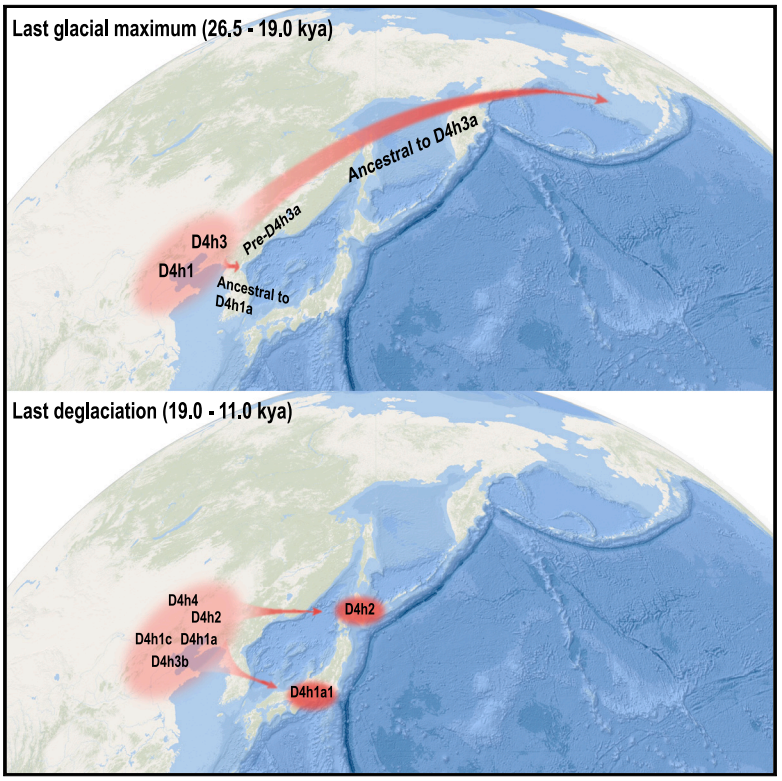
Cell Reports

Report

Mitogenome evidence shows two radiation events and dispersals of matrilineal ancestry from northern coastal China to the Americas and Japan

Graphical abstract



Highlights

- The Native American lineage D4h3a can trace its ancestry to northern coastal China
- Radiations of D4h contribute to gene pools of Native Americans and Japanese
- Coastal radiations of D4h support the coastal route of early Native Americans

Authors

Yu-Chun Li, Zong-Liang Gao, Kai-Jun Liu, ..., Ornella Semino, Antonio Torroni, Qing-Peng Kong

Correspondence

kongqp@mail.kiz.ac.cn

In brief

Li et al. analyze the contemporary and ancient mitogenomes of haplogroup D4h and trace the ancestral source of the Native American lineage D4h3a to northern coastal China. This ancestral source contributes to gene pools of both Native Americans and the Japanese and lends supports to the coastal route of early Native Americans.

Li et al., 2023, Cell Reports *42*, 112413
May 30, 2023
https://doi.org/10.1016/j.celrep.2023.112413

CellPress

Cell Reports

CellPress
OPEN ACCESS

Report

Mitogenome evidence shows two radiation events and dispersals of matrilineal ancestry from northern coastal China to the Americas and Japan

Yu-Chun Li,[1,4,5,10] Zong-Liang Gao,[1,4,5,6,7,10] Kai-Jun Liu,[2] Jiao-Yang Tian,[1,4,5] Bin-Yu Yang,[1,4,5] Zia Ur Rahman,[1,5,6,7] Li-Qin Yang,[1,4,5] Su-Hua Zhang,[8] Cheng-Tao Li,[8] Alessandro Achilli,[9] Ornella Semino,[9] Antonio Torroni,[9] and Qing-Peng Kong[1,3,4,5,11,*]

[1]State Key Laboratory of Genetic Resources and Evolution/Key Laboratory of Healthy Aging Research of Yunnan Province, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan 650223, China
[2]Chengdu 23Mofang Biotechnology Co., Ltd., Tianfu Software Park, Chengdu, Sichuan 610042, China
[3]CAS Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, Yunnan 650223, China
[4]KIZ/CUHK Joint Laboratory of Bioresources and Molecular Research in Common Diseases, Kunming, Yunnan, 650223, China
[5]Kunming Key Laboratory of Healthy Aging Study, Kunming, Yunnan 650223, China
[6]University of Chinese Academy of Sciences, Beijing 100049, China
[7]Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming, Yunnan 650204, China
[8]Shanghai Key Laboratory of Forensic Medicine, Shanghai Forensic Service Platform, Academy of Forensic Science, Ministry of Justice, Shanghai 200063, China
[9]Department of Biology and Biotechnology "L. Spallanzani", University of Pavia, 27100 Pavia, Italy
[10]These authors contributed equally
[11]Lead contact
*Correspondence: kongqp@mail.kiz.ac.cn
https://doi.org/10.1016/j.celrep.2023.112413

SUMMARY

Although it is widely recognized that the ancestors of Native Americans (NAs) primarily came from Siberia, the link between mitochondrial DNA (mtDNA) lineage D4h3a (typical of NAs) and D4h3b (found so far only in East China and Thailand) raises the possibility that the ancestral sources for early NAs were more variegated than hypothesized. Here, we analyze 216 contemporary (including 106 newly sequenced) D4h mitogenomes and 39 previously reported ancient D4h data. The results reveal two radiation events of D4h in northern coastal China, one during the Last Glacial Maximum and the other within the last deglaciation, which facilitated the dispersals of D4h sub-branches to different areas including the Americas and the Japanese archipelago. The coastal distributions of the NA (D4h3a) and Japanese lineages (D4h1a and D4h2), in combination with the Paleolithic archaeological similarities among Northern China, the Americas, and Japan, lend support to the coastal dispersal scenario of early NAs.

INTRODUCTION

As the last continent settled by modern humans, the peopling of the Americas and subsequent dispersals within the continent have been the focus of intense interest by geneticists.[1–6] Previous studies have shown that the ancestors of Indigenous Americans, also called Native Americans (NAs), originated in Asia, most likely in the eastern part of Asia,[3,6–9] and settled in the Americas by means of multiple dispersals through Siberia/Beringia[10] via the coastal route and possibly the inland ice-free corridor,[11] followed by later divergence into sub-groups.[12]

The origin of early NAs, to date, has been attributed to a complex process involving multiple dispersals from different source places. As indicated by substantial investigations, besides the widely recognized Siberian ancestry, ancestries from other places, although limited, have also been identified, including North Asia,[6,9] East Asia,[6,13] Southeast Asia,[14] and even Australo-Melanesia.[15] In agreement with these observations, evidence from uniparental markers further indicates that the majority of NAs show closer genetic affinity to Siberians, as manifested by NA founder types, e.g., mitochondrial DNA (mtDNA) haplogroups A2, B2, C1, C4c, D1, etc.,[16–19] and Y chromosome haplogroups Q-L54 (Q-Z780, Q-M848, and Q-M4303) and C-L1373 (C-MBP373),[19–24] and thus may trace their ancestral sources in Siberia. In contrast, a sister lineage of the NA matrilineal founder D4h3a,[25,26] viz., D4h3b, has been so far observed only in China[25] and Thailand,[27,28] suggesting that the ancestral maternal sources for early NAs were not restricted to Siberia but were from a much wider Asian geographic range.

To address this issue, an investigation integrating all available D4h data from a large-scale dataset covering the whole of Eurasia is needed. Given that D4h3 and its ancestor type D4h are relatively rare in contemporary populations (~0.5%),[29] we surveyed a total of 101,319 Eurasian individuals and identified the

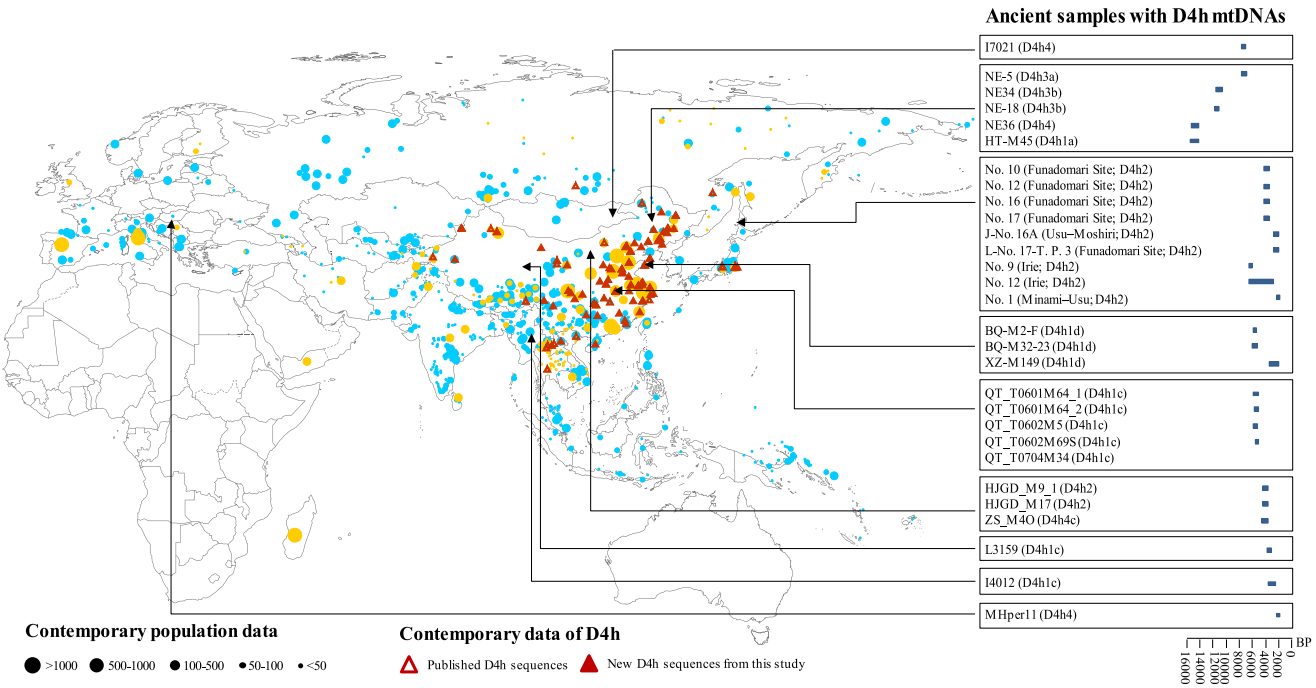


Figure 1. Geographic sources of mtDNA data employed in this study

Circles: populations surveyed for HVS variation are in light blue, while those surveyed for the variation of the entire mitogenome are in yellow. Only data from population surveys (99,722 samples from 1,135 populations) are shown. The remaining 1,597 mtDNAs are not shown on the map either because they were sporadic samples or because geographic information was lacking. For more details concerning the 101,319 samples, see Tables S1 and S2. Triangles: D4h samples, including published (hollow triangles) and newly sequenced samples (filled triangles). Ancient Asian samples harboring D4h mtDNAs were indicated by arrows, with the information shown on the right. The ancient samples from the Americas (see Table S4) are not shown.

mtDNAs belonging to D4h3 and its ancestral node D4h. These included 60,979 samples for which partial sequence data, mainly hypervariable segment (HVS) data (Table S1), were available and 40,340 samples with the complete (or almost complete) mitogenome sequence (Table S2; Figure 1). This survey identified 110 mtDNAs that could be assigned unambiguously to haplogroup D4h based on mitogenome information as well as 112 mtDNAs likely belonging to D4h based on their HVS or genotyping data (Table S3), whose complete sequencing revealed 106 additional D4h mitogenomes (Figure S1). Furthermore, to reconstruct the evolutionary history of D4h, we also searched this haplogroup in 15,460 ancient samples compiled by indo-european.eu (https://indo-european.eu/ancient-dna/),[30] thus covering virtually all global reported ancient mtDNA data, as well as additional 232 recently reported ancient mtDNA data from East Asia.[31,32] This survey yielded 39 ancient D4h samples (30 with the entire mitogenome and nine with HVS data) (Figure 1; Tables S4 and S5), which reflected the rarity of Dh4 in ancient times. Therefore, we integrated these ancient and contemporary data of this rare haplogroup to fully investigate its origin and expansion history.

RESULTS

Differentiation of D4h3 and D4h in Central and North China

To shed light on the origin of the NA founder D4h3a, we explored its ancestor D4h3. Our findings allowed an update of the D4h3 phylogeny and its branches (Figures 2A and S2). Specifically, to avoid any confusion, we kept the names of D4h3a and D4h3b and tentatively named their upstream nodes "pre-D4h3a" and "pre-D4h3b," respectively. Different from the NA founder D4h3a, the other branches of D4h3 are mainly distributed in China. In detail, D4h3b1 (root type in Hebei Province in North China) is found in North and Central China, while D4h3b2 (root type in Hubei Province) is mainly distributed in Central China. Coincidentally, among the reported ancient mtDNA data from different locations in Eurasia, we found three ancient samples belonging to D4h3 dated as early as 14–15 kilo years ago (kya) in the Amur River Valley (located in northern North China).[33] One of these mtDNAs, sample NE-5 (~14 kya), derives from pre-D4h3a and is phylogenetically the closest (six mutations apart; Figure S2) to the NA founder D4h3a mitogenome. The remaining two, samples NE34 (~14 kya) and NE-18 (~7 kya), are both members of pre-D4h3b. Overall these findings indicate that the ancestral homeland of D4h3 is most likely Central and North China and that both branches of D4h3 were there during the Paleolithic period. These branches locate in Central/North China and reflect the closest Asian matrilineal link to D4h3a, one of the founder pan-American mtDNA haplogroups.[25,26]

We then shifted our attention to haplogroup D4h, the most recent common ancestor of D4h3. Except for the NA D4h3a, the other D4h mtDNAs were predominantly found in China, mainly in North (48 out of 150 contemporary samples, discarding

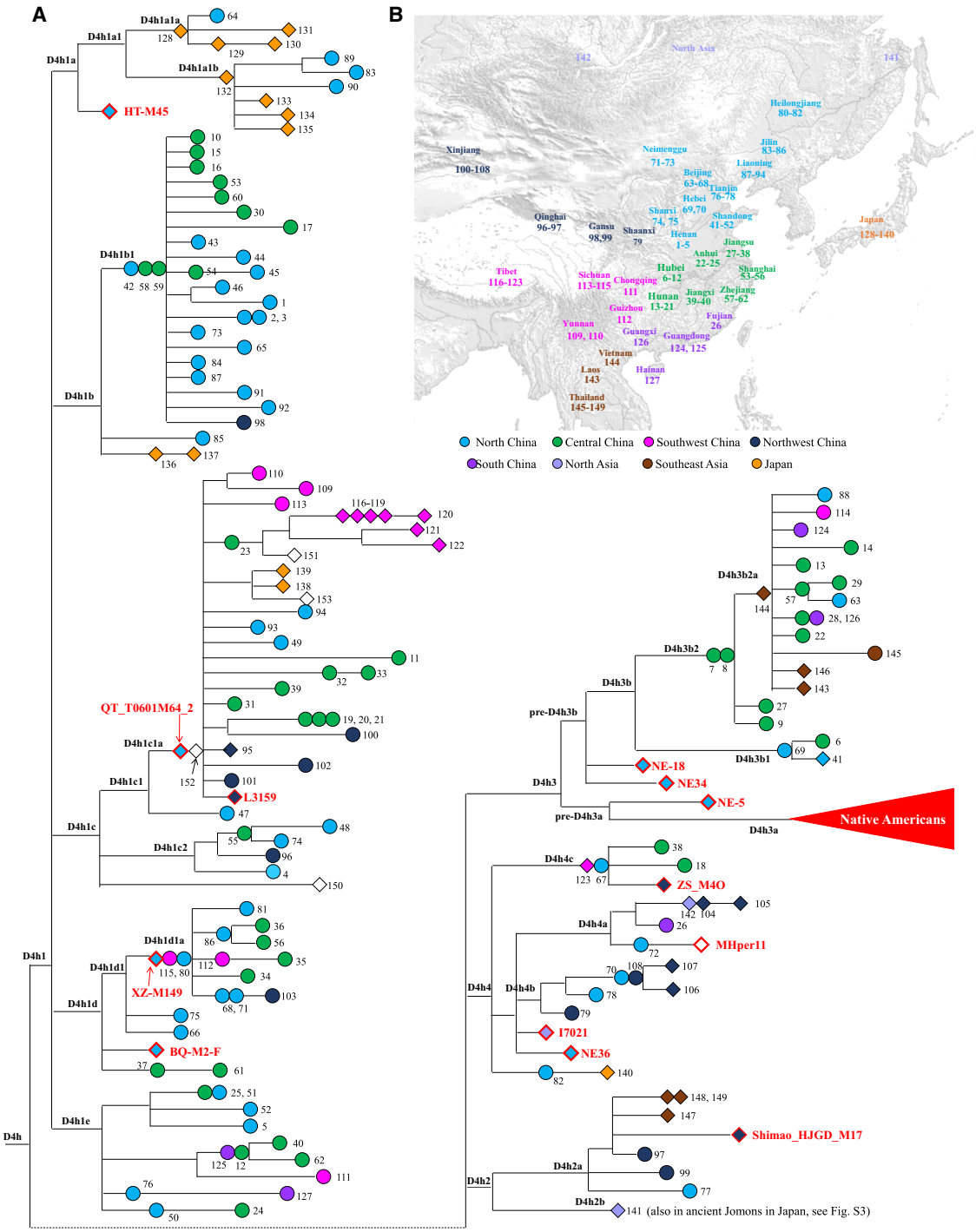


Figure 2. Phylogeography of haplogroup D4h and its sub-lineages

(A) Phylogenetic tree of D4h, with branch lengths proportional to number of variants. Circles: mitogenomes from this study; diamonds: previously published mitogenomes; black outlines: present-day samples; red outlines: ancient samples. The different colors, consistent with those in (B), refer to the different geographic source regions.

(B) Geographic sources of D4h mitogenomes in (A). Numbers on the map refer to the codes of samples and correspond to those in Figure 2A and Table S3.

four with unknown geographic information) and Central China (44 out of 150) (Table S4; Figure S3). A relatively small number of D4h mtDNAs were also identified in Northwest China (n = 14), Southwest China (n = 16), South China (n = 5), Japan (n = 13), Southeast Asia (n = 7), and North Asia (n = 2) (Table S4; Figure S3). Interestingly, the majority of the ancient D4h samples were detected in the northern regions of China (Figure 1), supporting a similar D4h distribution in the past. Further phylogeographic analyses revealed that the ancient and current samples from the same geographic region tend to cluster together in the same sub-branch, e.g., D4h1a, D4h1c, D4h1d, and D4h3. Meanwhile, most sub-haplogroups of D4h are predominant in North/Central China (i.e., D4h1b, D4h1d, D4h1e, and pre-D4h3b) or showed connections between North/Central China and other regions, including western China (D4h1c and D4h4), Japan (D4h1a), North Asia and Japan (D4h2), and even the Americas (D4h3a) (Figures 2, 3A, and S4). Moreover, samples from South China, Southwest China, Northwest China, Southeast Asia, and North Asia were sporadically distributed across the whole D4h haplogroup and primarily located on the terminal branches (Figures 2A and S2), most likely as a result of gene flow. Finally, the peculiar distributions of certain lineages, for instance D4h1a in Japan and D4h1c in Southwest China (Figures 2A, 3A, and S2), likely indicate founder events.

Given that some of the mitogenome data from literature are from phylogenetic rather than population studies, and given the relative scarcity of mitogenomes from Siberia, which will introduce bias to the phylogeographic analyses, we also collected and analyzed mtDNA HVS data from population studies (Table S1). Only few potential D4h samples were found in North Asian samples (n = 4,176) (for example, two belonging to D4h1d, which is defined by T16172C and C16174T, and one belonging to D4h1e, which is defined by C16174T and A16343G) (Figure S4), lending support to its rarity in North Asia. The median-joining network based on HVS data (Figure S4) revealed instead a much wider distribution range of D4h in Asia. Indeed, the majority of Asian D4h mtDNAs are observed in Central (58/228; 25.43%) and North (44/228; 21.05%) China, followed by Southwest China (35/228; 15.35%), Northwest China (15/228; 6.57%), Japan (29/228; 12.72%), Southeast Asia (11/228; 4.82%), South China (6/228; 2.63%), and North Asia (9/228; 3.94%). Moreover, the root types of the major branches, e.g., D4h1b, D4h1c, D4h1d, D4h1e, and D4h3b, are primarily found in Central and North China, while the terminal branches mainly contain samples from other regions, e.g., Southwest China, Northwest China, Southeast Asia, South Asia, and Central Asia. Finally, D4h1a and D4h2 are restricted to Japan and its surroundings, lending support to the founder events.

Taken together, these results indicate that the phylogenetic differentiation of D4h occurred somewhere in Central or North China, most likely in a region geographically close to the northern coast of China. In fact, among the North/Central China samples, more than half (64/92, 69.57%) were found in provinces along (Hebei, Liaoning, Tianjin, Shandong, Jiangsu, Shanghai, and Zhejiang) or near (Heilongjiang, Jilin, Beijing, Anhui, and Jiangxi) the northern coast of China (Table S4). Therefore, we propose that the northern coast of China might have played a critical role in the divergence and spread of D4h and its sub-haplogroups.

Two radiation events from northern coastal China contributed to NA and Japanese gene pools

Coalescent age estimations, updated by calibrated radiocarbon dates of ancient DNA samples using tip dating in BEAST (Tables 1 and S6), indicate that the radiations of D4h lineages (aged 32.39 kya, 95% highest probability density [HPD], 24.04–41.45 kya) occurred mainly within two time periods (Figure 3B). The first period fell within the Last Glacial Maximum (LGM; 26.5–19.0 kya),[34] during which D4h3 (26.39 kya, 95% HPD, 20.19–33.21 kya), pre-D4h3a (22.29 kya, 95% HPD, 17.24–27.68 kya), pre-D4h3b (21.55 kya, 95% HPD, 16.18–27.94 kya), D4h1 (21.83 kya, 95% HPD, 15.56–29.08 kya), and D4h2 (20.05 kya, 95% HPD, 12.12–29.48 kya) (Tables 1 and S6) differentiated into separate sub-haplogroups (Figure 3B). Among these sub-haplogroups, D4h3a further dispersed and became one of the pan-American haplogroup of NAs (Figure 4A). This radiation echoes well with the divergence of basal American branches from ancient eastern Asians 23–20 kya,[3] which was likely due to the LGM's inhospitable climate in the northern regions of Asia.[35] During the last deglaciation (19.0–11.5 kya), after the LGM, a second radiation of D4h occurred somewhere near the northern coast of China, as documented by D4h4 (18.11 kya, 95% HPD, 12.67–24.28 kya), D4h1c (16.17 kya, 95% HPD, 10.66–22.36 kya), D4h1a (15.59 kya, 95% HPD, 11.43–20.92 kya), D4h3b (13.22 kya, 95% HPD, 7.55–19.93 kya), D4h1c1 (12.77 kya, 95% HPD, 8.21–17.79 kya), and D4h1e (12.10, 95% HPD, 7.16–17.50 kya) (Figure 4B). Concordant with this phylogenetic radiation, a rapid increase in the effective population size of D4h ~15 kya was observed in the extended Bayesian skyline plot (EBSP) (Figure 3C), probably due to the post-LGM climate improvement. These results uncover two waves of previously unknown population dispersals along the northern coast of China during the LGM and last deglaciation, which led to the origin and expansion of different D4h lineages (Figure 4). The regions around the Bohai, Huanghai, and East China Seas, which were still connected by land along the northern coast before the Holocene,[36] probably allowed these expansions to occur.

Intriguingly, two haplogroups, D4h1a1 (12.24 kya, 95% HPD, 6.72–15.87 kya) and D4h2 (20.05 kya, 95% HPD, 12.12–29.48 kya), exhibited prevalent distributions in the Japanese Archipelago (Figures 2, 3A, and S4), suggesting that the expansions from the northern coast of China also exerted an influence in Japan. The discovery of D4h1a in ancient samples dated ~11 kya from the Nenjiang River valley[38] further supports its advent in regions close to Japan at least 11 kya. Similarly, D4h2 has been observed in ancient Jomons,[39] who are considered the descendants of Paleolithic settlers in the Japanese Archipelago.[40] The median-joining network (Table S5; Figure S4) showed that one branch of D4h2 (namely D4h2a) in China and Southeast Asia, while the other (D4h2b) is distributed in Siberians as well as the Ainu population (indigenous Japanese, 3 of 50 samples) and ancient Jomons. This further supports a genetic contribution possibly from China to different populations including Southeast Asians and ancient Japanese. Therefore, probably both D4h1a1

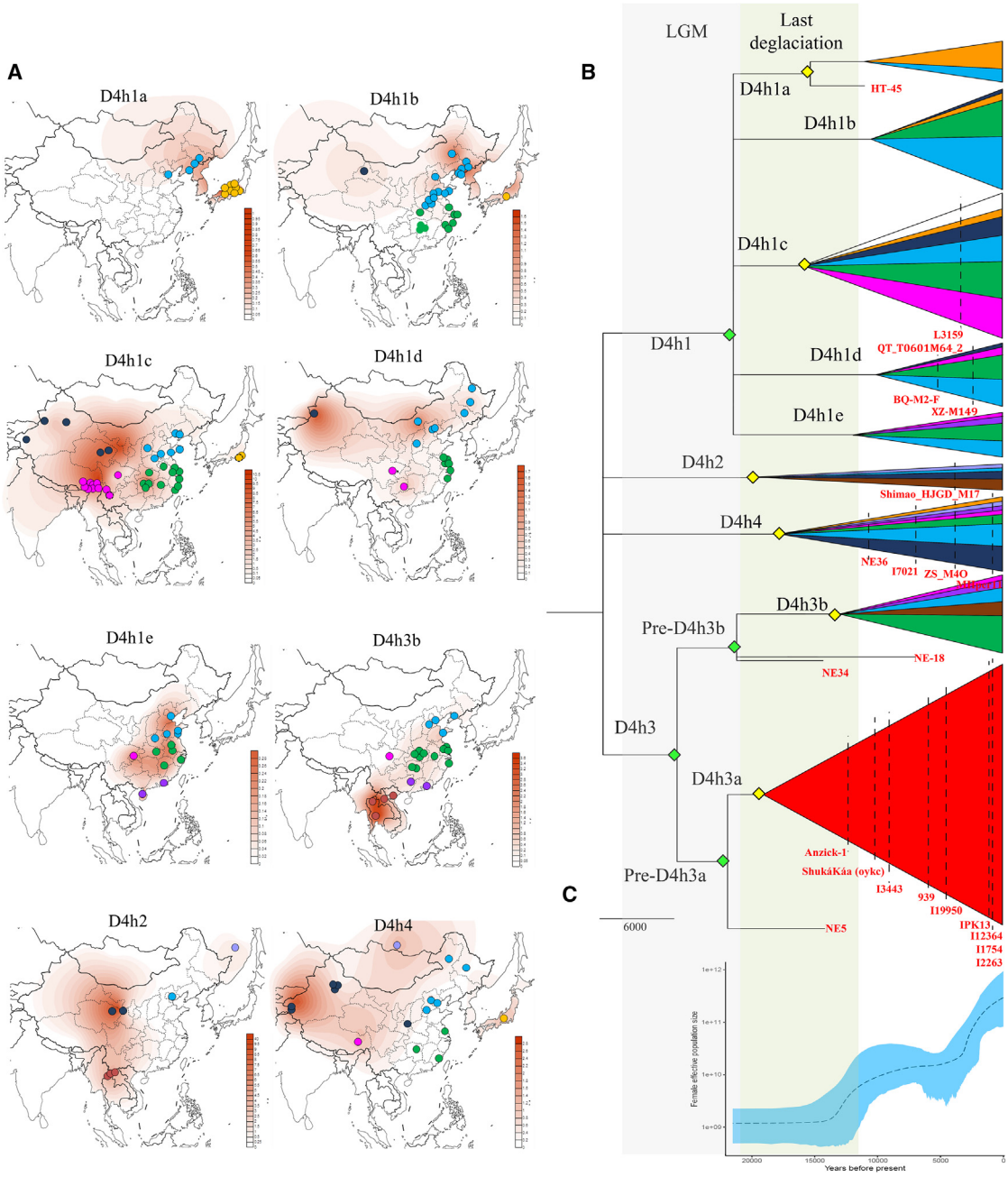


Figure 3. Geographic distribution and schematic tree of haplogroup D4h

(A) Geographic distributions of different branches of D4h. Each circle represents one sample, with geographic origin of samples shown by different colors, consistent with those in Figure 2B. Contour maps display spatial frequency distributions of haplogroups (see Table S7). Circles without outlines represent datasets from phylogenetic rather than population studies and thus were excluded in calculations of spatial frequencies.

(B) Bayesian age estimates using complete mitogenomes. Sizes of triangles are proportional to sub-haplogroup sample sizes. Colors represent different geographic regions, consistent with Figure 2B. Ancient samples are indicated in red. Green and yellow diamonds show the divergences within the LGM and the last deglaciation, respectively.

(C) Extended Bayesian skyline plot (EBSP) of D4h, showing effective population size changes through time.

and D4h2 dispersed from China to Japan after the LGM, possibly via the land bridges that connected China and the Japanese Archipelago until 12 kya.[41,42]

Potential supports from Y chromosome data

The origin of mtDNA D4h in northern coastal China of NAs echoes well also with the differentiation of Y chromosome

Table 1. Coalescent ages of D4h and its sublineages

Haplogroups/ sub-haplogroups	Number of mitogenomes[a]	Age (mean [95% HPD]) (kya)[b]
D4h	237	32.39 (24.04–41.45)
>D4h1	112	21.83 (15.56–29.08)
>>D4h1a	13	15.59 (11.43–20.92)
>>>D4h1a1	12	12.24 (6.72–15.87)
>>>>D4h1a1a	5	5.07 (1.83–8.56)
>>>>D4h1a1b	7	5.66 (2.87–8.87)
>>D4h1b	28	10.63 (6.26–15.53)
>>>D4h1b1	25	7.56 (4.41–11.06)
>>D4h1c	40	16.17 (10.66–22.36)
>>>D4h1c1	35	12.77 (8.21–17.79)
>>>>D4h1c1a	34	10.5 (7.01–14.61)
>>>D4h1c2	5	7.54 (3.43–12.46)
>>D4h1d	18	10.26 (6.47–14.26)
>>D4h1e	13	12.10 (7.16–17.50)
>D4h2	8	20.05 (12.12–29.48)
>>D4h2a	7	10.78 (6.57–15.46)
>D4h3	96	26.39 (20.19–33.21)
>>Pre-D4h3a	73	22.29 (17.24–27.68)
>>>D4h3a	71	19.40 (15.11–24.05)
>>Pre-D4h3b	24	21.55 (16.18–27.94)
>>>D4h3b	22	13.22 (7.55–19.93)
>>>>D4h3b1	3	1.93 (0.29–4.05)
>>>>D4h3b2	19	6.18 (3.31–9.56)
>D4h4	21	18.11 (12.67–24.28)

[a]Ancient mitogenome data were included in coalescent age estimations. Incomplete sequences were excluded from age estimations (see Table S4 for details).
[b]The mutation rate was recalibrated using the tip dating method. The best-fitting model was evaluated as previously described.[37]

haplogroup C2a-L1373 (ancestor to NA founder lineages C-MPB373 and C-P39) in low-latitude regions of northern Asia.[43] To further evaluate the potential radiation center of C2a-L1373, we assessed the frequencies of C2a-L1373 and its sub-lineages in different provinces of China based on Y chromosome genotyping data from 23Mofang Biotechnology Co., Ltd (totally 458,805 individuals, each with 33,000 Y chromosome SNPs genotyped). We detected the root type (C2a-L1373*) only in North China (including Beijing [0.020%], Tianjin [0.031%], Henan [0.004%], Heilongjiang [0.030%], Jilin [0.063%], Liaoning [0.071%], Shaanxi [0.035%]) and northwest China (Gansu [0.016%]; Table S8). It is worth underscoring that the highest C2a-L1373* frequencies were observed in Liaoning, Jilin, Heilongjiang, Tianjin, and Beijing (Table S8), which are all located close to northern coastal China. Moreover, the majority of other C2a-L1373 sub-lineages, including C-FGC28903, which is a sister branch of C-P39, harbor their highest frequencies in North China (Table S8). Moreover, samples belonging to C2a-L1373 in other places like South Asia, Central Asia, Europe, etc., were sporadically found or mainly occupied the terminal branches.[43] This evidence strongly suggests that C2a-L1373 differentiated in northern China, especially in the regions near the coast, similarly to mtDNA D4h.

In addition, two ancient samples from Songnen Plain in northern China, dated ~14,000 years ago, were found to belong to mtDNA D4h3 and Y chromosome C2a-L1373,[33] thus revealing the coexistence of both maternal and paternal ancestor lineages of NAs in northern coastal China. Interestingly, C2-M217 (~39.3 [34.7–44.5] kya)[22] and D4h (~32.39 [24.04–41.45] kya) had similar coalescent ages, and the divergence time of C2a-L1373 (about 21.6 [19.1–24.4] kya)[22] is similar to the time of the first D4h radiation estimated in this study, making it likely that an ancestral population from this region contributed to both the maternal and paternal gene pools of NAs. In fact, besides lineages of mtDNA D4h and Y chromosome C2-M217, substantial maternal and paternal lineages have also been observed in this region, e.g., Y chromosome lineages C-F1067[44] and mtDNA haplogroups A5, D4a, D4b, D4e, N9a, etc.,[29] most of which arose around the LGM.[44,45] This lends support to the scenario that this region was a differentiation center in East Asia after the LGM, which probably facilitated the expansions of different lineages including mtDNA D4h3 and Y chromosome C2a-L1373. Meanwhile, Y chromosome haplogroup C2-M217 has also been observed at a higher frequency in the Ainu (15%) than in other Japanese (3%).[46] Additionally, the coexistence of mtDNA D4h3 and Y chromosome C2 had also been reported in the same archaeological site in South America (~8 kya).[12] These observations collectively suggest that an ancestral population from northern China carrying mtDNA D4h and Y chromosome haplogroup C2 also spread into the Americas and the Japanese Archipelago.

DISCUSSION

In this study, by integrating ancient and contemporary mitogenomes of D4h from large-scale dataset covering virtually the whole of Eurasia, we traced the ancestry of one rare NA founder lineage (D4h3a) to a lower latitude region in northern coastal China around the Bohai and Huanghai Seas. This region is different from the geographical sources in Siberia hypothesized so far by the common maternal components, including mtDNA haplogroups A2, B2, C1, D4b1a2a1a, etc.[7,17,19] Our study thus uncovered an additional ancestral source for the ancestors of NAs beyond Siberia from the matrilineal perspective. This ancestry, although only contributed to a small proportion of the mtDNA gene pool of NAs (D4h3a),[25] would be important in complementing the whole picture of origination histories of early NAs.

Further support comes from the Y chromosome C2-M217, which harbors an age (~40 kya) that is similar to the one of mtDNA D4h and also probably radiated in northern coastal China during the LGM (as indicated by C2a-L1373), when the first radiation of D4h occurred. Interestingly, these uniparental ancestries echo well with the ancient ancestry in eastern Asia (~35 kya) that gave rise to East Asians, Siberians, and NAs at ~26 kya.[3] Meanwhile, it had also been inferred that about 40–23 kya, the ancestors of Jomons split from the ancient ancestry in eastern Asia.[47] This evidence strongly supports the existence of an old ancestry source, arising between 40 and 23 kya, that contributed to populations including East Asians, Jomons, Eastern Siberians, and

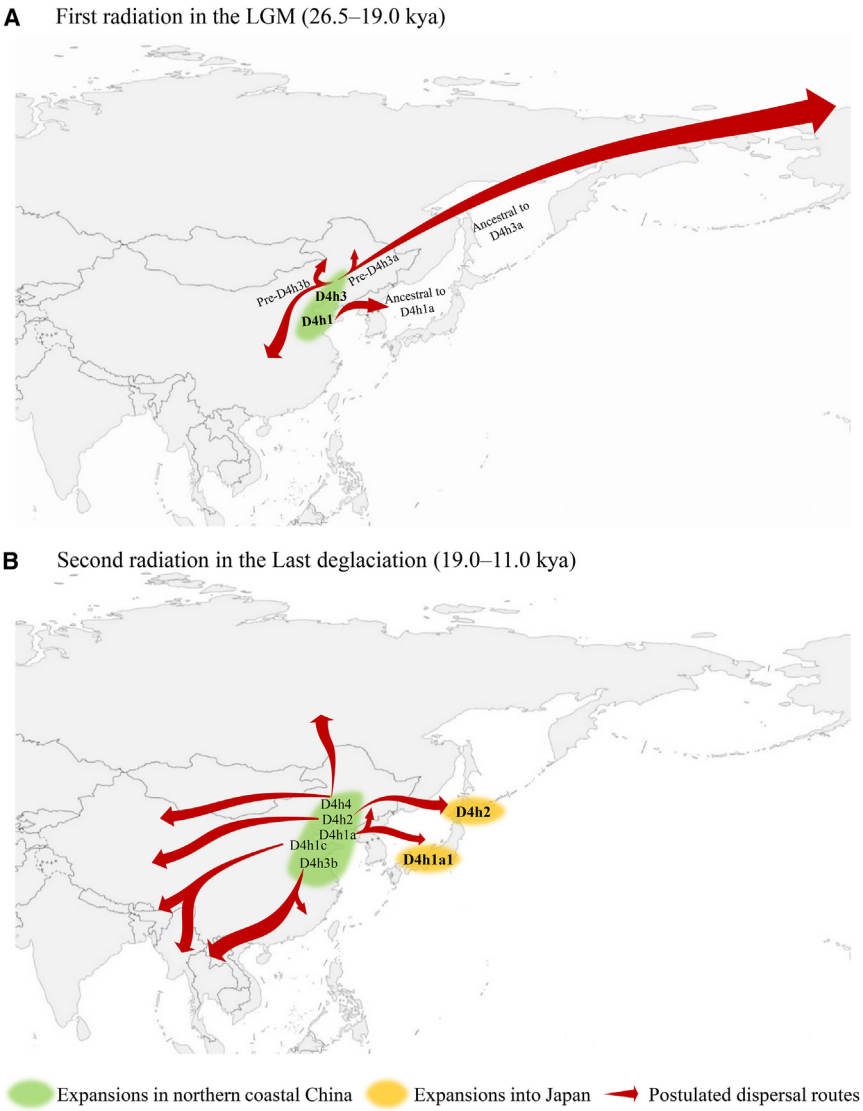


Figure 4. Two subsequent population radiations in the northern coastal regions of China contributed to NA and Japanese matrilineal ancestry

(A) The first radiation occurred during the LGM and involved D4h3, pre-D4h3b, and pre-D4h3a (from which D4h3a, typical of NAs, was derived).

(B) A later population expansion in the same general geographic area occurred in the deglaciation period and involved D4h1a1 and D4h2, whose derivatives are found in modern Japanese and ancient Jomons.

NAs (Figure S5). We propose that mtDNA D4h was one of the matrilineal lineages that witnessed these population splits and expansions. However, different from this East Asian ancestry contributing substantially to eastern Siberia,[47] D4h has been rarely found in this area. One explanation would be the loss of D4h during the expansions by genetic drift or matrilineal replacement.[48] More mtDNA data from Siberia will be of help to further assess this expansion process.

In addition, this genetic connection among China, the Americas, and Japan during the Pleistocene period parallels archaeological similarities, as early as Pleistocene period, among these regions. For instance, in the terminal Pleistocene period, Japanese microblades (18–17 kya), which exhibit similarities to those in Northeast Asia (including North China), display commonalities with contemporaneous stemmed points from incipient Jomon sites (~15.5 kya).[49] Importantly, stemmed points were well distributed around the Pacific rim from Japan to South America with close affinities with each other.[50] Recent findings on stemmed projectile points in North America (Cooper's Ferry site, ~15–16 kya) show closer affinity to the nonfluted projectile points in Japan than to those in North Asia.[51] We attribute this similarity in Paleolithic technology, as well as the phylogenetic relationships of D4h sub-lineages in China, the Americas, and Japan, to a probable Pleistocene connection among these regions (Figure 4B).

Our results also shed important light on the dispersal route of early NAs into the Americas. Given that mtDNA D4h radiated from northern coastal China, which is geographically close to Pacific coastal rim, we speculate that D4h would have documented LGM and post-LGM dispersals along the eastern Pacific coast. This echoes well with the dispersal D4h3a along the Pacific coastal path[25] when the ice-free corridor was closed.[11,52] Similarly, Y chromosome C-L1373, which probably radiated in parallel with mtDNA D4h, has also been reported in South Koreans (http://koreangenome.org/) and the Nivkh,[53] thus lending support to a coastal population expansion scenario initiated from northern coastal China. This, together with the Paleolithic cultural affinities along the Pacific, e.g., stemmed points,[51] and the palaeoecological feasibility of maritime dispersals (e.g., kelp highway hypothesis)[54] lends further support to the coastal route hypothesis of early NAs.[50,51,55]

Limitations of the study

By dissecting the origin of a rare NA founder lineage, our study revealed an ancestral root of both NAs and the Japanese in northern coastal China. However, some detailed expansions from this region into the Americas need to be further dissected. First, more data concerning mtDNA D4h from both ancient and contemporary samples are needed to elucidate the detailed expansion history of this lineage, especially from Siberia, where a relatively low number of mitogenomes have been assessed. Second, high-resolution Y chromosome data of C-L1373 from large-scale population dataset will help

to verify this radiation from paternal perspective. Third, investigations integrating mitogenomes, the Y chromosome, and autosomal genomes are also essential to explore whether there are differences between maternal, paternal, and autosomal markers and thus complement the whole picture of origination history of NAs.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Screening of D4h mitogenomes from dataset
  - DNA extraction, library construction, sequencing, and quality control
  - Ancient mtDNA acquisition
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Haplogroup affiliations and phylogenetic updates based on newly obtained and previously published data
  - Coalescent age estimations
  - Spatial geographic distribution
  - Extended Bayesian skyline plots
  - Median-joining network reconstruction

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2023.112413.

ACKNOWLEDGMENTS

We thank Dr. Qiaomei Fu for providing the ancient mtDNA sequencing data. This work was supported by the National Natural Science Foundation of China (31620103907, 81625013, 32170633, and 32150410355); the National Key R&D Program of China (2022YFC3302004); the Second Tibetan Plateau Scientific Expedition and Research (2019QZKK0607); the Strategic Priority Research Program (XDA20040102); Young Scientists in Basic Research (YSBR-076); the Key Research Program of Frontiers Science (QYZDB-SSW-SMC020); CAS "Light of West China" Program (Y.-C.L.) of the Chinese Academy of Sciences; the Digitalization, Development, and Application of Biotic Resource Program (202002AA100007); the Italian Ministry of Education, University and Research (MIUR) for Dipartimenti di Eccellenza Program (2018–2022) - Department of Biology and Biotechnology "L. Spallanzani," University of Pavia (A.A., A.T., and O.S.) and Progetti PRIN2017 20174BTC4R (A.A.); the High-level Talent Promotion and Training Project of Kunming (Spring City Plan; 2020SCP001); Yunling Scholar of the Yunnan Province (Q.-P.K.); the Yunnan Ten Thousand Talents Plan Young & Elite Talents Project (Y.-C.L.); and Yunnan Fundamental Research Projects (202201AW070012).

AUTHOR CONTRIBUTIONS

Q.-P.K. conceived the study; Z.-L.G., Y.-C.L., and Q.-P.K. performed the analyses; Z.-L.G., B.-Y.Y., and L.-Q.Y. performed the mitogenome sequencing; Y.-C.L., Z.-L.G., K.-J.L., J.-Y.T., and Z.U.R. collected the data; Y.-C.L., Z.-L.G., K.-J.L., Q.-P.K., S.-H.Z., C.-T.L., A.A., O.S., and A.T. contributed to interpretation of results; Y.-C.L. and Q.-P.K. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests. The authors from 23Mofang Biotechnology Co., Ltd, also declare no commercial or associative interests in connection with this work.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: July 28, 2022
Revised: January 5, 2023
Accepted: April 4, 2023
Published: May 9, 2023

REFERENCES

1. Brandini, S., Bergamaschi, P., Cerna, M.F., Gandini, F., Bastaroli, F., Bertolini, E., Cereda, C., Ferretti, L., Gómez-Carballa, A., Battaglia, V., et al. (2018). The Paleo-Indian entry into South America according to mitogenomes. Mol. Biol. Evol. *35*, 299–311. https://doi.org/10.1093/molbev/msx267.
2. Capodiferro, M.R., Aram, B., Raveane, A., Rambaldi Migliore, N., Colombo, G., Ongaro, L., Rivera, J., Mendizábal, T., Hernández-Mora, I., Tribaldos, M., et al. (2021). Archaeogenomic distinctiveness of the Isthmo-Colombian area. Cell *184*, 1706–1723.e24. https://doi.org/10.1016/j.cell.2021.02.040.
3. 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. https://doi.org/10.1038/nature25173.
4. Nakatsuka, N., Lazaridis, I., Barbieri, C., Skoglund, P., Rohland, N., Mallick, S., Posth, C., Harkins-Kinkaid, K., Ferry, M., Harney, É., et al. (2020). A paleogenomic reconstruction of the deep population history of the Andes. Cell *181*, 1131–1145.e21. https://doi.org/10.1016/j.cell.2020.04.015.
5. Potter, B.A., Baichtal, J.F., Beaudoin, A.B., Fehren-Schmitz, L., Haynes, C.V., Holliday, V.T., Holmes, C.E., Ives, J.W., Kelly, R.L., Llamas, B., et al. (2018). Current evidence allows multiple models for the peopling of the Americas. Sci. Adv. *4*, eaat5473. https://doi.org/10.1126/sciadv.aat5473.
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). Genomic evidence for the Pleistocene and recent population history of Native Americans. Science *349*, aab3884. https://doi.org/10.1126/science.aab3884.
7. 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. https://doi.org/10.1371/journal.pone.0000829.
8. Torroni, A., Schurr, T.G., Cabell, M.F., Brown, M.D., Neel, J.V., Larsen, M., Smith, D.G., Vullo, C.M., and Wallace, D.C. (1993). Asian affinities and continental radiation of the four founding Native American mtDNAs. Am. J. Hum. Genet. *53*, 563–590.
9. Yu, H., Spyrou, M.A., Karapetian, M., Shnaider, S., Radzevičiūtė, R., Nägele, K., Neumann, G.U., Penske, S., Zech, J., Lucas, M., et al. (2020). Paleolithic to bronze age Siberians reveal connections with first Americans and across Eurasia. Cell *181*, 1232–1245.e20. https://doi.org/10.1016/j.cell.2020.04.037.
10. Raghavan, M., Skoglund, P., Graf, K.E., Metspalu, M., Albrechtsen, A., Moltke, I., Rasmussen, S., Stafford, T.W., Jr., Orlando, L., Metspalu, E.,

et al. (2014). Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature *505*, 87–91. https://doi.org/10.1038/nature12736.

11. Lesnek, A.J., Briner, J.P., Lindqvist, C., Baichtal, J.F., and Heaton, T.H. (2018). Deglaciation of the Pacific coastal corridor directly preceded the human colonization of the Americas. Sci. Adv. *4*, eaar5040. https://doi.org/10.1126/sciadv.aar5040.

12. Posth, C., Nakatsuka, N., Lazaridis, I., Skoglund, P., Mallick, S., Lamnidis, T.C., Rohland, N., Nägele, K., Adamski, N., Bertolini, E., et al. (2018). Reconstructing the deep population history of Central and South America. Cell *175*, 1185–1197.e22. https://doi.org/10.1016/j.cell.2018.10.027.

13. Skoglund, P., and Reich, D. (2016). A genomic view of the peopling of the Americas. Curr. Opin. Genet. Dev. *41*, 27–35. https://doi.org/10.1016/j.gde.2016.06.016.

14. Zhang, X., Ji, X., Li, C., Yang, T., Huang, J., Zhao, Y., Wu, Y., Ma, S., Pang, Y., Huang, Y., et al. (2022). A late Pleistocene human genome from Southwest China. Curr. Biol. *32*, 3095–3109.e5. https://doi.org/10.1016/j.cub.2022.06.016.

15. Skoglund, P., Mallick, S., Bortolini, M.C., Chennagiri, N., Hünemeier, T., Petzl-Erler, M.L., Salzano, F.M., Patterson, N., and Reich, D. (2015). Genetic evidence for two founding populations of the Americas. Nature *525*, 104–108. https://doi.org/10.1038/nature14895.

16. Derenko, M., Malyarchuk, B., Grzybowski, T., Denisova, G., Dambueva, I., Perkova, M., Dorzhu, C., Luzina, F., Lee, H.K., Vanecek, T., et al. (2007). Phylogeographic analysis of mitochondrial DNA in northern Asian Populations. Am. J. Hum. Genet. *81*, 1025–1041. https://doi.org/10.1086/522933.

17. Starikovskaya, E.B., Sukernik, R.I., Derbeneva, O.A., Volodko, N.V., Ruiz-Pesini, E., Torroni, A., Brown, M.D., Lott, M.T., Hosseini, S.H., Huoponen, K., and Wallace, D.C. (2005). Mitochondrial DNA diversity in indigenous populations of the southern extent of Siberia, and the origins of Native American haplogroups. Ann. Hum. Genet. *69*, 67–89. https://doi.org/10.1046/j.1529-8817.2003.00127.x.

18. Volodko, N.V., Starikovskaya, E.B., Mazunin, I.O., Eltsov, N.P., Naidenko, P.V., Wallace, D.C., and Sukernik, R.I. (2008). Mitochondrial genome diversity in arctic Siberians, with particular reference to the evolutionary history of Beringia and Pleistocenic peopling of the Americas. Am. J. Hum. Genet. *82*, 1084–1100. https://doi.org/10.1016/j.ajhg.2008.03.019.

19. Dulik, M.C., Zhadanov, S.I., Osipova, L.P., Askapuli, A., Gau, L., Gokcumen, O., Rubinstein, S., and Schurr, T.G. (2012). Mitochondrial DNA and Y Chromosome variation provides evidence for a recent common ancestry between Native Americans and indigenous Altaians. Am. J. Hum. Genet. *90*, 229–246. https://doi.org/10.1016/j.ajhg.2011.12.014.

20. 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. (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. https://doi.org/10.1073/pnas.1118760109.

21. Grugni, V., Raveane, A., Ongaro, L., Battaglia, V., Trombetta, B., Colombo, G., Capodiferro, M.R., Olivieri, A., Achilli, A., Perego, U.A., et al. (2019). Analysis of the human Y-chromosome haplogroup Q characterizes ancient population movements in Eurasia and the Americas. BMC Biol. *17*, 3. https://doi.org/10.1186/s12915-018-0622-4.

22. Pinotti, T., Bergström, A., Geppert, M., Bawn, M., Ohasi, D., Shi, W., Lacerda, D.R., Solli, A., Norstedt, J., Reed, K., et al. (2019). Y chromosome sequences reveal a short Beringian standstill, rapid expansion, and early population structure of native American founders. Curr. Biol. *29*, 149–157.e3. https://doi.org/10.1016/j.cub.2018.11.029.

23. 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. https://doi.org/10.1093/molbev/msh009.

24. Colombo, G., Traverso, L., Mazzocchi, L., Grugni, V., Rambaldi Migliore, N., Capodiferro, M.R., Lombardo, G., Flores, R., Karmin, M., Rootsi, S., et al. (2022). Overview of the Americas' first peopling from a patrilineal perspective: New evidence from the Southern continent. Genes *13*, 220. https://doi.org/10.3390/genes13020220.

25. Perego, U.A., Achilli, A., Angerhofer, N., Accetturo, M., Pala, M., Olivieri, A., Hooshiar Kashani, B., Ritchie, K.H., Scozzari, R., Kong, Q.-P., et al. (2009). Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. Curr. Biol. *19*, 1–8. https://doi.org/10.1016/j.cub.2008.11.058.

26. Perego, U.A., Angerhofer, N., Pala, M., Olivieri, A., Lancioni, H., Hooshiar Kashani, B., Carossa, V., Ekins, J.E., Gómez-Carballa, A., Huber, G., et al. (2010). The initial peopling of the Americas: a growing number of founding mitochondrial genomes from Beringia. Genome Res. *20*, 1174–1179. https://doi.org/10.1101/gr.109231.110.

27. Kutanan, W., Kampuansai, J., Srikummool, M., Kangwanpong, D., Ghirotto, S., Brunelli, A., and Stoneking, M. (2017). Complete mitochondrial genomes of Thai and Lao populations indicate an ancient origin of Austroasiatic groups and demic diffusion in the spread of Tai–Kadai languages. Hum. Genet. *136*, 85–98. https://doi.org/10.1007/s00439-016-1742-y.

28. Kutanan, W., Kampuansai, J., Changmai, P., Flegontov, P., Schröder, R., Macholdt, E., Hübner, A., Kangwanpong, D., and Stoneking, M. (2018). Contrasting maternal and paternal genetic variation of hunter-gatherer groups in Thailand. Sci. Rep. *8*, 1536. https://doi.org/10.1038/s41598-018-20020-0.

29. Li, Y.C., Ye, W.J., Jiang, C.G., Zeng, Z., Tian, J.Y., Yang, L.Q., Liu, K.J., and Kong, Q.P. (2019). River valleys shaped the maternal genetic landscape of Han Chinese. Mol. Biol. Evol. *36*, 1643–1652. https://doi.org/10.1093/molbev/msz072.

30. Ancient Y-DNA and mtDNA. version number: all-ancient-dna-2-07-73. indo-european.eu, https://indo-european.eu/ancient-dna/.

31. Xue, J., Wang, W., Shao, J., Dai, X., Sun, Z., Gardner, J.D., Chen, L., Guo, X., Di, N., Pei, X., et al. (2022). Ancient mitogenomes reveal the origins and genetic structure of the Neolithic Shimao population in Northern China. Front. Genet. *13*, 909267. https://doi.org/10.3389/fgene.2022.909267.

32. Miao, B., Liu, Y., Gu, W., Wei, Q., Wu, Q., Wang, W., Zhang, M., Ding, M., Wang, T., Liu, J., et al. (2021). Maternal genetic structure of a neolithic population of the Yangshao culture. J. Genet. Genomics. *48*, 746–750. https://doi.org/10.1016/j.jgg.2021.04.005.

33. Mao, X., Zhang, H., Qiao, S., Liu, Y., Chang, F., Xie, P., Zhang, M., Wang, T., Li, M., Cao, P., et al. (2021). The deep population history of northern East Asia from the late Pleistocene to the Holocene. Cell *184*, 3256–3266.e13. https://doi.org/10.1016/j.cell.2021.04.040.

34. Clark, P.U., Dyke, A.S., Shakun, J.D., Carlson, A.E., Clark, J., Wohlfarth, B., Mitrovica, J.X., Hostetler, S.W., and McCabe, A.M. (2009). The Last Glacial Maximum. Science *325*, 710–714. https://doi.org/10.1126/science.1172873.

35. Willerslev, E., and Meltzer, D.J. (2021). Peopling of the Americas as inferred from ancient genomics. Nature *594*, 356–364. https://doi.org/10.1038/s41586-021-03499-y.

36. Li, G., Li, P., Liu, Y., Qiao, L., Ma, Y., Xu, J., and Yang, Z. (2014). Sedimentary system response to the global sea level change in the East China Seas since the Last Glacial Maximum. Earth Sci. Rev. *139*, 390–405. https://doi.org/10.1016/j.earscirev.2014.09.007.

37. Fu, Q., Mittnik, A., Johnson, P.L.F., Bos, K., Lari, M., Bollongino, R., Sun, C., Giemsch, L., Schmitz, R., Burger, J., et al. (2013). A revised timescale for human evolution based on ancient mitochondrial genomes. Curr. Biol. *23*, 553–559. https://doi.org/10.1016/j.cub.2013.02.044.

38. Ning, C., Fernandes, D., Changmai, P., Flegontova, O., Yüncü, E., Maier, B., Altınışık, E., Kassian, L., Krause, J., Lalueza, C., Manica, A., Potter, B., Robbeets, M., Sirak, K., Siska, V., Vajda, E.J., Vyazov, L.A., Wang, K., Wang, L.X., Wu, X.Y., Xiao, X.M., Zhang, F., Reich, D., Schiffels, S., Pinhasi, R., Cui, Y.Q., and Flegontov, P. (2020). The genomic formation of First

American ancestors in East and Northeast Asia. bioRxiv. https://doi.org/10.1101/2020.10.12.336628.

39. Adachi, N., Shinoda, K.I., Umetsu, K., Kitano, T., Matsumura, H., Fujiyama, R., Sawada, J., and Tanaka, M. (2011). Mitochondrial DNA Analysis of Hokkaido Jomon skeletons: Remnants of archaic maternal lineages at the southwestern edge of former Beringia. Am. J. Phys. Anthropol. *146*, 346–360. https://doi.org/10.1002/ajpa.21561.

40. Masami, I., and Yōsuke, K. (2015). The appearance and characteristics of the early Upper Paleolithic in the Japanese Archipelago. In Emergence and Diversity of Modern Human Behavior in Paleolithic Asia, K. Yōsuke, I. Masami, G. Ted, S. Hiroyuki, and O. Akira, eds. (Texas A&M University Press), pp. 289–313.

41. Lambeck, K., Rouby, H., Purcell, A., Sun, Y., and Sambridge, M. (2014). Sea level and global ice volumes from the Last Glacial Maximum to the Holocene. Proc. Natl. Acad. Sci. USA *111*, 15296–15303. https://doi.org/10.1073/pnas.1411762111.

42. Xiang, R., Sun, Y., Li, T., Oppo, D.W., Chen, M., and Zheng, F. (2007). Paleoenvironmental change in the middle Okinawa Trough since the last deglaciation: Evidence from the sedimentation rate and planktonic foraminiferal record. Palaeogeogr. Palaeoclimatol. Palaeoecol. *243*, 378–393. https://doi.org/10.1016/j.palaeo.2006.08.016.

43. Sun, J., Ma, P.C., Cheng, H.Z., Wang, C.Z., Li, Y.L., Cui, Y.Q., Yao, H.B., Wen, S.Q., and Wei, L.H. (2021). Post-Last Glacial Maximum expansion of Y-chromosome haplogroup C2a-L1373 in northern Asia and its implications for the origin of Native Americans. Am. J. Phys. Anthropol. *174*, 363–374. https://doi.org/10.1002/ajpa.24173.

44. Wu, Q., Cheng, H.Z., Sun, N., Ma, P.-C., Sun, J., Yao, H.-B., Xie, Y.-M., Li, Y.-L., Meng, S.-L., Zhabagin, M., et al. (2020). Phylogenetic analysis of the Y-chromosome haplogroup C2b-F1067, a dominant paternal lineage in Eastern Eurasia. J. Hum. Genet. *65*, 823–829. https://doi.org/10.1038/s10038-020-0775-1.

45. Behar, D.M., van Oven, M., Rosset, S., Metspalu, M., Loogväli, E.L., Silva, N.M., Kivisild, T., Torroni, A., and Villems, R. (2012). A "Copernican" reassessment of the human mitochondrial DNA tree from its root. Am. J. Hum. Genet. *90*, 675–684. https://doi.org/10.1016/j.ajhg.2012.03.002.

46. 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. https://doi.org/10.1371/journal.pgen.1003460.

47. Gakuhari, T., Nakagome, S., Rasmussen, S., Allentoft, M.E., Sato, T., Korneliussen, T., Chuinneagáin, B.N., Matsumae, H., Koganebuchi, K., Schmidt, R., et al. (2020). Ancient Jomon genome sequence analysis sheds light on migration patterns of early East Asian populations. Commun. Biol. *3*, 437. https://doi.org/10.1038/s42003-020-01162-2.

48. Sikora, M., Pitulko, V.V., Sousa, V.C., Allentoft, M.E., Vinner, L., Rasmussen, S., Margaryan, A., de Barros Damgaard, P., de la Fuente, C., Renaud, G., et al. (2019). The population history of northeastern Siberia since the Pleistocene. Nature *570*, 182–188. https://doi.org/10.1038/s41586-019-1279-z.

49. Takakura, J. (2020). Rethinking the disappearance of microblade technology in the Terminal Pleistocene of Hokkaido, Northern Japan: Looking at archaeological and palaeoenvironmental evidence. Quaternary *3*, 21. https://doi.org/10.3390/quat3030021.

50. Erlandson, J.M., and Braje, T.J. (2011). From Asia to the Americas by boat? Paleogeography, paleoecology, and stemmed points of the northwest Pacific. Quat. Int. *239*, 28–37. https://doi.org/10.1016/j.quaint.2011.02.030.

51. Davis, L.G., Madsen, D.B., Becerra-Valdivia, L., Higham, T., Sisson, D.A., Skinner, S.M., Stueber, D., Nyers, A.J., Keen-Zebert, A., Neudorf, C., et al. (2019). Late Upper Paleolithic occupation at Cooper's Ferry, Idaho, USA, ~16,000 years ago. Science *365*, 891–897. https://doi.org/10.1126/science.aax9830.

52. Pedersen, M.W., Ruter, A., Schweger, C., Friebe, H., Staff, R.A., Kjeldsen, K.K., Mendoza, M.L.Z., 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. https://doi.org/10.1038/nature19085.

53. Raghavan, M., Skoglund, P., Graf, K.E., Metspalu, M., Albrechtsen, A., Moltke, I., Rasmussen, S., Stafford, T.W., Jr., Orlando, L., Metspalu, E., et al. (2014). Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature *505*, 87–91. https://doi.org/10.1038/nature12736.

54. Erlandson, J.M., Graham, M.H., Bourque, B.J., Corbett, D., Estes, J.A., and Steneck, R.S. (2007). The kelp highway hypothesis: marine ecology, the coastal migration theory, and the peopling of the Americas. J. Island Coast. Archaeol. *2*, 161–174. https://doi.org/10.1080/15564890701628612.

55. 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. https://doi.org/10.1073/pnas.1620410114.

56. 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. https://doi.org/10.1038/nature13025.

57. Cui, Y., Lindo, J., Hughes, C.E., Johnson, J.W., Hernandez, A.G., Kemp, B.M., Ma, J., Cunningham, R., Petzelt, B., Mitchell, J., et al. (2013). Ancient DNA analysis of Mid-Holocene individuals from the Northwest coast of North America reveals different evolutionary paths for mitogenomes. PLoS One *8*, e66948. https://doi.org/10.1371/journal.pone.0066948.

58. de la Fuente, C., Ávila-Arcos, M.C., Galimany, J., Carpenter, M.L., Homburger, J.R., Blanco, A., Contreras, P., Cruz Dávalos, D., Reyes, O., San Roman, M., et al. (2018). Genomic insights into the origin and diversification of late maritime hunter-gatherers from the Chilean Patagonia. Proc. Natl. Acad. Sci. USA *115*, E4006–E4012. https://doi.org/10.1073/pnas.1715688115.

59. Nakatsuka, N., Luisi, P., Motti, J.M.B., Salemme, M., Santiago, F., D'Angelo del Campo, M.D., Vecchi, R.J., Espinosa-Parrilla, Y., Prieto, A., Adamski, N., et al. (2020). Ancient genomes in South Patagonia reveal population movements associated with technological shifts and geography. Nat. Commun. *11*, 3868. https://doi.org/10.1038/s41467-020-17656-w.

60. Kennett, D.J., Lipson, M., Prufer, K.M., Mora-Marín, D., George, R.J., Rohland, N., Robinson, M., Trask, W.R., Edgar, H.H.J., Hill, E.C., et al. (2022). South-to-north migration preceded the advent of intensive farming in the Maya region. Nat. Commun. *13*, 1530. https://doi.org/10.1038/s41467-022-29158-y.

61. Maár, K., Varga, G.I.B., Kovács, B., Schütz, O., Maróti, Z., Kalmár, T., Nyerki, E., Nagy, I., Latinovics, D., Tihanyi, B., et al. (2021). Maternal lineages from 10–11th century commoner cemeteries of the Carpathian basin. Genes *12*, 460. https://doi.org/10.3390/genes12030460.

62. Liu, J., Zeng, W., Sun, B., Mao, X., Zhao, Y., Wang, F., Li, Z., Luan, F., Guo, J., Zhu, C., et al. (2021). Maternal genetic structure in ancient Shandong between 9500 and 1800 years ago. Sci. Bull. *66*, 1129–1135. https://doi.org/10.1016/j.scib.2021.01.029.

63. Wang, C.C., Yeh, H.Y., Popov, A.N., Zhang, H.Q., Matsumura, H., Sirak, K., Cheronet, O., Kovalev, A., Rohland, N., Kim, A.M., et al. (2021). Genomic insights into the formation of human populations in East Asia. Nature *591*, 413–419. https://doi.org/10.1038/s41586-021-03336-2.

64. Ding, M., Wang, T., Ko, A.M.-S., Chen, H., Wang, H., Dong, G., Lu, H., He, W., Wangdue, S., Yuan, H., et al. (2020). Ancient mitogenomes show plateau populations from last 5200 years partially contributed to present-day Tibetans. Proc. Biol. Sci. *287*, 20192968. https://doi.org/10.1098/rspb.2019.2968.

65. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. *17*, 10–12. https://doi.org/10.14806/ej.17.1.200.

66. Renaud, G., Stenzel, U., and Kelso, J. (2014). leeHom: Adaptor trimming and merging for Illumina sequencing reads. Nucleic Acids Res. *42*, e141. https://doi.org/10.1093/nar/gku699.

67. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

68. 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. https://doi.org/10.1093/bioinformatics/btp352.

69. Renaud, G., Slon, V., Duggan, A.T., and Kelso, J. (2015). Schmutzi: Estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. Genome Biol. *16*, 224. https://doi.org/10.1186/s13059-015-0776-0.

70. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303. https://doi.org/10.1101/gr.107524.110.

71. Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., et al. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. *15*, e1006650. https://doi.org/10.1371/journal.pcbi.1006650.

72. Bouckaert, R.R., and Drummond, A.J. (2017). bModelTest: Bayesian phylogenetic site model averaging and model comparison. BMC Evol. Biol. *17*, 42. https://doi.org/10.1186/s12862-017-0890-6.

73. Russel, P.M., Brewer, B.J., Klaere, S., and Bouckaert, R.R. (2019). Model selection and parameter inference in phylogenetics using nested sampling. Syst. Biol. *68*, 219–233. https://doi.org/10.1093/sysbio/syy050.

74. Rambaut, A., Drummond, A.J., Xie, D., Baele, G., and Suchard, M.A. (2018). Posterior summarization in Bayesian phylogenetics using tracer 1.7. Syst. Biol. *67*, 901–904. https://doi.org/10.1093/sysbio/syy032.

75. Kong, Q.P., Salas, A., Sun, C., Fuku, N., Tanaka, M., Zhong, L., Wang, C.Y., Yao, Y.G., and Bandelt, H.J. (2008). Distilling artificial recombinants from large sets of complete mtDNA genomes. PLoS One *3*, e3016. https://doi.org/10.1371/journal.pone.0003016.

76. Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M., and Howell, N. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat. Genet. *23*, 147. https://doi.org/10.1038/13779.

77. Lipson, M., Cheronet, O., Mallick, S., Rohland, N., Oxenham, M., Pietrusewsky, M., Pryce, T.O., Willis, A., Matsumura, H., Buckley, H., et al. (2018). Ancient genomes document multiple waves of migration in Southeast Asian prehistory. Science *361*, 92–95. https://doi.org/10.1126/science.aat3188.

78. van Oven, M., and Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum. Mutat. *30*, E386–E394. https://doi.org/10.1002/humu.20921.

79. Edgar, R.C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797. https://doi.org/10.1093/nar/gkh340.

80. Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. *30*, 2725–2729. https://doi.org/10.1093/molbev/mst197.

81. 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. https://doi.org/10.1126/sciadv.1501385.

82. Rieux, A., and Khatchikian, C.E. (2017). tipdatingbeast: An r package to assist the implementation of phylogenetic tip-dating tests using beast. Mol. Ecol. Resour. *17*, 608–613. https://doi.org/10.1111/1755-0998.12603.

83. Forster, P., Harding, R., Torroni, A., and Bandelt, H.J. (1996). Origin and evolution of Native American mtDNA variation: A reappraisal. Am. J. Hum. Genet. *59*, 935–945.

84. Saillard, J., Forster, P., Lynnerup, N., Bandelt, H.J., and Nørby, S. (2000). mtDNA Variation among Greenland Eskimos: The edge of the Beringian expansion. Am. J. Hum. Genet. *67*, 718–726. https://doi.org/10.1086/303038.

85. Macaulay, V., Soares, P., and Richards, M.B. (2019). Rectifying long-standing misconceptions about the ρ statistic for molecular dating. PLoS One *14*, e0212311. https://doi.org/10.1371/journal.pone.0212311.

86. Heled, J., and Drummond, A.J. (2010). Bayesian inference of species trees from multilocus data. Mol. Biol. Evol. *27*, 570–580. https://doi.org/10.1093/molbev/msp274.

87. Atkinson, Q.D., Gray, R.D., and Drummond, A.J. (2008). mtDNA variation predicts population size in humans and reveals a major Southern Asian chapter in human prehistory. Mol. Biol. Evol. *25*, 468–474. https://doi.org/10.1093/molbev/msm277.

88. Atkinson, Q.D., Gray, R.D., and Drummond, A.J. (2009). Bayesian coalescent inference of major human mitochondrial DNA haplogroup expansions in Africa. Proc. Biol. Sci. *276*, 367–373. https://doi.org/10.1098/rspb.2008.0785.

89. Bandelt, H.J., Forster, P., and Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. *16*, 37–48. https://doi.org/10.1093/oxfordjournals.molbev.a026036.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
106 whole mitochondrial genomes of D4h samples	This paper	https://bigd.big.ac.cn/gwh/; accession number: PRJCA006291
Anzick-1 ancient mitochondrial genome	Rasmussen et al.[56]	N/A
Shuká Káa ancient mitochondrial genome	Lindo et al.[55]	N/A
Ancient 939 ancient mitochondrial genome	Cui et al.[57]	GenBank:KC998701.1
IPK13 ancient mitochondrial genome	de la Fuente et al.[58]	N/A
I12364 ancient mitochondrial genome	Nakatsuka et al.[59]	https://reichdata.hms.harvard.edu/pub/datasets/amh_repo/curated_releases/V52/V52.2/SHARE/public.dir/index_v52.2_MT.html
I1754 and I3443 ancient mitochondrial genomes	Posth et al.[12]	https://reichdata.hms.harvard.edu/pub/datasets/amh_repo/curated_releases/V52/V52.2/SHARE/public.dir/index_v52.2_MT.html
I19950 ancient mitochondrial genome	Kennett et al.[60]	https://reichdata.hms.harvard.edu/pub/datasets/amh_repo/curated_releases/V52/V52.2/SHARE/public.dir/index_v52.2_MT.html
I2263 ancient mitochondrial genome	Nakatsuka et al.[4]	https://reichdata.hms.harvard.edu/pub/datasets/amh_repo/curated_releases/V52/V52.2/SHARE/public.dir/index_v52.2_MT.html
MHper11 ancient mitochondrial genome	Maár et al.[61]	N/A
NE34,NE-5,NE36,NE-18 ancient mitochondrial genomes	Mao et al.[33]	NGDC:PRJCA003699 https://bigd.big.ac.cn/gwh/
BQ-M2-F and XZ-M149 ancient mitochondrial genomes	Liu et al.[62]	NGDC:PRJCA002947 https://bigd.big.ac.cn/gwh/
I7021 ancient mitochondrial genome	Wang et al.[63]	https://reichdata.hms.harvard.edu/pub/datasets/amh_repo/curated_releases/V52/V52.2/SHARE/public.dir/index_v52.2_MT.html
HT-M45 ancient mitochondrial genome	Ning et al.[38]	N/A
L3159 ancient mitochondrial genome	Ding et al.[64]	NGDC:PRJCA002243 https://bigd.big.ac.cn/gwh/
QT_T0601M64_2 ancient mitochondrial genome	Miao et al.[32]	NGDC:PRJCA004284 https://bigd.big.ac.cn/gwh/
Shimao_HJGD_M17 and ZS_M4O ancient mitochondrial genomes	Xue et al.[31]	NGDC:PRJCA009290 https://bigd.big.ac.cn/gwh/
Ancient Y-DNA and mtDNA	all-ancient-dna-2-07-73[30]	indo-european.eu (https://indo-european.eu/ancient-dna/)
Software and algorithms		
Cutadapt v1.16	Martin[65]	RRID:SCR_011841
leeHom v1.2.16	Renaud et al.[66]	RRID:SCR_002710
BWA v0.7.8	Li and Durbin[67]	RRID:SCR_010910
Samtools v1.13	Li et al.[68]	RRID:SCR_002105
schmutzi v1.5.6	Renaud et al.[69]	https://bioinf.eva.mpg.de/schmutzi/
GATK v4.2.2	McKenna et al.[70]	RRID:SCR_001876
BEAST 2.6.6	Bouckaert et al.[71]	RRID:SCR_010228
Surfer v8.0	N/A	https://www.goldensoftware.com/products/surfer
bmodeltest package	Bouckaert and Drummond[72]	https://taming-the-beast.org/
NS package	Russel et al.[73]	https://taming-the-beast.org/
Tracer v1.7.2	Rambaut et al.[74]	RRID:SCR_019121
FigTree v1.4.4	N/A	RRID:SCR_008515
R v4.1.2	R Core Team 2021	https://www.R-project.org/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Qing-Peng Kong (kongqp@mail.kiz.ac.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Complete mitogenome sequencing data (fasta format) of the 106 newly sequenced individuals have been deposited in the Genome Warehouse in the National Genomics Data Center (https://bigd.big.ac.cn/gwh/; NGDC: PRJCA006291). This paper does not report original code. Any additional information required to reanalyze the data reported in this work paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Complete mitogenomes of 106 samples belonging to D4h were sequenced in this study. The information of the subjects including the geographical origin, gender and age are shown in Table S3. The sample collection and experimental protocol were approved by the Ethics Committee at the Kunming Institute of Zoology, Chinese Academy of Sciences (Approval No. KIZRKX-2021-011), as well as by the Office of Human Genetic Resource Administration (OHGRA), The Ministry of Science and Technology (MOST), China (Approval No. 2022SLCJ0017). Informed consent was obtained from each individual before the study. For the genotyping data from 23 Mofang Inc., we only used Y chromosome data from customers who had signed the informed consents online to participate in this study and agree to share their genotyping information.

METHOD DETAILS

Screening of D4h mitogenomes from dataset

To unravel the evolutionary history of this haplogroup and the expansion of its sublineage pre-D4h3a into the Americas, we performed a search of D4h mtDNAs in a large-scale dataset: 101,319 Eurasian individuals, including 60,979 for which only hypervariable segment (HVS) data (Table S1) were available, and 40,340 samples with whole mtDNA (sequencing and genotyping) (Table S2). For the HVS and genotyping data, the motif-search strategy[75] was adopted to identify mtDNAs harboring diagnostic variants of D4h and its sublineages. This allowed the identification of 112 potential D4h mtDNAs (Table S3), which after complete mitogenome sequencing revealed 106 new Asian D4h mitogenomes (see below). These were added to 110 previously published D4h mitogenomes from contemporary populations and 30 D4h mitogenomes (Table S4) screened out from ancient samples for further phylogenetic analyses and coalescent age estimations.

DNA extraction, library construction, sequencing, and quality control

Total genomic DNA was isolated by using the genomic DNA extraction kit (Axygen). DNA yield and purity were measured via UV spectroscopy. Libraries were prepared with a standard library kit (MyGenostics Inc., Beijing, China). Sequencing was carried out using an Illumina HiSeq X Ten platform at MyGenostics, with sequencing depths ranging from 3 823.63× to 15 727.02× (average of 7 359.76×; Figure S1). The Cutadapt software[65] was used to trim adapters and to filter low quality sequences (including short reads, and reads with low mean quality score and many ambiguous (N) bases in fastq files. Reads were then aligned to the human reference genome version GRCH38 (which has the revised Cambridge Reference Sequence (rCRS)[76] as mtDNA reference) by "bwa mem" (v0.7.10) (http://bio-bwa.sourceforge.net/). Duplications were detected and removed using the MarkDuplicates module of GenomeAnalysisTK (GATK) and the GATK HaplotypeCaller module was employed to generate the variant file (vcf) using standard parameters. The final variants of each sample relative to the revised rCRS were recorded (Table S3).

Ancient mtDNA acquisition

Fasta files of ancient mtDNA sequences were downloaded from the literature or public database, with the exception of the mtDNA sequence of I4012,[77] which was extracted from the whole genome sequencing data of that sample. In detail, we downloaded raw fastqof I4012 from ENA (European Nucleotide Archive) and then trimed adapters using leeHom v1.2.16[66] and aligned to rCRS by using aln and samse commands of BWA v0.7.8[67] with parameters -n 0.01, -o 2, and -l 16500. Reads with mapping quality score (<30) were filtered by samtools v1.13.[68] Finally, we obtained endogenous mtDNA fasta file and allocated it to haplogroup D4h1c according to the variants. However, due to the contamination rate of about 0.99 (calculated by schmutzi v1.5.6),[69] we removed this sample in the subsequent analyses.

QUANTIFICATION AND STATISTICAL ANALYSIS

Haplogroup affiliations and phylogenetic updates based on newly obtained and previously published data

Haplogroup affiliation of each sample was carried out according to mtDNA tree Build 17 (http://phylotree.org/).[78] The phylogeny of D4h was reconstructed manually and checked using mtPhyl v5.003 (https://sites.google.com/site/mtphyl/). Many previously classified branches were confirmed, including D4h4, D4h1, D4h1a, and D4h1b, while others were updated, including D4h1c, D4h1c1, D4h1d, and pre-D4h3b. In addition, several novel branches were also defined, e.g., D4h3b1, D4h3b2, D4h1c2, and D4h1e (Figure S2).

Coalescent age estimations

Modern and ancient sequences (Table S4) were aligned using MUSCLE v3.8.3[79] in MEGA6.[80] Mutations including 309.1C(C), 315.1C(C), AC indels at 515–522, A16182C, A16183C, 16193.1C(C) and C16519T/T16519C were excluded in age estimations. bModelTest[72] package implemented in BEAST 2.6.6 was used to select the most appropriate substitution model for our data. As a result, TN93 model (121,131) with gamma rate heterogeneity (G) and proportion of invariant sites (I) was supported through visualization output in Tracer v1.7.2. The Bayes Factor (logBF = 15) computed by BEAST NS package (32 particles) indicated the strict clock model is suitable for our data than the uncorrelated lognormal relaxed clock model. The midpoints of calibrated radiocarbon dates or archaeological periods of the ancient samples (Table S4) were used as the tip date.[81] A date-randomization test[82] using BEAST 2.6.6 showed the clockRate parameter from the original dataset of 95% HPD intervals (highest posterior distribution) did not overlap the date-randomized datasets, indicating there was sufficient tip date signal to calibrate the clock rate. The Chain Monte Carlo (MCMC) runs of 100,000,000 steps were performed with a sampling of parameters every 10,000 steps and the initial 10% steps were discarded as burn-in. Coalescent Constant Population was adopted as tree Prior.[37] BEAUti within the package of BEAST was used to set the model and parameters. The convergences of MCMC were evaluated according to the effective sample size (ESS) by Tracer v1.7.2 (with ESS >200 as acceptable). As a result, whole mitogenomes without partitions into codon positions were adopted due to general higher ESS values (with only two ESS values between 100 and 200). The 95% HPD intervals of coalescent age estimates were recorded in FigTree v1.4.4.

In addition, Rho (ρ) statistics,[83,84] which provides unbiased and overlapping estimates of coalescent ages,[85] was also used to evaluate the coalescent ages of each clade in haplogroup D4h (Table S6).

Spatial geographic distribution

Geographic locations of mtDNAs belonging to D4h and its sublineages were plotted using Surfer v8.0 (Golden Software Inc. Golden, Colorado, USA). Contour maps of spatial frequencies were constructed using the Kriging algorithm in Surfer v8.0. Samples nonderiving from population studies were excluded.

Extended Bayesian skyline plots

An Extended Bayesian Skyline plot (EBSP)[86] for effective population size (Nef) through time was reconstructed using BEAST v2.6.6,[71] as described elsewhere.[87,88] The midpoints of calibrated radiocarbon dates or archaeological periods of the ancient samples (Table S4) were used as tip dates,[81] assuming 25 years for one generation. Each Markov Chain Monte Carlo (MCMC) simulation was run for 500,000,000 generations and sampled every 5,000 generations, with the first 50,000,000 generations discarded as burn-in. The EBSPs were reconstructed using EBSPAnalyser (10% buin-in) and visualized using an in-house R script.

Median-joining network reconstruction

A dataset of D4h HVS sequences (n = 62), which encompasses 53 mtDNAs from contemporary populations and nine from ancient samples (Table S5), together with the corresponding HVS sequences extracted from the complete mitogenomes, was used to reconstruct a D4h median-joining network (Figure S4). The median-joining network was firstly constructed by Network 4.510 (http://www.fluxus-engineering.com/sharenet.htm) and then checked and reconstructed manually.[89]