Reconstructing Native American population history.

David Reich, Nick Patterson, Desmond Campbell, Arti Tandon, Stéphane Mazieres, Nicolas Ray, Maria V Parra, Winston Rojas, Constanza Duque, Natalia Mesa, et al.

Reconstructing Native American Population History

David Reich[1,2,*], Nick Patterson[2], Desmond Campbell[3], Arti Tandon[1,2], Stéphane Mazieres[3,4], Nicolas Ray[5], Maria V. Parra[3,6], Winston Rojas[3,6], Constanza Duque[3,6], Claudio M. Bravi[3,7], Graciella Bailliet[7], Daniel Corach[8], Tábita Hünemeier[3,9], Maria-Cátira Bortolini[9], Francisco Salzano[9], María Luiza Petzl-Erler[10], Victor Acuña-Alonzo[11], Samuel Canizales-Quinteros[12,13], Carlos Aguilar-Salinas[12], Teresa Tusié-Luna[12], Laura Riba[12], Maricela Rodríguez-Cruz[14], Mardia Lopez-Alarcón[14], Ramón Coral-Vazquez[15], Thelma Canto-Cetina[16], Julio Molina[17], Ángel Carracedo[18], Antonio Salas[18], Carla Gallo[19], Giovanni Poletti[19], David B. Witonsky[20], Gorka Alkorta-Aranburu[20], Rem Sukernik[21], Ludmila Osipova[22], Sardana Fedorova[23], René Vasquez[24], Mercedes Villena[24], Damian Labuda[25], Ramiro Barrantes[26], Laurent Excoffier[27], Gabriel Bedoya[6], Francisco Rothhammer[28], Jean Michel Dugoujon[29], Georges Larrouy[29], David Pauls[30], William Klitz[31], Judith Kidd[32], Kenneth Kidd[32], Anna Di Rienzo[20], Nelson B. Freimer[33], Alkes L. Price[2,34] and Andrés Ruiz-Linares[3,*]

[1] Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA
[2] Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA
[3] Department of Genetics, Evolution and Environment. University College London, UK
[4] Anthropologie Bioculturelle, UMR 6578, Université de la Méditerranée/CNRS/EFS, Marseille, France
[5] EnviroSPACE Laboratory, Climate Change and Climatic Impacts, Institute for Environmental Sciences, University of Geneva, Carouge, Switzerland
[6] Laboratorio de Genética Molecular, Universidad de Antioquia, Medellín, Colombia,
[7] : Instituto Multidisciplinario de Biologia Celular, La Plata, Argentina
[8] Servicio de Huellas Digitales Genéticas, Universidad de Buenos Aires, Argentina
[9] Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil
[10] Departamento de Genética, Universidade Federal do Paraná, Curitiba Brazil
[11] National Institute of Anthropology and History, Mexico City, México
[12] Unit of Molecular Biology and Genomic Medicine, Instituto Nacional de Ciencias Médicas y Nutrición, México City, México
[13] Department of Biology, Facultad de Química, Universidad Nacional Autónoma de México, Mexico City, México
[14] Hospital de Pediatría, Centro Médico Nacional, MSS, México City, México.
[15] Sección de Posgrado, Escuela Superior de Medicina del Instituto Politécnico Nacional & C.M.N. 20 de Noviembre-ISSSTE, México City, México.
[16] Laboratorio de Biología de la Reproducción, Departamento de Salud Reproductiva y Genética, Centro de Investigaciones Regionales, Mérida Yucatán, México
[17] Centro de Investigaciones Biomédicas de Guatemala, Ciudad de Guatemala, Guatemala
[18] Instituto de Ciencias Forenses, Universidade de Santiago de Compostela, Fundación de Medicina Xenómica (SERGAS), CIBERER, Santiago de Compostela, Galicia, Spain.
[19] Laboratorios de Investigación y Desarrollo, Facultad de Ciencias y Filosofía, Universidad Peruana Cayetano Heredia, Lima, Perú.
[20] Department of Human Genetics, University of Chicago, Chicago, USA
[21] Laboratory of Human Molecular Genetics, Institute of Chemical Biology and Fundamental Medicine, Siberian Branch of the Russian Academy of Sciences, Novosibirsk Russia.

[22] Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, Novosibirsk Russia.
[23] Department of Molecular Genetics, Yakut Research Center of Complex Medical Problems, Yakutsk, Sakha (Yakutia), Russia
[24] Instituto Boliviano de Biología de la Altura. La Paz-Potosí, Bolivia.
[25] Département de Pédiatrie, Centre de Recherche du CHU Sainte-Justine, Université de Montréal, Montréal, Quebec, Canada
[26] Escuela de Biología, Universidad de Costa Rica, San José, Costa Rica
[27] Computational and Molecular Population Genetics Lab, Institute of Ecology and Evolution, University of Bern, Switzerland
[28] Facultad de Medicina, Universidad de Chile, Santiago and Instituto de Alta Investigación, Universidad de Tarapacá, Arica, Chile
[29] Anthropologie Moléculaire et Imagerie de Synthèse, CNRS UMR 5288, Université Paul Sabatier Toulouse III, Toulouse, France
[30] Center for Human Genetic Research, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA
[31] School of Public Health, University of California Berkeley, and Public Health Institute, Oakland, California, USA
[32] Department of Genetics, Yale University School of Medicine, New Haven, Connecticut, USA.
[33] Center for Neurobehavioral Genetics, University of California Los Angeles, Los Angeles, California, USA
[34] Departments of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, Massachusetts, USA

* To whom correspondence should be addressed: E-mail: reich@genetics.med.harvard.edu (D.R); a.ruizlin@ucl.ac.uk (A. R.-L.)

There is intense debate about whether all Native Americans stem from one migration or multiple waves of migration from Asia. In addition, little is known about the principal settlement routes and patterns of population diversification within the Americas. We assembled a dataset of 55 Native American and 19 Siberian populations typed at over 370,000 polymorphisms, the most comprehensive survey of genetic diversity in Native Americans to date, and masked out segments of recent European or African ancestry. Along with providing genetic support for controversial linguistic evidence for three episodes of migration from Asia, the data provide strong evidence for a southward population expansion (facilitated by the coast) with sequential splits and little gene flow after divergence. An important exception to this pattern is the history of Chibchan-speakers around the Panama isthmus, who our data suggest derive from a >5,000 year old mixture of South American and North American lineages, highlighting the isthmus as a region of genetic interaction between both hemispheres. Our results refute recent interpretations of mitochondrial DNA (mtDNA) positing a single settlement wave. They also highlight how genome-wide analyses of data directly accounting for the confounder of non-Native admixture can be used to document previously unknown historical events.

The initial peopling of the Americas occurred at least 15,000 years ago[1-3] through Beringia, a land bridge between Asia and America that existed during the ice ages, but there is controversy about whether Native Americans descend from a single[4-8], or multiple waves of migration[9-14], and even less is known about subsequent population movements. Most continent-wide analyses of Native American genetics have examined mtDNA[4-7] and the non-recombining portion of the Y-chromosome[11-13], but studies of large numbers of loci simultaneously can provide a much higher resolution view of history. We assembled samples of Native American populations from Canada to the southern tip of South America[15], genotyped them, and merged with five previously collected datasets. The final dataset consisted of >370,000 SNPs genotyped in 55 Native American populations with the lowest density being in the United States and Canada (475 samples; Figure 1 and Table S1), 19 Siberian populations (255 samples) (Figure S1 and Table S2) and 58 other populations (1,626 samples) (Note S1).

An immediate complication in studying the genetic history of Native Americans is gene flow from European and African immigrants in the last 500 years (Figure 1B and Figure S2). To address this confounder, we used the data to infer ancestry at each segment of the genome and

"masked" segments with non-Native American ancestry (Figure S3)[8]; the resulting dataset shows no evidence of African or European ancestry (Figure 1B; Figure S2). We applied a similar procedure to 19 Siberian and 2 Greenland Inuit populations (we did not apply it to the Aleutian populations who we found to be too admixed and thus excluded them from subsequent analyses) (Note S2). A potential concern is that the masking could bias the subsets of the genome we used for our analysis. Encouragingly, when we repeated a key analysis (population mixture in people around the isthmus of Panama) using unmasked data in which we explicitly modeled post-Colombian admixture, we obtained qualitatively identical inferences (Figure S4), encouraging us in the used of the masked data for subsequent analyses.

We first built a tree based on allele frequency differentiation (F_{ST} distances) between all pairs of populations (Table S3). This demonstrates remarkable agreement with geographic and linguistic classifications (Figure 1C). The first split (A) separates Asian populations from all New World populations along with the Siberian Inuit (Naukan). This monophyly agrees with mtDNA, Y-chromosome and other single-locus studies[16] that have identified pan-American variants of relatively recent origin, and is consistent with some shared Asian ancestry for all Native Americans[4-8]. Within the New World, an early split (B) separates Inuit from all other Native Americans. Among non-Arctic Native Americans, there follows a series of splits in an approximately north-to-south sequence, starting with a northern North American cluster and ending in a large group including four clusters from major geographic/linguistic subdivisions in lower Central America and South America. The first (#1) consists of Andean populations except the Inga. The second (#2) comprises populations from the Chaco region in southern South America. The third (#3) includes Equatorial-Tucanoan and Ge-Pano-Carib populations of eastern South America. The fourth (#4) includes predominantly Chibchan-Paezan-speaking populations of the Isthmo-Colombian area. This sequence of splits suggests settlement in a North-to-South expansion, which is also supported by a negative correlation between heterozygosity and distance from the Bering Strait ($r = -0.37$, $P = 0.04$). The correlation strengthens using "least cost distances" that consider the coasts as facilitators of migration[17-19] (Note S3; Figure S5). A second striking feature of the tree is the long population-specific branches, reflecting strong genetic drift. Analysis of linkage disequilibrium (LD) suggests recent bottlenecks explain part of the pattern: LD occurs on a scale that would be expected from bottlenecks 300-750 years ago especially in the Isthmo-Colombian and eastern South American areas (Note S4; Table S4).

Bifurcating trees provide a simplified view of history, in that they do not allow for the possibility of mixture across clades in the tree. To test whether the Neighbor Joining tree of Figure 1C provides an accurate description of the population relationships, we used the *4 Population Test*[20], which evaluates whether allele frequencies in any set of four populations are consistent with a proposed tree. We first tested the commonly held view that Native American and East Asian populations have a common origin with no migration since their split from Europeans and Africans by testing the tree ((Yoruba,French),(Han,Native American)) (Figure 2A). We reject this tree with high statistical significance for all 55 Native American populations: $|Z|>6.0$ ($P < 2\times10^{-9}$), with the sign of the *4 Population Test* statistic indicating that Europeans are more closely related to Native Americans than to East Asians. The values of the statistic are very similar for the 52 non-Arctic populations (0.027 ± 0.002), indicating that the signal does not reflect gene flow in the Americas (and hence we do not focus on it in this study), but instead, within Eurasia itself. Future studies that model the joint demographic history of Europeans, East Asians and Native Americans[21] need to take this complexity into account.

We next used the *4 Population Test* to evaluate whether Native American populations descend from a single, discrete, migration event[4-8]. We studied all possible pairs of 55 Native American populations, testing whether they represent sister groups after splitting from carefully chosen outgroups (Figure 2B). First, we evaluated whether the Inuit descend from the same Asian migration as all other Native American populations by testing ((Yoruba, Han),(Native American, Inuit)), and reject it at $|Z|>4.5$ for all pairs of Native American and Inuit populations that we tested, indicating that the Inuit are more closely related to Asians (Han) than the non-Arctic Native Americans (Figure 2B). Second, we evaluated whether data from the 52 non-Arctic Native Americans are consistent with descending from a discrete migration from Asia with no subsequent gene flow, by applying the *4 Population Test* to the tree ((*Outgroup1*, *Outgroup2*), (*NativeAmerican1*, *NativeAmerican2*)), using 10 different pairs of Asian and Arctic outgroups (Figure 2C and Table S5). The 47 most southern Native American populations are consistent with descending from a single peopling event (all statistics $|Z|<3$; Table S5). However, 5 Northern Native American (NNA) populations—Ojibwa, Cree, Algonquin, Cheyenne and Chipewyan—have Z-scores 3-6 standard errors from expectation, and are also outliers in population structure analyses (Figure 1B and Figure 2). Further examination of the values of the *4 Population Test* statistics demonstrates two distinct patterns of relationships to Arctic and East

Asian populations among these 5 NNA groups. The statistics for four of the NNA (Cheyenne, Cree, Ojibwa and Algonquin) are highly correlated (average r^2=0.72; Figure S6) and indicate a closer relationship of these populations to the Inuit than to any Asian group (Figure 2B and Table S6). By contrast, statistics involving the Chipewyan are not correlated to the other four NNA (r^2=0.05; Figure 2C; Table S6), suggesting distinct gene flows with Asians. Globally, these findings show that Native Americans break into three broad groups: the 47 Native American populations from Meso-America southward, the Inuit along with 4 NNA populations with whom they appear to have exchanged genes, and the Chipewyan who speak a Na-Dene language. This is consistent with the controversial[22] three migration wave model of Greenberg which views Inuit and Na-Dene languages as markers for distinct migrations from Asia[9], although not with the purest form of that model which would specify that the Inuit and Chipewyan represent sister groups to some Siberian populations, whereas in fact they cluster with Native Americans (Figure 1C), consistent with subequent admixture within the Americas. Intriguingly, Greenberg's hypothesis that Na-Dene marks a distinct migration with Asia has been supported by recent linguistic work that shows that Na-Dene language have a link with Siberian Yeniseian languages[23]. The group of Siberian populations with which the Chipewyan show the strongest genetic affinity includes the Ket, the sole living speakers of Yeniseian (Table S6).

We next sought to determine the timing of the migrations. While it is difficult to estimate dates of population splits using SNP array data subject to ascertainment bias, we obtain a minimum date for Inuit migrations by studying the decay of admixture LD in the Cheyenne, the Inuit-admixed NNA population with the largest sample size allowing the most accurate inference[24]. The extent of LD corresponds to a minimum of 1,500 years ago (95% confidence) (Note S5 and Figure S7), indicating the Inuit had already mixed with the NNA by that time.

To better understand the history of the 47 Native American populations from Meso-America southward who are consistent with a single founding event, we used Admixture Graphs (AG), which are generalizations of phylogenetic trees that allow for the possibility of discrete unidirectional population mixture events[20] (Note S6). We first identified a subset of populations with less evidence of admixture—to serve as a backbone for the AG—by applying the *4 Population Test* to the tree $((\text{Han},NA_i),(NA_j,NA_k))$ using Han as one outgroup and evaluating all possible triples of Native American (NA) populations consistent with Figure 1C. Only 15 of the 47 populations are poor fits in a substantial fraction of *4 Population Tests* (underlined). Of these,

10 correspond to a cluster of largely Chibchan speakers from the Isthmo-Colombian area. From the 32 populations with no evidence of admixture, we selected a subset that were geographically dispersed, included at least 4 samples, and remained a fit to the data when assessed using our more stringent AG fitting procedure (Note S6). We then added in populations modeling the possibility of a single admixture event involving other populations from the graph. The resulting AG of 18 populations provides an excellent fit to the data, in the sense that only 2 of the 11,781 statistics measuring patterns of allele frequency correlation predicted by the model are >3 standard errors from expectation (Note S6).

Three features of the AG are striking. First, the data suggest that some populations in Meso America have not experienced strong bottlenecks since arrival in the region. For example, the genetic drift between the Zapotec and the ancestors of all South Americans is estimated to be 0.004. Second, we fit a higher proportion of South American than Meso American populations using the AG approach. Specifically, we had difficulty fitting a Meso American population from a linguistic/geographic group into the AG once we had included another representative from that same group, but in South American populations, we were often able to fit multiple populations from any group. We hypothesize that this reflects "Isolation-by-Distance", in which populations bidirectionally and continuously exchange genes with neighbors, which is not modeled by AGs which specify unidirectional and discrete admixture events. The less extensive evidence for gene flow that we observe in the New World, and especially in South America, contrasts with analyses of the Old World where migration is prevalent[25]. Thus, cultural diffusion may have played a greater role in the spread of agriculture over long distances on the American continent than in the Old World where the long distance spread of farmers played a major role[26,27.]

The third striking finding is detection of population mixture events, demonstrating the power of genome-wide analyses of masked data to discover previously unappreciated events in Native American history. For example, the Inga can be modeled as having both Amazonian and Andean ancestry, consistent with speaking a Quechuan language but living in the eastern Andean slopes of Colombia with known exchanges with neighboring Amazonian lowlands. The Guarani and the Guahibo can be modeled as stemming from the admixture of differentiated strands of ancestry in eastern South America (Figure 3). The most finding is in diverse Chibchan-speaking populations from the Isthmo-Colombian area, who can only be fit into the AG if they are modeled as harboring a strand of ancestry from eastern South America and a strand of ancestry

more ancient than the separation of the Mexican Pima. Populations carrying this signal are present both to the north (Cabecar, Guaymi, Teribe, Zenu, Maleku and Bribri) and to the south (Kogi and Arhuaco) of the Panama isthmus, suggesting that the admixture occurred prior to the diversification of Chibchans and their spread across the isthmus (Note S6). For the Cabecar, the Chibchan-speaking group with the largest sample size, we used admixture LD to obtain a minimum 95% confidence date is >5,000 years ago (Figure 4) (consistent estimates were obtained for other Chibchan-speakers) (Figure S7; Table S7; Note S5). This is an entirely novel set of observations suggesting a major gene flow event across the Panama isthmus after the initial colonization of South America and before the advent of agriculture. It is also consistent with geography, emphasizing as it does the role of the Isthmo-Colombian region as a point of contact between the northern and southern hemispheres. As the origin of Chibchan culture is already the subject of long-standing controversies[28,29], existing linguistic and archaeological data may benefit from reanalysis in the light of this finding.

This study is the most comprehensive survey of genetic diversity in Native Americans to date, and also the first that directly accounts for the potential confounder of non-Native admixture. The approach taken here to account for recent admixture will also be applicable to whole genome sequences, which will provide data that is free of "ascertainment bias", thus for example allowing inference of divergence times and population size changes. Although here we focused on ethnically well-defined Native American populations, we believe that our approach is potentially applicable to other highly admixed populations that exist across the Americas[30]. Such work could increase the resolution of evolutionary analyses of the Americas, filling sampling gaps and allowing the study of regions where as a consequence of admixture no ethnically defined Native populations exist.

Methods

DNA Samples: The samples analyzed here were collected for previous studies over several decades using a range of informed consent and oversight procedures that were institutionally approved at the time each study was carried out. Ethical approval for the use of these samples in population genetic analyses was obtained prior to this study at Université de Montreal, University of California Berkeley, Universidad de Antioquia, Universidad Nacional Autonoma de Mexico, Centro de Investigaciones Biomédicas de Guatemala, Universidad de Costa Rica, Universidad Peruana Cayetano Heredia, Universidad de Chile, : Instituto Multidisciplinario de Biologia Celular and Universidad de Buenos Aires Argentina, Universidade Federal do Rio Grande do Sul, Universidade Federal do Paraná, Comitê Nacional de Ética em Pesquisa-Brazil, Universidad de Santiago de Compostela, CNRS - Université Paul Sabatier Toulouse 3 and Yale University. Special review panels convened at the request of the NIH re-reviewed some of the oldest collections genotyped for this study[31] and approved the use of the samples for population genetic studies. Ethical approval for the joint analyses of these data was provided by the NHS National Research Ethics Service, Central London REC 4 (Ref # 05/Q0505/31) after reviewing the proposed study as well as the informed consent and ethical review documents provided by the institutions contributing the samples. This study was also approved by the Harvard Medical School Institutional Review Board (protocol M11681-104). All DNA samples have been anonymized.

Genotyping: Genotyping was performed using Illumina arrays and standard protocols as detailed in Note S1. A subset of samples for which only small amounts of DNA were available were whole genome amplified using the Qiagen REPLI-g midi kit prior to genotyping.

Data curation: We required >95% completeness of genotyping for each SNP and >90% for each sample. We merged the data with five other datasets. We further removed samples that were outliers in PCA relative to others from their group, showed an excess rate of heterozygotes compared to the expected rate from the frequency in the population, or had evidence of being a second degree relative or closer to another sample in the study (Note S1).

Removal of genomic segments that might contain non-Native American ancestry: For each Native American individual in turn, we use HAPMIX [32] to model their haplotypes using two ancestral panels: (i) "Old World" populations, a pool of 392 Europeans and 134 West Africans, and (ii) "New World" populations, a pool of 628 Native Americans that were in our data set prior to our most aggressive filtering. Haplotype phase in the ancestral panel, which is necessary for HAPMIX, was determined by phasing both pools of samples together using fastPHASE [33]. We removed segments that had an expected number of more than 0.01 non-Native American chromosomes according to HAPMIX (SOM). For the PCA analysis of samples with non-Native American ancestry segments masked, we restricted to populations with at least 4 samples, and then filled in missing data based on the average genotype in the population.

Population structure analysis, F_{ST} and Neighbor Joining tree: We used EIGENSOFT to carry out PCA and compute F_{ST} [34]. Clustering was performed using ADMIXTURE [35]. A Neighbor Joining [36] tree based on F_{ST} was computed using POWERMARKER [37].

Admixture Graphs: We used the Admixture Graph framework [20] to fit models of population separation followed by mixture to the data. An Admixture Graph makes quantitative predictions about the correlations in allele frequency differentiation statistics (*f*-statistics) that will be observed among all pairs, triples, and quadruples of populations [20], and these can be compared to the observed values (along with a standard error from a Block Jackknife) to test hypotheses about the topology of population relationships (Note S6).

Estimating dates of admixture events: We used ROLLOFF [24] to estimate dates of population mixture. For each population in which we attempted to date admixture, we identified two other populations (or pools of populations) that we used as surrogates for the ancestral populations, guided by Figure 1C or Figure 3 (the surrogates that we used are listed in Table S7). We then binned SNP pairs by their genetic distance separation, and studied the correlation between the LD statistic and the expectation based on the frequency differences across populations if the LD was due to admixture. Dates were inferred based on the spatial scale of the decay of this correlation, which we fitted to an exponential function under the assumption of a single admixture event. A standard error on the date estimate was obtained by performing a weighted

jackknife over chromosomes. We determined 95% confidence intervals as the estimate ±1.96 standard error, and multiplied by 29 to convert from generations to years [38].

Estimating dates of founder events: To estimate the dates of population founder events, we used correlation of allele sharing as a measure of LD. We subtracted the LD within samples from a population to that between a population and a close relative (based on Figure 1C and Figure 3), thus identifying population-specific LD, and fitted the decay with an exponential (Note S4).

Correlating geography with population diversity: Euclidean distances from the Bering Strait (64.8N 177.8E) and the location of each population (Table S1) were calculated using great arc distances based on a Lambert azimuthal equal area projection of the American continent. Least-cost distances between the same points were computed using PATHMATRIX [17], and a spatial cost map incorporating the coastal outline of the Americas. We compared the following coastal/inland relative costs: 1:2, 1:5, 1:10, 1:20, 1:30, 1:40, 1:50, 1:100, 1:200, 1:300, 1:400, and 1:500. Pearson's correlation coefficient was estimated between mean heterozygosity for each population and their least cost distances from the Bering Strait (Note S3).

References

1 Goebel, T., Waters, M. R. & O'Rourke, D. H. The late Pleistocene dispersal of modern humans in the Americas. *Science* **319**, 1497-1502 (2008).

2 Dillehay, T. D. Probing deeper into first American studies. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 971-978 (2009).

3 O'Rourke, D. H. & Raff, J. A. The human genetic history of the Americas: the final frontier. *Curr. Biol.* **20**, R202-207 (2010).

4 Kitchen, A., Miyamoto, M. M. & Mulligan, C. J. A three-stage colonization model for the peopling of the Americas. *PLoS ONE* **3**, e1596 (2008).

5 Mulligan, C. J., Kitchen, A. & Miyamoto, M. M. Updated three-stage model for the peopling of the Americas. *PLoS ONE* **3**, e3199 (2008).

6 Tamm, E. *et al.* Beringian standstill and spread of Native American founders. *PLos ONE*, 1-6 (2007).

7 Fagundes, N. J. *et al.* Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. *Am. J. Hum. Genet.* **82**, 583-592 (2008).

8 Wall, J. D. *et al.* Genetic variation in Native Americans, inferred from Latino SNP and resequencing data. *Mol. Biol. Evol.* (in press).

9 Greenberg, J. H., Turner, C. G. & Zegura, S. L. The Settlement of the Americas - A Comparison of the Linguistic, Dental, and Genetic-Evidence. *Curr. Anthrop.* **27**, 477-497 (1986).

10 Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. *The History and Geography of Human Genes*. (Princeton U.P., 1994).

11 Karafet, T. M. *et al.* Ancestral Asian source(s) of new world Y-chromosome founder haplotypes. *Am.J.Hum.Genet.* **64**, 817-831 (1999).

12 Lell, J. T. *et al.* The dual origin and Siberian affinities of Native American Y chromosomes. *Am J.Hum.Genet.* **70**, 192-206 (2002).

13 Bortolini, M. C. *et al.* Y-chromosome evidence for differing ancient demographic histories in the Americas. *Am. J. Hum. Genet.* **73**, 524-539 (2003).

14 Ray, N. *et al.* A statistical evaluation of models for the initial settlement of the american continent emphasizes the importance of gene flow with Asia. *Mol. Biol. Evol.* **27**, 337-345 (2010).

15 Ruhlen, M. *A Guide to the World's Languages*. (Stanford University Press, 1991).

16 Schroeder, K. B. *et al.* Haplotypic background of a private allele at high frequency in the Americas. *Mol. Biol. Evol.* **26**, 995-1016 (2009).

17 Ray, N. PATHMATRIX: a geographical information system tool to compute effective distances among samples. *Mol. Ecol. Notes* **5**, 177-180 (2005).

18 Wang, S. *et al.* Genetic variation and population structure in native Americans. *PLoS Genet.* **3**, e185 (2007).

19 Yang, N. N. *et al.* Contrasting patterns of nuclear and mtDNA diversity in Native American populations. *Ann Hum Genet* **74**, 525-538 (2010).

20 Reich, D., Thangaraj, K., Patterson, N., Price, A. L. & Singh, L. Reconstructing Indian population history. *Nature* **461**, 489-494 (2009).

21 Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H. & Bustamante, C. D. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet* **5**, e1000695 (2009).

22 Campbell, L. *American Indian languages: the historical linguistics of Native America*. (Oxford University Press, 1997).

23 Kari, J. & Potter, B. *Anthropological papers of the Universityof Alaska Vol. 5*. (University of Alaska, 2010).

24 Moorjani, P. *et al.* The history of African gene flow into southern Europeans, Levantines and Jews. *PLoS Genet.* **7**, e1001373 (2011).
25 Novembre, J. *et al.* Genes mirror geography within Europe. *Nature* **456**, 98-101 (2008).
26 Diamond, J. & Bellwood, P. Farmers and their languages: the first expansions. *Science* **300**, 597-603 (2003).
27 Bramanti, B. *et al.* Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science* **326**, 137-140 (2009).
28 Barrantes, R. *et al.* Microevolution in lower Central America: genetic characterization of the Chibcha-speaking groups of Costa Rica and Panama, and a consensus taxonomy based on genetic and linguistic affinity. *Am.J.Hum.Genet.* **46**, 63-84 (1990).
29 Barrantes, R., Smouse, P. E., Neel, J. V., Mohrenweiser, H. W. & Gershowitz, H. Migration and genetic infrastructure of the Central American Guaymi and their Affinities with other tribal groups. *Am. J. Phys. Anthropol.* **58**, 201-214 (1982).
30 Bryc, K. *et al.* Colloquium paper: genome-wide patterns of population structure and admixture among Hispanic/Latino populations. *Proc. Natl. Acad. Sci. U. S. A.* **107 Suppl 2**, 8954-8961 (2010).
31 Cann, H. M. *et al.* A human genome diversity cell line panel. *Science* **296**, 261-262 (2002).
32 Price, A. L. *et al.* Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. *PLoS Genet.* **5**, e1000519 (2009).
33 Scheet, P. & Stephens, M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am J Hum Genet* **78**, 629-644 (2006).
34 Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, 2074-2093 (2006).
35 Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655-1664 (2009).
36 Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol.Biol.Evol.* **4**, 406-425 (1987).
37 Liu, K. & Muse, S. V. PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics.* **21**, 2128-2129 (2005).
38 Fenner, J. N. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *Am. J. Phys. Anthropol.* **128**, 415-423 (2005).

Acknowledgments. We are grateful to the volunteers who donated samples, to M. Metspalu, R. Villems and E. Willerslev for facilitating sharing of data from Siberian and Arctic populations, to C. Stevens for assistance with genotyping, and to P. Bellwood, K. Bryc, J. Diamond, T. Dillehay, R. Gonzalez-José, P. Moorjani, M. Ruhlen, and A. Williams for comments on the manuscript. Support was provided by NIH grants NS043538 (A.R.-L.), NS037484 (N.B.F.), GM079558 (A.D.), GM079558-S1 (A.D.) and GM057672 (K.K.K. & J.R.K.), an NSF HOMINID grant 1032255 (D.R.), a Canadian Institutes of Health Research grant (D.L.), a Universidad de Antioquia CODI grant (G.B.), a FIS grant PS09/02368 (A.C.), a Wenner-Gren Foundation Grant ICRG-65 (A.D. and R.S.), Russian Foundation for Basic Research Grants 06-04-048182 (R.S.) and 02-06-80524a (L.O.),. a Siberian Branch Russian Academy of Sciences Field Grant (L.O.), a CNRS grant (J.-M.D.), and discretionary funds from Harvard Medical School (D.R.) and the Harvard School of Public Health (A.L.P.).

Author contributions. D.R., N.B.F., A.L.P. and A.R-L. conceived the project; D.R., N.P., D.C., A.T., S.M., N.R. and A.R-L. performed analyses; D.R. and A. R.-L. wrote the paper with input from all the co-authors; and all other authors contributed to collection of samples and data.

Data access. The dataset is available on request from the corresponding authors.

Figure Legends

Figure 1: Geographic distribution and simple genetic analyses. (A) Sampling locations of 55 Native American populations based on the coordinates in Table S1, with colors corresponding to the linguistic categories of Ruhlen[15]. The numbered ellipses refer to the South American population groupings discussed in the text. (B) Masking of segments of non-Native American ancestry is allows examination of the relationship among Native American populations prior to European contact. We used HAPMIX [32] to filter out segments where the estimate of the number of non-Native American alleles was >0.01. Cluster-based analysis (k=4) using ADMIXTURE[35] shows evidence of Indo-European- and some Yoruba-related ancestry in most Native Americans prior to masking (top), but little afterward (bottom), and also hints at Siberian-related ancestry in some North Amerind-speaking groups. (C) Neighbor-Joining tree relating Native American to selected non-American populations (sample sizes in parentheses). All Native American and Siberian data were analyzed after masking of potentially non-Native American segments (except for the Aleutian Islanders), and branch lengths are proportional to F_{ST} (Table S3). The underlining indicates Native American populations that are a grossly poor fit to the tree, and red letters and numbers denote population splits or clusters discussed in the text.

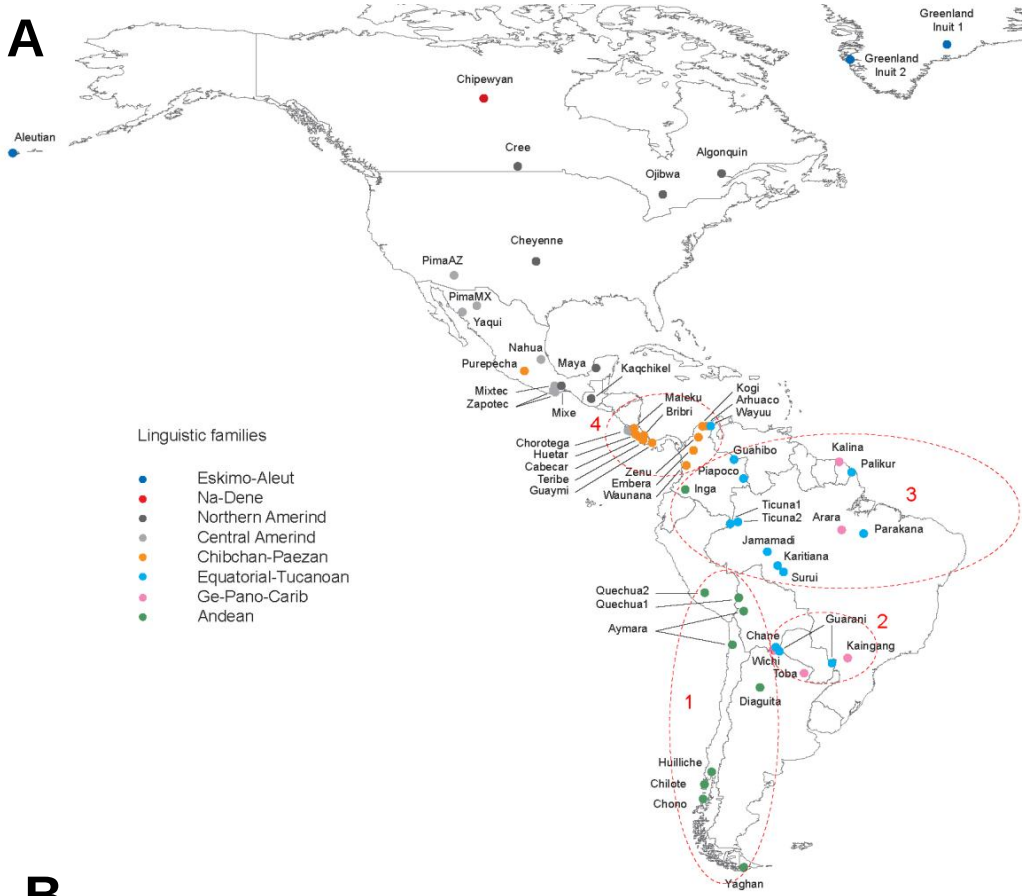
Figure 2: Migrations associated with the peopling of the Americas. Application of the *4 Population Test* reveals three complexities associated with the ancestry of Native Americans. (A) We first tested the hypothesis that Native Americans and East Asians are sister groups, but Europeans are significantly more closely related to Native Americans than to East Asians, invalidating many prevailing models of demographic history. (B) We found that 5 Native North American (NNA) populations do not form a clade with more southern Native Americans relative to diverse Asian and Arctic populations, as revealed by significantly non-zero *4 Population Test* f_4 statistics. The quantitative values of these statistics are highly correlated across the Cheyenne, Ojibwa, Cree and Agonquin, with the largest f_4 statistics seen when testing proximity to Inuit, suggesting that the pattern is due to gene flow from Inuit into the ancestors of these groups. (C) Principal Component Analysis shows that the 5 NNA are outliers relative to the 47 more southern Native American populations, with the Chipewyan being distinct from the other 4 NNA. *4 Population Test* analysis confirms a distinct relationship of the Na-Dene Chipewyan to Asians

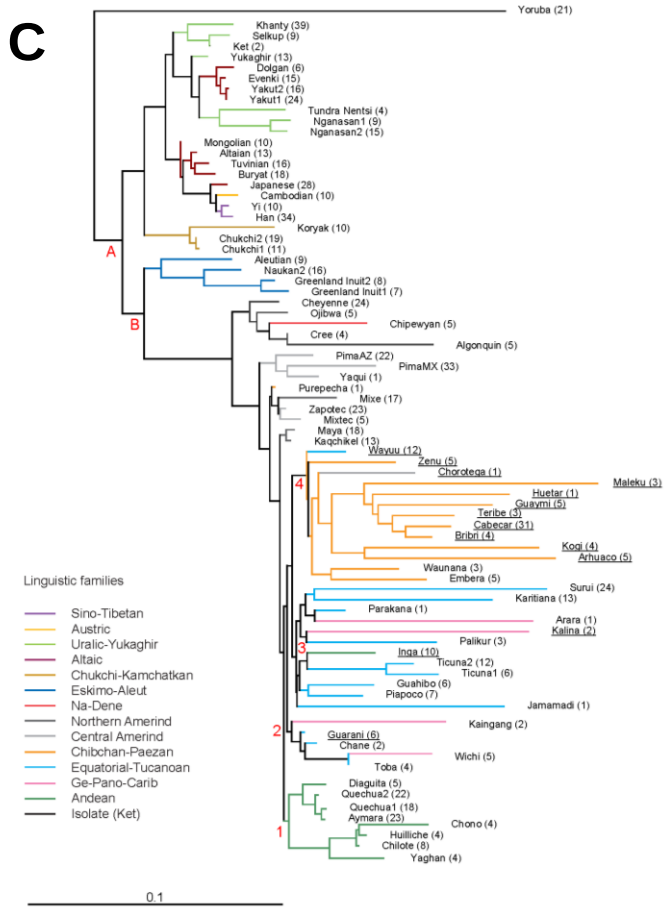
(uncorrelated test statistics). The Asian populations to which the Chipewyan show particular proximity are the Chukchi, Inuit, Nganasan and Ket (Table S6).

Figure 3: Admixture Graph analysis detects 4 novel population mixture events. This AG with 18 populations is the largest ever built and provides an excellent fit to the data as only 2 of the 11,781 *f*-statistics testing allele frequency correlations predicted by the model deviate >3 standard errors from expectation. Genetic drift estimated on each lineage is given in units proportional to $1000 \times F_{ST}$, and mixture events (dotted lines) are denoted by the inferred percentage of ancestry. The Arhuaco and Kogi (circled in green) are well modeled as a mixture between a strand of ancestry from eastern South America and a deep strand of Native American ancestry that is more ancient than the separation of the Mexican Pima (similar findings are obtained for other Chibchan-speakers; Note S6). The Inga (yellow) are modeled as a mixture of Andean and Amazonian ancestry; and the Guarani (blue) and the Guahibo (red) as mixtures of separate strands of ancestry from eastern South America. (Empty ellipses indicate ancestral populations that are inferred by the Admixture Graph model.) The colored lines indicate uncertainty: we show alternative insertion points for lineages involved in the four admixture events which are equally good fits.

Figure 4: Ancient admixture in the Cabecar >5,000 years ago. We binned SNPs based on their genetic distance separation, and computed the correlation of the observed LD to the sign that would be expected from mixture of a North American lineage (represented by a mixture of Pima, Maya, Cheyenne and Zapotec), and a lineage related to other populations in the primarily Chibchan-speaking clade of Figure 1C. We detect admixture between ancient North and South American lineages, with an extent of LD corresponding to 241 ± 41 generations (1 standard deviation), or 5,000-8,900 years ago assuming 29 years. (Black dots show the data; red line shows the fitted exponential decay.) No decay of admixture LD is detected when we do not use a mix of North and South American populations as surrogates for the ancestral populations.

Figure 1


A
Greenland Inuit 1
Greenland Inuit 2
Chipewyan
Aleutian
Cree
Algonquin
Ojibwa
Cheyenne
PimaAZ
PimaMX
Yaqui
Nahua
Purepecha
Maya
Kaqchikel
Mixtec
Zapotec
Mixe
Maleku
Bribri
Kogi
Arhuaco
Wayuu
Chorotega
Huetar
Cabecar
Teribe
Guaymi
Zenu
Embera
Waunana
Guahibo
Piapoco
Inga
Kalina
Palikur
Ticuna1
Ticuna2
Arara
Parakana
Jamamadi
Karitiana
Surui
Quechua2
Quechua1
Aymara
Guarani
Chane
Wichi
Toba
Kaingang
Diaguita
Huilliche
Chilote
Chono
Yaghan
1
2
3
4
Linguistic families
Eskimo-Aleut
Na-Dene
Northern Amerind
Central Amerind
Chibchan-Paezan
Equatorial-Tucanoan
Ge-Pano-Carib
Andean


C
Yoruba (21)
Khanty (39)
Selkup (9)
Ket (2)
Yukaghir (13)
Dolgan (6)
Evenki (15)
Yakut2 (16)
Yakut1 (24)
Tundra Nentsi (4)
Nganasan1 (9)
Nganasan2 (15)
Mongolian (10)
Altaian (13)
Tuvinian (16)
Buryat (18)
Japanese (28)
Cambodian (10)
Yi (10)
Han (34)
Koryak (10)
Chukchi2 (19)
Chukchi1 (11)
Aleutian (9)
Naukan2 (16)
Greenland Inuit2 (8)
Greenland Inuit1 (7)
Cheyenne (24)
Ojibwa (5)
Chipewyan (5)
Cree (4)
Algonquin (5)
PimaAZ (22)
PimaMX (33)
Yaqui (1)
Purepecha (1)
Mixe (17)
Zapotec (23)
Mixtec (5)
Maya (18)
Kaqchikel (13)
Wayuu (12)
Zenu (5)
Chorotega (1)
Maleku (3)
Huetar (1)
Guaymi (5)
Teribe (3)
Cabecar (31)
Bribri (4)
Kogi (4)
Arhuaco (5)
Waunana (3)
Embera (5)
Surui (24)
Karitiana (13)
Parakana (1)
Arara (1)
Kalina (2)
Palikur (3)
Inga (10)
Ticuna2 (12)
Ticuna1 (6)
Guahibo (6)
Piapoco (7)
Jamamadi (1)
Kaingang (2)
Guarani (6)
Chane (2)
Wichi (5)
Toba (4)
Diaguita (5)
Quechua2 (22)
Quechua1 (18)
Aymara (23)
Chono (4)
Huilliche (4)
Chilote (8)
Yaghan (4)
A
B
1
2
3
4
Linguistic families
Sino-Tibetan
Austric
Uralic-Yukaghir
Altaic
Chukchi-Kamchatkan
Eskimo-Aleut
Na-Dene
Northern Amerind
Central Amerind
Chibchan-Paezan
Equatorial-Tucanoan
Ge-Pano-Carib
Andean
Isolate (Ket)
0.1

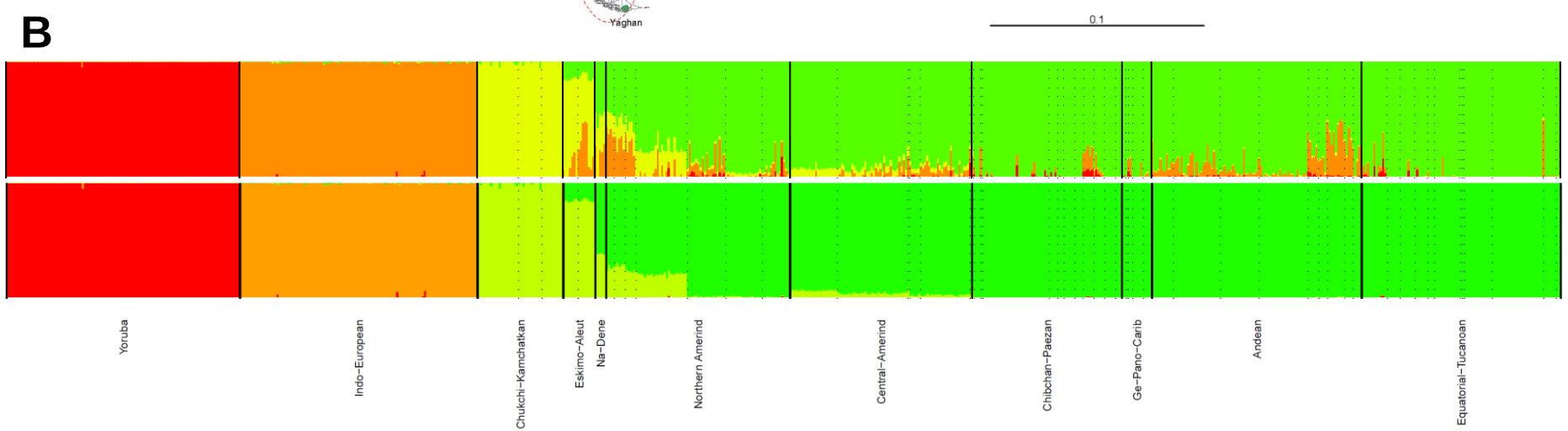

B
Yoruba
Indo-European
Chukchi-Kamchatkan
Eskimo-Aleut
Na-Dene
Northern Amerind
Central Amerind
Chibchan-Paezan
Ge-Pano-Carib
Andean
Equatorial-Tucanoan

Figure 2

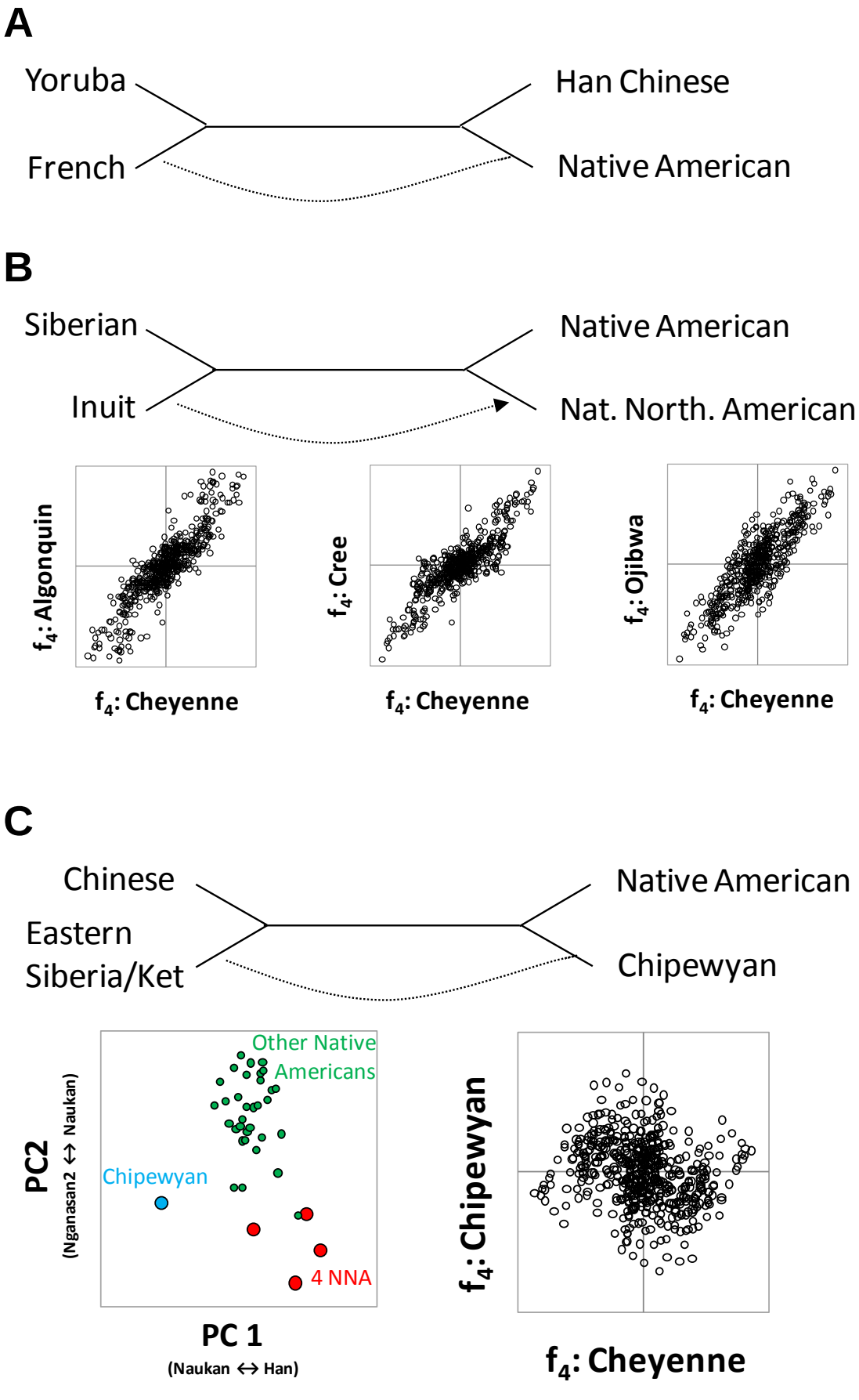

A
Yoruba
French
Han Chinese
Native American
B
Siberian
Inuit
Native American
Nat. North. American
f4: Algonquin
f4: Cheyenne
f4: Cree
f4: Cheyenne
f4: Ojibwa
f4: Cheyenne
C
Chinese
Eastern
Siberia/Ket
Native American
Chipewyan
Other Native
Americans
Chipewyan
4 NNA
PC2
(Nganasan2 ↔ Naukan)
PC 1
(Naukan ↔ Han)
f4: Chipewyan
f4: Cheyenne

Figure 3

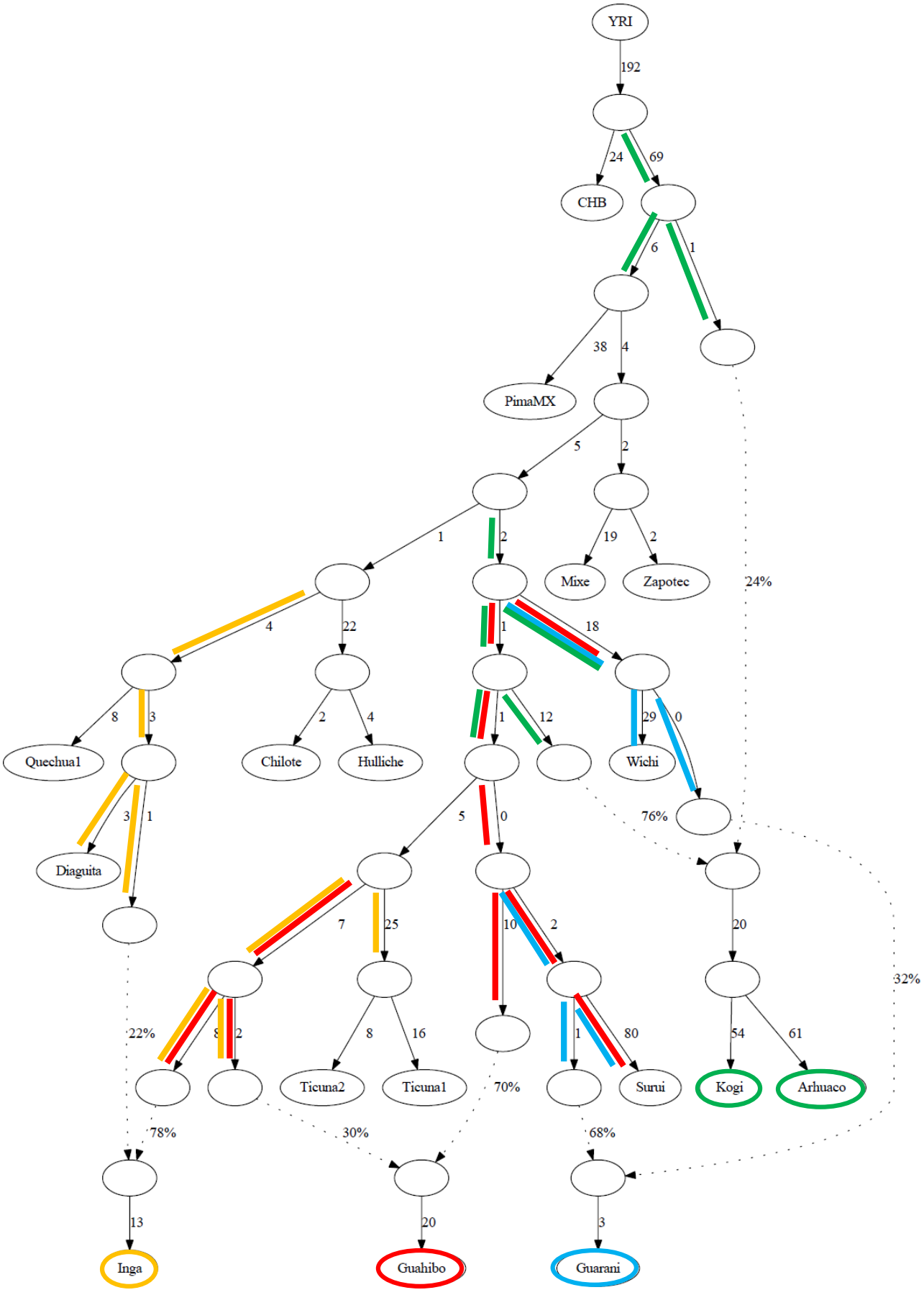
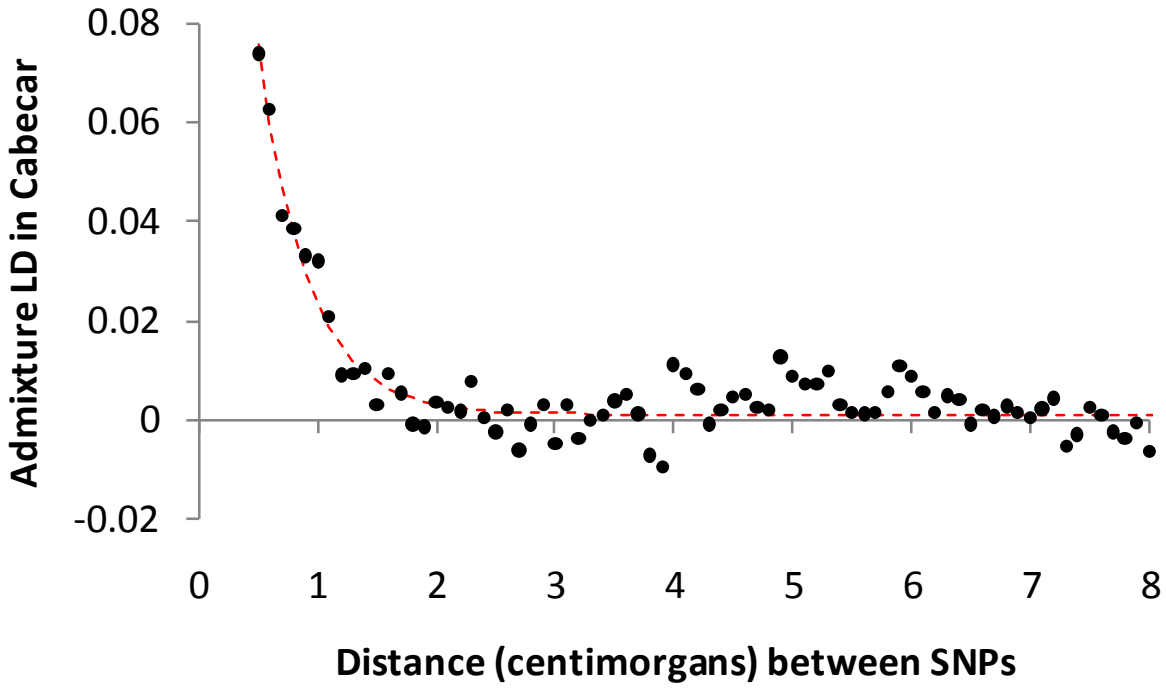

YRI
192
24
69
CHB
6
1
38
4
PimaMX
5
2
1
2
19
2
Mixe
Zapotec
24%
4
22
1
18
8
3
2
4
1
12
29
0
Quechua1
Chilote
Hulliche
Wichi
3
1
5
0
76%
Diaguita
7
25
10
2
20
32%
22%
8
2
8
16
1
80
54
61
Ticuna2
Ticuna1
70%
Surui
Kogi
Arhuaco
78%
30%
68%
13
20
3
Inga
Guahibo
Guarani

Figure 4


0.08
0.06
0.04
0.02
0
-0.02
Admixture LD in Cabecar
0
1
2
3
4
5
6
7
8
Distance (centimorgans) between SNPs

Supplementary Materials
Reconstructing Native American Population History

Table of Contents 1

Note S1 – Preparation of the data set 2-6
Note S2 – Masking segments of potential European or African ancestry 7-8
Note S3 – Correlation of genetic diversity with geographic distance from the Bering Strait 9-10
Note S4 – Dates of founder effects 11-13
Note S5 – Dates of admixture events 14-16
Note S6 – Inference of population relationships incorporating admixture 17-21

Figure S1 – Sampling locations of 19 Siberian and 5 East Asian populations 22
Figure S2 – PCA demonstrates the effectiveness of masking of non-Native American ancestry 23
Figure S3 – Examples of masking of segments of non-Native American ancestry 24
Figure S4 – The evidence of ancient admixture in Chibchans is not an artifact of masking 25
Figure S5 – Heterozygosity and geographic distance from the Bering Strait 2
Figure S6 – Native North Americans have a distinct relationship to Eurasians 27
Figure S7 – Dates of admixture events from the decay of admixture linkage disequilibrium 28-29

Table S1 – Summary information for 55 Native American populations 30
Table S2 – Summary information for 19 Siberian populations 31
Table S3 – F_{ST} for populations used to build the Neighbor Joining tree (masked data) 32
Table S4 – Estimates of bottleneck dates based on decay of allele sharing 33
Table S5 – Z-scores from *4 Population Tests* of the tree ((Out1,Out2), (NatAm1, NatAm2)) 34
Table S6 – f_4 statistics from *4 Population Tests* of the tree ((Zapotec, NNA), (Out1,Out2)) 35
Table S7 – Record of admixture dating analyses 36

Note S1

Preparation of the data set

(i) A merged dataset derived from six sources

We merged six datasets from samples genotyped on various Illumina SNP arrays (Table S1.1).

Table S1.1: Genotyping data sets that we merged for this study

Name of dataset	N	Comments
"Ruiz-Linares" (Native American and Siberian)	373	We attempted to genotype 509 samples from 49 populations on an Illumina 610-Quad array, and initially filtered out 3 samples that were genotyped twice, 9 samples due to inconsistency with a previous DNA fingerprint n the same sample, and 120 samples based on a call rate of <90%. We removed 59,163 SNPs with a call rate of <95% or no physical position.
"Kidd" (Native American and Siberian)	316	Genotyping was performed on an Illumina 650Y array, and we initially removed 16 samples that overlapped with the CEPH-HGDP samples or were outliers relatives to others from the same population in PCA.
"DiRienzo" (Siberian)	64	These data consisted of genotyping of 4 Siberian populations by Anna DiRienzo's laboratory on either an Illumina 610-Quad array (Nganasan and Yukaghir) or an Illumina 650Y array (Naukan and Chukchi[1]).
"Willerslev" (Arctic)	176	Previously published data[2]. We analyzed 12 Eurasian and 3 Native Arctic populations genotyped on an Illumina 650Y array (all from ref. 2 except for Na-Dene which did not have permissions appropriate for this analysis).
"HapMap3" (Worldwide)	1,184	Previously published data[3]. Genotyping was done on an Illumina 1M array.
"CEPH-HGDP" (Worldwide)	936	Previously published data[4]. Genotyping was done on an Illumina 650Y array. We restricted to individuals inferred to be unrelated up to 2nd degree relatives[5].

(ii) Data curation - Removal of Native American outlier samples

We performed data curation steps to remove outlier samples. This was important for the Native Americans, as there has been substantial mixture in the last five hundred years, both due to migration from Europe and Africa and due to recent gene flow among geographic neighbors.

We first ran HAPMIX (Note S2) to identify segments of the genome in Native Americans (excluding Arctic populations) that are of potentially West Eurasian or African ancestry. We subsequently treated the genotypes in these segments as if they were missing data. This "masking" prevented us from discarding all samples that had evidence of some post-Colombian European or African ancestry (if we had done this we would have lost the great majority of the samples). The estimates of non-Native American or non-Siberian ancestry, and the proportion of the genome that was masked in each population, is presented in Table S1 for Native American and Table S2 for Siberian populations. We then applied the following filters:

(1) 23 samples were removed due to a high missing genotype rate

We required that all samples had genotyping missing data rates of <10%.

(2) 33 samples were removed due to a high proportion of West Eurasian or African mixture
We removed samples with <22% of their genomes inferred to be of entirely Native American ancestry based on the masking analysis of Note S2.

(3) 80 samples were removed due to excess or deficiency of heterozygotes vs. expectation
In the Kidd dataset, all the Karitiana and most of the Ticuna had a significant excess of heterozygous genotypes compared with the allele frequency computed in the same samples (violations of Hardy-Weinberg equilibrium). We removed these populations. We also removed a handful of additional samples due to heterozygote excess or deficiency.

(4) 28 samples were removed due to evidence of being at least a 2nd degree relative to others
It was already known that the Surui sample contained relatives[6]. For all pairs of individuals in all populations that had evidence for >22% of their genome being shared, we removed one of the pair (in general we chose to remove the one with more missing data). For this purpose, we used the *SMARTREL* program, part of the EIGENSOFT package[7].

(5) 36 samples were removed as PCA outliers relative to others from the same population
To prepare the dataset for PCA-based outlier removal, we restricted to Native American populations with at least 3 samples, as outlier removal is impossible with fewer samples. Because many samples had substantial missing data (due to masking segments of potentially non-Native American ancestry), we filled in missing data at each SNP based on the mean allele frequency of other samples from the same population. For the PCA, we did not include SNPs that had entirely missing data for any of the population included in a particular PCA.

We divided the Native American populations into 5 geographic groupings (to make the visual inspection of the PCA plots tractable): North Americans, Meso-Americans, Andeans, North West South Americans and Eastern South Americans. We then performed PCA using EIGENSOFT[6]. We plotted samples on all eigenvectors that were statistically significant, as assessed using a Tracy-Widom distribution6. We iteratively removed samples that were outliers relative to others from the same population until the samples from each population appeared homogeneous. Some populations, such as the Cabecar, showed an over-dispersion in the PCA, likely reflecting recent admixture with neighboring populations affecting a substantial proportion of samples. We did not remove any samples in such populations.

The number of Native American samples in the merged dataset (excluding Siberians and Arctic Native Americans) before data curation was 623 and after was 451 (Table S1.2 reports results by population). Importantly, we performed the data curation entirely by visual and computational analysis of clusters in PCA, searching for individuals that were outliers with respect to their own population. Thus, if our data curation introduces bias, it would be to make populations more homogeneous, not to introduce correlations in ancestry across groups. In other words, we do not expect our curation to bias inferences about the topology of population relationships.

(iv) Data curation - Removal of Siberian and Arctic North American outlier samples
We performed a similar analysis in the 21 Siberian and 3 Arctic North Americans populations, after applying a similar masking procedure as for the non-Arctic Native Americans (Note S2; Table S1.3). This resulted in 19 Siberian and 3 Arctic North American populations, after we removed the Naukan1 and Yukaghir1 populations because so few samples were left from each after the data curation.

(1) 11 samples were removed due to evidence of being at least a 2nd degree relative to others For all pairs of individuals that had evidence for >22% of their genome being shared, we removed one of the pair (in general we chose to remove the one with more missing data). For this purpose, we use *smartrel*, which is part of the EIGENSOFT package6.

(2) 17 samples were removed due to being outliers in PCA relative to others from the same population. Since many samples had substantial missing data (corresponding to masked segments containing potential non-Native American ancestry), we filled in missing data at each SNP based on the mean allele frequency for others in the same population.

(3) 19 samples were removed due to less than 28% of the genome being available after masking. We removed samples from populations with limited data after masking, except for Aleutian Islanders where so much data was removed that we used unmasked data.

Table S1.2: Native American samples before and after data curation

Population	Before	After
CEPH-HGDP genotyping		
Maya	21	18
Piapoco	7	7
Kidd genotyping		
Cheyenne	47	24
PimaAZ	41	22
Quechua2	22	22
Ticuna2	34	12
Guahibo	10	6
CEPH-HGDP + Kidd genotyping		
PimaMX	46	33
Surui	30	24
Karitiana	35	13
Ruiz-Linares genotyping		
Cabecar	34	31
Zapotec	38	23
Aymara	24	23
Quechua1	18	18
Mixe	20	17
Ruiz-Linares genotyping (cont.)		
Kaqchikel	18	13
Wayuu	17	12
Inga	13	10
Chilote	10	8
Guarani	9	6
Ticuna1	6	6
Arhuaco	6	5
Algonquin	5	5
Ojibwa	5	5
Mixtec	5	5
Guaymi	5	5
Zenu	5	5
Diaguita	5	5
Wichi	5	5
Chipewyan	5	5
Embera	6	5
Kogi	6	4
Toba	5	4
Cree	5	4
Chono	4	4
Huilliche	4	4
Ruiz-Linares genotyping (cont.)		
Bribri	4	4
Yaghan	4	4
Waunana	5	3
Teribe	3	3
Palikur	3	3
Maleku	4	3
Chane	2	2
Kaingang	2	2
Kalina	2	2
Parakana	4	1
Arara	2	1
Jamamadi	2	1
Huetar	2	1
Purepecha	1	1
Yaqui	1	1
Chorotega	1	1
Ache	3	0
Pehuenche	1	0
Mekranoti	1	0

(v) Data curation - Removal of outlier samples from other populations

We also performed PCA to remove some outlier samples from non-Native American and non-Siberian populations. This analysis removed the entire MKK population3 (Masai from Kenya from HapMap3) because of many statistically significant eigenvectors that were difficult to interpret. We also removed 71 other samples that were outliers relative to their own populations.

(vi) Cases in which we had a pair of sample sets with the same population label

Four populations were genotyped in two different centers (Kidd and CEPH-HGDP) but were known to be from the same original sample collection: Yakut, Karitiana, Surui and PimaMX.

The Karitiana from the Kidd genotyping were dropped because of evidence for heterozygote excess (see above). PCA showed systematic differences in the two Yakut datasets, potentially reflecting a chance subdivision of the Yakut sample collection (which involved several urban collections of a small number of individuals). Therefore, both datasets were kept separate, and denoted Yakut1 and Yakut2. PCA indicate that the two Surui and PimaMX datasets were indistinguishable based on PCA, and so we merged them6. The labels we used were:

"PimaMX" (to designate Kidd PimaMX and the CEPH-HGDP Pima)
"Surui" (to designate Kidd Surui and CEPH-HGDP Surui)

We did not find evidence for relatives in these merged samples, as expected because *smartrel* had already been used to remove duplicate samples and close relatives across the entire data set.

There were six other examples of populations where there were two different sample collections, and we did not merge these either because PCA showed systematic differences or because we wished to separate the samples for historical reasons (e.g. the HapMap3 YRI and HGDP Yoruba were kept separate). Any observed genetic differences among these samples could reflect genuine substructure within these populations. The six pairs of populations in this category were:

Ticuna ("Ticuna1" and "Ticuna2")
Quechua ("Quechua1" and "Quechua2")
Pima ("PimaMX" and "PimaAZ")
Yoruba ("Yoruba" and "YRI")
Mongolian ("Mongolian" and "Mongola")
Nganasan ("Nganasan1" and "Nganasan2")

Table S1.3: Siberian and Arctic North American samples before and after data curation

Population	Before	After
CEPH-HGDP genotyping		
Yakut1	25	24
Kidd genotyping		
Khanty	47	39
Yakut2	20	16
Ruiz-Linares genotyping		
Naukan1 *	2	0
Tundra_Nentsi	4	4
DiRienzo genotyping		
Chukchi2	19	19
Nganasan2	15	15
Naukan2 *	16	16
Yukaghir2 †	14	13

Population	Before	After
Willerlev genotyping		
Aleutian	9	9
Altaian	13	13
Buryat	19	18
Chukchi1	14	11
Dolgan	7	6
GreenlandInuit1	10	8
GreenlandInuit2	10	7
Evenki	16	15
Ket	2	2
Koryak	17	10
Selkup	10	9
Nganasan1	15	9
Tuvinians	16	16
Yukaghir1 †	9	0

* The reduction in the number of Naukan1 samples due to data curation was so severe that only one was left, and we removed this sample from the dataset entirely and henceforward refer to "Naukan2" as "Naukan".
† The reduction in the number of Yukaghir1 samples due to data curation was so severe that only two were left, and we removed these two samples from the dataset and refer to "Yukaghir2" as "Yukaghir".

(vii) Removal of SNPs with inconsistent or potentially problematic genotyping

After merging data for all populations, we curated SNPs as follows:

(1) 16 SNPs were removed due to an excess or deficiency of heterozygous genotypes
6 SNPs in the Ruiz-Linares data, 6 in the Kidd data, 3 in the Willerslev data, and 1 in the CEPH-HGDP data, showed an excess or deficiency of heterozygotes compared with expectation given the frequency in their own populations (their chi-square statistics were visual outliers from the tail).

(2) 15 SNPs were removed due to inconsistency in frequency across data sets
For all SNPs, we compared the frequency across populations of similar ancestry. We found 9 SNPs from the Ruiz-Linares data set, and 6 SNPs from HapMap3, which were consistently much more differentiated from the other data sets than would be expected from the tail of the chi-square distribution, suggesting genotyping problems. These SNPs were removed.

(viii) Creation of merged datasets for analysis

We created two merged datasets. The first, "merge5," consists of all data except the Siberian populations from the Di Rienzo dataset for which there were substantially fewer SNPs typed. The second, "merge6," consists of all data (Table S1.4). Both the "merge5" and "merge6" datasets have two versions: ".unmasked" and ".masked". The "unmasked" version is the dataset after the data curation steps above. The "masked" dataset was obtained after running HAPMIX to define segments of potential African or West Eurasian ancestry due to admixture in the last few hundred years (Note S2). SNPs in such segments were then treated as missing.

Table S1.4: Merged datasets generated for this study

Name	Samples	Autosomal SNPS	Nat. Am. populations	Siberian populations	Other populations
merge5	2,289	470,949	55	14	58
merge6	2,356	378,659	55	19	58

Note: Each dataset has ".unmasked" and ".masked" versions. X chromosome data is included only for "merge5.unmasked".

References for Note S1

[1] Hancock AM, Witonsky DB, Alkorta-Aranburu G, Beall CM, Gebremedhin A, Sukernik R, Utermann G, Pritchard JK, Coop G, Di Rienzo A (2011) Adaptations to climate-mediated selective pressures in humans. *PLoS Genet.* **7,** e1001375.

[2] Rasmussen M. *et al.* Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463,** 757-762 (2010).

[3] International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* **467,** 52-58 (2010).

[4] Li J.Z. *et al.* Worldwide human relationships inferred from genome-wide patterns of variation. *Science* **319,** 1100-1104 (2008).

[5] Rosenberg NA. Standardized subsets of the HGDP-CEPH Human Genome Diversity Cell Line Panel, accounting for atypical and duplicated samples and pairs of close relatives. *Ann Hum Genet.* **70,** 841-847 (2006).

[6] Calafell F, Shuster A, Speed WC, Kidd JR, Black FL, Kidd KK (1999) Genealogy reconstruction from short tandem repeat genotypes in an Amazonian population. *American Journal of Physical Anthropology* **108**,137-146

[7] Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genet.* **2,** e190.

Note S2

Masking segments of potential European or African ancestry

Most Native American samples have inherited segments of their genomes from European and African ancestors who were immigrants to the New World since 1492. Since this study focuses on the pre-Columbian history of the Americas, these segments are confounders for our analyses.

To restrict analyses to segments of the genome that are likely to be of entirely Native American ancestry, we used methods that can infer the probability of different ancestral origins for each segment of the genome. We masked segments that are inferred to have a substantial probability of being of non-Native American ancestry (that is, we restricted analyses to segments of the genome that are inferred to be homozygous for Native American ancestry)[1]. The success of such a method relies on three ingredients: (i) admixture has occurred recently enough that there are multi-megabase genomic segments where it is possible to confidently infer ancestry; (ii) we have dense enough genotyping data to perform local ancestry inference over these segments, and (iii) appropriate methods are available for carrying out local ancestry analysis.

To perform local ancestry inference, we employed HAPMIX[2], which uses a haplotype Hidden Markov Model to model each segment of the genome as a mixture of two ancestral panels of haplotypes provided by the user. Our "non-Native American" ancestral panel consists of 526 samples representing both the European and African ancestral populations (24 Basque, 46 Bedouin, 112 CEU, 28 French, 12 Italian, 46 Palestinian, 28 Sardinian, 88 TSI, 8 Tuscan, 113 YRI and 21 Yoruba), and our "Native American" ancestral panel consists of 628 Native American samples. This is larger than the 451 samples that we had left after data curation (Note S1), because the masking procedure was performed prior to our most severe round of data curation including removal of outliers and removal of poorly performing samples. HAPMIX requires that the samples from the ancestral panels are phased2, and to achieve this we pooled all the samples in the parental panels and ran the fastPHASE software[3].

We ran HAPMIX on each of the Native American samples in turn, using the remaining Native American samples (all but the one being analyzed) as one parental panel and the 526 European and African samples as the other. For each sample, we used software settings corresponding to a prior hypothesis of an admixture proportion of 5%, and a number of generations since mixture of 10 (these prior hypotheses have minimal effect on ancestry inference for admixture in the last handful of generations2, which is the scenario that applies to Native Americans). The inferred proportion of non-Native American ancestry averaged over all loci is very similar to that generated by the ADMIXTURE clustering software when run with k=3 clusters[4] (corresponding to European, African and Native American). The main exceptions are Native North American populations where ADMIXTURE produces higher estimates of non-Native American ancestry, likely reflecting complex gene flows with Siberian populations as discussed in the main text.

At each locus, HAPMIX infers the probability that an individual has 0 (p_0), 1 (p_1) and 2 (p_2) alleles of non-Native American ancestry. Thus, the expected number of non-Native American alleles at any locus is $E = p_1 + 2p_2$. Running HAPMIX on the Native American samples, it infers that 21% of loci have a posterior estimate of $E > 0.01$ non-Native American alleles (averaging

across the genome and samples) (we note that this differs from the 14% of loci reported in the main text, because it was computed prior to removing samples with an extremely high proportion of non-Native American ancestry). We also explored using a less stringent threshold for the posterior estimate of the number of non-Native American alleles, but found that this only marginally increased the amount of loci (for example, increasing to $E \geq 0.1$ increases the amount of data we could analyze by only about one percent). Because we wished to be as confident as possible that we are analyzing Native American segments for studying history—and because we only lose a small amount of data by discarding segments with even a small probability of non-Native American ancestry—we chose the more stringent threshold. We also inspected the local ancestry inference for diverse samples, and found that in many cases, there were substantial stretches where HAPMIX confidently inferred no non-Native American ancestry (Figure S3).

It is likely that there are some biases in the segments of Native American genomes that we are successfully masking (or failing to mask). For example, it is likely that we are more often masking out segments at the telomeres where there is less confident ancestry inference. In addition, it is likely that there are segments of the genome where the haplotype structure is such that there is variable success in inferring local ancestry. In practice, what is important is whether such biases confound inferences of population relationships among Native Americans. The *4 Population Test* results reported in the main text, as well as the PCA and ADMIXTURE analyses reported in Figure 1B and Figure S2, suggest that after our local ancestry inference procedure, we have removed the great majority of non-Native American ancestry segments, to the point that we can perform meaningful population genetic analyses of the masked data.

References for Note S2

[1] Bryc K, Velez C, Karafet T, Moreno-Estrada A, Reynolds A, Auton A, Hammer M, Bustamante CD, Ostrer H (2010) Colloquium paper: genome-wide patterns of population structure and admixture among Hispanic/Latino populations. *Proc Natl Acad Sci USA* **107 Suppl 2,** 8954-8961.

[2] Price AL, Tandon A, Patterson N, Barnes KC, Rafaels N, Ruczinski I, Beaty TH, Mathias R, Reich D, Myers S (2009) Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. *PLoS Genet.* **5,** e1000519.

[3] Scheet P, Stephens M (2006) A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am J Hum Genet.* **78,** 629-644

[4] Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19,** 1655-1664

Note S3

Correlation of genetic diversity with geographic distance from the Bering Strait

For exploring the correlation of genetic diversity to distance, we used the "merge6.masked" dataset. We computed the observed heterozygosity for each individual and averaged across all individuals for each population. To reduce sampling variation, only populations with five or more individuals were included. Distances from the Bering Strait were computed using great arc routes from an Anadyr start point at 64.8N 177.8E, with the location of each population specified by the coordinates in Table S1. We computed a Pearson correlation coefficient between the mean observed population heterozygosity and the distance from Beringia. We evaluated statistical significance by using a t-distribution transformation (using the R-package[1]).

Table S3.1: Heterozygosity and distance from the Bering Strait

Population	N	Distance (meters)	Heterozygosity
Chipewyan	5	2,998,535	0.246
PimaAZ	22	4,904,611	0.251
Cheyenne	24	5,170,029	0.257
Ojibwa	5	5,184,797	0.260
PimaMX	33	5,432,128	0.240
Algonquin	5	5,619,796	0.239
Mixtec	5	7,105,459	0.248
Maya	18	7,138,397	0.253
Mixe	17	7,140,781	0.244
Zapotec	23	7,181,122	0.251
Kaqchikel	13	7,538,473	0.252
Cabecar	31	8,397,297	0.224
Guaymi	5	8,588,582	0.217
Arhuaco	5	8,746,097	0.211
Wayuu	12	8,788,814	0.242
Zenu	5	8,878,868	0.243
Embera	5	9,025,514	0.223
Guahibo	6	9,481,686	0.232
Inga	10	9,576,373	0.234
Piapoco	7	9,833,731	0.238
Ticuna1	6	10,391,952	0.228
Ticuna2	12	10,412,538	0.230
Quechua2	22	11,214,787	0.246
Karitiana	13	11,346,772	0.223
Quechua1	18	11,484,968	0.246
Surui	24	11,493,384	0.208
Aymara	23	11,941,135	0.246
Wichi	5	12,486,648	0.223
Guarani	6	12,739,695	0.249
Diaguita	5	12,960,201	0.245
Chilote	8	13,914,216	0.239

The distance and heterozygosity values that we used are shown in Table S3.1 and suggest a negative correlation between heterozygosity and distance from the Bering Strait (Figure S5, r = -0.37, P=0.04). Averaging heterozygosity for populations from major regions summarizes the trend: North Amerind: 0.253, Meso America: 0.249, North West South America/Lower Central America: 0.223, Andean: 0.241, Chaco: 0.242, East South America: 0.22.

A noticeable exception is the populations from North West South America/Lower Central America, which have a heterozygosity that is lower than expected based on geography. The low heterozygosity is consistent, however, with the tree of Figure 1C, which indicates that these populations are most closely related to populations from eastern South America, and thus may represent one of the last major population splits in the region. This agrees with a settlement model for South America involving an early migration southward along the Pacific coast, followed by a migration northward on the eastern side of the Andes and culminating in northern south America and the settlement of the Caribbean islands. Excluding the North West South America/Lower Central American populations from the analyses results in an increase of the heterozygosity-distance correlation to -0.481 (P=0.01). This correlation increases further when considering the coasts as facilitators of migration.

To include the effects of coasts, we also computed "effective", or "least-cost path" distances[2]. Compared to the standard geographic great arc distances, effective distances incorporate the effects of one or several landscape components. They are computed as least-cost paths on the basis of a spatial cost map that incorporates these landscape components. The effective distance is computed as the sum of costs ("cost distance") along the paths. Because the relative cost of landscape component is somewhat arbitrary, we tested a range of combinations. For example, a ratio of 1:10 coastline/land means that it is ten times more costly to go through land than through coastline. In addition to simple great arc distances, we used the following coastline/inland cost combinations: 1:2, 1:5, 1:10, 1:20, 1:30, 1:40, 1:50, 1:100, 1:200, 1:300, 1:400, 1:500.

The correlation peaks at -0.61 for a coastline/inland ration of 1:10 (Figure S5A,B). Excluding the 5 NNA populations with evidence of more recent gene flows from Asia/the Arctic (notes) the negative correlation persists (-0.40, P=0.076) and this correlation increases further when effective distances are considered (Figure S5C,D). These observations confirm that the trends observed in the full dataset are not solely the result of the higher diversity of the 5 NNA, which could be influenced by the more recent gene flows that has affected these populations.

References for Note S3

[1] R Development Core Team. R: A language and environment for statistical computing. (Vienna, Austria, 2010).
[2] Ray N (2005) PATHMATRIX: a geographical information system tool to compute effective distances among samples. *Molecular Ecology Notes* **5,** 177-180.

Note S4

Dates of founder events

(i) The POPSHARE method for estimating the dates of founder events

To estimate the dates of founder events in Native Americans, we updated the program POPSHARE[1]. The updated program eliminates a sample size dependence of the original test statistic that we have discovered since the original publication.

Within-population correlation of allele sharing

Suppose that we have n samples from a population ($n \geq 4$). At each SNP k, consider two individuals i and j ($i \neq j$), and write $g_k(i)$ and $g_k(j)$ as the number of variant alleles (0, 1 or 2) in that sample. We can define a function $S_k(i,j)$ equal to the number of alleles that two samples share. For example, $g_k(i)=0$, $g_k(j)=2 \Rightarrow S_k(i,j)=0$, $g_k(i)=1$, $g_k(j)=2 \Rightarrow S_k(i,j)=1$ and $g_k(i)=2$, $g_k(j)=2 \Rightarrow S_k(i,j)=2$. The only complicated case is $g_k(i)=1$, $g_k(j)=1$, and for this case we set $S_k(i,j)=1$, the expected number of shared alleles after phasing.

Given a sample of n individuals, we can compare all possible pairs of samples, and thus we have a vector S consisting of $n(n-1)/2$ values of $S_k(i,j)$ that captures the allele sharing pattern at the SNP. To compute the correlation of allele sharing as a function of distance, we compute the Pearson correlation coefficient of S for all possible pairs of SNPs and bin by genetic distance.

Across-population correlation of allele sharing

Consider two populations with n and m samples each. We define the $S_k(i,j)$ statistic as for the within-population case, with the modification that i and j are required to be from different populations, and thus the vector S has $n \times m$ entries. We can then similarly compute the Pearson correlation coefficient of S between all possible pairs of SNPs and bin by genetic distance.

Our statistic works provided that we have at least 4 samples. For the within population case we compute our correlation as above for 4 samples (within-population) and two pairs of samples (across population). We perform this computation in all possible ways and bin by genetic distance. This eliminates any sample size effect.

Estimating the dates of population-specific founder events

We aim to estimate the dates of population-specific founder events using the allele sharing due to descent from a limited number of ancestors since separation from other relatively closely related populations. A naïve way to estimate the date of a bottleneck would be to compute the extent of LD. However, LD reflects not just the most recent bottleneck in a population's history, but also other genetic drift events that occurred more anciently, including history shared with other populations (e.g., the bottlenecks that associated with the peopling of the Americas). Simply measuring the LD in a population and fitting its decay to an exponential distribution would result in a date that is an average of many LD-generating events including older ones not specific to the population, and would thus result in an overestimation of the date.

Our allele sharing statistics allow us to circumvent this problem, since we can compare the correlation in allele sharing within a population N to the correlation in allele sharing between

population *N* and its relative *M*. By subtracting these two curves, we hope to study the LD that has been generated since the separation of the two populations from each other.

To convert the subtraction of curves into time, we note that the average extent of LD should reflect the average time to the common ancestor of two alleles in population *N* that coalesce more recently than the separation from *M*. If all coalescence in population *N* is due to single founder event in the history of *N* since its split from *M*, the population-*N*-specific LD should decay exponentially and an exponential distribution fitted by least-squares should produce a decay constant that can be converted into a date of the founder event. Specifically, after the founder event, the correlation breaks down if a recombination event occurs on either side of a pair of shared haplotypes. Thus, for a pair of SNPs at genetic distance *d*, the expected correlation of allele sharing will be e^{-2nd} where *n* is the number of generations since the founder event.

Alternatively, if the population-specific LD is due to multiple bottlenecks or LD-generating events, the decay is expected to be non-exponential (a summation of exponentials), which may be possible to detect visually. If a single exponential distribution is fitted to a curve that is in fact a sum of exponentials, the date that will be obtained will be an average of the time depths of the LD-generating events that occurred in population *N* since its separation from population *M*.

Figure S4.1 shows an example of this procedure for the Wichi. The red curve shows the correlation in allele sharing within the Wichi without subtracting the LD shared with its neighboring populations. The curve shows both a fast rolloff and a long tail, and we hypothesize that the fast rolloff reflects LD-generating events in the common history of the Wichi and other Native American populations. The blue curve shows the correlation in allele sharing of the Wichi to the Guarani, who are closely related according to Figure 1C. As expected, there is a faster rolloff of LD for the across-population comparison since these populations are not expected to share recent LD-generating events. The green curve shows the subtraction of the blue from the red curves – the LD specific to the Wichi since their split from the Guarani – and this is relatively well fit by an exponential decay allowing for an affine (constant) term to account for residual familial relatedness. The rate constant corresponds to 22 generations (or 638 years, assuming a generation time of 29 years), suggesting a founder event in the Wichi around the time of the arrival of Old World populations in the Americas.

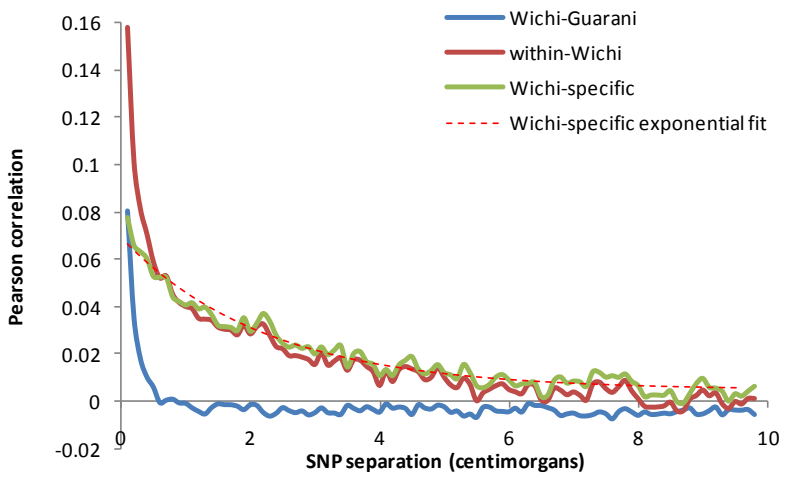


Figure S4.1: Estimating the date of founder events with POPSHARE. We compute the correlation in allele sharing in the Wichi (red) and subtract it from that between the Wichi and Guarani (blue) to obtain a Wichi-specific correlation (green). The decay of allele sharing specific to the Wichi is well fit by an exponential distribution, whose rate constant is what would be expected from a founder event 22 generations ago, specific to the history of the Wichi.

A potential pitfall of the strategy for subtracting background LD by a cross-population comparison is that it assumes that the tree in Figure 1C is correct; that is, the two compared populations are sister groups. However, if one or both of the populations are admixed, then the across population comparison will only eliminate some of the background LD. Then, we might expect to observe non-exponential decays, and indeed, we have observed patterns like this. For example, we believe that the negative asymptote for the Guarani-Wichi comparison may reflect such a phenomenon (although in this case, the negative value can be taken into account by including an affine term in the exponential fit). We only report dates for populations in which the decays look like a visually reasonable fit to an exponential decay with an affine term.

(ii) Analysis of within-population founder events in 40 Native American populations
We applied POPSHARE to the 40 Native American populations in Table S1 with at least 4 samples. We used Figure 1C to select the outgroup population to which we compared each population (Table S4). When populations had more than 10 samples, POPSHARE ran extremely slowly as it had to perform computations based on all possible 4-way subsets of samples (for populations with more than 4 samples we sub-sampled in order to compute autocorrelations). To speed up the runs, we either reduced the number of samples we used (choosing samples randomly) or only considered a fraction $1/n$ of pairs of SNPs, again using a random number generator to determine which SNPs to study.

We observe approximately exponential decays of population-specific LD in 23 populations (Table S4). The estimated dates are for the most part between 12-27 generations, although there is substantial error around the individual estimates reflecting the relatively small sample sizes. It is interesting that most of the dates are consistent with the approximately 18 generations that have elapsed since the arrival of Europeans in the Americans (520 years ago assuming 29 years per generation). Thus, these data suggest a history of recent population-specific demographic collapses approximately coinciding with the encounter with Europeans.

Our results also provide some evidence that these founder events were more extreme in some regions than in others. For Andeans, there is little evidence of recent founder events. Similarly, a number of Meso-American populations do not have evidence of founder events (e.g. Zapotec, Maya and Kaqchikel). Both the Andes and Meso-America had the largest Native American populations at the arrival of the Europeans and although they also suffered a major demographic collapse, it is likely that the absolute population sizes in many groups never became extraordinarily small. Some populations in Meso-America do show evidence of founder events (e.g. Pima and Mixe) but these have more recent estimated dates of 12-13 generations, perhaps reflecting more recent demographic events specific to these groups. By contrast, there is very strong evidence of bottlenecks consistent with the time shortly after European contact for most Chibchan-Paezan, Equatorial-Tucanoan and Ge-Pano-Carib populations (Table S4). These groups never reached the pre-Colombian population density of Meso-America and the Andes and it is likely that the demographic collapse associated with the arrival of the Europeans often resulted in very small absolute population numbers

References for Note S4

[1] Reich, D., Thangaraj, K., Patterson, N., Price, A.L. & Singh, L. Reconstructing Indian population history. *Nature* **461,** 489-494 (2009).

Note S5

Dates of admixture events

(i) ROLLOFF approach for estimating admixture dates

Having detected evidence for population mixture in a number of Native American populations, it is of interest to estimate the date of the admixture. To do this we used the ROLLOFF software, first reported in a study of admixture between groups of African and West Eurasian ancestry[1].

ROLLOFF analyzes pairs of SNPs on the autosomes, binned by their genetic distance separation which we estimate here using the Oxford LD-based genetic map[2]. ROLLOFF then studies how the signed linkage disequilibrium (LD) statistic *D* that is observed between SNPs, compares with the expected value under the assumption that the LD is due to admixture of two specified surrogates for the ancestral populations. If admixture occurred, there is expected to be a non-zero correlation. Under a "single pulse" model in which all the admixture occurred instantaneously, the decay of LD is expected to follow an exponential distribution, and the decay parameter can be translated into an estimate of the date of admixture. If the mixture was spread over many generations, the number obtained is expected to fall within the range of dates of admixture**1**.

ROLLOFF also computes a standard error on the estimated date, based on a Weighted Block Jackknife that removes each chromosome in turn, and studies the variation in the date estimate to obtain an approximately normally distributed standard error. ROLLOFF date estimates gain precision as sample size increases**1**, and thus the limited number of samples we have for a number of populations (Table S1) limits our ability to make precise estimates of some dates.

(ii) Positive control: Post-Colombian admixture in the Americas

To show that we can use ROLLOFF to estimate the dates of well-documented admixture events in Native Americans, we applied it to data from the Maya from the Yucatan in Mexico, using the merge5.unmasked dataset. Because non-Native American segments are unmasked, we can use the dataset to estimate the date of admixture between European and Native American ancestors in the Maya. ADMIXTURE analysis indicates that the Maya have ~13% Old World ancestry (Table S1), and thus if the method is working properly it should obtain a date this is within the post-Colombian range of admixture dates.

We ran ROLLOFF using French and Mixe (a Native American population from South Mexico with little evidence of non-Native American admixture; Table S1) as surrogates for the ancestral populations. Figure S5.1A shows a clear decay of LD with genetic distance. The inferred admixture date is 7.4 ± 0.7 generations, which given an average generation interval in humans of around 29 years[3] translates to about 215 years ago, a figure that is consistent with the period when Europeans and Native Americans were in contact in Mexico. When we repeat the analysis using French and Han as surrogates for the ancestral populations (Figure S5.1B), we observe a weaker correlation reflecting the fact that the Han are a poorer surrogate for the true ancestral population, but the scale of the decay of admixture LD is consistent leading to a date of 7.7 ± 1.1 generations, reflecting the fact that ROLLOFF date estimates are not very sensitive to use of the correct ancestral populations1. Finally, we ran ROLLOFF using two populations that are not thought to be related to the admixing populations as surrogates for the ancestral populations

(Yoruba and Mbuti Pygmy, both from Africa). We found no evidence of a decay of admixture LD (Figure S5.1C). This illustrates how ROLLOFF does not produce LD decay when the allele frequency differences between the populations that were actually involved in the admixture are not related to the allele frequency differences between the surrogates used in the analyses.

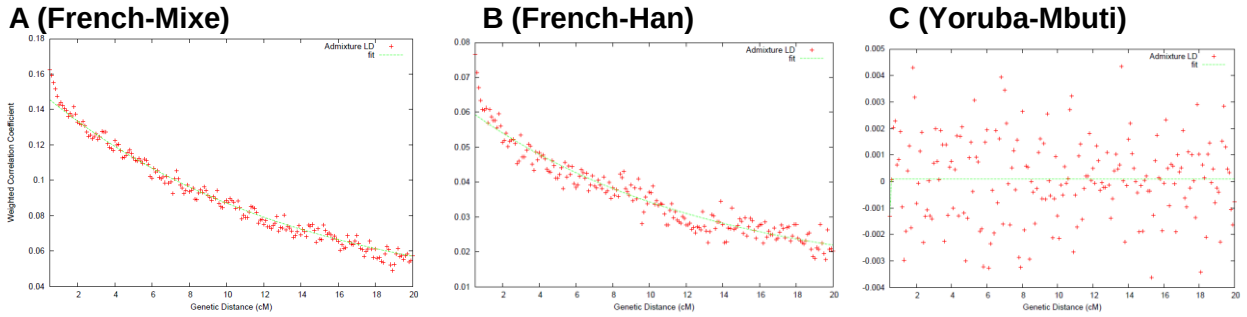


Figure S5.1: Analysis of post-Colombian mixture of European and Native American ancestry in the Maya demonstrates the usefulness of ROLLOFF for estimating admixture dates. We show the decay of admixture LD with distance in the Maya (red dots) and the best fitting exponential decay (green), compared with the expectation from admixture of (A) French and Mixe (a neighboring Native American population with little evidence of admixture; Table S1), (B) French and Chinese Han, and (C) Yoruba and Mbuti (the y-axis scales differ in the three panels). We observe a decay of admixture LD with distance for the first two scenarios, consistent with the Maya inheriting ancestry from a Native American population (ancestrally related to the Chinese) and a European population. Using Yoruba and Mbuti Africans as surrogates for the ancestral populations produces no decay, reflecting the fact that the known history of admixture in the Maya has nothing to do with the history separating these two African groups (or that the small amount of West African ancestry in the Maya is not enough to produce an observed decay in this plot). The estimated average date from ROLLOFF is 7.4 ± 0.7 when we use French and Mixe as surrogates for the ancestral populations, and 7.7 ± 1.1 when we use French and Han. The consistency of these dates reflects a useful property of ROLLOFF, in that we do not require accurate ancestral populations in order to obtain a reliable date.

(iii) Application of ROLLOFF to estimate dates of admixture in Native Americans

We applied ROLLOFF to estimate dates of admixture, focusing on populations for which we had evidence for historical admixture according to previous analyses (see Note S5 and main paper). Since we are focusing on admixture events unrelated to African and Europeans, we used the merge5.masked dataset for all cases except for the Chipewyan and Cheyenne, where we analyzed the merge6.masked dataset to include more data from Siberian populations. Because the power of ROLLOFF is improved by having more accurate estimates of allele frequency differences for the ancestral populations, we pooled data from a number of populations, based on the topology of the Neighbor Joining tree of Figure 1C and the Admixture Graph of Figure 3.

We observe clear decays of admixture LD in the Cheyenne, Inga, Guarani, Kogi and Cabecar (Figure S7). We hypothesize that for populations in which we detect a signal of an admixture by the *4 Population Test* but not by ROLLOFF analysis, this is due to limited sample size or poor surrogates for the ancestral populations. In each of the populations for which we detected admixture LD decay, we also performed a negative control in which we substituted the ancestral populations shown in Table S7 (chosen to be as related to the true ancestral populations based on the Neighbor Joining tree and Admixture Graph analyses) with French and Han. The allele frequency differences between French and Han should be unrelated to the allele frequency differences between Native American populations under the assumption that the Native American populations we are analyzing descend from a common ancestral population, and thus no decay of admixture LD should be observed in this analysis. As expected, there is no evident decay of admixture LD, indicating that the signal of admixture LD decay that we observe is due to mixture of populations related to the samples we are using to represent the ancestors.

The Cheyenne admixture date is confidently pre-Colombian: 182 ± 80 generations, corresponding to a 90% confidence interval of 1,500-9,100 years ago assuming 29 years per generation. This suggests that the gene flows between Native Americans and populations related to Inuit in the history of the Cheyenne are ancient. We do not detect visual evidence of admixture LD in the Chipewyan, which we hypothesize is due to our limited sample size and poor surrogates we have available for the ancestral populations.

The Chibchan admixture date is also of great interest. Examining a pool of the 9 Kogi and Arhuaco samples (motivated by the Admixture Graph of Figure 3 which suggests that they descend from the same historical admixture event), we observe a sharp exponential decay and obtain a date of 158 ± 38 generations, corresponding to a 90% confidence interval of 2,700-6,400 years. In the 31 Cabecar samples, we obtain a date of 241 ± 41 generations, corresponding to a 90% confidence interval of 5,000-8,900 years. In conjunction with the Admixture Graph analysis of Note S5, these results suggest that the deep lineages in at least some Chibchan populations reflect migration events in the ancestry of this population at least 5,000 years ago (with 95% confidence by a one-sided test).

References for Note S5

[1] Moorjani P, Patterson N, Hirschhorn JN, Keinan A, Hao L, Atzmon G, Burns E, Ostrer H, Price AL, Reich D (2011) The history of African gene flow into southern Europeans, Levantines and Jews. *PLoS Genetics* **7,** e1001373.

[2] Myers S, Bottolo L, Freeman C, McVean G, Donnelly P (2005) A fine-scale map of recombination rates and hotspots across the human genome. *Science* **310**, 321-324.

[3] Fenner JN. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *Am J Phys Anthropol.* **128,** 415-423 (2005).

Note S6

Inference of population relationships incorporating admixture

(i) Overview

In Figure 1C we present a Neighbor Joining tree relating all the Native American and selected Asian populations. However, such an analysis presents an oversimplified view of population relationships, as it presupposes that all populations descend from a common ancestor by a series of bifurcation events without subsequent admixture.

To obtain a richer picture of the relationships among the Native American populations, we focused on the 47 most southern populations that did not have evidence of more than one gene flow event with Eurasians. For the majority of the analyses below, we use the "merge5.masked" dataset, which includes 24% more SNPs and hence increases resolution compared with the "merge6.masked" dataset. Since we used "masked" datasets, all analyses are based on segments of the genome that are inferred to be solely of Native American ancestry (Note S2).

(ii) Pruning to 32 populations that approximately pass *4 Population Tests*

We identified a subset of the Native American populations that are roughly consistent with a tree. To do this, we applied the *4 Population Test* to all possible quartets of populations that are a subset of the 47 most southern Native American populations in the tree of Figure 1C, to assess whether they are consistent with being related as specified by that tree, without admixture. Consider a set of three Native American populations $\{N_i, N_j, N_k\}$, which according to Figure 1C are related according to the unrooted tree $((CHB,N_i),(N_j,N_k))$. If Figure 1C is accurate, then $f_4(CHB,N_i;N_j,N_k)$ is expected to be consistent with zero, which we can test by computing a standard error with a Block Jackknife[1]. We computed this *4 Population Test* statistic for all 16,215 = (47×46×45)/(3×2×1) possible triplets of Native American populations, in each case ordering the populations as specified as in Figure 1C. We then manually removed 15 populations that were involved in most of the violations of the null hypothesis (test statistics more than 3 standard errors from expectation). The populations that we removed are underlined in Figure 1C. This left 32 populations with few significant *4 Population Test* statistics: 3.7% at |Z|>2 standard error from zero, 0.18% at |Z|>3 standard errors from zero, and 0.01% at |Z|>4 standard errors from zero (the population with the highest proportion was the Maya with 8 of 282 statistics significant at |Z|>3 (2.8%)). The 15 populations that were removed fall into three categories:

(1) We removed the Kalina and Yaghan because these exhibited correlations to many other populations, in a way that was not obviously related to the structure of the tree of Figure 1C. A possible explanation is genotyping errors or data processing errors in these populations.

(2) We removed the Guarani and Inga, which show violations of the *4 Population Test* and are two populations that do not cluster with their linguistic neighbors (Equatorial-Tucanoan and Andean, respectively) but rather with their geographic neighbors. The inconsistency between the linguistic and geographic clusters could relate to ancient gene flow, a prediction that is also supported by our Admixture Graph analyses below.

(3) We removed 11 populations from a 13 population cluster around the Panama isthmus (Figure 1C). Of these 11 populations, 9 are Chibchan-Paezan speakers, the other two being the non-Chibchan-Paezan-speaking Wayuu and Chorotega. The two populations from this cluster that we

were able to include without introducing a large number of *4 Population Test* statistics at |Z|>3, are the two Paezan-speaking populations Embera and Waunana.

(iii) Strategy for building Admixture Graphs

We used Admixture Graphs to assess the fit of a proposed model of population relationships to the genetic data. Admixture Graphs[2] are representations of population relationships that can accommodate mixture, and which in the absence of population mixture, simplify to a bifurcating tree. The Admixture Graph fitting procedure is more stringent than the *4 Population Test* based pruning, since it computes the values of all possible *f*-statistics relating populations and assesses their fit to the data, rather than only ensuring that the f_4 statistics are consistent with Figure 1C. As a result, we had to remove many more populations to obtain a fit to the data.

To fit an Admixture Graph to data, it is necessary to specify the amount of genetic drift that occurred historically on each lineage, as well as admixture proportions. An Admixture Graph in which these quantities are specified makes quantitative predictions about the values of all possible *f*-statistics measuring the correlation in allele frequencies among two (f_2), three (f_3), and four (f_4) populations2. We can compare these predictions to the observed values (which have a standard error from a Block Jackknife) to assess the fit. A valuable feature of Admixture Graphs is that they are robust to ascertainment bias of SNPs (how the SNPs were chosen for inclusion in the study), making them useful for inferring topology tree topologies even using data from SNP arrays designed for medical genetics studies2.

To use the Admixture Graph framework to assess the fit of a proposed historical model to empirical data, we have written software that begins with a proposed topology, and then finds the combination of branch lengths and admixture proportions that best fit the data. A limitation is that we do not currently have a formal way to deal with the correlation in the *f*-statistics. In particular, while there are many possible *f*-statistics relating a given set of *N* populations—($N(N-1)/2$ f_2 statistics, $3N(N-1)(N-2)/6$ f_3 statistics, and $3N(N-1)(N-2)(N-3)/24$ f_4 statistics—in fact these are highly correlated. For example, all the f_3 and f_4 statistics can be written as linear combinations of the f_2 statistics. To deal with these correlations, we compute a chi-square statistic measuring the difference between all observed and predicted *f*-statistics taking into account the covariance structure (and using a standard error from a Block Jackknife). While this serves as a score that allows us to climb to a best fitting model, for the time being we do not understand its statistical distribution. Hence, while we can compute a nominal P-value, we do not consider it to be a formal goodness-of-fit test. As a secondary assessment of the fit, we can examine outlier *f*-statistics that are more than three standard errors from expectation. In practice in fitting Admixture Graphs, we view any graph that produces a substantial number of *f*-statistics more than |Z|>4 standard errors from expectation as a graph that we wish to avoid. For Admixture Graphs with a sufficient number of populations, |Z|>4 is expected by chance even if the graph is a correct representation of history so this is somewhat conservative. To further assess the graph, we count the number of *f*-statistics that are |Z|>3 standard errors from expectation, and attempt to minimize this quantity as well.

(iv) An Admixture Graph that fits the data for 16 Native American populations

To build up an Admixture Graph that fits the data, we first excluded the underlined populations in the Neighbor Joining tree of Figure 1C. We also restricted to the populations with at least 4

samples, motivated by the fact that the outlier removal procedure is less effective for populations with fewer samples (Note S1). A further benefit of requiring a minimum sample size is that populations with more samples are associated with *f*-statistics with smaller standard errors.

We fit our Admixture Graph using YRI and CHB as outgroups. We first identified a set of 11 Native American populations that fit a simple phylogenetic tree with no evidence of admixture. We then manually added five additional populations into the Admixture Graph by exploring all possible insertion points of a putative admixture event, and testing the fit. This resulted in the addition of the Kogi, Arhuaco, Guahibo, Guarani and Ingano. The resulting Admixture Graph of 16 Native American populations and 2 outgroups (Figure 3) provides a reasonable fit in the sense that there are only 2 *f*-statistics (out of 11,781) that are more than 3 standard errors from zero (the strongest is $|Z|=3.1$). The Admixture Graph fitting also produces estimates of genetic drift on each lineage (in units scaled to be comparable to $1000\times F_{ST}$), as well as admixture proportions. Standard errors in *f*-statistic values are around 0.001. Thus, short branches (e.g. of length $1 = 1000\times0.001$) are not reliably inferred, and the data are consistent with trifurcations at such nodes.

(v) Admixture events in the Inga, Guarani and Guahibo

The Admixture Graph analysis in Figure 3 suggests that the Inga, Guarani, and Guahibo, can be modeled as resulting from relatively simple admixture events.

We first explored the robustness of inference of admixture in the Inga, who in the tree of Figure 1C cluster with their geographic neighbors rather than with their linguistic neighbors, suggesting *a priori* that they may be the result of population mixture events. We began by testing the parsimonious hypothesis that the Inga are a sister group of the Ticuna (as suggested by Figure 1C). This model is strongly rejected with 143 *f*-statistics that are more than $|Z|>3$ standard errors from expectation including one at $|Z|=5.2$. However, a model of mixture between a Ticuna-related population and a Quechua-related population (as shown in the Admixture Graph of Figure 3) provides an excellent fit to the data. In Figure 3, we show all the possible places in the tree where the ancestral populations of the Inga could insert while being consistent with the data (chi-square statistic of <5 between the two fits to the data).

A similar situation applies to the Guarani, who in the tree of Figure 1C also cluster with their geographic neighbors rather than with their linguistic neighbors. This population can be fit into the Admixture Graph as an admixture of their immediate geographic neighbors and an Equatorial-Tucanoan speaking group (whose language group they share). The possible insertion points into the Admixture Graph that are consistent with the data are shown in Figure 3.

Finally, the Guahibo can be fit into the Admixture Graph as an admixture of the two deep branches of the Equatorial-Tucanoan cluster of Figure 1C, exemplified by the Surui on the one hand, and Ticuna on the other. The insertion of the Surui-related branch is not well specified, but the insertion of the Ticuna-related branch is well specified, as shown in Figure 3.

(vi) Deep admixture in the history of Chibchan-speakers

A striking finding is that 10 of the 13 populations in the primarily Chibchan-Paezan speaking clade in Figure 1C cannot be fit by a simple tree. The only populations that fit the tree are two Paezan-speaking groups, the Waunana and Embera. This suggests that the admixture may be due to events specific to Chibchan history. However, we were able to fit the Kogi and Arhuaco as an admixture of an Amazonian-related population and a Native American lineage that branched very anciently in the history of Native American populations (Figure 3).

To better understand the evidence for an ancient admixture event involving the Chibchan cluster, we used the Paezan-speaking Embera as a reference population based on Figure 1B (which suggested that it may not have the same history of admixture as many of the Chibchan-speakers. We then computed the statistic f_4*(CHB, Y; Embera, X)*, where *X* is any of the populations in the predominantly Chibchan-Paezan clade of Figure 1C (other than the Embera), and *Y* is a more distantly related population (we report the f_4 statistic rather than the Z-score because it has a quantitative interpretation in terms of mixture proportions and genetic drift). If Embera and *X* are sister groups with *Y* more distantly related, the allele frequency difference Embera-*X* should be uncorrelated to CHB-Y, and the expected value should be zero. However, this expectation is not fulfilled when Y is almost any Native American population, including Equatorial Tucanoan-speakers (e.g. Y=Ticuna2), Andean-speakers (e.g. Y=Quechua1), Northern Amerind-speakers (e.g. Y=Maya), or northern Mexicans (e.g. Y=PimaMX) (Figure S6.1). There is a strong correlation in their values whatever outgroup we choose to use (r^2=0.15-0.82; Figure S6.1), consistent with ancestry in many Chibchan-Paezan speaking groups that is actually from a deeper branch than all the tested outgroups Y. This analysis led to the hypothesis that Chibchan-speakers harbour ancestry form a population that roots deeply among Native North Americans.

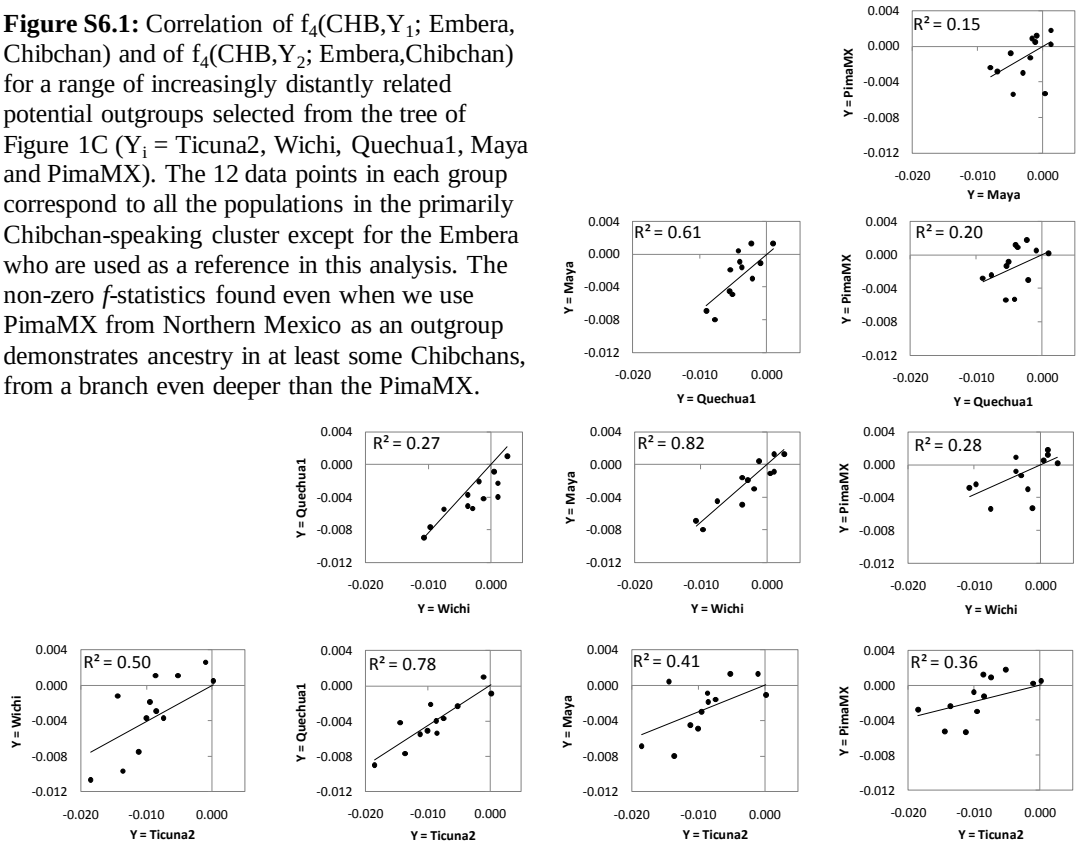


Figure S6.1: Correlation of f_4(CHB,Y_1; Embera, Chibchan) and of f_4(CHB,Y_2; Embera,Chibchan) for a range of increasingly distantly related potential outgroups selected from the tree of Figure 1C (Y_i = Ticuna2, Wichi, Quechua1, Maya and PimaMX). The 12 data points in each group correspond to all the populations in the primarily Chibchan-speaking cluster except for the Embera who are used as a reference in this analysis. The non-zero *f*-statistics found even when we use PimaMX from Northern Mexico as an outgroup demonstrates ancestry in at least some Chibchans, from a branch even deeper than the PimaMX.

To assess the generality of our finding that the Chibchan-Paezan populations are an admixture of a very deep Native American lineage and another lineage related to Amazonians, we modified the Admixture Graph of Figure 3 to remove the Kogi and Arhuaco, and added each population in the majority Chibchan-Paezan speaking clade of Figure 1C in turn. Table S6.1 shows the fit for each of these cases, as assessed by the number of f-statistics more that |Z|>3 standard errors from expectation, and the nominal P-value from a chi-square analysis (which we view with caution since it is not clear to us how many hypotheses we are testing given the correlation of the *f*-

statistics). The fit is good for multiple populations to the north and south of the isthmus of Panama, but the fit is poor for the Waunana, Embera, Huetar, and Chorotega (the fit to the Teribe and Bribri is of intermediate quality). We hypothesize that the poor fit to the Waunana and Embera (who fit reasonably well in the Neighbor Joining tree of Figure 1C) is due to more complex recent gene flows with other populations on the Admixture Graph. We conclude that the deep ancestry is shared in almost all Chibchan-speakers (but not the closely related Paezan-speakers), and hypothesize that the poor fits in some groups reflect additional admixture events.

Table S6.1: Fit of populations in the majority Chibchan-Paezan clade to a history involving admixture with a deep branch of Native Americans (in the position of the Arhuaco and Kogi in Figure 3)

	Sam-ples	No. outliers |Z|>3	Most extreme outlier (|Z|)	Nominal P-value*	Qualitative assessment of fit
Kogi	4	2	3.2	0.13	Good
Arhuaco	5	3	3.1	0.07	Good
Cabecar	31	3	3.2	0.09	Good
Wayuu	12	3	3.2	0.04	Good
Guaymi	5	5	3.1	0.03	Good
Teribe	3	5	3.4	0.01	OK
Zenu	5	6	3.2	0.05	Good
Maleku	3	6	3.4	0.05	Good
Bribri	4	8	3.5	0.03	OK
Chorotega	1	12	3.7	0.0002	Poor
Huetar	1	18	3.8	0.006	Poor
Embera	4	21	4.0	0.002	Poor
Waunana	3	71	4.3	0.00006	Poor

* The nominal P-value is computed based on the fit between all predicted and observed *f*-statistics, taking into account the standard errors and the covariance structure from a Block Jackknife.

We also explored how confident we can be, based on the Admixture Graph methodology, at inferring the insertion points of the two lineages contributing to the ancestry of the Chibchan-speakers. We tested inserting each of the two lineages ancestral to the Kogi and Arhauco at all possible positions in the Admixture Graph of Figure 3, and found that we were able to insert the lineages at all the edges highlighted in dark green in Figure 3 while still providing a fit to the data that had a chi-square of <5 from the best fitting location. The deep lineage was confidently inferred in these analyses to be above the split of the PimaMX from all other Native American groups. The other lineage was equally well fit as clustering with the two Amazonian clades: the majority Equatorial-Tucanoan speaking clade and the majority Ge-Pano-Carib speaking clade.

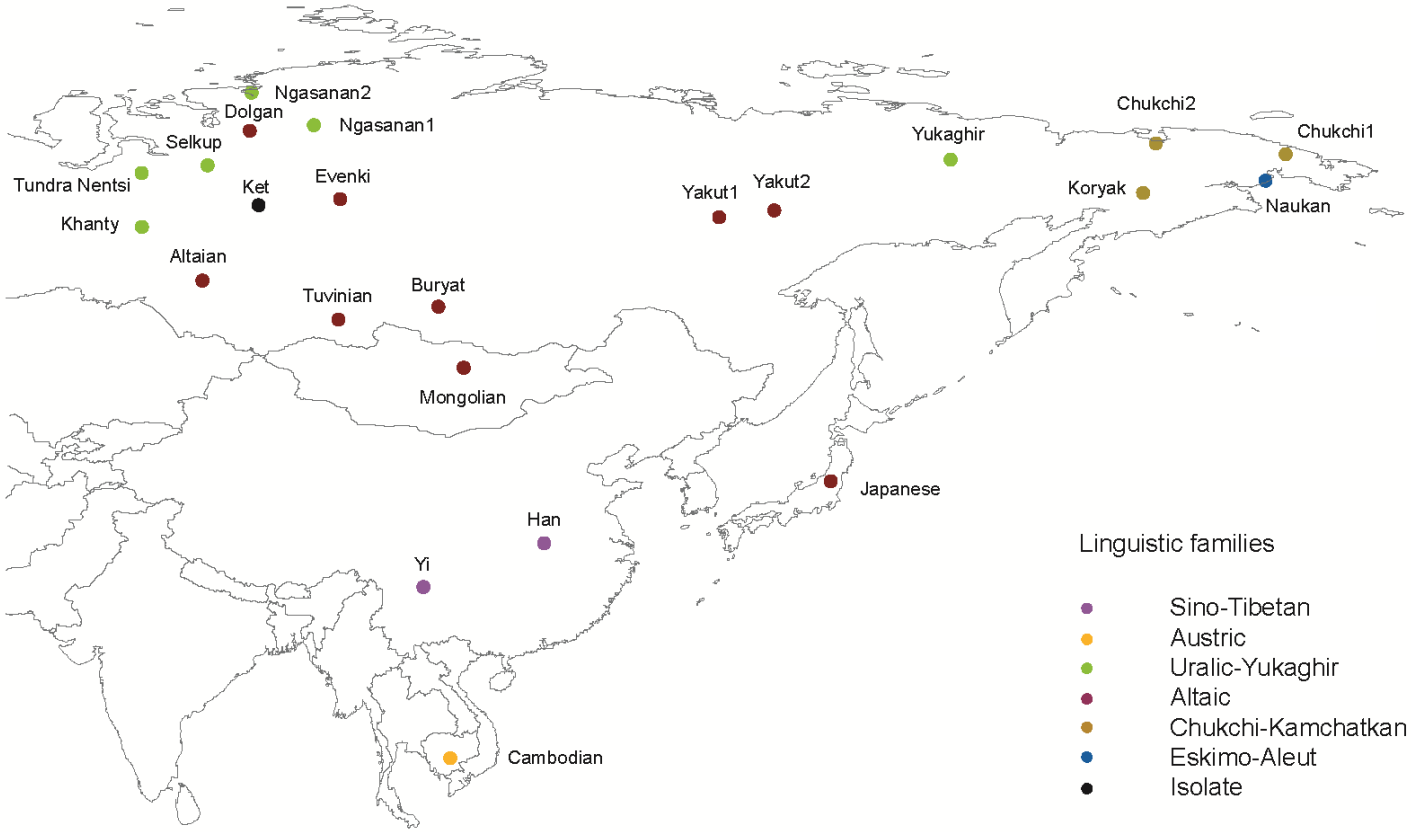
We conclude with an important clarification. While we have shown that Chibchan-speaking populations have likely inherited genetic material from a deep strand of Native American variation, we have no evidence for this ancestry deriving from a separate migration from Eurasia, since the Chibchan speaking groups are among the 47 Native American populations consistent with a single founding population (Table S5). Instead, we hypothesize that this deep ancestry is from a very early branch in the tree of Native American populations after the initial migration south of the North American ice sheets.

References for Note S6

[1] Künsch HR (1989) The jackknife and the bootstrap for general stationary observations. *Ann. Statist.*, **17,** 1217-1241 (1989).

[2] Reich, D., Thangaraj, K., Patterson, N., Price, A.L. & Singh, L. Reconstructing Indian population history. *Nature* **461,** 489-494 (2009).

<u>Figure S1.</u> Sampling locations of 19 Siberian and 5 East Asian populations
Color codes refer to linguistic family affiliation (according to Greenberg). The 5 populations designated as East Asian here are the Mongolian, Japanese, Han, Yi and Cambodian)



<u>Figure S2</u>. PCA demonstrates the effectiveness of masking of non-Native American ancestry.
We computed Principal Components using 3 Old World populations (West African, European American, and Nganasan Siberian), and projected Native American groups onto these PCs. Prior to masking we observe variation in the relatedness of Native Americans to the Old World groups, reflecting varying levels of admixture. However, after masking, we observe tight clustering of all Native American populations.

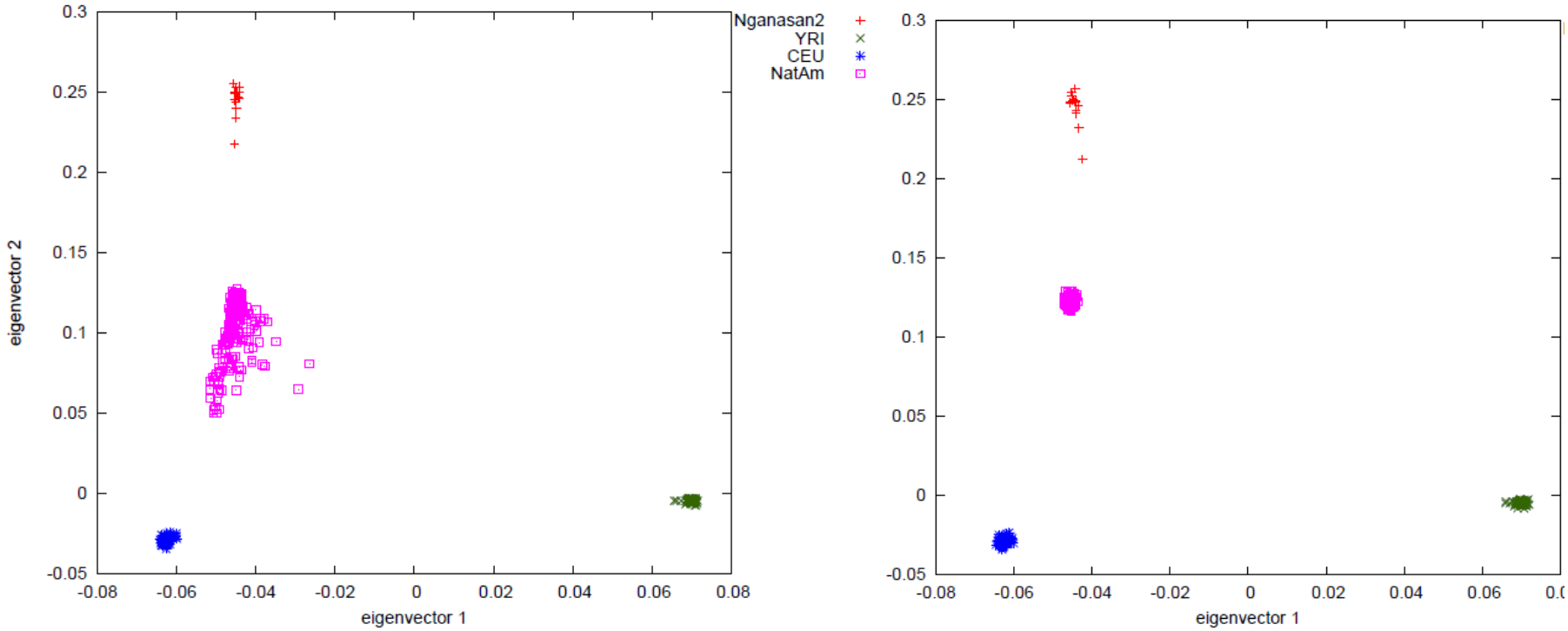


Figure S3. Examples of masking of segments of non-Native American ancestry.
Estimates from HAPMIX of the number of European or African alleles (*y* axis) across chromosome 7 (position on *x*-axis). Results are shown for selected (A) Native American and (B) Siberian or North American Arctic samples. The inferences in Native Americans in general show crisp transitions between segments of entirely Native American and likely admixed segments. Our main analyses restrict to loci where the expected number of European or African alleles is <0.01.

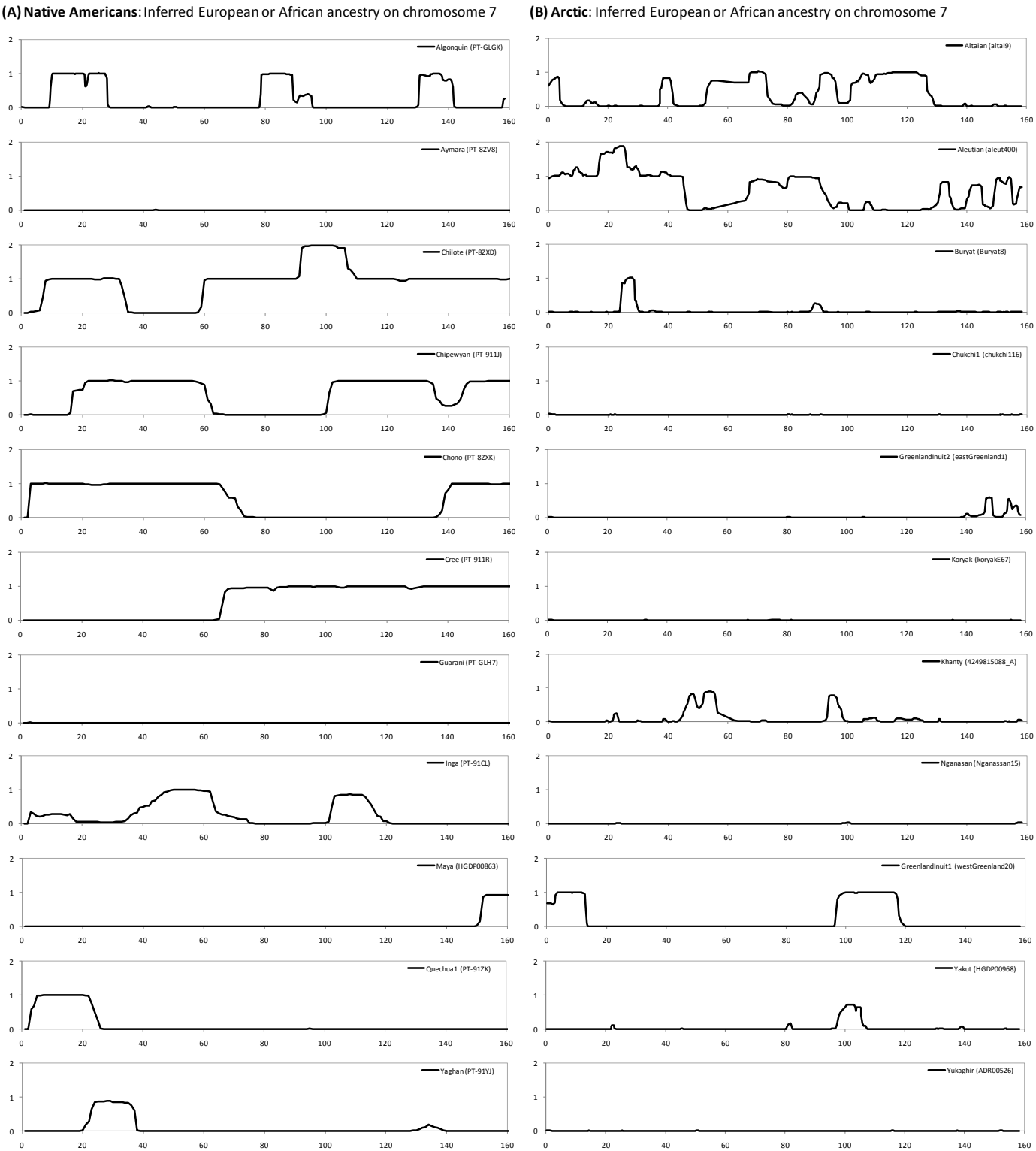
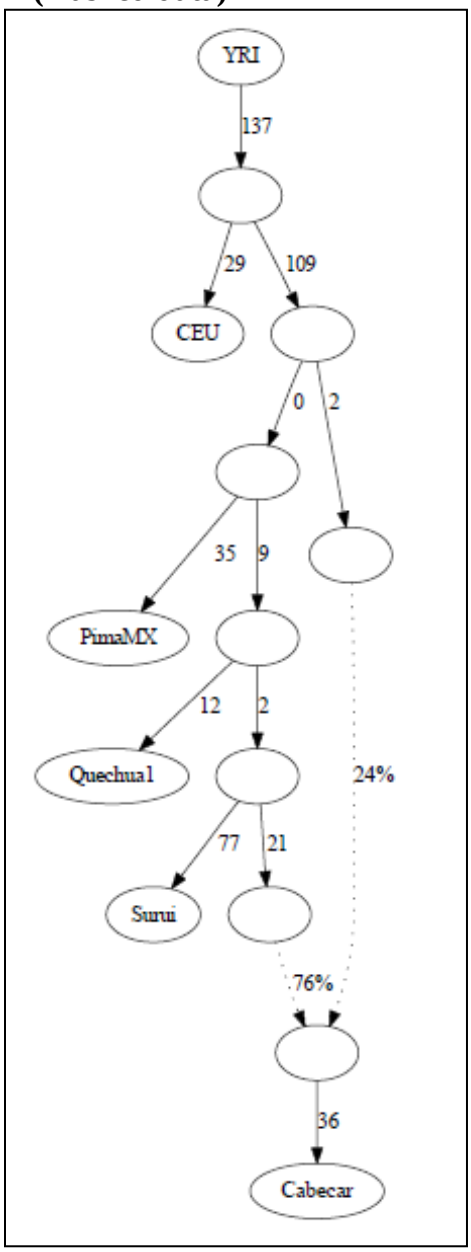
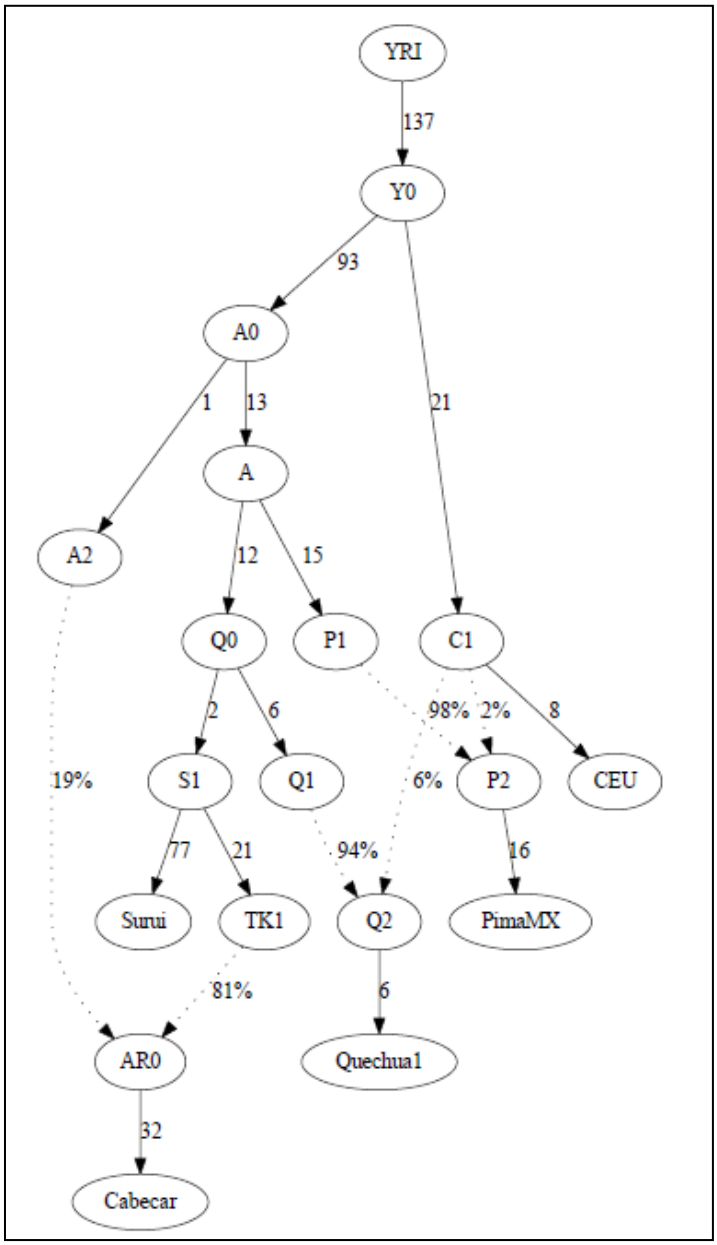


Figure S4. The evidence of ancient admixture in Chibchans is not an artifact of masking.
Admixture Graph analysis modeling how the Cabecar may relate to other Native Americans. Solid lines indicate genetic drift estimated to have occurred on each lineage (units proportional to F_{ST}×1000), and dotted lines indicate mixture proportions. Both graphs are excellent fits to the data (no *f*-statistics are more than |Z|>3 standard errors from expectation) as long as the Cabecar are considered to be an ancient admixture of a South American lineage and a lineage that roots deeply in the tree of Native North Americans. In contrast, if the Cabecar are modeled as unadmixed, we observe model failure: one statistic at |Z|>3 (=3.5) for the masked case, and 12 statistics at |Z|>3 (highest 6.7) for the unmasked case. (A) The Admixture Graph analysis on masked data focuses on a set of populations relevant to the mixture history in Chibchan-speakers. Two differences from Figure 3 include using CEU as a non-Native American outgroup (to replace CHB), and Cabecar as a representative of Chibchan-speakers (to replace Kogi and Arhuaco). (B) We obtained an equally good fit to the data (no |Z|-scores greater than 3, and similar estimates of genetic drift and mixture proportions) on the masked data, after modeling post-Colombian admixture in the PimaMX and the Quechua1, two populations that according to Table S1 have appreciable post-Colombian admixture.

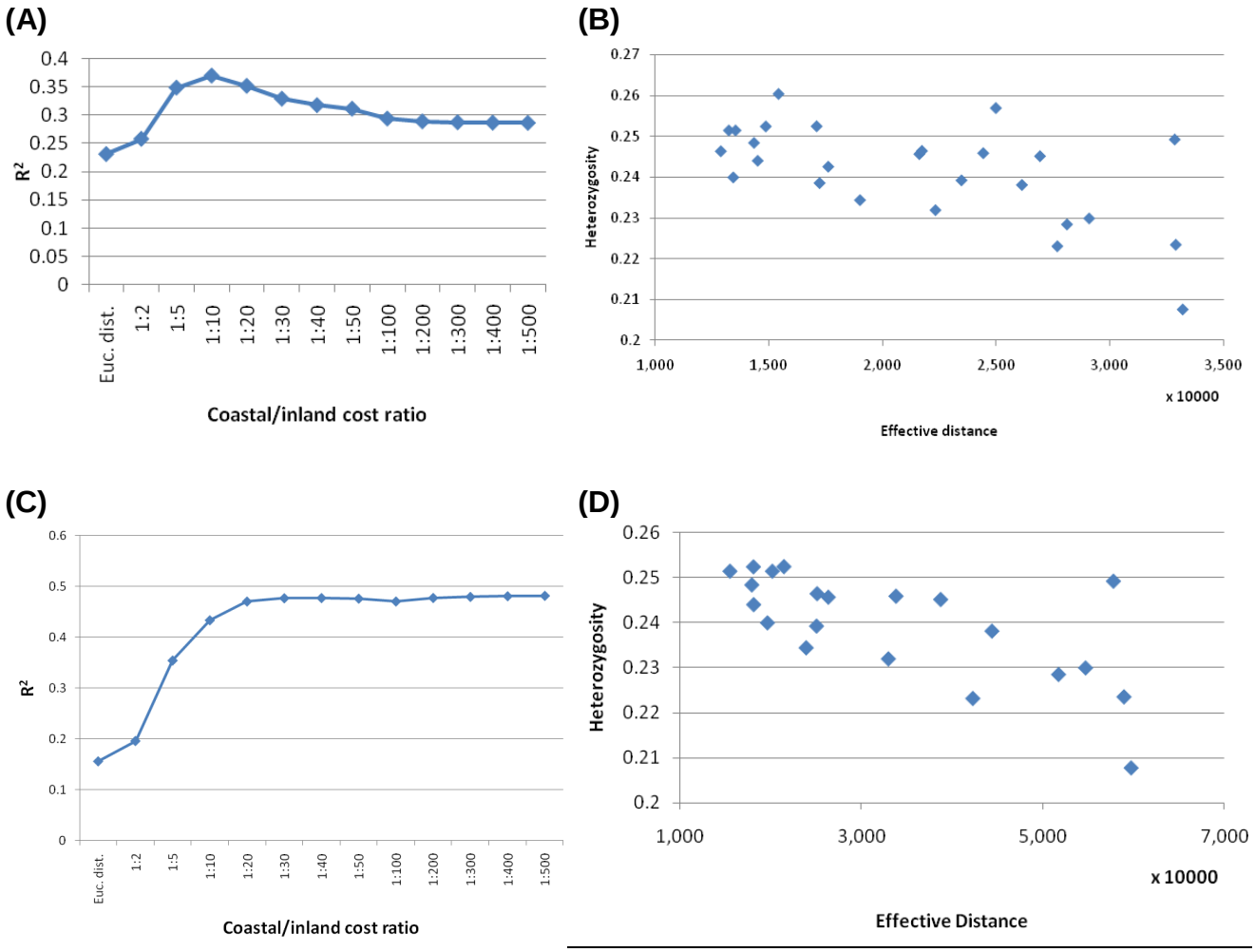
A (masked data)



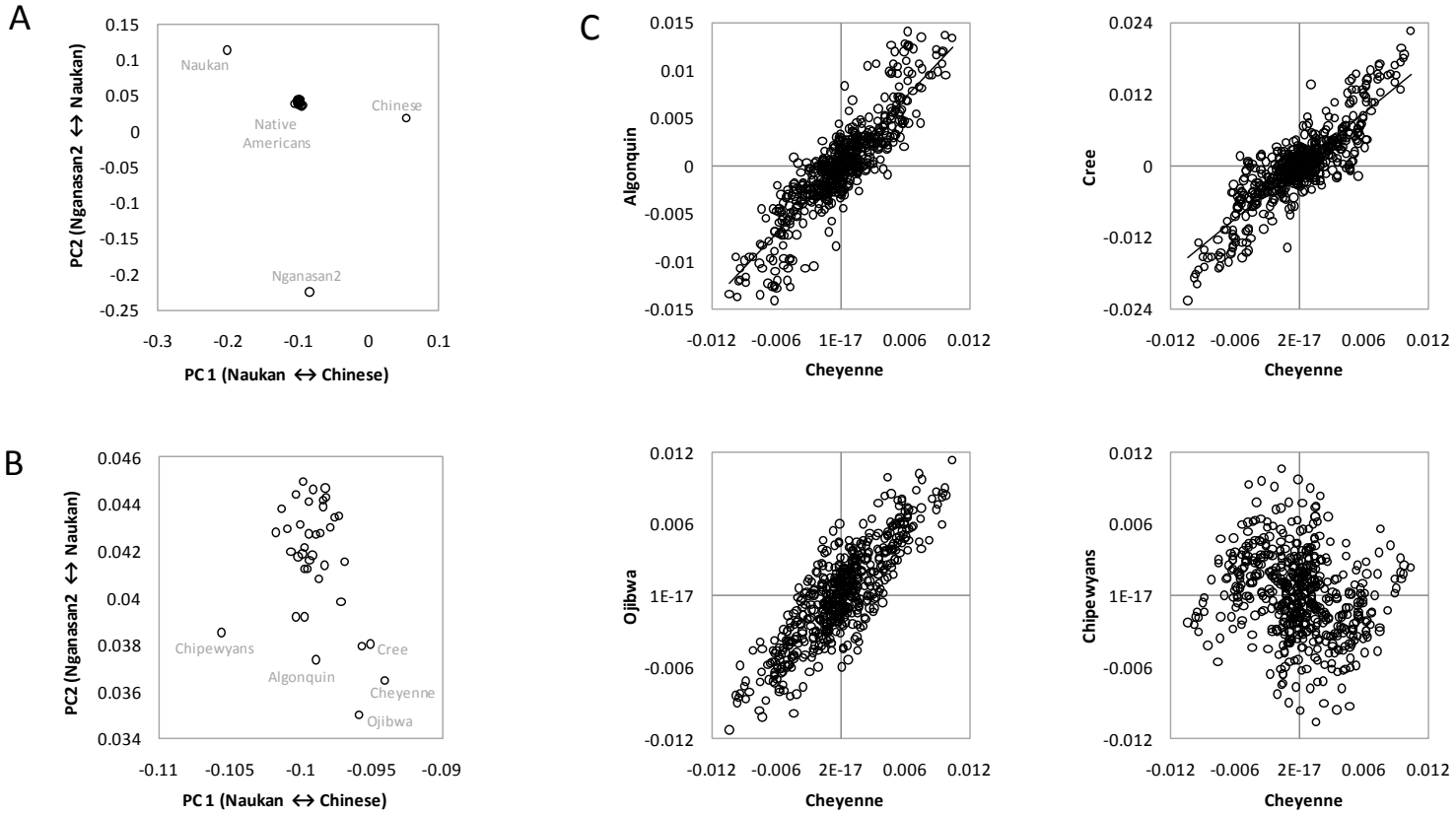
B (unmasked data)



<u>Figure S5</u>. Heterozygosity and geographic distance from the Bering Strait.
(A) We report R^2 (square of correlation) between mean population heterozygosity (for populations with 5 or more individuals genotyped) and distance from the Bering Strait (excluding populations in the Lower Central America/North-West South America cluster). Least-cost distances are based on coastal/inland cost ratios that assume greater permeability of the coasts relative to inland regions. All correlations are statistically significant, with P<0.05. The highest correlation is obtained when coastlines are set to be ten times more permeable than inland routes. (B) Scatter plot of heterozygosity and effective distance from the Bering Strait at the coastal/inland cost ratio of 1:10. The correlation is r = -0.60 (P =0.001) (C) R^2 when the 5 most Northern Native American populations (NNA - Ojibwa, Chipewyan, Cree, Algonquin and Cheyenne) are excluded. All correlations are significant, with P<0.05, except for the simple great arc distance (P=0.07). (D) Scatter plot of heterozygosity and effective distance from the Bering Strait at the coastal/inland cost ratio of 1:30 from panel C. Correlation is r = -0.69 (P =0.0005). The x-axis in panels B and D are in units of effective distance, with no meaning for absolute values.

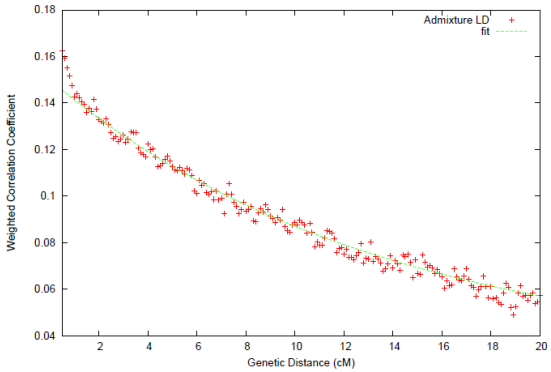


<u>Figure S6</u>. Native North Americans have a distinct relationship to Eurasians.
(A) We ran PCA on CEU European Americans, Naukan Inuit, and Nganasan2 Siberians, and projected on the top two PCs the mean scores for the 38 Native American populations with at least 4 samples. Most Native American populations are in one cluster, as would be expected if they descend from a homogeneous founding population. (B) At higher resolution (removing the CEU, Naukan and Nganasan2 from the plot), we observe that the 5 most Northern Native American (NNA) groups are outliers. The distinct relatedness of the NNA to Eurasian groups is confirmed by statistical analysis in Table S5 and Table S6. (C) To better understand how the NNA groups relate to Siberians, we computed statistics of the form f_4(Zapotec, NNA; Outgroup1, Outgroup2)—which have an expectation of zero if the Zapotec and NNA are sister groups relative to the two outgroups—for all possible pairs of 23 Siberian/Arctic populations and CHB as outgroups (the same quantities are tabulated in Table S6). Plotting the results, we see that the Cheyenne have a pattern that is highly correlated to that seen in the Algonquin (r^2=0.78), Cree (r^2=0.74), and Ojibwa (r^2=0.72). The correlation to the Chipewyan is poor (r^2=0.05), suggesting a very different relationship to Old World groups than to the four other NNA populations.

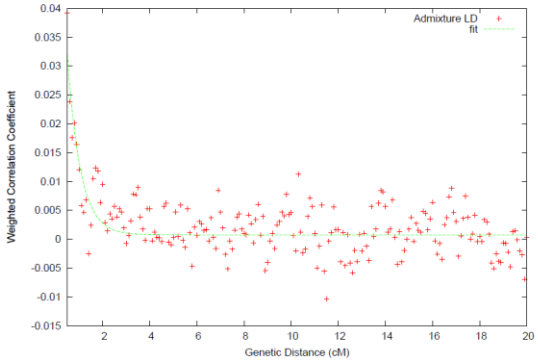


<u>Figure S7.</u> Dates of admixture events from the decay of admixture linkage disequilibrium.
ROLLOFF output along with the estimated number of generations since mixture plus or minus one standard error for populations in which there is a visually evident decay. The surrogate ancestral populations that are used are shown in Table S7. We observe an evident decay of admixture LD in the Maya (unmasked data) consistent with post-Colombian admixture. Using masked data (devoid of non-Native American ancestry), we find evident admixture LD in the Cheyenne, Inga, Guarani, and in many Chibchan populations. While the limited sample sizes often make it difficult to observe the exponential decay, the Cheyenne have a sufficiently clear decay that we can rule out post-Colombian dates (182 ± 80 generations, corresponding to a 90% confidence interval of 1,500-9,100 years assuming 29 years per generation). We also obtain confidently old dates in some Chibchan populations: for example in a pool of the Kogi and Arhuaco of 158 ± 38 (2,800-6,400 years) and the Cabecar of 241 ± 41 (5,000-8,900 years), suggesting that the deep admixture event described in Note S5 is likely to be a reflection of very old population migrations and admixture events.

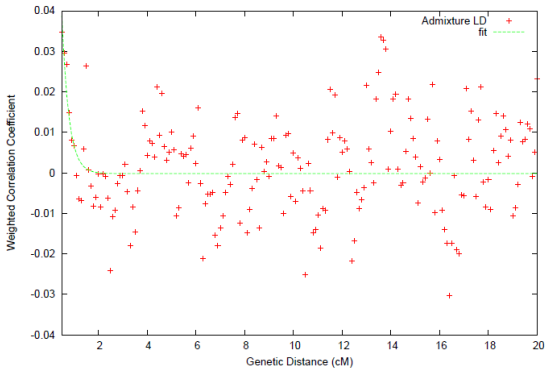
(A) Maya: 7.4 ± 0.7 generations



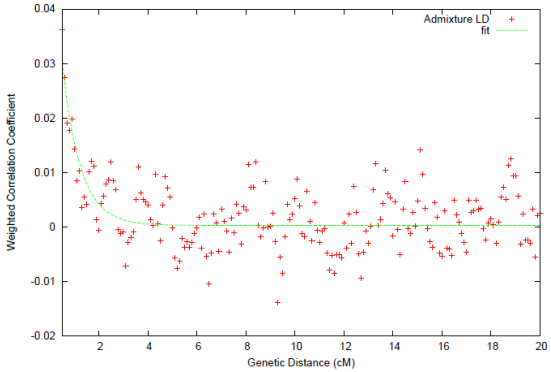
(B) Cheyenne: 182 ± 80 generations



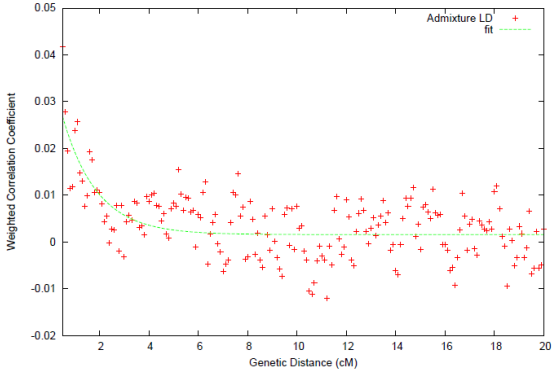
(C) Chipewyan: NO VISIBLE DECAY



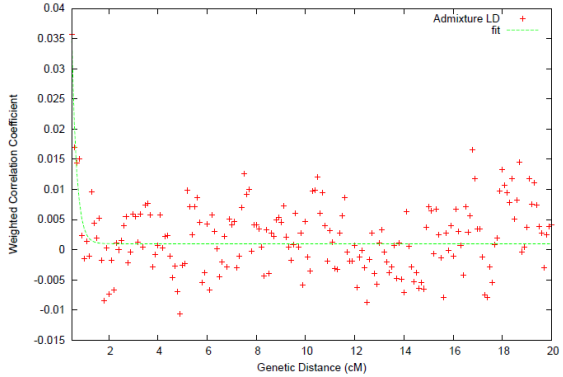
(D) Inga: 82 ± 95 generations



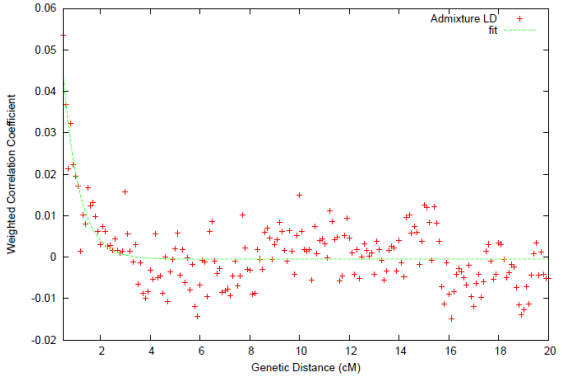
(E) Guarani: 39 ± 45 generations



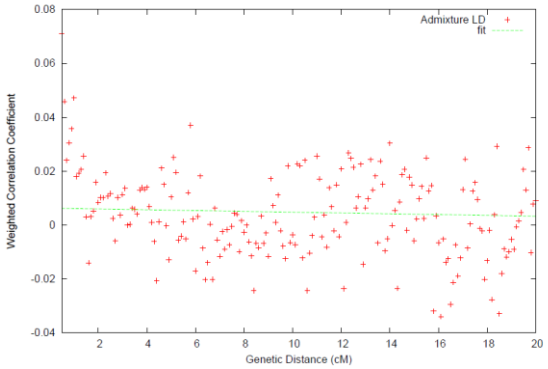
(F) Guahibo: NO VISIBLE DECAY



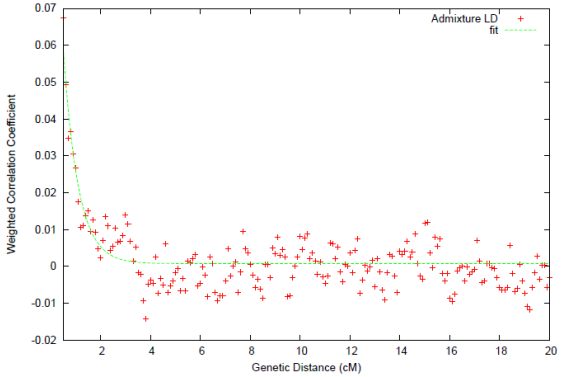
(G) Kogi: 140 ± 41 generations



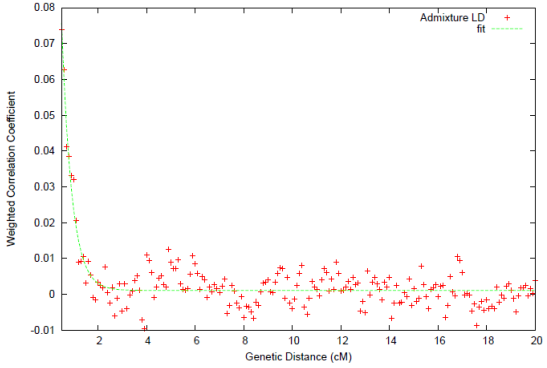
(H) Arhuaco: NO VISIBLE DECAY



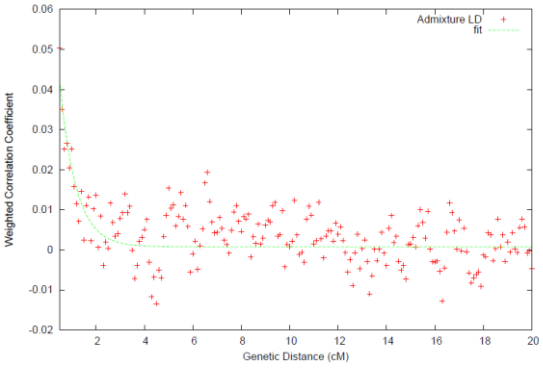
(I) Kogi+Arhuaco: 158 ± 38 generations



(J) Cabecar: 241 ± 41 generations



(K) Guaymi: 147 ± 49 generations



(L) Zenu: 272 ± 87 generations

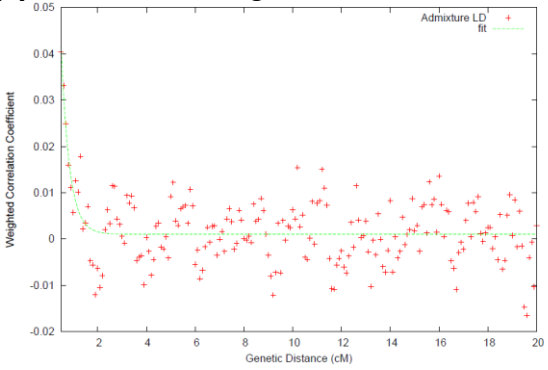


Table S1. Summary information for 55 Native American populations

Population	N	Language family[1]	Sampling location	Lat.	Long.	Data source[2]	%Non-Native[3]	%Masked[4]	4 Population Test statistic (Z-score)
Algonquin	5	Northern Amerind	Canada	48.4	-71.1	a	34%	49%	0.025 (Z=12.1)
Aleutian	9	Eskimo-Aleut	Aleutian Islands	52.0	-176.6	c	n/a	not masked	0.072 (Z=49.9)
Arara	1	Ge-Pano-Carib	Brazil	-4	-53.5	a	0%	14%	0.025 (Z=10.3)
Arhuaco	5	Chibchan-Paezan	Colombia	11	-73.8	a	23%	43%	0.022 (Z=10.2)
Aymara	23	Andean	Bolivia(&Chile)	16.5(-22)	-68.2(-70)	a	3%	8%	0.028 (Z=18.0)
Bribri	4	Chibchan-Paezan	Costa Rica	9.4	-83.1	a	3%	8%	0.027 (Z=14.5)
Cabecar	31	Chibchan-Paezan	Costa Rica	9.5	-84	a	2%	4%	0.027 (Z=16.2)
Chane	2	Equatorial-Tucanoan	Argentina	-22.3	-63.7	a	0%	2%	0.026 (Z=13.1)
Cheyenne	24	Northern Amerind	USA	35.5	-99	a	n/a	8%	0.027 (Z=18.5)
Chilote	8	Andean	Chile	-42.5	-73.9	a	39%	67%	0.029 (Z=14.7)
Chipewyan	5	Na-Dene	Canada	59.6	-107.3	a	33%	45%	0.027 (Z=14.4)
Chono	4	Andean	Chile	-45	-74	a	32%	56%	0.026 (Z=11.2)
Chorotega	1	Central-Amerind	Costa Rica	10.1	-85.5	a	25%	48%	0.021 (Z=6.1)
Cree	4	Northern Amerind	Canada	50.3	-102.5	a	44%	65%	0.026 (Z=12.1)
Diaguita	5	Andean	Argentina	-28.5	-65.8	a	25%	49%	0.028 (Z=14.1)
Embera	5	Chibchan-Paezan	Colombia	7	-76	a	0%	0%	0.027 (Z=14.6)
GreenlandInuit1	7	Eskimo-Aleut	Greenland	67.5	-37.9	c	n/a	55%	0.047 (Z=32.1)
GreenlandInuit2	8	Eskimo-Aleut	Greenland	65.3	-52.0	c	n/a	38%	0.028 (Z=18.5)
Guahibo	6	Equatorial-Tucanoan	Colombia	5.8	-69.5	a	0%	0%	0.027 (Z=15.6)
Guarani	6	Equatorial-Tucanoan	Paraguay(&Argentina)	-23(-22.5)	-54(-63.8)	a	8%	15%	0.026 (Z=15.2)
Guaymi	5	Chibchan-Paezan	Costa Rica	8.5	-82	a	0%	1%	0.028 (Z=15.0)
Huetar	1	Chibchan-Paezan	Costa Rica	9.7	-84.3	a	26%	47%	0.027 (Z=8.8)
Hulliche	4	Andean	Chile	-41	-73	a	12%	25%	0.029 (Z=15.9)
Inga	10	Andean	Colombia	1	-77	a	13%	33%	0.027 (Z=15.7)
Jamamadi	1	Equatorial-Tucanoan	Brazil	-8.5	-64.5	a	0%	0%	0.024 (Z=10.0)
Kaingang	2	Ge-Pano-Carib	Brazil	-24	-52.5	a	16%	34%	0.028 (Z=12.1)
Kalina	2	Ge-Pano-Carib	Guiana	5.7	-53.9	a	4%	6%	0.036 (Z=13.5)
Kaqchikel	13	Northern Amerind	Guatemala	15	-91	a	9%	18%	0.029 (Z=18.8)
Karitiana	13	Equatorial-Tucanoan	Brazil	-10	-63	b	0%	0%	0.028 (Z=15.7)
Kogi	4	Chibchan-Paezan	Colombia	11	-74	a	0%	0%	0.026 (Z=13.7)
Maleku	3	Chibchan-Paezan	Costa Rica	10.6	-84.8	a	3%	6%	0.026 (Z=12.6)
Maya	18	Northern Amerind	Mexico	20.3	-87.8	b	13%	25%	0.028 (Z=18.5)
Mixe	17	Northern Amerind	Mexico	17	-96	a	1%	3%	0.029 (Z=17.7)
Mixtec	5	Central-Amerind	Mexico	17	-97	a	5%	10%	0.029 (Z=16.6)
Ojibwa	5	Northern Amerind	Canada	46.5	-81	a	33%	50%	0.024 (Z=12.7)
Palikur	3	Equatorial-Tucanoan	Guiana	4	-51.8	a	1%	3%	0.028 (Z=15.1)
Parakana	1	Equatorial-Tucanoan	Brazil	-4.8	-50	a	0%	0%	0.025 (Z=11.0)
Piapoco	7	Equatorial-Tucanoan	Colombia	3	-68	b	2%	4%	0.024 (Z=14.0)
PimaAZ	22	Central-Amerind	USA	33.5	-111.8	a	0%	1%	0.028 (Z=18.5)
PimaMX	33	Central-Amerind	Mexico	29.3	-108.8	a&b	4%	8%	0.026 (Z=15.9)
Purepecha	1	Chibchan-Paezan	Mexico	19	-101.5	a	19%	34%	0.023 (Z=8.1)
Quechua1	18	Andean	Bolivia	-14.5	-69	a	5%	12%	0.029 (Z=18.5)
Quechua2	22	Andean	Peru	-14	-74	a	10%	22%	0.029 (Z=18.6)
Surui	24	Equatorial-Tucanoan	Brazil	-11	-62	b	0%	0%	0.028 (Z=15.5)
Teribe	3	Chibchan-Paezan	Costa Rica	9	-83.2	a	0%	1%	0.029 (Z=14.6)
Ticuna1	6	Equatorial-Tucanoan	Colombia	-3.81	-70.01	a	1%	5%	0.027 (Z=15.4)
Ticuna2	12	Equatorial-Tucanoan	Brazil	-3.5	-69	a	1%	3%	0.027 (Z=16.3)
Toba	4	Ge-Pano-Carib	Argentina	-26.5	-59.3	a	1%	4%	0.025 (Z=14.2)
Waunana	3	Chibchan-Paezan	Colombia	5	-77	a	0%	3%	0.027 (Z=14.7)
Wayuu	12	Equatorial-Tucanoan	Colombia	11	-73	a	10%	25%	0.025 (Z=15.2)
Wichi	5	Ge-Pano-Carib	Argentina	-22.5	-63.8	a	3%	5%	0.026 (Z=14.6)
Yaghan	4	Andean	Chile	-55	-68	a	25%	53%	0.027 (Z=13.6)
Yaqui	1	Central-Amerind	Mexico	28	-110.3	a	21%	47%	0.030 (Z=9.5)
Zapotec	23	Central-Amerind	Mexico	16.5(16)	-97.2(-97)	a	7%	16%	0.028 (Z=19.3)
Zenu	5	Chibchan-Paezan	Colombia	9	-75	a	8%	17%	0.028 (Z=15.1)

[1] Greenberg subdivides the "superfamily" Amerind into 7 subfamilies and this classification is used here (*27, 28*).
[2] Data sources are: (a) This study, (b) Li et al 2008 (*29*), (c) Rasmussen et al. 2010 (*31*).
[3] Estimate of non-Native American ancestry is based on ADMIXTURE with k=3 (the other two ancestries are European and West African).
[4] Percent of genome masked based on HAPMIX (where the posterior estimate of the number of non-Native American chromosomes is >0.01).

Table S2. Summary information for 19 Siberian populations

Population	N	Language Family	Location	Lat.	Long.	Data source[2]
Altaian	13	Altaic	Russia	56.3	82.8	1
Buryat	18	Altaic	Russia	52.6	104.3	1
Cambodian	10	Austric	Cambodia	12.0	105.0	2
Chukchi1	11	Chukchi-Kamchatkan	Russia	67.8	-178.4	1
Chukchi2	19	Chukchi-Kamchatkan	Russia	69	170	3
Dolgan	6	Altaic	Russia	69.8	88.1	1
Evenki	15	Altaic	Russia	64.1	95.4	1
Ket	2	Isolate	Russia	63.8	87.4	1
Khanty	39	Uralic-Yukaghir	Russia	63	76.5	3
Koryak	10	Chukchi-Kamchatkan	Russia	64.1	167.9	1
Naukan	16	Eskimo-Aleut	Russia	65	188	3
Nganasan1	9	Uralic-Yukaghir	Russia	73.3	88.0	3
Nganasan2	15	Uralic-Yukaghir	Russia	70	94	1
Selkup	9	Uralic-Yukaghir	Russia	66.4	84.9	1
Tundra Nentsi	4	Uralic-Yukaghir	Russia	66.1	76.5	3
Tuvinian	16	Altaic	Russia	52.0	94.4	1
Yakut1	24	Altaic	Russia	63	130	2
Yakut2	16	Altaic	Russia	63	135	3
Yukaghir	13	Uralic-Yukaghir	Russia	68	150	3

[1]Language classification follows Ruhlen 1991.
[2]Data sources: (1) Rasmussen et al. 2010; (2) Li et al. 2008; (3) this study.

Table S3. F_{ST} for populations used to build the Neighbor Joining tree (masked data)

	Cambodian	Han	Japanese	Mongolian	Yi	Yoruba	Aleutian	Algonquin	Arara	Arhuaco	Aymara	Bribri	Cabecar	Chane	Cheyenne	Chilote	Chipewyan	Chono	Chorotega	Cree	Diaguita	Embera	GreenlandI	GreenlandI	Guarani	Guaymi	Guihiba	Huetar	Huilliche	Ingano	Jamamadi	Kaingang	Kalina	Kaqchikel	Karitiana	Kogi	Maleku	Maya	Mixe	Mixtec	Ojibwa	Palikur	Parakana	Piapoco	PimaAZ	PimaMX	Purepecha	Quechua1	Quechua2	Surui	Teribe	Ticuna1	Ticuna2	Toba	Waunana	Wayuu	Wichi	Yaghan	Yaqui	Zapotec	Zenu	Altaian	Buryat	Chukchi1	Chukchi2	Dolgan	Evenki	Ket	Khanty	Koryak	Naukan	Nganasan1	Nganasan2	Selkup	Tundra_Ne	Tuvinians	Yakut1	Yakut2	Yukaghir
Cambodian		0.01	0.02	0.03	0.02	0.17	0.09	0.16	0.19	0.20	0.13	0.16	0.17	0.13	0.11	0.14	0.12	0.16	0.15	0.10	0.13	0.16	0.12	0.11	0.12	0.18	0.14	0.18	0.14	0.14	0.18	0.17	0.20	0.12	0.18	0.19	0.21	0.12	0.13	0.12	0.11	0.16	0.14	0.14	0.12	0.14	0.11	0.13	0.13	0.20	0.17	0.16	0.16	0.14	0.15	0.13	0.16	0.14	0.13	0.12	0.15	0.03	0.04	0.08	0.07	0.05	0.05	0.06	0.07	0.09	0.09	0.08	0.08	0.07	0.08	0.04	0.05	0.05	0.05
Han	0.01		0.01	0.02	0.01	0.19	0.09	0.15	0.20	0.20	0.13	0.16	0.17	0.13	0.11	0.14	0.12	0.16	0.15	0.10	0.13	0.16	0.11	0.11	0.12	0.18	0.14	0.18	0.14	0.14	0.18	0.17	0.20	0.12	0.18	0.19	0.21	0.12	0.13	0.12	0.11	0.16	0.13	0.14	0.12	0.14	0.11	0.13	0.13	0.20	0.17	0.16	0.15	0.13	0.15	0.13	0.16	0.14	0.12	0.11	0.15	0.03	0.03	0.07	0.07	0.04	0.04	0.06	0.07	0.09	0.09	0.08	0.07	0.06	0.08	0.03	0.04	0.04	0.05
Japanese	0.02	0.01		0.02	0.01	0.19	0.08	0.15	0.19	0.19	0.13	0.16	0.16	0.13	0.11	0.14	0.12	0.16	0.15	0.10	0.13	0.16	0.11	0.10	0.12	0.18	0.14	0.18	0.14	0.14	0.18	0.17	0.20	0.11	0.18	0.19	0.21	0.11	0.13	0.12	0.11	0.16	0.13	0.14	0.12	0.14	0.10	0.13	0.12	0.20	0.16	0.16	0.15	0.13	0.15	0.13	0.16	0.14	0.12	0.11	0.14	0.02	0.02	0.07	0.06	0.04	0.03	0.05	0.06	0.08	0.09	0.07	0.07	0.06	0.08	0.03	0.03	0.03	0.04
Mongolian	0.03	0.02	0.02		0.02	0.17	0.06	0.13	0.18	0.17	0.11	0.14	0.14	0.11	0.09	0.12	0.10	0.14	0.13	0.08	0.11	0.14	0.09	0.09	0.10	0.16	0.12	0.16	0.12	0.12	0.16	0.15	0.18	0.10	0.16	0.17	0.19	0.10	0.11	0.10	0.09	0.14	0.12	0.12	0.10	0.12	0.09	0.11	0.10	0.18	0.15	0.14	0.13	0.11	0.13	0.11	0.14	0.12	0.10	0.09	0.13	0.01	0.01	0.05	0.05	0.02	0.02	0.03	0.04	0.07	0.07	0.05	0.05	0.04	0.05	0.01	0.02	0.02	0.02
Yi	0.02	0.01	0.01	0.02		0.19	0.08	0.15	0.20	0.19	0.13	0.16	0.16	0.13	0.11	0.14	0.12	0.16	0.15	0.10	0.13	0.16	0.11	0.11	0.12	0.18	0.14	0.18	0.14	0.14	0.18	0.17	0.20	0.11	0.18	0.19	0.21	0.12	0.13	0.12	0.11	0.16	0.13	0.14	0.12	0.14	0.11	0.13	0.12	0.20	0.17	0.16	0.15	0.13	0.15	0.13	0.16	0.14	0.12	0.11	0.14	0.02	0.03	0.07	0.07	0.04	0.04	0.05	0.06	0.09	0.09	0.07	0.07	0.06	0.08	0.03	0.03	0.04	0.04
Yoruba	0.17	0.19	0.19	0.17	0.19		0.22	0.27	0.31	0.30	0.25	0.27	0.28	0.25	0.23	0.26	0.25	0.28	0.27	0.23	0.25	0.28	0.24	0.24	0.25	0.29	0.26	0.29	0.26	0.26	0.30	0.28	0.32	0.24	0.29	0.30	0.32	0.24	0.25	0.24	0.23	0.28	0.25	0.26	0.24	0.26	0.23	0.25	0.25	0.31	0.28	0.28	0.27	0.25	0.27	0.25	0.28	0.26	0.25	0.24	0.26	0.19	0.19	0.21	0.21	0.20	0.20	0.19	0.19	0.23	0.23	0.22	0.22	0.19	0.21	0.19	0.19	0.20	0.20
Aleutian	0.09	0.09	0.08	0.06	0.08	0.22		0.14	0.18	0.18	0.11	0.14	0.15	0.11	0.09	0.12	0.10	0.14	0.15	0.09	0.11	0.14	0.08	0.07	0.10	0.16	0.12	0.18	0.12	0.12	0.17	0.16	0.19	0.10	0.16	0.18	0.19	0.10	0.11	0.10	0.09	0.15	0.11	0.12	0.10	0.13	0.09	0.11	0.10	0.19	0.14	0.15	0.14	0.11	0.13	0.11	0.14	0.12	0.11	0.10	0.12	0.07	0.07	0.06	0.06	0.08	0.08	0.07	0.08	0.09	0.06	0.10	0.10	0.08	0.10	0.07	0.08	0.08	0.07
Algonquin	0.16	0.15	0.15	0.13	0.15	0.27	0.14		0.20	0.20	0.12	0.16	0.16	0.12	0.08	0.13	0.10	0.16	0.14	0.05	0.12	0.15	0.16	0.16	0.11	0.18	0.14	0.18	0.13	0.13	0.18	0.17	0.20	0.11	0.18	0.19	0.22	0.11	0.12	0.11	0.07	0.16	0.13	0.13	0.11	0.14	0.09	0.12	0.11	0.20	0.16	0.16	0.15	0.12	0.14	0.13	0.16	0.14	0.11	0.10	0.14	0.14	0.15	0.14	0.14	0.15	0.15	0.14	0.15	0.17	0.14	0.18	0.18	0.15	0.17	0.14	0.15	0.15	0.14
Arara	0.19	0.20	0.19	0.18	0.20	0.31	0.18	0.20		0.20	0.11	0.15	0.16	0.11	0.14	0.13	0.17	0.16	0.14	0.12	0.12	0.15	0.21	0.20	0.10	0.18	0.13	0.19	0.13	0.13	0.18	0.15	0.18	0.10	0.16	0.19	0.22	0.10	0.13	0.11	0.14	0.15	0.10	0.12	0.13	0.15	0.11	0.12	0.11	0.18	0.16	0.15	0.14	0.12	0.14	0.12	0.15	0.14	0.14	0.11	0.14	0.18	0.19	0.19	0.19	0.20	0.20	0.19	0.20	0.22	0.19	0.22	0.22	0.20	0.22	0.19	0.20	0.20	0.19
Arhuaco	0.20	0.20	0.19	0.17	0.19	0.30	0.18	0.20	0.20		0.12	0.12	0.13	0.12	0.14	0.14	0.17	0.16	0.13	0.13	0.12	0.14	0.20	0.20	0.11	0.15	0.14	0.16	0.14	0.13	0.19	0.16	0.19	0.11	0.18	0.14	0.19	0.11	0.13	0.12	0.14	0.16	0.13	0.13	0.13	0.15	0.11	0.12	0.12	0.20	0.13	0.16	0.15	0.13	0.13	0.11	0.16	0.15	0.14	0.11	0.13	0.18	0.19	0.19	0.18	0.20	0.19	0.19	0.19	0.22	0.19	0.22	0.22	0.19	0.22	0.19	0.19	0.20	0.19
Aymara	0.13	0.13	0.13	0.11	0.13	0.25	0.11	0.12	0.11	0.12		0.07	0.08	0.03	0.06	0.04	0.09	0.07	0.07	0.05	0.02	0.07	0.13	0.13	0.03	0.10	0.05	0.10	0.05	0.05	0.10	0.08	0.11	0.02	0.10	0.11	0.14	0.03	0.05	0.03	0.06	0.08	0.04	0.05	0.05	0.07	0.02	0.00	0.01	0.12	0.08	0.08	0.07	0.04	0.06	0.04	0.08	0.05	0.05	0.03	0.06	0.11	0.12	0.11	0.11	0.13	0.13	0.12	0.13	0.14	0.11	0.15	0.15	0.13	0.15	0.12	0.13	0.13	0.12
Bribri	0.16	0.16	0.16	0.14	0.16	0.27	0.14	0.16	0.15	0.12	0.07		0.03	0.07	0.09	0.09	0.13	0.11	0.07	0.09	0.07	0.08	0.16	0.16	0.07	0.07	0.09	0.08	0.09	0.09	0.14	0.12	0.16	0.05	0.14	0.11	0.12	0.06	0.09	0.07	0.10	0.11	0.08	0.09	0.09	0.11	0.06	0.08	0.07	0.16	0.04	0.11	0.10	0.08	0.07	0.05	0.11	0.10	0.09	0.07	0.07	0.15	0.15	0.15	0.15	0.16	0.16	0.15	0.16	0.18	0.15	0.18	0.18	0.16	0.18	0.15	0.16	0.16	0.15
Cabecar	0.17	0.17	0.16	0.14	0.16	0.28	0.15	0.16	0.16	0.13	0.08	0.03		0.08	0.10	0.09	0.13	0.12	0.08	0.10	0.08	0.09	0.17	0.17	0.07	0.07	0.10	0.08	0.10	0.09	0.15	0.12	0.16	0.06	0.14	0.12	0.13	0.07	0.09	0.08	0.10	0.12	0.09	0.09	0.09	0.12	0.07	0.08	0.08	0.17	0.05	0.12	0.11	0.09	0.08	0.06	0.12	0.10	0.10	0.07	0.08	0.15	0.16	0.15	0.15	0.17	0.16	0.16	0.17	0.18	0.15	0.19	0.19	0.16	0.19	0.16	0.16	0.16	0.16
Chane	0.13	0.13	0.13	0.11	0.13	0.25	0.11	0.12	0.11	0.12	0.03	0.07	0.08		0.06	0.04	0.09	0.07	0.05	0.05	0.03	0.07	0.13	0.12	0.01	0.09	0.04	0.10	0.05	0.04	0.10	0.07	0.10	0.02	0.09	0.11	0.14	0.02	0.05	0.03	0.06	0.07	0.03	0.04	0.05	0.07	0.03	0.03	0.03	0.11	0.08	0.07	0.06	0.02	0.06	0.04	0.05	0.05	0.04	0.03	0.05	0.11	0.12	0.11	0.11	0.13	0.12	0.12	0.13	0.14	0.11	0.15	0.15	0.13	0.15	0.12	0.12	0.13	0.12
Cheyenne	0.11	0.11	0.11	0.09	0.11	0.23	0.09	0.08	0.14	0.14	0.06	0.09	0.10	0.06		0.07	0.06	0.09	0.08	0.02	0.05	0.09	0.11	0.11	0.05	0.12	0.07	0.12	0.07	0.07	0.12	0.10	0.14	0.04	0.12	0.13	0.16	0.04	0.06	0.05	0.03	0.10	0.06	0.07	0.05	0.07	0.03	0.06	0.05	0.14	0.10	0.10	0.09	0.06	0.08	0.06	0.10	0.08	0.05	0.04	0.08	0.10	0.10	0.10	0.09	0.11	0.11	0.10	0.11	0.13	0.10	0.13	0.13	0.11	0.13	0.10	0.11	0.11	0.10
Chilote	0.14	0.14	0.14	0.12	0.14	0.26	0.12	0.13	0.13	0.14	0.04	0.09	0.09	0.04	0.07		0.11	0.03	0.08	0.06	0.04	0.08	0.14	0.14	0.04	0.11	0.07	0.12	0.01	0.06	0.11	0.10	0.12	0.04	0.11	0.13	0.15	0.04	0.06	0.05	0.07	0.09	0.05	0.06	0.06	0.09	0.04	0.04	0.04	0.13	0.10	0.09	0.08	0.05	0.07	0.05	0.09	0.03	0.06	0.04	0.07	0.12	0.13	0.12	0.12	0.14	0.14	0.13	0.14	0.15	0.12	0.16	0.16	0.14	0.16	0.13	0.13	0.14	0.13
Chipewya	0.12	0.12	0.12	0.10	0.12	0.25	0.10	0.10	0.17	0.17	0.09	0.13	0.13	0.09	0.06	0.11		0.13	0.12	0.04	0.09	0.13	0.12	0.11	0.09	0.15	0.11	0.15	0.11	0.11	0.16	0.13	0.17	0.08	0.15	0.17	0.19	0.08	0.10	0.08	0.06	0.13	0.10	0.11	0.08	0.11	0.07	0.09	0.09	0.18	0.14	0.13	0.12	0.10	0.12	0.10	0.13	0.11	0.09	0.08	0.11	0.10	0.11	0.10	0.10	0.12	0.12	0.11	0.12	0.13	0.10	0.14	0.14	0.12	0.14	0.11	0.12	0.12	0.11
Chono	0.16	0.16	0.16	0.14	0.16	0.28	0.14	0.16	0.16	0.16	0.07	0.11	0.12	0.07	0.09	0.03	0.13		0.11	0.09	0.07	0.11	0.17	0.16	0.07	0.14	0.09	0.16	0.03	0.09	0.15	0.12	0.16	0.06	0.14	0.16	0.18	0.07	0.09	0.07	0.10	0.12	0.08	0.09	0.09	0.11	0.07	0.07	0.07	0.16	0.12	0.12	0.11	0.08	0.10	0.08	0.11	0.06	0.09	0.07	0.10	0.15	0.16	0.15	0.15	0.16	0.16	0.15	0.16	0.18	0.15	0.19	0.19	0.16	0.18	0.15	0.16	0.16	0.15
Chorotega	0.15	0.15	0.15	0.13	0.15	0.27	0.15	0.14	0.14	0.13	0.07	0.07	0.08	0.05	0.08	0.08	0.12	0.11		0.08	0.07	0.09	0.16	0.16	0.06	0.10	0.08	0.12	0.08	0.08	0.11	0.12	0.15	0.05	0.13	0.13	0.14	0.05	0.07	0.06	0.09	0.11	0.07	0.08	0.07	0.10	0.05	0.07	0.06	0.15	0.09	0.11	0.10	0.07	0.08	0.06	0.11	0.09	0.07	0.05	0.08	0.14	0.15	0.14	0.14	0.15	0.15	0.15	0.15	0.17	0.14	0.18	0.17	0.15	0.17	0.14	0.15	0.15	0.14
Cree	0.10	0.10	0.10	0.08	0.10	0.23	0.09	0.05	0.12	0.13	0.05	0.09	0.10	0.05	0.02	0.06	0.04	0.09	0.08		0.05	0.09	0.11	0.10	0.05	0.11	0.07	0.12	0.07	0.07	0.11	0.10	0.14	0.04	0.12	0.13	0.15	0.04	0.06	0.04	0.01	0.09	0.06	0.07	0.04	0.07	0.04	0.05	0.05	0.13	0.10	0.09	0.08	0.06	0.06	0.04	0.09	0.07	0.05	0.04	0.07	0.08	0.09	0.08	0.08	0.10	0.09	0.09	0.10	0.11	0.09	0.12	0.12	0.10	0.12	0.09	0.10	0.10	0.09
Diaguita	0.13	0.13	0.13	0.11	0.13	0.25	0.11	0.12	0.12	0.12	0.02	0.07	0.08	0.03	0.05	0.04	0.09	0.07	0.07	0.05		0.07	0.13	0.12	0.03	0.10	0.05	0.11	0.04	0.05	0.10	0.07	0.11	0.02	0.10	0.12	0.14	0.03	0.05	0.03	0.06	0.07	0.04	0.05	0.05	0.07	0.02	0.02	0.02	0.12	0.08	0.08	0.07	0.04	0.06	0.04	0.07	0.05	0.05	0.03	0.06	0.11	0.12	0.11	0.11	0.13	0.13	0.12	0.13	0.14	0.11	0.15	0.15	0.13	0.15	0.12	0.12	0.13	0.12
Embera	0.16	0.16	0.16	0.14	0.16	0.28	0.14	0.15	0.15	0.14	0.07	0.08	0.09	0.07	0.09	0.08	0.13	0.11	0.09	0.09	0.07		0.16	0.16	0.06	0.11	0.08	0.12	0.09	0.08	0.14	0.12	0.15	0.06	0.13	0.13	0.15	0.06	0.09	0.07	0.10	0.11	0.08	0.08	0.08	0.11	0.06	0.07	0.07	0.16	0.09	0.11	0.10	0.08	0.06	0.06	0.11	0.09	0.08	0.07	0.08	0.14	0.15	0.15	0.14	0.16	0.16	0.15	0.16	0.18	0.15	0.18	0.18	0.16	0.18	0.15	0.16	0.16	0.15
GreenlandI	0.12	0.11	0.11	0.09	0.11	0.24	0.08	0.16	0.21	0.20	0.13	0.16	0.17	0.13	0.11	0.14	0.12	0.17	0.16	0.11	0.13	0.16		0.02	0.12	0.19	0.15	0.19	0.14	0.15	0.19	0.17	0.21	0.12	0.19	0.20	0.22	0.12	0.14	0.12	0.12	0.17	0.14	0.14	0.12	0.15	0.11	0.13	0.13	0.21	0.17	0.17	0.16	0.14	0.15	0.13	0.17	0.15	0.13	0.12	0.15	0.10	0.10	0.07	0.07	0.11	0.10	0.10	0.11	0.11	0.05	0.13	0.13	0.11	0.13	0.10	0.11	0.11	0.09
GreenlandI	0.11	0.11	0.10	0.09	0.11	0.24	0.07	0.16	0.20	0.20	0.13	0.16	0.17	0.12	0.11	0.14	0.11	0.16	0.16	0.10	0.12	0.16	0.02		0.12	0.18	0.14	0.18	0.14	0.14	0.19	0.17	0.21	0.11	0.18	0.20	0.22	0.12	0.13	0.12	0.11	0.17	0.13	0.14	0.12	0.14	0.10	0.13	0.12	0.21	0.17	0.16	0.15	0.13	0.15	0.13	0.16	0.14	0.12	0.11	0.14	0.09	0.10	0.06	0.06	0.10	0.10	0.09	0.11	0.10	0.04	0.12	0.12	0.10	0.12	0.09	0.10	0.10	0.08
Guarani	0.12	0.12	0.12	0.10	0.12	0.25	0.10	0.11	0.10	0.11	0.03	0.07	0.07	0.01	0.05	0.04	0.09	0.07	0.06	0.05	0.03	0.06	0.12	0.12		0.09	0.04	0.10	0.04	0.04	0.09	0.07	0.10	0.02	0.09	0.11	0.13	0.02	0.04	0.03	0.05	0.06	0.02	0.04	0.04	0.07	0.02	0.03	0.02	0.11	0.07	0.07	0.05	0.02	0.05	0.03	0.05	0.05	0.05	0.02	0.05	0.11	0.12	0.11	0.11	0.13	0.12	0.11	0.12	0.14	0.11	0.15	0.15	0.12	0.14	0.12	0.12	0.12	0.11
Guaymi	0.18	0.18	0.18	0.16	0.18	0.29	0.16	0.18	0.18	0.15	0.10	0.07	0.07	0.09	0.12	0.11	0.15	0.14	0.10	0.11	0.10	0.11	0.19	0.18	0.09		0.11	0.10	0.12	0.11	0.16	0.14	0.18	0.08	0.16	0.14	0.15	0.08	0.11	0.10	0.12	0.14	0.10	0.11	0.11	0.13	0.09	0.10	0.09	0.18	0.07	0.14	0.13	0.10	0.10	0.08	0.14	0.12	0.12	0.09	0.10	0.17	0.17	0.17	0.17	0.18	0.18	0.17	0.18	0.20	0.17	0.21	0.21	0.18	0.20	0.17	0.18	0.18	0.17
Guihiba	0.14	0.14	0.14	0.12	0.14	0.26	0.12	0.14	0.13	0.14	0.05	0.09	0.10	0.04	0.07	0.07	0.11	0.09	0.08	0.07	0.05	0.08	0.15	0.14	0.04	0.11		0.12	0.07	0.06	0.11	0.10	0.12	0.04	0.11	0.13	0.16	0.04	0.07	0.05	0.08	0.09	0.05	0.05	0.07	0.09	0.04	0.05	0.05	0.13	0.10	0.08	0.07	0.06	0.07	0.05	0.09	0.07	0.07	0.05	0.07	0.13	0.14	0.13	0.13	0.15	0.14	0.13	0.14	0.16	0.13	0.17	0.17	0.14	0.16	0.14	0.14	0.14	0.13
Huetar	0.18	0.18	0.18	0.16	0.18	0.29	0.18	0.18	0.19	0.16	0.10	0.08	0.08	0.10	0.12	0.12	0.15	0.16	0.12	0.12	0.11	0.12	0.19	0.18	0.10	0.10	0.12		0.12	0.12	0.18	0.14	0.19	0.09	0.17	0.15	0.15	0.09	0.11	0.10	0.13	0.15	0.10	0.11	0.12	0.14	0.08	0.10	0.10	0.19	0.08	0.14	0.14	0.11	0.11	0.08	0.14	0.12	0.10	0.09	0.10	0.17	0.18	0.17	0.17	0.18	0.18	0.17	0.18	0.20	0.17	0.21	0.20	0.18	0.21	0.17	0.18	0.18	0.17
Huilliche	0.14	0.14	0.14	0.12	0.14	0.26	0.12	0.13	0.13	0.14	0.05	0.09	0.10	0.05	0.07	0.01	0.11	0.03	0.08	0.07	0.04	0.09	0.14	0.14	0.04	0.12	0.07	0.12		0.07	0.12	0.10	0.13	0.04	0.12	0.13	0.16	0.04	0.06	0.05	0.07	0.09	0.06	0.06	0.06	0.09	0.04	0.05	0.04	0.14	0.10	0.09	0.08	0.05	0.08	0.06	0.09	0.04	0.07	0.04	0.07	0.12	0.13	0.13	0.13	0.14	0.14	0.13	0.14	0.16	0.13	0.17	0.16	0.14	0.16	0.13	0.14	0.14	0.13
Ingano	0.14	0.14	0.14	0.12	0.14	0.26	0.12	0.13	0.13	0.13	0.05	0.09	0.09	0.04	0.07	0.06	0.11	0.09	0.08	0.07	0.05	0.08	0.15	0.14	0.04	0.11	0.06	0.12	0.07		0.11	0.10	0.13	0.04	0.11	0.13	0.15	0.04	0.06	0.05	0.08	0.09	0.05	0.06	0.07	0.09	0.04	0.05	0.04	0.13	0.10	0.08	0.07	0.05	0.07	0.05	0.09	0.07	0.07	0.05	0.07	0.13	0.14	0.13	0.13	0.14	0.14	0.13	0.14	0.16	0.13	0.17	0.17	0.14	0.16	0.13	0.14	0.14	0.13
Jamamadi	0.18	0.18	0.18	0.16	0.18	0.30	0.17	0.18	0.18	0.19	0.10	0.14	0.15	0.10	0.12	0.11	0.16	0.15	0.11	0.11	0.10	0.14	0.19	0.19	0.09	0.16	0.11	0.18	0.12	0.11		0.15	0.18	0.09	0.16	0.18	0.21	0.09	0.12	0.10	0.13	0.14	0.10	0.11	0.11	0.14	0.10	0.10	0.10	0.18	0.15	0.14	0.13	0.11	0.13	0.10	0.14	0.12	0.10	0.10	0.13	0.17	0.18	0.17	0.17	0.19	0.18	0.17	0.18	0.20	0.17	0.21	0.21	0.18	0.20	0.18	0.18	0.18	0.18
Kaingang	0.17	0.17	0.17	0.15	0.17	0.28	0.16	0.17	0.15	0.16	0.08	0.12	0.12	0.07	0.10	0.10	0.13	0.12	0.12	0.10	0.07	0.12	0.17	0.17	0.07	0.14	0.10	0.14	0.10	0.10	0.15		0.16	0.07	0.14	0.16	0.18	0.07	0.10	0.08	0.10	0.12	0.08	0.10	0.10	0.12	0.07	0.08	0.08	0.16	0.13	0.12	0.11	0.08	0.10	0.08	0.11	0.10	0.10	0.08	0.10	0.15	0.16	0.16	0.16	0.17	0.17	0.16	0.17	0.19	0.16	0.19	0.19	0.17	0.19	0.16	0.16	0.17	0.16
Kalina	0.20	0.20	0.20	0.18	0.20	0.32	0.19	0.20	0.18	0.19	0.11	0.16	0.16	0.10	0.14	0.12	0.17	0.16	0.15	0.14	0.11	0.15	0.21	0.21	0.10	0.18	0.12	0.19	0.13	0.13	0.18	0.16		0.10	0.17	0.20	0.22	0.10	0.13	0.11	0.14	0.14	0.10	0.11	0.13	0.15	0.10	0.11	0.11	0.19	0.16	0.15	0.14	0.12	0.14	0.11	0.15	0.14	0.13	0.11	0.13	0.19	0.20	0.19	0.19	0.21	0.20	0.20	0.20	0.22	0.19	0.23	0.23	0.20	0.23	0.20	0.20	0.20	0.20
Kaqchikel	0.12	0.12	0.11	0.10	0.11	0.24	0.10	0.11	0.10	0.11	0.02	0.05	0.06	0.02	0.04	0.04	0.08	0.06	0.05	0.04	0.02	0.06	0.12	0.11	0.02	0.08	0.04	0.09	0.04	0.04	0.09	0.07	0.10		0.09	0.10	0.12	0.01	0.03	0.02	0.05	0.07	0.03	0.04	0.03	0.06	0.01	0.02	0.02	0.11	0.06	0.06	0.06	0.03	0.05	0.03	0.07	0.05	0.04	0.01	0.04	0.10	0.11	0.10	0.10	0.12	0.11	0.11	0.12	0.13	0.10	0.14	0.14	0.12	0.14	0.11	0.11	0.12	0.11
Karitiana	0.18	0.18	0.18	0.16	0.18	0.29	0.16	0.18	0.16	0.18	0.10	0.14	0.14	0.09	0.12	0.11	0.15	0.14	0.13	0.12	0.10	0.13	0.19	0.18	0.09	0.16	0.11	0.17	0.12	0.11	0.16	0.14	0.17	0.09		0.18	0.20	0.09	0.11	0.10	0.12	0.13	0.09	0.10	0.11	0.14	0.09	0.10	0.10	0.16	0.15	0.13	0.12	0.10	0.12	0.10	0.14	0.12	0.11	0.09	0.12	0.17	0.18	0.17	0.17	0.18	0.18	0.17	0.18	0.20	0.17	0.21	0.21	0.18	0.20	0.17	0.18	0.18	0.17
Kogi	0.19	0.19	0.19	0.17	0.19	0.30	0.18	0.19	0.19	0.14	0.11	0.11	0.12	0.11	0.13	0.13	0.17	0.16	0.13	0.13	0.12	0.13	0.20	0.20	0.11	0.14	0.13	0.15	0.13	0.13	0.18	0.16	0.20	0.10	0.18		0.19	0.10	0.13	0.11	0.13	0.16	0.12	0.13	0.13	0.15	0.11	0.12	0.11	0.20	0.12	0.15	0.14	0.12	0.12	0.10	0.15	0.13	0.12	0.11	0.12	0.18	0.19	0.18	0.18	0.19	0.19	0.18	0.19	0.21	0.18	0.22	0.22	0.19	0.21	0.18	0.19	0.19	0.18
Maleku	0.21	0.21	0.21	0.19	0.21	0.32	0.19	0.22	0.22	0.19	0.14	0.12	0.13	0.14	0.16	0.15	0.19	0.18	0.14	0.15	0.14	0.15	0.22	0.22	0.13	0.15	0.16	0.15	0.16	0.15	0.21	0.18	0.22	0.12	0.20	0.19		0.13	0.15	0.14	0.16	0.18	0.15	0.15	0.15	0.17	0.13	0.14	0.14	0.22	0.13	0.18	0.17	0.15	0.14	0.12	0.18	0.16	0.15	0.13	0.15	0.20	0.21	0.20	0.20	0.22	0.21	0.21	0.21	0.23	0.21	0.24	0.24	0.21	0.23	0.20	0.21	0.21	0.21
Maya	0.12	0.12	0.11	0.10	0.12	0.24	0.10	0.11	0.10	0.11	0.03	0.06	0.07	0.02	0.04	0.04	0.08	0.07	0.05	0.04	0.03	0.06	0.12	0.12	0.02	0.08	0.04	0.09	0.04	0.04	0.09	0.07	0.10	0.01	0.09	0.10	0.13		0.03	0.02	0.05	0.07	0.03	0.04	0.04	0.06	0.01	0.03	0.02	0.11	0.07	0.07	0.06	0.03	0.05	0.03	0.07	0.05	0.04	0.01	0.05	0.10	0.11	0.10	0.10	0.12	0.12	0.11	0.12	0.13	0.10	0.14	0.14	0.12	0.14	0.11	0.11	0.12	0.11
Mixe	0.13	0.13	0.13	0.11	0.13	0.25	0.11	0.12	0.13	0.13	0.05	0.09	0.09	0.05	0.06	0.06	0.10	0.09	0.07	0.06	0.05	0.09	0.14	0.13	0.04	0.11	0.07	0.11	0.06	0.06	0.12	0.10	0.13	0.03	0.11	0.13	0.15	0.03		0.03	0.07	0.09	0.06	0.06	0.05	0.08	0.03	0.05	0.05	0.14	0.10	0.09	0.08	0.06	0.08	0.05	0.09	0.07	0.05	0.03	0.07	0.12	0.13	0.12	0.12	0.13	0.13	0.12	0.13	0.15	0.12	0.16	0.15	0.13	0.15	0.12	0.13	0.13	0.12
Mixtec	0.12	0.12	0.12	0.10	0.12	0.24	0.10	0.11	0.11	0.12	0.03	0.07	0.08	0.03	0.05	0.05	0.08	0.07	0.06	0.04	0.03	0.07	0.12	0.12	0.03	0.10	0.05	0.10	0.05	0.05	0.10	0.08	0.11	0.02	0.10	0.11	0.14	0.02	0.03		0.05	0.08	0.04	0.05	0.04	0.06	0.01	0.03	0.03	0.12	0.08	0.08	0.07	0.04	0.06	0.04	0.08	0.06	0.04	0.01	0.06	0.11	0.11	0.10	0.10	0.12	0.12	0.11	0.12	0.14	0.10	0.14	0.14	0.12	0.14	0.11	0.12	0.12	0.11
Ojibwa	0.11	0.11	0.11	0.09	0.11	0.23	0.09	0.07	0.14	0.14	0.06	0.10	0.10	0.06	0.03	0.07	0.06	0.10	0.09	0.01	0.06	0.10	0.12	0.11	0.05	0.12	0.08	0.13	0.07	0.08	0.13	0.10	0.14	0.05	0.12	0.13	0.16	0.05	0.07	0.05		0.10	0.07	0.08	0.05	0.08	0.04	0.06	0.06	0.14	0.10	0.10	0.09	0.07	0.09	0.07	0.10	0.08	0.05	0.05	0.08	0.10	0.10	0.09	0.09	0.11	0.11	0.10	0.11	0.12	0.10	0.13	0.13	0.11	0.13	0.10	0.11	0.11	0.10
Palikur	0.16	0.16	0.16	0.14	0.16	0.28	0.15	0.16	0.15	0.16	0.08	0.11	0.12	0.07	0.10	0.09	0.13	0.12	0.11	0.09	0.07	0.11	0.17	0.17	0.06	0.14	0.09	0.15	0.09	0.09	0.14	0.12	0.14	0.07	0.13	0.16	0.18	0.07	0.09	0.08	0.10		0.07	0.08	0.09	0.12	0.07	0.08	0.08	0.15	0.12	0.11	0.10	0.08	0.10	0.08	0.12	0.10	0.10	0.07	0.10	0.15	0.16	0.15	0.15	0.17	0.16	0.16	0.17	0.18	0.15	0.19	0.19	0.17	0.19	0.16	0.16	0.16	0.16
Parakana	0.14	0.13	0.13	0.12	0.13	0.25	0.11	0.13	0.10	0.13	0.04	0.08	0.09	0.03	0.06	0.05	0.10	0.08	0.07	0.06	0.04	0.08	0.14	0.13	0.02	0.10	0.05	0.10	0.06	0.05	0.10	0.08	0.10	0.03	0.09	0.12	0.15	0.03	0.06	0.04	0.07	0.07		0.05	0.06	0.08	0.03	0.04	0.04	0.11	0.08	0.07	0.07	0.04	0.06	0.04	0.08	0.06	0.06	0.04	0.06	0.12	0.13	0.12	0.12	0.14	0.13	0.13	0.13	0.15	0.12	0.16	0.16	0.13	0.15	0.13	0.13	0.13	0.13
Piapoco	0.14	0.14	0.14	0.12	0.14	0.26	0.12	0.13	0.12	0.13	0.05	0.09	0.09	0.04	0.07	0.06	0.11	0.09	0.08	0.07	0.05	0.08	0.14	0.14	0.04	0.11	0.05	0.11	0.06	0.06	0.11	0.10	0.11	0.04	0.10	0.13	0.15	0.04	0.06	0.05	0.08	0.08	0.05		0.06	0.09	0.04	0.05	0.04	0.13	0.10	0.08	0.07	0.05	0.07	0.05	0.09	0.07	0.06	0.04	0.07	0.13	0.13	0.13	0.13	0.14	0.14	0.13	0.14	0.16	0.13	0.17	0.16	0.14	0.16	0.13	0.14	0.14	0.13
PimaAZ	0.12	0.12	0.12	0.10	0.12	0.24	0.10	0.11	0.13	0.13	0.05	0.09	0.09	0.05	0.05	0.06	0.08	0.09	0.07	0.04	0.05	0.08	0.12	0.12	0.04	0.11	0.07	0.12	0.06	0.07	0.11	0.10	0.13	0.03	0.11	0.13	0.15	0.04	0.05	0.04	0.05	0.09	0.06	0.06		0.05	0.02	0.05	0.04	0.13	0.09	0.09	0.08	0.05	0.07	0.05	0.09	0.07	0.03	0.03	0.07	0.11	0.11	0.11	0.11	0.12	0.12	0.11	0.12	0.14	0.11	0.15	0.14	0.12	0.14	0.11	0.12	0.12	0.11
PimaMX	0.14	0.14	0.14	0.12	0.14	0.26	0.13	0.14	0.15	0.15	0.07	0.11	0.12	0.07	0.07	0.09	0.11	0.11	0.10	0.07	0.07	0.11	0.15	0.14	0.07	0.13	0.09	0.14	0.09	0.09	0.14	0.12	0.15	0.06	0.14	0.15	0.17	0.06	0.08	0.06	0.08	0.12	0.08	0.09	0.05		0.05	0.07	0.07	0.16	0.12	0.12	0.11	0.08	0.10	0.08	0.11	0.09	0.05	0.06	0.10	0.13	0.14	0.13	0.13	0.14	0.14	0.14	0.14	0.16	0.13	0.17	0.17	0.14	0.16	0.13	0.14	0.14	0.13
Purepecha	0.11	0.11	0.10	0.09	0.11	0.23	0.09	0.09	0.11	0.11	0.02	0.06	0.07	0.03	0.03	0.04	0.07	0.07	0.05	0.04	0.02	0.06	0.11	0.10	0.02	0.09	0.04	0.08	0.04	0.04	0.10	0.07	0.10	0.01	0.09	0.11	0.13	0.01	0.03	0.01	0.04	0.07	0.03	0.04	0.02	0.05		0.02	0.02	0.11	0.07	0.07	0.06	0.03	0.05	0.03	0.06	0.05	0.03	0.01	0.05	0.09	0.10	0.09	0.09	0.11	0.11	0.10	0.11	0.13	0.09	0.13	0.13	0.11	0.13	0.10	0.11	0.11	0.10
Quechua1	0.13	0.13	0.13	0.11	0.13	0.25	0.11	0.12	0.12	0.12	0.00	0.08	0.08	0.03	0.06	0.04	0.09	0.07	0.07	0.05	0.02	0.07	0.13	0.13	0.03	0.10	0.05	0.10	0.05	0.05	0.10	0.08	0.11	0.02	0.10	0.12	0.14	0.03	0.05	0.03	0.06	0.08	0.04	0.05	0.05	0.07	0.02		0.00	0.12	0.08	0.08	0.07	0.04	0.06	0.04	0.08	0.05	0.05	0.03	0.06	0.11	0.12	0.11	0.11	0.13	0.13	0.12	0.13	0.14	0.11	0.15	0.15	0.13	0.15	0.12	0.13	0.13	0.12
Quechua2	0.13	0.13	0.12	0.10	0.12	0.25	0.10	0.11	0.11	0.12	0.01	0.07	0.08	0.03	0.05	0.04	0.09	0.07	0.06	0.05	0.02	0.07	0.13	0.12	0.02	0.09	0.05	0.10	0.04	0.04	0.10	0.08	0.11	0.02	0.10	0.11	0.14	0.02	0.05	0.03	0.06	0.08	0.04	0.04	0.04	0.07	0.02	0.00		0.12	0.08	0.07	0.06	0.04	0.06	0.04	0.07	0.05	0.05	0.02	0.05	0.11	0.12	0.11	0.11	0.13	0.12	0.12	0.13	0.14	0.11	0.15	0.15	0.12	0.15	0.12	0.12	0.12	0.12
Surui	0.20	0.20	0.20	0.18	0.20	0.31	0.19	0.20	0.18	0.20	0.12	0.16	0.17	0.11	0.14	0.13	0.18	0.16	0.15	0.13	0.12	0.16	0.21	0.21	0.11	0.18	0.13	0.19	0.14	0.13	0.18	0.16	0.19	0.11	0.16	0.20	0.22	0.11	0.14	0.12	0.14	0.15	0.11	0.13	0.13	0.16	0.11	0.12	0.12		0.17	0.16	0.15	0.12	0.14	0.12	0.16	0.14	0.14	0.11	0.14	0.19	0.20	0.19	0.19	0.20	0.20	0.19	0.20	0.22	0.19	0.23	0.23	0.20	0.22	0.19	0.20	0.20	0.19
Teribe	0.17	0.17	0.16	0.15	0.17	0.28	0.14	0.16	0.16	0.13	0.08	0.04	0.05	0.08	0.10	0.10	0.14	0.12	0.09	0.10	0.08	0.09	0.17	0.17	0.07	0.07	0.10	0.08	0.10	0.10	0.15	0.13	0.16	0.06	0.15	0.12	0.13	0.07	0.10	0.08	0.10	0.12	0.08	0.10	0.09	0.12	0.07	0.08	0.08	0.17		0.12	0.11	0.09	0.08	0.06	0.12	0.11	0.09	0.08	0.08	0.15	0.16	0.15	0.15	0.17	0.17	0.16	0.17	0.18	0.15	0.19	0.19	0.17	0.19	0.16	0.16	0.17	0.16
Ticuna1	0.16	0.16	0.16	0.14	0.16	0.28	0.15	0.16	0.15	0.16	0.08	0.11	0.12	0.07	0.10	0.09	0.13	0.12	0.11	0.09	0.08	0.11	0.17	0.16	0.07	0.14	0.08	0.14	0.09	0.08	0.14	0.12	0.15	0.06	0.13	0.15	0.18	0.07	0.09	0.08	0.10	0.11	0.07	0.08	0.09	0.12	0.07	0.08	0.07	0.16	0.12		0.03	0.08	0.10	0.08	0.12	0.10	0.10	0.07	0.10	0.15	0.16	0.15	0.15	0.17	0.16	0.15	0.16	0.18	0.15	0.19	0.19	0.16	0.18	0.16	0.16	0.16	0.16
Ticuna2	0.16	0.15	0.15	0.13	0.15	0.27	0.14	0.15	0.14	0.15	0.07	0.10	0.11	0.06	0.09	0.08	0.12	0.11	0.10	0.08	0.07	0.10	0.16	0.15	0.05	0.13	0.07	0.14	0.08	0.07	0.13	0.11	0.14	0.06	0.12	0.14	0.17	0.06	0.08	0.07	0.09	0.10	0.07	0.07	0.08	0.11	0.06	0.07	0.06	0.15	0.11	0.03		0.07	0.09	0.07	0.11	0.09	0.08	0.06	0.09	0.14	0.15	0.14	0.14	0.16	0.15	0.15	0.16	0.17	0.14	0.18	0.18	0.15	0.18	0.15	0.15	0.16	0.15
Toba	0.14	0.13	0.13	0.11	0.13	0.25	0.11	0.12	0.12	0.13	0.04	0.08	0.09	0.02	0.06	0.05	0.10	0.08	0.07	0.06	0.04	0.08	0.14	0.13	0.02	0.10	0.06	0.11	0.05	0.05	0.11	0.08	0.12	0.03	0.10	0.12	0.15	0.03	0.06	0.04	0.07	0.08	0.04	0.05	0.05	0.08	0.03	0.04	0.04	0.12	0.09	0.08	0.07		0.07	0.05	0.03	0.06	0.05	0.04	0.06	0.12	0.13	0.12	0.12	0.14	0.13	0.13	0.13	0.15	0.12	0.16	0.16	0.13	0.15	0.13	0.13	0.13	0.12
Waunana	0.15	0.15	0.15	0.13	0.15	0.27	0.13	0.14	0.14	0.13	0.06	0.07	0.08	0.06	0.08	0.07	0.12	0.10	0.08	0.08	0.06	0.06	0.15	0.15	0.05	0.10	0.07	0.11	0.08	0.07	0.13	0.10	0.14	0.05	0.12	0.12	0.14	0.05	0.08	0.06	0.09	0.10	0.06	0.07	0.07	0.10	0.05	0.06	0.06	0.14	0.08	0.10	0.09	0.07		0.05	0.10	0.08	0.08	0.05	0.07	0.14	0.14	0.14	0.14	0.15	0.15	0.14	0.15	0.17	0.14	0.17	0.17	0.15	0.17	0.14	0.15	0.15	0.14
Wayuu	0.13	0.13	0.13	0.11	0.13	0.25	0.11	0.13	0.12	0.11	0.04	0.05	0.06	0.04	0.06	0.05	0.10	0.08	0.06	0.06	0.04	0.06	0.13	0.13	0.03	0.08	0.05	0.08	0.06	0.05	0.10	0.08	0.11	0.03	0.10	0.10	0.12	0.03	0.05	0.04	0.07	0.08	0.04	0.05	0.05	0.08	0.03	0.04	0.04	0.12	0.06	0.08	0.07	0.05	0.05		0.08	0.06	0.06	0.03	0.05	0.12	0.13	0.12	0.12	0.13	0.13	0.12	0.13	0.15	0.12	0.16	0.16	0.13	0.15	0.12	0.13	0.13	0.12
Wichi	0.16	0.16	0.16	0.14	0.16	0.28	0.14	0.16	0.15	0.16	0.08	0.11	0.12	0.05	0.10	0.09	0.13	0.11	0.11	0.09	0.07	0.11	0.17	0.16	0.05	0.14	0.09	0.14	0.09	0.09	0.14	0.11	0.15	0.07	0.14	0.15	0.18	0.07	0.09	0.08	0.10	0.12	0.08	0.09	0.09	0.11	0.06	0.08	0.07	0.16	0.12	0.12	0.11	0.03	0.10	0.08		0.10	0.09	0.07	0.10	0.15	0.16	0.15	0.15	0.16	0.16	0.15	0.16	0.18	0.15	0.19	0.19	0.16	0.18	0.15	0.16	0.16	0.15
Yaghan	0.14	0.14	0.14	0.12	0.14	0.26	0.12	0.14	0.14	0.15	0.05	0.10	0.10	0.05	0.08	0.03	0.11	0.06	0.09	0.07	0.05	0.09	0.15	0.14	0.05	0.12	0.07	0.12	0.04	0.07	0.12	0.10	0.14	0.05	0.12	0.13	0.16	0.05	0.07	0.06	0.08	0.10	0.06	0.07	0.07	0.09	0.05	0.05	0.05	0.14	0.11	0.10	0.09	0.06	0.08	0.06	0.10		0.07	0.05	0.08	0.13	0.14	0.13	0.13	0.14	0.14	0.14	0.14	0.16	0.13	0.17	0.17	0.14	0.16	0.13	0.14	0.14	0.13
Yaqui	0.13	0.12	0.12	0.10	0.12	0.25	0.11	0.11	0.14	0.14	0.05	0.09	0.10	0.04	0.05	0.06	0.09	0.09	0.07	0.05	0.05	0.08	0.13	0.12	0.05	0.12	0.07	0.10	0.07	0.07	0.10	0.10	0.13	0.04	0.11	0.12	0.15	0.04	0.05	0.04	0.05	0.10	0.06	0.06	0.03	0.05	0.03	0.05	0.05	0.14	0.09	0.10	0.08	0.05	0.08	0.06	0.09	0.07		0.03	0.08	0.11	0.12	0.11	0.11	0.13	0.12	0.11	0.13	0.14	0.11	0.15	0.15	0.12	0.15	0.11	0.12	0.12	0.11
Zapotec	0.12	0.11	0.11	0.09	0.11	0.24	0.10	0.10	0.11	0.11	0.03	0.07	0.07	0.03	0.04	0.04	0.08	0.07	0.05	0.04	0.03	0.07	0.12	0.11	0.02	0.09	0.05	0.09	0.04	0.05	0.10	0.08	0.11	0.01	0.09	0.11	0.13	0.01	0.03	0.01	0.05	0.07	0.04	0.04	0.03	0.06	0.01	0.03	0.02	0.11	0.08	0.07	0.06	0.04	0.05	0.03	0.07	0.05	0.03		0.05	0.10	0.11	0.10	0.10	0.12	0.11	0.11	0.11	0.13	0.10	0.14	0.14	0.11	0.14	0.11	0.11	0.11	0.10
Zenu	0.15	0.15	0.14	0.13	0.14	0.26	0.12	0.14	0.14	0.13	0.06	0.07	0.08	0.05	0.08	0.07	0.11	0.10	0.08	0.07	0.06	0.08	0.15	0.14	0.05	0.10	0.07	0.10	0.07	0.07	0.13	0.10	0.13	0.04	0.12	0.12	0.15	0.05	0.07	0.06	0.08	0.10	0.06	0.07	0.07	0.10	0.05	0.06	0.05	0.14	0.08	0.10	0.09	0.06	0.07	0.05	0.10	0.08	0.08	0.05		0.13	0.14	0.13	0.13	0.15	0.14	0.14	0.15	0.16	0.13	0.17	0.17	0.14	0.17	0.14	0.14	0.15	0.14
Altaian	0.03	0.03	0.02	0.01	0.02	0.19	0.07	0.14	0.18	0.18	0.11	0.15	0.15	0.11	0.10	0.12	0.10	0.15	0.14	0.08	0.11	0.14	0.10	0.09	0.11	0.17	0.13	0.17	0.12	0.13	0.17	0.15	0.19	0.10	0.17	0.18	0.20	0.10	0.12	0.11	0.10	0.15	0.12	0.13	0.11	0.13	0.09	0.11	0.11	0.19	0.15	0.15	0.14	0.12	0.14	0.12	0.15	0.13	0.11	0.10	0.13		0.01	0.05	0.05	0.02	0.02	0.03	0.04	0.07	0.07	0.05	0.04	0.04	0.05	0.01	0.02	0.02	0.02
Buryat	0.04	0.03	0.02	0.01	0.03	0.19	0.07	0.15	0.19	0.19	0.12	0.15	0.16	0.12	0.10	0.13	0.11	0.16	0.15	0.09	0.12	0.15	0.10	0.10	0.12	0.17	0.14	0.18	0.13	0.14	0.18	0.16	0.20	0.11	0.18	0.19	0.21	0.11	0.13	0.11	0.10	0.16	0.13	0.13	0.11	0.14	0.10	0.12	0.12	0.20	0.16	0.16	0.15	0.13	0.14	0.13	0.16	0.14	0.12	0.11	0.14	0.01		0.05	0.05	0.02	0.02	0.03	0.05	0.07	0.08	0.05	0.05	0.04	0.06	0.01	0.02	0.02	0.02
Chukchi1	0.08	0.07	0.07	0.05	0.07	0.21	0.06	0.14	0.19	0.19	0.11	0.15	0.15	0.11	0.10	0.12	0.10	0.15	0.14	0.08	0.11	0.15	0.07	0.06	0.11	0.17	0.13	0.17	0.13	0.13	0.17	0.16	0.19	0.10	0.17	0.18	0.20	0.10	0.12	0.10	0.09	0.15	0.12	0.13	0.11	0.13	0.09	0.11	0.11	0.19	0.15	0.15	0.14	0.12	0.14	0.12	0.15	0.13	0.11	0.10	0.13	0.05	0.05		0.00	0.06	0.05	0.06	0.07	0.04	0.04	0.08	0.07	0.06	0.08	0.05	0.06	0.06	0.03
Chukchi2	0.07	0.07	0.06	0.05	0.07	0.21	0.06	0.14	0.19	0.18	0.11	0.15	0.15	0.11	0.09	0.12	0.10	0.15	0.14	0.08	0.11	0.14	0.07	0.06	0.11	0.17	0.13	0.17	0.13	0.13	0.17	0.16	0.19	0.10	0.17	0.18	0.20	0.10	0.12	0.10	0.09	0.15	0.12	0.13	0.11	0.13	0.09	0.11	0.11	0.19	0.15	0.15	0.14	0.12	0.14	0.12	0.15	0.13	0.11	0.10	0.13	0.05	0.05	0.00		0.06	0.05	0.06	0.07	0.04	0.04	0.08	0.07	0.06	0.08	0.05	0.06	0.06	0.03
Dolgan	0.05	0.04	0.04	0.02	0.04	0.20	0.08	0.15	0.20	0.20	0.13	0.16	0.17	0.13	0.11	0.14	0.12	0.16	0.15	0.10	0.13	0.16	0.11	0.10	0.13	0.18	0.15	0.18	0.14	0.14	0.19	0.17	0.21	0.12	0.18	0.19	0.22	0.12	0.13	0.12	0.11	0.17	0.14	0.14	0.12	0.14	0.11	0.13	0.13	0.20	0.17	0.17	0.16	0.14	0.15	0.13	0.16	0.14	0.13	0.12	0.15	0.02	0.02	0.06	0.06		0.01	0.03	0.05	0.08	0.08	0.04	0.04	0.04	0.05	0.02	0.01	0.01	0.02
Evenki	0.05	0.04	0.03	0.02	0.04	0.20	0.08	0.15	0.20	0.19	0.13	0.16	0.16	0.12	0.11	0.14	0.12	0.16	0.15	0.09	0.13	0.16	0.10	0.10	0.12	0.18	0.14	0.18	0.14	0.14	0.18	0.17	0.20	0.11	0.18	0.19	0.21	0.12	0.13	0.12	0.11	0.16	0.13	0.14	0.12	0.14	0.11	0.13	0.12	0.20	0.17	0.16	0.15	0.13	0.15	0.13	0.16	0.14	0.12	0.11	0.14	0.02	0.02	0.05	0.05	0.01		0.03	0.05	0.07	0.08	0.04	0.04	0.04	0.05	0.02	0.00	0.00	0.02
Ket	0.06	0.06	0.05	0.03	0.05	0.19	0.07	0.14	0.19	0.19	0.12	0.15	0.16	0.12	0.10	0.13	0.11	0.15	0.15	0.09	0.12	0.15	0.10	0.09	0.11	0.17	0.13	0.17	0.13	0.13	0.17	0.16	0.20	0.11	0.17	0.18	0.21	0.11	0.12	0.11	0.10	0.16	0.13	0.13	0.11	0.14	0.10	0.12	0.12	0.19	0.16	0.15	0.15	0.13	0.14	0.12	0.15	0.14	0.11	0.11	0.14	0.03	0.03	0.06	0.06	0.03	0.03		0.03	0.08	0.08	0.05	0.04	0.01	0.05	0.03	0.03	0.04	0.03
Khanty	0.07	0.07	0.06	0.04	0.06	0.19	0.08	0.15	0.20	0.19	0.13	0.16	0.17	0.13	0.11	0.14	0.12	0.16	0.15	0.10	0.13	0.16	0.11	0.11	0.12	0.18	0.14	0.18	0.14	0.14	0.18	0.17	0.20	0.12	0.18	0.19	0.21	0.12	0.13	0.12	0.11	0.17	0.13	0.14	0.12	0.14	0.11	0.13	0.13	0.20	0.17	0.16	0.16	0.13	0.15	0.13	0.16	0.14	0.13	0.11	0.15	0.04	0.05	0.07	0.07	0.05	0.05	0.03		0.09	0.09	0.06	0.06	0.03	0.05	0.04	0.05	0.05	0.05
Koryak	0.09	0.09	0.08	0.07	0.09	0.23	0.09	0.17	0.22	0.22	0.14	0.18	0.18	0.14	0.13	0.15	0.13	0.18	0.17	0.11	0.14	0.18	0.11	0.10	0.14	0.20	0.16	0.20	0.16	0.16	0.20	0.19	0.22	0.13	0.20	0.21	0.23	0.13	0.15	0.14	0.12	0.18	0.15	0.16	0.14	0.16	0.13	0.14	0.14	0.22	0.18	0.18	0.17	0.15	0.17	0.15	0.18	0.16	0.14	0.13	0.16	0.07	0.07	0.04	0.04	0.08	0.07	0.08	0.09		0.08	0.09	0.09	0.08	0.10	0.07	0.08	0.08	0.05
Naukan	0.09	0.09	0.09	0.07	0.09	0.23	0.06	0.14	0.19	0.19	0.11	0.15	0.15	0.11	0.10	0.12	0.10	0.15	0.14	0.09	0.11	0.15	0.05	0.04	0.11	0.17	0.13	0.17	0.13	0.13	0.17	0.16	0.19	0.10	0.17	0.18	0.21	0.10	0.12	0.10	0.10	0.15	0.12	0.13	0.11	0.13	0.09	0.11	0.11	0.19	0.15	0.15	0.14	0.12	0.14	0.12	0.15	0.13	0.11	0.10	0.13	0.07	0.08	0.04	0.04	0.08	0.08	0.08	0.09	0.08		0.10	0.10	0.08	0.10	0.07	0.08	0.08	0.06
Nganasan	0.08	0.08	0.07	0.05	0.07	0.22	0.10	0.18	0.22	0.22	0.15	0.18	0.19	0.15	0.13	0.16	0.14	0.19	0.18	0.12	0.15	0.18	0.13	0.12	0.15	0.21	0.17	0.21	0.17	0.17	0.21	0.19	0.23	0.14	0.21	0.22	0.24	0.14	0.16	0.14	0.13	0.19	0.16	0.17	0.15	0.17	0.13	0.15	0.15	0.23	0.19	0.19	0.18	0.16	0.17	0.16	0.19	0.17	0.15	0.14	0.17	0.05	0.05	0.08	0.08	0.04	0.04	0.05	0.06	0.09	0.10		0.01	0.05	0.05	0.04	0.05	0.05	0.04
Nganasan	0.08	0.07	0.07	0.05	0.07	0.22	0.10	0.18	0.22	0.22	0.15	0.18	0.19	0.15	0.13	0.16	0.14	0.19	0.17	0.12	0.15	0.18	0.13	0.12	0.15	0.21	0.17	0.20	0.16	0.17	0.21	0.19	0.23	0.14	0.21	0.22	0.24	0.14	0.15	0.14	0.13	0.19	0.16	0.16	0.14	0.17	0.13	0.15	0.15	0.23	0.19	0.19	0.18	0.16	0.17	0.16	0.19	0.17	0.15	0.14	0.17	0.04	0.05	0.07	0.07	0.04	0.04	0.04	0.06	0.09	0.10	0.01		0.05	0.05	0.04	0.04	0.04	0.04
Selkup	0.07	0.06	0.06	0.04	0.06	0.19	0.08	0.15	0.20	0.19	0.13	0.16	0.16	0.13	0.11	0.14	0.12	0.16	0.15	0.10	0.13	0.16	0.11	0.10	0.12	0.18	0.14	0.18	0.14	0.14	0.18	0.17	0.20	0.12	0.18	0.19	0.21	0.12	0.13	0.12	0.11	0.17	0.13	0.14	0.12	0.14	0.11	0.13	0.12	0.20	0.17	0.16	0.15	0.13	0.15	0.13	0.16	0.14	0.12	0.11	0.14	0.04	0.04	0.06	0.06	0.04	0.04	0.01	0.03	0.08	0.08	0.05	0.05		0.05	0.04	0.04	0.04	0.04
Tundra_Ne	0.08	0.08	0.08	0.05	0.08	0.21	0.10	0.17	0.22	0.22	0.15	0.18	0.19	0.15	0.13	0.16	0.14	0.18	0.17	0.12	0.15	0.18	0.13	0.12	0.14	0.20	0.16	0.21	0.16	0.16	0.20	0.19	0.23	0.14	0.20	0.21	0.23	0.14	0.15	0.14	0.13	0.19	0.15	0.16	0.14	0.16	0.13	0.15	0.15	0.22	0.19	0.18	0.18	0.15	0.17	0.15	0.18	0.16	0.15	0.14	0.17	0.05	0.06	0.08	0.08	0.05	0.05	0.05	0.05	0.10	0.10	0.05	0.05	0.05		0.05	0.06	0.06	0.05
Tuvinians	0.04	0.03	0.03	0.01	0.03	0.19	0.07	0.14	0.19	0.19	0.12	0.15	0.16	0.12	0.10	0.13	0.11	0.15	0.14	0.09	0.12	0.15	0.10	0.09	0.12	0.17	0.14	0.17	0.13	0.13	0.18	0.16	0.20	0.11	0.17	0.18	0.20	0.11	0.12	0.11	0.10	0.16	0.13	0.13	0.11	0.13	0.10	0.12	0.12	0.19	0.16	0.16	0.15	0.13	0.14	0.12	0.15	0.13	0.11	0.11	0.14	0.01	0.01	0.05	0.05	0.02	0.02	0.03	0.04	0.07	0.07	0.04	0.04	0.04	0.05		0.02	0.02	0.02
Yakut1	0.05	0.04	0.03	0.02	0.03	0.19	0.08	0.15	0.20	0.19	0.13	0.16	0.16	0.12	0.11	0.13	0.12	0.16	0.15	0.10	0.12	0.16	0.11	0.10	0.12	0.18	0.14	0.18	0.14	0.14	0.18	0.16	0.20	0.11	0.18	0.19	0.21	0.11	0.13	0.12	0.11	0.16	0.13	0.14	0.12	0.14	0.11	0.13	0.12	0.20	0.16	0.16	0.15	0.13	0.15	0.13	0.16	0.14	0.12	0.11	0.14	0.02	0.02	0.06	0.06	0.01	0.00	0.03	0.05	0.08	0.08	0.05	0.04	0.04	0.06	0.02		0.00	0.02
Yakut2	0.05	0.04	0.03	0.02	0.04	0.20	0.08	0.15	0.20	0.20	0.13	0.16	0.16	0.13	0.11	0.14	0.12	0.16	0.15	0.10	0.13	0.16	0.11	0.10	0.12	0.18	0.14	0.18	0.14	0.14	0.18	0.17	0.20	0.12	0.18	0.19	0.21	0.12	0.13	0.12	0.11	0.16	0.13	0.14	0.12	0.14	0.11	0.13	0.12	0.20	0.17	0.16	0.16	0.13	0.15	0.13	0.16	0.14	0.12	0.11	0.15	0.02	0.02	0.06	0.06	0.01	0.00	0.04	0.05	0.08	0.08	0.05	0.04	0.04	0.06	0.02	0.00		0.03
Yukaghir	0.05	0.05	0.04	0.02	0.04	0.20	0.07	0.14	0.19	0.19	0.12	0.15	0.16	0.12	0.10	0.13	0.11	0.15	0.14	0.09	0.12	0.15	0.09	0.08	0.11	0.17	0.13	0.17	0.13	0.13	0.18	0.16	0.20	0.11	0.17	0.18	0.21	0.11	0.12	0.11	0.10	0.16	0.13	0.13	0.11	0.13	0.10	0.12	0.12	0.19	0.16	0.16	0.15	0.12	0.14	0.12	0.15	0.13	0.11	0.10	0.14	0.02	0.02	0.03	0.03	0.02	0.02	0.03	0.05	0.05	0.06	0.04	0.04	0.04	0.05	0.02	0.02	0.03	

Table S4. Estimates of bottleneck dates based on decay of allele sharing

Population	N	Language group	Subtracted background LD	Generations
GreenlandInuit2	8	Eskimo-Aleutian	GreenlandInuit1-GreenlandInuit2	45
GreenlandInuit1	7	Eskimo-Aleutian	GreenlandInuit1-GreenlandInuit2	poor fit to exponential
Ojibwa	5	Northern Amerind	Algonquin-Ojibwa	260
Chipewyan	5	Na-Dene	Algonquin-Chipewyan	14
Cheyenne	24	Northern Amerind	Algonquin-Cheyenne	42
Algonquin	5	Northern Amerind	Algonquin-Cree	5
Cree	4	Northern Amerind	Algonquin-Cree	no visible decay
Mixe	17	Northern Amerind	Zapotec-Mixe	12
Maya	18	Northern Amerind	Maya-Kaqchikel	no visible decay
Kaqchikel	13	Northern Amerind	Maya-Kaqchikel	no visible decay
PimaAZ	22	Central-Amerind	PimaMX-PimaAZ	13
PimaMX	33	Central-Amerind	PimaMX-PimaAZ	13
Zapotec	23	Central-Amerind	Zapotec-Mixtec	no visible decay
Mixtec	5	Central-Amerind	Zapotec-Mixtec	28
Cabecar	31	Chibchan-Paezan	Cabecar:Bribri	8
Guaymi	5	Chibchan-Paezan	Guaymi-Bribri	25
Zenu	5	Chibchan-Paezan	Kogi-Zenu	15
Bribri	4	Chibchan-Paezan	Kogi-Bribri	28
Kogi	4	Chibchan-Paezan	Kogi-Arhuaco	26
Embera	5	Chibchan-Paezan	Embera-Waunana	7
Arhuaco	5	Chibchan-Paezan	Kogi-Arhuaco	poor fit to exponential
Wayuu	12	Equatorial-Tucanoan	Kogi-Wayuu	15
Piapoco	7	Equatorial-Tucanoan	Piapoco-Guahibo	23
Ticuna1	6	Equatorial-Tucanoan	Ticuna1-Ticuna2	8
Ticuna2	12	Equatorial-Tucanoan	Ticuna1-Ticuna2	no visible decay
Guahibo	6	Equatorial-Tucanoan	Piapoco-Guahibo	25
Surui	24	Equatorial-Tucanoan	Karitiana-Surui	7
Karitiana	13	Equatorial-Tucanoan	Karitiana-Surui	9
Guarani	6	Equatorial-Tucanoan	Wichi-Guarani	no visible decay
Wichi	5	Ge-Pano-Carib	Wichi-Toba	19
Toba	4	Ge-Pano-Carib	Wichi-Toba	no visible decay
Diaguita	5	Andean	Diaguita-Quechua2	no visible decay
Quechua1	18	Andean	Quechua1-Aymara	no visible decay
Quechua2	22	Andean	Quechua1-Quechua2	no visible decay
Aymara	23	Andean	Aymara-Quechua1	no visible decay
Inga	10	Andean	Inga-Ticuna2	26
Chono	4	Andean	Huilliche-Chono	no visible decay
Huilliche	4	Andean	Huilliche-Chono	no visible decay
Chilote	8	Andean	Yaghan-Chilote	no visible decay
Yaghan	4	Andean	Yaghan-Chilote	no visible decay

Table S5. Z-scores from *4 Population Tests* of the tree ((Outgroup1,Outgroup2), (NatAm1, NatAm2))

	NatAm1=Zapotec										NatAm1=Quechua2									
Outgroup1	CHB	CHB	CHB	CHB	Nganasan2	Nganasan2	Nganasan2	Naukan	Naukan	Chukchi2	CHB	CHB	CHB	CHB	Nganasan2	Nganasan2	Nganasan2	Naukan	Naukan	Chukchi2
Outgroup2	Nganasan2	Naukan	Chukchi2	Koryak	Naukan	Chukchi2	Koryak	Chukchi2	Koryak	Koryak	Nganasan2	Naukan	Chukchi2	Koryak	Naukan	Chukchi2	Koryak	Chukchi2	Koryak	Koryak
NatAm2																				
Algonquin	1.5	-1.9	0.3	1.4	-3.1	-1.1	0.0	3.0	3.5	1.5	0.7	-2.6	0.0	0.8	-3.1	-0.7	0.1	3.7	3.7	1.1
Arara	-1.7	-1.5	-0.8	0.8	-0.2	0.8	2.1	1.2	2.4	1.9	-2.3	-2.2	-0.9	0.3	-0.3	1.1	2.1	1.9	2.6	1.5
Arhuaco	-0.6	-0.1	-0.2	-0.2	0.3	0.4	0.4	0.0	0.0	0.0	-1.4	-0.9	-0.4	-0.6	0.2	0.9	0.5	0.8	0.3	-0.4
Aymara	2.1	1.2	1.3	1.7	-0.3	-0.4	0.0	0.0	0.4	0.6	0.9	-0.1	1.2	1.1	-0.8	0.3	0.3	1.6	1.3	0.0
Bribri	-1.4	0.5	-0.9	0.2	1.5	0.5	1.3	-1.7	-0.3	1.2	-2.1	-0.3	-1.1	-0.4	1.3	0.9	1.5	-0.9	0.0	0.9
Cabecar	-1.0	1.5	0.0	-0.1	2.3	0.9	0.7	-2.2	-1.7	-0.1	-2.0	0.6	-0.3	-0.7	2.1	1.4	0.9	-1.3	-1.4	-0.5
Chane	2.3	1.8	1.6	1.9	-0.2	-0.7	-0.1	-0.6	0.1	0.7	1.6	1.1	1.3	1.4	-0.3	-0.3	0.0	0.1	0.3	0.4
Cheyenne	2.1	-3.3	-2.2	-0.1	-4.7	-4.1	-1.8	1.9	3.3	2.6	0.8	-3.7	-2.3	-0.8	-4.2	-2.9	-1.4	2.6	3.5	1.8
Chilote	1.0	1.5	1.2	1.0	0.7	0.3	0.1	-0.6	-0.7	-0.2	0.1	0.8	0.9	0.3	0.7	0.8	0.2	0.0	-0.5	-0.7
Chipewyan	5.1	4.0	5.8	5.3	-0.1	0.6	0.8	0.8	0.9	0.3	4.2	3.1	5.2	4.4	-0.2	0.9	0.8	1.4	1.0	0.0
Chono	1.0	1.6	1.7	1.0	0.7	0.6	0.1	-0.3	-0.7	-0.6	0.5	1.1	1.5	0.7	0.7	1.0	0.3	0.3	-0.5	-0.9
Chorotega	-0.3	-1.0	0.1	0.4	-0.7	0.4	0.6	1.5	1.4	0.4	-0.6	-1.4	-0.1	0.0	-0.9	0.5	0.5	1.8	1.4	0.1
Cree	0.1	-1.2	-0.9	1.7	-1.2	-0.9	1.5	0.6	2.7	3.2	-0.5	-1.6	-0.9	1.5	-1.1	-0.3	1.8	1.2	3.1	2.9
Diaguita	0.0	2.0	1.9	0.4	2.0	1.9	0.4	-0.6	-1.8	-1.8	-0.8	1.1	1.7	-0.1	1.8	2.5	0.6	0.4	-1.3	-2.1
Embera	-0.1	0.9	0.0	0.5	1.0	0.1	0.6	-1.3	-0.4	0.7	-1.0	0.0	-0.3	-0.1	0.8	0.6	0.7	-0.4	-0.1	0.3
Guarani	1.6	1.8	1.6	1.5	0.5	0.1	0.1	-0.6	-0.5	0.0	0.5	0.8	1.2	0.7	0.3	0.7	0.3	0.4	-0.1	-0.5
Guaymi	-0.1	0.8	-0.3	-1.2	0.8	-0.2	-1.0	-1.5	-2.1	-1.2	-0.9	0.0	-0.6	-1.7	0.7	0.2	-0.9	-0.7	-1.7	-1.6
Guahibo	1.4	2.3	0.8	1.5	1.2	-0.4	0.4	-2.3	-1.0	1.0	0.4	1.4	0.4	0.8	1.0	0.1	0.5	-1.4	-0.6	0.5
Huetar	1.0	0.6	0.9	1.1	-0.3	-0.2	0.1	0.2	0.5	0.4	0.7	0.7	1.2	1.0	0.1	0.4	0.3	0.4	0.3	-0.1
Huilliche	0.9	1.0	0.3	-0.8	0.3	-0.5	-1.5	-1.0	-1.9	-1.5	0.1	0.2	0.1	-1.4	0.1	0.0	-1.4	-0.2	-1.5	-1.9
Inga	-1.0	1.1	0.3	-0.3	1.8	1.2	0.6	-1.3	-1.6	-0.8	-2.1	0.2	-0.1	-1.0	1.8	1.8	0.7	-0.4	-1.2	-1.2
Jamamadi	0.3	-0.7	0.1	-0.7	-0.9	-0.3	-0.9	1.0	0.1	-0.9	-0.2	-1.3	-0.2	-1.1	-1.0	0.0	-0.8	1.5	0.3	-1.2
Kaingang	-0.1	-0.7	-1.8	-1.7	-0.6	-1.5	-1.5	-1.1	-1.0	-0.3	-0.6	-1.2	-1.8	-2.1	-0.6	-1.0	-1.5	-0.4	-0.9	-0.8
Kalina	1.8	1.3	0.7	-0.2	0.1	-0.8	-1.8	-1.0	-1.7	-1.3	1.0	0.3	0.0	-1.0	-0.4	-0.9	-1.8	-0.4	-1.3	-1.3
Kaqchikel	-0.4	0.5	-1.4	-1.4	0.7	-0.9	-1.0	-2.3	-1.9	-0.2	-1.6	-0.8	-1.6	-2.0	0.4	-0.2	-0.7	-0.9	-1.3	-0.7
Karitiana	-0.6	-1.0	-0.7	-0.3	-0.5	-0.1	0.2	0.6	0.8	0.5	-1.4	-1.9	-1.0	-0.9	-0.6	0.4	0.4	1.4	1.1	0.1
Kogi	-0.3	0.2	-1.1	-1.0	0.4	-0.8	-0.7	-1.6	-1.1	0.0	-1.0	-0.5	-1.4	-1.4	0.3	-0.4	-0.6	-0.9	-0.9	-0.3
Maleku	-0.8	-2.1	-1.5	-0.8	-1.5	-0.7	-0.1	1.2	1.5	0.8	-1.5	-2.6	-1.6	-1.1	-1.6	-0.3	0.0	1.9	1.7	0.5
Maya	-0.3	-0.1	-1.6	-1.0	0.2	-1.2	-0.7	-1.8	-0.9	0.6	-1.6	-1.2	-1.8	-1.7	-0.1	-0.4	-0.4	-0.5	-0.4	0.0
Mixe	1.4	-0.9	-0.5	-1.1	-1.8	-1.8	-2.2	0.6	-0.2	-0.9	0.0	-1.8	-0.9	-1.7	-1.7	-0.8	-1.6	1.6	0.3	-1.2
Mixtec	0.5	1.0	0.3	1.0	0.5	-0.2	0.5	-1.0	0.0	1.0	-0.5	-0.1	-0.1	0.2	0.3	0.4	0.6	0.0	0.3	0.4
Ojibwa	2.2	-1.0	1.2	3.2	-2.6	-0.8	1.2	2.8	4.0	2.7	1.3	-1.7	0.8	2.4	-2.7	-0.5	1.2	3.4	4.3	2.2
Palikur	-1.0	0.9	0.4	-0.2	1.5	1.2	0.6	-0.7	-1.1	-0.8	-1.7	0.1	0.1	-0.7	1.4	1.6	0.7	0.0	-0.9	-1.1
Parakana	-0.5	1.0	1.5	0.9	1.3	1.9	1.3	0.4	-0.2	-0.7	-1.1	0.5	1.3	0.4	1.3	2.3	1.4	0.9	0.0	-1.0
Piapoco	-0.7	-0.2	-0.7	-1.5	0.3	0.0	-0.8	-0.5	-1.3	-1.1	-1.6	-1.2	-1.1	-2.2	0.2	0.5	-0.7	0.4	-0.8	-1.5
PimaAZ	2.6	-0.2	1.0	2.0	-2.0	-1.3	-0.2	1.5	2.1	1.4	1.0	-1.2	0.4	0.9	-2.0	-0.5	0.0	2.5	2.3	0.6
PimaMX	0.8	0.6	0.4	2.4	-0.1	-0.4	1.6	-0.3	1.6	2.5	-0.3	-0.4	0.0	1.3	-0.2	0.2	1.5	0.6	1.8	1.8
Purepecha	0.0	0.5	0.1	-0.8	0.5	0.0	-0.8	-0.7	-1.4	-1.1	-0.3	0.4	0.1	-0.9	0.6	0.4	-0.6	-0.5	-1.4	-1.2
Quechua1	1.6	0.9	0.9	1.0	-0.3	-0.5	-0.3	-0.2	0.1	0.3	0.4	-0.4	0.5	0.2	-0.8	0.2	-0.1	1.3	0.7	-0.4
Quechua2	1.3	1.3	0.5	0.9	0.2	-0.7	-0.2	-1.2	-0.5	0.6										
Surui	-0.8	0.9	-0.7	-1.1	1.4	0.0	-0.4	-2.2	-1.9	-0.5	-1.6	0.0	-1.1	-1.8	1.3	0.4	-0.3	-1.4	-1.7	-0.9
Teribe	1.7	1.3	1.0	0.8	0.0	-0.4	-0.5	-0.5	-0.5	-0.1	0.8	0.5	0.7	0.3	-0.1	0.0	-0.4	0.2	-0.2	-0.5
Ticuna1	-0.5	0.8	-1.5	-1.6	1.2	-1.0	-1.1	-2.9	-2.4	-0.3	-1.4	0.0	-1.8	-2.2	1.1	-0.5	-0.9	-2.3	-2.2	-0.7
Ticuna2	-1.3	-0.3	-1.7	-0.4	0.7	-0.4	0.8	-1.6	-0.1	1.5	-2.5	-1.2	-2.1	-1.1	0.6	0.1	0.9	-0.7	0.3	1.0
Toba	1.8	1.8	0.7	0.1	0.3	-1.2	-1.6	-1.9	-1.9	-0.7	1.0	0.9	0.3	-0.6	0.1	-0.6	-1.3	-1.0	-1.6	-1.0
Waunana	0.7	1.6	1.0	-0.1	1.0	0.3	-0.7	-1.1	-1.8	-1.3	-0.1	0.7	0.6	-0.7	0.8	0.6	-0.6	-0.4	-1.5	-1.6
Wayuu	-0.1	2.1	1.5	1.9	2.0	1.5	1.9	-1.0	-0.3	0.5	-1.2	0.9	1.1	1.0	1.8	2.0	2.0	0.1	0.1	0.0
Wichi	2.0	1.4	0.1	-0.5	-0.2	-1.9	-2.2	-2.0	-2.0	-0.7	1.2	0.6	-0.3	-1.1	-0.4	-1.4	-2.0	-1.1	-1.6	-1.0
Yaghan	1.4	0.6	0.7	0.3	-0.5	-0.7	-0.9	-0.1	-0.4	-0.5	0.7	-0.1	0.4	-0.2	-0.6	-0.3	-0.8	0.6	-0.1	-0.8
Yaqui	1.3	-1.0	-0.7	0.2	-2.1	-2.0	-0.9	0.6	1.2	1.2	0.7	-1.0	-1.0	-0.4	-1.5	-1.6	-1.0	0.2	0.6	0.6
Zapotec											-1.3	-1.3	-0.5	-0.9	-0.2	0.7	0.2	1.2	0.5	-0.6
Zenu	0.1	1.1	-0.5	0.0	0.9	-0.5	-0.1	-2.0	-1.1	0.5	-0.8	0.2	-0.8	-0.6	0.8	0.0	0.1	-1.2	-0.8	0.1

Note: We compute an f_4-statistic whose expected value is zero if the two Native American populations form a clade relative to the Outgroups, as well as a standard error from a Block Jackknife. We present the Z-score (standard errors from zero), rather than the f_4-statistic itself, to help in interpreting significance. Values of |Z|>3 are highlighted.

Table S6. f_4 statistics from *4 Population Tests* of the tree ((Zapotec, NNA), (Outgroup1, Outgroup2))

f_4(Zapotec, Chipewyan; Column outgroup, Row outgroup) x 1000

	CHB	Khanty	Tuvinians	Buryat	Aleutian	Altaian	Yakut1	Evenki	GreenlandInuit2	Dolgan	Yakut2	Selkup	Tundra_Nentsi	Nganasan1	Ket	Naukan	Nganasan2	Yukaghir	Chukchi2	GreenlandInuit1	Koryak	Chukchi1
CHB		-3	-4	-5	-5	-5	-6	-6	-6	-7	-7	-7	-7	-8	-8	-8	-8	-9	-9	-10	-10	-11
Khanty	3		0	-1	-3	-2	-3	-3	-3	-3	-3	-3	-4	-4	-4	-5	-5	-6	-6	-6	-6	-7
Tuvinians	4	0		-1	-1	-2	-2	-3	-3	-3	-3	-3	-4	-4	-4	-4	-5	-6	-6	-6	-6	-7
Buryat	5	1	1		1	-1	-1	-2	-2	-2	-2	-2	-3	-3	-3	-3	-4	-4	-5	-5	-5	-6
Aleutian	5	3	1	-1		3	-1	-1	3	-2	-2	0	-2	-3	-4	0	-3	-4	-3	0	-2	-6
Altaian	5	2	2	1	-3		-1	-1	-1	-1	-1	-1	-2	-2	-2	-3	-3	-4	-4	-5	-4	-6
Yakut1	6	3	2	1	1	1		0	-1	-1	-1	-1	-2	-2	-1	-2	-3	-3	-3	-4	-4	-5
Evenki	6	3	3	2	1	1	0		0	0	0	-1	-1	-2	-1	-2	-2	-3	-3	-4	-4	-5
GreenlandInui	6	3	3	2	-3	1	1	0		0	0	0	-1	-1	0	-2	-2	-3	-3	-4	-3	-5
Dolgan	7	3	3	2	2	1	1	0	0		0	0	-1	-1	-2	-2	-2	-3	-3	-3	-3	-4
Yakut2	7	3	3	2	2	1	1	0	0	0		0	-1	-1	-1	-2	-2	-3	-3	-3	-3	-4
Selkup	7	3	3	2	0	1	1	1	0	0	0		-1	-1	0	-1	-2	-2	-3	-3	-3	-4
Tundra_Nentsi	7	4	4	3	2	2	2	1	1	1	1	1		0	0	-1	-1	-2	-2	-3	-2	-3
Nganasan1	8	4	4	3	3	2	2	2	1	1	1	1	0		1	-1	-1	-2	-2	-2	-2	-3
Ket	8	4	4	3	4	2	1	1	0	2	1	0	0	-1		-2	-2	-1	-2	-3	-2	-4
Naukan	8	5	4	3	0	3	2	2	2	2	2	1	1	1	2		0	-1	-1	-2	-2	-3
Nganasan2	8	5	5	4	3	3	3	2	2	2	2	2	1	1	2	0		-1	-1	-2	-2	-3
Yukaghir	9	6	6	4	4	4	3	3	3	3	3	2	2	2	1	1	1		0	-1	-1	-2
Chukchi2	9	6	6	5	3	4	3	3	3	3	3	3	2	2	2	1	1	0		-1	-1	-2
GreenlandInui	10	6	6	5	0	5	4	4	4	3	3	3	3	2	3	2	2	1	1		0	-1
Koryak	10	6	6	5	2	4	4	4	3	3	3	3	2	2	2	2	2	1	1	0		-1
Chukchi1	11	7	7	6	6	6	5	5	5	4	4	4	3	3	4	3	3	2	2	1	1	

f_4(Zapotec, Cheyenne; Column outgroup, Row outgroup) x 1000

	CHB	Khanty	Tuvinians	Buryat	Aleutian	Altaian	Yakut1	Evenki	GreenlandInuit2	Dolgan	Yakut2	Selkup	Tundra_Nentsi	Nganasan1	Ket	Naukan	Nganasan2	Yukaghir	Chukchi2	GreenlandInuit1	Koryak	Chukchi1
CHB		0	-1	-1	7	0	-1	-2	4	-3	-2	-2	-2	-2	0	4	-2	-1	2	4	0	2
Khanty	0		-1	-1	8	0	-1	-2	4	-3	-1	-1	-1	-2	0	4	-2	-1	2	4	0	2
Tuvinians	1	1		-1	8	1	-1	-1	5	-2	-1	-1	-1	-1	0	5	-1	0	3	4	1	2
Buryat	1	1	1		9	1	0	0	6	-1	0	0	0	0	1	5	-1	0	4	5	2	3
Aleutian	-7	-8	-8	-9		-7	-9	-10	-3	-10	-9	-9	-10	-10	-7	-3	-9	-9	-6	-5	-7	-8
Altaian	0	0	-1	-1	7		-1	-2	4	-3	-1	-1	-1	-2	0	4	-2	-1	3	4	0	2
Yakut1	1	1	1	0	9	1		-1	6	-2	0	0	0	0	1	5	-1	0	4	5	1	3
Evenki	2	2	1	0	10	2	1		6	-1	0	0	0	0	2	6	0	1	4	5	2	3
GreenlandInui	-4	-4	-5	-6	3	-4	-6	-6		-7	-6	-6	-6	-6	-5	0	-6	-5	-2	-1	-4	-3
Dolgan	3	3	2	1	10	3	2	1	7		1	1	1	1	2	7	1	2	5	6	3	4
Yakut2	2	1	1	0	9	1	0	0	6	-1		0	0	0	1	6	-1	1	4	5	2	3
Selkup	2	1	1	0	9	1	0	0	6	-1	0		0	0	1	6	-1	1	4	5	2	3
Tundra_Nentsi	2	1	1	0	10	1	0	0	6	-1	0	0		0	1	5	-1	0	4	5	2	3
Nganasan1	2	2	1	0	10	2	0	0	6	-1	0	0	0		2	6	-1	1	4	5	2	3
Ket	0	0	0	-1	7	0	-1	-2	5	-2	-1	-1	-1	-2		4	-2	-1	3	4	1	2
Naukan	-4	-4	-5	-5	3	-4	-5	-6	0	-7	-6	-6	-5	-6	-4		-6	-5	-2	-1	-4	-3
Nganasan2	2	2	1	1	9	2	1	0	6	-1	1	1	1	1	2	6		1	4	6	2	4
Yukaghir	1	1	0	0	9	1	0	-1	5	-2	-1	-1	0	-1	1	5	-1		4	5	1	3
Chukchi2	-2	-2	-3	-4	6	-3	-4	-4	2	-5	-4	-4	-4	-4	-3	2	-4	-4		1	-2	-1
GreenlandInui	-4	-4	-4	-5	5	-4	-5	-5	1	-6	-5	-5	-5	-5	-4	1	-6	-5	-1		-4	-2
Koryak	0	0	-1	-2	7	0	-1	-2	4	-3	-2	-2	-2	-2	-1	4	-2	-1	2	4		2
Chukchi1	-2	-2	-2	-3	8	-2	-3	-3	3	-4	-3	-3	-3	-3	-2	3	-4	-3	1	2	-2	

Note: We compute an f_4 statistic measuring the affinity of the tested Northern North American population to one outgroup more than another, and present its value x1000 (here we are presenting f_4 statistics because they have a more quantitative interpretation, rather than Z-statistics as in Table S5). Values >0.004 = 4/1000 are highlighted. The patterns for the Algonquin, Cree, Cheyenne and Ojibwa are highly correlated (Figure S6), so only results for Cheyenne are shown. Populations are ordered by their f_4 statistic relative to CHB in the upper table (comparison to Chipewyan), to aid in visualization.

Table S7. Record of admixture dating analyses

Admixed Population	N	Surrogate ancestral population 1	N	Surrogate ancestral population 2	N	Dataset	Generations ± 1 std. err.	95% confidence interval in years*
Maya	28	French	31	Mixe	17	merge5. unmasked	7.4 ± 0.7	180-250
Cheyenne	24	Cree, Ojibwa, Zapotec, PimaAZ, Quechua1, Quechua2	94	Naukan, GreenlandInuit1, GreenlandInuit2	31	merge6. masked	182 ± 80	1500-9,100
Chipewyan	5	Cree, Ojibwa, Zapotec, PimaAZ, Quechua1, Quechua2, Cheyenne	118	Naukan, GreenlandInuit1, GreenlandInuit2, Chukchi1, Chukchi2, Yukaghir, Koryak	84	merge6. masked	no visible decay	no visible decay
Inga	10	Ticuna1, Ticuna2, Guahibo, Piapoco	31	Quechua1, Quechua2, Diaguita, Aymara	68	merge5. masked	82 ± 95	0-6,900
Guarani	6	Ticuna1, Ticuna2, Guahibo, Piapoco, Inga	41	Wichi, Toba, Chane, Kaingang	13	merge5. masked	39 ± 45	0-3,300
Guahibo	6	Ticuna1, Ticuna2, Piapoco, Inga	35	Quechua1, Quechua2, Diaguita, Zapotec	68	merge5. masked	no visible decay	no visible decay
Kogi	4	Maya, Zapotec, PimaAZ, Cheyenne	87	Maleku, Huetar, Guaymi, Teribe, Cabecar, Bribri, Zenu, Waunana, Embera	65	merge5. masked	140 ± 41	2,100-6,000
Arhuaco	5	Maya, Zapotec, PimaAZ, Cheyenne	87	Maleku, Huetar, Guaymi, Teribe, Cabecar, Bribri, Zenu, Waunana, Embera	64	merge5. masked	no visible decay	no visible decay
Arhuaco + Kogi	9	Maya, Zapotec, PimaAZ, Cheyenne	87	Maleku, Huetar, Guaymi, Teribe, Cabecar, Bribri, Zenu, Waunana, Embera	60	merge5. masked	158 ± 38	2,800-6,400
Cabecar	31	Maya, Zapotec, PimaAZ, Cheyenne	87	Maleku, Huetar, Guaymi, Teribe, Kogi, Arhuaco, Bribri, Zenu, Waunana, Embera	38	merge5. masked	241 ± 41	5,000-8,900
Guaymi	5	Maya, Zapotec, PimaAZ, Cheyenne	87	Maleku, Huetar, Cabecar, Teribe, Kogi, Arhuaco, Bribri, Zenu, Waunana, Embera	64	merge5. masked	147 ± 49	1,900-6,600
Bribri	4	Maya, Zapotec, PimaAZ, Cheyenne	87	Maleku, Huetar, Guaymi, Teribe, Kogi, Arhuaco, Cabecar, Zenu, Waunana, Embera	65	merge5. masked	184 ± 130	0-11,500
Zenu	5	Maya, Zapotec, PimaAZ, Cheyenne	87	Maleku, Huetar, Guaymi, Teribe, Kogi, Arhuaco, Bribri, Cabecar Waunana, Embera	64	merge5. masked	272 ± 87	3,700-12,000

Note: For the ancestral populations, we are guided by the structure of Figure 1C. We are sometimes using populations that we know are admixed for the ancestral populations, but simulations in Moorjani et al. 2011 suggests that ROLLOFF performs well in this case (what is important is only that the allele frequency differences between the true ancestral populations are correlated to the allele frequency differences between the surrogate ancestral populations).

* The 95% confidence interval is determined by taking the estimate plus or minus 1.645 standard errors, and multiplying by an assumed 29 years per generation.