Genomic evidence for the Pleistocene and recent population history of Native Americans

Abstract

How and when the Americas were populated remains contentious. Using ancient and modern genome-wide data, we find that the ancestors of all present-day Native Americans, including Athabascans and Amerindians, entered the Americas as a single migration wave from Siberia no earlier than 23 thousand years ago (KYA), and after no more than 8,000-year isolation period in Beringia. Following their arrival to the Americas, ancestral Native Americans diversified into two basal genetic branches around 13 KYA, one that is now dispersed across North and South America and the other is restricted to North America. Subsequent gene flow resulted in some Native Americans sharing ancestry with present-day East Asians (including Siberians) and, more distantly, Australo-Melanesians. Putative ‘Paleoamerican’ relict populations, including the historical Mexican Pericúes and South American Fuego-Patagonians, are not directly related to modern Australo-Melanesians as suggested by the Paleoamerican Model.

It is generally agreed that ancestral Native Americans are descendants of Siberian peoples who traversed the Bering Land Bridge (Beringia) from northeast Asia in Late Pleistocene times, and though consensus has yet to be reached, it is mostly conceded that the Clovis archaeological complex, dating to ca. 13 KYA, does not represent the first migration as long supposed (1–7). Archaeological evidence accumulated over the last two decades indicates that people were south of the North American continental ice sheets more than a millennium earlier and had reached as far south as southern South America by at least ca. 14.6 KYA (1–3). Interpretations differ, however, regarding the precise spatio-temporal dynamics of the peopling process, owing to archaeological claims for a significantly earlier human presence pre-dating the Last Glacial Maximum (LGM; ca. 20 KYA) (8–10), and conflicting interpretations of the number and timing of migrations from Beringia based on anatomical and genetic evidence (11–16). Much of the genetic evidence is from studies of mitochondrial DNA (mtDNA) and Y-chromosome, which as single, uniparentally inherited
loci are particularly subject to genetic drift and sex-biased demographic and cultural practices.

Among the principal issues still to be resolved regarding the Pleistocene and recent population history of Native Americans are: (i) the timing of their divergence from their Eurasian ancestors; (ii) whether the peopling was in a single wave or multiple waves, and, consequently, if the genetic differences seen between major subgroups of Native Americans (e.g., Amerindian and Athabaskan) result from different migrations or in situ diversification in the Americas (5, 6, 17, 18); (iii) if the migration involved ca. 15,000 years of isolation in the Bering Strait region, as proposed by the Beringian Incubation Model to explain the high frequency of unique and widespread American mitogenomes and private genetic variants (19–22); and, finally, (iv) if there was post-divergence gene flow from Eurasia and possibly even population replacement in the Americas, the latter suggested by the apparent differences in skull morphology between some early (‘Paleoamerican’) remains and those of more recent Native Americans (23–27). We address these issues using genomic data derived from modern populations, supplemented by ancient specimens that provide chronologically controlled snapshots of the genetics of the peopling process as it unfolded.

We sequenced 31 genomes from present-day individuals from the Americas, Siberia and Oceania to an average depth of ca. 20X: Siberians – Altai (n = 2), Buryat (n = 2), Ket (n=2), Koryak (n = 2), Sakha (n = 2), Siberian Yupik (n = 2); North American Native Americans – Tsimshian (n =); southern North American and Central and South American Natives – Pima (n = 1), Huichol (n = 1), Aymara (n = 1), Yukpa (n = 1); and, Oceanians – Papuan (n = 14) (28) (Table S1). All the genome-sequenced present-day individuals were previously genotyped using single nucleotide polymorphism (SNP) chips (4, 29–35) except for the Aymara individual that was SNP chip genotyped in this study (tables S3 and S4). They were selected on the basis of their ancestry profiles obtained with ADMIXTURE (36) to best represent their respective populations, and to minimize recent genetic admixture from populations of western Eurasian origin (28). For populations represented by more than one individual, we also verified from the genotype data that the sequenced individuals did not represent close relatives (28). We additionally sequenced 23 genomes from ancient individuals dating between ca. 0.2-6 KYA from North and South America, with an average depth ranging between 0.003X and 1.7X, including specimens affiliated to putative relict Paleoamerican groups such as the Pericúes from Mexico and Fuego-Patagonians from the southernmost tip of South America (23, 26–28) (table S5). Finally, we generated SNP chip genotype data from 79 present-day individuals belonging to 28 populations from the Americas and Siberia (28) (table S4). All the aforementioned datasets were analyzed together with previously published genomes and SNP chip genotype data (Tables S1, S3, and S4), masking the data for recent European admixture in some present-day Native American populations (28).

The structure of Native American populations and the timing of their initial divergence

We explored the genetic structure of Native American populations in the context of worldwide populations using ADMIXTURE (36), employing a reference panel consisting of
3,053 individuals from 169 populations (table S3) (28). The panel included SNP chip genotype data from present-day individuals generated in this study and previously published studies, as well as the 4,000 year-old Saqqaq individual from Greenland (29) and the 12,600 year-old Anzick-1 (Clovis culture) individual from Montana (5) (table S3). When assuming four ancestral populations \((K=4)\), we found a Native American-specific genetic component, indicating a shared genetic ancestry for all Native Americans including Amerindians and Athabascans (fig. S4). Assuming \(K=15\), there is structure within the Native Americans. Athabascans and northern Amerindians (primarily from Canada) differ from the rest of the Native Americans in sharing their own genetic component (fig. S4). As reported previously, Anzick-1 falls within the genetic variation of southern Native Americans (5), while the Saqqaq individual shares genetic components with Siberian populations (fig. S4) (29).

To ascertain the population history of present-day American populations in relation to worldwide populations, we generated admixture graphs with TreeMix (28, 37). All the modern Siberian and Native American genomes sequenced in this study, except for the North American Tsimshian genome that showed evidence of recent western Eurasian admixture (28), were used for this analysis, together with previously published genomes from Africa (Yoruba) (38), Europe (Sardinian, French) (38), East Asia (Dai, Han) (38), Siberia (Nivkh) (39) and the Americas (Karitiana, Athabascan, Greenlandic Inuit) (5, 38, 39) (table S1). The ancient individuals included in the analysis were Saqqaq, Anzick-1 and the 24,000 year-old Mal’ta child from south-central Siberia (4). TreeMix affirms that all Native Americans form a monophyletic group across all ten migration parameter values, with further diversification into two branches, one representing Amerindians (represented in this analysis by Amerindians from southern North America and Central and South America) and the other Athabascans (Fig. 1B and fig. S5). Paleo-Eskimos and Inuit were supported as a separate clade relative to the Native Americans, as reported previously (Fig. 1B and fig. S5) (29, 39). Our results show that the Siberian Yupik and Koryak are the closest Eurasian populations to the Americas, with the Yupik likely representing back-migration of the Inuit into Siberia (Fig. 1B and fig. S5).

To assess the pattern of the earliest human dispersal into the Americas, we estimated the timing of the divergence of ancestral Native Americans from East Asians (hereafter, including Siberians) using multiple methods. There is still some debate regarding mutation rates in the human genome (40), and this uncertainty could affect our estimates and results.

We applied \(diCal2.0\) (28) (Method 1), a new version of \(diCal\) (41) extended to handle complex demographic models involving multiple populations with migration (42), and an identity-by-state (IBS) tract method (43) (Method 2) to the modern genome dataset (28). With these, we first estimated divergence times between Native Americans and the Koryak of Siberia, one of the genetically closest sampled East Asian populations to Native Americans (fig. S5), using demographic models that reflect a clean split between the populations (28). With both \(diCal2.0\) and IBS tract method, the split of Native Americans (including Amerindians and Athabascans) from the Koryak dates to ca. 20 KYA (28) (tables S11A and S12 and fig. S15).
We further applied diCal2.0 to models with gene flow post-dating the split between Native Americans and Koryak (Fig. 2A) and found that they provided a better fit to the data than the models without gene flow (28). Overall, simulated data based on the models inferred using diCal2.0 and real data show very similar IBS tract length distributions (Fig. 2B) and relative cross coalescence rates (CCR) between pairs of individuals estimated using the Multiple Sequentially Markovian Coalescent (MSMC) method (Method 3) (28, 44) (Figs. 2, C and D). This serves as a confirmation for the model estimates from diCal2.0. We evaluated all the three methods using simulations under complex demographic models, and additionally investigated the effects of switch-errors in haplotype phasing on the estimates (28).

We then applied the diCal2.0 model that allows for gene flow between populations after their split to estimate divergence times for Native Americans from more geographically and genetically distant East Asian groups, including the Siberian Nivkh and Han Chinese. As before, the divergence estimates for Amerindians and Athabascans were very similar to one another, ca. 23 KYA (table S11B and figs. S18 and S21).

Hence, our results suggest that Amerindians and Athabascans were, by three different methods, consistently equidistant in time to populations that were sampled from different regions of East Asia, including some proximate to Beringia, and with varied population histories. This suggests that these two major Native American sub-groups are descendants of the same source population that split off from ancestral East Asians during the LGM. It is conceivable that harsh climatic conditions during the LGM may have contributed to the isolation of ancestral Native Americans, ultimately leading to their genetic divergence from their East Asian ancestors.

We also modeled the peopling of the Americas using a climate-informed spatial genetic model (CISGeM), in which the genetic history and local demography is informed by paleoclimatic and paleovegetation reconstructions (28, 45), and found the results to be in accordance with the conclusion of a single migration source for all Native Americans. Using present-day and ancient high coverage genomes, we found that Athabascans and Anzick-1, but not Greenlandic Inuit and Saqqaq (29, 39), belong to the same initial migration wave that also gave rise to present-day Amerindians from southern North America and Central and South America (Fig. 3), and that this migration likely followed a coastal route, given our current understanding of the glacial geological and paleoenvironmental parameters of the Late Pleistocene (fig. S31).

In all cases, the best fit of the demographic models to the IBS tract distribution and relative CCR by MSMC required gene flow between Siberian and Native American populations after their initial split (Figs. 2, B to D). We also found strong evidence for gene flow between Athabascans and the Inuit (table S11B) supported by results from ADMIXTURE (fig. S4), TreeMix (fig. S5), D-statistics employing both whole genome and SNP chip genotype data (28, 46, 47) (figs. S6 and S8A), and outgroup f3 statistics using whole genome data (28, 47) (Fig. S12). We attempted to estimate the divergence times between Inuit and Siberians as well as Inuit and Native Americans (table S11 and figs. S19 and S25 to S27).
but our analyses were complicated by gene flow between Inuit and Athabascans as well as complex admixture patterns among Arctic groups (fig. S5).

We tested the duration and magnitude of post-split gene flow between Native Americans and Siberians using diCal2.0 by introducing stopping time of gene flow as a free parameter (28). We still obtained the highest likelihood for a divergence time of 22 KYA between Amerindians and Siberians as well as Athabascans and Siberians, although estimates for gene flow rate and end of the gene flow differ (table S11C and fig. S22). Significant gene flow between Athabascans and Siberians seems to have stopped ca. 12 KYA (Table S11C), suggesting a link to the breaching of the Beringian Land Bridge by rising sea levels (48).

Overall, our results support a common Siberian origin for all Native Americans, contradicting claims for an early migration to the Americas from Europe (49), with their initial isolation and entrance into the Americas occurring no earlier than 23 KYA, but with subsequent admixture with East Asian populations. This additionally suggests that the Mal’ta-related admixture into the early Americans (4), representing ancestors of both Amerindians and Athabascans (Fig. 1 and fig. S5), occurred sometime after 23 KYA, following the Native American split from East Asians.

**Subsequent in situ diversification of Native American groups**

That Amerindian and Athabascan groups were part of the same migration implies that present-day genetic differences observed between them must have arisen later, after ca. 23 KYA. Using the clean-split model in diCal2.0 on the modern genomes dataset, we estimated that Athabascans and Karitiana diverged ca. 13 KYA (95% confidence interval of ca. 11.5-14.5 KYA, estimated from parametric bootstrap results) (table S11A, fig. S16), which is consistent with results from MSMC (fig. S27) (28).

Where the divergence between Karitiana and Athabascans occurred is not known. However, several independent lines of evidence suggest that it is more likely to have occurred in lower latitude North America instead of eastern Beringia (Alaska). These include the equidistant split times of Amerindians and Athabascans to Asian populations, the relatively brief interval between their estimated divergence date range and the age of Anzick-1 (12.6 KYA) (5), and lastly, the geographic location of Anzick-1 to the south of the North American ice sheets and its clear affiliation with the ‘southern branch’ of Native Americans (taken broadly to include Amerindians from southern North America and Central and South America) (5), as determined with outgroup f3 statistics using SNP chip genotype data from present-day worldwide populations (47) (Fig. 4 and figs. S13 and S14). Divergence in North America would also be consistent with the known pre-Clovis age sites in the Americas, such as Monte Verde (14.6 KYA) (50). The most parsimonious model would be that both Amerindians and Athabascans are descendants of the same ancestral Native American population that entered the Americas then subsequently diversified. However, we cannot discount alternative and more complex scenarios, which could be tested with additional ancient samples.

By the Clovis period (ca. 12.6 KYA), the ancestral Native American population had already diversified into 'northern' and 'southern' branches, with the former including ancestors of
present-day Athabascans and northern Amerindian groups such as Chipewyan, Cree and Ojibwa and the latter including Amerindians from southern North America and Central and South America (Fig. 4 and fig. S14). We tested whether later gene flow from East Asian sources, such as the Inuit, might explain the genetic differences between these two branches. Using $D$-statistics on SNP chip genotype data (47) masked for non-Native ancestry, we observed a signal of gene flow between the Inuit and northwest Pacific Coast Amerindians such as Coastal Tsimshian and Nisga’a, residing in the same region as the northern Athabascans (28) (fig. S8B). However, this signal of admixture with the Inuit, also detected in Athabascans (figs. S6 and S8A), was not evident among northern Amerindian populations located further east such as Cree, Ojibwa and Chipewyan (28) (fig. S8C). This suggests that the observed difference between the ‘northern’ and ‘southern’ branches is not a consequence of post-split East Asian gene flow into the ‘northern branch’, and also provides a possible explanation as to why the ‘southern branch’ Amerindians such as Karitiana are genetically closer to the northern Amerindians located further east than to northwest coast Amerindians and Athabascans (fig. S9).

In contrast to Anzick-1, several of the Holocene individuals from the Americas, including those sequenced in this study as well as the 8,500 year old Kennewick Man (51), are closely related to present-day Native American populations from the same geographical regions (Fig. 4 and figs. S13 and S14). This implies genetic continuity of ancient and modern populations in some parts of the Americas over at least the last 8.5 KYA, which is in agreement with recent results from Kennewick Man (51).

**Evidence of more distant Old World gene flow into some Native Americans**

When testing for gene flow between Athabascans and Inuit with masked SNP chip genotype data-based $D$-statistics (47) (fig. S8), we observed a weak tendency for the Inuit to be much closer to the Athabascans than to certain Amerindians like the North American Algonquin and Cree, and the Yaqui and Arhuaco of Central and South America (respectively), as compared to other Amerindians such as the Palikur and Surui of Brazil (fig. S8).

To further investigate this trend, we tested for additional gene flow from Eurasian populations into the Americas with $D$-statistics using the masked SNP chip genotype dataset (47). We found that some American populations, including the Aleutian Islanders, Surui, and Athabascans are closer to Australo-Melanesians compared to other Native Americans, such as North American Ojibwa, Cree and Algonquin, and the South American Purepecha, Arhuaco and Wayuu (fig. S10). The Surui are, in fact, one of closest Native American populations to East Asians and Australo-Melanesians, the latter including Papuans, non-Papuan Melanesians, Solomon Islanders, and South East Asian hunter-gatherers such as Aeta (fig. S10). We acknowledge that this observation is based on the analysis of a small fraction of the whole genome and SNP chip genotype datasets, especially for the Aleutian Islander data that is heavily masked due to recent admixture with Europeans (28), and that the trends in the data are weak.

Nonetheless, if it proves correct, these results suggest there may be a distant Old World signal related to Australo-Melanesians and East Asians in some Native Americans. The
widely scattered and differential affinity of Native Americans to the Australo-Melanesians, ranging from a strong signal in the Surui to much weaker signal in northern Amerindians such as Ojibwa, points to this gene flow occurring after the initial peopling by Native American ancestors.

However, how this signal may have ultimately reached South America remains unclear. One possible means is along a northern route via the Aleutian Islanders, previously found to be closely related to the Inuit (39), who have a relatively greater affinity to East Asians, Oceanians and Denisovan than Native Americans in both whole genome and SNP chip genotype data-based D-tests (table S10 and figs. S10 and S11). On the basis of archaeological evidence and mtDNA data from ancient and modern samples, the Aleutian Islands are hypothesized to have been peopled as early as ca. 9 KYA by ‘Paleo-Aleuts’ who were succeeded by the ‘Neo-Aleuts’, with present-day Aleutian Islanders potentially resulting from admixture between these two populations (52, 53). Perhaps their complex genetic history included input from a population related to Australo-Melanesians through an East Asian continental route, and this genomic signal might have been subsequently transferred to parts of the Americas, including South America, through past gene flow events (Fig. 1). Evidence for this gene flow is supported by diCal2.0 and MSMC analyses showing a weak but recent gene flow into South Americans from populations related to present-day Northeast Asians (Koryak) (Fig. 2C and table S11C), who might be considered a proxy for the related Aleutian Islanders.

**Testing the Paleoamerican model**

The detection of an Australo-Melanesian genetic signal in the Americas, however subtle, returns the discussion to the Paleoamerican model, which hypothesizes, on the basis of cranial morphology, that two temporally and source-distinct populations colonized the Americas. The earlier population reportedly originated in Asia in the Late Pleistocene and gave rise to both Paleoamericans and present-day Australo-Melanesians, whose shared cranial morphological attributes are presumed to indicate their common ancestry (23). The Paleoamericans were, in turn, thought to have been largely replaced by ancestors of present-day Amerindians, whose crania resemble modern East Asians and who are argued to be descendants of later arriving Mongoloid populations (14, 23, 26, 54). The presence of Paleoamericans is inferred primarily from ancient archaeological specimens in North and South America, and a few relict populations of more recent age, which include the extinct Pericúes and Fuego-Patagonians (24, 25, 55).

The Paleoamerican hypothesis predicts that these groups should be genetically closer to Australo-Melanesians than other Amerindians. Previous studies of mtDNA and Y chromosome data obtained from Fuego-Patagonian and Paleo-american skeletons have identified haplogroups similar to those of modern Native Americans (55–57). Although these results indicate some shared maternal and paternal ancestry with contemporary Native Americans, uniparental markers can be misleading when drawing conclusions about the demographic history of populations. To conclusively identify the broader population of ancestors who may have contributed to the Paleoamerican gene pool, autosomal genomic data are required.
We, therefore, sequenced 17 ancient individuals affiliated to the now-extinct Pericúes from Mexico and Fuego-Patagonians from Chile and Argentina (28), who, on the basis of their distinctive skull morphologies, are claimed to be relicts of Paleoamericans (23, 27, 58, 59). Additionally, we sequenced two pre-Columbian mummies from northern Mexico (Sierra Tarahumara) to serve as morphological controls, since they are expected to fall within the range of Native American morphological cranial variation (28). We found that the ancient samples cluster with other Native American groups and are outside the range of Oceanian genetic variation (28) (Fig. 5 and figs. S32, S33, and S34). Similarly, outgroup f\textsubscript{ST} statistics (47) reveal low shared genetic ancestry between the ancient samples and Oceanians (28) (Figs. S36, S37), and genome-based and masked SNP chip genotype data-based D-statistics (46, 47) show no evidence for gene flow from Oceanians into the Pericúes or Fuego-Patagonians (28) (fig. S39).

As the Paleoamerican model is based on cranial morphology (23, 27, 58, 59), we also measured craniometric data for the ancient samples and assessed their phenotypic affinities to supposed Paleoamericans, Amerindians and world-wide populations (28). The results revealed that the analyzed Fuego-Patagonians showed closest craniometric affinity to Arctic populations and the Paleoamericans, while the analyzed female Pericúes showed closest craniometric affinities to populations from North America, the Arctic region and Northern Japan (table S15). More importantly, our analyses demonstrated that the presumed ancestral ancient Paleoamerican reference sample from Lagoa Santa, Brazil (24) had closest affinities to Arctic and East Asian populations (table S15). Consequently, for the Fuego-Patagonians, the female Pericúes and the Lagoa Santa Paleoamerican sample, we were not able to replicate previous results (24) that report close similarity of Paleoamerican and Australo-Melanesian cranial morphologies. We note that male Pericúes samples displayed more craniometric affinities with populations from Africa and Australia relative to the female individuals of their population (fig. S41). The results of analyses based on craniometric data are, thus, highly sensitive to sample structure and the statistical approach and data filtering used (51). Our morphometric analyses suggest that these ancient samples are not true relicts of a distinct migration, as claimed, and hence do not support the Paleo-american model. Similarly, our genomic data also provide no support for an early migration of populations directly related to Australo-Melanesians into the Americas.

**Discussion**

That Native Americans diverged from their East Asian ancestors during the LGM and no earlier than 23 KYA provides an upper bound, and perhaps the climatic and environmental context, for the initial isolation of their ancestral population, and a maximum estimate for the entrance and subsequent spread into the Americas. This result is consistent with the model that people entered the Americas prior to the development of the Clovis complex and had reached as far as southern South America by 14.6 KYA. As archaeological evidence provides only a minimum age for human presence in the Americas, we can anticipate the possible discovery of sites that approach the time of the divergence of East Asians and Native Americans. However, our estimate for the initial divergence and entry of Native American ancestors does not support archaeological claims for an initial peopling significantly earlier than the LGM (8–10).
While our data cannot provide the precise geographical context for the initial peopling process, it has allowed us to more accurately estimate its temporal dynamics. This, in turn, has enabled us to re-assess the Beringian Incubation Model, which, based on mtDNA data and the timing and geographical distribution of archaeological sites, hypothesized a ca. 15,000 year-long period of isolation of ancestral Native Americans in Beringia during the LGM (19–21). Our results, along with recent findings of mtDNA haplogroup C1 in Iceland and ancient northwest Russia (60), do not fit with the proposed 15,000-year span of the Beringian Incubation Model (19–21). It is possible that a shorter period of isolation occurred (ca. 8 KYA), but whether it occurred in Siberia or Beringia will have to be determined by future ancient DNA and archaeological findings. Given the genetic continuity between Native Americans and some East Asian populations (figs. S4 and S5), other demographic factors, such as surfing during population expansions into unoccupied regions (61), may ultimately need to be taken into account to better understand the presence of a large number of high frequency private variants in the indigenous populations of the Americas.

The data presented here are consistent with a single initial migration of all Native Americans and with later gene flow from sources related to East Asians and, more distantly, Australo-Melanesians. From that single migration, there was a diversification of ancestral Native Americans leading to the formation of ‘northern’ and ‘southern’ branches, which appears to have taken place ca. 13 KYA within the Americas. This split is consistent with the patterns of uniparental genomic regions of mtDNA haplogroup X and some Y chromosome C haplotypes being present in northern, but not southern, populations in the Americas (18, 62). This diversification event coincides roughly with the opening of habitable routes along the coastal and the interior corridors into unglaciated North America some 16 KYA and 14 KYA, respectively (63, 64), suggesting a possible role of one or both these routes in the isolation and subsequent dispersal of Native Americans across the continent.

**Methods**

DNA was extracted from 31 present-day individuals from the Americas, Siberia and Oceania and 23 ancient samples from the Americas, and converted to Illumina libraries and shotgun-sequenced (28). Three of the ancient samples were radiocarbon dated, of which two were corrected for marine reservoir offset (28). SNP chip genotype data was generated from 79 present-day Siberians and Native Americans affiliated to 28 populations (28). Raw data from SNP chip and shotgun sequencing were processed using standard computational procedures (28). Error rate analysis, DNA damage analysis, contamination estimation, sex determination, mtDNA and Y chromosome haplogroup assignment, ADMIXTURE analysis, ancestry painting and admixture masking, Principal Component Analysis using SNP chip genotype data, *TreeMix* analysis on genomic sequence data, *D*-statistic and outgroup *f*<sub>T</sub>-statistic tests on SNP chip genotype and genomic sequence data, divergence time estimation using *diCal2.0*, an IBS tract method and MSMC, Climate-Informed Spatial Genetic Model analysis, and, craniometric analysis were performed as described (28).

**Supplementary Material**

Refer to Web version on PubMed Central for supplementary material.
Authors

Maanasa Raghavan¹, Matthias Steinrücken²,³,⁴, Kelley Harris⁵, Stephan Schiffels⁶, Simon Rasmussen⁷, Michael DeGiorgio⁸, Anders Albrechtsen⁹, Cristina Valdiosera¹⁰,¹¹, Maria C. Ávila-Arco¹,¹¹, Anna-Sapfo Malaspinas¹, Anders Eriksson¹²,¹³, Ida Moltke⁹, Mait Metspalu¹⁴,¹⁵, Julian R. Homburger¹¹, Jeff Wall¹⁶, Omar E. Cornejo¹⁷, J. Victor Moreno-Mayar¹, Thorfinn S. Korneliussen¹, Tracey Pierre¹, Morten Rasmussen¹,¹¹, Paula F. Campos¹,¹⁸, Peter de Barros Damgaard¹, Morten E. Allentoft¹, John Lindo¹⁹, Ene Metspalu¹⁴,¹⁵, Ricardo Rodríguez-Varela²⁰, Josefina Mansilla²¹, Celeste Henricksson²², Andaine Seguin-Orlando¹, Helena Malmström²³, Thomas Stafford Jr.¹,²⁴, Suyash S. Shringarpure¹¹, Andrés Moreno-Estrada¹¹,¹²,¹⁵, Monika Karmin¹⁴,¹⁵, Kristiina Tambets¹⁴, Anders Bergström⁶, Yali Xue⁶, Vera Warmuth²⁶,²⁷, Andrew D. Friend²⁸, Joy Singarayer²⁹, Paul Valdes³⁰, Francois Balloux²⁶, Ilán Leboeiro²¹, Jose Luis Vera³¹, Hector Rangel-Villalobos³², Davide Pettener³³, Donata Luiselli³³, Loren G. Davis³⁴, Evelyne Heyer³⁵, Christoph P. E. Zollikofer³⁶, Marcia S. Ponce de León³⁶, Colin I. Smith¹⁰, Vaughan Grimes³⁷,³⁸, Kelly-Anne Pike³⁷, Michael Deal³⁷, Benjamin T. Fuller³⁹, Bernardo Arriaza⁴⁰, Vivien Standen⁴¹, Maria F. Luz⁴², Francois Ricaut⁴³, Niede Guidon⁴⁴, Ludmila Osipova⁴⁴,⁴⁵, Mikhail I. Voevoda⁴⁴,⁴⁶, Olga L. Posukh⁴⁴,⁴⁵, Oleg Balanovsky⁴⁸,⁴⁹, Maria Lavryshina⁵⁰, Yuri Bogunov⁴⁸, Elza Khusnutdinova⁵¹,⁵², Marina Gubina⁴², Elena Balanovskaya⁴⁹, Sardana Fedorova⁵³,⁵⁴, Sergey Litvinov¹⁴,⁵¹, Boris Malyarchuk⁵⁵, Miroslava Derenko⁵⁵, M. J. Mosher⁵⁶, David Archer⁵⁷, Jerome Cybulski⁵⁸,⁵⁹,⁶⁰, Barbara Petzelt⁶¹, Joycelyn Mitchell⁶¹, Rosita Wor⁶², Paul J. Norman⁶³, Peter Parham⁶³, Brian M. Kemp¹⁷,⁶⁴, Toomas Kivisild¹⁴,⁶⁵, Chris Tyler-Smith⁶, Manjinder S. Sandhu⁶,⁶⁶, Michael Crawford⁶⁷, Richard Villems¹⁴,¹⁵, David Glenn Smith⁶⁸, Michael R. Waters⁶⁹,⁷⁰,⁷¹, Ted Goebel⁶⁹, John R. Johnson⁷², Ripan S. Malhi⁷³,⁷⁴, Mattias Jakobsson²³, David J. Meltzer¹,⁷⁴, Andrea Manica¹², Richard Durbin⁶, Carlos D. Bustamante¹¹, Yun S. Song²,³,⁷⁵,†, Rasmus Nielsen⁷⁵,†, and Eske Willerslev¹,†

Affiliations

¹Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5–7, 1350 Copenhagen, Denmark. ²Computer Science Division, University of California, Berkeley, CA 94720, USA. ³Department of Statistics, University of California, Berkeley, CA 94720, USA. ⁴Department of Biostatistics and Epidemiology, University of Massachusetts, Amherst, MA 01003, USA. ⁵Department of Mathematics, University of California, Berkeley, CA 94720, USA. ⁶Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SA, UK. ⁷Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, København, Building 208, 2800 Kongens Lyngby, Denmark. ⁸Departments of Biology and Statistics, Pennsylvania State University, 502 Wartik Laboratory, University Park, PA 16802, USA. ⁹The Bioinformatics Centre, Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, 2200 Copenhagen, Denmark. ¹⁰Department of Archaeology and History, La Trobe University, Melbourne, Victoria 3086, Australia. ¹¹Department of
Genetics, School of Medicine, Stanford University, 300 Pasteur Dr. Lane Bldg Room L331, Stanford, California 94305, USA. 12Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK. 13Integrative Systems Biology Laboratory, King Abdullah University of Science and Technology (KAUST), Thuwal, 23955-6900, Kingdom of Saudi Arabia. 14Estonian Biocentre, Evolutionary Biology Group, Tartu 51010, Estonia. 15Department of Evolutionary Biology, University of Tartu, Tartu 51010, Estonia. 16Institute for Human Genetics, University of California San Francisco, 513 Parnassus Avenue, San Francisco, CA 94143, USA. 17School of Biological Sciences, Washington State University, PO Box 644236, Heald 429, Pullman, Washington 99164, USA. 18CIMAR/CIIMAR, Centro Interdisciplinar de Investigação Marinha e Ambiental, Universidade do Porto, Rua dos Bragas 289, 4050-123 Porto, Portugal. 19Department of Anthropology, University of Illinois at Urbana-Champaign, 607 S. Mathews Ave, Urbana, IL 61801, USA. 20Centro Mixto, Universidad Complutense de Madrid–Instituto de Salud Carlos III de Evolución y Comportamiento Humano, Madrid, Spain. 21Instituto Nacional de Antropología e Historia, Moneda 13, Centro, Cuauhtémoc, 06060 Mexico City, Mexico. 22University of Utah, Department of Anthropology, 270 S 1400 E, Salt Lake City, Utah 84112, USA. 23Department of Evolutionary Biology and Science for Life Laboratory, Uppsala University, Norbyvägen 18D, SE-752 36 Uppsala, Sweden. 24AMS 14C Dating Centre, Department of Physics and Astronomy, Aarhus University, Ny Munkegade 120, 8000 Aarhus, Denmark. 25Laboratorio Nacional de Genómica para la Biodiversidad (LANGEBIO), CINVESTAV, Irapuato, Guanajuato 36821, Mexico. 26UCL Genetics Institute, Gower Street, London WC1E 6BT, UK. 27Evolutionsbiologiskt Centrum, Norbyvägen 18D, 75236 Uppsala, Sweden. 28Department of Geography, University of Cambridge, Downing Place, Cambridge CB2 3EN, UK. 29Centre for Past Climate Change and Department of Meteorology, University of Reading, Earley Gate, PO Box 243, Reading, UK. 30School of Geographical Sciences, University Road, Clifton, Bristol BS8 1SS, UK. 31Escuela Nacional de Antropología e Historia, Periférico Sur y Zapote s/n. Colonia Isidro Fabela, Tlalpan, Isidro Fabela, 14030 Mexico City, Mexico. 32Instituto de Investigación en Genética Molecular, Universidad de Guadalajara, Ocotlán, Mexico. 33Dipartimento di Scienze Biologiche, Geologiche e Ambientali (BiGeA), Università di Bologna, Via Selmi 3, 40126 Bologna, Italy. 34Department of Anthropology, Oregon State University, 238 Waldo Hall, Corvallis, OR, 97331 USA. 35Museum National d’Histoire Naturelle, CNRS, Université Paris 7 Diderot, Sorbonne Paris Cité, Sorbonne Universités, Unité Eco-Anthropologie et Ethnobiologie (UMR7206), Paris, France. 36Anthropological Institute and Museum, University of Zürich, Winterthurstrasse 190, 8057 Zürich, Switzerland. 37Department of Archaeology, Memorial University, Queen’s College, 210 Prince Philip Drive, St. John’s, Newfoundland, A1C 5S7, Canada. 38Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, Leipzig 04103, Germany. 39Department of Earth System Science, University of California, Irvine, Keck CCAMS Group, B321 Croul Hall, Irvine, California, 92697, USA. 40Instituto de Alta Investigación, Universidad de Tarapacá,
18 de Septiembre 2222, Carsilla 6-D Arica, Chile. 41Departamento de Antropologia, Universidad de Tarapacá, 18 de Septiembre 2222. Casilla 6-D Arica, Chile. 42Fundação Museu do Homem Americano, Centro Cultural Sérgio Motta, Campestre, 64770-000 Sao Raimundo Nonato, Brazil. 43Laboratoire d’Anthropologie Moléculaire et Imagérie de Synthèse UMR-5288, CNRS, Université de Toulouse, 31073 Toulouse, France. 44Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, Prospekt Lavrentyeva 10, 630090 Novosibirsk, Russia. 45Novosibirsk State University, 2 Pirogova Str., 630090 Novosibirsk, Russia. 46Institute of Internal Medicine, Siberian Branch of RAS, 175/1 ul. B. Bogatkova, Novosibirsk 630089, Russia. 47Novosibirsk State University, Laboratory of Molecular Epidemiology and Bioinformatics, 630090 Novosibirsk, Russia. 48Vavilov Institute of General Genetics, Gubkina 3, 119333 Moscow, Russia. 49Research Centre for Medical Genetics, Moskvorechie 1, 115478 Moscow, Russia. 50Kemerovo State University, Krasnaya 3, 650000 Kemerovo, Russia. 51Institute of Biochemistry and Genetics, Ufa Scientific Center of RAS, Prospekt Oktyabrya 71, 450054 Ufa, Russia. 52Department of Genetics and Fundamental Medicine, Bashkir State University, Zaki Validi 32, 450076 Ufa, Russia. 53Department of Molecular Genetics, Yakut Scientific Centre of Complex Medical Problems, Sergelyahskoe Shosse 4, 677010 Yakutsk, Russia. 54Laboratory of Molecular Biology, Institute of Natural Sciences, M.K. Ammosov North-Eastern Federal University, 677000 Yakutsk, Russia. 55Institute of Biological Problems of the North, Russian Academy of Sciences, Portovaya Street 18, Magadan 685000, Russia. 56Department of Anthropology, Western Washington University, Bellingham Washington 98225, USA. 57Department of Anthropology, Northwest Community College, 353 Fifth Street, Prince Rupert, British Columbia V8J 3L6, Canada. 58Canadian Museum of History, 100 Rue Laurier, Gatineau, Quebec K1A 0M8, Canada. 59University of Western Ontario, London, Ontario N6A 3K7, Canada. 60Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada. 61Metlakatla Treaty Office, PO Box 224, Prince Rupert, BC, Canada V8J 3P6. 62Sealaska Heritage Institute, 105 S. Seward Street, Juneau, Alaska 99801, USA. 63Department of Structural Biology, Stanford University School of Medicine, D100 Fairchild Science Building, Stanford, California 94305-5126, USA. 64Department of Anthropology, Washington State University, Pullman Washington 99163, USA. 65Division of Biological Anthropology, University of Cambridge, Henry Wellcome Building, Fitzwilliam Street, CB2 1QH, Cambridge, UK. 66Dept of Medicine, University of Cambridge, MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge CB2 0QH, UK. 67Laboratory of Biological Anthropology, University of Kansas, 1415 Jayhawk Blvd., 622 Fraser Hall, Lawrence, Kansas 66045, USA. 68Molecular Anthropology Laboratory, 209 Young Hall, Department of Anthropology, University of California, One Shields Avenue, Davis, California 95616, USA. 69Center for the Study of the First Americans, Texas A&M University, College Station, Texas 77843-4352, USA. 70Department of Anthropology, Texas A&M University, College Station, Texas 77843-4352, USA. 71Department of Geography, Texas A&M University, College
ACKNOWLEDGMENTS

We thank J. Valdés for providing craniometric measurements of the Pericúes at the National Museum of Anthropology in México; A. Monteverde from CINAH-Baja California Sur and V. Laborde at the Musée de l’Homme in Paris for providing documentation on Pericú and Fuego-Patagonian samples, respectively; T. Gilbert, M. McCoy, C. Sarkissian, M. Sikora, L. Orlando for helpful discussions and input; D. Yao and C. Barbieri for helping with the collection of the Aymara population sample; B. Henn and J. Kidd for providing early access to the Mayan sequencing data; Canadian Museum of History; Metlakatla and Lax Kw’alaams First Nations; Listuguj Mi’gmaq Band Council; A. Pye of TERRA Facility, Core Research Equipment & Instrument Training (CREAIT) Network at Memorial University; the Danish National High-throughput DNA Sequencing Centre (Copenhagen) for help with sequencing; and, Fondation Jean Dausset-Centre de Étude du Polymorphism Humain (CEPH) for providing DNA for the Human Genome Diversity Project (HGDP) samples that were genome-sequenced in this study. This study was supported by several funding bodies; Lundbeck Foundation and the Danish National Research (Centers for GeoGenetics members), Wellcome Trust grant 098051 (S.S., A.B., Y.X., C.T.-S., M.S.S., R.D.), Marie Curie Intra-European Fellowship FP7-People-PIEF-GA-2009-255503 and the Transforming Human Societies Research Focus Areas Fellowship from La Trobe University (C.V.), George Rosenkranz Prize for Health Care Research in Developing Countries and National Science Foundation award DMS-1201234 (M.C.A.A.), Swiss National Science Foundation (PBSSP3_143529) (A.-S.M.), Ministerio de Ciencia e Innovación (MICINN) Project CGL2009-12703-C03-03 and MICINN (BES-2010-030127) (R.R.-V.), Consejo Nacional de Ciencia y Tecnología (Mexico) (J.V.M.M.), Biotechnology and Biological Sciences Research Council BB/H005854/1 (V.W., F.B., A.M.), European Research Council and Marie Curie Actions Grant 300554 (M.E.A.), Wenner-Gren Foundations and the Australian Research Council Future Fellowship FT0992258 (C.I.S.), European Research Council ERC-2011-AdG 295733 (D.P. and D.L.), Bernice Peltier Huber Charitable Trust (C.H., L.G.D.), Russian Foundation for Basic Research grant 13-06-00670 (E.B.), Russian Foundation for Basic Research grant 14-0400725 (E.K.), European Union European Regional Development Fund through the Centre of Excellence in Genomics to Estonian Biocentre and Estonian Institutional Research_body grant IUT24-1 (E.M., K.T., M.M., K.R.V.), Estonian Science Foundation grant 8973 (M.M.), Stanford Graduate Fellowship (J.R.H.), Washington State University (B.M.K.), French National Research Agency grant ANR-14-CE31-0013-01 (F-X.R), European Research Council grant 262123 (T.K.), National Science Foundation BCS-1025139 (R.S.M.), Social Science Research Council of Canada (K.-A.P., V.G.), National Institutes of Health grants R01-GM094402 (M.S., Y.S.S.), R01- AI17892 (F.J.C., P.P.); R01 HG005229-09 (R.N., C.D.B.), Packard Fellowship for Science and Engineering (Y.S.S.), Russian Science Fund grant 14-04-00827 and Presidium of Russian Academy of Sciences Molecular and Cell Biology Programme (O.B.), and Russian Foundation for Basic Research grant 14-06-00384 (Y.B.). Informed consent was obtained for the sequencing of the modern individuals, with ethical approval from The National Committee on Health Research Ethics, Denmark (H-3-2012-FSP21). SNP chip genotype data and whole genome data for select present-day individuals are available only for demographic research under data access agreement with E.W. (see Tables S1 and S4 for a list of these samples). Raw reads from the ancient and the remainder of the present-day genomes are available for download through European Nucleotide Archive (ENA) accession no. PRJEB9733, and the corresponding alignment files are available at http://www.cbs.dtu.dk/suppl/NativeAmerican/. The remainder of the SNP chip genotype data can be accessed through Gene Expression Omnibus (GEO) series accession no. GSE70987 and at www.ebc.ee/free_data. C.D.B. is on the advisory board of Personalis, Inc.; Identify Genomics; Etalon DX; and Ancestry.com.

REFERENCES AND NOTES


28. Materials and methods are available as supplementary materials on Science Online.


37. Raghavan et al. Page 16


Science. Author manuscript; available in PMC 2016 February 01.


146. ISOGG. [http://www.isogg.org]


199. Morimoto N, de León M. S. Ponce, Zollikofer CP. Phenotypic variation in infants, not adults, reflects genotypic variation among chimpanzees and bonobos. PLOS ONE. 2014; 9:e102074.10.1371/journal.pone.0102074 [PubMed: 25013970]


Fig. 1. Origins and population history of Native Americans

(A) Our results show that the ancestors of all present-day Native Americans, including Amerindians and Athabascans, derived from a single migration wave into the Americas (purple), separate from the Inuit (green). This migration from East Asia occurred no later than 23 KYA and is in agreement with archaeological evidence from sites such as Monte Verde (50). A split between the northern and southern branches of Native Americans occurred ca. 13 KYA, with the former comprising Athabascans and northern Amerindians and the latter consisting of Amerindians in northern North America and Central and South America including the Anzick-1 individual (5). There is an admixture signal between Inuit and Athabascans and some northern Amerindians (yellow line); however, the gene flow
direction is unresolved due to the complexity of the admixture events (28). Additionally, we see a weak signal related to Australo-Melanesians in some Native Americans, which may have been mediated through East Asians and Aleutian Islanders (yellow arrows). Also shown is the Mal’ta gene flow into Native American ancestors some 23 KYA (yellow arrow) (4). It is currently not possible for us to ascertain the exact geographical locations of the depicted events; hence, the positioning of the arrows should not be considered a reflection of these. B. Admixture plot created on the basis of TreeMix results (fig. S5) shows that all Native Americans form a clade, separate from the Inuit, with gene flow between some Native Americans and the North American Arctic. The number of genome-sequenced individuals included in the analysis is shown in brackets.
Fig. 2. Divergence estimates between Native Americans and Siberian Koryak
(A) The demographic model used allows for continuous gene flow between populations 1 and 2, starting from the time $T_{DIV}$ of divergence and ending at $T_M$. The backward probability of migration per individual per generation is denoted by $m$. The bottleneck at $T_B$ captures the out-of-Africa event. (B) The red and black solid curves depict empirical distributions of IBS tracts shared between Karitiana-Koryak and Athabascan-Koryak, respectively. The orange, pink, dashed blue and dashed green curves depict IBS tracts shared between the two population pairs, simulated under two demographic models based on results from $diCal2.0$. Overall, for Karitiana-Koryak and Athabascan-Koryak, the migration scenarios (orange and pink, respectively) match the empirical curves (red and black, respectively) better than the clean split scenarios (dashed blue and dashed green, respectively), with more long IBS tracts showing evidence of recent common ancestry between Koryaks and Native Americans. (C and D) Relative cross coalescence rates (CCR) for the Karitiana-Koryak and Athabascan-Koryak divergence (red), respectively, including data simulated under the two demographic models in panel B. In both cases, the model with gene flow (orange) fits the data (red) better than the clean split model (blue). The migration model explains a broader CCR tail in the
case of Karitiana-Koryak and the relatively late onset of the CCR decay for Athabascan-Koryak.
Fig. 3. Testing migrations into the Americas using a climate-informed model
Estimates of difference in genetic divergence between Amerindians (from southern North America and Central and South America) or Koryak versus Athabascan and Greenlandic Inuit and the ancient Saqqaq and Anzick-1 genomes (black vertical lines), compared to posterior probability distribution predicted from a climate-informed spatial genetic model reconstructing a single wave into the Americas (curves, the colored part represents the 95% credibility interval). ΔT for population X is defined as T(X,Koryak)-T(X,Central and South Amerindians) (28). Both Anzick-1 and the Athabascans were part of the same wave into the Americas to which other Amerindian populations from southern North America and Central and South America belonged, while the Inuit and Saqqaq are the descendants of different waves (observed values outside the 95% credibility interval).
Fig. 4. Diversification within the Americas

SNP chip genotype data-based outgroup $f_3$ statistics (47) of the form $f_3(X, \text{Ancient}; \text{Yoruba})$ were used to estimate the shared ancestry between ancient samples from the Americas and a large panel of worldwide present-day populations (X), including Athabascan and Amerindian groups from North America (table S3), some of which were masked for non-Native ancestry prior to the analysis (28). The outgroup $f_3$ statistics are depicted as heat maps with the sampling location of the ancient sample marked by the dotted lines, and corresponding ranked plots with error bars are shown in fig. S14. BP refers to time before
present. We find the Anzick-1 sample to share most ancestry with the ‘southern’ branch of Native Americans when using multiple northern Native Americans sequenced in this study, consistent with (5). The seven Holocene aged samples share most ancestry with Native Americans, with a general tendency to be genetically closer to present-day Native American populations from the same geographical region.
Fig. 5. The Paleoamerican model

(A) Principal Component Analysis plot of 19 ancient samples combined with a worldwide reference panel, including 1,823 individuals from (6). Our samples plot exclusively with American samples. For plots with other reference panels consisting of Native American populations, see fig. S32. (B) Population structure in the ancient Pericú, Mexican mummy and Fuego-Patagonian individuals from this study. Ancestry proportions are shown when assuming six ancestral populations (K = 6). The top bar shows the ancestry proportions of the 19 ancient individuals, Anzick-1 (5), and two present-day Native American genomes.
from this study (Huichol and Aymara). The plot at the bottom illustrates the ancestry proportions for 1,823 individuals from (6). Our samples show primarily Native American (ivory, >92%) and Siberian (red, ca. 5%) ancestry. For the plot with $K$ =13, see fig. S33.