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[1,19] 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, Athabascan and Northern Amerind-speaking populations for which recent Asian gene flow has previously been documented[1,2,14,18] (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[8] (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[1,2,14,18] (Fig. 1b and Supplementary Fig. 7). We confirmed the phylogenetic placement of USR1 at a basal position in the Native American clade using TreeMix[23] 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[24,25] and the extant archaeological evidence (Supplementary Information section 20).

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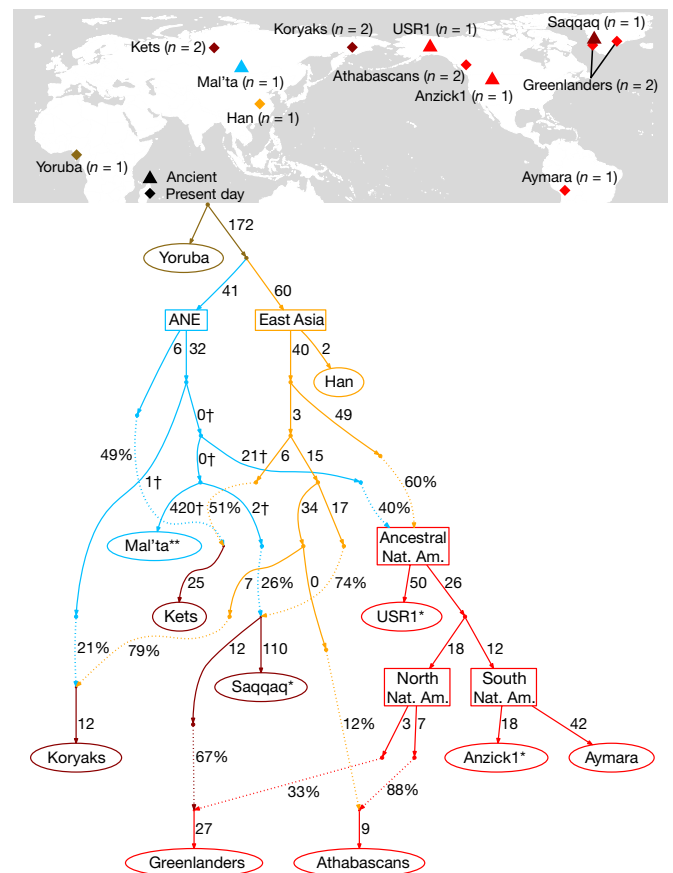

Figure 3 | A model for the formation of the different Native American populations. We fitted an admixture graph by sequentially adding admixed leaves to a 'seed' graph including the Yoruba, Han, Mal'ta, Ket, USR1, Anzick1 and Aymara genomes. For each 'non-seed' admixed group, we found the pair of edges 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 ancient 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.