Received: 4 July 2019 | Revised: 9 April 2020 | Accepted: 17 April 2020

DOI: 10.1002/ajpa.24076

American Journal of PHYSICAL ANTHROPOLOGY

WILEY

RESEARCH ARTICLE

Ancient DNA reveals two paternal lineages C2a1a1b1a/F3830 and C2b1b/F845 in past nomadic peoples distributed on the Mongolian Plateau

Jiawei Li[1] | Dawei Cai[1] | Ye Zhang[2] | Hong Zhu[1] | Hui Zhou[1,2]

[1]Ancient DNA Laboratory, Research Center for Chinese Frontier Archaeology, Jilin University, Changchun, People's Republic of China

[2]College of Life Science, Jilin University, Changchun, People's Republic of China

Correspondence
Hui Zhou, College of Life Science, Jilin University, Changchun, 130012, People's Republic of China.
Email: zhouhui@jlu.edu.cn

Funding information
China Postdoctoral Science Foundation, Grant/Award Number: 2019M661226; The Major Projects of the National Social Science Foundation of China, Grant/Award Number: 17ZDA221

Abstract

Objectives: Since the third century CE, a series of nomadic tribes have been active on the eastern part of the Mongolian Plateau. Characterizing the genetic compositions of past nomadic people is significant for research on the nomadic cultures of the Eurasian Steppe region. Ancient DNA analysis facilitates a deeper understanding of the relationship between historical and modern nomadic populations.

Materials and methods: Whole-genome shotgun sequencing and capture sequencing of the nonrecombining region of the Y chromosome were performed for six ancient Hg C2/M217 individuals. The individuals were interred at six separate sites on the Mongolian Plateau and represent dates spanning the late Neolithic to Yuan Dynasty (~3,500–700 BP).

Results: After NRY capture sequencing, three of the six ancient samples were attributed to C2b1b/F845 and the other three ancient samples belonged to C2a1a1b1a/F3830. Analysis of whole-genome shotgun sequencing data shows that the ancient C2b1b/F845 individuals are closely related to She, Han and other East Asian populations, while the ancient C2a1a1b1a/F3830 individuals are more similar to modern Northeast Asian peoples, such as the Ulchi and Yakut.

Discussion: Hg C2/M217, widely distributed in the eastern part of the Eurasian continent, was discovered in the ancient Central Steppe and Baikal region. This study shows that there were two important subclades of Hg C2/M217 among the ancient nomadic peoples: C2a1a1b1a/F3830, which has made important genetic contributions to modern Mongolic- and Manchu-speaking populations, and C2b1b/F845, which probably originated in the farming populations of southern East Asia and made certain genetic contributions to past nomadic peoples on the Mongolian Plateau.

KEYWORDS

ancient DNA, nomadic population, NRY capture, shotgun sequencing

1 | INTRODUCTION

The earliest nomadic group to be recorded historically was the Xiongnu: a people living on the eastern part of the Eurasian Steppe in an area that loosely corresponds to the territory of modern Mongolia and northern China. They first appear in historical texts dated as early as the fourth century CE (Minajev, 1996), and the formation of the powerful Xiongnu Confederacy presented a serious threat to the prosperous agricultural society of the Central Plains (Min, 2010). The confederacy captured parts of northern China from the Han Dynasty

(202 BCE–220 CE) and raided their borders, a situation that persisted until Emperor Wu of Han (141–87 BCE) launched several campaigns against the Xiongnu and eventually pushed them north of the Gobi Desert (Liu, 2008). After the disintegration of the Xiongnu Confederacy, the eastern Eurasian Steppe region played host to a series of historically recorded nomadic powers, including the Xianbei Confederation, Rouran Khaganate, Turkic Khaganate, Uighur Khaganate, and Khitan Empire (West, 2008). In the 13th century, another powerful nomadic state flourished on the Mongolian Plateau and established the largest contiguous land empire in the world: the Mongol Empire (1,206–1,294 CE). During this period in particular, goods, technologies, and ideas were exchanged across the Eurasian continent, which had a far-reaching impact on the movement of people. From this came the Yuan Dynasty (1,271–1,368 CE), the first dynasty to rule all of China that established by nomads (Bi, 2000; Franke & Twitchett, 1994). The huge territories of these nomadic states and the mobility of the people within them, particularly the later Mongol Empire and Yuan Dynasty, are visible in the biological data of modern nomadic peoples. Studying the genetic compositions of these past nomadic peoples allows for the characterization of the movement under past nomadic states, which has major significance for understanding the nomadic culture on the Eurasian Steppe as well as the role of nomadic societies in world history.

Thus far, several studies have focused on the genetic structure of ancient nomadic peoples on the Mongolian Plateau. In 2003, Keyser-Tracqui et al. analyzed three types of genetic markers to determine the genetic relationships among 62 individuals buried in the Egyin Gol necropolis (Mongolia) dated to the Xiongnu period (Keyser-Tracqui, Crubezy, & Ludes, 2003). This study was the first that used biparental, paternal, and maternal genetic systems to reconstruct partial genealogies in a protohistoric necropolis. To understand the maternal genetic structure of the Xianbei and trace its impact on the formation and development of the genetic composition of nomadic peoples in northern China, Yu et al. analyzed 17 sequences of hypervariable segment I (HVS-I) in the mitochondrial DNA (mtDNA) control region of Tuoba Xianbei remains from the Shangdu Dongdajing cemetery (Inner Mongolia, China), finding that the Tuoba Xianbei potentially contributed to the gene pool of some northern nomadic peoples and mixed with the Xiongnu in northern China (Yu, Zhao, & Zhou, 2014). To study the paternal genetic characteristics of Donghu-Xianbei people, Zhang et al. analyzed 40 ancient human remains that belonged to the Donghu (~2,500 years ago), Xianbei (~1,700 years ago), and Shiwei (~1,200 years ago), tracing the Y chromosome haplogroup (Hg) C2a1a1/F3918 (ISOGG 2019) to a 2,500-year-old Donghu nomadic group (Zhang et al., 2018). On this basis, by capturing the nonrecombining region of the Y chromosome (NRY) through sequencing, Li et al. found that Hg C2a1a1b1a/F3830 (ISOGG 2019), downstream of C2a1a1/F3918, was an important paternal lineage of Donghu-Xianbei nomads before the expansion of the Mongol Empire (Li, Zhang, Zhao, Chen, & Ochir, 2018).

The research presented here builds upon past studies characterizing the genetic characteristics of the Hg C2 paternal lineage. Hg C2/M217 is a commonly occurring branch of Hg C/M130, and it is widely distributed in the eastern part of the Eurasian continent, with the highest frequencies among the populations of Mongolia and the Russian Far East. Hg C2/M217 separated from its most closely related lineage approximately 50,000 years ago, whereas its subclades, C2a/L1373 and C2b/F1067, separated over 30,000 years ago (Karmin et al., 2015). Sublineage C2a/L1373, which mainly includes the three branches, C2a1a1/F3918, C2a1a2/M48, and C2a1a3/M504, is widely distributed in northern Asian regions, such as northern China and Mongolia. In addition, the subbranch C2a1a1a2a1/F3830 has been identified in samples from modern Mongolic- and Tungusic-speaking populations. Sublineage C2b/F1067, which mainly includes two branches, C2b1a/Z1300 and C2b1b/F845, appears in northeastern Asia, including northern China, Korea and Japan (Huang et al., 2018).

This article presents the results of research on the genetic characteristics of the Hg C2 paternal lineage, a genetic lineage demonstrated to have featured prominently in nomadic populations prior to the fourth century CE (Li et al., 2018). By analyzing NRY capture sequencing data of six individuals from six sites distributed on the Mongolian Plateau, we are able to examine genetic inheritance between these ancient Hg C2/M217 samples. In addition, we analyze autosomal data obtained from whole genome shotgun sequencing, which allows us to examine the genetic association between the ancient samples and modern populations, facilitating the exploration of the origins and movement of certain genes in ancient nomadic groups over time. The genetics research of past nomadic populations is a growing field, and this study presents analyses that contribute to gaining a full picture of the effects of historic mobility and long-distance interactions on the genetic compositions of nomadic peoples.

2 | MATERIALS AND METHODS

The individuals used in this study (TL1, ZHS5, GG3, YK15, QL11, and ZK3022) were interred at six separate sites (Khermen Tal, Zaan Khoshuu, Gangga, Yikeshu, Qilangshan, and Zhukaigou site, respectively) on the Mongolian Plateau in Mongolia and northern China and represent dates spanning the late Neolithic to Yuan Dynasty (~3,500–700 BP) (Figure 1, Table S1). Among the six sites (Figure 1), the Khermen Tal (TL1, Rouran period), the Zaan Khoshuu (ZHS5, Xianbei period), and the Gangga (GG3, Shiwei period) have been previously studied (Li et al., 2018). The other three sites are the Yikeshu (YK15), Qilangshan (QL11), and Zhukaigou (ZK3022), all of which are located in Inner Mongolia, China. The Yikeshu site lies northwest of ancient Shangdu City in Zhenglan Banner, Xilingol League. According to the structure of the tomb, characteristic funerary objects, and the presence of sacrificial sheep bones, the burial site is thought to have been a Mongol cemetery during the Yuan Dynasty (Wei, 2004). The Qilangshan site is located in Ulanchabu City, Chahar Right Middle Banner. According to the peculiarity of the tomb form, combination of funerary objects and characteristics of the buried objects, the Qilangshan site is considered to be a relic of the Tuoba Xianbei from the fourth–fifth century CE (Yu, Xie, Zhang, Zhou, & Zhu, 2006).

FIGURE 1 Geographic locations of the sites where the study samples were collected

The Zhukaigou site is located on the eastern part of the Ordos Plateau. On the basis of the funeral style, the Zhukaigou site is presumed to date to approximately 4,200–3,500 years ago (Wang et al., 2007).

The SNPs M130 (Hg C), M217 (Hg C2) and other markers that belong to the ISOGG 2019 Y DNA haplogroup tree, were selected to identify whether the ancient samples were attributed to Hg C2. All primers design, PCR amplification and Sanger sequencing setups were carried out as previously described (Zhang et al., 2018). NRY capture sequencing data of three of these individuals (TL1, ZHS5, and GG3) have been published in Li et al., 2018, and the three other samples (QL11, YK15, and ZK3022) were analyzed by NRY capture sequencing. Whole-genome shotgun sequencing was then carried out on these six ancient samples.

2.1 | Sample processing and DNA sequencing

All wet lab work for this study involved ancient DNA lab work, which was conducted in dedicated clean-room facilities at Jilin University according to strict aDNA standards (Sarkissian et al., 2013). DNA extraction and sequencing library construction were performed following established protocols (Li et al., 2018). Whole-genome shotgun sequencing of the ancient samples was performed by Novogene (Beijing, China), and NRY capture sequencing of male individuals was conducted in collaboration with Sierravast BioMedical (Shanghai, China) as previously described (Li et al., 2018). Raw read quality was analyzed using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), adaptor sequences were trimmed, and the forward/reverse reads of each read pair requiring an 11-base pair overlap were merged using AdapterRemoval (Schubert, Lindgreen, & Orlando, 2016). Next, the Burrows-Wheeler Aligner (BWA) and SAMtools software were used to map and splice the processed reads according to the hs37d5 human reference genome (Li et al., 2009; Li & Durbin, 2009). Then, Picard was used to remove repetitive sequences (http://picard.sourceforge.net).

2.2 | DNA authenticity

Authentication of the aDNA was carried out by examining the fragment length distributions and nucleotide substitution patterns characteristic of aDNA damage using mapDamage (Jonsson, Ginolhac, Schubert, Johnson, & Orlando, 2013). Considering the effect of uracil-DNA glycosylase (UDG) on aDNA damage, we also estimated the patterns of DNA damage in the CpG context using PMDtools (Skoglund et al., 2014). The levels of contamination were estimated for mtDNA sequences of all individuals using the Mapping Iterative Assembler (MIA) and contamMix (Fu et al., 2013), as well as for chromosome X of the male individuals using ANGSD software (Korneliussen, Albrechtsen, & Nielsen, 2014).

2.3 | MtDNA and Y chromosome analyses

MtDNA and Y chromosome sequences were extracted from final-analysis BAM files using SAMtools, and BCFtools was used to call positions when the base differed from that in the reference. Only bases with a quality score > 30, read depth > 3, and call rate > 90% were considered in the analysis. MtDNA Hgs were assigned with HaploGrep 2 (Weissensteiner et al., 2016), and Y chromosome Hgs were determined with yHaplo (Poznik, 2016). Then, 2,954 Y chromosome SNPs were filtered according to the data available from related published studies for SNP calling and subsequent analysis (Altshuler et al., 2010; Huang et al., 2018; Karmin et al., 2015; Li et al., 2018;

Lippold et al., 2014; Wei et al., 2017; Wei, Wang, et al., 2018; Wei, Yan, et al., 2018) (Table S1). We used BEAST 2 (Bouckaert et al., 2014) to estimate the coalescence times of Hg C2 and its sublineages. Coalescent Bayesian skyline priors were selected with the general time reversible (GTR) substitution model and a strict clock with a mutation rate of 0.74×10^{-9} per site per year (Karmin et al., 2015). The calculations were performed with 10 million iterations and sampling every 1,000 steps. Then, the results were visualized in Tracer v.1.7.1 and Figtree v1.4.4 with a burn-in of 10%, and it was confirmed that all effective sample sizes were above 200.

2.4 | Autosomal analyses

To prepare for autosomal genetic analyses, genotypes of 5,625 modern individuals obtained using the Affymetrix Human Origins array (https://reich.hms.harvard.edu/) were employed to construct a modern DNA background. Pseudohaploid genotypes of the six ancient samples were called by BCFtools based on a random read and filtered as previously described (Li et al., 2018). Population structure was investigated with principal component analysis (PCA) using smartpca in EIGENSOFT (Price et al., 2006) and plotted with the PC1 and PC2 vectors using R (https://www.r-project.org/). Outgroup-based estimation of admixture components was carried out using Admixture (Alexander, Novembre, & Lange, 2009). We performed model-based clustering analysis via the maximum-likelihood approach, assuming 2–20 ancestral populations (K = 2–20). Considering that the value of cross-validation (CV) error was maintained at 0.714–0.711 when K = 10–20, the model with 10 modeled clusters or ancestral populations (K = 10) was selected for the admixture analysis (Table S1). Outgroup f3 form statistics (Mbuti; ancient, East Asia) were computed with Mbuti as the outgroup using qp3Pop, as implemented in AdmixTools (Patterson et al., 2012). According to the admixture analysis and f3 statistics, 10 populations in East Asia were selected for inferring the historical patterns of population splits and admixture events ($m \leq 3$) using TreeMix (Pickrell & Pritchard, 2012). We rooted the Mbuti population as an outgroup, and 1,000 bootstrap replicates were generated to produce a consensus tree with a block size of 500 SNPs.

3 | RESULTS

The nucleotide substitution patterns characteristic of aDNA damage in the CpG context showed that the damage to all samples exceeded 20%. The distribution of read lengths showed that the mean read length of the shotgun library was 89–109 bp. MtDNA contamination of the aDNA library (excluding ZK3022) was estimated to be less than 4%, and X chromosome contamination for the other samples was 0.03–1.50%. The coverage of the shotgun sequencing data (0.02–0.65x) from ancient samples was low. Considering the few effective reads, low coverage, and high degree of damage to the ZK3022 sample, autosomal analysis was not carried out for this sample. Details of the aDNA library are provided in Table 1. After NRY capture sequencing, three of the six ancient samples (QL11, YK15, and ZK3022) were attributed to C2b1b/F845 (ISOGG 2019). The other three previously published ancient samples (TL1, ZHS5, and GG3) belonged to C2a1a1b1a/F3830 (ISOGG 2019).

3.1 | Phylogenetic analysis of the Y chromosome

The geographic distribution of modern individuals sharing haplotypes with ancient samples (C2a1a1a2a1/F3830 and C2b1b/F845) across Eurasia is shown in Figure 2 (Table S1). Ancient samples are indicated by red triangles (C2b1b/F845) and red squares (C2a1a1a2a1/F3830)

TABLE 1 List of samples used in this study, sequencing metrics for the ancient libraries and their mtDNA nucleotide changes as well as Y chromosome Haplogroups

	Shotgun sequencing								
Sample ID	Total reads	Mapped not clonal reads with $q \geq 30$	Mean read length (bp)	MT Contamin (%)	X Contamin. (Test1/Test2) (%)	C to T misincorporation at 5′ end	Coverage	mtDNA haplogroups	Y-haplogroups (ISOGG 2019)
ZK3022	30,286,856	750,978	93	5.2	0/6.06	0.39	0.02x	R (100%)	C2b1b/F845
QL11	88,286,382	19,691,286	104	3.4	0.14/0.24	0.34	0.65x	G1a1 (93.26%)	C2b1b/F845
YK15	38,974,089	8,278,290	91	3.9	0.14/0.47	0.25	0.24x	G1a1 (90.49%)	C2b1b/F845
ZHS5	66,222,326	1,031,069	96	NaN	NaN	0.26	0.03x	C4a2a1 (81.45%)	C2a1a1b1a/F3830
TL1	72,081,872	4,681,299	109	2.5	0.03/1.50	0.38	0.16x	D4b1a2a (95.64%)	C2a1a1b1a/F3830
GG3	59,373,236	3,002,378	89	2.4	0.14/0.47	0.3	0.09x	F1b (84.13%)	C2a1a1b1a/F3830

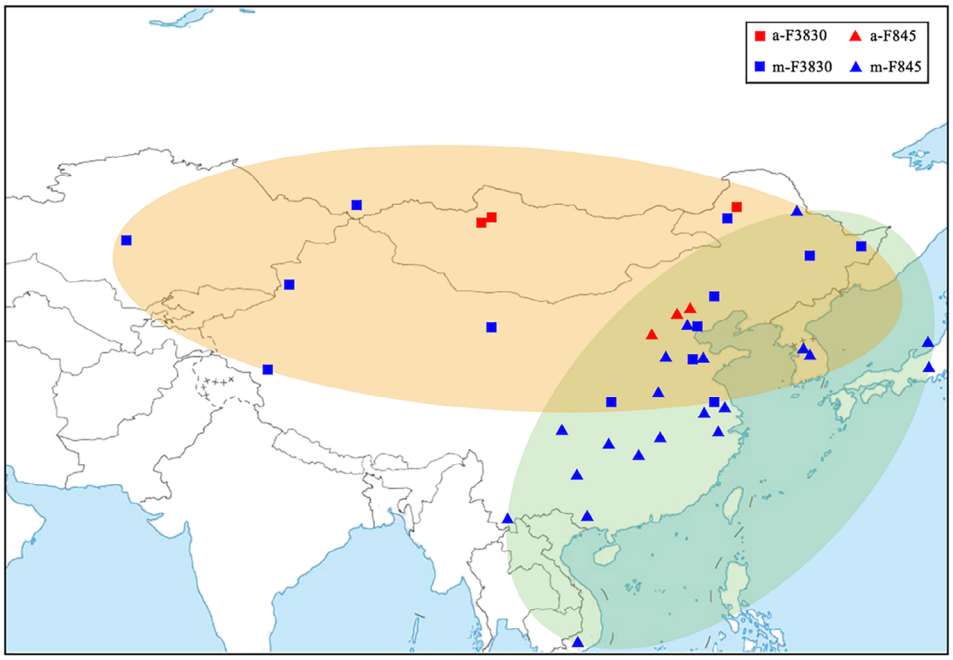

FIGURE 2 The geographic distribution of modern individuals sharing haplotypes with ancient individuals (C2a1a1a2a1/F3830 and C2b1b/F845). Ancient samples are indicated by red triangles (a-F845) or red squares (a-F3830), while modern individuals are represented by blue triangles (m-F845) or blue squares (m-F3830)

in Figure 2. As shown by the figure, the C2a1a1a2a1/F3830 samples were present in northern China, Mongolia, and Siberia, while the C2b1b/F845 samples were mainly observed in southern China and on the Korean Peninsula. The phylogenetic tree constructed using the NRY capture data from the six ancient samples and 31 modern reference samples shows that the ancient individuals are mainly distributed on two branches (Figure 3). Among the samples, TL1, ZHS5, and GG3 are located together on a branch (C2a1a1a2a1/F3830) with modern Mongolian, Mongolian-Buryat, Yugur, and Kazakh individuals. By contrast, QL11, YK15, and ZK3022 are located together on another branch (C2b1b/F845) with modern Dai, Han, and Japanese individuals. The divergence time between Hg C2a/L1373 (upstream of C2a1a1a2a1/F3830) and C2b/F1067 (upstream of C2b1b/F845) is estimated to be 33,682 years (95% confidence interval: 19,653–54,278 years).

3.2 | Genetic analysis of nuclear DNA

The PCA results (Figure 4) show that the ancient C2a1a1a2a1/F3830 individuals (TL1, ZHS5, and GG3) cluster away from the modern populations and the ancient C2b1b/F845 samples (YK15 and QL11) are scattered. QL11 clusters close to modern East Asian samples, and YK15, which is more closely related to the F3830 group, clusters together with Kusunda. The admixture analysis at K = 10 (Figure 5a; Table S1, Figure S1) reveals that both ancient C2a1a1a2a1/F3830 and C2b1b/F845 individuals have ancestral composition sources similar to those of most East Asian individuals. However, the ancient C2b1b/F845 samples (YK15 and QL11) have composition structures similar to those of modern Han, Miao, and She samples, while the ancient C2a1a1a2a1/F3830 samples (TL1, ZHS5, and GG3) are similar to modern Northeast Asian samples, especially Yakut samples, having greater Siberian and Far Eastern components. The f3 analyses (Mbuti; ancient, East Asia) (Figure 5b and Figure S2), which quantifies the amount of genetic drift shared between the East Asian populations and the ancient Hg C2/M217 individuals since they diverged from the Mbuti African outgroup, show that the largest amount of drift shared with the ancient C2a1a1a2a1/F3830 individuals (TL1, ZHS5, and GG3) came from the Ulchi. The largest amount of drift shared with the ancient C2b1b/F845 (YK15 and QL11) individuals came from the She. Using the probabilistic model of population splits and admixture events implemented in TreeMix (Figure 6; Figure S3), we found that the ancient C2a1a1a2a1/F3830 individuals (TL1, ZHS5, and GG3) cluster together, while the ancient C2b1b/F845 individuals (YK15 and QL11) are separate. One of the latter individuals (QL11, ~1,600 years ago) is on a branch with the Han, and the other one (YK15, ~700 years ago) clusters with the ancient C2a1a1a2a1/F3830 individuals. In addition, there was gene flow between GG3 and the branch of Mongol, Miao, and She individuals, as well as gene flow between QL11 and the branch of YK15 and two F3830 samples (ZHS5 and TL1) when two migration events occurred (m = 2) (Figure 6).

4 | DISCUSSION AND CONCLUSIONS

Previous aDNA studies have shown that an ancient individual from the Lokomotiv cemetery (~6,700 BP), part of the Early Neolithic Kitoi hunter-gatherer tradition in the Baikal region, belonged to the C2a1a1/F4015 (equivalent to C2a1a1/F3918) branch of the C2a/L1373 sublineage of the Hg C2/M217 branch (Damgaard et al., 2018). As one of the authors has established that the Donghu branch was downstream of C2a1a1/F3918 in an earlier work (Li et al., 2018), which implies that the Donghu originated from the hunter-gatherers of the Lake Baikal region. Additionally, male skeletal

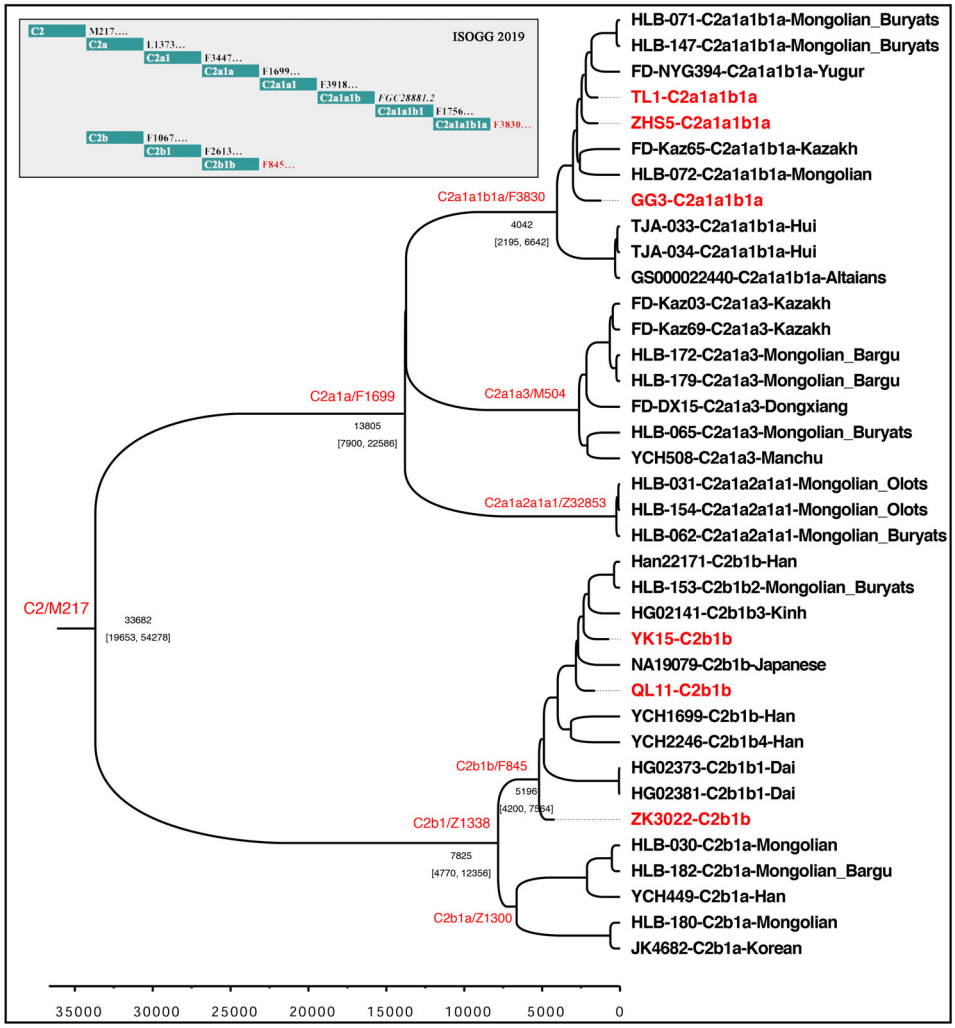

FIGURE 3 Y-DNA phylogenetic tree of haplogroup C2/M217 based on the Hg C2/M217 NRY data

remains from the sites of Jinggouzi (~2,500 years ago, Donghu), Chenwugou (~1,700 years ago, Xianbei), and Gangga (~1,200 years ago, Shiwei) belonged to the C2a1a1/F3918 branch, suggesting that the paternal inheritance of nomadic populations in the Mongolian grasslands was continuous up to the first millennium CE (Zhang et al., 2018). Research by one of the authors utilizing NRY capture sequencing also showed that TL1 (~1,400 years ago, Rouran), ZHS5 (~1,400 years ago, Xianbei), and GG3 (~1,200 years ago, Shiwei) belonged to C2a1a1b1a/F3830, a downstream branch of C2a1a1/F3918. This finding highlighted the important paternal genetic contribution by this branch to the current Mongolic- and Manchu-speaking populations on the eastern part of the Mongolian Plateau (Li et al., 2018).

This study aimed to expand upon previous research and contribute new data that could be used to map the movement and inheritance of genes through ancient nomadic populations and provide insights into the genetic inheritance of modern nomadic peoples. The mtDNA and autosomal DNA of the previously published C2a1a1b1a/F3830 samples (TL1, ZHS5, and GG3) were analyzed by whole-genome shotgun sequencing. Two of the C2a1a1b1a/F3830 individuals (TL1 and ZHS5) belonged to mtDNA haplotypes C4a2a1 and D4b1a2a, which are mostly found among the Yakut, Yukaghir, Buryat, and Tungusic peoples (Derenko et al., 2010). Admixture and f3 test analyses based on autosomal data also confirmed that the ancient C2a1a1b1a/F3830 samples display a close genetic relationship to Yakut, Ulchi, Oroqen, and Hezhen populations. The samples show a particularly strong relationship to the modern Ulchi, which are members of the southern Manchu-Tungusic language group reliant on a fishing and hunting economy located today in the Russian Far East. Moreover, gene flow from the Shiwei (GG3) to Mongol was detected through TreeMix analyses, suggesting that the ancient C2a1a1b1a/F3830 group most likely made a genetic contribution to modern Mongolians. These results indicated that the ancient samples (TL1, ZHS5, and GG3) belonging to Hg C2a1a1b1a/F3830 are closely genetically related to Manchu-Tungusic- and Mongolic-speaking populations in the Russian Far East and northeastern China, which is consistent with the previous results from the Y chromosome analyses (Li et al., 2018).

By contrast, three new samples—YK15 (~700 years ago, Mongol), QL11 (~1,600 years ago, Xianbei), and ZK3022 (~3,500 BP)—were identified as belonging to the C2b1b/F845 lineage by NRY capture sequencing, and this result represents the first time that this branch has been found in ancient populations, especially nomadic populations

FIGURE 4 PCA plot of the ancient F3830/F845 individuals and the modern individuals from the Affymetrix Human Origins array

on the eastern Mongolian Plateau. According to published articles, C2b1b/F845 is widely distributed but has a low frequency among East Asians, with a relatively low frequency in Koreans (2.4%) and Daur (4.8%) (Kwon, Lee, Lee, Yang, & Shin, 2015; Wang Ch & Li, 2018). Based on the geographical distribution of Hg C2/M217 subgroups, the C2b1b/F845 samples are distributed further south than the C2a1a1a2a1/F3830 samples in both ancient and modern populations. The C2b1b/F845 samples are mainly observed in southern China, Japan, Korea, Vietnam, and the Philippines, and these locations are much farther south than modern C2a1a1b1a/F3830 samples, which are present mainly in northeastern China, Mongolia, and Siberia. This finding suggests that C2b1b/F845 has a more southern origin than C2a1a1b1a/F3830. According to the phylogenetic tree, ancient C2b1b/F845 samples cluster on a branch with Dai, Tujia, and Han Chinese, showing that C2b1b/F845 was closely related to the genetic composition of peoples reliant on agriculture rather than mobile herding. Moreover, an ancient C2b1b/F845 sample was identified for the first time at the Neolithic site of Zhukaigou (~4,200–3,500 BP) in northern China. Zhukaigou is thought to have been a settled farming culture (Tian, 1988) that experienced an economic transformation from settled farming to animal husbandry approximately 3,000 years ago (Yang, 2012). Considering the time and distribution of the sites, we conclude that Hg C2b1b/F845 most likely originated from the agricultural population of southern East Asia.

In addition, two C2b1b/F845 samples (QL11 and YK15) both belonged to the mtDNA G1a1 subclade, which is found widely in East Asia, with the highest frequencies in a cluster that includes Japanese and Koreans (Chandrasekar et al., 2009). Admixture and f3 test analyses of the autosomal DNA indicates that ancient C2b1b/F845 samples have a greater amount of southern East Asian features than ancient C2a1a1b1a/F3830 branches, showing a very close genetic relationship to the She, a shifting-cultivation population in southern China. The PCA results further suggest that the ancient C2b1b/F845 group underwent change over time, and this interpretation is supported by the TreeMix analyses. The QL11 individual is genetically closer to East Asia populations, particularly Han Chinese people, and it exhibits gene flow to the YK15 individual, who is closer to the ancient C2a1a1b1a/F3830 group (including samples ZHS5 and TL1) and thus also East Asian nomadic peoples. These results do not conflict with the results of Y chromosome analyses, suggesting that the

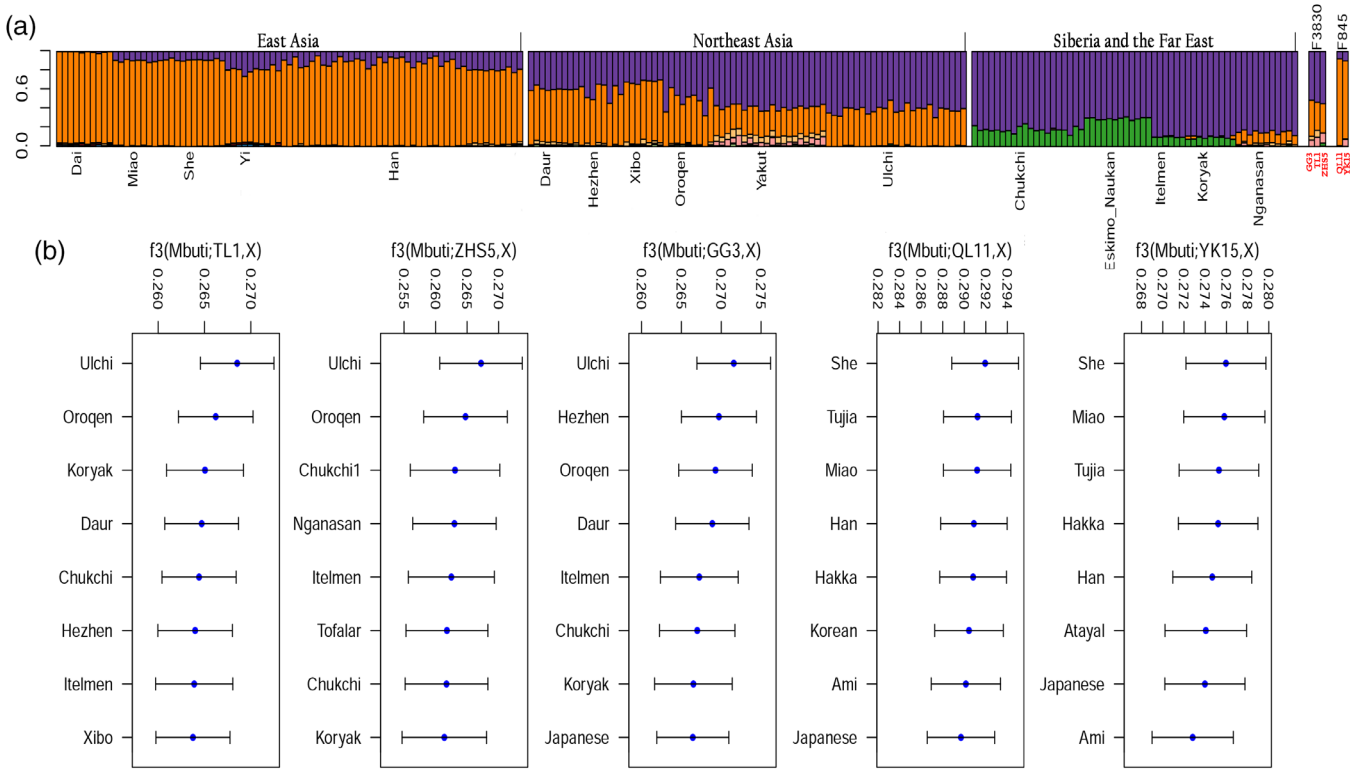

FIGURE 5 The genetic affinity of the ancient Hg C2/M217 individuals and the East Asia populations (a) Cluster analysis generated by Admixture for the ancient Hg C2/M217 individuals and 16 modern populations (K = 10). Full modern populations are available in Figure S1. (b) Ranked outgroup $f3$ statistics examining represents the amount of shared genetic drift between the ancient Hg C2/M217 individuals and each of the eight contemporary East Asia populations since their divergence with the Mbuti population

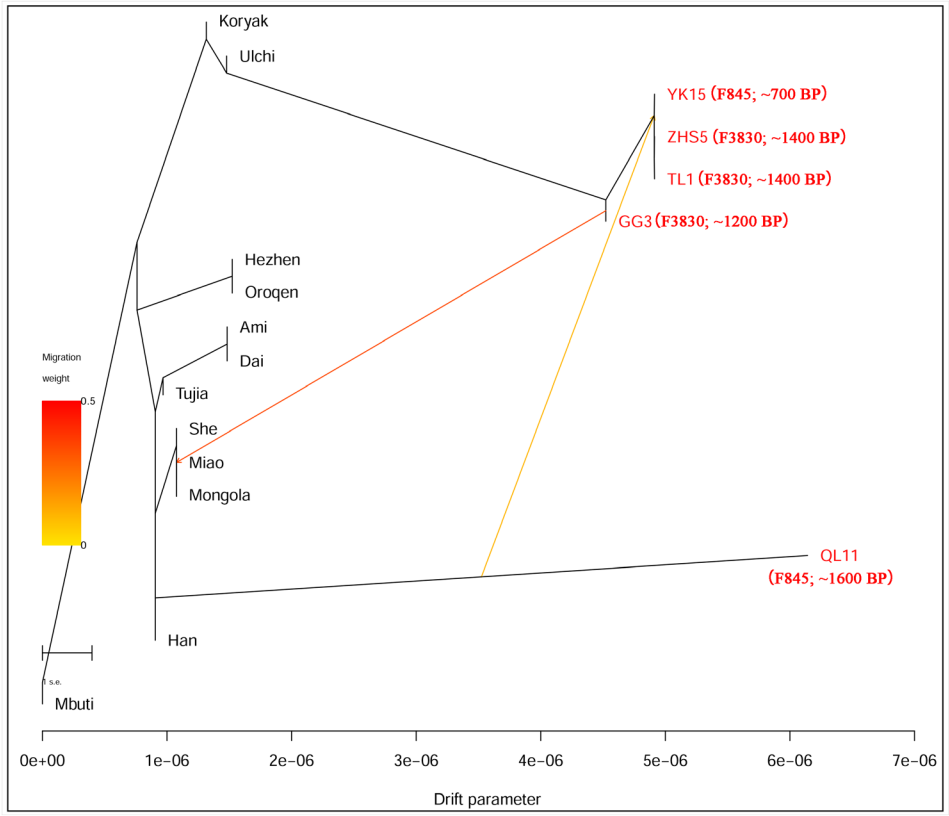

FIGURE 6 TreeMix phylogeny of the ancient Hg C2/M217 individuals and the East Asia populations representing the patterns of population divergence and migration events (m = 2)

C2a1a1b1a/F845 group originated from a farming people and made a genetic contribution to nomadic peoples.

In conclusion, by studying the inheritance of Hg C2/M217 in ancient nomadic populations, two paternal lineages—C2a1a1b1a/F3830 and C2b1b/F845—were identified in the past nomadic peoples who were interred on the Mongolian Plateau. The C2a1a1b1a/F3830 lineage is closely related to the Donghu branch, which probably originated earlier with the hunter-gatherers of the Lake Baikal region. This lineage made important genetic contributions to modern Manchu-Tungusic- and Mongolic-speaking populations on the eastern part of the Mongolian Plateau. In contrast, the C2b1b/F845 lineage originated from much further south geographically, most likely from the farming people of southern East Asia. Gene flow between samples indicates that C2b1b/F845 contributed to the gene pool of both the northern nomadic people and the southern farming people in East Asia.

ACKNOWLEDGMENTS

This work was supported by the China Postdoctoral Science Foundation funded project, Grant number: 2019 M661226 (Grant Recipient: Jiawei Li) and the Major Projects of the National Social Science Foundation of China under grant number 17ZDA221 (Grant Recipient: Daiwei Cai). Jiawei Li conceived and designed the study. Jiawei Li, Daiwei Cai, and Ye Zhang performed the analysis and wrote the manuscript. Hong Zhu and Hui Zhou helped in interpreting results and improving the manuscript. All authors read and approved the final manuscript. We thank Dr. Rebecca Victoria O'Sullivan for improving the readability of the manuscript, Dr. Philip Johnson for providing the contamMix program, and Pengfei Cheng from the Archeological Institute of Inner Mongolia, P. R. China for providing the background information related to anthropological factors.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in the European Nucleotide Archive (ENA) at https://www.ebi.ac.uk/ena, reference number PRJEB24670 and PRJEB33265.

ORCID


Hui Zhou https://orcid.org/0000-0001-5858-5636

REFERENCES

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664. https://doi.org/10.1101/gr.094052.109

Altshuler, D., Durbin, R. M., Abecasis, G. R., Bentley, D. R., Chakravarti, A., Clark, A. G., … Consortium, GP. (2010). A map of human genome variation from population-scale sequencing. *Nature*, *467*(7319), 1061–1073. https://doi.org/10.1038/nature09534

Bi, O. (2000). An Introduction to the Yuan Territories. *China's Borderland History and Geography Studies*, *10*(4), 3–23.

Bouckaert, R., Heled, J., Kuhnert, D., Vaughan, T., Wu, C. H., Xie, D., … Drummond, A. J. (2014). BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, *10*(4), 1–6. Retrieved from <Go to ISI>://WOS:000336507500022.

Chandrasekar, A., Kumar, S., Sreenath, J., Sarkar, B. N., Urade, B. P., Mallick, S., … Rao, V. R. (2009). Updating phylogeny of mitochondrial DNA Macrohaplogroup M in India: Dispersal of modern human in south Asian corridor. *Plos One*, *4*(10), e7447. https://doi.org/10.1371/journal.pone.0007447

Damgaard, P. D., Martiniano, R., Kamm, J., Moreno-Mayar, J. V., Kroonen, G., Peyrot, M., … Willerslev, E. (2018). The first horse herders and the impact of early bronze age steppe expansions into Asia. *Science*, *360*(6396), 1422. https://doi.org/10.1126/science.aar7711

Derenko, M., Malyarchuk, B., Grzybowski, T., Denisova, G., Rogalla, U., Perkova, M., … Zakharov, I. (2010). Origin and post-glacial dispersal of mitochondrial DNA Haplogroups C and D in northern Asia. *Plos One*, *5*(12), e15214. https://doi.org/10.1371/journal.pone.0015214

Franke, H., & Twitchett, D. C. (1994). *The Cambridge history of China: Volume 6: Alien regimes and border states, 907–1368* (Vol. 6). Cambridge: Cambridge University Press.

Fu, Q. M., Mittnik, A., Johnson, P. L. F., Bos, K., Lari, M., Bollongino, R., … Krause, J. (2013). A revised timescale for human evolution based on ancient mitochondrial genomes. *Current Biology*, *23*(7), 553–559. https://doi.org/10.1016/j.cub.2013.02.044

Huang, Y. Z., Wei, L. H., Yan, S., Wen, S. Q., Wang, C. C., Yang, Y. J., … Li, H. (2018). Whole sequence analysis indicates a recent southern origin of Mongolian Y-chromosome C2c1a1a1-M407. *Molecular Genetics and Genomics*, *293*(3), 657–663. https://doi.org/10.1007/s00438-017-1403-4

Jonsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F., & Orlando, L. (2013). mapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*, *29*(13), 1682–1684. https://doi.org/10.1093/bioinformatics/btt193

Karmin, M., Saag, L., Vicente, M., Sayres, M. A. W., Jarve, M., Talas, U. G., & Kivisild, T. (2015). A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Research*, *25*(4), 459–466. https://doi.org/10.1101/gr.186684.114

Keyser-Tracqui, C., Crubezy, E., & Ludes, B. (2003). Nuclear and mitochondrial DNA analysis of a 2,000-year-old necropolis in the Egyin Gol Valley of Mongolia. *American Journal of Human Genetics*, *73*(2), 247–260. https://doi.org/10.1086/377005

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics*, *15*, 356. https://doi.org/10.1186/S12859-014-0356-4

Kwon, S. Y., Lee, H. Y., Lee, E. Y., Yang, W. I., & Shin, K. J. (2015). Confirmation of Y haplogroup tree topologies with newly suggested Y-SNPs for the C2, O2b and O3a subhaplogroups. *Forensic Science International-Genetics*, *19*, 42–46. https://doi.org/10.1016/j.fsigen.2015.06.003

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*, *25*(14), 1754–1760 Retrieved from <Go to ISI>://WOS:000267665900006.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., & Proc, G. P. D. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079 Retrieved from <Go to ISI>://WOS:000268808600014.

Li, J., Zhang, Y., Zhao, Y. B., Chen, Y. Z., Ochir, A., Zhu, H., & Zhou, H. (2018). The genome of an ancient Rouran individual reveals an important paternal lineage in the Donghu population. *American Journal of Physical Anthropology*, *166*(4), 895–905. https://doi.org/10.1002/ajpa.23491

Lippold, S., Xu, H., Ko, A., Li, M., Renaud, G., Butthof, A., … Stoneking, M. (2014). Human paternal and maternal demographic histories: Insights from high-resolution Y chromosome and mtDNA sequences. *Investigative Genetics*, *5*(1), 13.

Liu, Y. (2008). The National Policy which adopts to Hun in the Western Han in dynasty. *Nationalities Research in Qinghai*, *19*(01), 148–151.

Min, H. (2010). *The study on the history of Xiongnu's development Doctoral dissertation* (). China: Lanzhou university.

Minajev, S. (1996). Les Xiongnu. *Dossiers d'Archéologie*, *212*, 74–85.

Patterson, N., Moorjani, P., Luo, Y. T., Mallick, S., Rohland, N., Zhan, Y. P., & Reich, D. (2012). Ancient admixture in human history. *Genetics*, *192*(3), 1065. https://doi.org/10.1534/genetics.112.145037

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, *8* (11), e1002967. https://doi.org/10.1371/journal.pgen.1002967

Poznik, G. D. (2016). *Identifying Y-chromosome haplogroups in arbitrarily large samples of sequenced or genotyped men*. *bioRxiv*. https://doi.org/10.1101/088716.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, *38*(8), 904–909. https://doi.org/10.1038/ng1847

Sarkissian, C. D., Balanovsky, O., Brandt, G., Khartanovich, V., Buzhilova, A., Koshel, S., & Consortium, G. (2013). Ancient DNA reveals prehistoric gene-flow from Siberia in the complex human population history of north East Europe. *PLoS Genetics*, *9*(2), e1003296. https://doi.org/10.1371/journal.pgen.1003296

Schubert, M., Lindgreen, S., & Orlando, L. (2016). AdapterRemoval v2: Rapid adapter trimming, identification, and read merging. *BMC Research Notes*, *9*, 88. https://doi.org/10.1186/s13104-016-1900-2

Skoglund, P., Northoff, B. H., Shunkov, M. V., Derevianko, A. P., Paabo, S., Krause, J., & Jakobsson, M. (2014). Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(6), 2229–2234. https://doi.org/10.1073/pnas.1318934111

Tian, G. (1988). Zhukaigou Site in Inner Mongolia. *Acta Archaeologia Sinica*, *1*(03), 301–332 369–376.

Wang Ch, S. M., & Li, H. (2018). Origin of Daur revealed by molecular anthropology. *Journal of Beifang University of Nationalities*, *143*(5), 110–117.

Wang, H., Chang, E., Cai, D., Zhang, Q., Zhou, H., & Zhu, H. (2007). Mitochondrial DNA analysis of remains from Zhukaigou archaeological site in Inner Mongolia. *Journal of Jilin University(Medicine Edition)*, *33* (01), 5–8.

Wei, J. (2004). *An archaeological research on Shangdu site of the yuan dynasty Doctoral dissertation* (). China: Jilin University.

Wei, L. H., Huang, Y. Z., Yan, S., Wen, S. Q., Wang, L. X., Du, P. X., & Li, H. (2017). Phylogeny of Y-chromosome haplogroup C3b-F1756, an important paternal lineage in Altaic-speaking populations. *Journal of Human Genetics*, *62*(10), 915–918. https://doi.org/10.1038/jhg.2017.60

Wei, L. H., Wang, L. X., Wen, S. Q., Yan, S., Canada, R., Gurianov, V., … Li, H. (2018). Paternal origin of paleo-Indians in Siberia: Insights from Y-chromosome sequences. *European Journal of Human Genetics*, *26* (11), 1687–1696. https://doi.org/10.1038/s41431-018-0211-6

Wei, L. H., Yan, S., Lu, Y., Wen, S. Q., Huang, Y. Z., Wang, L. X., … Li, H. (2018). Whole-sequence analysis indicates that the Y chromosome C2*-star cluster traces back to ordinary Mongols, rather than Genghis khan. *European Journal of Human Genetics*, *26*(2), 230–237. https://doi.org/10.1038/s41431-017-0012-3

Weissensteiner, H., Pacher, D., Kloss-Brandstatter, A., Forer, L., Specht, G., Bandelt, H. J., & Schonherr, S. (2016). HaploGrep 2: Mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Research*, *44*(W1), W58–W63. https://doi.org/10.1093/nar/gkw233

West, B. A. (2008). *Encyclopedia of the peoples of Asia and Oceania*. Facts On File Inc. NY, United States.

Yang, Z. (2012). *Interpretation of Zhukaigou culture*. Ordos Research Association. Ordos.

Yu, C. C., Xie, L., Zhang, X. L., Zhou, H., & Zhu, H. (2006). Genetic analysis on Tuoba Xianbei remains excavated from Qilang Mountain cemetery in Qahar right wing middle banner of Inner Mongolia. *FEBS Letters*, *580*(26), 6242–6246. https://doi.org/10.1016/j.febslet.2006.10.030

Yu, C. C., Zhao, Y. B., & Zhou, H. (2014). Genetic analyses of Xianbei populations about 1,500-1,800 years old. *Russian Journal of Genetics*, *50*(3), 308–314. https://doi.org/10.1134/S1022795414030119

Zhang, Y., Wu, X. Y., Li, J. W., Li, H. J., Zhao, Y. B., & Zhou, H. (2018). The Y-chromosome haplogroup C3*-F3918, likely attributed to the Mongol empire, can be traced to a 2500-year-old nomadic group. *Journal of Human Genetics*, *63*(2), 231–238. https://doi.org/10.1038/s10038-017-0357-z

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Li J, Cai D, Zhang Y, Zhu H, Zhou H. Ancient DNA reveals two paternal lineages C2a1a1b1a/F3830 and C2b1b/F845 in past nomadic peoples distributed on the Mongolian Plateau. *Am J Phys Anthropol*. 2020;172:402–411. https://doi.org/10.1002/ajpa.24076