Received: 5 November 2017 | Revised: 7 March 2018 | Accepted: 5 April 2018

DOI: 10.1002/ajpa.23491

RESEARCH ARTICLE

WILEY American Journal of PHYSICAL ANTHROPOLOGY

The genome of an ancient Rouran individual reveals an important paternal lineage in the Donghu population

Jiawei Li[1,2] | Ye Zhang[1] | Yongbin Zhao[3] | Yongzhi Chen[4] | A. Ochir[5] | Sarenbilige[6] | Hong Zhu[1] | Hui Zhou[1,2]

[1]Ancient DNA Laboratory, Research Center for Chinese Frontier Archaeology, Jilin University, Changchun 130012, People's Republic of China

[2]College of Life Science, Jilin University, Changchun 130012, People's Republic of China

[3]Life Science College, Jilin Normal University, Siping 136000, People's Republic of China

[4]Director, Inner Mongolian Museum, Hohhot 010011, People's Republic of China

[5]Coordinator, International Institute for Study of Nomadic Civilization, 210620A, Ulaanbaatar 11, Mongolia

[6]Editorial department, Cultural Relics and Archaeological Institute of Inner Mongolia, Hohhot 010010, 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
National Natural Science Foundation of China, Grant number: 31371266

Abstract

Objectives: Following the Xiongnu and Xianbei, the Rouran Khaganate (Rouran) was the third great nomadic tribe on the Mongolian Steppe. However, few human remains from this tribe are available for archaeologists and geneticists to study, as traces of the tombs of these nomadic people have rarely been found. In 2014, the IA-M1 remains (TL1) at the Khermen Tal site from the Rouran period were found by a Sino-Mongolian joint archaeological team in Mongolia, providing precious material for research into the genetic imprint of the Rouran.

Materials and methods: The mtDNA hypervariable sequence I (HVS-I) and Y-chromosome SNPs were analyzed, and capture of the paternal non-recombining region of the Y chromosome (NRY) and whole-genome shotgun sequencing of TL1 were performed. The materials from three sites representing the three ancient nationalities (Donghu, Xianbei, and Shiwei) were selected for comparison with the TL1 individual.

Results: The mitochondrial haplotype of the TL1 individual was D4b1a2a1. The Y-chromosome haplotype was C2b1a1b/F3830 (ISOGG 2015), which was the same as that of the other two ancient male nomadic samples (ZHS5 and GG3) related to the Xianbei and Shiwei, which were also detected as F3889; this haplotype was reported to be downstream of F3830 by Wei et al. (2017).

Discussion: We conclude that F3889 downstream of F3830 is an important paternal lineage of the ancient Donghu nomads. The Donghu-Xianbei branch is expected to have made an important paternal genetic contribution to Rouran. This component of gene flow ultimately entered the gene pool of modern Mongolic- and Manchu-speaking populations.

KEYWORDS
ancient DNA, nomadic population, NRY capture, Rouran Khaganate

1 | INTRODUCTION

Since the Xiongnu and Han Dynasty successfully established unified political power on the Eurasian Steppes and Central Plains of China at the end of the third century BC, conflicts, exchanges and interactions between these two great civilizations (i.e., nomadic and farming civilizations) occurred over more than 2000 years of Chinese history. Another nomadic tribe, the Donghu, lived in the east of Xiongnu. This tribe existed for ~1300 years, from the early Shang Dynasty (~1600 BC) to the Western Han Dynasty (~206 BC) (Zhang et al., 2014). After Donghu was destroyed by Xiongnu, the remaining Donghu population evolved into two lineages (Wuhuan and Xianbei) (Lin, 1989). The Xianbei lineage rapidly grew strong since the Xiongnu declined and moved westward. This lineage occupied the Mongolian Steppe, annexed the remaining tribes that once surrendered to Xiongnu, and eventually established the second largest nomadic Empire (Jing, 2013).

Following the Xiongnu and Xianbei, the Rouran Khaganate (Rouran) was the third great nomadic tribe in the Mongolian grasslands from the late fourth century to the mid-sixth century (Barbara, 2008). The Rouran controlled the area of Mongolia from the Manchurian border to Turpan, perhaps even to the east coast of Lake Balkhash, ranging from the Orkhon River to China. The core area was mainly located in

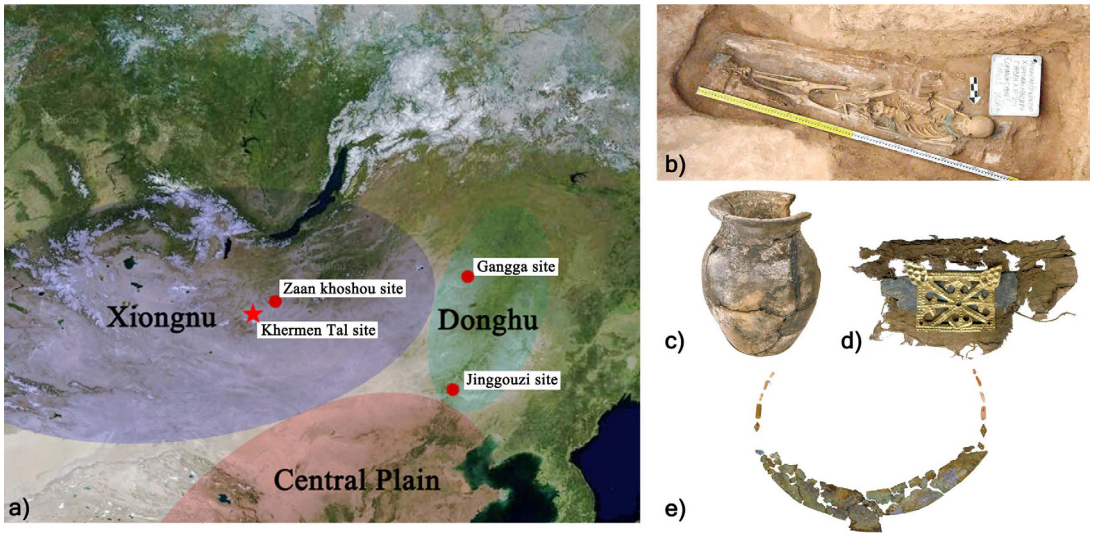

FIGURE 1 The geographic locations of the ancient groups used in this study (a) and the unearthed funerary objects of the Khermen Tal site (b, c, d, and e)

the territory of modern Mongolia. The territory sometimes expanded west to the ancient Western Regions (now part of Central Asia and Xinjiang, China) and south to the northern part of China's Inner Mongolia Autonomous Region. The Rouran set up a powerful nomadic empire and controlled the Silk Road, which ran through Yiwu, the Tianshan Mountains, Qiuci, and Pamirsfor ~150 years from the late fourth century to the mid-sixth century (Jing, 2013). Shiwei was another ancient nomadic group that might originated from Xianbei, and this population mainly inhabited the Greater Khingan Mountains, dating from the fifth century (J, 1985). During the 10th century, the Shiwei population began to migrate westward, finally reaching the region currently known as Mongolia (J, 1985; Zhang et al., 2018). However, the origins of nomadic tribes such as the Rouran that emerged after the Xianbei are puzzling because historical records concerning these tribes are rare. The customs and lifestyles differed between the nomadic population and the farming population. Most of the nomads lived in grasslands and did not settle, and there were no special signs indicating their tombs; thus, few remains are available for archaeologists and geneticists to study.

In 2014, a Sino-Mongolian joint archaeological team excavated a small earthen pit tomb at the Khermen Tal site in Ogiinuur Sum in Arkhangai Province, Mongolia (Figure 1a) (Chen et al., 2015a). The funerary objects unearthed from this tomb are abundant and include pottery, gold headdresses, arched bronze pendants, bronze bracelets, bronze rings, bronze makeup appliances, leather bags, silk fabric, iron belts, iron arrowheads, iron swords and bows made of bone (Figure 1b, d). These objects are relatively rare in tombs characterized by nomadic cultures in Mongolia. The pottery in this tomb exhibits special shape characteristics and displays iconic artifacts of the Xianbei culture (Figure 1c) (Zheng, 2009). The arched bronze pendant on the neck of the tomb's owner was similar to the pendant found at the Chenwugou Xianbei site in Huade county, Inner Mongolia Autonomous Region (Figure 1e) (Hu et al., 2014). The ^{14}C dating of the wood on the coffin board was confirmed both at Peking University, China (cal. AD 590–655) and the University of Cologne, Germany (cal. AD 335–535) (Chen et al., 2015a). The absolute chronology of the tomb ranges from the fourth to sixth centuries AD, which is a time period coinciding with the Rouran Khan Kingdom on the Mongolian Plateau. The article by Chen et al. (2015a) confirmed that this site was a Rouran site, based on the unearthed funerary objects, the absolute chronology, and the unique cultural attributes and connotations of this tomb. The tomb was the first of this period to be found in Mongolia and has provided us with precious material for research into the genetic imprint of the Rouran.

The mtDNA hypervariable sequence I (HVS-I) and Y-chromosome SNPs were analyzed, and capture of the paternal non-recombining region of the Y chromosome (NRY) and whole-genome shotgun sequencing were performed on the IA-M1 remains (TL1) from the Khermen Tal site. To study the genetic imprint of the Rouran Khaganate and the paternal genetic characteristics of the Donghu-Xianbei lineage, eight additional samples from three sites (the Jinggouzi, Zaan khoshuu, and Gangga sites) that are representative of three ancient nomadic tribes (Donghu, Xianbei, and Shiwei) were selected for comparison with the TL1 individual.

2 | MATERIALS AND METHODS

All nine samples used in this study came from four archaeological sites (the Khermen Tal, Jinggouzi, Zaan khoshuu, and Gangga sites) (Table 1, samples named by TL, JGZ, ZHS, and GG, respectively). The use of these samples was approved by the Archaeological Institute of Inner Mongolia and the International Institute for Study of Nomadic Civilization, Mongolia. This study was also supported by the National Natural Science Foundation of China. Among these sites, the human remains at the Jinggouzi site (118.14°E, 43.23°N) were associated with the Donghu, one of the ancient nomadic populations in Northeast China dating back to 2500 years BP and (Lin, 1989) considered to be the earliest ancestors of the Mongol nationality (Wang et al., 2012). The Zaan

TABLE 1 List of the samples used in this study and their MtDNA nucleotide changes and Y-chromosome haplogroups

Sample	Site	Period	Radiocarbon (14C) measurement	Sex identification (molecular)	Mutations in mitochondrial HVR-I (160001+)	mtDNA haplogroups	Y chromosome haplogroups	Reference
TL1	Khermen Tal site	Rouran Khaganate	590–655 AD	Male	93–129-173–223-319–362	D	C2b1a/F3918	This study
ZHS1	Zaan khoshuu Site	Khitan	1010–1060 AD	Female	203–256-291–304	F	—	This study
ZHS2	Zaan khoshuu Site	Uighur	—	Female	223–290-319–362	A	—	This study
ZHS3	Zaan khoshuu Site	Unknown	—	Male	172–223-362	D	xK	This study
ZHS4	Zaan khoshuu Site	Liao and Jin	1020–1160 AD	Female	223–224-291–362	D	—	This study
ZHS5	Zaan khoshuu Site	Xianbei	585–660 AD	Male	171–223-298–327-344–357	C	C2b1a/F3918	This study
ZHSII-1	Zaan khoshuu Site	Turkic	—	Female	223–362-369	D	—	This study
GG3	Gangga site	Shiwei	800–1000 AD	Male	249–304-311	F	C2b1a/F3918	This study
JGZ6	Jinggouzi site	Donghu	2440–2530 BP	Male	—	—	C2b1a/F3918	(Zhang et al., 2018)

khoshuu site (104.19°E, 47.59°N), which includes six tombs in different eras, namely, Xianbei, Rouran, Turkic, Uighur and Khitan periods, is of important academic value for the study of history and culture of modern Mongolia (Chen et al., 2015b). Lastly, the Gangga site is located in the Chen Barag Banner (119.45°E, 49.33°N) of the Inner Mongolia region in Northeast China and is regarded as representing the nobility of the Shiwei population (Zhang et al., 2018).

2.1 | Ancient DNA extraction and laboratory environment

The materials chosen for this study were teeth that were well preserved and had no caries. The teeth were soaked in a 5% sodium hypochlorite solution for 15 min, rinsed with ddH_2O and 95% ethanol, and dried in a UV-irradiation box. Then, 200 mg of tooth powder was obtained from the dentin using a dental drill (Strong 90, Korea) and mixed with 4.5 mL of EDTA (0.5 mol, pH 8.0) and 80 μL of proteinase K (100 mg mL^{-1}), followed by incubation for 12–16 hr in a rotating hybridization oven at 55°C. After centrifugation of the samples at 7,500 rpm for 5 min, the supernatant was condensed to ~100 μL at 7,500 rpm using an ultrafiltration tube (Centricon®YM-10). Ancient DNA was extracted and eluted in an 80-μL volume using the QIAquick PCR Purification Kit (Qiagen, Germany) following the manufacturer's protocol. The DNA extraction procedure was performed at least twice for each sample, and a blank control was included for every five ancient samples.

2.2 | Mitochondrial DNA amplification, sex identification and Y-chromosome SNP typing

The mtDNA HVS-I region was amplified with two overlapping primer pairs spanning 333 bp (nucleotide positions 16,051–16,384). Five haplogroup-defining SNPs (M/N, A, C, D, and F) in the coding region of the mitochondrial genome were typed to determine the mtDNA lineages. The X-Y amelogenin gene was used to identify the sexes of all eight individuals (Hummel & Herrmann, 1991). Then, five Y-chromosome SNPs (M9, M130, M217, L1373, and F3918) were detected. All primer sequences used in this study are shown in Supporting Information S1.

2.3 | Library preparation

The DNA library was prepared according to the manufacturer's instructions (NEBNext® Ultra™ DNA Library Prep Kit for Illumina®). The USER (uracil-DNA-glycosylase (UDG) and endonuclease VIII), applied after dA-tailing and ligation, causes a small residual signal of C-to-T substitutions to be detected at the 5′ end, with no influence on G-to-A substitutions at 3′ terminal positions (Star et al., 2014). The adaptor-ligated DNA was cleaned up with the MinElute Reaction Cleanup Kit instead of Vortex AMPure XP Beads. After PCR amplification, the PCR products were assessed via 2% agarose electrophoresis. Next, quantitative analysis of the DNA library was conducted with Qubit 2.0 (Invitrogen, USA) and an Agilent 2100 Bioanalyzer (Agilent, USA).

TABLE 2 Sequencing metrics for the four NRY capture and whole genome shotgun sequencing libraries

Sample ID	Sequencing methods	Total reads	Mapped not clonal reads with $q > 30$ & length ≥ 35 bp	Human DNA	Mean length (bp)	Average depth of coverage	mt depth of coverage	chrY depth of coverage	mt Contamin. (upper bound)	X Contamin. (Test1/Test2)	mtDNA haplogroups (Over Quality)	Y chromosome haplogroups
TL1	Whole genome shotgun sequencing	72,081,872	9,621,783	13.35%	101.58	0.32x	58.33x	0.04x	0.017 ± 0.0031	NA	D4b1a2a1 (96.46%)	NA
	NRY capture sequencing	35,037,918	11,093,747	31.66%	110.29	0.40x	10.54x	1.61x	0.019 ± 0.0029	0.69%/ 1.12%	D4b1a2a (58.53%)	F3889, downstream of C2b1a1b/F3830 (ISOGG 2015)
ZHS3	NRY capture sequencing	20,833,339	447,696	2.15%	101.49	0.01x	255.05x	0.04x	0.012 ± 0.002	NA	D4m2 (92.50%)	NA
ZHS5	NRY capture sequencing	39,758,418	5,617,930	14.13%	106.73	0.19x	5.84x	0.68x	0.017 ± 0.0021	NA	C (59.43%)	F3889, downstream of C2b1a1b/F3830 (ISOGG 2015)
GG3	NRY capture sequencing	49,682,501	13,246,503	26.66%	99.88	0.43x	15.40x	1.31x	0.025 ± 0.0029	NA	F (59.59%)	F3889, downstream of C2b1a1b/F3830 (ISOGG 2015)

Test1/Test2: Test 1, which uses all high-quality reads per sample and then evaluates the rate of contamination based on a maximum likelihood approach; and Test 2, which samples a single read per site to calculate the contamination rate.

2.4 | Whole-genome shotgun sequencing and NRY capture sequencing

Whole-genome shotgun sequencing of the TL1 individual (accession number: PRJEB24670) was performed in collaboration with Novogene (Beijing, China). The resulting FastQ files were processed using AdapterRemoval ver. 2.1.0 (Schubert, Lindgreen, & Orlando, 2016), bwa 0.7.5a-r405 (Li & Durbin, 2009), and SAMtools 0.1.19 (Li et al., 2009). The coverage and statistical calculations were performed using Qualimap 2.2.1 (Okonechnikov, Conesa, & Garcia-Alcalde, 2016), and indels were realigned using GATK 3.2-2 (McKenna et al., 2010).

In this study, 11.38 Mb of targeted unique regions of the NRY chromosome were used to design the array. The target sequenced regions (hg19) are shown in Supporting Information S2. The NRY capture analysis was conducted in collaboration with Sierravast Bio-Medical (Shanghai, China). The data analysis method used for NRY capture was similar to the method used for shotgun sequencing.

2.5 | DNA authenticity

In our DNA libraries, most damage accumulated at the ends of the molecules was repaired as a result of being partially UDG treated. Because terminal CpG dinucleotides are unaffected by the UDG treatment when methylated, we estimated patterns of DNA damage in the CpG context using PMDtools (Skoglund et al., 2014) and observed that the damage to all samples exceeded ~11%, as is to be expected for ancient DNA samples (Supporting Information S3). The distribution of read lengths calculated with MapDamage2 (Jonsson, Ginolhac, Schubert, Johnson, & Orlando, 2013) showed that the mean read length of the shotgun library was 101 bp, while the fragments sequenced from the NRY capture libraries ranged from 100 to 110 bp (Table 2, Supporting Information S3). Next, the MIA (Mapping Iterative Assembler) was used to generate a consensus file, and contamMix was used to estimate contamination across the mitochondrial genome compared with a dataset of 311 contaminant sequences from around the world (Fu et al., 2013). The mtDNA contamination could be estimated for five samples, with the upper boundary ranging from 1.2 to 2.5% (Table 2). Then, we used ANGSD software, which implements the method first described in Rasmussen et al. 2010, to estimate contamination based on the X chromosome in the male individuals (Korneliussen, Albrechtsen, & Nielsen, 2014). The X chromosome contamination could be estimated for the TL1 individual as 0.69% (test 1) or 1.12% (test 2) (Table 2).

2.6 | Mitochondrial analysis and haplogroup assignment

After alignment of the whole genome, we used SAMtools view to extract the reads mapped to the rCRS to a single bam file. Next, we used SAMtools mpileup to call positions at which the base differed from that in the reference mitogenome (NC_012920). Only bases with a quality score > 30, read depth > 3 and call rate > 90% were considered in the analysis (Supporting Information S4). Furthermore, we used Tablet (Milne et al., 2013) to visualize the alignments and checked all of the polymorphic positions reported in the vcf file by eye. Haplogroup

assignment was based on Haplogrep2 (http://haplogrep.uibk.ac.at/) (Table 2). As a comparator, 29 haplogroup D4b1a Mitochondrial genomes were collected using MitoTool (ref. rCRS) (Supporting Information S4) (Fan & Yao, 2011, 2013).

2.7 | Y chromosome analysis

Y chromosome variants were called from the bam files for the three male samples showing greater than 0.1× genome coverage using SAMtools. Considering other published data (Lippold et al., 2014; Wei et al., 2017) (Supporting Information S5), 2447 SNPs were employed for SNP calling and age calculation. The details of the target sequenced region and SNP calling list are provided in Supporting Information S2. Only bases with a quality score > 30, read depth > 3 and call rate > 90% were included in the analysis (Supporting Information S5). Haplogroup assignments for each individual sample were determined with yhaplo-1.0.13 (Poznik, 2016).

2.8 | Nuclear data analysis

To prepare for population genetic analyses, data from 932 individuals from 57 populations in America, Central South Asia, East Asia, Europe, the Middle East, Oceania, and Africa, obtained from the HGDP dataset, were used as the modern DNA background (Supporting Information S6). The datasets were merged and filtered to exclude SNPs with a minor allele frequency (MAF) ≤ 0.5% and SNPs lacking a genotype call in ≥ 10% of individuals (Gamba et al., 2014). The final dataset contained 538,631 autosomal SNPs genotyped in 932 individuals. For the low-coverage genome (0.32x coverage; TL1), genotypes were obtained for principal component analysis (PCA) using SAMtools mpileup. The resulting pileup files were filtered for a minimum quality of 30. For positions with more than 1× coverage, one allele was randomly chosen with a probability equal to the frequency of the base at that position. This allele was duplicated to form a homozygous diploid genotype. Plink software (Purcell et al., 2007) was used to merge the ancient individual and the modern DNA background without allowing missingness, resulting in a total of 86,618 shared SNPs across all samples. These SNPs were used in downstream analyses such as PCA, Admixture, Treemix and f3 test evaluations.

The PCA results were assessed with GCTA-1.26.0 software (Yang, Lee, Goddard, & Visscher, 2011) and plotted using the R language (https://www.r-project.org/) with the PC1 and PC2 vectors. The analysis of ancestral genetic components was carried out using Admixture-1.3.0 (Alexander, Novembre, & Lange, 2009). We performed model-based clustering analysis via the maximum-likelihood approach, implemented in Admixture, assuming 2–24 ancestral populations ($K = 2$–24). The lowest cross-validation (CV) error was found for the model with nine modeled clusters or ancestral populations ($K = 9$). The clustering results ($K = 5$–13) were then visualized using R (Supporting Information S7). We also performed cluster analysis generated by Admixture for the ancient TL1 individual and 428 modern Asians from 26 populations ($K = 3$ with lowest CV error). We applied TreeMix (Pickrell & Pritchard, 2012) to the East Asia populations from the modern DNA background employed in PCA to infer maximum likelihood trees and admixture graphs. Mbuti Pygmies (Mbuti) was specified as outgroup and 1,000 bootstrap replicates were generated to produce a consensus tree with block size of 500 SNPs. In addition, the outgroup f3 statistics of the f3 form (Mbuti; East Asia, ancient) were computed with Mbuti as the outgroup using the implementation distributed with TreeMix (Supporting Information S8).

2.9 | Phylogenetic analyses

We used the BEAST2 software package (Bouckaert et al., 2014) to reconstruct phylogenetic trees, estimate coalescent times and generate Bayesian skyline plots (BSPs) for the NRY data and the mitochondrial genomes. To reduce the computational load, the ChrY BEAST analysis included only the variable positions, and the phylogenetic tree of mitochondrial genomes was reconstructed with MEGA 5.05 (Hall, 2013) using the maximum likelihood tree model. The general time reversible (GTR) substitution model was selected as the best fit for the NRY data, and the HKY + I + G model was used for the mitochondrial genomes (Karmin et al., 2015). Then, the fossilized birth-death model (Heath, Huelsenbeck, & Stadler, 2014) was used to set up the divergence dating analysis, and the Bayesian coalescent skyline was used to obtain the prior sex-specific effective population sizes. BEAST analyses of the NRY data were run using a relaxed lognormal clock with a mutation rate of 0.74×10^{-9}/bp/year and a strict clock model with a rate of 1.665×10^{-8}/bp/year for the mtDNA (Karmin et al., 2015; Soares et al., 2009).

3 | RESULTS

3.1 | Genetic analysis of the samples

The mitochondrial haplogroups of the nine samples were A, C, D, and F based on Sanger sequencing (Table 1). The TL1 sample was further attributed to the mitochondrial haplogroup D4b1a2a1 (overall quality: 96.46%) through whole-genome sequencing analysis (Table 2). ZHS3, ZHS5, and GG3 were further attributed to the mitochondrial haplogroups D4m2 (overall quality: 92.50%), C (overall quality: 59.43%), and F (overall quality: 59.59%) through NRY capture sequencing analysis (Table 2). Five of the nine samples were identified as male. Four of the five male specimens (TL1, ZHS5, GG3, and JGZ6) were attributed to haplogroup C2b1a/F3918 [ISOGG 2015 (https://isogg.org/tree/2015/index15.html)] (Table 1) by Sanger sequencing. After NRY capture sequencing, three of the male samples (TL1, ZHS5, and GG3, excluding JGZ6, which was not sequenced in this study) were attributed to C2b1a1b/F3830 (ISOGG 2015) and further to F3889, which was reported to be downstream of C2b1a1b/F3830 by Wei et al. (2017) (Table 2). The ZHS3 individual failed to be detected by either Sanger or NRY capture sequencing.

3.2 | Genetic analysis of nuclear DNA

Principal component analysis showed that the ancient TL1 individual mainly clustered with the East Asia populations, being located closer to modern Mongola (ethnic Mongols from southern parts of Hulunbuir)

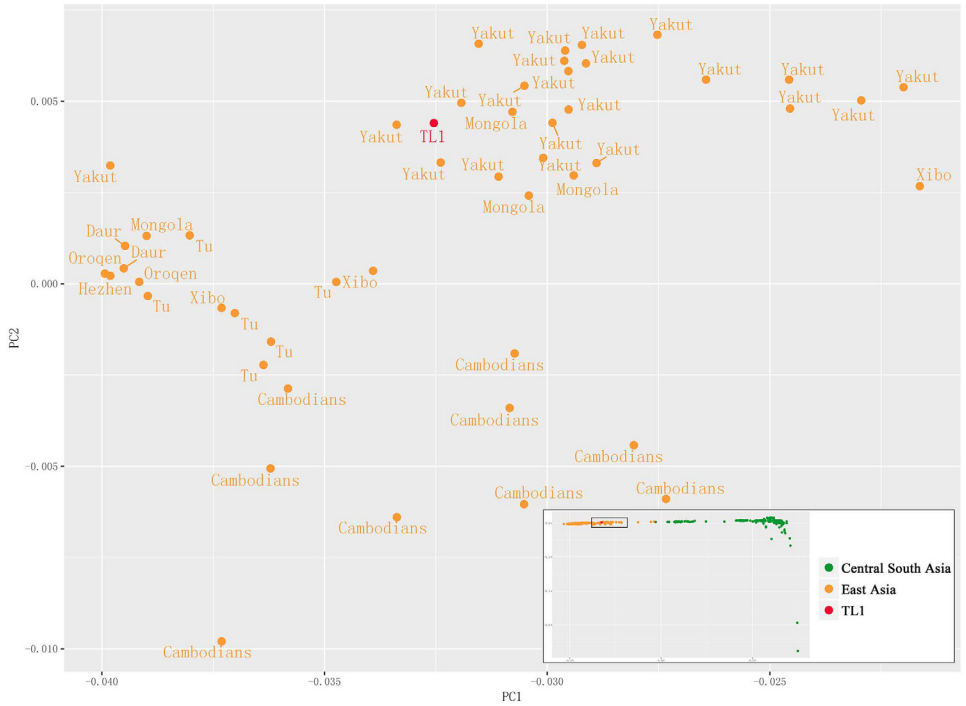

FIGURE 2 PCA plot of the TL1 individual (red) and 428 modern Asians in the HGDP databases

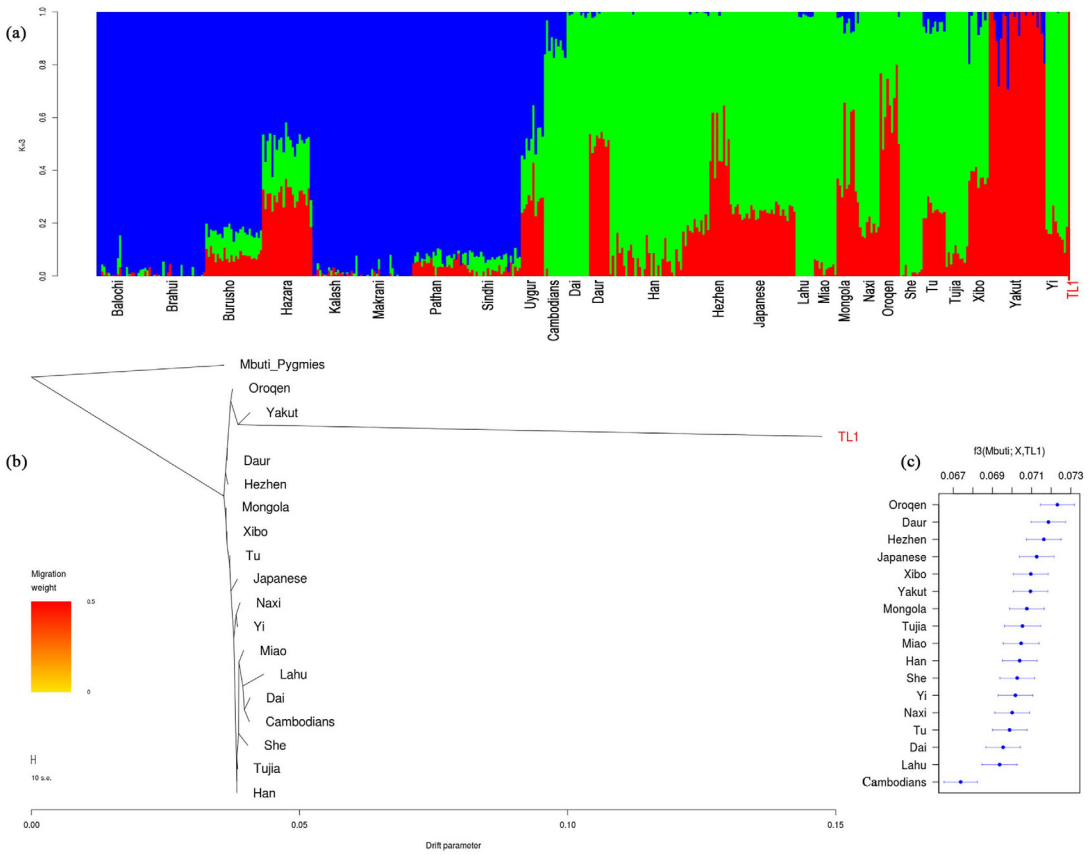

FIGURE 3 The genetic affinity of the ancient TL1 individual and the East Asia populations (a) Cluster analysis generated by Admixture for the ancient TL1 individual and 428 modern Asians from 26 populations ($K = 3$). (b) TreeMix phylogeny of TL1 along with the East Asia populations representing population divergence patterns. (c) Ranked outgroup f3 statistics examining represents the amount of shared genetic drift between TL1 and each of 17 contemporary East Asia populations since their divergence with the Mbuti Pygmies (Mbuti) population

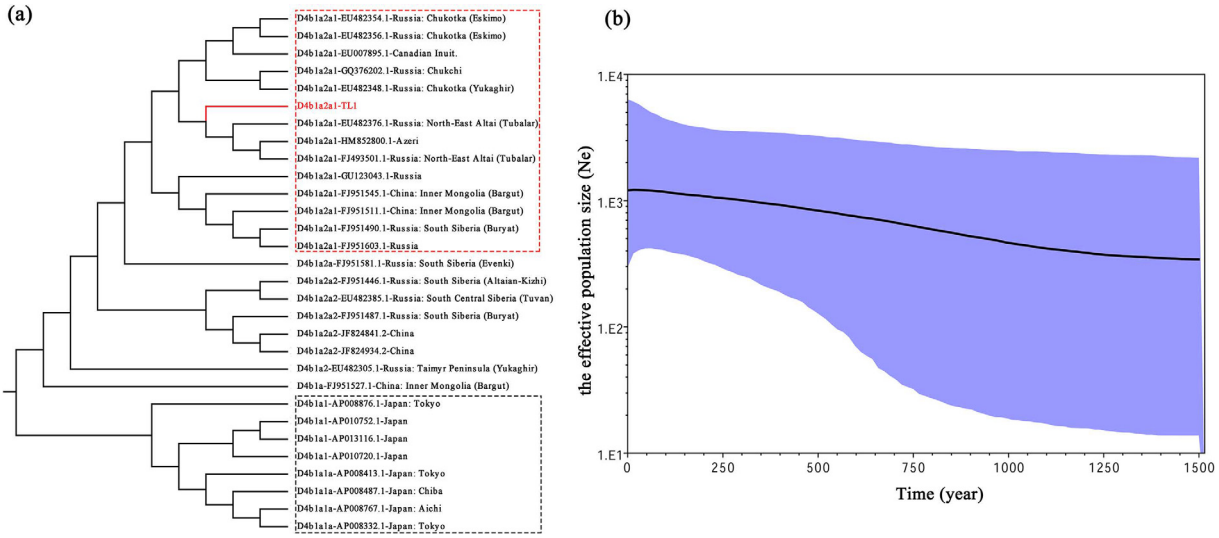

FIGURE 4 The mtDNA phylogenetic tree of haplogroup D4b1a (a) and the Bayesian Skyline Plot (BSP) based on the whole mitochondrial genome of the 14 samples under haplogroup D4b1a2a1 (b)

and Yakut than the others (Figure 2). Admixture analysis at $K = 9$ (Supporting Information S7), where we projected the TL1 individual based on modern reference data, revealed that the TL1 individual was similar to other East Asia samples predominantly exhibiting a mint green component and a small amount of violet. In addition, cluster analysis (Figure 3a) generated by Admixture for the TL1 individual and modern Asians ($K = 3$) shown that the Central South Asia populations were mainly composed of blue group while green group mainly occupied East Asia samples. The TL1 individual was most similar to Yakut people predominantly exhibiting red group, followed by Oroqen, Daur, Hezhen, and Mongola. Using the probabilistic model of population splits and divergence implemented in TreeMix, we found that the TL1 individual was very close to the Yakut and Oroqen (Figure 3b). The f3 analyses (Mbuti; East Asia, ancient), which quantified the amount of shared genetic drift between the Eastern Asia individuals and the ancient TL1 individual since they diverged from an African outgroup Mbuti, showed that the highest amount of shared drift with the TL1 individual came from Oroqen, followed by Daur (Figure 3c).

3.3 | Phylogenetic analysis of mitochondrial DNA

A phylogenetic tree was constructed using the whole mitochondrial genome, which showed that the D4b1a2a1 haplotype of the TL1 individual was distributed mainly across Russia and the Inner Mongolia Region

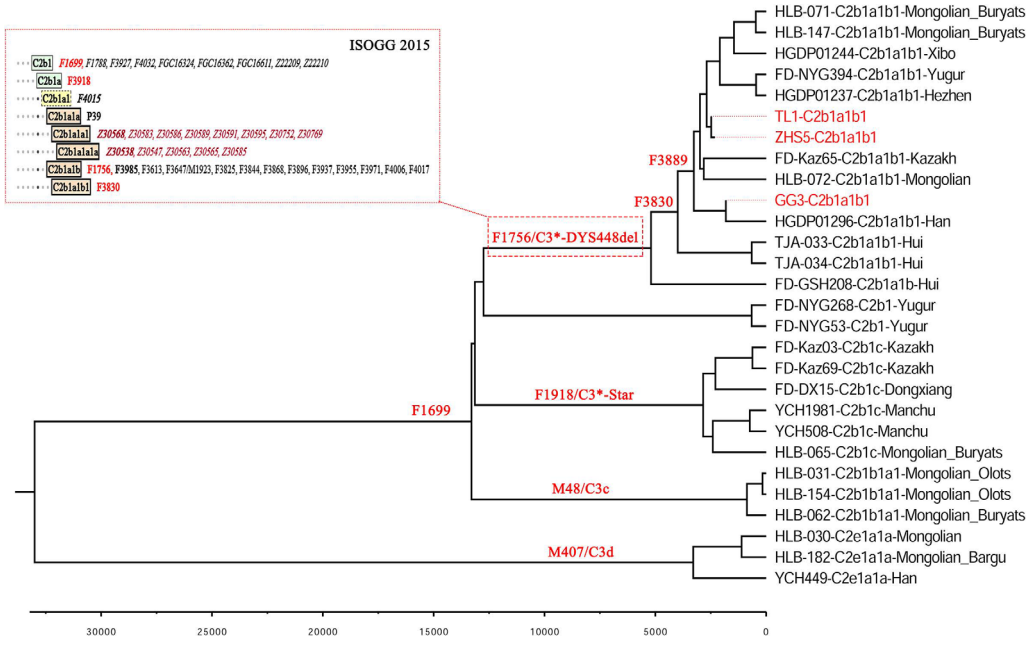

FIGURE 5 The ancient males (TL1, ZHS5, and GG3) was grouped under C2b1a1b1/F3880 on the Y-DNA haplogroup C lineage using BEAST

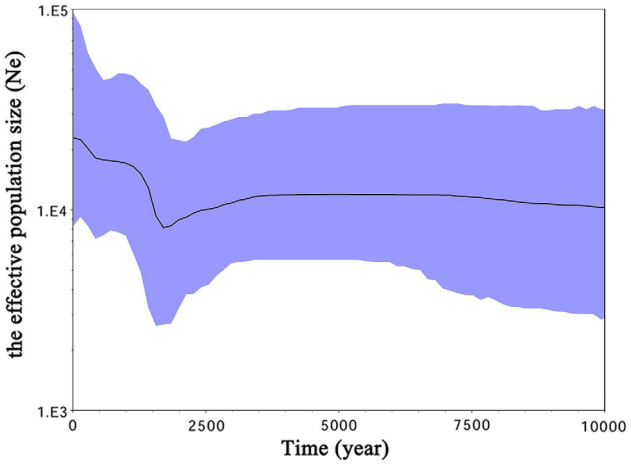

FIGURE 6 Bayesian Skyline Plot (BSP) based on the Hg C2 (proto C3) NRY data, indicating that the median of the hypothetical male effective population size changed over time

of China (Figure 4a, Supporting Information S4). By estimating the effective population size (Ne) using BEAST, we found that the D4b1a2a1 haplotype underwent a significant increase ~750 years ago (Figure 4b).

3.4 | Phylogenetic analysis of the Y chromosome

The phylogenetic tree constructed using the NRY capture data from the three ancient samples and 25 modern reference samples belonging to haplogroup C2 (proto C3) (Supporting Information S5) (Lippold et al., 2014; Wei et al., 2017) showed that the ancient individuals were located together on a branch (F3889) with modern Mongolian-Buryats, Xibo, Yugur, Hezhen, Kazakh, and Han individuals (Figure 5). Analysis of the BSP for the NRY data showed that Hg C2 (proto C3) experienced a bottleneck ~2000 years ago and that the effective male population size increased two times (~500-year-old and ~1500-year-old) over the subsequent 2000 years (Figure 6).

4 | DISCUSSION AND CONCLUSIONS

4.1 | Genetic imprint of the Rouran Khaganate

In our research, the mitochondrial haplotype of the TL1 individual was D4b1a2a1, which is downstream of haplogroup D4, the most frequently occurring mtDNA haplogroup among modern populations of northern East Asia, including the Japanese, Okinawans, Koreans, and Mongolic- or Tungusic-speaking populations of northern China (Derenko et al., 2012; Kong et al., 2003; Lee et al., 2006; Maruyama, Minaguchi, & Saitou, 2003; Umetsu et al., 2005; Zheng et al., 2011). Previous studies have also shown that haplogroup D4 was the most common haplogroup among the ancient populations of northern China, such as those associated with the Qilangshan, Dondajing, and Lamadong sites (Wang et al., 2007; Yu, Xie, Zhang, Zhou, & Zhu, 2006; Yu, Zhao, & Zhou, 2014). The research of Derenko et al. (2010) implied that D4b1a2a1, downstream of haplogroup D4b1a2, originated in southern Siberia, and the age of D4b1a2a1 was estimated at ~11kya (Derenko et al., 2010). From the phylogenetic tree of the mitochondrial genomes, we found that haplogroup D4b1a2a1 of the TL1 individual was mainly distributed in Russia and Inner Mongolia, China. Combined with the results of previous research, we speculate that mtDNA haplogroup D4b1a2a1 most likely originated from the Mongolian grasslands. In addition, PCA based on nuclear data showed that the ancient TL1 individual shared high similarity with contemporary Yakut and Mongola populations. We also observed that TL1 exhibited a very similar composition to the Yakut, Oroqen, Daur, Hezhen, and Mongola populations according to admixture analysis. Moreover, Treemix and f3 test analyses revealed that TL1 displayed a close genetic relationship to the Oroqen and Daur populations. All of these results indicated that the TL1 individual made an important genetic contribution to modern Mongolic- and Manchu-speaking peoples.

The Y-chromosome haplotype of the TL1 individual was F3889, positioned downstream of C2b1a1b/F3830 (ISOGG 2015), which accounts for a very low proportion of haplogroup C2b1a/F3918. The Y-chromosome haplotypes of ZHS5 (from the Zaan khoshuu site) and GG3 (from the Gangga site) also belonged to F3889. The ZHS5 sample, from the Zaan khoshuu site, came from the Xianbei period. The Gangga people belonged to the Shiwei culture, which might also have originated from the Xianbei population. All of these people were related to the ancient Donghu-Xianbei branch. These results suggested that TL1 likely presents a close paternal relationship to the Donghu people and may have even descended from a branch of the ancient Donghu-Xianbei people, based on the conclusion that haplogroup C2b1a/F3918 can be considered the paternal branch of the ancient Donghu people (Zhang et al., 2018). The Y-chromosome phylogenetic tree showed that TL1 shared a branch with modern Mongolian-Buryats, Hezhen, Xibo, Yugur, and Kazakh, suggesting that the TL1 individual from the Rouran period should also generally present close paternal genetic relationships with modern Mongolic- and Manchu-speaking peoples.

In general, the Rouran Khaganate originated from an alliance of the ancient Eurasian steppe nomads, which disintegrated and disappeared with the progress of history. This group was complex, and its origin cannot be explained based only on one individual. However, we can trace the genetic imprint of the Rouran people through genome analysis of the TL1 individual. On the basis of the comparison with other ancient nomadic people (Donghu, Xianbei, and Shiwei) and data on modern individuals from published articles (Lippold et al., 2014; Wei et al., 2017) (Supporting Information S5), we found that they all share the same haplotype implying shared paternal ancestry between the Donghu, Xianbei and Rouran populations. Furthermore, this gene flow (mainly haplogroup C2b1a/F3918) did not stop with the disappearance of the Rouran, and a portion was instead passed on in other groups, such as the ancient Shiwei people (later than Rouran), eventually reaching the gene pool of modern Mongolic- and Manchu-speaking populations (Mongolian-Buryats, Hezhen, Xibo, et al).

4.2 | Important paternal lineage in Donghu nomads: F3889

Hg C3* is widely distributed in Eurasia, as an important paternal lineage with a typical paternal haplogroup of Mongolic-speaking populations,

and shows a moderate distribution among other Tungusic peoples (Malyarchuk et al., 2010). This group is not monophyletic and has several sub-haplogroups, among which the C3*-Star cluster (C2b1c/F1918, ISOGG 2015) presents an extremely high frequency in the Mongolian tribes and Turkic peoples and is thought to be the paternal haplotype of Gen Gi Khan or his male relatives (Zerjal et al., 2003). The C3*-DYS448del (C2b1a1b/F1756, ISOGG 2015) cluster is widely distributed at a low frequency in almost all Mongolic- and Turkic-speaking populations. Wei et al. (2017) proposed that the C3*-DYS448del cluster might be a candidate for the paternal lineages of the ancient Donghu, Xianbei, and Shiwei tribes (Wei et al., 2017). However, these authors lacked data on ancient DNA to verify this hypothesis. Through the NRY capture analysis conducted in the present study, the TL1 individual and other two ancient samples [ZHS5 (~1400 years old, belonging to Xianbei) and GG3 (~1100 years old, belonging to Shiwei)] were allocated to haplotype F3889, which was the lower part of C2b1a1b/F1756 (proto C3*-DYS448del). Moreover, we found that C3*-Star separated early from C3*-DYS448del in the Y-chromosome phylogenetic tree, whereas the divergence time between C3*-DYS448del and its most closely related lineage was ~13 kya. The expansion time for branch F3889 (~3.3 kya), downstream of C3*-DYS448del, was much earlier than that of the C3*-Star cluster (~2.8 kya). This period corresponded only to the reign of the ancient Donghu nomads (from the early Shang Dynasty to the Western Han Dynasty) in northeast China. Our results supported the inference drawn by Wei et al. (2017) and further indicated that the F3889 branch, downstream of C2b1a1b/F1756 (proto C3*-DYS448del), was an important paternal lineage in Donghu-Xianbei nomads before the expansion of the Mongols.

4.3 | History of nomadic populations in the Mongolian grasslands

On the basis of the phylogenetic tree and BSPs of the NRY data, we found that the male effective population size of Hg C2 (proto C3) underwent three changes, which was consistent with the historical scenario. The male effective population size declined greatly ~2500 years ago, perhaps through decay of the Donghu and invasion of the Xiongnu. The population size subsequently increased ~1500 years ago, probably due to the establishment of the Xianbei and the Rouran. The male effective population size then underwent a second increase ~500 years ago. In addition, the female effective population size increased ~750 years ago according to the BSPs of the whole mitochondrial genomes. This period was also the timeframe in which the Mongol Empire (1206–1405) was established. We speculate that the increase in both the maternal and paternal effective population sizes was influenced by the establishment and expansion of the Mongolian Empire. When our results are combined with those of previous studies, we can infer that the main groups present in the eastern Mongolian grasslands belonged to ancient populations with the paternal lineage Hg C3, and especially the paternal lineage C3*-DYS448del. These groups were once the dominant populations in the eastern part of the Mongolian Plateau. As time went on, these people formed many different nomadic groups. These political groups were composed of different tribes and had similar genetic structures, although with different rulers and different names. All of these groups greatly influenced the genetic composition of modern nomadic populations in the Mongolian grasslands.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China, under grant number 31371266 (Grant Recipient: Hui Zhou). Jiawei Li and Ye Zhang conceived and designed the study. Jiawei Li, Ye Zhang, and Yongbin Zhao performed the analysis and wrote the manuscript. Yongzhi Chen, A. Ochir, and Sarenbilige participated in the archaeological dig and provided research materials. Hong Zhu and Hui Zhou helped in interpreting the results and improving the manuscript. All authors read and approved the final manuscript. The authors thank Dr. Philip Johnson for providing the contamMix program and Pengfei Cheng from the Archaeological Institute of Inner Mongolia, P.R. China for providing the relevant background anthropological information to them.

ORCID

Hui Zhou http://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*, 1655–1664.

Barbara, A. W. (2008). *Encyclopedia of the Peoples of Asia and Oceania*. New York, NY, USA: Facts on File.

Bouckaert, R., Heled, J., Kühnert, 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*, e1003537.

Chen, Y. Z., Sarenbilige, Cheng, P. F., Dantal, Ochir, A., Ankhabayar, B. … Angragsuren, (2015a). Excavation report on Khermen Tal IA-M1 site in Ugiinuur Sum, Arkhangai Province, Mongolia, 2014. *Steppe Cultural Rellcs*, *02*, 13. https://doi.org/10.16327/j.cnki.cn15-1361/k.2015.02.003

Chen, Y. Z., Sarenbilige, Cheng, P. F., Dantal, Ochir, A., Ankhabayar, B. … Angragsuren, (2015b). Excavation Report on Zaan khoshuu Site in Dashinchilen Sum of Bulgan Province, Mongolia. *Steppe Cultural Relics*, *02*, 25. https://doi.org/10.16327/j.cnki.cn15-1361/k.2015.02.002

Derenko, M., Malyarchuk, B., Denisova, G., Perkova, M., Rogalla, U., Grzybowski, T., … Zakharov, I. (2012). Complete mitochondrial DNA analysis of eastern Eurasian haplogroups rarely found in populations of Northern Asia and Eastern Europe. *Plos One*, *7*, e32179.

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*, e15214.

Fan, L., & Yao, Y. G. (2011). MitoTool: a web server for the analysis and retrieval of human mitochondrial DNA sequence variations. *Mitochondrion*, *11*, 351–356.

Fan, L., & Yao, Y. G. (2013). An update to MitoTool: using a new scoring system for faster mtDNA haplogroup determination. *Mitochondrion*, *13*, 360–363.

Fu, Q., 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*, 553–559.

Gamba, C., Jones, E. R., Teasdale, M. D., McLaughlin, R. L., Gonzalez-Fortes, G., Mattiangeli, V., ... Pinhasi, R. (2014). Genome flux and stasis in a five millennium transect of European prehistory. *Nature Communications*, *5*, 5257.

Hall, B. G. (2013). Building phylogenetic trees from molecular data with MEGA. *Molecular Biology and Evolution*, *30*, 1229–1235.

Heath, T. A., Huelsenbeck, J. P., & Stadler, T. (2014). The fossilized birth-death process for coherent calibration of divergence-time estimates. *Proc Natl Acad Sci USA*, *111*, E2957–E2966. https://doi.org/10.1073/pnas.1319091111

Hu, X. N., Bao, Q. C., Zhang, X. X., Zhang, T. Y., Wang, Y. S., Cheng, X. F., & Li, C. (2014). Archaeological excavation of the Chenwugou Xianbei site in Huade county. *Steppe Cultural Rellcs*, *1*, 23. https://doi.org/10.16327/j.cnki.cn15-1361/k.2014.01.007

Hummel, S., & Herrmann, B. (1991). Y-chromosome-specific DNA amplified in ancient human bone. *Die Naturwissenschaften*, *78*, 266–267.

J, S. (1985). The origin of Shiwei. *Seeking Truth*, *1*, 88–94.

Jing, W. (2013). *Rouran research*. (Master), Shanxi University.

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*, 1682–1684.

Karmin, M., Saag, L., Vicente, M., Sayres, M. A. W., Järve, M., Talas, U. G., ... Kivisild, T. (2015). A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Research*, *25*, 459–466.

Kong, Q. I.-P., Yao, Y. O.-G., Liu, M. U., Shen, S. H.-P., Chen, C., Zhu, C. H.-L., ... Zhang, Y. A.-P. (2003). Mitochondrial DNA sequence polymorphisms of five ethnic populations from northern China. *Human Genetics*, *113*, 391–405.

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics*, *15*, 356.

Lee, H. Y., Yoo, J. E., Park, M. J., Chung, U., Kim, C. Y., & Shin, K. J. (2006). East Asian mtDNA haplogroup determination in Koreans: haplogroup-level coding region SNP analysis and subhaplogroup-level control region sequence analysis. *Electrophoresis*, *27*, 4408–4418.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* , *25*, 2078–2079.

Lin, G. (1989). *A history of Donghu*. Hohhot: Inner Mongolia People's Publishing House.

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*, 13.

Malyarchuk, B., Derenko, M., Denisova, G., Wozniak, M., Grzybowski, T., Dambueva, I., & Zakharov, I. (2010). Phylogeography of the Y-chromosome haplogroup C in northern Eurasia. *Annals of Human Genetics*, *74*, 539–546.

Maruyama, S., Minaguchi, K., & Saitou, N. (2003). Sequence polymorphisms of the mitochondrial DNA control region and phylogenetic analysis of mtDNA lineages in the Japanese population. *International Journal of Legal Medicine*, *117*, 218–225.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., ... DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*, 1297–1303. https://doi.org/10.1101/gr.107524.110

Milne, I., Stephen, G., Bayer, M., Cock, P. J. A., Pritchard, L., Cardle, L., ... Marshall, D. (2013). Using tablet for visual exploration of second-generation sequencing data. *Briefings in Bioinformatics*, *14*, 193–202.

Okonechnikov, K., Conesa, A., & Garcia-Alcalde, F. (2016). Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*, *32*, 292–294.

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *Plos Genetics*, *8*, e1002967.

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

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., ... Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*, 559–575.

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 USA*, *111*, 2229–2234.

Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Röhl, A., ... Richards, M. B. (2009). Correcting for purifying selection: An improved human mitochondrial molecular clock. *American Journal of Human Genetics*, *84*, 740–759.

Star, B., Nederbragt, A. J., Hansen, M. H. S., Skage, M., Gilfillan, G. D., Bradbury, I. R., ... Jentoft, S. (2014). Palindromic sequence artifacts generated during next generation sequencing library preparation from historic and ancient DNA. *Plos One*, *9*, e89676.

Umetsu, K., Tanaka, M., Yuasa, I., Adachi, N., Miyoshi, A., Kashimura, S., ... Osawa, M. (2005). Multiplex amplified product-length polymorphism analysis of 36 mitochondrial single-nucleotide polymorphisms for haplogrouping of East Asian populations. *Electrophoresis*, *26*, 91–98.

Wang, H. J., Chen, L., Ge, B. W., Zhang, Y., Zhu, H., & Zhou, H. (2012). Genetic data suggests that the Jinggouzi people are associated with the Donghu, an Ancient Nomadic Group of North China. *Human Biology*, *84*, 365–378.

Wang, H., Ge, B., Mair, V. H., Cai, D., Xie, C., Zhang, Q., ... Zhu, H. (2007). Molecular genetic analysis of remains from Lamadong Cemetery, Liaoning, China. *American Journal of Physical Anthropology*, *134*, 404–411.

Wei, L. -H. A., Huang, Y. -Z. H., Yan, S. H., Wen, S. H.-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*, 915–918.

Yang, J. A., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics*, *88*, 76–82.

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*, 6242–6246.

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

Zerjal, T., Xue, Y., Bertorelle, G., Wells, R. S., Bao, W., Zhu, S., ... Tyler-Smith, C. (2003). The genetic legacy of the Mongols. *American Journal of Human Genetics*, *72*, 717–721.

Zhang, Q., Sun, Z. C., Li, M. C., Wang, L. S., Zhang, Q. C., & Zhu, H. (2014). A bioarchaeological study of human remains from Jinggouzi, a Bronze Age cemetery site in North China. *American Journal of Physical Anthropology*, *153*, 282–282.

Zhang, Y., Wu, X., Li, J., Li, H., Zhao, Y., & 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*, 231–238. https://doi.org/10.1038/s10038-017-0357-z

Zheng, H. X. (2009). A study of Xianbei tombs. *Acta Archaeologica Sinica*, *3*, 30.

Zheng, H. X., Yan, S., Qin, Z. D., Wang, Y., Tan, J. Z., Li, H., & Jin, L. (2011). Major population expansion of East Asians began before neolithic time: Evidence of mtDNA genomes. *Plos One*, *6*, e25835.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Li J, Zhang Y, Zhao Y, et al. The genome of an ancient Rouran individual reveals an important paternal lineage in the Donghu population. *Am. J. Phys. Anthropol.* 2018;166:895–905. https://doi.org/10.1002/ajpa.23491