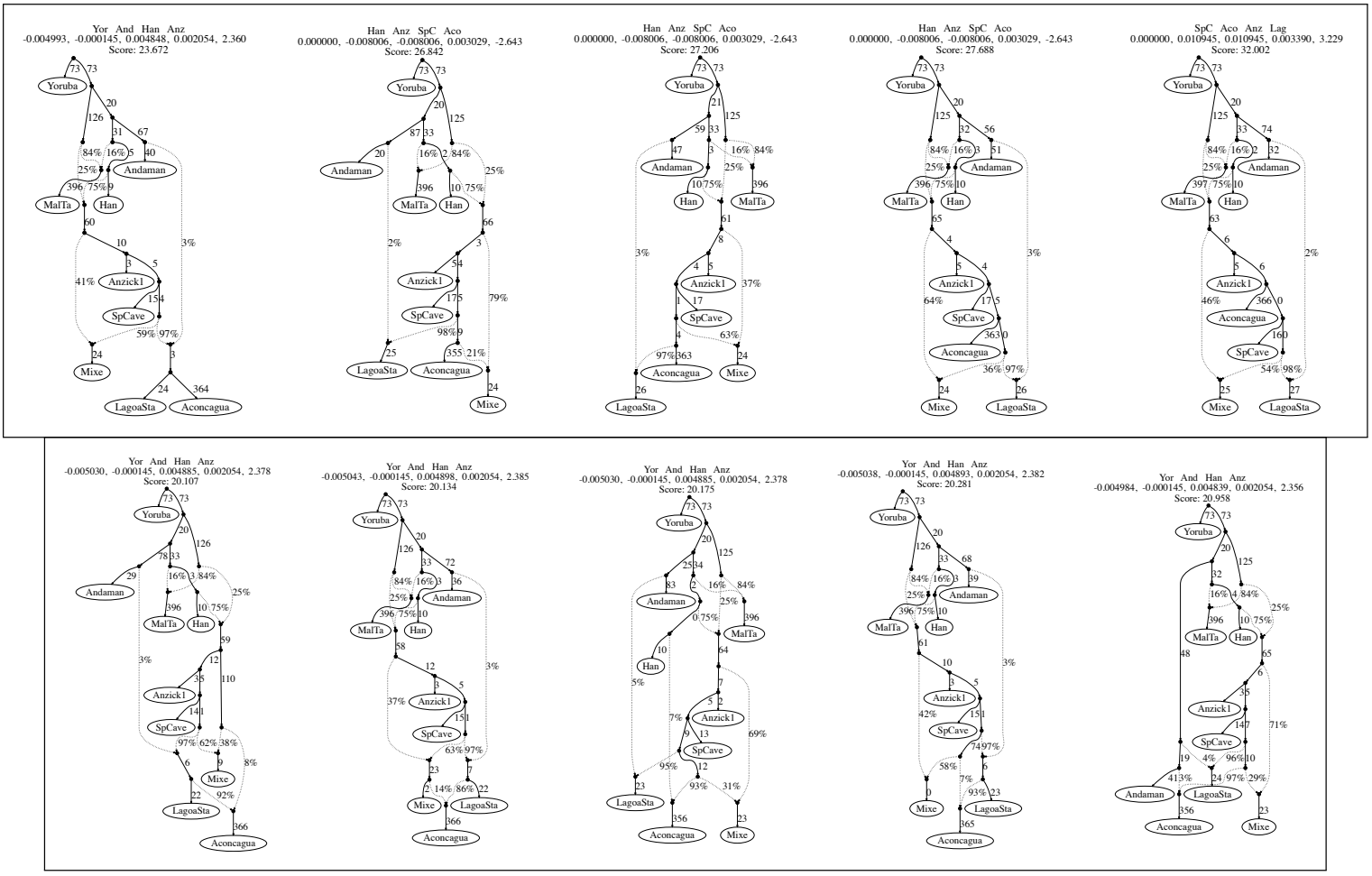


Figure S56. *f*-statistic-based admixture graphs exploring the placement of the Aconcagua ancient genome from the Andes with respect to other Native Americans. We enumerated all possible extensions of the seed graph shown in Fig. 3A where we added Aconcagua as either a 'non-admixed' or an admixed population, and optimized the parameters for each topology using qpGraph. The top panel shows the five best-fitting 'non-admixed' topologies and the bottom panel shows the 'admixed' topologies, sorted by their fit score. Above each graph, we show three text rows: 1. the four populations leading to the worst D-statistic residual after optimizing the model parameters; 2. the observed value for such statistic, the expected value under the fitted model, the residual, the standard error of the residuals and a Z-score for such residual; and 3. the model fit score. Numbers to the right of solid edges represent optimized drift parameters and percentages to the right of dashed edges represent admixture proportions.

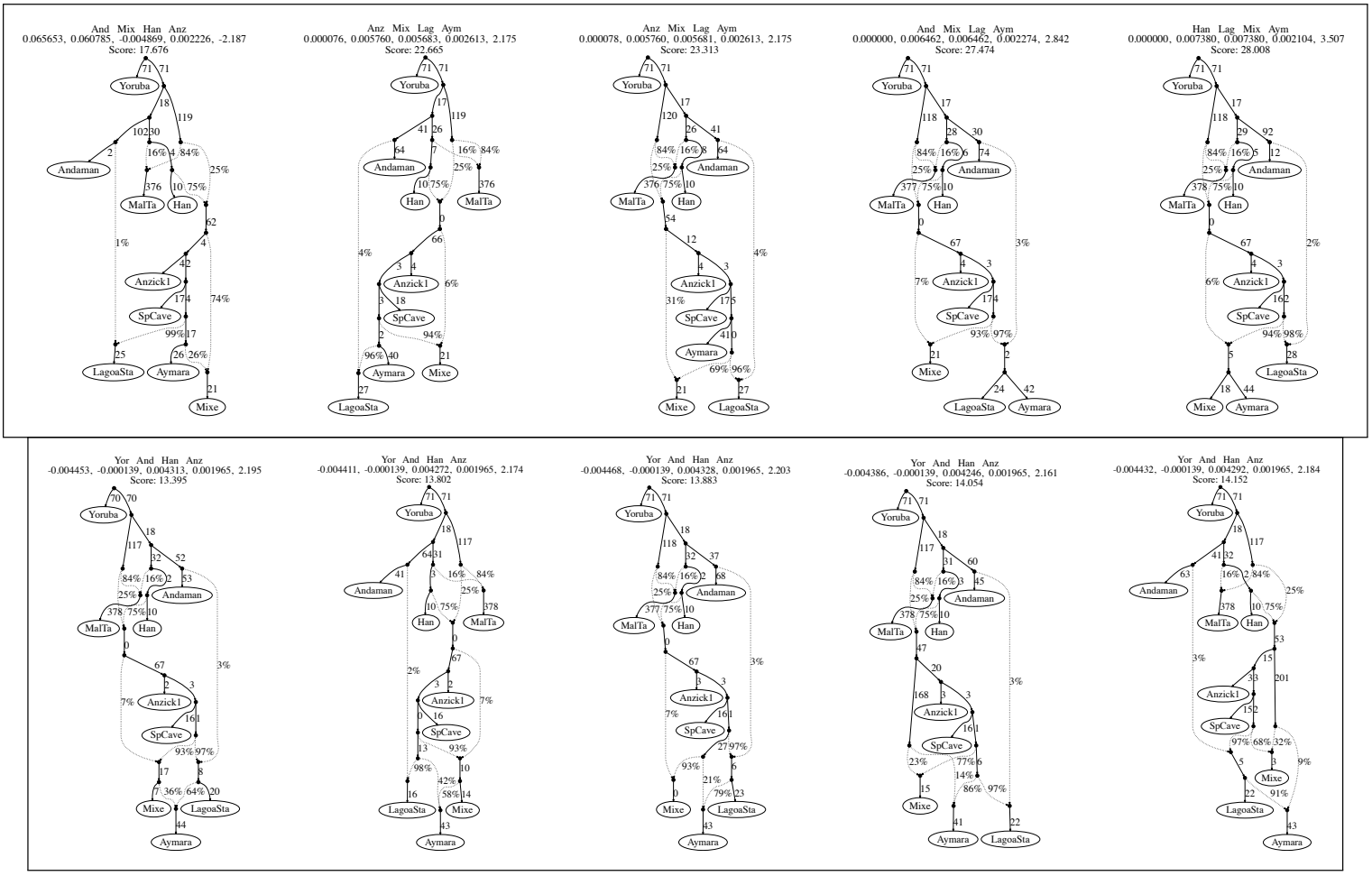


Figure S57. *f*-statistic-based admixture graphs exploring the placement of the present-day Aymara from Peru with respect to other Native Americans. We enumerated all possible extensions of the seed graph shown in Fig. 3A where we added Aymara as either a 'non-admixed' or an admixed population, and optimized the parameters for each topology using qpGraph. The top panel shows the five best-fitting 'non-admixed' topologies and the bottom panel shows the 'admixed' topologies, sorted by their fit score. Above each graph, we show three text rows: 1. the four populations leading to the worst D-statistic residual after optimizing the model parameters; 2. the observed value for such statistic, the expected value under the fitted model, the residual, the standard error of the residuals and a Z-score for such residual; and 3. the model fit score. Numbers to the right of solid edges represent optimized drift parameters and percentages to the right of dashed edges represent admixture proportions.

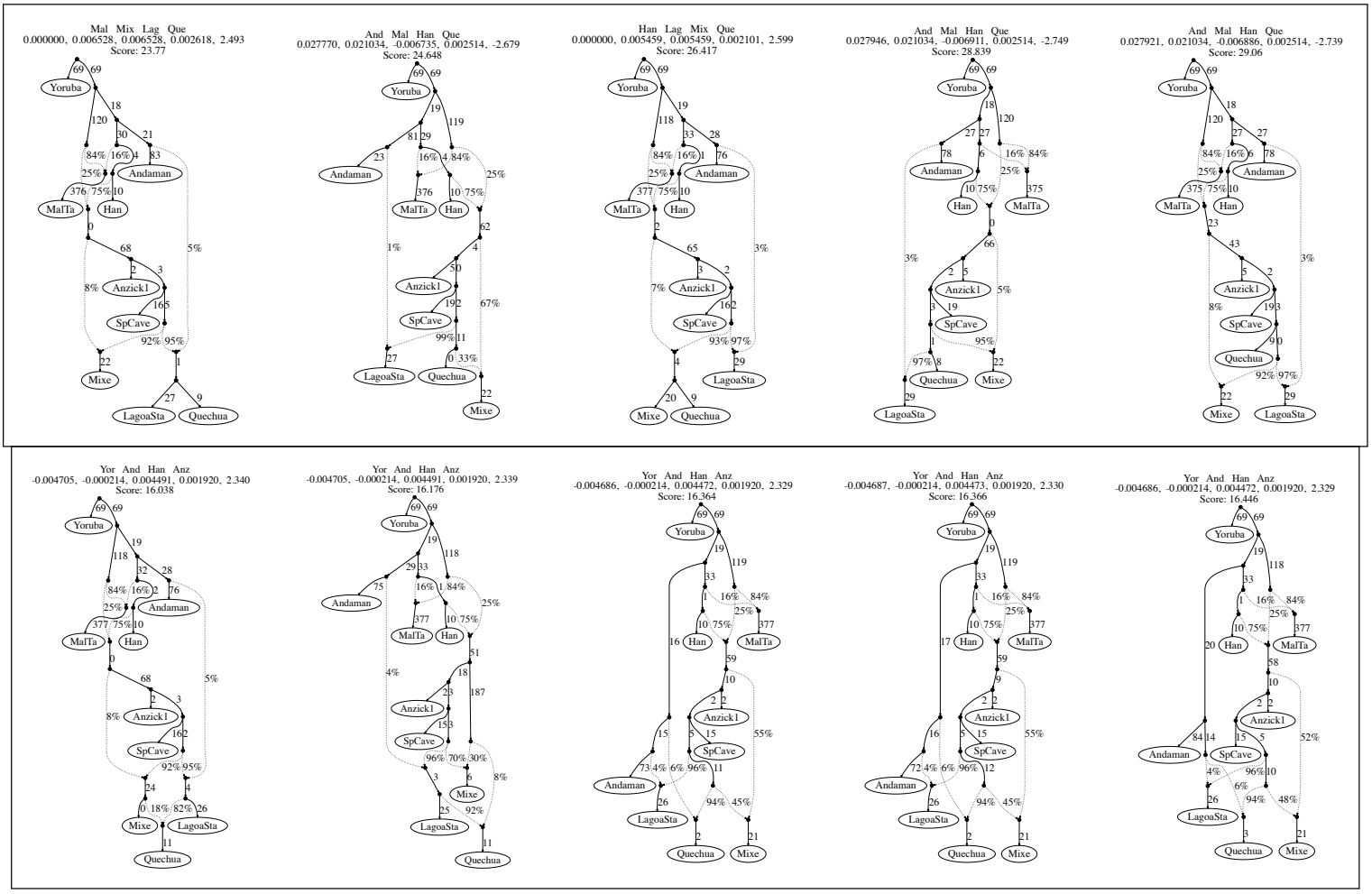
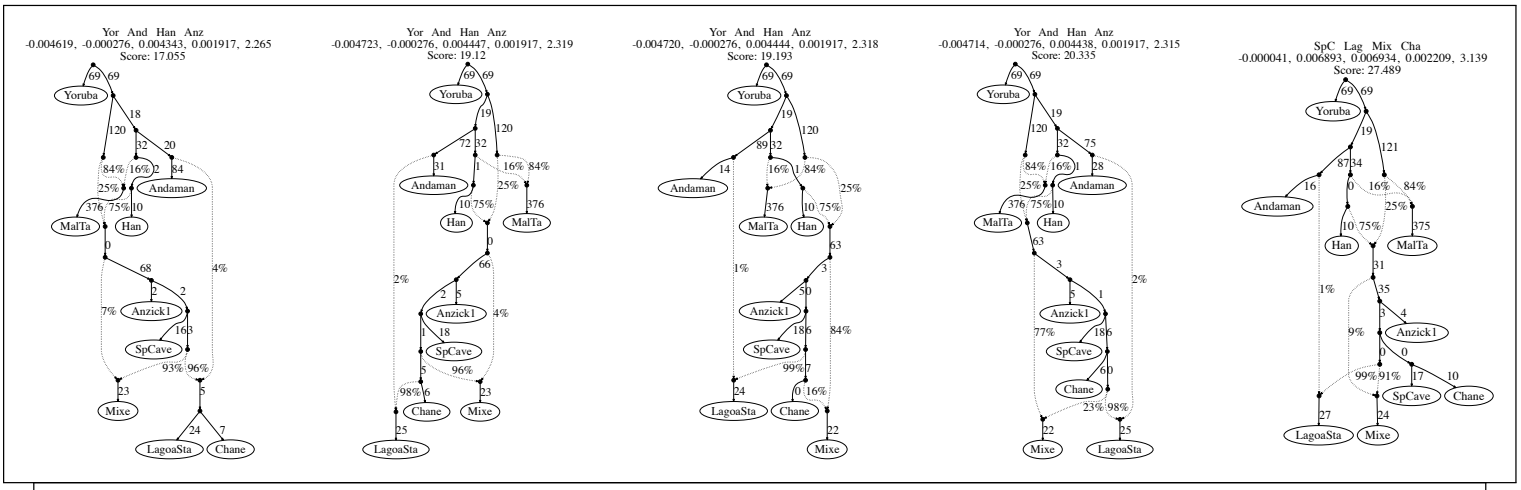


Figure S58. *f*-statistic-based admixture graphs exploring the placement of the present-day Quechua from Peru with respect to other Native Americans. We enumerated all possible extensions of the seed graph shown in Fig. 3A where we added Quechua as either a 'non-admixed' or an admixed population, and optimized the parameters for each topology using qpGraph. The top panel shows the five best-fitting 'non-admixed' topologies and the bottom panel shows the 'admixed' topologies, sorted by their fit score. Above each graph, we show three text rows: 1. the four populations leading to the worst D-statistic residual after optimizing the model parameters; 2. the observed value for such statistic, the expected value under the fitted model, the residual, the standard error of the residuals and a Z-score for such residual; and 3. the model fit score. Numbers to the right of solid edges represent optimized drift parameters and percentages to the right of dashed edges represent admixture proportions.



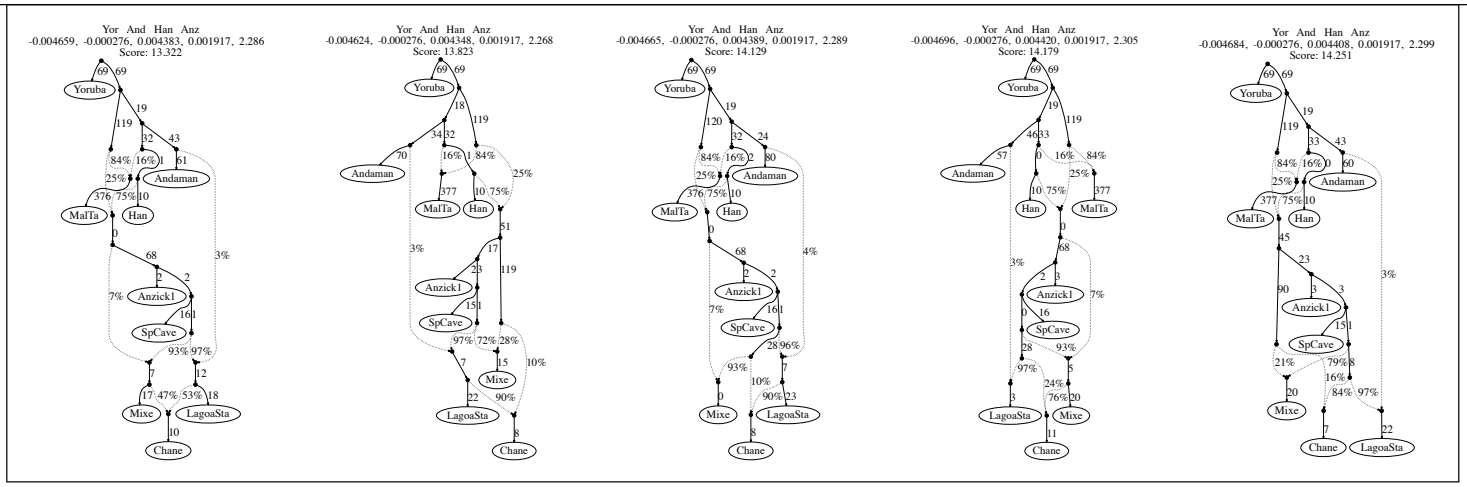


Figure S59. *f*-statistic-based admixture graphs exploring the placement of the present-day Chane from Argentina with respect to other Native Americans. We enumerated all possible extensions of the seed graph shown in Fig. 3A where we added Chane as either a 'non-admixed' or an admixed population, and optimized the parameters for each topology using qpGraph. The top panel shows the five best-fitting 'non-admixed' topologies and the bottom panel shows the 'admixed' topologies, sorted by their fit score. Above each graph, we show three text rows: 1. the four populations leading to the worst D-statistic residual after optimizing the model parameters; 2. the observed value for such statistic, the expected value under the fitted model, the residual, the standard error of the residuals and a Z-score for such residual; and 3. the model fit score. Numbers to the right of solid edges represent optimized drift parameters and percentages to the right of dashed edges represent admixture proportions.

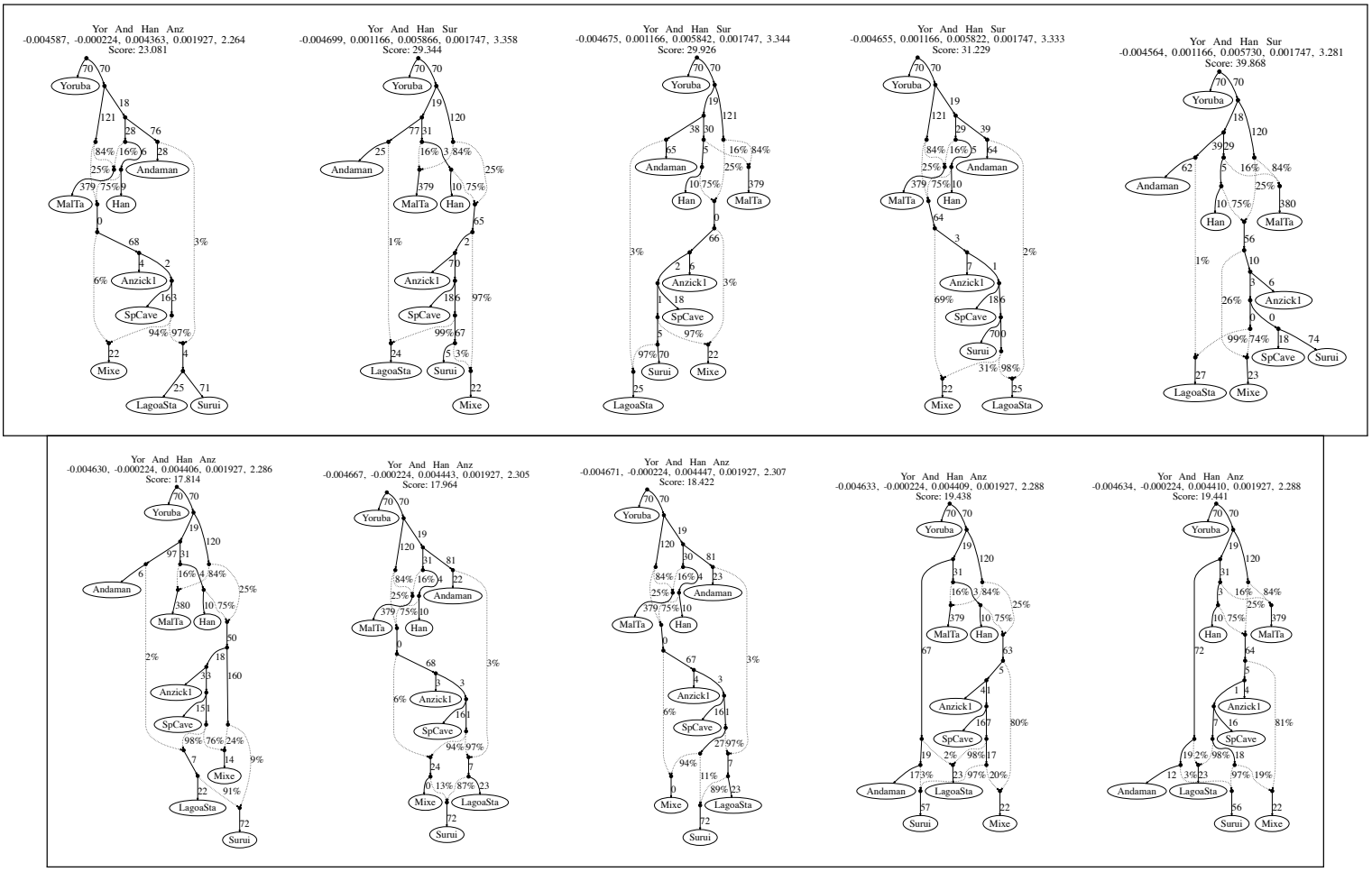


Figure S60. *f*-statistic-based admixture graphs exploring the placement of the present-day Surui from Brazil with respect to other Native Americans. We enumerated all possible extensions of the seed graph shown in Fig. 3A where we added Surui as either a 'non-admixed' or an admixed population, and optimized the parameters for each topology using qpGraph. The top panel shows the five best-fitting 'non-admixed' topologies and the bottom panel shows the 'admixed' topologies, sorted by their fit score. Above each graph, we show three text rows: 1. the four populations leading to the worst D-statistic residual after optimizing the model parameters; 2. the observed value for such statistic, the expected value under the fitted model, the residual, the standard error of the residuals and a Z-score for such residual; and 3. the model fit score. Numbers to the right of solid edges represent optimized drift parameters and percentages to the right of dashed edges represent admixture proportions.

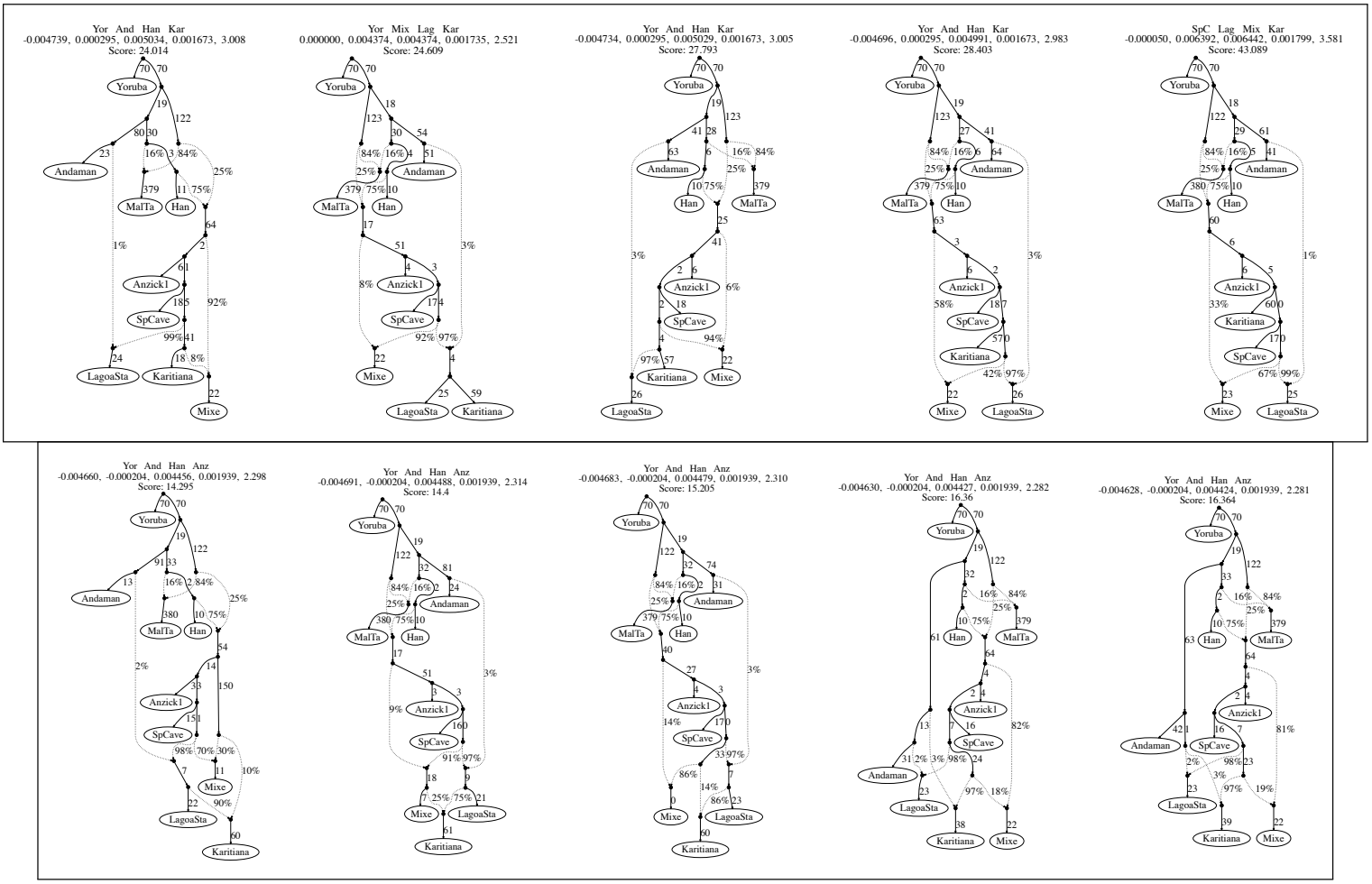


Figure S61. *f*-statistic-based admixture graphs exploring the placement of the present-day Karitiana from Brazil with respect to other Native Americans. We enumerated all possible extensions of the seed graph shown in Fig. 3A where we added Karitiana as either a 'non-admixed' or an admixed population, and optimized the parameters for each topology using qpGraph. The top panel shows the five best-fitting 'non-admixed' topologies and the bottom panel shows the 'admixed' topologies, sorted by their fit score. Above each graph, we show three text rows: 1. the four populations leading to the worst D-statistic residual after optimizing the model parameters; 2. the observed value for such statistic, the expected value under the fitted model, the residual, the standard error of the residuals and a Z-score for such residual; and 3. the model fit score. Numbers to the right of solid edges represent optimized drift parameters and percentages to the right of dashed edges represent admixture proportions.

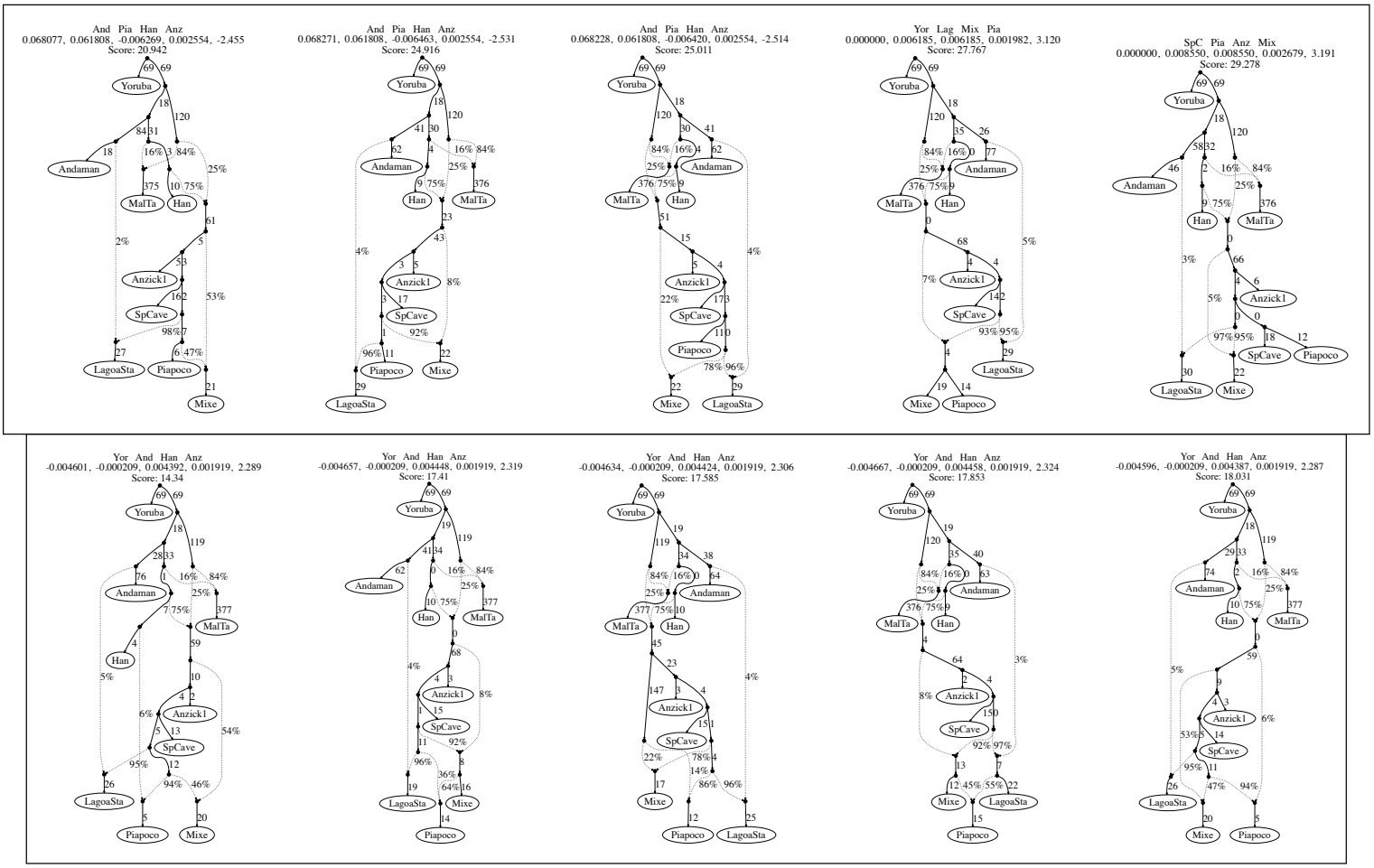


Figure S62. *f*-statistic-based admixture graphs exploring the placement of the present-day Piapoco from Brazil with respect to other Native Americans. We enumerated all possible extensions of the seed graph shown in Fig. 3A where we added Piapoco as either a 'non-admixed' or an admixed population, and optimized the parameters for each topology using qpGraph. The top panel shows the five best-fitting 'non-admixed' topologies and the bottom panel shows the 'admixed' topologies, sorted by their fit score. Above each graph, we show three text rows: 1. the four populations leading to the worst D-statistic residual after optimizing the model parameters; 2. the observed value for such statistic, the expected value under the fitted model, the residual, the standard error of the residuals and a Z-score for such residual; and 3. the model fit score. Numbers to the right of solid edges represent optimized drift parameters and percentages to the right of dashed edges represent admixture proportions.

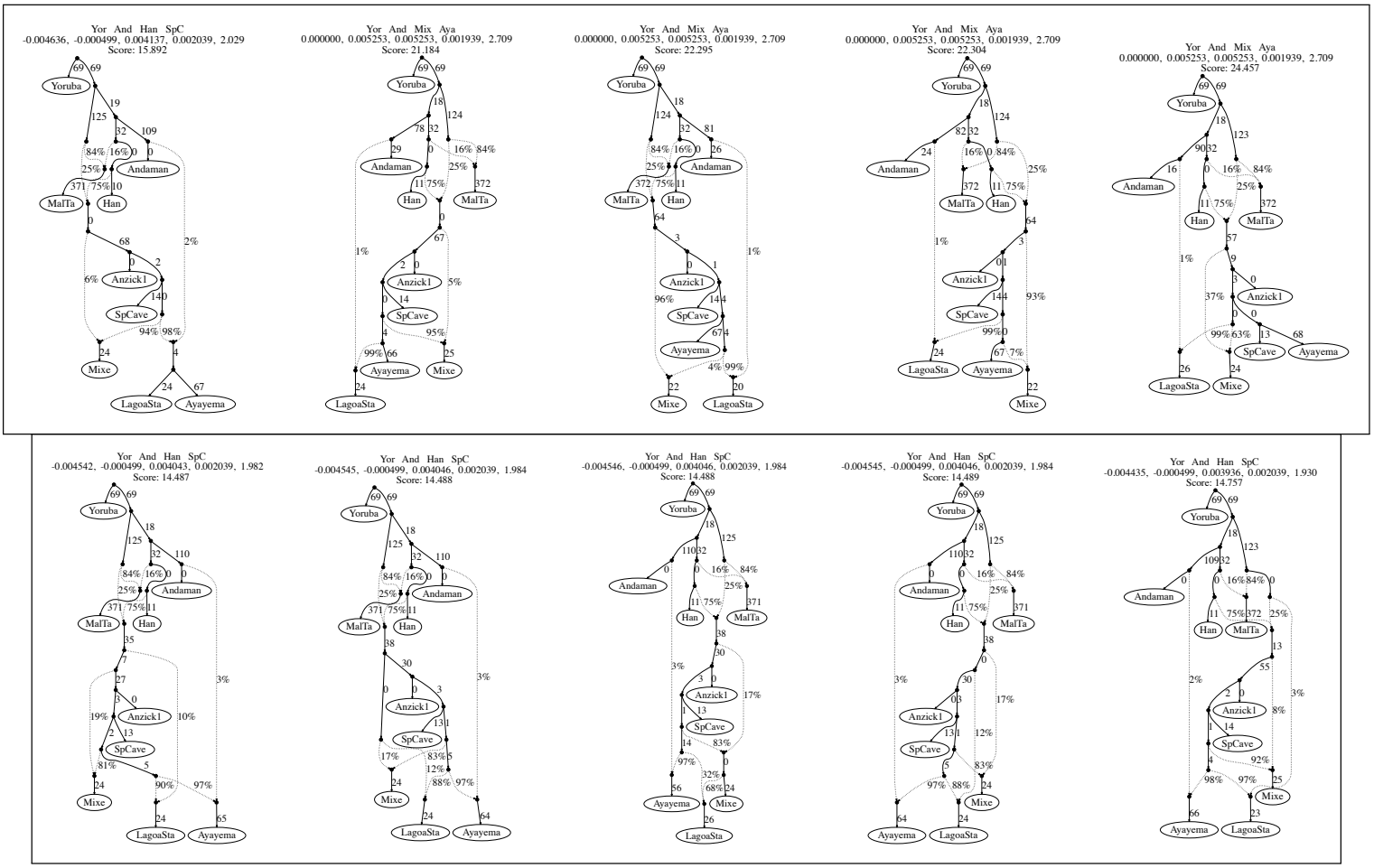


Figure S63. *f*-statistic-based admixture graphs exploring the placement of the Ayayema ancient genome from Patagonia with respect to other Native Americans. We enumerated all possible extensions of the seed graph shown in Fig. 3A where we added Ayayema as either a 'non-admixed' or an admixed population, and optimized the parameters for each topology using qpGraph. The top panel shows the five best-fitting 'non-admixed' topologies and the bottom panel shows the 'admixed' topologies, sorted by their fit score. Above each graph, we show three text rows: 1. the four populations leading to the worst D-statistic residual after optimizing the model parameters; 2. the observed value for such statistic, the expected value under the fitted model, the residual, the standard error of the residuals and a Z-score for such residual; and 3. the model fit score. Numbers to the right of solid edges represent optimized drift parameters and percentages to the right of dashed edges represent admixture proportions.

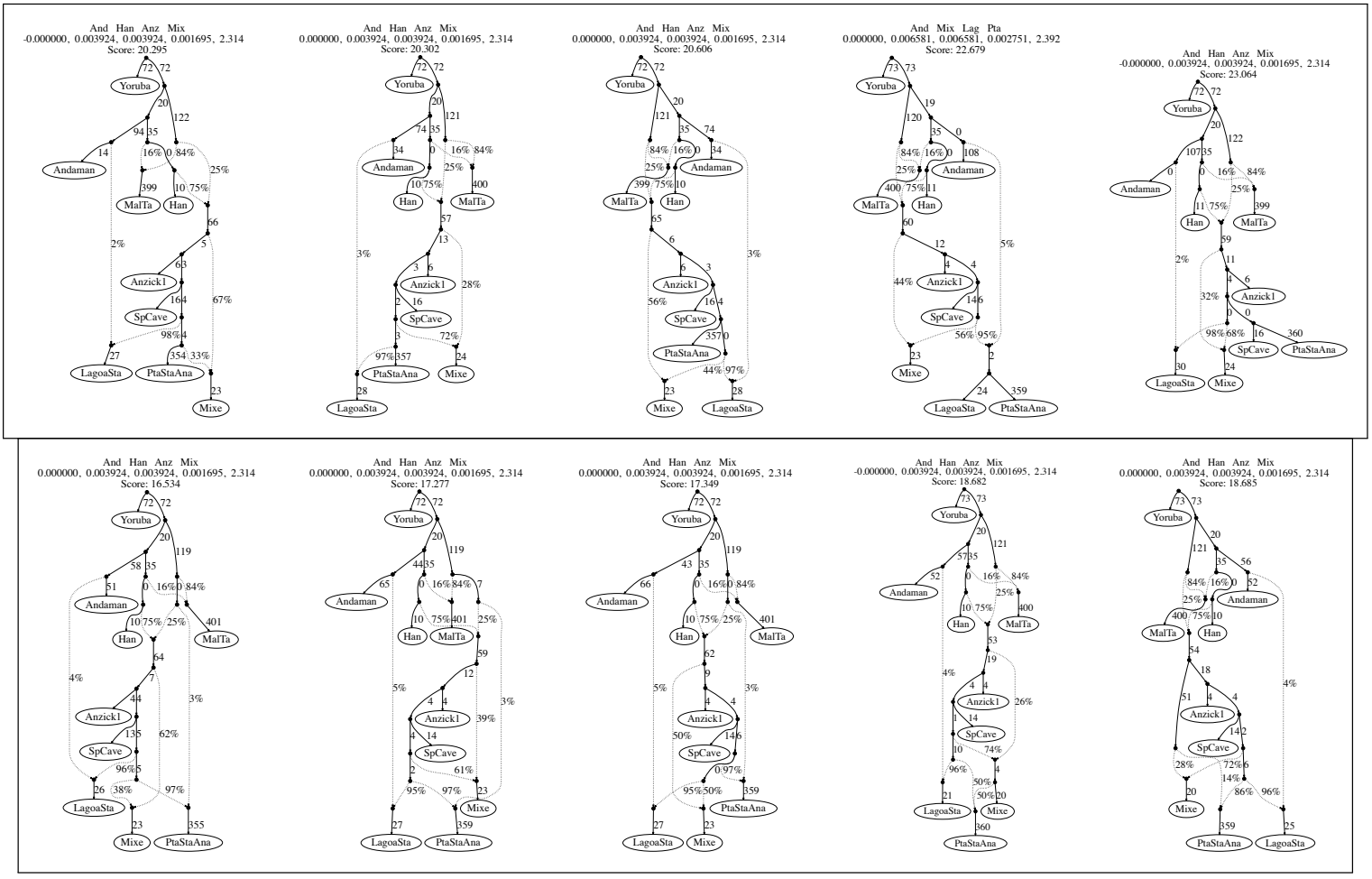


Figure S64. *f*-statistic-based admixture graphs exploring the placement of the Punta Santa Ana ancient genome from Patagonia with respect to other Native Americans. We enumerated all possible extensions of the seed graph shown in Fig. 3A where we added Punta Santa Ana as either a 'non-admixed' or an admixed population, and optimized the parameters for each topology using qpGraph. The top panel shows the five best-fitting 'non-admixed' topologies and the bottom panel shows the 'admixed' topologies, sorted by their fit score. Above each graph, we show three text rows: 1. the four populations leading to the worst D-statistic residual after optimizing the model parameters; 2. the observed value for such statistic, the expected value under the fitted model, the residual, the standard error of the residuals and a Z-score for such residual; and 3. the model fit score. Numbers to the right of solid edges represent optimized drift parameters and percentages to the right of dashed edges represent admixture proportions.

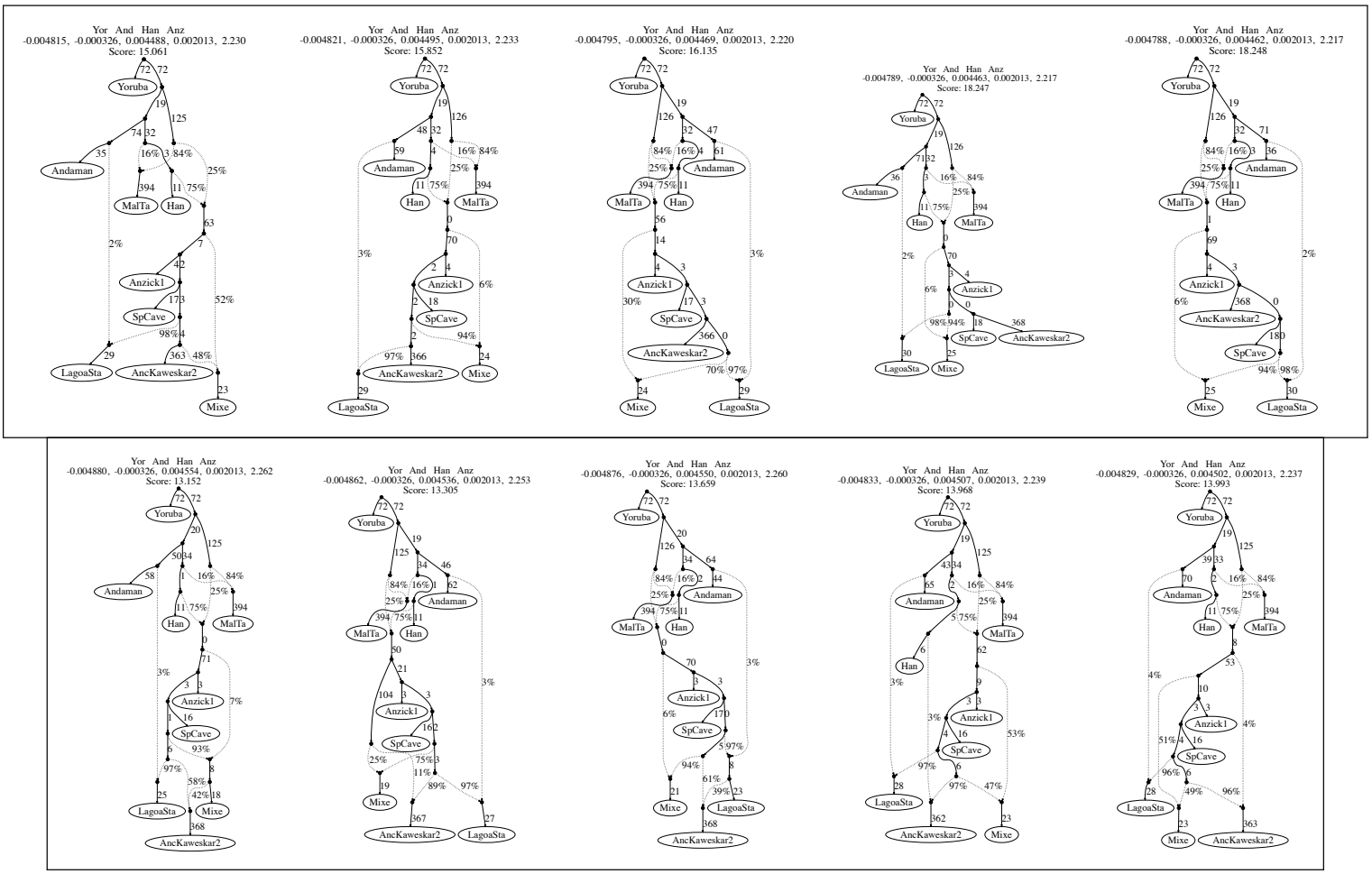


Figure S65. *f*-statistic-based admixture graphs exploring the placement of the Kaweskar2 ancient genome from Patagonia with respect to other Native Americans. We enumerated all possible extensions of the seed graph shown in Fig. 3A where we added Kaweskar2 as either a 'non-admixed' or an admixed population, and optimized the parameters for each topology using qpGraph. The top panel shows the five best-fitting 'non-admixed' topologies and the bottom panel shows the 'admixed' topologies, sorted by their fit score. Above each graph, we show three text rows: 1. the four populations leading to the worst D-statistic residual after optimizing the model parameters; 2. the observed value for such statistic, the expected value under the fitted model, the residual, the standard error of the residuals and a Z-score for such residual; and 3. the model fit score. Numbers to the right of solid edges represent optimized drift parameters and percentages to the right of dashed edges represent admixture proportions.

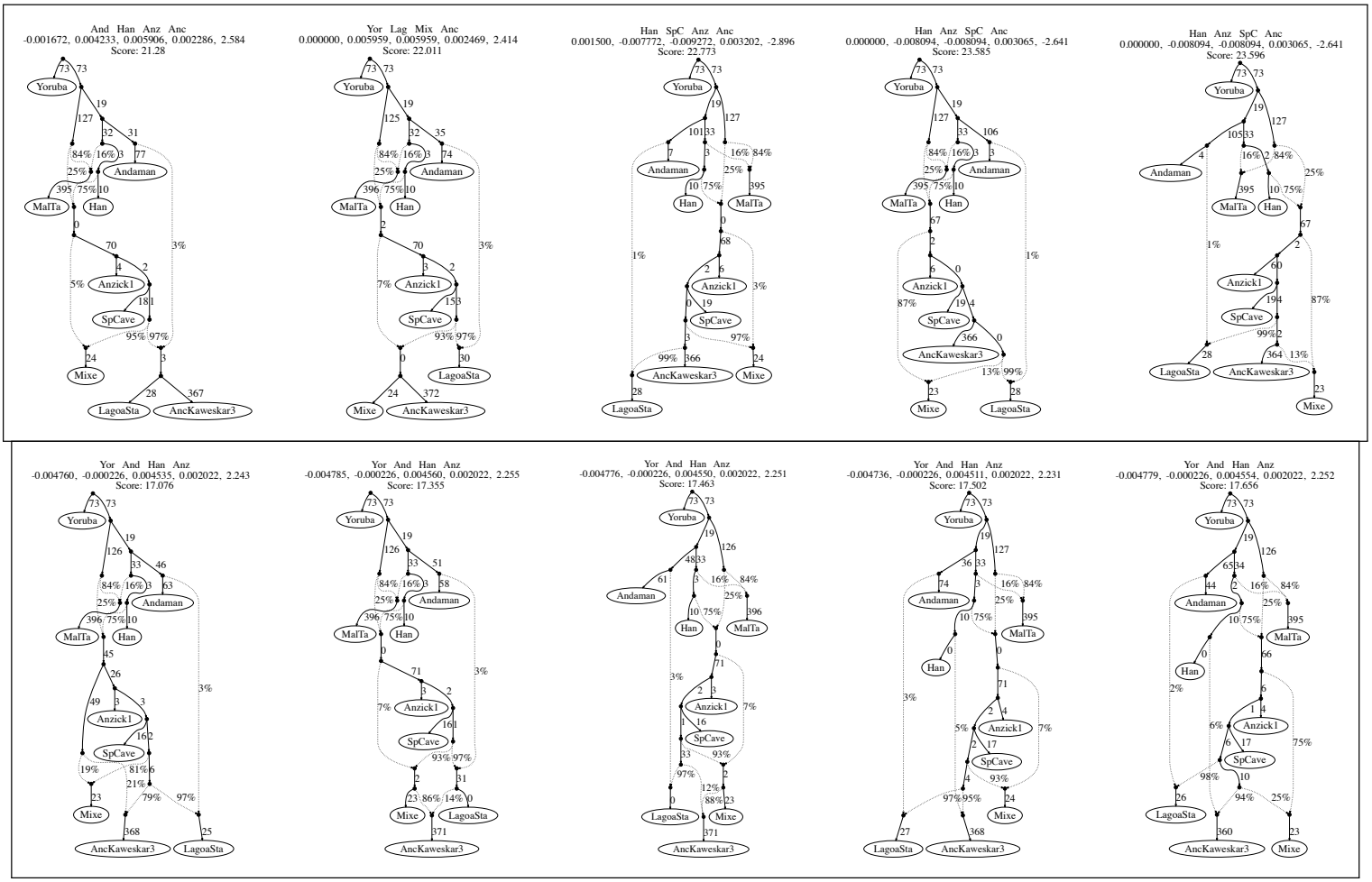


Figure S66. *f*-statistic-based admixture graphs exploring the placement of the Kaweskar3 ancient genome from Patagonia with respect to other Native Americans. We enumerated all possible extensions of the seed graph shown in Fig. 3A where we added Kaweskar3 as either a 'non-admixed' or an admixed population, and optimized the parameters for each topology using qpGraph. The top panel shows the five best-fitting 'non-admixed' topologies and the bottom panel shows the 'admixed' topologies, sorted by their fit score. Above each graph, we show three text rows: 1. the four populations leading to the worst D-statistic residual after optimizing the model parameters; 2. the observed value for such statistic, the expected value under the fitted model, the residual, the standard error of the residuals and a Z-score for such residual; and 3. the model fit score. Numbers to the right of solid edges represent optimized drift parameters and percentages to the right of dashed edges represent admixture proportions.

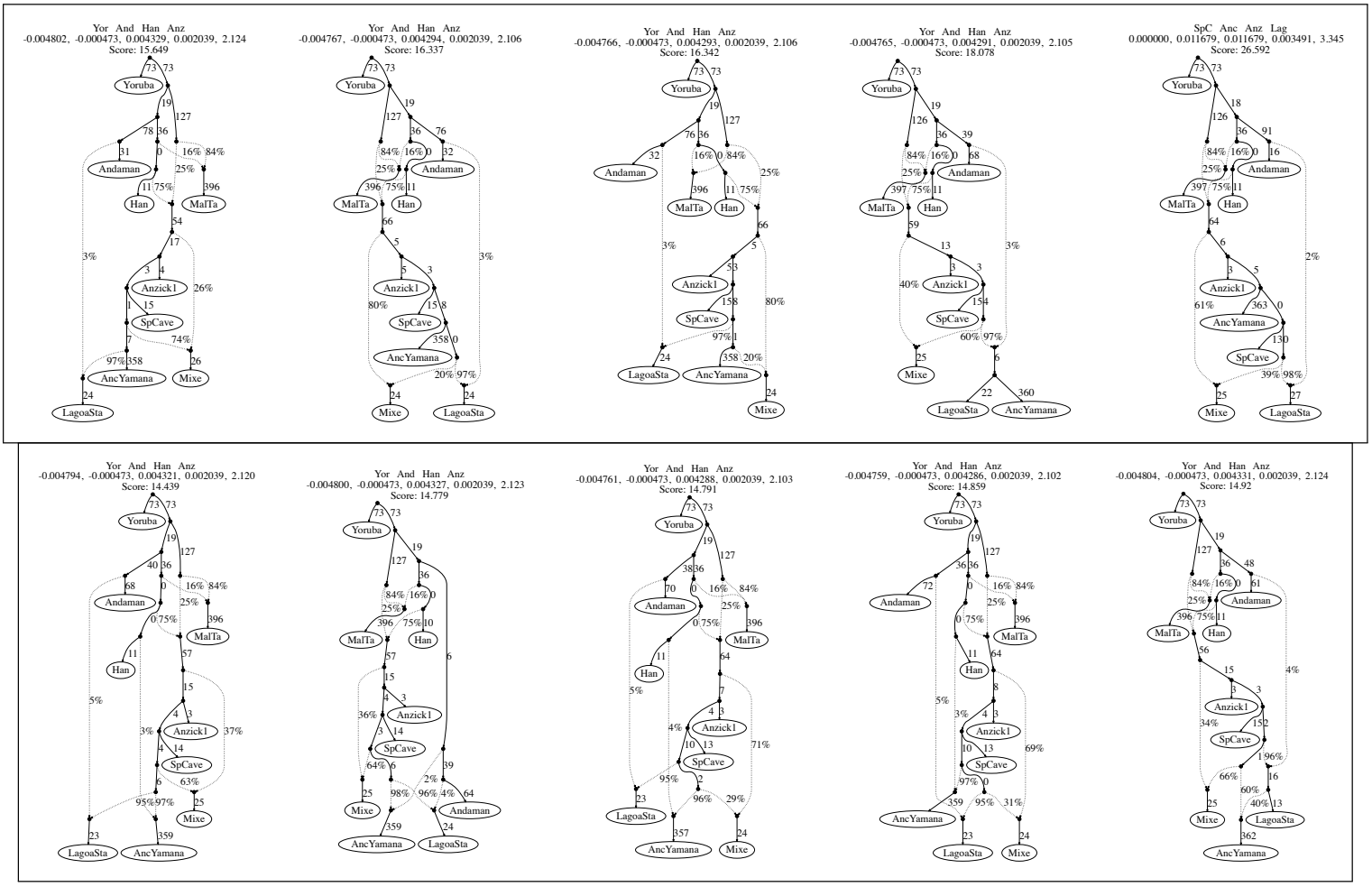


Figure S67. *f*-statistic-based admixture graphs exploring the placement of the Yamana ancient genome from Patagonia with respect to other Native Americans. We enumerated all possible extensions of the seed graph shown in Fig. 3A where we added Yamana as either a 'non-admixed' or an admixed population, and optimized the parameters for each topology using qpGraph. The top panel shows the five best-fitting 'non-admixed' topologies and the bottom panel shows the 'admixed' topologies, sorted by their fit score. Above each graph, we show three text rows: 1. the four populations leading to the worst D-statistic residual after optimizing the model parameters; 2. the observed value for such statistic, the expected value under the fitted model, the residual, the standard error of the residuals and a Z-score for such residual; and 3. the model fit score. Numbers to the right of solid edges represent optimized drift parameters and percentages to the right of dashed edges represent admixture proportions.

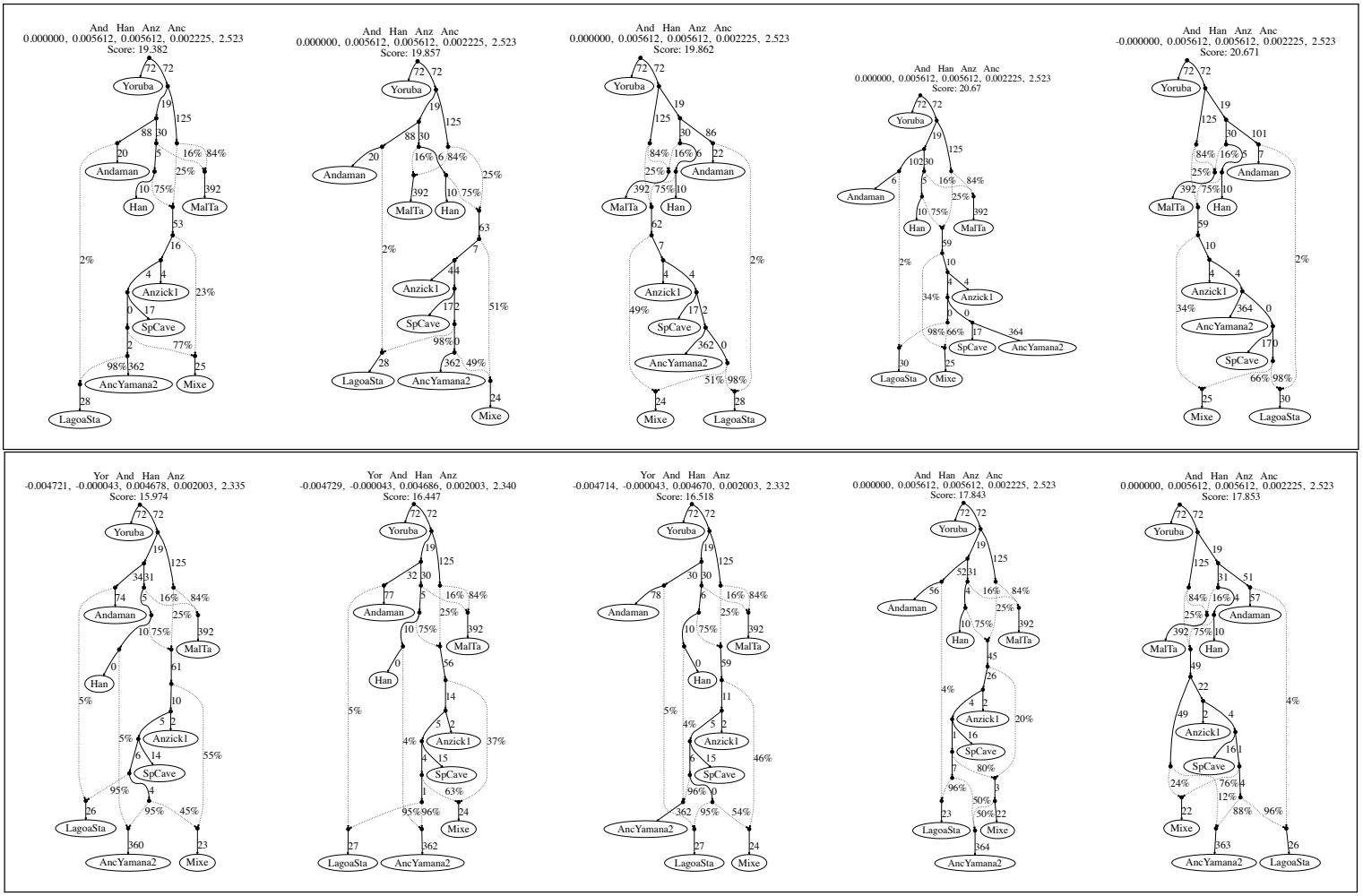


Figure S68. *f*-statistic-based admixture graphs exploring the placement of the Yamana2 ancient genome from Patagonia with respect to other Native Americans. We enumerated all possible extensions of the seed graph shown in Fig. 3A where we added Yamana2 as either a 'non-admixed' or an admixed population, and optimized the parameters for each topology using qpGraph. The top panel shows the five best-fitting 'non-admixed' topologies and the bottom panel shows the 'admixed' topologies, sorted by their fit score. Above each graph, we show three text rows: 1. the four populations leading to the worst D-statistic residual after optimizing the model parameters; 2. the observed value for such statistic, the expected value under the fitted model, the residual, the standard error of the residuals and a Z-score for such residual; and 3. the model fit score. Numbers to the right of solid edges represent optimized drift parameters and percentages to the right of dashed edges represent admixture proportions.

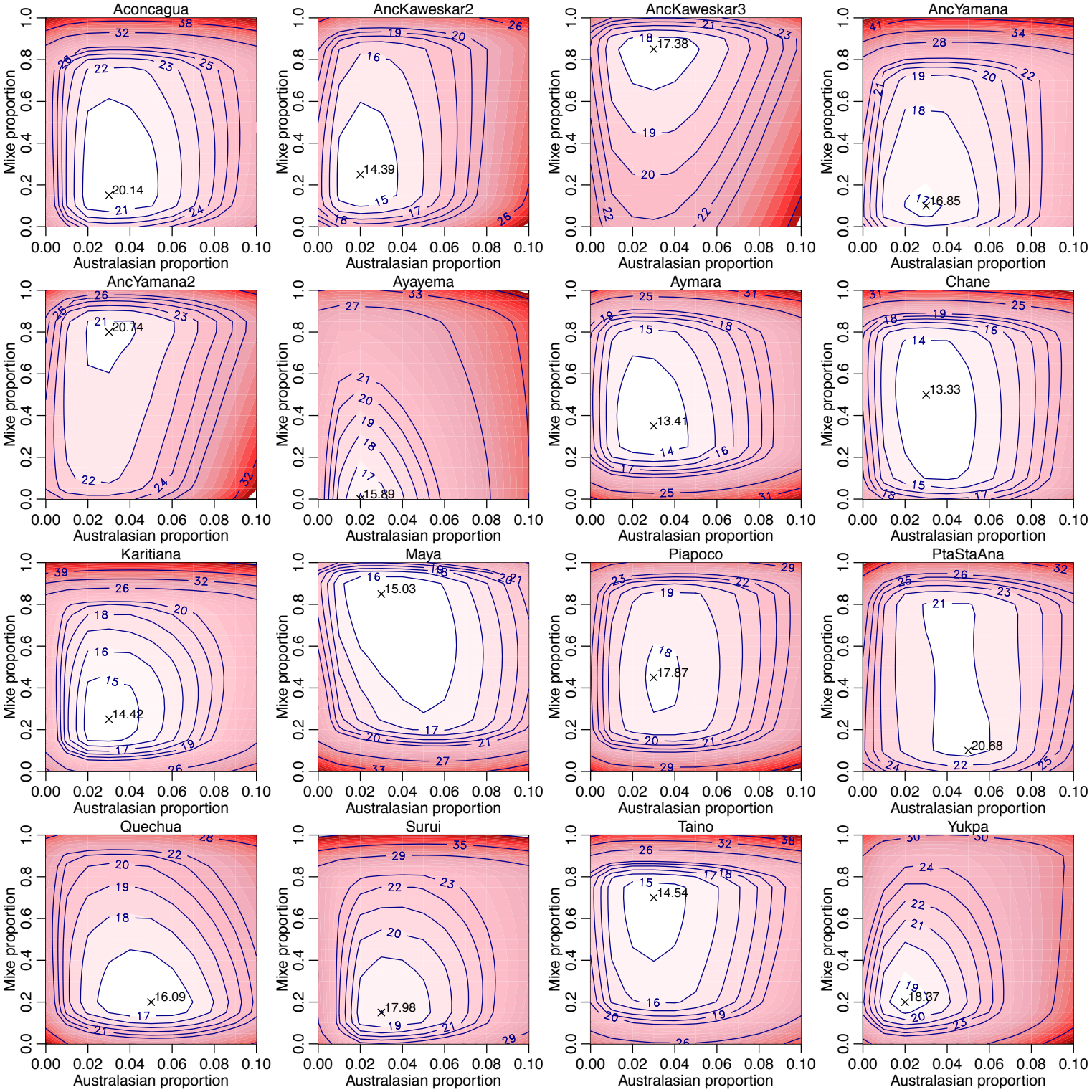


Figure S69. Fit score surfaces for the 'admixed' SNA model, with fixed Mixe and Australasian admixture proportions. For each SNA population, we fitted an f-statistic-based admixture graph similar to the shown in Fig. 3A. In this model, SNA populations derive from an admixture event between two populations most closely related to Lagoa Santa and Mixe, respectively. We explored the fit of the model across a grid of values for the Mixe proportion into SNA {0,0.05,...,1} and the Australasian contribution into Lagoa Santa {0,0.01,...0.1}. 'X' indicates the parameter combination yielding the best score for each population. Contour lines were drawn such that all parameter combinations contained within a given line, all yield a fit score lower than that indicated by the contour label. In practice, we considered a model to be a significant improvement compared to another if the absolute difference between fit scores is greater than ~4.6, which corresponds to a p-value~0.01.

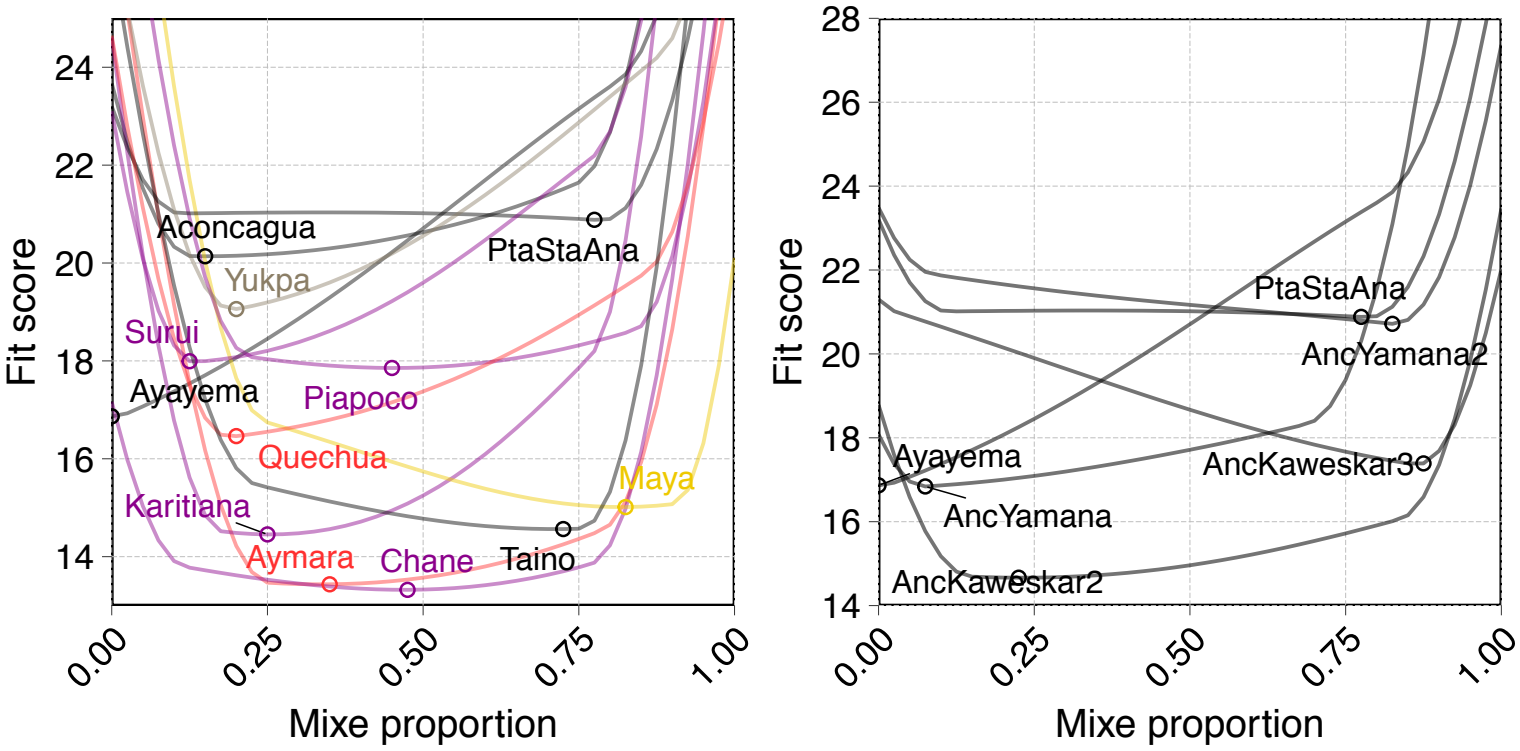


Figure S70. SNA admixed model fit score as a function of Mixe admixture into SNA. A one-dimensional representation of Fig. S70, where the Australasian contribution into Lagoa Santa was fixed to ~0.02. For each SNA population, we fitted an f-statistic-based admixture graph similar to the shown in Fig. 3A. In this model, SNA populations derive from an admixture event between two populations most closely related to Lagoa Santa and Mixe, respectively. We explored the fit of the model across a range values for the Mixe proportion into SNA {0,0.05,...,1}. Each line is labeled at the value that yields the best-fit score. In practice, we considered a model to be a significant improvement compared to another if the absolute difference between fit scores is greater than ~4.6, which corresponds to a p-value~0.01.

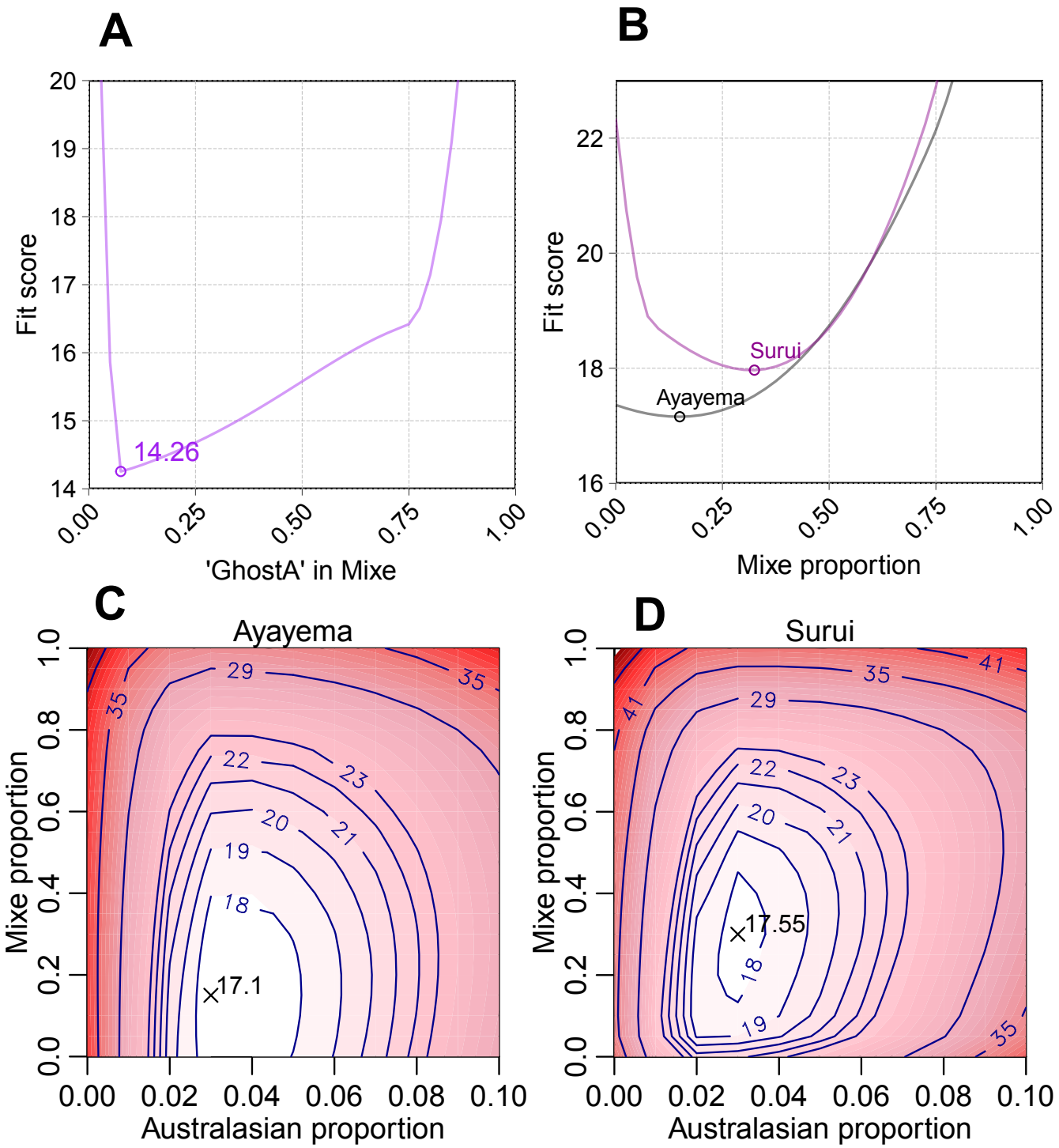


Figure S71. Fit score surfaces for Mixe, Surui and Ayayema, with fixed Mixe and Australasian admixture proportions, using the Human Origins dataset. We reproduced the results shown in Figures S47, S70 and S71, but using the Human Origins dataset. **a.** We show that a model where Mixe forms a clade with Lagoa Santa but also bears 'unsampled admixture' produces a significantly better fit than a 'non-admixed' model (related to Figure S47). **b.** We show that an admixed model is the best-fit for Surui whereas a 'non-admixed' model is the best fit for Ayayema (related to Figure S71). Each line is labeled at the value that yields the best-fit score. **c.** We show that an admixed model is the best-fit for Surui whereas a 'non-admixed' model is the best fit for Ayayema. Moreover, we find that Lagoa Santa requires non-zero Australasian admixture (related to Figure S70). 'X' indicates the parameter combination yielding the best score for each population. Contour lines were drawn such that all parameter combinations contained within a given line, all yield a fit score lower than that indicated by the contour label. In practice, we considered a model to be a significant improvement compared to another if the absolute difference between fit scores is greater than ~4.6, which corresponds to a p-value~0.01.

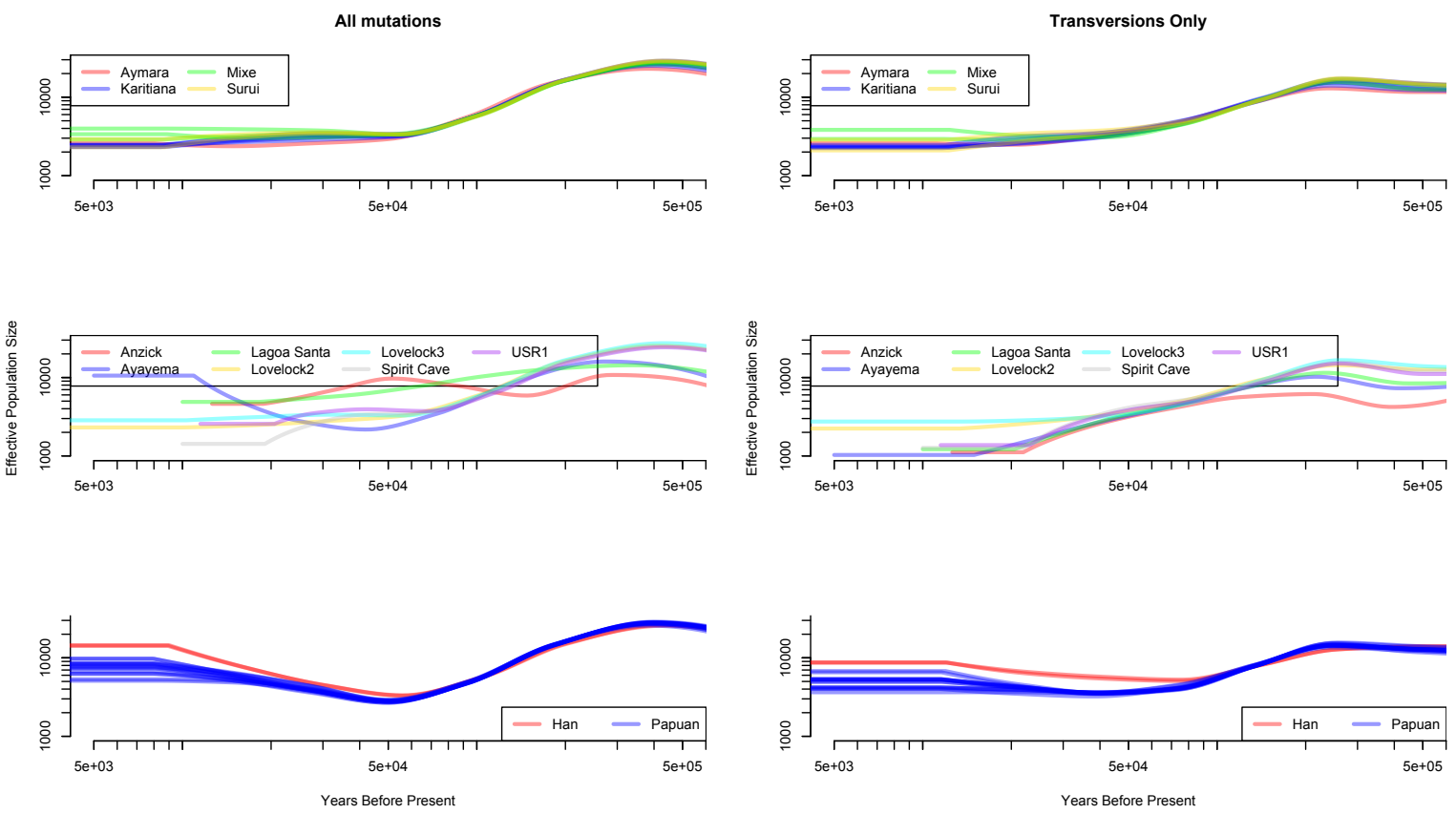


Figure S72. Population sizes inferred using *smc*++.

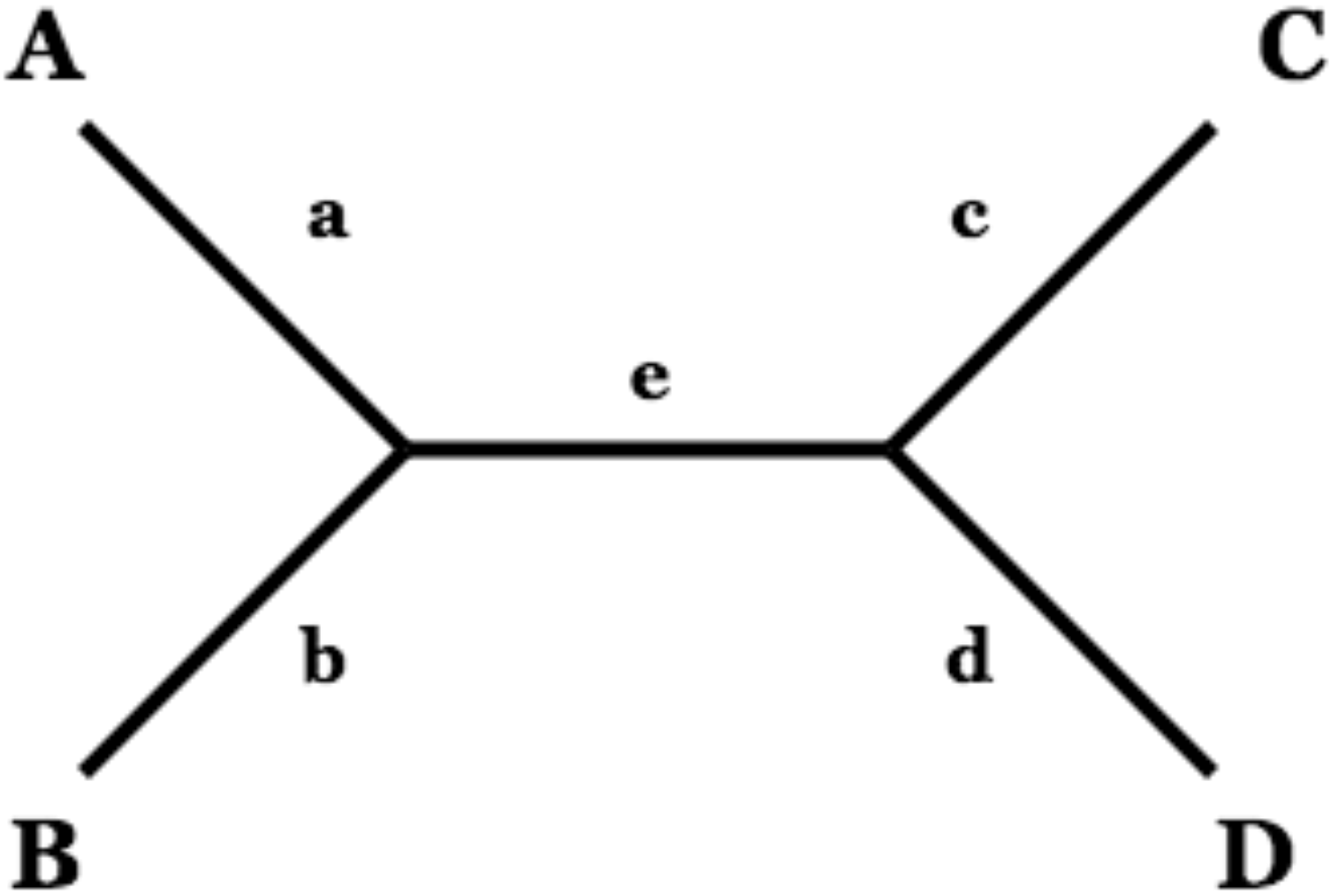


Figure S73. An unrooted labeled tree topology with 4 leaves. We show via *F3* statistics that (A = Spirit Cave; B = Anzick; C = Lagoa Santa; D = Mixe).

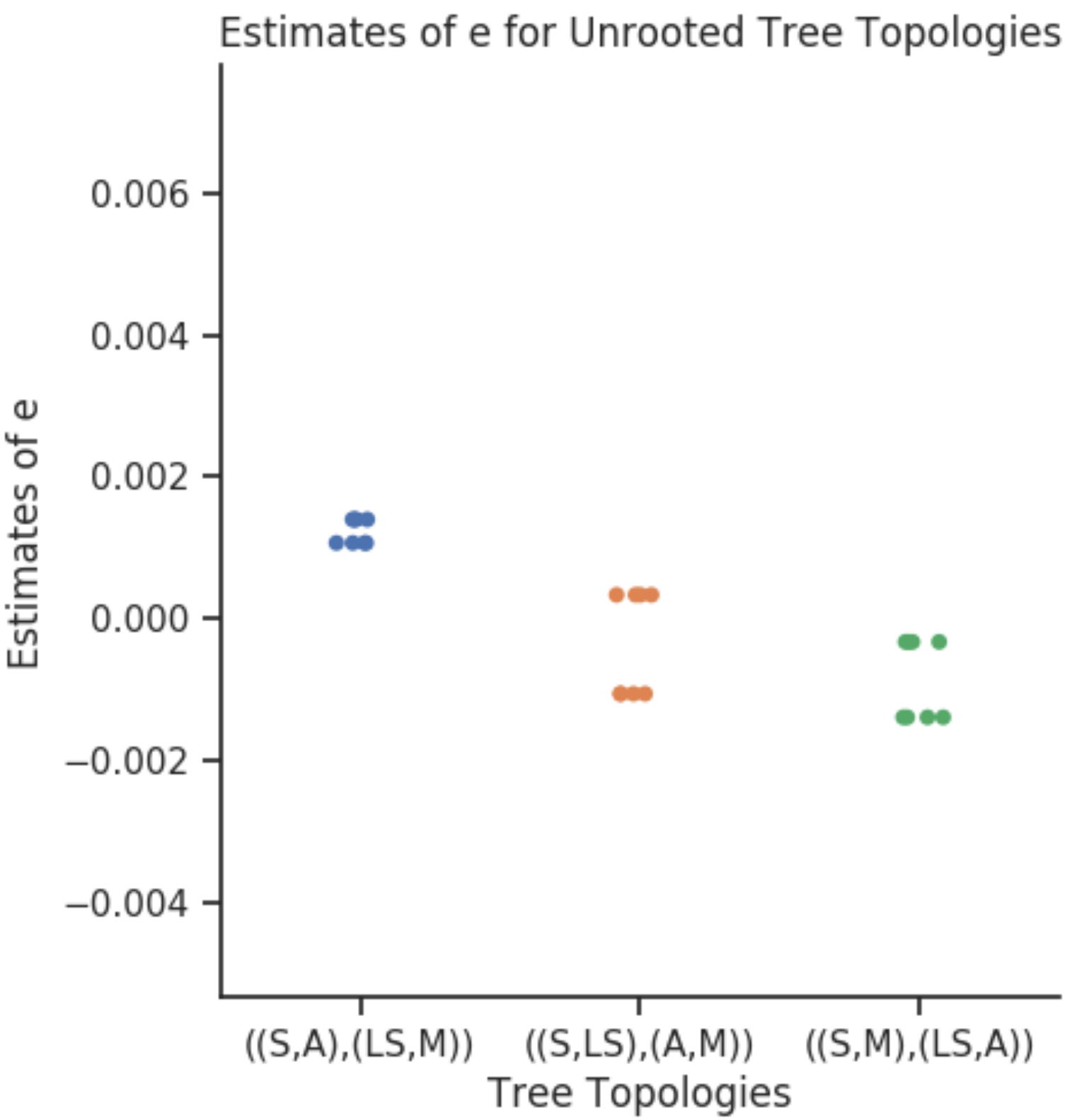


Figure S74. Estimates for $\hat{e}$ derived from *F3* statistics.

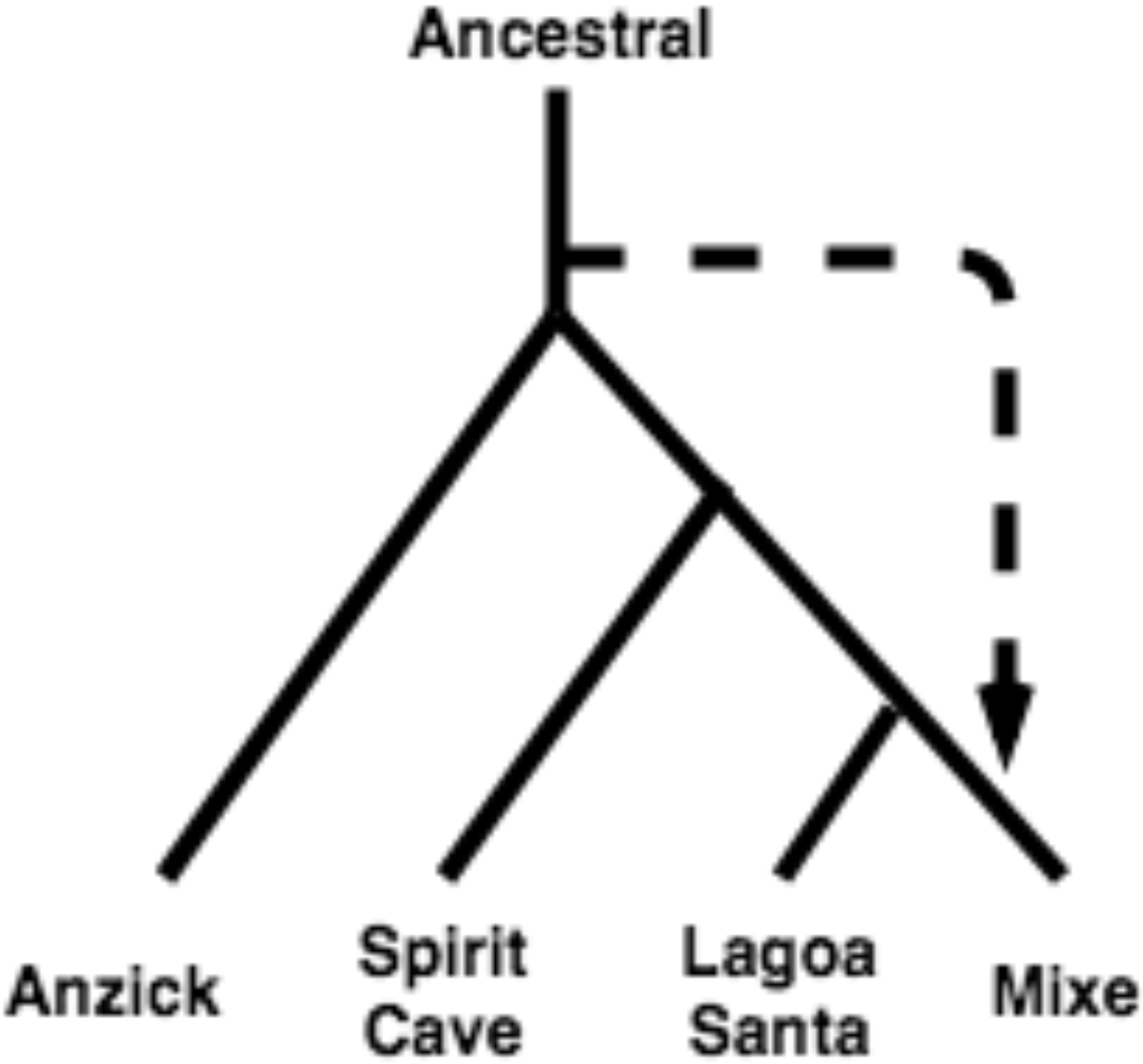


Figure S75. Admixture graph derived from *f*-statistics for Ancestral, Anzick, Spirit Cave, Lagoa Santa, and Mixe. The dashed line indicates an admixture event.

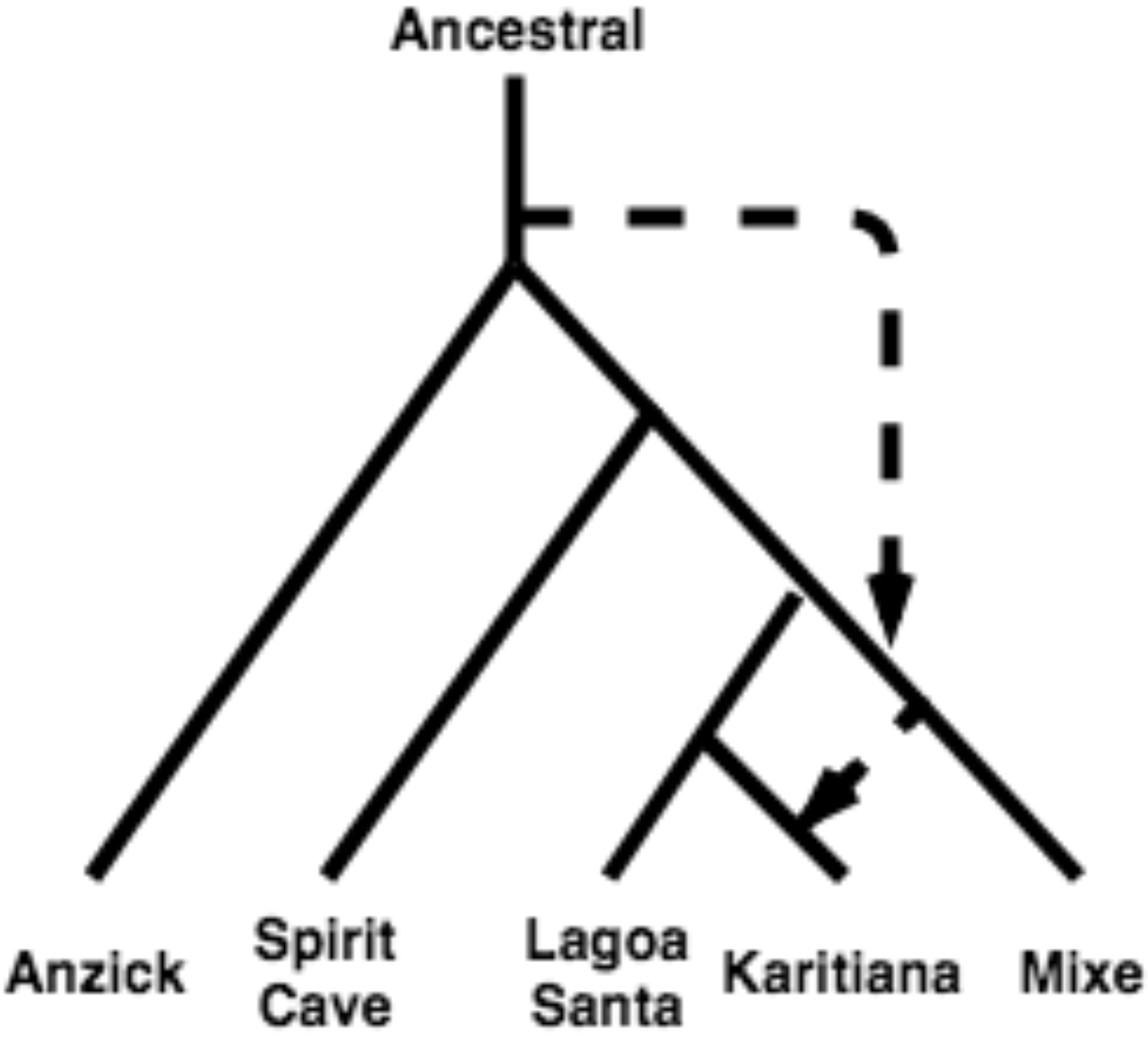
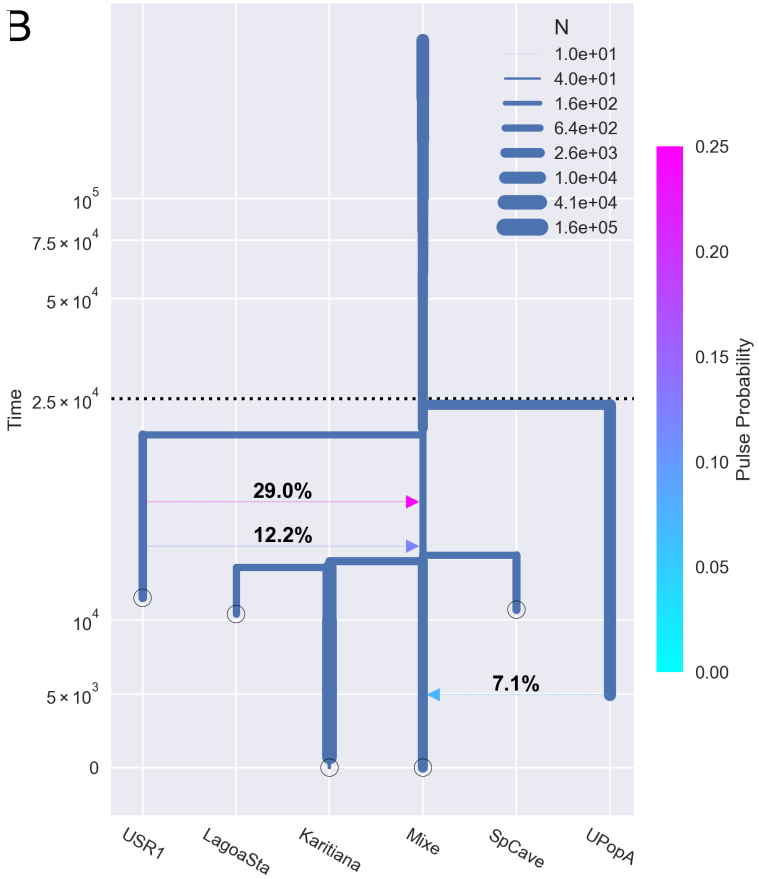
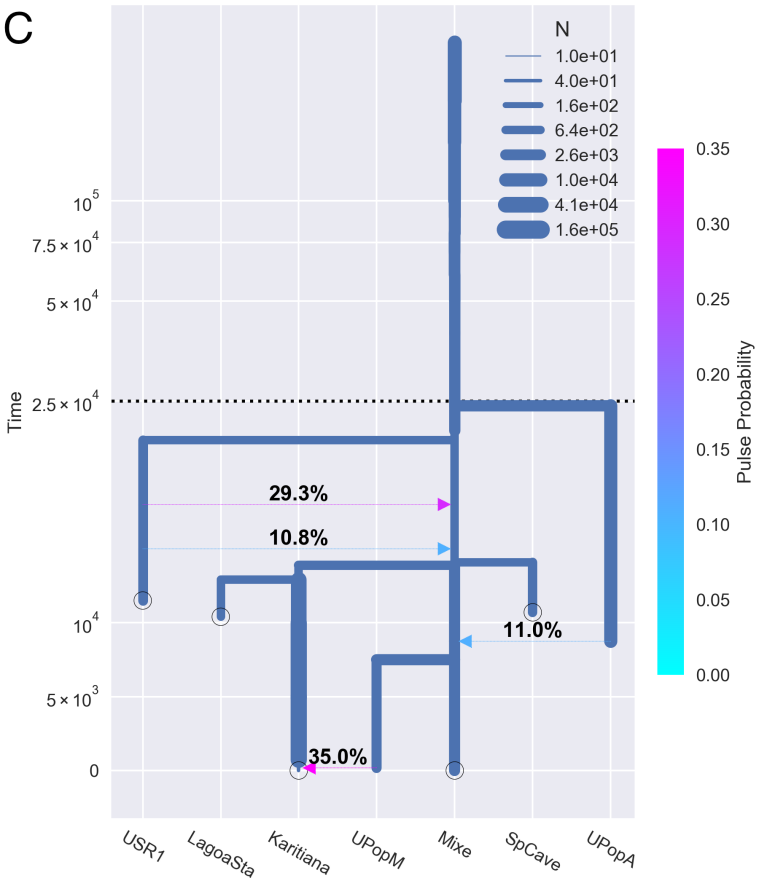


Figure S76. Admixture graph including Karitiana derived from *f*-statistics. The dashed line indicates an admixture event.





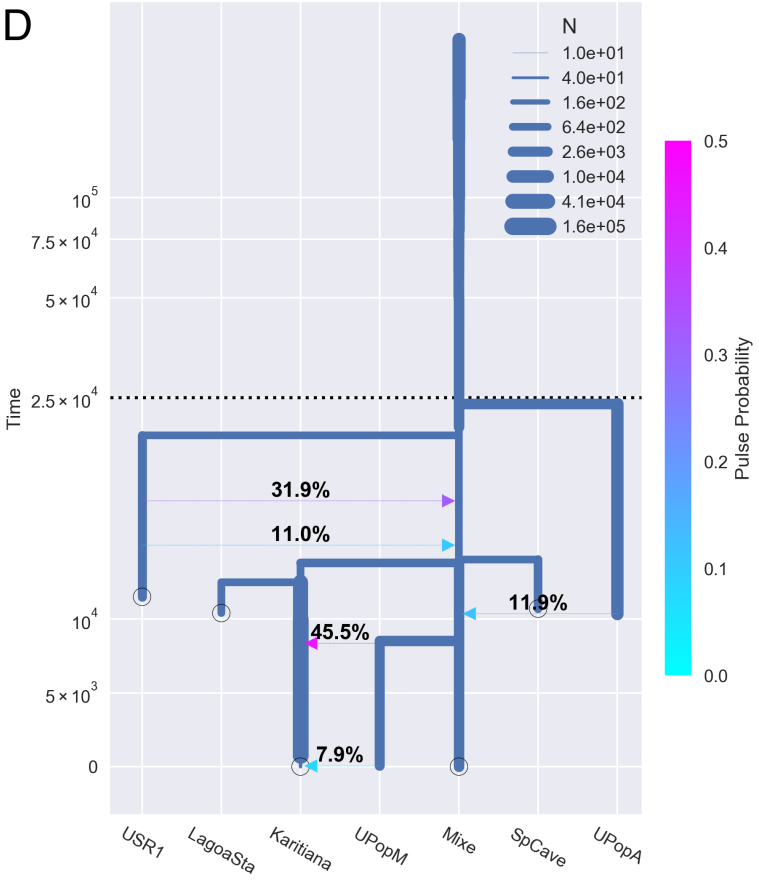


Figure S77. Nested demographic models. A. Model 1 with 5 populations and 2 pulses. **B.** Model 2 with 6 populations and 3 pulses. **C.** Model 3 with 7 populations and 4 pulses. **D.** Model 4 with 7 populations and 5 pulses.

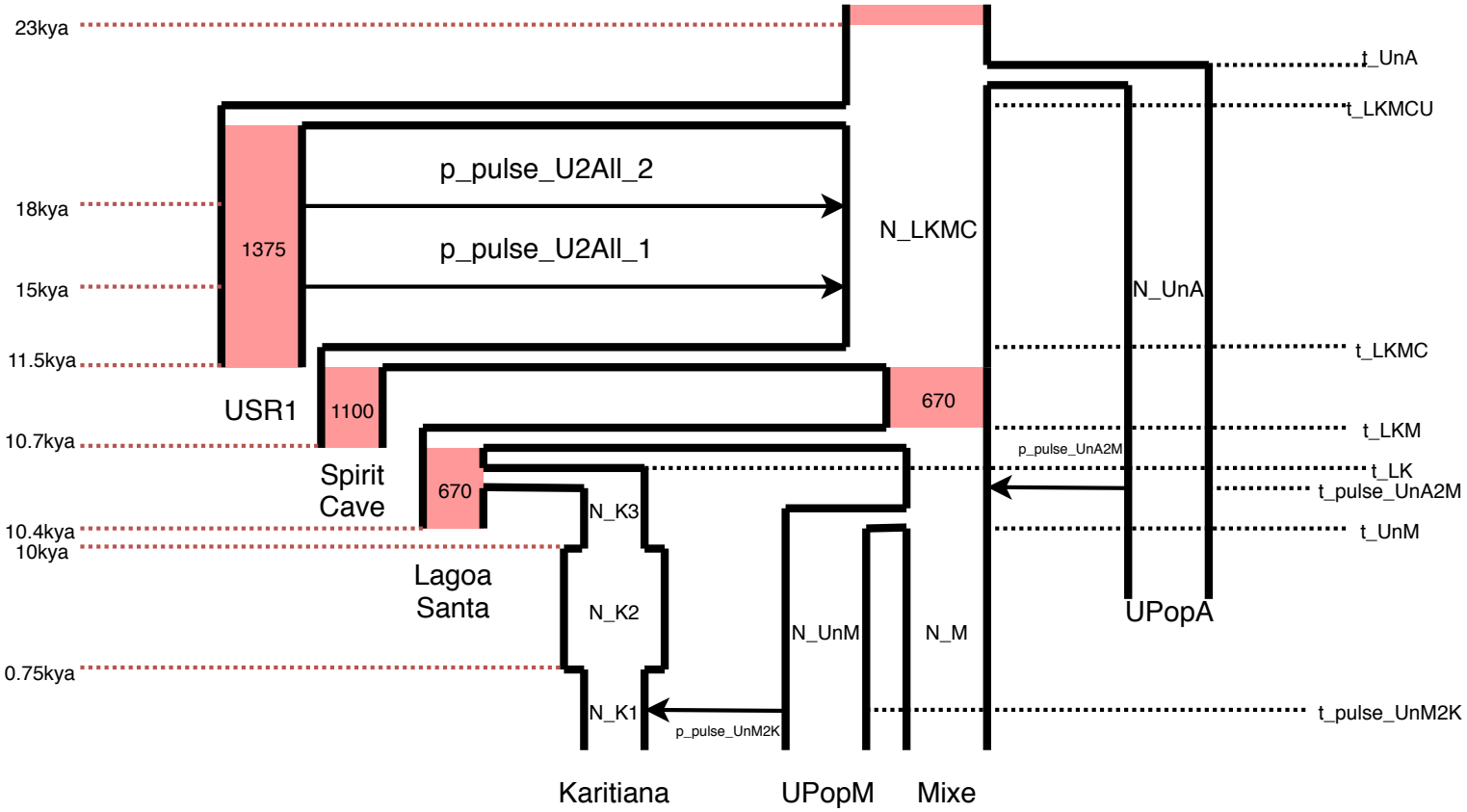


Figure S78. Diagram of parameters of model used for Model 3 in the *momi2* analysis. Red lines on the left hand size and shaded regions indicate times and population sizes, respectively, that were fixed and not inferred.

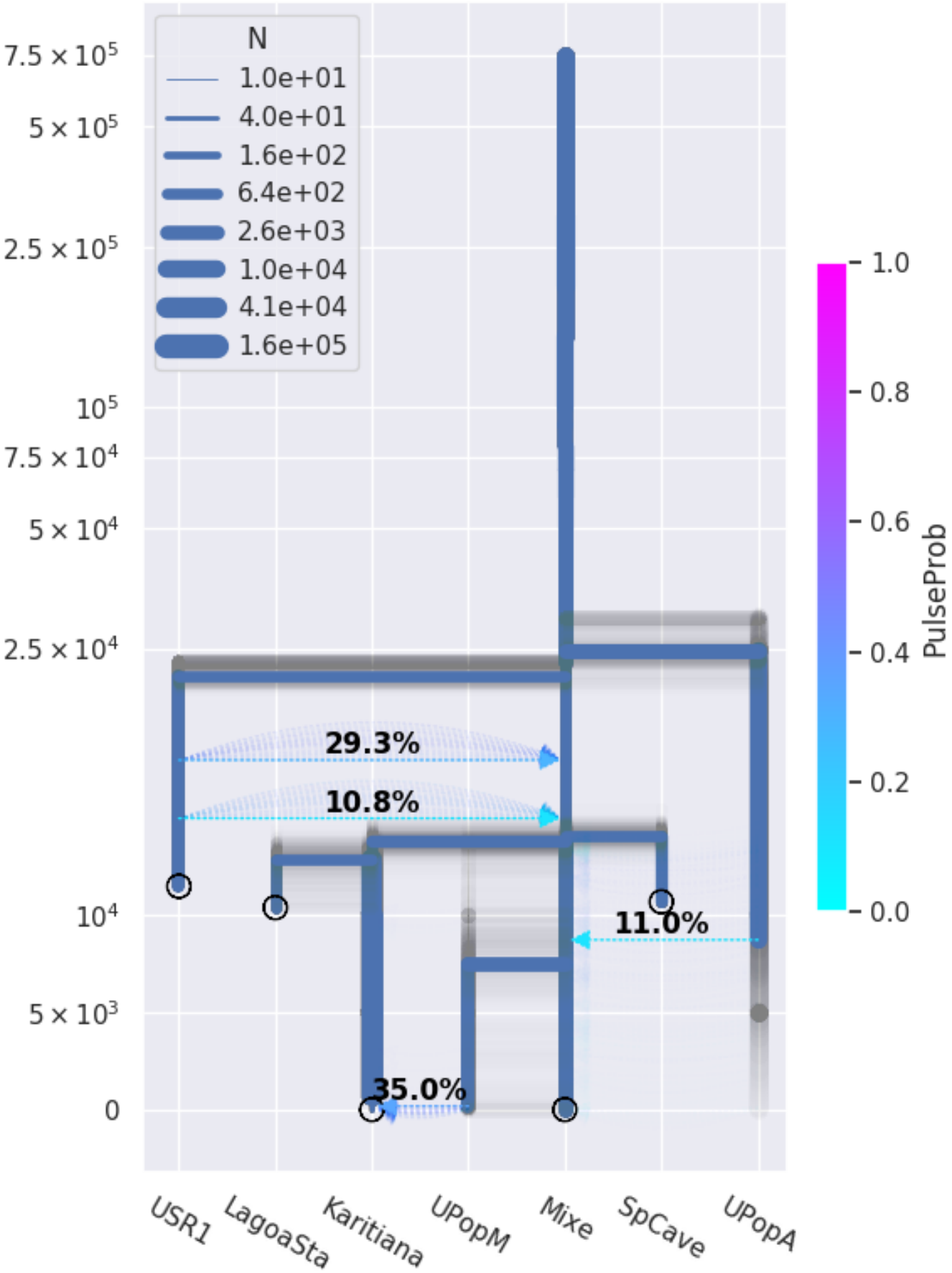


Figure S79. Nonparametric bootstrap for the demography inferred in Fig. S78C. We ran 228 bootstrap iterations. Each replicate produces a new demography plotted in gray.

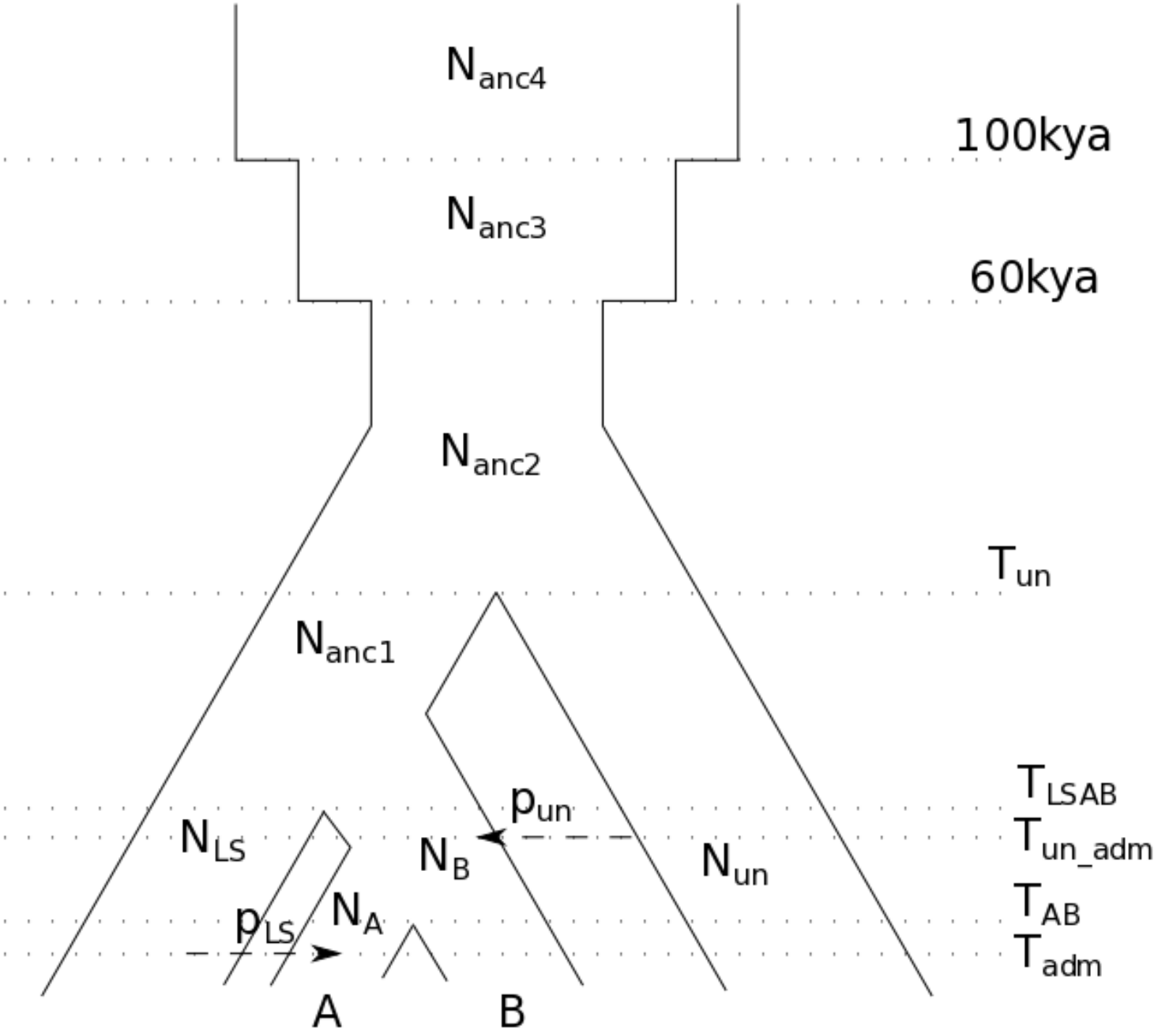


Figure S80. Model used for diCal analysis.

Table S1. **Radiocarbon ages and isotope data on the Trail Creek Cave human remains and select artifacts.** Calibrations done using Calib7.1.

Specimen number (Lab no.)	Material dated	Radiocarbon age BP	IntCal13 median age	IntCal13 age range (2 sigma)	Comment
OxA-37281	Human tooth	8085 ± 40	9020	8782-8832 (5.7%) 8862-8889 (2.5%) 8891-8918 (2.1%) 8957-8961 (0.2%) 8969-9130 (89.5%)	Layer III, 4 m section

Table S2. **Radiocarbon ages on the Big Bar Lake individual (from** (*80*)**).**
Calibrations done using Calib7.1.

Specimen number (Lab no.)	Radiocarbon age	IntCal13 median age	IntCal13 age range (2 sigma)	$\delta^{13}C$	Comments
CMC-1635 (Beta-194927)	5110 ± 40	5,823	5746-5834 (54.6%) 5838-5929 (45.4%)	-17.5‰ -17.7‰	fragments from three ribs
CMC-1666 (Beta-202619)	4690 ± 40	5,406	5317-5481 (84.9%) 5531-5578 (15.1%)	-18.5‰	left fibula
Composite of CMC-1635 & CMC-1666	4940 ± 40	5,664	5596-5742 (100%)		Re-runs from other two samples

Table S3. **Radiocarbon ages and isotope data on the Spirit Cave individuals.**
Calibrations done using Calib7.1.

Specimen number (Lab no.)	Age, sampled element	Radiocarbon age	IntCal13 median age	IntCal13 age range (2 sigma)	$\delta^{13}C$
Spirit Cave mummy (Ahur 2064; OxA-36200)	Adult male, [petrous]	9615 ± 50	10,946	10,770-11,170	-15.27
Spirit Cave 2 (Ahur 770; OxA-36201)	Adult female, [tooth]	9645 ± 65	10,981	10,772-11,198	-15.35
Spirit Cave 3 (Ahur 748; OxA-36202)	sub-Adult male, [innominate]	4682 ± 35	5400	5317-5475 (91.5%) 5546-5575 (8.5%)	-18.26

Table S4. Radiocarbon ages on the Lovelock Cave individuals. Calibrations done using Calib7.1.

Specimen number (Lab no.)	**Age, sampled element**	**Radiocarbon age**	**IntCal13 median age**	**IntCal13 age range (2 sigma)**	**$\delta^{13}C$**
Lovelock 1 (AAR-24197)	Adult, teeth & petrous	2001 ± 28	1950 (0 AD)	1884-2001 (100%)	-16.99
Lovelock 2 (AAR-24198)	Adult, molar & petrous	1925 ± 29	1874 (76 AD)	1818-1942 (100%)	-16.03
Lovelock 3 (AAR-24199)	Adult, teeth	703 ± 26	667 (1283 AD)	567-584 (10.6%) 648-687 (89.4%)	-14.87
Lovelock 4 (AAR-24200)	Adult, petrous	1836 ± 28	1,773 (177 AD)	1706-1830 (96%) 1842-1863 (4%)	-15.96
Lovelock 5 (AAR-24201)	Adult, teeth	588 ± 27	605 (1345 AD)	538-569 (28.9%) 582-650 (71.1%)	-17.59

Table S5. **Radiocarbon ages on the Lagoa Santa individuals**. All samples on petrous bones. The SH- numbers were assigned to the original 17 measurable skulls by Hansen in 1888 (*38*, *99*), and as that number series is used in other publications, e.g., (*5*) we denote it here (Specimen number). Calibrations done using Calib7.1.

Specimen number(Lab no.)	**Radiocarbon age**	**IntCal13 median age**	**IntCal13 age range (2 sigma)**	**$\delta^{13}C$**	**Seq. cover.**
ZMK1/1845:2352 (OxA-X-2677-49)	8785 ± 50	9806	9564-9571 (0.03%) 9594-9950 (90.2%) 9990-10,012 (1.3%) 10,025-10,038 (0.07%) 10,061-10,133 (7.5%)	-18.1‰	0.18X
SH-09 ZMK1/1845:15109 (OxA-X-2663-56)	9070 ± 45	10,229	10,173-10,296 (99.1%) 10,357-10,369 (0.9%)	-18.7‰	0.48X
SH-16 ZMK1/1845:15108 (OxA-33336)	9015 ± 45	10,202	9942-9990 (5.5%) 10,012-10,024 (0.08%) 10,039-10,060 (2.0%) 10,136-10,249 (91.7%)	-19.8‰	0.84X
SH-10 ZMK1/1845:2350 (OxA-33337)	8990 ± 50	10,175	9,924-10,067 (30.6%) 10,118-10,241 (69.4%)	-18.9‰	1.98X
SH-04 ZMK1/1845:15113 (OxA-X-33335)	9240 ± 50	10,408	10,258-10,524 (97.0%) 10,531-10,552 (3.0%)	-19.4‰	15.5X

Table S6: Chronometric data for the four samples dated at the ORAU. P-code refers to pretreatment code; '**AF***' is ultrafiltered collagen; '**HYP**' denotes the extraction of hydroxyproline from hydrolysed bone collagen (*113*, *114*); '**OO'** is carbon dioxide evolved from acid etched tooth enamel digested in orthophosphoric acid. This date should be considered as a minimum age since demonstrating complete decontamination on this material is challenging and it is low in carbon. %Yield is the percent yield of extracted collagen as a function of the starting weight of the bone analysed. When the HYP procedure was performed on previously extracted collagen, the yields are not reported. %C is the carbon present in the combusted sample (gelatin or hydroxyproline). C/N is the atomic ratio of carbon to nitrogen. Stable isotope ratios are expressed in per mil (‰) relative to vPDB with a mass spectrometric precision of ±0.2‰ (*213*). CRA is conventional radiocarbon age, expressed in years BP (*214*).

Sample no.	**P-code**	**% Yield**	**%C**	**C:N**	**δ^{13}C ‰**	**δ^{15}N ‰**	**CRA (BP)**	**±**	**OxA number**
P43584 CGG2_17374 (Spirit Cave)	AF*	8.6	44.1	3.2	-15.3	16.3	9615	50	OxA-36200
	HYP	/	39.4	5.0	-20.1	18.8	9600	55	OxA-X-2772-8
P43585 CGG2_17378 Spirit Cave)	AF*	8.6	46.0	3.5	-15.3	15.9	9645	65	OxA-36201
	HYP	/	43.3	5.1	-21.3	18.3	9540	50	OxA-X-2767-11
P43586 CGG2_17379 (Spirit Cave)	AF*	9.1	47.2	3.5	-18.3	12.7	4682	35	OxA-36202
	HYP	14.3	38,1	5.1	-22.6	15.5	4750	36	OxA-X-2767-12
P44821 CGG_2_19883 Trail Creek Cave 2)	OO	-	-	-	-9.3	-	8085	40	OxA-37281

Table S7. **Sampled materials for Spirit Cave, Lovelock, Big Bar and Trail Creek individuals.**

ID	**Alias**	**CGG2 number**	**Material**
Spirit Cave Man	AHUR_2064	017374	Petrous
Spirit Cave Man	AHUR_2064	017375	Tooth
Spirit Cave Man	AHUR_2064	017376	Tooth
Spirit Cave Man	AHUR_2064	017377	Tooth
Spirit Cave	AHUR_770, (c) 1-20-8	017378	Teeth, bone
Spirit Cave	AHUR_748, 1-20-10	017379	Bone
Lovelock1		017380	Petrous bone, Teeth
Lovelock2		017381	Teeth

Lovelock3		017382	Teeth
Lovelock4		017383	Petrous bone
Lovelock5		017384	Teeth
BigBar	EhRk-4:1	19651	Lower right M3, (246, 202, 224,130) mg
Trail creek		19883	Incisor, (54, 36) mg

Table S8. **Mapping statistics and uniparental haplogroups for the ancient genomes presented in this study.** This table, together with a description of each one of its fields is presented as a separate .xlsx file.

Table S9. **Y chromosome and mtDNA description of new ancient sequences.** Full description of ancient samples sequenced in this study, including Y and mtDNA coverage and variants defining each sample haplogroup assignment. This table is presented as a separate .xlsx file.

Table S10. **Y chromosome and mtDNA description of literature samples.** Full description of modern and ancient samples included in this study, including references and geographical location. This table is presented as a separate .xlsx file.

Table S11. **List of SNPs defining the deep branches downstream of K-M9 macro-haplogroup.** Full description of all SNPS defining deep Eurasian branches downstream of K-M9, including SNPs previously with uncertain placement. Different names of SNPs found on ISOGG are listed (SNP_ID and SNP_syn), and SNP_eq denotes a more well-known marker with an equivalent phylogenetical signal. All novel SNPs were named MPBxxx and sent to ISOGG. This table is presented as a separate .xlsx file.

Table S12. Whole genome sequencing dataset. Details about the whole genome sequences used in this study are shown. For each individual, we enumerate the following features: **ID**, a generic name used to refer to each individual; Type, the type of calls considered for each individual, called genotypes (0) or random alleles (1); **NoMiss**, the number of non-missing autosomal calls after implementing the 1000 Genomes Project strict accessibility mask; **%Het**, the proportion of heterozygous calls out of 'NoMiss'; Population, the population grouping used throughout this study; **Category**, a term describing broad continental ancestry and sample type, present-day (PD) or ancient (Anc); and **Reference**. For the latter, we only include the first author and publication year of each study, but we include full citations in the main and supplementary references.

ID	Type	NoMiss	%Het	Population	Category	Reference	ID	CallType	NoMiss	%Het	Population	Category	Reference
GB20	0	1913150724	0.06587	Mota	AfricaAnc	(203)	CAI6	Type	2052921311	0.0577	Cairns	OceaniaPd	(127)
DNK02	0	2055020261	0.07989	Dinka	AfricaPd	(126)	CAI7	0	2052669952	0.06055	Cairns	OceaniaPd	(127)
SS6004480	0	2057976911	0.08144	Dinka	AfricaPd	(202)	CAI8	0	2050408330	0.06231	Cairns	OceaniaPd	(127)
HGDP01284	0	2053674818	0.08001	Mandenka	AfricaPd	(126)	CAI9	0	2054783450	0.06439	Cairns	OceaniaPd	(127)
SS6004470	0	2058404826	0.08418	Mandenka	AfricaPd	(202)	EN1	0	2055427631	0.06486	EspNyun	OceaniaPd	(127)
HGDP00456	0	2047853174	0.08038	Mbuti	AfricaPd	(126)	EN2	0	2052177124	0.05483	EspNyun	OceaniaPd	(127)
SS6004471	0	2049805565	0.08548	Mbuti	AfricaPd	(202)	EN3	0	2050426307	0.06351	EspNyun	OceaniaPd	(127)
HGDP01029	0	2057858573	0.08658	San	AfricaPd	(126)	EN4	0	2051416012	0.06188	EspNyun	OceaniaPd	(127)
SS6004473	0	2058197550	0.0867	San	AfricaPd	(202)	EN5	0	2053336608	0.06793	EspNyun	OceaniaPd	(127)
HGDP00927	0	2057293909	0.08177	Yoruba	AfricaPd	(126)	EN6	0	2052554522	0.0631	EspNyun	OceaniaPd	(127)
SS6004475	0	2059250096	0.08336	Yoruba	AfricaPd	(202)	EN7	0	2052409099	0.06472	EspNyun	OceaniaPd	(127)
302	1	829104639	NA	302	AmericaAnc	(9)	EN8	0	2052754302	0.06246	EspNyun	OceaniaPd	(127)
443	1	1237395096	NA	443	AmericaAnc	(9)	NG002	0	2052199150	0.05521	Ngadju	OceaniaPd	(127)
939	1	684958835	NA	939	AmericaAnc	(2)	NG005	0	2053318189	0.06114	Ngadju	OceaniaPd	(127)
AM71	1	171273234	NA	AncKaweskar	AmericaAnc	(2)	NG006	0	2052586401	0.06041	Ngadju	OceaniaPd	(127)
Nr66	1	39975471	NA	AncKaweskar	AmericaAnc	(2)	NG1	0	2053076219	0.06427	Ngadju	OceaniaPd	(127)
Nr72	1	6419954	NA	AncKaweskar	AmericaAnc	(2)	NG3	0	2052935880	0.06253	Ngadju	OceaniaPd	(127)
Nr73	1	29836405	NA	AncKaweskar	AmericaAnc	(2)	NG4	0	2052196051	0.06261	Ngadju	OceaniaPd	(127)
Nr74	1	642177670	NA	AncKaweskar	AmericaAnc	(2)	13733_8	0	2045103695	0.03972	Papuan	OceaniaPd	(2)
MA572	1	15608969	NA	AncSelknam	AmericaAnc	(2)	13748_1	0	2040253509	0.0442	Papuan	OceaniaPd	(2)
MA575	1	5364767	NA	AncSelknam	AmericaAnc	(2)	13748_3	0	2040955005	0.04445	Papuan	OceaniaPd	(2)
MA577	1	1407925587	NA	AncSelknam	AmericaAnc	(2)	13748_5	0	2038103473	0.04489	Papuan	OceaniaPd	(2)
890	1	371243881	NA	AncYaghan	AmericaAnc	(2)	13748_6	0	2040188245	0.04362	Papuan	OceaniaPd	(2)
894	1	864129086	NA	AncYaghan	AmericaAnc	(2)	13748_7	0	2043335262	0.04433	Papuan	OceaniaPd	(2)
895	1	924204792	NA	AncYaghan	AmericaAnc	(2)	13748_8	0	2047026757	0.04375	Papuan	OceaniaPd	(2)
Clovis	0	1173069952	0.05603	Anzick1	AmericaAnc	(3)	13784_1	0	1997951139	0.04033	Papuan	OceaniaPd	(2)
A460	0	1922203635	0.03829	Ayayema	AmericaAnc	This study	13784_2	0	2043003271	0.04391	Papuan	OceaniaPd	(2)
EhRk41	1	1418327952	NA	BigBar	AmericaAnc	This study	13784_3	0	2039172986	0.04404	Papuan	OceaniaPd	(2)
Kennewick	1	1182458371	NA	Kennewick	AmericaAnc	(10)	13784_4	0	2020572636	0.04101	Papuan	OceaniaPd	(2)
Sumidouro5	0	1841613279	0.05265	LagoaSta	AmericaAnc	This study	13784_5	0	2017276512	0.03815	Papuan	OceaniaPd	(2)
Sumidouro4	1	1145477128	NA	LagoaStaLow	AmericaAnc	This study	HGDP00542	0	2050410340	0.04793	Papuan	OceaniaPd	(126)
Sumidouro6	1	1732584813	NA	LagoaStaLow	AmericaAnc	This study	SS6004472	0	2057768368	0.04854	Papuan	OceaniaPd	(202)
Sumidouro7	1	777123119	NA	LagoaStaLow	AmericaAnc	This study	PIL1	0	2052710519	0.04986	Pilbara	OceaniaPd	(127)

Sumidouro8	1	329138696	NA	LagoaStaLow	AmericaAnc	This study	PIL10	0	2053094364	0.05723	Pilbara	OceaniaPd	(127)
Lovelock2	0	2048854006	0.0463	Lovelock2	AmericaAnc	This study	PIL11	0	2052878185	0.06305	Pilbara	OceaniaPd	(127)
Lovelock3	0	2052618322	0.04833	Lovelock3	AmericaAnc	This study	PIL12	0	2052329830	0.06338	Pilbara	OceaniaPd	(127)
Lovelock1	1	1568835797	NA	LovelockLow	AmericaAnc	This study	PIL2	0	2053847489	0.06207	Pilbara	OceaniaPd	(127)
Lovelock4	1	847369339	NA	LovelockLow	AmericaAnc	This study	PIL3	0	2054227441	0.06278	Pilbara	OceaniaPd	(127)
BC23	1	14754972	NA	Pericu	AmericaAnc	(2)	PIL4	0	2054916315	0.05754	Pilbara	OceaniaPd	(127)
BC25	1	255840760	NA	Pericu	AmericaAnc	(2)	PIL5	0	2051340794	0.06157	Pilbara	OceaniaPd	(127)
BC27	1	14997137	NA	Pericu	AmericaAnc	(2)	PIL6	0	2054499265	0.05435	Pilbara	OceaniaPd	(127)
BC28	1	13836283	NA	Pericu	AmericaAnc	(2)	PIL7	0	2052117598	0.05343	Pilbara	OceaniaPd	(127)
BC29	1	76196041	NA	Pericu	AmericaAnc	(2)	PIL8	0	2052410917	0.06347	Pilbara	OceaniaPd	(127)
BC30	1	183683800	NA	Pericu	AmericaAnc	(2)	PIL9	0	2051877316	0.06263	Pilbara	OceaniaPd	(127)
5832	1	1529204681	NA	PtaStaAna	AmericaAnc	This study	RIV1	0	2056478906	0.06485	Riverine	OceaniaPd	(127)
Saqqaq	0	624920604	0.03663	Saqqaq	AmericaAnc	(144)	RIV2	0	2059890639	0.06414	Riverine	OceaniaPd	(127)
AHUR_2064	0	1988253967	0.04672	SpCave	AmericaAnc	This study	RIV3	0	2058811948	0.06524	Riverine	OceaniaPd	(127)
SpCave2	1	213575751	NA	SpCave2	AmericaAnc	This study	RIV4	0	2058057594	0.06822	Riverine	OceaniaPd	(127)
PC537	0	2002711863	0.04386	Taino	AmericaAnc	(44)	RIV5	0	2059120422	0.05761	Riverine	OceaniaPd	(127)
TrailCreek	1	684654457	NA	TrailCreek	AmericaAnc	This study	RIV6	0	2057938067	0.06385	Riverine	OceaniaPd	(127)
USR1	0	2025883785	0.04895	USR1	AmericaAnc	(1)	RIV7	0	2058861137	0.06336	Riverine	OceaniaPd	(127)
Aleutian_2_masked	0	1355518512	0.05877	Aleutian	AmericaPd	(2)	RIV8	0	2058452113	0.06555	Riverine	OceaniaPd	(127)
Athabascan_1	0	1622600857	0.05012	Athabascan	AmericaPd	(2)	WD1	0	2054470668	0.05203	WCDesert	OceaniaPd	(127)
Athabascan_2	0	1841824033	0.04931	Athabascan	AmericaPd	(2)	WD13a	0	2051890738	0.05256	WCDesert	OceaniaPd	(127)
TA6	0	1935028104	0.04027	Aymara	AmericaPd	(2)	WD2-1	0	2053107979	0.0534	WCDesert	OceaniaPd	(127)
LP6005519-DNA_D01	0	2055288172	0.05006	Chane	AmericaPd	(190)	WD3	0	2053854307	0.05182	WCDesert	OceaniaPd	(127)
HUI03	0	1939762402	0.0424	Huichol	AmericaPd	(2)	WD4-1	0	2052120362	0.05148	WCDesert	OceaniaPd	(127)
Greenlander_1	0	2061475551	0.05305	Inuit	AmericaPd	(2)	WD5	0	2052605395	0.0522	WCDesert	OceaniaPd	(127)
Greenlander_2	0	2060213915	0.05361	Inuit	AmericaPd	(2)	WD8	0	2053217716	0.05182	WCDesert	OceaniaPd	(127)
BI16	0	1986057261	0.0407	Karitiana	AmericaPd	(3)	WPA1	0	2051779033	0.06249	Weipa	OceaniaPd	(127)
HGDP00998	0	2050441887	0.0422	Karitiana	AmericaPd	(126)	WPA2	0	2053134742	0.05832	Weipa	OceaniaPd	(127)
LP6005441-DNA_G06	0	2057698713	0.04126	Karitiana	AmericaPd	(190)	WPA3	0	2051893276	0.06119	Weipa	OceaniaPd	(127)
LP6005441-DNA_H06	0	2057463097	0.04191	Karitiana	AmericaPd	(190)	WPA4	0	2051841481	0.06649	Weipa	OceaniaPd	(127)
SS6004476	0	2057156095	0.04325	Karitiana	AmericaPd	(202)	WPA5	0	2051797452	0.05585	Weipa	OceaniaPd	(127)
HGDP00877_masked	0	1719067560	0.0481	Maya	AmericaPd	(3)	WPA6	0	2053796017	0.06307	Weipa	OceaniaPd	(127)
LP6005441-DNA_G07	0	2056048819	0.04944	Maya	AmericaPd	(190)	WON1	0	2053523189	0.06199	Wongatha	OceaniaPd	(127)
LP6005441-DNA_H07	0	2057359488	0.04873	Maya	AmericaPd	(190)	WON10	0	2052269710	0.06078	Wongatha	OceaniaPd	(127)
LP6005443-DNA_E11	0	2056914218	0.04961	Mixe	AmericaPd	(6)	WON11	0	2052357237	0.06913	Wongatha	OceaniaPd	(127)
LP6005443-DNA_F11	0	2056684633	0.04883	Mixe	AmericaPd	(6)	WON2	0	2051826297	0.06827	Wongatha	OceaniaPd	(127)
SS6004479	0	2055570621	0.04929	Mixe	AmericaPd	(202)	WON3	0	2052399705	0.05174	Wongatha	OceaniaPd	(127)
LP6005443-DNA_G11	0	2057550395	0.0519	Mixtec	AmericaPd	(190)	WON4	0	2053853070	0.05968	Wongatha	OceaniaPd	(127)
LP6005443-DNA_H11	0	2057168328	0.05828	Mixtec	AmericaPd	(190)	WON5	0	2050058068	0.05813	Wongatha	OceaniaPd	(127)
LP6005441-DNA_A04	0	2054596804	0.04491	Piapoco	AmericaPd	(190)	WON6	0	2053517988	0.05453	Wongatha	OceaniaPd	(127)
LP6005441-DNA_B04	0	2055088998	0.04898	Piapoco	AmericaPd	(190)	WON7	0	2052803296	0.06024	Wongatha	OceaniaPd	(127)
CEPH_11_D12	0	1868099001	0.03442	Pima	AmericaPd	(2)	WON8	0	2052165778	0.06339	Wongatha	OceaniaPd	(127)
LP6005441-DNA_E10	0	2057673949	0.04859	Pima	AmericaPd	(190)	WON9a	0	2041334149	0.0589	Wongatha	OceaniaPd	(127)
LP6005441-DNA_F10	0	2056032841	0.04687	Pima	AmericaPd	(190)	NA04932	0	2049475160	0.06566	Maori	PolynesiaPd	(127)
LP6005519-DNA_G02	0	2055813297	0.05005	Quechua	AmericaPd	(190)	P2077	0	2049551653	0.05006	Rapanui	PolynesiaPd	(127)

LP6005677-DNA_E01	0	2058683109	0.05033	Quechua	AmericaPd	(190)	Andaman	0	1331424903	0.04603	HistAndaman	SEAsiaAnc	This study
LP6005677-DNA_F01	0	2056645318	0.05088	Quechua	AmericaPd	(190)	La368	1	847784856	NA	La368	SEAsiaAnc	(192)
LP6005441-DNA_A12	0	2056664098	0.04194	Surui	AmericaPd	(6)	Ma911	1	228188491	NA	Ma911	SEAsiaAnc	(192)
LP6005441-DNA_B12	0	2055695621	0.0402	Surui	AmericaPd	(6)	JAR-27	0	1933140912	0.04172	Jarawa	SEAsiaPd	(188)
LP6005443-DNA_D05	0	2054800370	0.05996	Tlingit	AmericaPd	(190)	JAR-32	0	2005339437	0.04244	Jarawa	SEAsiaPd	(188)
LP6005443-DNA_E05	0	2055211537	0.06291	Tlingit	AmericaPd	(190)	JAR-54	0	2001092340	0.04467	Jarawa	SEAsiaPd	(188)
Tsimshian_masked	0	1355536769	0.04767	Tsimshian	AmericaPd	(2)	JAR-61	0	2029633657	0.0443	Jarawa	SEAsiaPd	(188)
Y2040	0	1887987297	0.03975	Yukpa	AmericaPd	(2)	JHF05	0	2057541611	0.04719	Jehai	SEAsiaPd	(192)
Esk17_masked	0	934160391	0.04476	Yupik	AmericaPd	(2)	JHM06	0	2054703402	0.05417	Jehai	SEAsiaPd	(192)
Esk20_masked	0	1346191480	0.04862	Yupik	AmericaPd	(2)	ONG-1	0	1985433179	0.04344	Onge	SEAsiaPd	(188)
LP6005443-DNA_A12	0	2056731840	0.05178	Zapotec	AmericaPd	(190)	ONG-12	0	1951062019	0.04168	Onge	SEAsiaPd	(188)
LP6005677-DNA_D01	0	2056757113	0.05089	Zapotec	AmericaPd	(190)	ONG-14	0	1984619172	0.04313	Onge	SEAsiaPd	(188)
AltaiNea	0	2048409261	0.01503	AltaiNean	Archaic	(202)	ONG-4	0	2001527065	0.04437	Onge	SEAsiaPd	(188)
DenisovaPinky	0	2043104485	0.01677	Denisova	Archaic	(126)	ONG-8	0	1997699971	0.04309	Onge	SEAsiaPd	(188)
Avar	0	923711607	0.0494	Avar	CaucasusPd	(31)	ONG-9	0	2035880553	0.04891	Onge	SEAsiaPd	(188)
Kalmyk	0	2040359306	0.0531	Kalmyk	CaucasusPd	(199)	DA249	1	2039575528	NA	BaikalHG	SiberiaAnc	(138)
Kalmyks.1	0	2057528837	0.05607	Kalmyk2	CaucasusPd	(138)	NEO241	1	476265229	NA	Ekven_IA	SiberiaAnc	(189)
Kalmyks.2	0	2056910587	0.05471	Kalmyk2	CaucasusPd	(138)	NEO242	1	340442567	NA	Ekven_IA	SiberiaAnc	(189)
RISE505	0	1172663114	0.0461	Andronovo	CentralAsiaAnc	(138)	NEO243	1	105483901	NA	Ekven_IA	SiberiaAnc	(189)
BOT2016	0	2007667392	0.05417	Botai	CentralAsiaAnc	(138)	NEO246	1	553639695	NA	Ekven_IA	SiberiaAnc	(189)
EBA2	0	1852543157	0.05139	CSteppe_EMBA	CentralAsiaAnc	(138)	NEO247	1	1169171576	NA	Ekven_IA	SiberiaAnc	(189)
RISE493	0	817237790	0.05207	Karasuk	CentralAsiaAnc	(118)	NEO248	1	1120185680	NA	Ekven_IA	SiberiaAnc	(189)
RISE497	0	953563683	0.04234	Karasuk	CentralAsiaAnc	(118)	NEO249	1	1601344643	NA	Ekven_IA	SiberiaAnc	(189)
RISE509	0	545399486	0.03476	SteppeEMBA	CentralAsiaAnc	(138)	NEO250	1	500722252	NA	Ekven_IA	SiberiaAnc	(189)
RISE511	0	719648671	0.0412	SteppeEMBA	CentralAsiaAnc	(138)	NEO251	1	1312393673	NA	Ekven_IA	SiberiaAnc	(189)
Yamnaya	0	2020926790	0.05969	Yamnaya	CentralAsiaAnc	(138)	NEO253	1	145694437	NA	Ekven_IA	SiberiaAnc	(189)
Dungan.1	0	2058462156	0.05851	Dungan	CentralAsiaPd	(138)	NEO298	1	1730438364	NA	Ekven_IA	SiberiaAnc	(189)
Dungan.2	0	2058744381	0.05792	Dungan	CentralAsiaPd	(138)	NEO299	1	163973895	NA	Ekven_IA	SiberiaAnc	(189)
Karakalpaks.1	0	2055954972	0.05597	Karakalpak	CentralAsiaPd	(138)	Kolyma_River	0	1998302481	0.04739	Kolyma	SiberiaAnc	(189)
Karakalpaks.2	0	2058405042	0.06055	Karakalpak	CentralAsiaPd	(138)	MalTa	1	1332367737	NA	MalTa	SiberiaAnc	(31)
Karakalpaks.3	0	2057740702	0.05777	Karakalpak	CentralAsiaPd	(138)	RISE644	1	2027304568	NA	Okonevo	SiberiaAnc	(118)
Kazakhs.2	0	2057575130	0.05778	Kazakh	CentralAsiaPd	(138)	SII	0	627276384	0.02968	Sunghir	SiberiaAnc	(195)
Kazakhs.3	0	2057790787	0.05875	Kazakh	CentralAsiaPd	(138)	SIII	0	1937786376	0.04745	Sunghir	SiberiaAnc	(195)
Kazkahs.1	0	2057668193	0.05896	Kazakh	CentralAsiaPd	(138)	SIV	0	466558663	0.02728	Sunghir	SiberiaAnc	(195)
Kyrgyz.1	0	2057944356	0.05932	Kyrgyz	CentralAsiaPd	(138)	NEO233	1	6533311	NA	Uelen_IA	SiberiaAnc	(189)
Kyrgyz.2	0	1987525253	0.05756	Kyrgyz	CentralAsiaPd	(138)	NEO234	1	66923575	NA	Uelen_IA	SiberiaAnc	(189)
LP6005677-DNA_A02	0	2053142452	0.05992	Kyrgyz2	CentralAsiaPd	(190)	NEO230	1	823540692	NA	UstBelaya_EBA	SiberiaAnc	(189)
LP6005677-DNA_B02	0	2056232675	0.06144	Kyrgyz2	CentralAsiaPd	(190)	NEO231	1	799226197	NA	UstBelaya_EBA	SiberiaAnc	(189)
Mari	0	387378995	0.04426	Mari	CentralAsiaPd	(31)	NEO232	1	1364533893	NA	UstBelaya_EBA	SiberiaAnc	(189)
Tadjik	0	1635888424	0.05154	Tadjik	CentralAsiaPd	(31)	Ust_Ishim	0	2035347524	0.06303	UstIshim	SiberiaAnc	(152)
Tajiks.1	0	2057028329	0.05686	Tajik	CentralAsiaPd	(138)	Yana	0	2042608794	0.05501	Yana	SiberiaAnc	(189)
Tajiks.2	0	2057777739	0.0582	Tajik	CentralAsiaPd	(138)	Yana_old2	0	1446433870	0.0464	Yana2	SiberiaAnc	(189)
Tajiks.3	0	2057908665	0.05837	Tajik	CentralAsiaPd	(138)	YanaYng	1	1707802043	NA	YanaYng	SiberiaAnc	(189)
Turkmens.1	0	2058302496	0.06045	Turkmen	CentralAsiaPd	(138)	Alt1	0	1210165538	0.052	Altai	SiberiaPd	(2)
Turkmens.2	0	2058428305	0.06057	Turkmen	CentralAsiaPd	(138)	Alt2	0	2038026166	0.05778	Altai	SiberiaPd	(2)

Uzbeks.1	0	2057947559	0.05993	Uzbek	CentralAsiaPd	(138)	Altais.1	0	2056401421	0.05477	Altai2	SiberiaPd	(138)
Uzbeks.2	0	2056992215	0.05917	Uzbek	CentralAsiaPd	(138)	Altais.2	0	2055212141	0.0523	Altai2	SiberiaPd	(138)
Uzbeks.3	0	2058193047	0.06124	Uzbek	CentralAsiaPd	(138)	Altayan	0	2045375479	0.05602	Altai3	SiberiaPd	(199)
NEO236	1	1877738931	NA	DevilsG_N	EastAsiaAnc	(189)	LP6005442-DNA_F02	0	2056497672	0.05979	Altai4	SiberiaPd	(190)
NEO238	1	830685599	NA	DevilsG_N	EastAsiaAnc	(189)	Bur1	0	1927369444	0.0495	Buryat	SiberiaPd	(2)
NEO239	1	195278887	NA	DevilsG_N	EastAsiaAnc	(189)	Bur2	0	1931321643	0.05115	Buryat	SiberiaPd	(2)
NEO240	0	1376527222	0.03981	DevilsG_N	EastAsiaAnc	(189)	Buryats.1	0	2057824837	0.05487	Buryat2	SiberiaPd	(138)
Jomon	1	1609699538	NA	Jomon	EastAsiaAnc	(192)	Buryats.2	0	2057693667	0.05471	Buryat2	SiberiaPd	(138)
TianYuan	1	122351269	NA	Tianyuan	EastAsiaAnc	(191)	Buryat	0	2028218822	0.04973	Buryat3	SiberiaPd	(199)
HGDP01307	0	2053898690	0.05615	Dai	EastAsiaPd	(126)	LP6005443-DNA_D03	0	2055679713	0.05161	Chaplin	SiberiaPd	(190)
SS6004467	0	2058044579	0.05786	Dai	EastAsiaPd	(202)	LP6005443-DNA_C03	0	2055574216	0.06454	Chukchi	SiberiaPd	(190)
HGDP00778	0	2053891643	0.05637	Han	EastAsiaPd	(126)	Even	0	2040652334	0.04798	Even	SiberiaPd	(199)
SS6004469	0	2059180114	0.05757	Han	EastAsiaPd	(202)	LP6005443-DNA_B04	0	2054839836	0.05579	Even2	SiberiaPd	(190)
LP6005441-DNA_E08	0	2055371774	0.05916	Mongola	EastAsiaPd	(190)	LP6005443-DNA_C04	0	2054777293	0.05593	Even2	SiberiaPd	(190)
LP6005441-DNA_F08	0	2055359322	0.0577	Mongola	EastAsiaPd	(190)	LP6005592-DNA_F03	0	2056105713	0.058	Even2	SiberiaPd	(190)
LP6005442-DNA_G12	0	2056986989	0.0557	Ulchi	EastAsiaPd	(190)	Evenki	0	2038896496	0.04971	Evenki	SiberiaPd	(199)
LP6005442-DNA_H12	0	2056383801	0.05477	Ulchi	EastAsiaPd	(190)	IrtyshBarabinskTatars.1	0	2057249427	0.06037	IBTatars	SiberiaPd	(138)
Bichon	0	1327353334	0.03682	Bichon	EuropeAnc	(193)	IrtyshBarabinskTatars.2	0	2057170040	0.06012	IBTatars	SiberiaPd	(138)
8.4_BGI_WG	0	2036805942	0.06091	BR2	EuropeAnc	(194)	LP6005443-DNA_D04	0	2055072751	0.05176	Itelman	SiberiaPd	(190)
Karelian	0	1965594155	0.05146	Karelian	EuropeAnc	(199)	Ket1	0	1485001263	0.05611	Kets	SiberiaPd	(2)
KK1	0	1714633482	0.04231	KK1	EuropeAnc	(193)	Ket2	0	1880487167	0.04958	Kets	SiberiaPd	(2)
LBK	0	1959890785	0.05678	LBK	EuropeAnc	(202)	Khanty-1	0	2021264054	0.056	Khanty	SiberiaPd	(199)
Loschbour	0	1967301863	0.04552	Loschbour	EuropeAnc	(202)	Khanty-2	0	2018352467	0.04302	Khanty	SiberiaPd	(199)
10-3	0	2022212672	0.05912	NE1	EuropeAnc	(194)	Khanty-3	0	2048312720	0.05607	Khanty	SiberiaPd	(199)
Bashkirs.1	0	2058184919	0.0585	Bashkirs	EuropePd	(138)	Khanty-4	0	2049416228	0.05504	Khanty	SiberiaPd	(199)
Bashkirs.2	0	2058115593	0.05888	Bashkirs	EuropePd	(138)	Kor1	0	1893694913	0.04355	Koryaks	SiberiaPd	(2)
Bashkirs.3	0	2057846764	0.05925	Bashkirs	EuropePd	(138)	Kor2	0	1952020087	0.04315	Koryaks	SiberiaPd	(2)
Belarusian	0	1904421408	0.05205	Belarusian	EuropePd	(199)	Mansi_Child	0	1986005338	0.05586	Mansi	SiberiaPd	(199)
HGDP00521	0	2055515082	0.05948	French	EuropePd	(126)	Mansi_Father	0	2035540064	0.05607	Mansi	SiberiaPd	(199)
SS6004468	0	2057819126	0.06095	French	EuropePd	(202)	Mansi_Mother	0	2016528914	0.05733	Mansi	SiberiaPd	(199)
Komi_Izhma-1	0	2051905388	0.05528	KomiIzhma	EuropePd	(199)	LP6005443-DNA_F04	0	2055878310	0.05951	Mansi2	SiberiaPd	(190)
Komi_Izhma-2	0	2016536791	0.05516	KomiIzhma	EuropePd	(199)	LP6005443-DNA_G04	0	2055250344	0.06048	Mansi2	SiberiaPd	(190)
Komi_Obyachevo-1	0	2037237435	0.0543	KomiObyachevo	EuropePd	(199)	LP6005443-DNA_F03	0	2055803229	0.05305	Naukan	SiberiaPd	(190)
Komi_Obyachevo-2	0	2025296761	0.04808	KomiObyachevo	EuropePd	(199)	LP6005443-DNA_G03	0	2055910824	0.05153	Naukan	SiberiaPd	(190)
Russian_Andreapol-1	0	2016581015	0.05142	RussAndreapol	EuropePd	(199)	Nenets	0	2017507491	0.05472	Nenet	SiberiaPd	(199)
Russian_Andreapol-2	0	2053065899	0.05735	RussAndreapol	EuropePd	(199)	Nivh1	0	1919624025	0.04336	Nivkhs	SiberiaPd	(2)
Russian_Mezen-1	0	2039283620	0.05592	RussMezen	EuropePd	(199)	Nivh2	0	1965983090	0.04749	Nivkhs	SiberiaPd	(2)
Russian_Mezen-2	0	2036736458	0.04888	RussMezen	EuropePd	(199)	LP6005443-DNA_B03	0	2057363188	0.05131	Sireniki	SiberiaPd	(190)
Russian_Ustyuzhna-1	0	2040143975	0.05531	RussUstyuzhna	EuropePd	(199)	LP6005443-DNA_H03	0	2055976460	0.05331	Sireniki	SiberiaPd	(190)
Russian_Ustyuzhna-2	0	2047893266	0.05846	RussUstyuzhna	EuropePd	(199)	Teleuts.1	0	2057402803	0.05788	Teleut	SiberiaPd	(138)
LP6005592-DNA_C01	0	2055440066	0.05896	Saami	EuropePd	(190)	Teleuts.2	0	2057974337	0.05732	Teleut	SiberiaPd	(138)
LP6005592-DNA_D01	0	2054853760	0.05678	Saami	EuropePd	(190)	TomskTatars.1	0	2058462951	0.06133	TomskTatar	SiberiaPd	(138)
HGDP00665	0	2052104901	0.05789	Sardinian	EuropePd	(126)	TomskTatars.2	0	2058142393	0.05971	TomskTatar	SiberiaPd	(138)
SS6004474	0	2058705472	0.05836	Sardinian	EuropePd	(202)	LP6005442-DNA_E12	0	2056078116	0.05928	Tubalar	SiberiaPd	(190)
Veps-1	0	2048260009	0.05548	Veps	EuropePd	(199)	LP6005442-DNA_F12	0	2054725935	0.06042	Tubalar	SiberiaPd	(190)

Veps-2	0	2051753395	0.05741	Veps	EuropePd	(199)	Uyghurs.1	0	2057794358	0.05783	Uyghur	SiberiaPd	(138)
13748_4	0	2041878535	0.04964	AdmPapuan	OceaniaPd	(2)	Uyghurs.2	0	2057514575	0.05683	Uyghur	SiberiaPd	(138)
SS6004477	0	2055100997	0.05377	Australian	OceaniaPd	(202)	Uyghurs.3	0	2057811588	0.0592	Uyghur	SiberiaPd	(138)
SS6004478	0	2057501117	0.05428	Australian	OceaniaPd	(202)	VolgaTatars.1	0	2057618947	0.05798	VolgaTatar	SiberiaPd	(138)
BV1	0	2054085276	0.06811	Birdsville	OceaniaPd	(127)	VolgaTatars.2	0	2057662018	0.05837	VolgaTatar	SiberiaPd	(138)
BV10	0	2053962818	0.06435	Birdsville	OceaniaPd	(127)	Yakut	0	1951586284	0.04815	Yakut2	SiberiaPd	(199)
BV2	0	2055713865	0.0623	Birdsville	OceaniaPd	(127)	LP6005442-DNA_F01	0	2058090173	0.05847	Yakut3	SiberiaPd	(190)
BV4	0	2053632054	0.06127	Birdsville	OceaniaPd	(127)	LP6005443-DNA_D02	0	2054931573	0.05644	Yakut3	SiberiaPd	(190)
BV5	0	2054519586	0.06438	Birdsville	OceaniaPd	(127)	Yak1	0	1835271929	0.04744	Yakuts	SiberiaPd	(2)
BV6	0	2054536123	0.06997	Birdsville	OceaniaPd	(127)	Yak2	0	1852272255	0.0501	Yakuts	SiberiaPd	(2)
BV7	0	2054222164	0.06324	Birdsville	OceaniaPd	(127)	Hazaras.1	0	2057547952	0.05811	Hazara	SouthAsiaPd	(138)
BV8	0	2053746813	0.06398	Birdsville	OceaniaPd	(127)	Hazaras.2	0	2057977271	0.05912	Hazara	SouthAsiaPd	(138)
BV9	0	2053581272	0.06343	Birdsville	OceaniaPd	(127)	Hazaras.3	0	2058074297	0.06008	Hazara	SouthAsiaPd	(138)
CAI1	0	2053695678	0.055	Cairns	OceaniaPd	(127)	Manny	0	1660287445	0.04562	Indian	SouthAsiaPd	(31)
CAI10	0	2053236402	0.06067	Cairns	OceaniaPd	(127)	Kalash	0	1914793261	0.03617	Kalash	SouthAsiaPd	(200)
CAI2	0	2054062712	0.06459	Cairns	OceaniaPd	(127)	Bar8	0	1289869649	0.05048	Barcin_EN	WestAsiaAnc	(197)
CAI3	0	2050905124	0.06886	Cairns	OceaniaPd	(127)	Bon002	0	565192769	0.04674	Boncuklu_EN	WestAsiaAnc	(196)
CAI4	0	2051048767	0.06142	Cairns	OceaniaPd	(127)	WC1	0	1710721741	0.05286	Zagros	WestAsiaAnc	(198)
CAI5	0	2054336318	0.06231	Cairns	OceaniaPd	(127)							

Table S13: F4-statistics for all possible unrooted tree topologies containing Spirit Cave, Anzick, Lagoa Santa, and Mixe.

Unrooted Tree Topology	Observed f4 statistic	Estimated SD	Z-Score
((Spirit Cave, Anzick), (Lagoa Santa, Mixe))	3.32×10^{-4}	1.58×10^{-3}	0.21
((Spirit Cave, Lagoa Santa), (Anzick, Mixe))	1.40×10^{-3}	1.71×10^{-3}	0.817
((Spirit Cave, Mixe), (Lagoa Santa, Anzick))	1.06×10^{-3}	1.62×10^{-3}	0.658

Table S14: Least squares error of all possible pairs of ancestral placements using qpgraph.

Branch 1	Branch 2	Error
Spirit Cave	Anzick	1052.362
Spirit Cave	Lagoa Santa	0.059
Spirit Cave	Mixe	1052.089
Spirit Cave	Internal	1052.337
Anzick	Lagoa Santa	1052.091
Anzick	Mixe	0.07
Anzick	Internal	1052.26
Lagoa Santa	Mixe	1052.186
Lagoa Santa	Internal	1052.1
Mixe	Internal	1052.106

Table S15: F4 statistics to test placement of Karitiana population.

F4 Test	Z-Score
(Spirit Cave,Karitiana;Anzick,Lagoa Santa)	3.219
(Anzick, Karitiana; Spirit Cave, Lagoa Santa)	2.433
(Lagoa Santa, Karitiana; Anzick, Spirit Cave)	0.42
(Mixe, Karitiana; Anzick, Spirit Cave)	0.78
(Anzick, Karitiana; Mixe, Lagoa Santa)	0.895
(Spirit Cave, Karitiana; Mixe, Lagoa Santa)	0.998

Table S16: Final model (Model 3) inferred parameters.

Parameter	Estimate
N_M	3020
N_K1	40
N_K2	55900
N_K3	39100
N_LKMC	660
N_UnM	1680
N_UnA	9600
t_LKMCU	22300
t_LKMC	14100
t_LKM	13900
t_LK	12900
t_UnM	7500
t_pulse UnM2K	180
t_pulse UnA2M	8700
t_UnA	24700
p_pulse_UnM2K	0.35
p_pulse_UnA2M	0.11
p_pulse_U2All_1	0.11
p_pulse_U2All_2	0.29

Table S17: Estimates of the bias, standard deviation, and confidence intervals from nonparametric bootstrap for 228 iterations.

Parameter	Estimate	Bias	SD	2.50%	97.50%
N_M	3020	−104	371	2.44×10^3	3.81×10^3
N_K1	39.9	−13.9	27.5	10.4	106
N_K2	55900	-1.42×10^4	2.05×10^4	2.33×10^4	1.00×10^5
N_K3	39100	-4.44×10^3	8.38×10^4	2.14×10^4	5.00×10^4
N_LKMC	656	46.1	124	388	820
N_UnM	1680	-1.05×10^3	2.16×10^3	10.0	5.00×10^3
N_UnA	9600	-1.51×10^3	6.18×10^3	646	2.82×10^4
t_LKMCU	22300	−280	495	2.18×10^4	2.33×10^4
t_LKMC	14100	−33.0	453	1.32×10^4	1.49×10^4
t_LKM	13900	61.5	501	1.28×10^4	1.48×10^4
t_LK	12900	184	921	1.04×10^4	1.40×10^4
t_UnM	7490	−563	4.01×10^3	1.41	1.25×10^3
t_pulse UnM2K	176	−132	1.15×10^3	7.99×10^{-5}	1.39×10^4
t_pulse UnA2M	8740	433	4.28×10^3	36.9	1.39×10^4
t_UnA	24700	−865	2.68×10^3	2.20×10^4	3.0×10^4
p_pulse_UnM2K	0.35	0.0153	0.201	0.0529	0.71
p_pulse_UnA2M	0.11	0.0025	0.042	0.0445	0.198
p_pulse_U2All_1	0.108	−0.0401	0.0881	0.0067	0.335
p_pulse_U2All_2	0.293	0.0148	0.171	6.28×10^{-4}	0.5

Table S18: Parameter values fit using diCal.

PopA	PopB	T_{adm}	T_{AB}	T_{unadm}	T_{LSAB}	T_{un}	N_A	N_B	N_{LS}	N_{un}	N_{anc1}	N_{anc2}	N_{anc3}	N_{anc4}	p_{LS}	p_{un}
Aymara	Mixe	4.9ka	5.0ka	5.0ka	28.1ka	40.0ka	1,300	2,400	30	1,000	1,000	1,700	19,400	20,000	5%	84%
Karitiana	Mixe	0.6ka	4.7ka	9.8ka	18.0ka	34.8ka	600	2,000	200	1,000	4,400	2,700	9,300	22,900	51%	89%
Surui	Mixe	6.0ka	6.1ka	6.1ka	17.1ka	32.8ka	800	2,100	40	1,000	100,000	1,800	16,000	22,900	15%	83%

References and Notes

1. J. V. Moreno-Mayar, B. A. Potter, L. Vinner, M. Steinrücken, S. Rasmussen, J. Terhorst, J. A. Kamm, A. Albrechtsen, A.-S. Malaspinas, M. Sikora, J. D. Reuther, J. D. Irish, R. S. Malhi, L. Orlando, Y. S. Song, R. Nielsen, D. J. Meltzer, E. Willerslev, Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. *Nature* **553**, 203–207 (2018). doi:10.1038/nature25173 Medline

2. M. Raghavan, M. Steinrücken, K. Harris, S. Schiffels, S. Rasmussen, M. DeGiorgio, A. Albrechtsen, C. Valdiosera, M. C. Ávila-Arcos, A.-S. Malaspinas, A. Eriksson, I. Moltke, M. Metspalu, J. R. Homburger, J. Wall, O. E. Cornejo, J. V. Moreno-Mayar, T. S. Korneliussen, T. Pierre, M. Rasmussen, P. F. Campos, P. de Barros Damgaard, M. E. Allentoft, J. Lindo, E. Metspalu, R. Rodríguez-Varela, J. Mansilla, C. Henrickson, A. Seguin-Orlando, H. Malmström, T. Stafford Jr., S. S. Shringarpure, A. Moreno-Estrada, M. Karmin, K. Tambets, A. Bergström, Y. Xue, V. Warmuth, A. D. Friend, J. Singarayer, P. Valdes, F. Balloux, I. Leboreiro, J. L. Vera, H. Rangel-Villalobos, D. Pettener, D. Luiselli, L. G. Davis, E. Heyer, C. P. E. Zollikofer, M. S. Ponce de León, C. I. Smith, V. Grimes, K.-A. Pike, M. Deal, B. T. Fuller, B. Arriaza, V. Standen, M. F. Luz, F. Ricaut, N. Guidon, L. Osipova, M. I. Voevoda, O. L. Posukh, O. Balanovsky, M. Lavryashina, Y. Bogunov, E. Khusnutdinova, M. Gubina, E. Balanovska, S. Fedorova, S. Litvinov, B. Malyarchuk, M. Derenko, M. J. Mosher, D. Archer, J. Cybulski, B. Petzelt, J. Mitchell, R. Worl, P. J. Norman, P. Parham, B. M. Kemp, T. Kivisild, C. Tyler-Smith, M. S. Sandhu, M. Crawford, R. Villems, D. G. Smith, M. R. Waters, T. Goebel, J. R. Johnson, R. S. Malhi, M. Jakobsson, D. J. Meltzer, A. Manica, R. Durbin, C. D. Bustamante, Y. S. Song, R. Nielsen, E. Willerslev, Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* **349**, aab3884–aab3884 (2015). doi:10.1126/science.aab3884 Medline

3. M. Rasmussen, S. L. Anzick, M. R. Waters, P. Skoglund, M. DeGiorgio, T. W. Stafford Jr., S. Rasmussen, I. Moltke, A. Albrechtsen, S. M. Doyle, G. D. Poznik, V. Gudmundsdottir, R. Yadav, A.-S. Malaspinas, S. S. White 5th, M. E. Allentoft, O. E. Cornejo, K. Tambets, A. Eriksson, P. D. Heintzman, M. Karmin, T. S. Korneliussen, D. J. Meltzer, T. L. Pierre, J. Stenderup, L. Saag, V. M. Warmuth, M. C. Lopes, R. S. Malhi, S. Brunak, T. Sicheritz-Ponten, I. Barnes, M. Collins, L. Orlando, F. Balloux, A. Manica, R. Gupta, M. Metspalu, C. D. Bustamante, M. Jakobsson, R. Nielsen, E. Willerslev, The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* **506**, 225–229 (2014). doi:10.1038/nature13025 Medline

4. D. Reich, N. Patterson, D. Campbell, A. Tandon, S. Mazieres, N. Ray, M. V. Parra, W. Rojas, C. Duque, N. Mesa, L. F. García, O. Triana, S. Blair, A. Maestre, J. C. Dib, C. M. Bravi, G. Bailliet, D. Corach, T. Hünemeier, M. C. Bortolini, F. M. Salzano, M. L. Petzl-Erler, V. Acuña-Alonzo, C. Aguilar-Salinas, S. Canizales-Quinteros, T. Tusié-Luna, L. Riba, M. Rodríguez-Cruz, M. Lopez-Alarcón, R. Coral-Vazquez, T. Canto-Cetina, I. Silva-Zolezzi, J. C. Fernandez-Lopez, A. V. Contreras, G. Jimenez-Sanchez, M. J. Gómez-Vázquez, J. Molina, A. Carracedo, A. Salas, C. Gallo, G. Poletti, D. B. Witonsky, G. Alkorta-Aranburu, R. I. Sukernik, L. Osipova, S. A. Fedorova, R. Vasquez, M. Villena, C. Moreau, R. Barrantes, D. Pauls, L. Excoffier, G. Bedoya, F. Rothhammer, J.-

M. Dugoujon, G. Larrouy, W. Klitz, D. Labuda, J. Kidd, K. Kidd, A. Di Rienzo, N. B. Freimer, A. L. Price, A. Ruiz-Linares, Reconstructing Native American population history. *Nature* **488**, 370–374 (2012). doi:10.1038/nature11258 Medline

5. W. A. Neves, M. Hubbe, Cranial morphology of early Americans from Lagoa Santa, Brazil: Implications for the settlement of the New World. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 18309–18314 (2005). doi:10.1073/pnas.0507185102 Medline

6. P. Skoglund, S. Mallick, M. C. Bortolini, N. Chennagiri, T. Hünemeier, M. L. Petzl-Erler, F. M. Salzano, N. Patterson, D. Reich, Genetic evidence for two founding populations of the Americas. *Nature* **525**, 104–108 (2015). Medline

7. C. L. Scheib, H. Li, T. Desai, V. Link, C. Kendall, G. Dewar, P. W. Griffith, A. Mörseburg, J. R. Johnson, A. Potter, S. L. Kerr, P. Endicott, J. Lindo, M. Haber, Y. Xue, C. Tyler-Smith, M. S. Sandhu, J. G. Lorenz, T. D. Randall, Z. Faltyskova, L. Pagani, P. Danecek, T. C. O'Connell, P. Martz, A. S. Boraas, B. F. Byrd, A. Leventhal, R. Cambra, R. Williamson, L. Lesage, B. Holguin, E. Ygnacio-De Soto, J. Rosas, M. Metspalu, J. T. Stock, A. Manica, A. Scally, D. Wegmann, R. S. Malhi, T. Kivisild, Ancient human parallel lineages within North America contributed to a coastal expansion. *Science* **360**, 1024–1027 (2018). doi:10.1126/science.aar6851 Medline

8. B. Llamas, L. Fehren-Schmitz, G. Valverde, J. Soubrier, S. Mallick, N. Rohland, S. Nordenfelt, C. Valdiosera, S. M. Richards, A. Rohrlach, M. I. B. Romero, I. F. Espinoza, E. T. Cagigao, L. W. Jiménez, K. Makowski, I. S. L. Reyna, J. M. Lory, J. A. B. Torrez, M. A. Rivera, R. L. Burger, M. C. Ceruti, J. Reinhard, R. S. Wells, G. Politis, C. M. Santoro, V. G. Standen, C. Smith, D. Reich, S. Y. W. Ho, A. Cooper, W. Haak, Ancient mitochondrial DNA provides high-resolution time scale of the peopling of the Americas. *Sci. Adv.* **2**, e1501385–e1501385 (2016). doi:10.1126/sciadv.1501385 Medline

9. J. Lindo, A. Achilli, U. A. Perego, D. Archer, C. Valdiosera, B. Petzelt, J. Mitchell, R. Worl, E. J. Dixon, T. E. Fifield, M. Rasmussen, E. Willerslev, J. S. Cybulski, B. M. Kemp, M. DeGiorgio, R. S. Malhi, Ancient individuals from the North American Northwest Coast reveal 10,000 years of regional genetic continuity. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 4093–4098 (2017). doi:10.1073/pnas.1620410114 Medline

10. M. Rasmussen, M. Sikora, A. Albrechtsen, T. S. Korneliussen, J. V. Moreno-Mayar, G. D. Poznik, C. P. E. Zollikofer, M. P. de León, M. E. Allentoft, I. Moltke, H. Jónsson, C. Valdiosera, R. S. Malhi, L. Orlando, C. D. Bustamante, T. W. Stafford Jr., D. J. Meltzer, R. Nielsen, E. Willerslev, The ancestry and affiliations of Kennewick Man. *Nature* **523**, 455–458 (2015). Medline

11. J. Bardill, A. C. Bader, N. A. Garrison, D. A. Bolnick, J. A. Raff, A. Walker, R. S. Malhi; Summer internship for INdigenous peoples in Genomics (SING) Consortium, Advancing the ethics of paleogenomics. *Science* **360**, 384–385 (2018). doi:10.1126/science.aaq1131 Medline

12. K. G. Claw, M. Z. Anderson, R. L. Begay, K. S. Tsosie, K. Fox, N. A. Garrison; Summer internship for INdigenous peoples in Genomics (SING) Consortium, A framework for enhancing ethical genomic research with Indigenous communities. *Nat. Commun.* **9**, 2957 (2018). doi:10.1038/s41467-018-05188-3 Medline

13. See supplementary materials.

14. A. Gómez-Carballa, L. Catelli, J. Pardo-Seco, F. Martinón-Torres, L. Roewer, C. Vullo, A. Salas, The complete mitogenome of a 500-year-old Inca child mummy. *Sci. Rep.* **5**, 16462 (2015). doi:10.1038/srep16462 Medline

15. P. J. Reimer, E. Bard, A. Bayliss, J. W. Beck, P. G. Blackwell, C. B. Ramsey, C. E. Buck, H. Cheng, R. L. Edwards, M. Friedrich, P. M. Grootes, T. P. Guilderson, H. Haflidason, I. Hajdas, C. Hatté, T. J. Heaton, D. L. Hoffmann, A. G. Hogg, K. A. Hughen, K. F. Kaiser, B. Kromer, S. W. Manning, M. Niu, R. W. Reimer, D. A. Richards, E. M. Scott, J. R. Southon, R. A. Staff, C. S. M. Turney, J. van der Plicht, IntCal13 and Marine13 Radiocarbon Age Calibration Curves 0–50,000 Years cal BP. *Radiocarbon* **55**, 1869–1887 (2013). doi:10.2458/azu_js_rc.55.16947

16. P. Skoglund, D. Reich, A genomic view of the peopling of the Americas. *Curr. Opin. Genet. Dev.* **41**, 27–35 (2016). doi:10.1016/j.gde.2016.06.016 Medline

17. A. W. Briggs, U. Stenzel, M. Meyer, J. Krause, M. Kircher, S. Pääbo, Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res.* **38**, e87–e87 (2010). doi:10.1093/nar/gkp1163 Medline

18. P. Verdu, T. J. Pemberton, R. Laurent, B. M. Kemp, A. Gonzalez-Oliver, C. Gorodezky, C. E. Hughes, M. R. Shattuck, B. Petzelt, J. Mitchell, H. Harry, T. William, R. Worl, J. S. Cybulski, N. A. Rosenberg, R. S. Malhi, Patterns of admixture and population structure in native populations of Northwest North America. *PLOS Genet.* **10**, e1004530 (2014). doi:10.1371/journal.pgen.1004530 Medline

19. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009). doi:10.1101/gr.094052.109 Medline

20. A.-S. Malaspinas, O. Tange, J. V. Moreno-Mayar, M. Rasmussen, M. DeGiorgio, Y. Wang, C. E. Valdiosera, G. Politis, E. Willerslev, R. Nielsen, bammds: A tool for assessing the ancestry of low-depth whole-genome data using multidimensional scaling (MDS). *Bioinformatics* **30**, 2962–2964 (2014). doi:10.1093/bioinformatics/btu410 Medline

21. N. Patterson, P. Moorjani, Y. Luo, S. Mallick, N. Rohland, Y. Zhan, T. Genschoreck, T. Webster, D. Reich, Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012). doi:10.1534/genetics.112.145037 Medline

22. Q. Fu, C. Posth, M. Hajdinjak, M. Petr, S. Mallick, D. Fernandes, A. Furtwängler, W. Haak, M. Meyer, A. Mittnik, B. Nickel, A. Peltzer, N. Rohland, V. Slon, S. Talamo, I. Lazaridis, M. Lipson, I. Mathieson, S. Schiffels, P. Skoglund, A. P. Derevianko, N. Drozdov, V. Slavinsky, A. Tsybankov, R. G. Cremonesi, F. Mallegni, B. Gély, E. Vacca, M. R. G. Morales, L. G. Straus, C. Neugebauer-Maresch, M. Teschler-Nicola, S. Constantin, O. T. Moldovan, S. Benazzi, M. Peresani, D. Coppola, M. Lari, S. Ricci, A. Ronchitelli, F. Valentin, C. Thevenet, K. Wehrberger, D. Grigorescu, H. Rougier, I. Crevecoeur, D. Flas, P. Semal, M. A. Mannino, C. Cupillard, H. Bocherens, N. J. Conard, K. Harvati, V. Moiseyev, D. G. Drucker, J. Svoboda, M. P. Richards, D. Caramelli, R. Pinhasi, J. Kelso, N. Patterson, J. Krause, S. Pääbo, D. Reich, The genetic history of Ice Age Europe. *Nature* **534**, 200–205 (2016). doi:10.1038/nature17993 Medline

23. S. Soraggi, C. Wiuf, A. Albrechtsen, Powerful inference with the D-Statistic on low-coverage whole-genome data. *G3 (Bethesda)* **8**, 551–566 (2018). doi:10.1534/g3.117.300192 Medline

24. J. K. Pickrell, J. K. Pritchard, Inference of population splits and mixtures from genome-wide allele frequency data. *PLOS Genet.* **8**, e1002967 (2012). doi:10.1371/journal.pgen.1002967 Medline

25. J. A. Kamm, J. Terhorst, Y. S. Song, Efficient computation of the joint sample frequency spectra for multiple populations. *J. Comput. Graph. Stat.* **26**, 182–194 (2017). doi:10.1080/10618600.2016.1159212 Medline

26. J. A. Kamm, J. Terhorst, R. Durbin, Y. S. Song, Efficiently inferring the demographic history of many populations with allele count data. bioRxiv 287268 [Preprint]. 23 March 2018. https://doi.org/10.1101/287268.

27. J. Terhorst, J. A. Kamm, Y. S. Song, Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat. Genet.* **49**, 303–309 (2017). doi:10.1038/ng.3748 Medline

28. M. Steinrücken, J. A. Kamm, Y. S. Song, Inference of complex population histories using whole-genome sequences from multiple populations. bioRxiv 026591 [Preprint]. 16 September 2015. https://doi.org/10.1101/026591.

29. C. E. Holmes, in *From the Yenisei to the Yukon: Interpreting Lithic Assemblage Variability in Late Pleistocene/Early Holocene Beringia*, T. Goebel, I. Buvit, Eds. (Peopling of the Americas, Texas A&M Univ., ed. 1, 2011).

30. M. Lipson, D. Reich, A working model of the deep relationships of diverse modern human genetic lineages outside of Africa. *Mol. Biol. Evol.* **34**, 889–902 (2017). Medline

31. M. Raghavan, P. Skoglund, K. E. Graf, M. Metspalu, A. Albrechtsen, I. Moltke, S. Rasmussen, T. W. Stafford Jr., L. Orlando, E. Metspalu, M. Karmin, K. Tambets, S. Rootsi, R. Mägi, P. F. Campos, E. Balanovska, O. Balanovsky, E. Khusnutdinova, S. Litvinov, L. P. Osipova, S. A. Fedorova, M. I. Voevoda, M. DeGiorgio, T. Sicheritz-Ponten, S. Brunak, S. Demeshchenko, T. Kivisild, R. Villems, R. Nielsen, M. Jakobsson, E. Willerslev, Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87–91 (2014). doi:10.1038/nature12736 Medline

32. K. Leppälä, S. V. Nielsen, T. Mailund, admixturegraph: An R package for admixture graph manipulation and fitting. *Bioinformatics* **33**, 1738–1740 (2017). doi:10.1093/bioinformatics/btx048 Medline

33. P. Flegontov *et al.*, Na-Dene populations descend from the Paleo-Eskimo migration into America. bioRxiv 074476 [Preprint]. 13 September 2016. https://doi.org/10.1101/074476.

34. E. Tamm, T. Kivisild, M. Reidla, M. Metspalu, D. G. Smith, C. J. Mulligan, C. M. Bravi, O. Rickards, C. Martinez-Labarga, E. K. Khusnutdinova, S. A. Fedorova, M. V. Golubenko, V. A. Stepanov, M. A. Gubina, S. I. Zhadanov, L. P. Ossipova, L. Damba, M. I. Voevoda, J. E. Dipierri, R. Villems, R. S. Malhi, Beringian standstill and spread of Native American founders. *PLOS ONE* **2**, e829 (2007). doi:10.1371/journal.pone.0000829 Medline

35. M. C. Bortolini, R. González-José, S. L. Bonatto, F. R. Santos, Reconciling pre-Columbian settlement hypotheses requires integrative, multidisciplinary, and model-bound approaches. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E213–E214 (2014). doi:10.1073/pnas.1321197111 Medline

36. D. N. Harris, W. Song, A. C. Shetty, K. S. Levano, O. Cáceres, C. Padilla, V. Borda, D. Tarazona, O. Trujillo, C. Sanchez, M. D. Kessler, M. Galarza, S. Capristano, H. Montejo, P. O. Flores-Villanueva, E. Tarazona-Santos, T. D. O'Connor, H. Guio, Evolutionary genomic dynamics of Peruvians before, during, and after the Inca Empire. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E6526–E6535 (2018). doi:10.1073/pnas.1720798115 Medline

37. D. A. Bolnick, J. A. Raff, L. C. Springs, A. W. Reynolds, A. T. Miró-Herrans, Native American Genomics and Population Histories. *Annu. Rev. Anthropol.* **45**, 319–340 (2016). doi:10.1146/annurev-anthro-102215-100036

38. W. A. Neves, M. Hubbe, L. B. Piló, Early Holocene human skeletal remains from Sumidouro Cave, Lagoa Santa, Brazil: History of discoveries, geological and chronological context, and comparative cranial morphology. *J. Hum. Evol.* **52**, 16–30 (2007). doi:10.1016/j.jhevol.2006.07.012 Medline

39. R. L. Jantz, D. W. Owsley, Variation among early North American crania. *Am. J. Phys. Anthropol.* **114**, 146–155 (2001). doi:10.1002/1096-8644(200102)114:2<146:AID-AJPA1014>3.0.CO;2-E Medline

40. N. von Cramon-Taubadel, A. Strauss, M. Hubbe, Evolutionary population history of early Paleoamerican cranial morphology. *Sci. Adv.* **3**, e1602289 (2017). doi:10.1126/sciadv.1602289 Medline

41. W. A. Neves, J. F. Powell, A. Prous, E. G. Ozolins, M. Blum, Lapa Vermelha IV Hominid 1: Morphological affinities of the earliest known American. *Genet. Mol. Biol.* **22**, 461–469 (1999). doi:10.1590/S1415-47571999000400001

42. R. González-José, M. C. Bortolini, F. R. Santos, S. L. Bonatto, The peopling of America: Craniofacial shape variation on a continental scale and its interpretation from an interdisciplinary view. *Am. J. Phys. Anthropol.* **137**, 175–187 (2008). doi:10.1002/ajpa.20854 Medline

43. C. de la Fuente, M. C. Ávila-Arcos, J. Galimany, M. L. Carpenter, J. R. Homburger, A. Blanco, P. Contreras, D. Cruz Dávalos, O. Reyes, M. San Roman, A. Moreno-Estrada, P. F. Campos, C. Eng, S. Huntsman, E. G. Burchard, A.-S. Malaspinas, C. D. Bustamante, E. Willerslev, E. Llop, R. A. Verdugo, M. Moraga, Genomic insights into the origin and diversification of late maritime hunter-gatherers from the Chilean Patagonia. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E4006–E4012 (2018). doi:10.1073/pnas.1715688115 Medline

44. H. Schroeder, M. Sikora, S. Gopalakrishnan, L. M. Cassidy, P. Maisano Delser, M. Sandoval Velasco, J. G. Schraiber, S. Rasmussen, J. R. Homburger, M. C. Ávila-Arcos, M. E. Allentoft, J. V. Moreno-Mayar, G. Renaud, A. Gómez-Carballa, J. E. Laffoon, R. J. A. Hopkins, T. F. G. Higham, R. S. Carr, W. C. Schaffer, J. S. Day, M. Hoogland, A. Salas, C. D. Bustamante, R. Nielsen, D. G. Bradley, C. L. Hofman, E. Willerslev, Origins and genetic legacies of the Caribbean Taino. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 2341–2346 (2018). doi:10.1073/pnas.1716839115 Medline

45. L. A. Louderback, D. K. Grayson, M. Llobera, Middle-Holocene climates and human population densities in the Great Basin, western USA. *Holocene* **21**, 366–373 (2011). doi:10.1177/0959683610374888

46. D. K. Grayson, *The Great Basin: A Natural Prehistory* (Univ. of California, revised and expanded ed., 2011).

47. N. N. Johannsen, G. Larson, D. J. Meltzer, M. Vander Linden, A composite window into human history. *Science* **356**, 1118–1120 (2017). doi:10.1126/science.aan0737 Medline

48. D. B. Madsen, D. Rhode, Eds., *Across the West: Human Population Movement and the Expansion of the Numa* (Univ. of Utah, 1994).

49. L. Campbell, *American Indian Languages: The Historical Linguistics of Native America* (Oxford Univ., 1997).

50. L. Bueno, G. Politis, L. Prates, J. Steele, Eds., A Late Pleistocene/Early Holocene Archaeological 14C Database for Central and South America: Palaeoenvironmental Contexts and Demographic Interpretations. *Quat. Int.* **301**, 1–158 (2013).

51. M. R. Waters, T. W. Stafford Jr., Redefining the age of Clovis: Implications for the peopling of the Americas. *Science* **315**, 1122–1126 (2007). doi:10.1126/science.1137166 Medline

52. D. J. Meltzer, *First Peoples in a New World: Colonizing Ice Age America* (Univ. of California, 2009).

53. T. D. Dillehay, C. Ramírez, M. Pino, M. B. Collins, J. Rossen, J. D. Pino-Navarro, Monte Verde: Seaweed, food, medicine, and the peopling of South America. *Science* **320**, 784–786 (2008). doi:10.1126/science.1156533 Medline

54. G. R. Scott, K. Schmitz, K. N. Heim, K. S. Paul, R. Schomberg, M. A. Pilloud, Sinodonty, Sundadonty, and the Beringian Standstill model: Issues of timing and migrations into the New World. *Quat. Int.* **466**, 233–246 (2018). doi:10.1016/j.quaint.2016.04.027

55. M. I. Eren, B. Buchanan, M. J. O'Brien, Social learning and technological evolution during the Clovis colonization of the New World. *J. Hum. Evol.* **80**, 159–170 (2015). doi:10.1016/j.jhevol.2015.01.002 Medline

56. L. Becerra-Valdivia, M. R. Waters, T. W. Stafford Jr., S. L. Anzick, D. Comeskey, T. Devièse, T. Higham, Reassessing the chronology of the archaeological site of Anzick. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 7000–7003 (2018). doi:10.1073/pnas.1803624115 Medline

57. D. L. Jenkins, L. G. Davis, T. W. Stafford Jr., P. F. Campos, B. Hockett, G. T. Jones, L. S. Cummings, C. Yost, T. J. Connolly, R. M. Yohe 2nd, S. C. Gibbons, M. Raghavan, M. Rasmussen, J. L. A. Paijmans, M. Hofreiter, B. M. Kemp, J. L. Barta, C. Monroe, M. T. P. Gilbert, E. Willerslev, Clovis age Western Stemmed projectile points and human coprolites at the Paisley Caves. *Science* **337**, 223–228 (2012). doi:10.1126/science.1218443 Medline

58. C. E. Hughes, M. P. Rogers, A. C. Owings, B. Petzelt, J. Mitchell, H. Harry, T. Williams, D. Goldberg, D. Labuda, D. G. Smith, J. S. Cybulski, R. S. Malhi, Genetic Structure of First Nation Communities in the Pacific Northwest. *Hum. Biol.* **88**, 251–263 (2016). doi:10.13110/humanbiology.88.4.0251 Medline

59. R. Nielsen, J. M. Akey, M. Jakobsson, J. K. Pritchard, S. Tishkoff, E. Willerslev, Tracing the peopling of the world through genomics. *Nature* **541**, 302–310 (2017). doi:10.1038/nature21347 Medline

60. S. Lindgreen, AdapterRemoval: Easy cleaning of next-generation sequencing reads. *BMC Res. Notes* **5**, 337 (2012). doi:10.1186/1756-0500-5-337 Medline

61. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009). doi:10.1093/bioinformatics/btp324 Medline

62. M. Schubert, A. Ginolhac, S. Lindgreen, J. F. Thompson, K. A. S. Al-Rasheid, E. Willerslev, A. Krogh, L. Orlando, Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics* **13**, 178 (2012). doi:10.1186/1471-2164-13-178 Medline

63. Picard, http://picard.sourceforge.net.

64. M. A. DePristo, E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. del Angel, M. A. Rivas, M. Hanna, A. McKenna, T. J. Fennell, A. M. Kernytsky, A. Y. Sivachenko, K. Cibulskis, S. B. Gabriel, D. Altshuler, M. J. Daly, A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011). doi:10.1038/ng.806 Medline

65. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin; 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009). doi:10.1093/bioinformatics/btp352 Medline

66. B. N. Howie, P. Donnelly, J. Marchini, A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLOS Genet.* **5**, e1000529 (2009). doi:10.1371/journal.pgen.1000529 Medline

67. B. Howie, J. Marchini, M. Stephens, Genotype imputation with thousands of genomes. *G3 (Bethesda)* **1**, 457–470 (2011). doi:10.1534/g3.111.001198 Medline

68. SNPable Regions, http://lh3lh3.users.sourceforge.net/snpable.shtml.

69. The 1000 Genomes Project Consortium, A global reference for human genetic variation. *Nature* **526**, 68–74 (2015). doi:10.1038/nature15393 Medline

70. Q. Fu, A. Mittnik, P. L. F. Johnson, K. Bos, M. Lari, R. Bollongino, C. Sun, L. Giemsch, R. Schmitz, J. Burger, A. M. Ronchitelli, F. Martini, R. G. Cremonesi, J. Svoboda, P. Bauer, D. Caramelli, S. Castellano, D. Reich, S. Pääbo, J. Krause, A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* **23**, 553–559 (2013). doi:10.1016/j.cub.2013.02.044 Medline

71. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**, 356 (2014). doi:10.1186/s12859-014-0356-4 Medline

72. F. Racimo, G. Renaud, M. Slatkin, Joint Estimation of Contamination, Error and Demography for Nuclear DNA from Ancient Humans. *PLOS Genet.* **12**, e1005972 (2016). doi:10.1371/journal.pgen.1005972 Medline

73. A. S. Dyke, A. Moore, L. Robertson, “Deglaciation of North America” (Geological Survey of Canada Open File 1547, 2003).

74. J. Lindo *et al*., The genetic prehistory of the Andean highlands: 7000 years BP though European contact. *Sci. Adv.* 10.1126/sciadv.aau4921 (2018).

75. C. Posth *et al*., Reconstructing the Deep Population History of Central and South America. *Cell* 10.1016/j.cell.2018.10.027 (2018).

76. H. Larsen, Trail Creek: Final report on the excavation of two caves on Seward Peninsula, Alaska. *Acta Arctica* **15**, 7–79 (1968).

77. K. Pasda, *Seward Peninsula, Alaska: Trail Creek Caves 2 and 9 Revisited: The Skeletal Remains* (BAR International Series, Archaeopress, 2012).

78. S. J. AlQahtani, M. P. Hector, H. M. Liversidge, Brief communication: The London atlas of human tooth development and eruption. *Am. J. Phys. Anthropol.* **142**, 481–490 (2010). doi:10.1002/ajpa.21258 Medline

79. H. M. Liversidge, T. Molleson, Variation in crown and root formation and eruption of human deciduous teeth. *Am. J. Phys. Anthropol.* **123**, 172–180 (2004). doi:10.1002/ajpa.10318 Medline

80. J. S. Cybulski *et al.*, The Big Bar Lake burial: middle period human remains from the Canadian Plateau. *Can. J. Archaeol.*, 55–78 (2007).

81. R. S. Malhi, B. M. Kemp, J. A. Eshleman, J. Cybulski, D. G. Smith, S. Cousins, H. Harry, Mitochondrial haplogroup M discovered in prehistoric North Americans. *J. Archaeol. Sci.* **34**, 642–648 (2007). doi:10.1016/j.jas.2006.07.004

82. D. L. Pokotylo, D. Mitchell, in *Plateau*, D. E. Walker, Ed. (Handbook of North American Indians, Smithsonian Institution, 1998), pp. 81–102.

83. S. M. Wheeler, Cave burials near Fallon, Nevada. *Nev. Hist. Soc. Q.* **40**, 15–23 (1997).

84. B. Hockett, E. Palus, A Brief History and Perspective on Spirit Cave, Nevada. *PaleoAmerica* **4**, 1–7 (2018). doi:10.1080/20555563.2017.1412211

85. A. Dansie, Early Holocene burials in Nevada: Overview of localities, research and legal issues. *Nev. Hist. Soc. Q.* **40**, 4–14 (1997).

86. C. S. Fowler, E. M. Hattori, in *The Great Basin: People and Place in Ancient Times*, C. S. Fowler, D. D. Fowler, Eds. (Popular Southwestern Archaeology Books, School for Advanced Research Press, 2008), pp. 60–67.

87. P. Barker, C. Ellis, S. Damadio, “Determination of cultural affiliation of ancient human remains from Spirit Cave, Nevada” (Unpublished report on file, Bureau of Land Management, Nevada State Office, Reno, 2000).

88. H. J. Hecht Edgar, Paleopathology of the Wizards Beach Man (Ahur 2023) and the Spirit Cave mummy (Ahur 2064). *Nev. Hist. Soc. Q.* **40**, 57–61 (1997).

89. R. L. Jantz, D. W. Owsley, Pathology, taphonomy, and cranial morphometrics of the Spirit Cave mummy. *Nev. Hist. Soc. Q.* **40**, 62–82 (1997).

90. J. F. Powell, *The First Americans: Race, Evolution, and the Origin of Native Americans* (Cambridge Univ., 2005).

91. J. F. Powell, W. A. Neves, Craniofacial morphology of the first Americans: Pattern and process in the peopling of the New World. *Am. J. Phys. Anthropol.* **110** (suppl. 29), 153–188 (1999). Medline

92. D. W. Owsley, R. L. Jantz, Eds., *Kennewick Man: The Scientific Investigation of an Ancient American Skeleton* (Texas A&M Univ., ed. 1, 2014).

93. B. M. Auerbach, Skeletal variation among early Holocene North American humans: Implications for origins and diversity in the Americas. *Am. J. Phys. Anthropol.* **149**, 525–536 (2012). doi:10.1002/ajpa.22154 Medline

94. R. F. Heizer, L. K. Napton, *Archaeology and the Prehistoric Great Basin Lacustrine Subsistence Regime as Seen from Lovelock Cave, Nevada*, vol. 10 of *Contributions of the University of California Archaeological Research Facility* (Univ. of California, 1970).

95. W. A. Neves *et al*., in *Paleoamerican Odyssey*, K. E. Graf, C. V. Ketron, M. R. Waters, Eds. (Texas A&M Univ., Texas A&M Univ. ed. 1., 2014), pp. 397–412.

96. P. W. Lund, Carta escripta da Lagôa Santa (Minas Geraes), ao Sr. 1.o Secertario do Instituto, pelo socio honorario Sr. Dr. Lund. *Revista trimensal de Historia e Geographia* **4**, 80–87 (1842).

97. P. W. Lund, Carta do Dr. Lund escripta de Lagôa Santa (Minas Geraes) a 21 de abril de 1844. *Revista trimensal de Historia e Geographia* **6**, 326–334 (1844).

98. P. W. Lund, Notice sur des ossements humains fossiles, trouvés dans une caverne du Brésil. *Mémoires de la Societé Royale des Antiquaires du Nord* **2**, 49–77 (1843).

99. A. Hrdlička, *Early Man in South America* (Bureau of American Ethnology Bulletin 52, Government Printing Office, 1912).

100. O. R. Ortiz-Troncoso, Los yacimientos de Punta Santa Ana y Bahía Buena (Patagonia Austral). Excavaciones y fechados radiocarbónicos. *An. Inst. Patagon.* **7**, 93–122 (1975).

101. D. Legoupil, P. Sellier, La sepultura del la cueva Ayayema (isla Madre de Dios, archipielagos occidentales de Patagonia. *Magallania* **32**, 115–124 (2004).

102. F. Morello, J. Torres, I. Martínez, K. Rodriguez, M. Arroyo-Kalin, C. French, V. Sierpe, M. San Román, Arqueología de la Punta Santa Ana: Recontrucción de secuencias de ocupación de cazadores-recolectores marinos del Estrecho de Magallanes, Patagonia. *Magallania* **40**, 129–149 (2012). doi:10.4067/S0718-22442012000200008

103. A. G. Hogg, Q. Hua, P. G. Blackwell, M. Niu, C. E. Buck, T. P. Guilderson, T. J. Heaton, J. G. Palmer, P. J. Reimer, R. W. Reimer, C. S. M. Turney, S. R. H. Zimmerman, SHCal13 Southern Hemisphere Calibration, 0–50,000 Years cal BP. *Radiocarbon* **55**, 1889–1903 (2013). doi:10.2458/azu_js_rc.55.16783

104. M. Stuiver, P. J. Reimer, Extended 14C Data Base and Revised CALIB 3.0 14C Age Calibration Program. *Radiocarbon* **35**, 215–230 (1993). doi:10.1017/S0033822200013904

105. J. Schobinger, Los santuarios de altura incaicos y el Aconcagua: Aspectos generales e interpretativos. *Relaciones de la Sociedad Argentina de Antropología* **24**, 7–27 (1999).

106. S. Quevedo, E. Durán, Ofrendas a los dioses en las montañas: Santuarios de altura de la cultura inka. *Bol. Mus. Nac. Hist. Nat.* **43**, 193–206 (1992).

107. A. Salas, L. Catelli, J. Pardo-Seco, A. Gómez-Carballa, F. Martinón-Torres, J. Roberto-Barcena, C. Vullo, Y-chromosome Peruvian origin of the 500-year-old Inca child mummy sacrificed in Cerro Aconcagua (Argentina). *Sci. Bull.* (2018). doi:10.1016/j.scib.2018.08.009

108. A. R. Radcliffe-Brown, *The Andaman Islanders: A Study in Social Anthropology* (Forgotten Books, 2015).

109. R. S. Malhi, J. S. Cybulski, R. Y. Tito, J. Johnson, H. Harry, C. Dan, Brief communication: Mitochondrial haplotype C4c confirmed as a founding genome in the Americas. *Am. J. Phys. Anthropol.* **141**, 494–497 (2010). Medline

110. J. Ruhs, "Memorandum: Spirit Cave human remains and associated funerary objects – conclusions and determination concerning Native American status, cultural affiliation and disposition" (Unpublished report on file, Bureau of Land Management, Nevada State Office, Reno, 2016).

111. Department of the Interior, Notice of Inventory Completion: U.S. Department of the Interior, Bureau of Land Management, Nevada State Office, Reno, NV. *Fed. Regist.* **81**, 71760–71762 (2016).

112. E. Callaway, North America's oldest mummy returned to US tribe after genome sequencing. *Nature* **540**, 178–179 (2016). doi:10.1038/540178a

113. F. Brock, T. Higham, P. Ditchfield, C. B. Ramsey, Current Pretreatment Methods for AMS Radiocarbon Dating at the Oxford Radiocarbon Accelerator Unit (Orau). *Radiocarbon* **52**, 103–112 (2010). doi:10.1017/S0033822200045069

114. T. Deviese, D. Comeskey, J. McCullagh, C. Bronk Ramsey, T. Higham, New protocol for compound-specific radiocarbon analysis of archaeological bones. *Rapid Commun. Mass Spectrom.* **32**, 373–379 (2018). doi:10.1002/rcm.8047 Medline

115. M. T. P. Gilbert, H.-J. Bandelt, M. Hofreiter, I. Barnes, Assessing ancient DNA studies. *Trends Ecol. Evol.* **20**, 541–544 (2005). doi:10.1016/j.tree.2005.07.005 Medline

116. E. Willerslev, A. Cooper, Ancient DNA. *Proc. Biol. Sci.* **272**, 3–16 (2005). doi:10.1098/rspb.2004.2813 Medline

117. P. B. Damgaard, A. Margaryan, H. Schroeder, L. Orlando, E. Willerslev, M. E. Allentoft, Improving access to endogenous DNA in ancient bones and teeth. *Sci. Rep.* **5**, 11184 (2015). doi:10.1038/srep11184 Medline

118. M. E. Allentoft, M. Sikora, K.-G. Sjögren, S. Rasmussen, M. Rasmussen, J. Stenderup, P. B. Damgaard, H. Schroeder, T. Ahlström, L. Vinner, A.-S. Malaspinas, A. Margaryan, T. Higham, D. Chivall, N. Lynnerup, L. Harvig, J. Baron, P. Della Casa, P. Dąbrowski, P. R. Duffy, A. V. Ebel, A. Epimakhov, K. Frei, M. Furmanek, T. Gralak, A. Gromov, S. Gronkiewicz, G. Grupe, T. Hajdu, R. Jarysz, V. Khartanovich, A. Khokhlov, V. Kiss, J. Kolář, A. Kriiska, I. Lasak, C. Longhi, G. McGlynn, A. Merkevicius, I. Merkyte, M. Metspalu, R. Mkrtchyan, V. Moiseyev, L. Paja, G. Pálfi, D. Pokutta, Ł. Pospieszny, T. D. Price, L. Saag, M. Sablin, N. Shishlina, V. Smrčka, V. I. Soenov, V. Szeverényi, G. Tóth, S. V. Trifanova, L. Varul, M. Vicze, L. Yepiskoposyan, V. Zhitenev, L.

Orlando, T. Sicheritz-Pontén, S. Brunak, R. Nielsen, K. Kristiansen, E. Willerslev, Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–172 (2015). doi:10.1038/nature14507 Medline

119. I. Glocke, M. Meyer, Extending the spectrum of DNA sequences retrieved from ancient bones and teeth. *Genome Res.* **27**, 1230–1237 (2017). doi:10.1101/gr.219675.116 Medline

120. M. Meyer, M. Kircher, Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **2010**, t5448 (2010). doi:10.1101/pdb.prot5448 Medline

121. J. Dabney, M. Knapp, I. Glocke, M.-T. Gansauge, A. Weihmann, B. Nickel, C. Valdiosera, N. García, S. Pääbo, J.-L. Arsuaga, M. Meyer, Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 15758–15763 (2013). doi:10.1073/pnas.1314445110 Medline

122. L. Orlando, A. Ginolhac, G. Zhang, D. Froese, A. Albrechtsen, M. Stiller, M. Schubert, E. Cappellini, B. Petersen, I. Moltke, P. L. F. Johnson, M. Fumagalli, J. T. Vilstrup, M. Raghavan, T. Korneliussen, A.-S. Malaspinas, J. Vogt, D. Szklarczyk, C. D. Kelstrup, J. Vinther, A. Dolocan, J. Stenderup, A. M. V. Velazquez, J. Cahill, M. Rasmussen, X. Wang, J. Min, G. D. Zazula, A. Seguin-Orlando, C. Mortensen, K. Magnussen, J. F. Thompson, J. Weinstock, K. Gregersen, K. H. Røed, V. Eisenmann, C. J. Rubin, D. C. Miller, D. F. Antczak, M. F. Bertelsen, S. Brunak, K. A. S. Al-Rasheid, O. Ryder, L. Andersson, J. Mundy, A. Krogh, M. T. P. Gilbert, K. Kjær, T. Sicheritz-Ponten, L. J. Jensen, J. V. Olsen, M. Hofreiter, R. Nielsen, B. Shapiro, J. Wang, E. Willerslev, Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* **499**, 74–78 (2013). doi:10.1038/nature12323 Medline

123. C. Carøe, S. Gopalakrishnan, L. Vinner, S. S. T. Mak, M. H. S. Sinding, J. A. Samaniego, N. Wales, T. Sicheritz-Pontén, M. T. P. Gilbert, Single-tube library preparation for degraded DNA. *Methods Ecol. Evol.* **9**, 410–419 (2018). doi:10.1111/2041-210X.12871

124. S. S. T. Mak, S. Gopalakrishnan, C. Carøe, C. Geng, S. Liu, M. S. Sinding, L. F. K. Kuderna, W. Zhang, S. Fu, F. G. Vieira, M. Germonpré, H. Bocherens, S. Fedorov, B. Petersen, T. Sicheritz-Pontén, T. Marques-Bonet, G. Zhang, H. Jiang, M. T. P. Gilbert, Comparative performance of the BGISEQ-500 vs Illumina HiSeq2500 sequencing platforms for palaeogenomic sequencing. *Gigascience* **6**, 1–13 (2017). doi:10.1093/gigascience/gix049 Medline

125. M. Kircher, S. Sawyer, M. Meyer, Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3–e3 (2012). doi:10.1093/nar/gkr771 Medline

126. M. Meyer, M. Kircher, M.-T. Gansauge, H. Li, F. Racimo, S. Mallick, J. G. Schraiber, F. Jay, K. Prüfer, C. de Filippo, P. H. Sudmant, C. Alkan, Q. Fu, R. Do, N. Rohland, A. Tandon, M. Siebauer, R. E. Green, K. Bryc, A. W. Briggs, U. Stenzel, J. Dabney, J. Shendure, J. Kitzman, M. F. Hammer, M. V. Shunkov, A. P. Derevianko, N. Patterson, A. M. Andrés, E. E. Eichler, M. Slatkin, D. Reich, J. Kelso, S. Pääbo, A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222–226 (2012). doi:10.1126/science.1224344 Medline

127. A.-S. Malaspinas, M. C. Westaway, C. Muller, V. C. Sousa, O. Lao, I. Alves, A. Bergström, G. Athanasiadis, J. Y. Cheng, J. E. Crawford, T. H. Heupink, E. Macholdt, S. Peischl, S. Rasmussen, S. Schiffels, S. Subramanian, J. L. Wright, A. Albrechtsen, C. Barbieri, I. Dupanloup, A. Eriksson, A. Margaryan, I. Moltke, I. Pugach, T. S. Korneliussen, I. P. Levkivskyi, J. V. Moreno-Mayar, S. Ni, F. Racimo, M. Sikora, Y. Xue, F. A. Aghakhanian, N. Brucato, S. Brunak, P. F. Campos, W. Clark, S. Ellingvåg, G. Fourmile, P. Gerbault, D. Injie, G. Koki, M. Leavesley, B. Logan, A. Lynch, E. A. Matisoo-Smith, P. J. McAllister, A. J. Mentzer, M. Metspalu, A. B. Migliano, L. Murgha, M. E. Phipps, W. Pomat, D. Reynolds, F.-X. Ricaut, P. Siba, M. G. Thomas, T. Wales, C. M. Wall, S. J. Oppenheimer, C. Tyler-Smith, R. Durbin, J. Dortch, A. Manica, M. H. Schierup, R. A. Foley, M. M. Lahr, C. Bowern, J. D. Wall, T. Mailund, M. Stoneking, R. Nielsen, M. S. Sandhu, L. Excoffier, D. M. Lambert, E. Willerslev, A genomic history of Aboriginal Australia. *Nature* **538**, 207–214 (2016). doi:10.1038/nature18299 Medline

128. A. W. Briggs, U. Stenzel, P. L. F. Johnson, R. E. Green, J. Kelso, K. Prüfer, M. Meyer, J. Krause, M. T. Ronan, M. Lachmann, S. Pääbo, Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 14616–14621 (2007). doi:10.1073/pnas.0704665104 Medline

129. W. J. Kent, C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler, D. Haussler, The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002). doi:10.1101/gr.229102 Medline

130. K. Katoh, K. Misawa, K. Kuma, T. Miyata, MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002). doi:10.1093/nar/gkf436 Medline

131. K. Katoh, M. C. Frith, Adding unaligned sequences into an existing alignment using MAFFT and LAST. *Bioinformatics* **28**, 3144–3146 (2012). doi:10.1093/bioinformatics/bts578 Medline

132. A. Gelman, D. B. Rubin, Inference from iterative simulation using multiple sequences. *Stat. Sci.* **7**, 457–472 (1992). doi:10.1214/ss/1177011136

133. M. Rasmussen, X. Guo, Y. Wang, K. E. Lohmueller, S. Rasmussen, A. Albrechtsen, L. Skotte, S. Lindgreen, M. Metspalu, T. Jombart, T. Kivisild, W. Zhai, A. Eriksson, A. Manica, L. Orlando, F. M. De La Vega, S. Tridico, E. Metspalu, K. Nielsen, M. C. Ávila-Arcos, J. V. Moreno-Mayar, C. Muller, J. Dortch, M. T. P. Gilbert, O. Lund, A. Wesolowska, M. Karmin, L. A. Weinert, B. Wang, J. Li, S. Tai, F. Xiao, T. Hanihara, G. van Driem, A. R. Jha, F.-X. Ricaut, P. de Knijff, A. B. Migliano, I. Gallego Romero, K. Kristiansen, D. M. Lambert, S. Brunak, P. Forster, B. Brinkmann, O. Nehlich, M. Bunce, M. Richards, R. Gupta, C. D. Bustamante, A. Krogh, R. A. Foley, M. M. Lahr, F. Balloux, T. Sicheritz-Pontén, R. Villems, R. Nielsen, J. Wang, E. Willerslev, An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science* **334**, 94–98 (2011). doi:10.1126/science.1211177 Medline

134. D. M. Altshuler, R. A. Gibbs, L. Peltonen, D. M. Altshuler, R. A. Gibbs, L. Peltonen, E. Dermitzakis, S. F. Schaffner, F. Yu, L. Peltonen, E. Dermitzakis, P. E. Bonnen, D. M. Altshuler, R. A. Gibbs, P. I. de Bakker, P. Deloukas, S. B. Gabriel, R. Gwilliam, S. Hunt, M. Inouye, X. Jia, A. Palotie, M. Parkin, P. Whittaker, F. Yu, K. Chang, A. Hawes, L. R. Lewis, Y. Ren, D. Wheeler, R. A.

Gibbs, D. M. Muzny, C. Barnes, K. Darvishi, M. Hurles, J. M. Korn, K. Kristiansson, C. Lee, S. A. McCarrol, J. Nemesh, E. Dermitzakis, A. Keinan, S. B. Montgomery, S. Pollack, A. L. Price, N. Soranzo, P. E. Bonnen, R. A. Gibbs, C. Gonzaga-Jauregui, A. Keinan, A. L. Price, F. Yu, V. Anttila, W. Brodeur, M. J. Daly, S. Leslie, G. McVean, L. Moutsianas, H. Nguyen, S. F. Schaffner, Q. Zhang, M. J. Ghori, R. McGinnis, W. McLaren, S. Pollack, A. L. Price, S. F. Schaffner, F. Takeuchi, S. R. Grossman, I. Shlyakhter, E. B. Hostetter, P. C. Sabeti, C. A. Adebamowo, M. W. Foster, D. R. Gordon, J. Licinio, M. C. Manca, P. A. Marshall, I. Matsuda, D. Ngare, V. O. Wang, D. Reddy, C. N. Rotimi, C. D. Royal, R. R. Sharp, C. Zeng, L. D. Brooks, J. E. McEwen; International HapMap 3 Consortium, Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010). doi:10.1038/nature09298 Medline

135. P. Skoglund, J. Storå, A. Götherström, M. Jakobsson, Accurate sex identification of ancient human remains using DNA shotgun sequencing. *J. Archaeol. Sci.* **40**, 4477–4482 (2013). doi:10.1016/j.jas.2013.07.004

136. E. Cappellini, A. Prohaska, F. Racimo, F. Welker, M. W. Pedersen, M. E. Allentoft, P. de Barros Damgaard, P. Gutenbrunner, J. Dunne, S. Hammann, M. Roffet-Salque, M. Ilardo, J. V. Moreno-Mayar, Y. Wang, M. Sikora, L. Vinner, J. Cox, R. P. Evershed, E. Willerslev, Ancient Biomolecules and Evolutionary Inference. *Annu. Rev. Biochem.* **87**, 1029–1060 (2018). doi:10.1146/annurev-biochem-062917-012002 Medline

137. T. Kivisild, The study of human Y chromosome variation through ancient DNA. *Hum. Genet.* **136**, 529–546 (2017). doi:10.1007/s00439-017-1773-z Medline

138. P. de Barros Damgaard, R. Martiniano, J. Kamm, J. V. Moreno-Mayar, G. Kroonen, M. Peyrot, G. Barjamovic, S. Rasmussen, C. Zacho, N. Baimukhanov, V. Zaibert, V. Merz, A. Biddanda, I. Merz, V. Loman, V. Evdokimov, E. Usmanova, B. Hemphill, A. Seguin-Orlando, F. E. Yediay, I. Ullah, K.-G. Sjögren, K. H. Iversen, J. Choin, C. de la Fuente, M. Ilardo, H. Schroeder, V. Moiseyev, A. Gromov, A. Polyakov, S. Omura, S. Y. Senyurt, H. Ahmad, C. McKenzie, A. Margaryan, A. Hameed, A. Samad, N. Gul, M. H. Khokhar, O. I. Goriunova, V. I. Bazaliiskii, J. Novembre, A. W. Weber, L. Orlando, M. E. Allentoft, R. Nielsen, K. Kristiansen, M. Sikora, A. K. Outram, R. Durbin, E. Willerslev, The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science* **360**, eaar7711 (2018). doi:10.1126/science.aar7711 Medline

139. G. D. Poznik, B. M. Henn, M.-C. Yee, E. Sliwerska, G. M. Euskirchen, A. A. Lin, M. Snyder, L. Quintana-Murci, J. M. Kidd, P. A. Underhill, C. D. Bustamante, Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males versus females. *Science* **341**, 562–565 (2013). doi:10.1126/science.1237619 Medline

140. E. Garrison, G. Marth, Haplotype-based variant detection from short-read sequencing. arXiv:1207.3907 [q-bio.GN] (20 July 2012).

141. G. D. Poznik, Y. Xue, F. L. Mendez, T. F. Willems, A. Massaia, M. A. Wilson Sayres, Q. Ayub, S. A. McCarthy, A. Narechania, S. Kashin, Y. Chen, R. Banerjee, J. L. Rodriguez-Flores, M. Cerezo, H. Shao, M. Gymrek, A. Malhotra, S. Louzada, R. Desalle, G. R. S. Ritchie, E. Cerveira, T. W. Fitzgerald, E.

Garrison, A. Marcketta, D. Mittelman, M. Romanovitch, C. Zhang, X. Zheng-Bradley, G. R. Abecasis, S. A. McCarroll, P. Flicek, P. A. Underhill, L. Coin, D. R. Zerbino, F. Yang, C. Lee, L. Clarke, A. Auton, Y. Erlich, R. E. Handsaker, C. D. Bustamante, C. Tyler-Smith; 1000 Genomes Project Consortium, Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. *Nat. Genet.* **48**, 593–599 (2016). doi:10.1038/ng.3559 Medline

142. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014). doi:10.1093/bioinformatics/btu033 Medline

143. M. S. Jota, D. R. Lacerda, J. R. Sandoval, P. P. R. Vieira, D. Ohasi, J. E. Santos-Júnior, O. Acosta, C. Cuellar, S. Revollo, C. Paz-Y-Miño, R. Fujita, G. A. Vallejo, T. G. Schurr, E. M. Tarazona-Santos, S. Dj. Pena, Q. Ayub, C. Tyler-Smith, F. R. Santos; Genographic Consortium, New native South American Y chromosome lineages. *J. Hum. Genet.* **61**, 593–603 (2016). doi:10.1038/jhg.2016.26 Medline

144. M. Rasmussen, Y. Li, S. Lindgreen, J. S. Pedersen, A. Albrechtsen, I. Moltke, M. Metspalu, E. Metspalu, T. Kivisild, R. Gupta, M. Bertalan, K. Nielsen, M. T. P. Gilbert, Y. Wang, M. Raghavan, P. F. Campos, H. M. Kamp, A. S. Wilson, A. Gledhill, S. Tridico, M. Bunce, E. D. Lorenzen, J. Binladen, X. Guo, J. Zhao, X. Zhang, H. Zhang, Z. Li, M. Chen, L. Orlando, K. Kristiansen, M. Bak, N. Tommerup, C. Bendixen, T. L. Pierre, B. Grønnow, M. Meldgaard, C. Andreasen, S. A. Fedorova, L. P. Osipova, T. F. G. Higham, C. B. Ramsey, T. V. Hansen, F. C. Nielsen, M. H. Crawford, S. Brunak, T. Sicheritz-Pontén, R. Villems, R. Nielsen, A. Krogh, J. Wang, E. Willerslev, Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**, 757–762 (2010). doi:10.1038/nature08835 Medline

145. M. C. Dulik, A. C. Owings, J. B. Gaieski, M. G. Vilar, A. Andre, C. Lennie, M. A. Mackenzie, I. Kritsch, S. Snowshoe, R. Wright, J. Martin, N. Gibson, T. D. Andrews, T. G. Schurr, S. Adhikarla, C. J. Adler, E. Balanovska, O. Balanovsky, J. Bertranpetit, A. C. Clarke, D. Comas, A. Cooper, C. S. I. Der Sarkissian, A. GaneshPrasad, W. Haak, M. Haber, A. Hobbs, A. Javed, L. Jin, M. E. Kaplan, S. Li, B. Martinez-Cruz, E. A. Matisoo-Smith, M. Mele, N. C. Merchant, R. J. Mitchell, L. Parida, R. Pitchappan, D. E. Platt, L. Quintana-Murci, C. Renfrew, D. R. Lacerda, A. K. Royyuru, F. R. Santos, H. Soodyall, D. F. Soria Hernanz, P. Swamikrishnan, C. Tyler-Smith, A. V. Santhakumari, P. P. Vieira, R. S. Wells, P. A. Zalloua, J. S. Ziegle; Genographic Consortium, Y-chromosome analysis reveals genetic divergence and new founding native lineages in Athapaskan- and Eskimoan-speaking populations. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 8471–8476 (2012). doi:10.1073/pnas.1118760109 Medline

146. S. D. J. Pena, F. R. Santos, N. O. Bianchi, C. M. Bravi, F. R. Carnese, F. Rothhammer, T. Gerelsaikhan, B. Munkhtuja, T. Oyunsuren, A major founder Y-chromosome haplotype in Amerindians. *Nat. Genet.* **11**, 15–16 (1995). doi:10.1038/ng0995-15 Medline

147. T. D. Dillehay, C. Ocampo, J. Saavedra, A. O. Sawakuchi, R. M. Vega, M. Pino, M. B. Collins, L. Scott Cummings, I. Arregui, X. S. Villagran, G. A. Hartmann, M. Mella, A. González, G. Dix, New Archaeological Evidence for an Early

Human Presence at Monte Verde, Chile. *PLOS ONE* **10**, e0141923 (2015). doi:10.1371/journal.pone.0141923 Medline

148. J. F. O'Connell, J. Allen, The process, biotic impact, and global implications of the human colonization of Sahul about 47,000 years ago. *J. Archaeol. Sci.* **56**, 73–84 (2015). doi:10.1016/j.jas.2015.02.020

149. J. F. O'Connell, J. Allen, The restaurant at the end of the universe: Modelling the colonisation of Sahul. *Aust. Archaeol.* **74**, 5–17 (2012). doi:10.1080/03122417.2012.11681932

150. M. Hubbe, W. A. Neves, K. Harvati, Testing evolutionary and dispersion scenarios for the settlement of the new world. *PLOS ONE* **5**, e11105 (2010). doi:10.1371/journal.pone.0011105 Medline

151. M. Hubbe, A. Strauss, A. Hubbe, W. A. Neves, Early South Americans Cranial Morphological Variation and the Origin of American Biological Diversity. *PLOS ONE* **10**, e0138090 (2015). doi:10.1371/journal.pone.0138090 Medline

152. Q. Fu, H. Li, P. Moorjani, F. Jay, S. M. Slepchenko, A. A. Bondarev, P. L. F. Johnson, A. Aximu-Petri, K. Prüfer, C. de Filippo, M. Meyer, N. Zwyns, D. C. Salazar-García, Y. V. Kuzmin, S. G. Keates, P. A. Kosintsev, D. I. Razhev, M. P. Richards, N. V. Peristov, M. Lachmann, K. Douka, T. F. G. Higham, M. Slatkin, J.-J. Hublin, D. Reich, J. Kelso, T. B. Viola, S. Pääbo, Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**, 445–449 (2014). doi:10.1038/nature13810 Medline

153. Q. Fu, M. Hajdinjak, O. T. Moldovan, S. Constantin, S. Mallick, P. Skoglund, N. Patterson, N. Rohland, I. Lazaridis, B. Nickel, B. Viola, K. Prüfer, M. Meyer, J. Kelso, D. Reich, S. Pääbo, An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524**, 216–219 (2015). doi:10.1038/nature14558 Medline

154. T. M. Karafet, F. L. Mendez, H. Sudoyo, J. S. Lansing, M. F. Hammer, Improved phylogenetic resolution and rapid diversification of Y-chromosome haplogroup K-M526 in Southeast Asia. *Eur. J. Hum. Genet.* **23**, 369–373 (2015). doi:10.1038/ejhg.2014.106 Medline

155. A. Bergström, N. Nagle, Y. Chen, S. McCarthy, M. O. Pollard, Q. Ayub, S. Wilcox, L. Wilcox, R. A. H. van Oorschot, P. McAllister, L. Williams, Y. Xue, R. J. Mitchell, C. Tyler-Smith, Deep Roots for Aboriginal Australian Y Chromosomes. *Curr. Biol.* **26**, 809–813 (2016). doi:10.1016/j.cub.2016.01.028 Medline

156. M. Mondal, A. Bergström, Y. Xue, F. Calafell, H. Laayouni, F. Casals, P. P. Majumder, C. Tyler-Smith, J. Bertranpetit, Y-chromosomal sequences of diverse Indian populations and the ancestry of the Andamanese. *Hum. Genet.* **136**, 499–510 (2017). doi:10.1007/s00439-017-1800-0 Medline

157. G. Renaud, V. Slon, A. T. Duggan, J. Kelso, Schmutzi: Estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* **16**, 224 (2015). doi:10.1186/s13059-015-0776-0 Medline

158. M. van Oven, M. Kayser, Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* **30**, E386–E394 (2009). doi:10.1002/humu.20921 Medline

159. C. Posth, G. Renaud, A. Mittnik, D. G. Drucker, H. Rougier, C. Cupillard, F. Valentin, C. Thevenet, A. Furtwängler, C. Wißing, M. Francken, M. Malina, M. Bolus, M. Lari, E. Gigli, G. Capecchi, I. Crevecoeur, C. Beauval, D. Flas, M. Germonpré, J. van der Plicht, R. Cottiaux, B. Gély, A. Ronchitelli, K. Wehrberger, D. Grigorescu, J. Svoboda, P. Semal, D. Caramelli, H. Bocherens, K. Harvati, N. J. Conard, W. Haak, A. Powell, J. Krause, Pleistocene Mitochondrial Genomes Suggest a Single Major Dispersal of Non-Africans and a Late Glacial Population Turnover in Europe. *Curr. Biol.* **26**, 827–833 (2016). doi:10.1016/j.cub.2016.01.037 Medline

160. A. T. Duggan, A. J. T. Harris, S. Marciniak, I. Marshall, M. Kuch, A. Kitchen, G. Renaud, J. Southon, B. Fuller, J. Young, S. Fiedel, G. B. Golding, V. Grimes, H. Poinar, Genetic Discontinuity between the Maritime Archaic and Beothuk Populations in Newfoundland, Canada. *Curr. Biol.* **27**, 3149–3156.e11 (2017). doi:10.1016/j.cub.2017.08.053 Medline

161. S. Kumar, G. Stecher, K. Tamura, MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016). doi:10.1093/molbev/msw054 Medline

162. H. Weissensteiner, D. Pacher, A. Kloss-Brandstätter, L. Forer, G. Specht, H.-J. Bandelt, F. Kronenberg, A. Salas, S. Schönherr, HaploGrep 2: Mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* **44**, W58–W63 (2016). doi:10.1093/nar/gkw233 Medline

163. D. Vianello, F. Sevini, G. Castellani, L. Lomartire, M. Capri, C. Franceschi, HAPLOFIND: A new method for high-throughput mtDNA haplogroup assignment. *Hum. Mutat.* **34**, 1189–1194 (2013). doi:10.1002/humu.22356 Medline

164. P. Soares, L. Ermini, N. Thomson, M. Mormina, T. Rito, A. Röhl, A. Salas, S. Oppenheimer, V. Macaulay, M. B. Richards, Correcting for purifying selection: An improved human mitochondrial molecular clock. *Am. J. Hum. Genet.* **84**, 740–759 (2009). doi:10.1016/j.ajhg.2009.05.001 Medline

165. R. Bouckaert, J. Heled, D. Kühnert, T. Vaughan, C.-H. Wu, D. Xie, M. A. Suchard, A. Rambaut, A. J. Drummond, BEAST 2: A software platform for Bayesian evolutionary analysis. *PLOS Comput. Biol.* **10**, e1003537 (2014). doi:10.1371/journal.pcbi.1003537 Medline

166. T. G. Schurr, S. W. Ballinger, Y. Y. Gan, J. A. Hodge, D. A. Merriwether, D. N. Lawrence, W. C. Knowler, K. M. Weiss, D. C. Wallace, Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. *Am. J. Hum. Genet.* **46**, 613–623 (1990). Medline

167. A. Torroni, T. G. Schurr, M. F. Cabell, M. D. Brown, J. V. Neel, M. Larsen, D. G. Smith, C. M. Vullo, D. C. Wallace, Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am. J. Hum. Genet.* **53**, 563–590 (1993). Medline

168. A. Achilli, U. A. Perego, C. M. Bravi, M. D. Coble, Q.-P. Kong, S. R. Woodward, A. Salas, A. Torroni, H.-J. Bandelt, The phylogeny of the four pan-American MtDNA haplogroups: Implications for evolutionary and disease studies. *PLOS ONE* **3**, e1764 (2008). doi:10.1371/journal.pone.0001764 Medline

169. S. Brandini, P. Bergamaschi, M. F. Cerna, F. Gandini, F. Bastaroli, E. Bertolini, C. Cereda, L. Ferretti, A. Gómez-Carballa, V. Battaglia, A. Salas, O. Semino, A. Achilli, A. Olivieri, A. Torroni, The Paleo-Indian Entry into South America According to Mitogenomes. *Mol. Biol. Evol.* **35**, 299–311 (2018). doi:10.1093/molbev/msx267 Medline

170. M. Bodner, U. A. Perego, G. Huber, L. Fendt, A. W. Röck, B. Zimmermann, A. Olivieri, A. Gómez-Carballa, H. Lancioni, N. Angerhofer, M. C. Bobillo, D. Corach, S. R. Woodward, A. Salas, A. Achilli, A. Torroni, H.-J. Bandelt, W. Parson, Rapid coastal spread of First Americans: Novel insights from South America's Southern Cone mitochondrial genomes. *Genome Res.* **22**, 811–820 (2012). doi:10.1101/gr.131722.111 Medline

171. U. A. Perego, A. Achilli, N. Angerhofer, M. Accetturo, M. Pala, A. Olivieri, B. Hooshiar Kashani, K. H. Ritchie, R. Scozzari, Q.-P. Kong, N. M. Myres, A. Salas, O. Semino, H.-J. Bandelt, S. R. Woodward, A. Torroni, Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. *Curr. Biol.* **19**, 1–8 (2009). doi:10.1016/j.cub.2008.11.058 Medline

172. Y. Cui, J. Lindo, C. E. Hughes, J. W. Johnson, A. G. Hernandez, B. M. Kemp, J. Ma, R. Cunningham, B. Petzelt, J. Mitchell, D. Archer, J. S. Cybulski, R. S. Malhi, Ancient DNA analysis of mid-holocene individuals from the Northwest Coast of North America reveals different evolutionary paths for mitogenomes. *PLOS ONE* **8**, e66948 (2013). doi:10.1371/journal.pone.0066948 Medline

173. J. C. Chatters, D. J. Kennett, Y. Asmerom, B. M. Kemp, V. Polyak, A. N. Blank, P. A. Beddows, E. Reinhardt, J. Arroyo-Cabrales, D. A. Bolnick, R. S. Malhi, B. J. Culleton, P. L. Erreguerena, D. Rissolo, S. Morell-Hart, T. W. Stafford Jr., Late Pleistocene human skeleton and mtDNA link Paleoamericans and modern Native Americans. *Science* **344**, 750–754 (2014). doi:10.1126/science.1252619 Medline

174. K. Prüfer, M. Meyer, Comment on "Late Pleistocene human skeleton and mtDNA link Paleoamericans and modern Native Americans". *Science* **347**, 835–835 (2015). doi:10.1126/science.1260617 Medline

175. B. M. Kemp, J. Lindo, D. A. Bolnick, R. S. Malhi, J. C. Chatters, Response to Comment on "Late Pleistocene human skeleton and mtDNA link Paleoamericans and modern Native Americans". *Science* **347**, 835–835 (2015). doi:10.1126/science.1261188 Medline

176. C. de la Fuente, J. Galimany, B. M. Kemp, K. Judd, O. Reyes, M. Moraga, 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 (2015). doi:10.1002/ajpa.22815 Medline

177. M. de Saint Pierre, C. M. Bravi, J. M. B. Motti, N. Fuku, M. Tanaka, E. Llop, S. L. Bonatto, M. Moraga, An alternative model for the early peopling of southern South America revealed by analyses of three mitochondrial DNA haplogroups. *PLOS ONE* **7**, e43486 (2012). doi:10.1371/journal.pone.0043486 Medline

178. B. A. Potter, J. D. Irish, J. D. Reuther, H. J. McKinney, New insights into Eastern Beringian mortuary behavior: A terminal Pleistocene double infant burial at Upward Sun River. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 17060–17065 (2014). doi:10.1073/pnas.1413131111 Medline

179. J. C. Tackney, B. A. Potter, J. Raff, M. Powers, W. S. Watkins, D. Warner, J. D. Reuther, J. D. Irish, D. H. O'Rourke, Two contemporaneous mitogenomes from terminal Pleistocene burials in eastern Beringia. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 13833–13838 (2015). doi:10.1073/pnas.1511903112 Medline

180. F. A. Kaestle, D. G. Smith, Ancient mitochondrial DNA evidence for prehistoric population movement: The Numic expansion. *Am. J. Phys. Anthropol.* **115**, 1–12 (2001). doi:10.1002/ajpa.1051 Medline

181. G. S. Cabana, K. Hunley, F. A. Kaestle, Population continuity or replacement? A novel computer simulation approach and its application to the Numic expansion (Western Great Basin, USA). *Am. J. Phys. Anthropol.* **135**, 438–447 (2008). doi:10.1002/ajpa.20764 Medline

182. K. Thangaraj, G. Chaubey, T. Kivisild, A. G. Reddy, V. K. Singh, A. A. Rasalkar, L. Singh, Reconstructing the origin of Andaman Islanders. *Science* **308**, 996–996 (2005). doi:10.1126/science.1109987 Medline

183. S. S. Barik, R. Sahani, B. V. R. Prasad, P. Endicott, M. Metspalu, B. N. Sarkar, S. Bhattacharya, P. C. H. Annapoorna, J. Sreenath, D. Sun, J. J. Sanchez, S. Y. W. Ho, A. Chandrasekar, V. R. Rao, Detailed mtDNA genotypes permit a reassessment of the settlement and population structure of the Andaman Islands. *Am. J. Phys. Anthropol.* **136**, 19–27 (2008). doi:10.1002/ajpa.20773 Medline

184. M. G. Palanichamy, S. Agrawal, Y. G. Yao, Q. P. Kong, C. Sun, F. Khan, T. K. Chaudhuri, Y. P. Zhang, Comment on "Reconstructing the origin of Andaman islanders". *Science* **311**, 470, author reply 470 (2006). doi:10.1126/science.1120176 Medline

185. P. Endicott, M. Metspalu, C. Stringer, V. Macaulay, A. Cooper, J. J. Sanchez, Multiplexed SNP typing of ancient DNA clarifies the origin of Andaman mtDNA haplogroups amongst South Asian tribal populations. *PLOS ONE* **1**, e81 (2006). doi:10.1371/journal.pone.0000081 Medline

186. H.-W. Wang, B. Mitra, T. K. Chaudhuri, M. G. Palanichamy, Q.-P. Kong, Y.-P. Zhang, Mitochondrial DNA evidence supports northeast Indian origin of the aboriginal Andamanese in the Late Paleolithic. *J. Genet. Genomics* **38**, 117–122 (2011). doi:10.1016/j.jgg.2011.02.005 Medline

187. P. Marrero, K. K. Abu-Amero, J. M. Larruga, V. M. Cabrera, Carriers of human mitochondrial DNA macrohaplogroup M colonized India from southeastern Asia. *BMC Evol. Biol.* **16**, 246 (2016). doi:10.1186/s12862-016-0816-8 Medline

188. M. Mondal, F. Casals, T. Xu, G. M. Dall'Olio, M. Pybus, M. G. Netea, D. Comas, H. Laayouni, Q. Li, P. P. Majumder, J. Bertranpetit, Genomic analysis of Andamanese provides insights into ancient human migration into Asia and adaptation. *Nat. Genet.* **48**, 1066–1070 (2016). doi:10.1038/ng.3621 Medline

189. M. Sikora *et al*., The population history of northeastern Siberia since the Pleistocene. bioRxiv 448829 [Preprint]. 22 October 2018. https://doi.org/10.1101/448829.

190. S. Mallick, H. Li, M. Lipson, I. Mathieson, M. Gymrek, F. Racimo, M. Zhao, N. Chennagiri, S. Nordenfelt, A. Tandon, P. Skoglund, I. Lazaridis, S. Sankararaman, Q. Fu, N. Rohland, G. Renaud, Y. Erlich, T. Willems, C. Gallo, J. P. Spence, Y. S. Song, G. Poletti, F. Balloux, G. van Driem, P. de Knijff, I. G.

Romero, A. R. Jha, D. M. Behar, C. M. Bravi, C. Capelli, T. Hervig, A. Moreno-Estrada, O. L. Posukh, E. Balanovska, O. Balanovsky, S. Karachanak-Yankova, H. Sahakyan, D. Toncheva, L. Yepiskoposyan, C. Tyler-Smith, Y. Xue, M. S. Abdullah, A. Ruiz-Linares, C. M. Beall, A. Di Rienzo, C. Jeong, E. B. Starikovskaya, E. Metspalu, J. Parik, R. Villems, B. M. Henn, U. Hodoglugil, R. Mahley, A. Sajantila, G. Stamatoyannopoulos, J. T. S. Wee, R. Khusainova, E. Khusnutdinova, S. Litvinov, G. Ayodo, D. Comas, M. F. Hammer, T. Kivisild, W. Klitz, C. A. Winkler, D. Labuda, M. Bamshad, L. B. Jorde, S. A. Tishkoff, W. S. Watkins, M. Metspalu, S. Dryomov, R. Sukernik, L. Singh, K. Thangaraj, S. Pääbo, J. Kelso, N. Patterson, D. Reich, The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201–206 (2016). doi:10.1038/nature18964 Medline

191. M. A. Yang, X. Gao, C. Theunert, H. Tong, A. Aximu-Petri, B. Nickel, M. Slatkin, M. Meyer, S. Pääbo, J. Kelso, Q. Fu, 40,000-Year-Old Individual from Asia Provides Insight into Early Population Structure in Eurasia. *Curr. Biol.* **27**, 3202–3208.e9 (2017). doi:10.1016/j.cub.2017.09.030 Medline

192. H. McColl, F. Racimo, L. Vinner, F. Demeter, T. Gakuhari, J. V. Moreno-Mayar, G. van Driem, U. Gram Wilken, A. Seguin-Orlando, C. de la Fuente Castro, S. Wasef, R. Shoocongdej, V. Souksavatdy, T. Sayavongkhamdy, M. M. Saidin, M. E. Allentoft, T. Sato, A.-S. Malaspinas, F. A. Aghakhanian, T. Korneliussen, A. Prohaska, A. Margaryan, P. de Barros Damgaard, S. Kaewsutthi, P. Lertrit, T. M. H. Nguyen, H. C. Hung, T. Minh Tran, H. Nghia Truong, G. H. Nguyen, S. Shahidan, K. Wiradnyana, H. Matsumae, N. Shigehara, M. Yoneda, H. Ishida, T. Masuyama, Y. Yamada, A. Tajima, H. Shibata, A. Toyoda, T. Hanihara, S. Nakagome, T. Deviese, A.-M. Bacon, P. Duringer, J.-L. Ponche, L. Shackelford, E. Patole-Edoumba, A. T. Nguyen, B. Bellina-Pryce, J.-C. Galipaud, R. Kinaston, H. Buckley, C. Pottier, S. Rasmussen, T. Higham, R. A. Foley, M. M. Lahr, L. Orlando, M. Sikora, M. E. Phipps, H. Oota, C. Higham, D. M. Lambert, E. Willerslev, The prehistoric peopling of Southeast Asia. *Science* **361**, 88–92 (2018). doi:10.1126/science.aat3628 Medline

193. E. R. Jones, G. Gonzalez-Fortes, S. Connell, V. Siska, A. Eriksson, R. Martiniano, R. L. McLaughlin, M. Gallego Llorente, L. M. Cassidy, C. Gamba, T. Meshveliani, O. Bar-Yosef, W. Müller, A. Belfer-Cohen, Z. Matskevich, N. Jakeli, T. F. G. Higham, M. Currat, D. Lordkipanidze, M. Hofreiter, A. Manica, R. Pinhasi, D. G. Bradley, Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.* **6**, 8912 (2015). doi:10.1038/ncomms9912 Medline

194. C. Gamba, E. R. Jones, M. D. Teasdale, R. L. McLaughlin, G. Gonzalez-Fortes, V. Mattiangeli, L. Domboróczki, I. Kővári, I. Pap, A. Anders, A. Whittle, J. Dani, P. Raczky, T. F. G. Higham, M. Hofreiter, D. G. Bradley, R. Pinhasi, Genome flux and stasis in a five millennium transect of European prehistory. *Nat. Commun.* **5**, 5257 (2014). doi:10.1038/ncomms6257 Medline

195. M. Sikora, A. Seguin-Orlando, V. C. Sousa, A. Albrechtsen, T. Korneliussen, A. Ko, S. Rasmussen, I. Dupanloup, P. R. Nigst, M. D. Bosch, G. Renaud, M. E. Allentoft, A. Margaryan, S. V. Vasilyev, E. V. Veselovskaya, S. B. Borutskaya, T. Deviese, D. Comeskey, T. Higham, A. Manica, R. Foley, D. J. Meltzer, R. Nielsen, L. Excoffier, M. Mirazon Lahr, L. Orlando, E. Willerslev, Ancient

genomes show social and reproductive behavior of early Upper Paleolithic foragers. *Science* **358**, 659–662 (2017). doi:10.1126/science.aao1807 Medline

196. G. M. Kılınç, A. Omrak, F. Özer, T. Günther, A. M. Büyükkarakaya, E. Bıçakçı, D. Baird, H. M. Dönertaş, A. Ghalichi, R. Yaka, D. Koptekin, S. C. Açan, P. Parvizi, M. Krzewińska, E. A. Daskalaki, E. Yüncü, N. D. Dağtaş, A. Fairbairn, J. Pearson, G. Mustafaoğlu, Y. S. Erdal, Y. G. Çakan, İ. Togan, M. Somel, J. Storå, M. Jakobsson, A. Götherström, The Demographic Development of the First Farmers in Anatolia. *Curr. Biol.* **26**, 2659–2666 (2016). doi:10.1016/j.cub.2016.07.057 Medline

197. Z. Hofmanová, S. Kreutzer, G. Hellenthal, C. Sell, Y. Diekmann, D. Díez-Del-Molino, L. van Dorp, S. López, A. Kousathanas, V. Link, K. Kirsanow, L. M. Cassidy, R. Martiniano, M. Strobel, A. Scheu, K. Kotsakis, P. Halstead, S. Triantaphyllou, N. Kyparissi-Apostolika, D. Urem-Kotsou, C. Ziota, F. Adaktylou, S. Gopalan, D. M. Bobo, L. Winkelbach, J. Blöcher, M. Unterländer, C. Leuenberger, Ç. Çilingiroğlu, B. Horejs, F. Gerritsen, S. J. Shennan, D. G. Bradley, M. Currat, K. R. Veeramah, D. Wegmann, M. G. Thomas, C. Papageorgopoulou, J. Burger, Early farmers from across Europe directly descended from Neolithic Aegeans. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 6886–6891 (2016). doi:10.1073/pnas.1523951113 Medline

198. F. Broushaki, M. G. Thomas, V. Link, S. López, L. van Dorp, K. Kirsanow, Z. Hofmanová, Y. Diekmann, L. M. Cassidy, D. Díez-Del-Molino, A. Kousathanas, C. Sell, H. K. Robson, R. Martiniano, J. Blöcher, A. Scheu, S. Kreutzer, R. Bollongino, D. Bobo, H. Davudi, O. Munoz, M. Currat, K. Abdi, F. Biglari, O. E. Craig, D. G. Bradley, S. Shennan, K. Veeramah, M. Mashkour, D. Wegmann, G. Hellenthal, J. Burger, Early Neolithic genomes from the eastern Fertile Crescent. *Science* **353**, 499–503 (2016). doi:10.1126/science.aaf7943 Medline

199. E. H. M. Wong, A. Khrunin, L. Nichols, D. Pushkarev, D. Khokhrin, D. Verbenko, O. Evgrafov, J. Knowles, J. Novembre, S. Limborska, A. Valouev, Reconstructing genetic history of Siberian and Northeastern European populations. *Genome Res.* **27**, 1–14 (2017). doi:10.1101/gr.202945.115 Medline

200. Q. Ayub, M. Mezzavilla, L. Pagani, M. Haber, A. Mohyuddin, S. Khaliq, S. Q. Mehdi, C. Tyler-Smith, The Kalash genetic isolate: Ancient divergence, drift, and selection. *Am. J. Hum. Genet.* **96**, 775–783 (2015). doi:10.1016/j.ajhg.2015.03.012 Medline

201. P. B. Damgaard, N. Marchi, S. Rasmussen, M. Peyrot, G. Renaud, T. Korneliussen, J. V. Moreno-Mayar, M. W. Pedersen, A. Goldberg, E. Usmanova, N. Baimukhanov, V. Loman, L. Hedeager, A. G. Pedersen, K. Nielsen, G. Afanasiev, K. Akmatov, A. Aldashev, A. Alpaslan, G. Baimbetov, V. I. Bazaliiskii, A. Beisenov, B. Boldbaatar, B. Boldgiv, C. Dorzhu, S. Ellingvag, D. Erdenebaatar, R. Dajani, E. Dmitriev, V. Evdokimov, K. M. Frei, A. Gromov, A. Goryachev, H. Hakonarson, T. Hegay, Z. Khachatryan, R. Khaskhanov, E. Kitov, A. Kolbina, T. Kubatbek, A. Kukushkin, I. Kukushkin, N. Lau, A. Margaryan, I. Merkyte, I. V. Mertz, V. K. Mertz, E. Mijiddorj, V. Moiyesev, G. Mukhtarova, B. Nurmukhanbetov, Z. Orozbekova, I. Panyushkina, K. Pieta, V. Smrčka, I. Shevnina, A. Logvin, K.-G. Sjögren, T. Štolcová, A. M. Taravella, K. Tashbaeva, A. Tkachev, T. Tulegenov, D. Voyakin, L. Yepiskoposyan, S. Undrakhbold, V. Varfolomeev, A. Weber, M. A.

Wilson Sayres, N. Kradin, M. E. Allentoft, L. Orlando, R. Nielsen, M. Sikora, E. Heyer, K. Kristiansen, E. Willerslev, 137 ancient human genomes from across the Eurasian steppes. *Nature* **557**, 369–374 (2018). doi:10.1038/s41586-018-0094-2 Medline

202. K. Prüfer, F. Racimo, N. Patterson, F. Jay, S. Sankararaman, S. Sawyer, A. Heinze, G. Renaud, P. H. Sudmant, C. de Filippo, H. Li, S. Mallick, M. Dannemann, Q. Fu, M. Kircher, M. Kuhlwilm, M. Lachmann, M. Meyer, M. Ongyerth, M. Siebauer, C. Theunert, A. Tandon, P. Moorjani, J. Pickrell, J. C. Mullikin, S. H. Vohr, R. E. Green, I. Hellmann, P. L. F. Johnson, H. Blanche, H. Cann, J. O. Kitzman, J. Shendure, E. E. Eichler, E. S. Lein, T. E. Bakken, L. V. Golovanova, V. B. Doronichev, M. V. Shunkov, A. P. Derevianko, B. Viola, M. Slatkin, D. Reich, J. Kelso, S. Pääbo, The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014). doi:10.1038/nature12886 Medline

203. M. G. Llorente, E. R. Jones, A. Eriksson, V. Siska, K. W. Arthur, J. W. Arthur, M. C. Curtis, J. T. Stock, M. Coltorti, P. Pieruccini, S. Stretton, F. Brock, T. Higham, Y. Park, M. Hofreiter, D. G. Bradley, J. Bhak, R. Pinhasi, A. Manica, Ancient Ethiopian genome reveals extensive Eurasian admixture in Eastern Africa. *Science* **350**, 820–822 (2015). doi:10.1126/science.aad2879 Medline

204. I. Lazaridis, N. Patterson, A. Mittnik, G. Renaud, S. Mallick, K. Kirsanow, P. H. Sudmant, J. G. Schraiber, S. Castellano, M. Lipson, B. Berger, C. Economou, R. Bollongino, Q. Fu, K. I. Bos, S. Nordenfelt, H. Li, C. de Filippo, K. Prüfer, S. Sawyer, C. Posth, W. Haak, F. Hallgren, E. Fornander, N. Rohland, D. Delsate, M. Francken, J.-M. Guinet, J. Wahl, G. Ayodo, H. A. Babiker, G. Bailliet, E. Balanovska, O. Balanovsky, R. Barrantes, G. Bedoya, H. Ben-Ami, J. Bene, F. Berrada, C. M. Bravi, F. Brisighelli, G. B. J. Busby, F. Cali, M. Churnosov, D. E. C. Cole, D. Corach, L. Damba, G. van Driem, S. Dryomov, J.-M. Dugoujon, S. A. Fedorova, I. Gallego Romero, M. Gubina, M. Hammer, B. M. Henn, T. Hervig, U. Hodoglugil, A. R. Jha, S. Karachanak-Yankova, R. Khusainova, E. Khusnutdinova, R. Kittles, T. Kivisild, W. Klitz, V. Kučinskas, A. Kushniarevich, L. Laredj, S. Litvinov, T. Loukidis, R. W. Mahley, B. Melegh, E. Metspalu, J. Molina, J. Mountain, K. Näkkäläjärvi, D. Nesheva, T. Nyambo, L. Osipova, J. Parik, F. Platonov, O. Posukh, V. Romano, F. Rothhammer, I. Rudan, R. Ruizbakiev, H. Sahakyan, A. Sajantila, A. Salas, E. B. Starikovskaya, A. Tarekegn, D. Toncheva, S. Turdikulova, I. Uktveryte, O. Utevska, R. Vasquez, M. Villena, M. Voevoda, C. A. Winkler, L. Yepiskoposyan, P. Zalloua, T. Zemunik, A. Cooper, C. Capelli, M. G. Thomas, A. Ruiz-Linares, S. A. Tishkoff, L. Singh, K. Thangaraj, R. Villems, D. Comas, R. Sukernik, M. Metspalu, M. Meyer, E. E. Eichler, J. Burger, M. Slatkin, S. Pääbo, J. Kelso, D. Reich, J. Krause, Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413 (2014). doi:10.1038/nature13673 Medline

205. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, P. C. Sham, PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007). doi:10.1086/519795 Medline

206. E. Y. Durand, N. Patterson, D. Reich, M. Slatkin, Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* **28**, 2239–2252 (2011). doi:10.1093/molbev/msr048 Medline

207. B. M. Peter, Admixture, Population Structure, and F-Statistics. *Genetics* **202**, 1485–1501 (2016). doi:10.1534/genetics.115.183913 Medline

208. R. E. Green, J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, H. Li, W. Zhai, M. H. Y. Fritz, N. F. Hansen, E. Y. Durand, A. S. Malaspinas, J. D. Jensen, T. Marques-Bonet, C. Alkan, K. Prüfer, M. Meyer, H. A. Burbano, J. M. Good, R. Schultz, A. Aximu-Petri, A. Butthof, B. Höber, B. Höffner, M. Siegemund, A. Weihmann, C. Nusbaum, E. S. Lander, C. Russ, N. Novod, J. Affourtit, M. Egholm, C. Verna, P. Rudan, D. Brajkovic, Ž. Kucan, I. Gušic, V. B. Doronichev, L. V. Golovanova, C. Lalueza-Fox, M. de la Rasilla, J. Fortea, A. Rosas, R. W. Schmitz, P. L. F. Johnson, E. E. Eichler, D. Falush, E. Birney, J. C. Mullikin, M. Slatkin, R. Nielsen, J. Kelso, M. Lachmann, D. Reich, S. Pääbo, A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010). doi:10.1126/science.1188021 Medline

209. J. Kelleher, A. M. Etheridge, G. McVean, Efficient coalescent simulation and genealogical analysis for large sample sizes. *PLOS Comput. Biol.* **12**, e1004842 (2016). doi:10.1371/journal.pcbi.1004842 Medline

210. J. V. Moreno-Mayar, S. Rasmussen, A. Seguin-Orlando, M. Rasmussen, M. Liang, S. T. Flåm, B. A. Lie, G. D. Gilfillan, R. Nielsen, E. Thorsby, E. Willerslev, A.-S. Malaspinas, Genome-wide ancestry patterns in Rapanui suggest pre-European admixture with Native Americans. *Curr. Biol.* **24**, 2518–2525 (2014). doi:10.1016/j.cub.2014.09.057 Medline

211. A. Kong, M. L. Frigge, G. Masson, S. Besenbacher, P. Sulem, G. Magnusson, S. A. Gudjonsson, A. Sigurdsson, A. Jonasdottir, A. Jonasdottir, W. S. W. Wong, G. Sigurdsson, G. B. Walters, S. Steinberg, H. Helgason, G. Thorleifsson, D. F. Gudbjartsson, A. Helgason, O. T. Magnusson, U. Thorsteinsdottir, K. Stefansson, Rate of de novo mutations and the importance of father's age to disease risk. *Nature* **488**, 471–475 (2012). doi:10.1038/nature11396 Medline

212. H. Jónsson, P. Sulem, B. Kehr, S. Kristmundsdottir, F. Zink, E. Hjartarson, M. T. Hardarson, K. E. Hjorleifsson, H. P. Eggertsson, S. A. Gudjonsson, L. D. Ward, G. A. Arnadottir, E. A. Helgason, H. Helgason, A. Gylfason, A. Jonasdottir, A. Jonasdottir, T. Rafnar, M. Frigge, S. N. Stacey, O. Th Magnusson, U. Thorsteinsdottir, G. Masson, A. Kong, B. V. Halldorsson, A. Helgason, D. F. Gudbjartsson, K. Stefansson, Parental influence on human germline de novo mutations in 1,548 trios from Iceland. *Nature* **549**, 519–522 (2017). doi:10.1038/nature24018 Medline

213. T. B. Coplen, Reporting of stable hydrogen, carbon, and oxygen isotopic abundances (Technical Report). *Pure Appl. Chem.* **66**, 273–276 (1994). doi:10.1351/pac199466020273

214. M. Stuiver, H. A. Polach, Discussion Reporting of 14C Data. *Radiocarbon* **19**, 355–363 (1977). doi:10.1017/S0033822200003672

215. F. Mizuno, J. Gojobori, L. Wang, K. Onishi, S. Sugiyama, J. Granados, C. Gomez-Trejo, V. Acuña-Alonzo, S. Ueda, Complete mitogenome analysis of indigenous

populations in Mexico: Its relevance for the origin of Mesoamericans. *J. Hum. Genet.* **59**, 359–367 (2014). doi:10.1038/jhg.2014.35 Medline

216. S. Kumar, C. Bellis, M. Zlojutro, P. E. Melton, J. Blangero, J. E. Curran, Large scale mitochondrial sequencing in Mexican Americans suggests a reappraisal of Native American origins. *BMC Evol. Biol.* **11**, 293 (2011). doi:10.1186/1471-2148-11-293 Medline

217. C. M. Lewis Jr., B. Lizárraga, R. Y. Tito, P. W. López, G. C. Iannacone, A. Medina, R. Martínez, S. I. Polo, A. F. De La Cruz, A. M. Cáceres, A. C. Stone, Mitochondrial DNA and the peopling of South America. *Hum. Biol.* **79**, 159–178 (2007). doi:10.1353/hub.2007.0031 Medline

218. E. J. Lee, D. A. Merriwether, Identification of Whole Mitochondrial Genomes from Venezuela and Implications on Regional Phylogenies in South America. *Hum. Biol.* **87**, 29–38 (2015). doi:10.13110/humanbiology.87.1.0029 Medline

219. L. Arias, C. Barbieri, G. Barreto, M. Stoneking, B. Pakendorf, High-resolution mitochondrial DNA analysis sheds light on human diversity, cultural interactions, and population mobility in Northwestern Amazonia. *Am. J. Phys. Anthropol.* **165**, 238–255 (2018). doi:10.1002/ajpa.23345 Medline

220. R. S. Just, T. M. Diegoli, J. L. Saunier, J. A. Irwin, T. J. Parsons, Complete mitochondrial genome sequences for 265 African American and U.S. “Hispanic” individuals. *Forensic Sci. Int. Genet.* **2**, e45–e48 (2008). doi:10.1016/j.fsigen.2007.12.001 Medline

221. N. J. R. Fagundes, R. Kanitz, R. Eckert, A. C. S. Valls, M. R. Bogo, F. M. Salzano, D. G. Smith, W. A. Silva Jr., M. A. Zago, A. K. Ribeiro-dos-Santos, S. E. B. Santos, M. L. Petzl-Erler, S. L. Bonatto, 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 (2008). doi:10.1016/j.ajhg.2007.11.013 Medline

222. A. Hartmann, M. Thieme, L. K. Nanduri, T. Stempfl, C. Moehle, T. Kivisild, P. J. Oefner, Validation of microarray-based resequencing of 93 worldwide mitochondrial genomes. *Hum. Mutat.* **30**, 115–122 (2009). doi:10.1002/humu.20816 Medline

223. D. M. Behar, M. van Oven, S. Rosset, M. Metspalu, E.-L. Loogväli, N. M. Silva, T. Kivisild, A. Torroni, R. Villems, A “Copernican” reassessment of the human mitochondrial DNA tree from its root. *Am. J. Hum. Genet.* **90**, 675–684 (2012). doi:10.1016/j.ajhg.2012.03.002 Medline