

ARCTIC PREHISTORY THROUGH ANCIENT DNA

by

Justin C Tackney

A dissertation submitted to the faculty of  
The University of Utah  
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Anthropology

The University of Utah

December 2016

Copyright © Justin C Tackney 2016

All Rights Reserved



## ABSTRACT

Genetic information from human remains obtained from archaeological excavations can reveal the history of modern peoples, rediscover the signature of prehistoric populations, and track human variation across time and space. The events that led to the colonization of the Americas can be explored by sequencing geographically and temporally appropriate samples. Two distinct phases of Native American migrations are investigated here – the initial Pleistocene movement of people out of Beringia and into North America, and the much later Holocene migrations eastward across the North American Arctic. In the first chapter, whole mitochondrial genomes were sequenced from two contemporaneous human burials at a rare Pleistocene residential site at Upward Sun River in central Alaska, dating to ~11,500 cal B.P. One individual carried mitochondrial lineage C1b, whereas the other carried a root B2 haplotype. Today C1b and B2 are absent in modern populations of northern North America. These results validate the previously hypothesized expectation of higher levels of genetic diversity in the earlier, and now lost, Beringian gene pool. The second and third chapters concern the origins and migrations of the Neo-Eskimo Thule, the hypothesized ancestors of modern day Inupiat/Inuit. A population-scale sequencing project was initiated from a cemetery at Nuvuk, a long-term Thule village at Pt. Barrow, AK. The cemetery represents the oldest and largest number of Thule remains ever found in North America. The hypervariable segment I of the mitochondrial genome was amplified and sequenced in 44 individuals,

and Arctic haplogroups A2a, A2b, and D4b1a2a1a were identified. The haplogroup frequencies at Nuvuk were compared to populations across Siberia and North America, and were most similar to the modern Inuit communities of Canada and Greenland. This supports the ancient North Slope as the origin point for the Thule migrations. To further investigate phylogenetic relationships, additional human remains were sampled from Iqliqtsiugvigruaq, a 19<sup>th</sup> century interior Alaskan site located inside the boundary of Kobuk Valley National Park. Whole mitochondrial genomes were sequenced from three Nuvuk burials and three remains from Iqliqtsiugvigruaq, and these allowed a refinement of the haplogroup A2b and D4b1a2a1a phylogenies.

## CONTENTS

ABSTRACT .....	iii
LIST OF TABLES .....	vii
LIST OF FIGURES .....	viii
Chapters	
1. TWO CONTEMPORANEOUS MITOGENOMES FROM TERMINAL PLEISTOCENE BURIALS IN EASTERN BERINGIA.....	1
Abstract.....	2
Results.....	3
Discussion.....	6
Conclusion .....	6
Materials and Methods.....	7
Acknowledgements.....	7
References.....	7
Supporting Information.....	8
2. MOLECULAR ANALYSIS OF AN ANCIENT THULE POPULATION AT NUVUK, POINT BARROW, ALASKA.....	18
Introduction.....	19
Materials and Methods.....	28
Results.....	35
Discussion.....	42
Acknowledgements.....	51
Literature Cited .....	52
3. PHYLOGENETIC RECONSTRUCTION OF BERINGIAN HAPLOGROUPS A2B AND D4B1A2A1A USING DATED NEO-ESKIMO THULE REMAINS .....	70
Introduction.....	71
Materials and Methods.....	75
Results.....	82
Discussion.....	88

Acknowledgements.....	94
Literature Cited.....	95

## LIST OF TABLES

1S.1	Library and Torrent Suite sequencing metrics.....	17
1S.2	Alternative bioinformatics pipeline metrics.....	17
1S.3	Customized oligonucleotides used in this study .....	17
2.1	Summary statistics of Nuvuk aDNA samples.....	58
2.2	Nuvuk sample HVR-I polymorphisms and dates .....	59
2.3	Population sharing of Nuvuk haplotypes.....	60
2S.1	Oligonucleotides .....	61
2S.2	Minimum Beringian HVR-I motifs for population selection .....	62
2S.3	Beringian haplogroup frequencies across populations .....	63
2S.4	Excluded population frequency data.....	64
2S.5	Chronological appearance of Beringian haplogroups.....	65
3.1	Oligonucleotides used in this study .....	99
3.2	NGS library workflow .....	100
3.3	Sequencing and bioinformatics pipeline protocol .....	101
3.4	Sequencing and bioinformatics pipeline metrics .....	102

## LIST OF FIGURES

1.1	Geographic map of reported Native American populations with >40% C1 or B2 haplogroup frequencies, as well as locations of archaeological sites discussed.....	3
1.2	MP hand-curated phylogenetic trees of (A) C1b and (B) B4 mtDNA haplotypes ..	4
1.3	ML phylogenies of (A) haplogroup C1 and (B) haplogroup B2 .....	5
1S.1	Sequence coverage after the Torrent Suite pipeline across the mitochondrial genome for USR1 and USR2 on a 1-base sliding window .....	12
1S.2	(A) Read-length histograms. (B) Position-specific substitutions from the 5' end of reads post-Torrent Suite pipeline .....	13
1S.3	Sequence coverage after the alternative pipeline across the mitochondrial genome for USR1 and USR2 on a 1-base sliding window.....	14
1S.4	(A) Read-length histograms. (B) Position-specific substitutions from the 5' end of reads (Upper) and 3' end of read (Lower) following the alternative bioinformatics pipeline.....	15
1S.5	Plan view of burial and locations of aDNA samples, FS311 (Individual 1) and FS80 (Individual 2).....	16
1S.6	aDNA samples, FS311 (Individual 1, Left) and FS80 (Individual 2, Right).....	16
2.1	Beringian haplogroups at Nuvuk across populations.....	67
2.2	Principle component analysis of the Beringian haplogroup population frequencies .....	68
2.3	aDNA Beringian haplogroup appearance by date and location.....	69
3.1	Sample library mitochondrial sequence coverage on a 1-base sliding window ..	103
3.2	Maximum Parsimony phylogenetic tree and BEAST coalescence dates of the D4b1a2a mtDNA haplogroup .....	104

3.3	Maximum Parsimony phylogenetic tree and BEAST coalescence dates of the A2b mtDNA haplogroup .....	105
3.4	Damage patterns.....	106

## CHAPTER 1

### TWO CONTEMPORANEOUS MITOGENOMES FROM TERMINAL PLEISTOCENE BURIALS IN EASTERN BERINGIA

Proc. Natl. Acad. Sci, USA. 2015. 112(45):13833–13838. Two contemporaneous mitogenomes from terminal Pleistocene burials in eastern Beringia. Justin C. Tackney, Ben A. Potter, Jennifer Raff, Michael Powers, W. Scott Watkins, Derek Warner, Joshua D. Reuther, Joel D. Irish, and Dennis H. O'Rourke. Copyright and permission retained by the authors.



## Two contemporaneous mitogenomes from terminal Pleistocene burials in eastern Beringia

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

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

Edited by Richard G. Klein, Stanford University, Stanford, CA, and approved September 18, 2015 (received for review June 17, 2015)

**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 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, whereas 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.**

Pleistocene burials | ancient mitochondrial DNA | paleogenomics | peopling | Americas

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 (<14–16 kya), rapid colonization event following the Last Glacial Maximum. 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 before 13,250 cal B.P., 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 8,000 y (15).

Ancient DNA (aDNA) 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 although 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 B.P.) individuals have yielded mitochondrial genetic (mtDNA) data, although we highlight the seven sites with ancient human remains dating to >8,000-y-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 (the last four are reviewed in ref. 20) (Fig. 1).

In 2011 Potter et al. (21) reported on the discovery of a cremated 3-y-old child from a residential feature at Upward Sun River (USR) in eastern Beringia dating to 11,500 cal B.P. 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 and 11,270 cal B.P. (23). Based on dental and osteological aging methods, USR1 represents a late preterm fetus, and USR2 likely died within the first 6 wk 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

### Significance

**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 B.P.) that postdates 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.**

Author contributions: J.C.T., B.A.P., and D.H.O. designed research; J.C.T., B.A.P., J.R., M.P., J.D.R., and J.D.I. performed research; W.S.W. and D.W. contributed new reagents/analytic tools; J.C.T., W.S.W., and D.H.O. analyzed data; J.C.T., B.A.P., J.R., W.S.W., and D.H.O. wrote the paper.

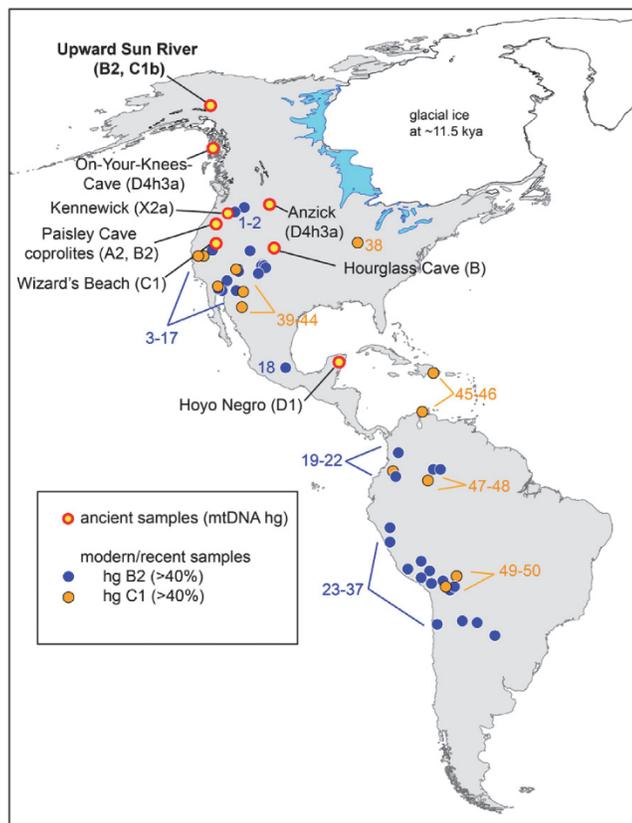
The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. KT891989 and KT891990).

<sup>1</sup>To whom correspondence should be addressed. Email: justin.tackney@anthro.utah.edu.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1511903112/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1511903112/-DCSupplemental).



**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 y B.P. with successfully genotyped human mitochondrial DNA lineages, are listed on the map (with reported haplotypes). Reported populations of  $\geq 20$  individuals with  $\geq 40\%$  C1 (yellow) or B2 (blue) are shown. Populations and frequencies specific to this figure (referenced by numbers 1–50) are available in the *SI Materials and Methods*.

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 PI chips with an Ion Proton System (Life Technologies). This is one of the first examples of the Ion Torrent technology applied to aDNA.

## Results

From 58.7 and 55.8 million sequencing reads, 20,004 and 32,979 unique mtDNA reads (MAPQ  $\geq 30$ ) from USR1 and USR2, respectively, were mapped to the human mtDNA reference (Table S1). We used 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 used in the aDNA literature.

Sequencing of the enriched mtDNA from samples USR1 and USR2 resulted in 100% coverage of the mtDNA genomes with average read depths of 117 $\times$  (geometric mean of 97 $\times$ ) for USR1 and 195 $\times$  (geometric mean of 180 $\times$ ) for USR2 (Fig. S1). Mean read lengths for the two samples were 98 and 99 bp. 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 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 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 aDNA 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

(30–120 bp), and quality, and we remapped (tmap) without 3' clipping. Following this alternative pipeline, 21,140 and 22,951 mtDNA reads at MAPQ  $\geq 70$  mapped to the mtDNA genome from USR1 and USR2, respectively (Table S2). One-hundred percent of the genome was covered, at average read depth of 113 $\times$  (geometric mean of 103 $\times$ ) for USR1 and 125 $\times$  (geometric mean of 119 $\times$ ) for USR2 (Fig. S3). Nucleotide mismatches now displayed the

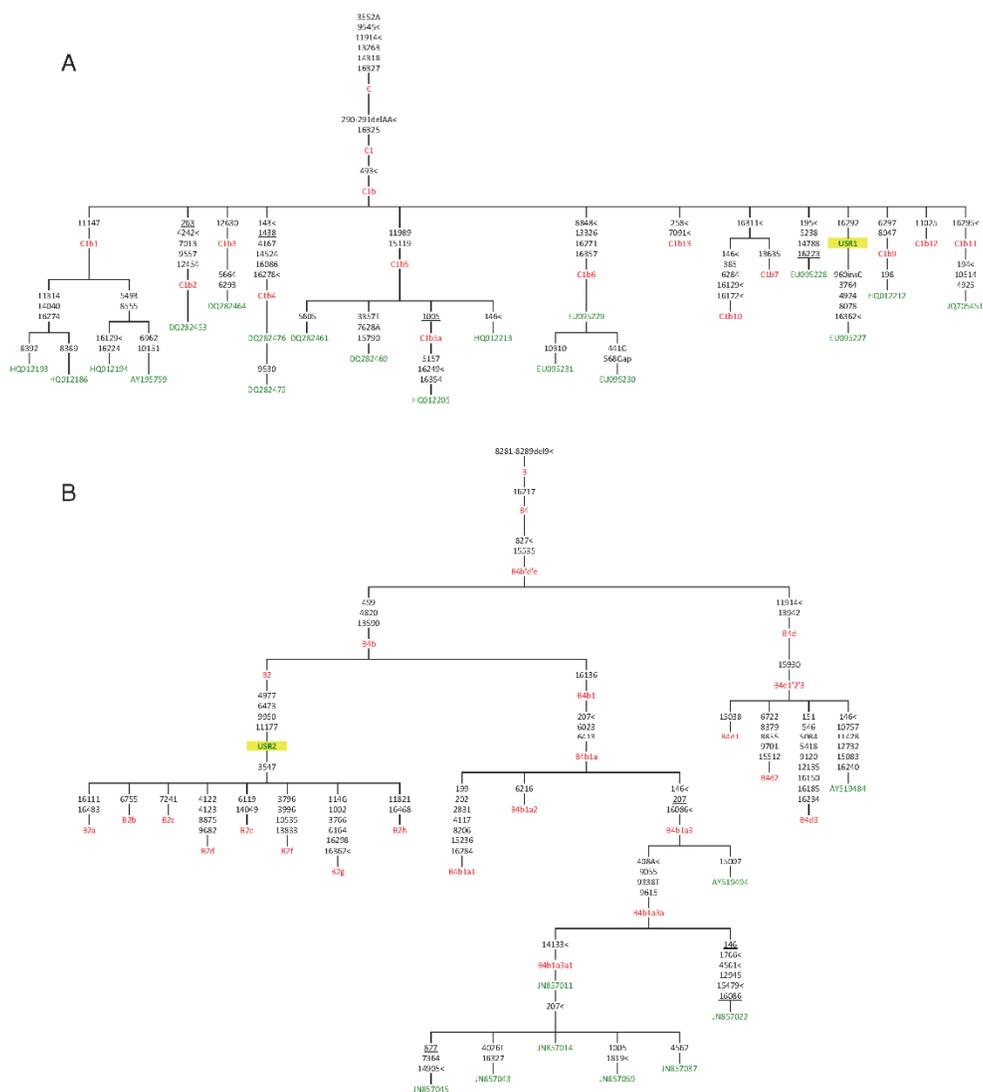


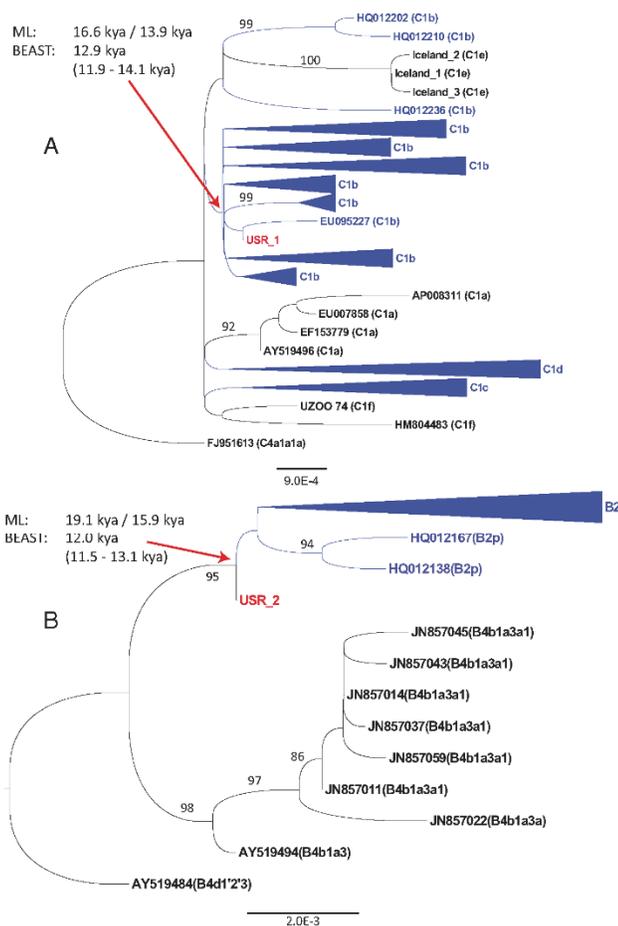
Fig. 2. MP hand-curated phylogenetic trees of (A) C1b and (B) B4 mtDNA haplotypes. Only a subset of the sequences analyzed in this study are shown, along with the placement of USR1 and USR2. Sequences used in this analysis are listed in green and node assignments are listed in red.

expected damage patterns for degraded samples, although the 5' read ends still showed some residual unexpected alternative signal (Fig. S4). Although 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, although masking expected DNA damage patterns at the ends of reads, did not bias the accurate calling of these two samples.

Maximum-likelihood (ML) 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, whereas USR2 was placed at the root of known Native American B2 diversity. Both samples exhibit branch length shortening relative to modern Native American sequences, because of their lower number of derived mutations, as expected for aDNA. The best tree by final likelihood score was compared with the results of 1,000 bootstrap runs. Nonparametric 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). Because USR1 and USR2 are 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 y ago, 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 y ago (a clearly too-recent date given the age of USR1). The Native American-specific B2 clade coalescence time was estimated at 19,100 or 15,900 y ago, respectively (27, 28). All of these dates fall within previously published estimates.



**Fig. 3.** ML 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. Nonparametric bootstrap support for branches are noted for uncollapsed clades with support  $\geq 80\%$ . The 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.

Because the ML-based estimates do not take into account the radiocarbon ages of USR1 and USR2, we used 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 y ago (11,853–14,079) [mean; 95% highest posterior density (HPD) interval], with tip dates of 11,500 y ago for USR1 and 8,300 y ago for UZOO-74 (see below). The C1b clade coalescence date is near the younger bounds of the timescales calculated in the literature, although 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 y ago (11,500–13,085), using a tip date of 11,500 y ago 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 (although 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 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 et al. (31) reported two individuals with haplogroup B2 in prehistoric (800 and 490 cal B.P.) 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: that is, at the ~8,800 cal B.P. burial at Hourglass Cave in Colorado and in three coprolites dated between 14,270 and 14,000 cal B.P. 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 and 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 y ago (32), three Native American-specific (C1b, C1c, C1d) branches previously molecularly dated to 19,000 y ago (12, 26), an Icelandic-specific C1e branch (33), and a novel C1f haplotype sequenced from an individual dated to ~8,300 cal B.P. (UZOO-74) at the Mesolithic site of Yuzhny Oleni Ostrov, North West Russia (25) (although, see Fig. 34 for a possibly related sequence, H1804483). 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: for example, the 10,400 cal B.P. individual from Wizard's Beach, NV and for further sequencing of any C1 lineages in Eurasia.

It is of interest that all five founding macrohaplogroups in Native American populations (A, B, C, D, and X) are represented in the small sample of individuals that lived more than 8,000 y ago in a geographic area stretching from subarctic Alaska to southern Mexico. Four of these macrohaplogroups are found at the three northern North American sites—Paisley Caves, Upward Sun River, and Anzick—dating to over 11,000 y ago. Mitochondrial lineage designation for the majority of the pre-8,000 y ago individuals were determined by low resolution methods of restriction fragment length polymorphism analysis and 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 postcolonization population dynamics in structuring modern genetic patterns. Cui et al. (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 postdates 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. Furthermore, 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 before 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 TVC, or reads were processed using offline tools to determine DNA damage patterns. Haplotypes of consensus mitochondrial genomes from these variants were identified by MP and phylogenetic trees of all known related sequences were created by 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 aDNA facility using established clean room protocols. Blanks were included at all steps in the process before sequencing and no laboratory personnel carry the haplotypes reported here. An expanded discussion of detailed materials and methods can be found in *SI Materials and Methods*, Figs. S1–S6, Tables S1–S3, and Datasets S1 and S2.

**ACKNOWLEDGMENTS.** We thank Lin Chen and Zhao Xu in the Life Technologies Next-Generation Sequencing Bioinformatics support team for long technical conversations; 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; Dr. Ryan Bohlander for being an excellent and patient de facto server administrator; the laboratory of Dr. Lynn B. Jorde for providing space for the preparation of the modern mtDNA bait; and the Healy Lake Tribal Council and Tanana Chiefs Conference representatives for their support. 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; and the College of Social and Behavioral Science and the Department of Anthropology at the University of Utah.

- de Acosta J (2002) *Natural and Moral History of the Indies* (Duke Univ Press, Durham, NC).
- Dixon EJ (1999) *Bones, Boats and Bison: Archaeology and the First Colonization of Western North America* (Univ of New Mexico Press, Albuquerque, NM).
- Meltzer DJ (2010) *First Peoples in a New World: Colonizing Ice Age America* (Univ of California Press, Berkeley, CA).
- Hoffecker JF, Elias SA, O'Rourke DH (2014) Anthropology. Out of Beringia? *Science* 343(6174):979–980.
- Goebel T, Waters MR, O'Rourke DH (2008) The late Pleistocene dispersal of modern humans in the Americas. *Science* 319(5869):1497–1502.
- Tamm E, et al. (2007) Beringian standstill and spread of Native American founders. *PLoS One* 2(9):e829.
- Reich D, et al. (2012) Reconstructing Native American population history. *Nature* 488(7411):370–374.
- Raghavan M, et al. (2014) The genetic prehistory of the New World Arctic. *Science* 345(6200):1258–1262.
- Raghavan M, et al. (2014) Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* 505(7481):87–91.
- Mulligan CJ, Kitchen A, Miyamoto MM (2008) Updated three-stage model for the peopling of the Americas. *PLoS One* 3(9):e3199.
- Dillehay TD, ed (1997) *The Archaeological Context and Interpretation* (Smithsonian Institution Press, Washington, DC), Vol 2.
- 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.
- 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):e6948.
- Fagundes NJ, et al. (2008) Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. *Am J Hum Genet* 82(3):583–592.
- Raghavan M, et al. (2015) POPULATION GENETICS. Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* 349(6250):aab3884.
- Pickrell J, Reich D (2014) Toward a new history and geography of human genes informed by ancient DNA. *Trends Genet* 30(9):377–389.
- Chatters JC, et al. (2014) Late Pleistocene human skeleton and mtDNA link Paleoamericans and modern Native Americans. *Science* 344(6185):750–754.
- 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.
- Rasmussen M, et al. (2015) The ancestry and affiliations of Kennewick Man. *Nature* 523(7561):455–458.
- Raff JA, Bolnick DA, Tackney J, O'Rourke DH (2011) Ancient DNA perspectives on American colonization and population history. *Am J Phys Anthropol* 146(4):503–514.
- 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.
- 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* 25:132–135.
- 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. *Proc Natl Acad Sci USA* 111(48):17060–17065.
- Marić T, Whitten M, Pääbo S (2010) Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One* 5(11):e14004.
- Der Sarkissian C, et al; Genographic Consortium (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.
- Perego UA, et al. (2010) The initial peopling of the Americas: A growing number of founding mitochondrial genomes from Beringia. *Genome Res* 20(9):1174–1179.
- Soares P, et al. (2009) Correcting for purifying selection: An improved human mitochondrial molecular clock. *Am J Hum Genet* 84(6):740–759.
- Fu Q, et al. (2013) A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr Biol* 23(7):553–559.
- Bouckaert R, et al. (2014) BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* 10(4):e1003537.
- Starikovskaya EB, et al. (2005) Mitochondrial DNA diversity in indigenous populations of the southern extent of Siberia, and the origins of Native American haplogroups. *Ann Hum Genet* 69(Pt 1):67–89.
- 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. *Hum Biol* 82(5-6):677–693.
- 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.
- Ebeneserdóttir SS, et al. (2011) A new subclade of mtDNA haplogroup C1 found in Icelanders: Evidence of pre-Columbian contact? *Am J Phys Anthropol* 144(1):92–99.
- Reuther JD (2013) *Late Glacial and Early Holocene Geochronology and Terrestrial Paleocology in the Lowlands of the Middle Tanana Valley, Subarctic Alaska* (Univ of Arizona, Tucson, AZ).
- Liversidge HM, Molleson T (2004) Variation in crown and root formation and eruption of human deciduous teeth. *Am J Phys Anthropol* 123(2):172–180.
- AlQahtani SJ (2008) *Atlas of Tooth Development and Eruption* (Queen Mary University of London, Barts and the London School of Medicine and Dentistry, London).
- Schutkowski H (1993) Sex determination of infant and juvenile skeletons: I. Morphognostic features. *Am J Phys Anthropol* 90(2):199–205.
- Turner CG, Nichol CR, Scott GR (1991) in *Advances in Dental Anthropology*, eds Kelly M, Larsen C (Wiley-Liss, New York), pp 13–32.
- Sciulli PW (1998) Evolution of the dentition in prehistoric Ohio Valley Native Americans: II. Morphology of the deciduous dentition. *Am J Phys Anthropol* 106(2):189–205.
- Rohland N, Hofreiter M (2007) Ancient DNA extraction from bones and teeth. *Nat Protoc* 2(7):1756–1762.
- Rohland N, Hofreiter M (2007) Comparison and optimization of ancient DNA extraction. *Biotechniques* 42(3):343–352.
- Krishnan A, Sweeney M, Vasic J, Galbraith D, Vasic B (2011) Barcodes for DNA sequencing with guaranteed error correction capability. *Electron Lett* 47(4):236–237.
- Gansauge M-T, Meyer M (2013) Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat Protoc* 8(4):737–748.
- Dabney J, et al. (2013) Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc Natl Acad Sci USA* 110(39):15758–15763.
- Li H, et al., 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
- DePristo MA, et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43(5):491–498.
- Thorvaldsdóttir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief Bioinform* 14(2):178–192.
- Jónsson H, Ginolhac A, Schubert M, Johnson PL, Orlando L (2013) mapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* 29(13):1682–1684.
- Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal* 17(1):10–12.
- Briggs AW, et al. (2007) Patterns of damage in genomic DNA sequences from a Neandertal. *Proc Natl Acad Sci USA* 104(37):14616–14621.
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol* 30(4):772–780.
- van Oven M, Kayser M (2009) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* 30(2):E386–E394.
- Lanfear R, Calcott B, Ho SYW, Guindon S (2012) Partitionfinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol Biol Evol* 29(6):1695–1701.
- Stamatakis A (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.
- Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24(8):1586–1591.

## Supporting Information

Tackney et al. 10.1073/pnas.1511903112

### SI 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 B.P. 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 the surface, within the burial pit; the upper cremated child was recovered at ~80 cm below the surface. Several continuous and discontinuous Ab horizons (Ab4) reflecting Typic Cryorthent shrub tundra-derived soils are expressed between 50 and 260 cm below the surface. Postdepositional 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 Upward Sun River site. The rapid burial of remains by wind-blown silt (loess) and very fine sand during the terminal Pleistocene and early Holocene (13,200–9,000 cal B.P.) 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 B.P. to present) (34). The pH values for the terminal Pleistocene and early Holocene deposits at the Upward Sun River site show relatively high alkaline characteristics range between 9.5 and 6.90, with an 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 the Upward Sun River site begin to trend toward more acidic values (6.90–5.60 in pH) around 1 m 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 [three 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  $9970 \pm 30$  B.P. (11,600–11,270 cal B.P.). 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 wk postnatal 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 nonmetric 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), whereas Individual 2 was more homogeneous in staining, generally 2.5YR 4/4 (yellowish red).

Excavation protocols are described in ref. 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 cataloged by J.D.I. at the University of Alaska Fairbanks Archaeology Laboratory.

**Legal and Ethical Issues Pertaining to the USR Specimens.** The Upward Sun River site burials were found on land owned by the State of Alaska. Before excavations were initiated in 2010, a Memorandum of Agreement 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 nonprofit Native organization, as invited signatories. This Memorandum of Agreement stipulated the process to be followed if human remains were conducted, following the Native American Graves Protection and Repatriation Act. 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. Of the sample, 80–120 mg was digested in a 1-mL buffer consisting of 0.5 M EDTA, 250  $\mu$ g/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  $\mu$ L of Ancient DNA GLASSMILK (silica suspension) components of the GENECLEAN For Ancient DNA Kit (MP Biomedicals), along with a final concentration of 0.05% Tween-20. This solution was incubated at 37 °C with rotation for 3 h. 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  $\mu$ L of TE<sup>-4</sup> (10 mM Tris, 0.1 mM 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 HVR1 was amplified and sequenced as described in ref. 31. The extraction blank and numerous water PCR blanks were processed at the same time. Variants present in nucleotide positions 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 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 a unique set of guaranteed error correcting codes that are redundant up to 2-bit errors (at most one nucleotide error in base space or two errors in flow space) (42) (Table S3). During adapter ligation and nick repair, final adapter concentrations in the reaction were reduced to 0.04  $\mu\text{M}$ . Unamplified libraries were eluted into 22  $\mu\text{L}$   $\text{TE}^{-}$  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- $\mu\text{L}$  primary library amplification reaction with AmpliTaq Gold 360 Master Mix (Life Technologies). This primary amplification was limited to  $\leq 15$  cycles. Final primary amplification products were eluted into 40  $\mu\text{L}$  of  $\text{TE}^{-}$  with UltraClean PCR Clean-Up Kits (MO BIO Laboratories). All previous and subsequent amplifications were done with Ion\_Aamp and Ion\_P1amp amplification primers at 0.4- $\mu\text{M}$  final concentrations (Table S3).

**Hybridization Capture and Sequencing of mtDNA.** Hybridization capture of mtDNA was performed as in ref. 24, using the lower hybridization and wash temperatures of ref. 44, for  $\sim 48$  h. In a modern genetics laboratory in a separate building, two long-range PCR amplicons were created (with Phusion Hotstart Flex; New England Biolabs) 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 100-bp target peaks. Biotinylated bait molecules were immobilized on Dynabeads MyOne Streptavidin C1 beads (Life Technologies). Each library was captured separately with  $\sim 270$  ng of bait, 320–613 ng of library (from multiple secondary amplifications with Q5 Hot Start mastermix; New England Biolabs), and blocking oligos appropriate to our Ion Torrent adapters (Table S3) at 1.9- $\mu\text{M}$  working concentrations. Libraries were released from the beads with a final 5-min incubation at 95  $^{\circ}\text{C}$ . Eluted captured molecules were quantified via qPCR, amplified with Q5 Hot Start mastermix into the exponential phase (43), and purified with silica columns.

Before sequencing, the libraries were assessed for concentration and fragment size distribution using a fragment analyzer (FA; Analytical Technologies). The FA results did not yield measurable 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 100 pM based upon the quantitative real-time PCR results. The individual libraries were further diluted to 10 pM before 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 2- $\mu\text{L}$  aliquot of the postemulsion PCR, 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 P1v2 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 suboptimally 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 TVC 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 used with the following changes: We omitted all sequences below 30 bp in length (postquality trimming) with additional Basecaller arguments (`-trim-min-read-len 30-min-read-length 30`; note default barcode settings allow for two 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 MAPQ < 30 using Samtools 0.1.19 (45). A stand-alone version of 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 igtools (47), replacing TVC FDP depth counts at called deletions. General sequencing QC metrics were analyzed with FastQC v. 0.11.2 ([www.bioinformatics.babraham.ac.uk/projects/fastqc/](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 *Authentication of aDNA Work*, below). 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 reanalyzed 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 reprocessed from both Ion P1v2 chips within Torrent Suite 4.0.2 without a reference genome (no mapping or related settings), without reads below 30 bp in length (`-trim-min-read-len`

30-min-read-length 30), and with default Ion quality trimming for v4.0.2. 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 ([sourceforge.net/projects/picard/](http://sourceforge.net/projects/picard/)). Of note, SamToFastq removes Ion flow space data, as Ion uses a nonstandard 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  $\geq 20$  (-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 30 bp and above length of 120 bp. We then used Samtools 0.1.19 (45) to remove reads with 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 are 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 finding 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  $\geq 30\%$  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 *Sanger Validation and Contamination Estimates*, below) 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 sites of low level damage. For USR2, many fewer additional substitutions were observed: 1 C > T, 1 G > A, two other nucleotide substitutions, and three other potential indels. Again, these were all near or within polynucleotide tracts, although 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  $\geq 30$  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%; see above for barcode issues and alternative processing) (Table S1). 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 0.65%, 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  $\geq 30$  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. S24).

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 (30–120 bp), 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 HVSI 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, and 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 nucleotide position 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](http://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. 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 C1c sequences (33), and one C1f sequence from the Mesolithic UZOO-74 individual (25). A haplotype C4a1a1a from the Teletut

of South Siberia (32) was used as the outgroup, for a total of 189 sequences (Dataset S1). 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 (Dataset S1).

Initially, a subset of these sequences were selected for an MP tree created using mtPhyl 4.015 (<https://sites.google.com/site/mtphyl/home>) and then manually edited for clarity (Fig. 2). Next, MAFFT (51) was used to align all of the sequences from each list with the highly accurate L-INS-i methodology. Once aligned, nucleotide positions representing C inserts between 303 and 315 (leaving any SNPs at 310), AC indels at 515–523, SNPs at 16182C and 16183C, C inserts between 16184 and 16193 (leaving any SNPs), and SNPs at 16519 were removed from the alignment. These sites are known mutational hotspots 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+ $\Gamma$ ), partitioning the two control regions separate from the coding region. ML phylogenetic trees were constructed using RAXML v8.1.15 (`raxmlHPC-PTHREADS-SSE3-T16-f-d-m GTRGAMMAI`) for 200 iterations (`-p 'random' -N 200`) and 1,000 nonparametric bootstrap replicates (`-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 ([tree.bio.ed.ac.uk/software/figtree/](http://tree.bio.ed.ac.uk/software/figtree/)) (Fig. 3).

ML 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+ $\Gamma$  mutation model (discrete distribution with 32 categories), and option G (the three partitions as above;  $\text{Malp} = 1$ ). Mutational distances were converted into years using a corrected molecular clock proposed by ref. 27 or a whole-genome substitution rate of  $2.67 \times 10^{-8}$  sub per site per year, determined by a Bayesian approach using 10 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 y ago for USR1 and USR2, and 8,300 y ago for UZOO-74. For the B2 dataset, 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+ $\Gamma$  site model, a Coalescent Bayesian Skyline tree prior (three populations), and a lognormal clockRate prior ( $M = 2.67\text{E-}8$ ,  $S = 1.4$ ). 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% HPD intervals from this tree. For the C1 dataset, 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 five populations.

**Authentication of aDNA Work.** Pre-PCR work was carried out in a dedicated aDNA 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 1,000) 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 airflow 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 personnel 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 laboratory 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 (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 aDNA 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 because of 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 next generation 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). Although 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, although not as smoothly distributed as some previously reported aDNA samples (48) (Fig. S4B).



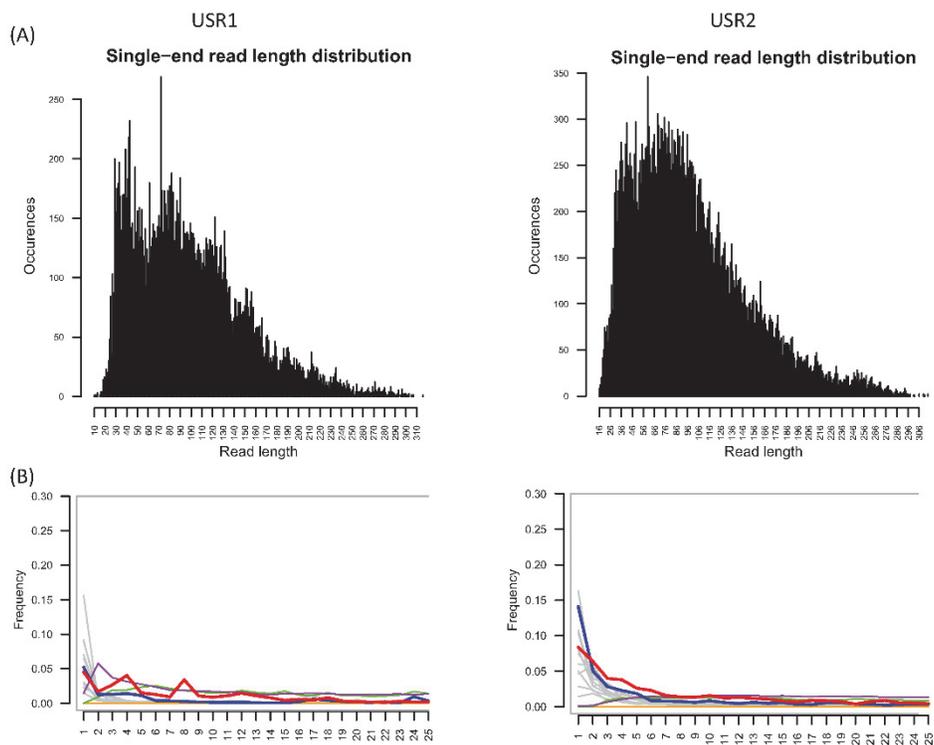


Fig. S2. (A) Read-length histograms. (B) Position-specific substitutions from the 5' end of reads post-Torrent Suite pipeline. All graphs produced by MapDamage v2.0.2-12; C-to-T substitutions are shown in red; G-to-A substitutions are shown in blue; insertions are shown in purple; deletions are shown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

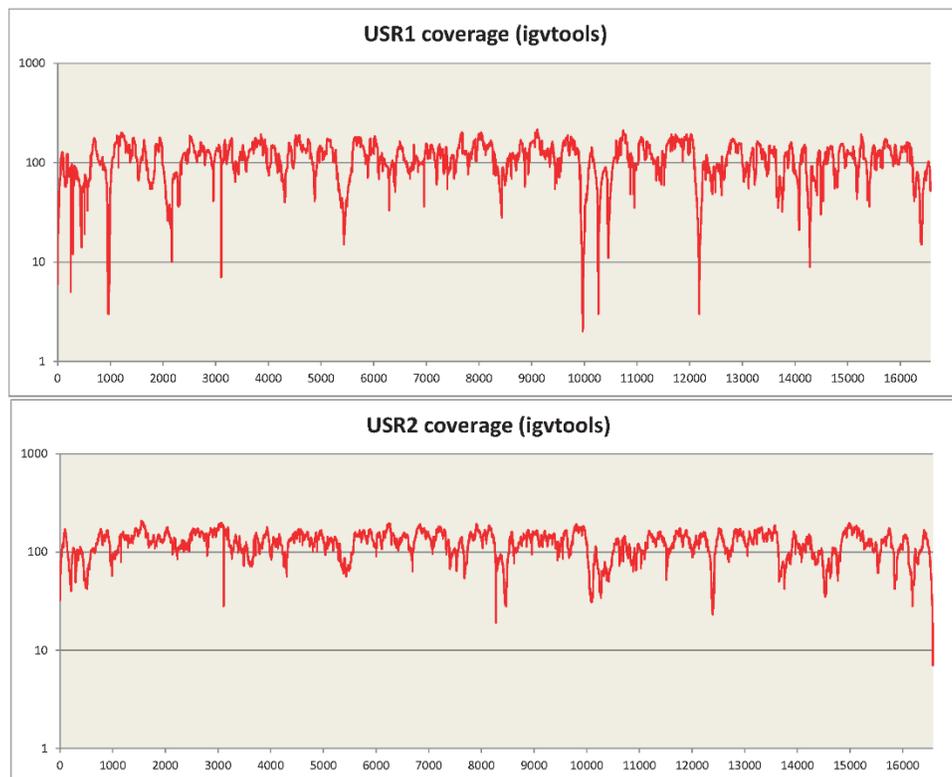
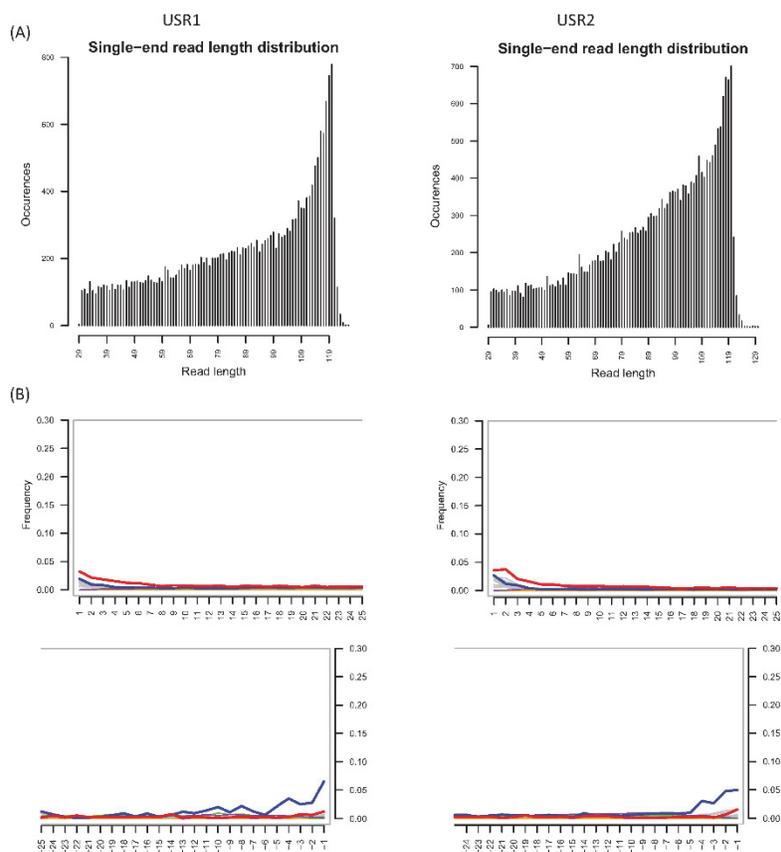


Fig. S3. Sequence coverage after the alternative pipeline across the mitochondrial genome for USR1 and USR2 on a 1-base sliding window. Coverage across indels are not corrected.



**Fig. S4.** (A) Read-length histograms. (B) Position-specific substitutions from the 5' end of reads (*Upper*) and 3' end of reads (*Lower*) following the alternative bio-informatics 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 gray.



**Table S1. Library and Torrent Suite sequencing metrics**

Torrent Suite workflow	USR1	USR2	Extraction blank	Library blank
1° Amplified library molecules per microliter	4.34E+08	2.04E+09	1.21E+08	1.02E+06
Post-enriched library molecules per microliter	3.74E+06	4.83E+06	3.47E+04	2.42E+04
Amplified post-enriched library molecules per microliter	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		
Percent relative to final ISPs	0.03%	0.06%		

**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	13,205,569	7,563,075
Mapped Reads of 30–120bp, MAPQ ≥ 70	5,895,852	3,668,765
Mapped Reads post-Picard Tools Mark Duplicates	21,140	22,951
% relative to Final ISPs	0.04%	0.04%

**Table S3. Customized oligonucleotides used in this study**

Name	Sequence	Purification	Barcode
Ion_A_bar1 <sup>†</sup>	C* C* A* T* CTCATCCCTGGCTGTCTCCGACTCAGTGC* G* G* G* C	HPLC	TGCCGGGC
Ion_A_bar1comp <sup>†</sup>	C* C* C* C* GCCACTGAGTCGGAGACACCCAGCGATCAGAT* G* G* T* T	HPLC	
Ion_A_bar3 <sup>‡</sup>	C* C* A* T* CTCATCCCTGGCTGTCTCCGACTCAGTGC* C* A* C* T	HPLC	TGTCCACT
Ion_A_bar3comp <sup>†</sup>	A* G* T* C* GACACTGAGTCGGAGACACCCAGCGATCAGAT* G* G* T* T	HPLC	
Ion_P1	C* C* A* C* TACGCTCCGCTTTCCTCTCTA* TGGCAGTCGG* T* G* A* T	HPLC	
Ion_P1comp	A* T* C* A* CCGACTGCCCATAGAGAGAAAGCCGACGGCTAGT* G* G* T* T	HPLC	
Ion_Aamp	CCATCTCATCCCTGGCTGTCTCCGACTCAGTGC	Standard desalting	
Ion_P1amp	CCACTAGCCCTCCGCTTTCCTCTCTATC	Standard desalting	
A1_Block <sup>‡</sup>	GCCCGGACTGAGTCGGAGACACCCAGCGATCAGATGGT/3SpC3/	IDT xGen blocking; HPLC	TGCCGGGC
A3_Block <sup>‡</sup>	AGTGCACACTGAGTCGGAGACACCCAGCGATCAGATGGT/3SpC3/	IDT xGen blocking; HPLC	TGTCCACT
P1_Block <sup>‡</sup>	ATCACCAGCTGCCCATAGAGAGAAAGCCGACGGCTAGTGGT/3SpC3/	IDT xGen blocking; HPLC	

<sup>†</sup>Adapters were incorrectly designed without a "GAT" barcode adapter sequence following the custom barcode.

<sup>‡</sup>Blocking oligos were incorrectly synthesized with unnecessary TT overhangs.

## Other Supporting Information Files

[Dataset S1 \(XLSX\)](#)

[Dataset S2 \(XLSX\)](#)

CHAPTER 2

MOLECULAR ANALYSIS OF AN ANCIENT THULE

POPULATION AT NUVUK, POINT

BARROW, ALASKA.

## Introduction

The Inuit, Iñupiat, and Eskimo peoples of the North American Arctic retain a large portion of their ancestry that is separate from all other Native American tribes. Beginning with population genetic surveys of single loci, discrete markers, and mitochondrial DNA (mtDNA), and continuing with high density single-nucleotide polymorphism (SNP) and microsatellite genotyping projects, Eskimo-Aleut speakers were shown to be genetically unique. They are the most closely related Native Americans to northeastern Asian people, and there is evidence of substantial population gene flow back and forth across the Bering Strait (Reich et al., 2012; Tackney et al., 2016; Wang et al., 2007). Recent whole genome sequencing of 31 individuals from within the Americas, Siberia, and Oceania included two Inuit from Greenland. The analysis placed the Inuit in a clade distinct from other Native Americans and shared with Siberian Yupik and Koryak people, who were themselves the most genetically similar Eurasian populations to Native Americans (Raghavan et al., 2015). As the consensus among geneticists is for a Late Pleistocene, post-Last Glacial Maximum (LGM) initial peopling of North and South America (“First Americans”; Goebel et al., 2008; Reich et al., 2012), the coalescence dates for Eskimo-Aleut mtDNA haplogroups additionally stand out for being firmly rooted in the past 10,000 years, and often substantially more recent (Dryomov et al., 2015; Tackney et al., 2016).

A long history of archaeological study supports the distinctness of Circum-Arctic Native Americans and their relatively recent Holocene migrations into northern North America (Jensen, 2014). Apart from central Beringian refugia potentially occupied during the initial peopling event (Hoffecker et al., 2014), during the LGM and

substantially following it most of the North American Arctic was inaccessible due to the massive Laurentide Ice Sheet. The first coastal inhabitants of Alaska, and the first to appear across what is now Canada and Greenland, were members of the Arctic Small Tool tradition (ASTt). Likely deriving from Siberian Neolithic cultures, they employed a multitude of microblades, created diagnostic small end and side blades for inserting into arrowheads, and were adapted for some maritime subsistence (Jensen, 2014). The oldest evidence for the ASTt is found at Kuzitrin Lake on the Seward Peninsula of Alaska, with an initial occupation of 5,500 cal BP (Harritt, 1998). In the subsequent millennia across the Arctic, ASTt developed into the “Paleo-Eskimo” traditions of Pre-Dorset/Dorset in Canada, the Independence I-Saqqaq in Greenland, and the Denbigh Flint Complex in Alaska. Further developments in Alaska saw the intermediate Choris and Old Whaling cultures (potentially direct descendants of the Denbigh Flint Complex) evolve into the Norton tradition. By 1,600 cal BP, the Ipiutak appear in the West with elaborate ivory carvings and a strong reliance on marine mammals, though there is some marked discontinuity of artifact remains with the earlier Norton. In particular, Ipiutak lacked whale-hunting equipment. From the type-site of Ipiutak near Point Hope, AK (Larsen and Rainey, 1948; Mason, 2006) and further south, a large number of Alaskan inland and coastal sites are known, though there are two northern locations at the inland Anaktuvuk Pass (Mills et al., 1999) and along the coast at Nuvuk, near Point Barrow (Jensen, 2009a; see below).

While the Paleo-Eskimo were present in the Arctic for thousands of years, archaeological and genetic studies (see below) suggest that modern Eskimo are in fact descendants of a second, separate and much more recent tradition known as the Northern

Maritime or Neo-Eskimo. Neo-Eskimo origins most likely lie within Old Bering Sea (OBS), a single culture existing from the last centuries BC for close to a millennium on the Chukchi Peninsula and St. Lawrence Island. OBS artifacts are characterized by richly ornamented harpoon heads, with kayak and umiak watercraft for proficient seal and walrus hunting. At around 400 AD, OBS people are contemporaries with the Ipiutak and, by 900 AD, they are supplanted both by the Neo-Eskimo Punuk tradition, which arose around St. Lawrence Island (and is possibly an expanding but simplified development from Old Bering Sea), and the Neo-Eskimo Birnirk tradition, which is restricted to the coasts of the Chukchi Sea (with its type-site of Pigniq/Birnirk near Point Barrow) (Jensen, 2014; Mason, 2009b). At the beginning of the Medieval Climatic Anomaly an increase in rougher seas leading to higher marine resource productivity seems to have encouraged expansions, with Birnirk entering the Bering Strait (particularly Kotzebue Sound) and Punuk expanding outward with occupation at least as far east as the coastal site of Nunagiak at Point Belcher. Interactions of Punuk and Birnirk seem to involve refinements in marine mammal hunting to allow large-scale whaling for the first time (Jensen, 2014; Mason, 2009a; Mason, 2009b). Following these expansions the final stage of Neo-Eskimo develops; the Thule. Historically viewed as a direct descendant of Birnirk, no site unequivocally provides evidence for Old Bering Sea to Birnirk to Thule, and the origins of the Thule have been variously placed within the Bering Strait, at the northern Siberian Cape Baranov, or at Point Barrow; alternatively Thule could have ancestry within Punuk (Mason, 2009a; Mason and Bowers, 2009). It is the Thule tradition that expands eastward across the Arctic and eventually regionalizes and diversifies into the Inuit, Iñupiat, and Eskimo of today (Jensen, 2014).

The Thule, as a prehistoric culture originating in Alaska, were first described and excavated by Therkel Mathiassen in 1927 as part of the 5<sup>th</sup> Thule Expedition (the expedition lending its name to the ancient people; Mathiassen, 1927). Mathiassen saw a prehistoric maritime and cold weather adapted culture living in semisubterranean houses and relying on whaling as well as other marine mammals for subsistence. The focus on whaling has the Thule alternatively known as the Arctic Whale Hunting culture (Larsen and Rainey, 1948). By this stage of the Neo-Eskimo tradition “the contrast between Ipiutak and Thule is so pronounced that it is almost inescapable to conclude that population replacement was involved” (Mason, 2009a). While scholars have long agreed that the Western Thule sites are antecedent to the Eastern (Greenlandic) Thule (Larsen and Rainey, 1948), the timing of the migration has not been so clear. For much of the last century the initial migration was believed to occur around 1000 AD (McGhee, 2000), corresponding with the aforementioned Medieval Climatic Anomaly, however, this date now appears much too old. The oldest securely dated and clearly Thule sites are located along the north coast of Alaska starting at 1000 AD (Jensen, 2007; Jensen, 2009a; Jensen, 2009b). Furthermore, early Thule sites to the east on the Beaufort Sea and Amundsen Gulf coasts re-date to no earlier than the 13<sup>th</sup> century AD, which importantly places their occupation almost simultaneously with sites on Ruin Island of northwest Greenland (Friesen and Arnold, 2008). This rapid movement from northern Alaska to the far eastern Arctic in a century or two is not only astonishing, but indicates a migration shortly before the onset of the Little Ice Age around 1400 AD. This climatic event had large repercussions for the classic Thule across the Arctic. Communities undergo profound social change including demographic increases, settlements away from prime whaling

coastal regions and into a growing number of inland sites, diversification of subsistence economies, and the development of an extensive traditional trading network across the Arctic, particularly with Siberian groups via St. Lawrence Island (Friesen and Arnold, 2008; Jensen, 2014; McGhee, 2000).

Ancient DNA (aDNA) from human remains associated with Paleo-Eskimo, Neo-Eskimo, and their putative ancestral cultures would help illuminate ancestor-descendent relationships and track prehistoric population migrations. Almost all modern Circum-Arctic population genetic studies have focused on mtDNA and they indicate a group of “Beringian-specific” haplogroups: A2a, A2b, D2a, and D4b1a2a1a. These haplogroups are present not only in North American Eskimo-Aleut speaking populations, but also in Athabaskans and native northeastern Siberian Chukotkans (Dryomov et al., 2015; Helgason et al., 2006; Raff et al., 2015; Tackney et al., 2016; Volodko et al., 2008). The first ancient whole mtDNA from the Arctic was the 4,170-3,600 cal BP remains from the Independence I-Saqqaq Greenlandic archaeological tradition. This individual carried a variant of D2a, and haplogroup D2a is currently found in modern Aleutian islanders and Siberians (Dryomov et al., 2015; Gilbert et al., 2008). Further ancient mtDNA data was provided by a low resolution but wide sampling of Dorset, Saqqaq, Thule, and Norse affiliated remains from across Canada and Greenland, as well as Siberian Birnirk samples and southern Alaskan late-prehistoric Thule samples. The Thule samples all carried haplogroups A2a, A2b, and D4b1a2a1a, while the Paleo-Eskimo samples all carried haplogroup D2a. The authors conclude not just genetic continuity from Neo-Eskimo to modern Eskimo, but also the lack of maternal gene-flow between Paleo-Eskimo and Neo-Eskimo. The five Siberian Birnirk samples all carried haplogroup A2a, indicative of

Neo-Eskimo, and none of the Norse samples carried any Native American haplogroups (Raghavan et al., 2014).

The full nuclear genomes of a small number of individuals have further increased our knowledge of the ancient Arctic. The first published was the Paleo-Eskimo Saqqaq high coverage genome (Rasmussen et al., 2010). Later came a handful of low coverage genomes from combined datasets of multiple individuals representing “Middle-Dorset,” “Late-Dorset,” “Canadian Thule,” “Greenlandic Thule,” and “Siberian Birnirk” (Raghavan et al., 2014). These whole nuclear genomes clustered the Thule and Birnirk samples with modern Greenlandic Inuit, but indicated gene flow with the earlier Paleo-Eskimo, Aleutian Islanders, and Siberian Chukotkans. Intriguingly the gene flow with the Paleo-Eskimo dated before any Neo-Eskimo tradition is known on the landscape and so likely occurred in Western Beringia. The Inuit additionally showed more Native American ancestry than the Paleo-Eskimo Saqqaq, with the former closer to Athabaskans and the latter closer to Asian Siberian groups (Raghavan et al., 2014; Raghavan et al., 2015). Both the mitochondrial and nuclear genomic data, therefore, suggest the relative isolation of the Paleo-Eskimo from other Native American populations, past or present.

Until recently, these population studies lacked any characterization of the Iñupiat people of the Alaskan North Slope. Coastal north Alaska is a likely origin point of the Thule (Mason and Bowers, 2009), yet Alaskan Eskimo-Aleut speaking populations are absent from previously published surveys of genetic variation in the Arctic (e.g., the Chukchi Peninsula and the eastern North American Arctic; Helgason et al., 2006; Raff et al., 2015; Tackney et al., 2016). To fill in this gap and to better understand the history of the people in this important geographic location Raff et al. (2015) sequenced the

mitochondrial D-loop from 137 individuals in eight Alaskan North Slope communities. Not only did the expected mtDNA haplogroups of A2a, A2b, and D4b1a2a1a appear, but haplogroup D2a was also typed. This was unusual because previously the most northern extent of D2a in North America was in the Aleuts of the Aleutian islands or in ancient Paleo-Eskimo remains (Helgason et al., 2006; see above). Additional matrilineal haplotypes were also found. One A2 sequence did not have diagnostic SNPs for either A2a or A2b and it might represent one of the southern Pan-American A2 lineages. Two remaining matrilineal haplogroups carried haplogroup C4, making them the most northern published evidence for a Pan-American mtDNA not easily accounted for by historic admixture. The North Slope is more diverse than expected (Raff et al., 2015). To complement this survey it would be necessary to sample from a coastal north Alaskan Thule site that dates from the Birnik-Thule transition to pre-European contact. The distinct lack of carefully excavated stratified sites in North Alaska prohibited such an endeavor until Dr. Anne Jensen began excavating the cemetery of Nuvuk in 2005 (Jensen, 2009a). Nuvuk provides the necessary archaeological and potentially ancient genetic information on the Thule before, during, and after their spread eastward.

Nuvuk, near Point Barrow and once the northernmost indigenous community of Alaska, represents a nearly one thousand year uninterrupted occupation from early Thule to postcontact Iñupiat Eskimo (Jensen, 2009b). First contact with European explorers around Point Barrow occurred with the expedition of the HMS *Blossom* in 1826 (Beechey, 1831), and the first dedicated scientific expedition was during the first International Polar Year from 1881-1884 (Ray, 1885). Nuvuk was situated near the related community of Utqiagvik, which would later become the modern city of Barrow.

Nuvuk appears to have been the larger of the two communities until the end of the 19<sup>th</sup> century. Around this time a combination of drastic beach erosion at the Point and an increase in European institutions at Utqiagvik eventually led to the abandonment of Nuvuk in the early 20<sup>th</sup> century (Jensen, 2009a; Maguire, 1988). Scientific expeditions of the 20<sup>th</sup> century included minimal salvage excavations at Nuvuk, with the focus primarily on other locations near Point Barrow, particularly Pigniq, the Birnirk type-site and only a few kilometers to the southwest of Nuvuk. It was believed that there was nothing left of archaeological value at Nuvuk and it represented a late prehistoric site at best, even if some Thule artifacts were occasionally discovered (Jensen, 2009a and references therein). With further erosion, however, the first reports of human remains from Nuvuk appeared in 1997. Designated “Nuvuk-01,” this individual was buried with an extensive array of artifacts including diagnostic harpoon heads and various wood, bone, and ivory hunting and manufacturing gear. With a calibrated AMS radiocarbon date of 810 - 1020 AD, this individual was clearly a member of the Early Thule / Western Thule tradition (Jensen, 2007; Jensen, 2009a). Subsequent appearance and loss of more human remains from erosion initiated the community driven Nuvuk Archaeology Project in 2005, with dual focus of scientific study and culturally appropriate reinterment of exposed burials further inland (Jensen, 2009a; Jensen, 2012).

Two distinct occupations are revealed at Nuvuk - the aforementioned Ipiutak site, with calibrated dates within the 4<sup>th</sup> century AD (making it the earliest Ipiutak site known) and a later, long lasting Thule occupation at least as early as Nuvuk-01 through the 16<sup>th</sup> century. People continue to use Nuvuk into the contact period as shown by an associated dated Peat Locus work area and grave markers from the 1920s (Jensen, 2009a). To date

85 graves have been excavated, all within the Thule occupation and, unlike the first burial discovered, the majority lack grave goods. However, the few artifacts recovered do indicate clear links to other Western and Eastern Thule sites (Jensen, 2009b). This makes Nuvuk the largest Thule cemetery ever excavated in North America, where Thule grave sites have been poorly represented compared to Siberian Chukotka (Jensen, 2009a). By necessity, most of the more than 159 dates from Nuvuk were taken from marine mammal products or wood. These indicate the cemetery was contemporaneous with known Punuk and Birnirk dated sites, and clearly preceded and overlapped dates from Thule sites in the Eastern Arctic (Friesen and Arnold, 2008; Jensen, 2009b). A recent reassessment of the temporal occupation at Nuvuk directly AMS dated 54 human remains (bone collagen) and recalibrated 31 burial-associated materials. The human remains (unfortunately lacking Nuvuk-01) had  $2\sigma$  calibrated, marine reservoir corrected, dates starting at minimum in 975 AD and continuing to 1885 AD. The associated burial artifacts had a range within that from 996 AD to 1631 AD (Coltrain et al., Under Review). In all, the Nuvuk burials confirm an unambiguous Early Thule presence around the city of Barrow and the site represents the earliest record of Thule on the North Slope (Jensen, 2009a). With the nearby Birnirk occupation, and the older Nuvuk Ipiutak artifacts, Point Barrow is of great importance for understanding the peopling of the Arctic.

The relationship of the individuals buried at Nuvuk to the modern Iñupiat of the North Slope and to other Circum-Arctic populations to the east and west can be revealed with comparative ancient DNA analyses. We selected 44 individuals from Nuvuk with calibrated median intercept dates between 1139 AD and 1669 AD (Coltrain et al., Under Review). We amplified and Sanger sequenced the hypervariable segment I (HVR-I) of

the control region of their mtDNA. The HVR-I is by far the most published sequence from the mitochondrial genome, and the Beringian-specific mtDNA haplogroups are easily characterized from polymorphisms in this small region. We compared the ancient population of Nuvuk with previously published HVR-I sequences of modern and ancient communities from across Asia and North America, including all populations showing at least one individual carrying an Arctic variant. We attempt a synthesis of the genetic data to support or reject coastal north Alaska as the origin point of the Thule.

## **Materials and Methods**

### *Sample Collection and Archaeological Context*

Point Barrow is currently a gravel spit at the northern end of the Arctic Coastal Plain, though at one point it might have been covered by vegetation. The highest point on the spit was 4 m above sea level at the start of excavations, but accelerating coastal erosion has decreased both the height and area of the Point since. Permafrost does exist, but it is found deeper than at other locations around Barrow (Jensen, 2009a). This has dual repercussions for archaeological sites on the spit: First, the permafrost did not prevent the erosion or site disturbance of burials and artifacts over the centuries. Second, human remains are found above, and not within, frozen ground. Any remaining DNA in the bones has been exposed to seasonal ground temperature and moisture content fluctuations, making the site not typically ideal for aDNA studies. The gravel above the burials was mostly dry at the surface and progressively wetter the deeper the excavation; ground temperature during the summer months was recorded at  $\sim 5^{\circ}\text{C}$ . The burials of Nuvuk were excavated starting with the most tenuously situated ones located on the beach ridge and working further inland.

Human remains were sampled as part of the archaeological excavations during the summer months of 2006-2011. Upon discovery of a burial, sediment was delicately removed until the first signs of the rib cage. At this point all researchers moved away from the burial and an individual garbed in sterile gloves, arm sleeves, and a face mask proceeded to reveal the rib bones with a dedicated brush and bleached tweezers. A small rib bone was selected and removed with the surrounding sediment into a sterile plastic bag. The bag was sealed, identified, and not opened again until within the aDNA cleanroom facility. At this point the other investigators returned to the burial to resume excavation. Forty four individuals were sampled in this manner.

#### *Sample Analysis Agreements*

Letters of support were requested and received by the Ukpeaġvik Iñupiat Corporation, the Native Village of Barrow, and the Barrow Senior Advisory Council. All understood that small samples of human remains (typically rib fragments) were to be collected during burial exposure, transported out of Alaska to Utah, and analyzed in a destructive manner to determine ancestral relationships with modern Iñupiat communities. No guarantees were made as to the success of the ancient DNA project, but annual presentations were made to the community of Barrow and other North Slope communities concerning progress.

Logistical support while in the field for collection was provided primarily through the Barrow Arctic Science Consortium (BASC) through a National Science Foundation (NSF) funded Cooperative Agreement.

### *DNA Extraction*

DNA was extracted from the rib fragments with a silica-based method (Rohland and Hofreiter, 2007a; Rohland and Hofreiter, 2007b). Ribs were manually broken into small sized pieces of ~0.5 mg total weight. The pieces were soaked in 20% bleach solution for 10 minutes and then washed three times in ultra-pure water. The bone fragments were dried under UV light for 40 minutes to complete the surface decontamination.

The bone mineral matrix was digested in a 15 mL Falcon tube with 4 mL of buffer consisting of 0.5M EDTA, 250 µg/ml proteinase K, and 40 mM DTT. The mixture was rotated overnight at 55°C. The extract was spun down at 5000 rpm for 5 min and 1 mL of the supernatant was mixed in a fresh 15 mL Falcon tube with 4 mL of the Guanidine Thiocyanate-based Dehybernation Solution A and 200 µl of the Ancient DNA GLASSMILK™ (silica suspension) components of the GENE CLEAN for Ancient DNA Kit (MP Biomedicals). The solution was incubated at room temperature with rotation for 2 hr and then spun down at 6000 rpm for 2 min. DNA was purified from the silica pellet as per the manufacturer's protocol for the rest of the GENE CLEAN kit. Final elution was in 110 µl of TE<sup>-4</sup> (10mM Tris, 0.1mM EDTA). DNA extracts were stored in LoBind tubes at -20 °C.

Duplicate extractions were performed on each bone sample, nonconcurrently, excluding six samples not duplicated due to time constraints. Water extraction blanks were processed through all steps at each day of extraction.

*Mitochondrial HVR-I Amplification and Sanger Sequencing*

Portions of the HVR-I were amplified with the Polymerase Chain Reaction (PCR) using AmpliTaq Gold DNA Polymerase (ABI) or AmpliTaq Gold 360 DNA Polymerase and the following reaction conditions: 2.5-3.5 mM Mg<sup>2+</sup>, 200 nM dNTPs w/dUTP, and 200 nM each primer in a 50 µl reaction. PCR cycling was completed in 40-45 cycles at annealing temperatures of 56-58°C. One PCR water blank was processed with each mastermix. Amplicons were visualized on agarose gels and successful amplifications were cleaned using QIAquick PCR Purification Kits (Qiagen) or UltraClean PCR Clean-Up kits (MoBio Laboratories). Purified PCR products were sequenced on both strands at the DNA Sequencing Core Facility, University of Utah.

All extracts were initially amplified using primers HVR1\_P1F and HVR1\_P1Rc (159 bp product), HVR1\_P2Fb and HVR1\_P2R (134 bp product), and HVR1\_P3F and HVR1\_P3R (157 bp product) (Table 2S.1). These products provide coverage of nucleotide positions (np) 16043-16161, 16183-16277, and 16288-16402 of the rCRS reference genome (NC\_012920). Following the identification of haplogroup D4b1a2a1a, it was clear that diagnostic SNP 16173(C>T) was internal to two primer binding sites. Further amplifications were performed using primers HVRI\_P4F and HVRI\_P4R (249 bp product; admittedly large for aDNA work) or Car\_P1F and Car\_P1R (105 bp product), providing total coverage of np 16043-16402 or 16043-16277 + 16288-16402, respectively (Table 2S.1). All extracts, including duplicates, were amplified for the first 3 fragments. Only one extract was chosen for each sample for amplification of HVRI\_P4F/HVRI\_P4R or Car\_P1F/Car\_P1R.

### *Authentication of Ancient DNA Work*

Pre-PCR work was carried out in a dedicated ancient DNA facility. 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 gauged pressure differentials between each subspace. Room-wide UV lighting provides daily surface/air sterilizations. All entering 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. All PCR reactions are setup with uracil-containing dNTPs as a failsafe against contaminating amplicons. Post-PCR work is located in a separate isolated laboratory space in the same building and strict cleanroom workflow requirements are followed by all personnel.

No personnel carry any of the HVR-I haplotypes reported in this study and all extractions and PCR mastermixes were run with extraction and amplification water blanks, respectively. No sequence was used unless both corresponding blanks were clean. All HVR-I SNPs typed from the first 3 fragments (see above) were validated in a duplicate series of extractions and amplifications; if necessary, a third series was completed or the sample was abandoned and its sequence not reported here.

No HVR-I haplogroups reported here are surprising to find in an Arctic population and all of the haplogroups generally make phylogenetic sense. We note, however, that many of the haplogroup defining SNPs are also C-to-T and G-to-A substitutions that are expected in ancient DNA sequences due to damage (Briggs et al.,

2007). SNPs were only called if the sequencing electropherograms were unambiguous. Double peaks were not-called and either the PCR or sequencing reaction was repeated until unambiguous calls could be made.

#### *Literature Review Dataset*

Published literature with mitochondrial genome sequence from populations across Asia and the Americas was examined for individuals carrying haplogroups A2a, A2b, D2a, D2b, D4b1a2a1a, D4b1a2a1b, or A2 (non-A2a or A2b) (We follow the naming convention for haplogroup D of Derenko et al., 2010). Datasets were required to consist of either full HVR-I sequence, whole mtDNA genome sequence, or targeted SNP genotypes indicating membership in these haplogroups (e.g., Raghavan et al., 2014). Sequences were scanned for haplogroup motifs (and derived variants) as indicated and described in Table 2S.2. Two subtypes of A2a, A2a4 and A2a5 (Achilli et al., 2013; Dryomov et al., 2015), were recorded separately due to their unique distribution in southern Athabaskan populations. Haplogroup motifs for the root of A2, present at low frequency in some Arctic populations (Helgason et al., 2006), were also recorded separately. Important caveats arose as a consequence of only using HVR-I sequence: First, it was impossible to distinguish D2a and D2b without further typing of mtDNA SNPs. When these SNPs were not additionally typed, educated assumptions were made for each population (see below). Second, there are some rare A2a and A2b mtDNA genome sequences that lack C16192T and A16265G, respectively (Dryomov et al., 2015; Volodko et al., 2008). Our HVR-I search by definition misses these individuals in other studies (see Table 2S.2).

Population frequencies of these haplogroups were recorded for each ancient or

modern population containing at least one individual with at least one of the above haplogroups (Table 2S.3). An exception was made for the root A2 motif - to avoid including the numerous Native American tribes that contain pan-American haplogroup A2, populations were only included if an additional Beringian haplogroup was present or if the population was an ancient Arctic or Sub-Arctic sample (e.g., aDNA from Brooks River). For reports that analyzed the same samples, the newest and largest sample set was used for haplogroup frequency determination. For reports that analyzed the same population but used different samples, a simple average was calculated among the frequencies for each haplogroup.

Over the course of the literature review, some populations and/or published datasets were set aside from the main analysis. In addition to removing populations that lack any Beringian haplogroups, a dataset was excluded if it did not allow the typing of all haplogroups (e.g., the dataset was targeted to only haplogroup D HVR-I sequences). These populations were not included in the subsequent Principle Components Analysis, but they are referenced in Table 2S.4. Some additional datasets were completely removed due to a bias in selection criteria - samples were initially typed with discrete markers and then a subset was selected for HVR-I sequencing. These studies would give a biased representation of the frequency of HVR-I motifs in the original population (e.g., Malhi et al., 2001; Stone and Stoneking, 1998; Torroni et al., 1993). Even if frequency data was excluded, however, the HVR-I sequences were recorded and compared to those at Nuvuk. Finally, the dataset from Shields et al. (1993) was ignored due to possible sequencing errors (Saillard et al., 2000).

### *Statistical Analyses*

Summary statistics for both haplotype frequencies and HVR-I diversity were calculated using Arlequin 3.5 (Excoffier and Lischer, 2010). Genetic distance was calculated from the covariance of allele frequencies between pairs of populations (Harpending and Jenkins, 1973) and the subsequent principal components were plotted to describe over 98% of the variance.

Figures were produced using ggplot2 in R version 3.2.3.

## **Results**

### *HVR-I Sequencing Summary*

Of the total 44 human remains selected for DNA extraction, we were able to type 39 individuals for reproducible, contamination-free sequences from a 360 bp fragment of the human mitochondrial HVR-I (a success rate of 89%). From 11 segregating sites, 8 distinct haplotypes were characterized with haplotype diversity of  $0.76 \pm 0.06$  and mean pairwise differences of  $2.15 \pm 1.22$  (Table 2.1). These metrics are comparable to other North American Arctic populations, particularly the Alaskan North Slope, but are less diverse than those reported in Siberian populations (Raff et al., 2015).

Table 2.2 lists the original archaeological sample IDs, the recalibrated and marine reservoir corrected  $2\sigma$  date ranges, and the HVR-I polymorphisms present in the sequenced fragment relative to positions 16043-16402 of the rCRS reference genome (NC\_012920; Andrews et al., 1999). All sequences from Nuvuk fell within the known Beringian-specific haplogroups of A2a, A2b, and D4b1a2a1a given their defining HVR-I polymorphisms (see Table 2S.2). There were no sequences with D2a-specific variants. Haplogroup frequencies were 25.6% (A2a), 66.7% (A2b), and 7.7% (D4b1a2a1a) in 39

individuals (Table 2.2). There are 4 haplotypes of A2b, 3 haplotypes of A2a, and a single haplotype of D4b1a2a1a in the samples sequenced.

*Beringian Haplogroups Across Asia and the Americas*

A robust literature search of populations containing at least one individual with mitochondrial DNA haplogroups A2a, A2b, D2a, D2b, D4b1a2a1a, or D4b1a2a1b revealed many outside of the expected geographic ranges of Arctic North America and Siberian Chukotka (Table 2S.3 and 2S.4). This was driven in large part by the presence of haplogroups A2a and D4b1a2a1a. Haplogroup A2a, present in all Eskimo-Aleut speaking populations, is also found in Na-Dené speakers, specifically the northern Tlingit and Dogrib, and the southern Athabaskan Apache and Navajo (Figure 2.1). While frequently in the form of haplotypes A2a4 and A2a5 (as previously described in Achilli et al., 2013), the A2a haplotype is also present. Intriguingly, non-Na-Dené speaking North American Northern Paiute and Zuni also have A2a4/A2a5, while the Cheyenne and Arapaho, as well as the South American Zaupes, all carry a low frequency of A2a. In Asia, outside of Chukotka/ N. Kamchatka, A2a is found at very low frequency in West Siberian Selkups and South Siberian Evenks (Table 2S.3). In all these cases, it is unclear if these low frequencies reflect recent/modern admixture or true population history.

Haplogroup D4b1a2a1a, found at 7.7% frequency at Nuvuk and from 2 - 26% across North American Inuit/Iñupiat, and Siberian Chukotkans, is additionally found in a wide number of Asian groups (Figure 2.1). Various populations in the Altai-Sayan region of South Siberia, the Uighurs, Karakalpaks, and Kirghiz of Central Asia, and even the distant Bashkirs and Kalmyks of Eastern Europe all have a very low (0.6 - 2.1%) frequency of this haplogroup. Additionally, it is noteworthy that there is a high

frequency of D4b1a2a1a in the North Altai Tubalar, who at 7.7% match the frequency at Nuvuk, and D4b1a2a1a was present in the ancient Yakuts of Sakha but not in the modern Yakuts. Sister haplogroup D4b1a2a1b, not found in Beringian populations, was present in various South Siberians, Eastern Asians, and Eastern Europeans and only overlapped with D4b1a2a1a in the Bashkirs and Uighurs (Table 2S.3 and 2S.4).

Within this distribution of past and present haplogroup frequencies across populations, we performed a Principle Components Analysis on the dataset from Table 2S.3 to visualize maternal genetic relationships with the Thule population at Nuvuk. The first principal component explains 37.1% of the variance and clearly separates those populations with a high proportion of Beringian-specific haplotypes from those with a low proportion. At intermediate frequencies of non-Beringian haplotypes are the ancient samples from Brooks River and Port Moller, the southern Athabaskan Apache, the Southeast Alaskan Haida and Tlingit, and the Chukchi. The second principal component pulls the Beringian ancient Paleo-Eskimo and modern Aleutians and Commander Island, all with very high or fixed frequencies of D2a, from the modern Inuit/Iñupiat and ancient Thule and Birnirk. Nuvuk clusters with the latter group (Figure 2.2A).

Principal Component 3 separates those populations in our dataset with a high proportion of pan-American haplogroup A2, namely the Na-Dené speaking Navajo, Apache, Tlingit, and Haida. Principal Component 4 pulls out those populations with a high frequency of A2a (excluding A2a4 and A2a5) from those with a high frequency of A2b (Figure 2.1). The Thule sampled from Nunalleq and the Birnirk samples from Siberia are fixed for A2a and represent one extreme. A more diffuse cluster of Nunavut and Northeast Canadian Thule, Nuvuk Thule, modern Nunavut Inuit, and modern

Siberian Naukan seems to cluster those populations with a high frequency of A2b and D4b1a2a1a (Figure 2.2B).

Principal Component 5 separates out populations with high frequencies of A2a5, which include the Apache, Dogrib, and Yakutat Tlingit, from those populations with pan-American haplogroup A2 but without A2a5, namely the Hoonah Tlingit, Hydaburg Haida, Yavapai, and the ancient samples from Brooks River and Port Moller. Principal Component 6, which only accounts for 2% of the variance, appears to pull out the modern Naukan with the ancient populations at Mink Island and Northeast Canadian Thule on one end with high frequencies of D4b1a2a1a, and on the other end, the Sireniki, the Inuit of Greenland and the Northwest Territories, and the ancient Thule in Nunavut, Greenland, and at Nuvuk with lower or absent frequencies of D4b1a2a1a but high frequencies of A2a and A2b (Figure 2.2C). All together the 6 components reported account for 98.3% of the variance.

#### *Shared Nuvuk HVR-I Haplotypes Across Asia and the Americas*

Given the HVRI-I haplotypes sequenced at Nuvuk (Table 2.2), we searched the published population datasets for individuals with exact matches within the covered nucleotide positions of 16043-16402. While a subset of our HVR-I sequences from Nuvuk lacked coverage at 10 bases between np 16278 and 16287 (see Materials and Methods), we only found two individuals in all the datasets (One with haplogroup A2a from E. Greenland and one with haplogroup A2b from Kitikmeot; Helgason et al., 2006; Saillard et al., 2000) with a SNP in this region (C16278T) and so the chances are low that any variation in our samples has been missed. Populations with individuals that share the HVR-I haplotypes at Nuvuk are listed in Table 2.3.

Modern populations of Chukotka and the North American Inuit and Iñupiat share with Nuvuk the root A2b haplotype, as do ancient Thule samples further east. The derived A2b + 16212 haplotype has been found only in the Chukchi and in the Canadian and Greenlandic Inuit; interestingly, this haplotype was not found in the modern populations of the North Slope (Raff et al., 2015). The root A2a haplotype, as expected, is found in modern Chukotko-Kamchatkan speakers, Eskimo-Aleut speakers, Athabaskan Dogrib, and ancient Thule from West Greenland. It is additionally found in the Cheyenne/Arapaho tribes and in the Ancient Oneota at Norris Farms. The derived A2a + 16311 haplotype is present in the Chukchi, Eskimo-Aleut speakers, and the ancient Thule from West Greenland. It was additionally found in the single ancient Birnirk sample (out of 5) that was whole-genome sequenced (Raghavan et al., 2014). The derived A2a + 16261 haplotype has not previously been found in any ancient samples, but is present in modern North American Eskimo-Aleut speakers, the Chukchi, and, likely due to modern admixture, the West Siberian Selkup and the South Siberian Evenks. The haplotype of D4b1a2a1a in the 3 burials from Nuvuk matches that found in the modern Inuit further east, as well as modern Chukotko-Kamchatkan speakers and Naukan Eskimo in Chukotka. It also matches a few much more distant Asian populations, as described previously. This haplotype has additionally been found in three separate ancient samples - the ancient Yakut, the ancient Aleuts on Mink Island, and the ancient Sadlermiut (Canadian) Thule (Table 2.3).

Two Nuvuk haplotypes (A2b + 16167 and A2b - 16319) were not found in any previously published datasets, nor were they found following a nucleotide Basic Local Alignment Search Tool (BLAST) query. They do, however, serve to highlight the

increased diversity at Nuvuk of haplogroup A2b compared to A2a. Haplotype A2b - 16319 was present in a single individual dated to 1424 AD, but haplotype A2b + 16167 was found in individuals from three burials of similar age: 1245, 1271, and 1249 AD (Table 2.2). This suggests that this haplotype might have arisen in a single matriline and then was subsequently lost at Nuvuk.

### *Chronological and Geographic Appearance of Beringian Haplogroups*

Over the course of the past two decades multiple ancient DNA studies from human remains in Asia and North America have typed individuals carrying HVR-I Beringian-specific haplogroups. Though the data are sparse, they can be used to rule out and/or support certain scenarios of population movement. Using supplied calibrated and marine reservoir corrected direct dates when available, Table 2S.5 lists the time and location for the first appearances of these haplogroups in 109 ancient individuals and Figure 2.3 visualizes the same data. As already revealed by Raghavan et al. (2014), the first appearance of haplogroup D2a is in Paleo-Eskimo individuals from the 3<sup>rd</sup> millennium BCE throughout Greenland and Canada. It is present in the ancient Arctic through late-Dorset individuals dating to the 12<sup>th</sup> century AD, and then in human remains from Mink Island, Alaska in the 15<sup>th</sup> and 16<sup>th</sup> centuries.

There are A2 root haplotypes (A2other in this study) in the subarctic at least as early as the 3<sup>rd</sup> century AD (an eastern Aleutian sample from Port Moller), though since A2 is a pan-American haplogroup we recognize that other ancient DNA studies have expectedly found much older appearances further south. The Beringian haplogroup A2a first appears in Birnirk samples from Chukotka in the 6<sup>th</sup>-7<sup>th</sup> centuries AD, but the oldest appearance of A2b reported in the literature comes from this study. An individual with a

median calibrated and corrected intercept date of 1139 AD, the oldest sampled individual from Nuvuk, was typed for A2b. A similarly aged Nuvuk individual carried A2a (Table 2S.5). This would indicate early A2a and A2b carriers clearly overlapped with middle and late Dorset D2a carriers (Figure 2.3). After the Birnirk and the early Nuvuk Thule samples, the youngest A2a sequences unexpectedly come from 14<sup>th</sup> century Oneota remains at the Norris Farms site in Illinois (Stone and Stoneking, 1998), though the dating method used is unclear. Additionally unclear are a handful of Canadian Thule samples that were typed to haplogroup A, but not A2a or A2b, in the 14<sup>th</sup> and 15<sup>th</sup> centuries AD (Raghavan et al., 2014). The next appearance in the Arctic of A2b and A2a, outside of Nuvuk, is in Silumiut Thule samples from Canada, with median intercept dates of 1402 AD and 1447 AD, respectively. This is followed by appearances in Thule from across Canada and Greenland, as well as the continued occupation at Nuvuk.

A single 5<sup>th</sup> century BCE sample of a Kurgan individual from South Siberia carried a possible D4b1a2a1b haplogroup, though the HVR-I sequence lacks C16173T (Ricaud et al., 2004). If correct, this is the only D4b1a2a1b found in the surveyed ancient DNA literature. The oldest appearance of haplogroup D4b1a2a1a is at Nuvuk, in an individual with a median calibrated and corrected intercept date of 1227 AD. Within a hundred years there are dated and typed individuals with D4b1a2a1a in Canada, Southeast Alaska, and Siberia; The Siberian “protohistoric” Yakut sample, however, was not directly dated and we use a minimum age of 1400 AD based on available information (Crubézy et al., 2010).

## Discussion

Phylogeographic and phylogenetic analyses of mitochondrial DNA control region sequences from ancient human inhabitants at Nuvuk and published datasets from across Asia and North America suggest possible scenarios of population movement from Beringia throughout the Holocene. Successful retrieval of HVR-I sequence from 39 individuals directly dated between 1139 and 1669 AD helps fill in a gap of knowledge about Thule genetic makeup in Alaska between the decline of the Paleo-Eskimo and the later appearance of modern Inuit and Iñupiat. While mtDNA is a single nonrecombining locus that is only inherited through the maternal line, it is more easily accessible for aDNA study than the nuclear genome, it retains a record of more recent population events, and a plethora of population data points are available (Tables 2S.3 and 2S.4). The Nuvuk HVR-I sequences allow hypotheses of Beringian population relationships across time and space, acknowledging that mtDNA is sensitive to genetic drift and sex-biased demographic movements, which might lead to discordant phylogenies with nuclear loci.

A set of mitochondrial DNA variants are at high frequency in Arctic and Subarctic populations of North America and Chukotka: A2a, A2b, D2a, and D4b1a2a1a (Tackney et al., 2016). Additional sister clades of D2b and D4b1a2a1b are not found in Native Americans but are in Asian or Eastern European populations (Table 2S.3), and were investigated here as they represent branches from the same Holocene expansions (Derenko et al., 2010; Dryomov et al., 2015; Volodko et al., 2008). A low frequency of haplogroup A2, a pan-American haplogroup, is also found in northern North America and Chukotka and might represent recent migration from further south or remnant Beringian or Na-Dene haplotypes from the post-LGM expansion of A2 into the Americas

(Dryomov et al., 2015; Raghavan et al., 2015).

The Nuvuk burials were typed with 66.7% A2b, 25.6% A2a, and 7.7% D4b1a2a1a, with no other haplogroups (Table 2.2; Figure 2.1). Principal Component Analysis of the Beringian haplogroups in ancient and modern populations across Asia and the Americas consistently clustered the Nuvuk population with the ancient Thule and modern Inuit of Greenland and Canada, as well as the Siberian Naukan Yupik Eskimo. While the first principal component, explaining 37.1% of the variance, separated out populations with a high proportion of Beringian-specific haplogroups from those with a low proportion, further clustering with Nuvuk was based on the presence of the triplet A2b/A2a/ D4b1a2a1a signature, an absence of D2a, and an absence or very low frequency of pan-American A2 (Figures 2.1 and 2.2). Distinct Beringian populations were the ancient Birnirk in Chukotka and the ancient Thule at Nunalleq in southwest Alaska fixed for A2a, while all known Paleo-Eskimo samples from Greenland and Canada are fixed for D2a (Raghavan et al., 2014). The Siberian Naukan are genetically the closest Asian population to Nuvuk, driven by the fact that the other modern populations of Chukchi, Chuvansti, Sireniki Eskimo, and Chaplinski Yupik Eskimo all have some frequency of D2a (as also noted by Dryomov et al., 2015). The same could also be said for the modern populations on Commander Island (fixed D2a) and the Aleutian Islands (A2a, D2a, pan-American A2) (Table 2S.3). Intriguingly, the modern communities of the North Slope were not the closest population to the Thule at Nuvuk, particularly because the North Slope is far more diverse with low frequencies of D2a, pan-American A2, and pan-American haplogroup C4 in addition to the triplet A2ab/A2a/D4b1a2a1a (Table 2S.3; Raff et al., 2015).

The high frequency of dated samples with haplogroups A2b and D4b1a2a1a at Nuvuk is informative. The Nuvuk Thule population was tied with the Thule from Nunavut, Canada with the second highest percentage of A2b in our dataset (67%; range: 0.3-71.4%, Figure 2.1), and Nuvuk shared the derived A2b+16212 haplotype with Siberian Chukchi, and Inuit further east (Table 2.3; We note here that haplotype comparisons are hampered by the fact that much of the Thule mtDNA data from Canada and Greenland contains genotyping results and not HVR-I sequence - see Materials and Methods). Other derived haplotypes of A2b are found sporadically outside of Nuvuk, but the root haplotype and 16212 haplotype are by far the most prevalent. The oldest evidence for A2b (as A2b root or A2b+16212) clearly comes from Nuvuk, followed by Thule in Canada and then Greenland (Figure 2.3), and the spread of A2b overall is limited; it is restricted to Northern Kamchatka, Chukotka, and across Arctic North America and the Alaskan peninsula (Table 2S.3). This at first suggests an origin for A2b in the Thule population on the North Slope, from a residual Beringian haplogroup A2 source. From full mitochondrial genomes, however, of 42 ancient and modern individuals, the coalescence of the A2b haplogroup (defined by A16265G) is dated to  $2.2 \pm 1.8$  kya, with a population expansion 2 to 1.5 kya (Dryomov et al., 2015). As these genetic dates are too early for Nuvuk or the 13th century AD expansion eastward (Friesen and Arnold 2008), and should the lack of A2b in the Siberian Birnirk hold up with further sampling, the Punuk as an alternative/additional immediate ancestor of the Thule needs to be reconsidered (Mason and Bowers, 2009). Unfortunately no human remains from the Punuk have been genetically characterized.

Nuvuk had an intermediate frequency of haplogroup D4b1a2a1a (7.7%; range:

0.6-28.6%, Figure 2.1; Tables 2S.3 and 2S.4), with the Inuit and ancient Thule of Canada (but not Greenland), and the Naukan of Chukotka containing higher proportions. Along with sister haplogroup D4b1a2a1b, the clade has roots in the Altai-Sayan region of southern Siberia at the Pleistocene-Holocene boundary (Derenko et al., 2010; Volodko et al., 2008), and so was not a part of the initial migrations of Native American ancestors (in contrast to haplogroup A2). The coalescence of D4b1a2a1a and D4b1a2a1b was dated by maximum likelihood methods from 18 whole mitochondrial genomes to 5.39 kya, though with a large confidence interval of 2.25 - 8.72 kya (Dryomov et al., 2015). The earliest evidence for D4b1a2a1a, as it was for A2b, comes from a Nuvuk burial in the 12<sup>th</sup> century AD, though a Thule individual from Cape Silumiut, Canada was similarly aged (Figure 2.3; Table 2S.5). The earliest appearance of D4b1a2a1b is tentatively ascribed to an ancient Kurgan burial around 450 BC (Ricaud et al., 2004). As modern distributions of D4b1a2a1b are limited to low frequencies in South Siberians, Eastern Asians, and Eastern Europeans, the old Kurgan finding in South Siberia is unsurprising. The spread of D4b1a2a1a, however, must be linked to the Beringian expansions, as the HVR-I haplotype defined by T16093C closely follows the distribution of A2b in the Arctic (Table 2S.3).

A complication to this hypothesis is that D4b1a2a1a is also present at low frequencies in populations of South Siberia, Central Asia, and E. Europe (Figure 2.1; Tables S2.3 and 2S.4). The only non-Beringian population with full genome data for D4b1a2a1a is the Tubalar, and these sequences suggest that a subclade, defined by polymorphisms at 11383 and 14122, is the true Beringian-specific haplotype with a much more recent coalescence confidence interval of 0.05 to 3.81 kya (Dryomov et al., 2015).

Confirmation awaits further sequencing of the Asian D4b1a2a1a lineages. The Thule from Canada and Greenland were all typed for the SNP at 11383 (Raghavan et al., 2014), but the Nuvuk burials were not, nor was an Yakut individual dated to sometime around the 14<sup>th</sup> century AD (Table 2S.5; Crubézy et al., 2010). How and when D4b1a2a1a arrived in Beringia cannot be answered by this study, though it is possible that the 11383 and 14122 variant arose in Chukotka (from standing Siberian D4b1a2a1a variation), and then spread east with the Thule in the 13<sup>th</sup> century AD. Another, perhaps separate concern is that the D4b1a2a1a HVR-I sequences from the modern communities of the North Slope lack T16093C (as well as the defining mutations for D4b1a2a1b), so where they fit in the phylogenetic tree is unclear (Raff et al., 2015). Further sequencing of whole ancient mitochondrial genomes will be required to untangle this clade's history.

The absence of the D2 lineage at Nuvuk supports the hypothesis of genetic separation between the earlier Paleo-Eskimo and the later Thule (Raghavan et al., 2014). All Thule samples, as well as earlier Birnik samples, have so far lacked any record of the D2 lineage, while all Paleo-Eskimo samples have been monomorphic for D2 (Table 2S.3). The first appearance of D2 occurs in the Saqqaq of Greenland (Figure 2.3). The genetic history of haplogroup D2a and its sister clade D2b is difficult to determine from this study given that the two clades cannot be distinguished with just HVR-I sequence. Some projects, however, have typed the necessary SNPs in certain populations and educated assumptions can be made for the remainder (Materials and Methods). The distribution of D2a and D2b do not appear to overlap (with the above caveats) in any population, with D2a limited to Chukotka and the Americas, and D2b limited to a low frequency (<2%) in a handful of populations in Sakha, South and West Siberia, Eastern

Europe (Kalmyks) (Table 2S.3). The coalescence date for D2a from 63 whole mitochondrial genomes was estimated at 4.4 kya (3.23 - 5.89 kya), which places it within the range calculated for the D4b1a2a1 clade (Dryomov et al., 2015) and suggests possible geographic and temporal links.

In the Americas, apart from high frequencies in populations in the Aleutians (including the Alaskan Peninsula) and Commander island, D2a (assumed given the typical lack of typing beyond HVR-I) is found at very low frequencies in the North Slope Iñupiat and in the southern Navajo and Yavapai (a non-Athabaskan group but likely admixed with the Navajo) (Table 2S.3; Monroe et al., 2013). D2a also was likely present in the southeastern Alaskan Tlingit until recently (Dryomov et al., 2015), and it was typed in ancient human remains from Mink Island in southeast Alaska (14<sup>th</sup> - 16<sup>th</sup> centuries AD) and Port Moller on the Alaskan Peninsula (undated) (Raff et al., 2015). In Chukotka, D2a is found in all populations except the Naukan (as mentioned earlier), and the variant found in the Sireniki appears the closest to the Paleo-Eskimo Saqqaq (Dryomov et al., 2015). Following the Paleo-Eskimo expansion from Beringia, separate later expansions from residual source populations possibly in southeast Alaska spread D2a into the Aleutians and into the Northern, and then Southern, Athabaskans. These could be linked to the Neo-Eskimo migrations (see below), even if the Neo-Eskimo never carried D2a.

At Nuvuk haplogroup A2a was at intermediate frequency in our dataset (25.6%; range: 0.3-100%, Figure 2.1; Table 2S.3) and less prevalent than in the modern Eskimo-Aleut communities of the North Slope, the Aleutians, Chukotka, and Greenland. The earliest appearance of A2a was in the Siberian Birnirk samples in the 7<sup>th</sup> century AD and

the next oldest was at Nuvuk in the 12<sup>th</sup> century (Figure 2.3). Intriguingly both the single Birnirk sample sequenced (out of 5 typed) and the earliest Nuvuk sample with A2a contained the A2a+16311 haplotype and not the A2a root (Table 2.2; Dryomov et al., 2015). While the whole A2a clade coalesces 3.9 kya (1.80 - 6.10 kya), the various clusters of haplotypes, including the 16311 variant (deemed A2a3) and the Na-Dene specific A2a4 and A2a5, all show much more recent dates of less than two thousand years (Dryomov et al., 2015).

The geographic distribution of A2a suggests an in situ differentiation within Beringia from the pan-America A2 haplogroup (Achilli et al., 2013; Volodko et al., 2008). In Asia, A2a is limited to Chukotka and to low frequencies in the Koryaks, Western Siberian Selkups, and South Siberian Evenks; recent admixture might explain these outliers (Table 2S.3). Ancient DNA evidence proves that A2a did not spread into the Americas with the Paleo-Eskimo expansion, despite its older time depth (in contrast to the hypotheses of Achilli et al., 2013). It was, however, fixed barring further sampling in the ancient Birnirk (Raghavan et al., 2014). At least two, and possibly three, expansions can be hypothesized to have spread A2a. The initial peopling of the Aleutians would be the first, including a demographic event from Paleo-Aleuts to Neo-Aleuts around 1000 AD (reviewed in Tackney et al., 2016). The modern Aleutian populations contain both A2a and A2 (and a majority of D2a), as well as A2b on the peninsula. Dependent both on the geographic location of the ancestral refugium (Chukotka or Alaska) for A2 and on the sequence identity of the Paleo-Aleut (A2a or A2; currently unknown), A2a might have spread into the Aleutians prior to the Neo-Eskimo (along with D2a; see above). Early sites on the Alaskan Peninsula, both at Port Moller in

the 5<sup>th</sup> century AD and Brooks River in the 12<sup>th</sup> century AD had human remains typed to A2 and not A2a (Table 2S.5; Raff et al., 2010). Additionally, the ancient Thule site at Nunalleq, dated to between 1300 and 1660 AD, was monomorphic for haplogroup A2a (Raghavan et al., 2014), though simple genetic drift could explain this population.

The largest expansion of A2a occurred with the Thule in the 13<sup>th</sup> century AD, from which the Birnirk and Nuvuk samples provide the earliest evidence. Not only did this expansion bring A2a (and A2b/ D4b1a2a1a) to the eastern Arctic, but it also possibly instigated the Athabaskan migration to the American Southwest a century later (Seymour, 2012). The Northern Na-Dene Yakutat Tlingit and Dogrib, as well as the southern Athabaskan Navajo and Apache have A2a, mostly in the form of derived haplotypes A2a5 or A2a4 (Table 2S.3). More recent admixture is presumed to have brought these haplotypes to the neighboring tribes of the Shuswap, Zuni, Paiute, and to the South American Vaupes (a noted outlier) (Achilli et al., 2013; Monroe et al., 2013; Tamm et al., 2007). An additional expansion, possibly related to the southward Athabaskan migration, would be required to explain the four A2a HVR-I sequences in the ancient Oneota specimens at the Norris Farms cemetery in Illinois, tentatively dated to 1300 AD (Stone and Stoneking, 1998; Table 2S.5). A2a was typed in single individuals from the Algonquin Cree (Achilli et al., 2013) and Cheyenne (Malhi et al., 2001), and these might represent remnant lineages of this more southeastern spread.

The scenarios presented here are limited foremost by the continued lack of dated, ancient whole mitochondrial and nuclear genomes from Beringia. In particular, human remains from Old Bering Sea, Punuk, and Birnirk (Alaskan) will need to be further sampled to clarify the source populations for the Neo-Eskimo. Earlier, but still

Holocene-dated, human remains from Chukotka are also needed to determine the arrival and differentiation of haplogroups D2a and D4b1a2a1a. In the Americas, modern sampling of Northern Na-Dene speaking populations, apart from the Tlingit, Haida, and Dogrib, is lacking and ancient sampling is completely absent. This prevents us from fully appreciating the amount of gene flow with Alaskan and Canadian Eskimo-Aleut speakers, or from fully understanding the subsequent Southern Athabaskan migration. Finally, the results from the ancient site at Nunalleq, Alaska highlight the lack of genetic data from modern Alaskan Yu'pik communities, who might have different genetic ancestry than the related Inuit and Iñupiat.

The ancient mitochondrial sequences from Nuvuk unambiguously confirm both the link between the North Slope and the Thule who later spread further east, and the discontinuity between the Neo-Eskimo (A2b/A2a/ D4b1a2a1a) and the Paleo-Eskimo (D2a) (Raff et al., 2015; Raghavan et al., 2014). The time depth of both the dated Nuvuk HVR-I sequences and the previously calculated coalescence and expansion times for the Beringian haplogroups hint at possible scenarios for the Holocene Arctic migrations. Haplogroups A2a, D2a and D4b1a2a1a all appear to coalesce 4-5 thousand years ago, prior to the Neo-Eskimo migrations, although it is important to note that these dates have so far not been calibrated with dated aDNA sequences. D2a and D4b1a2a1a likely had recent Holocene ancestry in populations outside of Beringia, while A2a and the later evolving A2b appear to have arisen within Beringia-proper from residual A2 lineages. Within the first wave of Paleo-Eskimo into North America, D2a became fixed in that population, whereas A2a and D4b1a2a1a only rose in frequency or, in the case of A2b, appeared within the ancestors of the Thule on the North Slope or Chukotka. Further

sequencing of ancient mitochondrial and full nuclear genomes from human remains in Siberia and Alaska are necessary to clarify these population migrations beyond the simple two wave scenario of Neo-Eskimo and Paleo-Eskimo.

### **Acknowledgements**

We would like to acknowledge The Barrow Senior Advisory Council, the Ukpeaġvik Iñupiat Corporation, and The Native Village of Barrow for providing permissions to sample from the Nuvuk burials. We would like to also acknowledge Dr. Anne Jensen, the senior Archaeologist at the Nuvuk site. Logistical support while in the field was provided by Glenn Sheehan, Alice Drake, Megan Bowen, Bryan Thomas, Laura Thomas, Lewis Brower, Nok Acker (BASC), Steve Hastings, and Faustine Mercer (CH2Hill Polar Services). This research was supported by an International Polar Year grant: Reconstruction of Human Genetic History Along the North Slope (OPP-0732846 and OPP-0732857), as well as further NSF funding sources: OPP-0637246, OPP-0535356, ARC-0726253, and OPP-0820790.

We would like to thank Caroline Kisielinski for designing Car\_P1F and Car\_P1R, and for running these primers on a subset of the Nuvuk samples to confirm SNPs in that region. We would like to thank Nathan Harris for designing a python script for the genetic distance calculations of Harpending and Jenkins (1973). We have appreciated many fruitful discussions with Dr. Alan Rogers and Dr. Henry Harpending, and we would finally like to thank Dr. Cara Monroe for providing HVR-I sequences from Monroe et al. (2013) and Budowle et al. (2002).

## Literature Cited

- Achilli A, Perego UA, Lancioni H, Olivieri A, Gandini F, Hooshiar Kashani B, Battaglia V, Grugni V, Angerhofer N, Rogers MP et al. 2013. Reconciling migration models to the Americas with the variation of North American native mitogenomes. *Proc Natl Acad Sci* 110(35):14308-14313.
- Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, and Howell N. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23(2):147-147.
- Beechey FWC. 1831. *Narrative of a Voyage to the Pacific and Beering's Strait*. Vol. 1. London: Henry Colburn and Richard Bentley.
- Briggs AW, Stenzel U, Johnson PL, Green RE, Kelso J, Prüfer K, Meyer M, Krause J, Ronan MT, and Lachmann M. 2007. Patterns of damage in genomic DNA sequences from a Neandertal. *Proc Natl Acad Sci* 104(37):14616-14621.
- Budowle B, Allard MW, Fisher CL, Isenberg AR, Monson KL, Stewart JE, Wilson MR, and Miller KW. 2002. HVI and HVII mitochondrial DNA data in Apaches and Navajos. *Int J Legal Med* 116(4):212-215.
- Coltrain J, Tackney J, and O'Rourke DH. Under Review. Thule whaling at Point Barrow, Alaska: The Nuvuk cemetery stable isotope and radiocarbon record. *J Arch Sci Reports*.
- Coltrain JB. 2009. Sealing, whaling and caribou revisited: additional insights from the skeletal isotope chemistry of eastern Arctic foragers. *J Arch Sci* 36(3):764-775.
- Coltrain JB. 2010. Alaska Peninsula stable isotope and radioisotope chemistry: A study in temporal and adaptive diversity. *Hum Biol* 82(5):613-627.
- Coltrain JB, Hayes MG, and O'Rourke DH. 2004. Sealing, whaling and caribou: the skeletal isotope chemistry of Eastern Arctic foragers. *J Arch Sci* 31(1):39-57.
- Crubézy E, Amory S, Keyser C, Bouakaze C, Bodner M, Gibert M, Röck A, Parson W, Alexeev A, and Ludes B. 2010. Human evolution in Siberia: from frozen bodies to ancient DNA. *BMC Evol. Biol.* 10(1):25.
- Derenko M, Malyarchuk B, Grzybowski T, Denisova G, Dambueva I, Perkova M, Dorzhu C, Luzina F, Lee HK, and Vanecek T. 2007. Phylogeographic analysis of mitochondrial DNA in northern Asian populations. *Am J Hum Genet* 81(5):1025-1041.
- Derenko M, Malyarchuk B, Grzybowski T, Denisova G, Rogalla U, Perkova M, Dambueva I, and Zakharov I. 2010. Origin and Post-Glacial Dispersal of Mitochondrial DNA Haplogroups C and D in Northern Asia. *PLoS ONE* 5(12):e15214.

- Dryomov SV, Nazhmidenova AM, Shalaurova SA, Morozov IV, Tabarev AV, Starikovskaya EB, and Sukernik RI. 2015. Mitochondrial genome diversity at the Bering Strait area highlights prehistoric human migrations from Siberia to northern North America. *Eur J Hum Genet* 23(10):1399-404.
- Dulik MC, Zhadanov SI, Osipova LP, Askapuli A, Gau L, Gokcumen O, Rubinstein S, and Schurr TG. 2012. Mitochondrial DNA and Y chromosome variation provides evidence for a recent common ancestry between Native Americans and Indigenous Altaians. *Am J Hum Genet* 90(2):229-246.
- Excoffier L, and Lischer HE. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Res* 10(3):564-567.
- Friesen TM, and Arnold CD. 2008. The Timing of the Thule Migration: New Dates from the Western Canadian Arctic. *Am Antiquity* 73(3):527-538.
- Gilbert MTP, Djurhuus D, Melchior L, Lynnerup N, Worobey M, Wilson AS, Andreasen C, and Dissing J. 2007. mtDNA from hair and nail clarifies the genetic relationship of the 15th century Qilakitsoq Inuit mummies. *Am J Phys Anthropol* 133(2):847-853.
- Gilbert MTP, Kivisild T, Grønnow B, Andersen PK, Metspalu E, Reidla M, Tamm E, Axelsson E, Götherström A, Campos PF et al. 2008. Paleo-Eskimo mtDNA Genome Reveals Matrilineal Discontinuity in Greenland. *Science* 320(5884):1787-1789.
- Goebel T, Waters MR, and O'Rourke DH. 2008. The Late Pleistocene Dispersal of Modern Humans in the Americas. *Science* 319(5869):1497-1502.
- Harpending H, and Jenkins T. 1973. Genetic distance among Southern African populations. In: Crawford M, and Workman PL, editors. *Methods and theories of anthropological genetics*. Albuquerque, NM: University of New Mexico Press. p 177-199.
- Harritt RK. 1998. Paleo-eskimo beginnings in North America: a new discovery at Kuzitrin lake, Alaska. *Études/Inuit/Studies*:61-81.
- Helgason A, Pálsson G, Pedersen HS, Angulalik E, Gunnarsdóttir ED, Yngvadóttir B, and Stefánsson K. 2006. mtDNA variation in Inuit populations of Greenland and Canada: migration history and population structure. *Am J Phys Anthropol* 130(1):123-134.
- Hoffecker JF, Elias SA, and O'Rourke DH. 2014. Out of Beringia? *Science* 343(6174):979-980.
- Jensen A. 2007. Nuvuk Burial: An Early Thule Hunter of High Status. *Alaska J Anthropol* 5(1):119-122.

- Jensen AM. 2009a. Nuvuk, Point Barrow, Alaska: The Thule Cemetery and Ipiutak Occupation [PhD Dissertation]. Bryn Mawr, PA: Bryn Mawr College.
- Jensen AM. 2009b. Radiocarbon dates from recent excavations at Point Barrow, Alaska and their implications for Neoeskimo prehistory. In: Gronnow B, editor. *On the track of the Thule culture, from Bering Strait to east Greenland, Proceedings of the SILA Conference "The Thule Culture - New Perspectives in Inuit Prehistory"*. Copenhagen: Publications from the National Museum Studies in Archaeology and History Vol. 15. p 45–62.
- Jensen AM. 2012. Culture and change: learning from the past through Community Archaeology on the North Slope. *Polar Geography* 35(3-4):211-227.
- Jensen AM. 2014. The archaeology of north Alaska: Point Hope in context. In: Hilton CE, Auerbach BM, and Cowgill LW, editors. *The Forages of Point Hope: The Biology and Archaeology of Humans on the Edge of the Alaskan Arctic*. Cambridge: Cambridge University Press. p 11-34.
- Keyser C, Bouakaze C, Crubézy E, Nikolaev VG, Montagnon D, Reis T, and Ludes B. 2009. Ancient DNA provides new insights into the history of south Siberian Kurgan people. *Hum Genet* 126(3):395-410.
- Koch A. 1996. In: Grønnow B, and Pind J, editors. *The Paleo-Eskimo Cultures of Greenland: New Perspectives in Greenlandic Archaeology*. Copenhagen: Danish Polar Center. p 35-38.
- Larsen H, and Rainey F. 1948. Ipiutak and the Arctic Whale Hunting Culture. *Anthropological Papers of the American Museum of Natural History*. New York.
- Maguire RC. 1988. *The Journal of Rochfort Maguire, 1852-1854: Two Years at Point Barrow, Alaska, Aboard HMS Plover in the search for Sir John Franklin*. London: The Hakluyt Society.
- Malhi RS, Schultz BA, and Smith DG. 2001. Distribution of mitochondrial DNA lineages among Native American tribes of Northeastern North America. *Hum Biol* 73(1):17-55.
- Mason OK. 2006. Ipiutak Remains Mysterious: A Focal Place Still Out of Focus. In: Arneborg J, and Grønnow B, editors. *Dynamics of Northern Societies: Proceedings of the SILA/NABO Conference on Arctic and North Atlantic Archaeology*. Copenhagen: National Museum of Denmark. p 103-119.
- Mason OK. 2009a. Flight from Bering Strait: Did Siberian Punuk/Thule military cadres conquer Northwest Alaska? In: Maschner H, Mason OK, and McGhee R, editors. *The Northern World AD 900-1400*. Salt Lake City, UT: University of Utah Press. p 76-130.
- Mason OK. 2009b. "The Multiplication of Forms": Bering Strait Harpoon Heads as a

- Demic and Macroevolutionary Proxy. In: Prentiss AM, Kuijt I, and Chatters JC, editors. *Macroevolution in Human Prehistory*. New York: Springer. p 73-107.
- Mason OK, and Bowers PM. 2009. The Origin of Thule is Always Elsewhere: Early Thule within Kotzebue Sound, Cul de sac or nursery. In: Gronnow B, editor. *On the track of the Thule culture, from Bering Strait to east Greenland, Proceedings of the SILA Conference "The Thule Culture - New Perspectives in Inuit Prehistory"*. Copenhagen: Publications from the National Museum Studies in Archaeology and History Vol. 15. p 25-44.
- Mathiassen T. 1927. *Archaeology of the Central Eskimos. Report of the Fifth Thule Expedition 1921-24, Vol 6*. Copenhagen: Gyldendalske.
- McGhee R. 2000. Radiocarbon dating and the timing of the Thule migration. In: Appelt M, Berglund J, and Gulløv HC, editors. *Identities and cultural contacts in the Arctic*. Copenhagen: Danish National Museum and Danish Polar Center. p 181-191.
- Mills R, Gerlach SC, Bowers PM, and MacIntosh SJ. 1999. *Final Report to the Cultural Resources Mitigation of the 1998 Anaktuvuk Pass Runway Realignment Project*. NLUR: LCMF, Inc.
- Monroe C, Kemp BM, and Smith DG. 2013. Exploring prehistory in the North American southwest with mitochondrial DNA diversity exhibited by Yumans and Athapaskans. *Am J Phys Anthropol* 150(4):618-631.
- Pakendorf B, Novgorodov IN, Osakovskij VL, Al'bina PD, Protod'jakonov AP, and Stoneking M. 2006. Investigating the effects of prehistoric migrations in Siberia: genetic variation and the origins of Yakuts. *Hum Genet* 120(3):334-353.
- Raff J, Tackney J, and O'Rourke DH. 2010. South from Alaska: a pilot aDNA study of genetic history on the Alaska Peninsula and the Eastern Aleutians. *Hum Biol* 82(5/6):677-693.
- Raff JA, Rzhetskaya M, Tackney J, and Hayes MG. 2015. Mitochondrial diversity of Iñupiat people from the Alaskan North Slope provides evidence for the origins of the Paleo-and Neo-Eskimo peoples. *Am J Phys Anthropol* 157(4):603-14.
- Raghavan M, DeGiorgio M, Albrechtsen A, Moltke I, Skoglund P, Korneliussen TS, Grønnow B, Appelt M, Gulløv HC, Friesen TM et al. 2014. The genetic prehistory of the New World Arctic. *Science* 345(6200):1255832.
- Raghavan M, Steinrücken M, Harris K, Schiffels S, Rasmussen S, DeGiorgio M, Albrechtsen A, Valdiosera C, Ávila-Arcos MC, Malaspina A-S et al. 2015. Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* 349(6250):aab3884.
- Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen A, Moltke I, Metspalu M,

- Metspalu E, Kivisild T, and Gupta R. 2010. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* 463(7282):757-762.
- Ray PHL. 1885. International Polar Expedition to Point Barrow, Alaska. Senate Committee on Military Affairs. Washington: Government Printing Office.
- Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, Ray N, Parra MV, Rojas W, Duque C, and Mesa N. 2012. Reconstructing native American population history. *Nature* 488(7411):370-374.
- Ricaux FX, Keyser-Tracqui C, Cammaert L, Crubézy E, and Ludes B. 2004. Genetic analysis and ethnic affinities from two Scytho-Siberian skeletons. *Am J Phys Anthropol* 123(4):351-360.
- Rohland N, and Hofreiter M. 2007a. Ancient DNA extraction from bones and teeth. *Nat Prot* 2(7):1756-1762.
- Rohland N, and Hofreiter M. 2007b. Comparison and optimization of ancient DNA extraction. *Biotechniques* 42(3):343.
- Rubicz R, Schurr TG, Babb PL, and Crawford MH. 2003. Mitochondrial DNA variation and the origins of the Aleuts. *Hum Biol* 75(6):809-35.
- Saillard J, Forster P, Lynnerup N, Bandelt H-J, and Nørby S. 2000. mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *Am J Hum Genet* 67(3):718-726.
- Schurr TG, Dulik MC, Owings AC, Zhadanov SI, Gaieski JB, Vilar MG, Ramos J, Moss MB, and Natkong F. 2012. Clan, language, and migration history has shaped genetic diversity in Haida and Tlingit populations from Southeast Alaska. *Am J Phys Anthropol* 148(3):422-435.
- Schurr TG, Sukernik RI, Starikovskaya YB, and Wallace DC. 1999. Mitochondrial DNA variation in Koryaks and Itel'men: population replacement in the Okhotsk Sea-Bering Sea region during the Neolithic. *Am J Phys Anthropol* 108(1):1-39.
- Seymour DJ. 2012. Gateways for Athabascan migration to the American Southwest. *Plains Anthropologist* 57(222):149-161.
- Shields GF, Schmiechen AM, Frazier BL, Redd A, Voevoda MI, Reed JK, and Ward RH. 1993. mtDNA sequences suggest a recent evolutionary divergence for Beringian and northern North American populations. *Am J Hum Genet* 53(3):549-562.
- Stone AC, and Stoneking M. 1998. mtDNA analysis of a prehistoric Oneota population: implications for the peopling of the New World. *Am J Hum Genet* 62(5):1153-1170.
- Sukernik RI, Volodko NV, Mazunin IO, Eltsov NP, Dryomov SV, and Starikovskaya EB.

2012. Mitochondrial genome diversity in the Tubalar, Even, and Ulchi: contribution to prehistory of native Siberians and their affinities to Native Americans. *Am J Phys Anthropol* 148(1):123-138.
- Tackney J, Coltrain JB, Raff J, and O'Rourke DH. 2016. Ancient DNA and Stable Isotopes: Windows on Arctic Prehistory. In: Friesen TM, and Mason OK, editors. *The Oxford Handbook of the Prehistoric Arctic*. New York: Oxford University Press.
- Tamm E, Kivisild T, Reidla M, Metspalu M, Smith DG, Mulligan CJ, Bravi CM, Rickards O, Martinez-Labarga C, and Khusnutdinova EK. 2007. Beringian standstill and spread of Native American founders. *PLoS ONE* 2(9):e829.
- Torrioni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, Vullo CM, and Wallace DC. 1993. Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am J Hum Genet* 53(3):563.
- Volodko NV, Starikovskaya EB, Mazunin IO, Eltsov NP, Naidenko PV, Wallace DC, and Sukernik RI. 2008. Mitochondrial Genome Diversity in Arctic Siberians, with Particular Reference to the Evolutionary History of Beringia and Pleistocenic Peopling of the Americas. *Am J Hum Genet* 82(5):1084-1100.
- Wang S, Lewis Jr CM, Jakobsson M, Ramachandran S, Ray N, Bedoya G, Rojas W, Parra MV, Molina JA, and Gallo C. 2007. Genetic variation and population structure in Native Americans. *PLoS Genetics* 3(11):e185.
- Zlojutro M, Rubicz R, and Crawford MH. 2009. Mitochondrial DNA and Y-chromosome variation in five eastern Aleut communities: Evidence for genetic substructure in the Aleut population. *Ann Hum Biol* 36(5):511-526.

Table 2.1 : Summary statistics of Nuvuk aDNA samples

Statistic	Value	S.D.
# of Samples (N)	39	
# of Haplotypes (K)	8	
# of Segregating Sites (S)	11	
Length of Sequence Coverage (basepairs)	360	
Haplotype Diversity (h)	0.7584	0.0612
Nucleotide Diversity ( $\pi$ )	0.00596	0.00376
Mean Pairwise Differences ( $\Pi$ )	2.1457	1.2183

Table 2.2: Nuvuk sample HVR-I polymorphisms and dates

Burial Sample Identifier	rCRS:	HVR-I Polymorphisms												Haplogroup	Haplotype	Duplicated?	ACRF No.	AMS Dates from Coltrain et al. Under Review	
		16093	16111	16167	16173	16192	16212	16223	16261	16265	16290	16311	16319					16362	AMS Cal 2 $\sigma$ Range AD
06A-185-XF2FS			T				G	T	G	T		A	C	A2b	A2b + 16212	Y	2238	1005-1299	1159
06A-213-QLYXH			T				T	T	G	T		C	C	A2b	A2b - 16319	Y	2240	1290-1584	1424
06A-111-18/19				T				T	G	T		A	C	A2b	A2b + 16167	Y	2236	1059-1402	1245
06B-113-MECCT			T					T	G	T		A	C	A2b	A2b	Y	2733	981-1341	1172
06C-116-FIOYJ			T				G	T	G	T		A	C	A2b	A2b + 16212	Y	2244	1001-1280	1139
07a-297-17/18			T					T	G	T		A	C	A2b	A2b	Y	2246	1319-1657	1484
07-HO-?			T			T		T		T	C	A	C	A2a	A2a + 16311	Y	2259	1115-1423	1273
07-C24-42/43		C			T			T		T		A	C	D4b1a2a1a	D4b1A2a1a	Y	2251	1046-1393	1227
07-A298-68			T					T	G	T		A	C	A2b	A2b	Y	2247	1055-1403	1242
07-D45-10			T			T		T		T		A	C	A2a	A2a	Y	2252	1123-1450	1303
07-B35-23			T			T		T		T	C	A	C	A2a	A2a + 16311	Y	2249	1037-1395	1220
07-EE67-XXBQT			T					T	G	T		A	C	A2b	A2b	Y	2254	1044-1411	1237
07-H39-38			T				G	T	G	T		A	C	A2b	A2b + 16212	Y	2260	1001-1283	1142
07-JJ22-YUCXV		C			T			T		T		A	C	D4b1a2a1a	D4b1A2a1a	Y	2261	1060-1432	1269
07A-513-OPZW4			T					T		T	C	A	C	A2a	A2a + 16311	Y	2248	1019-1274	1141
07-054-KLCXP			T			T		T		T		A	C	A2a	A2a + 16261	Y	2245	1067-1403	1251
07-LL28-WBGTR			T			T		T		T	C	A	C	A2a	A2a + 16311	Y	2262	1056-1407	1246
07-FF36-QZRSP			T					T	G	T		A	C	A2b	A2b	Y	2255	981-1316	1157
07-DD56-MQVRR			T			T		T		T		A	C	A2a	A2a	Y	2253	1414-1683	1547
08-625-27-LAMMK			T			T		T		T	C	A	C	A2a	A2a + 16311	N	2264	1146-1452	1308
09-EE55-111			T					T	G	T		A	C	A2b	A2b	N	2266	1104-1499	1328
08-D61-OCZPT			T					T	G	T		A	C	A2b	A2b	Y	2566	1153-1518	1346
08-D61-QSTL4			T					T	G	T		A	C	A2b	A2b	Y	2562	1072-1432	1275
08-099-71-NCVNO			T					T	G	T		A	C	A2b	A2b	Y	2564	1170-1486	1343
08-I50-MDBYZ			T					T	G	T		A	C	A2b	A2b	Y	2563	1146-1467	1319
08-K4-ZITFB			T					T	G	T		A	C	A2b	A2b	Y	2565	1116-1418	1269
08A-604-21-CTLAC		C			T			T		T		A	C	D4b1a2a1a	D4b1A2a1a	N		Not Dated	
09C-63-228-RVSN6			T					T	G	T		A	C	A2b	A2b	N	2555	1090-1458	1303
09-B96-FIRSQ			T					T	G	T		A	C	A2b	A2b	N	2557	1169-1412	1290
08-E87-KAMOR			T					T	G	T		A	C	A2b	A2b	Y	2554	1165-1533	1360
09-063-143/144-LBFKW/PXVG6			T					T	G	T		A	C	A2b	A2b	N	2570	1161-1523	1352
10-C78-FTTSD			T	T				T	G	T		A	C	A2b	A2b + 16167	Y	2568	1042-1455	1271
10A-928-BHASH			T				G	T	G	T		A	C	A2b	A2b + 16212	Y	2567	1115-1428	1277
10-175-ZQNPO			T	T				T	G	T		A	C	A2b	A2b + 16167	Y	2572	1028-1449	1249
11A-994-OPXSR			T					T	G	T		A	C	A2b	A2b	Y	2561	1036-1388	1213
B11A-949-38-FCYRR			T			T		T		T		A	C	A2a	A2a + 16261	Y	2560	1481-1885	1669
2011-38-JZHTC			T					T	G	T		A	C	A2b	A2b	Y	2558	1069-1427	1270
10A-927-EVARB			T			T		T		T		A	C	A2a	A2a	Y	2574	1039-1338	1217
B10-A918-WTIKR			T					T	G	T		A	C	A2b	A2b	Y	2575	1054-1415	1250
08-C32-108-VUNFJ									Sample Failed							Y	2263	1400-1708	1557
07-C-50-PMIXX									Sample Failed							Y	2250	1266-1519	1385
07-G94-20									Sample Failed							N	2258	1289-1626	1429
08-D15-THR4D									Sample Failed							Y	2556	1323-1699	1527
09C-25-206-TXN98									Sample Failed							N	2569	1126-1452	1306

Table 2.3: Population sharing of Nuvuk haplotypes

Nuvuk Haplotype	N (% in Nuvuk)	Present in Populations:
A2b	18 (46.2%)	<b>Chukchi, Naukan Eskimo, Chuvantsi, Sireniki Eskimo, Chaplin Eskimo</b> (Volodko et al. 2008, Dryomov et al. 2015), <b>Alaskan North Slope Inupiat</b> (Raff et al. 2015), <b>Northwest Territory Inuit, Kitikmeot Inuit, N. NE. E. S. W. NW. Greenlandic Inuit</b> (Saillard et al. 2000, Helgason et al. 2006, Schurr et al. 2012), <b>Ancient Sadlermiut Thule</b> (Raghavan et al. 2014, Dryomov 2015), <b>Ancient W. Greenlandic Thule</b> (Gilbert et al. 2007)
A2b + 16212	4 (10.3%)	<b>Chukchi</b> (Volodko et al. 2008, Dryomov et al. 2015), <b>Kitikmeot Inuit</b> (Helgason et al. 2006), <b>N. E. S. W. NW. Greenlandic Inuit</b> (Saillard et al. 2000, Helgason et al. 2006)
A2b + 16167	3 (7.7%)	Not previously reported
A2b - 16319	1 (2.6%)	Not previously reported
A2a	3 (7.7%)	<b>Chukchi, Naukan Eskimo, Chuvantsi, Sireniki Eskimo, Chaplin Eskimo</b> (Derenko et al. 2007, Volodko et al. 2008, Dryomov et al. 2015), <b>Koryaks</b> (Schurr et al. 1999), <b>Alaskan North Slope Inupiat</b> (Raff et al. 2015), <b>Aleutian Islanders</b> (Rubicz et al. 2003), <b>Dogrib</b> (Achilli et al. 2013), <b>N. E. S. W. NW. Greenlandic Inuit</b> (Saillard et al. 2000, Helgason et al. 2006), <b>Cheyenne/Arapaho</b> (Malhi et al. 2001), <b>Ancient W. Greenlandic Thule</b> (Gilbert et al. 2007), <b>Ancient Oneota</b> (Stone and Stoneking 1998)
A2a + 16311	5 (12.8%)	<b>Chukchi, Naukan Eskimo, Chuvantsi</b> (Volodko et al. 2008, Dryomov et al. 2015), <b>Aleutian Islanders</b> (Rubicz et al. 2003), <b>Alaskan North Slope Inupiat</b> (Raff et al. 2015), <b>Kitikmeot Inuit, NE. E. S. W. NW. Greenlandic Inuit</b> (Saillard et al. 2000, Helgason et al. 2006), <b>Ancient Birmirk</b> (Raghavan et al. 2014, Dryomov et al. 2015), <b>Ancient W. Greenlandic Thule</b> (Gilbert et al. 2007)
A2a + 16261	2 (5.1%)	<b>Chukchi</b> (Volodko et al. 2008, Dryomov et al. 2015), <b>Selkup, Evenks</b> (Tamm et al. 2007, Pakendorf et al. 2006), <b>Alaskan North Slope Inupiat</b> (Raff et al. 2015), <b>Aleutian Islanders</b> (Rubicz et al. 2003), <b>Northwest Territory Inuit</b> (Helgason et al. 2006), <b>S. W. NW. Greenlandic Inuit</b> (Saillard et al. 2000, Helgason et al. 2006)
D4b1a2a1a	3 (7.7%)	<b>Chukchi, Naukan Eskimo</b> (Tamm et al. 2007, Volodko et al. 2008, Dryomov et al. 2015), <b>Koryaks</b> (Schurr et al. 1999), <b>Kalmyks, Bashkirs, Karakalpaks</b> (Derenko et al. 2007, 2010), <b>Kitikmeot Inuit, S. W. NW. Greenlandic Inuit</b> (Helgason et al. 2006), <b>Ancient Yakut</b> (Crubezy et al. 2010), <b>Ancient Alaskan Mink Island</b> (Raff et al. 2010), <b>Ancient Sadlermiut Thule</b> (Raghavan et al. 2014, Dryomov et al. 2015)

Table 2S.1: Oligonucleotides

Primer ID	Sequence (5'->3')	Notes
HVR1_P1F	GTT CTT TCA TGG GGA AGC AG	Previously published in (Raff et al., 2010)
HVR1_P1Rc	TTG ATG TGG ATT GGG TTT TT	“ “
HVR1_P2Fb	AAA ACC CAA TCC ACA TCA AA	“ “
HVR1_P2R	GGG TGG GTA GGT TTG TTG G	“ “
HVR1_P3F	CCC ACT AGG ATA CCA ACA AAC C	“ “
HVR1_P3R	ATT GAT TTC ACG GAG GAT GG	“ “
HVRI_P4F	CCA GCC ACC ATG AAT ATT GT	
HVRI_P4R	GGG ATT TGA CTG TAA TGT GC	
Car_P1F	CGG TAC CAT AAA TAC TTG AC	
Car_P1R	GAT AGT TGA GGG TTG ATT G	

Table 2S.2: Minimum Beringian HVR-I motifs for population selection

Haplogroup	rCRS:	HVRI Polymorphisms												
		16093	16111	16129	16173	16192	16223	16233	16265	16271	16290	16319	16331	16362
A2a		T	C	G	C	C	C	A	A	T	C	G	G	T
A2a4	C	T				T	T <sup>α</sup>				T <sup>α</sup>	A <sup>α</sup>		C <sup>α</sup>
A2a5		T				T <sup>α</sup>	T	G <sup>α</sup>			T	A	A	C
A2b		T					T <sup>α</sup>		G		T <sup>α</sup>	A <sup>α</sup>		C <sup>α</sup>
D2a				A <sup>α</sup>			T <sup>α</sup>			C <sup>α</sup>				C
D2b				A <sup>α</sup>			T <sup>α</sup>			C <sup>α</sup>				C
D4b1a2a1a	C <sup>α</sup>				T		T					A <sup>α</sup>		C
D4b1a2a1b			A	T			T					A <sup>α</sup>		C
A2other*		T					T				T	A		C

\*Only reported when the population contains an additional Beringian haplogroup or the population is ancient and northern

α: The haplotype was recorded even if this SNP was absent, as long as all remaining SNPs were present

Table 2S.3: Beringian haplogroup frequencies across populations

Geographic Location	Population	Date Range	A2a (not A2a4 or A2a5)											N	Notes	References
			A2a4	A2a5	A2b	A2other	D2a	D2b	D4b1a2a1a	D4b1a2a1b	Non-Beringian					
Central Asia	Karakalpak - Kirghiz		0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	99.0	203		Derenko et al. 2010
Central Asia	Kazakh		0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	99.8	406		Tamm et al. 2007	
Central Asia	Uighurs		0.0	0.0	0.0	0.0	0.0	0.0	1.6	0.8	0.0	97.6	122		Derenko et al. 2010	
(N. Altain) Altai-Sayan (S. Siberia)	Shors		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.6	0.0	99.4	252		Tamm et al. 2007; Derenko et al. 2010	
(N. Altain) Altai-Sayan (S. Siberia)	Chelkans, Kumandins		0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.1	0.0	97.9	143		Dulik et al. 2012	
(N. Altain) Altai-Sayan (S. Siberia)	Tubalar		0.0	0.0	0.0	0.0	0.0	0.0	0.0	7.7	0.0	92.3	215		Dulik et al. 2012; Sukernik et al. 2012	
(S. Altain) Altai-Sayan (S. Siberia)	Altai-kizhi		0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.5	0.0	98.6	725		Derenko et al. 2010; Dulik et al. 2012	
S. Siberia	Khakassians		0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.6	0.0	98.4	185		Tamm et al. 2007	
S. Siberia	Buryats		0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.0	2.3	96.8	621		Derenko et al. 2010	
S. Siberia / E.Asia	Khantigians		0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	3.0	96.0	99		Derenko et al. 2007	
Tuva (S. Siberia)	Tuvians		0.0	0.0	0.0	0.0	0.0	0.0	2.9	0.0	0.0	97.1	307		Derenko et al. 2007; Tamm et al. 2007	
S. Siberia / E. Asia	East Evenks		0.3	0.0	0.0	0.0	0.0	0.0	1.7	0.0	0.0	98.0	369		Derenko et al. 2007; Tamm et al. 2007	
E. Europe	Kalmyks		0.0	0.0	0.0	0.0	0.0	0.0	1.8	0.9	0.0	97.3	110		Derenko et al. 2007	
Magadan Oblast / Sakha	Evens		0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	99.0	251		Tamm et al. 2007; Rubicz et al. 2010; Sukernik et al. 2012	
W. Siberia	Selkup		1.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	98.3	120		Tamm et al. 2007	
E. Asia	Mongolians		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.1	97.9	47		Derenko et al. 2007	
Sakha	Yakuts		0.0	0.0	0.0	0.0	0.0	1.8	0.0	0.0	0.0	98.2	459		Derenko et al. 2007; Tamm et al. 2007	
W. Siberia	Dolgan		0.0	0.0	0.0	0.0	0.0	1.3	0.0	0.0	0.0	98.7	157		Tamm et al. 2007	
Chukotka	Chukchi		33.5	0.0	0.0	13.9	1.3	8.2	0.0	3.8	0.0	39.3	182		Volodko et al. 2008	
N. Kamchatka	Koryaks		1.2	0.0	0.0	0.3	0.0	0.0	0.0	1.3	0.0	97.2	232		Schurr et al. 1999; Derenko et al. 2007; Tamm et al. 2007	
Chukotka	Yukaghir: Chuvantsi		18.8	0.0	0.0	6.3	0.0	3.1	0.0	6.3	0.0	65.5	32	D2a Assumed	Volodko et al. 2008	
Chukotka	Sireniki Eskimo		43.2	0.0	0.0	27.0	0.0	29.7	0.0	0.0	0.0	0.1	37	D2a assumed	Volodko et al. 2008	
Chukotka	Chaplin (Yuit/Yupik )		69.4	0.0	0.0	15.9	0.0	9.9	0.0	2.0	0.0	2.8	101		Tamm et al. 2007; Volodko et al. 2008	
Chukotka	Naukan (Yuit/Yupik )		33.4	0.0	0.0	41.0	0.0	0.0	0.0	25.6	0.0	0.0	39		Volodko et al. 2008	
Bering Sea	Commander Islands		0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	71	D2a assumed	Volodko et al. 2008; Zlojutro et al. 2009	
Bering Sea	Aleutian Island Aleuts		30.9	0.0	0.0	0.0	1.1	67.4	0.0	0.0	0.0	0.6	181	D2a assumed	Zlojutro et al. 2009	
Alaskan Peninsula	Alaskan Peninsula		30.6	0.0	0.0	2.0	11.2	43.9	0.0	1.0	0.0	11.3	98	D2a assumed	Zlojutro et al. 2009	
SE. Alaska	Aleuts Hoonah Tlingit		0.0	0.0	0.0	2.4	92.7	0.0	0.0	0.0	0.0	4.9	41	Inuit A2b	Schurr et al. 2012	
SE. Alaska	Yakutat Tlingit		4.8	0.0	42.9	0.0	28.6	4.8	0.0	0.0	0.0	18.9	21	Inupiat Ancestry (A2a), Athabaskan Ancestry (D2), D2a assumed	Schurr et al. 2012	
SE. Alaska	Hidalburg Haida		4.5	0.0	0.0	0.0	81.8	0.0	0.0	0.0	0.0	13.7	22	Tsimshian Ancestry (A2a)	Schurr et al. 2012	
N. Alaska	North Slope Inupiat		54.0	0.0	0.0	35.8	0.7	2.9	0.0	5.1	0.0	1.5	137	D2a assumed	Raff et al. 2015	
Northwest Territories, Canada	Dogrib		4.3	0.0	22.5	0.0	0.0	0.0	0.0	0.0	0.0	73.2	53		Tamm et al. 2007; Achilli et al. 2013	
Northwest Territories, Canada	Inuit		25.0	0.0	0.0	50.0	25.0	0.0	0.0	0.0	0.0	0.0	4		Helgason et al. 2006	
Nunavut, Canada	Inuit		13.3	0.0	0.0	63.3	10.0	0.0	0.0	13.3	0.0	0.0	90		Helgason et al. 2006	
Greenland	Inuit		44.6	0.0	0.0	51.3	0.8	0.0	0.0	3.3	0.0	0.0	392		Saillard et al. 2000; Helgason et al. 2006; Gilbert et al. 2008	
Southwestern USA	Apache		0.0	0.5	43.9	0.0	6.5	0.0	0.0	0.0	0.0	49.1	238		Budowle et al. 2002; Tamm et al. 2007; Achilli et al. 2013	
Southwestern USA	Western Apache		0.0	0.0	26.8	0.0	28.0	0.0	0.0	0.0	0.0	45.2	82		Monroe et al. 2013	
Southwestern USA	Navajo		1.4	3.8	10.0	0.0	26.0	1.4	0.0	0.0	0.0	57.4	210	D2a assumed	Budowle et al. 2002; Achilli et al. 2013	
Southwestern USA	Yavapai (Amerind-Yuman)		0.0	0.0	0.0	0.0	33.1	2.4	0.0	0.0	0.0	64.5	127	D2a assumed	Monroe et al. 2013	
Amerind	Northern Paiute, Zuni, Vaupes		1.9	1.9	3.8	0.0	0.0	0.0	0.0	0.0	0.0	92.4	53		Tamm et al. 2007	
Nunavut, Canada / Greenland	All Paleo-Eskimo	3000 BC - 1350 AD	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	22	One D4c Pre-Dorset sample included here as D2a	Raghavan et al. 2014	
Alaskan Peninsula	Hot Springs Site (Port Moller)	263 - 562 AD	0.0	0.0	0.0	0.0	25.0	25.0	0.0	0.0	0.0	50.0	4	D2a assumed	Raff et al. 2010	
Chukotka	Birnirk Brooks River	570 - 680 AD	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5		Raghavan et al. 2014	
Alaskan Peninsula	Nunalleq Thule	1039 - 1264 AD	0.0	0.0	0.0	0.0	50.0	0.0	0.0	0.0	0.0	50.0	8		Raff et al. 2010	
SW. Alaska	Mink Island (Prince William Sound)	1300 - 1660 AD	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	28		Raghavan et al. 2014	
SE. Alaska	Thule	1284 - 1658 AD	0.0	0.0	0.0	0.0	16.7	66.7	0.0	16.7	0.0	0.0	6	D2a assumed	Raff et al. 2010	
Nunavut, Canada	Thule	1210 - 1770 AD	4.2	0.0	0.0	66.7	20.8	0.0	0.0	8.3	0.0	0.0	24	"A" samples are listed here as A2other	Raghavan et al. 2014	
NE. Canada	Thule	1410 - 1900 AD	0.0	0.0	0.0	71.4	0.0	0.0	0.0	28.6	0.0	0.0	7		Raghavan et al. 2014	
Greenland	Thule	1300 - 1800 AD	57.1	0.0	0.0	42.9	0.0	0.0	0.0	0.0	0.0	0.0	28		Gilbert et al. 2007; Raghavan et al. 2014	
Sakha	Yakuts	1400 - 1900 AD	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.7	0.0	98.3	60		Crubczy et al. 2010	
N. Alaska	Nuyuk	1139 - 1669 AD	25.6	0.0	0.0	66.7	0.0	0.0	0.0	7.7	0.0	0.0	39		This study	

Table 2S.4: Excluded population frequency data

Geographic Location	Population	Date Range	A2a (not A2a4 or A2a5)	A2a4	A2a5	A2a5	A2b	A2other	D2a	D2b	D4b1a2a1a	D4b1a2a1b	N	References
E. Europe	Bashkirs		X	X	X	X	X	X	X	X	1.0	1.0	207	Derenko et al. 2010
E. Asia	Barghuts		X	X	X	X	X	X	X	X	3.4	3.4	149	Derenko et al. 2010
E. Europe	Udmurts		X	X	X	X	X	X	X	X	10.1	10.1	189	Derenko et al. 2010
E. Europe	Russian, Maris, Tatars, Belorussians, Poles		X	X	X	X	X	X	X	X	0.9	0.9	1000	Derenko et al. 2010
Khabarovsk Krai (Siberia / E.Asia)	Ulchi		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	160	Sukernik et al. 2012
S. Kamchatka	Itel'men		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	47	Schurr et al. 1999
Wyoming/Colorado	Cheyenne/Arapaho (Algonquian)		2.9	0.0	0.0	0.0	5.7	X	X	X	X	X	35	Achilli et al. 2013
Taimyr peninsula	Nganasan		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	146	Volodko et al. 2008, Tamm et al. 2007
Chukotka / Sakha	Yukaghir: Lower Kolyma		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	82	Volodko et al. 2008
Chukotka / Sakha	Yukaghir: Upper Kolyma		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	40	Volodko et al. 2008, Tamm et al. 2007
Greenland	Ancient Greenlandic Norse	920 - 1410 AD	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	35	Raghavan et al. 2014
Krasnoyarsk Krai (C. Siberia)	Ancient Kurgan	1800 BC - 400 AD	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	26	Keyser et al. 2009

X: No data available

Table 2S.5: Chronological appearance of Beringian haplogroups

Haplogroup	Site	Culture	Location	2 sigma range (AD)	Date <sup>a</sup>	Notes	Reference
D2a1	Qeqertasussuk	Saqqaq	NW Greenland	3340-2610 BCE	2975 BCE	β, δ	Raghavan et al. 2014, Koch 1996
D2a1	Qeqertasussuk	Saqqaq	NW Greenland	3040-2470 BCE	2755 BCE	β, δ	Raghavan et al. 2014, Koch 1996
D2a1	Qeqertasussuk	Saqqaq	NW Greenland	2264-1961 BCE	2113 BCE	β	Raghavan et al. 2014
D2a	Qeqertasussuk	Saqqaq	NW Greenland	2195-1922 BCE	2059 BCE	β	Raghavan et al. 2014
D4e	Rocky Point	Pre-Dorset	Nunavut, Canada	2142-1805 BCE	1974 BCE	β	Raghavan et al. 2014
D2a1	Itinnera	Saqqaq	W Greenland	1888-1565 BCE	1727 BCE	β	Raghavan et al. 2014
D4b1a2a1b	Sebystei	Kurgan	Altai-Sayan (S. Siberia)		450 BCE	ζ	Ricaud et al. 2004
D2a1	Buchanan	Middle-Dorset	Nunavut, Canada	173 BCE-232 AD	30 AD	β	Raghavan et al. 2014
D2a1	Tayara	Middle-Dorset	NE Canada	208-502	355 AD	β	Raghavan et al. 2014
D2a1	Gargamelle Cave	Middle-Dorset	NE Canada	312-549	386 AD	β	Raghavan et al. 2014
A2other	Port Moller	Eastern Aleut	AK Peninsula	263-562	421 AD		Raff et al. 2010, Coltrain 2010
D2a1	Gargamelle Cave	Middle-Dorset	NE Canada	312-563	438 AD	β	Raghavan et al. 2014
D2a1	Englee	Middle-Dorset	NE Canada	313-565	439 AD	β	Raghavan et al. 2014
D2a1	Eastern Point	Middle-Dorset	NE Canada	430-621	526 AD	β	Raghavan et al. 2014
D2a1	Philip's Garden	Middle-Dorset	NE Canada	426-628	527 AD	β	Raghavan et al. 2014
A2a	Paipelghak	Birnirk	Chukotka	572-674	623 AD	β	Raghavan et al. 2014
A2a	Paipelghak	Birnirk	Chukotka	576-676	626 AD	β	Raghavan et al. 2014
D2a	Saatut	Late-Dorset	Nunavut, Canada	578-675	627 AD	β	Raghavan et al. 2014
D2a	Alarnerk	Middle-Dorset	Nunavut, Canada	674-902	788 AD	β	Raghavan et al. 2014
A2b	Nuvuk	Thule	N. Alaska	1001-1280	1139 AD		This Study
A2a	Nuvuk	Thule	N. Alaska	1019-1274	1141 AD		This Study
A2b	Nuvuk	Thule	N. Alaska	1001-1283	1142 AD		This Study
A2other	Brooks River	Eastern Aleut(?)	AK Peninsula	1039-1258	1149 AD		Raff et al. 2010, Coltrain 2010
A2b	Nuvuk	Thule	N. Alaska	981-1316	1157 AD		This Study
A2other	Brooks River	Eastern Aleut(?)	AK Peninsula	1042-1264	1158 AD		Raff et al. 2010, Coltrain 2010
A2b	Nuvuk	Thule	N. Alaska	1005-1299	1159 AD		This Study
A2b	Nuvuk	Thule	N. Alaska	981-1341	1172 AD		This Study
D2a	Angekok	Late-Dorset	Nunavut, Canada	1051-1320	1181 AD	β	Raghavan et al. 2014
A2b	Nuvuk	Thule	N. Alaska	1036-1388	1213 AD		This Study
A2a	Nuvuk	Thule	N. Alaska	1039-1338	1217 AD		This Study
A2a	Nuvuk	Thule	N. Alaska	1037-1395	1220 AD		This Study
D4b1a2a1a	Nuvuk	Thule	N. Alaska	1046-1393	1227 AD		This Study
A2b	Nuvuk	Thule	N. Alaska	1044-1411	1237 AD		This Study
A2b	Nuvuk	Thule	N. Alaska	1055-1403	1242 AD		This Study
A2b	Nuvuk	Thule	N. Alaska	1059-1402	1245 AD		This Study
A2a	Nuvuk	Thule	N. Alaska	1056-1407	1246 AD		This Study
A2b	Nuvuk	Thule	N. Alaska	1028-1449	1249 AD		This Study
A2b	Nuvuk	Thule	N. Alaska	1054-1415	1250 AD		This Study
A2a	Nuvuk	Thule	N. Alaska	1067-1403	1251 AD		This Study
D4b1a2a1a	Nuvuk	Thule	N. Alaska	1060-1432	1269 AD		This Study
A2b	Nuvuk	Thule	N. Alaska	1116-1418	1269 AD		This Study
A2b	Nuvuk	Thule	N. Alaska	1069-1427	1270 AD		This Study
A2b	Nuvuk	Thule	N. Alaska	1042-1455	1271 AD		This Study
A2a	Nuvuk	Thule	N. Alaska	1115-1423	1273 AD		This Study
A2b	Nuvuk	Thule	N. Alaska	1072-1432	1275 AD		This Study
A2b	Nuvuk	Thule	N. Alaska	1115-1428	1277 AD		This Study
D4b1a2a1a	Silumiut	Thule	Nunavut, Canada	1212-1383	1288 AD		Raghavan et al. 2014, Coltrain et al. 2004
A2b	Nuvuk	Thule	N. Alaska	1169-1412	1290 AD		This Study
A2a	Norris Farms	Oneota	Illinois		1300 AD	γ	Stone and Stoneking 1998
A2a	Norris Farms	Oneota	Illinois		1300 AD	γ	Stone and Stoneking 1998
A2a	Norris Farms	Oneota	Illinois		1300 AD	γ	Stone and Stoneking 1998
A2a	Norris Farms	Oneota	Illinois		1300 AD	γ	Stone and Stoneking 1998
A2a	Nuvuk	Thule	N. Alaska	1123-1450	1303 AD		This Study
A2b	Nuvuk	Thule	N. Alaska	1090-1458	1303 AD		This Study
A2a	Nuvuk	Thule	N. Alaska	1146-1452	1308 AD		This Study
A	Kamarvik	Thule	Nunavut, Canada	1277-1411	1314 AD		Raghavan et al. 2014, Coltrain et al. 2004

Table 2S.5 continued

Haplogroup	Site	Culture	Location	2 sigma range (AD)	Date <sup>a</sup>	Notes	Reference
A2b	Nuvuk	Thule	N. Alaska	1146-1467	1319 AD		This Study
A2b	Nuvuk	Thule	N. Alaska	1104-1499	1328 AD		This Study
A2b	Nuvuk	Thule	N. Alaska	1170-1486	1343 AD		This Study
A2b	Nuvuk	Thule	N. Alaska	1153-1518	1346 AD		This Study
A2b	Nuvuk	Thule	N. Alaska	1161-1523	1352 AD		This Study
A2b	Nuvuk	Thule	N. Alaska	1165-1533	1360 AD		This Study
A	Silumiut	Thule	Nunavut, Canada	1296-1431	1361 AD		Raghavan et al. 2014, Coltrain 2009
D2a'b	Mink Island	unclear	SE AK	1284-1444	1362 AD		Raff et al. 2010, Coltrain 2010
A	Silumiut	Thule	Nunavut, Canada	1284-1429	1366 AD		Raghavan et al. 2014, Coltrain et al. 2004
D4b1a2a1a	Mink Island	unclear	SE AK	1303-1471	1395 AD		Raff et al. 2010, Coltrain 2010
D4b1a2a1a	n°3	Yakut	Kolyma (Sakha)		1400 AD	ε	Crubezy et al. 2010
A2b	Silumiut	Thule	Nunavut, Canada	1309-1442	1402 AD		Raghavan et al. 2014, Coltrain et al. 2004
A2b	Kamarvik	Thule	Nunavut, Canada	1319-1472	1412 AD		Raghavan et al. 2014, Coltrain 2009
A2b	Nuvuk	Thule	N. Alaska	1290-1584	1424 AD		This Study
D2a'b	Mink Island	unclear	SE AK	1319-1530	1436 AD		Raff et al. 2010, Coltrain 2010
A2b	Kamarvik	Thule	Nunavut, Canada	1326-1518	1438 AD		Raghavan et al. 2014, Coltrain 2009
A2a	Silumiut	Thule	Nunavut, Canada	1390-1526	1447 AD		Raghavan et al. 2014, Coltrain 2009
A2b	Silumiut	Thule	Nunavut, Canada	1413-1519	1451 AD		Raghavan et al. 2014, Coltrain et al. 2004
A2a	Qilakitsoq	Inuit	NW Greenland		1460 AD	δ	Gilbert 2007, Tauber 1989
A2a	Qilakitsoq	Inuit	NW Greenland		1460 AD	δ	Gilbert 2007, Tauber 1989
A2a	Qilakitsoq	Inuit	NW Greenland		1460 AD	δ	Gilbert 2007, Tauber 1989
A2a	Qilakitsoq	Inuit	NW Greenland		1460 AD	δ	Gilbert 2007, Tauber 1989
A2a	Qilakitsoq	Inuit	NW Greenland		1460 AD	δ	Gilbert 2007, Tauber 1989
A2a	Qilakitsoq	Inuit	NW Greenland		1460 AD	δ	Gilbert 2007, Tauber 1989
A2a	Qilakitsoq	Inuit	NW Greenland		1460 AD	δ	Gilbert 2007, Tauber 1989
A2b	Qilakitsoq	Inuit	NW Greenland		1460 AD	δ	Gilbert 2007, Tauber 1989
A2b	Imaha	Thule	NE Canada	1414-1533	1461 AD		Raghavan et al. 2014, Coltrain et al. 2004
A	Kamarvik	Thule	Nunavut, Canada	1416-1568	1476 AD		Raghavan et al. 2014, Coltrain 2009
A2b	Nuvuk	Thule	N. Alaska	1319-1657	1484 AD		This Study
D2a'b	Mink Island	unclear	SE AK	1407-1638	1491 AD		Raff et al. 2010, Coltrain 2010
A2b	Silumiut	Thule	Nunavut, Canada	1430-1625	1493 AD		Raghavan et al. 2014, Coltrain 2009
A	Kamarvik	Thule	Nunavut, Canada	1444-1636	1498 AD		Raghavan et al. 2014, Coltrain et al. 2004
A2b	Kamarvik	Thule	Nunavut, Canada	1439-1630	1505 AD		Raghavan et al. 2014, Coltrain 2009
A2b	Kamarvik	Thule	Nunavut, Canada	1441-1650	1510 AD		Raghavan et al. 2014, Coltrain et al. 2004
A2b	Kamarvik	Thule	Nunavut, Canada	1447-1649	1516 AD		Raghavan et al. 2014, Coltrain et al. 2004
A2b	Kamarvik	Thule	Nunavut, Canada	1446-1633	1517 AD		Raghavan et al. 2014, Coltrain 2009
D2a'b	Mink Island	unclear	SE AK	1446-1658	1541 AD		Raff et al. 2010, Coltrain 2010
A2b	Sadlermiut	Thule	Nunavut, Canada	1453-1650	1544 AD		Raghavan et al. 2014, Coltrain 2009
A2a	Nuvuk	Thule	N. Alaska	1414-1683	1547 AD		This Study
A2b	Sadlermiut	Thule	Nunavut, Canada	1481-1662	1568 AD		Raghavan et al. 2014, Coltrain 2009
D4b1a2a1a	Sadlermiut	Thule	Nunavut, Canada	1495-1666	1582 AD		Raghavan et al. 2014, Coltrain 2009
A2b	Sadlermiut	Thule	Nunavut, Canada	1489-1697	1604 AD		Raghavan et al. 2014, Coltrain 2009
A2b	Sadlermiut	Thule	Nunavut, Canada	1482-1720	1614 AD		Raghavan et al. 2014, Coltrain 2009
A2b	Port Hope, Ailek	Thule	NE Canada	1512-1723	1618 AD	β	Raghavan et al. 2014
A2b	Port Hope, Ailek	Thule	NE Canada	1513-1756	1635 AD	β	Raghavan et al. 2014
A2b	Labrador	Thule	NE Canada	1525-1759	1642 AD	β	Raghavan et al. 2014
A2b	Sadlermiut	Thule	Nunavut, Canada	1529-1774	1653 AD		Raghavan et al. 2014, Coltrain 2009
A2a	Nuvuk	Thule	N. Alaska	1481-1885	1669 AD		This Study
A2b	Hopedale	Thule	NE Canada	1610-1824	1717 AD	β	Raghavan et al. 2014
A2b	Sadlermiut	Thule	Nunavut, Canada	1624-1892	1736 AD		Raghavan et al. 2014, Coltrain 2009
A2b	Sadlermiut	Thule	Nunavut, Canada	1642-1892	1748 AD		Raghavan et al. 2014, Coltrain 2009
D4b1a2a1a	Labrador	Thule	NE Canada	1641-1881	1761 AD	β	Raghavan et al. 2014
D4b1a2a1a	Port Hope, Ailek	Thule	NE Canada	1635-1902	1769 AD	β	Raghavan et al. 2014

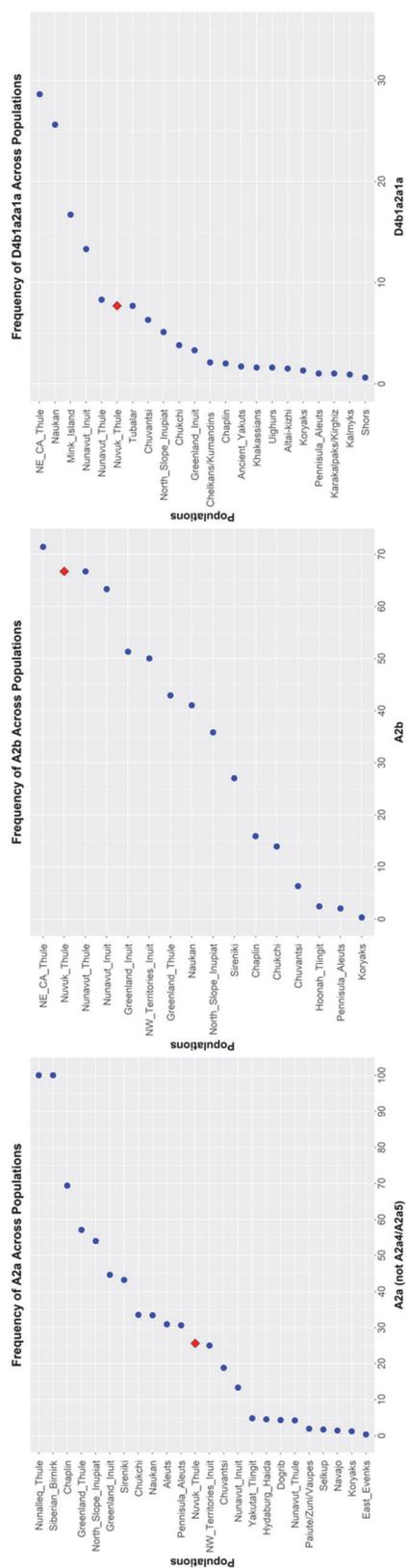


Figure 2.1 Beringian haplogroups at Nuvuk across populations. Ordered frequencies of the three haplogroups (A2a, A2b, and D4b1a2a1a) found at Nuvuk (marked here by a red diamond) and present in at least one individual sampled across the populations analyzed in this study. Data is drawn from Table 2S.3, minus those populations lacking the specified haplogroup.

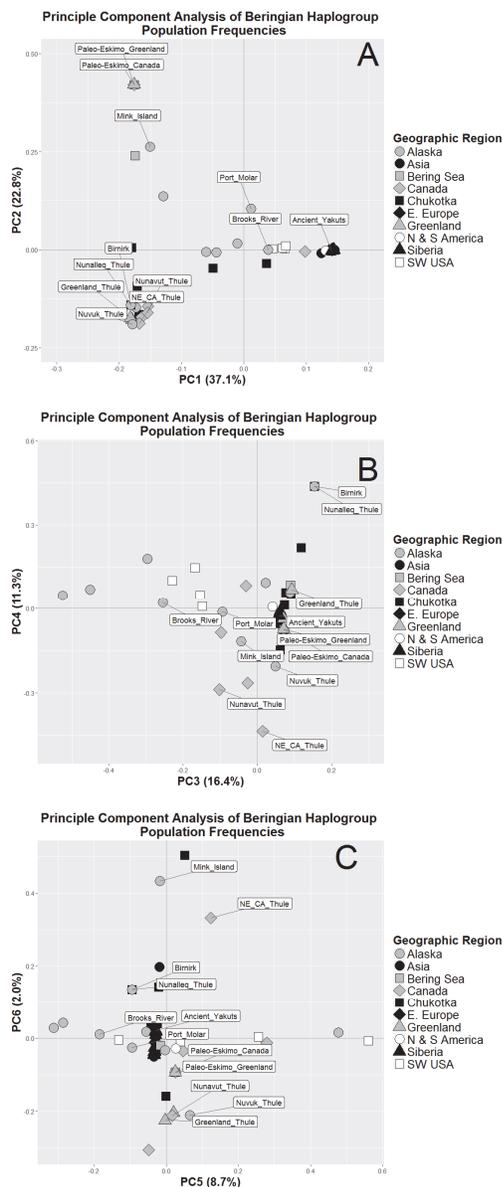


Figure 2.2 Principle component analysis of the Beringian haplogroup population frequencies. Labeled populations are the ancient sample sets under consideration, including the population at Nuvuk sampled here. Grey coloring indicates populations associated with northern North America, black coloring indicates Eurasian populations, and white coloring indicates southern populations in North and South America. All allele frequency data is drawn from Table 2S.3. (A) Principal Components 1 and 2, with associated proportion of variance explained (B) Principal Components 3 and 4, with associated proportion of variance explained (C) Principal Components 5 and 6, with associated proportion of variance explained. By Principal Component 6, 98.3% of the variance is explained.

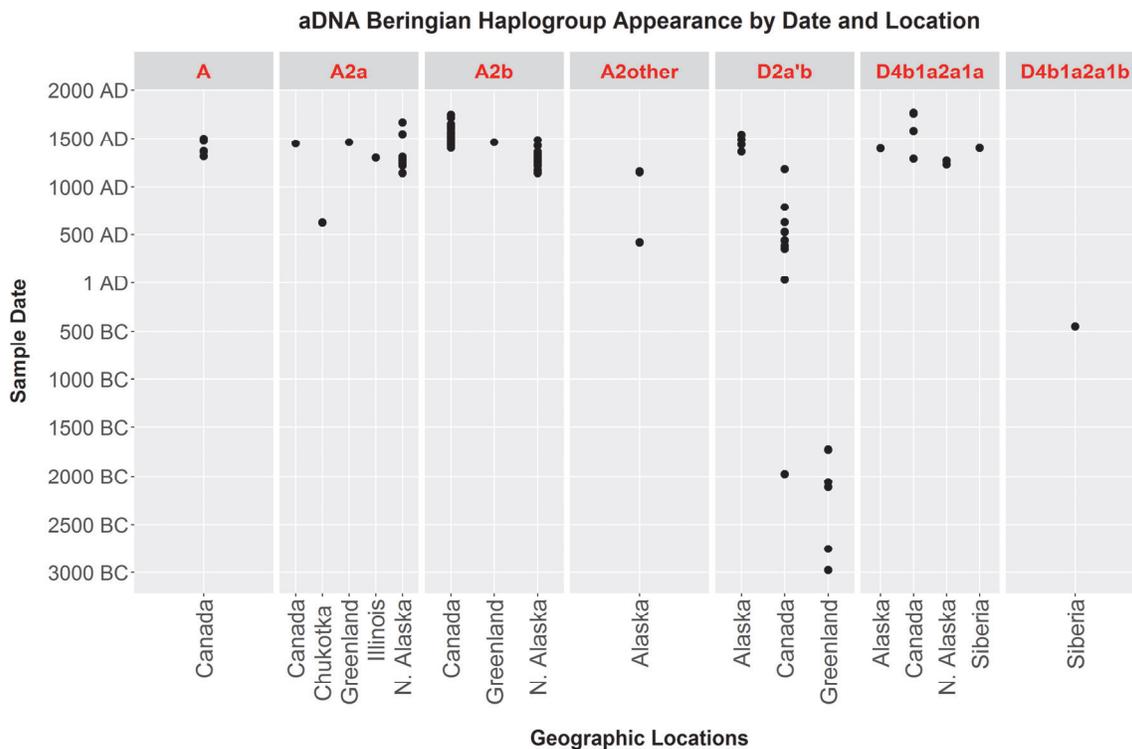


Figure 2.3 aDNA Beringian haplogroup appearance by date and location. Haplogroup assignment and geographic location are plotted for 109 dated human remains from the literature and from Nuvuk for which mtDNA has been typed. All data points are drawn from Supporting Table 5.

## CHAPTER 3

# PHYLOGENETIC RECONSTRUCTION OF BERINGIAN HAPLOGROUPS A2B AND D4B1A2A1A USING DATED NEO-ESKIMO THULE REMAINS

## Introduction

Modern populations that reside in what was once Beringia have high frequencies of a group of mitochondrial DNA haplogroups that are rare to absent elsewhere. In addition to the limited geographic range, these “Beringian-specific” haplogroups of A2a, A2b, D2a, and D4b1a2a1a all expanded in the Holocene, much more recently than the other Native American mitochondrial clades (Tackney et al., 2016). Haplogroup D2a has been linked to the first migrations into the North American Arctic by the Paleo-Eskimo, and the remaining haplogroups seem to be associated with the modern ancestors of the later Neo-Eskimo expansion and/or Native American Na-Dene speakers (Achilli et al., 2013; Raghavan et al., 2014; Chapter 2). There are details missing, however, of these two major population movements over the past six thousand years. In particular, exactly which clade arose with which archaeological tradition is unclear beyond the apparently fixed signature of D2a in the early Paleo-Eskimo (Raghavan et al., 2014). Ancient DNA analysis of dated samples in the region can help to clarify the first appearance of these maternal lineages and their sequences can inform calculations of the molecular clock in these clades.

Haplogroup A2b and a haplotype of haplogroup D4b1a2a1a (D4b1a2a1a1) are unique in having coalescence and expansion dates that closely align with the archaeological record of the Thule migration (Dryomov et al., 2015; Chapter 2). Haplogroup A2b has only been observed in the Inuit, Iñupiat, and Eskimo peoples of North America and Chukotka (and in a single Koryak individual). Haplogroup D4b1a2a1a is typed at moderate frequency in the same communities, but it also has an expanded range with low ( $\leq 2\%$ ) frequency in the populations of Sakha, the Altai-Sayan

region, central Asia, and Eastern Europe (see Chapter 2). Unfortunately, only the Hypervariable Region I (HVR-I) region of the mitochondrial genome is typically sequenced in population-wide surveys and, in this case, such limited information might indicate a false correspondence of mtDNA lineage. The only whole mitochondrial genome from a non-Beringian population for D4b1a2a1a is from a Tubalar. This single genome lacks the polymorphisms at np 11383 and 14122 (Volodko et al., 2008), while all the haplotypes from Beringian populations sequenced to date contain these variants (Dryomov et al., 2015). The haplotype with 11383 and 14122 is named D4b1a2a1a1 (Derenko et al., 2010). It is possible that D4b1a2a1a1, and not D4b1a2a1a, was carried with the Arctic migrations that followed the Paleo-Eskimo.

Modern Inuit, Iñupiat, and Eskimo people are likely descendants of a second, separate, and much more recent Arctic tradition known as the Neo-Eskimo. Beginning about 1,500 yr BP, two Neo-Eskimo cultures - Punuk, with origins around St. Lawrence island, and Birnirk, focused on the coasts of the Chukchi Sea - were contemporaries with descendants of the older Paleo-Eskimo, namely the Ipiutak around the Bering Strait (Jensen, 2014; Mason, 2009). Sometime prior to 1000 AD a third tradition, the Thule, appeared on the western North American Arctic archaeological landscape, though their origin location is unclear. In the 13<sup>th</sup> century (Friesen and Arnold, 2008) they expanded eastwards out of what was once Beringia to eventually regionalize and diversify into the modern North American Arctic peoples (Jensen, 2014; Mason and Bowers, 2009). The earliest appearance of both A2b and D4b1a2a1a is at the North Slope Thule site of Nuvuk in the 12<sup>th</sup> and 13<sup>th</sup> centuries, respectively (Chapter 2). Both haplogroups (specifically haplotypes A2b1 and D4b1a2a1a1) were typed in the later, more eastern Thule. Neither

haplogroup was found in a handful of early Birnirk samples from northeastern Chukotka, dated to between 570 - 680 AD, nor in a southwestern Alaskan Thule site at Nunalleq, dated to between 1300 and 1660 AD (Raghavan et al., 2014). Selected for study here are human remains from two locations: the early Thule cemetery of Nuvuk, and a late prehistoric Eskimo village at Igliqtiqsiugvigruaq.

Nuvuk was a long-term Thule village at Pt. Barrow, AK, and it is uniquely situated in time and space to investigate Thule origins. Nuvuk was once the northernmost indigenous community of Alaska, and the archaeological material at Nuvuk covers a nearly one thousand year uninterrupted occupation from early Thule to postcontact Iñupiat Eskimo. The few grave goods present amongst the burials show clear links to other Western and Eastern Thule locations, making Nuvuk by far the largest Thule cemetery ever excavated in North America (Jensen, 2009a; Jensen, 2009b). A recent reassessment of the temporal occupation at Nuvuk sampled from the human remains and the calibrated and marine reservoir corrected dates span 996 AD to 1631 AD (Coltrain et al., Under Review). Two of these samples (Labeled here as “Nuvuk 1” and “Nuvuk 2”; not to be confused with “Nuvuk 01” from Jensen, 2007), dating to the 13<sup>th</sup> century, were typed for HVR-I coding region variation and belong to haplogroup D4b1a2a1a. These two individuals provide a unique opportunity to explore this haplogroup at a time nearer to its calculated appearance. Another of these samples (“Nuvuk 3”), dating to the 16<sup>th</sup> century, was typed with PCR methods but gave conflicting genotypes. While potentially a member of the root A2 haplogroup, it was included here to see if next-generation sequencing methods could clarify its phylogenetic position.

The Iqliqtsiugvigruaq site is located just inside the Kobuk Valley National Park on a bluff adjacent to the Kobuk River. It was excavated by an archaeological team led by Dr. Doug Anderson of Brown University. It is near the modern Iñupiat community of Kiana, AK. The site represents a large precontact village dating to the turn of the 19<sup>th</sup> century with multiple semisubterranean dwellings of dimensions wider and longer than typical for Northwest Alaska at that time. While dating before the known arrival of Europeans, western trade goods have been found in some of the houses, particularly metal and glass beads. The origin of these western artifacts is unclear. Two houses have been excavated, and in one the remains of three individuals were discovered: a subadult (referenced here as “Kobuk 1”), and two adults (“Kobuk 2” and “Kobuk 3”). Isotopic analysis of the human remains suggests subsistence on freshwater fish (but not marine resources) and terrestrial diet (but not caribou). Local oral histories suggest that the village was abandoned sometime in the 1800s after a collapse in the fish stocks and currently the site is being eroded away by the Kobuk River. HVR-I coding variation of these three individuals indicated that they belong to haplogroup A2b.

The human remains at these two precontact Arctic sites in North Alaska can directly inform an investigation of the A2b and D4b1a2a1a distributions and phylogeny in an early Thule and late Iñupiat setting. They can also clarify ancestry with modern populations in the region, namely Barrow for Nuvuk and Kiana for Iqliqtsiugvigruaq. We attempted hybridization capture of the mitochondrial genome from each of the six individuals. Ancient DNA was extracted from small amounts of bone fragments and initial haplogroup identity was confirmed via PCR amplification and Sanger sequencing of the HVR-I region. Each sample was then converted into next-generation sequencing

libraries and sequenced with the Ion Proton system. Variants were typed from the resulting reads and consensus sequences were compared to known modern and ancient mitochondrial variation. We used the dated sequences to calculate the time to the most recent common ancestor for each of the haplogroups in order to determine what these new dates and phylogenetic trees can reveal about the Thule migration.

## **Materials and Methods**

### *Description of Samples*

The archaeological context for the three Nuvuk individuals has been given in Chapter 2. The library for Nuvuk 1 comes from sample ID 07-JJ22-YUCXV, which has been directly dated with AMS Cal  $2\sigma$  Range AD of 1060-1432, with an intercept of 1269 AD. The library for Nuvuk 2 comes from sample ID 07-C24-42/43, which has been directly dated to 1046-1393, with an intercept of 1227 AD (Coltrain et al., Under Review). Both Nuvuk 1 and Nuvuk 2 have HVR-I sequences indicating membership in haplogroup D4b1a2a1a. The library for Nuvuk 3 comes from sample ID 08-C32-108-VUNFJ. This sample was directly dated to 1400-1708, with an intercept of 1557 AD. HVR-I sequences from this sample were inconsistent, though some polymorphisms suggested that the haplogroup was A2root. This sample was chosen for next-generation sequencing to clarify the Sanger results.

The house at Igliqtiqsiugvigruaq with the three human remains dates to the 1790s-1800, just prior to formal contact. The dwelling has limited faunal remains, but it does have a wood-burning hearth, pieces from an oil lamp, and historic metal artifacts associated with the burials. The house contains almost no other debris. The subadult (Kobuk 1) and one of the adults (Kobuk 2) died close to each other in one corner of the

house, while the skeletal remains from the third adult (Kobuk 3) had a distinct and deeper distribution at another location inside the dwelling.

It is unclear if the three individuals are contemporaneous with the house. The isotope analysis of the bones suggests a more inland subsistence pattern, inconsistent with the location, and a lack of microcarbon in the dental calculus does not fit with the presence of a wood burning hearth. The individuals could represent temporary inhabitants who arrived following the collapse of Iqliqtsiugvigruaq. The metal objects associated with the remains could have been carried into the dwelling by these inhabitants, and not associated with the earlier village. Clarification might come from directly dating the hearth and the bones, and comparing these to the age of the dwelling(s), however, if the dates are too recent or close together no conclusions could be made.

#### *Sample Analysis Agreements*

For the Nuvuk samples letters of support were requested and received by the Ukpeaġvik Iñupiat Corporation, the Native Village of Barrow, and the Barrow Senior Advisory Council during the initiation of the project described in Chapter 2. For the Iqliqtsiugvigruaq samples, permissions were granted from both the local community of Kiana, AK, and the National Park Service. All involved parties understood that small samples of human remains were to be analyzed in a destructive manner to determine ancestral relationships with modern Iñupiat communities.

### *DNA Extraction and Mitochondrial HVR1 Sanger Sequencing*

DNA was extracted from human bone remains using the protocol described in Chapter 2. HVR-I sequencing also was performed as per Chapter 2. The preliminary HVR-I amplifications were limited to primer sets HVR1\_P1F and HVR1\_P1Rc (159 bp product), HVR1\_P2Fb and HVR1\_P2R (134 bp product), and HVR1\_P3F and HVR1\_P3R (157 bp product). The reactions were repeated on a duplicate set of extractions before proceeding to next-generation sequencing.

### *Ion Torrent Library Preparation*

One extraction for each sample was chosen for next-generation sequencing library preparation, along with its associated extraction blank. No DNA fragmentation or size selection was performed prior to library preparation, and a library blank (water input) was included with each library set. Barcoded Ion A Adapters were designed using laboratory-specific 8-base barcodes based on a 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) (Krishnan et al., 2011). Directionality for Barcode 2 and Adapter P1 was provided with “TT tails,” while directionality for the remaining Barcodes were provided by shortened complementary strands (Table 3.1). All samples were processed with a unique barcode (Table 3.2), while extraction blanks were processed with the same barcode as their associated sample, and library blanks were processed with a combination of all the barcodes used in that library set.

During adapter ligation final adapter concentrations in the reaction were reduced to .04  $\mu$ M. Library workflows changed between samples; details are outlined in Table 3.2. All primary amplifications were performed in AmpliTaq Gold 360 Master Mix (Life

Technologies), while all secondary amplifications were performed in Q5 Hot Start mastermix (NEB). Amplification primers used were Ion\_Aamp and Ion\_P1amp at 0.4  $\mu$ M final concentrations (Table 3.1). Cycles for the primary and secondary amplifications were determined from post-ligation and post-primary amplification quantitative PCR runs, respectively, using the GeneRead Library Quantification Kit (Qiagen).

For Sample Nuvuk 2, one library was processed up to the 1<sup>o</sup> amplification. Nuvuk 2B was then amplified, captured, and sequenced using a slightly different workflow than Nuvuk 2A (Table 3.2).

#### *Hybridization Capture and Sequencing of Mitochondrial DNA*

Hybridization capture of mitochondrial DNA molecules was completed as described in Tackney et al. (2015), with the following modifications: Hybridization times varied and are listed in Table 3.2. Blocking oligos used are listed in Table 3.1. Each library was captured separately using a different ratio of library ng: bait ng (though bait amounts were all between 200 - 250 ng; Table 3.2). Library blanks (both extraction and library) were also captured separately, but were only amplified the number of cycles as the lowest concentration associated sample, without regard for their final nanograms after 2<sup>o</sup> amplification. Library blanks were not sequenced if their molecule count following capture and postcapture amplification was more than an order of magnitude less than the associated sample.

Captured library concentrations in pM were calculated via qPCR using the GeneRead Library Quantification Kit (Qiagen). Libraries were templated individually on an Ion Torrent One Touch 2 (Life Technologies), using appropriate templating kits.

Libraries from Nuvuk 1, Nuvuk 2A, and Kobuk 2 were diluted to 100 pM prior to templating. Libraries from Kobuk 1 and Kobuk 2 were diluted to 110 pM (a 10% effective increase), and then the templating reactions were joined for one enrichment step. Libraries from the remaining samples were input at 400% the suggested concentration; previous runs at the core had indicated that chip loading might be improved with more templated molecules. Following templating an aliquot of the post emulsion PCR reaction was evaluated separately with the Ion Sphere Quality Control kit to determine the percentage preenrichment of the templating reaction. The manufacturer suggests a preenrichment percentage between 10% and 30% and all reactions fell within 10-23%. An enrichment step was performed on an Ion Torrent ES to enrich for positive Ion Sphere Particles (ISPs).

All libraries were sequenced on an Ion Torrent Proton using the Ion PI chip. Templating and Sequencing chemistry was enhanced during the course of the experiment, and so slightly differs between samples. Additionally, certain libraries were sequenced together on one chip, while other libraries were sequenced on individual chips. The libraries sequenced together and the chemistry used for each sample is shown in Table 3.2.

### *Read Mapping and Variant Calling*

For all samples, reads were initially base called within the Ion Torrent Suite, incorporating “Base Recalibration” with the included reference of rCRS (NC\_012920). Properly barcoded reads were isolated and trimmed for base quality and 3’ P1 adapter sequence (ATCACCGACTGCCCATAGAGAGGCTGAGAC) using cutadapt 1.8 (Martin, 2011). Given that the various samples were sequenced over a long length of

time with different versions of the Torrent Suite software package (we were unable to go back and reanalyze earlier runs with updated software), slight technical differences to the analysis pipeline up to this point are outlined in Table 3.3.

Trimmed reads were mapped against the rCRS with TMAP version 5.0.0 (-g 3 -n 12 -a 1 -o 2 -v -Y stage1 map2 --min-seq-length 30 --max-seq-length 120 map3 --min-seq-length 30 --max-seq-length 120 map4 --min-seq-length 30 --max-seq-length 120) without 3' soft-clipping and limiting mapped reads to sizes of 30 - 120 bases. Mapped BAMs were sorted and then processed to remove reads with mapping quality (MAPQ) <90 using Samtools 1.2 (Li et al., 2009). Duplicate reads were marked and removed with Picard Tools v1.91 (<http://sourceforge.net/projects/picard/>). Variants were called using Samtools mpileup and bcftools, filtering for indels closer than 10 bases, low quality variants ( $\leq 15$ ), depth less than 3, heterozygote calls with the variant allele at less than 50%, and strand bias with a Phred-scaled P-value  $\geq 40$ . VCF files were further manually edited to remove or make homozygote any remaining heterozygote calls after visual inspection of the supporting reads in IGV (Thorvaldsdóttir et al., 2012). Consensus sequences were created with the VCF file and the rCRS reference using bcftools consensus.

For Nuvuk Sample 2, reads from both Proton PI chips (Sample 2A and 2B) were filtered and mapped as above, but then the mapped BAMs were merged prior to variant calling. Reported coverage and damage patterns are based off of this merged BAM.

For all samples, genomic coverage depth was calculated at a 1 base window size with igvtools (Thorvaldsdóttir et al., 2012). Sequencing QC metrics were analyzed

with FastQC v. 0.11.2 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>).

Nucleotide misincorporation patterns were assessed using MapDamage v2.0.2-12 (Jónsson et al., 2013). Haplotypes were identified using Haplogrep2 (Kloss-Brandstätter et al., 2011).

### *Phylogenetic Analysis and Coalescence Time Estimates*

Currently available whole mitochondrial genome sequences for haplogroups A2b and D4b1a2a were drawn from Dryomov et al. (2015) recently published survey of Arctic mitochondrial diversity. Duplicate haplotypes were discarded unless samples were taken from distinct cultural populations. One ancient Canadian Sadlermiut Thule sample (XIV-C-748) and one Greenlandic Cape Irmenger Thule sample (KAL1245) were included in our analysis. These are the only two whole mitochondrial sequences, falling within these two haplogroups, sequenced by Raghavan et al. (2014) at 100% coverage. Consensus sequences were created from provided FASTQ files (Dryomov et al., 2015). Including these genomes, our dataset consisted of 17 D4b1a2a (which includes D4b1a2a1a and D4b1a2a1b) genomes and 28 A2b genomes.

Maximum Parsimony trees were created using mtPhyl 4.015 (<https://sites.google.com/site/mtphyl/home>) and then manually edited for clarity. The naming schemes of Dryomov et al. (2015) and Derenko et al. (2010) were followed. 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 trees. MAFFT (Katoh and Standley, 2013) was used to align all the sequences from each dataset with the highly accurate L-INS-i methodology. Once aligned, the same regions listed above were

removed from the alignment. These sites are known mutational hotspots and/or positions with recurrent sequencing errors (van Oven and Kayser, 2009).

PartitionFinder (Lanfear et al., 2012) was used to suggest a partitioning scheme and an appropriate nucleotide substitution model. This indicated that there was no justification for separating out the control region from the coding region in either of the two alignments. The HKY model was the best-fit substitution model according to the Bayesian information criterion. Bayesian estimated coalescence times for the A2b and D4b1a2a were calculated using BEAST v2.2.1 (Bouckaert et al., 2014). Tip dates for the ancient A2b samples were set at 200 yr BP (KAL1245) and 150 yr BP (Kobuk 1, 2, and 3). Tip dates for the ancient D4b1a2a samples were set at 368 yr BP (XIV-C-748), 681 yr BP (Nuvuk 1), and 723 yr BP (Nuvuk 2).

For both datasets, two Markov chain Monte Carlo runs of 30,000,000 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. The HKY site model, as per the PartitionFinder results, was selected as well as a Coalescent Bayesian Skyline tree prior (5 populations), and a lognormal clockRate prior ( $M=2.67E-8$ ,  $S=1.4$ ) with a starting rate taken from Fu et al. (2013b). TreeAnnotator v2.2.1 was used to produce the Maximum clade credibility tree with a posterior probability limit of 60%, and to calculate target clade divergence times and 95% highest posterior density (HPD) intervals. Mitochondrial genomes within the A2b1 and D4b1a2a1a1 clades were constrained in a prior to be monophyletic to aid in later dating.

## Results

### *Enriched Read Summary*

Mitochondrial genomes were successfully captured for all samples except Nuvuk 3. Coverage of the 16,569 bp genome was >99% for five samples while the Nuvuk 3 mitochondrial genome was only covered at 7.7%. The number of high quality and unique sequenced reads for each library, as well as other sequencing metrics, are given in Table 3.4.

The library from Nuvuk 1 was not amplified following hybridization capture (Table 3.2) and it is noteworthy that the total number of mapped reads from this library is 2 to 3 orders of magnitude less than that from postcapture amplified libraries. The final read count (Unique, MAPQ $\geq$ 90), was within an order of magnitude of the other samples; this implies that while amplification following capture does increase the final tally of captured fragments, the vast majority of reads from those fragments are PCR duplicates.

Sequencing coverage was fairly uniform in the 5 successfully captured samples, though Nuvuk 2, Kobuk 1, and Kobuk 3 all displayed sporadic regions of ultra-low or missing coverage (Figure 3.1). For the 5 samples, 0 to 47 bases (out of 16,569) will need to be Sanger sequenced to fill in the Proton read results (Table 3.4). None of these libraries appear to be sequenced to exhaustion and additional Proton sequencing runs could possibly increase the coverage (excluding Kobuk 2 that was 100% covered) and depth in each case. These additional runs might have diminishing returns, however: Nuvuk 2 was sequenced twice, from two library preparations (but only one extract), and still 100% coverage was not reached.

The extraction blanks associated with Nuvuk 2B (“Nuvuk 2B EB”) and the

Kobuk samples (“Kobuk EB”) both had high molecule counts following capture and amplification; these did not meet our criteria for an order of magnitude difference with true sample. When this happens either the cause is a true contaminating human mitochondrial DNA molecule, or environmental bacterial genome fragments were captured at a high level by the probes from the library. Many environmental bacterial strains have close homology to the human mitochondrial genome, so these “off-target” captures are expected with every hybridization.

Nuvuk 2B EB and Kobuk EB were sequenced together on one Proton P1 chip (“Shared C”). The reads from Nuvuk 2B EB had similar coverage of the mitochondrial genome as the failed Nuvuk 3, but with an average depth of only 1.4X. Both samples did not have enough high quality, unique reads to make any confident statement of if human DNA is present in the original bone sample (for Nuvuk 3) or what actual genome was captured (for Nuvuk 2B EB) (Table 3.4). Both samples also showed a typical pattern of “coverage islands” where certain regions of the captured sequence are covered to a large extent, with other regions not covered at all. In our experience, these islands are usually indicative of a failed or nonspecific capture.

The reads from the captured Kobuk EB sample blank covered 32% of the genome; this is far below that expected from a successful capture, but above that expected from a failed capture (see Table 3.4). Even with the large number of high quality, unique reads, this library still displayed coverage islands (Figure 3.1), and so it is unclear if these fragments actually represent a contaminating genome in the library or DNA extraction preparations. None of the captured fragments from Kobuk EB overlapped the HVR-I control region, which would explain the cleanliness of the earlier

PCR-based analyses on this extraction blank.

### *Consensus Sequence Haplotype Identification*

The variants typed in Nuvuk 1 and Nuvuk 2 placed these two individuals within haplogroup D4b1a2a1a, or more specifically haplotype D4b1a2a1a1 if the naming scheme of Derenko et al. (2010) is followed (D4b1a2a1a + 11383, 14122C) (Figure 3.2). Nuvuk 1 had a private variant G14198A, while Nuvuk 2 had three private variants C3487T, G11914A, and C13654T. Nuvuk 2 shares G11914A with a recently sequenced modern Chukchi individual (Dryomov et al., 2015). All three samples from Iglitqiugvigruaq had variants which placed them within haplogroup A2b, or more specifically A2b1 if the naming scheme of Dryomov et al. (2015) is followed (Figure 3.3). Kobuk 1 had private variants C535T, C541T, C12669T. Kobuk 2 had a private variant A14148G. Kobuk 3 had a private variant C16313T. None of these samples shared derived mutations with known modern haplotypes.

We did not attempt to variant call or create consensus sequences for Nuvuk 3 or Nuvuk 2B EB. From the 864 unique reads from Kobuk EB, however, some variants within the covered regions could be typed using our pipeline. There were heterozygote and homozygote SNPs at nps 2002, 2548, 4632, 4769, 4820, 7747, 7779, 12073, and 13595 (though see concerns above about the authenticity of these reads). Only the A4769G SNP was shared with the 5 successfully captured samples, and that is only because the derived state in rCRS is A, but G worldwide. No other SNPs in this sample are associated with A2b or D4b1a2a and we did not attempt a further determination of the haplotype of this possible contaminant.

### *A2b and D4b1a2a Coalescence Time Estimates*

The coalescence date for the D4b1a2a1a1 node from the D4b1a2a alignment had a mean value of 14,770 yr BP, a median value of 9,113 yr BP, and a 95% HPD of 952 - 46,481 yr BP (Figure 3.2). The coalescence date for the A2b root from the Maximum clade credibility tree of the A2b alignment following the BEAST analysis had a mean value of 4,659 yr BP, a median value of 2,707 yr BP, and a 95% HPD of 269 - 15,431 yr BP. The A2b1 node had a mean value of 4,408 yr BP, a median value of 2,572 yr BP, and a 95% HPD of 284 - 14,685 yr BP (Figure 3.3).

### *Authentication of Ancient DNA Work*

Pre-PCR work was carried out in a dedicated cleanroom, as described in Chapters 1 and 2. Adapter barcodes used for each sample have not previously been sequenced in our laboratory (nor have they been used at the sequencing core facility), so there is no chance of cross-contamination from a previous sequencing library with these barcodes. All mtDNA haplotypes are known for all laboratory personnel and none match those reported here. While we routinely sequence haplogroups A2b and D4b1a2a1a in our laboratory given our geographic area of focus, these are the first whole genomes we have sequenced on the Ion Proton with these haplotypes.

All extraction and PCR blanks for the HVR-I Sanger Sequencing genotyping were clean. All extraction and library blanks for the next-generation sequencing contained an order of magnitude less molecules following mitochondrial capture and postcapture amplification, except for the extraction blanks associated with Nuvuk 2B (“Nuvuk 2B EB”) and the Kobuk samples (“Kobuk EB”). When these two blanks were sequenced, they both captured much less mitochondrial fragments than a true sample.

For Kobuk EB, where 864 reads were retained after our mapping pipeline (covering 32% of the genome), the SNPs in those mitochondrial mapped reads were distinct from the SNPs in the associated sample libraries.

We evaluated the misincorporation patterns of the reads mapping in the five captured samples and in Kobuk EB. On the 5' end of reads we expected an increase of C-to-T substitutions and on the 3' end of reads we expected an increase in G-to-A substitutions, all from the deamination of cytosine to uracil in single-stranded DNA overhangs (Briggs et al., 2007). Our bioinformatics pipeline already limited mapped reads to sizes of 30 - 120 bp, fragment lengths appropriate for authentic ancient DNA molecules.

All the captured samples showed an increase in C-to-T substitutions on the 5' ends of reads, with the Kobuk 2 and 3 libraries displaying the clearest signal (Figure 3.4). All samples also had an increase of G-to-A substitutions at the 3' end of reads, however, the libraries from Nuvuk 2 and Kobuk 1 had unusual spikes and the library from Nuvuk 1 almost had no signal. The lack of smooth increases of both type of substitutions and the continued presence of other types of substitutions (the grey lines) in all libraries indicates that the low base quality issue seen in Tackney et al. (2015) has not been fully rectified even with the updated bioinformatics pipeline and oligo sequences. There is a continued read end quality issue stemming from the use of the Proton sequencer, which is unfortunately masking DNA damage patterns. The expected level of damage in these samples, however, is tempered by the still unclear understanding of damage accumulation over time and in specific environments. While all of the captured samples were <1,000 years in age, none of the bones were preserved in permafrost and likely both sites

experienced Arctic seasonal fluctuations in temperature and humidity. From the 864 reads mapped in Kobuk EB, the damage patterns were not at all expected for aDNA. Substitutions varied across read positions and neither read end showed signs of consistent cytosine deamination (Figure 3.4). These results further suggest the reads in Kobuk EB are modern contamination (likely human, but possibly bacterial).

## **Discussion**

Past and present human genetic variation from Alaska is underrepresented in the literature. Previous surveys across the Siberian and North American Arctic have sampled robustly from modern populations of Siberia and Chukotka (Derenko et al., 2010; Tamm et al., 2007; Volodko et al., 2008; Chapter 2), as well as the Inuit of Canada and Greenland (Gilbert et al., 2008; Helgason et al., 2006; Saillard et al., 2000; Chapter 2). The native people of Alaska were, until recently, poorly characterized genetically outside of the Aleutian Islands, the Alaskan Peninsula (Rubicz et al., 2003; Zlojutro et al., 2009), and the southeastern Na-Dene (Schurr et al., 2012). In 2014, a large number of Paleo-Eskimo and Neo-Eskimo human remains were genotyped, although the only Alaskan archaeological site with successful results was the precontact Thule site of Nunalleq near the modern Yup'ik Eskimo village of Quinhagak (Raghavan et al., 2014). In 2015 the first population wide sampling of the Iñupiat in the North Slope of Alaska was published, filling in a critical hole in modern Arctic sampling (Raff et al., 2015). In both of these investigations, however, the vast majority of individuals did not have their whole mitochondrial genomes sequenced. When investigating the Neo-Eskimo, in actuality only 4 ancient mitochondrial genomes have been produced with sufficient coverage, all from outside of Alaska: 1 Greenlandic Thule from the Cape Irminger site, 1 Canadian

Sadlermiut from Southampton Island, and 2 Siberian Birnirk from the settlement of Paipelghak in northeast Siberia (Raghavan et al., 2014). The Thule samples are further east and later in time than the archaeological evidence for the onset of the Thule, while the Birnirk samples have so far proven to be monomorphic for haplogroup A2a (Chapter 2).

Here, we have successfully extracted, captured, and whole mitochondrial genome sequenced 5 individuals from two ancient northern Alaskan Thule sites at Nuvuk and Iqliqtiqsiugvigruaq (Table 3.4). Two individuals from Nuvuk (Nuvuk 1 and 2) had sequences firmly within haplogroup D4b1a2a1a1 (Figure 3.2). They were sequenced with 99.85% and 99.72% coverage, at 15X and 73X depth, respectively (Figure 3.1, Table 3.4). Their genomes had private variants at G14198A (Nuvuk 1), and at C3487T, G11914A, and C13654T (Nuvuk 2). Nuvuk 2 shared G11914A with a recently sequenced modern Chukchi individual (Dryomov et al., 2015). The three individuals from Iqliqtiqsiugvigruaq (Kobuk 1, 2, and 3) were firmly within haplogroup A2b1 (Figure 3.3). They were sequenced with 99.9%, 100%, and 99.9% coverage, at 94X, 129X, and 80X depth, respectively (Figure 3.1, Table 3.4). Kobuk 1 had private variants at C535T, C541T, C12669T, Kobuk 2 had a private variant at A14148G, and Kobuk 3 had a private variant C16313T. All three A2b genomes are novel in the published literature.

The human remains from these two sites contribute not only to a greater understanding of Alaskan genetic diversity, but also to the Neo-Eskimo as a distinct Arctic migration event(s) by illuminating past variation in mitochondrial haplogroups. The samples from the precontact site of Iqliqtiqsiugvigruaq corroborate that the Neo-

Eskimo and their descendants, particularly the Iñupiat of northern Alaska, carried high frequencies of haplogroup A2b (Raff et al., 2015; Chapter 2). That the A2b genomes sequenced are unique (Figure 3.3) further supports the Bering Strait geographic region (remnant central Beringia) as the epicenter of A2b diversity. Of archaeological significance, the two adults and one subadult sampled from Iqliqtigsiugvigruaq all seem to carry different maternal lineages. They might be related paternally, but that analysis will need to wait for further genotyping of the Y chromosome (for males) or nuclear markers.

The oldest A2b HVR-I sequences come from the burials at Nuvuk, however, the age of the cemetery and the presence of early derived variants of A2b at Nuvuk suggests that the origin of the A2b haplogroup is elsewhere (Chapter 2). The Nuvuk and Iqliqtigsiugvigruaq results indicate almost certainly that it was not further south. The southwestern Alaskan Thule site at Nunalleq was monomorphic for A2a (Raghavan et al., 2014), and it is unclear if this is representative of modern Yu'pik communities nearby. To the southwest and the southeast of Alaska the genetic landscape is that of the Aleuts and the Na-Dene, respectively; both lack haplogroup A2b (Chapter 2). Other Neo-Eskimo human remains dating to before the 13<sup>th</sup> century Thule expansions (Friesen and Arnold, 2008) are lacking. Needed, for example, are sampled burials associated with the Neo-Eskimo Punuk, a putative alternative ancestor of the Thule (Mason and Bowers, 2009).

The two D4b1a2a1a1 whole genomes sequenced from Nuvuk, with median intercept dates of 1227 AD and 1269 AD, are the oldest known D4b1a2a1a mitochondrial genomes, even including individuals for whom only control region has been published

(Coltrain et al., Under Review; Chapter 2). A single HVR-I sequence from a Scytho-Siberian kurgan burial in the Altai Republic, dated by dendrochronology to the middle of the 5th century B.C (Ricaud et al., 2004), had variants indicating a possible haplotype of D4b1a2a1b, the sister clade of D4b1a2a1a (Figure 3.2). All three samples, however, post-date any calculated coalescence date for the greater D4b1a2a1 haplogroup. While D4b1a2 is hypothesized to have arisen within or immediately following the Last Glacial Maximum in the Altai-Sayan region, D4b1a2a1 only appears to have expanded into northern Siberia in the Holocene (Derenko et al., 2010). What the Nuvuk 1 and Nuvuk 2 genomes confirm is that the still further derived D4b1a2a1a1 (with variants at 16093, 11838, and 14122; Figure 3.2) was the clade that spread into the Americas with the Thule. A genetic bottleneck alone might explain why the Paleo-Eskimo expanded eastward monomorphic for haplogroup D2a. Should ancient human remains be discovered in Asia carrying ancestral D4b1a2a1 or D4b1a2 genomes, their age and location would help to clarify the origins of this clade.

Previous dates from whole mitochondrial sequences for haplogroups A2b and D4b1a2a1a were calculated by Dryomov et al. (2015) from both  $\rho$  statistics, using an ancient DNA calibrated mutation rate (Fu et al., 2013b), and a BEAST (Drummond et al., 2012) analysis. Dates from  $\rho$  statistics have been shown to vary under a range of demographic models (Cox, 2008), and so the current standard for dating is to use either Maximum Likelihood approaches or to run a Bayesian analysis through BEAST. The BEAST analysis of Dryomov et al. (2015) calibrated each of the alignments with an intra-species outgroup of a whole mitochondrial genome from a 39.5 kyr BP modern human from Tianyuan Cave, China genotyped to Asian haplogroup B (Fu et al., 2013a).

While this selection is certainly better than a fossil calibration point like the chimp-human split (see Molak et al., 2013), it is still a single external tip calibration outside of the known haplogroup variation for which dates are being calculated. These typically have a limited influence on rate estimation when compared to internal tip calibrations (Rieux et al., 2014).

We attempted to reanalyze the A2b and D4b1a2a phylogenetic trees using internal tip calibrations of our 5 whole genome sequences and the two previously published Thule individuals from further east. Even with the introduction of 4 dated tips in the A2b tree and 3 dated tips in the D4b1a2a tree, our estimated root dates had large confidence intervals: The coalescence for the D4b1a2a1a1 node had a median of 9,113 yr BP, with a 95% HPD of 952 - 46,481 yr BP (Figure 3.2), and the coalescence for the A2b root had a median of 2,707 yr BP, with a 95% HPD of 269 - 15,431 yr BP (Figure 3.3). In comparison, the median (95% HPD) values of Dryomov et al. (2015) for D4b1a2a1a1 were 1.92 ky BP (0.05; 3.81), and for A2b1 of 2.57 ky BP (0.93; 4.37). While all of our 95% highest posterior densities contained the previous estimates, and our A2b clade estimates were similar, our date ranges are realistically too large to be informative. The incorporation of multiple dated sequence tips in our BEAST analysis was limited by the fact that all of the dated samples were relatively young (from 150 yr BP to 723 yr BP) with respect to the mitochondrial mutation rate. Human mitochondrial sequences of less than 1,000 yrs in age are not expected to be that much different from their modern descendants. Unsurprisingly, none of our samples displayed the expected “branch shortening” typical of older ancient DNA sequences (Fu et al., 2013b; Molak et al., 2013; see Figures 3.2 and 3.3). Accurate estimations of the node dates in these two

haplogroups, from ancient DNA, will require sequences of greater time depth at internal points in the A2b and D4b1a2a phylogenetic trees. While 3 to 4 dated sequences can negatively affect the performance of rate estimation versus using greater numbers of tip dates (Molak et al., 2013), given older sampling tight estimates can be achieved even with small numbers of dated sequences (as demonstrated with the two Upward Sun River burials; Tackney et al., 2015). Our analysis was also limited by the small number of modern sequences available for these two haplogroups - only 25 A2b genomes and 15 D4b1a2a genomes have been previously published.

The DNA from the Nuvuk 3 individual, who has been directly dated to between 1400-1708 AD (Coltrain et al., Under Review), was previously amplified by PCR methods and sequencing results gave conflicting genotypes. The initial PCR-based results suggested membership in haplogroup A2, but the sample lacked the 16265 or 16192 variants defining A2b or A2a, respectively. The hybridization capture of the Nuvuk 3 mitochondrial genome failed, with only about 8% of the genome covered by the Proton sequencing reads (Table 3.4). This DNA extraction lacked sufficient quantity of endogenous human DNA molecules, but further extractions could be attempted from different areas of the bone sample. The identification of an ancient dated A2 root haplogroup in the Arctic would be significant; unfortunately no new information was provided here.

The successful capture and sequencing of these ancient genomes has provided 5 new Neo-Eskimo mitochondrial sequences, where as previously only 4 were available at sufficient depth and coverage. Future work will continue to focus on capturing and sequencing whole mitochondrial genomes, and potentially expanding into nuclear

genomic variation, of older and more diverse human remains around Beringia. We have noted the need for larger sampling of Birnirk individuals, initial sequencing of Punuk individuals, and more ancient Siberian and Alaskan Yupik associated human remains. Anthropological geneticists also need to expand their search for D4b1a2a1a or D4b1a2a1b genotyped modern and ancient individuals in Siberia/Chukotka. The timing and location of any isolation or split from an Asian source for the Beringian haplogroups (A2a, A2b, D2a, and D4b1a2a1a) is still undetermined. The focus here was on the Neo-Eskimo, but the potential of relic genetic signatures of the Paleo-Eskimo in Chukotka and the Aleutian islands has been noted by others (Dryomov et al., 2015) and needs to be explored further.

### **Acknowledgements**

We would like to acknowledge The Barrow Senior Advisory Council, the Ukpeaġvik Iñupiat Corporation, and The Native Village of Barrow for providing permissions to sample from the Nuvuk burials. We would like to also acknowledge Dr. Anne Jensen, the senior Archaeologist at the Nuvuk site. Logistical support while in the field was provided by Glenn Sheehan, Alice Drake, Megan Bowen, Bryan Thomas, Laura Thomas, Lewis Brower, Nok Acker (BASC), Steve Hastings, and Faustine Mercer (CH2MHill Polar Services). The Nuvuk research was supported by an International Polar Year grant: Reconstruction of Human Genetic History Along the North Slope (OPP-0732846 and OPP-0732857), as well as further NSF funding sources: OPP-0637246, OPP-0535356, ARC-0726253, and OPP-0820790. We would like to thank the local community of Kiana and the National Park Service for permissions to sample from Iqliqtqiugvigruaq. We acknowledge Dr. Doug Anderson for his work in excavating

Igliqtiqsiugvigruaq and for initiating this collaboration.

We would like to thank Ellie Fair for performing the DNA extractions on Kobuk 1, 2 and 3, and for amplifying and sequencing the HVR-I fragments from these samples. Sequencing, both Sanger and the Proton runs, was performed at the DNA Sequencing Core Facility, 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. Brendan O'Fallon and Dr. Remco Bouckaert for their support with BEAST 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.

### **Literature Cited**

- Achilli A, Perego UA, Lancioni H, Olivieri A, Gandini F, Hooshiar Kashani B, Battaglia V, Grugni V, Angerhofer N, Rogers MP et al. 2013. Reconciling migration models to the Americas with the variation of North American native mitogenomes. *Proc Natl Acad Sci* 110(35):14308-14313.
- Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A, and Drummond AJ. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comp Biol* 10(4):e1003537.
- Briggs AW, Stenzel U, Johnson PL, Green RE, Kelso J, Prüfer K, Meyer M, Krause J, Ronan MT, and Lachmann M. 2007. Patterns of damage in genomic DNA sequences from a Neandertal. *Proc Natl Acad Sci* 104(37):14616-14621.
- Coltrain J, Tackney J, and O'Rourke DH. Under Review. Thule whaling at Point Barrow, Alaska: The Nuvuk cemetery stable isotope and radiocarbon record. *J Arch Sci Reports*.
- Cox MP. 2008. Accuracy of molecular dating with the rho statistic: deviations from coalescent expectations under a range of demographic models. *Hum Biol* 80(4):335-357.
- Derenko M, Malyarchuk B, Grzybowski T, Denisova G, Rogalla U, Perkova M, Dambueva I, and Zakharov I. 2010. Origin and Post-Glacial Dispersal of Mitochondrial DNA Haplogroups C and D in Northern Asia. *PLoS ONE*

5(12):e15214.

- Drummond AJ, Suchard MA, Xie D, and Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 29(8):1969-1973.
- Dryomov SV, Nazhmidenova AM, Shalaurova SA, Morozov IV, Tabarev AV, Starikovskaya EB, and Sukernik RI. 2015. Mitochondrial genome diversity at the Bering Strait area highlights prehistoric human migrations from Siberia to northern North America. *Eur J Hum Genet* 23(10):1399-404.
- Friesen TM, and Arnold CD. 2008. The Timing of the Thule Migration: New Dates from the Western Canadian Arctic. *Am Antiquity* 73(3):527-538.
- Fu Q, Meyer M, Gao X, Stenzel U, Burbano HA, Kelso J, and Pääbo S. 2013a. DNA analysis of an early modern human from Tianyuan Cave, China. *Proc Natl Acad Sci* 110(6):2223-2227.
- Fu Q, Mittnik A, Johnson PL, Bos K, Lari M, Bollongino R, Sun C, Giemsch L, Schmitz R, and Burger J. 2013b. A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr Biol* 23(7):553-559.
- Gilbert MTP, Kivisild T, Grønnow B, Andersen PK, Metspalu E, Reidla M, Tamm E, Axelsson E, Götherström A, Campos PF et al. 2008. Paleo-Eskimo mtDNA Genome Reveals Matrilineal Discontinuity in Greenland. *Science* 320(5884):1787-1789.
- Helgason A, Pálsson G, Pedersen HS, Angulalik E, Gunnarsdóttir ED, Yngvadóttir B, and Stefánsson K. 2006. mtDNA variation in Inuit populations of Greenland and Canada: migration history and population structure. *Am J Phys Anthropol* 130(1):123-134.
- Jensen A. 2007. Nuvuk Burial: An Early Thule Hunter of High Status. *Alaska J Anthropol* 5(1):119-122.
- Jensen AM. 2009a. Nuvuk, Point Barrow, Alaska: The Thule Cemetery and Ipiutak Occupation [PhD Dissertation]. Bryn Mawr, PA: Bryn Mawr College.
- Jensen AM. 2009b. Radiocarbon dates from recent excavations at Point Barrow, Alaska and their implications for Neoeskimo prehistory. In: Grønnow B, editor. *On the track of the Thule culture, from Bering Strait to east Greenland, Proceedings of the SILA Conference "The Thule Culture - New Perspectives in Inuit Prehistory"*. Copenhagen: Publications from the National Museum Studies in Archaeology and History Vol. 15. p 45-62.
- Jensen AM. 2014. The archaeology of north Alaska: Point Hope in context. In: Hilton CE, Auerbach BM, and Cowgill LW, editors. *The Forages of Point Hope: The Biology and Archaeology of Humans on the Edge of the Alaskan Arctic*. Cambridge: Cambridge University Press. p 11-34.

- Jónsson H, Ginolhac A, Schubert M, Johnson PL, and Orlando L. 2013. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*:btt193.
- Katoh K, and Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol* 30(4):772-780.
- Kloss-Brandstätter A, Pacher D, Schönherr S, Weissensteiner H, Binna R, Specht G, and Kronenberg F. 2011. HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Hum Mutat* 32(1):25-32.
- Krishnan A, Sweeney M, Vasic J, Galbraith D, and Vasic B. 2011. Barcodes for DNA sequencing with guaranteed error correction capability. *Elect Let* 47(4):236-237.
- Lanfear R, Calcott B, Ho SYW, and Guindon S. 2012. PartitionFinder: Combined Selection of Partitioning Schemes and Substitution Models for Phylogenetic Analyses. *Mol Biol Evol* 29(6):1695-1701.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, and Durbin R. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078-2079.
- Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 17(1):pp. 10-12.
- Mason OK. 2009. "The Multiplication of Forms": Bering Strait Harpoon Heads as a Demic and Macroevolutionary Proxy. In: Prentiss AM, Kuijt I, and Chatters JC, editors. *Macroevolution in Human Prehistory*. New York: Springer. p 73-107.
- Mason OK, and Bowers PM. 2009. The Origin of Thule is Always Elsewhere: Early Thule within Kotzebue Sound, Cul de sac or nursery. In: Gronnow B, editor. *On the track of the Thule culture, from Bering Strait to east Greenland, Proceedings of the SILA Conference "The Thule Culture - New Perspectives in Inuit Prehistory"*. Copenhagen: Publications from the National Museum Studies in Archaeology and History Vol. 15. p 25-44.
- Molak M, Lorenzen ED, Shapiro B, and Ho SY. 2013. Phylogenetic estimation of timescales using ancient DNA: the effects of temporal sampling scheme and uncertainty in sample ages. *Mol Biol Evol* 30(2):253-262.
- Raff JA, Rzhetskaya M, Tackney J, and Hayes MG. 2015. Mitochondrial diversity of Iñupiat people from the Alaskan North Slope provides evidence for the origins of the Paleo- and Neo-Eskimo peoples. *Am J Phys Anthropol* 157(4):603-14.
- Raghavan M, DeGiorgio M, Albrechtsen A, Moltke I, Skoglund P, Korneliusen TS, Grønnow B, Appelt M, Gulløv HC, Friesen TM et al. 2014. The genetic prehistory of the New World Arctic. *Science* 345(6200):1255832.

- Ricaut FX, Keyser-Tracqui C, Cammaert L, Crubézy E, and Ludes B. 2004. Genetic analysis and ethnic affinities from two Scytho-Siberian skeletons. *Am J Phys Anthropol* 123(4):351-360.
- Rieux A, Eriksson A, Li M, Sobkowiak B, Weinert LA, Warmuth V, Ruiz-Linares A, Manica A, and Balloux F. 2014. Improved calibration of the human mitochondrial clock using ancient genomes. *Mol Biol Evol* 31(10):2780-2792.
- Rubicz R, Schurr TG, Babb PL, and Crawford MH. 2003. Mitochondrial DNA variation and the origins of the Aleuts. *Hum Biol* 75(6):809-35.
- Saillard J, Forster P, Lynnerup N, Bandelt H-J, and Nørby S. 2000. mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *Am J Hum Genet* 67(3):718-726.
- Schurr TG, Dulik MC, Owings AC, Zhadanov SI, Gaietski JB, Vilar MG, Ramos J, Moss MB, and Natkong F. 2012. Clan, language, and migration history has shaped genetic diversity in Haida and Tlingit populations from Southeast Alaska. *Am J Phys Anthropol* 148(3):422-435.
- Tackney JC, Potter BA, Raff J, Powers M, Watkins WS, Warner D, Reuther JD, Irish JD, and O'Rourke DH. 2015. Two contemporaneous mitogenomes from terminal Pleistocene burials in eastern Beringia. *Proc Natl Acad Sci* 112(45):13833-13838.
- Tackney J, Coltrain JB, Raff J, and O'Rourke DH. 2016. Ancient DNA and Stable Isotopes: Windows on Arctic Prehistory. In: Friesen TM, and Mason OK, editors. *The Oxford Handbook of the Prehistoric Arctic*. New York: Oxford University Press.
- Tamm E, Kivisild T, Reidla M, Metspalu M, Smith DG, Mulligan CJ, Bravi CM, Rickards O, Martinez-Labarga C, and Khusnutdinova EK. 2007. Beringian standstill and spread of Native American founders. *PLoS ONE* 2(9):e829.
- Thorvaldsdóttir H, Robinson JT, and Mesirov JP. 2012. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 14(2):178-92.
- van Oven M, and Kayser M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* 30(2):E386-E394.
- Volodko NV, Starikovskaya EB, Mazunin IO, Eltsov NP, Naidenko PV, Wallace DC, and Sukernik RI. 2008. Mitochondrial Genome Diversity in Arctic Siberians, with Particular Reference to the Evolutionary History of Beringia and Pleistocenic Peopling of the Americas. *Am J Hum Genet* 82(5):1084-1100.
- Zlojutro M, Rubicz R, and Crawford MH. 2009. Mitochondrial DNA and Y-chromosome variation in five eastern Aleut communities: Evidence for genetic substructure in the Aleut population. *Ann Hum Biol* 36(5):511-526.

Table 3.1: Oligonucleotides used in this study

Name	Sequence	Purification	Barcode
Ion_A_bar2 <sup>φ</sup>	C*C*A* <sup>φ</sup> T*CTCATCCCTGCCGTCTCCGACTCAGTGCG*A*A*G*T	HPLC	TGCGAAGT
Ion_A_bar2comp <sup>φ</sup>	A*C*T* <sup>φ</sup> T*CGCACTGAGTCGGAGACACGCAGGATGAGAT*G*G*T*T	HPLC	TGTTGTGCC
Ion_A_bar4	C*CA* <sup>φ</sup> TCTCATCCCT*G*CGTGTCTCCGACTCAGTGTGTGCCGA*T	HPLC	TGAACGAC
Ion_A_bar4comp	A*TCGGCACACACTGAGTCGGAGACACG*C	HPLC	TGATTAAT
Ion_A_bar5	C*CA* <sup>φ</sup> TCTCATCCCT*G*CGTGTCTCCGACTCAGTGAACGACGGA*T	HPLC	TGGAGATT
Ion_A_bar5comp	A*TCGTGTTCACTGAGTCGGAGACACG*C	HPLC	TGGTAGTC
Ion_A_bar6	C*CA* <sup>φ</sup> TCTCATCCCT*G*CGTGTCTCCGACTCAGTGAATTAATGA*T	HPLC	
Ion_A_bar6comp	A*TCATTAATCACTGAGTCGGAGACACG*C	HPLC	
Ion_A_bar7	C*CA* <sup>φ</sup> TCTCATCCCT*G*CGTGTCTCCGACTCAGTGGAGATTGA*T	HPLC	
Ion_A_bar7comp	A*TCAAATCTCCACTGAGTCGGAGACACG*C	HPLC	
Ion_A_bar8	C*CA* <sup>φ</sup> TCTCATCCCT*G*CGTGTCTCCGACTCAGTGGTAGTCGA*T	HPLC	
Ion_A_bar8comp	A*TCGACTACCACTGAGTCGGAGACACG*C	HPLC	
Ion_P1	C*C*A* <sup>φ</sup> C*TACGCCCTCCGCTTTCCTCTCTATGGGCAGTCGG*T*G*A*T	HPLC	
Ion_P1comp	A*T*C*A* <sup>φ</sup> CCGACTGCCCATAGAGAGAAAGCGGAGCCGTAGT*G*G*T*T	Standard Desalting	
Ion_Aamp	CCATCTCATCCCTGCCGTGC	Standard Desalting	
Ion_P1amp	CCACTACGCCCTCCGCTTTCCTCTCTATG	Standard Desalting	
A2_Block <sup>φ</sup>	ACTTCGCACTGAGTCGGAGACACGCAGGATGAGATGGT/3SpC3/	IDT xGen Blocking; HPLC	TGCGAAGT
A4_Block	ATCGGCACACACTGAGTCGGAGACACGCAGGATGAGATGG/3SpC3/	IDT xGen Blocking; HPLC	TGTTGTGCC
A5_Block	ATCGTCGTTCACTGAGTCGGAGACACGCAGGATGAGATGG/3SpC3/	IDT xGen Blocking; HPLC	TGAACGAC
A6_Block	ATCATTAACTCACTGAGTCGGAGACACGCAGGATGAGATGG/3SpC3/	IDT xGen Blocking; HPLC	TGATTAAT
A7_Block	ATCAATCTCCACTGAGTCGGAGACACGCAGGATGAGATGG/3SpC3/	IDT xGen Blocking; HPLC	TGGAGATT
A8_Block	ATCGACTACCACTGAGTCGGAGACACGCAGGATGAGATGG/3SpC3/	IDT xGen Blocking; HPLC	TGGTAGTC
P1_Block_new	ATCACCGACTGCCCATAGAGAGAAAGCGGAGCCGTAGTGG/3SpC3/	IDT xGen Blocking; HPLC	
P1_Block	ATCACCGACTGCCCATAGAGAGAAAGCGGAGCCGTAGTGGT/3SpC3/	IDT xGen Blocking; HPLC	

<sup>φ</sup> Adapters were designed without a 'GAT' barcode adapter sequence following the custom barcode

Table 3.2: NGS library workflow

Sample	Sample ID	Barcode Set	Library Kit	Post-Ligation cleanup			1° cleanup*			2° cleanup*			Library(ng):Probe(ng)	Capture Time <sup>a</sup>	post-Capture 1° cleanup*	Ion PI Chip <sup>β</sup>	Sequencing Kit
				.7X/.7X dSPRI	silica column												
Nuvuk 1	07-J122-YUCXV	2 (TT adapter design)	Ion Plus Fragment	.7X/.7X dSPRI	silica column	850:200	~45 hours	no amplification		200 v3							
Nuvuk 2A	07-C24-42/43	6	Ion Plus Fragment	silica column	silica column	silica column	silica column	silica column	silica column	silica column	silica column	5136:220	~45 hours	1.8X SPRI		200 v3	
Nuvuk 2B	07-C24-42/43	6	Ion Plus Fragment	silica column	silica column	silica column	silica column	1.6X SPRI	1.6X SPRI	1.6X SPRI	1.6X SPRI	2008:250	~70 hours	1.6X SPRI	Shared A	Hi-Q 200	
Nuvuk 2B EB	Extraction Blank	6	Ion Plus Fragment	silica column	silica column	silica column	silica column	1.6X SPRI	1.6X SPRI	1.6X SPRI	1.6X SPRI	2008:250	~70 hours	1.6X SPRI	Shared C	Hi-Q 200	
Nuvuk 3	08-C32-108-YUNFJ	7	Ion Plus Fragment	silica column	silica column	silica column	silica column	1.6X SPRI	1.6X SPRI	1.6X SPRI	1.6X SPRI	736:250	~70 hours	1.6X SPRI	Shared A	Hi-Q 200	
Kobuk 1	K_B1_1	4	NEBNext Fast DNA	1.4X SPRI	1.2X SPRI	1.2X SPRI	1.2X SPRI	1.2X SPRI	1.2X SPRI	1.2X SPRI	1.6X SPRI	540:250	~70 hours	1.6X SPRI	Shared B	Hi-Q 200	
Kobuk 2	K_B2_1	5	NEBNext Fast DNA	1.4X SPRI	1.2X SPRI	1.2X SPRI	1.2X SPRI	1.2X SPRI	1.2X SPRI	1.2X SPRI	1.6X SPRI	860:250	~70 hours	1.6X SPRI		Hi-Q 200	
Kobuk 3	K_B3_1	8	NEBNext Fast DNA	1.4X SPRI	1.2X SPRI	1.2X SPRI	1.2X SPRI	1.2X SPRI	1.2X SPRI	1.2X SPRI	1.6X SPRI	788:250	~70 hours	1.6X SPRI	Shared B	Hi-Q 200	
Kobuk EB	Extraction Blank	4, 5, 8	NEBNext Fast DNA	1.4X SPRI	1.2X SPRI	1.2X SPRI	1.2X SPRI	1.2X SPRI	1.2X SPRI	1.2X SPRI	1.6X SPRI	788:250	~70 hours	1.6X SPRI	Shared C	Hi-Q 200	

\* All primary amplifications with Amplitaq Gold 360 and all secondary/post-capture amplifications with NEB Q5 Hot Start High-Fidelity

<sup>a</sup> All captures done at 60 °C

<sup>β</sup> Samples were sequenced individually per Proton P1 chip, unless otherwise indicated

Table 3.3: Sequencing and bioinformatics pipeline protocol

Sample	Torrent Suite	Barcode Errors in FlowSpace	Min Read Length	trim-qual-cutoff, trim-qual-window-size	cutadapt quality trim
Nuvuk 1	4.0.2	2	30	15,30	20
Nuvuk 2A	4.0.2	2	30	15,30	20
Nuvuk 2B	5.0	1	30	20,20	24
Nuvuk 2B EB	5.0	1	30	20,20	24
Nuvuk 3	5.0	1	30	20,20	24
Kobuk 1	5.0	1	30	20,20	24
Kobuk 2	5.0	1	30	20,20	24
Kobuk 3	5.0	1	30	20,20	24
Kobuk EB	5.0	1	30	20,20	24

Table 3.4: Sequencing and bioinformatics pipeline metrics

Workflow Metric	Novuk 1 LB	Novuk 1 EB	Novuk 2A LB	Novuk 2A EB	Novuk 2A 1B	Novuk 2B	Novuk 2B EB	Novuk 2B & 3 1B	Novuk 2B & 3 EB	Novuk 3	Novuk 3 EB	Novuk 3 1B	Novuk 3 EB						
Post-Enriched library molecules/d	6.06E+08	3.51E+04	7.51E+06	6.13E+06	1.25E+06	8.10E+05	1.09E+04	1.61E+05	2.09E+03	1.61E+05	2.09E+03	4.64E+03	1.91E+06	5.43E+06	2.33E+06	2.24E+06	2.24E+06	2.33E+06	2.24E+06
Amplified post-enriched library molecules/d	-	-	1.28E+10	6.99E+08	1.87E+08	1.13E+09	1.37E+08	4.68E+09	1.29E+07	4.68E+09	1.29E+07	3.52E+07	2.73E+09	2.25E+09	5.04E+09	3.91E+09	3.91E+09	2.73E+09	2.25E+09
Ion PI Final ISPs	81,307,809	-	71,423,411	-	-	86,370,596	75,510,676	86,370,596	-	86,370,596	-	-	61,845,964	36,477,310	61,845,964	75,510,676	75,510,676	61,845,964	36,477,310
Ion PI Total Bases	10.2 G	-	8.5 G	-	-	8.6 G	8.13 G	8.6 G	-	8.6 G	-	-	7.12 G	4.13 G	7.12 G	8.13 G	8.13 G	7.12 G	4.13 G
Barcoded Read Count	7,80E+07	-	7,09E+07	-	-	5,30E+07	3,66E+07	3,27E+07	-	3,27E+07	-	-	3,39E+07	3,59E+07	2,70E+07	3,66E+07	3,66E+07	3,39E+07	3,59E+07
Filtered Reads (cundapp)	7,60E+07	-	6,78E+07	-	-	4,99E+07	3,52E+07	3,15E+07	-	3,15E+07	-	-	3,24E+07	3,48E+07	2,58E+07	3,52E+07	3,52E+07	3,24E+07	3,48E+07
Mapped Reads to rCRS	16,318	-	2,540,222	-	-	1,478,896	2,703	1,477,566	-	1,477,566	-	-	15,776,181	19,669,303	13,322,420	15,598,864	15,598,864	15,776,181	19,669,303
Mapped Reads (Unique, MAPQ≥90)	3,031	-	7,454	-	-	8,440	13	306	-	306	-	-	17,927	23,427	15,598	864	864	17,927	23,427
Coverage	99.85%	-	See merged	-	-	See merged	5.56%	15.894	-	15.894	-	-	99.93%	100.00%	99.91%	32.17%	32.17%	99.93%	100.00%
Bases Missing	25	-	See merged	-	-	See merged	99.72%	47	-	99.72%	-	-	12	0	15	11,239	11,239	12	0
Avg Depth	15.2	-	See merged	-	-	See merged	15.648	72.6	-	72.6	-	-	95.5	129	79.7	12.7	12.7	95.5	129

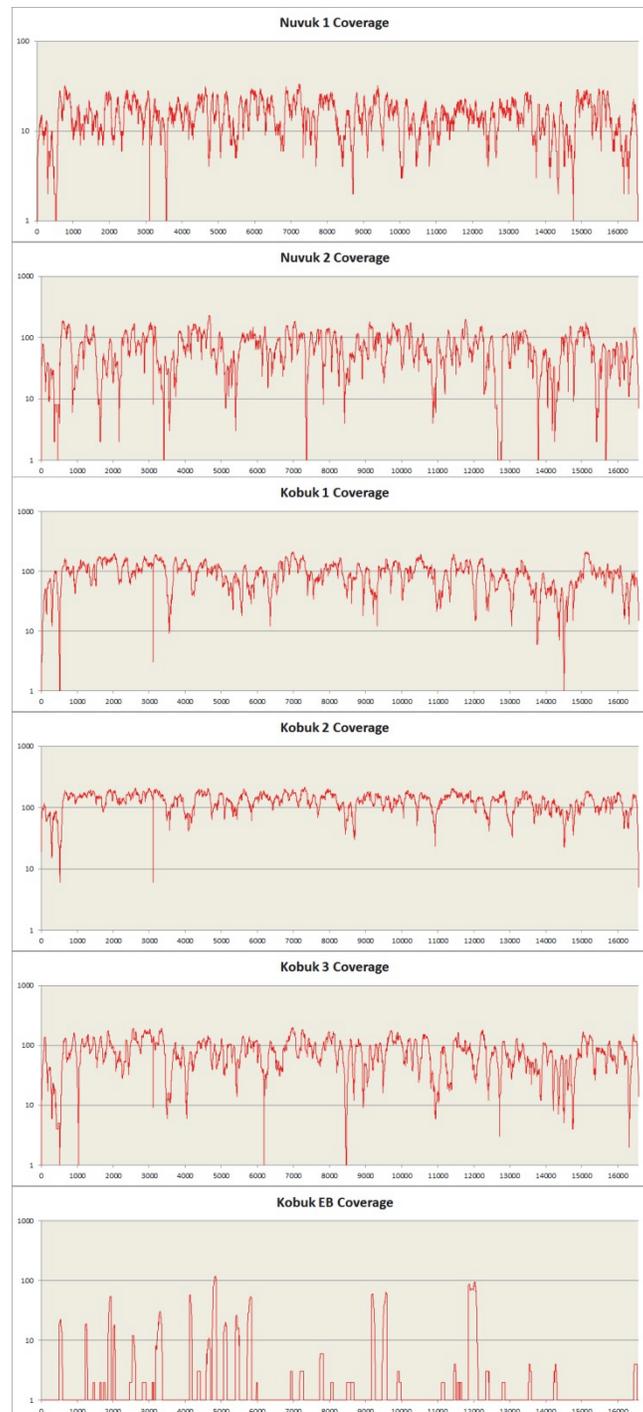


Figure 3.1 Sample library mitochondrial sequence coverage on a 1-base sliding window. Sequence coverage for each labeled captured sample across the mitochondrial genome on a 1-base sliding window, calculated by igvtools.

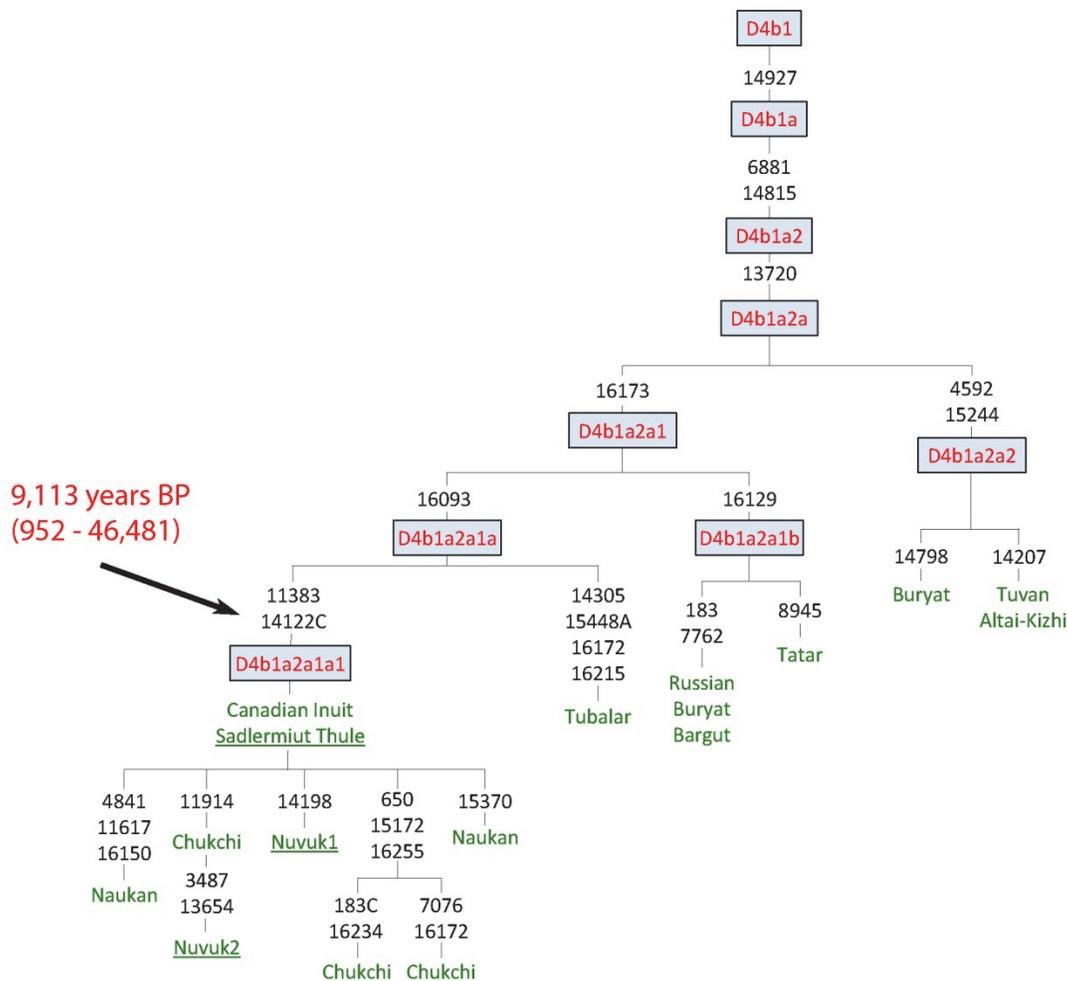


Figure 3.2 Maximum Parsimony phylogenetic tree and BEAST coalescence dates of the D4b1a2a mtDNA haplogroup. The phylogeny of the D4b1a2a haplogroup and various daughter clades. All variants are transitions relative to the rCRS (NC\_012920), except when the base is listed. Back mutations are underlined. Ancient and dated sample names are underlined; these samples provided dated tip sequences for the BEAST analysis. Node dates are displayed as median (95% HPD) from the BEAST analysis.

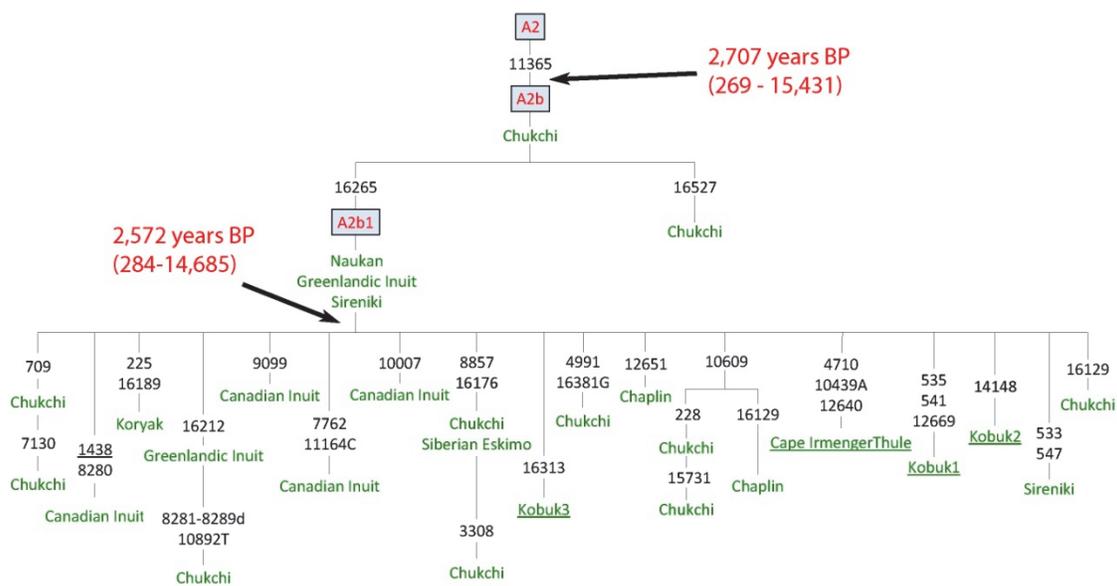


Figure 3.3 Maximum Parsimony phylogenetic tree and BEAST coalescence dates of the A2b mtDNA haplogroup. The phylogeny of the A2b haplogroup. All variants are transitions relative to the rCRS (NC\_012920), except when the base is listed. Back mutations are underlined. Ancient and dated sample names are underlined; these samples provided dated tip sequences for the BEAST analysis. Node dates are displayed as median (95% HPD) from the BEAST analysis.

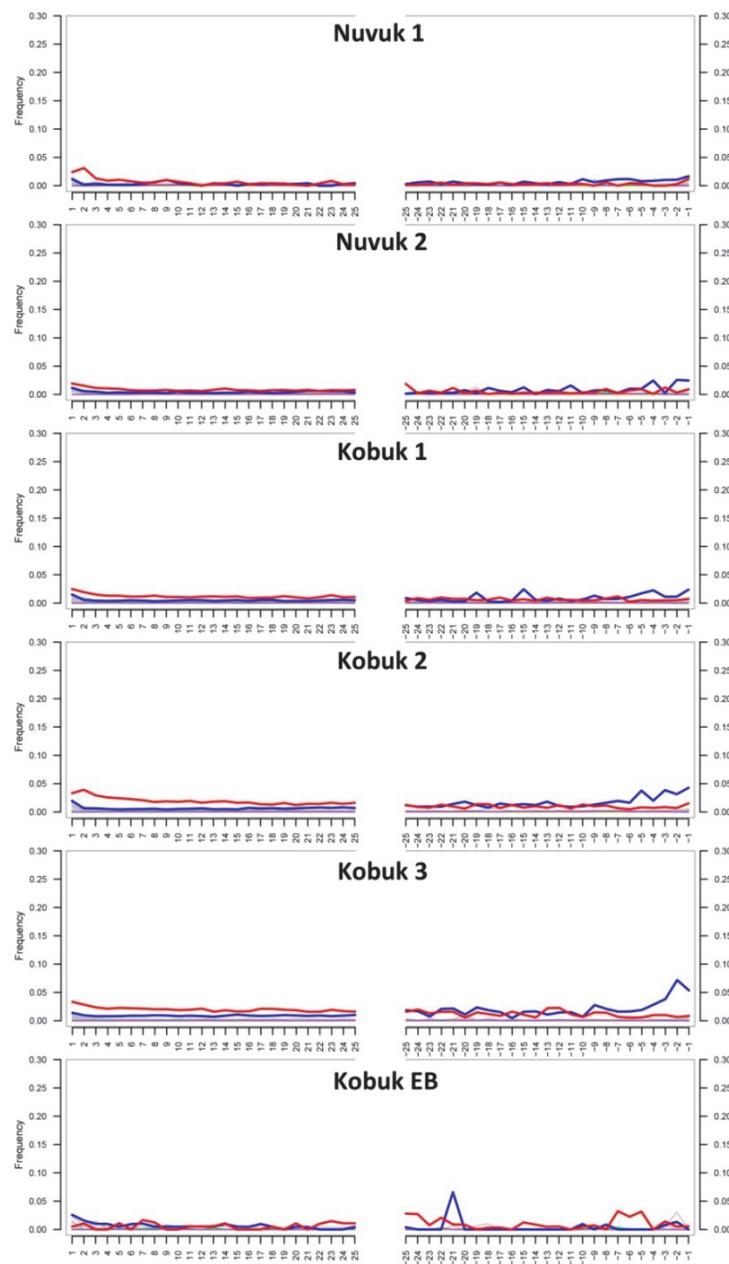


Figure 3.4 Damage patterns. Position specific substitutions from the 5' end of reads (left) and 3' end of reads (right) following the bioinformatics mapping pipeline, for each labeled sample. All graphs produced by MapDamage v2.0.2-12; C-to-T substitutions are shown in red; G-to-A substitutions are shown in blue; insertions are shown in purple; deletions are shown in green; all other substitutions are shown in grey.