Science Advances

AAAS

advances.sciencemag.org/cgi/content/full/2/4/e1501385/DC1

Supplementary Materials for

Ancient mitochondrial DNA provides high-resolution time scale of the peopling of the Americas

Bastien Llamas, Lars Fehren-Schmitz, Guido Valverde, Julien Soubrier, Swapan Mallick, Nadin Rohland, Susanne Nordenfelt, Cristina Valdiosera, Stephen M. Richards, Adam Rohrlach, Maria Inés Barreto Romero, Isabel Flores Espinoza, Elsa Tomasto Cagigao, Lucía Watson Jiménez, Krzysztof Makowski, Ilán Santiago Leboreiro Reyna, Josefina Mansilla Lory, Julio Alejandro Ballivián Torrez, Mario A. Rivera, Richard L. Burger, Maria Constanza Ceruti, Johan Reinhard, R. Spencer Wells, Gustavo Politis, Calogero M. Santoro, Vivien G. Standen, Colin Smith, David Reich, Simon Y. W. Ho, Alan Cooper, Wolfgang Haak

Published 1 April 2016, *Sci. Adv.* **2**, e1501385 (2016)
DOI: 10.1126/sciadv.1501385

The PDF file includes:

- Table S7. Highest predictive accuracy of the PCMLR model for each population structure model when model C is removed from the analysis.
- Table S8. Prediction of the population structure model that fits the observed data when model C is removed from the analysis.
- Fig. S1. Location of archaeological sites (see table S2 for detailed information about sites and samples).
- Fig. S2. Phylogenetic tree of newly sequenced ancient mitogenomes (haplogroup A).
- Fig. S3. Phylogenetic tree of newly sequenced ancient mitogenomes (haplogroup B).
- Fig. S4. Phylogenetic tree of newly sequenced ancient mitogenomes (haplogroup C).
- Fig. S5. Phylogenetic tree of newly sequenced ancient mitogenomes (haplogroup D).
- Fig. S6. Maximum parsimony phylogenetic trees of the 93 ancient mitogenomes and three replicate data sets of the 87 modern Native American mitogenomes (see table S3).
- Fig. S7. Extended Bayesian skyline plots of female effective population size for the three replicate data sets.
- Fig. S8. Bayesian estimates of node ages for haplogroups A2, B2, C1, D1, and D4h3a.
- Fig. S9. Results of the date-randomization test for temporal sampling structure.
- Fig. S10. PCA plot of summary statistics for the 15,000 simulations for each of the seven models simulated with BayeSSC (see Fig. 4).
- Fig. S11. PCA plot of summary statistics for the 15,000 simulations for each of the models simulated with BayeSSC, when model C is removed from the PCA.
- References (*48–86*)

Supplementary Materials and Methods

Section S1. Archaeological samples, radiocarbon dating, and DNA contamination control

Archaeological samples and radiocarbon dating

Metadata about all archaeological samples can be found in fig. S1 and table S2, including site location, actual collection location, cultural assignation, and specimen and sample description.

Permissions to collect, export and analyze ancient specimens were granted by the relevant government organizations. In Peru, the Ministry of Culture (the former National Institute of Cultural Heritage—INC) delivered the permits Resolución Directional Nacional No. 1346 and CR-8907, and Resolución Viceministerial No. 120-2010-VMPCIC-MC, 474-2011-VPMCIC-MC and 545-2011-VMPCIC-MC. In Bolivia, the National Institute of Archaeology, a specialized agency of the Deputy Minister of Culture, delivered the permit UDAM–Autorización 015/2012. In Chile, a formal research agreement was established between the Corporación Regional de Desarrollo Científico y Tecnológico del Hombre en el Desierto (CODECITE) and the Australian Centre for Ancient DNA. In Mexico, permits for DNA analysis and export of ancient human remains were granted by the Consejo Arqueológico and Consejo Jurídico committees of the National Institute of Anthropology and History (INAH)—Oficio Número 401-3-4319 and Expediente AA-19-09 A/2134, respectively. In Argentina, the Cultural Patrimony Office of the Ministry of Education provided the permit #07-1999.

The two samples from Arroyo Seco 2, Argentina (B9S14 and B2S19 in table S2) were radiocarbon dated as part of a previous study (*40*) (table S1). Samples from Lauricocha (ACAD12686–9 in table S2) were sent to the Curt Engelhorn Center for Archaeometry–MAMS for AMS radiocarbon dating as part of a reexamination of the Lauricocha archaeological site (*39*) (table S1). Eight samples from the Huaca Pucllana archaeological site, Lima, Peru (ACAD10709, 10722, 10734, 10754, 10791, 10805, 10810, 10817) and two samples from Tiwanaku, Bolivia (ACAD13241, 13264) were sent to the Oxford Radiocarbon Unit for AMS radiocarbon dating in order to confirm the archaeological assignation based on stratification and examination of artifacts (table S1).

Contamination control

All samples were collected using surgical gloves, a facemask, and a full-body suit (or alternatively a cap/hair net and protective long-sleeved clothes). Working surface and tools were decontaminated with 5% bleach before each sample was collected. The large majority of samples were teeth extracted directly from their alveolar sockets. The few bone samples were collected using a Dremel tool and diamond cutting discs in a well-ventilated room.

All ancient DNA work was conducted in clean-room facilities at the University of Adelaide's Australian Centre for Ancient DNA (ACAD), the University of California at Santa Cruz's Human Paleogenomics lab, and the Reich lab at Harvard Medical School (HMS). During DNA extractions, multiple extraction blank controls were included for every sample batch analyzed.

Capture hybridization and sequencing of 21 samples (ACAD10709, 10722, 10725, 10729, 10734, 10754, 10773, 10774, 10800, 10805, 10814, 10817, 12686–12690, 13240, 13241, 13264, 13272) were independently replicated at ACAD and HMS (table S2).

Section S2. DNA Extraction, library preparation, and mitochondrial capture (ACAD)

Bone and tooth samples were UV irradiated (260 nm) on all surfaces for 30 min. Sample surface was wiped with 3% bleach, then ~1 mm was removed using a Dremel tool and abrasive disks. Each sample was ground to a fine powder using a Mikro-Dismembrator (Sartorius). DNA extraction was based on the protocol from Brotherton et al. 2013 (*22*). For each individual, 200 mg of powder were demineralized overnight under constant rotation at 37°C in 4 mL of 0.5M EDTA, pH 8.0. The next day, lysis was achieved by adding 1.2 mg of proteinase K and incubating for 2 hours under constant rotation at 55°C. After lysis, samples were centrifuged at 4,600 rpm for 1 min and the supernatant was mixed with 125 µL of medium-sized silica suspension and the following modified binding buffer: 13.5 mL QG buffer (Qiagen), 2.86 mL of 1X Triton (Sigma-Aldrich), 20 mM NaCl (Sigma-Aldrich), 0.2 M ammonium acetate (Sigma-Aldrich). DNA was left to bind to silica for 1 hour at room temperature under constant rotation. Silica particles were pelleted by centrifuging at 4,600 rpm for 1 min and the supernatant was discarded. The pellet was washed three times by adding 1 mL 80% ethanol, centrifuging for 1 min at 13,000 rpm and discarding the supernatant. The pellet was left to dry for 30 min, resuspended in 200 µL of pre-warmed TE buffer (10 mM Tris, 1 mM EDTA) with 0.05% Tween-20, and incubated for 10 min. After centrifugation for 1 min at 13,000 rpm, the supernatant was aliquoted and stored at -18°C until further use.

Double-stranded Illumina libraries were built from 20 µL of each DNA extract following the protocol from Meyer & Kircher 2010 (*42*), using truncated Illumina adapters with dual 5-mer internal barcodes (*41*). No DNA repair was performed when preparing the DNA libraries. Most libraries were amplified by PCR using AmpliTaq Gold (Life Technologies), but some libraries were amplified using isothermal amplification (TwistAmp® Basic kit, TwistDx Ltd) towards the end of the study (see table S2). Detailed amplification protocols as in (*41*).

Mitochondrial RNA baits for DNA capture by hybridization were prepared in-house. In a first step, the whole mitochondrial genome (WMG) of BL (haplogroup H3k1) was amplified in three overlapping fragments using the Expand Long Range dNTPack kit (Roche), and following the manufacturer's protocol with an annealing temperature of 60ºC and the primers in table S4. From this step onward, all commercial kits were used according to the manufacturer's protocols. For each amplicon, one of the primers was 5'-tailed with a T7 promoter sequence (5'–AATTGTAATACGACTCACTATAGGG–3') in order to perform an *in vitro* transcription using the T7 High Yield RNA Synthesis Kit (New England BioLabs). DNA was degraded with DNase after the *in vitro* transcription was completed. Resulting RNA was purified using the Ambion MEGAclear kit (Life Technology), and eluted in 2 x 50 µL of RNase-free molecular grade water. RNA was quantified using a Nanodrop spectrophotometer (FisherScientific), and RNA integrity was assessed using a TapeStation (Agilent). For each mitochondrial fragment, 40 µg of RNA was then fragmented using the NEBNext Magnesium RNA Fragmentation Module (New England BioLabs). Fragmented RNA was purified using the RNeasy Minelute Cleanup kit (Qiagen), and eluted in 15 µL of RNase-free molecular grade water. RNA was quantified using a Nanodrop spectrophotometer, and RNA fragmentation was assessed using a TapeStation. Finally, fragmented RNA was randomly biotinylated using the

EZ-Link Psoralen-PEG3-Biotin kit (Thermo Scientific), performing as many assays as was possible given the concentration of starting template. Biotinylated RNA was purified using the RNeasy Minelute Cleanup kit (Qiagen), and eluted in 15 µL of RNase-free molecular grade water. Biotinylated RNA baits from the initial three mitochondrial fragments were pooled at a concentration of 50 ng/µL each (150 ng total) for capture by hybridization.

For the capture assay, 200–250 ng of barcoded DNA library were mixed with 2.5 µg of salmon sperm DNA (Invitrogen) and 2.5 µg of Cot1 DNA (Invitrogen) in a volume of 8 µL. We prepared a RNA mix consisting of 800 ng of biotinylated RNA baits, 25 pmol of blocking RNA oligonucleotides (matching the sequence of the truncated Illumina P5 and P7 adapters), and 20 U of SUPERase-In (Ambion) in a volume of 6.5 µL. The hybridization buffer contained 5X SSPE, 10 mM EDTA pH=8.0, 10X Denhart's solution, and 0.2% SDS in a volume of 18.4 µL. The DNA mix was denatured at 94ºC for 5 min. Hybridization buffer, RNA and DNA were incubated at 65ºC for 5 min separately, then mixed and incubated as follows: 65ºC for 9h, 63ºC for 9h, 61ºC for 9h, 59ºC for 9h, 57ºC for 9h, 60ºC for 3h.

For each hybridization reaction, 50 µL of Dynabeads® MyOne Streptavidin C1 (Life Technologies) were washed two times with 500 µL of 2X SSC + 0.05% Tween-20 at room temperature. Beads were then incubated with 100 µg of yeast tRNA in 500 µL of 2X SSC + 0.05% Tween-20 for 30 min under constant rotation at room temperature, to saturate all sites on the beads that bind nucleic acids in a non-specific manner. Beads were then washed with 500 µL of 2X SSC + 0.05% Tween-20 at room temperature and resuspended in 200 µL of 2X SSC + 0.05% Tween-20. The hybridization reaction was incubated with the beads for 30 min under constant rotation at room temperature to allow binding of the biotin to the streptavidin. A series of increasingly stringent washes was then performed: one wash with 500 µL of 2X SSC + 0.05% Tween-20 for 10 min at room temperature, two washes with 500 µL of 0.75X SSC + 0.05% Tween-20 for 10 min at 60ºC, and one washes with 500 µL of 0.2X SSC + 0.05% Tween-20 for 10 min at 60ºC. After the last wash, beads were directly used for off-bead PCR amplification, using full-length 7-mer indexed Illumina adapters as described in Meyer & Kircher 2010 (*42*).

Indexed libraries (with the exception of ACAD10730 and ACAD10794, see below) were pooled and sequenced on one lane of Illumina HiSeq2000 with the v3 chemistry, using 2x100 cycles, at the Australian Cancer Research Foundation (ACRF) Cancer Genomics Facility, Adelaide, South Australia. Reads were automatically demultiplexed by index using Casava and a strict matching of the 7-mer indices. An additional demultiplexing based on a strict matching of the two 5-mer internal barcodes and trimming of the barcodes were performed using the program Sabre 1.0 (https://github.com/najoshi/sabre). AdapterRemoval v1.5.2 (*48*) was used with default parameters to trim residual adapter sequences, merge overlapping reads, and discard reads shorter than 25 nt. Merged reads were aligned to the Reconstructed Sapiens Reference Sequence (RSRS) sequence (*49*) with bwa 0.7.5a-r405 (*50*), using the parameter space recommended for ancient DNA (no seed, one gap opening, relaxed edit distance) (*51*). Duplicate reads were removed using FilterUniqueSAMCons.py (*52*) and final pileup statistics were calculated using SAMtools. Characteristic damage patterns were assessed using MapDamage v0.3.6 (*53*).

Sample ACAD10730 was sequenced on an Ion Torrent PGM at ACAD and sequencing data were analyzed as described in (*54*). Sample ACAD13245 was sequenced on an Ion Torrent PGM at ACAD as well as on a HiSeq2000 at the ACRF Cancer Genomics Facility. Data from both

sequencing runs were mapped independently and pileup files were merged to remove duplicates.

Finally, sample ACAD10810 was not used for mitochondrial capture. Instead, shotgun sequencing was performed on two lanes of HiSeq2000 with the v2 chemistry, using 1x50 cycles, at the Beijing Genomics Institute (BGI), Shenzhen, Guangdong, China. Processing and mapping of reads were performed as described above.

Information on sequencing coverage and damage estimates is in table S2.

Section S3. DNA extraction and library preparation (UCSC Human Paleogenomic laboratory and HMS Reich laboratory)

DNA extraction was based on the protocol from Dabney et al. 2013 (*55*) using 75 mg of bone or tooth powder. DNA was eluted twice in 16–30 µL 1X TE buffer (with 0.05% Tween-20). Aliquots were stored at -18°C until further use.

Double-stranded Illumina libraries were built from 30 µL of each DNA extract following the protocol from Rohland et al. 2014 (*56*), using truncated Illumina adapters with dual 7-mer internal barcodes. Depending on the libraries, DNA was either not repaired, subject to partial UDG treatment (*56*), or subject to USER repair (New England Biolabs) (see table S2 for details). DNA libraries were then sent to the Reich lab at HMS for mitochondrial capture by hybridization and sequencing.

Section S4. Mitochondrial DNA capture (HMS Reich laboratory)

Hybridization enrichment was performed as previously described (*57*) with DNA baits targeting the human mitochondrial genome (3 bp tiling based on NC_001807 as in (*58*)) using a semi-automated protocol in a 96-well plate setup on an Evolution P3 (Perkin Elmer, Waltham, USA) for two consecutive rounds. For each hybridization reaction, we used 500 ng of single-stranded bait library together with 500 ng of DNA library. The oligonucleotide blockers are specified in (*56*). All other parameters of the hybridization, capture, and washing steps and amplifications can be found in the original article (*57*); hybridization and washing temperatures for stringent washes were lowered by 5°C over the course of the study. Each sample was subsequently indexed prior to sequencing with a unique index (pair), using 7-mer index sequences as in Meyer & Kircher 2010 (*42*).

Indexed libraries were pooled and sequenced on a MiSeq or a NextSeq500 instrument at HMS with the v2 chemistry, using 2x75 cycles or 2x76 with the standard TruSeq HT protocol that reads both indices. Reads were automatically demultiplexed by index using Illumina BaseSpace allowing up to one mismatch per index. SeqPrep was used to trim residual adapter sequences and merge reads that had at least 15 overlapping nucleotides (allowing one mismatch). SeqPrep was modified to a more conservative setting to ensure that quality scores in the merged regions use the best score rather than aggregating the two inferred reliabilities of the base call. An additional demultiplexing based on the two 7-mer internal barcodes (allowing up to two mismatches) and trimming of the barcodes were performed. Merged reads were aligned to i) the human reference genome (hg19) which contains the revised

Cambridge Reference Sequence (rCRS) mitochondrial genome, and ii) the Reconstructed Sapiens Reference Sequence (RSRS) sequence (*49*), with bwa 0.6.1-r104 (*50*). Final target coverage was calculated using BEDtools v2.16.2 (*59*). Characteristic damage patterns were assessed using MapDamage2.0 (*60*). Mitochondial DNA contamination was estimated using a Markov Chain Monte Carlo based estimator (SI 5 in (*21*)). More precisely, a consensus sequence was built using a minimum base quality of 30 and a minimum coverage of 5, stripping gaps and ignoring any spurious heterozygote positions. All reads were realigned to this consensus, and the resulting alignments were used for estimation of contamination (SI 5 in (*21*)), trimming the first and last 5 bases from every read to minimize errors due to ancient DNA degradation.

Information on sequencing coverage, damage estimates, and contamination ratio is given in table S2.

Section S5. Bayesian phylogenetic analysis of the mitochondrial data

Ancient sequences

Reads pileups were visualized in Geneious v7.1.3 (Biomatters; available from http://www.geneious.com/). SNPs were called in Geneious for all polymorphisms with minimum coverage 2 and a minimum variant frequency 0.7. The assembly and the resulting list of SNPs were verified by eye and compared to SNPs reported at phylotree.org (mtDNA tree Build 16 [19 Feb 2014]) (*49*). Following recommendations in van Oven & Kayser 2009 (*24*) we excluded common indels and mutation hotspots at nucleotide positions 309.1C(C), 315.1C, AC indels at 515–522, 16182C, 16183C, 16193.1C(C), and C16519T.

We embedded the ancient WMGs in the existing mitochondrial tree (mtDNA tree Build 16 [19 Feb 2014]). This was done manually by following the hierarchical structure of the tree up to the most derived SNP shared with an existing sequence haplotype. Additional SNPs were considered to be private mutations. All ancient sequences represent unique branches along the stem of Native American haplogroups A2, B2, C1, and D1 (figs. S2–5).

Modern sequences

A total of 370 WMGs corresponding to Native Central and South American populations were collected from published studies. Of note, the large dataset of 215 Mexican WMGs from Kumar et al. 2011 (*61*) was not used to prevent an over-representation of Mexican samples.

As Bayesian analysis of more than 400 WMGs would be too computationally intensive we randomly selected three replicate datasets of 87 sequences, all of which contained exclusively the five Native American haplogroups A2, B2, C1, D1, and D4h3a (table S3). We added 20 East Siberian individuals from sister lineages within haplogroups A, B, C and D in order to estimate the time of the split between Siberian and American lineages. We also added the 92 ancient WMGs from this study, as well as the Anzick-1 mitogenome (*45*), to obtain three replicate datasets of 200 sequences. The Anzick-1 WMG was included as it added a deep time calibration at 12.6 kya, was the only ancient D4h3a sequence available, and because the Anzick-1 individual can reasonably be considered a descendant of the first wave of colonizers.

Sequence alignment

Sequences were aligned to RSRS (*49*) using the program Geneious v7.1.3 (Biomatters; available from http://www.geneious.com/) and adjusted manually. We excluded common indels and mutation hotspots at nucleotide positions 309.1C(C), 315.1C, AC indels at 515–522, 16182C, 16183C, 16193.1C(C), and C16519T (*24*).

Partitioning of the datasets

The partitioning scheme and the substitution model for each replicate dataset were simultaneously selected using the Bayesian information criterion as implemented by PartitionFinder v1.1.1 (*62*). Nine candidate partitions were considered: all three codon positions of the protein-coding genes (CP1, CP2, CP3) and ND6 (ND6_CP1, ND6_CP2, ND6_CP3), tRNA genes, rRNA genes, and the D-loop. The best scheme identified by PartitionFinder included five partitions: CP1 combined with rRNA and tRNA, CP2, CP3, all three codon positions of ND6 combined, and the D-loop. Since there is no *a priori* biological support for the combination of CP1 and the RNA genes, we separated CP1 and tRNA/rRNA in two partitions. This led to the following scheme of six partitions: CP1, CP2, CP3, all three codon positions of ND6 combined, rRNA and tRNA genes combined, and the D-loop.

Molecular dating analyses

A Bayesian approach was used to estimate coalescence times from the three replicate WMG datasets, as implemented in BEAST 1.8.0 (*43*). The ages of the ancient sequences were used as tip calibrations in the phylogenetic analysis. When available, the calibrated radiocarbon date was used as the tip date (table S1). For the majority of samples, the midpoint of the relevant archaeological period was used as tip date (table S1). We used point estimates for tip calibrations to avoid accounting for the estimation error in each of the 92 tip dates, and over-parameterizing the Bayesian coalescent analysis. Studies of simulated and real data have shown that accounting for uncertainty in tip dates has little impact on the resulting estimates (*63*, *64*).

For each replicate dataset, the six partitions were assigned independent substitution models as identified by PartitionFinder, independent clock models, and a shared genealogy. Independent uncorrelated lognormal relaxed clocks were tested for all partitions in preliminary analyses to account for potential rate variations between different branches of the tree (*65*). However, strict clocks could not be rejected for any of the partitions in all three replicate datasets, and independent strict clocks were used in the subsequent analyses. Based on the estimates of coalescence times, the demographic history was reconstructed using the extended Bayesian skyline method (*44*), which infers the effective population size through time and also estimates the number of demographic changes from the data. Three Markov Chain Monte Carlo (MCMC) runs of 100 million steps were performed for each replicate dataset, with sampling of parameters every 10,000 steps. Parameter traces were monitored in Tracer v1.6 to ensure convergence of the MCMC runs and effective sample sizes higher than 200. The initial 10 million steps were discarded as burn-in before sampled trees and parameter traces from the three independent runs were combined to summarize the results. The demographic plot was reconstructed using a Java program (*35*) and visualized using an in-house R script, using a generation time of 25 years (Fig. 3B and fig. S7). Node ages, or the times to most recent common ancestor (TMRCAs), were estimated by BEAST (fig. S8). The three replicate datasets produced very similar estimates of the tree and demographic history.

We used replicate dataset 1 for illustration (Fig. 3 and fig. S7) as well as for the additional tests and calculations detailed below.

Date randomization test

A date randomization test was conducted on replicate dataset 1 to verify that the number and the temporal distribution of dated sequences were sufficiently informative to calibrate estimates of molecular rates (*66*). Molecular rate estimates from the original data (with the true sample dates) and ten analyses with randomized dates were compared for the six mitochondrial partitions. For all partitions, the mean posterior rate estimate from the original dataset was not included in any of the 95% highest posterior density (HPD) intervals from the date-randomized datasets, indicating the test was passed successfully (fig. S9). However, the 95% HPD intervals did not reach extreme low values and resembled the results obtained for the bowhead whale data reported in (*66*), indicating a potential low sequence diversity. When the same date randomization test was conducted on the unpartitioned WMGs, the test was again passed successfully and 95% HPD intervals reached extreme low values, in a fashion similar to successful tests in (*66*) (fig. S9).

Studies of simulated data have shown that the date randomization test generally performs well in identifying data sets that lack temporal signal (*67*). Our rate estimate from whole mitochondrial genomes also passes the date randomization test based on the more conservative criterion proposed by Duchêne and colleagues (*67*), which calls for non-overlapping 95% HPD intervals between the rates estimated from the original and date-randomized data sets.

Estimation of the start of the separation between Siberian and Native American ancestral populations

Based on the assumption that the divergence between Siberian and Native American ancestral populations is a unique event, we could use the 95% HPD intervals to obtain a more precise estimate of the timing of the event. For each replicate dataset, we recorded the latest point at which any of the haplogroups A, B, C or D coalesced for all steps in the MCMC. Hence, we obtained only one sample of deepest coalescence times for all haplogroups, avoiding the need for any post hoc adjustments, and then calculated the 95% HPD for that posterior sample using the *HPDinterval* function in R (*47*). We could estimate the last divergence started around 24.9kya (Fig. 2B). This last genetic divergence marks a decrease in gene flow between ancestors of Native Americans and their Siberian counterparts. It is likely that the real population divergence occurred after 24.9kya, making this point a minimum estimate for the isolation of Native American ancestral populations.

Precise estimation of the Native American node ages and consequences for the timing of the population isolation

We observed that the node ages for the Native American founder haplogroups A2, B2, C1, D1, and D4h3a were highly synchronous, as previously reported (*25*). Based on the assumption that the TMRCA for each of the Native American haplogroups provides an independent estimate of the isolation time of a unique small population, we could use the 95% HPD intervals to obtain a more precise estimate of the timing of the event. We chose not to combine the posterior samples for the TMRCA for each Native American haplogroup, as the TMRCA for each haplogroup targets the time until a unique individual. However, by finding

the intersection of upper and lower 95% HPD intervals for each haplogroup, we were able to identify a time interval that allowed for a synchronous radiation event. That is, the lowest and highest bounds of the TMRCA estimates could be defined as the highest lower 95% HPD bound and the lowest higher 95% HPD bound among the five Native American haplogroups, respectively (Fig. 2A).

Alternatively, using the same method as for the Siberian/Native American divergence, we were able to refine the time interval that allowed for a synchronous radiation event. That is, the lowest and highest bounds of the TMRCA estimates could be defined as the highest lower 95% HPD bound (18.4 kya) and the lowest higher 95% HPD bound (12.5 kya, constrained by the age of the Anzick-1 remains) among the five Native American haplogroups, respectively. As a consequence, the lowest bound of the TMRCA estimates could also be used as a proxy for the complete isolation of the Native American ancestral population from their Siberian counterparts (18.4kya; Fig. 2B).

Section S6. Ecological assessment of population sustainability in Beringia

Whether the ancestors of the Native Americans resided in Beringia for a prolonged period of time before entering the Americas or were just sojourners is highly dependent on the potential of the Beringian biome to sustain human populations during the Last Glacial Maximum (LGM). To calculate net productivity and carrying capacity of Beringia, we need to take into account the interplay of several factors such as the characteristics of the ecosystem, the climatic factors, and latitude. We also need to know the human population size during the initial peopling process of the Americas as well as their economy and social structure. Based on the archaeology, it is known that the first people entering the Americas were bands of hunter-gatherers presumably living at low population density (*1*). In the present study, the effective population size that entered the Americas is estimated to be ~2,000 females (Fig. 3C). Although this number cannot be directly translated into a census size, it suggests that the population residing in Beringia during the standstill was relatively small, probably not exceeding a few tens of thousands of people. The presence of large numbers of megafauna in eastern Beringia during the late Pleistocene, including the LGM, indicates an ice-free region. Recent archaeological studies also found that climatic conditions in Beringia must have been relatively mild during the LGM compared with the adjacent Siberian and Alaskan regions, and the landscape was dominated by shrub tundra, offering a rich fauna and a likely source of wood fuel to its inhabitants (*29*).

The net productivity and carrying capacity of the prehistoric Beringian ecosystem can be inferred from modern tundra environments. These include a high diversity of vegetation classes with a variable above-ground biomass production of ~100–4000 g/m^2 (*68*) sustaining a faunal biomass of 440–800 kg/km^2 (*69*). To the extent of our knowledge, no reports exist on the hunter-gatherer vegetable diet in tundra environments, and thus we cannot estimate the nutritional calorie gain using vegetable dietary components. As a consequence, the available reports of primary nutritional calorie estimates in a tundra environment (~400–700 kcal/m^2/year) rely on hunting activity (*69*). Paleobotanic evidence indicates that central Beringia was dominated by shrub tundra (*29*). These resulting biomass estimates are based on undifferentiated, less productive tundra environments (*69*) and the actual carrying capacity of the biome may have been higher. The existence of megafauna in Beringia during the LGM might also have constituted a much higher faunal resource than in modern tundra

biomes. Thus, it is likely that the carrying capacity of Beringia is currently being underestimated.

To calculate a maximum availability of nutritional calories via hunting activities, we used a model of sustainable hunting where only 5% of existing prey will be extracted to allow for a natural replenishment of stock lost through disease, predation, and age (*70*, *71*), and only 50% of the animal carcass has food value. Ethnographic and ecological studies show that under these considerations hunter-gatherers could have extracted a maximal sustainable yield of 11–20 kg /km^2/year of meat weight from a tundra environment, with each kilogram providing about 2,700 kcal (*69*, *70*). Thus, 1 km^2 could supply approximately 54,000 kcal/year. Based on activity patterns in foraging societies, climate, and latitude, we assumed a daily calorie usage per individual of ~2,900 kcal (*72*). If resources and human individuals were equally distributed in the ecosystem and all of the human diet was based on animal meat, that would mean that ~18 km^2 of tundra could sustain one human per year, or a sustainable population density of 0.05 individuals/km^2/day. Based on ethnographic observations, it is unlikely that the diet of the inhabitants of Beringia was entirely based on terrestrial animal meat, and at least 20% of the calorie intake must have come from plants and marine resources. Thus, we considered a diet based on 80% terrestrial animals and estimated a final sustainable population density of 0.06 individuals/km^2.

The complete landmass of Beringia spanned about 1,502,200 km^2. Resources throughout Beringia were probably not distributed equally, especially wood fuel (*29*), suggesting that specific areas were preferentially used by humans. This correlates well with the observation that hunter-gatherer groups in tundra environments often exhibit aggregated distribution patterns (as opposed to uniform distribution) across space (*72*). The relationship between resource availability and spatial use links to a suggested optimum population of 20–60% of the theoretical maximum for hunter-gatherers (*72*). This results in an estimated minimum sustainable population of 18,000–54,000 humans in Beringia, which sits well with the founding effective population size from our Bayesian demographic reconstruction when applying a conventional N_e/N ratio of 0.11 (*73*, *74*). In combination with the local abundance of wood fuel (*29*), the carrying capacity estimates strongly support a scenario in which human populations lived in the Beringian biome for several thousand years before entering the Americas.

Section S7. Bayesian coalescent simulations and evaluation of demographic scenarios

We observed from the mitogenomic tree (Fig. 3B and figs. S2–S6) that ancient and modern sequences form clearly separate clades with long branches within each of the major haplogroups, and that none of the ancient lineages was shared with or gave rise to lineages sampled from present-day individuals. We hypothesize that this pattern could be explained by: i) drastic under-sampling of both modern and ancient individuals; ii) poor geographic overlap between ancient and modern sampling locations; or iii) the result of an early differentiation into geographic regions (with minimal gene flow), in which the Native American lineages diversified further. Upon inspection of the geographic locations/origins of ancient and modern samples, respectively, we noticed very little overlap between regions. However, while this implies that our samples were drawn from different geographic regions, it also suggests that ancient lineages from particular regions went extinct and were likely to have been replaced by surviving lineages from other regions, which led to the diversity observed today.

In an effort to formally test whether or not the data had come from one large, continuous, panmictic population, or a population comprising two separate sub-populations representing geographic differentiation, we used BayeSSC (*36*) to simulate data for seven population scenarios. We used best estimates of model parameters (population size and substitution model parameters) resulting from the extended Bayesian skyline statistics as implemented in BEAST (see previous section) to set up the parameter files for seven alternative models, as shown in Fig. 4.

- In line with Fig. 3, model A assumes one panmictic population of constant size (2000 females), which undergoes a strong exponential growth starting 600 generations ago (15,000 years ago assuming a generation time of 25 years) and resulting in 100,000 females 500 generations ago, after which the population size remains constant until the present day (time 0).
- Model B builds on model A, but accounts for a drastic bottleneck 20 generations ago as a consequence of the contact with European colonizers. This model tests whether the loss of ancient lineages could have been random and explained by an effective population size reduction of 50%.
- Model C also assumes one panmictic population of constant size (2,000 females), and the same growth between 500–600 generations ago as in Model A. Following the massive increase in Ne we assumed a split into two demes (0 and 1) of equal size (50,000 females each with no migration between demes) around 360 generations ago (~9,000 years ago or the last time we observe an ancestor common to modern and ancient lineages in hg A2; black triangle in Fig. 3B), and a constant population size for both ('geographically') isolated demes. Ancient lineages were exclusively sampled from deme 1 by converting radiocarbon or archeological dates into number of generations. To account for an extinction of ancient lineages as a consequence of the contact with European colonizers, we modeled a drastic population decline from 50,000 to 1 female (virtually extinct) for deme 1 between 20 generations ago (~1500 AD) and to the present day (time 0). In contrast, we assumed an exponential growth for deme 0 to 100,000 females over the last 20 generations to account for the present-day mitochondrial diversity.
- Model D builds on model C, but with survival of ancient lineages to the present day, meaning that both deme 0 and deme 1 maintain a constant effective population of 50,000 females.
- Model E builds on model D, but accounts for a drastic bottleneck (effective population size reduction of 50% in both demes) 20 generations ago as a consequence of the contact with European colonizers. This model serves as a 'true' null hypothesis for model C and tests whether the loss of ancient lineages could have been random and explained by a combination of population structure and a bottleneck.
- Model F builds on model C, but with a mild and progressive loss of lineages starting when demes split. Then the extinction of ancient lineages as a consequence of the contact with European colonizers European is modeled as a drastic population decline from 35,000 to 1 female for deme 1.
- Model G builds on model F, but with a progressive loss of lineages starting when demes split until the present. There is no dramatic effect of the contact with European colonizers.

All seven models were simulated using 15,000 genealogies in BayeSSC (*36*) with the following parameters of mitochondrial sequence evolution for all simulations: a generation time of 25

years, a fixed mutation rate of 0.01023 substitutions per mitogenome per generation (or 2.4694×10^{-8} substitutions per site per year), a transition/transversion ratio of 0.9841 (*36*), and a gamma distribution of rates with parameter values of 0.05 (theta) and 6 (k) (*22*).

We then calculated and recorded summary statistics from our observed data (same three replicates of 93 ancient and 87 modern sequences used for the BEAST analysis) for each genealogy, using a custom script and the complete suite of within-deme (Tajima's D, haplotype diversity, number of haplotypes, etc.) and between-deme (Fst, average pairwise distances) statistics from BayeSSC, and the same parameters and sampling times as for the simulations.

The Akaike information criterion for model selection implemented in the BayeSSC package estimates the quality of the models relatively to each other, but it does not provide any information about the absolute quality of the models. To estimate both the relative and the absolute qualities of the tested models, we used an approach based on principal component estimation of functional logistic regression (*37*). First, we performed a principal component analysis to remove co-linearity between the summary statistics, since they are likely to be correlated (*37*, *46*), using the *princomp* function in R (*47*). Upon inspection of the within-component variances (scree plot), we retained the first five principal components. We then fit a multiple logistic regression model in which the response variable was the model under which the data were simulated, and the explanatory variables were the first five principal components. A full interaction model was fit, and a backward step AIC method was used to find the best predictive model. We avoided the use of likelihood ratio scores, as the population models were not nested in any sense. Instead we employed a Bayesian model verification step, which is both robust against sensitive likelihood calculations and in concordance with the Bayesian analysis performed in BEAST.

The accuracy of the principal component multiple logistic regression (PCMLR) model in the selected principal components is reported in Table S5: the higher the accuracy value, the better the PCMLR model can predict a given population structure model. The accuracy was extremely high (99.87%) for model G because the cluster overlaps only marginally with the other clusters. However, the overall quality of model G was poor as none of the observed data points was close to the model G cluster (Fig. 4 and fig. S10). The second best accuracy was for demographic model C: the PCMLR model could predict model C at 75.35% in the principal components, despite a very strong overlap with clusters for population structure models A, B, D, E and F (Fig. 4 and fig. S10).

We took the three observed data sets and transformed them to obtain the principal components, and using the fitted PCMLR model, predicted which of the population structure models the data came from. In all cases, the data were most likely to have come from a population with two separate sub-populations and a rapid and complete extinction of one of the sub-populations after the European colonization (model C, minimum probability 0. 995) (Table S6, Fig. 4, and fig. S10).

If we removed the demographic model C from the analysis, the principal components remained mostly unchanged with respect to direction, though the scale was altered marginally (fig. S11). This observation supports the robustness of the PCMLR model in

correctly discriminating between the different demographic models (*38*). However, the highest predictive accuracy was rather low among all population models, except for model G (Table S7), which was still overlapping only marginally with other clusters (fig. S11). In addition, the PCMLR model did not predict all observed data points with strong probability in the absence of demographic model C (maximum probability 0.865; Table S8), and the observed data points were remotely located from any population model cluster (fig. S11). Most importantly, the PCMLR model did not agree on the same demographic model for all observed data points when model C was removed (Table S8). This is further evidence that model C strongly fit the observed data.

Supplementary Tables

table S1. Date calibrations for ancient mitogenome sequences.

Source	N	Minimum age	Maximum age	Mean age (ya)
Pueblo Viejo	11	1438 AD	1572 AD	509
Llullaillaco	2	1430 AD	1520 AD	539
San Sebastian	2	600 BP	500 BP	550
Cueva Candelaria	5	1000 AD	1600 AD	714
Huaca Pucllana (Ychsma)	15	1100 AD	1440 AD	744
ACAD10709$		*706 calBP*	*662 calBP*	*684*
ACAD10722$		*729 calBP*	*672 calBP*	*701*
ACAD10805$		*727 calBP*	*670 calBP*	*699*
ACAD10810$		*801 calBP*	*701 calBP*	*751*
Pica-8 (PO3, PO4)	2	1000 BP	500 BP	750
Pica-8 (ACAD11116, 11118, 11121)	3	1000 AD	1476 AD	776
Chancay	7	1000 AD	1470 AD	779
Chullpa Botirigiayocc (BC12)	1	800 BP	800 BP	800
Chullpa Botirigiayocc (ACAD12690)	1	850 AD	850 AD	1164
Huaca Pucllana (Wari)	10	500 AD	1000 AD	1264
ACAD10734$		*1174 calBP*	*982 calBP*	*1078*
ACAD10754$		*976 calBP*	*730 calBP*	*853*
Tiwanaku	9	400 AD	1000 AD	1314
ACAD13241$		*1049 calBP*	*926 calBP*	*988*
ACAD13264$		*933 calBP*	*795 calBP*	*864*
Huaca Prieta	1	1800 BP	1000 BP	1400
Palpa	1	325 AD	440 AD	1632
Huaca Pucllana (Lima)	9	100 AD	650 AD	1639
ACAD10791$		*1366 calBP*	*1290 calBP*	*1328*
ACAD10817$		*1416 calBP*	*1308 calBP*	*1362*
Jauranga	1	1700 BP	1600 BP	1650
Los Molinos	1	2000 BP	1800 BP	1900
Lauricocha (LAU3)*	1	3631 calBP	3513 calBP	3572
Camarones	1	4000 BP	3500 BP	3750
La Galgada	3	4000 BP	4000 BP	4000
Lauricocha (LAU4)*	1	5936 calBP	5905 calBP	5921
Arroyo Seco 2 (B9S14)*	1	7928 calBP	7581 calBP	7755

Arroyo Seco 2 (B2S19)*	1	7928 calBP	7591 calBP	7760
Lauricocha (LAU1)*	1	8589 calBP	8482 calBP	8536
Lauricocha (LAU2)*	1	8589 calBP	8482 calBP	8536
Lauricocha (LAU5)*	1	8700 calBP	8599 calBP	8650

*: these sequences have direct radiocarbon dates, all other dates are based on the relevant archaeological period. calBP: calibrated radiocarbon years before present (or calendar *years before* 1950 AD); ya: years ago. See fig. S1 and table S2 for details about geographical location, archaeological information, samples description and sequencing results. $: The radiocarbon dates are only shown to confirm the cultural assignation; the relevant archaeological period was used as date calibration for all Huaca Pucllana and Tiwanaku samples.

table S2. List of ancient samples, associated metadata, and sequencing results (separate Excel document).

table S3. Complete list of accession numbers for modern mitogenome sequences.

Accession Number	hg	Group	Origin	Ref
AF347012	C1d1	1,3	Warao, Venezuela	(*75*)
AF347013	C1d1	2	Warao, Venezuela	(*75*)
AY519494	B1	1,2,3	Tubalar, Siberia	(*76*)
DQ282389	A2	2	Hispanic American, USA	(*77*)
DQ282391	A2	3	Hispanic American, USA	(*77*)
DQ282393	A2	2,3	Hispanic American, USA	(*77*)
DQ282398	A2	3	Hispanic American, USA	(*77*)
DQ282402	A2	3	Hispanic American, USA	(*77*)
DQ282404	A2	2	Hispanic American, USA	(*77*)
DQ282405	A2	2	Hispanic American, USA	(*77*)
DQ282406	A2	1	Hispanic American, USA	(*77*)
DQ282407	A2	2,3	Hispanic American, USA	(*77*)
DQ282408	A2	1	Hispanic American, USA	(*77*)
DQ282409	A2	1,3	Hispanic American, USA	(*77*)
DQ282410	A2	1	Hispanic American, USA	(*77*)
DQ282411	A2	2	Hispanic American, USA	(*77*)
DQ282414	A2	1	Hispanic American, USA	(*77*)
DQ282415	A2	1,3	Hispanic American, USA	(*77*)
DQ282416	A2	1,2,3	Hispanic American, USA	(*77*)
DQ282417	A2	2	Hispanic American, USA	(*77*)
DQ282418	A2	1	Hispanic American, USA	(*77*)
DQ282420	A2	3	Hispanic American, USA	(*77*)
DQ282423	A2	2	Hispanic American, USA	(*77*)
DQ282424	A2	2	Hispanic American, USA	(*77*)
DQ282426	A2	1	Hispanic American, USA	(*77*)
DQ282427	A2	3	Hispanic American, USA	(*77*)
DQ282428	A2	1	Hispanic American, USA	(*77*)
DQ282434	B2	1	Hispanic American, USA	(*77*)
DQ282436	B2	1,3	Hispanic American, USA	(*77*)
DQ282438	B2	2,3	Hispanic American, USA	(*77*)
DQ282439	B2	1	Hispanic American, USA	(*77*)
DQ282443	B2	2	Hispanic American, USA	(*77*)
DQ282446	B2	1	Hispanic American, USA	(*77*)
DQ282449	C1	2	Hispanic American, USA	(*77*)
DQ282451	C1	1	Hispanic American, USA	(*77*)
DQ282453	C1	2	Hispanic American, USA	(*77*)
DQ282462	C1	1	Hispanic American, USA	(*77*)
DQ282464	C1	2,3	Hispanic American, USA	(*77*)
DQ282465	C1	2,3	Hispanic American, USA	(*77*)
DQ282467	C1	2	Hispanic American, USA	(*77*)
DQ282470	C1	1	Hispanic American, USA	(*77*)
DQ282471	C1	1	Hispanic American, USA	(*77*)
DQ282472	C1	1,2	Hispanic American, USA	(*77*)
DQ282474	C1	1	Hispanic American, USA	(*77*)
DQ282476	C1	2	Hispanic American, USA	(*77*)
DQ282477	D1	1	Hispanic American, USA	(*77*)

Table S3 (Continued)

Accession Number	hg	Group	Origin	Ref
DQ282478	D1	2	Hispanic American, USA	(*77*)
DQ282481	D1	1	Hispanic American, USA	(*77*)
DQ282483	D1	2	Hispanic American, USA	(*77*)
DQ282485	D1	2	Hispanic American, USA	(*77*)
DQ282486	D1	2,3	Hispanic American, USA	(*77*)
DQ282487	D1	1	Hispanic American, USA	(*77*)
EF079874	B2	2	Dominican Republic	(*26*)
EF153771	A4b	1,2,3	Buryat, Siberia	(*78*)
EF153795	A4c	1,2,3	Buryat, Siberia	(*78*)
EF153799	A4a1	1,2,3	Buryat, Siberia	(*78*)
EF153801	A5c	1,2,3	Buryat, Siberia	(*78*)
EF153806	A2b	1,2,3	Chukchi, East Siberia	(*78*)
EU007835	D	1,2,3	Chukchi, East Siberia	(*79*)
EU007838	A	1,2,3	Chukchi, East Siberia	(*79*)
EU007839	C	1,2,3	Chukchi, East Siberia	(*79*)
EU007840	A	1,2,3	Chukchi, East Siberia	(*79*)
EU007842	A	1,2,3	Chukotka, East Siberia	(*79*)
EU007861	C	1,2,3	Yakut, Siberia	(*79*)
EU095196	A2	2	Waiwai, Brazil	(*25*)
EU095197	A2	3	Zoro, Brazil	(*25*)
EU095198	A2	1	Surui, Brazil	(*25*)
EU095200	A2	1,3	Yanomama, Brazil	(*25*)
EU095202	A2	3	Yanomama, Brazil	(*25*)
EU095203	A2	1	Kayapo, Brazil	(*25*)
EU095206	B2	3	Ache, Paraguay	(*25*)
EU095207	B2	1	Gaviao, Brazil	(*25*)
EU095210	B2	3	Xavante, Brazil	(*25*)
EU095211	B2	2	Xavante, Brazil	(*25*)
EU095213	B2	2	Quechua, Bolivia	(*25*)
EU095214	B2	1	Quechua, Bolivia	(*25*)
EU095215	B2	3	Guarani, Brazil	(*25*)
EU095216	B2	2	Kayapo, Brazil	(*25*)
EU095218	B2	1	Kayapo, Brazil	(*25*)
EU095219	B2	1	Guarani, Brazil	(*25*)
EU095221	B2	2	Yanomama, Brazil	(*25*)
EU095222	C1	1	Waiwai, Brazil	(*25*)
EU095224	C1	1,3	Zoro, Brazil	(*25*)
EU095225	C1	1	Quechua, Bolivia	(*25*)
EU095226	C1	2	Quechua, Bolivia	(*25*)
EU095227	C1	3	Arara, Brazil	(*25*)
EU095228	C1	3	Poturujara, Brazil	(*25*)
EU095229	C1	1,2	Yanomama, Brazil	(*25*)
EU095230	C1	1	Yanomama, Brazil	(*25*)
EU095231	C1	2	Yanomama, Brazil	(*25*)
EU095232	D1	3	Gaviao, Brazil	(*25*)
EU095233	D1	1	Gaviao, Brazil	(*25*)

Table S3 (Continued)

Accession Number	hg	Group	Origin	Ref
EU095234	D1	1,2	Surui, Brazil	(*25*)
EU095235	D1	3	Waiwai, Brazil	(*25*)
EU095238	D1	1	Katuena, Brazil	(*25*)
EU095239	D1	3	Poturujara, Brazil	(*25*)
EU095240	D1	1	Tiryo, Brazil	(*25*)
EU095527	C1c	1	Arsario, Colombia	(*25*)
EU095528	A2	1,2	Arsario, Colombia	(*25*)
EU095530	A2	1	Cayapa, Ecuador	(*25*)
EU095531	D4	3	Cayapa, Ecuador	(*25*)
EU095532	B2	3	Cayapa, Ecuador	(*25*)
EU095534	D2a	1,2,3	Chukchi, East Siberia	(*25*)
EU095535	B2	1,3	Coreguaje, Colombia	(*25*)
EU095537	C1d	1	Coreguaje, Colombia	(*25*)
EU095541	D2b	1,2,3	Evenk, Siberia	(*25*)
EU095544	C1c	1,2	Kogui, Colombia	(*25*)
EU095545	A2	2,3	Kogui, Colombia	(*25*)
EU095546	B2	1,3	Ngoebe, Panama	(*25*)
EU095548	B2	1	Waunana, Colombia	(*25*)
EU482344	A2a	1,2,3	Chukchi, East Siberia	(*80*)
EU597486	A2	2	Colombia	(*81*)
EU597498	D1	2	Karitiana, Brazil	(*81*)
EU597510	D1	1	Karitiana, Brazil	(*81*)
EU597542	C5e	1,2,3	Yakut, Siberia	(*81*)
EU597545	C1b	1	Pima, Mexico	(*81*)
EU597557	C1b	1,3	Pima, Mexico	(*81*)
EU597569	B2	3	Colombia	(*81*)
EU597580	B2	2,3	Colombia	(*81*)
FJ168714	D4h3a1	1	Coquimbo, Chile	(*15*)
FJ168715	D4h3a1	1	O'Higgins, Chile	(*15*)
FJ168718	D4h3a1	1	Coquimbo, Chile	(*15*)
FJ168720	D4h3a1	2,3	Santiago, Chile	(*15*)
FJ168721	D4h3a1	1,3	Los Lagos, Chile	(*15*)
FJ168722	D4h3a1	1,2	Bío-Bío, Chile	(*15*)
FJ168723	D4h3a1	1	Bío-Bío, Chile	(*15*)
FJ168729	D4h3a3	2,3	Nuevo León, Mexico	(*15*)
FJ168733	D4h3a3	3	Chihuahua, Mexico	(*15*)
FJ168734	D4h3a3	2	Tarahumara, Mexico	(*15*)
FJ168736	D4h3a4	2,3	Arequipa, Peru	(*15*)
FJ168737	D4h3a4	1,3	Arequipa, Peru	(*15*)
FJ168739	D4h3a5	1,2	Maule, Chile	(*15*)
FJ168740	D4h3a5	2	Los Lagos, Chile	(*15*)
FJ168741	D4h3a5	3	Santiago, Chile	(*15*)
FJ168742	D4h3a	1,2	Veracruz, Mexico	(*15*)
FJ168744	D4h3a	3	Arequipa, Peru	(*15*)
FJ168747	D4h3a	3	Loreto, Peru	(*15*)
FJ168748	D4h3a	1,2	Loreto, Peru	(*15*)

Table S3 (Continued)

Accession Number	hg	Group	Origin	Ref
FJ168752	D4h3a	2	Ancash, Peru	(*15*)
FJ168753	D4h3a	3	San Luis Potosi, Mexico	(*15*)
FJ168754	D4h3a	2	Maranhão, Brazil	(*15*)
FJ951438	C5b1	1,2,3	Yakut, Siberia	(*82*)
FJ951618	C4b1	1,2,3	Yakut, Siberia	(*82*)
GQ398480	A2	3	Waorani, Ecuador	(*83*)
GQ398482	A2	2,3	Waorani, Ecuador	(*83*)
GQ398483	A2	2	Waorani, Ecuador	(*83*)
GQ398485	A2	1	Waorani, Ecuador	(*83*)
GQ398487	A2	2	Waorani, Ecuador	(*83*)
GQ398490	A2	2	Waorani, Ecuador	(*83*)
HM107308	C1d	1	Chihuahua, Mexico	(*13*)
HM107309	C1d	3	Salta, Argentina	(*13*)
HM107311	C1d2a	1	Mestizo, Colombia	(*13*)
HM107312	C1d2a	2	Mestizo, Colombia	(*13*)
HM107327	C1d1b1	3	Buenos Aires, Argentina	(*13*)
HM107328	C1d1b1	1	Rio Negro, Argentina	(*13*)
HM107329	C1d1b1	1	Buenos Aires, Argentina	(*13*)
HM107338	C1d1	2	Rio Grande do Sul, Brazil	(*13*)
HM107340	C1d1	2	Buenos Aires, Argentina	(*13*)
HM107341	C1d1	2	Loreto, Peru	(*13*)
HM107346	C1d1d	3	Buenos Aires, Argentina	(*13*)
HM107347	C1d1d	3	Rio Grande do Sul, Brazil	(*13*)
HM107348	C1d1d	2	Uruguay	(*13*)
HM107349	C1d1	2	Minas Gerais, Brazil	(*13*)
HM107350	C1d1e	2	Biobio, Chile	(*13*)
HM107351	C1d1e	1,2	Rio Negro, Argentina	(*13*)
HM107352	C1d1	3	Cajamarca, Peru	(*13*)
HM107353	C1d1	3	Huanuco, Peru	(*13*)
HM107354	C1d1	3	Puca Pucara, Peru	(*13*)
HM107357	C1d1	1,3	Chaco, Paraguay	(*13*)
HM107360	C1d1	2,3	Huancavelica, Peru	(*13*)
HM107363	C1d1	2	Los Lagos, Chile	(*13*)
HM107365	C1d1	3	Panama	(*13*)
HM107366	C1d1	2,3	Panama	(*13*)
JN253392	D1	3	Rio Negro, Argentina	(*18*)
JN253394	D1	3	Buenos Aires, Argentina	(*18*)
JN253395	D1	3	Rio Negro, Argentina	(*18*)
JN253398	D1	1,3	Tarapaca, Chile	(*18*)
JN253402	D1	1,3	Buenos Aires, Argentina	(*18*)
JN253404	D1	3	Rio Negro, Argentina	(*18*)
JN253406	D1	1,3	Rio Negro, Argentina	(*18*)
JN253407	D1	1	Maule, Chile	(*18*)
JN253412	D1	1,2,3	Atacama, Chile	(*18*)
JN253414	D1	3	Buenos Aires, Chile	(*18*)
JN253416	D1	2,3	Rio Negro, Argentina	(*18*)

Table S3 (Continued)

Accession Number	hg	Group	Origin	Ref
JN253418	D1	3	Mapuche, Argentina	(*18*)
JN253419	D1	2	Biobio, Chile	(*18*)
JN253421	D1	2	Buenos Aires, Argentina	(*18*)
JN253423	D1	2,3	Cordoba, Argentina	(*18*)
JN253425	D1	3	Valparaiso, Chile	(*18*)
JN253430	D1	2,3	Buenos Aires, Argentina	(*18*)
JN253431	D1	1	Buenos Aires, Argentina	(*18*)
JN253432	D1	2	Buenos Aires, Argentina	(*18*)
JN857011	B4b1a3a1	1,2,3	South Siberia	(*84*)
JN857014	B4b1a3a1a	1,2,3	South Siberia	(*84*)
JX413011	B2l1	2	Rio Negro, Argentina	(*85*)
JX413013	B2l1	2,3	Mapuche, Argentina	(*85*)
JX413015	B2l1	1,3	Chiloe Islands, Chile	(*85*)
JX413017	B2l1	1	Aconcagua, Chile	(*85*)
JX413020	B2l1	1,2,3	Chile	(*85*)
JX413022	B2l1	3	Chile	(*85*)
JX413024	B2l2	2	Aconcagua, Chile	(*85*)
JX413026	B2l2	3	Chiloe Islands, Chile	(*85*)
JX413028	B2l2	3	Chile	(*85*)
JX413030	B2l2	1	Chiloe Islands, Chile	(*85*)
JX413033	B2l2	1,2	Aconcagua, Chile	(*85*)
JX413035	B2l	1,3	Santiago, Chile	(*85*)
JX413037	C1b13a	2	Aconcagua, Chile	(*85*)
JX413039	C1b13a	2	Chile	(*85*)
JX413040	C1b13b	2	Aconcagua, Chile	(*85*)
JX413043	C1b13b	3	Talagante, Chile	(*85*)
JX413044	C1b13c	1,2	Mapuche, Argentina	(*85*)
JX413045	C1b13c	1,2	Chiloe Islands, Chile	(*85*)
JX413046	C1b13c	1,3	Chile	(*85*)
JX413049	C1b13d	3	Aconcagua, Chile	(*85*)
JX413050	C1b13d	2	Chile	(*85*)
JX413052	C1b13e	3	Chile	(*85*)
JX413056	C1b13	1	Chiloe Islands, Chile	(*85*)
KC711005	A2a4	3	Chihuahua, Mexico	(*86*)
KC711006	A2a4	3	Chihuahua, Mexico	(*86*)
KC711022	B2a	2	Chihuahua, Mexico	(*86*)
KC711024	B2a1a	2	Chihuahua, Mexico	(*86*)
KC711025	B2a1b	1	Chihuahua, Mexico	(*86*)
KC711031	B2a3	3	Chihuahua, Mexico	(*86*)
KC711033	B2a4a	1	Sinaloa, Mexico	(*86*)
KC711034	B2a4a1	2	Chihuahua, Mexico	(*86*)
KC711036	B2a4a1	1	Durango, Mexico	(*86*)

hg: mitochondrial haplogroup; Group: random replicate group for BEAST analysis; Ref: reference.

table S4. Polymerase chain reaction (PCR) primers used for mitochondrial long-range PCR.

Primer Name	Primer sequence (5'–3')	Expected length (bp)
H2698-tailT7 L8351	AATTGTAATACGACTCACTATAGGGTTGACCTGCCCGTGAAGAGG TTGGGGCATTTCACTGTAAAGAGG	5721
H7801-tailT7 L14267	AATTGTAATACGACTCACTATAGGGCTATCCTGCCCGCCATCATC GAGGGGTCAGGGTTGATTCG	6530
H13305 L3407-tailT7	TCGGCATCAACCAACCACAC AATTGTAATACGACTCACTATAGGGTACAACGTTGGGGCCTTTGC	6735

table S5. Predictive accuracy of the PCMLR model.

		Predictive accuracy of the MLR model in the principal components						
		A	**B**	**C**	**D**	**E**	**F**	**G**
Demographic model tested	**A**	45.74%	42.59%	0.43%	3.60%	2.77%	4.86%	0.01%
	B	42.24%	45.23%	0.79%	4.13%	4.74%	2.87%	0.01%
	C	0.21%	0.28%	75.35%	8.41%	8.40%	7.33%	0.02%
	D	4.20%	4.12%	3.65%	33.38%	30.56%	24.08%	0.00%
	E	3.38%	3.78%	3.73%	32.48%	37.13%	19.50%	0.00%
	F	4.02%	3.77%	3.41%	24.81%	22.97%	40.97%	0.05%
	G	0.01%	0.01%	0.01%	0.03%	0.03%	0.05%	99.87%

table S6. Prediction of the population structure model that fits the observed data.

	A	**B**	**C**	**D**	**E**	**F**	**G**
Replicate 1	0.01%	0.01%	99.56%	0.08%	0.28%	0.03%	0.04%
Replicate 2	0.01%	0.01%	99.89%	0.02%	0.06%	0.01%	0.00%
Replicate 3	0.01%	0.02%	99.54%	0.01%	0.04%	0.01%	0.37%

table S7. Highest predictive accuracy of the PCMLR model for each population structure model when model C is removed from the analysis.

A	**B**	**D**	**E**	**F**	**G**
45.93%	45.21%	35.37%	38.14%	42.98%	99.87%

table S8. Prediction of the population structure model that fits the observed data when model C is removed from the analysis.

	A	**B**	**D**	**E**	**F**	**G**
Replicate 1	2.06%	4.20%	21.25%	57.80%	6.40%	8.29%
Replicate 2	6.76%	12.91%	21.62%	51.26%	6.90%	0.55%
Replicate 3	2.05%	4.10%	1.74%	4.84%	0.85%	86.42%

Supplementary Figures

fig. S1. Location of archaeological sites (see table S2 for detailed information about sites and samples). Camarones and Pica-8 have the same coordinates, as well as Jauranga and Los Molinos.

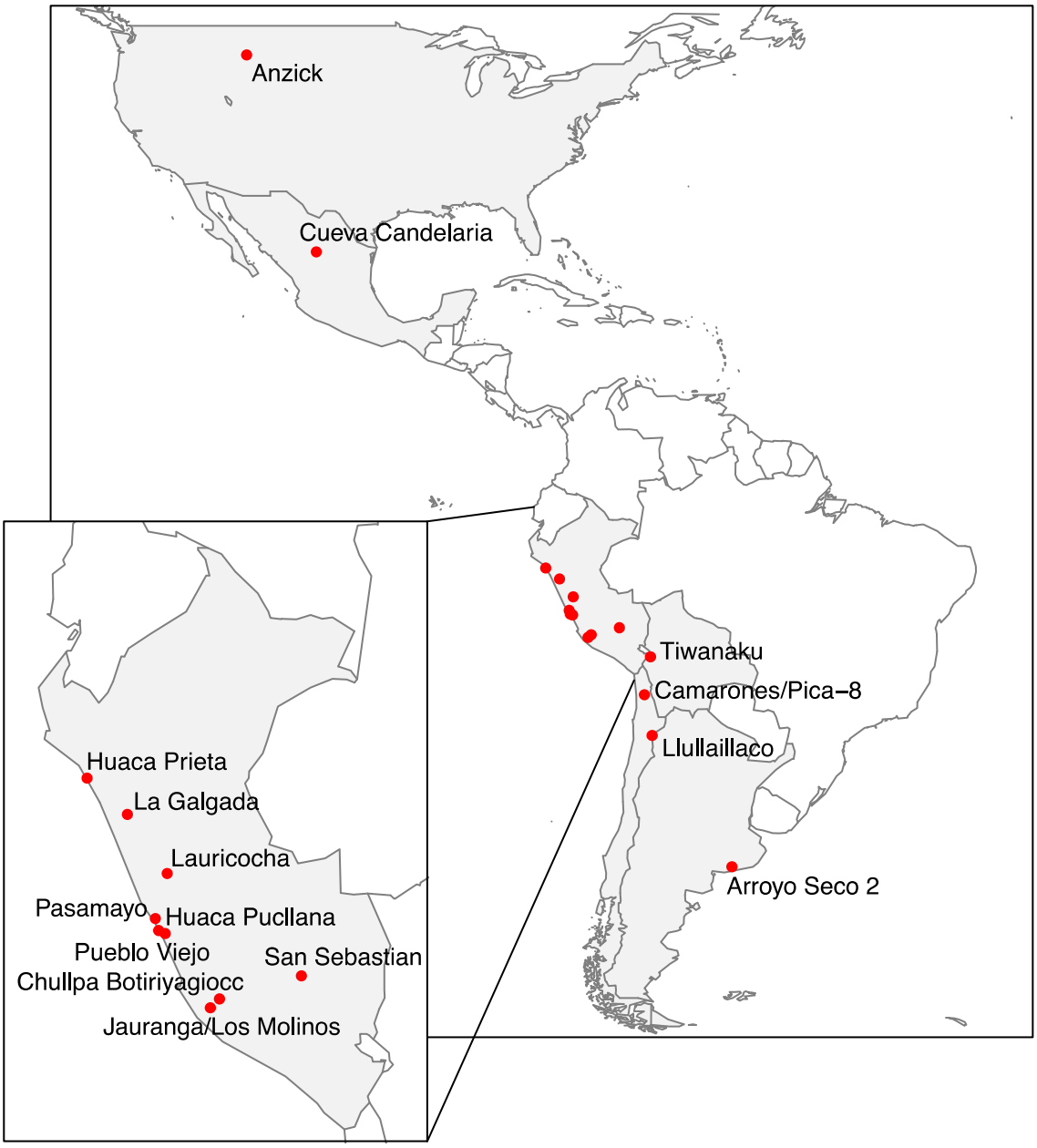


fig. S2. Phylogenetic tree of newly sequenced ancient mitochondrial genomes (haplogroup A). Nomenclature and topology are based on the PhyloTree mtDNA tree build 16 (19 Feb 2014).

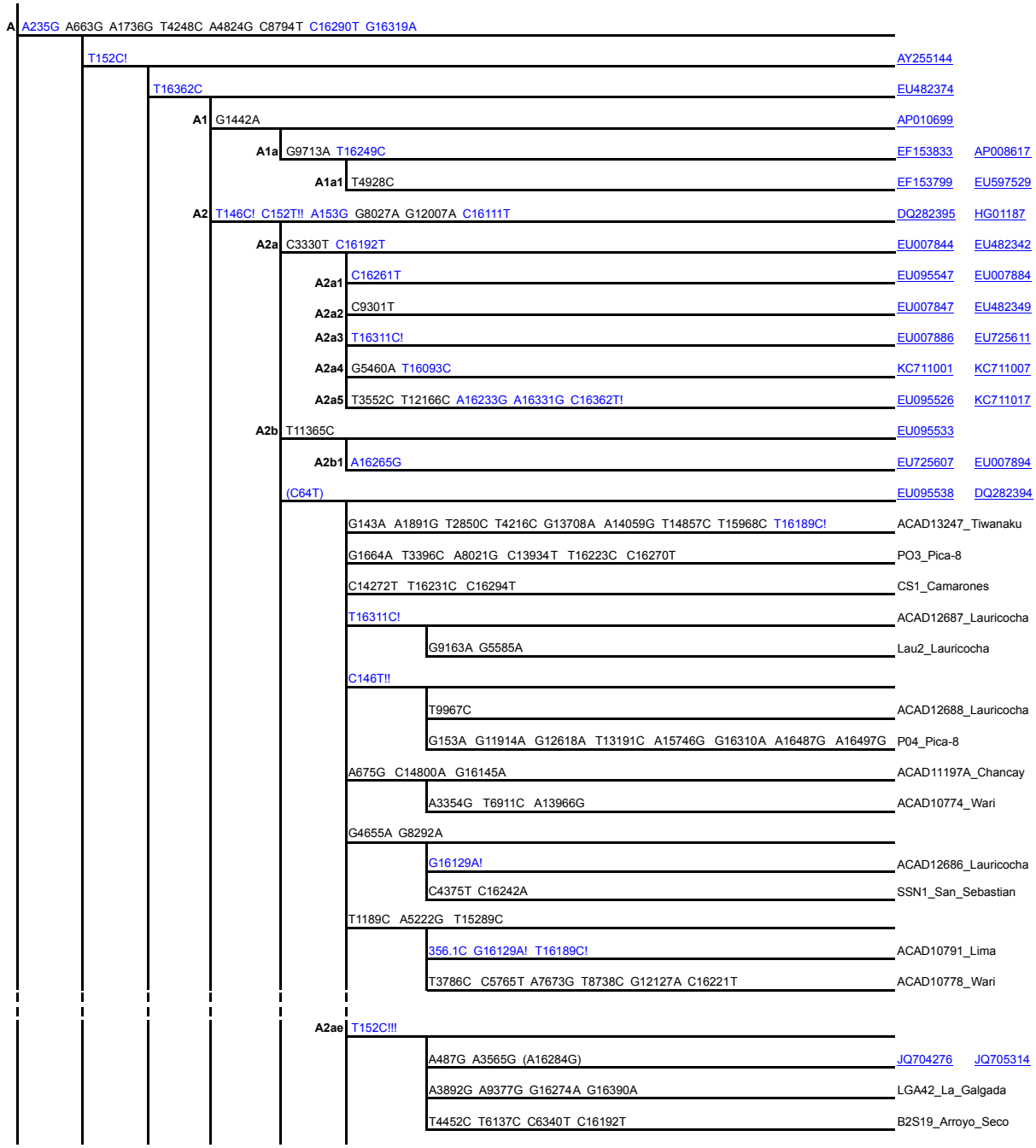


fig. S3. Phylogenetic tree of newly sequenced ancient mitochondrial genomes (haplogroup B). Nomenclature and topology are based on the PhyloTree mtDNA tree build 16 (19 Feb 2014).

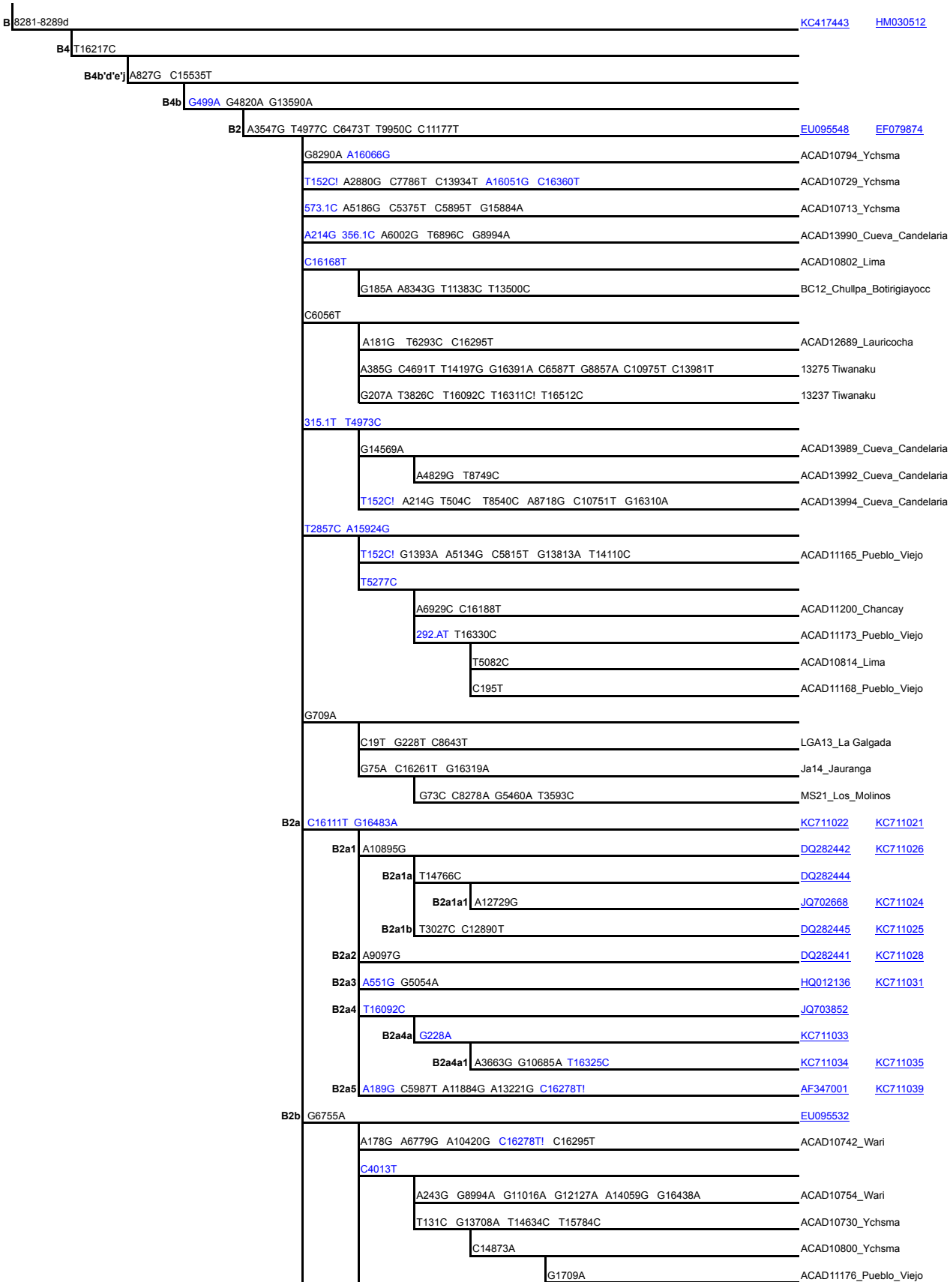


fig. S3. (continued)

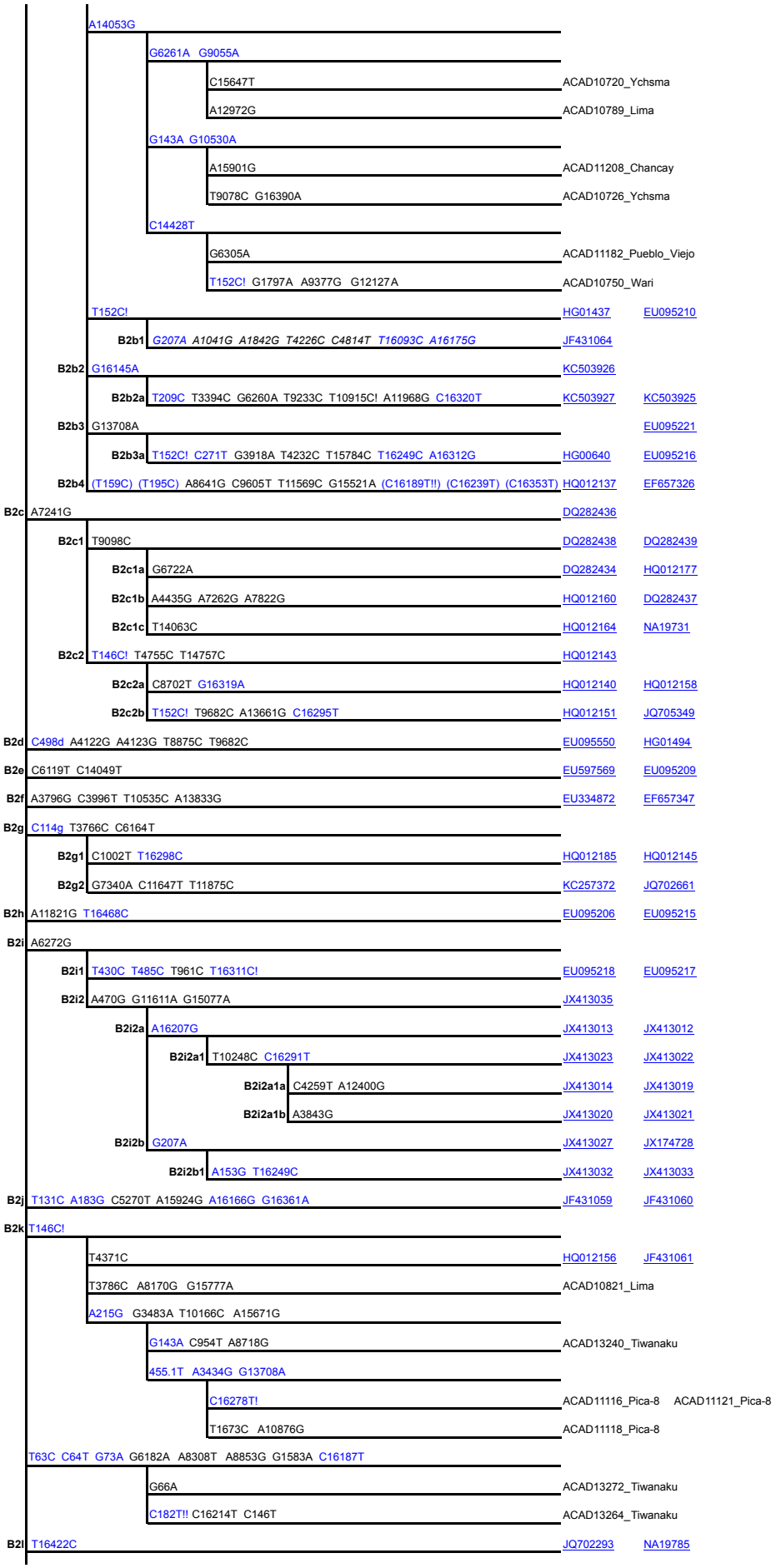


fig. S4. Phylogenetic tree of newly sequenced ancient mitochondrial genomes (haplogroup C). Nomenclature and topology are based on the PhyloTree mtDNA tree build 16 (19 Feb 2014).

fig. S4. (continued)

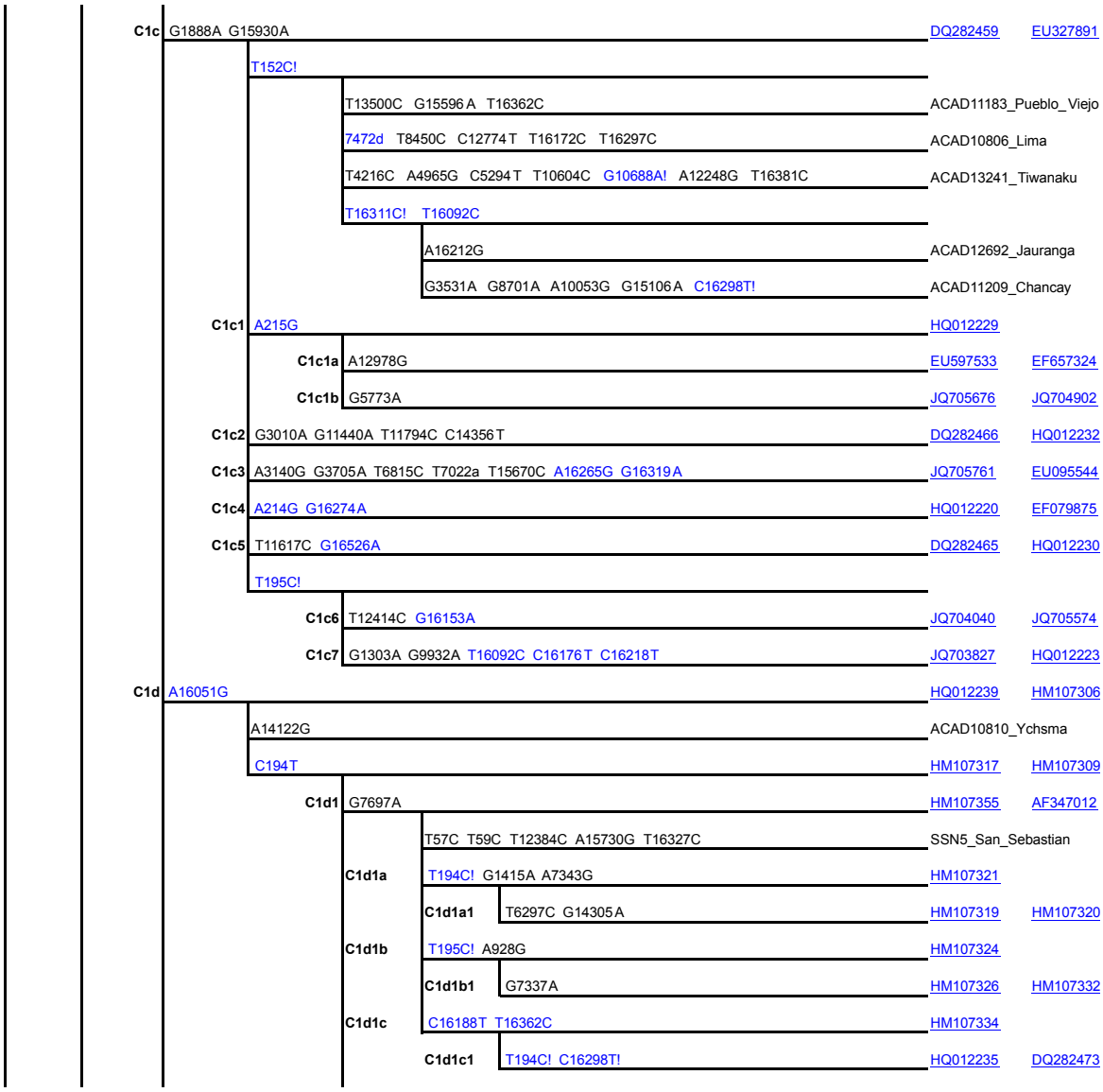


fig. S5. Phylogenetic tree of newly sequenced ancient mitochondrial genomes (haplogroup D). Nomenclature and topology are based on the PhyloTree mtDNA tree build 16 (19 Feb 2014).

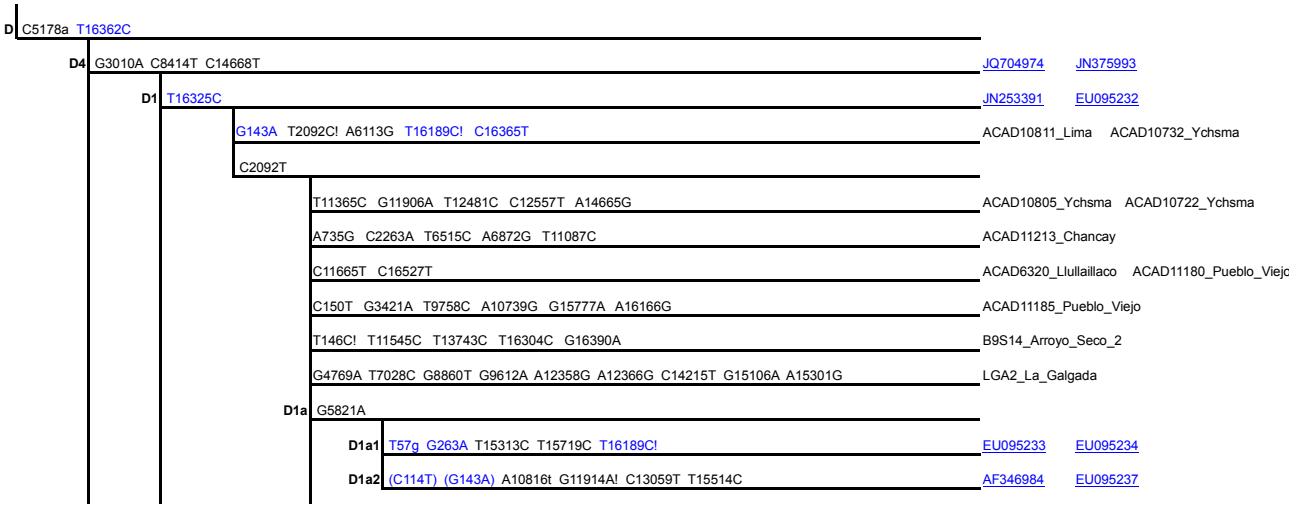


fig. S6. Maximum parsimony phylogenetic trees of the 93 ancient mitogenomes and the three replicate datasets of 87 modern Native American mitogenomes (see table S3). Ancient mitogenomes are in red, modern mitogenomes are in blue.

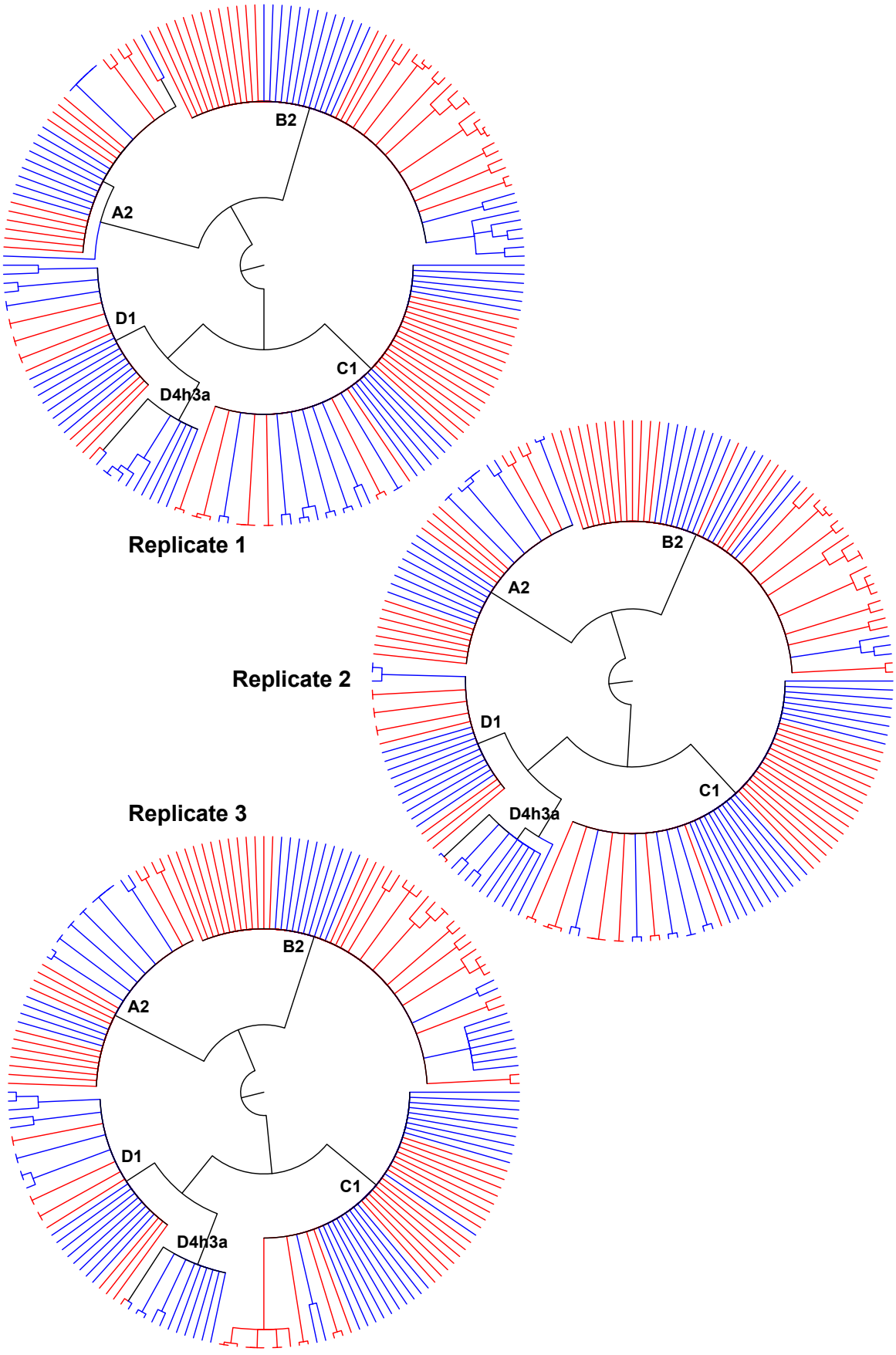


fig. S7. Extended Bayesian skyline plots of female effective population size for the three replicate data sets.

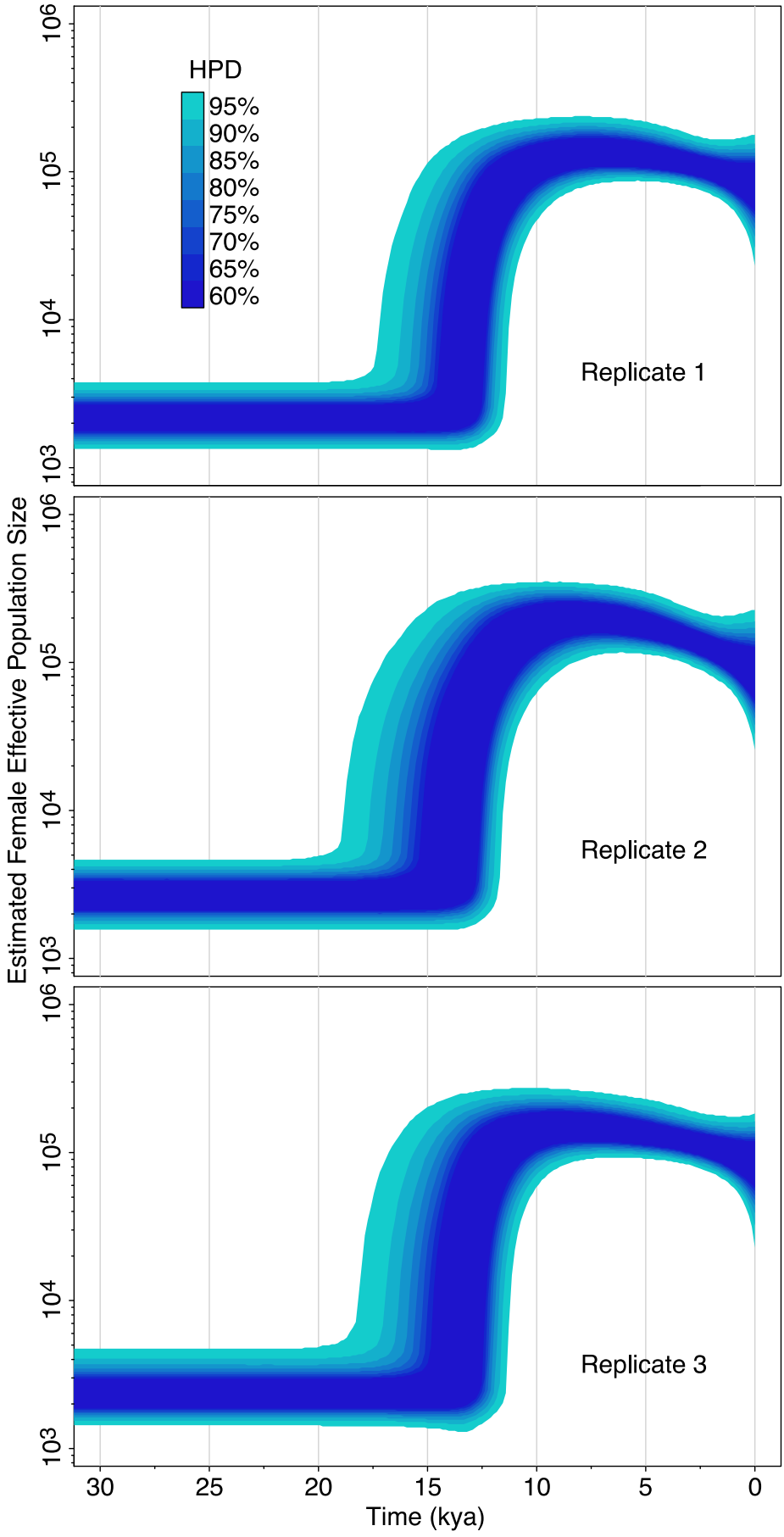


fig. S8. Bayesian estimates of node ages for haplogroups A2, B2, C1, D1, and D4h3a. Error bars denote 95% HPD intervals.

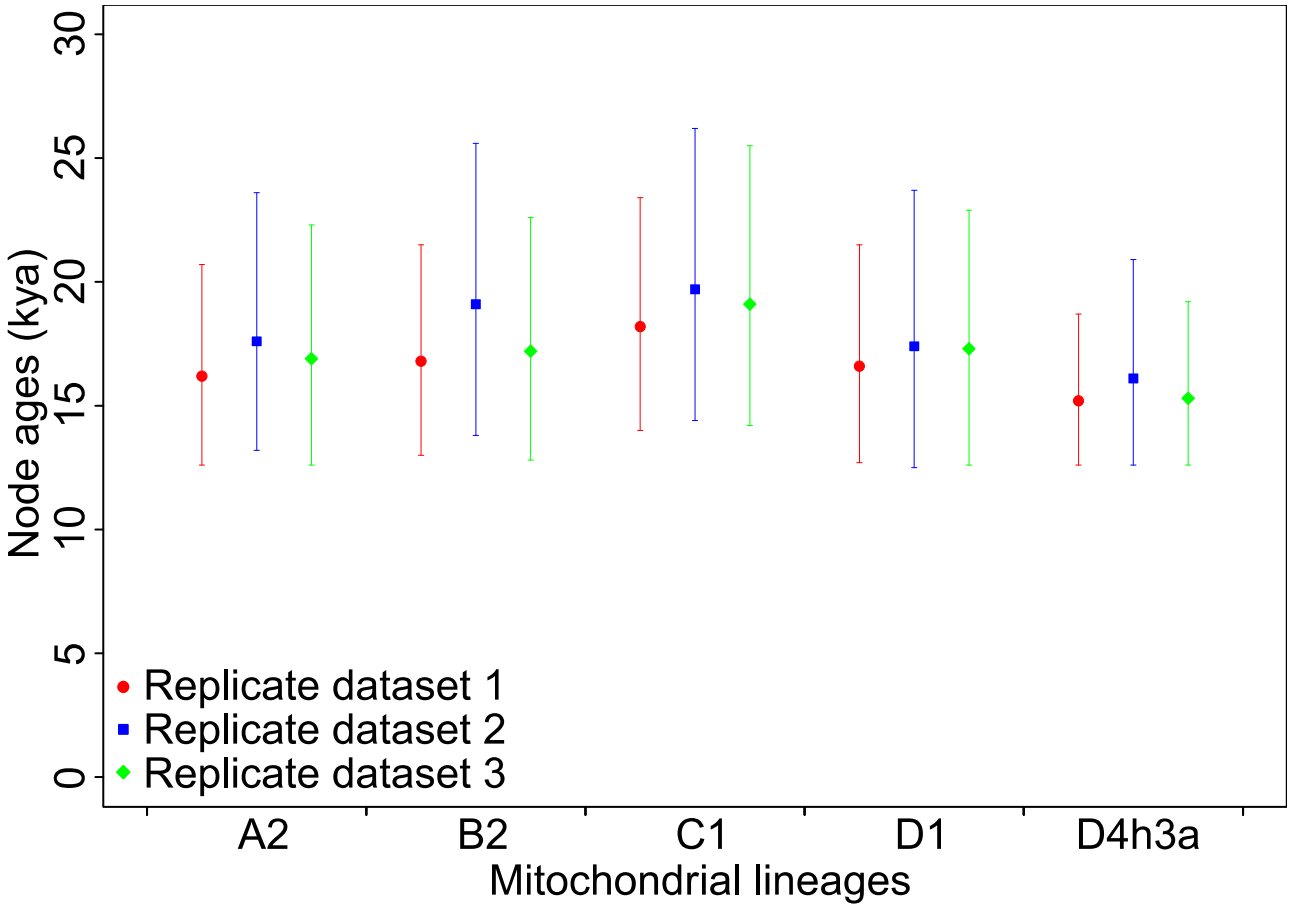
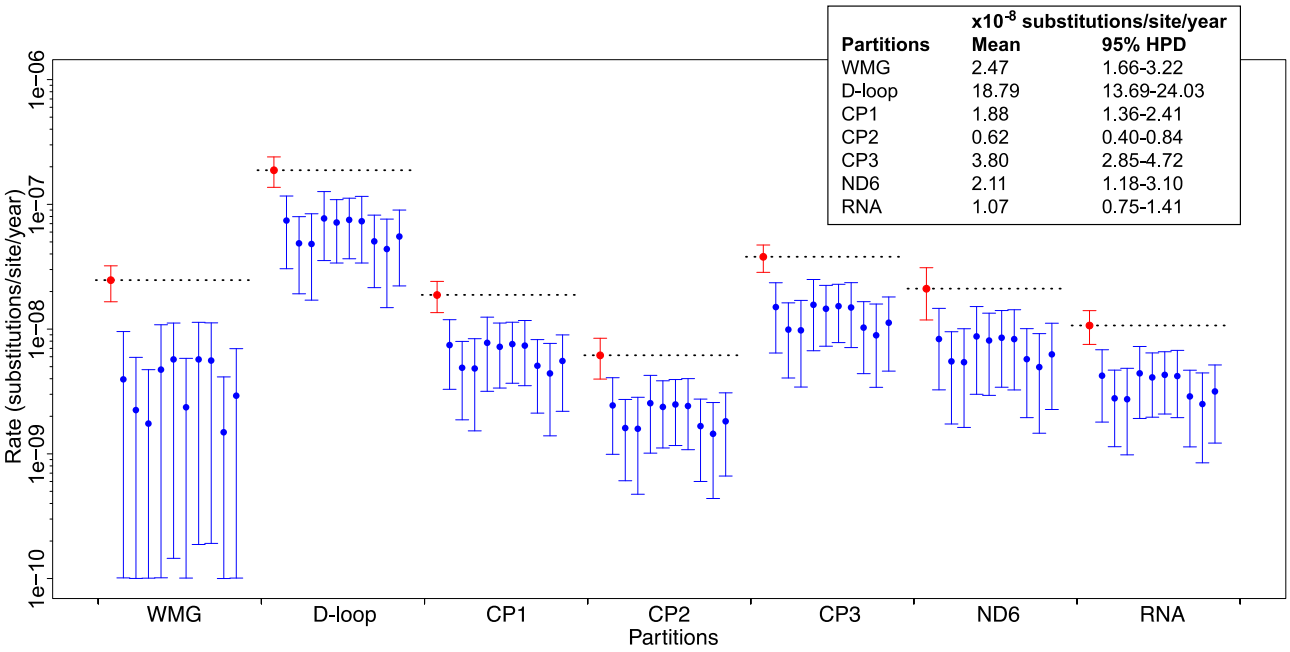


fig. S9. Results of the date-randomization test for temporal sampling structure. For the whole mitogenome (WMG) and for each data partition, the 95% HPD intervals of the rate estimates from the analyses with randomized dates (blue) do not overlap with the mean rate estimate (dotted line) from the analysis of the data with the true sample dates (red). Mean substitution rate and 95% HPD interval for each partition are given in the inset table. CP: codon position.



Partitions	x10^{-8} substitutions/site/year Mean	95% HPD
WMG	2.47	1.66-3.22
D-loop	18.79	13.69-24.03
CP1	1.88	1.36-2.41
CP2	0.62	0.40-0.84
CP3	3.80	2.85-4.72
ND6	2.11	1.18-3.10
RNA	1.07	0.75-1.41

fig. S10. PCA plot of summary statistics for the 15,000 simulations for each of the seven models simulated with BayeSSC (see Fig. 4). Observed data from the three replicates are the black circles.

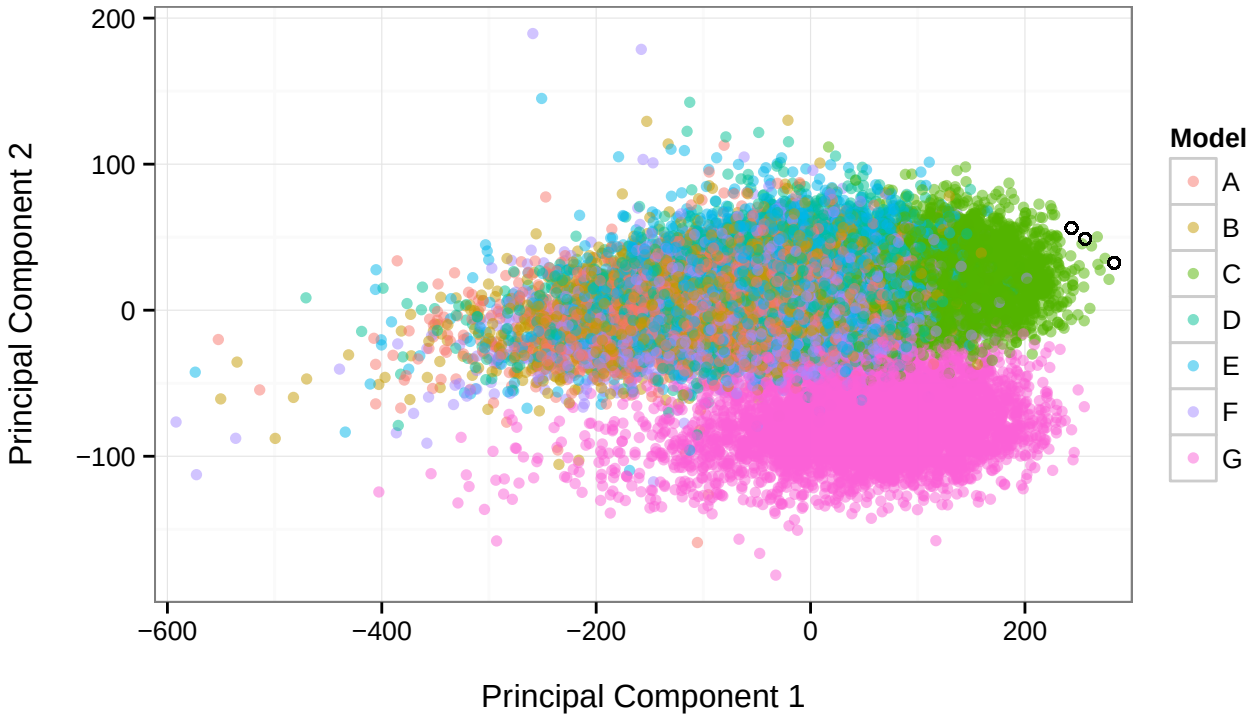


fig. S11. PCA plot of summary statistics for the 15,000 simulations for each of the models simulated with BayeSSC, when model C is removed from the PCA. Observed data from the three replicates are the black circles.