12-17-2019

Mitochondrial Genome of an 8,400-Year-Old Individual from Northern China Reveals a Novel Sub-Clade under C5d

Xiyan Wu
Jilin University, Changchun, China, wuxy17@mails.jlu.edu.cn

Chao Ning
Max Planck Institute for the Science of Human History, Jena, Germany, ningchao.jlu@gmail.com

Qingchuan Bao
The Institute of Cultural Relics and Archaeology, Inner Mongolia Autonomous Region, China, 2530655097@qq.com

Shizhu Gao
Jilin University, Changchun, China, gaosz@jlu.edu.cn

Fan Zhang
Jilin University, Changchun, China, zf1992cyx@gmail.com

See next page for additional authors

Authors

Xiyan Wu, Chao Ning, Qingchuan Bao, Shizhu Gao, Fan Zhang, Sihao Wu, Tianjiao Li, Linyuan Fan, Tao Li, Xuan Yang, Dawei Cai, and Yinqiu Cui

Mitochondrial Genome of an 8,400-Year-Old Individual from Northern China Reveals a Novel Sub-Clade 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,*] Yinqiu Cui,[1,2,8,*]

[1]Research Center for Chinese Frontier Archaeology, Jilin University, Changchun, P. R. China.

[2]School of Life Sciences, Jilin University, Changchun, P. R. China.

[3]Max Planck Institute for the Science of Human History, Jena, Germany.

[4]The Institute of Cultural Relics and Archaeology, Inner Mongolia Autonomous Region, China.

[5]College of Pharmacia Sciences, Jilin University, Changchun, P. R. China.

[6]Department of Archaeology, College of History, Wuhan University, Wuhan, P. R. China.

[7]School of Archaeology, Jilin University, Changchun, P. R. China.

[8]Key Laboratory for Evolution of Past Life and Environment in Northeast Asia, Jilin University, Ministry of Education, Changchun, China.

*Correspondence to:

Yinqiu Cui, Life Science Building, Jilin University, Qianjin Street, No. 2699, Changchun, P. R. China, 130012. E-mail: cuiyq@jlu.edu.cn.

Dawei Cai, Research Center for Chinese Frontier Archaeology, Jilin University, Changchun P. R. China, 130012. E-mail: caidw@jlu.edu.cn.

Short Title: Phylogenetic Analysis of Mitochondrial Genome from an 8,400-Year-Old Individual, China

KEY WORDS: ANCIENT DNA, MITOCHONDRIAL GENOME, MTDNA HAPLOGROUP C5D, INNER MONGOLIA, MILLET FARMING, PHYLOGENETIC ANALYSIS.

Abstract

Ancient DNA studies have always refreshed our understanding of the human past that can't be tracked by modern DNA alone. Until recently, ancient mitochondrial genomic studies in East Asia are 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 was most probably originated in northern Asia and may have a very low frequency in extant populations that is not yet 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 from around 4,000 years before present (BP). The time when intensive millet farming was built 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 added this farming technology to their original way of life and that the various subsistence may provide abundant food sources and may further contribute to the increase of the population size.

Mitochondrial DNA possesses several favorable characteristics, including the 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 most frequently used markers to explore population genetic diversity and structures (Gonzalez-Martin 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 applying the short fragment of mitochondrial hypervariable regions (HVR) and the ever-increasing whole mitochondrial genomes of present-day populations. However, based on what we have seen for ancient DNA studies in other regions, ancient DNA has always refreshed our understanding of the human past that can't be tracked from the modern DNA alone because of frequent population replacements and interbreeding were happened (Reich 2018; Sikora et al. 2019; Slatkin and Racimo 2016). In this sense, ancient DNA will provide 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 team conducted by 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 in the shape of round and semi-subterranean,

with a round hearth standing in the middle, without the traces of gateway and obvious posthole. The ash ditch was at the southwest of the excavation area, in the direction of southeast and northwest and the tomb was under the floor of the house F1. It is an earthen shaft pit tomb with a completely well-preserved human bones and no grave goods were found and was dated to 8,400 years before present (BP) (Fig. 1). The number of cultural relics unearthed from YM was relatively less (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 grey and black mudstone (the major), sandstone (second major), flint, quartzite and etc. (Fig. 1). The shapes of the stone tools vary from semicircular shovel, flake choppers, spear-shaped tools, stone core, stone ball and etc. Few of the pottery could be restored, and most of which were small pottery shard, including sand inclusion yellow and brown, black and brown pottery, made in the way of using mud-piece pasting technique. The shapes of the pottery included round-bottom and barrel shaped jar, sharp and round-bottom Fu (caldron), flake object, flared-mouth jar, etc. (Dang 2014; Dang 2017).

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

Here we target-enriched and sequenced the complete mitochondrial genome of the only human remain from the YM site. The individual is a female based on physical anthropology

analysis and was dated to 8,400 years before present (BP) and was characterized as assigning to the subclade of mitochondrial haplogroup C5. By comparing with a large dataset of mitochondrial DNA publicly available that belonged to haplogroup C5, we reconstructed the phylogeny and the demographic history of haplogroup C5 and further track the relations between the ancient and present-day populations in East Asia.

Material and Methods

Archeological 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 team consisted 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 (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 ancient DNA (aDNA) analyses. All aDNA lab works were carried out following the strict standards specially designed to minimize the potential modern DNA contaminations (Adler et al. 2011; Llamas et al. 2016). Specifically, the teeth samples were immersed in a 5% liquid sodium hypochlorite (bleach) for

10 minutes, and were then washed by ultra-pure water and 100% ethanol followed by the UV irradiation on each side of the teeth. Then 150 mg fine teeth power 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. 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). Mitochondrial DNA enrichment was performed using the MyGenostics Human Mitochondria Capture Kit (MyGenostics Inc., Beijing, China). The post-capture 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 v1.92.50 program, an automated computational pipeline specially designed for ancient DNA data processing (Peltzer et al. 2016). Quality filtering was performed with FastQC software (Andrews 2010) and the adapters were trimmed with AdapterRemoval v2.2.0 (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 30bp are considered and the duplications were removed using the default parameter of the Dedup software (Li and Durbin 2009; Li et al. 2009). Both 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 View (IGV)(Thorvaldsdottir et al. 2013). We then applied the default parameters of MapDamage 2.0 to determine the molecular damage that is typical of ancient DNA (Jonsson et al. 2013).

Phylogenetic Analysis and Coalescent Simulations of the HDY DNA

We prepared a large dataset of complete mitochondrial sequences that belong to haplogroup C5 from the MitoTool (http://mitotool.kiz.ac.cn) and the MitoMap (https://www.mitomap.org/MITOMAP) databases as well as those from the literature (Derenko et al. 2012; Derenko et al. 2010; Dryomov et al. 2019; Li et al. 2019; Mielnik-Sikorska et al. 2013). The sequences were then aligned to the revised Cambridge Reference Sequence (rCRS) using the Bioedit v14.0 (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 the NETWORK v.4.5.1.16 (Bandelt et al. 1999). 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 to 16,409), using ARLEQUIN v3.11.11 (Excoffier et al. 2007). The MultiDimensional Scaling analysis was performed based on the Fst matrix calculated between pairs of selected populations, which the Fst values with a P-value smaller than 0.05 was discarded. The isofrequency map of haplogroup C5 was generated by using Surfer 12 of Golden Software (Golden Software Inc., Golden, Colorado), following the Kriging procedure. The information of haplogroup C5d was summarized in Table S1(Supplementary information).

A total of 111 complete mitochondrial sequences that belong to haplogroup C5 were collected and by using these sequences we next constructed a Bayes Skyline Plot to estimate the effective population size changes across time as implemented in BEAST v2.4.1 (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 from the jModeltest v2.1.415 software (Darriba et al. 2012). We ran 40,000,000 generations of the Markov Chain Monte Carlo 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) (1.708×10^{-8} substitutions per site per year) and noncoding region (16024-576) (9.883×10^{-8} substitutions per site per year). The BEAST outputs were then analyzed with the software Tracer v1.4 software (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 (Fig. S1 in Supplementary information) as well as the relatively short fragment length of 113 base pairs, which are characteristic of ancient DNA. We then estimate the mitochondrial contamination rate by ContaMix (Fu et al. 2013), a likelihood-based method to predict the sequencing error rate present in the dataset as a uniform per base error rate. As a result, a low contamination rate of 1.2% (95% CI: 1%-1.4%) was identified. All these analyses confirm the authentication of our aDNA data.

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

To evaluate the genetic affinity of the HDY individual with present-day worldwide populations, we compare the HDY mitochondrial sequences with a large dataset of present-day populations by focusing only on the HVR-I region (393bp, spanning 16,017-16,049). The dataset includes Turkish, Japanese, Korean, Evenk, Yakut, Buryat, Oroqen, Mongolian, Han Chinese (including the northern Han and southern Han), the Tibeto-Burman-speaking populations as well as those

from the Central Asia and the 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; Wen et al. 2004b; Yao et al. 2004; Yao et al. 2002; Zhao et al. 2009). We carried out a Multidimensional Scaling analysis (MDS) based on a matrix of Fst distances (Fig. 2). Three different clusters are found in the first two dimensions as represented by the west Eurasians (Europeans, Turkish and central Asians), the 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 Sub-Clade 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 the mitochondrial haplogroup C5d, in accordance with the current phylogeny. Haplogroup C5 is one of the principal sub-clades 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 Altai

region, 1.2% in the Lake Baikal region, 4.5% in west 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) (Fig. S2 in Supplementary information). All of those populations are mainly located in northeast Asia, consistent with what we have observed in the MDS analysis.

So far, two sub-haplogroups, C5d1 as defined by the transitions at G1415A, A8188G, G16390A and C5d2 defined by the transitions at A10682G, G13968A under haplogroup C5d have been identified. The median-joining network shows that the HDY individual doesn't 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 that the HDY individual is assigned to the sub-clade of haplogroup C5d and can't be further assigned to either C5d1 or C5d2 because of a private mutation at A13105G, two deletions at 248 and 3106 as well as one insertion at 594 (Fig. 4). Thus, we conclude that the HDY individual belongs to a previously undescribed sub-clade under C5d that may present in very low frequency in extant populations which has not yet sampled.

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 from around 4,000 BP (Fig. 5a). A time period that populations associate with the lower Xiajiadian culture in northern China who had engaged in intensive millet farming and millet farming had been widely applied across East Asia (Xuelian et al. 2017). Meanwhile, wheat was also introduced to northern China from the Near East together with other crop packages (Betts et al. 2014). We argue that populations carrying the haplogroup C5 may learn those farming technologies and add them to their original way of life, which various food strategies provided abundant food resources and finally facilitated the increasing in populations size. In contrast, a relatively small and stable effective population size was observed for C5d, which may indicate that populations related to this clade didn't contribute to the increase of the population size (Fig. 5b). Although no ancient individual published in the nearby region belongs to haplogroup C5, this may partly because of the relatively limited ancient individuals sequenced in this region and it is highly likely that this haplogroup also present in 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% HPD for haplogroup C5d lineage. The estimated coalescence time for C5d varies from 9 to 14 kya with a median coalescence time of 11.5 kya, the time that is consistent with the previous study (Derenko et al. 2010). Given that the HDY individual is 8,400 BP and belonging

to a sub-clade of C5d, this provide an ideal time-stamped signal to understand the divergence and phylogenetic history for the formation of C5d.

Taken together, we show 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 evidences can provide a more extensive picture of prehistoric human populations. Further studies on large numbers of samples, in particular those from different time period will provide us greater insight into the demographic process 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 no.41472024). Chao Ning 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 Transeurasian family (granted to M.R., grant no. 646612) and the Max Planck Society. We thank Xiaonong Hu for providing us the archaeology background.

Author Contributions

Yinqiu Cui planned and designed this study. Dawei Cai, Yinqiu Cui and Qingchuan Bao collected the sample. Xiyan Wu, Fan Zhang, Shizhu Gao, Tianjiao Li, Linyuan Fan, Xuan Yang

performed the wet lab works and bioinformatics analysis. Xiyan Wu, Tao Li, Chao Ning and Yinqiu Cui wrote the manuscript.

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

Literature Cited

Adler, C. J., W. Haak, D. Donlon et al. 2011. Survival and recovery of DNA from ancient teeth and bones. *J. Archaeol. Sci.* 38:956–964.

Babraham Bioinformatics. 2010. FastQC. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Bandelt, H. J., P. Forster, and A. Röhl. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16:37–48.

Betts, A., P. W. Jia, and J. Dodson. 2014. The origins of wheat in China and potential pathways for its introduction: A review. *Quat. Int.* 348:158–168.

Cui, Y., H. Li, C. Ning et al. 2013. Y chromosome analysis of prehistoric human populations in the West Liao River Valley, Northeast China. *BMC Evol. Biol.* 13:1–10.

Dang, Y. 2014. The summary of archaeological discoveries of the Inner Mongolia Autononomous Region Institute of Cultural Relics and Archaeology in 2014. *Steppe Cultur. Relics* (1):2.

Dang, Y. 2017. The summary of archaeological discoveries of the Inner Mongolia Autonomous Region Institute of Cultural Relics and Archaeology in 2016. *Steppe Cultur. Relics* (1):2.

Darriba, D., G. L. Taboada, R. Doallo et al. 2012. jModelTest 2: More models, new heuristics and parallel computing. *Nat. Methods* 9:772.

Der Sarkissian, C., P. Brotherton, O. Balanovsky et al. 2014. Mitochondrial genome sequencing in Mesolithic north east Europe unearths a new sub-clade within the broadly distributed human haplogroup C1. *PLoS One* 9:1–10.

Derbeneva, O. A., R. I. Sukernik, N. V. Volodko et al. 2002. Analysis of mitochondrial DNA diversity in the Aleuts of the Commander Islands and its implications for the genetic history of Beringia. *Am. J. Hum. Genet*. 71:415–421.

Derenko, M. V., T. Grzybowski, B. A. Malyarchuk et al. 2003. Diversity of mitochondrial DNA lineages in South Siberia. *Ann. Hum.Genet*. 67:391–411.

Derenko, M., B. Malyarchuk, G. Denisova et al. 2012. Complete mitochondrial DNA analysis of Eastern Eurasian haplogroups rarely found in populations of Northern Asia and Eastern Europe. *PLoS One* 7:1–12.

Derenko, M., B. Malyarchuk, T. Grzybowski et al. 2007. Phylogeographic analysis of mitochondrial DNA in northern Asian populations. *Am. J. Hum. Genet*. 81:1,025–1,041.

Derenko, M., B. Malyarchuk, T. Grzybowski et al. 2010. Origin and post-glacial dispersal of mitochondrial DNA haplogroups C and D in Northern Asia. *PLoS One* 5:e1–9.

Derenko, M. V., and G. F. Shields. 1997. [Diversity of mitochondrial DNA nucleotide sequences in three groups of aboriginal inhabitants of Northern Asia]. *Mol. Biol. (Mosk.)* 31:784–789.

Drummond, A. J., M. A. Suchard, D. Xie et al. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol*. 29:1,969–1,973.

Dryomov, S. V., A. M. Nazhmidenova, E. B. Starikovskaya et al. 2019. Mitochondrial genome diversity in the Central Siberian Plateau with particular reference to prehistory of northernmost Eurasia. *bioRxiv* 1–33.

Duggan, A. T., M. Whitten, V. Wiebe et al. 2013. Investigating the prehistory of Tungusic peoples of Siberia and the Amur-Ussuri region with complete mtDNA genome sequences and Y-chromosomal markers. *PLoS One* 8:1–19.

Eltsov, N., and N. Volodko. 2011. mtPhyl: Software tool for human mtDNA analysis and phylogeny reconstruction. https://sites.google.com/site/mtphyl/home.

Excoffier, L., G. Laval, and S. Schneider. 2007. Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol. Bioinform. Online* 1:47–50.

Fu, Q., A. Mittnik, P. L. F. Johnson et al. 2013. A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* 23:553–559.

González-Martín, A., A. Gorostiza, L. Regalado-Liu et al. 2015. Demographic history of Indigenous populations in Mesoamerica based on mtDNA sequence data. *PLoS One* 10:1–24.

Green, P. J. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82:711–732.

Hall, T. 2011. BioEdit: An important software for molecular biology. *GERF Bull. Biosci.* 2:60–61.

Horai, S., K. Murayama, K. Hayasaka et al. 1996. mtDNA polymorphism in East Asian populations, with special reference to the peopling of Japan. *Am. J. Hum. Genet*. 59:579–590.

Hu, X. 2017. An original village 8400 years ago. *Inner Mongolia Daily* 11.

Jónsson, H., A. Ginolhac, M. Schubert et al. 2013. mapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* 29:1,682–1,684.

Kivisild, T., P. Shen, D. P. Wall et al. 2006. The role of selection in the evolution of human mitochondrial genomes. *Genetics* 172:373–387.

Kivisild, T. 2015. Maternal ancestry and population history from whole mitochondrial genomes. *Investig. Genet.* 6:1–10.

Kolman, C. J., N. Sambuughin, and E. Bermingham. 1996. Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders. *Genetics* 142:1,321–1,334.

Kong, Q.-P., Y.-G. Yao, M. Liu et al. 2003. Mitochondrial DNA sequence polymorphisms of five ethnic populations from northern China. *Hum. Genet*. 113:391–405.

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.

Li, H., X. Cai, E. R. Winograd-Cort et al. 2007. Mitochondrial DNA diversity and population differentiation in southern East Asia. *Am. J. Phys. Anthropol*. 134:481–488.

Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1,754–1,760.

Li, H., B. Handsaker, A. Wysoker et al. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2,078–2,079.

Li, H., X. Zhao, Y. Zhao et al. 2011. Genetic characteristics and migration history of a bronze culture population in the West Liao-River valley revealed by ancient DNA. *J. Hum. Genet.* 56:815–822.

Liu, J., L.-D. Wang, Y.-B. Sun et al. 2011. Deciphering the signature of selective constraints on cancerous mitochondrial genome. *Mol. Biol. Evol.* 29:1,255–1,261.

Li, Y.-C., W.-J. Ye, C.-G. Jiang et al. 2019. River valleys shaped the maternal genetic landscape of Han Chinese. *Mol. Biol. Evol.* 36:1,643–1,652.

Llamas, B., G. Valverde, L. Fehren-Schmitz et al. 2017. From the field to the laboratory: Controlling DNA contamination in human ancient DNA research in the high-throughput sequencing era. *Sci. Technol. Archaeol. Res.* 3:1–14.

Mielnik-Sikorska, M., P. Daca, B. Malyarchuk et al. 2013. The history of Slavs inferred from complete mitochondrial genome sequences. *PLoS One* 8:1–11.

Ning, C., S. Gao, B. Deng et al. 2016. Ancient mitochondrial genome reveals trace of prehistoric migration in the east Pamir by pastoralists. *J. Hum. Genet.* 61:103–108.

Pakendorf, B., V. Wiebe, L. A. Tarskaia et al. 2003. Mitochondrial DNA evidence for admixed origins of central Siberian populations. *Am. J. Phys. Anthropol.* 120:211–224.

Peltzer, A., G. Jäger, A. Herbig et al. 2016. EAGER: Efficient ancient genome reconstruction. *Genome Biol*. 17:1–14.

Postillone, M. B., and S. I. Perez. 2017. Mitochondrial-DNA phylogenetic information and the reconstruction of human population history: The South American case. *Hum. Biol*. 89:229–250.

Rambaut, A., and A. Drummond. 2007. Tracer v1. 4: MCMC trace analyses tool. http://beast.bio.ed.ac.uk/Tracer.

Reich, D. 2018. *Who We Are and How We Got Here: Ancient DNA and the New Science of the Human Past*. New York: Pantheon Books.

Ricaut, F.-X., M. Bellatti, and M. M. Lahr. 2006. Ancient mitochondrial DNA from Malaysian hair samples: Some indications of Southeast Asian population movements. *Am. J. Hum. Biol.* 18:654–667.

Schubert, M., S. Lindgreen, and L. Orlando. 2016. AdapterRemoval v2: Rapid adapter trimming, identification, and read merging. *BMC Res. Notes* 9:1–7.

Sikora, M., V. V. Pitulko, V. C. Sousa et al. 2019. The population history of northeastern Siberia since the Pleistocene. *Nature* 570:182–188.

Slatkin, M., and F. Racimo. 2016. Ancient DNA and human history. *Proc. Natl. Acad. Sci. U. S. A.* 113:6,380–6,387.

Starikovskaya, E. B., R. I. Sukernik, O. A. Derbeneva et al. 2005. Mitochondrial DNA diversity in indigenous populations of the southern extent of Siberia, and the origins of Native American haplogroups. *Ann. Hum. Genet.* 69:67–89.

Stoneking, M. 1994. Mitochondrial DNA and human evolution. *J. Bioenerg. Biomembr.* 26:251–259.

Tamm, E., T. Kivisild, M. Reidla et al. 2007. Beringian standstill and spread of Native American founders. *PLoS One* 9:1–6.

Thorvaldsdóttir, H., J. T. Robinson, and J. P. Mesirov. 2013. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief. Bioinform.* 14:178–192.

Volodko, N. V., E. B. Starikovskaya, I. O. Mazunin et al. 2008. Mitochondrial genome diversity in Arctic Siberians, with particular reference to the evolutionary history of Beringia and Pleistocenic peopling of the Americas. *Am. J. Hum. Genet.* 82:1,084–1,100.

Wei, Z., W. Wang, P. Hu et al. 2011. SNVer: A statistical tool for variant calling in analysis of pooled or individual next-generation sequencing data. *Nucleic Acids Res.* 39:1–13.

Wen, B., H. Li, D. Lu et al. 2004a. Genetic evidence supports demic diffusion of Han culture. *Nature* 431:302–305.

Wen, B., X. Xie, S. Gao et al. 2004b. Analyses of genetic structure of Tibeto-Burman populations reveals sex-biased admixture in southern Tibeto-Burmans. *Am. J. Hum. Genet.* 74:856–865.

Zhang, X. L., G. X. Liu, M. H. Wang et al. 2017. Stable istopic analysis of human remains from Xinglonggou site. *Cult. Relics South. China* (4):185–195.

Yang, D. Y., B. Eng, J. S. Waye et al. 1998. Technical note: Improved DNA extraction from ancient bones using silica-based spin columns. *Am. J. Phys. Anthropol*. 105:539–543.

Yao, Y.-G., Q.-P. Kong, C.-Y. Wang et al. 2004. Different matrilineal contributions to genetic structure of ethnic groups in the Silk Road region in China. *Mol. Biol. Evol*. 21:2,265–2,280.

Yao, Y.-G., L. Nie, H. Harpending et al. 2002. Genetic relationship of Chinese ethnic populations revealed by mtDNA sequence diversity. *Am. J. Phys. Anthropol*. 118:63–76.

Zhang, Y., X. Wu, J. Li et al. 2018. The Y-chromosome haplogroup C3*-F3918, likely attributed to the Mongol Empire, can be traced to a 2500-year-old nomadic group. *J. Hum. Genet.* 63:231–238.

Zhao, M., Q.-P. Kong, H.-W. Wang et al. 2009. Mitochondrial genome evidence reveals successful Late Paleolithic settlement on the Tibetan Plateau. *Proc. Natl. Acad. Sci. U. S. A.* 106:21,230–21,235.

Table 1. Summary of the HDY Sequence Data

Sample name	HDY
Average sequence depth	383X
Total reads aligned to reference	56946
Average read length of trimmed reads	113.4bp
Mitochondrial haplogroup	Previous undefined sub-clade of C5d
Mutation sites	73G, 248d, 263G, 489C, 594+C, 750G, 1438G, 2706G, 3106d, 3552A, 4715G, 4769G, 7028T, 7196A, 8584A, 8701G, 8860G, 9540C, 9545G, 10398G, 10400T, 10873C, 11719A, 11914A, 12705T, 13105G, 13263G, 14318C, 14766T, 14783C, 15043A, 15080G, 15301A, 15326G, 15487T, 16093C, 16223T, 16288C, 16298C, 16327T, 16519C

Supplementary Table S1. The Information of Haplogroup C5d

Haplogroup	Accession number	Location/Population	References
C5d1	EU482303.1	East Siberia /Yukaghir	Volodko et al. 2008
	EU482307.1	East Siberia /Yukaghir	Volodko et al. 2008
	EU482329.1	East Siberia /Yukaghir	Volodko et al. 2008
	EU828637.1	South Siberia /Tuvan	Starikovskaya et al. 2005
	FJ951440.1	Altai /Altaians	Derenko et al. 2010
	KF148223.1	Siberia /Evenk	Duggan et al. 2013
	KF148224.1	Siberia /Evenk	Duggan et al. 2013
	KF148238.1	Siberia /Evenk	Duggan et al. 2013
	KF148248.1	Siberia /Evenk	Duggan et al. 2013
	KF148274.1	Siberia /Even	Duggan et al. 2013
	KF148318.1	Siberia /Even	Duggan et al. 2013
	KF148574.1	East Siberia /Yukaghir	Duggan et al. 2013
	FJ951576.1	South Siberia /Khamnigan	Derenko et al. 2010
C5d2	JF824872.2	China /Chinese	Liu et al. 2011
	JF824965.2	China /Chinese	Liu et al. 2011
	DQ112787.3	China /Chinese	Kivisild et al. 2006

	A76LAJ36I31S (Sample number)	China /Chinese	Li et al. 2019

Figure Captions

Figure 1. (A) Grographic location of Yumin site. (B) Human skeletion and archaeological relics excavated from the Yumin Site. The individual was the only human skeleton excavated in an earthen shaft pit tomb, with flexed and squat burial. Two fundamentally different types of stone tools including chipped stone and ground stone coexisting in Yumin site.

Figure 2. A multidimensional scaling analysis (MDS) plotted based on the Fst matrix calculated from the HVR-I sequences of the HDY individual and the modern populations. Turkish, Japanese, Korean, Evenki, Central Asians, Yakut, Buryat, Oroqen, Mongolian, Europeans, northern Han Chinese, southern Han Chinese and Tibeto-Burman-speaking populations were abbreviated as TURK, JAP, KOR, EWK, CA, YAK, BUR, ORO, MG, EUR, NH, SH and TB individually.

Figure 3. Median-joining network based on all available complete mitochondrial sequences of haplogroup C5d in the dataset. The HDY individual forms an unique clade under haplogroup C5d.

Figure 4. Phylogeny of mitochondrial haplogroup C5.

Figure 5. Bayesian skyline plots showing changes in effective population size through time (year). The x-axis indicates the years before present, 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.

Supplementary Figure S1. Ancient DNA damage patterns for HDY individual. The mismatch frequency is relative to the reference as a function of read position, which C to T misincorporation is colored in red and G to A in blue. The lower misincorporation in 5' end than 3' end might be caused by using the Q5 enzyme when applied the DNA library.

Supplementary Figure S2. Spatial frequency distribution map of mitochondrial haplogroup C5.

Figure 1.

Figure 2.

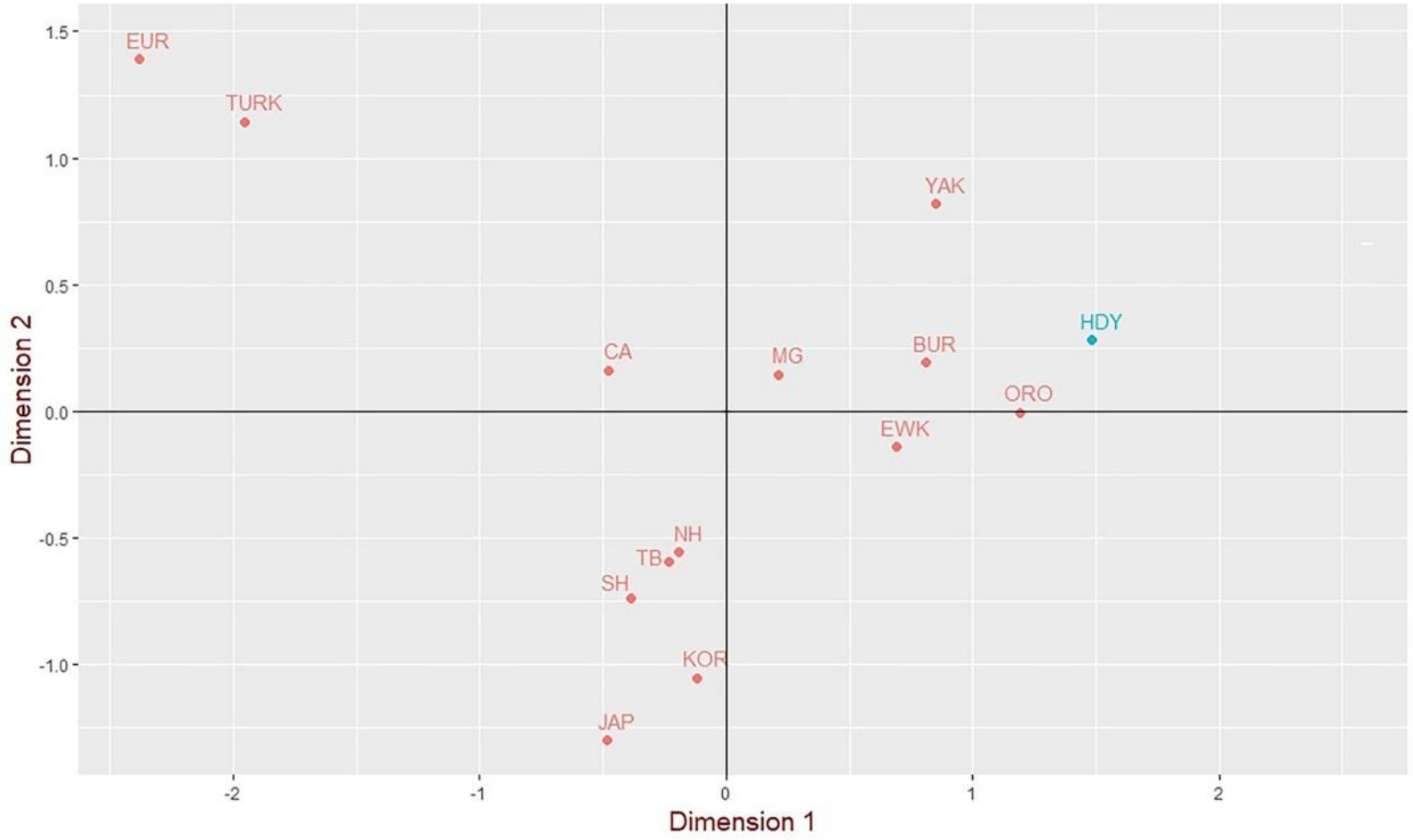


Figure 3.

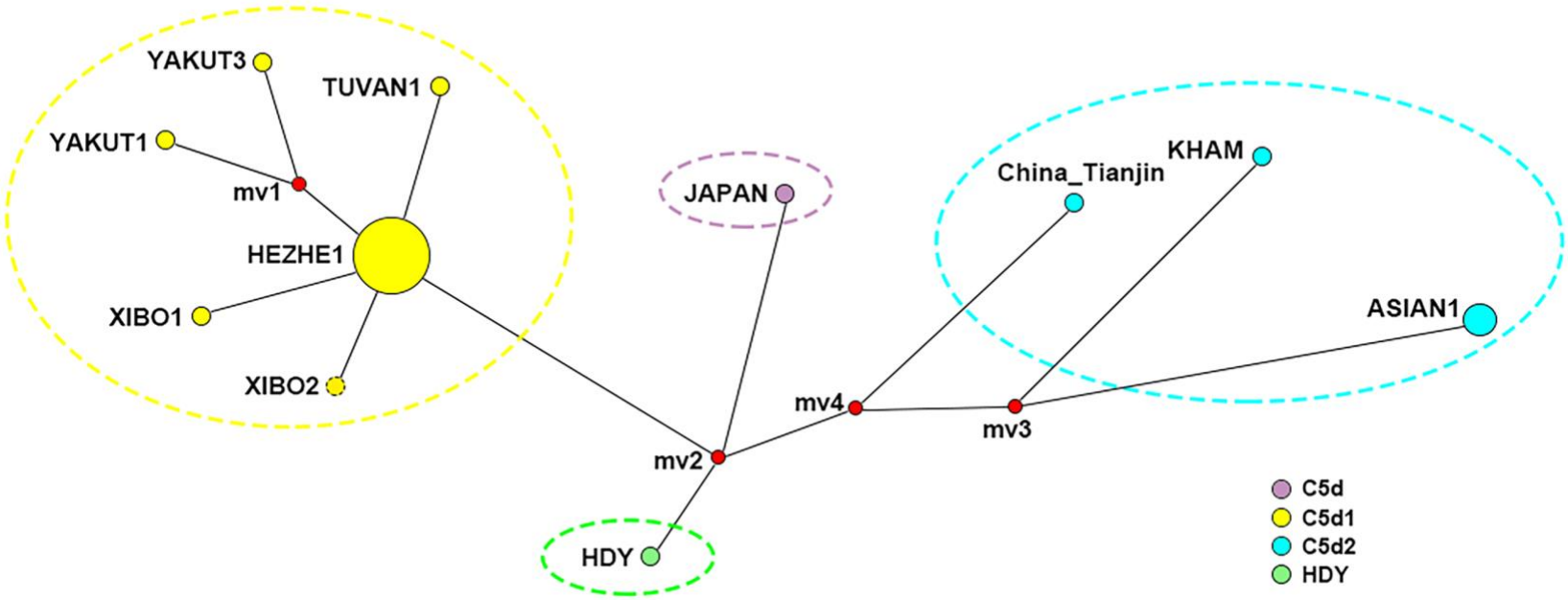

YAKUT3
TUVAN1
YAKUT1
mv1
HEZHE1
XIBO1
XIBO2
JAPAN
KHAM
China_Tianjin
ASIAN1
mv4
mv3
mv2
HDY
C5d
C5d1
C5d2
HDY

Figure 4.

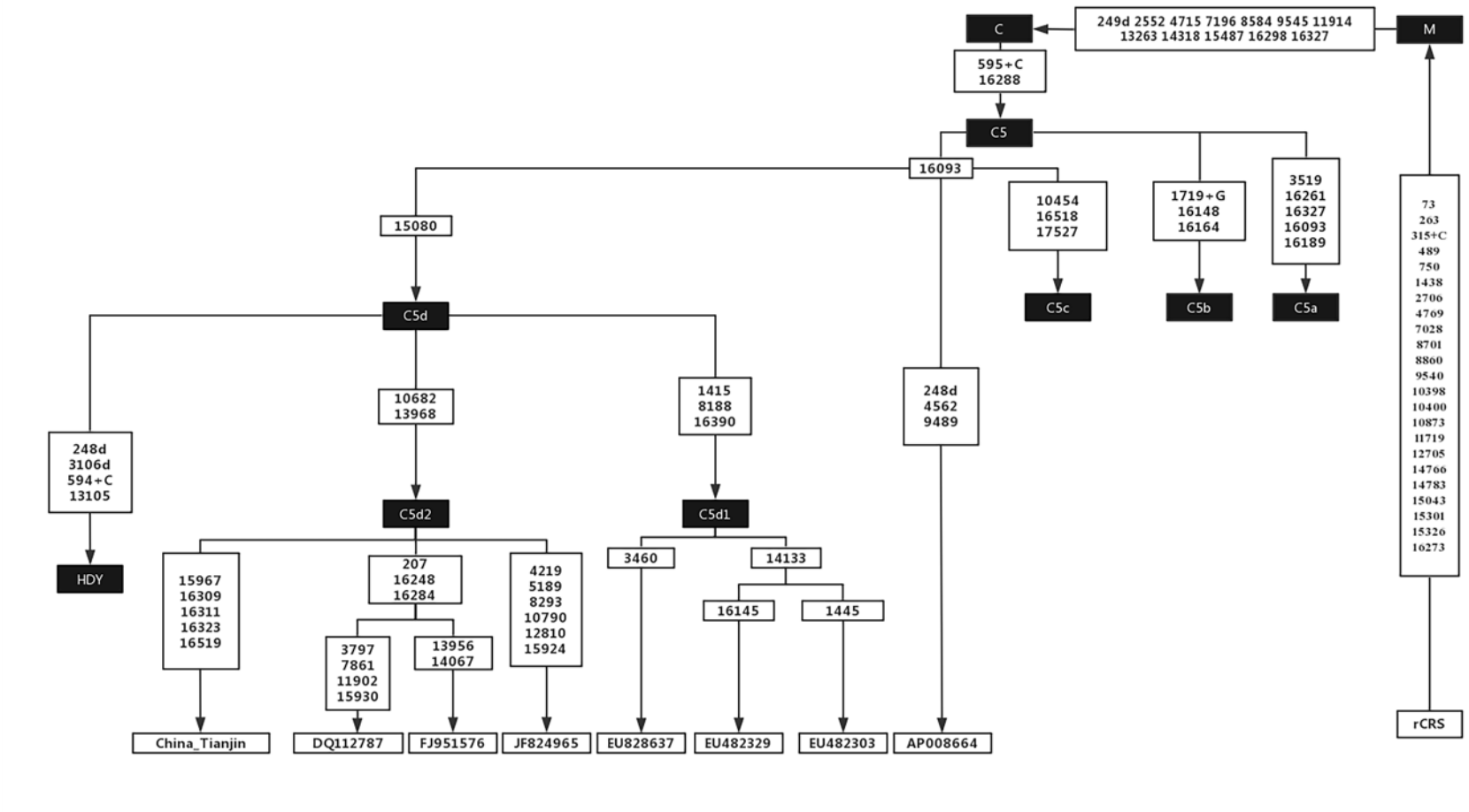
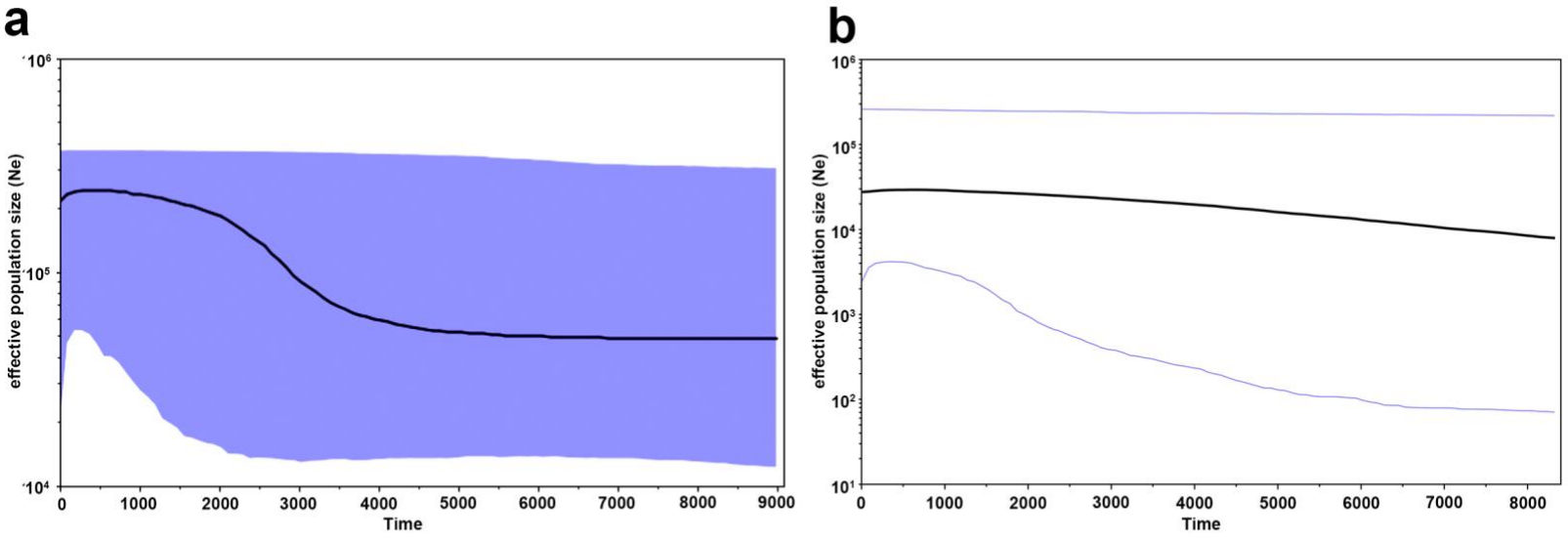
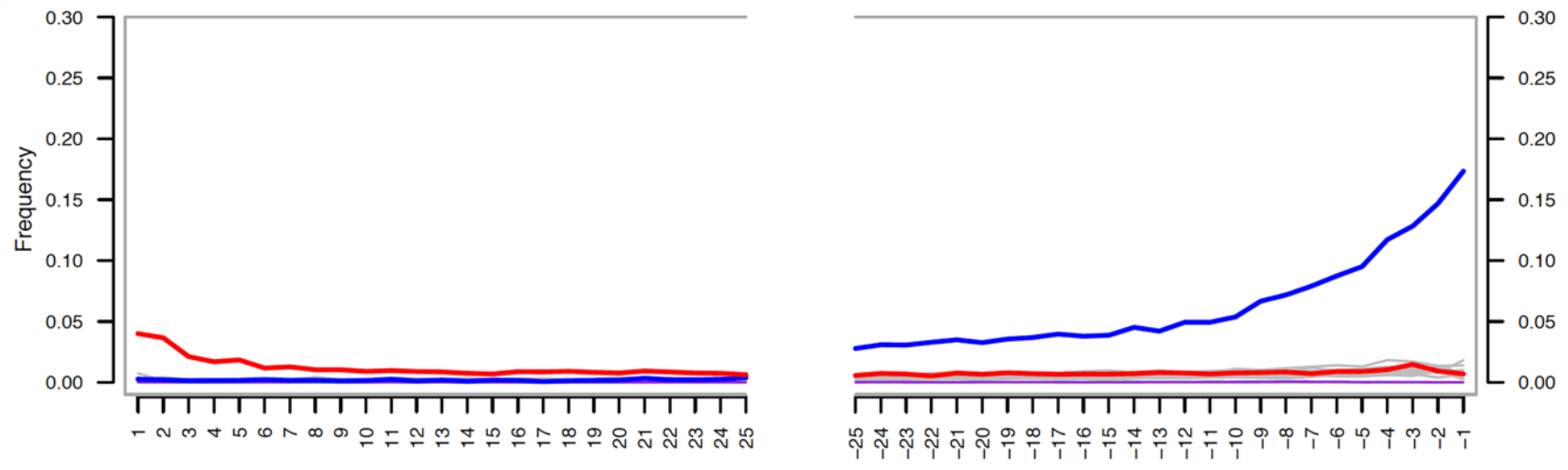

C
249d 2552 4715 7196 8584 9545 11914
13263 14318 15487 16298 16327
M
595+C
16288
C5
16093
10454
16518
17527
1719+G
16148
16164
3519
16261
16327
16093
16189
15080
C5c
C5b
C5a
73
263
315+C
489
750
1438
2706
4769
7028
8701
8860
9540
10398
10400
10873
11719
12705
14766
14783
15043
15301
15326
16273
C5d
10682
13968
1415
8188
16390
248d
4562
9489
248d
3106d
594+C
13105
C5d2
C5d1
HDY
15967
16309
16311
16323
16519
207
16248
16284
4219
5189
8293
10790
12810
15924
3460
14133
3797
7861
11902
15930
13956
14067
16145
1445
rCRS
China_Tianjin
DQ112787
FJ951576
JF824965
EU828637
EU482329
EU482303
AP008664

Figure 5.



Supplementary Figure S1.



Supplementary Figure S2.