Journal of Human Genetics
https://doi.org/10.1038/s10038-020-0796-9

BRIEF COMMUNICATION

Paleolithic genetic link between Southern China and Mainland Southeast Asia revealed by ancient mitochondrial genomes

Fan Bai[1,2,3] · Xinglong Zhang[4] · Xueping Ji[5,6] · Peng Cao[1] · Xiaotian Feng[1] · Ruowei Yang[1] · Minsheng Peng[3,7,8] · Shuwen Pei[1,2] · Qiaomei Fu[1,2,3]

Received: 18 May 2020 / Revised: 21 June 2020 / Accepted: 27 June 2020

Abstract

The genetic history of Southern East Asians is not well-known, especially prior to the Neolithic period. To address this, we successfully sequenced two complete mitochondrial genomes of 11,000-year-old human individuals from Southern China, thus generating the oldest ancient DNA sequences from this area. Integrating published mitochondrial genomes, we characterized M71d, a new subhaplogroup of haplogroup M71. Our results suggest a possible early migration between Southern China and mainland Southeast Asia by at least 22,000 BP.

Introduction

Since Southern China connects to Northern China and Southeast Asia, it is a critical geographic crossroads for early human migrations [1]. In recent years, the analysis of ancient DNA (aDNA) has been employed to investigate population history. Multiple waves of migrations have been reported from East Asia into Southeast Asia since the Middle Holocene [2–4]. The hot and humid climate of Southern China and Southeast Asia is detrimental to the preservation of DNA, which contributes to the poor success rate in recovering genetic information from ancient remains in this region. As a consequence, the genetic connections between East and Southeast Asia before the Neolithic period are still poorly characterized.

To investigate the genetic characteristics of modern humans in Southern China before the Neolithic period, we obtained and sequenced the mitochondrial genomes of two 11,000-year-old ancient humans found in Southern China. Their complete mitochondrial genomes were analyzed together with the associated 4784 present-day sequences retrieved from GenBank and 82 ancient individuals from Southern China and Southeast Asia [2–4]. We aimed to identify the genetic relationship between early modern humans in Southern China and present-day Southeast/East Asians.

Materials and methods

The bone powder of two individuals over 11,000 years old were obtained from Longlin Laomocao Cave (24°38′31″N, 105°09′56″E, 11,510 Before Present (BP)) of Guangxi Zhuang Autonomous Region and Qingshuiyuan Dadong (QSYDD) (26°04′31.8″N, 106°49′53.8″E, 11,201–11,079

These authors contributed equally: Fan Bai, Xinglong Zhang

✉ Qiaomei Fu
fuqiaomei@ivpp.ac.cn

1 Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences, Institute of Vertebrate Paleontology and Paleoanthropology, CAS, 100044 Beijing, China

2 Center for Excellence in Life and Paleoenvironment, Chinese Academy of Sciences, 100044 Beijing, China

3 University of Chinese Academy of Sciences, 100049 Beijing, China

4 Guizhou Institute of Cultural Relics and Archaeology, Guiyang 550003, China

5 Yunnan Institute of Cultural Relics and Archaeology, Kunming 650118, China

6 Yunnan Key Laboratory of Earth System Science, Yunnan University, Kunming 650500, China

7 State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China

8 KIZ/CUHK Joint Laboratory of Bioresources and Molecular Research in Common Diseases, Kunming 650223, China

Published online: 11 July 2020

Table 1 Location, age, and sequencing quality of ancient individuals in this study

Sample	Skeletons	Site	Age BP	Coverage	5′CT%	3′CT%	Contamination ratio %	G6257A		G11518A	
								Coverage	Suppported	Coverage	Suppported
L0911	Petrous	Longlin, Guangxi	11,510[a]	61	30.1	47.4	23.7				
L2034	Petrous	Longlin, Guangxi	11,510[a]	268	29.3	52.1	9.9				
L2043	Petrous	Longlin, Guangxi	11,510[a]	87	25.7	54	17.1				
Longlin_d[b]	Petrous	Longlin, Guangxi	11,510[a]	99	61.8	74	4.4	72	94.4%	88	94.3%
L5068	Tooth	QSYDD, Guizhou	11,201–11,079[c]	359	46.9	31.8	2.7	324	99.1%	233	99.6%

[a]Indirect date based on associated archeological layers [20].

[b]Information was determined using the damage-restricted fragments from the Longlin libraries

[c]Calibrated radiocarbon date range of the skeleton

BP, dating) [5] of Guizhou Province in Southern China (Table 1 and Fig. 1a). The skeletons were curated by Yunnan and Guizhou Institute of Cultural Relics and Archaeology, respectively, and their sampling was performed with appropriate permissions. Both samples were then processed in a dedicated aDNA laboratory at the Institute of Vertebrate Paleontology and Paleoanthropology (IVPP), Chinese Academy of Sciences in Beijing, China. The DNA was extracted as described previously [6]. Libraries were constructed by a single-strand library preparation method and without treatment with the uracil-specific excision reagent enzyme mix. Hybridization capture with mitochondrial DNA (mtDNA) probes was performed to enrich for mitochondrial sequences [7], and enriched mtDNA libraries were sequenced using Illumina Miseq, generating 2×76 bp paired-end reads.

Merged reads with a length of at least 30 bp were aligned to the revised Cambridge Reference Sequence (rCRS) [8]. We removed duplicates and estimated the level of contamination by present-day human DNA in each library using ContamMix [9]. Table 1 shows the QSYDD library has low contamination (2.7%). However, the Longlin libraries show a higher contamination level (9.9–23.7%), so for this sample, only deaminated fragments were used in our analyses, which resulted in 99-fold coverage of the mtDNA genome. To facilitate comparison with the ancient samples, an additional 4784 present-day East Asian/Southeast Asian mtDNAs retrieved from the literature and analyzed [10–15]. Haplogroups were assigned using HaploGrep2 [16] with PhyloTree mtDNA tree Build 17 (released Feb 2016). The phylogenetic tree of the complete mtDNA sequences, disregarding the C-stretch length polymorphism found in regions 303–315, was reconstructed by mtPhyl (https://sites.google.com/site/mtphyl/home). The Bayesian phylogeny of complete mtDNA sequences was constructed by using BEAST 1.10.4 [17] with the GTR + I + G substitution model. A strict clock and piecewise-linear tree prior were used. The most recent common ancestor (TMRCA) of the mtDNA lineage clade was estimated using tip-calibration of two ancient samples. The median-joining network of 81 complete mtDNA genomes was constructed by PopArt1.7 (http://popart.otago.ac.nz/index.shtml). Age of novel subhaplogroup was also estimated by using the statistics rho [18] and sigma [19], whereas the recently proposed calibration rates for mtDNA were adopted [9].

Results and discussion

Both sequences from the two ancient humans (QSYDD and Longlin) included the diagnostic variants of mtDNA haplogroup M71, which is presently found in Southern China and mainland Southeast Asia [12]. According to the latest

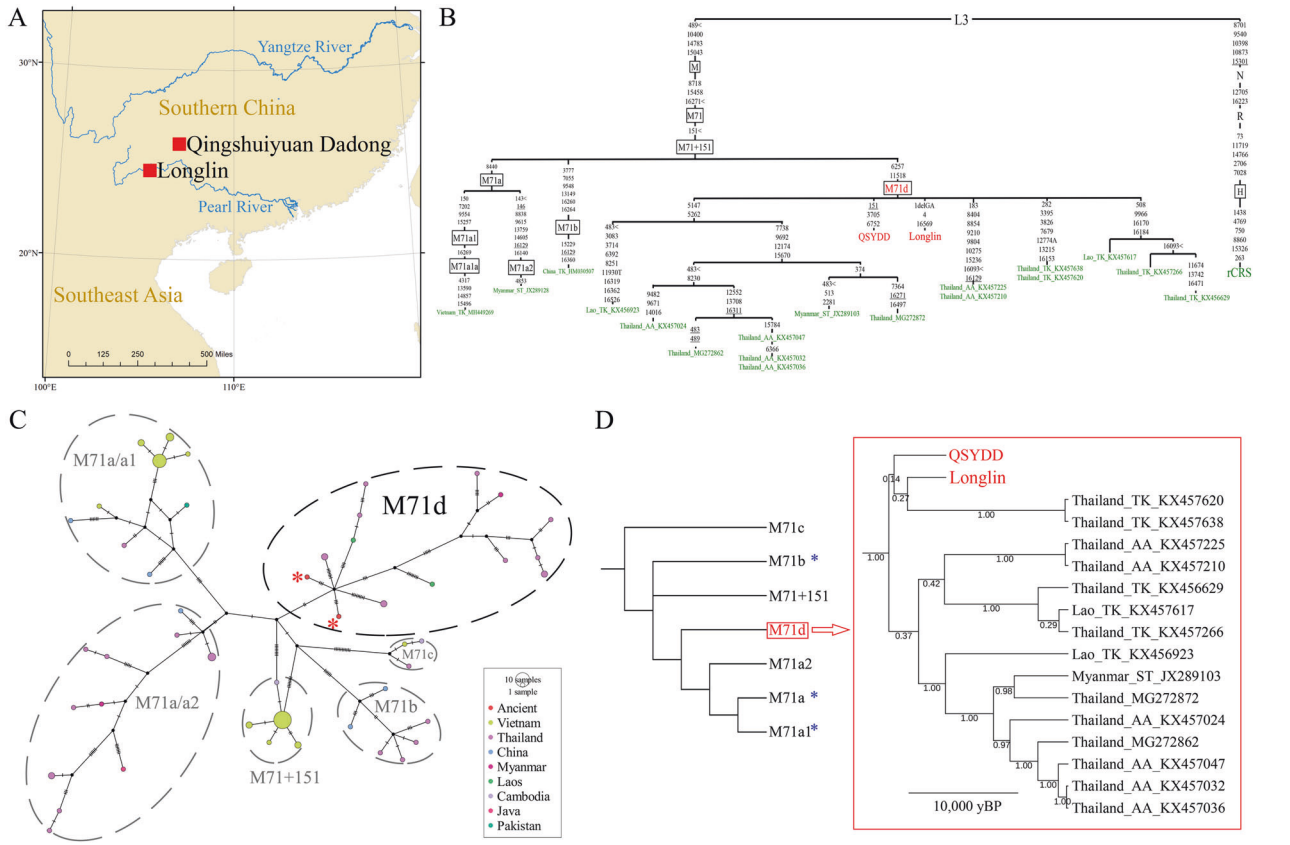

Fig. 1 **a** Geographic sampling locations of Longlin and QSYDD from Southern China; **b** phylogenetic tree of the new M71d lineage. This tree includes 17 complete mtDNA sequences and illustrates that QSYDD and Longlin are basal branches within M71d. The nucleotide positions of the sequences were scored relative to the revised Cambridge Reference Sequence (rCRS) [8]. The deletions and insertions are shown as "del" and "ins", respectively. Parallel mutations are indicated by '>' and back mutations are underlined. **c** Median-joining network of haplogroup M71 with respect to 81 present-day complete mtDNA genomes (geographic location indicated by colors) and the nested clade of M71d that includes the two ancient sequences (indicated by red asterisk); **d** Left, phylogenetic tree of the haplogroup M71 using 2 ancient and 81 present-day complete mtDNA genomes (blue stars indicate haplotypes found in present-day Chinese samples). Right, Detailed tree of the new M71d clade, where its estimated coalescence age is 22,210 yBP (95% HPD, 14,337–31,515). Posterior values are shown. *Abbreviations for language family: Tai-Kadai (TK), Austroasiatic (AA), Sino-Tibetan (ST), Austronesian (AN)

version of PhyloTree build 17 (released Feb 2016), the nomenclature of the M71 lineage includes three subclades: M71a, M71b, and M71c. We found that the two ancient individuals from Southern China shared two common substitutions in the protein coding regions G6257A and G11518A. We scanned previously published data and identified 15 out of 81 M71 whole mtDNA genomes sharing these same variations. All these samples were from mainland Southeast Asia: 12 samples from Thailand, 2 samples from Laos, and 1 sample from Myanmar. We named the subhaplogroup M71d for the cluster of 15 modern and 2 ancient sequences (Fig. 1b) to update the phylogeny of haplogroup M71. M71d shares a single variant (C151T) with M71a′b and includes the additional variants G6257A and G11518A in the coding regions. Our analysis shows the ancient individuals QSYDD and Longlin to have basal lineages of the mitochondrial subhaplogroup M71d (Fig. 1b).

In order to better understand the demographic history of M71, we constructed a median-joining network of the M71 haplogroup using all 81 whole mtDNA representative sequences (Fig. 1c). All 17 sequences of the M71d subhaplogroup reported here form a distinct lineage, with the 15 present-day individuals all located in mainland Southeast Asia and the two ancient samples from Southern China. Two ancient samples are basal to all 15 present-day samples within this clade. This pattern shows a genetic connection between ancient Southern China and present-day mainland Southeast Asians.

We used Bayesian phylogenetic methods to estimate the coalescence ages of M71d node based on complete mtDNA sequences. The early splits separated the two aDNA sequences as basal lineages from other modern sequences, although the posterior possibilities (<0.5) for the nodes were low (Fig. 1d). The TMRCA of M71d was dated to ~22.21 kya (14.34–31.52 kya, 95% HPD). "Rho" method

gave a similar age for M71d with a narrower time interval: 22.75 (18.90, 26.60) kya, $\rho \pm \sigma$:10.07 ± 1.70. Our results reveal a Paleolithic genetic link between Southern China and mainland Southeast Asia, reflecting ancient dispersal. Surprisingly, the M71d has not been detected in modern populations from Southern China. This could be explained as either genetic drift or population replacement, especially given the massive ancient dispersals between Southern China and mainland Southeast Asia around the late Upper Pleistocene and the early Holocene [4]. Since the coalescence time of M71d is much older than the two ancient samples (~11 kya), we cannot reject the scenerio that M71d originated in Southeast Asia and dispersed northward into southern China prior to the early Neolithic period. More aDNA data, including ancient nuclear genomes, will provide a better understanding of the connection between mainland Southeast Asia and Southern China during the early Holocene.

Data availability

The whole mitochondrial genome sequence data reported in this paper have been deposited in the Genome Warehouse in National Genomics Data Center, Beijing Institute of Genomics (China National Center for Bioinformation), Chinese Academy of Sciences, under accession number GWHANOT01000000 and GWHANOU01000000 that is publicly accessible at https://bigd.big.ac.cn/.

Acknowledgements We would like to thank A.M. Ko, and E.A. Bennett for comments.

Funding This work was supported by the CAS (XDB26000000) and NSFC (91731303, 41925009, and 41630102).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Jin L, Su B. Natives or immigrants: modern human origin in east Asia. Nat Rev Genet. 2000;1:126–33.
2. McColl H, Racimo F, Vinner L, Demeter F, Gakuhari T, Moreno-Mayar JV, et al. The prehistoric peopling of Southeast Asia. Science. 2018;361:88–92.
3. Zhang X, Li C, Zhou Y, Huang J, Yu T, Liu X, et al. A matrilineal genetic perspective of hanging coffin custom in Southern China and Northern Thailand. iScience. 2020;23:101032.
4. Yang MA, Fan X, Sun B, Chen C, Lang J, Ko Y-C, et al. Ancient DNA indicates human population shifts and admixture in northern and southern China. Science. 2020;eaba0909.
5. Zhang X, Pei S, Wu X, Bar-Yosef O. Qingshuiyuan Dadong: a newly discovered Late Paleolithic site in Guizhou province, China. Archaeol Res Asia. 2017;9:13–21.
6. Dabney J, Knapp M, Glocke I, Gansauge M-T, Weihmann A, Nickel B, et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proc Natl Acad Sci USA. 2013;110:15758–63.
7. Fu Q, Meyer M, Gao X, Stenzel U, Burbano HA, Kelso J, et al. DNA analysis of an early modern human from Tianyuan Cave, China. Proc Natl Acad Sci USA. 2013;110:2223–7.
8. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet. 1999; 23:147.
9. Fu Q, Mittnik A, Johnson PLF, Bos K, Lari M, Bollongino R, et al. A revised timescale for human evolution based on ancient mitochondrial genomes. Curr Biol. 2013;23:553–9.
10. Kutanan W, Kampuansai J, Brunelli A, Ghirotto S, Pittayaporn P, Ruangchai S, et al. New insights from Thailand into the maternal genetic history of Mainland Southeast Asia. Eur J Hum Genet. 2018;26:898–911.
11. Duong NT, Macholdt E, Ton ND, Arias L, Schröder R, Phong NV, et al. Complete human mtDNA genome sequences from Vietnam and the phylogeography of Mainland Southeast Asia. Sci Rep. 2018;8:11651.
12. Kong Q-P, Sun C, Wang H-W, Zhao M, Wang W-Z, Zhong L, et al. Large-scale mtDNA screening reveals a surprising matrilineal complexity in East Asia and its implications to the peopling of the region. Mol Biol Evol. 2011;28:513–22.
13. Kutanan W, Kampuansai J, Srikummool M, Kangwanpong D, Ghirotto S, Brunelli A, et al. Complete mitochondrial genomes of Thai and Lao populations indicate an ancient origin of Austroasiatic groups and demic diffusion in the spread of Tai–Kadai languages. Hum Genet. 2017;136:85–98.
14. Summerer M, Horst J, Erhart G, Weißensteiner H, Schönherr S, Pacher D, et al. Large-scale mitochondrial DNA analysis in Southeast Asia reveals evolutionary effects of cultural isolation in the multi-ethnic population of Myanmar. BMC Evol Biol. 2014;14:17.
15. Yao Y-G, Kong Q-P, Bandelt H-J, Kivisild T, Zhang Y-P. Phylogeographic differentiation of mitochondrial DNA in Han Chinese. Am J Hum Genet. 2002;70:635–51.
16. Weissensteiner H, Pacher D, Kloss-Brandstätter A, Forer L, Specht G, Bandelt H-J, et al. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. Nucleic Acids Res. 2016;44(W1):W58–63.
17. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol. 2018. https://academic.oup.com/ve/article/4/1/vey016/5035211
18. Forster P, Harding R, Torroni A, Bandelt HJ. Origin and evolution of Native American mtDNA variation: a reappraisal. Am J Hum Genet. 1996;59:935–45.
19. Saillard J, Forster P, Lynnerup N, Bandelt H-J, Nørby S. mtDNA variation among Greenland Eskimos: the edge of the Beringian Expansion. Am J Hum Genet. 2000;67:718–26.
20. Curnoe D, Xueping J, Herries AIR, Kanning B, Taçon PSC, Zhende B, et al. Human Remains from the Pleistocene-Holocene Transition of Southwest China Suggest a Complex Evolutionary History for East Asians. PLoS ONE 2012;7:e31918.