

# Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans

J. Víctor Moreno-Mayar<sup>1\*</sup>, Ben A. Potter<sup>2\*</sup>, Lasse Vinner<sup>1\*</sup>, Matthias Steinrücken<sup>3,4</sup>, Simon Rasmussen<sup>5</sup>, Jonathan Terhorst<sup>6,7</sup>, John A. Kamm<sup>6,8</sup>, Anders Albrechtsen<sup>9</sup>, Anna-Sapfo Malaspinas<sup>1,10,11</sup>, Martin Sikora<sup>1</sup>, Joshua D. Reuther<sup>2</sup>, Joel D. Irish<sup>12</sup>, Ripan S. Malhi<sup>13,14</sup>, Ludovic Orlando<sup>1</sup>, Yun S. Song<sup>6,15,16</sup>, Rasmus Nielsen<sup>1,6,17</sup>, David J. Meltzer<sup>1,18</sup> & Eske Willerslev<sup>1,8,19</sup>

**Despite broad agreement that the Americas were initially populated via Beringia, the land bridge that connected far northeast Asia with northwestern North America during the Pleistocene epoch, when and how the peopling of the Americas occurred remains unresolved<sup>1–5</sup>. Analyses of human remains from Late Pleistocene Alaska are important to resolving the timing and dispersal of these populations. The remains of two infants were recovered at Upward Sun River (USR), and have been dated to around 11.5 thousand years ago (ka)<sup>6</sup>. Here, by sequencing the USR1 genome to an average coverage of approximately 17 times, we show that USR1 is most closely related to Native Americans, but falls basal to all previously sequenced contemporary and ancient Native Americans<sup>1,7,8</sup>. As such, USR1 represents a distinct Ancient Beringian population. Using demographic modelling, we infer that the Ancient Beringian population and ancestors of other Native Americans descended from a single founding population that initially split from East Asians around  $36 \pm 1.5$  ka, with gene flow persisting until around  $25 \pm 1.1$  ka. Gene flow from ancient north Eurasians into all Native Americans took place 25–20 ka, with Ancient Beringians branching off around 22–18.1 ka. Our findings support a long-term genetic structure in ancestral Native Americans, consistent with the Beringian ‘standstill model’<sup>9</sup>. We show that the basal northern and southern Native American branches, to which all other Native Americans belong, diverged around 17.5–14.6 ka, and that this probably occurred south of the North American ice sheets. We also show that after 11.5 ka, some of the northern Native American populations received gene flow from a Siberian population most closely related to Koryaks, but not Palaeo-Eskimos<sup>1</sup>, Inuits or Kets<sup>10</sup>, and that Native American gene flow into Inuits was through northern and not southern Native American groups<sup>1</sup>. Our findings further suggest that the far-northern North American presence of northern Native Americans is from a back migration that replaced or absorbed the initial founding population of Ancient Beringians.**

The details of the peopling of the Americas, and particularly the population history of Beringia, remain unresolved<sup>2,3</sup>. Humans were present in the Americas south of the continental ice sheets by around 14.6 ka<sup>11</sup>, indicating that they traversed Beringia earlier. During the Last Glacial Maximum (LGM), this region was marked by harsh climates and glacial barriers<sup>5</sup>, which may have led to the isolation of populations for extended periods, and at times complicated dispersal across the region<sup>12</sup>. It remains unknown whether and for how long

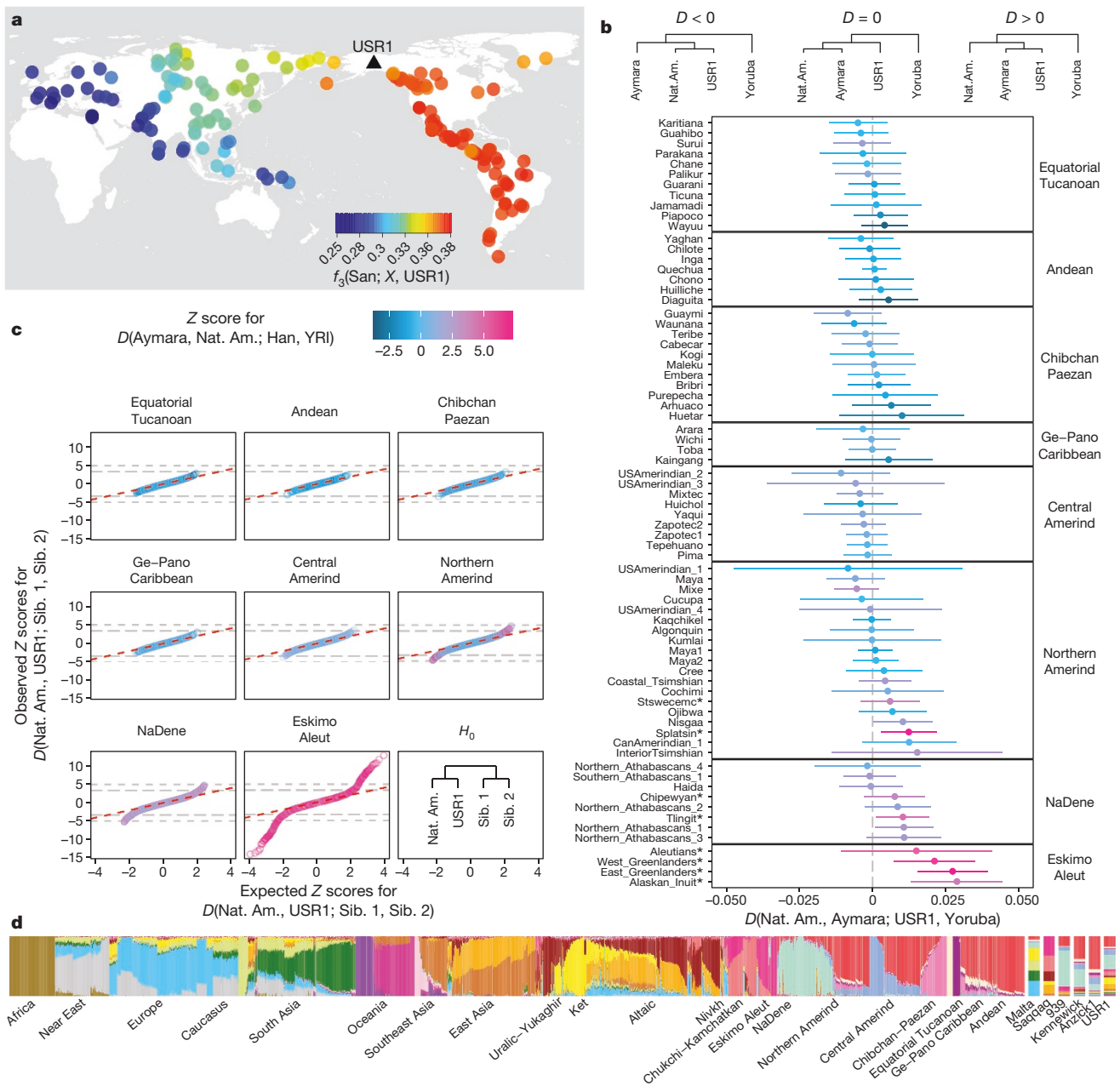
Native American ancestors were isolated from Asian groups in Beringia before entering the Americas<sup>2,9,13</sup>; whether one or more early migrations gave rise to the founding population of Native Americans<sup>1–4,7,14</sup> (it is commonly agreed that the Palaeo-Eskimos and Inuit populations represent separate and later migrations<sup>1,15,16</sup>); and when and where the basal split between southern and northern Native American (SNA and NNA, respectively) branches occurred. It also remains unresolved whether the genetic affinity between some SNA groups and indigenous Australasians<sup>2,3</sup> reflects migration by non-Native Americans<sup>3,4,14</sup>, early population structure within the first Americans<sup>3</sup> or later gene flow<sup>2</sup>. To resolve these uncertainties, a better understanding of the population history of Beringia, the entryway for the Pleistocene peopling of the Americas, is needed.

Genomic insight into that population history has now become available with the recently recovered infant remains (USR1 and USR2) from the Upward Sun River site, Alaska (eastern Beringia), which have been dated to approximately 11.5 ka<sup>6,17</sup>. Mitochondrial DNA sequences (haplogroups C1 and B2, respectively) were previously acquired from these individuals<sup>6,17</sup> (Supplementary Information sections 1, 4.5). We have since obtained whole-genome sequence data, which provide a broader opportunity to investigate the number, source(s) and structure of the initial founding population(s) and the timing and location of their subsequent divergence. We sequenced the genome of USR1 to an average depth of approximately 17×, on the basis of eight sequencing libraries from uracil-specific excision reagent-treated extracts that had previously been confirmed to contain DNA fragments with characteristic ancient DNA misincorporation patterns (Supplementary Information sections 2–4). We estimated modern human contamination to be around 0.14% based on the nuclear genome and about 0.15% based on mitochondrial DNA (Supplementary Information section 4). As expected, the error rate in the uracil-specific excision reagent-treated sequencing data was low (0.09% errors per base), and comparable to other high-coverage contemporary genomes, based on called genotypes (Supplementary Information section 4). Although USR2<sup>6</sup> did not have sufficient endogenous DNA for high-coverage genome sequencing, we found that both individuals were close relatives (Supplementary Information section 5), equally related to worldwide present-day populations (Supplementary Fig. 4g).

We assessed the genetic relationship between USR1, a set of ancient genomes<sup>2,7,8,14,16</sup> and a panel of 167 worldwide populations genotyped for 199,285 single-nucleotide polymorphisms<sup>1,2,18</sup> (Supplementary

<sup>1</sup>Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, 1350 Copenhagen, Denmark. <sup>2</sup>Department of Anthropology, University of Alaska, Fairbanks, Alaska 99775, USA. <sup>3</sup>Department of Biostatistics and Epidemiology, University of Massachusetts, Amherst, Massachusetts 01003, USA. <sup>4</sup>Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637, USA. <sup>5</sup>Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, 2800 Kongens Lyngby, Denmark. <sup>6</sup>Department of Statistics, University of California, Berkeley, California 94720, USA. <sup>7</sup>Department of Statistics, University of Michigan, Ann Arbor, Michigan 48109, USA. <sup>8</sup>Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, UK. <sup>9</sup>The Bioinformatics Centre, Department of Biology, University of Copenhagen, 2200 Copenhagen, Denmark. <sup>10</sup>Department of Computational Biology, University of Lausanne, Lausanne, Switzerland. <sup>11</sup>Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland. <sup>12</sup>Research Centre in Evolutionary Anthropology and Palaeoecology, Liverpool John Moores University, Liverpool L3 3AF, UK. <sup>13</sup>Department of Anthropology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. <sup>14</sup>Carle R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. <sup>15</sup>Computer Science Division, University of California, Berkeley, California 94720, USA. <sup>16</sup>Chan Zuckerberg Biohub, San Francisco, California 94158, USA. <sup>17</sup>Department of Integrative Biology, University of California, Berkeley, California 94720, USA. <sup>18</sup>Department of Anthropology, Southern Methodist University, Dallas, Texas 75275, USA. <sup>19</sup>Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK.

\*These authors contributed equally to this work.

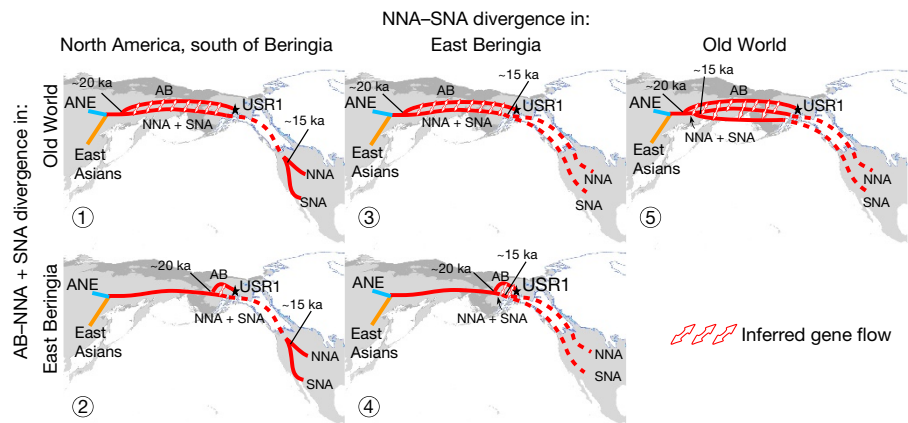


**Figure 1 | Genetic affinities between USR1, present-day Native Americans and world-wide populations.** **a**,  $f_3$  statistics of the form  $f_3(\text{San}; X, \text{USR1})$ , for each population in the genotype panel. Warmer colours represent greater shared drift between a population ( $X$ ) and USR1. **b**,  $D$  statistics of the form  $D(\text{Native American}, \text{Aymara}; \text{USR1}, \text{Yoruba})$  (points). The Andean Aymara were used to represent SNA. \*Native American populations with Asian admixture ( $|Z|$  for  $D(H1, \text{Aymara}; \text{Han}, \text{Yoruba}) > 3.3$ ) (Supplementary Fig. 5a). Error bars represent 1 and approximately 3.3 standard errors ( $P \approx 0.001$ ). Native American populations were grouped by language family<sup>1</sup>. **c**, Quantile–quantile plot comparing observed Z scores to the expected normal distribution under the null hypothesis ( $H_0$ ), for all possible  $D(\text{Native American}, \text{USR1}; \text{Siberian1},$

Siberian2). Colours correspond to the Z score obtained for  $D(H1, \text{Aymara}; \text{Han}, \text{Yoruba})$ . The expected normal distribution under the null hypothesis was computed for all groups jointly (Supplementary Information section 10.4). Thick and thin lines represent a Z score of approximately 3.3 ( $P \approx 0.001$ ) and a Z score of approximately 4.91 ( $P \approx 0.01$  after applying a Bonferroni correction for 11,322 tests). The bottom-right panel shows the expected tree under the null hypothesis. **d**, Admixture proportions estimated by ADMIXTURE<sup>20</sup> assuming  $K = 20$  ancestral populations. Bars represent individuals, and colours represent admixture proportions from each ancestral component. Admixture proportions in ancient genomes (wider bars) were estimated using a genotype likelihood-based approach<sup>21</sup>. Nat. Am., Native American; Sib., Siberian.

Information section 6), using outgroup  $f_3$  statistics<sup>19</sup>, model-based clustering<sup>20,21</sup> and multidimensional scaling<sup>22</sup> (Supplementary Information section 7–9). Outgroup  $f_3$  statistics of the form  $f_3(\text{San}; X, \text{USR1})$  revealed that USR1 is more closely related to present-day Native Americans than to any other tested population, followed by Siberian and East Asian populations<sup>1,2</sup> (Fig. 1a). Pairwise comparisons of the  $f_3$  statistics for USR1 and a set of ancient and contemporary Native American genomes<sup>2,7,14</sup> (Supplementary Information section 6)

showed that all are similarly related to Eurasian, Australasian and African populations, although other Native American genomes (Aymara<sup>2</sup>, Athabaskan<sup>15</sup>, 939<sup>2</sup>, Anzick<sup>17</sup> and Kennewick<sup>14</sup>) have a higher affinity for contemporary Native Americans than does USR1 (Supplementary Information section 9). Multidimensional scaling and ADMIXTURE analysis showed that the USR1 genome did not cluster with any specific Native American group (Fig. 1d and Supplementary Fig. 3b). These results imply that USR1 belonged to a previously



**Figure 2 | Possible geographic locations for the USR1 and NNA-SNA splits.** We propose two possible locations for the split between USR1 and other Native Americans: the Old World (scenarios 1, 3, 5) and Beringia (scenarios 2, 4); and three possible locations for the NNA-SNA split: the Old World (scenario 5), Beringia (scenarios 3, 4), and North America south of Beringia (scenarios 1, 2). Schematics show estimated glacial extent around 14.8 ka. Dashed lines represent the Native American

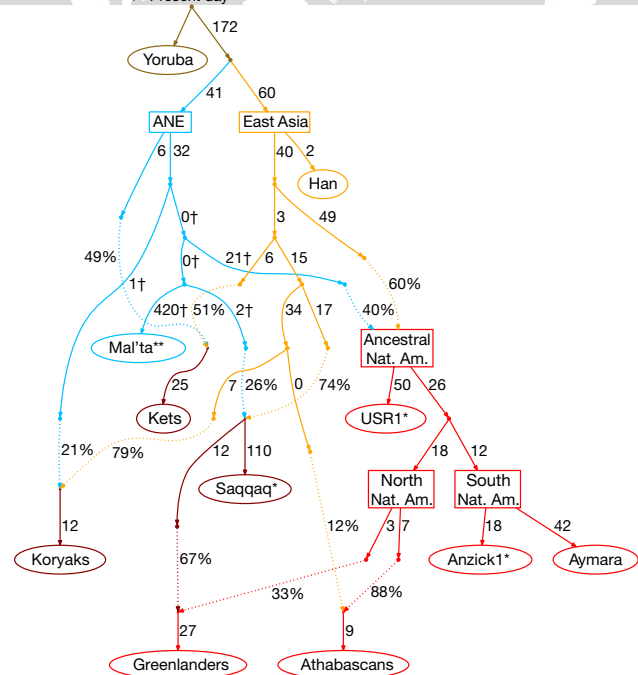
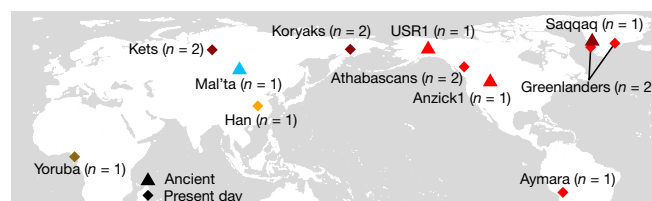
migration south of eastern Beringia, but they do not correspond to a specific migration route. Model discussion (Supplementary Information section 20) is based on extant archaeological evidence and inferred demographic parameters: a USR1-NNA and SNA split about 20 ka with ensuing moderate gene flow and a NNA-SNA split around 15 ka (Supplementary Information sections 18, 19). AB, Ancient Beringian; ANE, Ancient North Eurasian.

unknown Native American population that was not represented in the reference dataset, and which is herein identified as Ancient Beringians (Supplementary Information section 8.3).

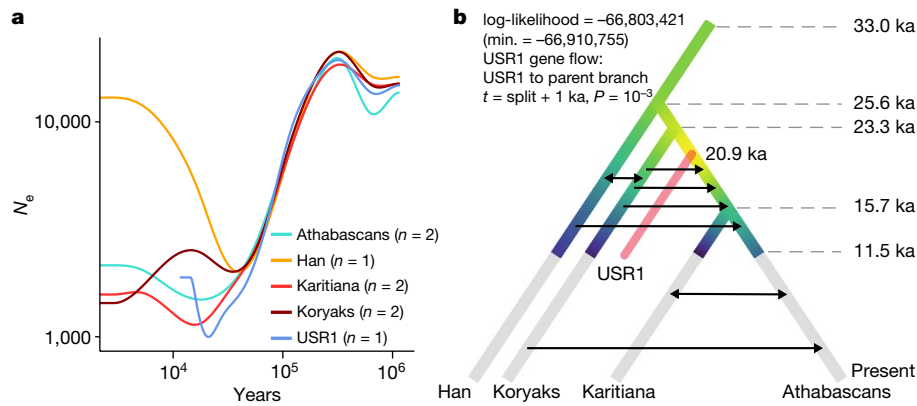
To investigate whether USR1 derived from the same source population that gave rise to contemporary Native Americans, we computed 11,322 allele frequency-based *D*-statistics<sup>1,19</sup> of the form *D*(Native American, USR1; Siberian1/Han, Siberian2/Han) (Supplementary Information section 10.4). The resulting *Z*-score distribution corresponds qualitatively to the expected normal distribution under the null hypothesis that USR1 forms a clade with Native Americans to the exclusion of Siberians and East Asians—except for a set of Eskimo-Aleut, Athabaskan and Northern Amerind-speaking populations for which recent Asian gene flow has previously been documented<sup>1,2,14,18</sup> (Fig. 1c and Supplementary Figs 5a, 6). Additionally, we found that present-day Native Americans and USR1 yield similar results for *D*(Native American/USR1, Han; Mal'ta, Yoruba), suggesting that they are equally related to the ancient north Eurasian population represented by the 24-thousand-year-old Mal'ta individual<sup>8</sup> (Supplementary Information section 10.5). These results confirm that USR1 and present-day Native Americans derived from the same ancestral source, which carried a mixture of East Asian- and Mal'ta-related ancestry. We infer that descendants of this source represent the basal group that first migrated into the Americas.

To explore the relationship between USR1 and present-day Native Americans, we computed allele frequency-based and genome-wide *D* statistics of the form *D*(Native American, Aymara; USR1, Yoruba). We could not reject the null hypothesis that USR1 is an outgroup to any pair of Native Americans, with the exception of a set of populations bearing recent Asian gene flow<sup>1,2,14,18</sup> (Fig. 1b and Supplementary Fig. 7). We confirmed the phylogenetic placement of USR1 at a basal position in the Native American clade using TreeMix<sup>23</sup> and two methods to estimate average genomic divergence and genetic drift, respectively (Supplementary Information sections 14–16). These results support the branching of USR1 within the Native American clade, but with USR1 being equidistant to NNA and SNA. Below we discuss the potential geographic locations of the split between USR1 and the common ancestor of NNA and SNA, and the NNA-SNA split (Fig. 2) on the basis of genetic results, the glacial geography of terminal Pleistocene North America<sup>24,25</sup> and the extant archaeological evidence (Supplementary Information section 20).

Recent detection of an Australasian-derived genetic signature in some Native American groups<sup>2,3</sup> led us to explore whether USR1 also bears this signature (Supplementary Information sections 10.7, 11–13).



**Figure 3 | A model for the formation of the different Native American populations.** We fitted an admixture graph by sequentially adding admixed leaves to a 'seed' graph including the Yoruba, Han, Mal'ta, Ket, USR1, Anzick1 and Aymara genomes. For each 'non-seed' admixed group, we found the pair of edges that produced the best-fitting graph, based on the fitting and maximum  $|Z|$  scores (3.27 for this graph). Ellipse-shaped nodes: sampled populations; box-shaped nodes: metapopulations. \*Single high-depth ancient genome; \*\*single low-depth genome. †Subgraphs with a structure that we were unable to resolve due to sequencing and genotyping error in the Saqqaq genome (Supplementary Information section 17). Sample sizes and locations are shown at the top.



**Figure 4 | USR1 demographic history in the context of East Asians, Siberians and other Native Americans. a**, SMC++-inferred effective population sizes with respect to time for Athabascans (NNA), Karitiana (SNA), Han, Koryaks and USR1 (Supplementary Information section 19.1). We used these demographic histories as a basis for fitting a joint model for these populations. **b**, A ‘backbone demography’ was fitted excluding USR1 using momi2, a maximum likelihood approach based

Using frequency-based and ‘enhanced’  $D$  statistics, we found no support for USR1 being closer to Papuans (a proxy for Australasians) than other Native Americans.

We leveraged the position of USR1 on the Native American branch before the NNA–SNA split to re-assess the origins of Athabaskan and Eskimo populations by fitting admixture graphs. We considered a whole-genome dataset, including Siberian, East Asian, Native American and Eskimo groups, as well as Mal’ta (Supplementary Information section 17). The heuristic approach in TreeMix<sup>23</sup> showed that the best proxies for the Asian component in Athabascans and Greenlandic Inuit are Koryaks and the Saqqaq individual, respectively. We then used an incremental approach to fit an  $f$ -statistic-based admixture graph<sup>19</sup>, including the Kets, which have previously been suggested to share a linguistic and perhaps a genetic link with Athabascans<sup>10,26</sup>. This approach recapitulated the TreeMix results, and yielded a model in which both Athabascans and Greenlandic Inuit derive from the NNA branch. However, the Asian ancestry in Athabascans is most closely related to the Asian component in Koryaks, whereas the Saqqaq genome is the best proxy for the Siberian component in the Greenlandic Inuit (Fig. 3). We infer the latter is a consequence of Palaeo- and Neo-Eskimos having been derived from a similar Siberian population<sup>1,15</sup>. This model appears to be a good fit to the data, as the observed  $f$  statistic that deviated the most from the model prediction yielded  $Z = 3.27$ . We also tested the robustness of this model and these predictions by computing individual  $D$  statistics and by re-fitting the model using alternative datasets (Supplementary Information section 17.3).

Finally, we inferred the demographic history of USR1 with respect to Native Americans, Siberians and East Asians, using two independent methods: diCal2<sup>27</sup> and momi2<sup>28</sup> (Supplementary Information sections 18, 19). diCal2 results indicate that the founding population of USR1, Native Americans and Siberians had a very weak structure from around 36 ka up to about 24.5 ka (Supplementary Table 7), which is when the ancestors of USR1 and Native Americans began to diverge substantially from Siberians. USR1 diverged from other Native Americans around 20.9 ka, with a period of ensuing moderate gene flow between them (Supplementary Tables 6 and 7), as indicated by a simulation study that showed a significant increase in likelihood when comparing a ‘clean split’ model to an ‘isolation with migration’ model (Supplementary Information section 18.4). Using momi2 and SMC++<sup>29</sup>, we estimated a backbone demography in which Karitiana and Athabascans split around 15.7 ka, whereas their ancestral population split from Koryaks about 23.3 ka (Fig. 4). With momi2, we inferred the most likely branch (the population immediately ancestral to NNA and SNA) and time (around 21 ka) for the USR1 population to join the backbone demography,

on a site frequency spectrum (Supplementary Fig. 27), along with the most likely join-on point for USR1 onto the backbone demography (Supplementary Information section 19). We show the likelihood heat map for the latter; warmer colours correspond to a higher likelihood of USR1 joining at a given point. These estimates agree with those obtained using diCal2, a method based on haplotype data (Supplementary Information section 18).

while allowing for possible gene flow between USR and other populations (Fig. 4b and Supplementary Information section 19); results that are consistent with ref. 13 and the diCal2 inference.

These new findings, along with existing data, allow us to place Ancient Beringians within the broader context of the Pleistocene peopling of the Americas. The founding population of Native Americans (consisting of Ancient Beringians and NNA and SNA) began to diverge from ancestral Asians as early as around 36 ka, probably in northeast Asia, as there is no evidence of people in Beringia or northwest North America at this period. A high level of gene flow was maintained between them and other Asians until as late as around 25 ka<sup>2,13</sup>. The subsequent isolation of the Native American founding population about 24 ka roughly corresponds to a decline in archaeological evidence for a human presence in Siberia<sup>30</sup>. Both changes may result from the same underlying cause: the onset of harsh climatic conditions during the LGM<sup>2</sup>. These findings, coupled with a divergence date of around 20.9 ka between USR1 and other Native Americans, are in agreement with the Beringian standstill model<sup>9</sup> (Supplementary Information section 21). Ancient Beringians and the common ancestor of NNA and SNA began to diverge around 20.9 ka, after which gene flow ensued, although whether this only involved the latter or the already differentiated NNA and SNA branches cannot be determined owing to the shallow divergence times among groups.

These findings allow us to consider possible scenarios regarding where ancient Native American populations diverged (Fig. 2 and Supplementary Information sections 20, 21). Scenarios 3–5 require extended periods of strong population structure marking Ancient Beringians, NNA and SNA as separate groups, for which we do not see compelling genetic evidence; these can therefore be rejected. Scenarios 1 and 2 are compatible with our evidence of continuous gene flow among these groups, but differ as to the location of the Ancient Beringians versus NNA and SNA split at 20.9 ka, whether in northeast Asia (scenario 1) or eastern Beringia (scenario 2). Each has strengths and weaknesses relative to genetic and archaeological evidence: scenario 1 best fits the archaeological and palaeoecological evidence, as the earliest securely dated sites in Beringia are no older than around 15–14 ka, and the LGM cold period is unlikely to be associated with northward-expanding populations<sup>30</sup>. Scenario 2 is genetically most parsimonious, given evidence of continuous gene flow between the Ancient Beringians and NNA and SNA, suggesting their geographical proximity 20.9–11.5 ka, and that all three were isolated from Asian and/or Siberian groups after about 24 ka and form a clade.

Scenarios 1 and 2 are both consistent with the NNA–SNA split at around 15 ka<sup>2</sup> having occurred in a region south of eastern Beringia.

The ice sheets were at that time still a substantial barrier to movement that would have helped to maintain separation from the Ancient Beringian population. Although members of the SNA branch were not been documented in regions that were once north of the Pleistocene glaciers<sup>1,18</sup>, NNA groups (including Athabaskan speakers) are present in Alaska today. Therefore, NNA are likely to be descendants of a population that moved north sometime after 11.5 ka<sup>25</sup>.

The USR1 results provide direct genomic evidence that all Native Americans can be traced back to the same source population from a single Late Pleistocene founding event. Descendants of that population were present in eastern Beringia until at least 11.5 ka. By that time, however, a separate branch of Native Americans had already established itself in unglaciated North America, and diverged into the two basal groups that ultimately became the ancestors of most of the indigenous populations of the Americas.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 29 March; accepted 26 November 2017.**

**Published online 3 January 2018.**

- Reich, D. *et al.* Reconstructing Native American population history. *Nature* **488**, 370–374 (2012).
- Raghavan, M. *et al.* Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* **349**, aab3884 (2015).
- Skoglund, P. *et al.* Genetic evidence for two founding populations of the Americas. *Nature* **525**, 104–108 (2015).
- von Cramon-Taubadel, N., Strauss, A. & Hubbe, M. Evolutionary population history of early Paleoamerican cranial morphology. *Sci. Adv.* **3**, e1602289 (2017).
- Hoffecker, J. F., Elias, S. A., O'Rourke, D. H., Scott, G. R. & Bigelow, N. H. Beringia and the global dispersal of modern humans. *Evol. Anthropol.* **25**, 64–78 (2016).
- Potter, B. A., Irish, J. D., Reuther, J. D. & McKinney, H. J. New insights into Eastern Beringian mortuary behavior: a terminal Pleistocene double infant burial at Upward Sun River. *Proc. Natl Acad. Sci. USA* **111**, 17060–17065 (2014).
- Rasmussen, M. *et al.* The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* **506**, 225–229 (2014).
- Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87–91 (2014).
- Tamm, E. *et al.* Beringian standstill and spread of Native American founders. *PLoS ONE* **2**, e829 (2007).
- Flegontov, P. *et al.* Na-Dene populations descend from the Paleo-Eskimo migration into America. Preprint at <https://doi.org/10.1101/074476> (2016).
- Dillehay, T. D. *et al.* Monte Verde: seaweed, food, medicine, and the peopling of South America. *Science* **320**, 784–786 (2008).
- Goebel, T. & Potter, B. A. In *The Oxford Handbook of the Prehistoric Arctic* (eds Friesen, T. M. & Mason, O. K.) 223–252 (Oxford Univ. Press, 2016).
- Llomas, B. *et al.* Ancient mitochondrial DNA provides high-resolution time scale of the peopling of the Americas. *Sci. Adv.* **2**, e1501385 (2016).
- Rasmussen, M. *et al.* The ancestry and affiliations of Kennewick Man. *Nature* **523**, 455–458 (2015).
- Raghavan, M. *et al.* The genetic prehistory of the New World Arctic. *Science* **345**, 1255832 (2014).
- Rasmussen, M. *et al.* Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**, 757–762 (2010).
- Tackney, J. C. *et al.* Two contemporaneous mitogenomes from terminal Pleistocene burials in eastern Beringia. *Proc. Natl Acad. Sci. USA* **112**, 13833–13838 (2015).
- Verdu, P. *et al.* Patterns of admixture and population structure in native populations of northwest North America. *PLoS Genet.* **10**, e1004530 (2014).
- Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
- Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
- Skotte, L., Korneliussen, T. S. & Albrechtsen, A. Estimating individual admixture proportions from next generation sequencing data. *Genetics* **195**, 693–702 (2013).
- Malaspinas, A.-S. *et al.* bammds: a tool for assessing the ancestry of low-depth whole-genome data using multidimensional scaling (MDS). *Bioinformatics* **30**, 2962–2964 (2014).
- Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**, e1002967 (2012).
- Dyke, A. S., Moore, A. & Robertson, L. Deglaciation of North America. (Natural Resources Canada, 2003).
- Pedersen, M. W. *et al.* Postglacial viability and colonization in North America's ice-free corridor. *Nature* **537**, 45–49 (2016).
- Kari, J. M. & Potter, B. A. (eds) *The Dene-Yeniseian Connection* (Univ. Alaska, 2011).
- Steinrücken, M., Kamm, J. A. & Song, Y. S. Inference of complex population histories using whole-genome sequences from multiple populations. Preprint at <https://doi.org/10.1101/026591> (2015).
- Kamm, J. A., Terhorst, J. & Song, Y. S. Efficient computation of the joint sample frequency spectra for multiple populations. *J. Comput. Graph. Stat.* **26**, 182–194 (2017).
- Terhorst, J., Kamm, J. A. & Song, Y. S. Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat. Genet.* **49**, 303–309 (2017).
- Goebel, T. The 'microblade adaptation' and recolonization of Siberia during the late Upper Pleistocene. *Archaeol. Pap. Am. Anthropol. Assoc.* **12**, 117–131 (2002).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** The Upward Sun River excavations and analysis were conducted under a Memorandum of Agreement (MOA) signed by the State of Alaska, the National Science Foundation, the Healy Lake Tribal Council and the Tanana Chiefs Conference. We appreciate the cooperation of all parties. We thank M. Allentoft, S. Gopalakrishnan, T. Korneliussen, P. Librado, J. Ramos-Madrigo, G. Renaud and F. Vieira for discussions, and the Danish National High-throughput Sequencing Centre for assistance with data generation. GeoGenetics members were supported by the Lundbeck Foundation and the Danish National Research Foundation (DNRF94) and KU2016. J.V.M.-M. was supported by Conacyt (Mexico). Samples were recovered during excavations by B.A.P. supported by NSF Grants 1138811 and 1223119. Research was supported in part by NIH grant R01-GM094402 (M.St., J.T., J.A.K. and Y.S.S.) and a Packard Fellowship for Science and Engineering (Y.S.S.). Y.S.S. is a Chan Zuckerberg Biohub investigator. D.J.M. is supported by the Quest Archaeological Research Fund. A.-S.M. is supported by the Swiss National Science Foundation and the ERC.

**Author Contributions** The project was conceived by E.W. and B.A.P. and headed by E.W. L.V. processed ancient DNA. J.V.M.-M. and S.R. assembled datasets. J.V.M.-M., M.St., J.T., J.A.K. and A.A. analysed genetic data. B.A.P. led the USR field investigation and B.A.P. and D.J.M. provided anthropological contextualization. B.A.P., J.D.R. and J.D.I. conducted archaeological and bioanthropological work. R.N., Y.S.S., M.Si., A.-S.M., and L.O. supervised bioinformatic and statistical analyses. B.A.P. engaged with indigenous communities. J.V.M.-M., B.A.P., D.J.M. and E.W. wrote the manuscript with input from L.V., A.-S.M., M.Si., R.S.M., L.O., Y.S.S., R.N. and the other authors.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to E.W. ([ewillerslev@snm.ku.dk](mailto:ewillerslev@snm.ku.dk)).

## METHODS

**Laboratory procedures.** Ancient DNA work was conducted in dedicated clean laboratory facilities at the Centre for GeoGenetics, Natural History Museum, University of Copenhagen. We prepared bone powder from remains of the pars petrosa of both USR individuals and extracted DNA following previously published protocols<sup>31</sup>. Double-stranded dual-indexed Illumina libraries were built from uracil-specific excision reagent (USER)- and non-USER-treated extracts and were paired-end sequenced ( $2 \times 75$  bp) on Illumina HiSeq 2500 instruments (Supplementary Information section 2).

**Sequence data processing.** Raw reads were trimmed for Illumina adaptor sequences and overlapping pairs were collapsed into single reads using AdapterRemoval<sup>32</sup>. Collapsed reads were mapped to the human reference genome build 37 using BWA v.0.6.2-r126<sup>33</sup>; seeding ( $-I$  parameter) was disabled in order to prevent 5' terminal substitutions characteristic of ancient DNA to bias the mapping<sup>34</sup>. Reads with mapping quality lower than 30 were discarded, PCR duplicates were removed using MarkDuplicates (<http://picard.sourceforge.net>) and local realignment was performed using GATK<sup>35</sup>. We called USR1 genotypes using SAMtools mpileup<sup>36</sup> and applied the standard filters described in ref. 2. Called genotypes were phased with shapeit2-r727<sup>37</sup> using the 1,000 genomes phased variant panel (phase 3) as a reference and the HapMap recombination rates as a proxy for the genetic map of the human genome. Sites not included in the 1,000 genomes reference panel were kept as 'unphased' genotypes. Finally, we masked the dataset using a 35-mer 'snppability' mask with a stringency of 0.5 (<http://lh3lh3.users.sourceforge.net/snppable.shtml>) (Supplementary Information section 3).

**Ancient DNA data authentication.** We assessed the authenticity of the ancient DNA data by examining the fragment length distributions and the base substitution patterns across non-USER-treated reads using bamdamage<sup>22</sup>. We estimated mtDNA contamination using contamMix<sup>38</sup> on the basis of a majority rule mtDNA consensus sequence and an alignment of 311 worldwide mtDNA sequences<sup>39</sup>. Nuclear contamination was estimated using the two-population model implemented in DICE<sup>40</sup>, for which we used the 1,000 Genomes Project 'CEU' population as the putative contaminant and the 'YRI' population as the 'anchor'. Sequencing and genotyping error rates relative to a 'high-quality' sample were obtained following the method described in ref. 41 (Supplementary Information section 4).

**Relatedness between USR individuals.** We explored the familial relationship between both USR individuals by using NGSrelate<sup>42</sup> and relate<sup>43</sup>. Given the unavailability of allele frequency data for the Ancient Beringian population, we used allele frequencies from the 1,000 Genomes Project 'PEL' population as a proxy, which limited the resolution of these analyses (Supplementary Information section 5).

**Reference datasets.** We compared the genomes of the USR individuals to a set of 49 worldwide contemporary and ancient genomes and a SNP array dataset comprising 2,537 contemporary individuals from 167 ethnic groups (enriched in Native Americans), genotyped across 199,285 SNP sites. For the latter, European and African ancestry tracts were masked in Native American individuals (Supplementary Information 6).

**Population structure analyses.** We investigated the relationship between USR1, a set of ancient genomes and the SNP array reference dataset using multidimensional scaling as implemented in bammms<sup>22</sup>. Additionally, we explored the genetic ancestry components in the reference panel using ADMIXTURE<sup>20</sup>. We obtained the most likely ancestry proportions in the ancient genomes on the basis of allele frequencies inferred by ADMIXTURE, through the genotype likelihood-based optimization method described in ref. 21 (Supplementary Information sections 7, 8).

**$f_3$  statistics.** We computed  $f_3$  statistics to measure the shared drift between by two particular populations or genomes, and used 'basic' and 'enhanced'  $D$  statistics to formally test hypotheses of treeness and gene-flow. We used admixtools<sup>19</sup> for allele-frequency-based tests and ANGSD<sup>44</sup> for single genome tests. For both tools, standard errors were estimated through a weighted block jackknife approach over approximately 5-Mb blocks (Supplementary Information sections 9–13).

**Admixture graph fitting using TreeMix.** We used the heuristic approach in TreeMix<sup>23</sup> to assess the phylogenetic placement of USR1 in the broader context of Eurasian and Native American populations and to explore the origin of the Na-Dene and Inuit (see 'Admixture graph fitting using qpGraph'). We restricted the analysis to transversion sites where all considered populations have at least one individual with a non-missing genotype call. We grouped the resulting number of SNPs into approximately 5-Mb blocks to account for linkage disequilibrium, and for each number of migrations we ran 1,000 replicates with random seeds and kept the run with the highest likelihood. We estimated the support for internal nodes and migration edges through a bootstrap procedure (Supplementary Information sections 14, 17).

**Pairwise branch lengths and genomic divergence.** We used the method from ref. 7 to measure the amount of drift leading to different pairs of genomes after their split. We restricted this analysis to sites that are variable in five African genomes and obtained the counts for each of the five possible genotype configurations between a given pair of genomes, after which we used numerical optimization to infer maximum likelihood parameters (Supplementary Information section 15). We computed the average DNA divergence between pairs of genomes using the triangulation method from ref. 45, and estimated standard errors using a weighted block jackknife approach over 5-Mb blocks (Supplementary Information section 16).

**Admixture graph fitting using qpGraph.** We used a two-step approach to assess the origin of the Na-Dene and Inuit. First, we found the most likely Eurasian ancestry sources for these groups by using TreeMix. We then fitted  $f$ -statistics-based admixture graphs<sup>1,19</sup> incrementally, such that for each new 'admixed leaf' we enumerated all possible pairs of edges using ref. 46 and kept the admixture event that produced the graph with the best maximum  $|Z|$  and fitting scores. We assessed the robustness of this model and its predictions using pooled  $D$  statistics and by fitting the model using alternative datasets (Supplementary Information section 17).

**Demographic inference using the sequentially Markov coalescent.** We used diCal2<sup>27</sup> to estimate the key demographic parameters relating pairs of genomes including USR1 (sample dated to 11.5 ka) and a set of present-day Asian and Native American genomes. We analysed these pairs under different models, including a clean split, isolation with migration until the present, isolation with migration with a stopping time and isolation with migration with a stopping time and a second contact. We tested competing models through a simulation study and obtained confidence intervals for the inferred parameters through a parametric bootstrap strategy (Supplementary Information section 18).

**Demographic inference using the site frequency spectrum.** We used a combination of SMC++<sup>29</sup> and momi2<sup>28</sup> to infer demographic parameters for USR1 and a set of present-day genomes. We estimated the marginal sizes over time for each population using SMC++. We used these demographic histories as a basis for fitting a joint 'backbone demography' for the present-day populations using momi2. We then inferred the most likely join-on point for USR1 onto the backbone demography using momi2. Confidence intervals were obtained through a parametric bootstrap strategy (Supplementary Information section 19).

**Data availability.** Sequence data were deposited in the ENA under accession: PRJEB20398.

- Allentoft, M. E. *et al.* Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–172 (2015).
- Lindgreen, S. AdapterRemoval: easy cleaning of next-generation sequencing reads. *BMC Res. Notes* **5**, 337 (2012).
- Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
- Schubert, M. *et al.* Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics* **13**, 178 (2012).
- DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
- Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
- Delaneau, O., Zagury, J.-F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* **10**, 5–6 (2013).
- Fu, Q. *et al.* A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* **23**, 553–559 (2013).
- Green, R. E. *et al.* A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* **134**, 416–426 (2008).
- Racimo, F., Renaud, G. & Slatkin, M. Joint estimation of contamination, error and demography for nuclear DNA from ancient humans. *PLoS Genet.* **12**, e1005972 (2016).
- Orlando, L. *et al.* Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* **499**, 74–78 (2013).
- Korneliussen, T. S. & Moltke, I. NgsRelate: a software tool for estimating pairwise relatedness from next-generation sequencing data. *Bioinformatics* **31**, 4009–4011 (2015).
- Albrechtsen, A. *et al.* Relatedness mapping and tracts of relatedness for genome-wide data in the presence of linkage disequilibrium. *Genet. Epidemiol.* **33**, 266–274 (2009).
- Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* **15**, 356 (2014).
- Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010).
- Leppälä, K., Nielsen, S. V. & Mailund, T. admixturegraph: an R package for admixture graph manipulation and fitting. *Bioinformatics* **33**, 1738–1740 (2017).

## Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see [Reporting Life Sciences Research](#). For further information on Nature Research policies, including our [data availability policy](#), see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### ▶ Experimental design

#### 1. Sample size

Describe how sample size was determined.

No statistical methods were used to predetermine sample sizes. Sample size in this study was limited by the fact that to date, USR1 is the only human skeleton from Terminal Pleistocene East Beringia that has allowed for high coverage genome sequencing.

#### 2. Data exclusions

Describe any data exclusions.

We sequenced libraries from both UDG- and non-UDG-treated extracts. Yet, in order to minimize the bias introduced by post-mortem DNA damage, genotype calls were based only on UDG-treated libraries. Note however, that all UDG-treated libraries come from extracts from which we were able to identify aDNA-characteristic damage patterns, by sequencing non-UDG-treated libraries (SI Section 3).

#### 3. Replication

Describe whether the experimental findings were reliably reproduced.

No specific attempts were made to replicate the experimental findings. While the USR1 remains represent the only Terminal Pleistocene East Beringia sample suited for whole genome sequence to date, the USR1 genome derives from different extracts and libraries, all of which carry similar population genetic signatures. Moreover, this study relies on genomic data which implies that each site in the genome represents an 'independent' realization of the USR1 evolutionary history. These provide 'internal' replicates.

#### 4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

Samples were not randomized for the experiments. Only one ancient genome (USR1) is presented in this study.

#### 5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

No blinding techniques were implemented. Group allocation is not relevant to population history studies of this kind.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

## 6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

- |                                     |  |
|-------------------------------------|--|
| n/a                                 | Confirmed  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> The <u>exact sample size</u> ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A statement indicating how many times each experiment was replicated   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section) |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as an adjustment for multiple comparisons  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The test results (e.g. $P$ values) given as exact values whenever possible and with confidence intervals noted   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A clear description of statistics including <u>central tendency</u> (e.g. median, mean) and <u>variation</u> (e.g. standard deviation, interquartile range)  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> Clearly defined error bars   |

See the web collection on [statistics for biologists](#) for further resources and guidance.

## ► Software

Policy information about [availability of computer code](#)

### 7. Software

Describe the software used to analyze the data in this study.

Computer programs and software used in this work are publicly available and they are referred to their respective publications in the main text and supplementary material. List of software and respective versions:

CASAVA v1.8.2  
 AdapterRemoval v1.5.3  
 bwa v0.6.2-r126  
 picard tools v2.7.1  
 samtools v1.3.1  
 GATK v3.4-0-3.3  
 shapeit2-r727  
 bammds v20140402  
 ANGSD v0.915  
 contamMix v1.0-5  
 HAPLOFIND  
 DICE  
 NGSrelate  
 relate v0.9993  
 plink v1.90b3.42  
 ADMIXTURE v1.23  
 admixtools v4.1  
 TreeMix v. 1.12  
 R 3.2.3  
 python 2.7.12  
 perl v5.22.1  
 admixturegraph v1.0.2 (R package)  
 dical2  
 momi2

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). [Nature Methods guidance for providing algorithms and software for publication](#) provides further information on this topic.

## ► Materials and reagents

Policy information about [availability of materials](#)

### 8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

No unique materials were used. Materials and Reagents are detailed in Supplementary Section 2. Laboratory Procedures



## 9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

NA

## 10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

NA

b. Describe the method of cell line authentication used.

NA

c. Report whether the cell lines were tested for mycoplasma contamination.

NA

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by [ICLAC](#), provide a scientific rationale for their use.

NA

## ► Animals and human research participants

---

Policy information about [studies involving animals](#); when reporting animal research, follow the [ARRIVE guidelines](#)

### 11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

NA

Policy information about [studies involving human research participants](#)

### 12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

NA