Classification: Biological Sciences - Anthropology

Two Contemporaneous Mitogenomes from Terminal Pleistocene Burials in Eastern Beringia

Short Title: Mitogenomes from Pleistocene Burials in Beringia

Justin Tackney[a,1], Ben A. Potter[b], Jennifer Raff[c], Michael Powers[d], W. Scott Watkins[e], Derek Warner[d], Joshua D. Reuther[b,f], Joel D. Irish[g], and Dennis H. O'Rourke[a]

[a] Department of Anthropology, University of Utah, Salt Lake City, UT 84112.
[b] Department of Anthropology, University of Alaska, Fairbanks, AK 99775.
[c] Department of Anthropology, University of Kansas, Lawrence, KS 66045.
[d] DNA Sequencing Core, University of Utah, Salt Lake City, UT 84112.
[e] Department of Human Genetics, University of Utah, Salt Lake City, UT 84112.
[f] Archaeology Department, University of Alaska Museum of the North, Fairbanks, AK 99775.
[g] Research Centre in Evolutionary Anthropology and Paleoecology, Liverpool John Moores University, Liverpool, UK L33AF.

[1] To whom correspondence should be addressed. 270S 1400E RM 102, Stewart Building, Salt Lake City, UT 84112; Phone: 908-399-5470; Email: Justin.Tackney@anthro.utah.edu

Keywords: Pleistocene burials, ancient mitochondrial DNA, peopling of Americas

Abstract

Pleistocene residential sites with multiple contemporaneous human burials are extremely rare in the Americas. We report mitochondrial genomic variation in the first multiple mitochondrial genomes from a single prehistoric population: two infant burials (USR1 and USR2) from a common interment at the Upward Sun River Site in central Alaska dating to ~11,500 calendar years before present (cal B.P.). Using a targeted capture method and next-generation sequencing we determined that the USR1 infant possessed variants that define mitochondrial lineage C1b, while the USR2 genome falls at the root of lineage B2, allowing us to refine younger coalescence age estimates for these two clades. C1b and B2 are rare to absent in modern populations of Northern North America. Documentation of these lineages at this location in the Late Pleistocene provides evidence for the extent of mitochondrial diversity in early Beringian populations, which supports the expectations of the Beringian Standstill Model.

Significance Statement

Beringia gave rise to the first Western Hemisphere colonists, although the genetic characterization of that source population has remained obscure. We report two mitogenomes from human remains within Beringia, with an age (~11,500 cal BP) that post-dates the end of the initial colonization by only a few millennia. The mitochondrial lineages identified (B2, C1b) are rare to absent in modern northern populations, indicating greater genetic diversity in early Beringia than in modern populations of the region. The antiquity and geographic location of these two burials, and the combined genomic and archaeological analyses, provide new perspectives on the link between Asia and the Americas, and the genetic makeup of the first Americans.

\body

Text

The colonization of the Western Hemisphere has been of interest to scholars since 1590 when Jose de Acosta postulated a northeast Asian origin of the indigenous populations of the Americas (1). Both the archaeological (2, 3) and genetic (4-10) records consistently indicate a primary entry point from Asia to the Americas via the Bering Land Bridge sometime during the Late Pleistocene. However, there are unfortunate lacunae in both records. The archaeological record indicates a relatively late (<14-16kya), rapid colonization event following the Last Glacial Maximum (LGM). This temporal scale supports the clear northeastward geographical expansion of late Upper Paleolithic (Diuktai) populations from southern and central Siberia to Beringia after 16 kya (5). However, archaeological evidence is accumulating that shows people had penetrated parts of North and South America prior to 13,250 cal BP, the earliest date associated with Clovis, the first widespread cultural tradition in North America (2-5, 11).

The genetic record is equally problematic. Continental scale analyses of genetic variation rely heavily on Central and South American population data, as well as data from Arctic populations (6-9, 12, 13). Few data exist for North American populations south of the Arctic. Recent surveys of contemporary genetic variation in the Americas are consistent with a period of population isolation during which the distinctive composition of Native American genomes differentiated from ancestral Asian genomes, followed by a rapid colonization; this scenario has been deemed the 'Beringian Standstill Model' (6, 7, 10). How early the Native American gene pool diverged remains uncertain, but estimates of up to 30 kya have been postulated (5, 6, 10, 12, 14, 15). Most geneticists argue for at least a several thousand-year period of isolation and genetic differentiation in Beringia before a southward dispersal, despite the absence of

supporting archaeological evidence (2, 4, 5, 10). Recently, Raghavan, *et al.* (15), using genome-wide low coverage data, suggested the dates of this isolation began no earlier than 23 kya and lasted no longer than eight thousand years (15).

Ancient DNA samples from early inhabitants of the Americas would be important for linking the modern genetic and archaeological records (16), but few exist. The Mal'ta child from South Central Siberia indicates an early origin (>24 kya) of some signal of Native American ancestry (9), but while a few Pleistocene-aged remains have been recovered in central North America (below the Laurentide Ice Sheet) or along the Northwest Coast, no similarly aged Beringian human remains have previously been available for genetic comparison. Very few Late Pleistocene (>10,000 cal BP) individuals have yielded mitochondrial genetic (mtDNA) data, though we highlight the seven sites with ancient human remains dating to >8,000 years old that have been characterized for mtDNA lineages: Hoyo Negro, Mexico(17); Anzick, MT (18); Kennewick, WA(19); On-Your-Knees Cave, AK; Wizard's Beach, NV; Hourglass Cave, CO; and, indirectly through coprolite analysis, Paisley Cave, OR (last four are reviewed in (20); **Fig 1**).

In 2011 Potter et al (21) reported on the discovery of a cremated 3-year old child from a residential feature at Upward Sun River (USR) in eastern Beringia dating to 11,500 cal BP. Additional excavation at this deeply stratified and well-dated site (22) recently yielded two additional infant burials (**Fig 1**; USR1 and USR2) (23). A series of radiocarbon ages securely date the three individuals between 11,600-11,270 cal BP (23). Based on dental and osteological aging methods, USR1 represents a late pre-term fetus, while USR2 likely died within the first six weeks of life (23). The proximity of these three burials, their context within the same feature, and radiocarbon analyses presented in Potter et al. (23) strongly suggest that all three burials

represent nearly contemporaneous events, and that the three individuals were members of a single population.

We attempted to extract and sequence the mitochondrial genomes from these three Late Pleistocene burials. From burnt bone fragments of the cremated infant and well preserved samples of the petrous portion of the parietal bone, DNA was extracted using a silica-based method and attempts were made to Sanger sequence three overlapping fragments of the mitochondrial hypervariable region 1 (HVR1). From USR1 and USR2 all three HVR1 fragments were successfully amplified, and from the cremated infant only one fragment amplified, albeit inconsistently. DNA samples and applicable blank controls from USR1 and USR2 were converted to Ion Torrent Ion Plus Fragment libraries with laboratory-unique barcodes. We targeted the mitochondrial genomes by hybridization capture (24) and sequenced the libraries on two P1 chips with an Ion Proton System (Life Technologies). This is one of the first examples of the Ion Torrent technology applied to ancient DNA.

Results

From 58.7 and 55.8 million sequencing reads, 20,004 and 32,979 unique mtDNA reads (MAPQ ≥ 30) from USR1 and USR2, respectively, were mapped to the human mtDNA reference **(Table S1).** We utilized the Torrent Suite analytical pipeline to take advantage of flow space information, base recalibration, read realignment, and an Ion-optimized mapping (tmap) and duplicate filtering approach. This pipeline also allowed variant calling with the Torrent Variant Caller (TVC), providing a range of variant quality metrics identical to current best-practices approaches for next-generation sequencing of modern samples. This pipeline is

optimized for Ion Torrent reads, unlike most methodologies currently employed in the ancient DNA literature.

Sequencing of the enriched mtDNA from samples USR1 and USR2 resulted in 100% coverage of the mtDNA genomes with average read depths of 117x (geometric mean of 97x) for USR1 and 195x (geometric mean of 180x) for USR2 **(Fig S1)**. Mean read lengths for the two samples were 98 and 99 base pairs. Contamination estimates were made by dividing the reference allele counts at called variants by the total coverage from the TVC output; contamination rates were estimated at 3.5% and 4.9% for the two samples, respectively. Maximum Parsimony (MP) analysis of single-nucleotide polymorphisms (SNPs) and insertion/deletions (indels) in the full genomes indicated membership in mtDNA lineages C1b (USR1) **(Fig 2A)** and B2 (USR2) **(Fig 2B)**. The mtDNA genome of USR1 had a private variant in the form of SNP C16292T. The B2 lineage carried by USR2 revealed a single back mutation at nucleotide position (np) 3547 to an ancestral adenine. A subset of called variants, in addition to the previously typed HVR1, were validated by Sanger sequencing.

From the initial Torrent Suite bioinformatics pipeline we observed an irregular pattern of DNA damage expected from ancient DNA samples **(Fig S2)**. The 5' ends of these reads had unexpected low quality base calls, likely from our custom adapters lacking a spacer sequence after the barcodes, and we were not able to investigate 3' damage patterns. We initiated an alternative pipeline for reads from both Ion P1 chips: We performed additional read trimming for adapter sequence, length (30 to 120bp), and quality, and we re-mapped (tmap) without 3' clipping. Following this alternative pipeline, 21,140 and 22,951 mtDNA reads at MAPQ ≥ 70 mapped to the mtDNA genome from USR1 and USR2, respectively **(Table S2)**. 100% of the genome was covered, at average read depth of 113x (geometric mean of 103x) for USR1 and

125x (geometric mean of 119x) for USR2 **(Fig S3).** Nucleotide mismatches now displayed the expected damage patterns for degraded samples, though the 5' read ends still showed some residual unexpected alternative signal **(Fig S4)**. While this pipeline lost the necessary flow-space information to make variant calls from Ion Torrent data, visual inspection of the aligned reads confirmed all variants called earlier by TVC. This suggests that the previous quality issue, while masking expected DNA damage patterns at the ends of reads, did not bias the accurate calling of these two samples.

Maximum Likelihood trees were created from curated alignments of 189 haplogroup C **(Fig 3A)** and 147 haplogroup B sequences (**Fig 3B**). USR1 was placed within a large clade shared with C1b, while USR2 was placed at the root of known Native American B2 diversity. Both samples exhibit branch length shortening relative to modern Native American sequences, owing to their lower number of derived mutations, as expected for ancient DNA. The best tree by final likelihood score was compared to the results of 1000 bootstrap runs. Non-parametric bootstrap support on the trees was poor within the Native American specific haplotypes, given the relatively small number of characters providing signal in otherwise highly similar, and polytomous, mtDNA clades (25). As USR1 and USR2 are believed to be contemporaneous, and modern Native American B2 and C1 sequences are observed to have similar coalescence times (6, 12, 26), we investigated the effect of these new sequences on the molecular dates of these clades.

We calculated the coalescence times using an ML-based approach and either a molecular clock corrected for purifying selection (27) or a faster, Bayesian-determined molecular clock based on ancient mitochondrial genomes (28). The C1b clade divergence time was estimated at 16,600 or 13,900 ya, respectively with the two rates. USR1 was most closely related to an

individual of the Arara people of Brazil (EU095227), with an estimated divergence date of 8,200 or 7,000 ya (a clearly too-recent age given the archaeological age estimate of USR1). The Native American-specific B2 clade coalescence time was estimated at 19,100 or 15,900 ya, respectively (27, 28). All of these dates fall within previously published estimates.

As the ML-based estimates do not take into account the radiocarbon ages of USR1 and USR2, we utilized the Bayesian Markov Chain Monte Carlo framework of BEAST 2.2 (29). This Bayesian phylogenetic method uses temporal information from dated sequences to calibrate a molecular clock without relying on geological or paleontological information. Using this approach, we calculated the C1b clade coalescence time at 12,854 ya (11,853-14,079) (mean; 95% highest posterior density (HPD) interval), with tip dates of 11,500 ya for USR1 and 8,300 ya for UZOO-74 (see below). The C1b clade coalescence date is near the younger bounds of the timescales calculated in the literature, though the 95% HPD overlaps with the date previously determined using the faster aDNA-calibrated substitution rate. The B2 clade coalescence time was estimated at 12,024 ya (11,500-13,085), using a tip date of 11,500 ya for USR2. This B2 date is also on the later end of previously reported timescales (28). As these estimates are derived from only one (B2) or two (C1) point estimated sequence ages, the analysis can be improved with increased whole genome sequencing of ancient samples specifically within these clades. The general agreement, however, with the faster Bayesian molecular clock supports relatively young clade coalescence dates.

Discussion

The presence of mtDNA haplotype B2 is somewhat unexpected in this geographic location. This lineage is absent in northern and eastern Siberia (though it is found in the southern

periphery; 30), and the pan-American B2 haplotype has not been reported in high latitude populations of modern indigenous North Americans **(Fig 1)**. This unusual geographic distribution, coupled with lower RFLP (restriction fragment length polymorphism) haplotype genetic diversity estimates, led to the hypothesis that the B2 lineage was introduced by a later, separate colonization event that did not pass through Beringia. However, following increased sampling and whole mitochondrial genome sequencing, haplogroup B2 phylogenies were shown to have similar star-like phylogenies and coalescence times to the other pan-American founding lineages (12). Moreover, Raff and colleagues (31) reported two ancient individuals with haplogroup B2 in prehistoric (800 and 490 cal BP) populations on the upper Alaska peninsula.

Haplogroup B2 in subarctic interior Alaska at the Upward Sun River site at such an early date suggests it was likely present and polymorphic in the Beringian population that gave rise to the initial dispersal south into the interior of the American continents. Importantly, the finding of haplogroup B2 in far northern interior populations shortly after initial colonization negates the need to postulate models of independent introduction of this mitochondrial lineage through alternative colonization routes. Its absence from modern high latitude populations now appears consistent with the action of migration and genetic drift in small, dispersed early populations (4) rather than selection or independent introduction. It is noteworthy that haplogroup B was identified at two of the oldest sites in the Americas mentioned earlier, i.e. at the ~8800 cal BP burial at Hourglass Cave in Colorado and in three coprolites dated between 14,270-14,000 cal BP at Paisley 5 Mile Point Caves in south-central Oregon. Neither site has yielded full mitochondrial genome data.

Haplogroup C is one of the two most common mitochondrial DNA clades throughout northern, eastern, and central Asia (the other being haplogroup D). The wide distribution of

haplogroup C suggests it was a component of most migrations in northern Eurasia, with an origin between 30-50 kya (32). One daughter clade of the haplogroup is C1, which is composed of an Asian-specific C1a branch previously molecularly dated to 8,500 ya (32), three Native American-specific (C1b, C1c, C1d) branches previously molecularly dated to 19,000 ya (12, 26), an Icelandic-specific C1e branch (33), and a novel C1f haplotype sequenced from an individual dated to ~8,300 cal BP (UZOO-74) at the Mesolithic site of Yuzhnyy Oleni Ostrov, North West Russia (25) (though see **Fig 3A** for a possibly related sequence, HM804483). Unlike the case for UZOO-74, the USR1 C1b sequence has a clear origin and evolutionary history in the Americas. This result highlights the need for further genomic sequencing of comparably aged C1 lineages, e.g. the 10,400 cal BP individual from Wizard's Beach, NV and for further sequencing of any C1 lineages in Eurasia.

It is of interest that all five founding macro-haplogroups in Native American populations (A, B, C, D, and X) are represented in the small sample of individuals that lived more than 8,000 years ago in a geographic area stretching from subarctic Alaska to southern Mexico. Four of these macro-haplogroups are found at the three northern North American sites - Paisley Caves, Upward Sun River, and Anzick - dating to over 11,000 years ago. Mitochondrial lineage designation for the majority of the pre-8,000 years ago individuals were determined by low resolution methods of RFLP analysis and/or direct sequencing of PCR products. Only three of the included studies (17-19) used genomic approaches, in addition to the Upward Sun River individuals reported here. Collectively, these results indicate a broad base of mitochondrial diversity in the earliest populations in North America and suggest the importance of post-colonization population dynamics in structuring modern genetic patterns. Cui and colleagues (13) recently bolstered this inference by reporting four mtDNA genomes from mid-Holocene

individuals from coastal British Columbia. The persistence of two unique A2a lineages but the extinction of the D4h3a lineage observed in the transition from ancient to modern Native American populations emphasizes that extant genetic patterns alone can be inadequate indicators of prehistoric population diversity.

Although the Upward Sun River population post-dates the end of the original dispersal of populations into North and South America by a few thousand years, it is temporally and geographically the closest known to the larger interior Beringian population that was the source of that earlier migration. Further, if the Beringian population was subdivided in refugia as recently suggested (4), the geographic structure seen in modern indigenous North Americans may reflect early population differentiation and multiple dispersals of small, isolated groups in interior Beringia to interior North America. Available archaeological and genetic data from Late Pleistocene contexts in North America are consistent with the origin of Native American mitochondrial genomes in populations resident in interior Beringia with subsequent dispersal southward sometime prior to 14-16 kya. The distribution of founding mitochondrial lineages in ancient samples of the Americas suggests an early movement of interior Beringian peoples southward at colonization, followed shortly by similar dispersal along the Pacific coast. The ancient mitochondrial genomes of the two contemporaneous Upward Sun River infant burials provide an important anchor between modern patterns of genetic variation and the inferences that may be drawn from retrospective population genetic analyses.

Conclusion

The genomic results on the Upward Sun River infants are significant for several reasons. First, they not only double the number of late Pleistocene burials that have been characterized

genetically, but they are also the only example to date of multiple burials from a single North American Pleistocene-aged archaeological site. Second, the genomic results from the USR infants support the Beringian route into the Americas and imply substantial interior Beringian genetic variation in the Late Pleistocene, consistent with expectations of the Beringian Standstill Model. Phylogenetic coalescent dates informed by the sample radiocarbon ages suggest more recent expansions for the Native American C1 and B2 clades than has previously been suggested. Third, these results clarify the infants' biological relationship to one another, something that morphological data could not do (23). Fourth, the fact that the infants are contemporaneous in time and buried together in a single act speaks to population diversity in ways that single sample reports cannot. And fifth, the dual burial of maternally unrelated infants (although perhaps paternally related) suggests additional hypotheses regarding mortuary practices and social and ceremonial behaviors present at this early time; this line of investigation may be addressed in the future by both nuclear genomic analyses of the infants, as well as continued elaboration of the archaeological context of the site.

Materials and Methods

USR1 and USR2 were complete and located 8-10 cm apart at the bottom of the pit feature at the Upward Sun River site, located in the middle Tanana River valley. Two petrous specimens were selected for DNA analyses given their overall mass and high density. Details on site formation, chronology, site disturbance, and excavation protocols are reported elsewhere (21-23, 34). Destructive analysis and genetic sequencing of the material was formally allowed by a Memorandum of Agreement with all interested parties. DNA was extracted using a silica-based method and initially amplified using established protocols. Extracts were prepared into Ion Plus

Fragment libraries (Life Technologies) with no DNA fragmentation or size selection. Fragments were blunt-end ligated with adapters containing laboratory-specific custom barcodes. Mitochondrial DNA was captured by hybridization (24) and each sample library was sequenced on its own Ion PIv2 chip (Life Technologies). Read processing was completed either within Torrent Suite, with variants called using Torrent Variant Caller, or reads were processed using offline tools to determine DNA damage patterns. Haplotypes of consensus mitochondrial genomes from these variants were identified by Maximum Parsimony and phylogenetic trees of all known related sequences were created by Maximum Likelihood (ML). Coalescence dates for the clades within these trees were calculated using ML-based or Bayesian-based phylogenetic methods. Work was performed in a dedicated ancient DNA facility using established clean room protocols. Blanks were included at all steps in the process prior to sequencing and no laboratory personnel carry the haplotypes reported here. An expanded discussion of detailed Materials and Methods can be found in **Supporting Information**.

Acknowledgements

Sequencing was performed at the DNA Sequencing Core Facility, University of Utah. This project was funded in part by National Science Foundation Grants OPP-0732846, OPP-1137078, OPP-1138811, and OPP-1223119. This project was additionally funded by the College of Social and Behavioral Science and the Department of Anthropology at the University of Utah. We would like to acknowledge the long technical conversations with Lin Chen and Zhao Xu in the Life Technologies NGS Bioinformatics support team. We would like to thank Dr. Chad Huff for his expertise with PAML, Dr. Brendan O'Fallon and Dr. Remco Bouckaert for their support with BEAST, two anonymous reviewers for their helpful critiques, and Dr. Ryan Bohlender as an excellent and patient de facto server administrator. We acknowledge the laboratory of Dr. Lynn B. Jorde for providing space for the preparation of the modern mtDNA bait. Finally, we thank Healy Lake Tribal Council and Tanana Chiefs Conference representatives for their support.

References

1. de Acosta J (2002) *Natural and Moral History of the Indies.* (Duke University Press, Durham).
2. Dixon EJ (1999) *Bones, Boats and Bison: Archaeology and the First Colonization of Western North America* (University of New Mexico Press, Albuquerque).
3. Meltzer DJ (2010) *First peoples in a new world: colonizing ice age America* (Univ of California Press).
4. Hoffecker JF, Elias SA, & O'Rourke DH (2014) Out of Beringia? *Science* 343(6174):979-980.
5. Goebel T, Waters MR, & O'Rourke DH (2008) The Late Pleistocene Dispersal of Modern Humans in the Americas. *Science* 319(5869):1497-1502.
6. Tamm E*, et al.* (2007) Beringian standstill and spread of Native American founders. *PLoS One* 2(9):e829.
7. Reich D*, et al.* (2012) Reconstructing native American population history. *Nature* 488(7411):370-374.
8. Raghavan M*, et al.* (2014) The genetic prehistory of the New World Arctic. *Science* 345(6200):1255832.
9. Raghavan M*, et al.* (2013) Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature*.
10. Mulligan CJ, Kitchen A, & Miyamoto MM (2008) Updated three-stage model for the peopling of the Americas. *PLoS One* 3(9):e3199.
11. Dillehay TD ed (1997) *The Archaeological Context and Interpretation* (Smithsonian Institution Press, Washington, DC), Vol 2.
12. Achilli A*, et al.* (2008) The phylogeny of the four pan-American MtDNA haplogroups: implications for evolutionary and disease studies. *PLoS One* 3(3):e1764.
13. Cui Y*, et al.* (2013) Ancient DNA Analysis of Mid-Holocene Individuals from the Northwest Coast of North America Reveals Different Evolutionary Paths for Mitogenomes. *PLoS One* 8(7):e66948.
14. Fagundes NJ*, et al.* (2008) Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. *The American Journal of Human Genetics* 82(3):583-592.
15. Raghavan M*, et al.* (2015) Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science*.
16. Pickrell J & Reich D (2014) Towards a new history and geography of human genes informed by ancient DNA. *bioRxiv*.
17. Chatters JC*, et al.* (2014) Late Pleistocene Human Skeleton and mtDNA Link Paleoamericans and Modern Native Americans. *Science* 344(6185):750-754.
18. Rasmussen M*, et al.* (2014) The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* 506(7487):225-229.
19. Rasmussen M*, et al.* (2015) The ancestry and affiliations of Kennewick Man. *Nature* 523(7561):455-458.
20. Raff JA, Bolnick DA, Tackney J, & O'Rourke DH (2011) Ancient DNA perspectives on American colonization and population history. *American journal of physical anthropology* 146(4):503-514.
21. Potter BA, Irish JD, Reuther JD, Gelvin-Reymiller C, & Holliday VT (2011) A terminal Pleistocene child cremation and residential structure from eastern Beringia. *Science* 331(6020):1058-1062.
22. Potter BA, Reuther JD, Bowers PM, & Gelvin-Reymiller C (2008) Little Delta Dune site: A Late Pleistocene multi-component site in Central Alaska. *Curr. Res. Pleistocene* 25:132.

23. Potter BA, Irish JD, Reuther JD, & McKinney HJ (2014) New insights into Eastern Beringian mortuary behavior: A terminal Pleistocene double infant burial at Upward Sun River. *Proceedings of the National Academy of Sciences* 111(48):17060-17065.
24. Maricic T, Whitten M, & Pääbo S (2010) Multiplexed DNA Sequence Capture of Mitochondrial Genomes Using PCR Products. *PLoS ONE* 5(11):e14004.
25. Der Sarkissian C*, et al.* (2014) Mitochondrial Genome Sequencing in Mesolithic North East Europe Unearths a New Sub-Clade within the Broadly Distributed Human Haplogroup C1. *PLoS ONE* 9(2):e87612.
26. Perego UA*, et al.* (2010) The initial peopling of the Americas: a growing number of founding mitochondrial genomes from Beringia. *Genome research* 20(9):1174-1179.
27. Soares P*, et al.* (2009) Correcting for purifying selection: an improved human mitochondrial molecular clock. *The American Journal of Human Genetics* 84(6):740-759.
28. Fu Q*, et al.* (2013) A revised timescale for human evolution based on ancient mitochondrial genomes. *Current Biology* 23(7):553-559.
29. Bouckaert R*, et al.* (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS computational biology* 10(4):e1003537.
30. Starikovskaya EB*, et al.* (2005) Mitochondrial DNA diversity in indigenous populations of the southern extent of Siberia, and the origins of Native American haplogroups. *Annals of human genetics* 69(1):67-89.
31. Raff J, Tackney J, & O'Rourke DH (2010) South from Alaska: a pilot aDNA study of genetic history on the Alaska Peninsula and the Eastern Aleutians. *Human biology* 82(5/6):677-693.
32. Derenko M*, et al.* (2010) Origin and Post-Glacial Dispersal of Mitochondrial DNA Haplogroups C and D in Northern Asia. *PLoS ONE* 5(12):e15214.
33. Ebenesersdóttir SS*, et al.* (2011) A new subclade of mtDNA haplogroup C1 found in icelanders: Evidence of pre-columbian contact? *American journal of physical anthropology* 144(1):92-99.
34. Reuther JD (2013) Late glacial and early Holocene geoarchaeology and terrestrial paleoecology in the lowlands of the middle Tanana Valley, subarctic Alaska. (University of Arizona).
35. Liversidge H & Molleson T (2004) Variation in crown and root formation and eruption of human deciduous teeth. *American journal of physical anthropology* 123(2):172-180.
36. AlQahtani SJ (2008) *Atlas of tooth development and eruption.* (Queen Mary University of London, Barts and the London School of Medicine and Dentistry).
37. Schutkowski H (1993) Sex determination of infant and juvenile skeletons: I. Morphognostic features. *American journal of physical anthropology* 90(2):199-205.
38. Turner CG, Nichol CR, & Scott GR (1991) *Advances in Dental Anthropology*, eds Kelly M & Larsen C (Wiley-Liss, New York), pp 13-32.
39. Sciulli PW (1998) Evolution of the dentition in prehistoric Ohio Valley Native Americans: II. Morphology of the deciduous dentition. *American journal of physical anthropology* 106(2):189-205.
40. Rohland N & Hofreiter M (2007) Ancient DNA extraction from bones and teeth. *Nature protocols* 2(7):1756-1762.
41. Rohland N & Hofreiter M (2007) Comparison and optimization of ancient DNA extraction. *Biotechniques* 42(3):343.
42. Krishnan A, Sweeney M, Vasic J, Galbraith D, & Vasic B (2011) Barcodes for DNA sequencing with guaranteed error correction capability. *Electronics letters* 47(4):236-237.
43. Gansauge M-T & Meyer M (2013) Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nature protocols* 8(4):737-748.

44. Dabney J*, et al.* (2013) Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences*.
45. Li H*, et al.* (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078-2079.
46. DePristo MA*, et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics* 43(5):491-498.
47. Thorvaldsdóttir H, Robinson JT, & Mesirov JP (2012) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in bioinformatics*:bbs017.
48. Jónsson H, Ginolhac A, Schubert M, Johnson PL, & Orlando L (2013) mapDamage2. 0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*:btt193.
49. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal* 17(1):pp. 10-12.
50. Briggs AW*, et al.* (2007) Patterns of damage in genomic DNA sequences from a Neandertal. *Proceedings of the National Academy of Sciences* 104(37):14616-14621.
51. Katoh K & Standley DM (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* 30(4):772-780.
52. van Oven M & Kayser M (2009) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human Mutation* 30(2):E386-E394.
53. Lanfear R, Calcott B, Ho SYW, & Guindon S (2012) PartitionFinder: Combined Selection of Partitioning Schemes and Substitution Models for Phylogenetic Analyses. *Molecular Biology and Evolution* 29(6):1695-1701.
54. Stamatakis A (2014) RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics*.
55. Yang Z (2007) PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution* 24(8):1586-1591.

Figure Legends

Fig 1. Geographic map of reported Native American populations with >40% C1 or B2 haplogroup frequencies, as well as locations of archaeological sites discussed.

The locations of the Upward Sun River site, as well as the seven previously reported archaeological sites dated at >8,000 yr bp with successfully genotyped human mitochondrial DNA lineages, are listed on the map (with reported haplotypes). Reported populations of ≥ 20 individuals with ≥ 40% C1 (yellow) or B2 (blue) are shown. Populations and frequencies specific to this figure (referenced by numbers 1 to 50) are available in the **Supporting Information**.

Fig 2. Maximum Parsimony phylogenetic trees of (A) C1b and (B) B4 mtDNA haplotypes.

Hand curated MP trees of a subset of the sequences analyzed in this study, showing the placement of USR1 and USR2 in the known human mtDNA tree. Sequences used in our analysis are shown in green, while node assignments are show in red.

Fig 3: Maximum Likelihood Phylogenies of (A) haplogroup C1 and (B) haplogroup B2.

Native American tips and clades are highlighted in blue. USR1 and USR2 are highlighted in red. Some clades have been collapsed for space. Non-parametric bootstrap support for branches are noted for uncollapsed clades with support ≥ 80%. The Maximum Likelihood (ML) clade coalescence times using two alternative molecular clocks (27, 28; see text) or using BEAST 2.2 (29; mean (95% HPD interval)) are noted for clades C1b and B2. Lower scale bars represent branch length in average nucleotide substitutions per site.

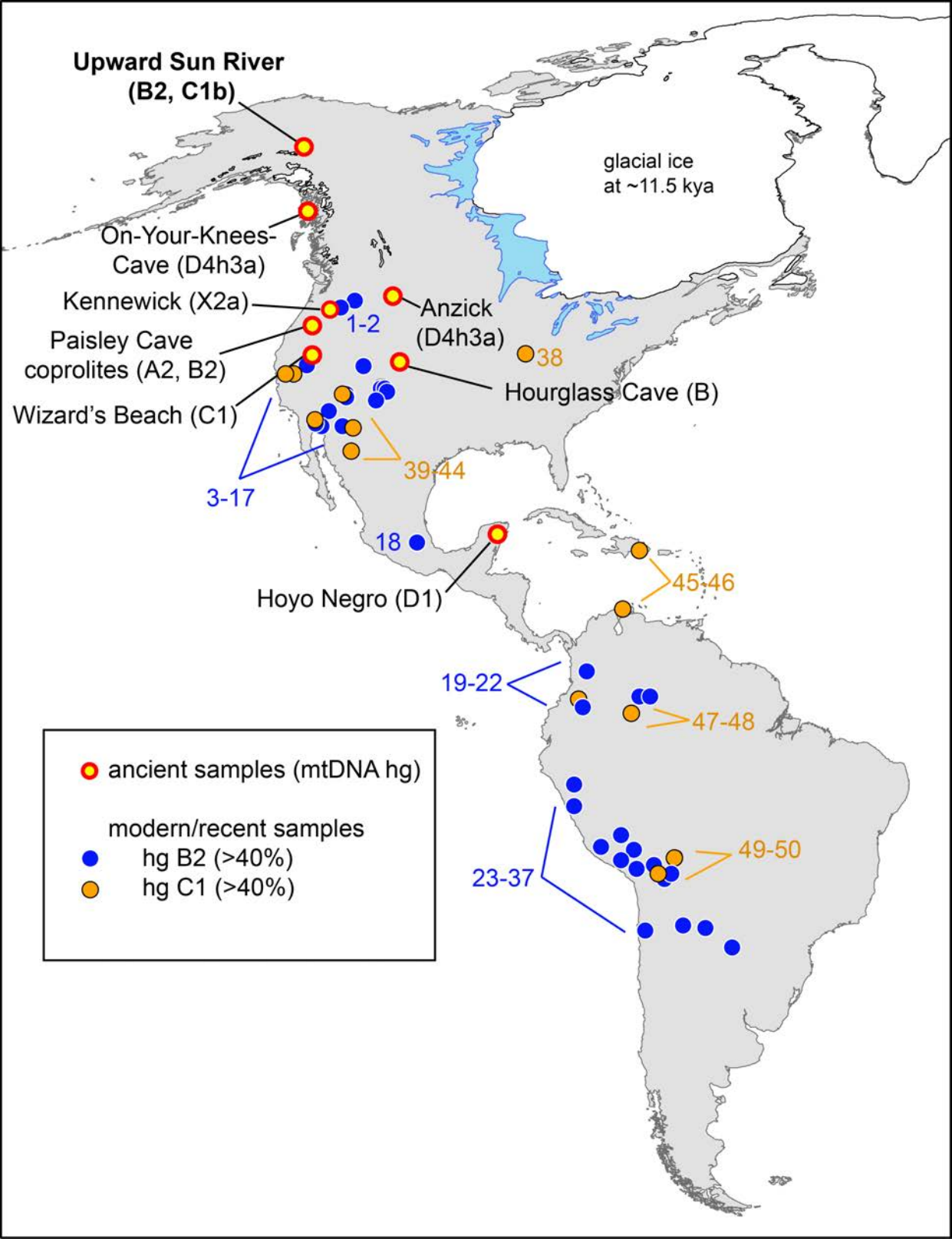
Supporting Information

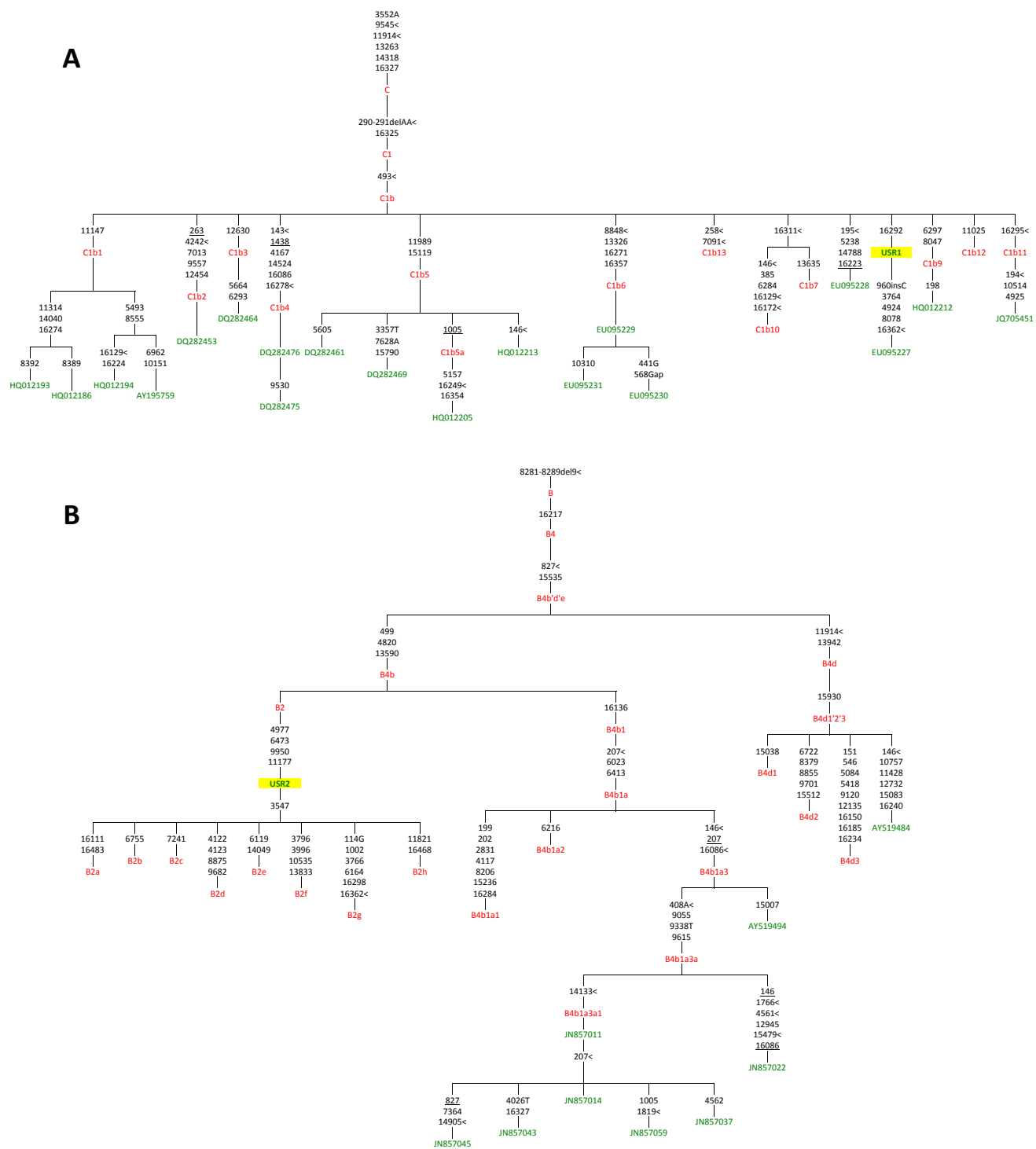
Detailed Materials and Methods

Figures. S1 to S6

Tables S1 to S5

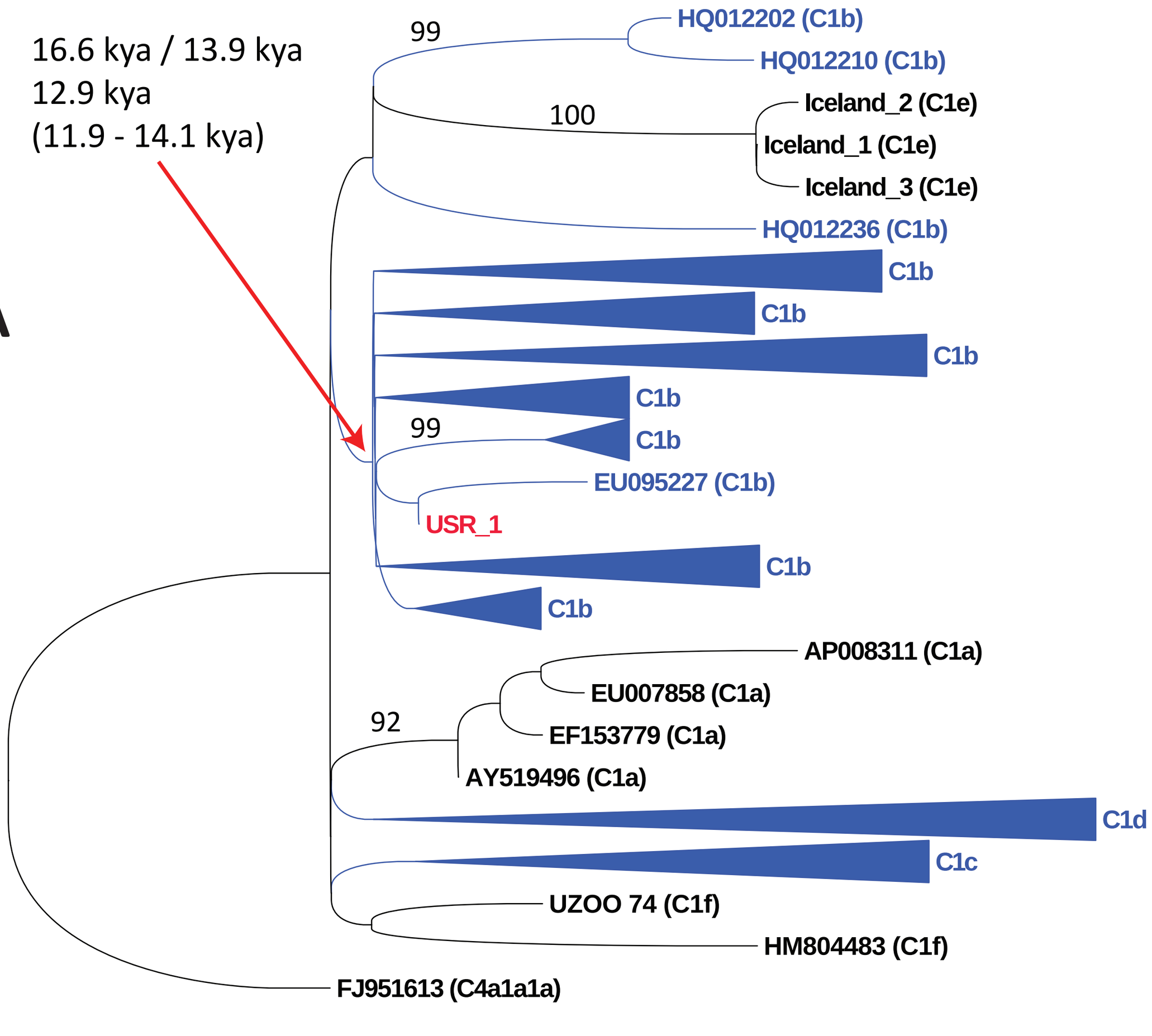
References (35 to 53)


Upward Sun River
(B2, C1b)
glacial ice
at ~11.5 kya
On-Your-Knees-
Cave (D4h3a)
Kennewick (X2a)
1-2
Anzick
(D4h3a)
Paisley Cave
coprolites (A2, B2)
38
Hourglass Cave (B)
Wizard's Beach (C1)
39-44
3-17
18
Hoyo Negro (D1)
45-46
19-22
47-48
49-50
23-37
ancient samples (mtDNA hg)
modern/recent samples
hg B2 (>40%)
hg C1 (>40%)


A
3552A
9545<
11914<
13263
14318
16327
C
290-291delAA<
16325
C1
493<
C1b
11147
C1b1
11314
14040
16274
8392
HQ012193
8389
HQ012186
5493
8555
16129<
16224
HQ012194
6962
10151
AY195759
263
4242<
7013
9557
12454
C1b2
DQ282453
12630
C1b3
5664
6293
DQ282464
143<
1438
4167
14524
16086
16278<
C1b4
DQ282476
9530
DQ282475
11989
15119
C1b5
5605
DQ282461
3357T
7628A
15790
DQ282469
1005
C1b5a
5157
16249<
16354
HQ012205
146<
HQ012213
8848<
13326
16271
16357
C1b6
EU095229
10310
EU095231
441G
568Gap
EU095230
258<
7091<
C1b13
16311<
146<
385
6284
16129<
16172<
C1b10
13635
C1b7
195<
5238
14788
16223
EU095228
16292
USR1
960insC
3764
4924
8078
16362<
EU095227
6297
8047
C1b9
198
HQ012212
11025
C1b12
16295<
C1b11
194<
10514
4925
JQ705451
B
8281-8289del9<
B
16217
B4
827<
15535
B4b'd'e
499
4820
13590
B4b
B2
4977
6473
9950
11177
USR2
3547
16111
16483
B2a
6755
B2b
7241
B2c
4122
4123
8875
9682
B2d
6119
14049
B2e
3796
3996
10535
13833
B2f
114G
1002
3766
6164
16298
16362<
B2g
11821
16468
B2h
16136
B4b1
207<
6023
6413
B4b1a
199
202
2831
4117
8206
15236
16284
B4b1a1
6216
B4b1a2
146<
207
16086<
B4b1a3
408A<
9055
9338T
9615
B4b1a3a
15007
AY519494
14133<
B4b1a3a1
JN857011
207<
827
7364
14905<
JN857045
4026T
16327
JN857043
JN857014
1005
1819<
JN857059
4562
JN857037
146
1766<
4561<
12945
15479<
16086
JN857022
11914<
13942
B4d
15930
B4d1'2'3
15038
B4d1
6722
8379
8855
9701
15512
B4d2
151
546
5084
5418
9120
12135
16150
16185
16234
B4d3
146<
10757
11428
12732
15083
16240
AY519484

A

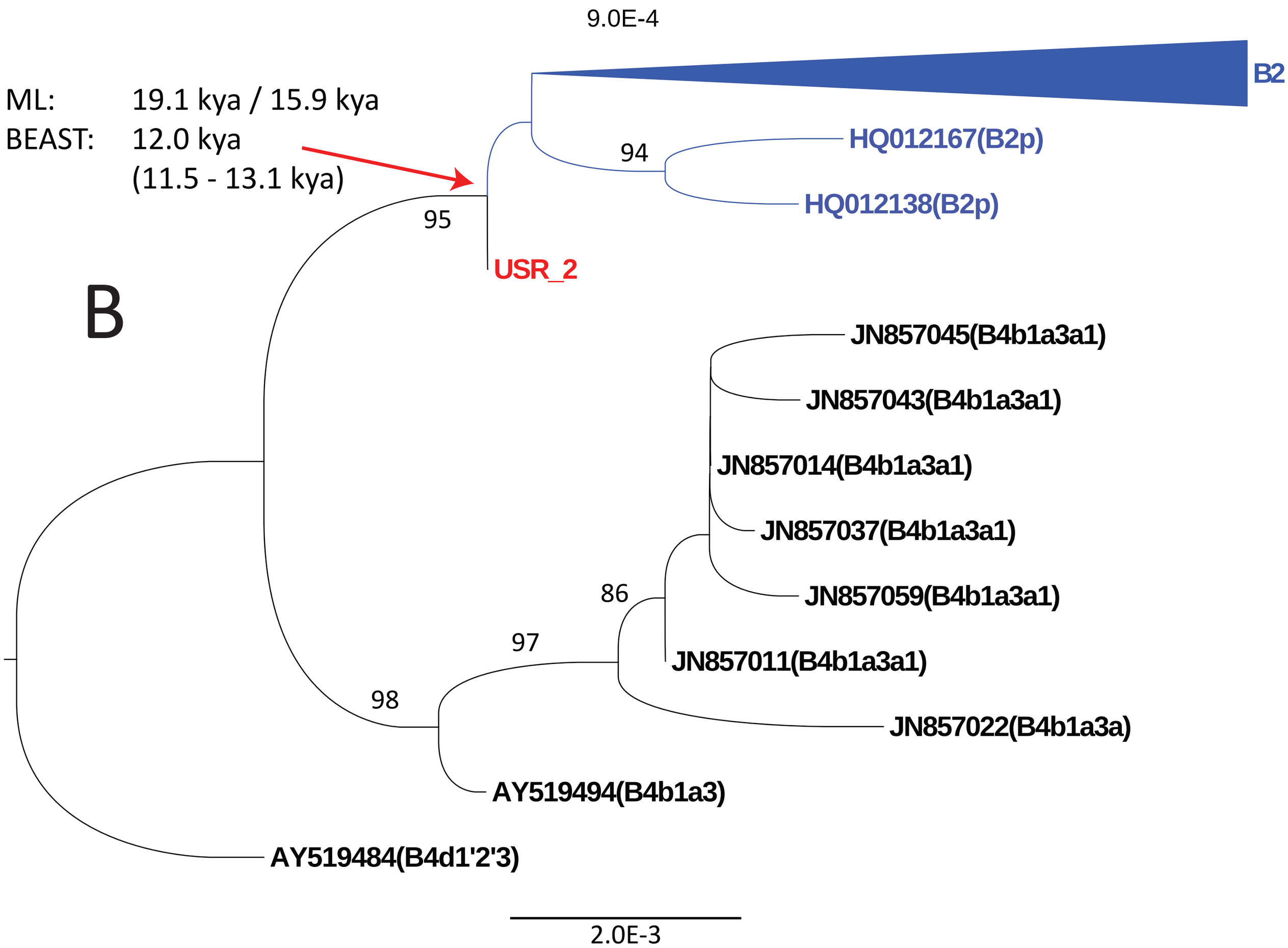
ML: 16.6 kya / 13.9 kya
BEAST: 12.9 kya
(11.9 - 14.1 kya)



9.0E-4

B

ML: 19.1 kya / 15.9 kya
BEAST: 12.0 kya
(11.5 - 13.1 kya)



2.0E-3

Supporting Information

Detailed Materials and Methods

Description of samples and archaeological context

Both infants were recovered from a burial pit at the Upward Sun River site (USR), located in the middle Tanana River valley (49XBD-298) (**Fig S5**). Details on site formation, chronology, and evaluation of site disturbance have been previously reported (21-23, 34). Four components dating between ~13,200 and ~10,000 cal BP are securely dated through a suite of 27 feature and stratigraphic dates (21, 23). The infants were recovered within a massive aeolian silt (Unit IV) at ~160-170 cm below surface, within the burial pit; the upper cremated child was recovered at ~80 cm below surface. Several continuous and discontinuous Ab horizons (Ab4) reflecting Typic Cryorthent shrub tundra-derived soils are expressed between 50 and 260 cm below surface. Post-depositional natural disturbance is interpreted to be minimal given the Ab horizons are level across the site with smooth and very abrupt horizon boundaries and limited evidence of faunal burbation and microfaulting that did not intersect cultural features. Component integrity is considered high given a thin vertical distribution of cultural materials and sharp hearth feature boundaries.

Organic preservation is excellent within the lower deposits of the USR. The rapid burial of remains by wind-blown silt (loess) and very fine sand during the terminal Pleistocene and early Holocene (13,200 to 9,000 cal BP) created a buffer of over 1 m of sediment between the interred individuals and the cremation and the more acidic coniferous dominated forest soils of the middle and late Holocene (8,000 cal BP to present; (34)). The pH values for the terminal Pleistocene and early Holocene deposits at USR show relatively high alkaline characteristics range between 9.5 and 6.90 with average of 8.78. The pH values for the sediments surrounding the burial and cremation pit range between 9.26 and 7.34, and average 7.95 in pH. The alkaline nature of the sediments surrounding the burials and cremation is also evident in the presence of calcium carbonate root casts. The sediments and soils at USR begin to trend towards more acidic values (6.90 to 5.60 in pH) around 1 meter above the burials.

All three sets of human remains are associated with Feature F2010-5/2011-13 from Component 3 at USR, dating to the terminal Pleistocene – Holocene boundary (23). The cremated child (Individual 3, not analyzed here) was found within F2010-5, a pit hearth associated with dense charcoal concentrations, burned bone and oxidized sediment extending to 80 cm BD (~43 cm below the occupation surface). The occupation surface is composed of numerous lithics and charcoal fragments in a thin, unimodal vertical distribution. Two charcoal samples from the base of this hearth (Beta-280585, 280586, both *Populus balsamifera*), and one sample from the top of the backfilled pit hearth (Beta-280584, *Populus/Salix*) were statistically of the same age. Pit fill (designated F2011-13) was encountered below the oxidized sediment. The two infants and grave goods (3 bifaces and four antler rods comprising a hunting toolkit of two hafted dart/spear projectile points and foreshafts (23)) were found at the base of the pit at 124-128 cm BD, or 44-48 cm below the upper hearth. A piece of charcoal adhering to one of the antler rods yielded a statistically similar date (Beta-371567, *Betula* sp.) to the three other dates ($\chi^2 = 7.81$, df=3, $p<0.05$), with a mean pooled radiocarbon age of 9970+/-30 BP (11,600-11,270 cal BP). These ages are consistent with the dating on other hearths and are intermediate between Component 2 and Component 4 radiocarbon dates (21, 22).

Individuals 1 and 2 were complete and located 8-10 cm apart at the bottom of the pit feature. Deciduous crown development (35, 36) indicates that Individual 1 died at 6-12 weeks post-natal and Individual 2 died at >30 prenatal weeks. Morphological analyses may suggest that both individuals are female(37).

No skeletal pathologies were evident. Results from dental non-metric trait analysis in Individual 1 are consistent with a Native American population affinity (38, 39).

Two petrous specimens were selected for aDNA analyses given their overall mass and high density (**Fig S6**). Specimen #58-311 from Individual 1 had an initial weight of 3.4 g and specimen #58-80 from Individual 2 had an initial weight of 1.2 g. Ochre covered the remains, and was variable for Individual 1 specimens, including 58-311, ranging from 10R 3/2 (dusky red) to 5YR 4 (reddish brown), while Individual 2 was more homogeneous in staining, generally 2.5YR 4/4 (yellowish red).

Excavation protocols are described in (23). Field protocols for collection of human remains included powder-free nitrile gloves and facemasks and minimal contact with remains between excavation and bagging. Each specimen was identified and catalogued by Joel Irish at the University of Alaska Fairbanks Archaeology Lab.

Legal and Ethical Issues Pertaining to the USR specimens

The USR burials were found on land owned by the State of Alaska. Before excavations were initiated in 2010, a Memorandum of Agreement (MOA) was signed by the State of Alaska and the lead federal agency (National Science Foundation) with Healy Lake Tribal Council, the local BIA-recognized tribal authority and the Tanana Chiefs Conference, the regional non-profit Native organization, as invited signatories. This MOA stipulated the process to be followed if human remains were conducted, following the Native American Graves Protection and Repatriation Act (NAGPRA). After the remains were discovered, an amendment was signed by all parties that allowed for destructive analysis on very small portions of the skeletal remains to determine age, dietary evidence through stable isotope analyses and genetic relationships through aDNA analyses.

DNA Extraction and mitochondrial HVR1 Sanger sequencing

DNA was extracted using a silica-based method, as is typically applied in the field (40, 41). The original specimens were either already highly fragmented or brittle/burnt, so no drilling was performed. 80-120 mg of sample was digested in a 1mL buffer consisting of 0.5M EDTA, 250 µg/ml proteinase K, and 40 mM DTT at 37°C overnight with rotation. The extraction buffer was spun down, and the released DNA molecules in the supernatant were mixed with 4 mL of Guanidine Thiocyanate-based Dehybernation Solution A and 200 µl of Ancient DNA GLASSMILK^{TM} (silica suspension) components of the GENECLEAN For Ancient DNA Kit (MP Biomedicals), along with a final concentration of .05% Tween-20. This solution was incubated at 37°C with rotation for 3 hours. The silica particles were collected and purified as per the manufacturer's protocol for the rest of the GENECLEAN kit. Final elution using two rounds of 30 µl of TE^{-4}(10mM Tris, 0.1mM EDTA) + 0.02% Tween-20 was performed. DNA extracts were stored in LoBind tubes at -20 °C. One water extraction blank was processed at the same time as the samples.

A portion of the mtDNA hypervariable region 1 (HVR1) was amplified and sequenced as described in (31). The extraction blank and numerous water PCR blanks were processed at the same time. Variants present in np 16043-16161, 16183-16277, and 16288-16402 were typed. Because these variants suggested Native American haplotypes, and no product was evident at any point in the blanks, these extracts were chosen for next-generation sequencing.

Ion Torrent Library Preparation

Libraries were prepared as per the Ion Plus Fragment Library Kit (Life Technologies) with the following modifications: No DNA fragmentation or size selection at any point was performed. All solid-phase reversible immobilization (SPRI) bead purification steps were replaced with silica-column clean-ups (Clean & Concentrator™-5; Zymo Research). Ion A Adapters were created with the suggested TT tails and with laboratory-specific 8-base barcodes, based on an unique set of guaranteed error correcting codes that are redundant up to 2-bit errors (at most 1 nucleotide error in base space or 2 errors in flow space) (42) (**Table S3**). During adapter ligation and nick repair, final adapter concentrations in the reaction were reduced to .04 µM. Unamplified libraries were eluted into 22 µl TE^{-4} and initially quantified by qPCR (GeneRead Library Quantification Kit; NGTF-ITZ-F Qiagen) to both determine molecule concentrations and optimal cycles for amplification (43). Ten microliters of the unamplified libraries were used in a 100 ul primary library amplification reaction with AmpliTaq Gold 360 Master Mix (Life Technologies). This primary amplification was limited to ≤ 15 cycles. Final primary amplification products were eluted into 40 µl of TE^{-4} with UltraClean PCR Clean-Up Kits (MO BIO Laboratories, Inc). All previous and subsequent amplifications were done with Ion_Aamp and Ion_P1amp amplification primers at 0.4 µM final concentrations (**Table S3).**

Hybridization Capture and Sequencing of mitochondrial DNA

Hybridization capture of mitochondrial DNA was performed as in (24), using the lower hybridization and wash temperatures of (44), for ~48 hours. In a modern genetics laboratory in a separate building, two long range PCR amplicons were created (with Phusion Hotstart Flex; NEB) from an African mtDNA with haplotype L2a4a, with private mutations (514T, 516T, 573.XC!, 6254G, 16188.C, 16319A, 16519C). The amplicons were mixed in equimolar amounts and fragmented with a Covaris S2 down to 100bp target peaks. Biotinylated bait molecules were immobilized on Dynabeads MyOne Streptavidin C1 beads (Life Technologies). Each library was captured separately with ~270 ng of bait, 320-613 ng of library (from multiple secondary amplifications with Q5 Hot Start mastermix; NEB), and blocking oligos appropriate to our Ion Torrent adapters (**Table S3**) at 1.9 µM working concentrations. Libraries were released from the beads with a final 5 minute incubation at 95°C. Eluted captured molecules were quantified via qPCR, amplified with Q5 Hot Start mastermix into the exponential phase (43), and purified with silica columns.

Prior to sequencing, the libraries were assessed for concentration and fragment size distribution using a fragment analyzer (FA; Analytical Technologies Inc). The FA results did not yield measureable amounts of DNA. A quantitative real time PCR analysis generated using the GeneRead Kit, however, did detect properly ligated libraries. The libraries were diluted to 100pM based upon the quantitative real time PCR results. The individual libraries were further diluted to 10pM prior to amplification in the templating reaction using the Ion Torrent One Touch 2 (Life Technologies) and the Ion PI Templating OT2 200 v3 kit (Life Technologies). After the amplification step by emulsion PCR, an enrichment step was performed on the Ion Torrent ES to enrich for positive Ion Sphere Particles (ISPs). While the enrichment step was being performed, a 2uL aliquot of the post emulsion PCR reaction, taken before the enrichment step, was evaluated separately with the Ion Sphere Quality Control kit to determine the percentage pre-enrichment of the templating reaction. The manufacture suggests a pre-enrichment percentage between 10% and 30% ; these libraries were 10.48% and 10.57% for USR1 and USR2, respectively. After capturing enriched ISPs, sequencing was performed on the Ion Torrent Proton using the Ion PI Sequencing 200 v3 Kit and Ion PIv2 chip (Life Technologies).

Torrent Suite Data Processing and Mapping

Torrent Suite 4.0.2 was used for the initial read processing off of the Proton sequencer. In our experience next-generation sequencing tools available online are customized for Illumina sequencing chemistry and error profiles. These tools perform sub-optimally on exported FASTQ files from Ion Torrent reads. Additionally, FASTQ files do not contain flow space information (flow order and flow signal), used by

sequencing-by-synthesis methods, which is necessary to fully leverage Proton read processing, as well as the Torrent Variant Caller software package. We initially chose this analytical pipeline to take advantage of these strengths of the Ion Proton technology, while retaining some of the customizability that is offered by typical freeware programs and scripts.

For Torrent Suite 4.0.2, default parameters were utilized with the following changes: We omitted all sequences below 30bp in length (post quality trimming) with additional Basecaller arguments (--trim-min-read-len 30 --min-read-length 30; note default barcode settings allow for 2 errors in flow space). We mapped against rCRS (NC_012920) using TMAP (stage1 map2 map3 map4, and allowing the default 3' soft clipping function). We incorporated 'Base Recalibration', 'Mark as Duplicates', and 'Enable Realignment', and we finished by running the FilterDuplicates plugin. The Torrent Suite uses an Ion-optimized duplicate filtering approach that takes into account not only the 5' alignment start site but the 3' adapter flow position (if the read extends into it) as well. We have found that this approach retains more unique sequences than the Samtools (45) rmdup function, which is optimized for paired-end Illumina sequencing reads.

Mapped BAMs (after duplicate filtration) were processed to remove reads with mapping quality (MAPQ) <30 using Samtools 0.1.19 (45). A stand-alone version of Torrent Variant Caller (TVC) 4.2.3 (optimized for Ion Torrent reads with flow space information) was used to call variants, with a custom parameter file at a high-stringency setting to minimize false-positive calls and optimize for a haploid genome. A consensus file was created from the vcf and the rCRS (NC_012920) using the reference utility FastaAlternateReferenceMaker of the GATK(46) version found within TVC 4.2.3. In two situations (the SNPs at 16182/16183 and A9545G) the consensus file was manually edited based on the produced VCF to properly note complex variants (namely SNPs in close proximity to other SNPs or indels) that were not called by TVC or properly translated by GATK.

Genomic coverage depth was calculated at a 1 base window size with igvtools (47), replacing TVC FDP depth counts at called deletions. General sequencing QC metrics were analyzed with FastQC v. 0.11.2 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Read length histograms and nucleotide misincorporation patterns were assessed using MapDamage v2.0.2-12 (48) (--length 120 --seq-length 20 --forward). Haplotypes were initially assigned using mtPhyl 4.015 (https://sites.google.com/site/mtphyl/home) and then manually confirmed or clarified following the latest nomenclature on PhyloTree.org (Build 16 (19 Feb 2014)).

After Torrent Suite analysis, it became apparent that faulty adapter design issues caused low quality base calls at the 5' end of all reads and prevented us from investigating typical aDNA damage patterns (see later Authentication). This is likely the result of our custom adapters lacking a 'GAT' barcode adapter sequence between the barcode and the start of the ligated fragment. A joining sequence is suggested to avoid a two-mer (or more) incorporation at the end of the barcode during Ion semiconductor sequencing, which would result in the software not correctly identifying and clipping the barcode, and potentially cause low quality base calling in the Torrent Suite software package. We therefore re-analyzed the reads outside of the Torrent Suite, which we outline in the subsequent section. This also provided us an opportunity to validate the Torrent Suite analytical pipeline.

Alternative Bioinformatics Data Processing and Mapping

Reads were re-processed from both Ion PIv2 chips within Torrent Suite 4.0.2 without a reference genome (no mapping or related settings), without reads below 30bp in length (--trim-min-read-len 30 --min-read-length 30), and with default Ion quality trimming for version 4.0.2. Both of the output BAM files were converted to FASTQ files - the barcoded (detected and subsequently trimmed) reads and the no-barcode reads - with Picard Tools v1.91 (http://sourceforge.net/projects/picard/). Of note, SamToFastq removes

Ion flow space data, as Ion uses a non-standard SAM tag to store this information. Cutadapt v1.8 (49) was next used to trim the FASTQ files in the following order: barcodes (for just the no-barcode reads and requiring a perfect match; -n 5 --overlap 8), 3' adapter sequences (10% error tolerance; -n 2 --overlap 6), and read end base quality of ≥20 (--minimum-length 30 -q 20,20). The processed FASTQ files for each sample were merged and were mapped against rCRS (NC_012920) using TMAP (stage1 map2 map3 map4, and without 3' clipping). We used a custom perl script to remove all mapped reads below length of 30bp and above length of 120bp. We then used Samtools 0.1.19 (45) to remove reads with mapping quality (MAPQ) <70 . Finally, we processed the mapped BAMs with Picard Tools MarkDuplicates to mark and remove duplicates. Read metrics were processed as in the original Torrent Suite pipeline.

We are unable to do proper variant calling on BAMs that lack flow space information (Ion does not currently offer the tools to do read manipulations while editing and retaining flow space information; this field cannot be trimmed like quality scores). However, we did view these final BAM files in IGV (47). We note that alignment viewers like IGV are limited when viewing Ion data - the reads reflect the sequences as originally called, not as finally evaluated after flow space re-evaluation. In particular, many variants in and around polynucleotide tracts visible in IGV are actually false calls and disappear after flow space data is taken into account. We were, however, able to visually check each nucleotide position along the rCRS reference genome.

For both USR1 and USR2, all variants called by the earlier pipeline with TVC were unambiguously confirmed in these new mapped BAM files, except for an indel at position 14342 in USR1, which was completely absent in the alternatively processed BAM. All other sites had >90% variant support (going by the simple ratio of derived over reference alleles from the IGV summary). This indicates that the consensus sequences created by the initial Torrent Suite pipeline are likely correct, even after correcting for low quality 5' read ends. For USR1, if we limit variant discovery to derived allele frequencies of ≥ 30% and without taking into account base quality, 25 C>T, 4 G>A, and 4 other nucleotide substitutions are observed. As C-to-T and G-to-A substitutions are expected in ancient DNA sequences (50), observation of these additional SNPs is expected. Eight of these C>T potential variants were within regions that were Sanger sequenced (see later Validation section) and all eight were sequenced as the reference base. All of the remaining observed substitutions were found and discarded in the earlier described high-stringency TVC calling pipeline (albeit with a different processed BAM). The majority of these are near or within polynucleotide tracts. This indicates to us that these sites are all false calls (fixed later by flow space) and/or sites of low level damage. For USR2, many fewer additional substitutions were observed: 1 C>T, 1 G>A, 2 other nucleotide substitutions, and 3 other potential indels. Again, these were all near or within polynucleotide tracts, though none were Sanger sequenced to verify. No additional substitutions in either sample had the visual level of support within IGV as the originally called variants.

Enriched Read Summary

Following enrichment, amplification, and templating of the two libraries, Ion Proton sequencing read counts passing default filters (and ≥ 30 bases) were on the low end of the expected 60-80 million reads specified for the chip. Additionally, of these reads, a higher than expected percentage failed barcode identification (19% and 9.4%; **Table S1**; see above for barcode issues and alternative processing). After the Torrent Suite pipeline, a large percentage of sequences mapped to human mtDNA - 39% for USR1 and 24% for USR2. Following duplicate removal these recoveries dropped to 0.40% and .65%, respectively. This achieved 20,044 high quality reads for the USR1 library and 32,979 high quality reads for the USR2 library (**Table S1**). The large amount of duplicate amplicons in the libraries suggests that both libraries have been sequenced to exhaustion. Read length histograms of the unique, MAPQ ≥30 enriched libraries show mean read lengths of 98 and 99 bp, and median read lengths of 90 and 89 bp, for USR1 and USR2 respectively (**Fig S2A**).

In the alternative bioinformatics pipeline, unaligned reads (lacking 'Base Recalibration' given the absence of a reference genome) with the additional read trimming resulted in 40.3 million and 45.3 million reads for USR1 and USR2, respectively. After mapping 33% for USR1 and 17% for USR2 mapped to the mtDNA genome. Following duplicate removal, length restrictions (30 to 120 base pairs), and a MAPQ threshold of 70, 21,140 and 22,951 reads mapped to USR1 and USR2 respectively. These represent recoveries of 0.04% each **(Table S2**).

Sanger Validation and Contamination Estimates

In addition to the portion of HVS1 previously sequenced (see above), we selected a subset of variants called by the TVC to validate using Sanger sequencing. For USR1 the following variants were validated: T489C, A493G, 523delAC, T3552A, T9540C, A9545G, T14318C, C16223T, C16292T, T16298C, T16325C, C16327T. For USR2 the following variants were validated: G499A, 3547A (ancestral A), 8281-8289d, A16183C, T16189C, T16217C. Additionally, each validated variant was sequenced in the other ancient sample and, as expected based on the assigned haplogroup, found to be the rCRS reference base. The single insertion of a T at np 14342 in USR1 was called by TVC but was not Sanger validated. Apart from indel calling being suboptimal to SNP calling in most variant callers, and the lack of this variant in any known mtDNA sequence on PhyloTree.org or in the alternatively processed BAM, this insertion had the lowest phred-quality score of all called variants. Because this was the only case of a false positive in the variants we validated, we believe our high stringency TVC parameters are working as designed.

Mitochondrial contamination estimates were made by taking advantage of the TVC produced VCF files, specifically the reference allele and read depth observation counts at each called locus in the BAM file as determined by freebayes after flow space evaluation (TVC info tags FRO and FDP). This gives us a rough approximation of possible contaminant reads. For all variants (excluding 14342) in each VCF we calculated the mean, median, and range of the percentage of FRO:FDP. For USR1 these values were 3.5%, 1.64%, and (0 - 23.8%), with an average FDP of 94 and for USR2 these values were 4.9%, 3.4%, and (0-35.9%), with an average FDP of 166. The 23.8% reference allele ratio for USR1 was from the SNP at np 493, with a FDP of 21. This SNP was the second lowest scoring variant in the USR1 VCF. The 35.85% reference allele ratio for USR2 was from the SNP at np 16182, with a FDP of 53. This SNP was the lowest scoring variant in the USR2 VCF, and it is further complicated by the adjacent SNP at 16183; the statistics from the TVC are therefore unreliable at this locus. Out of the 64 called variants in the two samples, only 8 showed reference allele percentages above 5%. Using this metric, the apparent genome-wide contamination rate for both samples is <5%, with expected variation.

Phylogenetic Trees and Coalescence Time Estimates

Two curated lists of previously published whole mtDNA genomes were made, in addition to USR1 and USR2, following manual removal of duplicates and incomplete sequence. For haplogroup C1, 187 C1 sequences included the Asian C1a branch, three Icelandic C1e sequences (33), and one C1f sequence from the Mesolithic UZOO-74 individual (25). A haplotype C4a1a1a from the Teleut of South Siberia(32) was used as the outgroup, for a total of 189 sequences **(Table S4)**. For haplogroup B2, 137 Native American B2 sequences and 8 closely related northern Asian B4b1a3 sequences were included. One haplotype B4d1'2'3 from the Buryats of southern Siberia was used as the outgroup, for a grand total of 147 sequences **(Table S4)**.

Initially a subset of these sequences were selected for a Maximum Parsimony tree created using mtPhyl 4.015 (https://sites.google.com/site/mtphyl/home) and then manually edited for clarity (**Fig 2A & 2B**). Next, MAFFT(51) was used to align all the sequences from each list with the highly accurate L-INS-i

methodology. Once aligned, nucleotide positions representing C inserts between 303-315 (leaving any SNPs at 310), AC indels at 515-523, SNPs at 16182C and 16183C, C inserts between 16184-16193 (leaving any SNPs), and SNPs at 16519 were removed from the alignment. These sites are known mutational hotspots and/or positions with recurrent sequencing errors (52). An appropriate partitioning scheme was chosen using PartitionFinder (53) with the -raxml option and the three partitions of 1-576, 577-16023, and 16024-16569. The general time reversible substitution model with invariant sites and a gamma distribution correction for rate heterogeneity was selected (GTR+I+Γ), partitioning the two control regions separate from the coding region. Maximum Likelihood (ML) phylogenetic trees were constructed using RAxML v. 8.1.15 (raxmlHPC-PTHREADS-SSE3 -T 16 -f d -m GTRGAMMAI) for 200 iterations (-p 'random' -N 200) and 1000 non-parametric bootstrap replicates (-p 'random' -b 'random' -N 1000)(54). Bootstrap support values were written onto the best ML tree (-f b -z RAxML_bootstrap.xxx -t RAxML_bestTree.xxx) and the tree was visualized and formatted with FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/) **(Fig 3A & 3B)**.

Maximum likelihood estimates of coalescence times for the major clades within each of the RAxML trees were calculated with PAML 4.7 (55) using settings that included a global clock, a GTR+Γ mutation model (discrete distribution with 32 categories), and option G (the three partitions as above; Malpha=1). Mutational distances were converted into years using a corrected molecular clock proposed by (27) or a whole genome substitution rate of 2.67×10^{-8} sub/site/year, determined by a Bayesian approach utilizing ten securely dated ancient mitochondrial genomes (28).

Bayesian estimated coalescence times for the C1 and B2 clades were calculated using BEAST v2.2.1 (29). Tip dates were set at 11,500 ya for USR1 and USR2, and 8,300 ya for UZOO-74. For the B2 data set, two Markov chain Monte Carlo runs of 40,000,000 generations each, with samples taken every 5,000 generations, were performed. The runs were combined using LogCombiner v2.2.1, with 10% discarded as burn-in, for a final 72,000,000 total generations. We selected the GTR+I+Γ site model, a Coalescent Bayesian Skyline tree prior (3 populations), and a lognormal clockRate prior (M= 2.67E-8, S = 1.4). The two control partitions were combined (1-576; 16024-16569) and the resulting control and coding partitions were linked with a strict clock model and tree model. We used TreeAnnotator v2.2.1 to produce the Maximum clade credibility tree with a posterior probability limit of 60%, and calculated target clade divergence times (node height) 95% highest posterior density (HPD) intervals from this tree. For the C1 data set, the same workflow was followed, except the tree was fixed at the RAxML best tree and the Coalescent Bayesian Skyline tree prior was estimated with 5 populations.

Authentication of Ancient DNA work

Pre-PCR work was carried out in a dedicated ancient DNA facility, physically isolated from any room with post-PCR amplicons, and found in a building where no modern human DNA work has ever been processed. The laboratory is a state of the art cleanroom that consists of one ISO class 7 (Fed class 10,000) gowning area, two ISO class 6 (Fed class 1000) laboratory spaces, and numerous dedicated laminar flow hoods (ISO class 5/Fed class 100). The entire space is under positive pressure from ceiling mounted HEPA filters, with air flow directed from the 'cleanest' pre-PCR room to the main extraction room to the gowning room to the outside environment. Room-wide UV lighting provides daily surface/air sterilizations. Upon entering all personal must garb in full 'tyvek' cleanroom suits, which are subsequently bleached. The active workspaces in the laboratory are bleached and washed as used, with full lab cleaning scheduled as necessary.

Sample libraries chosen for the Ion Torrent templating reaction (clonal amplification) were required to have at least an order of magnitude more molecules than either of the library blanks; USR1 and USR2 libraries exceeded that standard (see **Table S1**). Each of the library blanks were created with adapters

containing all sample barcodes. These custom barcodes had never previously been used in any run on the Ion Proton machine at the core sequencing facility.

All mtDNA haplotypes are known for all laboratory personnel and none match those determined for USR1 or USR2. Additionally, samples containing haplogroups C and B have been analyzed exceedingly rarely in our ancient DNA facilities and we see no clear contamination source from previously processed samples. The variants posited for URS1 and USR2 do not match our African bait DNA beyond those expected from the human mtDNA tree.

Following the Torrent Suite pipeline, we evaluated the misincorporation patterns of the reads mapping from USR1 and USR2. We were unable to observe any putative damage-induced misincorporations on the 3' end of reads as our TMAP alignment allowed 3' soft-clipping of bases. On the 5' end of reads, we expected an increase of C-to-T substitutions due to the deamination of cytosine to uracil in single-stranded DNA overhangs (50). Instead, we observed an irregular pattern of misincorporations in both samples (**Fig S2B**). Unfortunately, the barcode for USR1 ended in a cytosine and the barcode for USR2 ended in a thymine, exactly the bases involved in typical aDNA damage. The base quality scores on the 5' ends of these reads also showed a drop in quality relative to the remainder of the read, unusual for NGS sequencing. It appeared that this issue was masking any true damage patterns in this region.

Following the alternative bioinformatics pipeline, we were able to minimize the bias from these 5' low quality bases, and reveal true DNA damage patterns at both ends of our sequenced reads **(Fig S4B)**. While the 5' ends still show some non-C-to-T substitutions, and a lower than expected relative frequency of C-to-T substitutions, the expected damage still makes up the majority of substitutions observed. At the 3' ends, we observe the expected rise in G-to-A substitutions, though not as smoothly distributed as some previously reported ancient DNA samples (48)**(Fig S4B)**.

Supporting Figure Legends

Supporting Figure S1 – Sequence coverage after the Torrent Suite pipeline across the mitochondrial genome for USR1 and USR2 on a 1-base sliding window. TVC FDP depth counts at called deletions.

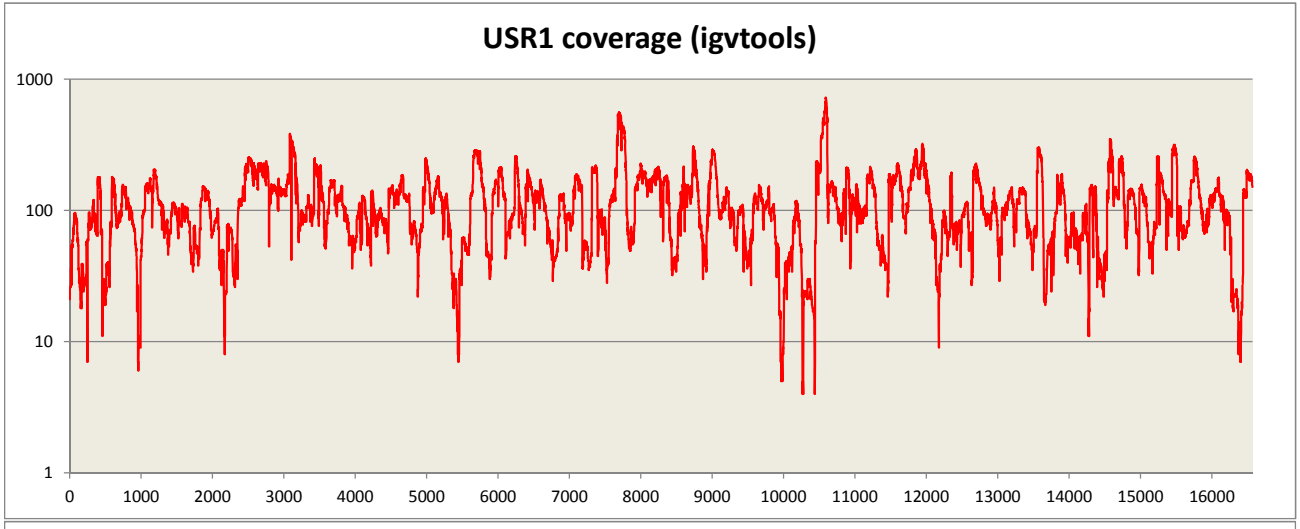
Supporting Figure S2 – (A) Read Length Histograms (B) Position specific substitutions from the 5' end of reads post-Torrent Suite pipeline. All graphs produced by MapDamage v2.0.2-12; C-to-T substitutions are shown in red; G-to-A substitutions are shown in blue; insertions are shown in purple; deletions are shown in green; all other substitutions are shown in grey.

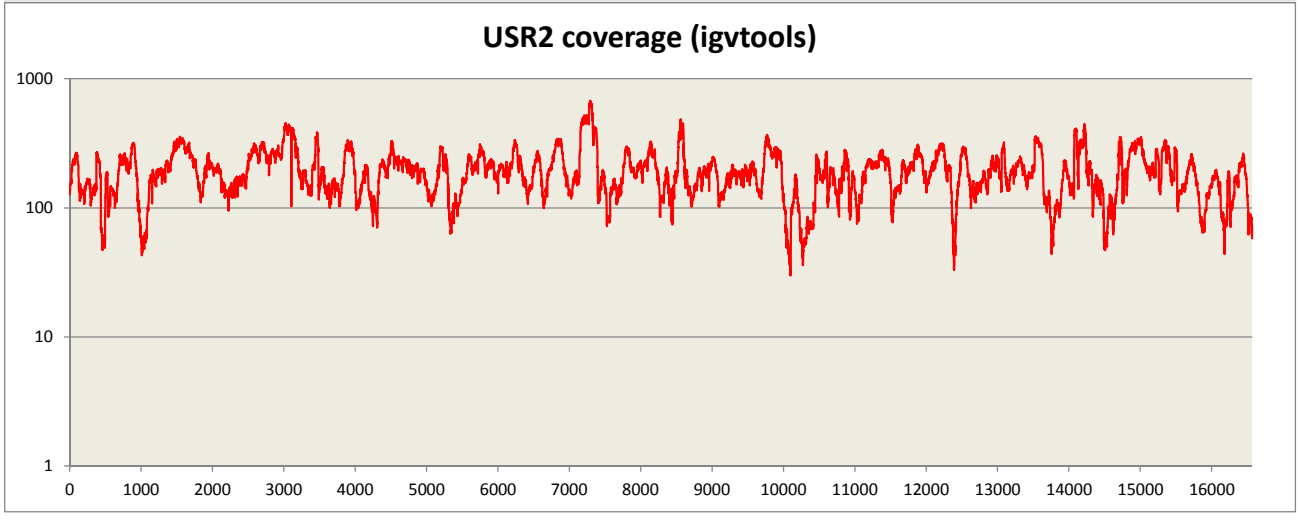
Supporting Figure S3 – Sequence coverage after the alternative pipeline across the mitochondrial genome for USR1 and USR2 on a 1-base sliding window. Coverage across indels are not corrected

Supporting Figure S4 – (A) Read Length Histograms (B) Position specific substitutions from the 5' end of reads (top) and 3' end of reads (bottom) following the alternative bioinformatics pipeline. All graphs produced by MapDamage v2.0.2-12; C-to-T substitutions are shown in red; G-to-A substitutions are shown in blue; insertions are shown in purple; deletions are shown in green; all other substitutions are shown in grey.

Supporting Figure S5 - Plan view of burial and locations of aDNA samples, FS311 (Individual 1) and FS80 (Individual 2). Inset. Location of Upward Sun River.

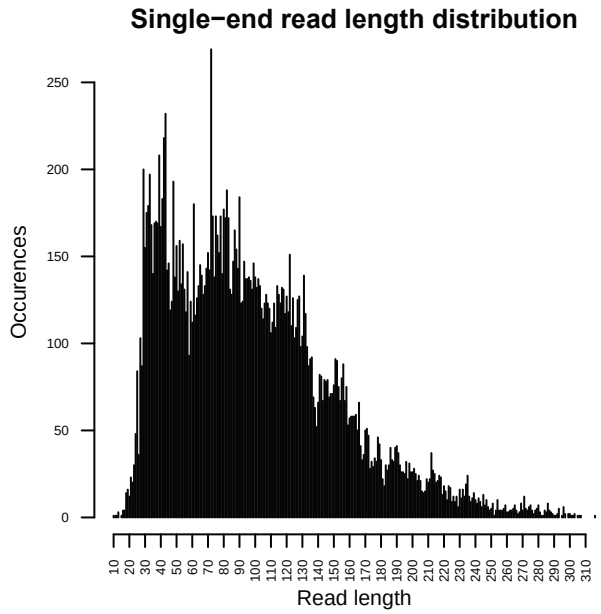
Supporting Figure S6 - aDNA samples, FS311 (Individual 1), left, and FS80 (Individual 2), right. Scale bar = 2 cm.


USR1 coverage (igvtools)
1000
100
10
1
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
11000
12000
13000
14000
15000
16000

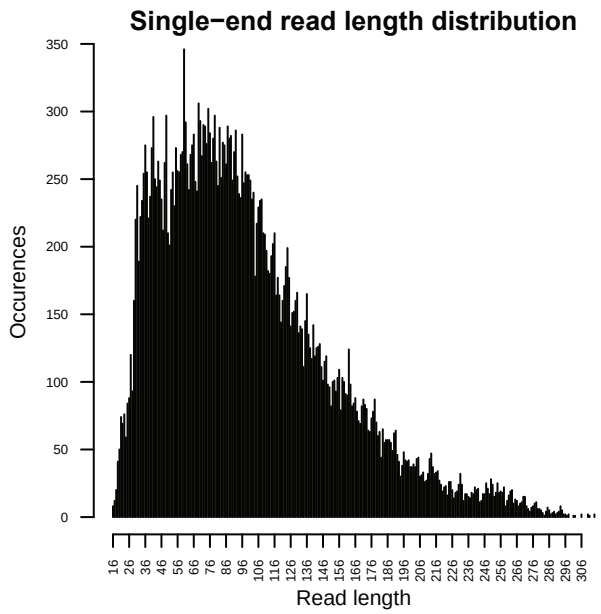

USR2 coverage (igvtools)
1000
100
10
1
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
11000
12000
13000
14000
15000
16000

(A)

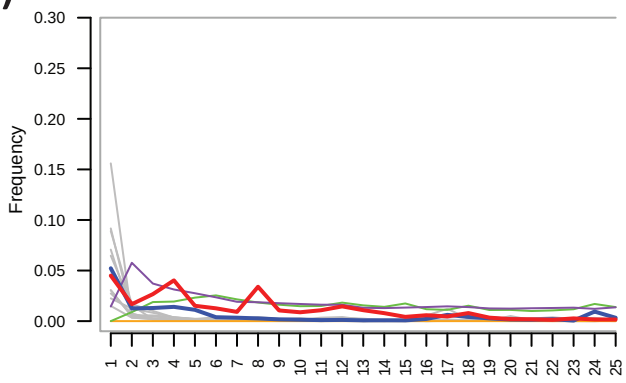
USR1

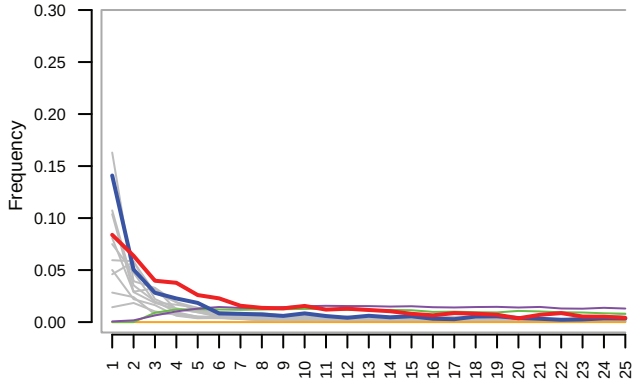

Single−end read length distribution
Occurences
Read length

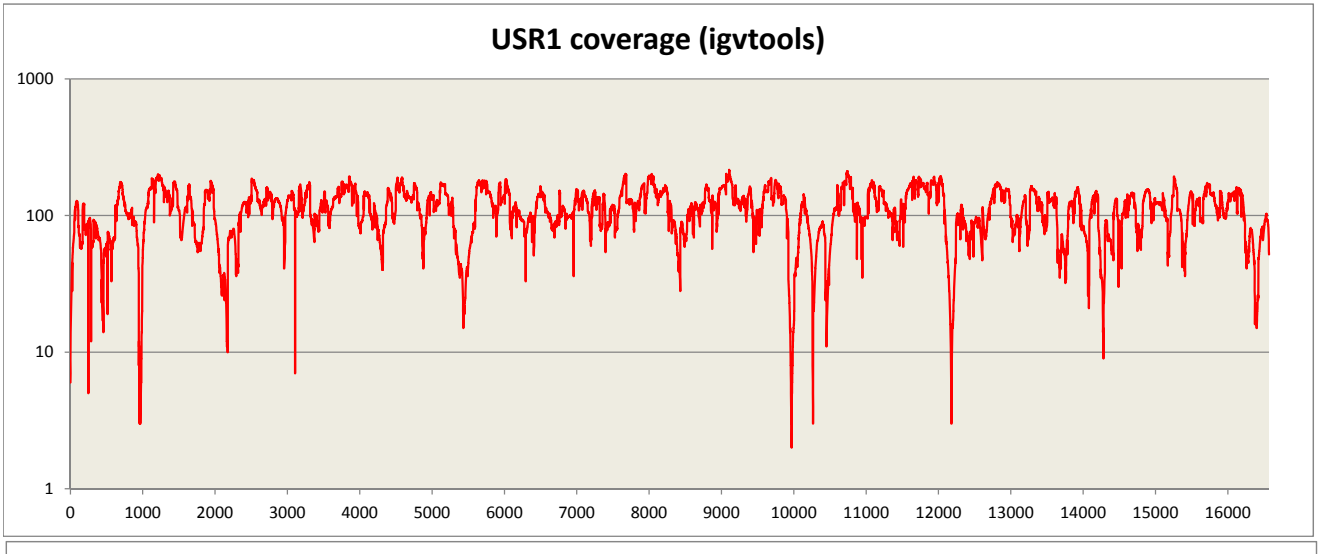
USR2

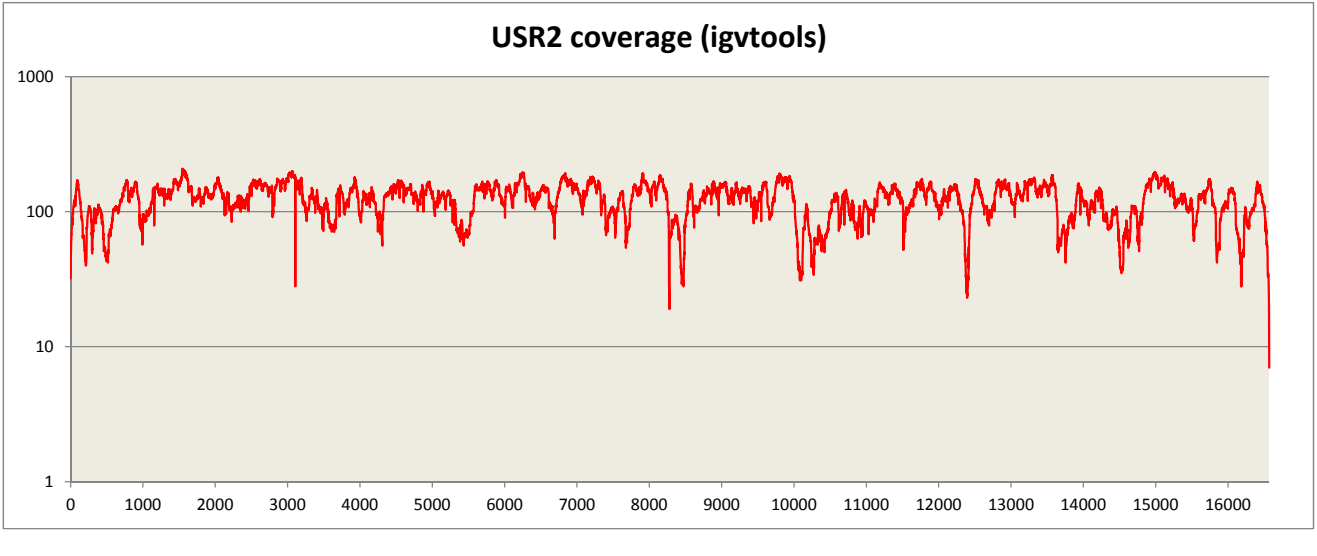

Single−end read length distribution
Occurences
Read length

(B)


Frequency

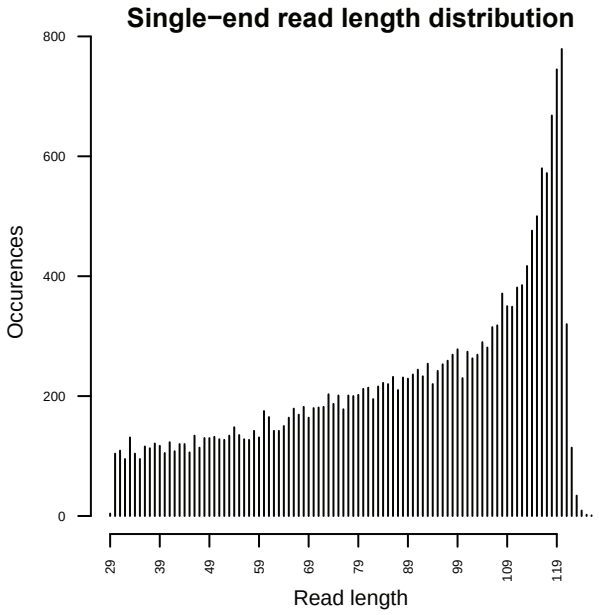

Frequency


USR1 coverage (igvtools)
1000
100
10
1
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
11000
12000
13000
14000
15000
16000

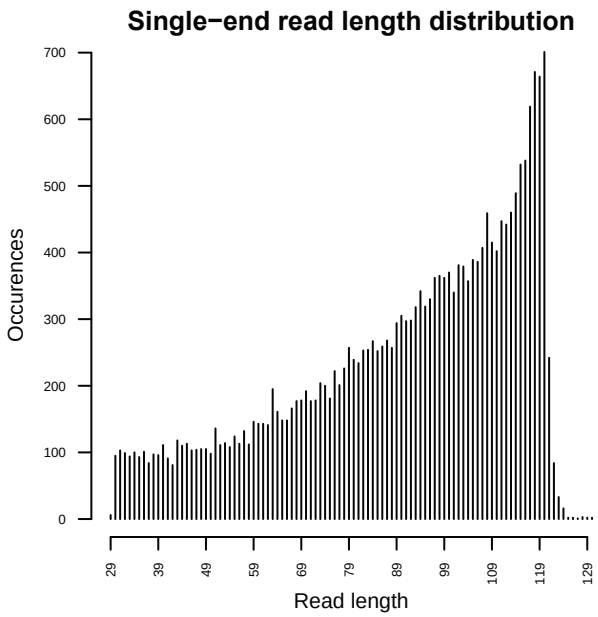

USR2 coverage (igvtools)
1000
100
10
1
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
11000
12000
13000
14000
15000
16000

(A)

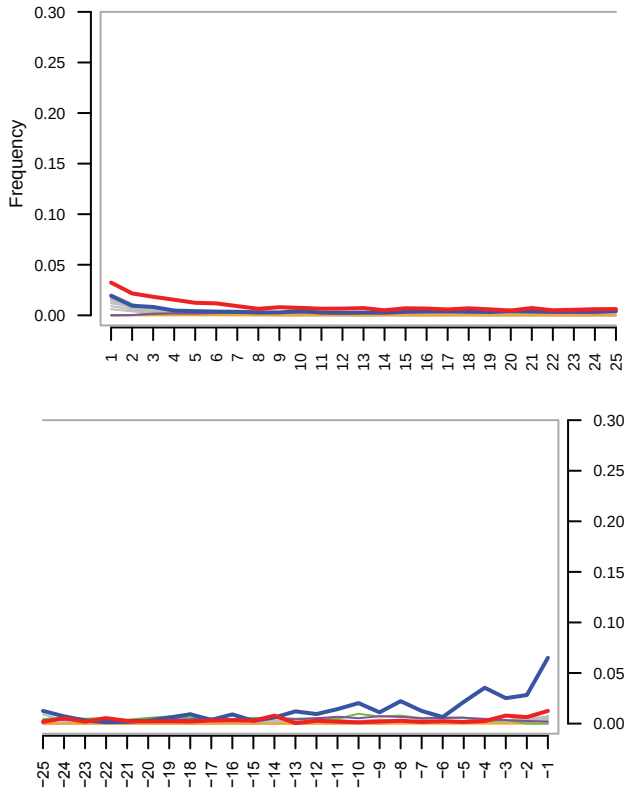
USR1

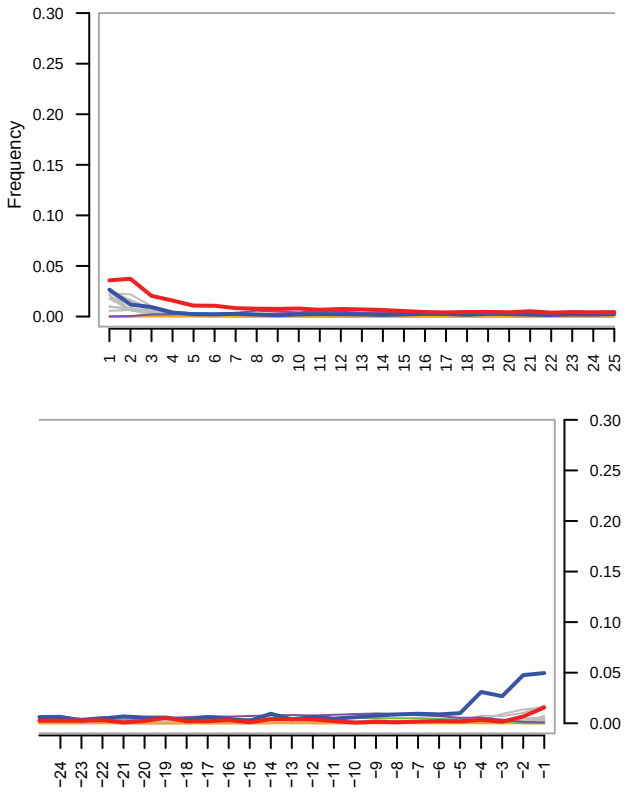

Single−end read length distribution
Occurences
Read length

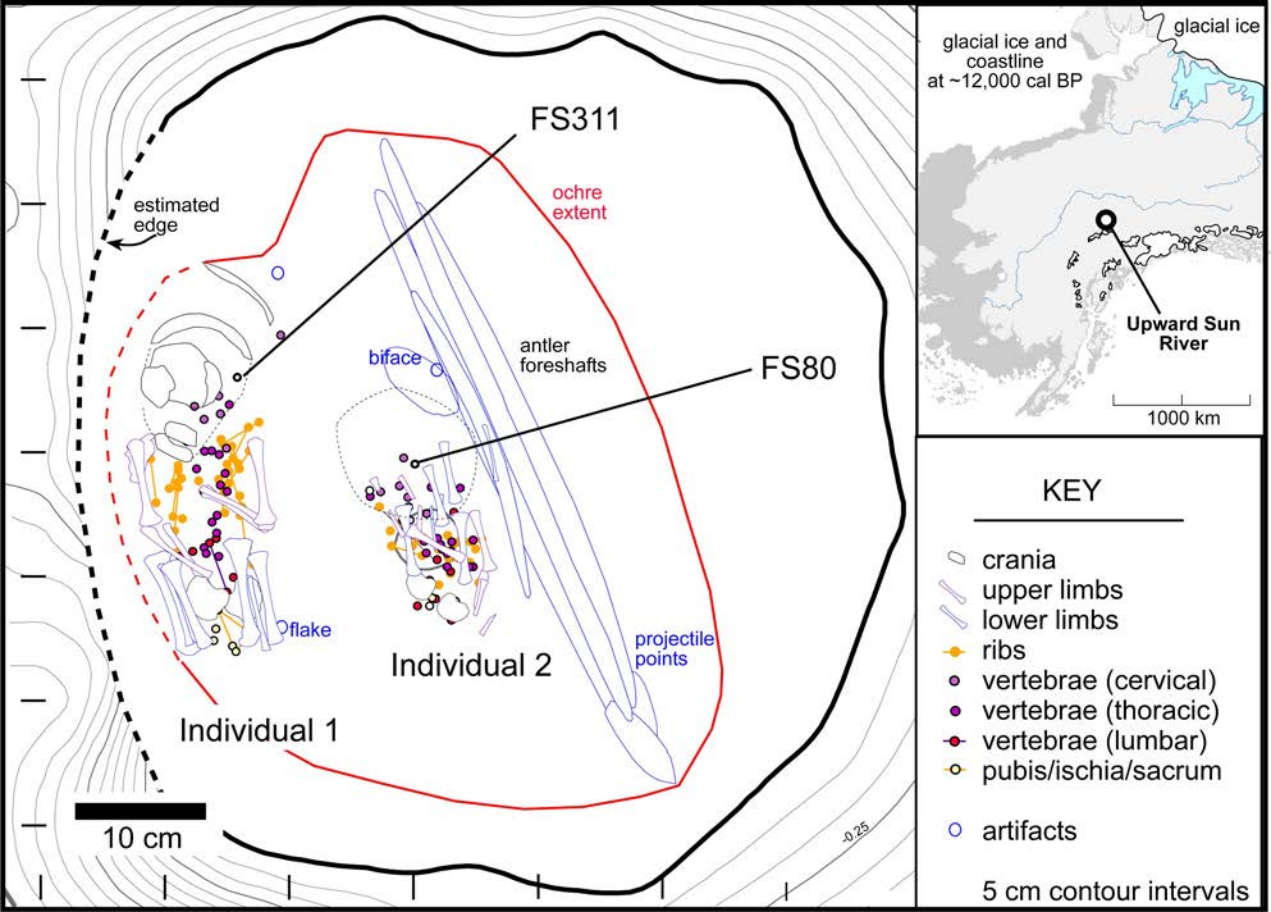
USR2


Single−end read length distribution
Occurences
Read length

(B)


Frequency


Frequency


FS311
ochre extent
estimated edge
biface
antler foreshafts
FS80
flake
Individual 2
projectile points
Individual 1
10 cm
-0.25
glacial ice
glacial ice and coastline at ~12,000 cal BP
Upward Sun River
1000 km
KEY
crania
upper limbs
lower limbs
ribs
vertebrae (cervical)
vertebrae (thoracic)
vertebrae (lumbar)
pubis/ischia/sacrum
artifacts
5 cm contour intervals

Supporting Table S1 - Library and Torrent Suite Sequencing Metrics

Torrent Suite Workflow	USR1	USR2	Extraction Blank	Library Blank
1° Amplified library molecules/ul	4.34E+08	2.04E+09	1.21E+08	1.02E+06
Post-Enriched library molecules/ul	3.74E+06	4.83E+06	3.47E+04	2.42E+04
Amplified post-enriched library molecules/ul	5.89E+08	5.15E+08	6.31E+06	1.36E+06
Ion P1 Final ISPs (--min-read-length 30)	58,711,675	55,834,961		
Ion P1 Total Bases (--min-read-length 30)	7.2 G	6.9 G		
Read Count With Barcode	47,616,189	50,545,121		
Mapped Reads (3' soft clipped) to rCRS	18,777,320	11,986,192		
Mapped Reads post-FilterDuplicates	189,790	326,865		
Mapped Reads MAPQ ≥ 30	20,044	32,979		
% relative to Final ISPs	0.03%	0.06%		

Supporting Table S2 - Alternative Bioinformatics Pipeline Metrics

Alternative Workflow	**USR1**	**USR2**
Ion P1 Final ISPs (--min-read-length 30)	55,460,151	54,469,446
Ion P1 Total Bases (--min-read-length 30)	6.5 G	6.6 G
post-cutadapt Quality, Barcode, and Adapter Filtering	40,320,121	45,333,921
Mapped Reads to rCRS	13,206,569	7,563,075
Mapped Reads of 30-120bp, MAPQ ≥ 70	5,895,852	3,668,765
Mapped Reads post-Picard MarkDuplicates	21,140	22,951
% relative to Final ISPs	0.04%	0.04%

Supporting Table S3 - Customized oligonucleotides used in this study

Name	Sequence	Purification	Barcode
Ion_A_bar1 [Φ]	C*C*A*T*CTCATCCCTGCGTGTCTCCGACTCAGTGCC*G*G*G*C	HPLC	TGCCGGGC
Ion_A_bar1comp [Φ]	G*C*C*C*GGCACTGAGTCGGAGACACGCAGGGATGAGAT*G*G*T*T	HPLC	
Ion_A_bar3 [Φ]	C*C*A*T*CTCATCCCTGCGTGTCTCCGACTCAGTGTC*C*A*C*T	HPLC	TGTCCACT
Ion_A_bar3comp [Φ]	A*G*T*G*GACACTGAGTCGGAGACACGCAGGGATGAGAT*G*G*T*T	HPLC	
Ion_P1	C*C*A*C*TACGCCTCCGCTTTCCTCTCTATGGGCAGTCGG*T*G*A*T	HPLC	
Ion_P1comp	A*T*C*A*CCGACTGCCCATAGAGAGGAAAGCGGAGGCGTAGT*G*G*T*T	HPLC	
Ion_Aamp	CCATCTCATCCCTGCGTGTC	Standard Desalting	
Ion_P1amp	CCACTACGCCTCCGCTTTCCTCTCTATG	Standard Desalting	
A1_Block*	GCCCGGCACTGAGTCGGAGACACGCAGGGATGAGATGGTT/3SpC3/	IDT xGen Blocking; HPLC	TGCCGGGC
A3_Block*	AGTGGACACTGAGTCGGAGACACGCAGGGATGAGATGGTT/3SpC3/	IDT xGen Blocking; HPLC	TGTCCACT
P1_Block*	ATCACCGACTGCCCATAGAGAGGAAAGCGGAGGCGTAGTGGTT/3SpC3/	IDT xGen Blocking; HPLC	

[Φ]Adapters were incorrectly designed without a 'GAT' barcode adapter sequence following the custom barcode

*Blocking oligos were incorrectly synthesized with unnecessary TT overhangs

Haplogroup	Original ID or origin	GenBank ID	Haplogroup	Original ID or origin	GenBank ID
B2a	Pi_26_27	AF347001	C1d	Wa_SPACH	AF347012
B2d	Na1B	AY195749	C1d	Wa_RML	AF347013
B4b1a3	TubalarB1	AY519494	C1b	Canary	AF382009
B2c1	B2-1-01	DQ282434	C1a	Japanese	AP008311
B2	B2-1-02	DQ282435	C1b1	Na4C	AY195759
B2c	B2-1-03	DQ282436	C1a	Ulchi	AY519496
B2c1	B2-1-04	DQ282437	C1b	Hispanic	DQ282449
B2c1	B2-1-05	DQ282438	C1b	Hispanic	DQ282453
B2c1	B2-1-06	DQ282439	C1b	Hispanic	DQ282458
B2c1	B2-1-07	DQ282440	C1c	Hispanic	DQ282459
B2a2	B2-2-01	DQ282441	C1c	Hispanic	DQ282460
B2a1	B2-2-02	DQ282442	C1b	Hispanic	DQ282461
B2a1	B2-2-03	DQ282443	C1c2	C1-2-04	DQ282462
B2a1a	B2-2-04	DQ282444	C1c	Hispanic	DQ282463
B2a1b	B2-2-05	DQ282445	C1b	Hispanic	DQ282464
B2a1a	B2-2-06	Dq282446	C1c	Hispanic	DQ282465
B2	Tor23	EF079874	C1b	Hispanic	DQ282469
B2	Sinixt	EF648602	C1d	Hispanic	DQ282472
B2	ACHE78	EU095206	C1d	Hispanic	DQ282473
B2	GAV123	EU095207	C1b	Hispanic	DQ282475
B2	POMO01	EU095208	C1b	Hispanic	DQ282476
B2	WAI24	EU095209	C1c	Tor24	EF079875
B2	XAV04	EU095210	C1a	Buryat324	EF153779
B2	XAV12	EU095211	C1a	Nanaitci	EU007858
B2	Quechua	EU095212	C1	WAI16	EU095222
B2	Quechua	EU095213	C1	ZOR19	EU095223
B2	Quechua	EU095214	C1	ZOR31	EU095224
B2	GRC169	EU095215	C1	Quechua	EU095225
B2	KBK23	EU095216	C1	Quechua	EU095226
B2	KBK39	EU095217	C1	ARL58	EU095227
B2	KKT01	EU095218	C1	PTJ68	EU095228
B2	KRC33	EU095219	C1	Y591	EU095229
B2	KTN209	EU095220	C1	Y650	EU095230
B2	Y637	EU095221	C1	Y669	EU095231
B2b	Cayapa602	EU095532	C1c	Arsario5	EU095527
B2	Coreguaje1-30	EU095535	C1d	Coreguaje1-54	EU095537
B2d	Ngoebe14	EU095546	C1c	Kogui12	EU095544
B2	Waunana2-8	EU095548	C1b	Wayuu4	EU095549
B2	Wayuu17	EU095551	C1c		EU327891
B2a1	B2#11	EU431083	C1c		EU327973
B2	IA_G1	EU431084	C1c	IA_A3	EU431086
B2	MA003	HQ012134	C1c	IA_A7	EU431087
B2	MA006	HQ012135	C1b	Pima	EU597557
B2	MA013	HQ012136	C1c		EU617323
B2	MA015	HQ012137	C1d	S-644177	HM107307
B2	MA016	HQ012138	C1d	S-683498	HM107308
B2	MA022	HQ012140	C1d	SA51	HM107310
B2	MA027	HQ012141	C1d2a	Mst42	HM107311
B2	MA029	HQ012142	C1d2a	Mst61	HM107313

B2	MA032	HQ012143	C1d2a	Mst50	HM107314
B2	MA050	HQ012144	C1d2	Mst64	HM107315
B2	MA051	HQ012145	C1d	ABS174	HM107316
B2	MA055	HQ012146	C1d	S-635878	HM107317
B2	MA064	HQ012147	C1d1a1	S-632547	HM107318
B2	MA071	HQ012148	C1d1a1	S-914766	HM107319
B2	MA074	HQ012149	C1d1a1	S-677163	HM107320
B2	MA082	HQ012150	C1d1a1	S-678282	HM107321
B2	MA085	HQ012151	C1d1a1	S-631499	HM107322
B2	MA087	HQ012152	C1d1b	SA27	HM107323
B2	MA088	HQ012153	C1d1b	D123	HM107324
B2	MA089	HQ012154	C1d1b	SA40	HM107325
B2	MA092	HQ012155	C1d1b1	D122	HM107326
B2	MA094	HQ012156	C1d1b1	ABS228	HM107327
B2	MA101	HQ012157	C1d1b1	ABS228	HM107328
B2	MA104	HQ012158	C1d1b1	ABS299	HM107329
B2	MA105	HQ012159	C1d1b1	ACO388	HM107330
B2	MA109	HQ012160	C1d1b1	S-629812	HM107331
B2	MA122	HQ012161	C1d1b1	ABS284	HM107332
B2	MA123	HQ012162	C1d1b1	SA53	HM107333
B2	MA126	HQ012163	C1d1c	S-934519	HM107334
B2	MA127	HQ012164	C1d1c1	S-658745	HM107335
B2	MA128	HQ012165	C1d1c1	S-635687	HM107336
B2	MA133	HQ012166	C1d1c1	S-643125	HM107337
B2	MA135	HQ012167	C1d1	S-689881	HM107338
B2	MA139	HQ012168	C1d1	S-923879	HM107339
B2	MA141	HQ012169	C1d1	ABS229	HM107340
B2	MA142	HQ012170	C1d1	S-987167	HM107341
B2	MA144	HQ012171	C1d1	S-919781	HM107342
B2	MA153	HQ012172	C1d1	S-935165	HM107343
B2	MA158	HQ012173	C1d1	Mst43	HM107345
B2	MA160	HQ012174	C1d1d	ABS326	HM107346
B2	MA162	HQ012175	C1d1d	S-681747	HM107347
B2	MA169	HQ012176	C1d1d	S-681241	HM107348
B2	MA170	HQ012177	C1d1	S-649798	HM107349
B2	MA175	HQ012178	C1d1e	S-686135	HM107350
B2	MA179	HQ012179	C1d1e	ARN112	HM107351
B2	MA186	HQ012180	C1d1	S-919735	HM107352
B2	MA187	HQ012181	C1d1	S-984386	HM107353
B2	MA191	HQ012182	C1d1	S-657487	HM107354
B2	MA195	HQ012183	C1d1	ABS155	HM107355
B2	MA208	HQ012184	C1d1	S-681199	HM107356
B2	MA213	HQ012185	C1d1	S-932862	HM107357
B2a1a	B2#05	JQ702668	C1d1	SA11	HM107358
B2a1	B2#18	JQ702855	C1d1	S-938246	HM107359
B2a2	B2#23	JQ703828	C1d1	S-939485	HM107360
B2a1a	B2#09	JQ703834	C1d1	ACO394	HM107361
B2a4	B2#30	JQ703852	C1d1	S-686784	HM107362
B2a1b	B2#15	JQ705598	C1d1	S-686788	HM107363
B2i2a	ARN083	JX413011	C1d1	S-635448	HM107364

B2i2a	Mco32	JX413012	C1d1	S-996796	HM107365
B2i2a	Mco32	JX413013	C1d1	S-915151	HM107366
B2i2a	D04	JX413014	C1d1	S-915473	HM107367
B2i2a	CA007	JX413015	C1f		HM804483
B2i2a	CA012	JX413016	C1b	MA004	HQ012186
B2i2a	XL060	JX413017	C1b	MA011	HQ012187
B2i2a	686289	JX413018	C1b	MA018	HQ012188
B2i2a	ARN086	JX413019	C1b	MA024	HQ012189
B2i2a	686571	JX413020	C1b	MA030	HQ012190
B2i2a	XL119	JX413021	C1b	MA031	HQ012191
B2i2a	933177	JX413022	C1b	MA034	HQ012192
B2i2a	XL144	JX413023	C1b	MA036	HQ012193
B2i2b	XL058	JX413024	C1b	MA037	HQ012194
B2i2b	CA028	JX413026	C1b	MA040	HQ012195
B2i2b	XL050	JX413027	C1b	MA041	HQ012196
B2i2b	686246	JX413028	C1b	MA044	HQ012197
B2i2b	XL156	JX413029	C1b	MA049	HQ012198
B2i2b	CA046	JX413030	C1b	MA052	HQ012199
B2i2b	ARN109	JX413031	C1b	MA056	HQ012200
B2i2b	XL061	JX413032	C1b	MA059	HQ012202
B2i2b	XL012	JX413033	C1b	MA066	HQ012203
B2i2b	H05	JX413034	C1b	MA068	HQ012204
B2i2	MSP	JX413035	C1b	MA076	HQ012205
B2a	B2#01	KC711021	C1b	MA078	HQ012206
B2a	B2#02	KC711022	C1b	MA090	HQ012207
B2a1a	B2#06	KC711023	C1b	MA110	HQ012208
B2a1a	B2#08	KC711024	C1b	MA131	HQ012209
B2a1b	B2#13	KC711025	C1b	MA184	HQ012210
B2a1	B2#16	KC711026	C1b	MA185	HQ012211
B2a1	B2#17	KC711027	C1b	MA188	HQ012212
B2a3	B2#26	KC711031	C1b	MA192	HQ012213
B2a4a	B2#31	KC711033	C1b	MA194	HQ012214
B2a4a1	B2#32	KC711034	C1b	MA203	HQ012215
B2a4a1	B2#33	KC711035	C1b	MA205	HQ012216
B2a4a1	B2#34	KC711036	C1b	MA215	HQ012217
B2a5	B2#36	KC711037	C1c	MA008	HQ012219
B2a5	B2#37	KC711038	C1c	MA026	HQ012220
B2a5	B2#38	KC711039	C1c	MA073	HQ012221
B4d1'2'3	Buryat(B3)	AY519484	C1c	MA112	HQ012223
B4b1a3a1	Alt_124	JN857011	C1c	MA113	HQ012224
B4b1a3a1	Alt_175	JN857014	C1c	MA121	HQ012225
B4b1a3a	Bur_417	JN857022	C1c	MA174	HQ012227
B4b1a3a1	Khm_62	JN857045	C1c	MA177	HQ012228
B4b1a3a1	Khm_44	JN857043	C1c	MA183	HQ012229
B4b1a3a1	Sh_15'	JN857059	C1c	MA189	HQ012230
B4b1a3a1	CT_55	JN857037	C1c	MA198	HQ012231
B2f	B4b1	EU334872	C1c	MA204	HQ012232

High % B2 Published Populations

Fig Reference	Site or Tribe	Date (YBP)	N (≥ 20 Required)	%A	%B	%C	%D	%X	Reference
1	Yakama	Contemp	42	4.8	66.7	7.1	16.7	4.8	Eshleman JA, Malhi RS, Johnson, JR, Kaestle FA, Lorenz J, Smith, DG. Mitochondrial DNA
2	Wishram	Contemp	33	21.2	51.5	0	27.3	0	Eshleman JA, Malhi RS, Johnson, JR, Kaestle FA, Lorenz J, Smith, DG. Mitochondrial DNA
3	N. Paiute/Shoshoni	Contemp	94	0	42.6	9.6	47.9	0	Malhi RS, Mortensen HM, Eshleman JA, Kemp BM, Lorenz JG, Kaestle FA, Johnson JR, Gor
4	Washo	Contemp	28	0	53.6	35.7	10.7	0	Lorenz JG, Smith DG. Distribution of four founding mtDNA haplogroups among Native North
5	Fremont	500-1500	30	0	80	13.3	6.7	0	Carlyle SW, Parr RL, Hayes MG, O'Rourke DH. The context of maternal lineages in the Great
6	Tommy Site	850-1150	36	2.8	69.4	13.9	13.9	0	Snow MH, Durand KR, Smith DG. Ancestral Puebloan mtDNA in context of the greater south
7	Anasazi	1010-2010	38	10.7	71.4	17.9	0	0	Carlyle, Shawn W. *Geographical and Temporal Lineage Stability Among the Anasazi* . Diss.
8	Navajo	Contemp	64	51.6	40.6	4.7	0	3.1	LeBlanc SA, Kreisman LSC, Kemp BM, Smiley FE, Carlyle SW, Dhody AN, Benjamin T. Qu
9	Jemez	Contemp	107	0	86	3.7	0	10	Snow MH, Durand KR, Smith DG. Ancestral Puebloan mtDNA in context of the greater south
10	Hualapai	Contemp	76	1.3	50	48.7	0	0	LeBlanc SA, Kreisman LSC, Kemp BM, Smiley FE, Carlyle SW, Dhody AN, Benjamin T. Qu
11	Pai Yuman	Contemp	27	7.9	66.7	25.9	0	0	Malhi RS, Mortensen HM, Eshleman JA, Kemp BM, Lorenz JG, Kaestle FA, Johnson JR, Gor
12	Zuni	Contemp	70	17.1	74.3	8.6	0	0	Snow MH, Durand KR, Smith DG. Ancestral Puebloan mtDNA in context of the greater south
13	River Yuman	Contemp	22	0	63.6	36.4	0	0	Malhi RS, Mortensen HM, Eshleman JA, Kemp BM, Lorenz JG, Kaestle FA, Johnson JR, Gor
14	Delta Yuman	Contemp	23	0	56.5	43.5	0	0	Malhi RS, Mortensen HM, Eshleman JA, Kemp BM, Lorenz JG, Kaestle FA, Johnson JR, Gor
15	Tohono O'odham (Papago)	Contemp	42	7.1	57.1	35.7	0	0	Kemp BM, Gonzalez-Oliver A, Malhi RS, Monroe C, Schroeder KB, McDonough J, Rhett G,
16	Akimal O'odham (Pima)	Contemp	146	4.8	46.6	48	0.7	0	LeBlanc SA, Kreisman LSC, Kemp BM, Smiley FE, Carlyle SW, Dhody AN, Benjamin T. Qu
17	Quechan/Cocopa	Contemp	22	0	68.2	31.8	0	0	Lorenz JG, Smith DG. Distribution of four founding mtDNA haplogroups among Native North
18	Nahua-Atocpan	Contemp	50	38	40	18	4	0	Sandoval K, Buentello-Malo L, Peñaloza-Espinosa R, Avelino H, Salas A, Calafell F, Comas I
19	Embera	Contemp	21	9.5	52.4	28.6	9.5	0	Usme-Romero S, Alonso M, Hernandez-Cuervo H, Yunis EJ, Yunis JJ. Genetic differences be
20	Puinave	Contemp	61	8.2	50.8	32.8	6.6	1.6	Usme-Romero S, Alonso M, Hernandez-Cuervo H, Yunis EJ, Yunis JJ. Genetic differences be
21	Curriperco	Contemp	22	4.5	40.9	36.4	14.7	4.5	Usme-Romero S, Alonso M, Hernandez-Cuervo H, Yunis EJ, Yunis JJ. Genetic differences be
22	Ingano	Contemp	48	39.6	35.4	22.9	2.1	0	Usme-Romero S, Alonso M, Hernandez-Cuervo H, Yunis EJ, Yunis JJ. Genetic differences be
23	Yungay	Contemp	36	3	47	36	14	0	Lewis CM, Lizárraga B, Tito RY, López PW, Iannacone GC, Medina A, Martínez R, Polo SI,
24	San Martin	Contemp	21	10	57	5	29	0	Fuseilli S, Tarazona-Santos E, Dupanloup I, Soto A, Luiselli D, Pettener D. Mitochondrial DN
25	Peruvian Highlanders	550-450	35	8.6	65.7	22.9	2.9	0	Shinoda K, Adachi N, Guillen S, Shimada I. Mitochondrial DNA analysis of ancient Peruvian
26	Yacotogia	1187	25	4	56	40	0	0	Fehren-Schmitz L, Reindel M, Cagigao ET, Hummel S, Herrmann B. Pre-Columbian populatic
27	Ancash	Contemp	33	9	52	18	21	0	Lewis CM, Tito RY, Lizárraga B, Stone AC. Land, language, and loci: mtDNA in Native Ame
28	Arequpa	Contemp	22	9	68	14	9	0	Fuseilli S, Tarazona-Santos E, Dupanloup I, Soto A, Luiselli D, Pettener D. Mitochondrial DN
29	Chimane	Contemp	40	40	55	5	0	0	Bert F, Corella A, Gené M, Pérez-Peréz A, Turbón D. Major mitochondrial DNA haplotype he
30	Puno (Quecha)	Contemp	30	7	60	23	10	0	Lewis CM, Lizárraga B, Tito RY, López PW, Iannacone GC, Medina A, Martínez R, Polo SI,
31	Quechua 2	Contemp	32	15.6	75	9.4	0	0	Bert F, Corella A, Gené M, Pérez-Peréz A, Turbón D. Major mitochondrial DNA haplotype he
32	Aymara 2	Contemp	33	0	93.9	3	3	0	Bert F, Corella A, Gené M, Pérez-Peréz A, Turbón D. Major mitochondrial DNA haplotype he
33	Trinitario	Contemp	33	15.2	42.4	39.4	3	0	Bert F, Corella A, Gené M, Pérez-Peréz A, Turbón D. Major mitochondrial DNA haplotype he
34	Quebrada de Humahuaca	Contemp	46	10.9	67.4	17.4	4.3	0	Dipierri JE, Alfaro E, Martinez-Marignac VL, Bailliet G, Bravi CM, Cejas S, Bianchi NO. Pat
35	Atacamenos	Contemp	113	13.3	71.7	9.7	5.3	0	Bailliet et al. 1994, Merriwether et al. 1995
36	Chorote	Contemp	20	15	40	30	15	0	Goicoechea AS, Carnese FR, Dejean CB, Avena SA, Weimer TA, Franco MH, Callegari-Jacqu
37	Gram Chaco	Contemp	202	16	42	13	28	0	Cabana GS, Merriwether DA, Hunley K, Demarchi DA. 2006. Is the genetic structure of Gran

High % C1 Published Populations

Fig Reference	Site or Tribe	Date (YBP)	N (≥ 20 Required)	%A	%B	%C	%D	%X	Reference
38	Norris Farms	700	108	31.5	12	42.6	8.3	5.6	Stone AC and Stoneking M. mtDNA analysis of a prehistoric Oneota population: Implications
39	Cecil	3600-2860	23	0	6.3	56.3	37.5	0	Eshleman JA. *Mitochondrial DNA and prehistoric population movements in Western North A*
40	Cook	2000	23	4.3	8.7	43.5	43.5	0	Eshleman JA. *Mitochondrial DNA and prehistoric population movements in Western North A*
41	Hualapai	Contemp	76	1.3	50	48.7	0	0	LeBlanc SA, Kreisman LSC, Kemp BM, Smiley FE, Carlyle SW, Dhody AN, Benjamin T. Qu
42	Delta Yuman	Contemp	23	0	56.5	43.5	0	0	Malhi RS, Mortensen HM, Eshleman JA, Kemp BM, Lorenz JG, Kaestle FA, Johnson JR, Gor
43	Akimal O'odham (Pima)	Contemp	146	4.8	46.6	48	0.7	0	LeBlanc SA, Kreisman LSC, Kemp BM, Smiley FE, Carlyle SW, Dhody AN, Benjamin T. Qu
44	Pima	contemp	97	11.3	3.1	84.5	1	0	Sandoval K, Buentello-Malo L, Peñaloza-Espinosa R, Avelino H, Salas A, Calafell F, Comas I
45	La Caleta (Tainos)	1330-320	24	0	0	75	25	0	Lalueza-Fox C, Calderon FL, Calafell F, Morera B, Bertranpetit J. MtDNA from extinct Taino
46	Arawaken	Contemp	29	28	28	44	0	0	Melton PE, Briceno I, Gomez A, Devor EJ, Bernal JE, Crawford MH. Biological relationship b
47	Guambiano	Contemp	24	4.2	12.5	66.6	16.7	0	Keyeux G, Rodas C, Gelvez N, Carter D. Possible migration routes into South America deduce
48	Desano	Contemp	20	15	15	45	25	0	Usme-Romero S, Alonso M, Hernandez-Cuervo H, Yunis EJ, Yunis JJ. Genetic differences be
49	Movima	contemp	22	9.1	9.1	63.6	18.2	0	Bert F, Corella A, Gené M, Pérez-Peréz A, Turbón D. Major mitochondrial DNA haplotype he
50	Ignaciano	Contemp	21	19	38.1	42.9	0	0	Bert F, Corella A, Gené M, Pérez-Peréz A, Turbón D. Major mitochondrial DNA haplotype he