Journal of Genetics and Genomics 48 (2021) 899–907

Contents lists available at ScienceDirect

Journal of Genetics and Genomics

Journal homepage: www.journals.elsevier.com/journal-of-genetics-and-genomics/

Original research

Maternal genetic history of southern East Asians over the past 12,000 years

Yalin Liu [a, b, c, 1], Tianyi Wang [a, c, d, 1], Xichao Wu [e, 1], Xuechun Fan [f, g], Wei Wang [h], Guangmao Xie [i, j], Zhen Li [i], Qingping Yang [i], Peng Cao [a], Ruowei Yang [a], Feng Liu [a], Qingyan Dai [a], Xiaotian Feng [a], Wanjing Ping [a], Bo Miao [a, d], Yun Wu [k, l], Yichen Liu [a, *], Qiaomei Fu [a, c, *]

[a] Key Laboratory of Vertebrate Evolution and Human Origins, Institute of Vertebrate Paleontology and Paleoanthropology, Center for Excellence in Life and Paleoenvironment, Chinese Academy of Sciences, Beijing 100044, China
[b] Sino-Danish Center, University of the Chinese Academy of Sciences, Beijing 100049, China
[c] University of the Chinese Academy of Sciences, Beijing 100049, China
[d] Northwest University, Xi'an 710069, China
[e] Fujian Longyan Museum, Longyan 364000, China
[f] International Research Center for Austronesian Archaeology, Pingtan 350000, China
[g] Fujian Museum, Fuzhou 350001, China
[h] Institute of Cultural Heritage, Shandong University, Qingdao 266237, China
[i] Guangxi Institute of Cultural Relic Protection and Archaeology, Nanning 530022, China
[j] College of History, Culture and Tourism, Guangxi Normal University, Guilin 541001, China
[k] Yunnan Institute of Cultural Relics and Archaeology, Kunming 650118, China
[l] Archaeological Institute for Yangtze Civilization, Wuhan University, Wuhan 430072, China

ARTICLE INFO

Article history:
Received 19 April 2021
Received in revised form
4 June 2021
Accepted 5 June 2021
Available online 21 June 2021

Keywords:
Ancient DNA
Southern East Asians
Population history
Maternal genetic structure

ABSTRACT

Southern East Asia, including Guangxi and Fujian provinces in China, is home to diverse ethnic groups, languages, and cultures. Previous studies suggest a high complexity regarding population dynamics and the history of southern East Asians. However, large-scale genetic studies on ancient populations in this region are hindered by limited sample preservation. Here, using highly efficient DNA capture techniques, we obtain 48 complete mitochondrial genomes of individuals from Guangxi and Fujian in China and reconstruct their maternal genetic history over the past 12,000 years. We find a strong connection between southern East Asians dating to ~12,000–6000 years ago and present-day Southeast Asians. In addition, stronger genetic affinities to northern East Asians are observed in historical southern East Asians than Neolithic southern East Asians, suggesting increased interactions between northern and southern East Asians over time. Overall, we reveal dynamic connections between ancient southern East Asians and populations located in surrounding regions, as well as a shift in maternal genetic structure within the populations over time.

Introduction

Southern East Asia, the southern part of East Asia divided by Qinling-Huaihe (not including Southeast Asia), including Guangxi and Fujian provinces in China, is a hub for demic and cultural interactions between East Asians and Southeast Asians (Stoneking and Delfin, 2010). Genetic and archaeological evidence suggests immense complexity regarding the population dynamics and the history of this region (O'Connell et al., 2018; Wang et al., 2021). On the one hand, southern East Asians were genetically connected to various populations in East Asia and Southeast Asia (McColl et al., 2018; Yang et al., 2020). Specifically, southern East Asians from the Fujian region were proposed to be the origin of proto-Austronesians, and Neolithic southern East Asians showed genetic connections to northern East Asians, western East Asians on the Plateau, and Siberians (Yang et al., 2020). On the other hand, the past dynamics

* Corresponding authors.
E-mail addresses: yichen.liu@ivpp.ac.cn (Y. Liu), fuqiaomei@ivpp.ac.cn (Q. Fu).
[1] These authors contributed equally to this work as co-first authors.

https://doi.org/10.1016/j.jgg.2021.06.002

between southern East Asians and Southeast Asians remains controversial. The modern humans who migrated out of Africa settled in Southeast Asia and formed hunter-gathers at least 43,500 years before present (BP, BP is before 1950 A.D.) (Lahr and Foley, 1994; Ji et al., 2016). Based on cranial and dental morphologies, it has been hypothesized that East Asian farmers replaced the indigenous hunter-gatherers in Southeast Asia (Matsumura et al., 2019), whereas genetic evidence suggested that both East Asian farmers and indigenous hunter-gatherers contributed to genetic diversities of present-day Southeast Asians (McColl et al., 2018).

In addition, southern East Asia is home to various cultures. Diversified cultural traditions were found in populations from southern East Asia (e.g., Guangxi and Fujian), as suggested by the archaeological evidence (Fu, 1988; Mei, 1989). Burial practices also play an important role in ancient cultures, and many groups have their burial traditions. Cave Burial and Hanging Coffin are two common burial practices in southern East Asia and Southeast Asia, the former buries dead in caves in mountains, and the latter hangs coffins over the cliff (Chen, 1989a; Luo, 2000). Hanging Coffin originated in Fujian ~3600 BP (Lin et al., 1980), and it is earlier than the Cave Burial in Guangxi, which has been widely considered to appear ~2000 BP (Peng, 2001). Archaeological evidence shows that both the burial practices are related to the cliff and practiced by Hmong-Mien populations, who are widely distributed in southern China and Southeast Asia (Chen, 1989b; Wen et al., 2004). However, the genetic relationship between the groups practicing these two types of burial remains unclear.

The ancient DNA (aDNA) technique is a powerful tool for recovering genetic information from ancient materials, thereby offering a better understanding of human population history and evolution from a genetic perspective (Zhang and Fu, 2020). The quality and quantity of obtained aDNA data largely depend on the preservation of samples. Hot and humid regions, such as southern East Asia, are unfavorable for DNA preservation, which hinders the retrieval of aDNA from samples collected from such regions. One solution is using highly efficient DNA capture techniques to enrich target DNA fragments to increase the yield. Due to the small genomic size and high copy number in cells, mitochondrial DNA (mtDNA) serves as a good candidate for DNA capture and investigating the maternal genetic structure of ancient populations (Fu et al., 2013a). Here, we obtain whole mitochondrial genomes from 48 ancient southern East Asian individuals to examine (1) the maternal genetic structure of ancient southern East Asians, (2) the maternal genetic dynamics of southern East Asians over the past 12,000 years, (3) the maternal genetic relationships between ancient southern East Asians and populations located in surrounding regions, and (4) the maternal genetic relationships between the individuals practicing Cave Burial and Hanging Coffin.

Results

Sample and grouping information

To gain a better genetic perspective on the history of southern East Asians, we obtained complete mitochondrial genomes from 48 individuals dated from 11,747 to 294 calibrated years before present (cal BP) from 20 sites in Guangxi and Fujian of southern China (Table S1). DNA was extracted from about 100 mg of bone or tooth powder, and single-stranded libraries were constructed (Dabney et al., 2013; Gansauge and Meyer, 2013). DNA libraries were enriched for human mitochondrial DNA using designed probes and then sequenced at Illumina MiSeq platform, resulting in mitochondrial genomes between 15-fold and 560-fold coverage (the average 191-fold coverage; Table S1). Contamination rates were estimated across all the samples (Table S1). To avoid potential biases introduced from contamination, only reads with typical aDNA damage (deamination at the end of the fragments) were kept for follow-up analyses for the 10 samples that with a contamination rate higher than 4% (Table S1; the rate 3%–4% is an empirical cutoff widely used in aDNA studies) (Rohland et al., 2015; Fu et al., 2016). We also excluded an individual (BalongKD06) with a probable matrilineal kinship (BalongKD06 and BalongKD10 share an identical mitochondrial genomic sequence and are from the same tomb), resulting in 47 sequences for analysis (Table S1).

To increase the temporal and geographical coverage, we additionally included two individuals from previous studies into our ancient southern East Asian data set: Longlin (11,765–11,255 cal BP, in Guangxi) (Bai et al., 2020) and Liangdao (8320–8060 cal BP, in the Fujian surrounding area) (Ko et al., 2014), resulting in a total of 49 samples that representative of populations distributed in southern East Asia over the past 12,000 years. To avoid artificially low genetic diversity introduced from insufficient sampling and the availability of samples, only populations covered by no less than five individuals were included. Individuals dating to around 12,000–6000 BP were not sufficient to group into populations and thus were analyzed separately. With the highest percentage variance among groups, sampled individuals after ~5000 BP were combined into two main chronological groups: (1) the Late Neolithic group (the LN group): including 10 individuals dated to 4644–4225 cal BP; (2) the Historical group: including 32 individuals dated to 1688–294 cal BP (Fig. 1A and 1B; Table S1; see Supplementary Data).

A total of 37 haplotypes belonging to 17 haplogroups were identified from these 49 individuals (Fig. 1C; Tables S2 and S3). We then obtained 174 ancient and 6859 present-day whole mitochondrial genomes of individuals from East Asia (EA), Southeast Asia (SEA), Central Asia (CA), South Asia (SA), North Asia (NA), and Melanesia from previous studies for follow-up analyses (Tables S4 and S5). The haplotype diversity of ancient southern Chinese is 0.9906, which is just lower than the present-day Han (0.9945) among all southern Chinese populations, and all of them are higher than 0.9, suggesting the genetic diversity did not change a lot over time (Table S6). We first carried out principal component analysis (PCA) using population haplogroup frequencies to assess the genetic relationships of these populations included in this study (Fig. 1D; Table S7). Overall, the sampled ancient southern Asians (red) after 5000 BP are clustered closely to East Asians. Late Neolithic southern East Asians cluster with present-day southern East Asians, whereas historical southern East Asians are located in the overlap of present-day southern and northern East Asians, suggesting a shift from southern East Asians to northern East Asians over time. In the following sections, we will discuss the maternal genetics of these Late Neolithic and historical southern East Asians in detail, including their mitochondrial haplotypes and haplogroups profiles, as well as the networks and the Bayesian phylogenetic trees built using whole mitochondrial sequences.

Early southern East Asians in Guangxi and Fujian

We first examined the maternal genetics of early southern East Asians, including four individuals (Longlin, Dushan, Baojianshan5_M1, and Baojianshan5_M2) from Guangxi and three individuals (Qihe2, Qihe3, and Liangdao) from Fujian and the surrounding area (Table S1).

Among seven early southern East Asians, three haplotypes were identified from four Guangxi individuals: M71d (Longlin, 11,765–11,255 cal BP), B4a1e (Dushan, 8974–8593 cal BP), and M75 (Baojianshan_M1, Baojianshan_M2, ~8335–6400 BP). Although B4a1e and M75 are found in a small proportion of present-day southern East Asians, all three haplotypes are predominantly distributed in present-day Southeast Asians (Kong et

al., 2011; Kutanan et al., 2017; Larruga et al., 2017). Identification of these haplotypes in early southern East Asians may reflect a connection between the early Guangxi population and present-day Southeast Asians. To further investigate this connection, we constructed Bayesian phylogenetic trees and networks of the M75 lineage and B4a1e lineage (Figs. 2A, 2B, S1A, S1B) and compared our results with previously published results of M71d lineage (Bai et al., 2020). The B4a1e lineage forms two clades: one consists of three Southeast Asian individuals and the other consists of one early individual (Dushan), one historical individual (QinchangKD13), two present-day southern East Asians, and three present-day Southeast Asians, in which the ~9000-year-old Dushan individual locates at a basal position (Fig. 2A and 2B). All B4a1e individuals share a common ancestor 12,077 BP (16,368–9285 BP, highest posterior density [HPD] 95%; Fig. 2B; Table S8). Haplotype M75 is only identified from one Southeast Asian among the 7033 individuals surveyed in our study. The estimated coalescence time of M75 is 36,514 BP (52, 197–24,320 BP, HPD 95%; Fig. S2B; Table S8), suggesting it represents a lineage that much older than B4a1e lineage and the previously published M71d lineage (~22,210 BP, ~14,340–31,520 BP, HPD 95%) (Bai et al., 2020). Based on the sampled Baojianshan individuals, the dating of this extremely old lineage likely suggests a deeper connection between southern East Asians and Southeast Asians than previously expected. However, it is also possible that the haplotype M75 was carried by ancient individuals from southern East Asia to Southeast Asia. Therefore, further ancient DNA evidence is needed to confirm this hypothesis.

Similarly, the haplotypes identified from three individuals from Fujian and the surrounding area (Qihe3 [11,747–11,356 cal BP]: E; Liangdao [8320–8060 cal BP]: E1; Qihe2 [8428–8359 cal BP]: R9c1b) are present in Southeast Asians and southern East Asians today (Ko et al., 2014; Soares et al., 2008). Interestingly, these haplotypes seem to reflect their affinities to present-day Austronesians, with haplogroup E (including the subhaplogroup E1) prevalent among Austronesians (Ko et al., 2014), and the haplogroup R9 is considered to be the Austronesian ancestry (Tofanelli et al., 2009). We further carried out the phylogenetic analysis on these lineages. Qihe2 locates at the basal position of haplogroup R9c1b, and the estimated coalescence time of this lineage is 15,579 BP (21,766–10,719 BP, HPD 95%) (Figs. S2C and S2D; Table S8). Interestingly, Liangdao and Qihe3 locate at the basal positions of haplogroups E1 and E2, respectively (Fig. 2C). Although Liangdao has been suggested to have the most ancestral E sequence (Ko et al., 2014), here Qihe3, ~3400 years older than Liangdao, represents a distinct ancestral lineage that even older than Liangdao in the Bayesian phylogenetic tree (Fig. 2D), and these two individuals are five mutations away with differences at nucleotide positions 5513-10029-13254-14577-14783 (Figs. 2C and S2). The coalescence time of lineage E has been previously estimated to be ~9000 BP (~11,000–8000 BP, HPD 95%, relax model) or ~15,000 BP (~18,000–11,000 BP, HPD 95%, strict model) (Ko et al., 2014). Here, we included the ~12,000-year-old Qihe3 (in the lineage E), the oldest individual that sampled from Fujian and the surrounding area so far, and the newly estimated coalescence time of lineage E is 14,816 BP (19,898–11,982 BP, HPD 95%, relax model; Table S8) or 19,313 BP (25,560–14,470 BP, HPD 95%, strict model; Fig. 2D; Table S8), ~5000 years earlier than previously estimated.

Furthermore, a total of six different haplotypes are identified from seven early southern East Asians in Guangxi and Fujian (including the surrounding area; Fig. 1C), which is suggestive of high maternal genetic diversity. The observation that no haplotype is shared between Guangxi and Fujian individuals dating to ~12,000–6000 BP likely indicates a differentiation in maternal genetics between the early Guangxi and Fujian southern East Asians. However, these speculations are limited by the number of samples. Further sampling is required to confirm the genetic structure of early southern East Asians.

The Late Neolithic (LN) group in Fujian

The LN group includes 10 individuals from Fujian (Tanshishan, $n = 3$; Xitoucun, $n = 7$; Table S1). From these individuals, we identified 10 haplotypes belonging to five haplogroups (A, B4, B5, F, and R+16189C; Fig. 1C; Table S1), and all of them are common in East Asians and Southeast Asians today (Wall et al., 2019; Kutanan et al., 2020).

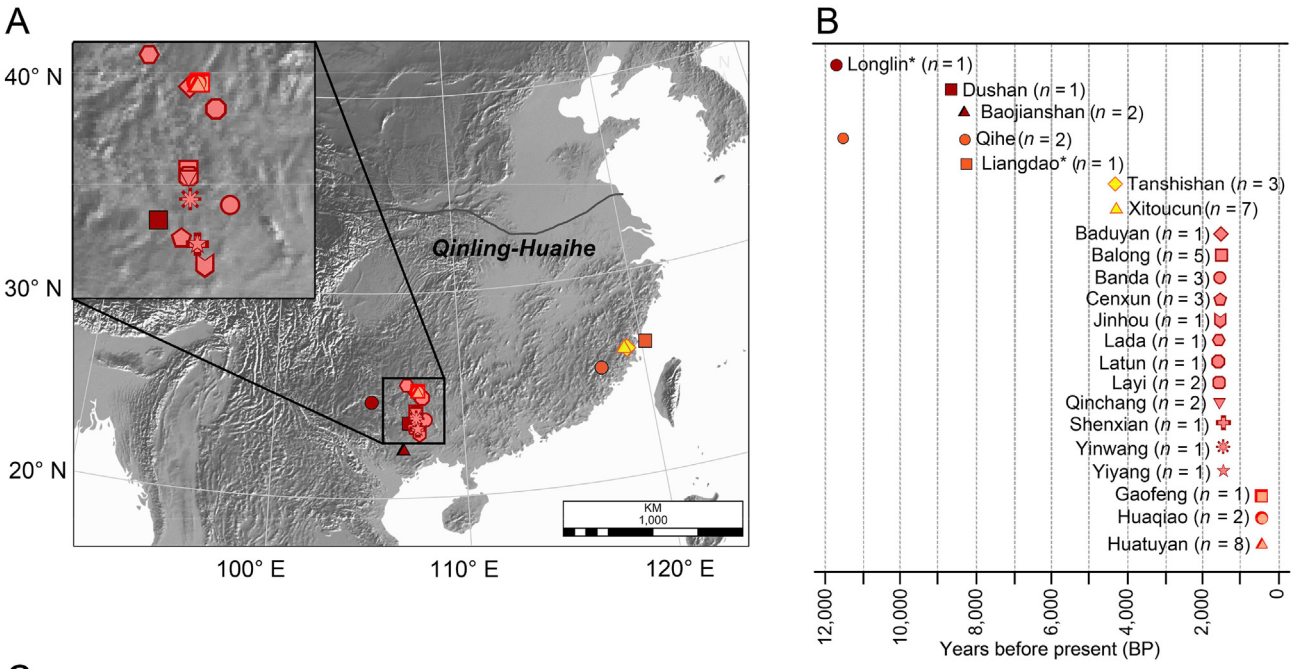
Among these haplotypes, four of them (F4b, F3b1, F3a1, and F3b1a+16093C) belong to the haplogroup R9, and three (B4b1, B4b1c1, and B4c1b2a) belong to haplogroup B4 (Table S1). Haplogroups R9 and B4 have been identified from early individuals in Fujian and Guangxi, respectively (Fig. 1C). This result is suggestive of a link regarding maternal genetics between the LN group and early individuals in Guangxi and Fujian. Furthermore, three haplogroups (A, B5, and R+16189C) that absent in early southern East Asians are identified from the LN group (Fig. 1C). These haplogroups are widely found across East Asians and Southeast Asians, and our observation may indicate the relatively high maternal diversity of southern East Asians in Late Neolithic. However, the haplogroups (E, M71, and M75) that are mainly distributed in present-day Southeast Asians are absent from the LN group (Fig. 1C), suggesting a weaker link between the southern East Asians from this period and present-day Southeast Asians comparing to early individuals.

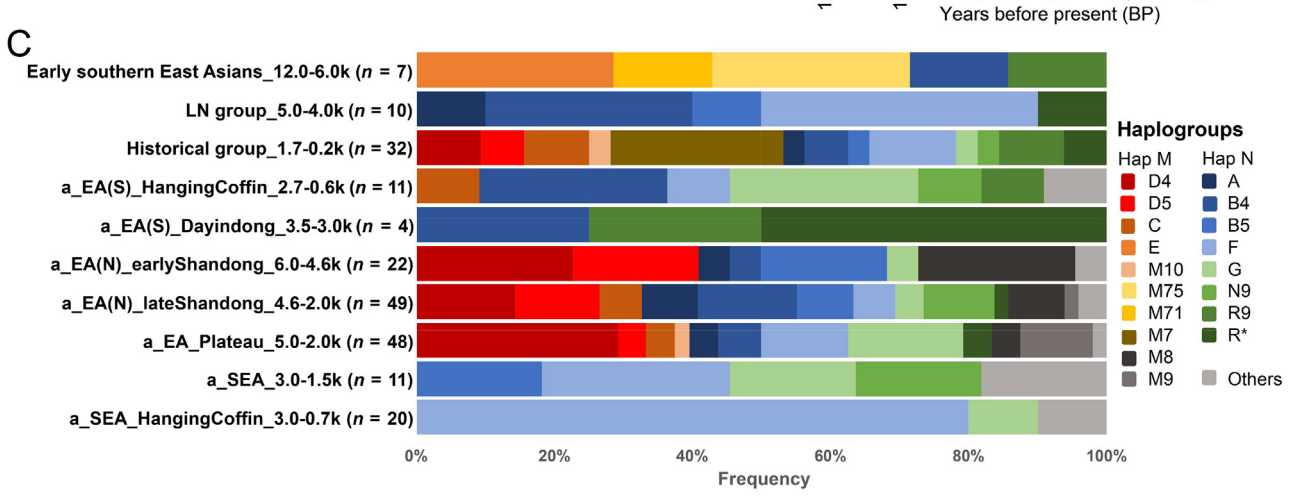
To further explore the genetic relationships between the LN group and other populations, we calculated the pairwise genetic distances (Φ_{ST}), and the pattern of the multidimensional scaling (MDS) plot (Fig. S3) built of Φ_{ST} supported the PCA result (Fig. 1D). As expected, the LN group has a smaller Φ_{ST} with ancient East Asians than ancient Southeast Asians (Fig. 3A; Table S9). When compared with present-day populations, the LN group is more closely related to southern East Asians and Southeast Asians than northern East Asians. Haplogroup sharing shows a similar result that the LN group shares the highest level of haplogroups with southern East Asians (15.93%), followed by Southeast Asians (excluding Myanmar populations; 12.83%), and northern East Asians (9.29%) (Fig. 3B; Table S10).

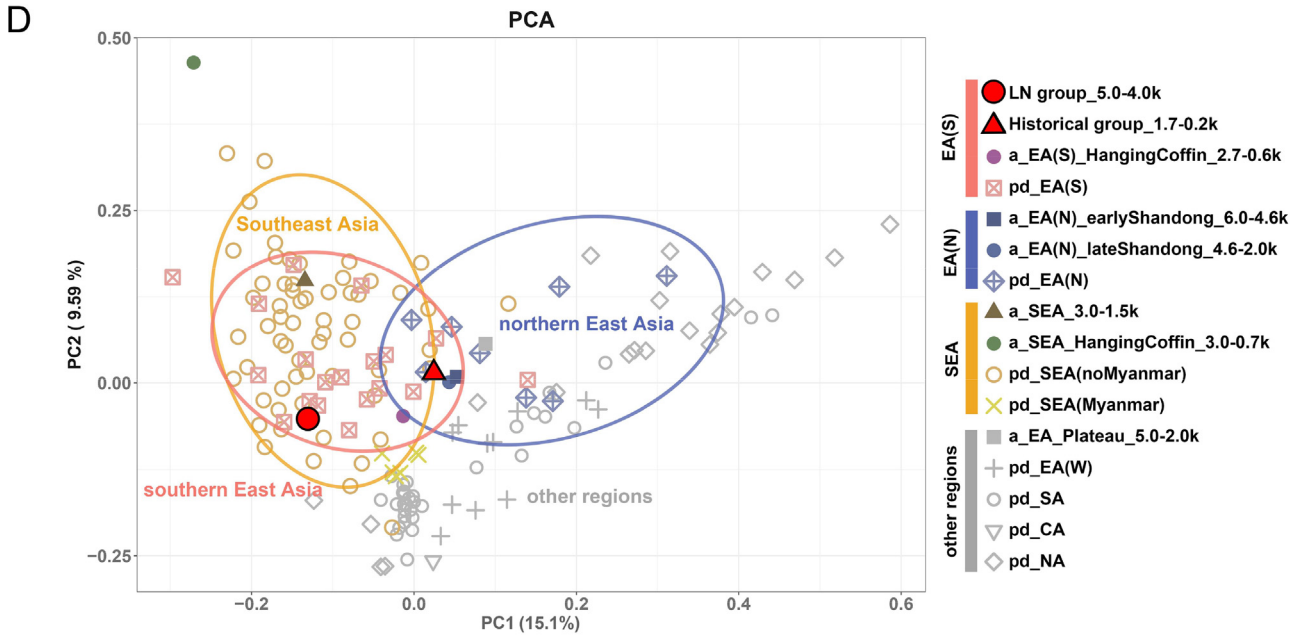
Although the connection between the LN group and northern East Asians is not as very strong, a haplogroup that more prevalent in northern East Asians, B5b2, is identified in the LN group. Therefore, we constructed the network and the Bayesian phylogenetic tree using our Xitoucun individual (Xitoucun_M13: B5b2a2a2) with ancient and present-day individuals carrying the haplogroup B5b (Fig. 2E and 2F). The Xitoucun individual and three present-day individuals from Malaysia and Solomons cluster together, and both the Bayesian phylogenetic tree and network support sharing a common B5b2 ancestor that is related to the ~5500- to 4600-year-old early Shandong individuals as well as the ~9500-year-old Bianbian individual. This is consistent with the hypothesis that northern East Asian ancestry spread across southern East Asia after the Neolithic, transforming the genetic ancestry of southern China (Yang et al., 2020). In addition, the coalescence time of the B5b2 lineage estimated using our data set is 19,070 BP (24,966–14,479 BP, HPD 95%; Fig. 2F; Table S8), narrowing down the coalescence time of this lineage for over 6000 years (Liu et al., 2021).

The Historical group in Guangxi

We also explored maternal genetics in the Historical group, including 32 individuals from 15 archaeological sites in Guangxi (Table S1). These individuals are allocated into 24 haplotypes belonging to 13 haplogroups A, B4, B5a, C, D4, D5, F, G, M10, M7, N9, R+16189c, and R9 (Fig. 1C; Table S1). These haplogroups are


A
40° N
30° N
20° N
100° E
110° E
120° E
Qinling-Huaihe
KM
1,000
B
Longlin* (n = 1)
Dushan (n = 1)
Baojianshan (n = 2)
Qihe (n = 2)
Liangdao* (n = 1)
Tanshishan (n = 3)
Xitoucun (n = 7)
Baduyan (n = 1)
Balong (n = 5)
Banda (n = 3)
Cenxun (n = 3)
Jinhou (n = 1)
Lada (n = 1)
Latun (n = 1)
Layi (n = 2)
Qinchang (n = 2)
Shenxian (n = 1)
Yinwang (n = 1)
Yiyang (n = 1)
Gaofeng (n = 1)
Huaqiao (n = 2)
Huatuyan (n = 8)
12,000
10,000
8,000
6,000
4,000
2,000
0
Years before present (BP)


C
Early southern East Asians_12.0-6.0k (n = 7)
LN group_5.0-4.0k (n = 10)
Historical group_1.7-0.2k (n = 32)
a_EA(S)_HangingCoffin_2.7-0.6k (n = 11)
a_EA(S)_Dayindong_3.5-3.0k (n = 4)
a_EA(N)_earlyShandong_6.0-4.6k (n = 22)
a_EA(N)_lateShandong_4.6-2.0k (n = 49)
a_EA_Plateau_5.0-2.0k (n = 48)
a_SEA_3.0-1.5k (n = 11)
a_SEA_HangingCoffin_3.0-0.7k (n = 20)
0%
20%
40%
60%
80%
100%
Frequency
Haplogroups
Hap M
D4
D5
C
E
M10
M75
M71
M7
M8
M9
Hap N
A
B4
B5
F
G
N9
R9
R*
Others


D
PCA
0.50
0.25
0.00
−0.25
PC2 (9.59 %)
−0.2
0.0
0.2
0.4
0.6
PC1 (15.1%)
Southeast Asia
northern East Asia
southern East Asia
other regions
EA(S)
LN group_5.0-4.0k
Historical group_1.7-0.2k
a_EA(S)_HangingCoffin_2.7-0.6k
pd_EA(S)
EA(N)
a_EA(N)_earlyShandong_6.0-4.6k
a_EA(N)_lateShandong_4.6-2.0k
pd_EA(N)
SEA
a_SEA_3.0-1.5k
a_SEA_HangingCoffin_3.0-0.7k
pd_SEA(noMyanmar)
pd_SEA(Myanmar)
other regions
a_EA_Plateau_5.0-2.0k
pd_EA(W)
pd_SA
pd_CA
pd_NA

widely present in East Asians and Southeast Asians today (Yao et al., 2002; Larruga et al., 2017; Wall et al., 2019). Haplogroups A, C, D4, D5, G, M10, and N9 are found in higher proportions of northern East Asians, whereas haplogroups B4, B5a, F, M7, and R9 are more likely found in southern East Asians (Tanaka et al., 2004; Derenko et al., 2010; Duong et al., 2018).

Among these haplogroups, B4 and R9 are shared by early southern East Asians, the LN group, and the Historical group; A, B5, and R+16189C are only shared between LN and Historical groups (Fig. 1C). These shared haplogroups may reflect maternal genetic continuity in southern East Asians from ~12,000 BP to the Historical period. At least Dushan (8974–8593 cal BP) and QinchangKD13 (1520–1363 cal BP) represent a highly likely case, in which the two individuals share the same haplotype (B4a1e) and QinchangKD13 clusters with Dushan as well as four present-day individuals from southern East Asia or Southeast Asia in the network and the Bayesian phylogenetic tree (Fig. 2A and 2B). Besides, 13 haplotypes belonging to seven haplogroups (C, D4, D5, M10, M7, G, and N9) are identified only in the Historical group (Table S1). All these newly emerged haplogroups, except for M7, are more frequently found in northern East Asians. Specifically, D4, D5, and C are present in both ancient Shandong populations and present-day northern East Asians with relatively high frequencies (Fig. 1C; Table S7). Also, Φ_{ST} results show that the Historical group is most closely related to the late Shandong population ($\Phi_{ST} = 0.02$, $P < 0.05$) among all ancient populations (Table S9). Together with the emergence of haplogroup D4, D5, and C in the Historical group, our observation may be a result of the increased influence of northern populations on southern East Asians from the Late Neolithic to the Historical Period. Compared with present-day populations, the Historical group is closely related to East Asians and Southeast Asians, with the lowest Φ_{ST} detected between the Historical group and present-day Miao ($\Phi_{ST} = 0.02$, $P > 0.05$), albeit the average Φ_{ST} between the Historical group and tested populations are close to each other (Fig. 3A; Table S9).

From archaeological evidence, the sampled Historical group was practicing Cave Burial, which is one of the important cultural traditions of southern East Asians (Luo, 2000). It remains unclear that if this population is closely related to a population practicing a similar burial tradition—Hanging Coffin (Chen, 1989b). Therefore, we compared our samples with historical southern East Asians and Southeast Asians practicing the Hanging Coffin tradition. Analysis of molecular variance (AMOVA) suggests that the sampled historical Cave Burial individuals cannot be assigned to the same population with either Hanging Coffin individuals from China or Thailand (Table S11). We further quantified the genetic difference between these populations using Φ_{ST}. The results show that although individuals practicing Cave Burials and Hanging Coffin are from different populations, the historical Cave Burial individuals are closer to Hanging Coffin individuals from East Asia ($\Phi_{ST} = 0.07$, $P = 0.02$) than Hanging Coffin individuals from Southeast Asia ($\Phi_{ST} = 0.25$, $P = 0.00$; Table S9). In China, Cave Burial individuals share haplotypes (C7a, and F3a1, 4.88%) with Hanging Coffin individuals, which is of higher level than with all the other ancient populations (0.00%–3.66%) and most of the present-day East Asian populations (10 of 15 populations: 0.00%–3.66%; Table S10). Compared with other ancient populations, Chinese Hanging Coffin population clusters with Cave Burial population and late Shandong population in the heatmap built using Φ_{st} values (Fig. S4). The previously mentioned result indicates that ancient populations in China practicing Cave Burial and Hanging Coffin have the genetic connection.

Discussion

By reconstructing the maternal genetic history of southern East Asians over the past 12,000 years, we reveal dynamic connections between ancient individuals in Guangxi and Fujian and populations located in surrounding regions, as well as a shift in maternal genetic structure within the populations over time. Adjacent to Southeast Asia, the genetic influence from southern East Asians to Southeast Asians has long been suggested, including the Paleolithic dispersal of southern East Asians and the expansion of East Asian farmers ~4000 BP in Southeast Asia (Zhang et al., 2015; McColl et al., 2018). Here, we find early southern East Asians (~12,000–6000 BP) share a strong connection with present-day Southeast Asians, hinting at a matrilineal influence from Southeast Asians to southern East Asians. However, due to a lack of chronological maternal genetic history of Southeast Asians, currently, we cannot rule out the possibility that this observation is from more recent interactions, and genome-wide analyses on ancient East and Southeast Asians can provide further evidence. Furthermore, mitochondrial haplotypes that are widely distributed in present-day Austronesians are identified in ancient southern East Asians (Qihe2 and Qihe3), which is consistent with the previously proposed southern East Asian origin of Austronesians (Yang et al., 2020). The nuclear genome information of some individuals (Tanshishan and Xitoucun) analyzed in this study suggested an admixture between southern and northern East Asians (Yang et al., 2020), whereas these individuals possess typical southern East Asian mitochondrial haplotypes, suggesting further complexity of the interactions between past East Asians. During the historical period, mitochondrial haplogroups typically found in northern East Asians emerged. These observations are likely resulted from increased interactions between northern and southern East Asians over time, as suggested earlier (Yang et al., 2020).

High-resolution characterization of extremely old lineages is critical for the reconstruction of population history. Here, we calibrated the coalescence time of two maternal lineages using newly sequenced samples. Specifically, the newly estimated coalescence time of lineage E is 14,816 BP, and the coalescence time of the B5b2 lineage estimated using our data set is 19,070 BP. The former (E) is a haplogroup widely distributed in present-day Austronesians, while the latter (B5b2) is frequently found in ancient northern East Asians (Ko et al., 2014; Liu et al., 2021). The estimated coalescence time of both is extremely old, and future studies tracking these lineages' back time will be necessary for a better understanding of the maternal genetic structure of early East Asians.

The profiling of ancient populations is difficult and requires multidisciplinary efforts. In this study, we leveraged the archaeological background of the historical individuals and explored the

Fig. 1. Geographic, temporal information, and genetic analysis of ancient individuals in Guangxi and Fujian (including the surrounding area). **A**: Map showing the locations from which ancient individuals were sampled for this study. Each symbol represents one site: reddish symbols are sampled in Guangxi, and yellowish symbols are sampled in Fujian and the surrounding area. **B**: Chronology of ancient individuals sampled and grouped in this study, where individuals date to around 12,000 BP to 200 BP. Longlin* ($n = 1$), Dushan ($n = 1$), Baojianshan ($n = 2$), Qihe ($n = 2$), Liangdao* ($n = 1$), Tanshishan ($n = 3$), Xitoucun ($n = 7$), Baduyan ($n = 1$), Balong ($n = 5$), Banda ($n = 3$), Cenxun ($n = 3$), Jinhou ($n = 1$), Lada ($n = 1$), Latun ($n = 1$), Layi ($n = 1$), Qinchang ($n = 2$), Shenxian ($n = 1$), Yinwang ($n = 1$), Yiyang ($n = 1$), Gaofeng ($n = 1$), Huaqiao ($n = 2$), and Huatuyan ($n = 8$). Numbers in parentheses indicate the number of samples from each population, and "*" represents published data. **C**: The haplogroup frequency plot of ancient individuals of various regions and periods. **D**: The PCA plot of haplogroup frequencies of populations. The percentage following PC represents the variance interpretation of this PC. Ancient populations are represented by solid symbols whose labels are combined with "a", regions, and sample ages ("k" represents thousand years before the present day), as well as present-day populations are represented by hollow symbols whose labels are combined with "pd" and regions. EA, East Asia; EA(S), southern East Asia; EA(N), northern East Asia; EA(W), western East Asia; SEA, Southeast Asia; SEA(noMyanmar), Southeast Asia without Myanmar; SEA(Myanmar), Myanmar; SA, South Asia; CA, Central Asia; NA, North Asia; BP, before present; PCA, principal component analysis; PC, principal component.

maternal genetic characteristics of individuals practicing two types of burial traditions. We found that despite the similarities between Hanging Coffin and Cave Burial (Chen, 1989b), ancient individuals practicing these burial traditions cannot be assigned to one population. However, we observed genetic connections (shared haplotypes C7a and F3a1 [4.88%]) between these two groups of individuals in China, whereas it remains to be further explored if this observation is related to cultural interaction. In addition, these two burial populations are genetically close to present-day Hmong-Mien populations. Haplogroups B, R9, N9a, and M7, which account for more than half of Hmong-Mien populations (Wen et al., 2004), are also identified in 48.48% of Cave Burial individuals (Table S1). A close relationship between Cave Burial population and present-day Miao (one Hmong-Mien population from China) is observed, and Hanging Coffin population also shows the affinity to one present-day Hmong-Mien population from Vietnam (Zhang et al., 2015).

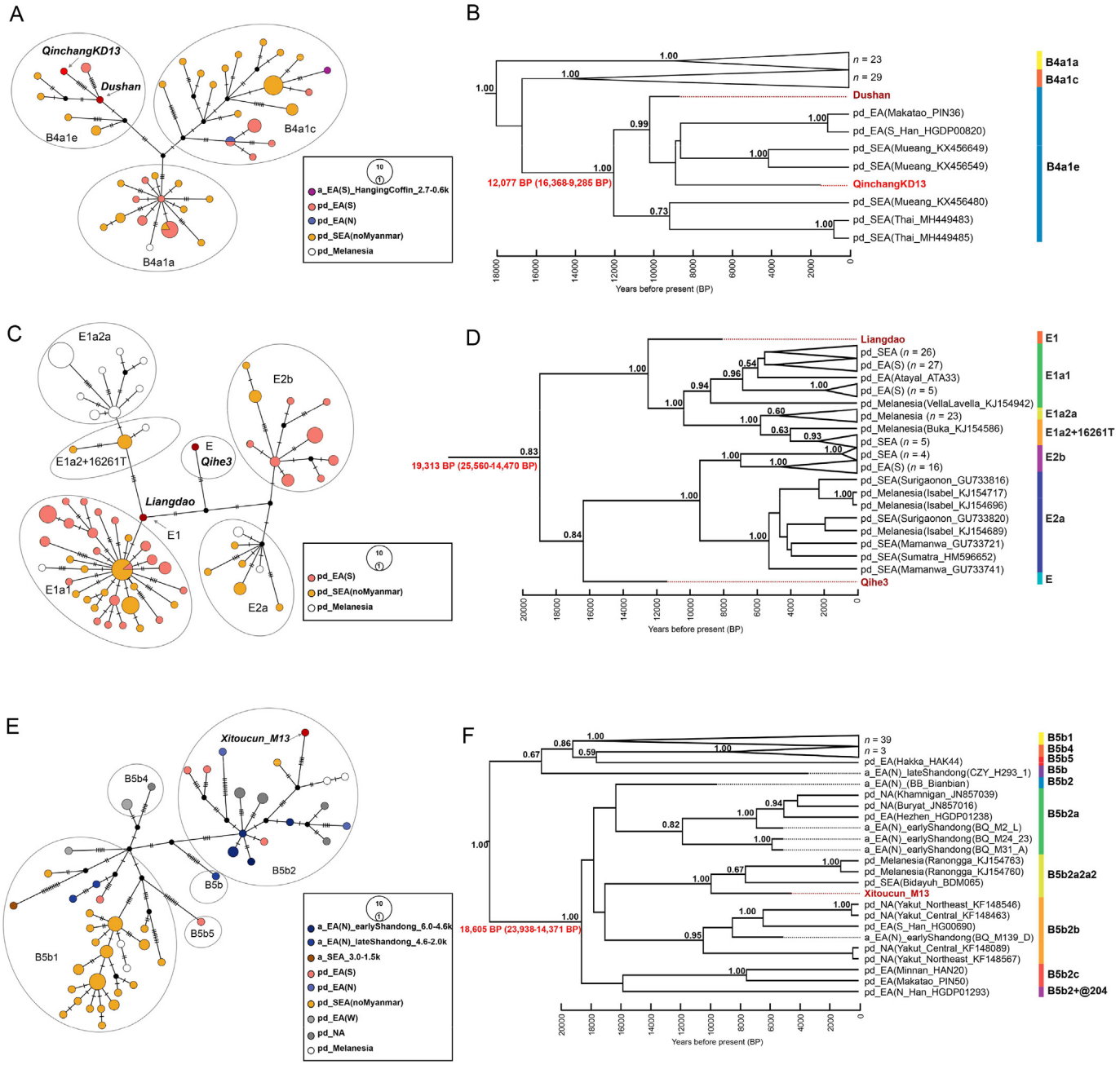


Fig. 2. Genetic analysis of populations in the haplogroups related to ancient individuals from Guangxi and Fujian (including the surrounding area). **A**: The network of lineage B4a1. **B**: The Bayesian phylogenetic tree of lineage B4a1. The subhaplogroup B4a1a includes pd_EA(S) ($n = 12$), pd_SEA ($n = 10$), and pd_Melanesia ($n = 1$). The subhaplogroup B4a1c includes a_EA(S)_HangingCoffin_2.7-0.6k ($n = 1$), pd_EA(S) ($n = 6$), pd_EA(N) ($n = 1$), and pd_SEA ($n = 21$). **C**: The network of lineage E. **D**: The Bayesian phylogenetic tree of lineage E. **E**: The network of lineage B5b2. **F**: The Bayesian phylogenetic tree of lineage B5b2. The subhaplogroup B5b1 includes a_EA(N)_lateShandong_4.6-2.0k ($n = 2$), a_SEA_3.0-1.5k ($n = 1$), pd_EA(S) ($n = 1$), pd_SEA ($n = 32$), pd_EA(W) ($n = 1$), and pd_Melanesia ($n = 2$). The subhaplogroup B5b4 includes pd_EA(W) ($n = 2$), and pd_NA ($n = 1$). The studied ancient individuals from Guangxi or Fujian including the surrounding area were highlighted as red. In networks, populations are colored by periods and regions. In Bayesian phylogenetic trees, all individuals are labeled with ancient ("a")/present-day ("pd"), nations, and the sample IDs. The posterior rates higher than 0.5 are shown on the nodes. The coalescence time with its 95% HPD of each haplogroup is under the node. HPD, highest probability density.

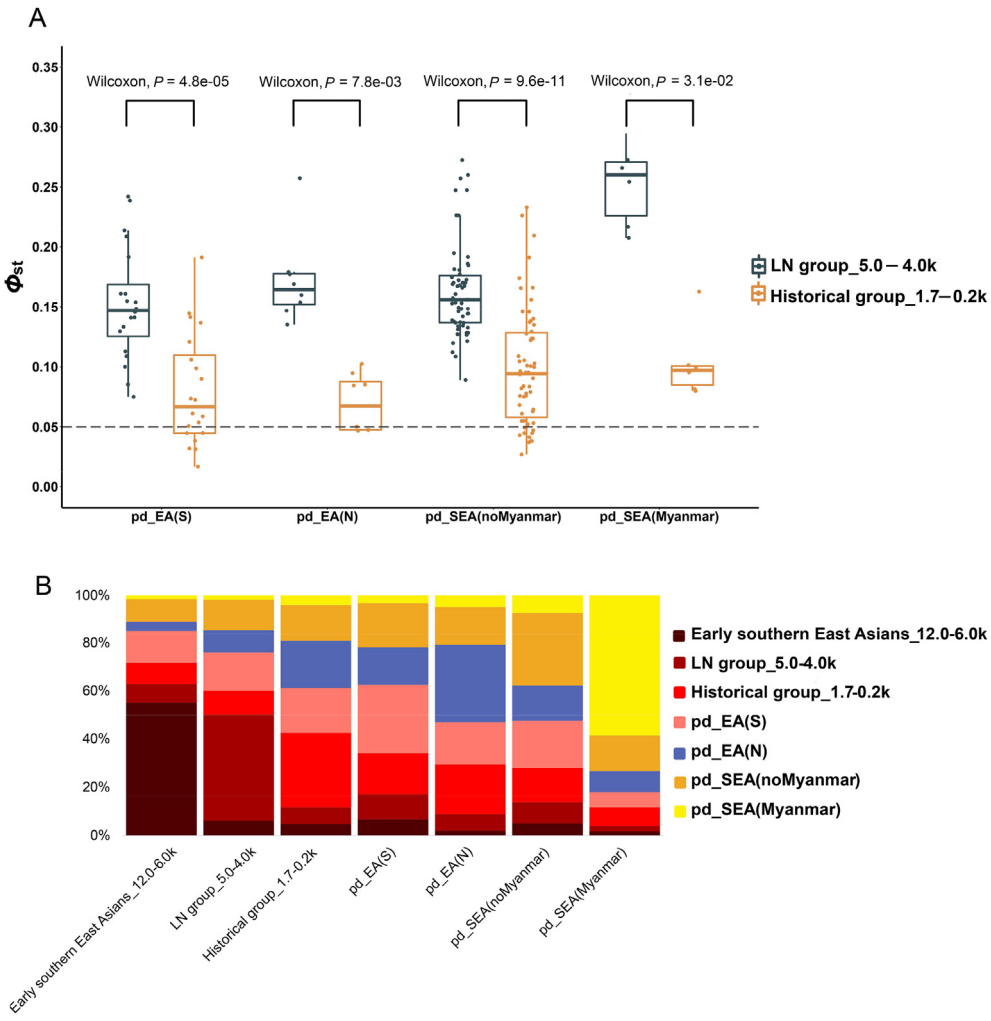


Fig. 3. Genetic relationship between ancient southern East Asians of different periods and present-day populations. **A**: The boxplot of genetic distances (Φ_{ST}) between the ancient southern East Asian populations and present-day populations. The statistical significance was calculated in Wilcoxon test. **B**: Haplogroup sharing between ancient southern East Asians and present-day individuals from various regions. Proportion of shared and unique haplogroups is denoted by the color of other groups and the group itself.

Obtaining aDNA information from ancient samples collected in southern East Asia is extremely difficult. This is mainly because the preservation of skeleton remains in humid and not climate is not ideal; thus, samples preserved well enough for aDNA studies are scarce. Despite great efforts been putting into sample collection and application of highly efficient DNA capture techniques, the temporal and geographical coverage of our samples can be further improved in the future. We are aware that the availability of samples can introduce potential biases to our study; thus, multiple measures were taken to minimize such bias, such as setting a cutoff that balancing the sample availability and statistical power for samples to be included in population studies. In addition, we acknowledge that our samples only cover a small subset of ancient southern East Asian populations, as well as the LN group from Fujian and the Historical group from Guangxi would lead to the regional deviations. Future studies including additional samples would be necessary for a comprehensive understanding of detailed population dynamics in this region. Nevertheless, the high consistency between the results generated from this study and previous archaeological and genetic studies suggests that our data offer a glimpse into the genuine maternal genetic structures of the ancient southern East Asian populations. Future studies with genome-wide data obtained from skeletons or sediments will offer a better understanding of the genetic history and connections across East Asia.

Materials and methods

Ancient sample collection

We collected 48 samples from 17 sites in Guangxi and three sites in Fujian (Table S1). The archaeological details are provided in the sites and specimen description (see supplementary data).

mtDNA extraction and library preparation

DNA was extracted from ~100 mg bone or tooth powder of 48 ancient samples (Table S1) following a previously described protocol (Dabney et al., 2013). The extracted DNA was used to construct 48 libraries with the single-stranded protocol and was denoted as "sslibrary" (Gansauge and Meyer, 2013). Eleven of 48 libraries were partially treated with uracil DNA glycosylase (UDG), which retains the characteristic damage of the aDNA in the first position at the 5′-end as well as the last position at the 3′-end (Rohland et al., 2015), and this process is also called "half.UDG" (Table S1). P5 and P7 adaptors were used to add sequence (5′-GTCT-3′) to the aDNA to avoid present-day DNA contamination (Kircher et al., 2012). Libraries were amplified for 35 cycles using AccuPrime Pfx DNA polymerase under conditions described in the previous study (Dabney and Meyer, 2012). After amplification, the concentration was determined using NanoDrop 2000.

Capture of mtDNA

The percentage of endogenous DNA is low, and we used DNA capture approaches to enrich genomes. Libraries and oligonucleotide probes overlapping the mtDNA were hybridized for enriching aDNA sequences and get better quality for sequencing. The complete mitochondrial genomes were captured with the protocol (Fu et al., 2013a). Briefly, we used the DNA capture method to recover the mitochondrial DNA.

Sequencing

High-throughput DNA sequencing for the enriched library pools was carried out on an Illumina MiSeq platform with 2 × 76 base pairs (bp) paired-end reads. The program leeHom was used to merge overlapping mate-pairs (>11 bp) with the parameter "–ancientdna" and to trim adapter sequences (Renaud et al., 2014). Fragments (≥30 bp) were mapped to the mtDNA revised Cambridge Reference Sequence (Andrews et al., 1999), and duplicates were removed. The sequences were mapped to the reference using the Burrows-Wheeler algorithm (BWA 0.7.17) (Li and Durbin, 2009) with the following command: bwa -n 0.01 and -l 16500. Reads with mapping quality below 30 were removed.

Test for contamination

The contamination rate was estimated after running the ContamMix (Fu et al., 2013a). Three hundred and eleven worldwide mtDNA sequences were used as references. If the percentage of fragments matching the consensus better than any of 311 worldwide mtDNA sequences is less than 95%, it means that the library may be contaminated (Fu et al., 2016). For those libraries with substantial contamination (rate > 4%) (Rohland et al., 2015), we restricted our analyses to only the fragments having characteristics typical of aDNA damage to retain as many individuals as possible for analysis (Briggs et al., 2007). Damaged fragments were retrieved by filtering out fragments with at least one C to T substitution in the first three positions at the 5′-end and the last three positions at the 3′-end using pmdtools 0.60 (Skoglund et al., 2014) with the –customterminus parameter. These libraries were referred to as damage-restricted libraries in Table S1.

Data analysis

Additional 174 ancient and 6859 present-day sequences were added into our database to be compared with our samples in this study. Haplotypes were assigned using HaploGrep2 (Weissensteiner et al., 2016) with PhyloTree mtDNA tree Build 17 (released Feb 2016) (van Oven and Kayser, 2009). The haplotype diversity of a group is the frequency at which two different haplotypes are randomly selected from that group. The Bayesian phylogenetic trees were constructed by BEAST v1.10.4 (Suchard et al., 2018) using a Bayesian skyline with a piecewise-linear/constant tree model for 100,000,000 interactions sampling every 10,000 steps, and the substitution models were calculated by jModelTest v2.1.10 (http://code.google.com/p/jmodeltest2). Two different models (a strict clock and an uncorrelated lognormal-distributed relaxed clock) of rate variation among branches were constructed (Fu et al., 2013b), and all phylogenetic trees were built in the strict clock. All effective sample size values exceeded 1×10^3 and were manually checked in Tracer v1.7.1 (Rambaut et al., 2018). The phylogenetic tree was visualized in FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/) with 10% of the trees discarded as burn-in using TreeAnnotator v1.10.4. The median-joining networks were constructed by PopArt v1.7 (Leigh and Bryant, 2015). Pairwise genetic distances (Φ_{ST}) with corresponding *P* and Analysis of Molecular Variance (AMOVA) were analyzed by Arlequin v3.5 (http://cmpg.unibe.ch/software/arlequin3). The principle component analysis (PCA) plot is drawn based on the haplogroup frequencies, and the multidimensional scaling (MDS) plot is drawn based on the Φ_{ST}. The haplogroup sharing calculated the proportions of haplogroups unique to a given group and those shared with other groups (Ding et al., 2020).

Data availability

All newly reported mtDNA genome sequences are available in the Genome Warehouse in National Genomics Data Center (China National Center for Bioinformation - National Genomics Data Center Members and Partners, 2020), Beijing Institute of Genomics (China National Center for Bioinformation), Chinese Academy of Science under accession number PRJCA004858. Of the newly sequenced data from this study, thirty nuclear genomes from the same individuals from Dushan, Baojianshan, Qihe and Guangxi historical sites have been published in (Wang et al., 2021).

Credit authorship contribution statement

Yalin Liu, **Tianyi Wang**: Formal analysis, Writing - Original draft, Writing - Review & Editing. **Xichao Wu**, **Xuechun Fan**, **Wei Wang**, **Guangmao Xie**, **Zhen Li**, **Qingping Yang**, **Yun Wu**: Archaeological materials and dating. **Peng Cao**, **Ruowei Yang**, **Feng Liu**, **Qingyan Dai**, **Xiaotian Feng**, **Wanjing Ping**: Wet laboratory work. **Bo Miao**: Writing - Original draft, Writing - Review & Editing. **Yichen Liu**: Conceptualization, Supervision, Writing - Original draft, Writing - Review & Editing. **Qiaomei Fu**: Conceptualization, Funding acquisition, Project administration, Supervision, Writing - Original draft, Writing - Review & Editing.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgments

We thank everyone in the laboratory for comments. We also thank our anonymous peer reviewers for constructive comments that have greatly helped us improve the quality of this study. This work was supported by the Chinese Academy of Sciences (CAS; XDB26000000, YSBR-019, XDA1905010, QYZDB-SSW-DQC003), National Natural Science Foundation of China (41925009, 41630102, 41672021), "Research on the roots of Chinese civilization" of Zhengzhou University (XKZDJC202006), Tencent Foundation through the XPLORER PRIZE, and the Howard Hughes Medical Institute (55008731).

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jgg.2021.06.002.

References

Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M., Howell, N., 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat. Genet. 23, 147-147.

Bai, F., Zhang, X., Ji, X., Cao, P., Feng, X., Yang, R., Peng, M., Pei, S., Fu, Q., 2020. Paleolithic genetic link between Southern China and Mainland Southeast Asia revealed by ancient mitochondrial genomes. J. Hum. Genet. 65, 1125–1128.

Briggs, A.W., Stenzel, U., Johnson, P.L.F., Green, R.E., Kelso, J., Prüfer, K., Meyer, M., Krause, J., Ronan, M.T., Lachmann, M., et al., 2007. Patterns of

damage in genomic DNA sequences from a Neandertal. Proc. Natl. Acad. Sci. U. S. A. 104, 14616–14621.
Chen, M., 1989a. A review of research on Hanging Coffin (in Chinese). Ethno-National Stud. 1, 98–103.
Chen, M., 1989b. A study of the comparison between hanging coffin and cave burial in southern China (in Chinese). J. Minzu Univ. China (Phil. Soc. Sci. Ed.) 5, 42–49.
China National Center for Bioinformation, National Genomics Data Center Members and Partners, 2020. Database resources of the national genomics data center, China national center for bioinformation in 2021. Nucleic Acids Res. 49, D18–D28.
Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., García, N., Pääbo, S., Arsuaga, J.-L., et al., 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proc. Natl. Acad. Sci. U. S. A. 110, 15758–15763.
Dabney, J., Meyer, M., 2012. Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. Biotechniques 52, 87–94.
Derenko, M., Malyarchuk, B., Grzybowski, T., Denisova, G., Rogalla, U., Perkova, M., Dambueva, I., Zakharov, I., 2010. Origin and post-glacial dispersal of mitochondrial DNA haplogroups C and D in northern Asia. PLoS One 5, e15214.
Ding, M., Wang, T., Ko, A.M.-S., Chen, H., Wang, H., Dong, G., Lu, H., He, W., Wangdue, S., Yuan, H., et al., 2020. Ancient mitogenomes show plateau populations from last 5200 years partially contributed to present-day Tibetans. Proc. Biol. Sci. 287, 20192968.
Duong, N.T., Macholdt, E., Ton, N.D., Arias, L., Schröder, R., Van Phong, N., Thi Bich Thuy, V., Ha, N.H., Thi Thu Hue, H., Thi Xuan, N., et al., 2018. Complete human mtDNA genome sequences from Vietnam and the phylogeography of Mainland Southeast Asia. Sci. Rep. 8, 11651.
Fu, Q., Meyer, M., Gao, X., Stenzel, U., Burbano, H.A., Kelso, J., Pääbo, S., 2013a. DNA analysis of an early modern human from Tianyuan Cave, China. Proc. Natl. Acad. Sci. U. S. A. 110, 2223–2227.
Fu, Q., Mittnik, A., Johnson, Philip L.F., Bos, K., Lari, M., Bollongino, R., Sun, C., Giemsch, L., Schmitz, R., Burger, J., et al., 2013b. A revised timescale for human evolution based on ancient mitochondrial genomes. Curr. Biol. 23, 553–559.
Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., Furtwängler, A., Haak, W., Meyer, M., Mittnik, A., et al., 2016. The genetic history of ice age Europe. Nature 534, 200–205.
Fu, X., 1988. On stepped adze and shoulder stone tools (in Chinese). Acta Archaeol. Sinica 1, 1–36.
Gansauge, M.-T., Meyer, M., 2013. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. Nat. Protoc. 8, 737–748.
Ji, X., Kuman, K., Clarke, R.J., Forestier, H., Li, Y., Ma, J., Qiu, K., Li, H., Wu, Y., 2016. The oldest Hoabinhian technocomplex in Asia (43.5 ka) at Xiaodong rockshelter, Yunnan Province, southwest China. Quat. Int. 400, 166–174.
Kircher, M., Sawyer, S., Meyer, M., 2012. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Res. 40, e3.
Ko, Albert M.-S., Chen, C.-Y., Fu, Q., Delfin, F., Li, M., Chiu, H.-L., Stoneking, M., Ko, Y.-C., 2014. Early Austronesians: into and out of Taiwan. Am. J. Hum. Genet. 94, 426–436.
Kong, Q.-P., Sun, C., Wang, H.-W., Zhao, M., Wang, W.-Z., Zhong, L., Hao, X.-D., Pan, H., Wang, S.-Y., Cheng, Y.-T., et al., 2011. Large-scale mtDNA screening reveals a surprising matrilineal complexity in East Asia and its implications to the peopling of the region. Mol. Biol. Evol. 28, 513–522.
Kutanan, W., Kampuansai, J., Srikummool, M., Kangwanpong, D., Ghirotto, S., Brunelli, A., Stoneking, M., 2017. 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. 136, 85–98.
Kutanan, W., Shoocongdej, R., Srikummool, M., Hübner, A., Suttipai, T., Srithawong, S., Kampuansai, J., Stoneking, M., 2020. Cultural variation impacts paternal and maternal genetic lineages of the Hmong-Mien and Sino-Tibetan groups from Thailand. Eur. J. Hum. Genet. 28, 1563–1579.
Lahr, M.M., Foley, R., 1994. Multiple dispersals and modern human origins. Evol. Anthropol. Issues News Rev. 3, 48–60.
Larruga, J.M., Marrero, P., Abu-Amero, K.K., Golubenko, M.V., Cabrera, V.M., 2017. Carriers of mitochondrial DNA macrohaplogroup R colonized Eurasia and Australasia from a southeast Asia core area. BMC Evol. Biol. 17, 115-115.
Leigh, J.W., Bryant, D., 2015. popart: full-feature software for haplotype network construction. Meth. Ecol. Evol. 6, 1110–1116.
Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760.
Lin, Z., Wu, Y., Lin, Z., Mei, H., 1980. A brief report on the clearance of Baiyan hanging coffin in Wuyi mountain, Chong 'an, Fujian (in Chinese). Cult. Relics 14–22.
Liu, J., Zeng, W., Sun, B., Mao, X., Zhao, Y., Wang, F., Li, Z., Luan, F., Guo, J., Zhu, C., et al., 2021. Maternal genetic structure in ancient Shandong between 9500 and 1800 years ago. Sci. Bull. 66, 1129–1135.
Luo, E., 2000. Cave burial in southern China (in Chinese). Archaeology 6, 79–87.
Matsumura, H., Hung, H.-c., Higham, C., Zhang, C., Yamagata, M., Nguyen, L.C., Li, Z., Fan, X.-c., Simanjuntak, T., Oktaviana, A.A., et al., 2019. Craniometrics reveal "two layers" of prehistoric human dispersal in eastern Eurasia. Sci. Rep. 9, 1451.
McColl, H., Racimo, F., Vinner, L., Demeter, F., Gakuhari, T., Moreno-Mayar, J.V., van Driem, G., Gram Wilken, U., Seguin-Orlando, A., de la Fuente Castro, C., et al., 2018. The prehistoric peopling of Southeast Asia. Science 361, 88.
Mei, H., 1989. On the pottery pots unearthed in Fujian, guangdong and Guangxi (in Chinese). Archaeology 11, 1022–1026.
O'Connell, J.F., Allen, J., Williams, M.A.J., Williams, A.N., Turney, C.S.M., Spooner, N.A., Kamminga, J., Brown, G., Cooper, A., 2018. When did *Homo sapiens* first reach Southeast Asia and Sahul? Proc. Natl. Acad. Sci. U. S. A. 115, 8482–8490.
Peng, C., 2001. Preliminary study on early cave burials in Guangxi (in Chinese). Guangxi Ethnic Stud. 84–90.
Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior summarization in Bayesian phylogenetics using tracer 1.7. Syst. Biol. 67, 901–904.
Renaud, G., Stenzel, U., Kelso, J., 2014. leeHom: adaptor trimming and merging for Illumina sequencing reads. Nucleic Acids Res. 42, e141.
Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., Reich, D., 2015. Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. Phil. Trans. Biol. Sci. 370, 20130624.
Skoglund, P., Northoff, B.H., Shunkov, M.V., Derevianko, A.P., Pääbo, S., Krause, J., Jakobsson, M., 2014. Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. Proc. Natl. Acad. Sci. U. S. A. 111, 2229.
Soares, P., Trejaut, J.A., Loo, J.-H., Hill, C., Mormina, M., Lee, C.-L., Chen, Y.-M., Hudjashov, G., Forster, P., Macaulay, V., et al., 2008. Climate change and postglacial human dispersals in Southeast Asia. Mol. Biol. Evol. 25, 1209–1218.
Stoneking, M., Delfin, F., 2010. The human genetic history of East Asia: Weaving a complex tapestry. Curr. Biol. 20, R188–R193.
Suchard, M.A., Lemey, P., Baele, G., Ayres, D.L., Drummond, A.J., Rambaut, A., 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol. 4, vey016.
Tanaka, M., Cabrera, V.M., González, A.M., Larruga, J.M., Takeyasu, T., Fuku, N., Guo, L.J., Hirose, R., Fujita, Y., Kurata, M., et al., 2004. Mitochondrial genome variation in eastern Asia and the peopling of Japan. Genome Res. 14, 1832–1850.
Tofanelli, S., Bertoncini, S., Castrì, L., Luiselli, D., Calafell, F., Donati, G., Paoli, G., 2009. On the origins and admixture of Malagasy: new evidence from high-resolution analyses of paternal and maternal lineages. Mol. Biol. Evol. 26, 2109–2124.
van Oven, M., Kayser, M., 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum. Mutat. 30, E386–E394.
Wall, J.D., Stawiski, E.W., Ratan, A., Kim, H.L., Kim, C., Gupta, R., Suryamohan, K., Gusareva, E.S., Purbojati, R.W., Bhangale, T., et al., 2019. The GenomeAsia 100K Project enables genetic discoveries across Asia. Nature 576, 106–111.
Wang, T., Wang, W., Xie, G., Li, Z., Fan, X., Yang, Q., Wu, X., Cao, P., Liu, Y., Yang, R., et al., 2021. Human population history at the crossroads of East and Southeast Asia since 11,000 years ago. Cell 184, 3829–3841. e3821.
Wang, C.-C., Yeh, H.-Y., Popov, A.N., Zhang, H.-Q., Matsumura, H., Sirak, K., Cheronet, O., Kovalev, A., Rohland, N., Kim, A.M., et al., 2021. Genomic insights into the formation of human populations in East Asia. Nature 591, 413–419.
Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.J., Kronenberg, F., Salas, A., Schönherr, S., 2016. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. Nucleic Acids Res. 44, W58–W63.
Wen, B., Li, H., Gao, S., Mao, X., Gao, Y., Li, F., Zhang, F., He, Y., Dong, Y., Zhang, Y., et al., 2004. Genetic structure of Hmong-mien speaking populations in East Asia as revealed by mtDNA lineages. Mol. Biol. Evol. 22, 725–734.
Yang, M.A., Fan, X., Sun, B., Chen, C., Lang, J., Ko, Y.-C., Tsang, C.-h., Chiu, H., Wang, T., Bao, Q., et al., 2020. Ancient DNA indicates human population shifts and admixture in northern and southern China. Science 369, 282–288.
Yao, Y.G., Kong, Q.P., Bandelt, H.J., Kivisild, T., Zhang, Y.P., 2002. Phylogeographic differentiation of mitochondrial DNA in Han Chinese. Am. J. Hum. Genet. 70, 635–651.
Zhang, M., Fu, Q., 2020. Human evolutionary history in Eastern Eurasia using insights from ancient DNA. Curr. Opin. Genet. Dev. 62, 78–84.
Zhang, X., Liao, S., Qi, X., Liu, J., Kampuansai, J., Zhang, H., Yang, Z., Serey, B., Sovannary, T., Bunnath, L., et al., 2015. Y-chromosome diversity suggests southern origin and Paleolithic backwave migration of Austro-Asiatic speakers from eastern Asia to the Indian subcontinent. Sci. Rep. 5, 15486.