Report

Current Biology

Y Chromosome Sequences Reveal a Short Beringian Standstill, Rapid Expansion, and early Population structure of Native American Founders

Highlights

- We sequenced 20 Native American Y chromosomes chosen for their genetic diversity
- A Beringian Standstill of <4,600 years led to both Siberian and American Y-lineages
- Y-lineage split times rule out occupation of the Americas before 19,500 years ago
- Present-day male population structure in South America arose before 12,000 years ago

Authors

Thomaz Pinotti, Anders Bergström, Maria Geppert, ..., Lutz Roewer, Fabrício R. Santos, Chris Tyler-Smith

Correspondence

fsantos@icb.ufmg.br (F.R.S.), cts@sanger.ac.uk (C.T.-S.)

In Brief

Pinotti et al. provide a genetic analysis of male history in the Americas that reveals three or four founding lineages, an occupation of Beringia for no longer than 4,600 years, entry south of the ice sheets after 19,500 years ago, and the establishment of the present-day male population structure in South America by 12,000 years ago.

Pinotti et al., 2019, Current Biology *29*, 149–157
January 7, 2019
https://doi.org/10.1016/j.cub.2018.11.029

CellPress

Current Biology
Report

Y Chromosome Sequences Reveal a Short Beringian Standstill, Rapid Expansion, and early Population structure of Native American Founders

Thomaz Pinotti,[1,2] Anders Bergström,[2] Maria Geppert,[3] Matt Bawn,[4,5] Dominique Ohasi,[1] Wentao Shi,[2,6] Daniela R. Lacerda,[1] Arne Solli,[7,8] Jakob Norstedt,[7] Kate Reed,[7] Kim Dawtry,[7] Fabricio González-Andrade,[9] Cesar Paz-y-Miño,[10] Susana Revollo,[11] Cinthia Cuellar,[11] Marilza S. Jota,[1] José E. Santos, Jr.,[1] Qasim Ayub,[2,12,13] Toomas Kivisild,[14,15,16] José R. Sandoval,[4] Ricardo Fujita,[4] Yali Xue,[2] Lutz Roewer,[3] Fabrício R. Santos,[1,*] and Chris Tyler-Smith[2,17,*]

[1]Laboratório de Biodiversidade e Evolução Molecular (LBEM), Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Av. Antonio Carlos 6627, 31270-010 Belo Horizonte, Brazil
[2]The Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK
[3]Institute of Legal Medicine and Forensic Sciences, Department of Forensic Genetics, Charité—Universitätsmedizin Berlin, Berlin, Germany
[4]Centro de Genética y Biología Molecular (CGBM), Instituto de Investigación, Facultad de Medicina Humana, Universidad de San Martin de Porres, 15009 Lima, Peru
[5]The Earlham Institute, NR4 7UG Norwich, UK
[6]Department of Genetics, School of Basic Medical Sciences, Tianjin Medical University, 300070 Tianjin, China
[7]Q Nordic Independent Researchers
[8]Department of Archaeology, History, Cultural Studies and Religion (AHKR), University of Bergen, Norway
[9]Translational Medicine Unit, Central University of Ecuador, Faculty of Medical Sciences, Iquique N14-121 y Sodiro-Itchimbía, Sector El Dorado, 170403 Quito, Ecuador
[10]Universidad de las Americas, Av. de los Granados E12-41, 170513 Quito, Ecuador
[11]Universidad Mayor de San Andrés, Av. Villazón 1995, 2008 La Paz, Bolivia
[12]Monash University Malaysia Genomics Facility, Tropical Medicine and Biology Multidisciplinary Platform, 47500 Bandar Sunway, Selangor Darul Ehsan, Malaysia
[13]School of Science, Monash University Malaysia, 47500 Bandar Sunway, Selangor Darul Ehsan, Malaysia
[14]Department of Archaeology and Anthropology, University of Cambridge, CB2 1QH Cambridge, UK
[15]Estonian Biocentre, 51010 Tartu, Estonia
[16]Present address: Department of Human Genetics, O&N I Herestraat 49, Box 604, 3000 Leuven, Belgium
[17]Lead Contact
*Correspondence: fsantos@icb.ufmg.br (F.R.S.), cts@sanger.ac.uk (C.T.-S.)
https://doi.org/10.1016/j.cub.2018.11.029

SUMMARY

The Americas were the last inhabitable continents to be occupied by humans, with a growing multidisciplinary consensus for entry 15–25 thousand years ago (kya) from northeast Asia via the former Beringia land bridge [1–4]. Autosomal DNA analyses have dated the separation of Native American ancestors from the Asian gene pool to 23 kya or later [5, 6] and mtDNA analyses to ~25 kya [7], followed by isolation ("Beringian Standstill" [8, 9]) for 2.4–9 ky and then a rapid expansion throughout the Americas. Here, we present a calibrated sequence-based analysis of 222 Native American and relevant Eurasian Y chromosomes (24 new) from haplogroups Q and C [10], with four major conclusions. First, we identify three to four independent lineages as autochthonous and likely founders: the major Q-M3 and rarer Q-CTS1780 present throughout the Americas, the very rare C3-MPB373 in South America, and possibly the C3-P39/Z30536 in North America. Second, from the divergence times and Eurasian/American distribution of lineages, we estimate a Beringian Standstill duration of 2.7 ky or 4.6 ky, according to alternative models, and entry south of the ice sheet after 19.5 kya. Third, we describe the star-like expansion of Q-M848 (within Q-M3) starting at 15 kya [11] in the Americas, followed by establishment of substantial spatial structure in South America by 12 kya. Fourth, the deep branches of the Q-CTS1780 lineage present at low frequencies throughout the Americas today [12] may reflect a separate out-of-Beringia dispersal after the melting of the glaciers at the end of the Pleistocene.

RESULTS AND DISCUSSION

Y Chromosome Sequencing of Modern South Americans and Comparative Data

We generated new genomic sequences for 20 Native South American Y chromosomes from across the continent: 19 from haplogroup Q, chosen non-randomly to cover the breadth of known Y-STR diversity, and one from haplogroup C3. We compared these to 65 published Native

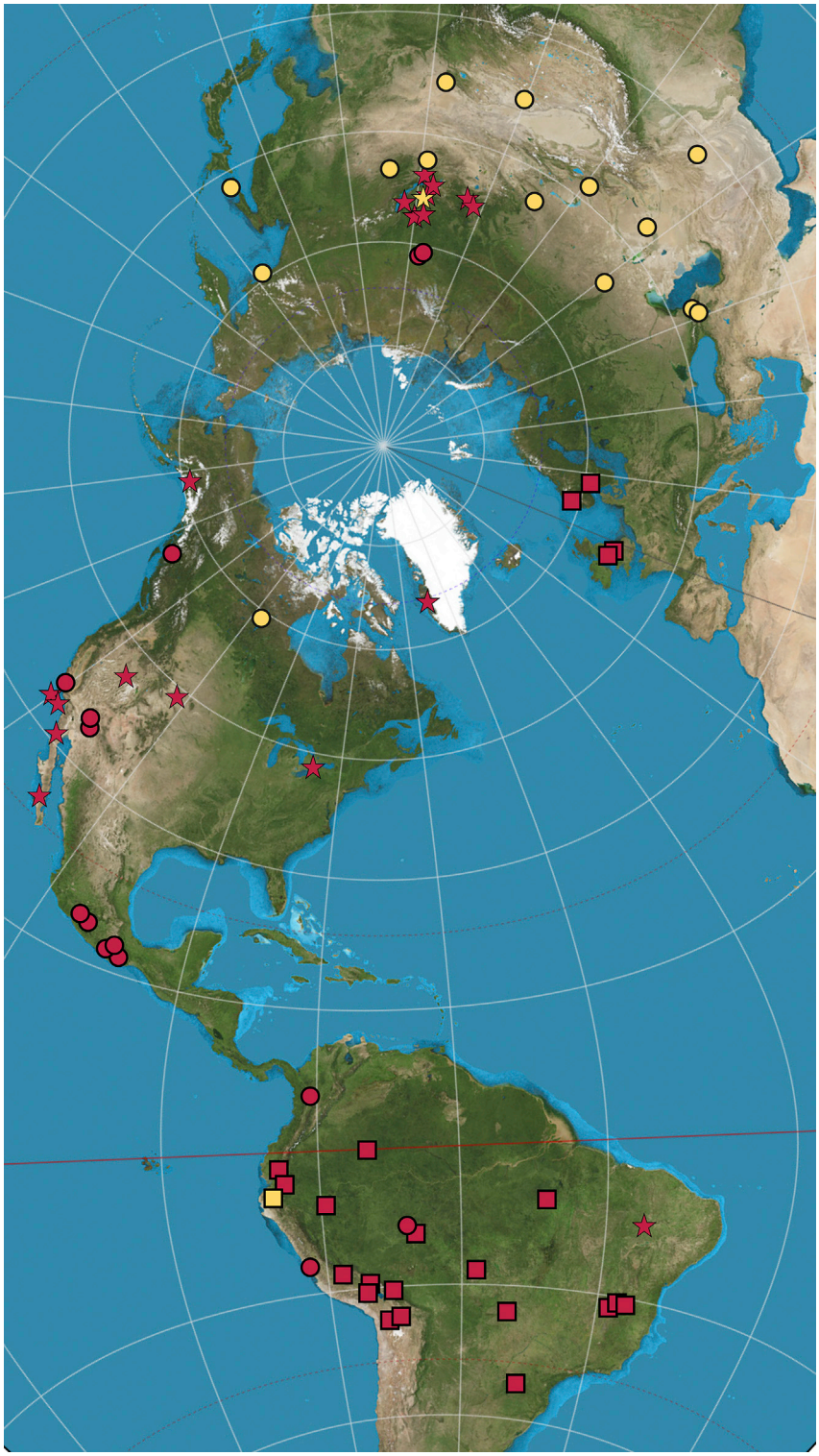

Figure 1. Geographical Distribution of Samples

Location of relevant Eurasian and American samples used in this study. Samples in red represent haplogroup Q-L54 lineages; samples in yellow represent haplogroup C3-L1373 lineages. Squares denote modern sequences published here for the first time, circles are modern sequences from the literature, and stars are ancient sequences from the literature. Symbols are not proportional to sample size. For full listing, information, and references regarding the samples, see Data S1. The approximate sample locations were plotted by the authors on an edited map in Peirce Quincuncial Projection made by Daniel R. Strebe, who is not involved in this work, and whom we thank.

American Y chromosome sequences (including 16 from ancient samples) and 137 from relevant worldwide samples, including four new sequences from Northern European individuals who carry Q-L804 Y chromosomes—a very rare sister haplogroup to Native American Q-M3—and 82 from the literature, resulting in a combined dataset of 222 Y chromosomes (Figure 1 and Data S1).

We analyzed SNPs from the accessible regions of the Y chromosome [13] that are covered in all 184 modern sequences, resulting in 13,221 high-confidence SNPs from 7.6 Mb of sequence (STAR Methods). Despite detectable platform biases that influence the fine-scale definition of branches, we were able to reconstruct the full calibrated phylogenetic structure described previously [11]. The new data greatly increase the resolution of the relevant branches of the haplogroups C3 and Q (Figures 2 and S1). We present here 452 novel informative SNPs in these two haplogroups, identifying 14 previously undescribed monophyletic sub-haplogroups (nine downstream of Q-M848) and providing further support for 19 other lineages. New SNPs were named MPB001–MPB452 and have been submitted to ISOGG (Data S2). Strikingly, only 2 of the 20 new Native American samples could be assigned to previously described sub-haplogroups, highlighting how understudied the genetic diversity in the South American continent has been.

A Calibrated Phylogeny of Y Chromosome Haplogroup C3

Haplogroup C is one of the early diverging non-African branches of the Y chromosome phylogeny [11], a feature that likely explains its very wide geographical distribution throughout modern peoples of Australia [14, 15], Asia [16–18], and the Americas [10, 19], as well as ancient Europeans [20–23]. All Native American C chromosomes examined thus far fall within haplogroup C3 (defined by marker M217) [10, 19]. We found this haplogroup to carry deep divergences estimated to have begun around 38 thousand years ago (kya) (95% CI: 34.7–44.5

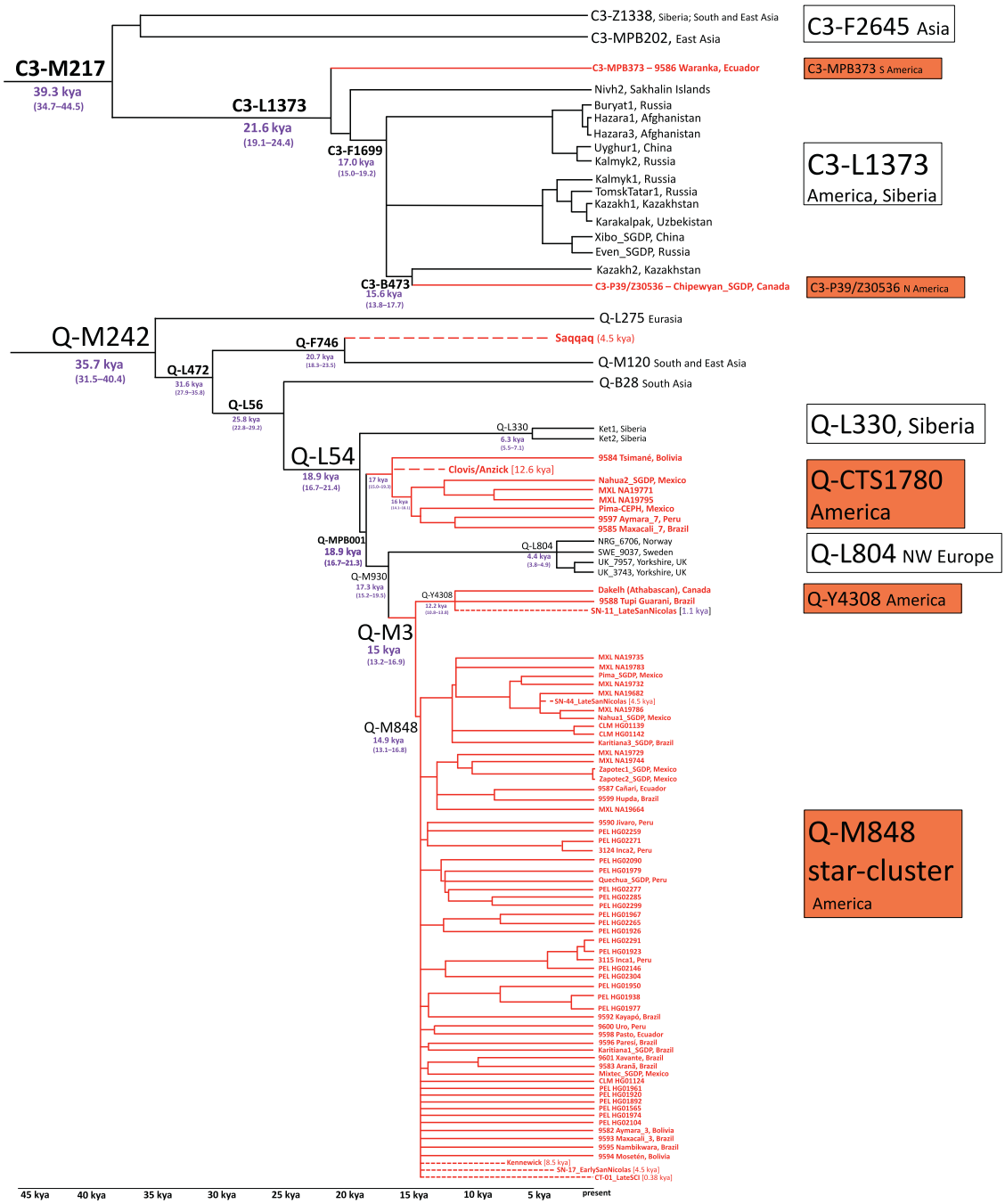

Figure 2. Calibrated Phylogeny of Y Haplogroups C3-M217 and Q-M242

Time-calibrated phylogeny showing the main lineages discussed in the text. Branches in red are American, and hashed lines are used in the final branch of ancient samples.

For full list of samples and references, see Data S1. For a phylogeny calibrated by SNP number, see Figure S1.

kya), clustering all American and some Siberian samples into one branch (a clade we name C3-L1373), while all other Asian C samples belong to a separate branch of C3 [16] (Figures 2 and S1).

Only 12 published individuals (~5%) within our American/Siberian dataset, plus the new Ecuadorian C3* reported here, fall within C3-L1373. The South American C3*, henceforth named C3-MBP373, is the first branch to diverge within C3-L1373 from the other, mainly Siberian clade C3-F3914, at 21.6 (19.1–24.4) kya (Figure 2). The Athabaskan C3-P39/Z30536 individual from Canada clusters inside the C3-F3914 clade downstream to C3-B473, where he shares a most-recent common ancestor with an individual from Kazakhstan at 15.6 (13.8–17.7) kya. Although in theory, this split time is still fully compatible with C3-P39/Z30536 status as initial founding

lineage (and an entry around ~15 kya), its very northern geographical distribution in the Americas, exclusively in populations known to harbor extra Siberian ancestry [5], makes it possibly a result of later Holocene gene flow instead.

Since only a single South American C3-MBP373 chromosome was sequenced, we wished to determine whether or not it was representative of C3* lineages previously reported from Ecuador [19]. However, sequencing of these additional samples was not possible because of limited sample quantity and quality, so we genotyped 18 of the newly discovered Y-SNPs specific to the sequenced C3-MBP373 individual in 21 C3* Native Americans from Ecuador (Kichwas and Waoranis) using PCR and Sanger sequencing of whole-genome amplified DNA. 11 variants, including MPB373, were shared by all these C3* chromosomes, two by all except one individual, and five were found to be private to the sequenced chromosome (Data S3). These results show that the sequenced C3-MBP373 chromosome is indeed representative of the major, if not the only, clade of South American C3 chromosomes (Figure S2) and confirm its previously proposed status as a founder lineage [24]. Since there was similarly only a single C3-P39/Z30536 chromosome available [25], we sought to compare it with previously genotyped samples to determine if it could be used as a proxy for the North American C3 lineage. The ISOGG phylogeny identifies two branches within C3-P39/Z30536, defined by the SNPs BY1360 and Z38874 (https://isogg.org). The sequenced chromosome is derived for one of these, Z38874, and for ISOGG markers downstream of Z38874, demonstrating that the sequenced C3-P39/Z30536 chromosome clusters within the known diversity of this North American lineage.

A Calibrated Phylogeny of Y Chromosome Haplogroup Q

Haplogroup Q branches are currently found in many indigenous peoples of Asia and the Americas [11]. Most Native American Y chromosomes belong to haplogroup Q and lie within the branch Q-L54, which coalesces at 18.9 (16.7–21.4) kya but also contains Northern Eurasian chromosomes (Figures 2 and S1). We therefore focused on the structure of this part of the phylogeny and identified a previously unknown G → A transition (MPB001) shared by Native Americans (both Q-M3 and Q-CTS1780) and a rare branch of Northern European Y chromosomes (Q-L804), but not by Siberian Kets. Downstream of Q-MPB001, the Q-CTS1780 branch (coalescence at 17.0 [15.0–19.3] kya) is exclusively found throughout the Americas, including in the 12.7–12.9-kya Clovis sample Anzick 1 [26]. The other branch (MRCA 17.3 [15.2–19.5] kya) contains the vast majority of Native American Y chromosomes (Q-M3, MRCA 15.0 [13.2–16.9] kya) and the Northern European Q-L804 lineage. Within the Q-M3 lineage, there is an enormous star-like expansion (Q-M848) identified previously and here estimated to start at 14.9 (13.1–16.8) kya, leading to 24 branches (Figures 2 and S1), one of which is found in the 8.3–9.2-kya Kennewick sample [27].

We note that the ancient Saqqaq Palaeo-Eskimo Y chromosome from a 4-kya Greenland individual belonging to the Arctic Small Tool archaeological tradition lies on a different branch of the haplogroup Q phylogeny (Figure 2), which represents a much later (Holocene) entry into the Americas [28] not considered further here.

Founding Lineages and the Beringian Standstill

Some previous mtDNA studies [7, 8] have assumed that the beginning of the Beringian Standstill and divergence of initial founding lineages in Beringia took place only after all American autochthonous lineages split from non-American lineages. Therefore, this hypothesis—a subset of Beringian Standstill models that we call here the Beringian-American hypothesis—assumes that the original Beringian diversity gave rise exclusively to Native Americans, but not to Siberians or Northern Europeans. Following this scenario, we can infer the maximum date of the start of the Beringian Standstill from our time-calibrated phylogeny using the latest time American lineages split from non-American ones. While divergence times should be not taken as population split times, because we cannot rule out the possibility that a lineage existed long before it entered the Americas (as for the Saqqaq Palaeo-Eskimo lineage [28, 29]), relevant lineage split dates in the context of the geographical distribution of the clades can serve as upper or lower brackets of dispersal events under appropriate assumptions. The genomic evidence from the oldest Native American remains demonstrates the early presence of the Q-CTS1780 and Q-M3 lineages in the American continent as far back as 12.6 and 10.3 kya, respectively [26, 30]. We estimate the split date of the Native American Q-M3 lineage from the Northern European Q-L804 at 17.3 (15.2–19.5) kya. Therefore, we can infer that Beringian isolation may not have started before 19.5 kya under the assumptions of the Beringian-American hypothesis. On this basis, we identify two clear founding lineages in Beringia [8]: Q-CTS1780 and Q-M3. We note that a similar conclusion could be drawn using haplogroup C lineages: a split after 17.4 kya (Figure 2). However, as we have only one chromosome from each autochthonous American lineage (North and South) and no ancient DNA evidence, their status as initial founding lineages is less clear, although their geographical locations and other genetic evidence [24] support founder status, particularly for C3-MPB373.

Following the logic of the mtDNA studies further, we can also infer the length of the Beringian Standstill using the time at which the America-specific lineages began to expand as a measure of its end. The early diversification (before 12 kya) within Q-CTS1780 could have occurred in Beringia, but the very striking expansion of the Q-M848 lineage beginning at 14.9 (13.1–16.8) kya within Q-M3 provides a compelling end date for the Beringian Standstill, when part of this population dispersed southward, perhaps via the ice-free Pacific coast [31, 32]. Taking into account the uncertainty in the phylogenetic dates, the duration of the Beringian Standstill can be estimated at a maximum 2.7 (19.5 minus 16.8) ky—considerably shorter than originally proposed [8, 9] (Figure 3).

However, another possible scenario can also be considered where the ancient Beringian diversity gave rise to populations in both the Americas and Asia that we call the out-of-Beringia hypothesis. In this scenario, ancient Beringians were ancestors of both the early genetic pool of the American continent, and some Asian populations, represented by Y lineages found currently in Northern Eurasia. This scenario implies an earlier diversification in Beringia, at around 18.9 (16.7–21.4) kya (the coalescence time of the American and northern Siberian and European Q lineages), and a Beringian origin for the Eurasian lineages Q-L330 and Q-L804, as previously argued for some

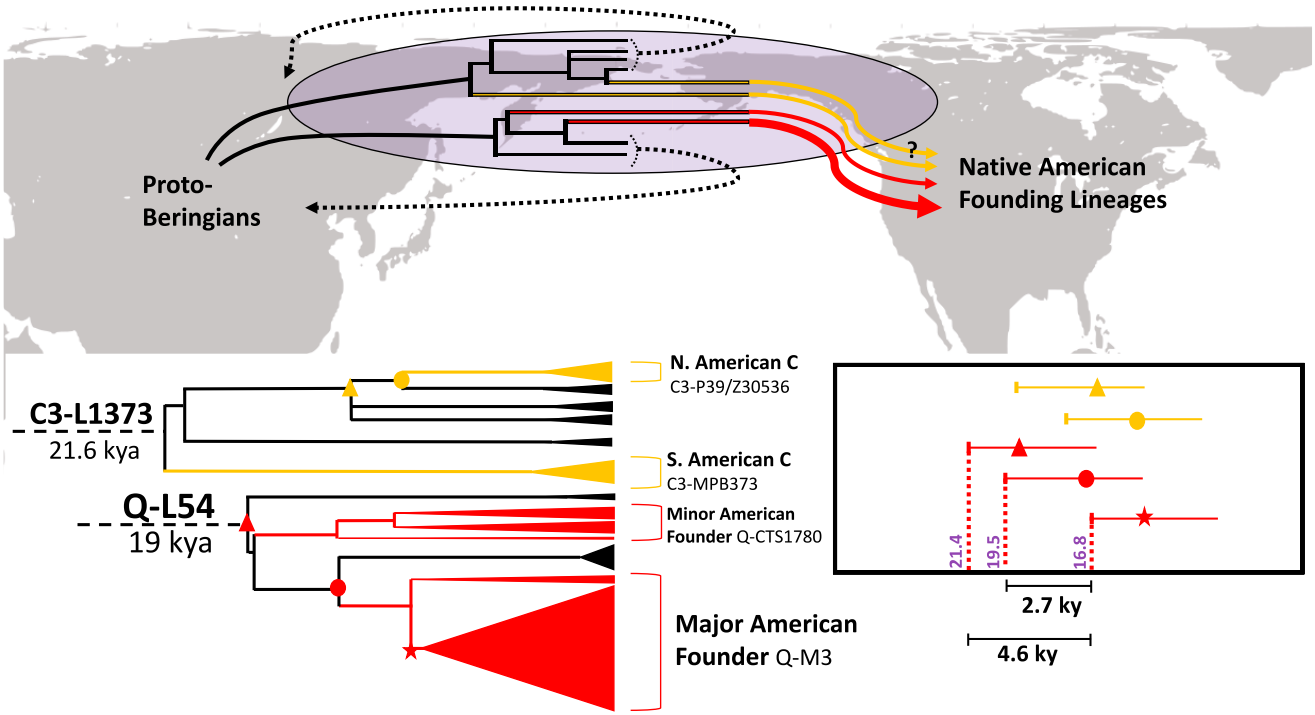

Figure 3. Proposed Model for Male-Lineage Expansions out of Beringia

Using present-day locations of lineages, ancient DNA evidence, and lineage divergence times, we propose a model where the Beringian Standstill lasted a maximum 4.6 kya and gave rise to the out-of-Beringia population carrying the three or four autochthonous founding lineages and the back-from-Beringia population carrying very rare paternal lineages now found in Siberian Kets and some Northern Eurasians (top). Arrows do not represent actual migration routes. A schematic phylogeny with relevant split dates discussed in the text is shown on the bottom left, and the range of confidence intervals used to calculate the length of the Beringian Standstill under different models is show on the bottom right.
See also STAR Methods.

mtDNA lineages [9], and even for the Dene-Yeniseian languages [33]. With this initial date for the start of the Beringian Standstill, it could have lasted up to 4.6 (21.4 minus 16.8) ky if we again consider the oldest date for the Q-M848 lineage expansion as the end of the Standstill period. The split time of 17.3 (15.2–19.5) ky between the American Q-M3 lineage and the Northern European Q-L804 would mark the date for separation of these lineages in Beringia. Because Q-M3 is most certainly an American-autochthonous lineage carried by the population that first expanded beyond the glaciers into the Americas (probably through a coastal route), we can establish 19.5 kya, the upper bound of this divergence, as the earliest date for the occupation of the Americas (south of the glaciers) by their current descendants.

Expansion and Geographical Structure within the Americas

The population expansion following the entry into the Americas, here identified by the male lineage expansion of haplogroup Q-M848 within Q-M3, is probably the largest-scale and most rapid demographic expansion in human history after the out-of-Africa event (which is identified by the CT-M168 lineage ancestral to all non-African Y chromosomes). The signal of this ~15 kya expansion has been detected in every sequence-based study of randomly chosen Native American Y chromosomes [11, 16], but here, the increased number of Y chromosomes examined and their ascertainment to represent the known genetic diversity within Native American Y chromosomes reveal an additional feature of this expansion. Of the 24 Q-M848 sub-lineages derived from the initial "starburst," 10 are represented by more than one chromosome, and we can compare the geographical locations of these Q-M848 descendant sub-lineages, which are scattered throughout Mexico and South America (Figure 4). Strikingly, after 12.3 (10.8–13.9) kya, all members of individual sub-lineages are clustered geographically, strongly suggesting that spatial structure among the paternal lineages in South America arose as early as 12.3 kya. The restriction of the C3-MPB373 lineage to parts of Ecuador, where it is found among Kichwa and Waorani speakers [19], also fits this pattern, and has in addition been proposed for North America on the basis of low-coverage ancient DNA data [34]. If this ancient history of occupation and population structure of Native South Americans is corroborated, it could partially explain the huge diversity of linguistic families and the difficulty of establishing relationships between them [35].

The inferred presence of Q-CTS1780 in Beringia and the evidence for it in ancient individuals [26] contrast sharply with its present-day distribution, rarity, ancient diversification, and the absence of a clear phylogenetic expansion signal. This could reflect genetic drift alone or perhaps another layer of structure

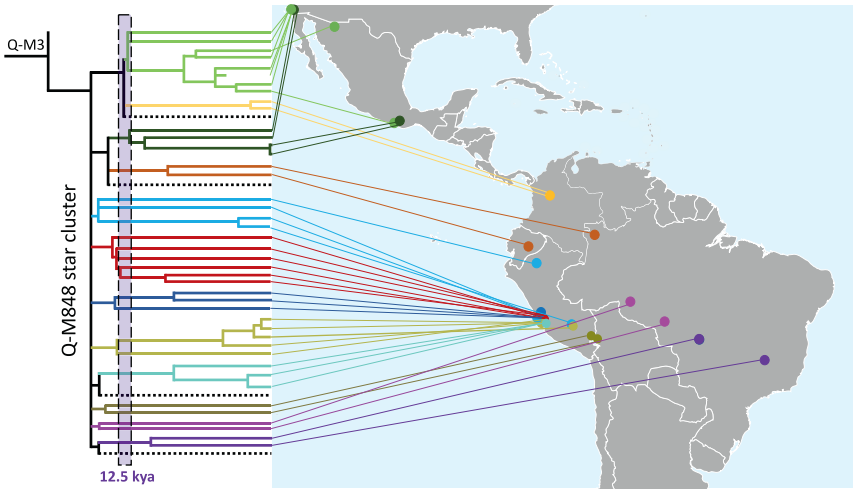

Figure 4. South American Y Chromosome Phylogeography

Comparison of the spatial distribution of samples and their phylogenetic relationship. Downstream of the Q-M848 star expansion, members of sub-lineages arising after 12.5 kya are geographically close in present-day populations, suggesting strong population structure that arose as early as the Late Pleistocene. Each branch is given a different color to facilitate the spatial visualization, and dotted lines indicate lineages represented by only one chromosome after 12.5 kya and therefore not placed on the map. Geographical locations are only approximate, and the phylogenetic branching pattern is the same as in Figure 2.

For full-sample listing and locations, see Data S1.

in Beringia [4, 5]. Early population structure in Beringia contributing to at least two major early dispersals toward the Americas, first via a coastal route (Pacific) and subsequently following the melting of the ice sheets, has been proposed in an interdisciplinary model considering both cranial morphology and genetic evidence [2, 36]. Autosomal analyses have proposed a deep split between the northernmost Native Americans and all other groups [6], but this signal is unclear in the Y chromosome phylogeny, as only two modern Y sequences from our dataset represent those northern populations. However, if there were some gene flow between those populations in the past, as recently proposed [37], it would be more difficult to use present-day Y chromosomes to establish the relationship between those ancestry components. Q-CTS1780 might thus be associated with a more complex out-of-Beringia settlement scenario into the Americas.

The pre-Columbian settlement of the Americas is even more complex if we consider the much later arrival (after the Bering Strait formation) of Palaeo-Eskimos [28, 29] and Inuits [38], as well as evidence from very rare Y chromosome lineages (e.g., Q-L53*, Q-L275, Q-NWT01, Q-P89) that are difficult to explain by an out-of-Beringia origin [12, 38], and the Holocene change in skull morphology of Native Americans after 12 kya [2]. Indeed, a scenario of a circum-Arctic gene flow through the Bering Strait during the Holocene has been proposed to explain the spread of many derived traits of the skulls found in Native Asians and Americans, as well as some rare Y and mtDNA lineages [36].

This study thus greatly increases the resolution of the paternal phylogeny in the Americas with a detailed analysis of a large quantity of ancient and modern Native American Y chromosomes that includes the phylogenetic placement of very rare (<0.5%) lineages from South America and Eurasia. Our dating strategy (see Data S4 and STAR Methods) relies on a mutation rate described elsewhere [39], and we highlight the fact that alternative rates and additional complexities exist [16, 25, 40, 41] that could lead to time estimates up to 15% later. The combination of present-day distributions of the lineages, ancient DNA data, and the timing of lineage divergences and expansion events nevertheless allows us to build a robust model for the initial settlement of the Americas (Figure 3), as well as to define two major founding paternal lineages (Q-M3 and Q-CTS1780) and two minor and rare lineages that need further study (C3-P39/Z30536 and C3-MPB373).

Conclusions

We conclude that the out-of-Beringia event into the Americas could only have happened after 19.5 kya and was shaped by a short Beringian Standstill no longer than 4.6 kya, assuming that Beringia also gave rise to some Northern Eurasian lineages (Q-L330, Q-L804) as a result of a back-from-Beringia dispersal. Human groups dispersed southward likely through an ice-free coastal route [31, 32], expanding rapidly after 15 kya and reaching all of South America soon after, as supported by widely accepted archaeological dates [42, 43]. We also infer that male population structure was present as early as 12 kya in South America, a conclusion also supported by the high heterogeneity in South American archaeological traditions [44] in contrast to their North American counterpart, the contemporaneous Clovis tradition, which was widespread and relatively uniform. We believe the data on Y chromosome variation in the American continents compiled here will prove a useful tool for guiding future research on pre-Columbian events in South America [45–50] and for ancient DNA analyses that will undoubtedly further change our understanding of the deep history that shaped the present-day biological and cultural diversity of the South American continent.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Sample selection and sequencing
  - Y chromosome coverage estimation
  - Variant calling and phylogenetic inference
  - Tree topology and novel SNPs

- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Dating of nodes
  - Estimating the length of the Beringian Standstill
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures and four data files and can be found with this article online at https://doi.org/10.1016/j.cub.2018.11.029.

ACKNOWLEDGMENTS

Foremost, the authors would like to thank all Native Americans who collaborated in the Genographic Project of South America and the EQ Genetics Project, which made this historical genetics research possible. This work is dedicated to them. The authors would also like to thank Pedro Paulo R. Vieira, Oscar Acosta, Donaldo Pinedo, and Paulo Robles-Ruiz for participating in the fieldwork in Peru and Ecuador and Marc Haber, Elena Arciero, Maria Cátira Bortolini, Rafael Bisso-Machado, Tábita Hünemeier, David Comas, and Ray Banks for helpful comments and discussion. We also thank Vladimir Gurianov, Vadim Urasin, and the YFull team for their detailed phylogeny online and Rui Martiniano, Christiana Scheib, Swapan Mallick, Michelle Lee, and David Reich for facilitating access to published sequences. T.P. and F.R.S. were supported by the National Geographic Society of the United States, FAPEMIG, CAPES, and CNPq of Brazil; A.B., Y.X., and C.T.-S. were supported by Wellcome grant number 098051.

Portuguese: Acima de tudo, os autores gostariam de agradecer a todas as comunidades indígenas participantes do Projeto Genográfico da América do Sul, que permitiram a realização desta pesquisa em genética histórica. Este trabalho é dedicado a eles.

Spanish: Principalmente, los autores les gustarían agradecer a todas las comunidades originarias que colaboraron en el Proyecto Genográfico de América del Sur, que han permitido la ejecución de esta investigación en genética histórica. Este trabajo está dedicado a ellos.

AUTHOR CONTRIBUTIONS

Conceptualization, Y.X., F.R.S., and C.T.-S.; Methodology, T.P., A.B., Y.X., F.R.S., and C.T.-S.; Formal Analysis, T.P. and A.B.; Investigation, T.P., A.B., M.G., W.S., and Q.A.; Resources, A.S., J.N., K.R., K.D., F.G.-A., C.P.-y.-M., S.R., C.C., J.R.S., R.F., Y.X., L.R., F.R.S., and C.T.-S.; Data Curation, T.P., M.B., M.G., D.O., D.R.L., A.S., J.N., K.R., K.D., M.S.J., J.E.S.J., L.R., and F.R.S.; Writing – Original Draft, T.P., A.B., Y.X., F.R.S., and C.T.-S.; Writing – Review & Editing, all authors; Visualization, T.P.; Supervision, T.K., Y.X., L.R., F.R.S., and C.T.-S.; Project Administration, Y.X., F.R.S., and C.T.-S., Funding Acquisition, F.R.S. and C.T.-S.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 3, 2018
Revised: September 3, 2018
Accepted: November 9, 2018
Published: December 20, 2018

REFERENCES

1. Goebel, T., Waters, M.R., and O'Rourke, D.H. (2008). The late Pleistocene dispersal of modern humans in the Americas. Science *319*, 1497–1502.
2. González-José, R., Bortolini, M.C., Santos, F.R., and Bonatto, S.L. (2008). The peopling of America: craniofacial shape variation on a continental scale and its interpretation from an interdisciplinary view. Am. J. Phys. Anthropol. *137*, 175–187.
3. Hoffecker, J.F., Elias, S.A., O'Rourke, D.H., Scott, G.R., and Bigelow, N.H. (2016). Beringia and the global dispersal of modern humans. Evol. Anthropol. *25*, 64–78.
4. Skoglund, P., and Reich, D. (2016). A genomic view of the peopling of the Americas. Curr. Opin. Genet. Dev. *41*, 27–35.
5. Moreno-Mayar, J.V., Potter, B.A., Vinner, L., Steinrücken, M., Rasmussen, S., Terhorst, J., Kamm, J.A., Albrechtsen, A., Malaspinas, A.S., Sikora, M., et al. (2018). Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. Nature *553*, 203–207.
6. Raghavan, M., Steinrücken, M., Harris, K., Schiffels, S., Rasmussen, S., DeGiorgio, M., Albrechtsen, A., Valdiosera, C., Ávila-Arcos, M.C., Malaspinas, A.S., et al. (2015). POPULATION GENETICS. Genomic evidence for the Pleistocene and recent population history of Native Americans. Science *349*, aab3884.
7. Llamas, B., Fehren-Schmitz, L., Valverde, G., Soubrier, J., Mallick, S., Rohland, N., Nordenfelt, S., Valdiosera, C., Richards, S.M., Rohrlach, A., et al. (2016). Ancient mitochondrial DNA provides high-resolution time scale of the peopling of the Americas. Sci. Adv. *2*, e1501385.
8. Bonatto, S.L., and Salzano, F.M. (1997). A single and early migration for the peopling of the Americas supported by mitochondrial DNA sequence data. Proc. Natl. Acad. Sci. USA *94*, 1866–1871.
9. Tamm, E., Kivisild, T., Reidla, M., Metspalu, M., Smith, D.G., Mulligan, C.J., Bravi, C.M., Rickards, O., Martinez-Labarga, C., Khusnutdinova, E.K., et al. (2007). Beringian standstill and spread of Native American founders. PLoS ONE *2*, e829.
10. Zegura, S.L., Karafet, T.M., Zhivotovsky, L.A., and Hammer, M.F. (2004). High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas. Mol. Biol. Evol. *21*, 164–175.
11. Poznik, G.D., Xue, Y., Mendez, F.L., Willems, T.F., Massaia, A., Wilson Sayres, M.A., Ayub, Q., McCarthy, S.A., Narechania, A., Kashin, S., et al.; 1000 Genomes Project Consortium (2016). Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. Nat. Genet. *48*, 593–599.
12. Jota, M.S., Lacerda, D.R., Sandoval, J.R., Vieira, P.P., Ohasi, D., Santos-Júnior, J.E., Acosta, O., Cuellar, C., Revollo, S., Paz-Y-Miño, C., et al.; Genographic Consortium (2016). New native South American Y chromosome lineages. J. Hum. Genet. *61*, 593–603.
13. Poznik, G.D., Henn, B.M., Yee, M.C., Sliwerska, E., Euskirchen, G.M., Lin, A.A., Snyder, M., Quintana-Murci, L., Kidd, J.M., Underhill, P.A., and Bustamante, C.D. (2013). Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males versus females. Science *341*, 562–565.
14. Bergström, A., Nagle, N., Chen, Y., McCarthy, S., Pollard, M.O., Ayub, Q., Wilcox, S., Wilcox, L., van Oorschot, R.A., McAllister, P., et al. (2016). Deep roots for Aboriginal Australian Y chromosomes. Curr. Biol. *26*, 809–813.
15. Hudjashov, G., Kivisild, T., Underhill, P.A., Endicott, P., Sanchez, J.J., Lin, A.A., Shen, P., Oefner, P., Renfrew, C., Villems, R., and Forster, P. (2007). Revealing the prehistoric settlement of Australia by Y chromosome and mtDNA analysis. Proc. Natl. Acad. Sci. USA *104*, 8726–8730.
16. Karmin, M., Saag, L., Vicente, M., Wilson Sayres, M.A., Järve, M., Talas, U.G., Rootsi, S., Ilumäe, A.M., Mägi, R., Mitt, M., et al. (2015). A recent bottleneck of Y chromosome diversity coincides with a global change in culture. Genome Res. *25*, 459–466.
17. Xue, Y., Zerjal, T., Bao, W., Zhu, S., Shu, Q., Xu, J., Du, R., Fu, S., Li, P., Hurles, M.E., et al. (2006). Male demography in East Asia: a north-south contrast in human population expansion times. Genetics *172*, 2431–2439.
18. Zerjal, T., Xue, Y., Bertorelle, G., Wells, R.S., Bao, W., Zhu, S., Qamar, R., Ayub, Q., Mohyuddin, A., Fu, S., et al. (2003). The genetic legacy of the Mongols. Am. J. Hum. Genet. *72*, 717–721.
19. Roewer, L., Nothnagel, M., Gusmão, L., Gomes, V., González, M., Corach, D., Sala, A., Alechine, E., Palha, T., Santos, N., et al. (2013). Continent-wide decoupling of Y-chromosomal genetic variation from language and geography in native South Americans. PLoS Genet. *9*, e1003460.

20. Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., Furtwängler, A., Haak, W., Meyer, M., Mittnik, A., et al. (2016). The genetic history of Ice Age Europe. Nature *534*, 200–205.
21. Olalde, I., Allentoft, M.E., Sánchez-Quinto, F., Santpere, G., Chiang, C.W., DeGiorgio, M., Prado-Martinez, J., Rodríguez, J.A., Rasmussen, S., Quilez, J., et al. (2014). Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. Nature *507*, 225–228.
22. Seguin-Orlando, A., Korneliussen, T.S., Sikora, M., Malaspinas, A.S., Manica, A., Moltke, I., Albrechtsen, A., Ko, A., Margaryan, A., Moiseyev, V., et al. (2014). Paleogenomics. Genomic structure in Europeans dating back at least 36,200 years. Science *346*, 1113–1118.
23. Sikora, M., Seguin-Orlando, A., Sousa, V.C., Albrechtsen, A., Korneliussen, T., Ko, A., Rasmussen, S., Dupanloup, I., Nigst, P.R., Bosch, M.D., et al. (2017). Ancient genomes show social and reproductive behavior of early Upper Paleolithic foragers. Science *358*, 659–662.
24. Mezzavilla, M., Geppert, M., Tyler-Smith, C., Roewer, L., and Xue, Y. (2015). Insights into the origin of rare haplogroup C3* Y chromosomes in South America from high-density autosomal SNP genotyping. Forensic Sci. Int. Genet. *15*, 115–120.
25. Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N., Nordenfelt, S., Tandon, A., et al. (2016). The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature *538*, 201–206.
26. Rasmussen, M., Anzick, S.L., Waters, M.R., Skoglund, P., DeGiorgio, M., Stafford, T.W., Jr., Rasmussen, S., Moltke, I., Albrechtsen, A., Doyle, S.M., et al. (2014). The genome of a Late Pleistocene human from a Clovis burial site in western Montana. Nature *506*, 225–229.
27. Rasmussen, M., Sikora, M., Albrechtsen, A., Korneliussen, T.S., Moreno-Mayar, J.V., Poznik, G.D., Zollikofer, C.P.E., de León, M.P., Allentoft, M.E., Moltke, I., et al. (2015). The ancestry and affiliations of Kennewick Man. Nature *523*, 455–458.
28. Rasmussen, M., Li, Y., Lindgreen, S., Pedersen, J.S., Albrechtsen, A., Moltke, I., Metspalu, M., Metspalu, E., Kivisild, T., Gupta, R., et al. (2010). Ancient human genome sequence of an extinct Palaeo-Eskimo. Nature *463*, 757–762.
29. Raghavan, M., DeGiorgio, M., Albrechtsen, A., Moltke, I., Skoglund, P., Korneliussen, T.S., Grønnow, B., Appelt, M., Gulløv, H.C., Friesen, T.M., et al. (2014). The genetic prehistory of the New World Arctic. Science *345*, 1255832.
30. Kemp, B.M., Malhi, R.S., McDonough, J., Bolnick, D.A., Eshleman, J.A., Rickards, O., Martinez-Labarga, C., Johnson, J.R., Lorenz, J.G., Dixon, E.J., et al. (2007). Genetic analysis of early holocene skeletal remains from Alaska and its implications for the settlement of the Americas. Am. J. Phys. Anthropol. *132*, 605–621.
31. Braje, T.J., Dillehay, T.D., Erlandson, J.M., Klein, R.G., and Rick, T.C. (2017). Finding the first Americans. Science *358*, 592–594.
32. Pedersen, M.W., Ruter, A., Schweger, C., Friebe, H., Staff, R.A., Kjeldsen, K.K., Mendoza, M.L., Beaudoin, A.B., Zutter, C., Larsen, N.K., et al. (2016). Postglacial viability and colonization in North America's ice-free corridor. Nature *537*, 45–49.
33. Sicoli, M.A., and Holton, G. (2014). Linguistic phylogenies support back-migration from Beringia to Asia. PLoS ONE *9*, e91722.
34. Lindo, J., Achilli, A., Perego, U.A., Archer, D., Valdiosera, C., Petzelt, B., Mitchell, J., Worl, R., Dixon, E.J., Fifield, T.E., et al. (2017). Ancient individuals from the North American Northwest Coast reveal 10,000 years of regional genetic continuity. Proc. Natl. Acad. Sci. USA *114*, 4093–4098.
35. Campbell, L. (2015). Do Languages and Genes Correlate? Some Methodological Issues. Lang. Dyn. Change *5*, 202–226.
36. Bortolini, M.C., González-José, R., Bonatto, S.L., and Santos, F.R. (2014). Reconciling pre-Columbian settlement hypotheses requires integrative, multidisciplinary, and model-bound approaches. Proc. Natl. Acad. Sci. USA *111*, E213–E214.
37. Scheib, C.L., Li, H., Desai, T., Link, V., Kendall, C., Dewar, G., Griffith, P.W., Mörseburg, A., Johnson, J.R., Potter, A., et al. (2018). Ancient human parallel lineages within North America contributed to a coastal expansion. Science *360*, 1024–1027.
38. Dulik, M.C., Owings, A.C., Gaieski, J.B., Vilar, M.G., Andre, A., Lennie, C., Mackenzie, M.A., Kritsch, I., Snowshoe, S., Wright, R., et al.; Genographic Consortium (2012). Y-chromosome analysis reveals genetic divergence and new founding native lineages in Athapaskan- and Eskimoan-speaking populations. Proc. Natl. Acad. Sci. USA *109*, 8471–8476.
39. Fu, Q., Li, H., Moorjani, P., Jay, F., Slepchenko, S.M., Bondarev, A.A., Johnson, P.L., Aximu-Petri, A., Prüfer, K., de Filippo, C., et al. (2014). Genome sequence of a 45,000-year-old modern human from western Siberia. Nature *514*, 445–449.
40. Harris, K. (2015). Evidence for recent, population-specific evolution of the human mutation rate. Proc. Natl. Acad. Sci. USA *112*, 3439–3444.
41. Helgason, A., Einarsson, A.W., Guðmundsdóttir, V.B., Sigurðsson, Á., Gunnarsdóttir, E.D., Jagadeesan, A., Ebenesersdóttir, S.S., Kong, A., and Stefánsson, K. (2015). The Y-chromosome point mutation rate in humans. Nat. Genet. *47*, 453–457.
42. Bueno, L., Dias, A.S., and Steele, J. (2013). The Late Pleistocene/Early Holocene archaeological record in Brazil: A geo-referenced database. Quat. Int. *301*, 74–93.
43. Dillehay, T.D., Ocampo, C., Saavedra, J., Sawakuchi, A.O., Vega, R.M., Pino, M., Collins, M.B., Scott Cummings, L., Arregui, I., Villagran, X.S., et al. (2015). New archaeological evidence for an early human presence at Monte Verde, Chile. PLoS ONE *10*, e0141923.
44. Araujo, A.G.M. (2015). On vastness and variability: cultural transmission, historicity, and the Paleoindian record in eastern South America. An. Acad. Bras. Cienc. *87*, 1239–1258.
45. Barbieri, C., Heggarty, P., Yang Yao, D., Ferri, G., De Fanti, S., Sarno, S., Ciani, G., Boattini, A., Luiselli, D., and Pettener, D. (2014). Between Andes and Amazon: the genetic profile of the Arawak-speaking Yanesha. Am. J. Phys. Anthropol. *155*, 600–609.
46. Barbieri, C., Sandoval, J.R., Valqui, J., Shimelman, A., Ziemendorff, S., Schröder, R., Geppert, M., Roewer, L., Gray, R., Stoneking, M., et al. (2017). Enclaves of genetic diversity resisted Inca impacts on population history. Sci. Rep. *7*, 17411.
47. de la Fuente, C., Galimany, J., Kemp, B.M., Judd, K., Reyes, O., and Moraga, M. (2015). Ancient marine hunter-gatherers from Patagonia and Tierra Del Fuego: Diversity and differentiation using uniparentally inherited genetic markers. Am. J. Phys. Anthropol. *158*, 719–729.
48. Fehren-Schmitz, L., Warnberg, O., Reindel, M., Seidenberg, V., Tomasto-Cagigao, E., Isla-Cuadrado, J., Hummel, S., and Herrmann, B. (2011). Diachronic investigations of mitochondrial and Y-chromosomal genetic markers in pre-Columbian Andean highlanders from South Peru. Ann. Hum. Genet. *75*, 266–283.
49. Sandoval, J.R., Lacerda, D.R., Acosta, O., Jota, M.S., Robles-Ruiz, P., Salazar-Granara, A., Vieira, P.P.R., Paz-Y-Miño, C., Fujita, R., and Santos, F.R.; Genographic Project Consortium (2016). The genetic history of Peruvian Quechua-Lamistas and Chankas: uniparental DNA patterns among autochthonous Amazonian and Andean populations. Ann. Hum. Genet. *80*, 88–101.
50. Sandoval, J.R., Lacerda, D.R., Jota, M.S., Elward, R., Acosta, O., Pinedo, D., Danos, P., Cuellar, C., Revollo, S., Santos, F.R., and Fujita, R. (2018). Genetic ancestry of families of putative Inka descent. Mol. Genet. Genomics *293*, 873–881.
51. Damgaard, P.B., Marchi, N., Rasmussen, S., Peyrot, M., Renaud, G., Korneliussen, T., Moreno-Mayar, J.V., Pedersen, M.W., Goldberg, A., Usmanova, E., et al. (2018). 137 ancient human genomes from across the Eurasian steppes. Nature *557*, 369–374.
52. Pritchard, J.K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics *155*, 945–959.
53. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

54. Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. arXiv, arXiv:1207.3907.

55. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313.

56. Jota, M.S., Lacerda, D.R., Sandoval, J.R., Vieira, P.P., Santos-Lopes, S.S., Bisso-Machado, R., Paixão-Cortes, V.R., Revollo, S., Paz-Y-Miño, C., Fujita, R., et al.; Genographic Consortium (2011). A new subhaplogroup of native American Y-Chromosomes from the Andes. Am. J. Phys. Anthropol. *146*, 553–559.

57. Balanovsky, O., Gurianov, V., Zaporozhchenko, V., Balaganskaya, O., Urasin, V., Zhabagin, M., Grugni, V., Canada, R., Al-Zahery, N., Raveane, A., et al. (2017). Phylogeography of human Y-chromosome haplogroup Q3-L275 from an academic/citizen science collaboration. BMC Evol. Biol. *17* (*Suppl 1*), 18.

58. Ilumäe, A.-M., Reidla, M., Chukhryaeva, M., Järve, M., Post, H., Karmin, M., Saag, L., Agdzhoyan, A., Kushniarevich, A., Litvinov, S., et al. (2016). Human Y chromosome haplogroup N: a non-trivial time-resolved phylogeography that cuts across language families. Am. J. Hum. Genet. *99*, 163–173.

59. Hallast, P., Batini, C., Zadik, D., Maisano Delser, P., Wetton, J.H., Arroyo-Pardo, E., Cavalleri, G.L., de Knijff, P., Destro Bisol, G., Dupuy, B.M., et al. (2015). The Y-chromosome tree bursts into leaf: 13,000 high-confidence SNPs covering the majority of known clades. Mol. Biol. Evol. *32*, 661–673.

60. Wei, W., Ayub, Q., Xue, Y., and Tyler-Smith, C. (2013). A comparison of Y-chromosomal lineage dating using either resequencing or Y-SNP plus Y-STR genotyping. Forensic Sci. Int. Genet. *7*, 568–572.

61. Kivisild, T. (2017). The study of human Y chromosome variation through ancient DNA. Hum. Genet. *136*, 529–546.

62. Sims, D., Sudbery, I., Ilott, N.E., Heger, A., and Ponting, C.P. (2014). Sequencing depth and coverage: key considerations in genomic analyses. Nat. Rev. Genet. *15*, 121–132.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
20 Native American Y chromosomes	S Revollo, C Paz-y-Miño, R Fujita & FR Santos	This study
4 Northern European Y chromosomes	J Norstedt; A Solli; K Dawtry; K Reed & Q Nordic Family Tree DNA group project, http://www.familytreedna.com/groups/qnordic/about	This study
SNP genotypes from 21 Native American Y chromosomes from Ecuador	F González-Andrade, M Geppert & L Roewer – from EQ-Genetics Project	This study Data S3
1000 Genomes Y chromosome data	1000 Genomes Consortium; http://www.internationalgenome.org/	N/A
Saqqaq ancient genome	[28]	SRA: SRA010102
Dakelh (Athabascan) and Nivh2 genome	[29]; http://www.cbs.dtu.dk/suppl/arctic/	ENA: PRJEB6516
Anzick ancient genome	[26]	SRA: SRX381032
Kennewick ancient genome	[27]	SRA: SRS937952
Ket1, Ket2, CEPH_11_D12 modern genomes; Enoque65 ancient genome	[6], http://www.cbs.dtu.dk/suppl/NativeAmerican/	ENA: PRJEB9733
Simons Genome Diversity Project genomes	[25], some require signed consent letter	ENA: PRJEB9586 ENA: ERP010710
13 ancient genomes from North America	[37]	ENA: PRJEB25445
22 ancient genomes from Siberia	[51]	ENA: PRJEB26349
10 modern genomes from Central Asia	[51]	ENA: PRJEB26349
Software and Algorithms		
STRUCTURE 2.3.4	https://web.stanford.edu/group/pritchardlab/structure.html	[52]
SAMtools 1.6	http://www.htslib.org	[53]
FreeBayes v. 0.9.18	https://github.com/ekg/freebayes	[54]
BCFtools v. 1.3.1	http://www.sanger.ac.uk/science/tools/samtools-bcftools-htslib	N/A
vcflib	https://github.com/vcflib/vcflib	N/A
RAxML v. 8.2.10	https://sco.h-its.org/exelixis/software.html	[55]
FigTree	http://tree.bio.ed.ac.uk/software/figtree/	N/A
ISOGG (International Society for Genetic Genealogy)	https://isogg.org	N/A

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Chris Tyler-Smith (cts@sanger.ac.uk).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

The novel Native American samples were collected from participants in Brazil, Peru, Bolivia and Ecuador during the South American Genographic Project, and ethical approval was provided by the Brazilian National Ethics Committee (CONEP Resolution number 763/2009) and by local ethical committees for the non-Brazilian samples. The DNA was extracted from mouth swabs using standard procedures and had previously been typed to identify Native American haplogroups using a TaqMan RT-PCR approach [12, 56]. Previously-described C3 samples from Ecuador (Kichwas and Waoranis) [19] were collected by F González-Andrade, M Geppert and L Roewer within the EQ-Genetics Project.

The four new Northern European BigY data were obtained through a collaboration with a citizen-science initiative, Q Nordic, and were sequenced prior to our study. The sample donors have consented to all analyses done in this study and the non-commercial usage of the data.

METHOD DETAILS

Sample selection and sequencing

Using Y-STR data from over 2000 Native American participants from South America [12], a STRUCTURE (v. 2.3.4) analysis was performed to identify the most divergent chromosomes within Native American diversity belonging to lineages Q-M3 and Q-CTS1780. Eighteen of those highly divergent Y chromosome samples were chosen and, together with two Peruvians potentially related to the Inca ruler's descendants [50], sequenced to > 60x Y chromosome coverage using BigY, a commercial method from Family Tree DNA (Gene by Gene; https://www.familytreedna.com). This method uses thousands of capture probes to sequence approximately 10 Mb of the male-specific non-recombining portion of the Y chromosome, followed by Illumina sequencing (paired-end, 100 bp reads). The reads were mapped to GRCh37 human reference using its proprietary software Arpeggi.

Y chromosome coverage estimation

The regions of the Y chromosome targeted by BigY do not fully overlap with the regions typically used for non-targeted Illumina sequencing data, e.g., the approximately 10 Mb of accessible sequence defined in [13]. To get a better understanding of possible coverage bias from our compiled dataset, we calculated the Y chromosome coverage in this region in all 222 samples using SAMtools depth function (arguments '–r Y' to target the Y chromosome only, '–b' to specify the bed file with the 10 Mb region and '-aa' so SAMtools output all sites from the selected region). We plotted this information for the shotgun samples with its reported mean whole-genome coverage and the expected Y chromosome coverage – this is, half of the autosomes (Figure S4).

Variant calling and phylogenetic inference

Several strategies have been employed in previous studies using BigY to try to avoid platform bias [57, 58], which we initially found to heavily influence the fine-scale definition of branches in phylogenetic analyses, likely related to patterns of missing data between platforms. We performed joint variant calling across all 179 samples using FreeBayes v. 0.9.18 (arguments "–ploidy 1" and "–report-monomorphic") in the accessible 10 Mb, after unifying the header sequence dictionaries of the BAM files from the various sources using SAMtools. We then used a Perl script to set any genotype supported by less than two reads to missing and used BCFtools v. 1.3.1 to retain only positions that were called in at least 22 of the 24 BigY samples. This reduced our dataset to 7.6 Mb of overlapping sequence and 13,221 high-confidence SNPs, but avoided regions of systematic differences in genotype missingness between samples obtained from the literature and those sequenced through BigY capture. The vcfallelicprimitives tool from the vcflib package was used to decompose multi-nucleotide variants, and indels were removed. Those positions that were polymorphic were then extracted and transposed to fasta format to build a maximum-likelihood phylogeny using RAxML v. 8.2.10, with the "ASC_GTRGAMMA" substitution model and the "stamatakis" ascertainment correction. Statistical support for the clades in the tree was assessed by running 100 bootstrap replicates, and trees were visualized using the FigTree software and manually rooted between the A0 and A1 haplogroups.

Tree topology and novel SNPs

The maximum likelihood inference was not capable of detecting fine-scale substructure inside the star-shaped phylogenetic clade of Q-M848. We therefore proceeded with a more classical approach, using shared variants to resolve the tree topology by parsimony, producing a phylogram. All SNPs, private and shared, can be found in Data S2, together with any pre-existing ISOGG labels. SNPs with a phylogenetic signal equivalent to that of a more well-known variant were given tags of the form '<SNPID > _eq' to simplify their handling and interpretation. Novel SNPs were named with a prefix of MPB and deposited in the ISOGG database. Branches were only named after new SNPs if no other previously-known SNPs supported them. The new nomenclature for the C haplogroup was not used in this study due to synonymy and potential confusion with previous haplogroup names. Previously haplogroups were named C1, C2 to C7, but recent developments led to a refinement of the tree, in a bifurcating C phylogeny with C3 as a sister branch of all other Cs. The new nomenclature then proposed in ISOGG uses C1 to identify all C haplogroups with the exception of C3, identified as C2. Therefore, the C3 clade described here is the one defined by the marker M217. The two new C clades described here were called C3-MPB373 and C3-MPB202. C3-MPB373 represents the South American C3* from Ecuador (see Main Text), and C3-MPB202 is a lineage represented by two C3* samples from China (1000 Genomes CHS HG00628 and one Dai individual from [25]). C3-MPB202 is a deeply divergent branch in C3-F2645, and sister branch to the common C3-Z1338 lineage from South, Southeast and East Asia. The marker P39 itself [10] was filtered out of our analyses as it is located in a palindromic region of the Y chromosome outside the 10 Mb accessible region. Therefore, we recommend here the usage of marker Z30536 (G>A transition in position chrY:7,601,110 in GRCh37) to define the clade, and the joint name C3-P39/Z30536 for it. Details of SNP calls are tabulated in Data S2.

QUANTIFICATION AND STATISTICAL ANALYSIS

Dating of nodes

Nodes in the phylogeny were dated using the ρ (rho) statistic, using the Y chromosome mutation rate reported in [39], obtained from the genome of the upper Palaeolithic Ust'-Ushim man, currently the oldest modern human genome sequenced to date. This rate is 0.76×10^{-9} mutations per site per year, with a 95% confidence interval of 0.67×10^{-9} to 0.86×10^{-9}. Variant sites were only counted if confidently called in both of the two samples used for a given pairwise dating calculation, and a rate was obtained by dividing the

number of differing genotypes by the total number of confidently called sites, including those that did not vary [14]. Such divergence estimates based on complete sequencing have been demonstrated to be considerably more robust than those based on Y-STRs [59, 60], and less prone to artifacts due to fluctuation in variant counts between samples. When the dating of a node involved one ancient sample, because of low coverage and DNA damage, its genotypes were only used to count the derived alleles in the modern sample (for a review, see [61]), and no attempt was made to count the number of derived alleles in the ancient sample.

Both the number of called sites with derived genotypes and the number of called sites with ancestral genotypes vary between high and low coverage samples. Theoretically, we would expect variation in coverage to affect these two numbers proportionally, after filtering [62] [59], though in practice we still find very low coverage leads to less reliable results, likely because of imperfect filtering. Because BigY is a targeted capture method, its coverage is not uniform, with some positions within the targeted regions still having very low coverage, and therefore high coverage non-capture samples from the literature were preferentially used for node dating whenever possible. If more than one sample was available downstream of a given node, estimates across them were averaged. Therefore, the confidence intervals reported throughout the text between parentheses after the point estimate aim to represent two sources of uncertainty: 1) error in the estimated mutation rate from [39], by using also its 95% confidence interval estimates, and 2) the stochastic nature of mutations and its impact in variant number across samples.

Estimating the length of the Beringian Standstill

Following the logic of previous mtDNA work [7], the Beringian Standstill took place at a time (1) after the last observable divergence event between Eurasian and American lineages, and (2) before the oldest date all autochthonous American lineages formed. We point out that parameter (2) assumes there are no Eurasian lineages which will nest inside what we think today to be exclusively American lineages. Although this is a reasonable assumption, we reasoned a better parameter to measure the end of the Standstill would be the oldest time estimated for the explosive population expansion inferred from the phylogeny at the Q-M848 node, as this most certainly took place south of the glaciers.

However, the length of the Beringian Standstill also depends on the demographic scenario assumed after its end: whether the Beringian diversity gave rise only to Native American lineages (which we call the 'Beringian-American' hypothesis) or also to Eurasian lineages (the 'out-of-Beringia' hypothesis). In the first model, the oldest time the start of the diversification of Native American lineages could take place is around 19.5 kya, after the last split between an Eurasian and an American lineage (Q-L804 x Q-M3), leading to a length of the Standstill of only 2.7 ky (between 19.5 kya and 16.8 kya). If we assume a Beringian origin for Q-L804 and maybe also Q-L330 (as the split times between Q-L54 and Q-MPB001 are almost identical), then the likely time of isolation and maturation of Beringian lineages can be estimated to start around 21.3 kya, yielding a longer Standstill of around 4.6 ky (between 21.4 kya and 16.8). We note that, for these calculations, we did not used the time estimates of the C3 lineages because we only had one chromosome for each relevant American lineage, and additionally we believe it is not clear whether C3-P39/Z30536 represents an initial founding lineage or a later entry.

DATA AND SOFTWARE AVAILABILITY

DRYAD https://doi.org/10.5061/dryad.h38853n.