INTRODUCTION: Genetic studies of the Pleistocene peopling of the Americas have focused on the timing and number of migrations from Siberia into North America. They show that ancestral Native Americans (NAs) diverged from Siberians and East Asians ~23,000 years (~23 ka) ago and that a split within that ancestral lineage into northern NA (NNA) branches ~15.5 ka ago, a split inferred to have taken place south of eastern Beringia (present-day Alaska and western Yukon Territory), occurred ~21 ka ago. Subsequently, NAs diverged into northern NA (NNA) and southern NA (SNA) branches ~15.5 ka ago, a split inferred to have taken place south of eastern Beringia (present-day Alaska and western Yukon Territory). RATIONALE: Claims of migrations into the Americas by people related to Australasians or by bearers of a distinctive cranial morphology (“Paleoamericans”) before the divergence of NAs from Siberians and East Asians have created controversy. Likewise, the speed by which the Americas were populated; the number of basal divergences; and the degrees of isolation, admixture, and continuity in different regions are poorly understood. To address these matters, we sequenced 15 ancient human genomes recovered from sites spanning from Alaska to Patagonia; six are ≥10 ka old (up to ~18× coverage).

RESULTS: All genomes are most closely related to NAs, including those of two morphologically distinct Paleoamericans and an AB individual. However, we also found that the previous model is just a rough outline of the peopling process: NA dispersal gave rise to more complex serial splitting and early population structure—including that of a population that diverged before the NNA-SNA split—as well as admixture with an earlier unsampled population, which is neither AB nor NNA or SNA. Once in the Americas, SNAs spread widely and rapidly, as evidenced by genetic similarity, despite differences in material cultural, between >10-ka-old genomes from North and South America. Soon after arrival in South America, groups diverged along multiple geographic paths, and before 10.4 ka ago, these groups admixed with a population that harbored Australasian ancestry, which may have been widespread among early South Americans. Later, Mesoamerican-related population(s) expanded north and south, possibly marking the movement of relatively small groups that did not necessarily swamp local populations genetically or culturally.

CONCLUSION: NAs radiated rapidly and gave rise to multiple groups, some visible in the genetic record only as unsampled populations. At different times these groups expanded to different portions of the continent, though not as extensively as in the initial peopling. That the early population spread widely and rapidly suggests that their access to large portions of the hemisphere was essentially unrestricted, yet there are genomic and archaeological hints of an earlier human presence. How these early groups are related or structured, particularly those with Australasian ancestry, remains unknown. Rapid expansion, compounded by the attenuating effect of distance and, in places, by geographic and social barriers, gave rise to complex population histories. These include strong population structure in the Pacific Northwest; isolation in the North American Great Basin, followed by long-term genetic continuity and ultimately an episode of admixture predating ~0.7 ka ago; and multiple independent, geographically uneven migrations into South America. One such migration provides clues of Late Pleistocene Australasian ancestry in South America, whereas another represents a Mesoamerican-related expansion; both contributed to present-day South American ancestry.

HUMAN GENOMICS

Early human dispersals within the Americas


INTRODUCTION: Genetic studies of the Pleistocene peopling of the Americas have focused on the timing and number of migrations from Siberia into North America. They show that ancestral Native Americans (NAs) diverged from Siberians and East Asians ~23,000 years (~23 ka) ago and that a split within that ancestral lineage into northern NA (NNA) branches ~15.5 ka ago, a split inferred to have taken place south of eastern Beringia (present-day Alaska and western Yukon Territory).
Early human dispersals within the Americas

J. Victor Moreno-Mayar1,6, Lasse Vinner1,6, Peter de Barros Damgaard1,6, Constanza de la Fuente14, Jeffrey Chan24, Jeffrey P. Spence1,6, Morten E. Allentoft1, Tharsika Vimala1, Fernando Racimo1, Thomaz Pinotti4, Simon Rasmussen5, Ashot Margaryan1,6, Miren Iraeta Ortegozko1, Dorothea Mylopotamitaki1, Matthew Woollery, Clement Bataille8, Lorena Becerra-Valdivia9, David Chivall9, Jeremi Cybulski16,17,18, Carlos Vullo19, Flavia Morello20, Miguel Vilar21, Daniel Comeskey9, Thibaut Devièse9, Donald K. Grayson10, Len George11, Harold Harry12, Verner Alexandersson13, Charlotte Primeau13, Jon Erlandson14, Claudia Rodrigues-Carvalho15, Silvia Reis15, Murilo R. R. Bastos15, Jeremi Cybulski16,17,18, Carlos Vello19, Flavia Morello20, Miguel Vilar21, Spencer Wells22, Kristian Gregersen1, Kasper Lykke Hansen1, Niels Lynnerup13, Marta Mirazón Lahr23, Kurt Kjær1, André Strauss24,25, Marta Alfonso-Durruty26, Antonio Salas27,28, Hannes Schroeder1, Thomas Higham9, Ripan S. Malhi29, Jeffrey T. Rasic30, Luiz Souza31, Fabricio R. Santos4, Anna-Sapfo Malaspinas32, Martin Sikora1, Rasmus Nielsen1,3,34, Yun S. Song33,35,1, David J. Meltzer1,36, Eske Willerslev1,37,38.

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 human genomes spanning from Alaska to Patagonia; six are ≥10,000 years old (up to ~18× coverage). All are most closely related to Native Americans, including those from an Ancient Beringian individual and two morphologically distinct “Paleoamericans.” We found evidence of rapid dispersal and early diversification that included previously unknown groups as people moved south. This resulted in multiple independent, geographically uneven migrations, including one that provides clues of a Late Pleistocene (c. ~25,000 ± 1100 years ago; 25 ± 1.1 ka ago) Paleoamerican signal, as well as a later Mesoamerican-related expansion. These led to complex and dynamic population histories from North to South America.

Early human dispersals within the Americas

HUMAN GENOMICS

Early human dispersals within the Americas

J. Victor Moreno-Mayar1,6, Lasse Vinner1,6, Peter de Barros Damgaard1,6, Constanza de la Fuente14, Jeffrey Chan24, Jeffrey P. Spence1,6, Morten E. Allentoft1, Tharsika Vimala1, Fernando Racimo1, Thomaz Pinotti4, Simon Rasmussen5, Ashot Margaryan1,6, Miren Iraeta Ortegozko1, Dorothea Mylopotamitaki1, Matthew Woollery, Clement Bataille8, Lorena Becerra-Valdivia9, David Chivall9, Jeremi Cybulski16,17,18, Carlos Vullo19, Flavia Morello20, Miguel Vilar21, Daniel Comeskey9, Thibaut Devièse9, Donald K. Grayson10, Len George11, Harold Harry12, Verner Alexandersson13, Charlotte Primeau13, Jon Erlandson14, Claudia Rodrigues-Carvalho15, Silvia Reis15, Murilo R. R. Bastos15, Jeremi Cybulski16,17,18, Carlos Vello19, Flavia Morello20, Miguel Vilar21, Spencer Wells22, Kristian Gregersen1, Kasper Lykke Hansen1, Niels Lynnerup13, Marta Mirazón Lahr23, Kurt Kjær1, André Strauss24,25, Marta Alfonso-Durruty26, Antonio Salas27,28, Hannes Schroeder1, Thomas Higham9, Ripan S. Malhi29, Jeffrey T. Rasic30, Luiz Souza31, Fabricio R. Santos4, Anna-Sapfo Malaspinas32, Martin Sikora1, Rasmus Nielsen1,3,34, Yun S. Song33,35,1, David J. Meltzer1,36, Eske Willerslev1,37,38.

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 human genomes spanning from Alaska to Patagonia; six are ≥10,000 years old (up to ~18× coverage). All are most closely related to Native Americans, including those from an Ancient Beringian individual and two morphologically distinct “Paleoamericans.” We found evidence of rapid dispersal and early diversification that included previously unknown groups as people moved south. This resulted in multiple independent, geographically uneven migrations, including one that provides clues of a Late Pleistocene (c. ~25,000 ± 1100 years ago; 25 ± 1.1 ka ago) Paleoamerican signal, as well as a later Mesoamerican-related expansion. These led to complex and dynamic population histories from North to South America.

P revious genomic studies have estimated that ancestral Native Americans (NA) diverged from Siberian and East Asian populations ~25,000 ± 1100 years ago (25 ± 1.1 ka ago) (1, 2), with a subsequent split 22 to 18.1 ka ago within that ancestral lineage between later NAs and Ancient Beringians (ABs). NAs then diverged into two branches, northern NAs (NNAs) and southern NAs (SNAs), ~17.5 to 14.6 ka ago (2–4), a process inferred to have taken place south of eastern Beringia (present Alaska and western Yukon Territory). All contemporary and ancient NA individuals for whom genome-wide data have been generated before this study derive from either the NNA or SNA branch. However, disagreement exists over claims of earlier migrations into the Americas by people possibly related to Australasians or by bearers of a distinctive cranial form (“Paleoamericans”) (5, 6). Whether additional splits occurred within the Americas, how many migratory movements north and south took place, and the speed of human dispersal at different times and regions are also contentious. In contrast to models based on contemporary and Pleistocene-age genetic data (3, 4), genomic studies of later Holocene human remains indicate postdivergence admixture between basal NA 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 the Late Pleistocene and Early Holocene are rare. If we are to resolve how the peopling process occurred, more sequences are needed beyond the three currently available: Anzıkől from Montana (~12.8 ka old) (3), Kennewick Man/Ancient One from Washington (~9 ka old) (10), and USRI from Alaska (~11.5 ka old) (7).

Dataset and method summary

We engaged and sought feedback from Indigenous groups linked to the ancestral individuals analyzed in this study by using the recommendations for genomics research with Indigenous communities (11–13). We obtained genome sequences from 15 ancient human remains (Fig. 1A). These include remains from Trail Creek Cave 2, Alaska (radiocarbon dated to ~9 ka ago; ~0.4× genomic depth of coverage); Big Bar Lake, British Columbia (~5.6 ka old; ~1.2× coverage); and Spirit Cave, Nevada (~10.7 ka old; ~18× coverage); four individuals from Lovelock Cave, Nevada (ranging in age from ~1.95 to 0.6 ka old; ~0.5× to ~18.7× coverage); five individuals from Lagoa Santa, Brazil (~10.4 to ~9.8 ka old; ~0.18× to ~15.5× coverage); one individual each from the Punta Santa Ana and Ayaymi sites in Patagonian Chile (~7.2 and ~5.1 ka old, with ~15× and ~10.6× coverage, respectively); and an Incan mummy from Mendoza, Argentina (estimated to be ~0.5 ka old; ~2.5× coverage) (14) [all 1³⁶C ages are in calibrated years (13, 15)] (Fig. 1A and B). We also sequenced a ~15× genome from a ~19th-century Andaman islander, used as a proxy for Australasian ancestry in models involving admixture into NAs (2, 6, 13, 16).

All DNA extracts were confirmed to contain fragments with characteristic ancient DNA misincorporation patterns and low contamination levels (<3%) (20). The Spirit Cave, Lovelock 2, and Lovelock 3 genomes were generated solely from uracil-specific excision reagent–treated (USER) extracts, confirmed to contain characteristic ancient DNA misincorporation patterns before treatment (13, 17).

To assess the genetic relationships among 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 single-nucleotide polymorphism (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 SNAs, before the divergence of South Americas (4), which lacks the Australasian ancestry signal documented among some Amazonian groups (2, 6, 16).

We explored the ancient individuals’ broad genetic affinities initially by using model-based clustering (19) and multidimensional scaling (MDS) (Fig. 1, C to F) (31). MDS was applied to both the identity-by-state distance matrix for all individuals (20) and the f3 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 by using the joint site frequency spectrum (SJFS) (25, 26) and linkage disequilibrium (27, 28) information (Fig. 5). These efforts enabled us to explore finer-scale complex models by using whole-genome data (19). The 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.

Insights into early eastern Beringian populations from an Alaskan genome

Although the earliest archaeological evidence for a human presence in eastern Beringia remains
disputed, people were present in Alaska by at least 14.4 ka ago (29). Genomic insights from the USR1 genome indicate that ABs (I) remained isolated in Interior Alaska until at least the Pleistocene and were an outgroup to NNAs and SNAs. It was inferred that the NNA-SNA population split occurred outside of eastern Beringia (1, 2). By contrast, recent findings suggest that the ancestral population of NNAs 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 (21). This individual carried ancient ancestry, adjacent to USR1 in MDS analyses (Fig. 1C) (13) and carries a similar distribution of ancestry components (Fig. 1F) (13). The Trail Creek individual and USR2 (found with and a close relative to USR1) harbor the same mitochondrial DNA (mtDNA) haplogroup, B2, but not the derived B2 variant found elsewhere in the Americas (1, 13). Genotype-based D statistics of the form D(AYama, NA; TrailCreek, Yoruba) and D(USR1, TrailCreek; NA, Yoruba) suggest that Trail Creek forms a clade with USR1 that represents an outgroup to other NNAs (13). This placement was supported by fitting f-statistic–based admixture graphs (13, 21).

The procedure described here, which was also used for other samples, relies on a “seed graph” that incorporates the formation of the ancestral NA group and its three basal branches (ABs, NNAs, and SNAs) (I, 30, 31). The seed graph includes the following leaves: Yoruba (representing Africans), Mal’ta (ancient north Eurasians), Andaman (Australasians), Han (East Asians), USR1 (ABs), Athabascan (NNAs), and Spirit Cave (SNAs) (see below) (13). We enumerated all possible extensions of the seed graph where an individual genome, Trail Creek in this case, was added as either a nonadmixed or an admixed population (32). We optimized the parameters for each topology by 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 with its nonadmixed counterpart only if the absolute difference between fit scores (log likelihoods) was greater than ~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 the Trail Creek and USR1 individuals form a clade to be the most likely (Fig. 2A) (13).

These results suggest that the USR1 and Trail Creek individuals were members of an AB metropolitan population that occupied eastern Beringia and remained isolated from other NA populations during the Late Pleistocene and Early Holocene. Finding two members of the AB population, from sites ~750 km apart, with similar artifact technologies (33) supports the inference that the SNA-NNN split occurred south of eastern Beringia (1, 9). The alternative, that NNAs and SNAs split in Alaska, seems less likely; it would have required several thousand years of strong population structure prior to ~16 ka ago to differentiate those groups from each other and from ABs, as well as a separate SNA presence, which has yet to be found (I). These data indicate that the Athabascans and Inuit, who inhabit Alaska today and are NNAs but with additional Siberian-related ancestry (1, 4, 18, 33), presumably moved north into the region sometime after ~9 ka ago, 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 to 14.6 ka ago (1, 2). Members of the SNA branch ultimately reached southern South America, and on the basis of mtDNA, Y chromosome, and some-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 (107 ka old) and Lagoa Santa (10.4 ka old), 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, whereas Lagoa Santa is closest to southeastern South American. 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 the Spirit Cave genome (Q-M848) (13). Nonetheless, MDS transformations restricted to SNAs (Fig. 1, D and E) (13), together with TreeMix graphs including admixture (I5), suggest that these ancient North and South American 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 the Anzick1, Spirit Cave, and Lagoa Santa individuals are separated by ~2 ka and thousands of kilometers, genomes from these three individuals can be modeled as a clade to the exclusion of the Mesoamerican Mixe (33). Although we did not find evidence rejecting this clade by using TreeMix and D statistics (13), further SFS-based modeling indicates that the Mixe most likely carry gene flow from an unsampled outgroup and form a clade with Lagoa Santa. Including nonzero outgroup admixture into the 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 NNAs ~24.7 ka ago, with an age range between 30 and 22 ka ago (95% confidence interval; 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 NAs from Siberians and East Asians 26.1 to 23.9 ka ago (1) and the divergence of USR1 from other NAs (23.3 to 21.2 ka ago). This temporal overlap, which cannot be fully resolved into a relative sequence with current data, suggests that multiple splits took place in Beringia within a short span of time. Depending on how close these splits ultimately prove to be, they could imply that moderate structure existed within Beringia (1, 37), possibly along with indirect gene flow from Siberians, perhaps via other NA populations. Under a model with a

Fig. 1. Ancient genome overview and broad genetic affinities. (A) Sampling locations for ancient genomes (circles) (newly reported genomes are labeled in bold) and present-day NAs [triangles colored by the grouping used in (2, 4)]. NNA and SNAs were classified by following (1). Present-day whole-genome data are labeled in dark blue. Broad geographic features mentioned in the text are shown in dark red; the extent of glacial ice sheets ~15.5 ka ago (kya) (73) is shown in light blue. Anc, ancient; Pta Sta Ana, Punta Santa Ana. (B) Calibrated radiocarbon ages for ancient genomes. Open circles, previously published genomes; filled circles with depth of coverage, genomes from this study. SpCave, Spirit Cave; LagoaSta, Lagoa Santa. (C) MDS plot from the \( f_2 \) distance matrix computed from a subset of the SNP array dataset (~200,000 sites), including Siberian and NA populations. Dim, dimension. (D and E) MDS plots similar to the plot in (C), showing the first three dimensions for SNA groups only. (F) ADMIXTURE proportions with the assumption of \( K = 16 \) ancestral populations. Bars represent individuals; colors represent ancestral components. For clarity, we show only NAs. Three individuals are represented for populations with \( n > 3 \) genomes, and single genomes are represented as wider bars. Siberians and NAs are organized according to (4). WGS, whole-genome sequence; Hist, historic.
pulselike gene flow, we inferred a probability of ~11% gene flow from UPoPa into the Mixe ~8.7 ka ago (95% CI, 0.4 to 13.9 ka ago; the wide interval potentially reflects unmodeled continuous migration) (Fig. 5) (13). Thus, we favor a model where the common ancestor of the Anzick and Spirit Cave individuals diverged from the common ancestor of the Lagoa Santa and Mixe individuals ~14.1 ka ago (95% CI, 13.2 to 14.9 ka ago), perhaps as the Lagoa Santa–Mixe ancestral population was moving southward. We infer that the Lagoa Santa population diverged from the Mixe shortly thereafter, ~13.9 ka ago (95% CI, 12.8 to 14.8 ka ago) (Fig. 5) (13). The proximity of these estimated divergence times suggests that 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 NAs. Interpretations of this pattern range from its being the result of a separate earlier migration into the Americas to its arising from population continuity with in situ differentiation owing to factors such as isolation and drift (33, 40–42). We examined whether 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 NAs, we computed D statistics of the form $D_{NA, NA}$ (6, 39), where NA represents all newly sequenced and reference high-depth NA genomes (13). In agreement with previous results (6), we found that the Amazonian Suruí share a larger proportion of alleles with Australasian groups (represented by Papuans, Australians, and Andaman Islanders) than do the Mixe (13). Lagoa Santa yielded results similar to those obtained for the Suruí. The analyzed Lagoa Santa genome also shares a larger proportion of alleles with Australasian groups, but not with other Eurasians, than do Mesoamerican groups (the Mixe and Huichol) (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 “contamination-corrected” $f_2$ statistics (Fig. 4B) (13). The presence of the Australasian genomic signature in Brazil 10.4 ka ago and its absence in all genomes tested to date that are as old or older and located farther north present a challenge in accounting for its presence in Lagoa Santa.

**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)]. By contrast, analysis of later Holocene genomes suggests that South Americans derived from one or more admixture events between two ancestral NA groups, possibly via multiple movements into South America (7).

To test these competing scenarios, we performed an exhaustive admixture graph search, as described above. We fitted a seed graph involving Yoruba, Maťa, Andaman, Han, Anzick, Spirit Cave, Lagoa Santa, and Mixe (present-day Mesoamerican)
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 lines are color-coded as in subsequent panels. Meso., Mesoamerican. (B) Fit score of the graph shown in (A) (excluding South Americans) as a function of “unsampled admixture” in the Mixe. The point indicates the unsampled admixture proportion that yields the best fit score. (C) Error-corrected D statistics showing that Lagoa Santa (LagoaS), Mixe, and most SNA genomes cannot be modeled by using a simple tree. (Top) 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 (labels on the right). Points represent D statistics, and error bars represent ±3.3 SEs (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” SNM model with fixed Mixe and Australasian admixture proportions. For the Ayayema and Surui, we explored the fit of the model shown in (A) across a grid of values for the Mixe proportion in SNAs (0.005...1) and the Australasian contribution to Lagoa Santa (0.001...1). “X” indicates the parameter combination yielding the best score. Contour lines were drawn such that all parameter combinations contained within a given line 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 to Lagoa Santa at 3%. Each line is labeled at the value that yields the best fit score. We compared different models on the basis of 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 test SNA population across a range of values (23).
Fig. 4. Allele sharing symmetry tests for pairs of NAs, relative to present-day Eurasian groups. (A) We computed $D$ statistics of the form $D$(NA, NA; Eurasian, Y oruba) to test whether a given NA group carries excess “non-NA” ancestry compared with other NAs. For each statistic, we obtained a $z$-score (diamonds) on the basis of a weighted block jackknife procedure over 5-Mb blocks. Vertical lines represent −3.3 and −3.3 (which correspond to a $P$ value of ~0.001). In this case, we show only results for present-day Eurasian populations (13). Purple, Oceanians; pink, Southeast Asians; gray, non-Australasians. (B) Contamination-corrected $f_4$ statistics of the form $f_4$(Mixe, Lagoa Santa; Australasian, Yoruba). For each statistic, we subtracted the value of $f_4$(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 −3.3 SEs. We observe that the apparent allele sharing between Lagoa Santa and Australians increases as a function of the correction. As a reference, we show the values of $f_4$(Mixe, Surui; Australasian, Yoruba) as solid vertical lines. All tests are from the whole-genome dataset described in (13) and excluding transition polymorphisms. $H_3$, $D$-statistic term (see the supplementary materials); WCDesert, Ngaanyatjarra from western central desert. (C) Approximate sampling locations for Australasian groups highlighted in (A).
Fig. 5. Demographic history of SNAs. 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 genomes) and Karitiana (n = 5 genomes). Demographic parameters were inferred by using momi2 (13). This model features a quick north-to-south splitting pattern for SNAs over a period shorter than 2 ka, with later admixture from an outgroup (UPopA) into the Mesoamerican Mixe. In addition, we found 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 NAs follow the inference in (1).

every case a significant improvement when modeling Australasian admixture into SNA groups through Lagoa Santa (13). This result suggests that this ancestry was 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., the Suruí) to bear more of this ancestry than Andean groups (e.g., the Aymara) (Fig. 3F) (13). A possible explanation for this difference is that greater Mesoamerican-related admixture occurred on the western side of the Andes.

Lastly, we explored the demographic history of present-day South Americans by using both joint SFS (momii2) (25, 26) and linkage disequilibrium information [SMC++ (27) and diCal2 (28)]. We seek to understand these groups’ relationships to Lagoa Santa—which also provides an indirect means of assessing the effects of admixture on the Australasian signature. For the SFS analysis, we selected the Karitiana, the only SNA population for which a sufficient number of unadmixed genomes are publicly available (n = 5); for the diCal analysis, we used the Karitiana, Aymara (n = 1 genome), and Suruí (n = 2 genomes) (13).

From SFS analysis, we infer that the ancestors of Lagoa Santa and Karitiana diverged from each other ~12.9 ka ago (95% CI, 10.4 to 14.0 ka ago). Subsequently, the latter received gene flow from a Mesoamerican-related population, which already carried admixture from the outgroup UPopA (Fig. 5) (13). With the assumption of pulselike 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. 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 populations, showing that their demographic histories involved a mixture between a Lagoa Santa–related source 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 decreased effective precipitation and increased aridity during the Middle Holocene, which led to a human population decline (45, 46). By ~5 ka ago, regional populations were rebounding, but whether these were descendants of the previous inhabitants is unknown. 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. Linguistic evidence suggests that ancestors of Numic speakers presently inhabiting the region today arrived recently, perhaps ~1 ka ago. 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 whether 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 so-called Numic Expansion hypothesis has been highly debated (46, 48); we address the population aspect by comparing genomes from Spirit Cave and Lovelock Cave (Fig. 1A) (13). We were not able to reject the model without Mesoamerican-related admixture for Lovelock 2 (~1.9 ka old). However, the ~0.7-ka-old Lovelock 3 individual received Mesoamerican admixture from a group that was likely not present in the region just ~1.2 ka earlier, at the time of Lovelock 2. 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)). Notably, 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 that of early SNA groups (1, 2, 9, 18, 33). To explore the population history and the relationship of regional populations to NNA and SNAs, we assessed the genetic affinities between the 5.6-ka-old Big Bar Lake individual from the Fraser Plateau of central British Columbia and other NAs. 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). As these genomes have been deemed representatives of NNAs, we also included genomic data from ancient 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. 1C and F) (23). However, we observed genetic differentiation between these individuals and other North American populations. Whereas the ancient coastal British Columbian individuals clustered together with present-day
Athabaskan and Tsimshian speakers from the region, the ASO group and Kennewick Man/Ancient One were placed in an intermediate position between NNAs and SNAs. Although the Big Bar individual was placed close to NNA populations not carrying recent Siberian admixture (Fig. 1, C and F) (23), D statistics of the form D(Aymara, NA; BigBar, Yoruba) and D(USR1, NA; BigBar, Yoruba) suggest that Big Bar represents a previously undetected outgroup to non-AB NAs, one that diverged before the NNA-SNA split (13).

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 Columbian individuals are best modeled as a clade with Athabascans, who bear Siberian-related admixture (Fig. 2C). However, the best-fitting model suggests that the Big Bar individual represents a population that split before the NNA-SNA divergence but after AB divergence and without Siberian admixture (Fig. 2B) (39). Lastly, in accordance with their placement in both MDS transformations, we observed that Kennewick Man/Ancient One 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. 2, D and E) (23). We confirm this through error-corrected D statistics (Fig. 2F) (33) suggesting gene flow between ASO individuals and an SNA group that diverged after the split of Anzick-1 and did not bear recent Mesoamerican-related ancestry.

The broader population history in this region was evidently marked by admixture between the NNA and SNA branches that most likely gave rise to the ancestors of Kennewick Man/Ancient One and ASO individuals and by isolation between groups in coastal British Columbia (represented by the 939 individual) and interior British Columbia (represented by Big Bar).

**Discussion**

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

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

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

The Australasian signal is not present in USR1 or Spirit Cave and appears only in Lagoa Santa. None of these individuals have UPopA- or Mesoamerican-related admixture, which apparently dampened the Australasian signature in South American groups, such as the Karitiana (Figs. 4 and 5). These findings suggest that the Australasian signal, possibly present in a structured ancestral NA population (16), was absent in NA before the Spirit Cave–Lagoa Santa split. Either groups carrying this signal were 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 ago (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 that 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 detected the Australasian signal in one of the Lagoa Santa individuals identified as a Paleoamerican, it is absent in other Paleoamericans (2, 10), including the Spirit Cave genome with its strong genetic affinities to Lagoa Santa. This indicates that the Paleoamerican cranial form is not associated with the Australasian genetic signal, as previously suggested (6), or any other specific NA clade (2). The Paleoamerican cranial form, if it is representative of broader population patterns, evidently did not result from separate ancestry but likely from multiple factors, including isolation, drift, and nonstochastic 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 Poraries in the interior (Big Bar) were 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 existed despite fluctuating human population densities and the cultural and linguistic changes that occurred over a 9-ka span (46). In the Pacific Northwest, 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 population 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 (38) (Fig. 6A). How or whether such differences map to the region’s rich linguistic complexity and material culture differences is not known (43, 58). Previous research on mtDNA and Y sequences hypothesized a shared origin for Pacific Northwest populations, followed by divergence due to isolation and drift (12). That Big Bar represents a previously unseen, isolated population supports its ancestral isolation and drift but implies that the initial peopling of the region was complex and structured.

We also found evidence of a later Mesoamerican admixture, which though 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 Lovelock 3 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 are related to NAs past and present remain to be resolved.

Materials and methods

Laboratory procedures

Ancient DNA work was performed in dedicated clean laboratory 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

Sequencing reads were trimmed for Illumina adaptors by using AdapterRemoval (60) and mapped to the human reference genome build 37 by using BWA v0.6.2-r126 (61) with disabled seeding (–I parameter) (62). Reads with mapping quality lower than 30 were discarded, polymerase chain reaction duplicates were identified by using MarkDuplicates (63), and local realignment was carried out by using GATK (64). Genotype calls from high-coverage samples were generated by using SAMtools mpileup (65) and filtered according to the method of (2). Called genotypes were phased with impute2 (66, 67) by using the 1000 Genomes phased-variant panel (phase 3) as a reference and the HapMap recombination rates. The final call set was marked by using a 35-mer “snipability” 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 by using bamdamage (20). We estimated mtDNA, X chro-

mosome, and nuclear contamination by using contamMix (70), ANGSDD (71), and DICE (72), respectively.

Population structure analyses

We investigated the broad relationships between ancient and present-day genomes by using model-based clustering as implemented in ADMIXTURE (39) and MDS applied to the identity-by-state (20) and f3 distance (21, 22) matrices.

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 were computed according to the method of (23). For both approaches, standard errors were estimated through a weighted block jackknife approach over 5-Mb blocks.

Admixture graph fitting

We used qpGraph to fit f statistics–based admixture graphs (4, 21). We implemented an exhaustive admixture graph search where we considered a seed graph onto which a test population was added as either a nonadmixed or an admixed group in every possible position. Extensions of the seed graph were enumerated by using the admixturegraph R package (32). We evaluated each topology on the basis of its fit score, the z-score of the worst residual between the observed and predicted D statistics, and the presence of zero-length internal edges and carried out likelihood ratio tests by following (30). For all tests, we considered only transversion polymorphisms.

Demographic inference

We estimated marginal population sizes over time for different NA groups by using SMC++ (27). We then used momi2 (25, 26) to infer demographic parameters for a number of models with the joint SFS. CIs were obtained through a nonparametric bootstrap procedure. We confirmed the SFS-based inference by using diCal2 (29), which relies on linkage disequilibrium information, to infer key demographic parameters relating pairs of NA populations. A detailed description of laboratory and analytical methods is provided in (13).

REFERENCES AND NOTES


179. D. J. Moseley, First Peoples in a New World: Colonizing Ice Age America (Univ. of California Press, 2009).


SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6419/eaav2621/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S30
Tables S1 to S18
References (76–223)
3 September 2018; accepted 30 October 2018
Published online 8 November 2018
10.1126/science.aav2621
Early human dispersals within the Americas


Science 362 (6419), eaav2621.
DOI: 10.1126/science.aav2621originally published online November 8, 2018

Complex processes in the settling of the Americas

The expansion into the Americas by the ancestors of present day Native Americans has been difficult to tease apart from analyses of present day populations. To understand how humans diverged and spread across North and South America, Moreno-Mayar et al. sequenced 15 ancient human genomes from Alaska to Patagonia. Analysis of the oldest genomes suggests that there was an early split within Beringian populations, giving rise to the Northern and Southern lineages. Because population history cannot be explained by simple models or patterns of dispersal, it seems that people moved out of Beringia and across the continents in a complex manner.

Science, this issue p. eaav2621