Early human dispersals within the Americas

J. Víctor Moreno-Mayar†, Lasse Vinner†, Peter de Barros Damgaard‡, Constanza de la Fuente*, Jeffrey Chan*, Jeffrey P. Spence*, Morten E. Allentoft†, Tharsika Vimala†, Fernando Racimo†, Thomaz Pinotti†, Simon Rasmussen§, Ashot MargaryanⅥ, Miren Iraeta Orbegozo†, Dorothea Mylopotamitaki†, Matthew Wooller‡, Clement Bataille*, Lorena Becerra-Valdivia*, David Chivall‡, Daniel Comeskey§, Thibaut Devièse†, Donald K. Grayson†, Len George†, Harold Harry†, Verner Alexandersen†, Charlotte Primeau†, Jon Erlandsen†, Claudia Rodrígues-Carvalho†, Silvia Reis†, Murilo Q. R. Bastos†, Jerome CybulskiⅫ,ⅩⅢ, Charles Vullo†, Flavia Morello†, Miguel Vilar†, Spencer Wells‡, Kristian Gregersen†, Kasper Lykke Hansen†, Niels Lynnerup†, Marta Mirazón Lahr‡, Kurt Kjær†, André Strauss‡,‡, Marta Alfonso-Durruty‡, Antonio Salas‡,‡, Hannes Schroeder†, Thomas Higham†, Ripan S. Malhi†, Jeffrey T. Rasicvos, Luiz Souza‡, Fabricio R. Santos‡, Anna-Sapfo Malaspina‡, Martin Sikora‡, Rasmus NielsenⅠ,‖, Yun S. SongⅡ,Ⅴ,Ⅴ†, David J. MeltzerⅠ,‖, Eske WillerslevⅠ,‖,ⅩⅣ,ⅩⅣ

*These authors contributed equally to this work.
†Corresponding author. Email: ewillerslev@snm.ku.dk (E.W.); dmeltzer@smu.edu (D.J.M.); yss@berkeley.edu (Y.S.S.)

Studies of the peopling of the Americas have focused on the timing and number of initial migrations. Less attention has been paid to the subsequent spread of people within the Americas. We sequenced 15 ancient genetic signal, and a later Mesoamerican-related expansion. These led to complex and dynamic population histories from North to South America.

Previous genomic studies have estimated that ancestral Native Americans (NA) diverged from Siberian and East Asian populations around 25,000 ± 1100 years ago (25 ± 1.1 ka) (1, 2), followed by a split 22-18.1 ka within that ancestral lineage between later NA and Ancient Beringians (AB). Subsequently, NA diverged into two branches, Northern Native Americans (NNA) and Southern Native Americans (SNA) ~17.5-14.6 ka (2–4), a process inferred to have taken...
place south of eastern Beringia (present day Alaska and western Yukon Territory). All contemporary and ancient NA individuals for whom genome-wide data have been generated prior to this study, derive from either the NNA or SNA branch.

However, there is disagreement over claims of earlier migrations into the Americas, possibly related to Australasians or by bearers of a distinctive cranial form (“Paleoamericans”) (5, 6). Whether there were additional splits within the Americas, how many migratory movements north and south took place, and the speed of human dispersal at different times and regions, are contentious. In contrast to models based on contemporary and Pleistocene age genetic data (3, 4), genomic studies of later Holocene human remains indicate there was post-divergence admixture between basal Native American groups (7). Overall, the degree of population isolation, admixture, or continuity in different geographic regions of the Americas after initial settlement is poorly understood (7–9).

Genome sequences from Late Pleistocene/Early Holocene are rare. If we are to resolve how the peopling process occurred, more are needed than the three currently available: Anzick1, Montana (~12.8 ka) (3); Kennewick Man/Ancient One, Washington (~9 ka) (10), and USR1, Alaska (~11.5 ka) (7).

**Dataset and method summary**

We engaged and sought feedback from Indigenous groups linked to the ancestral individuals analyzed in this study using the recommendations for genomics research with Indigenous communities (11–13). We obtained genome sequences from 15 ancient human remains (Fig. 1A). These include Trail Creek Cave 2, Alaska (radiocarbon dated to ~9 ka, ~0.4× genomic depth of coverage); Big Bar Lake, British Columbia (~5.6 ka, ~1.2×); Spirit Cave, Nevada (~10.7 ka, ~18×); four individuals from Lovelock Cave, Nevada (ranging in age from ~1.95 to 0.6 ka, and ~0.5× to ~18.7×); five individuals from Lagoa Santa, Brazil (~10.4 to ~9.8 ka, ~0.18× to ~15.5×); two individuals from the Punta Santa Ana and Ayayema sites in Patagonian Chile (~7.2 ka and ~5.1 ka, ~1.5× and ~10.6×, respectively); and, an Inca mummy from Mendoza, Argentina (estimated to be ~0.5 ka, ~2.5×) (14) [all ^13C ages are in calibrated years (13, 15)] (Fig. 1, A and B). We also sequenced a ~15.9× genome from a ~19th century Andaman islander, used as a proxy for Australasian ancestry in models involving admixture into Native Americans (2, 6, 13, 16). All DNA extracts were confirmed to contain fragments with characteristic ancient DNA misincorporation patterns and low contamination levels (<3%) (13). The Spirit Cave, Lovelock 2, and Lovelock 3 genomes were generated solely from USER-treated extracts, confirmed to contain characteristic ancient DNA misincorporation patterns prior to treatment (13, 17).

To assess the genetic relationships between these and other ancient and contemporary human genomes, we compiled a whole-genome comparative dataset of 378 individuals (13). Additionally, we merged these data with a SNP panel of 167 worldwide populations genotyped for 199,285 SNPs, enriched in NA populations whose European and African ancestry components have been identified and masked (2, 4, 13, 18) (Fig. 1A). Of particular interest are the Mixe, a Mesoamerican reference group representing an early internal branch within SNA, prior to the divergence of South Americans (4), and that lacks the Australasian ancestry signal documented among some Amazonian groups (2, 6, 16).

We explored the ancient individuals’ broad genetic affinities initially using model-based clustering (19) and multidimensional scaling (MDS) (Fig. 1, C to F (13)). MDS was applied to both the identity-by-state distance matrix for all individuals (20) and the f2-distance matrix over populations included in the SNP array dataset (13, 21, 22). We then tested specific hypotheses by computing error-corrected and genotype-based D-statistics (13, 21, 23) and fitting admixture graphs (4, 13, 21, 24) (Figs. 2, 3, and 4). Furthermore, we inferred demographic and temporal parameters using the joint site frequency spectrum (SFS) (25, 26) and linkage disequilibrium (27, 28) information (Fig. 5). These efforts enabled us to explore finer scale complex models, using whole-genome data (13). Average depth of coverage of the genomes presented in this study ranges widely, which meant that not all genomes could be used in all analyses, as specified (13).

Our aim is to understand broad patterns in the dispersal, divergence and admixture of people throughout the Americas. Given the highly uneven distribution of genome samples in time and space, our results are expressed as much as possible chronologically from oldest to youngest, and geographically from north to south, to mirror how the peopling of the Americas proceeded.

**Alaskan genome informs on early eastern Beringian populations**

While the earliest archaeological evidence for a human presence in eastern Beringia remains disputed, people were present in Alaska by at least 14.4 ka (29). Genomic insights from the USR1 genome indicate that AB (1) remained isolated in interior Alaska until at least the terminal Pleistocene, and was an outgroup to NNA and SNA. It was inferred the NNA/SNA population split occurred outside of eastern Beringia (1, 2). In contrast, recent findings suggest the ancestral population of NNA may have existed north of the continental ice sheets (9).

The Trail Creek Cave genome is from a tooth of a young child recovered from Alaska’s Seward Peninsula (13). This individual clusters adjacent to USR1 in MDS analyses (Fig. 1C) (13) and carries a similar distribution of ancestry components (Fig. 1F) (13). Trail Creek and USR2 (found with and a close relative to USR1), harbor the same mtDNA haplogroup, B2,
but not the derived B2 variant found elsewhere in the Americas (1, 13). Genotype-based D-statistics of the form D(Aymara, NA; TrailCreek, Yoruba) and D(USR1, TrailCreek; NA, Yoruba) suggest that Trail Creek forms a clade with USR1 that represents an outgroup to other Native Americans (13). This placement was supported by fitting f-statistic-based admixture graphs (13, 21).

This procedure used here and for other samples, relies on a “seed graph” that incorporates the formation of ancestral NA and their three basal branches (AB, NNA, SNA) (1, 30, 31). The seed graph includes the following leaves: Yoruba (representing Africans), Mal’ta (Ancient North Eurasians), Andaman (Australasians), Han (East Asians), USR1 (AB), Athabaskan (NNA) and Spirit Cave (SNA, see below) (13). We then enumerated all possible extensions of the seed graph where an individual genome, Trail Creek in this case, was added as either a non-admixed or an admixed population (32). We optimized the parameters for each topology using qpGraph (21) and favored the graph producing the best likelihood and the lowest residuals between observed and predicted f-statistics. Given that admixed models yield better likelihood scores (because of the additional parameters being optimized), we considered an admixed model to be an improvement compared to its non-admixed counterpart only if the absolute difference between fit scores (log-likelihoods) was >~4.6, corresponding to a p-value of <0.01 in a standard likelihood ratio test (30). In agreement with the exploratory analyses, we found the model in which Trail Creek and USR1 form a clade to be the most likely (Fig. 2A) (13).

These results suggest that USR1 and Trail Creek were members of an AB metapopulation that occupied eastern Beringia and remained isolated from other NA populations during the Late Pleistocene/Early Holocene. Finding two members of the AB population, from sites ~750 km apart, with a similar artifact technology (13), supports the inference that the SNA/NNA split occurred south of eastern Beringia (1, 9). The alternative, that NNA/SNA split in Alaska, seems less likely; it would have required several thousand years of strong population structure prior to ~16 ka to differentiate those groups from each other and from AB, as well as a separate SNA presence, which has yet to be found (1). These data indicate that the Athabascans and Inuit, who inhabit Alaska today and are NNA, but with additional Siberian-related ancestry (1, 4, 18, 33), presumably moved north into the region sometime after ~9 ka, the age of the Trail Creek individual (1, 13).

Rapid dispersal of the SNA population across the Rockies and into South America

The NNA/SNA split is estimated to have taken place ~17.5-14.6 ka (1, 2). Members of the SNA branch ultimately reached southern South America, and based on mtDNA, Y chromosome and genome-wide evidence this likely occurred quickly (2, 7, 8, 34, 35). This movement gave rise to serial splitting and early population structure, with Mesoamericans being the most deeply divergent group, followed by South Americans east and west of the Andes (4, 36). However, genomic data from Spirit Cave (10.7 ka) and Lagoa Santa (10.4 ka), the oldest sites in this study, show that the SNA dispersal pattern south of the continental ice sheets involved complex admixture events between earlier-established populations.

MDS and ADMIXTURE, as well as a TreeMix tree focused on SNA genomes, reveal that the Spirit Cave and Lagoa Santa individuals were members of the SNA branch (Fig. 1, C and F) (13). Within that branch, Spirit Cave is closest to Anzick1, while Lagoa Santa is closest to southern SNA groups. Interestingly, two of the Lagoa Santa individuals carry the same mtDNA haplogroup (D4h3a) as Anzick1, yet three of the Lagoa Santa individuals harbor the same Y chromosome haplogroup as Spirit Cave (Q-M848) (13). Nonetheless, MDS transformations restricted to SNA (Fig. 1, D and E) (13) together with TreeMix graphs including admixture (13), suggest these North and South American ancient individuals are closely related, regardless of Lagoa Santa’s affinity to present-day South American groups.

We formally tested this scenario by fitting f-statistics-based admixture graphs, and found that even though Anzick1, Spirit Cave and Lagoa Santa are separated by ~2000 years and thousands of kilometers, these three individuals can be modeled as a clade to the exclusion of the Mesoamerican Mixe (13). While we did not find evidence rejecting this clade using TreeMix and D-statistics (13), further SFS-based modeling indicates that Mixe most likely carry gene flow from an unsampled outgroup, and form a clade with Lagoa Santa. Including nonzero outgroup admixture into Mixe when fitting an f-statistics-based admixture graph resulted in a significantly better fit (likelihood ratio test, P < 0.05) (Fig. 3, A and B) (13). Hereafter we refer to that outgroup as ‘Unsampled population A’ (UPopA), which is neither AB, NNA or SNA, and which we infer split off from Native Americans ~24.7 ka, ranging from 30-22 ka (95% CI; this large range is a result of the analytical challenge of estimating divergence and admixture times in the absence of UPopA genome data). This age range overlaps with the inferred split of Native Americans from Siberians and East Asians 26.1-23.9 ka (1), and the divergence of USR1 from other NA (23.3-21.2 ka). This temporal overlap, which cannot be fully resolved into a relative sequence with current data, suggests there were multiple splits that took place in Beringia close in time. Depending on how close they ultimately prove to be, they could imply there was moderate structure within Beringia (1, 37), possibly along with indirect gene flow from Siberians, perhaps via other NA populations. Under a model with a pulse-like gene flow, we inferred a probability of ~11% gene flow from UPopA into Mixe ~8.7 ka (95% CI: 0.4-13.9 ka; the wide interval...
potentially reflects unmodeled continuous migration) (Fig. 5) (13). Thus, we favor a model where the common ancestor of Anzick1 and Spirit Cave diverged from the common ancestor of Lagoa Santa and Mixe ~14.1 ka (95% CI: 13.2-14.9 ka), perhaps as the latter were moving southward. We infer that the Lagoa Santa population diverged from Mixe shortly thereafter, ~13.9 ka (95% CI: 12.8-14.8 ka) (Fig. 5) (13). The proximity of these estimated divergence times suggests the dispersal process was very rapid on an archaeological time scale, as populations expanded across North America perhaps in a matter of centuries, and then into eastern South America within a millennium or two.

**Australasian ancestry in Early Holocene South America and claims of "Paleoamericans"**

Both the Spirit Cave and Lagoa Santa individuals have been identified as "Paleoamericans" (38, 39), connoting a cranial morphology distinct from that of modern Native Americans. Interpretations of this pattern range from it being the result of a separate earlier migration into the Americas, to population continuity and in situ differentiation owing to factors such as isolation and drift (13, 40–42). We examined if this morphology might be associated with ancient Australasian genetic ancestry found in present-day Amazonian groups (2, 6). However, no morphometric data are available for present-day peoples with this genetic signal (6), nor has this signal been detected in any ancient skeleton with this morphology (2, 10).

To test for the Australasian genetic signal in Native Americans, we computed $D$-statistics of the form $D(NA, NA; Eurasian, Yoruba)$, where NA represents all newly sequenced and reference high-depth Native American genomes (13). In agreement with previous results (6), we find that the Amazonian Suruí share a larger proportion of alleles with Australasian groups (represented by Papuans, Australians and Andaman Islanders), with respect to Mixe (13). Lagoa Santa yielded similar results to those obtained for the Suruí. When compared to Mesoamerican groups (Mixe and Huichol), Lagoa Santa also shares a larger proportion of alleles with Australasian groups but not with other Eurasians (Fig. 4) (13). However, the Australasian signal is not present in the Spirit Cave individual, and we include this distinction in the admixture graph modeling (Figs. 3A and 4A) (13). We inferred less than 3% European contamination in the Lagoa Santa genome (<3%) (13) and show that this finding is robust to potential European contamination by computing "correction-corrected" $f_2$-statistics (Fig. 4B) (13). The fact that the Australasian genomic signature was present in Brazil 10.4 ka, but absent in all genomes tested to date as old or older and further north, presents a challenge in accounting for its presence in Lagoa Santa.

Importantly, all sequenced "Paleoamericans" (including Kennewick/Ancient One) (2, 10), are genetically closer to contemporary Native Americans, than to any other ancient or contemporary group sequenced to date.

**Multiple dispersals into South America**

Genome-wide data from contemporary populations suggested a single expansion wave into South America with little gene flow between groups (4), but see (36). In contrast, analysis of later Holocene genomes suggests South Americans derived from one or more admixture events between two ancestral Native American groups, possibly via multiple movements into South America (7).

To test these competing scenarios, we performed an exhaustive admixture graph search, as described above. Here we fitted a "seed graph" involving Yoruba, Mal’ta, Andaman, Han, Anzick1, Spirit Cave, Lagoa Santa and Mixe (present-day Mesoamericans), and tested all possible "non-admixed" and "admixed" models for SNA: the Mesoamerican Maya and Yukpa of Venezuela; groups east (Suruí, Karitiana, Piapoco, Chane) and west of the spine of the Andes (Aymara, Quechua); six ancient Patagonians [Ayayema, Punta Santa Ana and four individuals from (43)], the ancient Taino (44) and the Aconcagua Inca mummy (14) (Fig. 1B) (13). This analysis indicates most present-day South American populations do not form a clade with Lagoa Santa, but instead derive from a mixture of Lagoa Santa and Mesoamerican-related ancestries (Fig. 3A) (13). We confirmed these results by computing standard and error-corrected $D$-statistics of the form $D(LagoaSanta, SNA; Mixe, Yoruba)$ and $D(Mixe, SNA; LagoaSanta, Yoruba)$ (Fig. 3B) (13). For most groups, these statistics are inconsistent with a simple tree and indicate multiple dispersals into South America.

The ~5.1 ka Patagonian Ayayema genome is an exception; it forms a clade with the Lagoa Santa population. This suggests the arrival of the Mesoamerican-related ancestry occurred post-5.1 ka, and/or that it did not reach the remote region inhabited by the Ayayema individual’s ancestors (Fig. 3C) (13). This result is qualitatively mirrored by the 7.2 ka Punta Santa Ana individual (both cluster with present-day Patagonians, and form a clade with Lagoa Santa). However, the low coverage of Punta Santa Ana may reduce our power to detect possible Mesoamerican-related admixture (Fig. 3C) (13).

We further explored the fit of the model (Fig. 3A) for each South American group by fixing the Australasian contribution into Lagoa Santa and the Mesoamerican contribution (Fig. 3, D and E) into the test SNA population, across a range of values (13). While an Australasian contribution <1% and >6% results in a significant decrease in likelihood (likelihood ratio test, $P < 0.05$), the Mesoamerican contribution has a wider range of plausible values (Fig. 3E) (13). Yet, modelling each SNA group with little to no Mesoamerican-related admixture consistently yields significantly lower fit scores ($P < 0.05$) (13), except for the Ayayema individual (Fig. 3D) (13).
The Australasian contribution into Lagoa Santa is consistently nonzero when modeling South Americans, although we did not observe in every case a significant improvement when modeling Australasian admixture into SNA groups through Lagoa Santa (13). This result suggests this ancestry may have been widespread among early South Americans. Although we are unable to estimate the Lagoa Santa-related admixture proportion for these groups with confidence, we observe a general trend for populations east of the Andes (e.g., Suruí) to bear more of this ancestry than Andean groups (e.g., Aymara) (Fig. 3F) (13). A possible explanation for this difference is that on the western side of the Andes there was greater Mesoamerican-related admixture.

Finally, we explore the demographic history of present day South Americans using both joint SFS (momi2) (25, 26) and linkage disequilibrium information [smc++ (27) and dl-Cal2 (28)]. We seek to understand these groups’ relationships to Lagoa Santa – which also provides an indirect means of assessing the impact of admixture on the Australasian signature. For the SFS analysis, we selected the Karitiana, the only SNA population for which there are sufficient unadmixed publicly-available genomes (n = 5); for the diCal analysis, we used the Karitiana, Aymara (n = 1) and Suruí (n = 2) (13). From SFS analysis we infer that the ancestors of Lagoa Santa and Karitiana diverged from each other ~12.9 ka (95% CI: 10.4-14.0 ka). Subsequently, the latter received gene flow from a Mesoamerican-related population, which already carried admixture from the outgroup UPopA (Fig. 5) (13). Assuming pulse-like migration, this points to recent gene flow (~35%) from the Mesoamerican-related group into Karitiana (Fig. 5), possibly suggesting ongoing admixture over an extended period. Indeed, when we allowed for two pulses, we inferred substantial gene flow in both the recent and distant past (13). The diCal2 results are consistent for the Karitiana, Aymara, and Suruí showing their demographic histories involved a mixture between a Lagoa Santa-related and a Mixe-related source (13).

Overall, our findings suggest that soon after arrival, South Americans diverged along multiple geographic paths (36). That process was further complicated by the arrival of a second independent migration and gene flow in Middle to Late Holocene times. Later admixture potentially reduced the Australasian signature that might have been carried by earlier inhabitants.

**Long-term population continuity in the North American Great Basin and the Numic Expansion**

Mesoamerican-related expansion possibly had a bearing on a later, unresolved pattern seen in North America. In the western Great Basin of North America paleoenvironmental evidence indicates there was decreased effective precipitation and increased aridity during the Middle Holocene, which led to a human population decline (45, 46). By ~5 ka regional populations were rebounding, but whether these were descendants of the previous inhabitants is not known. Unclear also is the relationship between those later Holocene groups and the people present in the region at the time of European Contact and today. There is linguistic evidence that ancestors of Numic speakers presently inhabiting the region today arrived recently, perhaps ~1 ka. There is also archaeological evidence of changes in material culture around that time, though how those relate to the linguistic turnover is uncertain. Nor is it known if these changes are related to population admixture or replacement. Patterns and changes in language, material culture and genetics need not be congruent or causally linked (47). Thus, the question of what has been called the Numic Expansion has been highly debated (46, 48); we address the population aspect by comparing genomes from Spirit Cave and Lovelock Cave (Fig. 1A) (13).

MDS and ADMIXTURE analyses, as well as D-statistics of the form D(SpiritCave, Lovelock2/3; NA, Yoruba) suggest that despite the ~9,000 years separating the Spirit Cave and Lovelock individuals, they form a clade with respect to other NA (Fig. 1, C to F) (13). We tested that topology through the same admixture graph search implemented for SNA (13). We were not able to reject the model without Mesoamerican-related admixture for Lovelock2 (~1.9 ka). However, the ~0.7 ka Lovelock3 individual received Mesoamerican-admixture, from a group that was likely not present in the region just ~1.2 ka earlier in Lovelock2. Because we do not know the language(s) that may have been involved, we cannot securely attribute this admixture to arriving Numic speakers [the Mixe, whom we use as a proxy for Mesoamerican ancestry, fall in a separate language family from Numic (49)]. Importantly, we also observe genetic continuity, suggesting that there was not a complete population replacement.

Present-day Pima from northern Mexico can also best be modeled as a Mesoamerican-related mixture. However, the Pima require admixture from a branch splitting above the Mixe-Spirit Cave divergence, likely an NNA population (13). We cannot specify a particular source population. These patterns indicate that complex population movements and mixture occurred after the initial settlements of the Great Basin and Southwest from both the north and south.

**Long-term complex population history in the Pacific Northwest**

Pacific Northwest groups had a late Pleistocene demographic history argued to be distinct from early SNA groups (1, 2, 9, 18, 33). To explore the population history and the relationship of regional populations to NNA and SNA, we assessed the genetic affinities between the 5.6 ka Big Bar Lake individual from the Fraser Plateau of central British Columbia, and other NA. Given their relative geographic proximity, we included the 939, 302, and 443 individuals from coastal British Columbia (9), and the Kennewick Man/Ancient One
(10). Since these genomes have been deemed representatives of NNA, we also included genomic data from Southwestern Ontario (ASO) individuals, who are closely related to Algonquin (NNA) populations (7).

These ancient North American individuals clustered separately from SNA populations in both MDS transformations, and their ancestry component distribution closely resembles that of NNA populations (Fig. 1, C and F) (13). However, we observed genetic differentiation between these individuals and other North American populations. While the coastal British Columbia ancient individuals clustered together with present-day Athabascan and Tsimshian speakers from the region, ASO and Kennewick were placed in an intermediate position between NNA and SNA. Although the Big Bar individual was placed close to NNA populations not carrying recent Siberian admixture (Fig. 1, C and F) (13), D-statistics of the form D(Aymara, NA; BigBar, Yoruba) and D(USR1, NA; BigBar, Yoruba) suggest Big Bar represents a previously undetected outgroup to non-AB Native Americans, one that diverged prior to the NNA/SNA split (13).

In order to describe the genetic ancestry of these individuals, we used the admixture graph search strategy (13). In agreement with previous results, the ancient coastal British Columbia individuals are best modeled as a clade to Athabascans, who bear Siberian-related admixture (Fig. 2C). However, the best-fitting model suggests the Big Bar individual represents a population that split prior to the NNA-SNA divergence, but after AB and without Siberian admixture (Fig. 2B) (13). Finally, in accordance with their placement in both MDS transformations, we observed that Kennewick and ASO individuals are best modeled as deriving a fraction of their ancestry from an SNA-related source, represented by Spirit Cave in this case (Fig. 2C). We confirm this through error-corrected D-statistics (Fig. 2F) (13), suggesting gene flow between ASO and an SNA group that diverged after the split of Anzick1, and that did not bear recent Mesoamerican-related ancestry.

Thus, the broader population history in this region was evidently marked by admixture between the NNA and SNA branches which most likely gave rise to the ancestors of Kennewick and ASO, and by isolation between groups in coastal (e.g., the 939 individual) and interior (Big Bar) British Columbia.

Discussion

The genomes described here do not undermine the previously established tree of AB splitting from ancestral Native Americans, followed by the basal NNA/SNA split south of eastern Beringia. However, they show it is at best a rough outline of the process. We now find that once south of eastern Beringia, Native Americans radiated rapidly and gave rise to multiple populations, some of which are visible in the genetic record only as unsampled populations, and who at different times expanded to different portions of the continent, though not as extensively as the initial peopling (Fig. 6).

Rapid movement from North to South America is evident genetically (Fig. 6, A and B), and had been anticipated by the ‘archaeologically-instantaneous’ appearance of sites throughout the hemisphere just after 13 ka (50, 51). It suggests the mechanism of movement was not simply gradual population growth and incremental geographic expansion, but rather was likely more akin to leap-frogging across large portions of the diverse intervening landscape (52). If this result holds, it predicts that additional terminal Pleistocene samples will fit on a star-like pattern as observed here.

That the early population evidently spread widely, rapidly, but if somewhat unevenly across the Americas, in turn suggests their access to large portions of the hemisphere was essentially unrestricted (52). Yet, there are hints in the genetic record of early unsampled populations (6) (Fig. 5), and the material culture associated with that rapid spread (Clovis and later) is distinct from and post-dates the earliest secure archaeological presence in the Americas at 14.6 ka (53). How these early groups are related, particularly ones with excess Australasian ancestry, and their degree of structure, remains largely unknown.

The Australasian signal is not present in USRI or Spirit Cave, but only appears in Lagoa Santa. None of these individuals have UPoA/Mesoamerican-related admixture, which apparently dampened the Australasian signature in South American groups, such as the Karitiana (Figs. 4 and 5). These findings suggest the Australasian signal, possibly present in a structured ancestral NA population (16), was absent in NA prior to the Spirit Cave/Lagoa Santa split. Groups carrying this signal were either already present in South America when the ancestors of Lagoa Santa reached the region, or Australasian-related groups arrived later but before 10.4 ka (the Lagoa Santa 14C age). That this signal has not been previously documented in North America implies that an earlier group possessing it had disappeared, or a later-arriving group passed through North America without leaving any genetic trace (Fig. 6, A and C). If such a signal is ultimately detected in North America it could help determine when groups bearing Australasian ancestry arrived, relative to the divergence of SNA groups.

Although we detect the Australasian signal in one of the Lagoa Santa individuals identified as a “Paleoamerican,” it is absent in other “Paleoamericans” (2, 10), including Spirit Cave with its strong genetic affinities to Lagoa Santa. This indicates the “Paleoamerican” cranial form is not associated with the Australasian genetic signal, as previously suggested (6), or any other specific NA clade (2). The cause of this cranial form, if it is representative of broader population patterns, evidently did not result from separate ancestry, but likely multiple factors, including isolation and drift and non-
stochastic mechanisms (2, 10, 13, 54).

The attenuating effects of distance, compounded (in places) by geographic barriers, led to cultural drift and regional adaptations, even early in the peopling process (52, 55). It was previously surmised that Clovis (Anzick1) and Western Stemmed (Spirit Cave) technologies (46, 56) represented “genetically divergent, founding groups” (57). It appears instead that the divergence is principally cultural, between genetically close populations living on opposite sides of the Rocky Mountains. This result affirms the point that archaeological, anatomical, and genetic records are not necessarily congruent (47).

That one of the principal isolating mechanisms was likely geographic helps explain the long-term population continuity in the Great Basin. Continuity occurred despite fluctuating human population densities and the cultural and linguistic changes that occurred over a 9,000 year span (46). In the Pacific NW, geographic barriers were less formidable, but we surmise that the region’s natural richness and diversity may have led groups to inhabit different environmental niches, which resulted in the emergence of social boundaries that maintained populations separation. In this region’s long history, we find evidence that groups on the coast (e.g., 939) and their contemporaries in the interior (Big Bar), were as genetically distinct as are present-day groups (18) (Fig. 6A). It is not known how or whether such differences map on to the region’s rich linguistic complexity and material culture differences (13, 58). Previous research on mtDNA and Y sequences hypothesized a shared origin for Pacific Northwest populations, followed by divergence due to isolation and drift (18). That Big Bar represents a previously unseen, isolated population supports the latter, but implies that the initial peopling of the region was complex and structured.

We also find evidence of a later Mesoamerican-admixture, which while geographically extensive, was not associated with a “wave” throughout the Americas, nor did it inevitably lead to replacement. Rather, it appears to mark the movement north and south (Fig. 6C), of what may have been relatively small groups that did not necessarily swamp local populations genetically or culturally, as illustrated by admixture in the Lovelock3 individual. Regardless of whether this marks the “Numic Expansion,” it was associated with evidence of cultural continuity as well as change; it was not an instance of population replacement. How or whether this Mesoamerican-related expansion is expressed culturally in South America is not known.

The genomes reported here fill gaps in our temporal and spatial coverage, and are valuable anchor points that reveal that the human population history of the Americas. As has long been expected (52), and is characteristic of human population histories around the world (59), the peopling process was marked by complex local and long-range demographic processes over time. The peopling of the Americas will likely prove more complicated still. As we have found, there was a previously unknown population in the Americas (UPopA), as well as one that harbored an Australasian signal in the Late Pleistocene and reached South America, yet left no apparent traces in North America. In addition, all of our evidence of the peopling process is from archaeologically known groups: Clovis (Anzick) and later populations. Yet, there is archaeological evidence of an earlier, pre-Clovis presence in the Americas, one for which we have yet to recover any ancient DNA. How these various population threads may ultimately come together, and how these populations were related to Native Americans past and present, remains to be resolved.

Materials and methods
Laboratory procedures
Ancient DNA work was performed in dedicated clean laborator-ory facilities at the Centre for GeoGenetics, Natural History Museum, University of Copenhagen. Extraction, treatment and library build protocols followed for each sample are detailed in (13). Sequencing was carried out in Illumina HiSeq instruments.

Data processing
Sequence reads were trimmed for Illumina adaptors using AdapterRemoval (60) and mapped to the human reference genome build 37 using BWA v.0.6.2-r126 (61) with disabled seeding (-l parameter) (62). Reads with mapping quality lower than 30 were discarded, PCR duplicates were identified using MarkDuplicates (63) and local realignment was carried out using GATK (64). Genotype calls for high coverage samples were generated using SAMtools mpileup (65) and filtered following (2). Called genotypes were phased with impute2 (66, 67) using the 1,000 genomes phased variant panel (phase 3) as a reference and the HapMap recombination rates. The final callset was masked using a 35-mer “snappability” mask with a stringency of 0.5 (68), and the strict accessible regions from the 1000 Genomes Project (69).

Ancient DNA data authentication
We examined the fragment length distributions and the base substitution patterns using bamdamage (20). We estimated mtDNA, X-chromosome and nuclear contamination using contamMix (70), ANGSD (71) and DICE (72), respectively.

Population structure analyses
We investigated the broad relationships between ancient and present-day genomes using model-based clustering as implemented in ADMIXTURE (19) and multidimensional scaling applied to the identity-by-state (20) and f(distance matrices (21, 22).

D-statistics
We computed D-statistics to formally test hypotheses of treeness and gene-flow. Genotype-based D-statistics were computed as detailed in (21), and error-corrected D-statistics
Confidence intervals were obtained through a nonparametric number of models, using the joint site frequency spectrum. Information, to infer key demographic parameters relating to ancestry, which relies on linkage disequilibrium and admixture graphs. First release: 8 November 2018 www.sciencemag.org (Page numbers not final at time of first release)
First release: 8 November 2018
help and support; and the Danish National High-throughput Sequencing Centre for assistance with data generation. St. John’s College, Cambridge University, provided E.W. and D.J.M. a congenial venue for scientific discussions. **Funding:** GeoGenetics members were supported by the Lundbeck Foundation and the Danish National Research Foundation (DNRF94) and KU2016. Study of the Lagoa Santa human remains was partly funded by the Augustinus Foundation. J.C. is supported in part by an NSF graduate research fellowship. J.C., J.P.S., and Y.S.S. are supported in part by the National Institutes of Health grant R01-GM094402, a Microsoft Azure research award, and a Packard Fellowship for Science and Engineering. A.S. was funded by FAPESP (2017/16495-2). Study of Patagonian samples by M.A.P.D. was sponsored by the National Geographic Society, Genographic Program, grant 14-01. H.S. was supported by HERA and the European Union’s Horizon2020 Research and Innovation Programme (grant 649307). J.R.R. was supported by funding from the National Park Service Shared Beringian Heritage Program. A.S.M. thanks the Swiss National Science Foundation and the European Research Council (ERC starting grant) for funding. Y.S.S. is a Chan Zuckerberg Biohub Investigator. D.J.M. is supported by the Quest Archaeological Research Fund and the Potts and Sibley Foundation. **Author contributions:** The project was conceived by E.W., J.V.M.-M., and D.J.M. and headed by E.W. L.V., P.B.D., C.F., M.E.A., A.M., M.I.O., D.M., and H.S. processed ancient DNA. J.V.M.-M., S.R., and T.P. assembled datasets. J.V.M.-M., J.C., J.P.S., T.V., F.R., T.P., and Y.S.S. analyzed genetic data with input from P.B.D. and C.F. R.N., Y.S.S., M.Si., and A.-S.M. supervised bioinformatic and statistical analyses. Radiocarbon analyses were done by L.B.-V., D.Ch., D.Co., T.D., and T.H.C.V., F.M., M.V., S.W., K.G., N.L., M.M.L., K.K., and L.S. excavated, curated, sampled, or described human remains. D.J.M., D.K.G., V.A., C.P., J.E., C.R.-C., S.R., M.Q.R.B., J.C., M.W., C.B., K.L.H., M.M.L., A.St., A.Sa., F.M., M.A.-D., and J.T.R conducted archaeological and bioanthropological analyses and provided contextualization. L.G. and H.H. partnered with R.M. and E.W. to engage with indigenous communities. All authors contributed to data interpretation. J.V.M.-M., D.J.M., Y.S.S., and E.W. wrote the manuscript, with input from J.C., J.P.S., M.A.-D., R.M., J.T.R., T.P., F.R.S., A.-S.M., and the other authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Sequence data are available at the ENA under accession PRJEB29074.

**SUPPLEMENTARY MATERIALS**
www.sciencemag.org/cgi/content/full/science.aav2621/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S80
Tables S1 to S18
References (76–223)

3 September 2018; accepted 30 October 2018
Published online 8 November 2018
10.1126/science.aav2621
Fig. 1. Ancient genome overview and their broad genetic affinities. (A) Sampling locations for ancient genomes (circles; new ones in bold) and present-day Native Americans [triangles colored by the grouping used in (2, 4)]. NNA and SNA were classified following (1). Present-day whole-genome data are labeled in dark blue. Broad geographic features mentioned in the text are shown in brown; the extent of glacial ice sheets ~15.5 ka (73) in light blue. (B) Calibrated radiocarbon ages for ancient genomes; previously published (open circles), this study (filled circles with depth of coverage). (C) Multidimensional scaling plot from the $f_3$-distance matrix computed from a subset of the SNP array dataset (~200,000 sites), including Siberian and Native American populations. (D and E) MDS plots similar to C, showing the first three dimensions for SNA groups only. (F) ADMIXTURE proportions assuming $K = 16$ ancestral populations. Bars represent individuals; colors represent ancestral components. For clarity, we only show Native Americans, three individuals for populations with $n > 3$, and single genomes are represented as wider bars. Siberians and Native Americans are organized according to (4).
Fig. 2. Admixture graphs modelling the ancestry of ancient North American genomes. We enumerated all possible extensions of the seed graph (13) where we added Trail Creek (A), Big Bar (B), 939 (C), Kennewick (D), and ASO (E) as a “non-admixed” or an admixed population, and optimized the parameters for each topology using qpGraph. In each graph, the test population is shown in blue. We show the best-fitting model for each genome as inferred from the final fit score. Above each graph, we show: the four populations leading to the worst $D$-statistic residual; the observed value for such statistic, and the expected values under the fitted model, the residual, the standard error of the residuals and a $Z$-score for such a residual; and the model fit score. Numbers to the right of solid edges are proportional to optimized drift; percentages to the right of dashed edges represent admixture proportions. In each graph we highlight the test genome in blue. (F) Error-corrected $D$-statistics restricted to transversion polymorphisms testing the genetic affinity between ASO and different SNA pairs. Points represent $D$-statistics and error bars represent $\pm 3.3$ standard errors ($P < 0.001$). For each test, we show the absolute $Z$-score beside its corresponding $D$ value. A pool of the five sequenced individuals represents the Lagoa Santa population.
Fig. 3. \(f\)-statistics–based tests show a rapid dispersal into South America, followed by Mesoamerican-related admixture. (A) Schematic representation of a model for SNA formation. This model represents a reasonable fit to most present-day populations (13). “UPopA”-, Mixe- and Australasian-related admixture edges are color-coded as in (C) to (F). (B) Fit score of the graph shown in (A) (excluding South Americans), as a function of “unsampled admixture” in Mixe. The point indicates the “unsampled admixture” proportion that yields the best fit score. (C) Error-corrected \(D\)-statistics showing that Lagoa Santa, Mixe and most SNA genomes cannot be modeled using a simple tree. In the top panel, we show the tested null hypothesis, together with an indication of the pair of populations with excess allele sharing, depending on the sign of \(D\). SNA populations are organized according to their sampling location (right labels). Points represent \(D\)-statistics and error bars represent \(-3.3\) standard errors (which corresponds to a \(P\) value of \(-0.001\) in a \(Z\)-test). For each test, we show the absolute \(Z\)-score beside its corresponding \(D\) value. (D and E) Fit score surfaces for the “admixed” SNA model with fixed Mixe and Australasian admixture proportions. For Ayayema and Suruí, we explored the fit of the model shown in a., across a grid of values for the Mixe proportion into SNA \(\{0,0.05,...,1\}\) and the Australasian contribution into Lagoa Santa \(\{0,0.01,...,0.1\}\). “X” indicates the parameter combination yielding the best score. Contour lines were drawn such that all parameter combinations contained within a given line, all yield a fit score lower than that indicated by the contour label. (F) A one-dimensional representation of (D) and (E) for all SNA populations. In this case, we fixed the Australasian contribution into Lagoa Santa to 3%. Each line is labeled at the value that yields the best fit score. We compare different models based on their fit scores, where a difference of \(-3\) corresponds to a \(P\) value of \(-0.05\) and a difference of \(-4.6\) corresponds to a \(P\) value of \(-0.01\). For (B) a pool of the five sequenced individuals represents the Lagoa Santa population. For (C) to (F), we considered the called-genotype dataset excluding transitions (13), and used the high-depth “Sumidouro5” individual as a representative of the Lagoa Santa population.
Fig. 4. Allele sharing symmetry tests for pairs of Native Americans, relative to present-day Eurasian groups. (A) We computed $D$-statistics of the form $D(\text{Native American, Native American; Eurasian, Yoruba})$ to test whether a given Native American group carries excess “non-Native American” ancestry compared to other Native Americans. For each statistic, we obtained a $Z$-score (diamonds) based on a weighted block-jackknife procedure over 5-Mb blocks. Vertical lines represent $-3.3$ and $-3.3$ (which correspond to a $P$ value of $\sim 0.001$). Purple: Oceanians; pink: Southeast Asians; gray: non-Australasians. In this case we only show results for present-day Eurasian populations (13). (B) “Contamination-corrected” $f_4$-statistics of the form $f_4(\text{Mixe, Lagoa Santa; Australasian, Yoruba})$. For each statistic, we subtracted the value of $f_4(\text{Mixe, French; Australasian, Yoruba})$, weighted by an assumed contamination fraction $c$ ranging between 0 and 10% ($y$-axis). Points represent $f_4$-statistics and error bars represent $\pm 3.3$ standard errors. We observe that the apparent allele sharing between Lagoa Santa and Australasians increases as a function of the correction. As a reference, we show the value of $f_4(\text{Mixe, Surui; Australasian, Yoruba})$ as solid vertical lines. All tests are from the whole-genome dataset described in (13) and excluding transition polymorphisms. (C) Approximate sampling location for Australasian groups highlighted in (A).
**Fig. 5. Demographic history of South Native Americans.** A schematic representation of the most likely model relating the ancient USR1, Anzick1, Spirit Cave and Lagoa Santa genomes and the present-day Mixe (n = 3) and Karitiana (n = 5). Demographic parameters were inferred using momi2 (13). This model features a quick north to south splitting pattern of South Native Americans over a period shorter than 2,000 years, with later admixture from an outgroup ("UPopA") into the Mesoamerican Mixe. In addition we find evidence of gene flow from the latter into present-day South Americans, represented in this case by the Karitiana. Admixture pulses from USR1 into the ancestors of other Native Americans follow the inference in (I).
Fig. 6. Schematic depiction of the processes of human dispersal and divergence in the Americas, arranged chronologically. (A) Initial entry into eastern Beringia then into unglaciated North America, ~25 ka to ~13 ka, during which there are multiple splits: first in Beringia (UPopA and AB from the NNA+SNA line), followed by the Big Bar ancestral population splitting from NNA+SNA, then lastly by the NNA/SNA split south of eastern Beringia. NNA groups remained in northern North America, while SNA groups began to disperse across the North American continent. (B) Period of dispersal hemisphere-wide, ~14 ka to ~6 ka, during which time SNA moved rapidly from North into South America, resulting in the close affinities of the nearly contemporaneous Spirit Cave and Lagoa Santa individuals. Early South American populations possibly carried an Australasian-related admixture, as seen in the Lagoa Santa individual, and diverged west and east of the Andes. There was also admixture in North America between NNA and SNA groups prior to 9 ka that formed the population of which Kennewick/Ancient One was a member. It is inferred that during this period but after 9 ka (the age of the Trail Creek AB individual) NNA groups moved north into Alaska. (C) Population expansion out of Mesoamerica sometime after ~8.7 ka. These groups moved north into the Great Basin, resulting in a population turnover after 2 ka, evidenced by the difference between the Lovelock 2 and Lovelock 3 individuals. In South America that expansion contributed to the ancestry of most South American groups, but did not reach Patagonia by 5.1 ka, the age of the Ayayema individual.
Early human dispersals within the Americas


published online November 8, 2018

ARTICLE TOOLS
http://science.sciencemag.org/content/early/2018/11/07/science.aav2621
SUPPLEMENTARY MATERIALS
http://science.sciencemag.org/content/suppl/2018/11/07/science.aav2621.DC1
RELATED CONTENT
http://science.sciencemag.org/content/sci/362/6415/627.full
REFERENCES
This article cites 199 articles, 49 of which you can access for free
http://science.sciencemag.org/content/early/2018/11/07/science.aav2621#BIBL
PERMISSIONS
http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service