Mitochondrial Genome of an 8,400-Year-Old Individual from Northern China Reveals a Novel Subclade Under C5d

Authors: Wu, Xiyan, Ning, Chao, Bao, Qingchuan, Gao, Shizhu, Zhang, Fan, et. al.

Source: Human Biology, 91(1) : 21-30

Published By: Wayne State University Press

URL: https://doi.org/10.13110/humanbiology.91.1.04

BioOne Complete (complete.BioOne.org) is a full-text database of 200 subscribed and open-access titles in the biological, ecological, and environmental sciences published by nonprofit societies, associations, museums, institutions, and presses.

Your use of this PDF, the BioOne Complete website, and all posted and associated content indicates your acceptance of BioOne’s Terms of Use, available at www.bioone.org/terms-of-use.

Usage of BioOne Complete content is strictly limited to personal, educational, and non-commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

BioOne sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.
Mitochondrial Genome of an 8,400-Year-Old Individual from Northern China Reveals a Novel Subclade under C5d

Xiyan Wu,1,2 Chao Ning,3 Qingchuan Bao,4 Shizhu Gao,5 Fan Zhang,2 Sihao Wu,2 Tianjiao Li,2 Linyuan Fan,2 Tao Li,6 Xuan Yang,2 Dawei Cai,1,7* and Yinqiu Cui 1,2,8*

Abstract

Ancient DNA studies have always refreshed our understanding of the human past that cannot be tracked by modern DNA alone. Until recently, ancient mitochondrial genomic studies in East Asia were still very limited. Here, we retrieved the whole mitochondrial genome of an 8,400-year-old individual from Inner Mongolia, China. Phylogenetic analyses show that the individual belongs to a previously undescribed clade under haplogroup C5d that most probably originated in northern Asia and may have a very low frequency in extant populations that have not yet been sampled. We further characterized the demographic history of mitochondrial haplogroups C5 and C5d and found that C5 experienced a sharp increase in population size starting around 4,000 years before present, the time when intensive millet farming was developed by populations who are associated with the Lower Xiajiadian culture and was widely adopted in northern China. We caution that people related to haplogroup C5 may have added this farming technology to their original way of life and that the various forms of subsistence may have provided abundant food sources and further contributed to the increase in population size.

Mitochondrial DNA (mtDNA) possesses several favorable characteristics, including strictly maternal inheritance, multiple copies in the cell, high mutation rate compared with nuclear DNA, small genome size, and high level of sequence polymorphisms. All these features make mtDNA a unique and frequently used marker to explore population genetic diversity and structures (González-Martín et al. 2015; Kivisild 2015; Postillone and Perez 2017; Ricaut et al. 2006; Stoneking 1994). In recent years, much insight has been gained into the prehistory of populations in East Asia by studying short fragments of mtDNA (hypervariable regions [HVRs]) and, increasingly, whole mitochondrial genomes of present-day populations. However, studies of ancient DNA (aDNA) in other regions have always refreshed our understanding of the human past that cannot
be tracked from the modern DNA alone, because of frequent population replacements and interbreeding (Reich 2018; Sikora et al. 2019; Slatkin and Racimo 2016). In this sense, aDNA provides a direct time transect and plays a pivotal role in understanding East Asian prehistory, where only very limited ancient mitochondrial genomes are available (Fu et al. 2013; Ning et al. 2016).

The Yumin (YM) site is located in Ulanchap, Inner Mongolia, China (Fig. 1). The site was first excavated in 2010 by a joint effort of the Inner Mongolia Autonomous Region Institute of Cultural Relics and Archaeology, Ulanchap Municipal Museum, and Huade County Administration of Cultural Relics; later, between 2014 and 2016, a second season of excavation was carried out by the same group. A total of 14 house foundations, 1 ash ditch, and 1 tomb were found inside the site. The house foundations were round and semisubterranean, with a round hearth standing in the middle, without traces of gateway or obvious postholes. The ash ditch was at the southwest of the excavation area, in the direction of southeast to northwest, and the tomb was under the floor of house F1. It is an earthen shaft pit tomb with completely well-preserved human bones and no grave goods, dated to 8,400 years before present (BP) (Fig. 1).

The number of cultural relics unearthed from YM was relatively small (1,500 pieces in total), most of which are stone tools, followed by some pottery pieces and bone tools. The stone tools were made of gray and black mudstone (the majority), sandstone (second largest number), flint, quartzite, and other types (Fig. 1). The shapes of the stone tools vary, including semicircular shovels, flake choppers, spear-shaped tools, stone cores, and stone balls. Few of the pottery pieces could be restored; most pottery relics were small pottery shards, including sand inclusion yellow and brown pottery and black and brown pottery, made using the mud-piece pasting technique. The shapes of the pottery included round-bottom and barrel-shaped jars, sharp and round-bottom *fu* (caldrons), flake objects, and flared-mouth jars (Dang 2015, 2017).

All the above evidence shows that the YM people lived a relatively primitive way of life in the transition period from the Mesolithic to Neolithic and belonged to a culture that is different from other contemporaneous and later cultures from the nearby regions. They thus provide unique and valuable material in understanding the prehistory of populations in this region (Hu 2017).

Here we targeted- enriched and sequenced the complete mitochondrial genome of the only...
human remain from the YM site. The individual is female, based on physical anthropology analysis; was dated to 8,400 years BP; and was assigned to mitochondrial haplogroup subclade C5. By comparison with a large data set of mtDNA publicly available that belonged to haplogroup C5, we reconstructed the phylogeny and the demographic history of haplogroup C5 and further tracked the relations between the ancient and present-day populations in East Asia.

Materials and Methods

Archaeological Context
The YM site is located in the Huade County, Ulanchap, Inner Mongolia Autonomous Region, China (42°3′ N, 114°18′ E). Between 2010 and 2016, two seasons of excavations were conducted by a joint effort of the Inner Mongolia Autonomous Region Institute of Cultural Relics and Archaeology, the Ulanchap Municipal Museum, and the Huade County Administration of Cultural Relics. Only one intact human remain (individual HDY) was excavated in the site and was radiocarbon dated to 8,400 BP.

Ancient DNA Lab Works
Two well-preserved molars were collected from the HDY individual for aDNA analyses. All aDNA analyses were carried out following the strict standards specially designed to minimize the potential modern DNA contamination (Adler et al. 2011; Llamas et al. 2016). Specifically, the teeth samples were immersed in a 5% liquid sodium hypochlorite (bleach) for 10 min and were then washed by ultrapure water and 100% ethanol, followed by the ultraviolet irradiation on each side of the teeth. Then 150-mg fine tooth powder was prepared by using a dental drill (STRONG 90) (Li et al. 2011; Zhang et al. 2018). DNA was extracted from the teeth in a dedicated clean room specially designed for aDNA studies at the aDNA lab of Jilin University, using an in-solution silica-based protocol (Li et al. 2010; Yang et al. 1998).

The double-stranded libraries were built by the NEBNext Ultra DNA Library Prep Kit (New England Biolabs, Beijing, China) following the manufacturer’s instructions, but with minor corrections. Specifically, we diluted the adaptor (15 μM) to 1.5 μM with a 10-fold dilution (1:9) in sterile water for immediate use. The libraries were then purified using Agencourt Ampure XP Bead (Beckman Coulter) 1:1.5 DNA-to-bead ratio and were quantified using the Qubit fluorometer (Life Technologies, Paisley, UK) and the Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA). mtDNA enrichment was performed using the MyGenostics Human Mitochondria Capture Kit (MyGenostics Inc., Beijing, China). The postcapture libraries were then amplified for 15 PCR cycles and were sequenced on an Illumina HiSeq X10 platform.

Genetic Data Processing
The raw fastq data were processed by EAGER, version 1.92.50, an automated computational pipeline specially designed for aDNA data processing (Peltzer et al. 2016). Quality filtering was performed with FastQC software (Babraham Bioinformatics; http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and the adapters were trimmed with AdapterRemoval, version 2.2.0 (Schubert et al. 2016). The reads were then aligned to the revised Cambridge Reference Sequence (rCRS; accession NC_012920) using BWA 0.7.12, with only sequence lengths larger than 30 bp considered, and the duplications were removed using the default parameter of the DeDup software (Li and Durbin 2009; Li et al. 2009). Single-nucleotide polymorphisms (SNPs), insertions, and deletions (INDELs) were called using SNVer-0.5.2 (Wei et al. 2011). The SNPs and the INDELs were then double confirmed by visual inspection with the Integrative Genomics Viewer (Thorvaldsdóttir et al. 2013). We then applied the default parameters of mapDamage 2.0 to determine the molecular damage that is typical of aDNA (Jónsson et al. 2013).

Phylogenetic Analysis and Coalescent Simulations of the HDY DNA
We prepared a large data set of complete mitochondrial sequences that belong to haplogroup C5 from the MitoTool database (http://mitotool.kiz.ac.cn) and the MitoMap database (https://www.mitomap.org/MITOMAP), as well as those from the literature (Derenko et al. 2010, 2012; Dryomov et al. 2019; Li et al. 2019; Mielnik-Sikorska et al. 2013). The sequences were then aligned to the rCRS using BioEdit, version 14.0 (Hall 2011). We constructed the phylogenetic tree of haplogroup...
C5 with mtPhyl software (Eltsov and Volodko 2011; https://sites.google.com/site/mtphyl/home). A network was constructed by reduced median-joining method in Network, version 4.5.1.16 (Bandelt et al. 1999; http://www.fluxus-engineering.com). Mutations A16182C, A16183C, T16189C, and C16519T were systematically ignored, as they are known to represent either the mutational hotspots or recurrent sequencing artifacts that may create reticulations in the following phylogenetic analysis (Der Sarkissian et al. 2014). We further performed the molecular variance analysis by focusing on the 393-bp HVR-I sequences (np 16,017–16,409), using Arlequin, version 3.11.11 (Excoffier et al. 2007). Multidimensional scaling (MDS) analysis was performed based on the $F_{ST}$ matrix calculated between pairs of selected populations, with the $F_{ST}$ values with a $p < 0.05$ discarded. The isofrequency map of haplogroup C5 was generated using Surfer 12 (Golden Software Inc., Golden, CO, USA), following the kriging procedure. The information of haplogroup C5d is summarized in Supplementary Table S1.

A total of 111 complete mitochondrial sequences that belong to haplogroup C5 were collected. Using these sequences, we next constructed a Bayes skyline plot to estimate the effective population size changes across time, as implemented in BEAST version 2.4.1 (Drummond et al. 2012; Green 1995). A general time-reversible sequence-evolution model with a fixed fraction of invariable sites was determined to be the best-fit model as estimated with jModelTest, version 2.1.415 (Darriba et al. 2012). We ran 40,000,000 generations of the Markov chain Monte Carlo simulation, with the first 4,000,000 generations discarded as burn-in. We applied different clock rates for nucleotides in the coding region (np 577–16023; $1.708 \times 10^{-8}$ substitutions per site per year) and noncoding region (np 16024–576; $9.883 \times 10^{-8}$ substitutions per site per year). The BEAST outputs were then analyzed with Tracer, version 1.4 (Rambaut and Drummond 2007).

Results and Discussion

Authentication and Data Statistics of the HDY Individual

We generated a high-quality mitochondrial genome for the HDY individual. A total of 56,946 unique reads were aligned to the rCRS. After removing duplicated fragments, we obtained an average of 383-fold coverage across the complete mitochondrial genome. We verified the authentication of our data by identifying the high deamination rate at both 3' and 5' ends (Supplementary Fig. S1), as well as the relatively short fragment length of 113 base pairs, which are characteristic of aDNA. We then estimate the mitochondrial contamination rate by ContamMix (Fu et al. 2013), a likelihood-based method to predict the sequencing error rate present in the data set as a uniform per base error rate. As a result, a low contamination rate of 1.2% (95% confidence interval, 1–1.4%) was identified. All these analyses confirm the authentication of our aDNA data.

The Maternal Genetic Relationship between the HDY Individual and Present-Day Worldwide Populations

To evaluate the genetic affinity of the HDY individual with present-day worldwide populations, we compared the HDY mitochondrial sequences with a large data set of present-day populations by focusing only on the HVR-I region (393 bp, spanning np 16,017–16,049). The data set includes Turkish, Japanese, Korean, Evenk, Yakut, Buryat, Oroqen, Mongolian, and Han Chinese (including the northern Han and southern Han) populations, the Tibeto-Burman-speaking populations, and populations from Central Asia and Europe (Derenko et al. 2003; Horai et al. 1996; Kolman et al. 1996; Kong et al. 2003; Li et al. 2007; Pakendorf et al. 2003; Wen et al. 2004a, 2004b; Yao et al. 2002, 2004; Zhao et al. 2009). We carried out MDS analysis based on a matrix of $F_{ST}$ distances (Fig. 2). Three different clusters are found in the first two dimensions, comprising west Eurasians (Europeans, Turkish, and Central Asians), Northeast Asians (Yakut, Buryat, Oroqen, Evenk, and Mongolian), and the Sino-Tibetan-speaking groups (including the Tibetan-Burmans, Han Chinese, and the Tibetans, but not Japanese and Korean). The HDY individual forms a cluster with the Tungusic (Oroqen and Evenk) and the Mongolic speakers (Buryat, Yakut, and Mongolian), who are all from Northeast Asia. This suggests that the HDY individual has a close relationship with populations from East Asia, in particular with the northern ones, which mirrors the geographic location of the HDY individual.
Phylogenetic Analysis Reveals the HDY Individual Belongs to a Novel Subclade of Haplogroup C5d

By aligning against the rCRS (PhyloTree build 17), we identified 38 polymorphisms, 1 insertion, and 2 deletions (Table 1) for the mitochondrial sequence of the HDY individual, which assigned it to mitochondrial haplogroup C5d, in accordance with the current phylogeny. Haplogroup C5 is one of the principal subclades of haplogroup C (Derenko et al. 2010). Today haplogroup C5 has a wide distribution: 7.9% in Northeast Asia, 2.6% in Russian Far East, 3.3% in the Altai region, 1.2% in the Lake Baikal region, 4.5% in western Siberia, 0.82% in Southeast Asia, and 0.8% in Eastern Asia, including the Korean (0.1%), Japanese (0.1%), and Chinese (0.8%), as well as in the historical Xiongnu populations from Mongolia and Transbaikal region (Derenko et al. 2010). However, haplogroup C5 reaches its highest frequency in some Tungusic-speaking populations, including the east Evenk (26.7%), Negidal (6.1%), Evens (11.5%), and Ulchi (5.8%), as well as in Koryak (15.4%) and Itelmen (13%) (Derbeneva et al. 2002; Derenko et al. 2007; Derenko and Shields 1997; Starikovskaya et al. 2005; Tamm et al. 2007) (Supplementary Fig. S2). All of those populations are mainly located in Northeast Asia, consistent with what we have observed in the MDS analysis.

So far, two subhaplogroups, C5d1 as defined by the transitions at G1415A, A8188G, and G16390A and C5d2 defined by the transitions at A10682G and G13968A under haplogroup C5d have been identified. The median-joining network shows that the HDY individual does not form a clade with either C5d1 or C5d2 but instead lies in a separate branch (Fig. 3). This observation is further confirmed by the phylogenetic tree constructed by the mtPhyl analysis: the HDY individual is assigned to the subclade of haplogroup C5d and cannot be further assigned to either C5d1 or C5d2 because of a private mutation at A13105G, two deletions at 248 and 3,106, and one insertion at 594 (Fig. 4). Thus, we conclude that the HDY individual belongs to a previously undescribed subclade under C5d that may be present in very low frequency in extant populations that have not yet been sampled.

Table 1. Summary of the HDY Sequence Data

<table>
<thead>
<tr>
<th>Measure</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average sequence depth</td>
<td>383-fold</td>
</tr>
<tr>
<td>Total reads aligned to reference</td>
<td>56,946</td>
</tr>
<tr>
<td>Average read length of trimmed reads</td>
<td>113.4 bp</td>
</tr>
<tr>
<td>Mitochondrial haplogroup</td>
<td>Previous undefined subclade of C5d</td>
</tr>
</tbody>
</table>
FIGURE 3. Median-joining network based on all available complete mitochondrial sequences of haplogroup C5d in the data set. The HDY individual forms a unique clade under haplogroup C5d.

FIGURE 4. Phylogeny of mitochondrial haplogroup C5.

Demographic History of HDY Inferred from Whole Mitochondrial Genomes

We reconstructed population demographic histories based on the mitochondrial sequences of haplogroup C5 and C5d, respectively. The skyline plot indicates that populations related to haplogroup C5 experienced a sharp increase in size beginning around 4,000 BP (Fig. 5a), a time period when populations associated with the Lower Xiajiadian culture in northern China had engaged in intensive millet farming and when millet farming had been widely applied across East Asia (Zhang et al. 2017). Meanwhile, wheat was also introduced to northern China from the Near East, together with other forms of crops (Betts et al. 2014). We argue that populations carrying haplogroup C5 may have learned those farming technologies and added them to their original way of life, and that these various food strategies provided abundant food resources and finally facilitated the increased population size. In contrast, a relatively small and stable effective population size was observed for C5d, which may indicate that populations related to this clade did not contribute to the increase of the population size (Fig. 5b). Although no ancient individual in the nearby region published in the

Downloaded From: https://bioone.org/journals/Human-Biology on 24 Feb 2020
Terms of Use: https://bioone.org/terms-of-use
Access provided by Max Planck Digital Library
Phylogenetic Analysis of Mitochondrial Genome from an 8,400-Year-Old Individual

The literature belongs to haplogroup C5, this may be partly because of the relatively limited ancient individuals sequenced in this region, and it is highly likely that this haplogroup was also present at a very low frequency in ancient populations as it is in present-day populations (Cui et al. 2013; Li et al. 2011). We further estimate the coalescence time and associated 95% highest posterior density for haplogroup C5d lineage. The estimated coalescence time for C5d varies from 9,000 to 14,000 years BP, with a median coalescence time of 11,500 years BP, a time consistent with the previous study (Derenko et al. 2010). Given that the HDY individual is dated to 8,400 years BP and belongs to a subclade of C5d, this provides an ideal time-stamped signal to understand the divergence and phylogenetic history for the formation of C5d.

Taken together, we describe a clear case study of a single ancient mitochondrial genome in tracing the population prehistory, and this study shows that interdisciplinary research combining genetic and archaeological evidence can provide a more extensive picture of prehistoric human populations. Further studies on large numbers of samples, in particular, those from different time periods, will provide greater insight into the demographics of population genetic history of East Asia.

ACKNOWLEDGMENTS

We are grateful to the Institute of Cultural Relics and Archaeology, Inner Mongolia Autonomous Region, for providing us with the sample. This work is supported by the National Natural Science Foundation of China (grant 41472024). C.N. is supported by the European Union’s Horizon 2020 Research and Innovation Programme “Millet and Beans, Language and Genes: The Origin and Dispersal of the Trans-eurasian Family” (grant 646612 to Dr. Martine Robbeets) and by the Max Planck Society. We thank Xiaomong Hu for providing us the archaeology background.

Received 12 July 2019; accepted for publication 4 October 2019.

LITERATURE CITED


FIGURE 5. Bayesian skyline plots showing changes in effective population size through time. The x-axis indicates years before present (years BP), and the y-axis shows the effective population size with the 95% posterior probability. (a) Demographic history of haplogroup C5. (b) Demographic history of haplogroup C5d.


Li, C., H. Li, Y. Cui et al. 2010. Evidence that a West-East admixed population lived in the Tarim Basin as early as the early Bronze Age. *BMC Biol.* 8:1–12.


Liu, J., L.-D. Wang, Y.-B. Sun et al. 2011. Deciphering the
Supplementary Table S1. Information on Haplogroup C5d

<table>
<thead>
<tr>
<th>Haplogroup / Accession Number</th>
<th>Location / Population</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>EU482303.1</td>
<td>East Siberia / Yukaghir</td>
<td>Volodko et al. 2008</td>
</tr>
<tr>
<td>EU482307.1</td>
<td>East Siberia / Yukaghir</td>
<td>Volodko et al. 2008</td>
</tr>
<tr>
<td>EU482329.1</td>
<td>East Siberia / Yukaghir</td>
<td>Volodko et al. 2008</td>
</tr>
<tr>
<td>EU828637.1</td>
<td>South Siberia / Tuvan</td>
<td>Starikovskaya et al. 2005</td>
</tr>
<tr>
<td>FJ951440.1</td>
<td>Altai / Altaians</td>
<td>Derenko et al. 2010</td>
</tr>
<tr>
<td>KF148223.1</td>
<td>Siberia / Evenk</td>
<td>Duggan et al. 2013</td>
</tr>
<tr>
<td>KF148224.1</td>
<td>Siberia / Evenk</td>
<td>Duggan et al. 2013</td>
</tr>
<tr>
<td>KF148238.1</td>
<td>Siberia / Evenk</td>
<td>Duggan et al. 2013</td>
</tr>
<tr>
<td>KF148248.1</td>
<td>Siberia / Evenk</td>
<td>Duggan et al. 2013</td>
</tr>
<tr>
<td>KF148268.1</td>
<td>Siberia / Evenk</td>
<td>Duggan et al. 2013</td>
</tr>
<tr>
<td>KF148274.1</td>
<td>Siberia / Evenk</td>
<td>Duggan et al. 2013</td>
</tr>
<tr>
<td>KF148318.1</td>
<td>Siberia / Evenk</td>
<td>Duggan et al. 2013</td>
</tr>
<tr>
<td>KF148574.1</td>
<td>East Siberia / Yukaghir</td>
<td>Duggan et al. 2013</td>
</tr>
<tr>
<td>FJ951576.1</td>
<td>South Siberia / Khamnigan</td>
<td>Derenko et al. 2010</td>
</tr>
</tbody>
</table>

C5d2

<table>
<thead>
<tr>
<th>Accession Number</th>
<th>Location / Population</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>JF824872.2</td>
<td>China / Chinese</td>
<td>Liu et al. 2011</td>
</tr>
<tr>
<td>JF824965.2</td>
<td>China / Chinese</td>
<td>Liu et al. 2011</td>
</tr>
<tr>
<td>DQ112787.3</td>
<td>China / Chinese</td>
<td>Kivisild et al. 2006</td>
</tr>
<tr>
<td>A76LAJ36I31S (sample no.)</td>
<td>China / Chinese</td>
<td>Li et al. 2019</td>
</tr>
</tbody>
</table>

**SUPPLEMENTARY FIGURE S1**
(above). Ancient DNA damage patterns for the HDY individual. The mismatch frequency is relative to the reference as a function of read position; C to T misincorporation is shown in red, and G to A, in blue. The lower misincorporation in the 5’ end than in the 3’ end might be caused by using the Q5 enzyme when applying the DNA library.

**SUPPLEMENTARY FIGURE S2**
(right). Spatial frequency distribution map of mitochondrial haplogroup C5.